WEBVTT

NOTE duration:”01:23:27”

NOTE recognizability:0.862

NOTE language:en-us

NOTE Confidence: 0.963834928333333

00:00:00.000 --> 00:00:03.120 It is my pleasure to introduce

NOTE Confidence: 0.963834928333333

00:00:03.120 --> 00:00:06.320 the Hanan Jay Bhaskar Jay.

NOTE Confidence: 0.963834928333333

00:00:06.320 --> 00:00:08.378 He is a past doctoral researcher

NOTE Confidence: 0.963834928333333

00:00:08.378 --> 00:00:10.213 in the Department of Genetics

NOTE Confidence: 0.963834928333333

00:00:10.213 --> 00:00:12.158 at Yale School of Medicine.

NOTE Confidence: 0.963834928333333

00:00:12.160 --> 00:00:15.240 He has background strong quantitative

NOTE Confidence: 0.963834928333333

00:00:15.240 --> 00:00:17.704 background both in mathematical

NOTE Confidence: 0.963834928333333

00:00:17.704 --> 00:00:20.099 modeling and machine learning,

NOTE Confidence: 0.963834928333333

00:00:20.099 --> 00:00:21.878 anthropological data analysis

NOTE Confidence: 0.963834928333333

00:00:21.878 --> 00:00:24.250 with applications in biophysics

NOTE Confidence: 0.963834928333333

00:00:24.324 --> 00:00:26.040 and biomedical research.

NOTE Confidence: 0.963834928333333

00:00:26.040 --> 00:00:28.464 He received his PhD in Biomedical

NOTE Confidence: 0.963834928333333

00:00:28.464 --> 00:00:31.094 Engineering and Data and Master's degree

NOTE Confidence: 0.963834928333333

00:00:31.094 --> 00:00:33.914 in Data Science from Brown University,

NOTE Confidence: 0.963834928333333

00:00:33.920 --> 00:00:36.308 and for that he studied computer

NOTE Confidence: 0.963834928333333

00:00:36.308 --> 00:00:37.900 science and applied mathematics

NOTE Confidence: 0.963834928333333

00:00:37.964 --> 00:00:40.079 at University of British Columbia.

NOTE Confidence: 0.963834928333333

00:00:40.080 --> 00:00:44.128 This is going to be 4 parts series and

NOTE Confidence: 0.963834928333333

00:00:44.128 --> 00:00:47.440 we have a group of presenters who I will

NOTE Confidence: 0.963834928333333

00:00:47.527 --> 00:00:51.680 introduce for each section separately.

NOTE Confidence: 0.963834928333333

00:00:51.680 --> 00:00:53.759 And Jay, I'm going to give it to you.

NOTE Confidence: 0.963834928333333

00:00:53.760 --> 00:00:56.384 So if you want to add anything about

NOTE Confidence: 0.963834928333333

00:00:56.384 --> 00:00:58.291 overall research, you're welcome to.

NOTE Confidence: 0.963834928333333

00:00:58.291 --> 00:00:59.959 And if you guys,

NOTE Confidence: 0.963834928333333

00:00:59.960 --> 00:01:01.217 anybody has questions,

NOTE Confidence: 0.963834928333333

00:01:01.217 --> 00:01:04.150 you're welcome to type them in the

NOTE Confidence: 0.963834928333333

00:01:04.226 --> 00:01:06.360 chat or in question and answers.

NOTE Confidence: 0.963834928333333

00:01:06.360 --> 00:01:08.012 And Jay said that he will respond

NOTE Confidence: 0.963834928333333

00:01:08.012 --> 00:01:09.320 to them as they come.

NOTE Confidence: 0.752189074444444

00:01:10.160 --> 00:01:12.008 And thank you, Helen,

NOTE Confidence: 0.752189074444444

00:01:12.008 --> 00:01:14.318 for the very kind introduction.

NOTE Confidence: 0.752189074444444

00:01:14.320 --> 00:01:15.826 And welcome, everyone,

NOTE Confidence: 0.752189074444444

00:01:15.826 --> 00:01:19.220 to the first workshop in this series.

NOTE Confidence: 0.752189074444444

00:01:19.220 --> 00:01:22.440 Today, my goal is to introduce you

NOTE Confidence: 0.752189074444444

00:01:22.539 --> 00:01:25.809 to a methodology called topological

NOTE Confidence: 0.752189074444444

00:01:25.809 --> 00:01:29.079 data analysis and machine learning.

NOTE Confidence: 0.752189074444444

00:01:29.080 --> 00:01:31.840 Both of these techniques are,

NOTE Confidence: 0.752189074444444

00:01:31.840 --> 00:01:33.404 you know, very broad.

NOTE Confidence: 0.752189074444444

00:01:33.404 --> 00:01:35.750 They encompass a number of different

NOTE Confidence: 0.752189074444444

00:01:35.821 --> 00:01:38.301 methods and so there's no way I can

NOTE Confidence: 0.752189074444444

00:01:38.301 --> 00:01:40.649 cover both of them in any amount of

NOTE Confidence: 0.752189074444444

00:01:40.649 --> 00:01:42.562 detail in just a single session.

NOTE Confidence: 0.752189074444444

00:01:42.562 --> 00:01:45.770 My goal is to give you a broad

NOTE Confidence: 0.752189074444444

00:01:45.862 --> 00:01:48.417 overview of these techniques and

NOTE Confidence: 0.752189074444444

00:01:48.417 --> 00:01:51.728 to build some intuition and for

NOTE Confidence: 0.752189074444444

00:01:51.728 --> 00:01:55.160 us to share a common vocabulary.

NOTE Confidence: 0.752189074444444

00:01:55.160 --> 00:01:57.278 And then in the subsequent workshops,

NOTE Confidence: 0.752189074444444

00:01:57.280 --> 00:01:59.158 we are going to take in,

NOTE Confidence: 0.752189074444444

00:01:59.160 --> 00:02:01.800 take bits and pieces of topological

NOTE Confidence: 0.752189074444444

00:02:01.800 --> 00:02:04.291 data analysis, TDA in short,

NOTE Confidence: 0.752189074444444

00:02:04.291 --> 00:02:07.513 and machine learning that we need

NOTE Confidence: 0.752189074444444

00:02:07.520 --> 00:02:10.640 to analyze some neuroimaging data.

NOTE Confidence: 0.752189074444444

00:02:10.640 --> 00:02:13.799 And as Helen mentioned, my name is Tanan Jay.

NOTE Confidence: 0.752189074444444

00:02:13.800 --> 00:02:16.040 I generally go by Jay.

NOTE Confidence: 0.752189074444444

00:02:16.040 --> 00:02:18.280 I wear many different hats,

NOTE Confidence: 0.752189074444444

00:02:18.280 --> 00:02:21.796 but most relevant is I'm a

NOTE Confidence: 0.752189074444444

00:02:21.800 --> 00:02:24.312 postdoctoral fellow in neuroscience.

NOTE Confidence: 0.752189074444444

00:02:24.312 --> 00:02:28.028 I also work, I wanted to disclose,

NOTE Confidence: 0.752189074444444

00:02:28.028 --> 00:02:29.399 with Bohinger Ingelheim,

NOTE Confidence: 0.752189074444444

00:02:29.400 --> 00:02:31.998 which is a German pharmaceutical company,

NOTE Confidence: 0.752189074444444

00:02:32.000 --> 00:02:34.025 and I still maintain some

NOTE Confidence: 0.752189074444444

00:02:34.025 --> 00:02:35.240 affiliations with Brown,

NOTE Confidence: 0.752189074444444

00:02:35.240 --> 00:02:37.720 having completed my PhD there.

NOTE Confidence: 0.752189074444444

00:02:37.720 --> 00:02:41.280 So first, I wanted to just talk a little bit,

NOTE Confidence: 0.752189074444444

00:02:41.280 --> 00:02:42.576 give you a little bit of

NOTE Confidence: 0.752189074444444

00:02:42.576 --> 00:02:43.440 an introduction to myself.

NOTE Confidence: 0.752189074444444

00:02:43.440 --> 00:02:45.316 So as Helen mentioned,

NOTE Confidence: 0.752189074444444

00:02:45.316 --> 00:02:48.130 I come from a very quantitative

NOTE Confidence: 0.752189074444444

00:02:48.226 --> 00:02:49.879 dry lab background.

NOTE Confidence: 0.752189074444444

00:02:49.880 --> 00:02:52.448 I received my undergraduate degree in

NOTE Confidence: 0.752189074444444

00:02:52.448 --> 00:02:54.840 computer science and math and did a

NOTE Confidence: 0.752189074444444

00:02:54.840 --> 00:02:57.320 master's degree in applied mathematics.

NOTE Confidence: 0.752189074444444

00:02:57.320 --> 00:02:59.516 And in those years, a long,

NOTE Confidence: 0.752189074444444

00:02:59.520 --> 00:03:01.041 long time ago,

NOTE Confidence: 0.752189074444444

00:03:01.041 --> 00:03:03.576 I was interested in modelling

NOTE Confidence: 0.752189074444444

00:03:03.576 --> 00:03:05.879 biophysics and in particular,

NOTE Confidence: 0.752189074444444

00:03:05.880 --> 00:03:08.105 I was interested in developmental

NOTE Confidence: 0.752189074444444

00:03:08.105 --> 00:03:09.440 and cancer biology.

NOTE Confidence: 0.752189074444444

00:03:09.440 --> 00:03:12.000 So I spent a lot of my formative

NOTE Confidence: 0.752189074444444

00:03:12.000 --> 00:03:15.200 years putting together agent based

NOTE Confidence: 0.752189074444444

00:03:15.200 --> 00:03:18.108 models to simulate cell migration

NOTE Confidence: 0.752189074444444

00:03:18.108 --> 00:03:21.390 and cell morphology and emergence of

NOTE Confidence: 0.752189074444444

00:03:21.483 --> 00:03:24.378 different types of migratory patterns

NOTE Confidence: 0.752189074444444

00:03:24.378 --> 00:03:27.740 in normal healthy tissue and also

NOTE Confidence: 0.752189074444444

00:03:27.740 --> 00:03:32.026 various kinds of tumors and you know,

NOTE Confidence: 0.752189074444444

00:03:32.026 --> 00:03:35.038 across development and embryogenesis.

NOTE Confidence: 0.752189074444444

00:03:35.040 --> 00:03:35.700 Subsequently,

NOTE Confidence: 0.752189074444444

00:03:35.700 --> 00:03:40.320 I I moved over to Brown University

NOTE Confidence: 0.752189074444444

00:03:40.320 --> 00:03:43.504 where I was in the data science and

NOTE Confidence: 0.752189074444444

00:03:43.504 --> 00:03:45.200 biomedical engineering departments.

NOTE Confidence: 0.752189074444444

00:03:45.200 --> 00:03:47.960 And it was at Brown University

NOTE Confidence: 0.752189074444444

00:03:47.960 --> 00:03:50.944 where I became fascinated with some

NOTE Confidence: 0.752189074444444

00:03:50.944 --> 00:03:54.040 mathematical concepts to do with shape.

NOTE Confidence: 0.752189074444444

00:03:54.040 --> 00:03:58.048 And I realized during these years

NOTE Confidence: 0.752189074444444

00:03:58.048 --> 00:04:01.246 that learning the shape of the data

NOTE Confidence: 0.752189074444444

00:04:01.246 --> 00:04:04.230 and being able to quantify the shape

NOTE Confidence: 0.752189074444444

00:04:04.230 --> 00:04:07.775 of the data can be a really powerful

NOTE Confidence: 0.752189074444444

00:04:07.775 --> 00:04:10.920 tool for biomedical data analysis.

NOTE Confidence: 0.752189074444444

00:04:10.920 --> 00:04:12.220 And so for instance,

NOTE Confidence: 0.752189074444444

00:04:12.220 --> 00:04:14.684 if you have a bunch of data

NOTE Confidence: 0.752189074444444

00:04:14.684 --> 00:04:17.270 points that are arranged in this

NOTE Confidence: 0.752189074444444

00:04:17.270 --> 00:04:19.479 weird double Taurus like shape,

NOTE Confidence: 0.752189074444444

00:04:19.480 --> 00:04:22.210 being able to actually take these

NOTE Confidence: 0.752189074444444

00:04:22.210 --> 00:04:24.449 individual data points and be

NOTE Confidence: 0.752189074444444

00:04:24.449 --> 00:04:26.836 able to fill in the empty spaces.

NOTE Confidence: 0.752189074444444

00:04:26.840 --> 00:04:29.010 And to be able to recognize that

NOTE Confidence: 0.752189074444444

00:04:29.010 --> 00:04:31.462 there are two big holes in the data

NOTE Confidence: 0.752189074444444

00:04:31.462 --> 00:04:34.112 and that our data has a loop like

NOTE Confidence: 0.752189074444444

00:04:34.112 --> 00:04:36.374 structure where there's a bigger loop

NOTE Confidence: 0.752189074444444

00:04:36.374 --> 00:04:38.716 around which our data points are organized.

NOTE Confidence: 0.752189074444444

00:04:38.720 --> 00:04:40.970 And then there are smaller loops

NOTE Confidence: 0.752189074444444

00:04:40.970 --> 00:04:42.920 that surround those bigger loops.

NOTE Confidence: 0.752189074444444

00:04:42.920 --> 00:04:45.573 Being able to recognize these types of

NOTE Confidence: 0.752189074444444

00:04:45.573 --> 00:04:48.119 patterns can be extremely powerful,

NOTE Confidence: 0.752189074444444

00:04:48.120 --> 00:04:50.195 especially when we are dealing

NOTE Confidence: 0.752189074444444

00:04:50.195 --> 00:04:51.440 with biological data.

NOTE Confidence: 0.752189074444444

00:04:51.440 --> 00:04:56.036 And then recently moved to postdoc

NOTE Confidence: 0.752189074444444

00:04:56.040 --> 00:04:58.679 in in genetics and also I guess

NOTE Confidence: 0.752189074444444

00:04:58.679 --> 00:04:59.810 in computer science

NOTE Confidence: 0.842413505

00:04:59.891 --> 00:05:01.519 at at Yale University.

NOTE Confidence: 0.842413505

00:05:01.520 --> 00:05:04.536 And I spent my postdoc thinking a lot

NOTE Confidence: 0.842413505

00:05:04.536 --> 00:05:07.797 about data that is structured like a graph.

NOTE Confidence: 0.842413505

00:05:07.800 --> 00:05:10.504 And So what I mean by a graph

NOTE Confidence: 0.842413505

00:05:10.504 --> 00:05:13.560 here is a set of nodes and edges.

NOTE Confidence: 0.842413505

00:05:13.560 --> 00:05:16.059 So the nodes being represented as the

NOTE Confidence: 0.842413505

00:05:16.059 --> 00:05:18.719 circles and edges being the lines that

NOTE Confidence: 0.842413505

00:05:18.719 --> 00:05:20.746 are connecting different nodes together.

NOTE Confidence: 0.842413505

00:05:20.746 --> 00:05:24.179 There's lots and lots of data out there

NOTE Confidence: 0.842413505

00:05:24.179 --> 00:05:26.993 that can be represented in this format.

NOTE Confidence: 0.842413505

00:05:27.000 --> 00:05:29.256 For example, if you are looking

NOTE Confidence: 0.842413505

00:05:29.256 --> 00:05:30.760 to do drug discovery,

NOTE Confidence: 0.842413505

00:05:30.760 --> 00:05:33.705 you can represent molecules using

NOTE Confidence: 0.842413505

00:05:33.705 --> 00:05:36.650 nodes and and edges corresponding

NOTE Confidence: 0.842413505

00:05:36.744 --> 00:05:39.559 to atoms and bonds respectively.

NOTE Confidence: 0.842413505

00:05:39.560 --> 00:05:41.360 You can take protein sequences,

NOTE Confidence: 0.842413505

00:05:41.360 --> 00:05:44.048 fold them with alpha fold and then

NOTE Confidence: 0.842413505

00:05:44.048 --> 00:05:46.199 represent protein structure in this manner.

NOTE Confidence: 0.842413505

00:05:46.200 --> 00:05:49.815 But also you can take neuroscience

NOTE Confidence: 0.842413505

00:05:49.815 --> 00:05:53.560 data such as brain imaging data and

NOTE Confidence: 0.842413505

00:05:53.560 --> 00:05:56.429 divide brain into different parcels

NOTE Confidence: 0.842413505

00:05:56.429 --> 00:05:58.840 and learn to represent brain imaging

NOTE Confidence: 0.842413505

00:05:58.840 --> 00:06:01.525 data in this format where the nodes

NOTE Confidence: 0.842413505

00:06:01.525 --> 00:06:03.460 are going to represent different

NOTE Confidence: 0.842413505

00:06:03.460 --> 00:06:05.518 parcels or regions of the brain.

NOTE Confidence: 0.842413505

00:06:05.520 --> 00:06:08.544 And the edges could be anatomical

NOTE Confidence: 0.842413505

00:06:08.544 --> 00:06:10.056 connectivity between those

NOTE Confidence: 0.842413505

00:06:10.056 --> 00:06:11.599 parcels in the brain,

NOTE Confidence: 0.842413505

00:06:11.600 --> 00:06:14.160 or it could be due to functional connectivity

NOTE Confidence: 0.842413505

00:06:14.160 --> 00:06:15.999 between those parcels in the brain.

NOTE Confidence: 0.842413505

00:06:16.000 --> 00:06:17.800 Maybe at an even higher level,

NOTE Confidence: 0.842413505

00:06:17.800 --> 00:06:20.565 one can think of like taking biomedical

NOTE Confidence: 0.842413505

00:06:20.565 --> 00:06:23.292 data in general and representing it in

NOTE Confidence: 0.842413505

00:06:23.292 --> 00:06:26.000 in the format of a knowledge graph,

NOTE Confidence: 0.842413505

00:06:26.000 --> 00:06:28.436 where you can bring in data from,

NOTE Confidence: 0.842413505

00:06:28.440 --> 00:06:29.890 you know,

NOTE Confidence: 0.842413505

00:06:29.890 --> 00:06:33.515 publications from single cell sequencing

NOTE Confidence: 0.842413505

00:06:33.515 --> 00:06:35.356 experiments and other modalities

NOTE Confidence: 0.842413505

00:06:35.356 --> 00:06:37.960 and represent all of that data in

NOTE Confidence: 0.842413505

00:06:38.029 --> 00:06:40.439 a large biomedical knowledge graph.

NOTE Confidence: 0.842413505

00:06:40.440 --> 00:06:41.892 So during my postdoc,

NOTE Confidence: 0.842413505

00:06:41.892 --> 00:06:43.707 I've developed techniques to not

NOTE Confidence: 0.842413505

00:06:43.707 --> 00:06:45.679 only represent data as graphs,

NOTE Confidence: 0.842413505

00:06:45.680 --> 00:06:49.360 but also to develop machine learning

NOTE Confidence: 0.842413505

00:06:49.360 --> 00:06:51.760 techniques to learn to reason about

NOTE Confidence: 0.842413505

00:06:51.760 --> 00:06:53.676 these kinds of graphs and represent

NOTE Confidence: 0.842413505

00:06:53.676 --> 00:06:56.040 them in a way that a computer can

NOTE Confidence: 0.842413505

00:06:56.040 --> 00:06:57.770 understand the structure of the

NOTE Confidence: 0.842413505

00:06:57.770 --> 00:06:59.886 graph and take advantage of that

NOTE Confidence: 0.842413505

00:06:59.886 --> 00:07:01.800 to answer all kinds of questions.

NOTE Confidence: 0.842413505

00:07:01.800 --> 00:07:05.137 And so today I'm going to talk

NOTE Confidence: 0.842413505

00:07:05.137 --> 00:07:08.653 to you about a technique that is

NOTE Confidence: 0.842413505

00:07:08.653 --> 00:07:11.318 utilizing this graph structure and

NOTE Confidence: 0.842413505

00:07:11.318 --> 00:07:14.559 combining it with aspects of topology,

NOTE Confidence: 0.842413505

00:07:14.560 --> 00:07:16.936 which is essentially the technique that

NOTE Confidence: 0.842413505

00:07:16.936 --> 00:07:19.997 allows us to recognize the shape of our data.

NOTE Confidence: 0.842413505

00:07:20.000 --> 00:07:22.560 And so to motivate this,

NOTE Confidence: 0.842413505

00:07:22.560 --> 00:07:24.228 this technique that I'm going to

NOTE Confidence: 0.842413505

00:07:24.228 --> 00:07:25.986 talk about and we're going to

NOTE Confidence: 0.842413505

00:07:25.986 --> 00:07:27.756 develop over the next few workshops,

NOTE Confidence: 0.842413505

00:07:27.760 --> 00:07:30.328 I wanted to share with you some

NOTE Confidence: 0.842413505

00:07:30.328 --> 00:07:32.768 time lapse microscopy images that

NOTE Confidence: 0.842413505

00:07:32.768 --> 00:07:35.599 were taken a long time ago.

NOTE Confidence: 0.842413505

00:07:35.600 --> 00:07:38.762 And these are calcium imaging data

NOTE Confidence: 0.842413505

00:07:38.762 --> 00:07:42.600 sets of a developing zebrafish embryo.

NOTE Confidence: 0.842413505

00:07:42.600 --> 00:07:46.278 And if I could play these,

NOTE Confidence: 0.842413505

00:07:46.280 --> 00:07:49.160 not sure if I can hang on a second,

NOTE Confidence: 0.928219111666666

00:07:52.320 --> 00:07:55.116 OK, if I play these images,

NOTE Confidence: 0.928219111666666

00:07:55.120 --> 00:07:57.559 What you'll notice here is that on the left,

NOTE Confidence: 0.928219111666666

00:07:57.560 --> 00:07:59.620 you have a zebrafish embryo

NOTE Confidence: 0.928219111666666

00:07:59.620 --> 00:08:01.680 that's early in its development.

NOTE Confidence: 0.928219111666666

00:08:01.680 --> 00:08:04.560 In the middle, it's grown a little bit more.

NOTE Confidence: 0.928219111666666

00:08:04.560 --> 00:08:05.856 And on the right,

NOTE Confidence: 0.928219111666666

00:08:05.856 --> 00:08:07.476 the zebrafish embryo is much

NOTE Confidence: 0.928219111666666

00:08:07.476 --> 00:08:09.360 further along in its development.

NOTE Confidence: 0.928219111666666

00:08:09.360 --> 00:08:11.064 What you're going to notice is

NOTE Confidence: 0.928219111666666

00:08:11.064 --> 00:08:12.200 that the signalling patterns,

NOTE Confidence: 0.928219111666666

00:08:12.200 --> 00:08:14.910 the calcium signalling patterns across

NOTE Confidence: 0.928219111666666

00:08:14.910 --> 00:08:17.078 development look very different.

NOTE Confidence: 0.928219111666666

00:08:17.080 --> 00:08:19.480 In the video on the left,

NOTE Confidence: 0.928219111666666

00:08:19.480 --> 00:08:21.676 early in development we see that

NOTE Confidence: 0.928219111666666

00:08:21.680 --> 00:08:24.240 we see individual spiking events,

NOTE Confidence: 0.928219111666666

00:08:24.240 --> 00:08:26.016 so individual calcium signaling

NOTE Confidence: 0.928219111666666

00:08:26.016 --> 00:08:28.236 events that are not really

NOTE Confidence: 0.928219111666666

00:08:28.240 --> 00:08:30.896 correlated temporarily or specially.

NOTE Confidence: 0.928219111666666

00:08:30.896 --> 00:08:33.516 A little bit further along in development,

NOTE Confidence: 0.928219111666666

00:08:33.520 --> 00:08:36.358 we start to see patches of

NOTE Confidence: 0.928219111666666

00:08:36.358 --> 00:08:38.960 synchronous activity in the embryo.

NOTE Confidence: 0.928219111666666

00:08:38.960 --> 00:08:40.436 So you see these small patches,

NOTE Confidence: 0.928219111666666

00:08:40.440 --> 00:08:43.394 but they don't really travel very far.

NOTE Confidence: 0.928219111666666

00:08:43.400 --> 00:08:45.362 And even later in development you

NOTE Confidence: 0.928219111666666

00:08:45.362 --> 00:08:47.577 start to see these waves traveling

NOTE Confidence: 0.928219111666666

00:08:47.577 --> 00:08:50.103 wave like patterns where you have

NOTE Confidence: 0.928219111666666

00:08:50.103 --> 00:08:51.956 calcium signaling starting at a

NOTE Confidence: 0.928219111666666

00:08:51.956 --> 00:08:53.840 small group of cells and that

NOTE Confidence: 0.928219111666666

00:08:53.840 --> 00:08:56.086 really kind of expands and goes all

NOTE Confidence: 0.928219111666666

00:08:56.086 --> 00:08:58.359 across the embryo of the zebrafish.

NOTE Confidence: 0.928219111666666

00:08:58.360 --> 00:08:59.791 And even today,

NOTE Confidence: 0.928219111666666

00:08:59.791 --> 00:09:02.176 although we have really nice

NOTE Confidence: 0.928219111666666

00:09:02.176 --> 00:09:04.779 techniques for being able to capture

NOTE Confidence: 0.928219111666666

00:09:04.779 --> 00:09:07.160 this kind of imaging techniques,

NOTE Confidence: 0.928219111666666

00:09:07.160 --> 00:09:10.760 we don't really have good quantitative

NOTE Confidence: 0.928219111666666

00:09:10.760 --> 00:09:14.567 tools to be able to analyze the

NOTE Confidence: 0.928219111666666

00:09:14.567 --> 00:09:16.575 spatial temporal patterns that

NOTE Confidence: 0.928219111666666

00:09:16.575 --> 00:09:18.959 we see in these videos.

NOTE Confidence: 0.928219111666666

00:09:18.960 --> 00:09:20.832 Likewise for brain imaging,

NOTE Confidence: 0.928219111666666

00:09:20.832 --> 00:09:23.172 we have really well developed

NOTE Confidence: 0.928219111666666

00:09:23.172 --> 00:09:26.260 tools through fMRI nears EEG,

NOTE Confidence: 0.928219111666666

00:09:26.260 --> 00:09:32.919 all kinds of tools to be to image the brain.

NOTE Confidence: 0.928219111666666

00:09:32.920 --> 00:09:36.742 And we see here for example in this in

NOTE Confidence: 0.928219111666666

00:09:36.742 --> 00:09:39.346 this example that the brain activity

NOTE Confidence: 0.928219111666666

00:09:39.346 --> 00:09:42.007 patterns that we get in a healthy

NOTE Confidence: 0.928219111666666

00:09:42.007 --> 00:09:44.152 typically developed human and an

NOTE Confidence: 0.928219111666666

00:09:44.152 --> 00:09:46.960 individual who's suffering from Alzheimer's,

NOTE Confidence: 0.928219111666666

00:09:46.960 --> 00:09:48.340 they're very different.

NOTE Confidence: 0.928219111666666

00:09:48.340 --> 00:09:51.560 We don't really have a good tool,

NOTE Confidence: 0.928219111666666

00:09:51.560 --> 00:09:54.430 tool set to be able to analyze

NOTE Confidence: 0.928219111666666

00:09:54.430 --> 00:09:56.645 the spatial temporal dynamics that

NOTE Confidence: 0.928219111666666

00:09:56.645 --> 00:09:59.357 we are seeing across the brain.

NOTE Confidence: 0.928219111666666

00:09:59.360 --> 00:10:02.756 So this problem of like quantifying

NOTE Confidence: 0.928219111666666

00:10:02.760 --> 00:10:06.000 dynamics both spatially and temporally,

NOTE Confidence: 0.928219111666666

00:10:06.000 --> 00:10:09.304 this exists not just at a cellular

NOTE Confidence: 0.928219111666666

00:10:09.304 --> 00:10:11.576 and tissue scale in biology,

NOTE Confidence: 0.928219111666666

00:10:11.576 --> 00:10:14.572 but also at a systems and organ

NOTE Confidence: 0.928219111666666

00:10:14.572 --> 00:10:15.850 scale in neuroscience.

NOTE Confidence: 0.928219111666666

00:10:15.850 --> 00:10:17.975 And this is something that

NOTE Confidence: 0.928219111666666

00:10:17.975 --> 00:10:19.960 we wish to address.

NOTE Confidence: 0.928219111666666

00:10:19.960 --> 00:10:23.360 And So what are some of the challenges

NOTE Confidence: 0.928219111666666

00:10:23.360 --> 00:10:26.033 in these data sets and how do we go

NOTE Confidence: 0.928219111666666

00:10:26.033 --> 00:10:28.514 from these noisy high dimensional

NOTE Confidence: 0.928219111666666

00:10:28.514 --> 00:10:31.676 neuroimaging data sets to neural insights?

NOTE Confidence: 0.928219111666666

00:10:31.680 --> 00:10:34.578 And what I mean by neural insights

NOTE Confidence: 0.928219111666666

00:10:34.578 --> 00:10:37.030 here is really figuring out

NOTE Confidence: 0.928219111666666

00:10:37.030 --> 00:10:39.820 patterns of activity both spatially

NOTE Confidence: 0.928219111666666

00:10:39.820 --> 00:10:43.111 and temporally in the brain that

NOTE Confidence: 0.928219111666666

00:10:43.111 --> 00:10:46.159 correspond to various kinds of stimuli,

NOTE Confidence: 0.928219111666666

00:10:46.160 --> 00:10:49.280 various kinds of diseases,

NOTE Confidence: 0.928219111666666

00:10:49.280 --> 00:10:51.797 and various kinds of like tasks.

NOTE Confidence: 0.928219111666666

00:10:51.797 --> 00:10:54.016 And so ideally what we want to

NOTE Confidence: 0.928219111666666

00:10:54.016 --> 00:10:55.358 be able to do is,

NOTE Confidence: 0.928219111666666

00:10:55.360 --> 00:10:58.752 is to build a network that can take

NOTE Confidence: 0.928219111666666

00:10:58.752 --> 00:11:02.035 in patterns of brain activity and say

NOTE Confidence: 0.928219111666666

00:11:02.035 --> 00:11:05.132 that this pattern of brain activity

NOTE Confidence: 0.928219111666666

00:11:05.132 --> 00:11:08.437 corresponds to somebody who's maybe

NOTE Confidence: 0.928219111666666

00:11:08.440 --> 00:11:11.998 clicking their right thumb like this.

NOTE Confidence: 0.928219111666666

00:11:12.000 --> 00:11:14.200 And so the challenge is,

NOTE Confidence: 0.928219111666666

00:11:14.200 --> 00:11:16.588 is enormous because when we look

NOTE Confidence: 0.928219111666666

00:11:16.588 --> 00:11:18.680 at this kind of data,

NOTE Confidence: 0.928219111666666

00:11:18.680 --> 00:11:20.920 and this is again a brain imaging

NOTE Confidence: 0.928219111666666

00:11:20.920 --> 00:11:21.880 data set here,

NOTE Confidence: 0.928219111666666

00:11:21.880 --> 00:11:23.200 if you visualize the data,

NOTE Confidence: 0.928219111666666

00:11:23.200 --> 00:11:26.900 we see that there is a lot of noise

NOTE Confidence: 0.928219111666666

00:11:26.900 --> 00:11:28.720 in this data set.

NOTE Confidence: 0.928219111666666

00:11:28.720 --> 00:11:31.030 If we take just one voxel of

NOTE Confidence: 0.928219111666666

00:11:31.030 --> 00:11:32.020 this brain imaging

NOTE Confidence: 0.859633768333333

00:11:32.097 --> 00:11:34.833 data set and we visualize it over time,

NOTE Confidence: 0.859633768333333

00:11:34.840 --> 00:11:38.233 we see that we don't really see this nice

NOTE Confidence: 0.859633768333333

00:11:38.240 --> 00:11:39.920 clean line that we would like to see.

NOTE Confidence: 0.859633768333333

00:11:39.920 --> 00:11:41.204 In fact, we see that the

NOTE Confidence: 0.859633768333333

00:11:41.204 --> 00:11:42.639 data is all over the place.

NOTE Confidence: 0.859633768333333

00:11:42.640 --> 00:11:45.076 So we have to learn to be

NOTE Confidence: 0.859633768333333

00:11:45.076 --> 00:11:47.200 able to denoise this data set.

NOTE Confidence: 0.937663298333333

00:11:49.320 --> 00:11:51.096 The second thing we want to do is

NOTE Confidence: 0.937663298333333

00:11:51.096 --> 00:11:53.536 we want to be, we want to learn

NOTE Confidence: 0.937663298333333

00:11:53.536 --> 00:11:55.960 salient features of the data set.

NOTE Confidence: 0.937663298333333

00:11:55.960 --> 00:11:59.040 So in these neuroimaging data sets and also

NOTE Confidence: 0.937663298333333

00:11:59.040 --> 00:12:01.878 in calcium imaging and other data sets,

NOTE Confidence: 0.937663298333333

00:12:01.880 --> 00:12:04.545 not all features of the

NOTE Confidence: 0.937663298333333

00:12:04.545 --> 00:12:06.677 image are equally important.

NOTE Confidence: 0.937663298333333

00:12:06.680 --> 00:12:08.962 There are some features of the image

NOTE Confidence: 0.937663298333333

00:12:08.962 --> 00:12:11.238 that are salient to the task at hand,

NOTE Confidence: 0.937663298333333

00:12:11.240 --> 00:12:13.450 whether it's to diagnose individuals

NOTE Confidence: 0.937663298333333

00:12:13.450 --> 00:12:17.070 or to learn what kind of stimulus they

NOTE Confidence: 0.937663298333333

00:12:17.070 --> 00:12:19.722 they they're experiencing or to learn

NOTE Confidence: 0.937663298333333

00:12:19.722 --> 00:12:22.918 how to decode their brain activity into

NOTE Confidence: 0.937663298333333

00:12:22.920 --> 00:12:26.080 whatever stimulus that they experienced.

NOTE Confidence: 0.937663298333333

00:12:26.080 --> 00:12:29.125 So distilling the state space of the

NOTE Confidence: 0.937663298333333

00:12:29.125 --> 00:12:31.433 brain and learning salient features

NOTE Confidence: 0.937663298333333

00:12:31.433 --> 00:12:34.632 of this data set is very important.

NOTE Confidence: 0.937663298333333

00:12:34.640 --> 00:12:38.636 And finally, in neuroimaging in particular,

NOTE Confidence: 0.937663298333333

00:12:38.640 --> 00:12:42.396 we are always challenged by spatial

NOTE Confidence: 0.937663298333333

00:12:42.396 --> 00:12:44.274 versus temporal resolution.

NOTE Confidence: 0.937663298333333

00:12:44.280 --> 00:12:47.112 So we have techniques such as

NOTE Confidence: 0.937663298333333

00:12:47.112 --> 00:12:49.000 EEG which have very,

NOTE Confidence: 0.937663298333333

00:12:49.000 --> 00:12:52.200 very good temporal resolution but

NOTE Confidence: 0.937663298333333

00:12:52.200 --> 00:12:55.400 have very poor spatial resolution.

NOTE Confidence: 0.937663298333333

00:12:55.400 --> 00:12:56.716 On the other hand,

NOTE Confidence: 0.937663298333333

00:12:56.716 --> 00:12:59.202 we have techniques such as fMRI where

NOTE Confidence: 0.937663298333333

00:12:59.202 --> 00:13:01.397 the spatial resolution is amazing.

NOTE Confidence: 0.937663298333333

00:13:01.400 --> 00:13:03.400 We get thousands and thousands

NOTE Confidence: 0.937663298333333

00:13:03.400 --> 00:13:05.400 of voxels across the brain,

NOTE Confidence: 0.937663298333333

00:13:05.400 --> 00:13:08.970 but the temporal resolution of fMRI at

NOTE Confidence: 0.937663298333333

00:13:08.970 --> 00:13:13.235 around .5 Hertz is very low compared to EEG.

NOTE Confidence: 0.937663298333333

00:13:13.240 --> 00:13:15.664 So we want to develop techniques

NOTE Confidence: 0.937663298333333

00:13:15.664 --> 00:13:18.376 that can bridge bridge the gap

NOTE Confidence: 0.937663298333333

00:13:18.376 --> 00:13:20.404 between high spatial resolution

NOTE Confidence: 0.937663298333333

00:13:20.404 --> 00:13:22.432 and high temporal resolution.

NOTE Confidence: 0.937663298333333

00:13:22.440 --> 00:13:25.008 And we want to develop techniques

NOTE Confidence: 0.937663298333333

00:13:25.008 --> 00:13:27.211 that can perhaps integrate multiple

NOTE Confidence: 0.937663298333333

00:13:27.211 --> 00:13:30.151 modalities of data together so we can

NOTE Confidence: 0.937663298333333

00:13:30.151 --> 00:13:33.008 benefit from both high spatial resolution

NOTE Confidence: 0.937663298333333

00:13:33.008 --> 00:13:35.518 and also high temporal resolution.

NOTE Confidence: 0.9529839

00:13:38.480 --> 00:13:41.424 So those were just some of the motivating

NOTE Confidence: 0.9529839

00:13:41.424 --> 00:13:43.750 factors that in our lab led to the

NOTE Confidence: 0.9529839

00:13:43.750 --> 00:13:45.200 development of a technique called

NOTE Confidence: 0.852120611428571

00:13:47.800 --> 00:13:50.724 GSTHGSTH stands for geometric

NOTE Confidence: 0.852120611428571

00:13:50.724 --> 00:13:52.917 scattering trajectory Homology.

NOTE Confidence: 0.852120611428571

00:13:52.920 --> 00:13:54.838 And it's a bit of a mouthful.

NOTE Confidence: 0.852120611428571

00:13:54.840 --> 00:13:59.200 And over the next two or three workshops,

NOTE Confidence: 0.852120611428571

00:13:59.200 --> 00:14:01.920 we are going to go into all the

NOTE Confidence: 0.852120611428571

00:14:01.920 --> 00:14:05.360 components that form this methodology.

NOTE Confidence: 0.852120611428571

00:14:05.360 --> 00:14:06.998 And so today, just to begin with,

NOTE Confidence: 0.852120611428571

00:14:07.000 --> 00:14:09.555 I'll just give you a very short

NOTE Confidence: 0.852120611428571

00:14:09.555 --> 00:14:11.080 introduction for for what,

NOTE Confidence: 0.852120611428571

00:14:11.080 --> 00:14:13.957 for how this methodology kind of works.

NOTE Confidence: 0.852120611428571

00:14:13.960 --> 00:14:15.680 And so in this method,

NOTE Confidence: 0.852120611428571

00:14:15.680 --> 00:14:20.072 we start by creating a graph from our data.

NOTE Confidence: 0.852120611428571

00:14:20.080 --> 00:14:21.960 If we are dealing with

NOTE Confidence: 0.852120611428571

00:14:21.960 --> 00:14:24.072 some calcium imaging data,

NOTE Confidence: 0.852120611428571

00:14:24.072 --> 00:14:26.712 like imagine you are imaging

NOTE Confidence: 0.852120611428571

00:14:26.712 --> 00:14:28.712 calcium from the primary visual

NOTE Confidence: 0.852120611428571

00:14:28.712 --> 00:14:31.204 cortex of of a mouse and you're

NOTE Confidence: 0.852120611428571

00:14:31.204 --> 00:14:33.279 maybe imaging in like layer 4.

NOTE Confidence: 0.852120611428571

00:14:33.280 --> 00:14:36.165 Let's say you're going to

NOTE Confidence: 0.852120611428571

00:14:36.165 --> 00:14:38.240 get a sequence of images.

NOTE Confidence: 0.852120611428571

00:14:38.240 --> 00:14:41.354 And what you can do is you can use

NOTE Confidence: 0.852120611428571

00:14:41.360 --> 00:14:43.754 existing tools to segment those images

NOTE Confidence: 0.852120611428571

00:14:43.754 --> 00:14:46.997 so you know where the cells are located.

NOTE Confidence: 0.852120611428571

00:14:47.000 --> 00:14:49.317 And then you can build a graph.

NOTE Confidence: 0.852120611428571

00:14:49.320 --> 00:14:50.812 And by graph again,

NOTE Confidence: 0.852120611428571

00:14:50.812 --> 00:14:53.937 I mean nodes and edges by using the

NOTE Confidence: 0.852120611428571

00:14:53.937 --> 00:14:57.121 centroids of all the cells as nodes in

NOTE Confidence: 0.852120611428571

00:14:57.210 --> 00:15:00.059 the graph and putting an edge between

NOTE Confidence: 0.852120611428571

00:15:00.059 --> 00:15:02.880 any pair of nodes that share an edge.

NOTE Confidence: 0.852120611428571

00:15:02.880 --> 00:15:05.393 So any two cells that are adjacent

NOTE Confidence: 0.852120611428571

00:15:05.393 --> 00:15:07.892 to each other will be two nodes

NOTE Confidence: 0.852120611428571

00:15:07.892 --> 00:15:10.440 connected by an edge in the graph.

NOTE Confidence: 0.852120611428571

00:15:10.440 --> 00:15:10.868 Similarly,

NOTE Confidence: 0.852120611428571

00:15:10.868 --> 00:15:14.292 if you have some neuro imaging data set

NOTE Confidence: 0.852120611428571

00:15:14.292 --> 00:15:17.600 that you're looking to analyze with GSTH,

NOTE Confidence: 0.852120611428571

00:15:17.600 --> 00:15:20.579 what you can do is you can take the

NOTE Confidence: 0.852120611428571

00:15:20.579 --> 00:15:23.656 brain and you can convert it into

NOTE Confidence: 0.852120611428571

00:15:23.656 --> 00:15:26.000 parcels using your favorite Atlas.

NOTE Confidence: 0.852120611428571

00:15:26.000 --> 00:15:29.486 And so those individual parcels of the

NOTE Confidence: 0.852120611428571

00:15:29.486 --> 00:15:33.197 brain will form the nodes in the graph.

NOTE Confidence: 0.852120611428571

00:15:33.200 --> 00:15:36.215 And we are going to put an edge between

NOTE Confidence: 0.852120611428571

00:15:36.215 --> 00:15:39.292 any pair of parcels that are anatomically

NOTE Confidence: 0.852120611428571

00:15:39.292 --> 00:15:42.158 close to each other in the brain.

NOTE Confidence: 0.852120611428571

00:15:42.160 --> 00:15:44.554 So we start with a graph construction.

NOTE Confidence: 0.852120611428571

00:15:44.560 --> 00:15:47.656 Now each node in the graph will have

NOTE Confidence: 0.852120611428571

00:15:47.656 --> 00:15:50.785 a signal assigned to it and the signal

NOTE Confidence: 0.852120611428571

00:15:50.785 --> 00:15:54.280 is going to be a time varying signal.

NOTE Confidence: 0.852120611428571

00:15:54.280 --> 00:15:56.596 In the case of calcium imaging,

NOTE Confidence: 0.852120611428571

00:15:56.600 --> 00:15:57.456 for example,

NOTE Confidence: 0.852120611428571

00:15:57.456 --> 00:16:00.880 we are going to have as our signal

NOTE Confidence: 0.852120611428571

00:16:00.975 --> 00:16:03.800 the calcium activity over time.

NOTE Confidence: 0.852120611428571

00:16:03.800 --> 00:16:05.800 In the case of neuro,

NOTE Confidence: 0.852120611428571

00:16:05.800 --> 00:16:07.468 neuro imaging data sets,

NOTE Confidence: 0.852120611428571

00:16:07.468 --> 00:16:10.892 we are going to have averaged voxel

NOTE Confidence: 0.852120611428571

00:16:10.892 --> 00:16:14.222 activations within each parcel as

NOTE Confidence: 0.852120611428571

00:16:14.222 --> 00:16:18.438 our time lapse signal on the graph.

NOTE Confidence: 0.852120611428571

00:16:18.440 --> 00:16:19.660 And so in GSTH,

NOTE Confidence: 0.852120611428571

00:16:19.660 --> 00:16:22.399 what we do is we take that graph

NOTE Confidence: 0.852120611428571

00:16:22.400 --> 00:16:25.382 and we use some techniques in graph

NOTE Confidence: 0.852120611428571

00:16:25.382 --> 00:16:27.922 signal processing to convert the time

NOTE Confidence: 0.852120611428571

00:16:27.922 --> 00:16:30.687 lapse signal on the graph into some

NOTE Confidence: 0.852120611428571

00:16:30.769 --> 00:16:33.277 kind of numerical representation.

NOTE Confidence: 0.852120611428571

00:16:33.280 --> 00:16:35.242 So think of like taking this

NOTE Confidence: 0.852120611428571

00:16:35.242 --> 00:16:37.255 time lapse signal on the graph

NOTE Confidence: 0.852120611428571

00:16:37.255 --> 00:16:39.157 and coming up with a vector,

NOTE Confidence: 0.852120611428571

00:16:39.160 --> 00:16:42.160 which is nothing but a sequence

NOTE Confidence: 0.852120611428571

00:16:42.160 --> 00:16:45.464 of numbers that represent how that

NOTE Confidence: 0.852120611428571

00:16:45.464 --> 00:16:48.920 signal is distributed in the graph.

NOTE Confidence: 0.852120611428571

00:16:48.920 --> 00:16:51.706 And we're going to cover how this

NOTE Confidence: 0.852120611428571

00:16:51.706 --> 00:16:53.378 graph signal processing happens

NOTE Confidence: 0.852120611428571

00:16:53.378 --> 00:16:54.998 in the next workshop.

NOTE Confidence: 0.852120611428571

00:16:55.000 --> 00:16:56.836 But assuming you can do that,

NOTE Confidence: 0.852120611428571

00:16:56.840 --> 00:17:00.400 the next step in our in our methodology

NOTE Confidence: 0.852120611428571

00:17:00.400 --> 00:17:04.960 is to construct a trajectory of

NOTE Confidence: 0.852120611428571

00:17:04.960 --> 00:17:08.400 the dynamics using some non linear

NOTE Confidence: 0.852120611428571

00:17:08.400 --> 00:17:09.840 dimensionality reduction techniques.

NOTE Confidence: 0.852120611428571

00:17:09.840 --> 00:17:10.376 And again,

NOTE Confidence: 0.852120611428571

00:17:10.376 --> 00:17:12.252 this is something that we will cover

NOTE Confidence: 0.852120611428571

00:17:12.252 --> 00:17:14.237 in detail in subsequent workshops.

NOTE Confidence: 0.852120611428571

00:17:14.240 --> 00:17:16.712 But what's happening here is that

NOTE Confidence: 0.852120611428571

00:17:16.712 --> 00:17:18.808 we are representing the time

NOTE Confidence: 0.852120611428571

00:17:18.808 --> 00:17:21.040 lapse data that we started with

NOTE Confidence: 0.885165698333333

00:17:21.040 --> 00:17:24.358 in through a low dimensional trajectory.

NOTE Confidence: 0.885165698333333

00:17:24.360 --> 00:17:27.780 So in this case, I'm showing you a 3D

NOTE Confidence: 0.885165698333333

00:17:27.780 --> 00:17:29.880 trajectory and it's colored by time.

NOTE Confidence: 0.885165698333333

00:17:29.880 --> 00:17:32.320 And so we are saying that we start

NOTE Confidence: 0.885165698333333

00:17:32.320 --> 00:17:35.000 over here and we kind of move around,

NOTE Confidence: 0.885165698333333

00:17:35.000 --> 00:17:37.547 we go in a circle and we end up

NOTE Confidence: 0.885165698333333

00:17:37.547 --> 00:17:39.997 in this region of the space.

NOTE Confidence: 0.885165698333333

00:17:40.000 --> 00:17:42.944 And so recall how I talked to you

NOTE Confidence: 0.885165698333333

00:17:42.944 --> 00:17:45.357 earlier about denoising and learning

NOTE Confidence: 0.885165698333333

00:17:45.357 --> 00:17:48.549 the state space as being important

NOTE Confidence: 0.885165698333333

00:17:48.549 --> 00:17:51.438 challenges in in neural in neuroscience.

NOTE Confidence: 0.885165698333333

00:17:51.440 --> 00:17:54.155 Well, this graph signal processing

NOTE Confidence: 0.885165698333333

00:17:54.155 --> 00:17:56.870 in Step 2 effectively denoises

NOTE Confidence: 0.885165698333333

00:17:56.958 --> 00:17:59.800 the data set that we started with.

NOTE Confidence: 0.885165698333333

00:17:59.800 --> 00:18:01.789 And these trajectories,

NOTE Confidence: 0.885165698333333

00:18:01.789 --> 00:18:05.104 these low dimensional trajectories allow

NOTE Confidence: 0.885165698333333

00:18:05.104 --> 00:18:10.039 us to quantify where in state space we are.

NOTE Confidence: 0.885165698333333

00:18:10.040 --> 00:18:10.880 In particular,

NOTE Confidence: 0.885165698333333

00:18:10.880 --> 00:18:13.820 what I want to emphasize is that

NOTE Confidence: 0.885165698333333

00:18:13.820 --> 00:18:15.799 within these trajectories,

NOTE Confidence: 0.885165698333333

00:18:15.800 --> 00:18:19.560 anytime you get a loop in the trajectory,

NOTE Confidence: 0.885165698333333

00:18:19.560 --> 00:18:22.265 that means that your underlying

NOTE Confidence: 0.885165698333333

00:18:22.265 --> 00:18:24.970 signaling pattern has some kind

NOTE Confidence: 0.885165698333333

00:18:25.061 --> 00:18:27.636 of periodicity attached to it.

NOTE Confidence: 0.885165698333333

00:18:27.640 --> 00:18:30.979 Because we a loop structure in this

NOTE Confidence: 0.885165698333333

00:18:30.979 --> 00:18:33.484 low dimension indicates that we end

NOTE Confidence: 0.885165698333333

00:18:33.484 --> 00:18:36.694 up at the same state or close to the

NOTE Confidence: 0.885165698333333

00:18:36.694 --> 00:18:39.118 same state where we started from.

NOTE Confidence: 0.885165698333333

00:18:39.120 --> 00:18:42.342 So these trajectories are really quite

NOTE Confidence: 0.885165698333333

00:18:42.342 --> 00:18:45.055 informative and we can interpret the

NOTE Confidence: 0.885165698333333

00:18:45.055 --> 00:18:47.590 shape of these trajectories by looking

NOTE Confidence: 0.885165698333333

00:18:47.590 --> 00:18:50.355 at looking at our data through the

NOTE Confidence: 0.885165698333333

00:18:50.355 --> 00:18:53.759 lens of periodicity and quasi periodicity.

NOTE Confidence: 0.885165698333333

00:18:53.760 --> 00:18:56.161 So we recognize that the shape of

NOTE Confidence: 0.885165698333333

00:18:56.161 --> 00:18:58.200 these trajectories is very important.

NOTE Confidence: 0.885165698333333

00:18:58.200 --> 00:19:00.880 And in order to be able to compare

NOTE Confidence: 0.885165698333333

00:19:00.880 --> 00:19:03.280 across different data sets and across

NOTE Confidence: 0.885165698333333

00:19:03.280 --> 00:19:05.360 different subjects in an experiment,

NOTE Confidence: 0.885165698333333

00:19:05.360 --> 00:19:08.640 we need to find a way of quantifying

NOTE Confidence: 0.885165698333333

00:19:08.640 --> 00:19:11.159 the shape of the trajectory.

NOTE Confidence: 0.885165698333333

00:19:11.160 --> 00:19:14.000 And to quantify the shape of the trajectory,

NOTE Confidence: 0.885165698333333

00:19:14.000 --> 00:19:16.640 we use our topological data

NOTE Confidence: 0.885165698333333

00:19:16.640 --> 00:19:19.280 analysis as our main tool.

NOTE Confidence: 0.885165698333333

00:19:19.280 --> 00:19:21.485 And so topological data analysis is a

NOTE Confidence: 0.885165698333333

00:19:21.485 --> 00:19:23.679 technique that I'm going to cover today,

NOTE Confidence: 0.885165698333333

00:19:23.680 --> 00:19:26.520 which takes point cloud data.

NOTE Confidence: 0.885165698333333

00:19:26.520 --> 00:19:29.000 What I mean by point cloud data is just a

NOTE Confidence: 0.885165698333333

00:19:29.060 --> 00:19:31.594 bunch of points sitting in some dimension.

NOTE Confidence: 0.885165698333333

00:19:31.600 --> 00:19:32.494 In this case,

NOTE Confidence: 0.885165698333333

00:19:32.494 --> 00:19:35.077 these points are all in like 3 dimensional

NOTE Confidence: 0.885165698333333

00:19:35.077 --> 00:19:38.066 space and it it converts them into

NOTE Confidence: 0.885165698333333

00:19:38.066 --> 00:19:40.480 something called a persistence diagram.

NOTE Confidence: 0.885165698333333

00:19:40.480 --> 00:19:43.480 And this persistence diagram quantifies

NOTE Confidence: 0.885165698333333

00:19:43.480 --> 00:19:46.480 how connected those points are.

NOTE Confidence: 0.885165698333333

00:19:46.480 --> 00:19:49.399 And it also quantifies the shape of

NOTE Confidence: 0.885165698333333

00:19:49.399 --> 00:19:52.981 this data in the sense that it measures

NOTE Confidence: 0.885165698333333

00:19:52.981 --> 00:19:55.825 how loopy the trajectory is and whether

NOTE Confidence: 0.885165698333333

00:19:55.825 --> 00:19:58.959 or not that trajectory has any holes in it.

NOTE Confidence: 0.885165698333333

00:19:58.960 --> 00:19:59.600 So again,

NOTE Confidence: 0.885165698333333

00:19:59.600 --> 00:20:02.160 this might sound very abstract at this stage,

NOTE Confidence: 0.885165698333333

00:20:02.160 --> 00:20:03.904 but this is a technique that I'm going

NOTE Confidence: 0.885165698333333

00:20:03.904 --> 00:20:05.676 to talk about in more detail today,

NOTE Confidence: 0.885165698333333

00:20:05.680 --> 00:20:06.775 topological data analysis.

NOTE Confidence: 0.885165698333333

00:20:06.775 --> 00:20:10.304 And what we do then is we can take

NOTE Confidence: 0.885165698333333

00:20:10.304 --> 00:20:12.549 these topological features that are

NOTE Confidence: 0.885165698333333

00:20:12.549 --> 00:20:15.238 capturing the shape of our trajectory,

NOTE Confidence: 0.885165698333333

00:20:15.240 --> 00:20:17.601 and we can put them through some

NOTE Confidence: 0.885165698333333

00:20:17.601 --> 00:20:20.128 machine learning in order to be able

NOTE Confidence: 0.885165698333333

00:20:20.128 --> 00:20:22.880 to use GSTH as a diagnostic tool,

NOTE Confidence: 0.885165698333333

00:20:22.880 --> 00:20:23.534 for example.

NOTE Confidence: 0.885165698333333

00:20:23.534 --> 00:20:26.150 So what machine learning will do is it

NOTE Confidence: 0.885165698333333

00:20:26.222 --> 00:20:28.597 will take these topological features.

NOTE Confidence: 0.885165698333333

00:20:28.600 --> 00:20:31.834 And it will classify whether or not

NOTE Confidence: 0.885165698333333

00:20:31.834 --> 00:20:35.070 the individual that we are looking at

NOTE Confidence: 0.885165698333333

00:20:35.070 --> 00:20:37.380 is a typically developed individual

NOTE Confidence: 0.885165698333333

00:20:37.461 --> 00:20:39.611 or whether they have schizophrenia

NOTE Confidence: 0.885165698333333

00:20:39.611 --> 00:20:42.800 or they have OCD or Alzheimer's.

NOTE Confidence: 0.885165698333333

00:20:42.800 --> 00:20:44.848 What you can also do is you can

NOTE Confidence: 0.885165698333333

00:20:44.848 --> 00:20:46.816 use this technique to figure out

NOTE Confidence: 0.885165698333333

00:20:46.816 --> 00:20:48.556 whether or not the brain,

NOTE Confidence: 0.763471135384615

00:20:48.560 --> 00:20:50.541 is it in the resting state or

NOTE Confidence: 0.763471135384615

00:20:50.541 --> 00:20:52.320 is it engaged in some task.

NOTE Confidence: 0.763471135384615

00:20:52.320 --> 00:20:55.144 You can learn to figure out what task

NOTE Confidence: 0.763471135384615

00:20:55.144 --> 00:20:57.645 an individual is doing by quantifying

NOTE Confidence: 0.763471135384615

00:20:57.645 --> 00:20:59.800 the shape of these trajectories.

NOTE Confidence: 0.763471135384615

00:20:59.800 --> 00:21:01.424 And there are many,

NOTE Confidence: 0.763471135384615

00:21:01.424 --> 00:21:03.860 many other application areas that I'm

NOTE Confidence: 0.763471135384615

00:21:03.938 --> 00:21:06.714 sure you can think of applying this to.

NOTE Confidence: 0.763471135384615

00:21:06.720 --> 00:21:10.222 So just want to start with the

NOTE Confidence: 0.763471135384615

00:21:10.222 --> 00:21:12.088 workshop organization briefly.

NOTE Confidence: 0.763471135384615

00:21:12.088 --> 00:21:16.084 So we have two other fantastic speakers

NOTE Confidence: 0.763471135384615

00:21:16.084 --> 00:21:18.674 for our workshop, Rahul Singh,

NOTE Confidence: 0.763471135384615

00:21:18.674 --> 00:21:21.816 he's in the audience today and Brian

NOTE Confidence: 0.763471135384615

00:21:21.816 --> 00:21:24.112 Zabowski is also in the audience today.

NOTE Confidence: 0.763471135384615

00:21:24.120 --> 00:21:27.036 Rahul is a WOOSAI postdoctoral fellow.

NOTE Confidence: 0.763471135384615

00:21:27.040 --> 00:21:30.030 He will be talking to you next week

NOTE Confidence: 0.763471135384615

00:21:30.030 --> 00:21:32.640 and he'll be talking about graph

NOTE Confidence: 0.763471135384615

00:21:32.640 --> 00:21:34.872 signal processing methods that form

NOTE Confidence: 0.763471135384615

00:21:34.872 --> 00:21:37.434 the second step of our methodology.

NOTE Confidence: 0.763471135384615

00:21:37.440 --> 00:21:39.280 And then the following week,

NOTE Confidence: 0.763471135384615

00:21:39.280 --> 00:21:40.510 me and Brian,

NOTE Confidence: 0.763471135384615

00:21:40.510 --> 00:21:43.380 we will jointly present to you the

NOTE Confidence: 0.763471135384615

00:21:43.467 --> 00:21:46.750 entirety of the GSTH technique and we'll

NOTE Confidence: 0.763471135384615

00:21:46.750 --> 00:21:49.344 share with you several applications

NOTE Confidence: 0.763471135384615

00:21:49.344 --> 00:21:53.166 of GSTH both for cellular imaging data

NOTE Confidence: 0.763471135384615

00:21:53.166 --> 00:21:56.198 sets and also for neuro imaging data sets.

NOTE Confidence: 0.763471135384615

00:21:56.200 --> 00:21:58.920 And then of course, Helen will be around

NOTE Confidence: 0.763471135384615

00:21:58.920 --> 00:22:01.158 to facilitate all of these workshops.

NOTE Confidence: 0.763471135384615

00:22:01.160 --> 00:22:04.359 She's really the brains behind the operation.

NOTE Confidence: 0.763471135384615

00:22:04.360 --> 00:22:07.800 And so we have the 1st 3 workshops to cover

NOTE Confidence: 0.763471135384615

00:22:07.880 --> 00:22:11.240 different aspects of the GSTH methodology.

NOTE Confidence: 0.763471135384615

00:22:11.240 --> 00:22:12.920 We are starting from the end.

NOTE Confidence: 0.763471135384615

00:22:12.920 --> 00:22:14.957 So I'm going to talk about topological

NOTE Confidence: 0.763471135384615

00:22:14.957 --> 00:22:17.078 data analysis and machine learning today.

NOTE Confidence: 0.763471135384615

00:22:17.080 --> 00:22:18.884 Rahul within talked about

NOTE Confidence: 0.763471135384615

00:22:18.884 --> 00:22:20.237 graph signal processing.

NOTE Confidence: 0.763471135384615

00:22:20.240 --> 00:22:21.944 In the third workshop,

NOTE Confidence: 0.763471135384615

00:22:21.944 --> 00:22:24.074 we'll bring these things together

NOTE Confidence: 0.763471135384615

00:22:24.074 --> 00:22:26.896 and go over the complete GSTH

NOTE Confidence: 0.763471135384615

00:22:26.896 --> 00:22:28.756 methodology and its applications.

NOTE Confidence: 0.763471135384615

00:22:28.760 --> 00:22:31.240 And then the final week of the workshop,

NOTE Confidence: 0.763471135384615

00:22:31.240 --> 00:22:34.294 we are going to do a hands on

NOTE Confidence: 0.763471135384615

00:22:34.294 --> 00:22:36.976 tutorial where you'll get to load

NOTE Confidence: 0.763471135384615

00:22:36.976 --> 00:22:39.819 some neuro imaging data set and

NOTE Confidence: 0.763471135384615

00:22:39.819 --> 00:22:42.603 also cellular imaging data set and

NOTE Confidence: 0.763471135384615

00:22:42.696 --> 00:22:45.396 analyze it using GSTH in Python.

NOTE Confidence: 0.763471135384615

00:22:45.400 --> 00:22:47.080 And at the moment I think

NOTE Confidence: 0.763471135384615

00:22:47.080 --> 00:22:48.200 we're planning to make,

NOTE Confidence: 0.763471135384615

00:22:48.200 --> 00:22:50.084 we're planning to hold our 4th

NOTE Confidence: 0.763471135384615

00:22:50.084 --> 00:22:52.245 workshop as like a hybrid workshop

NOTE Confidence: 0.763471135384615

00:22:52.245 --> 00:22:54.658 that might have in person component.

NOTE Confidence: 0.763471135384615

00:22:54.658 --> 00:22:57.785 So we'll get back to you on on that and

NOTE Confidence: 0.763471135384615

00:22:57.785 --> 00:23:00.396 the location for that in subsequent weeks.

NOTE Confidence: 0.825900406470588

00:23:00.720 --> 00:23:03.289 Yes, I will send the series of

NOTE Confidence: 0.825900406470588

00:23:03.289 --> 00:23:05.606 emails where people can sign up

NOTE Confidence: 0.825900406470588

00:23:05.606 --> 00:23:07.520 for in person component. Great.

NOTE Confidence: 0.77538607875

00:23:10.160 --> 00:23:13.480 All right. So we have few live participants.

NOTE Confidence: 0.77538607875

00:23:13.480 --> 00:23:16.007 I understand that the majority of these

NOTE Confidence: 0.77538607875

00:23:16.007 --> 00:23:18.326 workshops get viewed online over a period

NOTE Confidence: 0.77538607875

00:23:18.326 --> 00:23:20.520 of like weeks and months and years.

NOTE Confidence: 0.77538607875

00:23:20.520 --> 00:23:23.110 So please feel free to stop me

NOTE Confidence: 0.77538607875

00:23:23.110 --> 00:23:25.080 anytime and to ask questions.

NOTE Confidence: 0.77538607875

00:23:25.080 --> 00:23:28.644 And so as I mentioned, I'm going to start

NOTE Confidence: 0.77538607875

00:23:28.644 --> 00:23:30.599 with with topological data analysis.

NOTE Confidence: 0.77538607875

00:23:30.600 --> 00:23:32.903 And depending on how much time I

NOTE Confidence: 0.77538607875

00:23:32.903 --> 00:23:34.890 have available to me, I might,

NOTE Confidence: 0.77538607875

00:23:34.890 --> 00:23:37.080 I will also cover some fundamentals

NOTE Confidence: 0.77538607875

00:23:37.080 --> 00:23:39.579 of machine learning just to make sure

NOTE Confidence: 0.77538607875

00:23:39.579 --> 00:23:41.992 that everybody is on the same page

NOTE Confidence: 0.77538607875

00:23:41.992 --> 00:23:44.267 and we all share the same vocabulary

NOTE Confidence: 0.77538607875

00:23:44.267 --> 00:23:46.428 in the weeks going forward.

NOTE Confidence: 0.77538607875

00:23:46.428 --> 00:23:48.713 So let's start with TDA.

NOTE Confidence: 0.77538607875

00:23:48.720 --> 00:23:51.240 And I wanted to start by just showing

NOTE Confidence: 0.77538607875

00:23:51.240 --> 00:23:53.998 you some of these point cloud examples.

NOTE Confidence: 0.77538607875

00:23:54.000 --> 00:23:56.160 And so when I look at these data sets,

NOTE Confidence: 0.77538607875

00:23:56.160 --> 00:23:59.049 what I see is that maybe in the first

NOTE Confidence: 0.77538607875

00:23:59.049 --> 00:24:01.284 data set, we have two variables.

NOTE Confidence: 0.77538607875

00:24:01.284 --> 00:24:03.570 Maybe we have an independent variable

NOTE Confidence: 0.77538607875

00:24:03.636 --> 00:24:06.300 and a dependent variable that are

NOTE Confidence: 0.77538607875

00:24:06.300 --> 00:24:07.632 strongly correlated together.

NOTE Confidence: 0.77538607875

00:24:07.640 --> 00:24:09.808 And this to me looks kind of like

NOTE Confidence: 0.77538607875

00:24:09.808 --> 00:24:11.200 a linear correlation,

NOTE Confidence: 0.77538607875

00:24:11.200 --> 00:24:13.958 like a regression type of data set.

NOTE Confidence: 0.77538607875

00:24:13.960 --> 00:24:16.714 When I look at the second data set here,

NOTE Confidence: 0.77538607875

00:24:16.720 --> 00:24:19.080 what I'm recognizing is that

NOTE Confidence: 0.77538607875

00:24:19.080 --> 00:24:21.440 the data set is clustered.

NOTE Confidence: 0.77538607875

00:24:21.440 --> 00:24:24.352 We have a bunch of points that are

NOTE Confidence: 0.77538607875

00:24:24.352 --> 00:24:26.759 grouped together and we have kind of

NOTE Confidence: 0.77538607875

00:24:26.759 --> 00:24:29.439 three clusters of data in this data set.

NOTE Confidence: 0.77538607875

00:24:29.440 --> 00:24:31.280 The third data set,

NOTE Confidence: 0.77538607875

00:24:31.280 --> 00:24:33.120 to me looks cyclical.

NOTE Confidence: 0.77538607875

00:24:33.120 --> 00:24:36.558 I can spot a circle in this data set,

NOTE Confidence: 0.77538607875

00:24:36.560 --> 00:24:38.798 and that might indicate that perhaps

NOTE Confidence: 0.77538607875

00:24:38.798 --> 00:24:41.398 this is some time lapse data set.

NOTE Confidence: 0.77538607875

00:24:41.400 --> 00:24:43.386 Maybe there's some kind of oscillatory

NOTE Confidence: 0.77538607875

00:24:43.386 --> 00:24:45.000 nature to this data set,

NOTE Confidence: 0.77538607875

00:24:45.000 --> 00:24:47.328 and maybe we're going around in

NOTE Confidence: 0.77538607875

00:24:47.328 --> 00:24:49.420 circles and the last data set

NOTE Confidence: 0.77538607875

00:24:49.420 --> 00:24:51.840 here has this kind of Y shape.

NOTE Confidence: 0.77538607875

00:24:51.840 --> 00:24:55.080 It looks like it's kind of branching out.

NOTE Confidence: 0.77538607875

00:24:55.080 --> 00:24:57.166 This could be maybe some stem cells

NOTE Confidence: 0.77538607875

00:24:57.166 --> 00:24:59.330 down here that are, you know,

NOTE Confidence: 0.77538607875

00:24:59.330 --> 00:25:01.755 differentiating into two different lineages.

NOTE Confidence: 0.77538607875

00:25:01.760 --> 00:25:04.469 It seems to have this tree like

NOTE Confidence: 0.77538607875

00:25:04.469 --> 00:25:05.920 hyperbolic structure to it.

NOTE Confidence: 0.77538607875

00:25:05.920 --> 00:25:08.080 And so our brains are really,

NOTE Confidence: 0.77538607875

00:25:08.080 --> 00:25:10.790 really good at recognizing the

NOTE Confidence: 0.77538607875

00:25:10.790 --> 00:25:12.958 shape of the data,

NOTE Confidence: 0.77538607875

00:25:12.960 --> 00:25:15.012 especially when the data is presented

NOTE Confidence: 0.77538607875

00:25:15.012 --> 00:25:17.440 to us in these low dimensions.

NOTE Confidence: 0.77538607875

00:25:17.440 --> 00:25:19.376 And we understand fundamentally

NOTE Confidence: 0.77538607875

00:25:19.376 --> 00:25:22.280 that any data that we have,

NOTE Confidence: 0.77538607875

00:25:22.280 --> 00:25:24.480 that data has some shape,

NOTE Confidence: 0.77538607875

00:25:24.480 --> 00:25:27.276 and the shape carries some meaning.

NOTE Confidence: 0.77538607875

00:25:27.280 --> 00:25:30.334 And this really is the central

NOTE Confidence: 0.77538607875

00:25:30.334 --> 00:25:33.280 tenet of topological data analysis,

NOTE Confidence: 0.77538607875

00:25:33.280 --> 00:25:36.292 which is a branch of applied

NOTE Confidence: 0.77538607875

00:25:36.292 --> 00:25:38.300 mathematics and computer science

NOTE Confidence: 0.77538607875

00:25:38.390 --> 00:25:41.354 that has to do with understanding

NOTE Confidence: 0.77538607875

00:25:41.354 --> 00:25:44.160 fundamentally the shape of our data.

NOTE Confidence: 0.77538607875

00:25:44.160 --> 00:25:46.946 And underlying all of this is what

NOTE Confidence: 0.77538607875

00:25:46.946 --> 00:25:49.959 we call the manifold hypothesis.

NOTE Confidence: 0.77538607875

00:25:49.960 --> 00:25:53.152 The idea being that any scientific

NOTE Confidence: 0.77538607875

00:25:53.152 --> 00:25:56.637 data that we collect in our lab is

NOTE Confidence: 0.77538607875

00:25:56.640 --> 00:25:58.754 it might look very noisy and it

NOTE Confidence: 0.77538607875

00:25:58.754 --> 00:26:00.880 might be very high dimensional.

NOTE Confidence: 0.77538607875

00:26:00.880 --> 00:26:03.736 But quite often that scientific data

NOTE Confidence: 0.77538607875

00:26:03.736 --> 00:26:06.658 is sampled from some low dimensional

NOTE Confidence: 0.77538607875

00:26:06.658 --> 00:26:09.927 manifold And what we are really after

NOTE Confidence: 0.77538607875

00:26:09.927 --> 00:26:12.819 is to understand what that manifold

NOTE Confidence: 0.77538607875

00:26:12.819 --> 00:26:15.740 looks like and what the intrinsic

NOTE Confidence: 0.77538607875

00:26:15.740 --> 00:26:18.640 dimension of that manifold is.

NOTE Confidence: 0.77538607875

00:26:18.640 --> 00:26:21.335 So in this example here our manifold

NOTE Confidence: 0.77538607875

00:26:21.335 --> 00:26:24.663 looks to be kind of saddle shaped and

NOTE Confidence: 0.77538607875

00:26:24.663 --> 00:26:27.320 it has these two curvature areas.

NOTE Confidence: 0.77538607875

00:26:27.320 --> 00:26:29.960 So it has a direction of positive curvature,

NOTE Confidence: 0.88705654

00:26:29.960 --> 00:26:32.000 a direction of negative curvature,

NOTE Confidence: 0.88705654

00:26:32.000 --> 00:26:34.265 and our data is simply

NOTE Confidence: 0.88705654

00:26:34.265 --> 00:26:36.077 sampled from this manifold.

NOTE Confidence: 0.88705654

00:26:36.080 --> 00:26:37.865 So what we really want to understand

NOTE Confidence: 0.88705654

00:26:37.865 --> 00:26:39.520 is the shape of the manifold.

NOTE Confidence: 0.88705654

00:26:39.520 --> 00:26:42.159 Another way to look at this is

NOTE Confidence: 0.88705654

00:26:42.159 --> 00:26:45.155 what we get in our experiments

NOTE Confidence: 0.88705654

00:26:45.155 --> 00:26:47.515 are individual data points,

NOTE Confidence: 0.88705654

00:26:47.520 --> 00:26:49.560 and those data points all

NOTE Confidence: 0.88705654

00:26:49.560 --> 00:26:52.160 together form some kind of shape.

NOTE Confidence: 0.88705654

00:26:52.160 --> 00:26:54.169 And what we really want to see

NOTE Confidence: 0.88705654

00:26:54.169 --> 00:26:56.240 is what that shape looks like.

NOTE Confidence: 0.88705654

00:26:56.240 --> 00:26:57.408 So in this case,

NOTE Confidence: 0.88705654

00:26:57.408 --> 00:26:59.160 all these data points form a

NOTE Confidence: 0.88705654

00:26:59.230 --> 00:27:00.916 torus and this is kind of,

NOTE Confidence: 0.88705654

00:27:00.920 --> 00:27:02.948 this is the realization that we

NOTE Confidence: 0.88705654

00:27:02.948 --> 00:27:05.287 are going to come to is that

NOTE Confidence: 0.88705654

00:27:05.287 --> 00:27:07.111 our data is arranged in the

NOTE Confidence: 0.88705654

00:27:07.111 --> 00:27:09.278 shape of a doughnut or a Taurus.

NOTE Confidence: 0.88705654

00:27:09.280 --> 00:27:12.439 So how do we actually go about doing that?

NOTE Confidence: 0.88705654

00:27:12.440 --> 00:27:15.728 Let me share with you the methodology

NOTE Confidence: 0.88705654

00:27:15.728 --> 00:27:19.256 using some very simple data sets that

NOTE Confidence: 0.88705654

00:27:19.256 --> 00:27:22.252 are easy to plot in A2 dimensional slide.

NOTE Confidence: 0.88705654

00:27:22.252 --> 00:27:24.928 And so we'll be working with these two

NOTE Confidence: 0.88705654

00:27:24.928 --> 00:27:27.077 data sets for the next few slides.

NOTE Confidence: 0.88705654

00:27:27.080 --> 00:27:28.556 The data set on the left,

NOTE Confidence: 0.88705654

00:27:28.560 --> 00:27:31.008 I'm going to call the concentric

NOTE Confidence: 0.88705654

00:27:31.008 --> 00:27:33.567 circles data set and and that's

NOTE Confidence: 0.88705654

00:27:33.567 --> 00:27:36.105 simply in recognition of the fact

NOTE Confidence: 0.88705654

00:27:36.105 --> 00:27:38.260 that these points are sampled

NOTE Confidence: 0.88705654

00:27:38.260 --> 00:27:40.708 from 2 circles where one circle

NOTE Confidence: 0.88705654

00:27:40.708 --> 00:27:42.296 is within another circle.

NOTE Confidence: 0.88705654

00:27:42.296 --> 00:27:45.040 And the data set on the right,

NOTE Confidence: 0.88705654

00:27:45.040 --> 00:27:48.152 I'm going to call the half moons data

NOTE Confidence: 0.88705654

00:27:48.152 --> 00:27:50.799 set simply because both of these,

NOTE Confidence: 0.88705654

00:27:50.800 --> 00:27:53.056 we have kind of two arcs in our

NOTE Confidence: 0.88705654

00:27:53.056 --> 00:27:55.460 data and they both look like kind

NOTE Confidence: 0.88705654

00:27:55.460 --> 00:27:57.720 of half moons or Crescent moons.

NOTE Confidence: 0.88705654

00:27:57.720 --> 00:27:59.970 And So what we want to do is we

NOTE Confidence: 0.88705654

00:27:59.970 --> 00:28:02.342 want to use a technique to recognize

NOTE Confidence: 0.88705654

00:28:02.342 --> 00:28:05.137 the fact that our data on the left

NOTE Confidence: 0.88705654

00:28:05.137 --> 00:28:06.837 is arranged in two circles.

NOTE Confidence: 0.88705654

00:28:06.840 --> 00:28:08.316 And the data on the right,

NOTE Confidence: 0.88705654

00:28:08.320 --> 00:28:10.200 it looks kind of circular,

NOTE Confidence: 0.88705654

00:28:10.200 --> 00:28:13.596 but it's not really two circles

NOTE Confidence: 0.88705654

00:28:13.600 --> 00:28:15.520 or one circle for that matter.

NOTE Confidence: 0.88705654

00:28:15.520 --> 00:28:16.745 And so one thing you might want

NOTE Confidence: 0.88705654

00:28:16.745 --> 00:28:17.840 to do is you want to,

NOTE Confidence: 0.88705654

00:28:17.840 --> 00:28:20.087 you might consider it like using a

NOTE Confidence: 0.88705654

00:28:20.087 --> 00:28:22.157 clustering method to see if that works,

NOTE Confidence: 0.88705654

00:28:22.160 --> 00:28:22.488 right?

NOTE Confidence: 0.88705654

00:28:22.488 --> 00:28:24.784 So you could take those data points

NOTE Confidence: 0.88705654

00:28:24.784 --> 00:28:26.959 and throw them into an algorithm,

NOTE Confidence: 0.88705654

00:28:26.960 --> 00:28:29.120 maybe something similar to K means.

NOTE Confidence: 0.88705654

00:28:29.120 --> 00:28:30.080 And you might see like,

NOTE Confidence: 0.88705654

00:28:30.080 --> 00:28:31.960 OK, does the data cluster?

NOTE Confidence: 0.88705654

00:28:31.960 --> 00:28:32.300 Well,

NOTE Confidence: 0.88705654

00:28:32.300 --> 00:28:35.360 if you run this data set through K means,

NOTE Confidence: 0.88705654

00:28:35.360 --> 00:28:37.238 you'll end up with these clusters,

NOTE Confidence: 0.88705654

00:28:37.240 --> 00:28:39.795 the blue cluster and the audience cluster.

NOTE Confidence: 0.88705654

00:28:39.800 --> 00:28:41.990 And these two clusters don't really

NOTE Confidence: 0.88705654

00:28:41.990 --> 00:28:45.039 tell you the true story behind the data.

NOTE Confidence: 0.88705654

00:28:45.040 --> 00:28:45.740 In particular,

NOTE Confidence: 0.88705654

00:28:45.740 --> 00:28:47.840 they don't recognize the fact that

NOTE Confidence: 0.88705654

00:28:47.840 --> 00:28:50.239 these data are arranged in two circles.

NOTE Confidence: 0.88705654

00:28:50.240 --> 00:28:50.944 And we,

NOTE Confidence: 0.88705654

00:28:50.944 --> 00:28:53.056 we even get some miss clustering

NOTE Confidence: 0.88705654

00:28:53.056 --> 00:28:55.559 happening in the data set on the right.

NOTE Confidence: 0.88705654

00:28:55.560 --> 00:28:57.360 Now you might then go back and say that,

NOTE Confidence: 0.88705654

00:28:57.360 --> 00:28:57.780 OK,

NOTE Confidence: 0.88705654

00:28:57.780 --> 00:29:00.300 I should use a different sort

NOTE Confidence: 0.88705654

00:29:00.300 --> 00:29:01.560 of clustering technique.

NOTE Confidence: 0.88705654

00:29:01.560 --> 00:29:04.998 Maybe I can cluster the data by its density.

NOTE Confidence: 0.88705654

00:29:05.000 --> 00:29:08.200 And so when you employ a density based

NOTE Confidence: 0.88705654

00:29:08.200 --> 00:29:10.200 clustering methods such as DB scan,

NOTE Confidence: 0.88705654

00:29:10.200 --> 00:29:12.762 you do indeed get the correct

NOTE Confidence: 0.88705654

00:29:12.762 --> 00:29:14.960 cluster labels for your data.

NOTE Confidence: 0.88705654

00:29:14.960 --> 00:29:16.472 You are able to separate data

NOTE Confidence: 0.88705654

00:29:16.472 --> 00:29:18.100 points in the inner circle from

NOTE Confidence: 0.88705654

00:29:18.100 --> 00:29:19.798 data points in the outer circle,

NOTE Confidence: 0.88705654

00:29:19.800 --> 00:29:22.131 and you are able to separate the

NOTE Confidence: 0.88705654

00:29:22.131 --> 00:29:24.220 data points belonging to the upper

NOTE Confidence: 0.88705654

00:29:24.220 --> 00:29:26.600 Crescent moon and the lower Crescent moon.

NOTE Confidence: 0.826497284285714

00:29:26.600 --> 00:29:29.869 Even then, the machine doesn't really know

NOTE Confidence: 0.826497284285714

00:29:29.869 --> 00:29:33.279 that the data is arranged as circles.

NOTE Confidence: 0.826497284285714

00:29:33.280 --> 00:29:34.960 It has no recognition of that.

NOTE Confidence: 0.826497284285714

00:29:34.960 --> 00:29:36.955 It has simply learned that your data

NOTE Confidence: 0.826497284285714

00:29:36.955 --> 00:29:38.760 is clustered into these two groups,

NOTE Confidence: 0.826497284285714

00:29:38.760 --> 00:29:41.640 but it doesn't fundamentally understand.

NOTE Confidence: 0.826497284285714

00:29:41.640 --> 00:29:44.712 What we can tell immediately is that this

NOTE Confidence: 0.826497284285714

00:29:44.712 --> 00:29:47.996 data is arranged in a circular pattern.

NOTE Confidence: 0.826497284285714

00:29:48.000 --> 00:29:51.640 And so this is where topology comes in.

NOTE Confidence: 0.826497284285714

00:29:51.640 --> 00:29:53.656 And so I'm going to talk

NOTE Confidence: 0.826497284285714

00:29:53.656 --> 00:29:55.000 to you about topology.

NOTE Confidence: 0.826497284285714

00:29:55.000 --> 00:29:57.639 And because I'm a very visual learner,

NOTE Confidence: 0.826497284285714

00:29:57.640 --> 00:30:00.258 I'm going to use some animations and

NOTE Confidence: 0.826497284285714

00:30:00.258 --> 00:30:03.440 some figures to kind of demonstrate how

NOTE Confidence: 0.826497284285714

00:30:03.440 --> 00:30:05.920 topology works without necessarily going

NOTE Confidence: 0.826497284285714

00:30:05.920 --> 00:30:09.238 into all the math and all the code behind it.

NOTE Confidence: 0.826497284285714

00:30:09.240 --> 00:30:11.606 We'll get to, we'll use some of

NOTE Confidence: 0.826497284285714

00:30:11.606 --> 00:30:14.198 the code in our third workshop.

NOTE Confidence: 0.826497284285714

00:30:14.200 --> 00:30:16.012 But honestly, like the code is

NOTE Confidence: 0.826497284285714

00:30:16.012 --> 00:30:17.720 something you import and you use.

NOTE Confidence: 0.826497284285714

00:30:17.720 --> 00:30:20.870 And so I think it's much more important to

NOTE Confidence: 0.826497284285714

00:30:20.870 --> 00:30:24.080 kind of build intuition around topology.

NOTE Confidence: 0.826497284285714

00:30:24.080 --> 00:30:27.230 So what we do in in in topology is we

NOTE Confidence: 0.826497284285714

00:30:27.324 --> 00:30:31.079 build something called simplicial complexes.

NOTE Confidence: 0.826497284285714

00:30:31.080 --> 00:30:34.088 And there's a number of different kinds of

NOTE Confidence: 0.826497284285714

00:30:34.088 --> 00:30:36.037 simplicial complexes that one can build.

NOTE Confidence: 0.826497284285714

00:30:36.040 --> 00:30:39.376 But I'm going to talk about the viatorius

NOTE Confidence: 0.826497284285714

00:30:39.376 --> 00:30:42.180 ribs simplicial complex to begin with today.

NOTE Confidence: 0.826497284285714

00:30:42.180 --> 00:30:44.786 And so to create a Viatorius ribs

NOTE Confidence: 0.826497284285714

00:30:44.786 --> 00:30:47.076 simplicial complex from your data,

NOTE Confidence: 0.826497284285714

00:30:47.080 --> 00:30:50.941 what you do is you start with a given

NOTE Confidence: 0.826497284285714

00:30:50.941 --> 00:30:55.030 data point and you imagine a disk of some

NOTE Confidence: 0.826497284285714

00:30:55.030 --> 00:30:57.840 radius epsilon around that data point.

NOTE Confidence: 0.826497284285714

00:30:57.840 --> 00:31:00.017 And you do this for every other

NOTE Confidence: 0.826497284285714

00:31:00.017 --> 00:31:01.918 data point in the data set.

NOTE Confidence: 0.826497284285714

00:31:01.920 --> 00:31:05.004 And you're going to grow this

NOTE Confidence: 0.826497284285714

00:31:05.004 --> 00:31:07.640 epsilon radius disk over time.

NOTE Confidence: 0.826497284285714

00:31:07.640 --> 00:31:10.752 And what you're going to do is when

NOTE Confidence: 0.826497284285714

00:31:10.752 --> 00:31:13.226 two epsilon radius disks intersect

NOTE Confidence: 0.826497284285714

00:31:13.226 --> 00:31:16.196 with each other, so they overlap,

NOTE Confidence: 0.826497284285714

00:31:16.196 --> 00:31:19.360 you're going to draw an edge between

NOTE Confidence: 0.826497284285714

00:31:19.452 --> 00:31:23.036 those two data points creating A1 simplex.

NOTE Confidence: 0.826497284285714

00:31:23.040 --> 00:31:26.020 When you have three points shown

NOTE Confidence: 0.826497284285714

00:31:26.020 --> 00:31:27.920 here as AB and C,

NOTE Confidence: 0.826497284285714

00:31:27.920 --> 00:31:30.608 and they their epsilon discs all

NOTE Confidence: 0.826497284285714

00:31:30.608 --> 00:31:33.520 intersect in a pair wise manner,

NOTE Confidence: 0.826497284285714

00:31:33.520 --> 00:31:36.152 then we're going to draw a filled

NOTE Confidence: 0.826497284285714

00:31:36.152 --> 00:31:38.706 in triangle which we are going to

NOTE Confidence: 0.826497284285714

00:31:38.706 --> 00:31:39.717 call A2 simplex.

NOTE Confidence: 0.826497284285714

00:31:39.720 --> 00:31:41.840 And then in higher dimensions,

NOTE Confidence: 0.826497284285714

00:31:41.840 --> 00:31:44.381 when we have four data points all

NOTE Confidence: 0.826497284285714

00:31:44.381 --> 00:31:46.760 intersecting in a pair wise manner,

NOTE Confidence: 0.826497284285714

00:31:46.760 --> 00:31:49.640 then we're going to draw a three simplex.

NOTE Confidence: 0.826497284285714

00:31:49.640 --> 00:31:51.674 So we are going to take our data set.

NOTE Confidence: 0.826497284285714

00:31:51.680 --> 00:31:54.976 In this case the data set happens to

NOTE Confidence: 0.826497284285714

00:31:54.976 --> 00:31:57.938 be two-dimensional and we are we are

NOTE Confidence: 0.826497284285714

00:31:57.938 --> 00:31:59.998 constructing these simplices from our

NOTE Confidence: 0.826497284285714

00:32:00.076 --> 00:32:02.616 data by expanding these epsilon radius

NOTE Confidence: 0.826497284285714

00:32:02.616 --> 00:32:05.880 discs around each point in the data set.

NOTE Confidence: 0.826497284285714

00:32:05.880 --> 00:32:08.838 And so in this visualization here,

NOTE Confidence: 0.826497284285714

00:32:08.840 --> 00:32:10.568 I'm simply showing you the 0

NOTE Confidence: 0.826497284285714

00:32:10.568 --> 00:32:12.370 simplex which which are the data

NOTE Confidence: 0.826497284285714

00:32:12.370 --> 00:32:13.835 points that we started from,

NOTE Confidence: 0.826497284285714

00:32:13.840 --> 00:32:16.514 and the one simplex which are all

NOTE Confidence: 0.826497284285714

00:32:16.514 --> 00:32:20.007 the edges that get created as we are

NOTE Confidence: 0.826497284285714

00:32:20.007 --> 00:32:22.272 expanding this epsilon radius disk.

NOTE Confidence: 0.826497284285714

00:32:22.280 --> 00:32:24.752 I'm not showing you the disk and I'm

NOTE Confidence: 0.826497284285714

00:32:24.752 --> 00:32:27.396 not showing you the field in triangles

NOTE Confidence: 0.826497284285714

00:32:27.396 --> 00:32:29.752 or the tetrahedrons simply because the

NOTE Confidence: 0.826497284285714

00:32:29.752 --> 00:32:34.040 the figure gets very, very crowded.

NOTE Confidence: 0.826497284285714

00:32:34.040 --> 00:32:37.768 So why do we want to construct this

NOTE Confidence: 0.826497284285714

00:32:37.768 --> 00:32:40.119 via Torres Ribs complex?

NOTE Confidence: 0.826497284285714

00:32:40.120 --> 00:32:40.528 Well,

NOTE Confidence: 0.826497284285714

00:32:40.528 --> 00:32:44.328 it turns out if you have some data shown

NOTE Confidence: 0.826497284285714

00:32:44.328 --> 00:32:48.552 here as these red dots that are sampled,

NOTE Confidence: 0.826497284285714

00:32:48.560 --> 00:32:50.760 this is your experimental data,

NOTE Confidence: 0.826497284285714

00:32:50.760 --> 00:32:52.595 and you imagine that this

NOTE Confidence: 0.826497284285714

00:32:52.595 --> 00:32:54.430 data is coming from some

NOTE Confidence: 0.877120092857143

00:32:54.504 --> 00:32:56.560 kind of underlying manifold.

NOTE Confidence: 0.877120092857143

00:32:56.560 --> 00:32:58.756 So there's a recognition here that

NOTE Confidence: 0.877120092857143

00:32:58.756 --> 00:33:00.905 whatever data we sample comes from

NOTE Confidence: 0.877120092857143

00:33:00.905 --> 00:33:03.068 a manifold that has maybe two holes

NOTE Confidence: 0.877120092857143

00:33:03.068 --> 00:33:05.660 in the middle of it, it turns out,

NOTE Confidence: 0.877120092857143

00:33:05.660 --> 00:33:08.320 and there's a theorem to prove this,

NOTE Confidence: 0.877120092857143

00:33:08.320 --> 00:33:09.559 although we're not going to go through.

NOTE Confidence: 0.877120092857143

00:33:09.560 --> 00:33:12.638 The proof of the theorem is that if your

NOTE Confidence: 0.877120092857143

00:33:12.638 --> 00:33:15.612 data is well sampled, so all these X,

NOTE Confidence: 0.877120092857143

00:33:15.612 --> 00:33:17.610 the points in X are sampled

NOTE Confidence: 0.877120092857143

00:33:17.689 --> 00:33:20.399 throughout the manifold quite well,

NOTE Confidence: 0.877120092857143

00:33:20.400 --> 00:33:23.405 then when you construct the

NOTE Confidence: 0.877120092857143

00:33:23.405 --> 00:33:26.410 viatorus rips complex from this

NOTE Confidence: 0.877120092857143

00:33:26.516 --> 00:33:29.876 data set for some radius epsilon,

NOTE Confidence: 0.877120092857143

00:33:29.880 --> 00:33:33.184 then this viatorus rips complex is basically

NOTE Confidence: 0.877120092857143

00:33:33.184 --> 00:33:36.199 equivalent to the underlying manifold.

NOTE Confidence: 0.877120092857143

00:33:36.200 --> 00:33:39.513 So in kind of more intuitive terms,

NOTE Confidence: 0.877120092857143

00:33:39.513 --> 00:33:42.219 what this theorem is saying is

NOTE Confidence: 0.877120092857143

00:33:42.219 --> 00:33:44.983 that if you want to learn the

NOTE Confidence: 0.877120092857143

00:33:44.983 --> 00:33:47.387 shape of your manifold where the

NOTE Confidence: 0.877120092857143

00:33:47.387 --> 00:33:49.477 data is being sampled from,

NOTE Confidence: 0.877120092857143

00:33:49.480 --> 00:33:53.127 it is sufficient to construct a Vietorisrips

NOTE Confidence: 0.877120092857143

00:33:53.127 --> 00:33:56.079 complex at some radius epsilon.

NOTE Confidence: 0.877120092857143

00:33:56.080 --> 00:33:58.969 And you will be able to find the manifold

NOTE Confidence: 0.877120092857143

00:33:58.969 --> 00:34:01.165 underneath the data and you'll be able

NOTE Confidence: 0.877120092857143

00:34:01.165 --> 00:34:03.496 to recognize the fact that your data

NOTE Confidence: 0.877120092857143

00:34:03.496 --> 00:34:05.608 is forming this one connected object

NOTE Confidence: 0.877120092857143

00:34:05.608 --> 00:34:09.920 which has two holes punched into it.

NOTE Confidence: 0.877120092857143

00:34:09.920 --> 00:34:13.000 OK, so let's get back to our example,

NOTE Confidence: 0.877120092857143

00:34:13.000 --> 00:34:14.972 the concentric circles example

NOTE Confidence: 0.877120092857143

00:34:14.972 --> 00:34:17.437 and the half moons example.

NOTE Confidence: 0.877120092857143

00:34:17.440 --> 00:34:20.485 And so here I'm showing you those

NOTE Confidence: 0.877120092857143

00:34:20.485 --> 00:34:23.158 epsilon radius discs around the data.

NOTE Confidence: 0.877120092857143

00:34:23.160 --> 00:34:26.440 So we have epsilon equals

NOTE Confidence: 0.877120092857143

00:34:26.440 --> 00:34:28.324 .05 at the beginning,

NOTE Confidence: 0.877120092857143

00:34:28.324 --> 00:34:30.679 we increase our epsilon value.

NOTE Confidence: 0.877120092857143

00:34:30.680 --> 00:34:32.717 And as we increase the epsilon value,

NOTE Confidence: 0.877120092857143

00:34:32.720 --> 00:34:35.275 these discs that I'm plotting in grade,

NOTE Confidence: 0.877120092857143

00:34:35.280 --> 00:34:37.866 they get bigger and bigger until

NOTE Confidence: 0.877120092857143

00:34:37.866 --> 00:34:40.160 they cover the whole space.

NOTE Confidence: 0.877120092857143

00:34:40.160 --> 00:34:42.932 And So what you can recognize here

NOTE Confidence: 0.877120092857143

00:34:42.932 --> 00:34:46.238 is that when our disk is quite small,

NOTE Confidence: 0.877120092857143

00:34:46.240 --> 00:34:49.080 even at epsilon equals .05,

NOTE Confidence: 0.877120092857143

00:34:49.080 --> 00:34:51.240 all the little points that

NOTE Confidence: 0.877120092857143

00:34:51.240 --> 00:34:53.400 are in the inner circle,

NOTE Confidence: 0.877120092857143

00:34:53.400 --> 00:34:55.360 they all get connected together

NOTE Confidence: 0.877120092857143

00:34:55.360 --> 00:34:57.827 because all of those disks are

NOTE Confidence: 0.877120092857143

00:34:57.827 --> 00:34:59.559 overlapping with each other.

NOTE Confidence: 0.877120092857143

00:34:59.560 --> 00:35:02.986 Then when we increase our epsilon to .15,

NOTE Confidence: 0.877120092857143

00:35:02.986 --> 00:35:05.516 the inner circles are still

NOTE Confidence: 0.877120092857143

00:35:05.516 --> 00:35:07.034 all connected together,

NOTE Confidence: 0.877120092857143

00:35:07.040 --> 00:35:10.799 but now the outer circles outer point.

NOTE Confidence: 0.877120092857143

00:35:10.800 --> 00:35:12.744 The points in the outer circle

NOTE Confidence: 0.877120092857143

00:35:12.744 --> 00:35:14.040 are also connected together.

NOTE Confidence: 0.877120092857143

00:35:14.040 --> 00:35:17.760 So we observe 2 loops in our data.

NOTE Confidence: 0.877120092857143

00:35:17.760 --> 00:35:20.640 As epsilon increases even more,

NOTE Confidence: 0.877120092857143

00:35:20.640 --> 00:35:23.909 these loops get closed in and they

NOTE Confidence: 0.877120092857143

00:35:23.909 --> 00:35:27.026 merge with each other until at the

NOTE Confidence: 0.877120092857143

00:35:27.026 --> 00:35:29.318 end when epsilon is really big,

NOTE Confidence: 0.877120092857143

00:35:29.320 --> 00:35:32.434 all of the disks intersect with

NOTE Confidence: 0.877120092857143

00:35:32.434 --> 00:35:35.255 each other and everything collapses

NOTE Confidence: 0.877120092857143

00:35:35.255 --> 00:35:38.515 into just one connected component

NOTE Confidence: 0.877120092857143

00:35:38.520 --> 00:35:41.677 in the two half moons data set.

NOTE Confidence: 0.877120092857143

00:35:41.680 --> 00:35:43.930 What we see is that there is a value

NOTE Confidence: 0.877120092857143

00:35:43.930 --> 00:35:46.371 of epsilon indeed where there is a

NOTE Confidence: 0.877120092857143

00:35:46.371 --> 00:35:48.523 small circle that forms as these

NOTE Confidence: 0.877120092857143

00:35:48.523 --> 00:35:50.193 points all get connected together

NOTE Confidence: 0.877120092857143

00:35:50.193 --> 00:35:52.260 in a pair wise manner.

NOTE Confidence: 0.877120092857143

00:35:52.260 --> 00:35:54.910 But that little circle quickly

NOTE Confidence: 0.877120092857143

00:35:54.910 --> 00:35:57.089 disappears when epsilon increases

NOTE Confidence: 0.877120092857143

00:35:57.089 --> 00:36:00.125 further and these two arcs get

NOTE Confidence: 0.877120092857143

00:36:00.125 --> 00:36:02.440 connected together into one hole.

NOTE Confidence: 0.877120092857143

00:36:02.440 --> 00:36:06.000 And so this is this technique

NOTE Confidence: 0.877120092857143

00:36:06.000 --> 00:36:08.800 is called persistence homology.

NOTE Confidence: 0.877120092857143

00:36:08.800 --> 00:36:11.968 And what it gives us is what we

NOTE Confidence: 0.877120092857143

00:36:11.968 --> 00:36:14.960 call a topological barcode.

NOTE Confidence: 0.877120092857143

00:36:14.960 --> 00:36:17.856 So there is code out there that will

NOTE Confidence: 0.877120092857143

00:36:17.856 --> 00:36:20.517 take these data points as an input,

NOTE Confidence: 0.877120092857143

00:36:20.520 --> 00:36:22.030 doesn't have to be two-dimensional

NOTE Confidence: 0.877120092857143

00:36:22.030 --> 00:36:22.634 or three-dimensional,

NOTE Confidence: 0.877120092857143

00:36:22.640 --> 00:36:24.660 could be high dimensional data

NOTE Confidence: 0.877120092857143

00:36:24.660 --> 00:36:27.509 and it will perform this kind of

NOTE Confidence: 0.877120092857143

00:36:27.509 --> 00:36:29.384 computation and give you back

NOTE Confidence: 0.877120092857143

00:36:29.384 --> 00:36:31.759 a visual that looks like this,

NOTE Confidence: 0.686840408

00:36:31.760 --> 00:36:34.200 which is the topological barcode.

NOTE Confidence: 0.686840408

00:36:34.200 --> 00:36:36.496 So let's kind of go through the

NOTE Confidence: 0.686840408

00:36:36.496 --> 00:36:37.896 topological barcode and learn

NOTE Confidence: 0.686840408

00:36:37.896 --> 00:36:39.636 how to interpret the barcode.

NOTE Confidence: 0.686840408

00:36:39.640 --> 00:36:41.716 The barcode consists of two parts.

NOTE Confidence: 0.686840408

00:36:41.720 --> 00:36:44.726 The top half I'm going to call H sub

NOTE Confidence: 0.686840408

00:36:44.726 --> 00:36:47.749 zero for dimension 0 homology and the

NOTE Confidence: 0.686840408

00:36:47.749 --> 00:36:50.813 lower part I'm going to call edge

NOTE Confidence: 0.686840408

00:36:50.813 --> 00:36:53.315 sub one for dimension 1 homology.

NOTE Confidence: 0.686840408

00:36:53.320 --> 00:36:56.080 And so in dimension 0 homology,

NOTE Confidence: 0.686840408

00:36:56.080 --> 00:36:58.440 we are measuring connectedness of

NOTE Confidence: 0.686840408

00:36:58.440 --> 00:37:01.600 our data and we generally call this

NOTE Confidence: 0.686840408

00:37:01.600 --> 00:37:03.396 number of connected components.

NOTE Confidence: 0.686840408

00:37:03.396 --> 00:37:06.899 And So what you can see here is

NOTE Confidence: 0.686840408

00:37:06.899 --> 00:37:09.440 that when epsilon is close to 0,

NOTE Confidence: 0.686840408

00:37:09.440 --> 00:37:10.772 where my cursor is,

NOTE Confidence: 0.686840408

00:37:10.772 --> 00:37:14.477 we see lots and lots of bars in our data set.

NOTE Confidence: 0.686840408

00:37:14.480 --> 00:37:17.240 And these bars correspond to how

NOTE Confidence: 0.686840408

00:37:17.240 --> 00:37:19.080 many connected components there

NOTE Confidence: 0.686840408

00:37:19.160 --> 00:37:20.600 are in our data set.

NOTE Confidence: 0.686840408

00:37:20.600 --> 00:37:22.480 So when epsilon is 0,

NOTE Confidence: 0.686840408

00:37:22.480 --> 00:37:25.476 all the points are sitting by themselves,

NOTE Confidence: 0.686840408

00:37:25.480 --> 00:37:26.900 none of the points are

NOTE Confidence: 0.686840408

00:37:26.900 --> 00:37:28.036 connected to each other.

NOTE Confidence: 0.686840408

00:37:28.040 --> 00:37:30.768 So we get as many bars as the

NOTE Confidence: 0.686840408

00:37:30.768 --> 00:37:33.198 number of points in our data.

NOTE Confidence: 0.686840408

00:37:33.200 --> 00:37:35.400 As epsilon starts increasing,

NOTE Confidence: 0.686840408

00:37:35.400 --> 00:37:38.150 we start merging together points

NOTE Confidence: 0.686840408

00:37:38.150 --> 00:37:40.749 by connecting them with an

NOTE Confidence: 0.686840408

00:37:40.749 --> 00:37:43.194 edge and forming A1 simplex.

NOTE Confidence: 0.686840408

00:37:43.200 --> 00:37:45.432 So as epsilon is increasing here

NOTE Confidence: 0.686840408

00:37:45.432 --> 00:37:48.054 you can see that the number of

NOTE Confidence: 0.686840408

00:37:48.054 --> 00:37:50.504 bars is fewer and fewer until at

NOTE Confidence: 0.686840408

00:37:50.580 --> 00:37:53.030 high values of epsilon we end up

NOTE Confidence: 0.686840408

00:37:53.030 --> 00:37:56.672 with just one bar in our barcode.

NOTE Confidence: 0.686840408

00:37:56.672 --> 00:37:59.912 So this dimension 0 homology,

NOTE Confidence: 0.686840408

00:37:59.920 --> 00:38:02.326 this is capturing the connectivity of

NOTE Confidence: 0.686840408

00:38:02.326 --> 00:38:05.670 our data and by looking at the slope

NOTE Confidence: 0.686840408

00:38:05.670 --> 00:38:08.999 by which these bars are decreasing in number,

NOTE Confidence: 0.686840408

00:38:09.000 --> 00:38:11.484 we can figure out how connected

NOTE Confidence: 0.686840408

00:38:11.484 --> 00:38:14.079 data our data set really is.

NOTE Confidence: 0.686840408

00:38:14.080 --> 00:38:15.193 In dimension 1,

NOTE Confidence: 0.686840408

00:38:15.193 --> 00:38:18.520 which is at the bottom of this barcode,

NOTE Confidence: 0.686840408

00:38:18.520 --> 00:38:20.608 what we are measuring is the

NOTE Confidence: 0.686840408

00:38:20.608 --> 00:38:23.000 presence of loops in our data set.

NOTE Confidence: 0.686840408

00:38:23.000 --> 00:38:26.440 So at epsilon equal to 0 on the very left,

NOTE Confidence: 0.686840408

00:38:26.440 --> 00:38:28.477 we have no bars in the lower

NOTE Confidence: 0.686840408

00:38:28.477 --> 00:38:30.000 part of this diagram,

NOTE Confidence: 0.686840408

00:38:30.000 --> 00:38:32.472 which means there are no loops

NOTE Confidence: 0.686840408

00:38:32.472 --> 00:38:35.079 present at that value of epsilon.

NOTE Confidence: 0.686840408

00:38:35.080 --> 00:38:37.642 At a later value of epsilon we

NOTE Confidence: 0.686840408

00:38:37.642 --> 00:38:39.897 see the occurrence of this 1st

NOTE Confidence: 0.686840408

00:38:39.897 --> 00:38:41.692 loop from these orange points

NOTE Confidence: 0.686840408

00:38:41.692 --> 00:38:44.279 in the inner concentric circle.

NOTE Confidence: 0.686840408

00:38:44.280 --> 00:38:48.600 That loop persists for a long period of time.

NOTE Confidence: 0.686840408

00:38:48.600 --> 00:38:51.280 What I mean by time is it persists

NOTE Confidence: 0.686840408

00:38:51.280 --> 00:38:54.557 for a large range of epsilon values.

NOTE Confidence: 0.686840408

00:38:54.560 --> 00:38:55.868 During this process,

NOTE Confidence: 0.686840408

00:38:55.868 --> 00:38:58.920 there is a second loop that forms,

NOTE Confidence: 0.686840408

00:38:58.920 --> 00:39:00.918 indicated by the second red bar.

NOTE Confidence: 0.686840408

00:39:00.920 --> 00:39:04.420 Here it emerges at at a higher

NOTE Confidence: 0.686840408

00:39:04.420 --> 00:39:05.920 value of epsilon,

NOTE Confidence: 0.686840408

00:39:05.920 --> 00:39:08.368 and this outer loop dies sooner

NOTE Confidence: 0.686840408

00:39:08.368 --> 00:39:10.480 than the inner loop does.

NOTE Confidence: 0.686840408

00:39:10.480 --> 00:39:13.154 The inner loop persists for even longer.

NOTE Confidence: 0.686840408

00:39:13.160 --> 00:39:17.320 So by looking at the bars in our bar code,

NOTE Confidence: 0.686840408

00:39:17.320 --> 00:39:20.400 we can learn that our data has so

NOTE Confidence: 0.686840408

00:39:20.400 --> 00:39:23.108 many points simply by counting the

NOTE Confidence: 0.686840408

00:39:23.108 --> 00:39:26.400 number of bars at epsilon equal to 0.

NOTE Confidence: 0.686840408

00:39:26.400 --> 00:39:28.871 We can learn how connected our data

NOTE Confidence: 0.686840408

00:39:28.871 --> 00:39:32.489 set is by looking at how these bars

NOTE Confidence: 0.686840408

00:39:32.489 --> 00:39:34.473 disappear as epsilon increases.

NOTE Confidence: 0.686840408

00:39:34.480 --> 00:39:37.200 And then in the lower part of the bar code,

NOTE Confidence: 0.686840408

00:39:37.200 --> 00:39:39.720 by looking at these bars,

NOTE Confidence: 0.686840408

00:39:39.720 --> 00:39:41.562 we can learn how many loops

NOTE Confidence: 0.686840408

00:39:41.562 --> 00:39:43.200 are present in our data.

NOTE Confidence: 0.686840408

00:39:43.200 --> 00:39:44.154 In particular,

NOTE Confidence: 0.686840408

00:39:44.154 --> 00:39:47.493 the bars that are longer in length

NOTE Confidence: 0.686840408

00:39:47.493 --> 00:39:49.450 actually represent actual loops

NOTE Confidence: 0.686840408

00:39:49.450 --> 00:39:51.400 that are present in our data.

NOTE Confidence: 0.686840408

00:39:51.400 --> 00:39:53.780 There are indeed some smaller

NOTE Confidence: 0.686840408

00:39:53.780 --> 00:39:56.160 bars which are small noisy

NOTE Confidence: 0.925586017777778

00:39:56.251 --> 00:40:00.075 loops that form as we perform this procedure.

NOTE Confidence: 0.925586017777778

00:40:00.080 --> 00:40:01.940 And what's apparent from these

NOTE Confidence: 0.925586017777778

00:40:01.940 --> 00:40:04.774 two barcodes is that in our first

NOTE Confidence: 0.925586017777778

00:40:04.774 --> 00:40:07.034 example with the concentric circles,

NOTE Confidence: 0.925586017777778

00:40:07.040 --> 00:40:10.600 there are two clear loops in that data.

NOTE Confidence: 0.925586017777778

00:40:10.600 --> 00:40:13.000 And in our second example, there is

NOTE Confidence: 0.925586017777778

00:40:13.000 --> 00:40:16.080 indeed a small loop that emerges here,

NOTE Confidence: 0.925586017777778

00:40:16.080 --> 00:40:18.140 but it quickly disappears,

NOTE Confidence: 0.925586017777778

00:40:18.140 --> 00:40:20.715 so there's no really topologically

NOTE Confidence: 0.925586017777778

00:40:20.715 --> 00:40:22.053 significant loops present

NOTE Confidence: 0.925586017777778

00:40:22.053 --> 00:40:23.918 in the second data set.

NOTE Confidence: 0.925586017777778

00:40:23.920 --> 00:40:26.362 And so these bar codes capture

NOTE Confidence: 0.925586017777778

00:40:26.362 --> 00:40:28.520 the shape of our data.

NOTE Confidence: 0.925586017777778

00:40:28.520 --> 00:40:33.134 You can continue to plot H2 and H3

NOTE Confidence: 0.925586017777778

00:40:33.134 --> 00:40:35.402 which are going to capture higher

NOTE Confidence: 0.925586017777778

00:40:35.402 --> 00:40:37.157 dimensional holes in your data.

NOTE Confidence: 0.925586017777778

00:40:37.160 --> 00:40:40.528 So H2 is going to capture 3 dimensional

NOTE Confidence: 0.925586017777778

00:40:40.528 --> 00:40:43.040 holes or voids in the data.

NOTE Confidence: 0.925586017777778

00:40:43.040 --> 00:40:44.580 H4 will capture even higher

NOTE Confidence: 0.925586017777778

00:40:44.580 --> 00:40:46.120 dimensional holes in the data.

NOTE Confidence: 0.925586017777778

00:40:46.120 --> 00:40:48.598 So topology captures the shape of our

NOTE Confidence: 0.925586017777778

00:40:48.598 --> 00:40:50.473 data by measuring connectedness and

NOTE Confidence: 0.925586017777778

00:40:50.473 --> 00:40:53.077 the presence of loops in the data.

NOTE Confidence: 0.925586017777778

00:40:53.080 --> 00:40:54.040 Are there any questions?

NOTE Confidence: 0.803696451818182

00:40:54.600 --> 00:40:56.917 Yes, Jay, just to kind of translate

NOTE Confidence: 0.803696451818182

00:40:56.917 --> 00:40:58.640 math into more intuition.

NOTE Confidence: 0.803696451818182

00:40:58.640 --> 00:41:01.520 When you say you have holes or loops,

NOTE Confidence: 0.803696451818182

00:41:01.520 --> 00:41:03.624 you're pretty much talking

NOTE Confidence: 0.803696451818182

00:41:03.624 --> 00:41:05.702 about some impossible states,

NOTE Confidence: 0.803696451818182

00:41:05.702 --> 00:41:09.279 meaning that your state cannot have this,

NOTE Confidence: 0.803696451818182

00:41:09.280 --> 00:41:11.440 like cannot be in the specific

NOTE Confidence: 0.803696451818182

00:41:11.440 --> 00:41:13.120 state for whatever reasons, right?

NOTE Confidence: 0.83345938

00:41:13.120 --> 00:41:14.600 Yeah, that's a great question.

NOTE Confidence: 0.83345938

00:41:14.600 --> 00:41:16.910 So I'm talking indeed about

NOTE Confidence: 0.83345938

00:41:16.910 --> 00:41:18.758 impossible states because these

NOTE Confidence: 0.83345938

00:41:18.758 --> 00:41:20.974 points are derived from experiments

NOTE Confidence: 0.83345938

00:41:20.974 --> 00:41:23.892 and they represent the state of our

NOTE Confidence: 0.83345938

00:41:23.892 --> 00:41:26.160 brain or the state of our tissue.

NOTE Confidence: 0.83345938

00:41:26.160 --> 00:41:28.648 And therefore if we have a hole in

NOTE Confidence: 0.83345938

00:41:28.648 --> 00:41:30.808 our data set, that means there's no

NOTE Confidence: 0.83345938

00:41:30.808 --> 00:41:32.560 data points present in the middle.

NOTE Confidence: 0.83345938

00:41:32.560 --> 00:41:35.038 And that means that there is that

NOTE Confidence: 0.83345938

00:41:35.038 --> 00:41:37.802 state is impossible as far as we can

NOTE Confidence: 0.83345938

00:41:37.802 --> 00:41:39.437 tell from our experimental data.

NOTE Confidence: 0.83345938

00:41:39.440 --> 00:41:41.524 So that's first conclusion.

NOTE Confidence: 0.83345938

00:41:41.524 --> 00:41:45.799 The 2nd conclusion which we can get is this.

NOTE Confidence: 0.83345938

00:41:45.800 --> 00:41:50.200 H1 measures kind of holes in two dimensions.

NOTE Confidence: 0.83345938

00:41:50.200 --> 00:41:53.650 And so that necessarily means that there

NOTE Confidence: 0.83345938

00:41:53.650 --> 00:41:56.059 is data that surrounds the hole, right?

NOTE Confidence: 0.83345938

00:41:56.059 --> 00:41:58.432 There must be some surrounding data and

NOTE Confidence: 0.83345938

00:41:58.432 --> 00:42:00.838 whenever there is data that surrounds a hole,

NOTE Confidence: 0.83345938

00:42:00.840 --> 00:42:02.985 that might indicate some kind

NOTE Confidence: 0.83345938

00:42:02.985 --> 00:42:05.520 of periodicity in the data set.

NOTE Confidence: 0.83345938

00:42:05.520 --> 00:42:08.166 So you can imagine that if you have data

NOTE Confidence: 0.83345938

00:42:08.166 --> 00:42:10.395 points that are arranged in a circle,

NOTE Confidence: 0.83345938

00:42:10.400 --> 00:42:12.377 doesn't have to be a perfect circle, it

NOTE Confidence: 0.83345938

00:42:12.377 --> 00:42:15.073 could be like an elliptical or skewed circle.

NOTE Confidence: 0.83345938

00:42:15.080 --> 00:42:16.456 This technique still works.

NOTE Confidence: 0.83345938

00:42:16.456 --> 00:42:18.520 But that that tells you that

NOTE Confidence: 0.8178842675

00:42:18.600 --> 00:42:20.720 there is there is some sort of process.

NOTE Confidence: 0.842100112142857

00:42:21.080 --> 00:42:23.102 Yeah, there's a process that goes

NOTE Confidence: 0.842100112142857

00:42:23.102 --> 00:42:26.000 around in in a kind of periodic way.

NOTE Confidence: 0.842100112142857

00:42:26.000 --> 00:42:28.394 So you can navigate those that state

NOTE Confidence: 0.842100112142857

00:42:28.394 --> 00:42:31.333 space in a way that's periodic or

NOTE Confidence: 0.842100112142857

00:42:31.333 --> 00:42:33.633 almost periodic or quasi periodic.

NOTE Confidence: 0.842100112142857

00:42:33.640 --> 00:42:34.966 So impossible state.

NOTE Confidence: 0.842100112142857

00:42:34.966 --> 00:42:37.618 Well, as periodic states are being

NOTE Confidence: 0.842100112142857

00:42:37.618 --> 00:42:39.826 captured through dimension 1 homology

NOTE Confidence: 0.842100112142857

00:42:39.826 --> 00:42:42.358 in the status in this technique,

NOTE Confidence: 0.842100112142857

00:42:42.360 --> 00:42:44.160 indeed, I

NOTE Confidence: 0.961629856666667

00:42:44.160 --> 00:42:45.280 also have one question.

NOTE Confidence: 0.961629856666667

00:42:45.280 --> 00:42:48.420 So when we are increasing epsilon,

NOTE Confidence: 0.961629856666667

00:42:48.420 --> 00:42:52.018 yeah, are there some loops

NOTE Confidence: 0.961629856666667

00:42:52.018 --> 00:42:53.716 that are disappearing?

NOTE Confidence: 0.961629856666667

00:42:53.720 --> 00:42:56.080 Because if we increase epsilon,

NOTE Confidence: 0.961629856666667

00:42:56.080 --> 00:42:57.932 loops should not discover

NOTE Confidence: 0.961629856666667

00:42:57.932 --> 00:42:59.760 disappear, right loops

NOTE Confidence: 0.937421992333333

00:42:59.760 --> 00:43:01.944 can disappear. So the the way

NOTE Confidence: 0.937421992333333

00:43:01.944 --> 00:43:03.880 this outer loop is disappearing

NOTE Confidence: 0.937421992333333

00:43:03.880 --> 00:43:07.096 here is when there is a value of

NOTE Confidence: 0.937421992333333

00:43:07.096 --> 00:43:09.939 epsilon when one of the disks from

NOTE Confidence: 0.937421992333333

00:43:09.939 --> 00:43:11.817 the outer loop intersects with

NOTE Confidence: 0.937421992333333

00:43:11.817 --> 00:43:13.599 the disk from the inner loop.

NOTE Confidence: 0.937421992333333

00:43:13.600 --> 00:43:16.150 As soon as these two discs

NOTE Confidence: 0.937421992333333

00:43:16.150 --> 00:43:17.000 start intersecting,

NOTE Confidence: 0.937421992333333

00:43:17.000 --> 00:43:18.953 we draw an edge that goes from

NOTE Confidence: 0.937421992333333

00:43:18.953 --> 00:43:21.329 a point in the inner loop to a

NOTE Confidence: 0.937421992333333

00:43:21.329 --> 00:43:23.722 point on the outer loop and that

NOTE Confidence: 0.937421992333333

00:43:23.722 --> 00:43:25.867 effectively connects that those two

NOTE Confidence: 0.937421992333333

00:43:25.867 --> 00:43:28.255 loops together and the the enclosing

NOTE Confidence: 0.937421992333333

00:43:28.255 --> 00:43:30.810 space between the inner loop and the

NOTE Confidence: 0.937421992333333

00:43:30.883 --> 00:43:33.319 outer loop disappears at that stage.

NOTE Confidence: 0.937421992333333

00:43:33.320 --> 00:43:35.939 To get rid of the inner loop you have

NOTE Confidence: 0.937421992333333

00:43:35.939 --> 00:43:38.373 to increase epsilon a lot higher

NOTE Confidence: 0.937421992333333

00:43:38.373 --> 00:43:41.248 because you have two points that are

NOTE Confidence: 0.937421992333333

00:43:41.248 --> 00:43:44.152 opposite to each other in the inner loop,

NOTE Confidence: 0.937421992333333

00:43:44.160 --> 00:43:45.500 and when those two points

NOTE Confidence: 0.937421992333333

00:43:45.500 --> 00:43:46.840 get connected to each other,

NOTE Confidence: 0.937421992333333

00:43:46.840 --> 00:43:50.438 the inner loop closes in and disappears.

NOTE Confidence: 0.937421992333333

00:43:50.440 --> 00:43:52.920 So yes, loops can disappear,

NOTE Confidence: 0.937421992333333

00:43:52.920 --> 00:43:55.512 and in fact at a high enough value

NOTE Confidence: 0.937421992333333

00:43:55.512 --> 00:43:58.640 of epsilon, all loops will close up,

NOTE Confidence: 0.937421992333333

00:43:58.640 --> 00:44:00.760 and when epsilon is Infinity,

NOTE Confidence: 0.937421992333333

00:44:00.760 --> 00:44:02.775 all the points are necessarily

NOTE Confidence: 0.937421992333333

00:44:02.775 --> 00:44:05.173 intersecting with each other and there

NOTE Confidence: 0.937421992333333

00:44:05.173 --> 00:44:07.357 are no loops present in the data.

NOTE Confidence: 0.937421992333333

00:44:07.360 --> 00:44:09.488 Maybe this is a little bit more apparent

NOTE Confidence: 0.937421992333333

00:44:09.488 --> 00:44:11.714 in this previous animation when I was

NOTE Confidence: 0.937421992333333

00:44:11.714 --> 00:44:13.663 drawing the one simplex where you

NOTE Confidence: 0.937421992333333

00:44:13.663 --> 00:44:15.958 can see that at a higher value of epsilon,

NOTE Confidence: 0.937421992333333

00:44:15.960 --> 00:44:18.544 we do indeed end up closing the outer

NOTE Confidence: 0.937421992333333

00:44:18.544 --> 00:44:21.555 loop by connecting it to the inner loop.

NOTE Confidence: 0.937421992333333

00:44:21.560 --> 00:44:23.621 So you see that there's a bridge that forms

NOTE Confidence: 0.937421992333333

00:44:23.621 --> 00:44:25.477 from the outer loop to the inner loop,

NOTE Confidence: 0.937421992333333

00:44:25.480 --> 00:44:27.601 and that empty space here closes in

NOTE Confidence: 0.937421992333333

00:44:27.601 --> 00:44:30.440 and at a very high value of epsilon,

NOTE Confidence: 0.937421992333333

00:44:30.440 --> 00:44:32.320 there will be edges that

NOTE Confidence: 0.937421992333333

00:44:32.320 --> 00:44:34.200 go across the inner loop,

NOTE Confidence: 0.937421992333333

00:44:34.200 --> 00:44:36.320 closing the inner loop entirely.

NOTE Confidence: 0.937421992333333

00:44:36.320 --> 00:44:37.890 Although that doesn't happen in

NOTE Confidence: 0.937421992333333

00:44:37.890 --> 00:44:39.460 this animation because I didn't

NOTE Confidence: 0.937421992333333

00:44:39.513 --> 00:44:40.957 increase epsilon high enough.

NOTE Confidence: 0.63813168

00:44:43.040 --> 00:44:44.320 I suppose the question

NOTE Confidence: 0.847241107

00:44:44.840 --> 00:44:47.520 there is a, there is a question in the chat.

NOTE Confidence: 0.847241107

00:44:47.520 --> 00:44:49.000 And Zachary, would you like

NOTE Confidence: 0.847241107

00:44:49.000 --> 00:44:50.776 to ask your question live or

NOTE Confidence: 0.847241107

00:44:50.776 --> 00:44:52.435 would you like me to read that?

NOTE Confidence: 0.939641918333333

00:44:53.120 --> 00:44:54.278 I think I can read it.

NOTE Confidence: 0.939641918333333

00:44:54.280 --> 00:44:56.686 So the question says earlier you

NOTE Confidence: 0.939641918333333

00:44:56.686 --> 00:44:58.748 mentioned sufficient sampling of points

NOTE Confidence: 0.939641918333333

00:44:58.748 --> 00:45:01.076 as necessary to interpret a manifold.

NOTE Confidence: 0.939641918333333

00:45:01.080 --> 00:45:02.592 What you're showing now seems to

NOTE Confidence: 0.939641918333333

00:45:02.592 --> 00:45:04.040 address properties of the manifold.

NOTE Confidence: 0.939641918333333

00:45:04.040 --> 00:45:06.944 Is there a property to address how much

NOTE Confidence: 0.939641918333333

00:45:06.944 --> 00:45:10.173 variability is or isn't accounted for by

NOTE Confidence: 0.939641918333333

00:45:10.173 --> 00:45:12.117 this manifold characterization process,

NOTE Confidence: 0.939641918333333

00:45:12.120 --> 00:45:14.240 possibly due to under sampling?

NOTE Confidence: 0.939641918333333

00:45:14.240 --> 00:45:16.352 That's a great question.

NOTE Confidence: 0.939641918333333

00:45:16.352 --> 00:45:19.220 So what I'm presenting at the

NOTE Confidence: 0.939641918333333

00:45:19.220 --> 00:45:21.600 moment is under the assumption

NOTE Confidence: 0.939641918333333

00:45:21.600 --> 00:45:24.480 that our manifold is well sampled.

NOTE Confidence: 0.939641918333333

00:45:24.480 --> 00:45:27.726 So indeed, if your experimental

NOTE Confidence: 0.939641918333333

00:45:27.726 --> 00:45:29.736 procedure failed to sample a

NOTE Confidence: 0.939641918333333

00:45:29.736 --> 00:45:31.879 data point in the middle here,

NOTE Confidence: 0.939641918333333

00:45:31.880 --> 00:45:34.142 my conclusion would be that your

NOTE Confidence: 0.939641918333333

00:45:34.142 --> 00:45:36.400 data set has cellular states

NOTE Confidence: 0.939641918333333

00:45:36.400 --> 00:45:38.640 organized into two circles,

NOTE Confidence: 0.939641918333333

00:45:38.640 --> 00:45:39.940 and they're kind of

NOTE Confidence: 0.939641918333333

00:45:39.940 --> 00:45:41.240 independent of each other.

NOTE Confidence: 0.939641918333333

00:45:41.240 --> 00:45:43.160 There is indeed an outer circle,

NOTE Confidence: 0.939641918333333

00:45:43.160 --> 00:45:45.386 and that might be completely wrong

NOTE Confidence: 0.939641918333333

00:45:45.386 --> 00:45:47.752 simply because we never sampled data

NOTE Confidence: 0.939641918333333

00:45:47.752 --> 00:45:50.200 that exists between these two circles.

NOTE Confidence: 0.939641918333333

00:45:50.200 --> 00:45:53.280 So I am indeed operating under the

NOTE Confidence: 0.939641918333333

00:45:53.280 --> 00:45:56.718 assumption that the manifold is well sampled.

NOTE Confidence: 0.939641918333333

00:45:56.720 --> 00:45:59.667 Another aspect of the question was what

NOTE Confidence: 0.939641918333333

00:45:59.667 --> 00:46:02.929 what kind of like what properties of

NOTE Confidence: 0.939641918333333

00:46:02.929 --> 00:46:05.797 the manifold am I really capturing?

NOTE Confidence: 0.939641918333333

00:46:05.800 --> 00:46:08.065 And so I'm capturing topological

NOTE Confidence: 0.939641918333333

00:46:08.065 --> 00:46:09.877 properties of the manifold.

NOTE Confidence: 0.939641918333333

00:46:09.880 --> 00:46:12.000 And by topological properties I

NOTE Confidence: 0.939641918333333

00:46:12.000 --> 00:46:14.120 mean how connected that manifold

NOTE Confidence: 0.939641918333333

00:46:14.192 --> 00:46:16.278 is and whether or not there are

NOTE Confidence: 0.939641918333333

00:46:16.278 --> 00:46:17.840 holes in that manifold.

NOTE Confidence: 0.939641918333333

00:46:17.840 --> 00:46:20.423 And so this technique is like invariant

NOTE Confidence: 0.939641918333333

00:46:20.423 --> 00:46:23.239 to things like translation of the data.

NOTE Confidence: 0.939641918333333

00:46:23.240 --> 00:46:25.452 So if I take these circles and

NOTE Confidence: 0.939641918333333

00:46:25.452 --> 00:46:27.480 I translate them somewhere else,

NOTE Confidence: 0.939641918333333

00:46:27.480 --> 00:46:28.720 that doesn't impact it.

NOTE Confidence: 0.939641918333333

00:46:28.720 --> 00:46:30.867 If I take this diagram and I

NOTE Confidence: 0.939641918333333

00:46:30.867 --> 00:46:32.651 rotate it by 45° or 30°,

NOTE Confidence: 0.939641918333333

00:46:32.651 --> 00:46:34.536 that's not going to change

NOTE Confidence: 0.939641918333333

00:46:34.536 --> 00:46:36.360 the bar code at all.

NOTE Confidence: 0.939641918333333

00:46:36.360 --> 00:46:38.484 So it's translation invariant,

NOTE Confidence: 0.939641918333333

00:46:38.484 --> 00:46:40.077 it's rotationally invariant.

NOTE Confidence: 0.939641918333333

00:46:40.080 --> 00:46:42.906 And it's also invariant to certain

NOTE Confidence: 0.939641918333333

00:46:42.906 --> 00:46:45.689 kinds of deformations where if I take

NOTE Confidence: 0.939641918333333

00:46:45.689 --> 00:46:48.520 this arc and I deform it a little bit,

NOTE Confidence: 0.939641918333333

00:46:48.520 --> 00:46:50.977 that's not going to change my barcode

NOTE Confidence: 0.939641918333333

00:46:50.977 --> 00:46:53.534 and it's not going to change the

NOTE Confidence: 0.939641918333333

00:46:53.534 --> 00:46:55.682 fact that the data is connected

NOTE Confidence: 0.939641918333333

00:46:55.754 --> 00:46:57.719 in this Crescent moon shake.

NOTE Confidence: 0.939641918333333

00:46:57.720 --> 00:47:00.163 We'll get into more of the details

NOTE Confidence: 0.939641918333333

00:47:00.163 --> 00:47:02.772 of what aspects of the manifold we

NOTE Confidence: 0.939641918333333

00:47:02.772 --> 00:47:04.424 are capturing as we progress further.

NOTE Confidence: 0.939641918333333

00:47:04.424 --> 00:47:06.735 But I hope that kind of goes a little

NOTE Confidence: 0.939641918333333

00:47:06.735 --> 00:47:08.240 way towards answering your question.

NOTE Confidence: 0.701104385454546

00:47:10.760 --> 00:47:13.640 OK, so we have introduced this

NOTE Confidence: 0.701104385454546

00:47:13.640 --> 00:47:16.400 idea of a topological barcode.

NOTE Confidence: 0.701104385454546

00:47:16.400 --> 00:47:19.775 Next I want to show you a more convenient

NOTE Confidence: 0.701104385454546

00:47:19.775 --> 00:47:22.479 way of representing this barcode

NOTE Confidence: 0.701104385454546

00:47:22.479 --> 00:47:25.833 which is called a persistence diagram.

NOTE Confidence: 0.701104385454546

00:47:25.840 --> 00:47:27.640 And a persistence diagram is

NOTE Confidence: 0.701104385454546

00:47:27.640 --> 00:47:29.440 is very easy to construct.

NOTE Confidence: 0.701104385454546

00:47:29.440 --> 00:47:32.920 What you do is you draw two axes,

NOTE Confidence: 0.701104385454546

00:47:32.920 --> 00:47:35.034 the the X axis is called the

NOTE Confidence: 0.701104385454546

00:47:35.034 --> 00:47:37.442 birth axis and the Y axis is

NOTE Confidence: 0.701104385454546

00:47:37.442 --> 00:47:39.232 called the death axis generally.

NOTE Confidence: 0.701104385454546

00:47:39.240 --> 00:47:42.509 And these axes are going to represent

NOTE Confidence: 0.701104385454546

00:47:42.509 --> 00:47:45.279 when points start in the barcode.

NOTE Confidence: 0.701104385454546

00:47:45.280 --> 00:47:47.534 When, when do the bar start in

NOTE Confidence: 0.701104385454546

00:47:47.534 --> 00:47:49.837 the barcode and where do they end?

NOTE Confidence: 0.701104385454546

00:47:49.840 --> 00:47:53.360 And so you cannot complete or you cannot

NOTE Confidence: 0.701104385454546

00:47:53.360 --> 00:47:56.917 end a barcode before starting it.

NOTE Confidence: 0.701104385454546

00:47:56.920 --> 00:47:58.472 And because of that,

NOTE Confidence: 0.701104385454546

00:47:58.472 --> 00:48:01.293 all the points in this persistence diagram

NOTE Confidence: 0.701104385454546

00:48:01.293 --> 00:48:04.198 are going to happen above the diagonal.

NOTE Confidence: 0.701104385454546

00:48:04.200 --> 00:48:06.118 And there's two kinds of points here.

NOTE Confidence: 0.701104385454546

00:48:06.120 --> 00:48:07.891 I hope you can see that there's

NOTE Confidence: 0.701104385454546

00:48:07.891 --> 00:48:09.629 points that are represented as tiny

NOTE Confidence: 0.701104385454546

00:48:09.629 --> 00:48:11.471 circles and there are points that

NOTE Confidence: 0.701104385454546

00:48:11.471 --> 00:48:13.400 are represented as tiny diamonds,

NOTE Confidence: 0.701104385454546

00:48:13.400 --> 00:48:15.836 and so the circles are coming out

NOTE Confidence: 0.701104385454546

00:48:15.836 --> 00:48:18.878 of the H 0 dimension 0 homology,

NOTE Confidence: 0.701104385454546

00:48:18.880 --> 00:48:21.280 representing the connectedness of the data.

NOTE Confidence: 0.701104385454546

00:48:21.280 --> 00:48:23.576 They are all present here on the

NOTE Confidence: 0.701104385454546

00:48:23.576 --> 00:48:25.823 left side of the persistence diagram

NOTE Confidence: 0.701104385454546

00:48:25.823 --> 00:48:29.018 because all of these bars in H zero

NOTE Confidence: 0.701104385454546

00:48:29.018 --> 00:48:31.882 start at epsilon zero and then they end

NOTE Confidence: 0.701104385454546

00:48:31.882 --> 00:48:34.400 with some positive value of epsilon.

NOTE Confidence: 0.701104385454546

00:48:34.400 --> 00:48:34.888 Therefore,

NOTE Confidence: 0.701104385454546

00:48:34.888 --> 00:48:37.816 all of these points here represent

NOTE Confidence: 0.701104385454546

00:48:37.816 --> 00:48:41.092 H 0 and the points that are over

NOTE Confidence: 0.701104385454546

00:48:41.092 --> 00:48:43.820 here further away from zero,

NOTE Confidence: 0.701104385454546

00:48:43.820 --> 00:48:46.518 these are all representing H1.

NOTE Confidence: 0.701104385454546

00:48:46.518 --> 00:48:48.666 So I'm simply taking the starting

NOTE Confidence: 0.701104385454546

00:48:48.666 --> 00:48:50.808 coordinate of the bar and the

NOTE Confidence: 0.701104385454546

00:48:50.808 --> 00:48:52.473 ending coordinate of the bar,

NOTE Confidence: 0.701104385454546

00:48:52.480 --> 00:48:54.390 and I'm just representing it

NOTE Confidence: 0.701104385454546

00:48:54.390 --> 00:48:55.918 along these two axes.

NOTE Confidence: 0.701104385454546

00:48:55.920 --> 00:48:58.170 And this is what's called

NOTE Confidence: 0.701104385454546

00:48:58.170 --> 00:48:59.520 a persistence diagram.

NOTE Confidence: 0.701104385454546

00:48:59.520 --> 00:49:03.160 The conventional wisdom in persistence

NOTE Confidence: 0.701104385454546

00:49:03.160 --> 00:49:05.364 homology and topological data

NOTE Confidence: 0.701104385454546

00:49:05.364 --> 00:49:08.670 analysis is that points that are

NOTE Confidence: 0.701104385454546

00:49:08.763 --> 00:49:11.518 further away from the diagonal,

NOTE Confidence: 0.701104385454546

00:49:11.520 --> 00:49:13.895 they correspond to longer bars

NOTE Confidence: 0.701104385454546

00:49:13.895 --> 00:49:15.795 in the bar code,

NOTE Confidence: 0.701104385454546

00:49:15.800 --> 00:49:18.872 and those are the topologically more

NOTE Confidence: 0.701104385454546

00:49:18.872 --> 00:49:21.400 significant features in our data.

NOTE Confidence: 0.701104385454546

00:49:21.400 --> 00:49:22.660 So in this case,

NOTE Confidence: 0.701104385454546

00:49:22.660 --> 00:49:24.235 for the concentric circles example,

NOTE Confidence: 0.701104385454546

00:49:24.240 --> 00:49:26.424 we see two red diamonds that

NOTE Confidence: 0.701104385454546

00:49:26.424 --> 00:49:28.999 are far away from the diagonal,

NOTE Confidence: 0.701104385454546

00:49:29.000 --> 00:49:32.354 indicating the presence of two loops

NOTE Confidence: 0.701104385454546

00:49:32.354 --> 00:49:36.038 or two circles in our data set.

NOTE Confidence: 0.701104385454546

00:49:36.040 --> 00:49:37.564 This is the first,

NOTE Confidence: 0.701104385454546

00:49:37.564 --> 00:49:39.469 the topological barcode for the

NOTE Confidence: 0.701104385454546

00:49:39.469 --> 00:49:40.959 Half Moons data set,

NOTE Confidence: 0.701104385454546

00:49:40.960 --> 00:49:43.160 and this is the corresponding

NOTE Confidence: 0.701104385454546

00:49:43.160 --> 00:49:44.040 persistence diagram.

NOTE Confidence: 0.701104385454546

00:49:44.040 --> 00:49:46.280 You'll notice that this highlighted

NOTE Confidence: 0.701104385454546

00:49:46.280 --> 00:49:48.520 red diamond corresponding to that

NOTE Confidence: 0.701104385454546

00:49:48.592 --> 00:49:50.867 tiny loop that emerged for a bit

NOTE Confidence: 0.701104385454546

00:49:50.867 --> 00:49:53.127 is actually quite close to the

NOTE Confidence: 0.701104385454546

00:49:53.127 --> 00:49:55.157 diagonal in this persistence diagram,

NOTE Confidence: 0.701104385454546

00:49:55.160 --> 00:49:58.478 indicating that it's not very significant.

NOTE Confidence: 0.701104385454546

00:49:58.480 --> 00:49:58.854 Now,

NOTE Confidence: 0.701104385454546

00:49:58.854 --> 00:50:01.098 there are ways to compute statistics

NOTE Confidence: 0.701104385454546

00:50:01.098 --> 00:50:03.920 here and to figure out more quantity

NOTE Confidence: 0.701104385454546

00:50:03.920 --> 00:50:06.680 quantitatively whether or not a given

NOTE Confidence: 0.701104385454546

00:50:06.680 --> 00:50:08.640 topological feature is significant,

NOTE Confidence: 0.701104385454546

00:50:08.640 --> 00:50:10.872 but I'm not going to get into that today.

NOTE Confidence: 0.701104385454546

00:50:10.880 --> 00:50:12.924 There are bootstrapping methods

NOTE Confidence: 0.701104385454546

00:50:12.924 --> 00:50:15.479 that give you a confidence

NOTE Confidence: 0.701104385454546

00:50:15.479 --> 00:50:17.600 interval around the diagonal,

NOTE Confidence: 0.701104385454546

00:50:17.600 --> 00:50:20.498 and anything that falls inside of

NOTE Confidence: 0.701104385454546

00:50:20.498 --> 00:50:23.063 that confidence interval is going

NOTE Confidence: 0.701104385454546

00:50:23.063 --> 00:50:25.279 to be insignificant features.

NOTE Confidence: 0.701104385454546

00:50:25.280 --> 00:50:26.925 And anything that falls outside

NOTE Confidence: 0.701104385454546

00:50:26.925 --> 00:50:28.570 of that confidence interval is

NOTE Confidence: 0.701104385454546

00:50:28.631 --> 00:50:30.276 further away from the diagonal,

NOTE Confidence: 0.701104385454546

00:50:30.280 --> 00:50:32.740 and it's going to be topologically

NOTE Confidence: 0.701104385454546

00:50:32.740 --> 00:50:33.560 significant features.

NOTE Confidence: 0.701104385454546

00:50:33.560 --> 00:50:36.276 So those those kinds of tools exist,

NOTE Confidence: 0.701104385454546

00:50:36.280 --> 00:50:38.478 but I'm not getting into that today,

NOTE Confidence: 0.701104385454546

00:50:38.480 --> 00:50:41.800 just to build intuition.

NOTE Confidence: 0.826091706842105

00:50:41.800 --> 00:50:44.491 OK, so here's a small quiz that I like

NOTE Confidence: 0.826091706842105

00:50:44.491 --> 00:50:48.183 to do when I maybe present this also to

NOTE Confidence: 0.826091706842105

00:50:48.183 --> 00:50:50.500 undergraduates where I have three data

NOTE Confidence: 0.826091706842105

00:50:50.500 --> 00:50:52.400 sets and three persistence diagrams,

NOTE Confidence: 0.826091706842105

00:50:52.400 --> 00:50:54.560 but I've kind of jumbled them all up.

NOTE Confidence: 0.826091706842105

00:50:54.560 --> 00:50:56.960 And so let me just quickly tell you what the

NOTE Confidence: 0.826091706842105

00:50:57.016 --> 00:50:59.200 data sets are and what the diagrams are.

NOTE Confidence: 0.826091706842105

00:50:59.200 --> 00:51:00.635 I'll give you a moment to think,

NOTE Confidence: 0.826091706842105

00:51:00.640 --> 00:51:01.700 I'll drink some water,

NOTE Confidence: 0.826091706842105

00:51:01.700 --> 00:51:03.840 and then we'll go over the solution.

NOTE Confidence: 0.826091706842105

00:51:03.840 --> 00:51:06.636 So the first data set here,

NOTE Confidence: 0.826091706842105

00:51:06.640 --> 00:51:08.200 it's very hard to tell,

NOTE Confidence: 0.826091706842105

00:51:08.200 --> 00:51:11.496 but these are two spheres that where I

NOTE Confidence: 0.826091706842105

00:51:11.496 --> 00:51:14.983 have data sample from an inner sphere and

NOTE Confidence: 0.826091706842105

00:51:14.983 --> 00:51:18.678 a data sample from the outer sphere here.

NOTE Confidence: 0.826091706842105

00:51:18.680 --> 00:51:20.680 So it's in three dimensions.

NOTE Confidence: 0.826091706842105

00:51:20.680 --> 00:51:23.249 The second example are three circles where

NOTE Confidence: 0.826091706842105

00:51:23.249 --> 00:51:26.054 I have two circles that are concentric

NOTE Confidence: 0.826091706842105

00:51:26.054 --> 00:51:29.005 with one another and a separate circle

NOTE Confidence: 0.826091706842105

00:51:29.005 --> 00:51:31.555 that's outside of these two circles.

NOTE Confidence: 0.826091706842105

00:51:31.560 --> 00:51:33.296 And the third one is really hard

NOTE Confidence: 0.826091706842105

00:51:33.296 --> 00:51:35.200 to tell in this visualization,

NOTE Confidence: 0.826091706842105

00:51:35.200 --> 00:51:37.402 but this is data that's samples

NOTE Confidence: 0.826091706842105

00:51:37.402 --> 00:51:39.724 from the surface of a doughnut

NOTE Confidence: 0.826091706842105

00:51:39.724 --> 00:51:41.674 or a Taurus in mathematics.

NOTE Confidence: 0.826091706842105

00:51:41.680 --> 00:51:44.599 So this is again A3 dimensional object.

NOTE Confidence: 0.826091706842105

00:51:44.600 --> 00:51:46.120 It in the lower half,

NOTE Confidence: 0.826091706842105

00:51:46.120 --> 00:51:49.012 I'm showing you persistence diagrams where

NOTE Confidence: 0.826091706842105

00:51:49.012 --> 00:51:52.326 the black dots are representing dimension

NOTE Confidence: 0.826091706842105

00:51:52.326 --> 00:51:56.238 0 homology and that's connected components.

NOTE Confidence: 0.826091706842105

00:51:56.240 --> 00:51:59.636 The red triangles are representing dimension

NOTE Confidence: 0.826091706842105

00:51:59.636 --> 00:52:03.559 1 homology which are loops in our data.

NOTE Confidence: 0.826091706842105

00:52:03.560 --> 00:52:06.702 And now we have blue diamonds,

NOTE Confidence: 0.826091706842105

00:52:06.702 --> 00:52:09.412 and the blue diamonds are

NOTE Confidence: 0.826091706842105

00:52:09.412 --> 00:52:12.464 representing H2 dimension 2 homology,

NOTE Confidence: 0.826091706842105

00:52:12.464 --> 00:52:14.342 which are three-dimensional

NOTE Confidence: 0.826091706842105

00:52:14.342 --> 00:52:17.440 holes or voids in our data.

NOTE Confidence: 0.826091706842105

00:52:17.440 --> 00:52:20.440 And so I'd like you to think about

NOTE Confidence: 0.826091706842105

00:52:20.440 --> 00:52:23.627 matching these data sets to their

NOTE Confidence: 0.826091706842105

00:52:23.627 --> 00:52:25.397 corresponding persistence diagrams.

NOTE Confidence: 0.826091706842105

00:52:25.400 --> 00:52:27.717 A hint would be to look at

NOTE Confidence: 0.826091706842105

00:52:27.717 --> 00:52:29.240 the blue diamonds first,

NOTE Confidence: 0.826091706842105

00:52:29.240 --> 00:52:31.140 because blue diamonds are

NOTE Confidence: 0.826091706842105

00:52:31.140 --> 00:52:33.040 indicating 3D empty space.

NOTE Confidence: 0.826091706842105

00:52:33.040 --> 00:52:34.822 I'm going to take a quick drink of water

NOTE Confidence: 0.826091706842105

00:52:34.822 --> 00:52:36.520 and then we'll go over the solution.

NOTE Confidence: 0.891803057142857

00:52:49.280 --> 00:52:54.284 OK, So hopefully folks have realized

NOTE Confidence: 0.891803057142857

00:52:54.284 --> 00:52:59.356 that this persistence diagram on the left

NOTE Confidence: 0.891803057142857

00:52:59.360 --> 00:53:01.754 doesn't have any blue diamonds in it,

NOTE Confidence: 0.891803057142857

00:53:01.760 --> 00:53:04.840 doesn't have any 3D empty space in it,

NOTE Confidence: 0.891803057142857

00:53:04.840 --> 00:53:07.878 and therefore it corresponds to the second

NOTE Confidence: 0.891803057142857

00:53:07.878 --> 00:53:11.598 data set of the three concentric circles.

NOTE Confidence: 0.891803057142857

00:53:11.600 --> 00:53:14.201 You can see that there is 2 red triangles

NOTE Confidence: 0.891803057142857

00:53:14.201 --> 00:53:17.080 that are further away from the diagonal here,

NOTE Confidence: 0.891803057142857

00:53:17.080 --> 00:53:19.334 and there's one red triangle that's a

NOTE Confidence: 0.891803057142857

00:53:19.334 --> 00:53:21.758 little bit away from the diagonal here.

NOTE Confidence: 0.891803057142857

00:53:21.760 --> 00:53:25.066 And those three, those 3 triangles

NOTE Confidence: 0.891803057142857

00:53:25.066 --> 00:53:27.920 correspond to these three loops.

NOTE Confidence: 0.891803057142857

00:53:27.920 --> 00:53:29.663 One of the triangles is quite close

NOTE Confidence: 0.891803057142857

00:53:29.663 --> 00:53:31.310 to the diagonal because of the

NOTE Confidence: 0.891803057142857

00:53:31.310 --> 00:53:32.715 fact that these are concentric,

NOTE Confidence: 0.891803057142857

00:53:32.720 --> 00:53:35.128 so you can kind of bridge across

NOTE Confidence: 0.891803057142857

00:53:35.128 --> 00:53:36.160 them quite easily.

NOTE Confidence: 0.891803057142857

00:53:36.160 --> 00:53:38.920 Now we have 2 persistence diagrams

NOTE Confidence: 0.891803057142857

00:53:38.920 --> 00:53:40.959 that have diamonds in them,

NOTE Confidence: 0.891803057142857

00:53:40.959 --> 00:53:44.120 and to figure out which one is which,

NOTE Confidence: 0.891803057142857

00:53:44.120 --> 00:53:46.608 I think you have to look at some

NOTE Confidence: 0.891803057142857

00:53:46.608 --> 00:53:48.520 of these triangles again.

NOTE Confidence: 0.891803057142857

00:53:48.520 --> 00:53:50.704 And So what distinguishes the right one

NOTE Confidence: 0.891803057142857

00:53:50.704 --> 00:53:53.115 from the left one is on the right one.

NOTE Confidence: 0.891803057142857

00:53:53.120 --> 00:53:55.166 I'm not really seeing any triangles

NOTE Confidence: 0.891803057142857

00:53:55.166 --> 00:53:57.758 that are very far from the diagonal.

NOTE Confidence: 0.891803057142857

00:53:57.760 --> 00:53:59.068 But in this one,

NOTE Confidence: 0.891803057142857

00:53:59.068 --> 00:54:01.030 I see one triangle here that's

NOTE Confidence: 0.891803057142857

00:54:01.106 --> 00:54:02.638 far from the diagonal.

NOTE Confidence: 0.891803057142857

00:54:02.640 --> 00:54:04.490 And maybe there's another triangle

NOTE Confidence: 0.891803057142857

00:54:04.490 --> 00:54:07.060 here that's kind of separated from all

NOTE Confidence: 0.891803057142857

00:54:07.060 --> 00:54:09.034 the noise over here that's slightly

NOTE Confidence: 0.891803057142857

00:54:09.034 --> 00:54:10.799 further away from the diagonal.

NOTE Confidence: 0.891803057142857

00:54:10.800 --> 00:54:14.193 And so the way you can get these two

NOTE Confidence: 0.891803057142857

00:54:14.193 --> 00:54:17.439 triangles is because when you have a Taurus,

NOTE Confidence: 0.891803057142857

00:54:17.440 --> 00:54:20.077 if you think about a Taurus as a doughnut,

NOTE Confidence: 0.891803057142857

00:54:20.080 --> 00:54:23.925 you have a a circle that goes

NOTE Confidence: 0.891803057142857

00:54:23.925 --> 00:54:25.080 across the torus,

NOTE Confidence: 0.891803057142857

00:54:25.080 --> 00:54:27.060 like a horizontal circle

NOTE Confidence: 0.891803057142857

00:54:27.060 --> 00:54:29.040 going across the donor.

NOTE Confidence: 0.891803057142857

00:54:29.040 --> 00:54:30.996 And then you have another loop,

NOTE Confidence: 0.891803057142857

00:54:31.000 --> 00:54:33.538 another circle that goes kind of

NOTE Confidence: 0.891803057142857

00:54:33.538 --> 00:54:36.063 perpendicular to the 1st circle that

NOTE Confidence: 0.891803057142857

00:54:36.063 --> 00:54:38.716 goes around the donor in this way.

NOTE Confidence: 0.891803057142857

00:54:38.720 --> 00:54:40.240 So then in a donor,

NOTE Confidence: 0.891803057142857

00:54:40.240 --> 00:54:43.078 there are two loops and there's

NOTE Confidence: 0.891803057142857

00:54:43.078 --> 00:54:46.199 one empty space or one 3D hole.

NOTE Confidence: 0.891803057142857

00:54:46.200 --> 00:54:49.356 Whereas in the concentric spheres example,

NOTE Confidence: 0.891803057142857

00:54:49.360 --> 00:54:53.280 you just have the empty space between the

NOTE Confidence: 0.891803057142857

00:54:53.280 --> 00:54:57.037 two spheres that's shown by this diamond.

NOTE Confidence: 0.891803057142857

00:54:57.040 --> 00:54:57.720 So sorry,

NOTE Confidence: 0.891803057142857

00:54:57.720 --> 00:54:59.080 there's 22 empty spaces.

NOTE Confidence: 0.891803057142857

00:54:59.080 --> 00:55:02.398 There's empty space inside the inner sphere,

NOTE Confidence: 0.891803057142857

00:55:02.400 --> 00:55:04.740 and there's empty space between the

NOTE Confidence: 0.891803057142857

00:55:04.740 --> 00:55:07.280 outer sphere and the inner sphere.

NOTE Confidence: 0.891803057142857

00:55:07.280 --> 00:55:09.548 And that shows up here because you

NOTE Confidence: 0.891803057142857

00:55:09.548 --> 00:55:11.680 have one blue diamond up here,

NOTE Confidence: 0.891803057142857

00:55:11.680 --> 00:55:13.619 and then maybe you have one blue

NOTE Confidence: 0.891803057142857

00:55:13.619 --> 00:55:14.862 diamond here that's slightly

NOTE Confidence: 0.891803057142857

00:55:14.862 --> 00:55:16.637 further away from the diagonal.

NOTE Confidence: 0.891803057142857

00:55:16.640 --> 00:55:19.419 And so those correspond to the space

NOTE Confidence: 0.891803057142857

00:55:19.419 --> 00:55:22.139 inside the inner sphere and the

NOTE Confidence: 0.891803057142857

00:55:22.139 --> 00:55:25.037 interstitial space between the two spheres.

NOTE Confidence: 0.891803057142857

00:55:25.040 --> 00:55:26.780 So hopefully that helps you build

NOTE Confidence: 0.891803057142857

00:55:26.780 --> 00:55:27.360 some intuition.

NOTE Confidence: 0.891803057142857

00:55:27.360 --> 00:55:29.264 This is a lot easier to do when

NOTE Confidence: 0.891803057142857

00:55:29.264 --> 00:55:31.159 I put confidence intervals,

NOTE Confidence: 0.891803057142857

00:55:31.160 --> 00:55:33.274 but if I do put confidence intervals,

NOTE Confidence: 0.891803057142857

00:55:33.280 --> 00:55:35.520 we'll have to compute them separately for

NOTE Confidence: 0.891803057142857

00:55:35.520 --> 00:55:38.760 dimension 0, dimension 1, and dimension 2.

NOTE Confidence: 0.891803057142857

00:55:38.760 --> 00:55:40.200 And so that makes things.

NOTE Confidence: 0.891803057142857

00:55:40.200 --> 00:55:40.524 Also,

NOTE Confidence: 0.891803057142857

00:55:40.524 --> 00:55:42.468 I don't really have enough space

NOTE Confidence: 0.891803057142857

00:55:42.468 --> 00:55:45.074 to to draw 3 persistence diagrams

NOTE Confidence: 0.891803057142857

00:55:45.074 --> 00:55:47.154 with three confidence intervals,

NOTE Confidence: 0.891803057142857

00:55:47.160 --> 00:55:48.480 but hopefully that makes sense.

NOTE Confidence: 0.891803057142857

00:55:48.480 --> 00:55:50.872 If you have any question about how to

NOTE Confidence: 0.891803057142857

00:55:50.872 --> 00:55:52.234 interpret these persistence diagrams

NOTE Confidence: 0.891803057142857

00:55:52.234 --> 00:55:55.034 or like the solution to this little quiz,

NOTE Confidence: 0.891803057142857

00:55:55.040 --> 00:55:56.720 please feel free to chime in.

NOTE Confidence: 0.954864083333333

00:55:59.440 --> 00:56:02.560 OK, so now at this point,

NOTE Confidence: 0.954864083333333

00:56:02.560 --> 00:56:05.376 we we have figured out how to take

NOTE Confidence: 0.954864083333333

00:56:05.376 --> 00:56:08.440 our point cloud data set and convert

NOTE Confidence: 0.954864083333333

00:56:08.440 --> 00:56:10.720 it into a topological barcode,

NOTE Confidence: 0.954864083333333

00:56:10.720 --> 00:56:12.520 which we can represent

NOTE Confidence: 0.954864083333333

00:56:12.520 --> 00:56:14.320 as a persistence diagram.

NOTE Confidence: 0.954864083333333

00:56:14.320 --> 00:56:16.860 So the next thing we want to do is we

NOTE Confidence: 0.954864083333333

00:56:16.931 --> 00:56:19.675 want to compare two different data sets.

NOTE Confidence: 0.954864083333333

00:56:19.680 --> 00:56:21.710 And so comparing two different

NOTE Confidence: 0.954864083333333

00:56:21.710 --> 00:56:24.320 point clouds can be quite tricky.

NOTE Confidence: 0.954864083333333

00:56:24.320 --> 00:56:27.435 You don't really know like how to

NOTE Confidence: 0.954864083333333

00:56:27.435 --> 00:56:30.219 distinguish a torus from a sphere

NOTE Confidence: 0.954864083333333

00:56:30.219 --> 00:56:32.479 from some other blobby thing.

NOTE Confidence: 0.954864083333333

00:56:32.480 --> 00:56:34.880 And so one way in which you can

NOTE Confidence: 0.954864083333333

00:56:34.880 --> 00:56:37.042 compare these kinds of data to

NOTE Confidence: 0.954864083333333

00:56:37.042 --> 00:56:40.251 different point clouds is by instead

NOTE Confidence: 0.954864083333333

00:56:40.251 --> 00:56:43.039 comparing their persistence diagrams.

NOTE Confidence: 0.954864083333333

00:56:43.040 --> 00:56:45.995 And so there are multiple

NOTE Confidence: 0.954864083333333

00:56:45.995 --> 00:56:48.359 techniques to compute distances

NOTE Confidence: 0.954864083333333

00:56:48.359 --> 00:56:50.960 between persistence diagrams.

NOTE Confidence: 0.954864083333333

00:56:50.960 --> 00:56:53.207 And one of those techniques is what's

NOTE Confidence: 0.954864083333333

00:56:53.207 --> 00:56:54.720 called the bottleneck distance.

NOTE Confidence: 0.954864083333333

00:56:54.720 --> 00:56:57.716 And So what happens in the bottleneck

NOTE Confidence: 0.954864083333333

00:56:57.716 --> 00:57:00.584 distance is that you paired up.

NOTE Confidence: 0.954864083333333

00:57:00.584 --> 00:57:01.640 And again,

NOTE Confidence: 0.954864083333333

00:57:01.640 --> 00:57:03.884 I should apologize here because I'm

NOTE Confidence: 0.954864083333333

00:57:03.884 --> 00:57:05.720 using color slightly differently here.

NOTE Confidence: 0.954864083333333

00:57:05.720 --> 00:57:09.059 So the the blue colored dots are

NOTE Confidence: 0.954864083333333

00:57:09.059 --> 00:57:11.280 from first persistence diagram,

NOTE Confidence: 0.954864083333333

00:57:11.280 --> 00:57:13.480 diagram one and the red

NOTE Confidence: 0.954864083333333

00:57:13.480 --> 00:57:15.680 squares are from diagram 2.

NOTE Confidence: 0.954864083333333

00:57:15.680 --> 00:57:17.480 And so we want to compare

NOTE Confidence: 0.954864083333333

00:57:17.480 --> 00:57:18.880 diagram 1 to diagram 2.

NOTE Confidence: 0.954864083333333

00:57:18.880 --> 00:57:21.832 And the way we do that is by first

NOTE Confidence: 0.954864083333333

00:57:21.832 --> 00:57:23.442 matching features in diagram

NOTE Confidence: 0.954864083333333

00:57:23.442 --> 00:57:25.836 1 to features in diagram 2,

NOTE Confidence: 0.954864083333333

00:57:25.840 --> 00:57:29.460 where we also allow to allow ourselves to

NOTE Confidence: 0.954864083333333

00:57:29.460 --> 00:57:32.400 map certain features to the diagonal itself.

NOTE Confidence: 0.954864083333333

00:57:32.400 --> 00:57:35.158 So that's a matching process that happens.

NOTE Confidence: 0.954864083333333

00:57:35.160 --> 00:57:37.645 And once you have matched the features

NOTE Confidence: 0.954864083333333

00:57:37.645 --> 00:57:40.078 to each other or to the diagonal,

NOTE Confidence: 0.954864083333333

00:57:40.080 --> 00:57:42.782 you find the two paired features that

NOTE Confidence: 0.954864083333333

00:57:42.782 --> 00:57:45.671 are furthest away from each other and

NOTE Confidence: 0.954864083333333

00:57:45.671 --> 00:57:48.155 you compute this distance between them.

NOTE Confidence: 0.954864083333333

00:57:48.160 --> 00:57:50.638 This is called the bottleneck distance,

NOTE Confidence: 0.954864083333333

00:57:50.640 --> 00:57:53.180 and there's ways to represent

NOTE Confidence: 0.954864083333333

00:57:53.180 --> 00:57:54.196 that mathematically.

NOTE Confidence: 0.954864083333333

00:57:54.200 --> 00:57:55.840 Here that's not so important.

NOTE Confidence: 0.954864083333333

00:57:55.840 --> 00:57:57.872 The intuition is probably

NOTE Confidence: 0.954864083333333

00:57:57.872 --> 00:57:59.396 what's most important,

NOTE Confidence: 0.954864083333333

00:57:59.400 --> 00:58:02.081 and there is a very important theorem

NOTE Confidence: 0.954864083333333

00:58:02.081 --> 00:58:04.980 in the field that guarantees stability

NOTE Confidence: 0.954864083333333

00:58:04.980 --> 00:58:07.760 is called the stability theorem.

NOTE Confidence: 0.954864083333333

00:58:07.760 --> 00:58:10.680 And what it says is that if I have a

NOTE Confidence: 0.954864083333333

00:58:10.764 --> 00:58:13.531 point cloud X and a point cloud Y,

NOTE Confidence: 0.954864083333333

00:58:13.531 --> 00:58:15.790 if my point cloud X is just a slight

NOTE Confidence: 0.954864083333333

00:58:15.863 --> 00:58:17.798 perturbation of point cloud Y.

NOTE Confidence: 0.954864083333333

00:58:17.800 --> 00:58:19.648 So I've just moved the points

NOTE Confidence: 0.954864083333333

00:58:19.648 --> 00:58:20.880 around a little bit.

NOTE Confidence: 0.954864083333333

00:58:20.880 --> 00:58:23.610 Then the bottleneck distance between

NOTE Confidence: 0.954864083333333

00:58:23.610 --> 00:58:25.794 the persistence diagrams computed

NOTE Confidence: 0.954864083333333

00:58:25.794 --> 00:58:28.235 through the Beatrice ribs complex of

NOTE Confidence: 0.954864083333333

00:58:28.235 --> 00:58:31.038 X and the Beatrice ribs complex of Y.

NOTE Confidence: 0.954864083333333

00:58:31.040 --> 00:58:33.104 This bottleneck distance is

NOTE Confidence: 0.954864083333333

00:58:33.104 --> 00:58:36.200 guaranteed to be small because if

NOTE Confidence: 0.954864083333333

00:58:36.293 --> 00:58:38.915 X is slightly different from Y,

NOTE Confidence: 0.954864083333333

00:58:38.920 --> 00:58:40.510 the right hand side of this

NOTE Confidence: 0.954864083333333

00:58:40.510 --> 00:58:42.399 equation is going to be close to 0,

NOTE Confidence: 0.954864083333333

00:58:42.400 --> 00:58:44.150 and therefore the bottleneck distance

NOTE Confidence: 0.954864083333333

00:58:44.150 --> 00:58:46.640 is going to be very very small.

NOTE Confidence: 0.954864083333333

00:58:46.640 --> 00:58:48.590 So this basically guarantees the fact

NOTE Confidence: 0.954864083333333

00:58:48.590 --> 00:58:50.918 that if you have one point cloud,

NOTE Confidence: 0.954864083333333

00:58:50.920 --> 00:58:52.672 maybe points arranged as a circle

NOTE Confidence: 0.954864083333333

00:58:52.672 --> 00:58:54.760 and you tweak the point slightly,

NOTE Confidence: 0.954864083333333

00:58:54.760 --> 00:58:56.496 so you add a little bit of noise

NOTE Confidence: 0.954864083333333

00:58:56.496 --> 00:58:57.360 to those points,

NOTE Confidence: 0.954864083333333

00:58:57.360 --> 00:58:59.360 then the bottleneck distance is

NOTE Confidence: 0.954864083333333

00:58:59.360 --> 00:59:01.360 not going to change much.

NOTE Confidence: 0.954864083333333

00:59:01.360 --> 00:59:03.866 So it means that this bottleneck distance

NOTE Confidence: 0.954864083333333

00:59:03.866 --> 00:59:07.036 is a stable way of comparing point clouds.

NOTE Confidence: 0.954864083333333

00:59:07.040 --> 00:59:09.119 Again, I don't care so much about the math.

NOTE Confidence: 0.954864083333333

00:59:09.120 --> 00:59:10.240 That's kind of the, the,

NOTE Confidence: 0.954864083333333

00:59:10.240 --> 00:59:13.320 the main result is that topology is,

NOTE Confidence: 0.954864083333333

00:59:13.320 --> 00:59:14.520 is able,

NOTE Confidence: 0.954864083333333

00:59:14.520 --> 00:59:15.120 it's,

NOTE Confidence: 0.954864083333333

00:59:15.120 --> 00:59:18.150 it's robust to these kinds of

NOTE Confidence: 0.954864083333333

00:59:18.150 --> 00:59:20.170 noise and perturbations in

NOTE Confidence: 0.865437954166667

00:59:20.267 --> 00:59:22.490 our data. Now one of the problems

NOTE Confidence: 0.865437954166667

00:59:22.490 --> 00:59:24.500 with the ball neck distance is that

NOTE Confidence: 0.865437954166667

00:59:24.500 --> 00:59:26.120 we are doing all this matching,

NOTE Confidence: 0.865437954166667

00:59:26.120 --> 00:59:28.136 but ultimately we are only really

NOTE Confidence: 0.865437954166667

00:59:28.136 --> 00:59:30.230 looking at the distance between points

NOTE Confidence: 0.865437954166667

00:59:30.230 --> 00:59:32.624 that are matched but are furthest away

NOTE Confidence: 0.865437954166667

00:59:32.624 --> 00:59:34.651 from each other and we're ignoring

NOTE Confidence: 0.865437954166667

00:59:34.651 --> 00:59:37.046 all the other points that got matched.

NOTE Confidence: 0.865437954166667

00:59:37.046 --> 00:59:40.110 So maybe you want to be more sensitive

NOTE Confidence: 0.865437954166667

00:59:40.191 --> 00:59:42.613 to how well the matching works and

NOTE Confidence: 0.865437954166667

00:59:42.613 --> 00:59:45.531 and the way to actually use that

NOTE Confidence: 0.865437954166667

00:59:45.531 --> 00:59:47.786 information is to compute what's

NOTE Confidence: 0.865437954166667

00:59:47.786 --> 00:59:50.260 called the washer steam distance,

NOTE Confidence: 0.865437954166667

00:59:50.260 --> 00:59:53.680 where you perform the matching process 1st.

NOTE Confidence: 0.865437954166667

00:59:53.680 --> 00:59:56.578 And then you compute this washer steam

NOTE Confidence: 0.865437954166667

00:59:56.578 --> 00:59:59.079 distance between diagram one and diagram 2,

NOTE Confidence: 0.865437954166667

00:59:59.080 --> 01:00:01.957 where you sum over the distance between

NOTE Confidence: 0.865437954166667

01:00:01.957 --> 01:00:04.797 all the points that are matched to

NOTE Confidence: 0.865437954166667

01:00:04.797 --> 01:00:06.755 each other and you ignore the points

NOTE Confidence: 0.865437954166667

01:00:06.755 --> 01:00:08.440 that get matched to the diagonal.

NOTE Confidence: 0.865437954166667

01:00:08.440 --> 01:00:12.193 So this is a much more even more stable

NOTE Confidence: 0.865437954166667

01:00:12.200 --> 01:00:15.560 way of comparing 2 persistence diagrams.

NOTE Confidence: 0.865437954166667

01:00:15.560 --> 01:00:17.876 And it also has similar stability

NOTE Confidence: 0.865437954166667

01:00:17.876 --> 01:00:20.506 properties that I was talking about

NOTE Confidence: 0.865437954166667

01:00:20.506 --> 01:00:23.554 previously in the Washington steam distance.

NOTE Confidence: 0.865437954166667

01:00:23.560 --> 01:00:24.493 If you have,

NOTE Confidence: 0.865437954166667

01:00:24.493 --> 01:00:26.670 if you have any kind of experience

NOTE Confidence: 0.865437954166667

01:00:26.741 --> 01:00:28.976 with like optimal transport theory

NOTE Confidence: 0.865437954166667

01:00:28.976 --> 01:00:30.317 or like statistics,

NOTE Confidence: 0.865437954166667

01:00:30.320 --> 01:00:32.567 then I just wanted to highlight that

NOTE Confidence: 0.865437954166667

01:00:32.567 --> 01:00:34.658 the washer steam distance that we're

NOTE Confidence: 0.865437954166667

01:00:34.658 --> 01:00:36.794 talking about here is similar to,

NOTE Confidence: 0.865437954166667

01:00:36.800 --> 01:00:38.600 it's actually exactly the same as

NOTE Confidence: 0.865437954166667

01:00:38.600 --> 01:00:40.358 the washer steam distance that you

NOTE Confidence: 0.865437954166667

01:00:40.358 --> 01:00:42.101 would be familiar with in the sense

NOTE Confidence: 0.865437954166667

01:00:42.101 --> 01:00:43.940 that you have a transport map and

NOTE Confidence: 0.865437954166667

01:00:43.940 --> 01:00:45.640 you're kind of moving mounds of

NOTE Confidence: 0.865437954166667

01:00:45.640 --> 01:00:47.680 earth from one place to another.

NOTE Confidence: 0.865437954166667

01:00:47.680 --> 01:00:49.514 Or the Washington steam distance that you

NOTE Confidence: 0.865437954166667

01:00:49.514 --> 01:00:51.679 used to compare to probability distributions.

NOTE Confidence: 0.865437954166667

01:00:51.680 --> 01:00:53.828 You can think about these topological

NOTE Confidence: 0.865437954166667

01:00:53.828 --> 01:00:55.623 features as like probability distributions

NOTE Confidence: 0.865437954166667

01:00:55.623 --> 01:00:58.359 and you're learning to map 1 to the other.

NOTE Confidence: 0.865437954166667

01:00:58.360 --> 01:01:01.560 So this is just an aside for folks who might

NOTE Confidence: 0.865437954166667

01:01:01.642 --> 01:01:04.558 be more familiar with optimal transport.

NOTE Confidence: 0.865437954166667

01:01:04.560 --> 01:01:06.680 OK,

NOTE Confidence: 0.865437954166667

01:01:06.680 --> 01:01:10.306 I wanted to put this like review

NOTE Confidence: 0.865437954166667

01:01:10.306 --> 01:01:13.318 article in because I have talked

NOTE Confidence: 0.865437954166667

01:01:13.318 --> 01:01:16.028 so far about some about extracting

NOTE Confidence: 0.865437954166667

01:01:16.028 --> 01:01:18.298 topological features from point cloud

NOTE Confidence: 0.865437954166667

01:01:18.298 --> 01:01:21.471 data set and then learning how to

NOTE Confidence: 0.865437954166667

01:01:21.471 --> 01:01:23.207 interpret those topological features

NOTE Confidence: 0.865437954166667

01:01:23.207 --> 01:01:25.989 and to compare two different data

NOTE Confidence: 0.865437954166667

01:01:25.989 --> 01:01:28.294 sets by computing the bottleneck

NOTE Confidence: 0.865437954166667

01:01:28.294 --> 01:01:31.200 distance or the washer steam distance

NOTE Confidence: 0.865437954166667

01:01:31.200 --> 01:01:33.152 between their topological features.

NOTE Confidence: 0.865437954166667

01:01:33.160 --> 01:01:36.639 But it is possible to also compute

NOTE Confidence: 0.865437954166667

01:01:36.640 --> 01:01:39.160 to use topology in a different way,

NOTE Confidence: 0.865437954166667

01:01:39.160 --> 01:01:42.562 where you use topology to inform the

NOTE Confidence: 0.865437954166667

01:01:42.562 --> 01:01:47.040 training of a machine learning architecture.

NOTE Confidence: 0.865437954166667

01:01:47.040 --> 01:01:49.392 And so for folks who are familiar

NOTE Confidence: 0.865437954166667

01:01:49.392 --> 01:01:50.400 with machine learning,

NOTE Confidence: 0.865437954166667

01:01:50.400 --> 01:01:52.500 I just wanted to point out that

NOTE Confidence: 0.865437954166667

01:01:52.500 --> 01:01:55.042 there are ways in which you can use

NOTE Confidence: 0.865437954166667

01:01:55.042 --> 01:01:57.220 topology inside the loss function

NOTE Confidence: 0.865437954166667

01:01:57.220 --> 01:01:59.080 of your neural network.

NOTE Confidence: 0.865437954166667

01:01:59.080 --> 01:02:01.784 So you can have a topology informed loss

NOTE Confidence: 0.865437954166667

01:02:01.784 --> 01:02:04.949 and a good example of this that is a

NOTE Confidence: 0.865437954166667

01:02:04.949 --> 01:02:08.120 paper by Channel in 2019 that I can point to.

NOTE Confidence: 0.865437954166667

01:02:08.120 --> 01:02:10.760 You can also in machine learning

NOTE Confidence: 0.865437954166667

01:02:10.760 --> 01:02:13.273 use topology to compare two

NOTE Confidence: 0.865437954166667

01:02:13.273 --> 01:02:15.277 different model architectures.

NOTE Confidence: 0.865437954166667

01:02:15.280 --> 01:02:17.224 One way of doing this would be to

NOTE Confidence: 0.865437954166667

01:02:17.224 --> 01:02:18.940 have two different like machine

NOTE Confidence: 0.865437954166667

01:02:18.940 --> 01:02:20.492 learning architectures where you

NOTE Confidence: 0.865437954166667

01:02:20.492 --> 01:02:22.875 look at the activations in all

NOTE Confidence: 0.865437954166667

01:02:22.875 --> 01:02:24.795 the layers of those architectures,

NOTE Confidence: 0.865437954166667

01:02:24.800 --> 01:02:27.360 treat that as a point cloud and compare

NOTE Confidence: 0.865437954166667

01:02:27.360 --> 01:02:29.904 them against each other using Washington

NOTE Confidence: 0.865437954166667

01:02:29.904 --> 01:02:32.199 distance of their persistence features.

NOTE Confidence: 0.865437954166667

01:02:32.200 --> 01:02:34.390 And that's in a good example of that is Zo

NOTE Confidence: 0.725452529166667

01:02:34.447 --> 01:02:35.622 ET al. In 2021.

NOTE Confidence: 0.725452529166667

01:02:35.622 --> 01:02:38.639 And then the what I'm talking about,

NOTE Confidence: 0.725452529166667

01:02:38.640 --> 01:02:41.080 which is similar to this paper in 2017,

NOTE Confidence: 0.725452529166667

01:02:41.080 --> 01:02:44.192 is to actually just take your data and

NOTE Confidence: 0.725452529166667

01:02:44.192 --> 01:02:46.798 use topology to featurize that data,

NOTE Confidence: 0.725452529166667

01:02:46.800 --> 01:02:48.600 which is to extract topological

NOTE Confidence: 0.725452529166667

01:02:48.600 --> 01:02:50.040 features of that data,

NOTE Confidence: 0.725452529166667

01:02:50.040 --> 01:02:52.012 learn to interpret those

NOTE Confidence: 0.725452529166667

01:02:52.012 --> 01:02:52.998 topological features,

NOTE Confidence: 0.725452529166667

01:02:53.000 --> 01:02:55.202 and then perhaps pass them into

NOTE Confidence: 0.725452529166667

01:02:55.202 --> 01:02:56.670 some machine learning framework

NOTE Confidence: 0.725452529166667

01:02:56.733 --> 01:02:58.599 to generate some kind of output.

NOTE Confidence: 0.725452529166667

01:02:58.600 --> 01:03:01.344 So there are different places in machine

NOTE Confidence: 0.725452529166667

01:03:01.344 --> 01:03:03.959 learning that where one can use topology.

NOTE Confidence: 0.725452529166667

01:03:03.960 --> 01:03:04.797 In our case,

NOTE Confidence: 0.725452529166667

01:03:04.797 --> 01:03:07.218 we are going to focus on ways in

NOTE Confidence: 0.725452529166667

01:03:07.218 --> 01:03:09.213 which we use topological features

NOTE Confidence: 0.725452529166667

01:03:09.213 --> 01:03:11.539 extracted from our data and pass

NOTE Confidence: 0.725452529166667

01:03:11.539 --> 01:03:13.531 them into machine learning in order

NOTE Confidence: 0.725452529166667

01:03:13.531 --> 01:03:17.480 to do some kind of downstream task.

NOTE Confidence: 0.725452529166667

01:03:17.480 --> 01:03:21.140 OK, so next I wanted to cover a few ways

NOTE Confidence: 0.725452529166667

01:03:21.239 --> 01:03:24.539 of taking these topological features

NOTE Confidence: 0.725452529166667

01:03:24.539 --> 01:03:27.839 and converting them into summaries.

NOTE Confidence: 0.725452529166667

01:03:27.840 --> 01:03:29.844 And the reason for doing that

NOTE Confidence: 0.725452529166667

01:03:29.844 --> 01:03:32.565 is because we want to use these

NOTE Confidence: 0.725452529166667

01:03:32.565 --> 01:03:34.755 topological features as input for

NOTE Confidence: 0.725452529166667

01:03:34.755 --> 01:03:36.678 machine learning down the line.

NOTE Confidence: 0.725452529166667

01:03:36.680 --> 01:03:38.555 And these diagrams that I've

NOTE Confidence: 0.725452529166667

01:03:38.555 --> 01:03:40.800 been drawing for you so far,

NOTE Confidence: 0.725452529166667

01:03:40.800 --> 01:03:43.236 they're easy to draw on a screen,

NOTE Confidence: 0.725452529166667

01:03:43.240 --> 01:03:45.388 but they're not really great for

NOTE Confidence: 0.725452529166667

01:03:45.388 --> 01:03:47.187 machine learning because if you

NOTE Confidence: 0.725452529166667

01:03:47.187 --> 01:03:49.280 have a bunch of different data sets,

NOTE Confidence: 0.725452529166667

01:03:49.280 --> 01:03:51.793 you're going to get different number of

NOTE Confidence: 0.725452529166667

01:03:51.793 --> 01:03:53.719 topological features for each data set.

NOTE Confidence: 0.725452529166667

01:03:53.720 --> 01:03:55.912 And you don't really know a way of

NOTE Confidence: 0.725452529166667

01:03:55.912 --> 01:03:57.505 like converting this into something

NOTE Confidence: 0.725452529166667

01:03:57.505 --> 01:03:59.479 that can go into machine learning.

NOTE Confidence: 0.725452529166667

01:03:59.480 --> 01:04:02.511 So folks have found various ways of

NOTE Confidence: 0.725452529166667

01:04:02.511 --> 01:04:04.433 taking these persistence diagrams

NOTE Confidence: 0.725452529166667

01:04:04.433 --> 01:04:07.685 and converting them into even more

NOTE Confidence: 0.725452529166667

01:04:07.685 --> 01:04:10.198 convenient representations that can be

NOTE Confidence: 0.725452529166667

01:04:10.198 --> 01:04:12.634 used for either with medical analysis,

NOTE Confidence: 0.725452529166667

01:04:12.640 --> 01:04:14.415 but more importantly for machine

NOTE Confidence: 0.725452529166667

01:04:14.415 --> 01:04:15.835 learning down the road.

NOTE Confidence: 0.725452529166667

01:04:15.840 --> 01:04:18.125 And one such representation is

NOTE Confidence: 0.725452529166667

01:04:18.125 --> 01:04:19.953 called the persistence landscape,

NOTE Confidence: 0.725452529166667

01:04:19.960 --> 01:04:22.792 where you are taking a diagram like this

NOTE Confidence: 0.725452529166667

01:04:22.792 --> 01:04:24.919 and converting that into a function.

NOTE Confidence: 0.725452529166667

01:04:24.920 --> 01:04:26.474 And the way you do that is,

NOTE Confidence: 0.725452529166667

01:04:26.480 --> 01:04:27.988 is quite simple really.

NOTE Confidence: 0.725452529166667

01:04:27.988 --> 01:04:31.025 You take each point and you draw a

NOTE Confidence: 0.725452529166667

01:04:31.025 --> 01:04:33.461 10th function based off of that point

NOTE Confidence: 0.725452529166667

01:04:33.461 --> 01:04:36.277 by connecting it to its X coordinate

NOTE Confidence: 0.725452529166667

01:04:36.277 --> 01:04:39.422 and connecting it to its Y coordinate

NOTE Confidence: 0.725452529166667

01:04:39.422 --> 01:04:41.478 intersected with the diagonal.

NOTE Confidence: 0.725452529166667

01:04:41.480 --> 01:04:43.478 It takes more words to describe,

NOTE Confidence: 0.725452529166667

01:04:43.480 --> 01:04:45.293 so you just you can just simply

NOTE Confidence: 0.725452529166667

01:04:45.293 --> 01:04:46.919 see it from this picture.

NOTE Confidence: 0.725452529166667

01:04:46.920 --> 01:04:49.554 You draw this little tent function

NOTE Confidence: 0.725452529166667

01:04:49.554 --> 01:04:53.226 and then tilt the diagram by 45° and

NOTE Confidence: 0.725452529166667

01:04:53.226 --> 01:04:56.677 there you end up getting this function

NOTE Confidence: 0.725452529166667

01:04:56.677 --> 01:04:59.519 representation of your persistence diagram.

NOTE Confidence: 0.725452529166667

01:04:59.520 --> 01:05:02.200 You can now treat this as a function,

NOTE Confidence: 0.725452529166667

01:05:02.200 --> 01:05:05.812 and you can use tools from functional

NOTE Confidence: 0.725452529166667

01:05:05.812 --> 01:05:08.662 analysis to analyze this persistence diagram.

NOTE Confidence: 0.725452529166667

01:05:08.662 --> 01:05:09.096 Again,

NOTE Confidence: 0.725452529166667

01:05:09.096 --> 01:05:12.378 you can kind of formalize this with a

NOTE Confidence: 0.725452529166667

01:05:12.378 --> 01:05:14.802 bunch of math by drawing out what A10

NOTE Confidence: 0.725452529166667

01:05:14.873 --> 01:05:17.281 function looks like and how you take

NOTE Confidence: 0.725452529166667

01:05:17.281 --> 01:05:20.920 a diagram and convert it into a function,

NOTE Confidence: 0.725452529166667

01:05:20.920 --> 01:05:24.718 but this is all just notation.

NOTE Confidence: 0.725452529166667

01:05:24.720 --> 01:05:27.317 I simply want to convey to you

NOTE Confidence: 0.725452529166667

01:05:27.317 --> 01:05:29.287 the intuition behind taking a

NOTE Confidence: 0.725452529166667

01:05:29.287 --> 01:05:31.639 diagram and converting it into a

NOTE Confidence: 0.725452529166667

01:05:31.639 --> 01:05:33.959 function for downstream analysis.

NOTE Confidence: 0.725452529166667

01:05:33.960 --> 01:05:36.312 There's some important reasons why one

NOTE Confidence: 0.725452529166667

01:05:36.312 --> 01:05:39.319 might want to convert this into a function.

NOTE Confidence: 0.725452529166667

01:05:39.320 --> 01:05:41.280 One of them is that you can use

NOTE Confidence: 0.725452529166667

01:05:41.280 --> 01:05:43.000 tools from functional analysis.

NOTE Confidence: 0.725452529166667

01:05:43.000 --> 01:05:45.080 Another thing that's important is

NOTE Confidence: 0.725452529166667

01:05:45.080 --> 01:05:47.800 that this is an injective mapping,

NOTE Confidence: 0.725452529166667

01:05:47.800 --> 01:05:50.464 and it satisfies the same properties

NOTE Confidence: 0.725452529166667

01:05:50.464 --> 01:05:52.240 that persistence diagrams satisfy.

NOTE Confidence: 0.956275205

01:05:54.280 --> 01:05:57.105 Another convenient way of converting

NOTE Confidence: 0.956275205

01:05:57.105 --> 01:05:59.365 persistence diagrams into something

NOTE Confidence: 0.956275205

01:05:59.365 --> 01:06:02.085 that's useful for machine learning is

NOTE Confidence: 0.956275205

01:06:02.085 --> 01:06:04.879 to convert the diagram into an image.

NOTE Confidence: 0.956275205

01:06:04.880 --> 01:06:07.211 The reason you might want to do this is

NOTE Confidence: 0.956275205

01:06:07.211 --> 01:06:09.517 because we have architectures that are very,

NOTE Confidence: 0.956275205

01:06:09.520 --> 01:06:12.238 very good at dealing with images.

NOTE Confidence: 0.956275205

01:06:12.240 --> 01:06:14.000 We know how to classify

NOTE Confidence: 0.956275205

01:06:14.000 --> 01:06:15.760 cats and dogs and horses.

NOTE Confidence: 0.956275205

01:06:15.760 --> 01:06:17.797 We also know how to generate images.

NOTE Confidence: 0.956275205

01:06:17.800 --> 01:06:20.140 So we can take advantage of all the tools

NOTE Confidence: 0.956275205

01:06:20.140 --> 01:06:22.620 we have developed for dealing with images

NOTE Confidence: 0.956275205

01:06:22.620 --> 01:06:25.232 in machine learning if we can convert

NOTE Confidence: 0.956275205

01:06:25.232 --> 01:06:27.077 persistence diagrams into an image.

NOTE Confidence: 0.956275205

01:06:27.080 --> 01:06:29.834 And the way one goes about doing that is

NOTE Confidence: 0.956275205

01:06:29.834 --> 01:06:32.718 you take your input persistence diagram,

NOTE Confidence: 0.956275205

01:06:32.720 --> 01:06:35.680 you tilt it again by 45°.

NOTE Confidence: 0.956275205

01:06:35.680 --> 01:06:37.600 So now you're measuring the birth

NOTE Confidence: 0.956275205

01:06:37.600 --> 01:06:39.017 coordinate and you're measuring

NOTE Confidence: 0.956275205

01:06:39.017 --> 01:06:41.330 distance from the diagonal, which we

NOTE Confidence: 0.956275205

01:06:41.330 --> 01:06:43.520 call persistence on the Y coordinate.

NOTE Confidence: 0.956275205

01:06:43.520 --> 01:06:44.693 So nothing fancy,

NOTE Confidence: 0.956275205

01:06:44.693 --> 01:06:47.039 just kind of tilting the diagram.

NOTE Confidence: 0.956275205

01:06:47.040 --> 01:06:50.505 Then what you do is at each point in

NOTE Confidence: 0.956275205

01:06:50.505 --> 01:06:53.572 the diagram, you drop a Gaussian,

NOTE Confidence: 0.956275205

01:06:53.572 --> 01:06:55.837 so like a 2D Gaussian,

NOTE Confidence: 0.956275205

01:06:55.840 --> 01:06:58.384 and you weigh the Gaussians by

NOTE Confidence: 0.956275205

01:06:58.384 --> 01:07:00.720 distance away from the X axis.

NOTE Confidence: 0.956275205

01:07:00.720 --> 01:07:02.560 So points that are higher

NOTE Confidence: 0.956275205

01:07:02.560 --> 01:07:04.400 up get a brighter Gaussian.

NOTE Confidence: 0.956275205

01:07:04.400 --> 01:07:06.200 The points that are lower down,

NOTE Confidence: 0.956275205

01:07:06.200 --> 01:07:08.918 they get a lower amplitude Gaussian.

NOTE Confidence: 0.956275205

01:07:08.920 --> 01:07:09.239 Again,

NOTE Confidence: 0.956275205

01:07:09.239 --> 01:07:11.472 the rationale for doing that is that

NOTE Confidence: 0.956275205

01:07:11.472 --> 01:07:14.142 points that are further away are points

NOTE Confidence: 0.956275205

01:07:14.142 --> 01:07:16.112 that are more topologically significant.

NOTE Confidence: 0.956275205

01:07:16.120 --> 01:07:18.353 Points that are closed are kind of

NOTE Confidence: 0.956275205

01:07:18.353 --> 01:07:20.479 derived from some noise in our data,

NOTE Confidence: 0.956275205

01:07:20.480 --> 01:07:22.280 and we want to be robust to noise.

NOTE Confidence: 0.956275205

01:07:22.280 --> 01:07:24.387 So it makes sense to weigh things

NOTE Confidence: 0.956275205

01:07:24.387 --> 01:07:26.438 by distance away from the diagonal.

NOTE Confidence: 0.956275205

01:07:26.440 --> 01:07:28.036 This is called a persistence image.

NOTE Confidence: 0.956275205

01:07:28.040 --> 01:07:30.440 This is still a continuous object.

NOTE Confidence: 0.956275205

01:07:30.440 --> 01:07:32.933 And So what you can do then is you

NOTE Confidence: 0.956275205

01:07:32.933 --> 01:07:35.427 can take this surface and you can

NOTE Confidence: 0.956275205

01:07:35.427 --> 01:07:37.791 just divide it into smaller pixels

NOTE Confidence: 0.956275205

01:07:37.791 --> 01:07:40.794 and convert this into an image format.

NOTE Confidence: 0.956275205

01:07:40.800 --> 01:07:42.915 And once you have this in an image format,

NOTE Confidence: 0.956275205

01:07:42.920 --> 01:07:45.205 you can use convolutional neural

NOTE Confidence: 0.956275205

01:07:45.205 --> 01:07:48.027 networks and other kinds of like

NOTE Confidence: 0.956275205

01:07:48.027 --> 01:07:50.441 generative AI tools to take advantage

NOTE Confidence: 0.956275205

01:07:50.441 --> 01:07:53.418 of like all of those tools to to

NOTE Confidence: 0.956275205

01:07:53.418 --> 01:07:55.758 work with these persistence images.

NOTE Confidence: 0.956275205

01:07:55.760 --> 01:07:56.026 Again,

NOTE Confidence: 0.956275205

01:07:56.026 --> 01:07:57.888 there's a bunch of math that one

NOTE Confidence: 0.956275205

01:07:57.888 --> 01:08:00.077 can write down to kind of formally

NOTE Confidence: 0.956275205

01:08:00.077 --> 01:08:01.034 describe this process,

NOTE Confidence: 0.956275205

01:08:01.040 --> 01:08:03.600 but I think the visuals do a much better job.

NOTE Confidence: 0.903809147857143

01:08:05.720 --> 01:08:08.210 Finally, you can convert your

NOTE Confidence: 0.903809147857143

01:08:08.210 --> 01:08:10.700 persistence diagrams into what are

NOTE Confidence: 0.903809147857143

01:08:10.787 --> 01:08:13.359 called smooth persistence curves.

NOTE Confidence: 0.903809147857143

01:08:13.360 --> 01:08:15.820 And the way this works is

NOTE Confidence: 0.903809147857143

01:08:15.820 --> 01:08:18.400 you walk along the diagonal,

NOTE Confidence: 0.903809147857143

01:08:18.400 --> 01:08:20.619 and as you're walking along the diagonal

NOTE Confidence: 0.903809147857143

01:08:20.619 --> 01:08:22.877 from the bottom left to the top right,

NOTE Confidence: 0.903809147857143

01:08:22.880 --> 01:08:25.360 you look at a window.

NOTE Confidence: 0.903809147857143

01:08:25.360 --> 01:08:26.336 And you construct the

NOTE Confidence: 0.903809147857143

01:08:26.336 --> 01:08:27.556 window by looking at this,

NOTE Confidence: 0.903809147857143

01:08:27.560 --> 01:08:29.732 like this rectangular section

NOTE Confidence: 0.903809147857143

01:08:29.732 --> 01:08:32.990 that's to the top left of

NOTE Confidence: 0.903809147857143

01:08:33.097 --> 01:08:36.355 wherever you are on the diagonal.

NOTE Confidence: 0.903809147857143

01:08:36.360 --> 01:08:38.440 And what you do is you compute some

NOTE Confidence: 0.903809147857143

01:08:38.440 --> 01:08:40.558 kind of statistic of points that

NOTE Confidence: 0.903809147857143

01:08:40.558 --> 01:08:42.473 exist within this little window.

NOTE Confidence: 0.903809147857143

01:08:42.480 --> 01:08:44.844 So one simple statistic would be

NOTE Confidence: 0.903809147857143

01:08:44.844 --> 01:08:47.250 simply counting the number of points

NOTE Confidence: 0.903809147857143

01:08:47.250 --> 01:08:49.155 that exist within this window.

NOTE Confidence: 0.903809147857143

01:08:49.160 --> 01:08:52.450 So that allows you to construct a

NOTE Confidence: 0.903809147857143

01:08:52.450 --> 01:08:54.933 function as you're walking from

NOTE Confidence: 0.903809147857143

01:08:54.933 --> 01:08:57.618 left to right, construct a curve,

NOTE Confidence: 0.903809147857143

01:08:57.618 --> 01:08:58.836 a continuous curve,

NOTE Confidence: 0.903809147857143

01:08:58.840 --> 01:09:00.920 which you can then analyze.

NOTE Confidence: 0.937590552222222

01:09:03.280 --> 01:09:05.512 I don't have the curve here for some reason,

NOTE Confidence: 0.937590552222222

01:09:05.520 --> 01:09:06.702 but you can imagine like as

NOTE Confidence: 0.937590552222222

01:09:06.702 --> 01:09:07.880 you're walking along the diagonal,

NOTE Confidence: 0.937590552222222

01:09:07.880 --> 01:09:09.670 just counting how many objects

NOTE Confidence: 0.937590552222222

01:09:09.670 --> 01:09:11.865 exist within this window over time

NOTE Confidence: 0.937590552222222

01:09:11.865 --> 01:09:13.555 gives you a continuous curve,

NOTE Confidence: 0.937590552222222

01:09:13.560 --> 01:09:15.204 which you can describe, but medically

NOTE Confidence: 0.937590552222222

01:09:15.204 --> 01:09:17.159 you can prove things about that curve.

NOTE Confidence: 0.937590552222222

01:09:17.160 --> 01:09:19.184 And one of the reasons why you might

NOTE Confidence: 0.937590552222222

01:09:19.184 --> 01:09:21.347 want to use this curve is that there

NOTE Confidence: 0.937590552222222

01:09:21.347 --> 01:09:23.855 are ways to speed this process up a lot

NOTE Confidence: 0.937590552222222

01:09:23.855 --> 01:09:25.632 because these are all like Gaussians.

NOTE Confidence: 0.937590552222222

01:09:25.632 --> 01:09:27.984 And so there's ways to compute

NOTE Confidence: 0.937590552222222

01:09:27.984 --> 01:09:30.376 this curve very, very fast.

NOTE Confidence: 0.937590552222222

01:09:30.376 --> 01:09:33.416 So that's also helpful for

NOTE Confidence: 0.937590552222222

01:09:33.416 --> 01:09:35.240 machine learning purposes.

NOTE Confidence: 0.937590552222222

01:09:35.240 --> 01:09:38.496 OK, Finally I just wanted to give you

NOTE Confidence: 0.937590552222222

01:09:38.496 --> 01:09:41.581 some sense of how you do topology for

NOTE Confidence: 0.937590552222222

01:09:41.581 --> 01:09:45.077 data that is not point cloud based data.

NOTE Confidence: 0.937590552222222

01:09:45.080 --> 01:09:47.216 What if you have like images

NOTE Confidence: 0.937590552222222

01:09:47.216 --> 01:09:49.519 that you want to work with?

NOTE Confidence: 0.937590552222222

01:09:49.520 --> 01:09:52.080 You can compute topology

NOTE Confidence: 0.937590552222222

01:09:52.080 --> 01:09:54.000 directly from images.

NOTE Confidence: 0.937590552222222

01:09:54.000 --> 01:09:56.100 So think of an image as nothing

NOTE Confidence: 0.937590552222222

01:09:56.100 --> 01:09:58.083 but a matrix of values, right?

NOTE Confidence: 0.937590552222222

01:09:58.083 --> 01:10:00.344 And the values are going to depict

NOTE Confidence: 0.937590552222222

01:10:00.344 --> 01:10:02.117 how bright a given pixel is.

NOTE Confidence: 0.937590552222222

01:10:02.120 --> 01:10:03.440 So if I have one here,

NOTE Confidence: 0.937590552222222

01:10:03.440 --> 01:10:05.200 that pixel is quite dark,

NOTE Confidence: 0.937590552222222

01:10:05.200 --> 01:10:07.408 5 is brighter, three.

NOTE Confidence: 0.937590552222222

01:10:07.408 --> 01:10:07.960 Sorry,

NOTE Confidence: 0.937590552222222

01:10:07.960 --> 01:10:10.424 this is not really the color here

NOTE Confidence: 0.937590552222222

01:10:10.424 --> 01:10:12.318 doesn't really correspond to the value,

NOTE Confidence: 0.937590552222222

01:10:12.320 --> 01:10:14.399 but five would be a brighter pixel.

NOTE Confidence: 0.937590552222222

01:10:14.400 --> 01:10:16.199 Three would be slightly dimmer than five,

NOTE Confidence: 0.937590552222222

01:10:16.200 --> 01:10:17.280 but brighter than one.

NOTE Confidence: 0.937590552222222

01:10:17.280 --> 01:10:19.968 So you can think of an image as nothing

NOTE Confidence: 0.937590552222222

01:10:19.968 --> 01:10:22.320 but a matrix of image intensity values.

NOTE Confidence: 0.937590552222222

01:10:22.320 --> 01:10:25.576 And you can perform the same kind of

NOTE Confidence: 0.937590552222222

01:10:25.576 --> 01:10:28.293 filtration that we did previously by

NOTE Confidence: 0.937590552222222

01:10:28.293 --> 01:10:31.017 expanding that epsilon radius disk by

NOTE Confidence: 0.937590552222222

01:10:31.104 --> 01:10:34.206 going through this matrix of values

NOTE Confidence: 0.937590552222222

01:10:34.206 --> 01:10:36.610 and simply deleting everything that's

NOTE Confidence: 0.937590552222222

01:10:36.610 --> 01:10:38.920 above a value or below a value.

NOTE Confidence: 0.937590552222222

01:10:38.920 --> 01:10:40.894 And these are called sub level set

NOTE Confidence: 0.937590552222222

01:10:40.894 --> 01:10:42.719 or super level set filtrations,

NOTE Confidence: 0.937590552222222

01:10:42.720 --> 01:10:45.004 where in this example,

NOTE Confidence: 0.937590552222222

01:10:45.004 --> 01:10:47.812 any value that's only values that are

NOTE Confidence: 0.937590552222222

01:10:47.812 --> 01:10:50.056 less than or equal to one are shown

NOTE Confidence: 0.937590552222222

01:10:50.056 --> 01:10:52.240 where we spot two holes in our data.

NOTE Confidence: 0.937590552222222

01:10:52.240 --> 01:10:54.520 So five and three become holes.

NOTE Confidence: 0.937590552222222

01:10:54.520 --> 01:10:57.278 Then you increase your threshold to three.

NOTE Confidence: 0.937590552222222

01:10:57.280 --> 01:10:59.116 So now three gets filled in,

NOTE Confidence: 0.937590552222222

01:10:59.120 --> 01:11:01.433 but there's only one hole in the data set.

NOTE Confidence: 0.937590552222222

01:11:01.440 --> 01:11:03.198 And then as you increase your

NOTE Confidence: 0.937590552222222

01:11:03.198 --> 01:11:05.004 threshold to five is then both

NOTE Confidence: 0.937590552222222

01:11:05.004 --> 01:11:06.720 of those holes get filled in.

NOTE Confidence: 0.937590552222222

01:11:06.720 --> 01:11:09.480 So when you're working with images,

NOTE Confidence: 0.937590552222222

01:11:09.480 --> 01:11:11.555 you can construct what are

NOTE Confidence: 0.937590552222222

01:11:11.555 --> 01:11:12.800 called cubicle complexes,

NOTE Confidence: 0.937590552222222

01:11:12.800 --> 01:11:14.730 where you define a threshold

NOTE Confidence: 0.937590552222222

01:11:14.730 --> 01:11:16.274 value for your image,

NOTE Confidence: 0.937590552222222

01:11:16.280 --> 01:11:18.704 and by applying that threshold you

NOTE Confidence: 0.937590552222222

01:11:18.704 --> 01:11:21.610 can count in your pixels how many

NOTE Confidence: 0.937590552222222

01:11:21.610 --> 01:11:24.438 holes exist and you can quantify the

NOTE Confidence: 0.937590552222222

01:11:24.516 --> 01:11:27.078 shape of an image in that manner.

NOTE Confidence: 0.937590552222222

01:11:27.080 --> 01:11:29.510 I'll show you how this can be used in a

NOTE Confidence: 0.937590552222222

01:11:29.578 --> 01:11:33.200 very powerful way in a subsequent workshop.

NOTE Confidence: 0.937590552222222

01:11:33.200 --> 01:11:36.593 I just want to point to the paper here.

NOTE Confidence: 0.937590552222222

01:11:36.600 --> 01:11:39.120 So there's a nudist paper came out

NOTE Confidence: 0.937590552222222

01:11:39.120 --> 01:11:41.944 in 2020 by Bastian Riek and and

NOTE Confidence: 0.937590552222222

01:11:41.944 --> 01:11:44.515 other folks in the topology community

NOTE Confidence: 0.937590552222222

01:11:44.515 --> 01:11:47.320 where they took F MRI images,

NOTE Confidence: 0.937590552222222

01:11:47.320 --> 01:11:49.798 which were volumetric F MRI images.

NOTE Confidence: 0.937590552222222

01:11:49.800 --> 01:11:51.914 So there's lots and lots of data.

NOTE Confidence: 0.937590552222222

01:11:51.920 --> 01:11:54.395 And they performed this cubicle

NOTE Confidence: 0.937590552222222

01:11:54.395 --> 01:11:56.375 complex filtration of these

NOTE Confidence: 0.937590552222222

01:11:56.375 --> 01:11:58.890 volumetric images to construct a

NOTE Confidence: 0.937590552222222

01:11:58.890 --> 01:12:00.834 sequence of persistence diagrams,

NOTE Confidence: 0.937590552222222

01:12:00.840 --> 01:12:03.918 which they converted into persistence images.

NOTE Confidence: 0.937590552222222

01:12:03.920 --> 01:12:05.800 And from these persistence images,

NOTE Confidence: 0.937590552222222

01:12:05.800 --> 01:12:07.792 they were able to use many

NOTE Confidence: 0.937590552222222

01:12:07.792 --> 01:12:09.120 for learning techniques to

NOTE Confidence: 0.82620466

01:12:09.191 --> 01:12:10.345 categorize different

NOTE Confidence: 0.82620466

01:12:10.345 --> 01:12:12.076 brain state trajectories.

NOTE Confidence: 0.82620466

01:12:12.080 --> 01:12:14.440 So this is a very cool paper that

NOTE Confidence: 0.82620466

01:12:14.440 --> 01:12:16.951 combines a lot of the stuff that we

NOTE Confidence: 0.82620466

01:12:16.951 --> 01:12:19.610 talked about of going from images to

NOTE Confidence: 0.82620466

01:12:19.610 --> 01:12:21.410 persistence diagrams to persistence

NOTE Confidence: 0.82620466

01:12:21.410 --> 01:12:23.957 images and those images being used

NOTE Confidence: 0.82620466

01:12:23.957 --> 01:12:26.351 as input for machine learning to

NOTE Confidence: 0.82620466

01:12:26.351 --> 01:12:28.518 classify brain state trajectories.

NOTE Confidence: 0.82620466

01:12:28.520 --> 01:12:30.795 You can also take directly

NOTE Confidence: 0.82620466

01:12:30.795 --> 01:12:32.160 the persistence diagram,

NOTE Confidence: 0.82620466

01:12:32.160 --> 01:12:34.310 compute summaries of the persistence

NOTE Confidence: 0.82620466

01:12:34.310 --> 01:12:36.030 diagram such as persistence

NOTE Confidence: 0.82620466

01:12:36.030 --> 01:12:37.840 landscapes and persistence curves.

NOTE Confidence: 0.82620466

01:12:37.840 --> 01:12:39.796 This is a persistence curve here,

NOTE Confidence: 0.82620466

01:12:39.800 --> 01:12:42.038 and you can use those persistence

NOTE Confidence: 0.82620466

01:12:42.038 --> 01:12:43.993 curves directly to perform regression

NOTE Confidence: 0.82620466

01:12:43.993 --> 01:12:46.345 tasks such as estimating the severity

NOTE Confidence: 0.82620466

01:12:46.345 --> 01:12:48.959 of the disease in these F MRI images.

NOTE Confidence: 0.898670131

01:12:51.240 --> 01:12:53.568 Lastly, something that's going to be

NOTE Confidence: 0.898670131

01:12:53.568 --> 01:12:56.274 highly relevant to us going forward

NOTE Confidence: 0.898670131

01:12:56.274 --> 01:13:00.313 is doing TDA on time series data.

NOTE Confidence: 0.898670131

01:13:00.320 --> 01:13:02.020 So here's an example of

NOTE Confidence: 0.898670131

01:13:02.020 --> 01:13:03.720 a time series data set.

NOTE Confidence: 0.898670131

01:13:03.720 --> 01:13:05.920 These are just two sinusoidal

NOTE Confidence: 0.898670131

01:13:05.920 --> 01:13:07.503 curves, F1 and F2.

NOTE Confidence: 0.898670131

01:13:07.503 --> 01:13:11.280 You can see F1 has a higher amplitude,

NOTE Confidence: 0.898670131

01:13:11.280 --> 01:13:14.675 F2 has a smaller amplitude over time.

NOTE Confidence: 0.898670131

01:13:14.680 --> 01:13:16.320 And So what you can do is you

NOTE Confidence: 0.898670131

01:13:16.320 --> 01:13:18.077 can plot them against each other.

NOTE Confidence: 0.898670131

01:13:18.080 --> 01:13:20.754 So you can plot F1 against F2.

NOTE Confidence: 0.898670131

01:13:20.760 --> 01:13:22.728 And this is one way of

NOTE Confidence: 0.898670131

01:13:22.728 --> 01:13:24.040 taking time series data.

NOTE Confidence: 0.898670131

01:13:24.040 --> 01:13:26.400 You have to discretize in time of course,

NOTE Confidence: 0.898670131

01:13:26.400 --> 01:13:27.840 and converting them into

NOTE Confidence: 0.898670131

01:13:27.840 --> 01:13:29.640 a point cloud data set.

NOTE Confidence: 0.898670131

01:13:29.640 --> 01:13:31.710 And you can compute topology directly

NOTE Confidence: 0.898670131

01:13:31.710 --> 01:13:33.679 from this point cloud data set.

NOTE Confidence: 0.898670131

01:13:33.680 --> 01:13:35.619 So this works when you have two

NOTE Confidence: 0.898670131

01:13:35.619 --> 01:13:37.793 time series data sets, F1 and F2.

NOTE Confidence: 0.898670131

01:13:37.793 --> 01:13:40.640 You can convert that into a point cloud.

NOTE Confidence: 0.898670131

01:13:40.640 --> 01:13:43.952 If you have just one time series data set,

NOTE Confidence: 0.898670131

01:13:43.960 --> 01:13:46.084 what you can do is you can do a

NOTE Confidence: 0.898670131

01:13:46.084 --> 01:13:47.720 sliding window transformation.

NOTE Confidence: 0.898670131

01:13:47.720 --> 01:13:49.876 So you take a small sliding window,

NOTE Confidence: 0.898670131

01:13:49.880 --> 01:13:51.840 so a small chunk of the data,

NOTE Confidence: 0.898670131

01:13:51.840 --> 01:13:55.116 move that window forward 1 by 1 by 1.

NOTE Confidence: 0.898670131

01:13:55.120 --> 01:13:57.000 And for within that window,

NOTE Confidence: 0.898670131

01:13:57.000 --> 01:13:59.562 you can construct this phase portrait or

NOTE Confidence: 0.898670131

01:13:59.562 --> 01:14:02.358 this time delay embedding as it's called.

NOTE Confidence: 0.898670131

01:14:02.360 --> 01:14:05.069 And then you can take that loop and you

NOTE Confidence: 0.898670131

01:14:05.069 --> 01:14:08.236 can convert that into a persistence diagram.

NOTE Confidence: 0.898670131

01:14:08.240 --> 01:14:10.781 So I wanted to show you some

NOTE Confidence: 0.898670131

01:14:10.781 --> 01:14:13.079 examples of kind of doing this,

NOTE Confidence: 0.898670131

01:14:13.080 --> 01:14:13.613 right.

NOTE Confidence: 0.898670131

01:14:13.613 --> 01:14:16.811 So this is combining both cubicle

NOTE Confidence: 0.898670131

01:14:16.811 --> 01:14:20.056 homology and this time delay embedding

NOTE Confidence: 0.898670131

01:14:20.056 --> 01:14:23.320 the sliding window embedding and using

NOTE Confidence: 0.898670131

01:14:23.411 --> 01:14:26.556 that to compute persistence diagrams.

NOTE Confidence: 0.898670131

01:14:26.560 --> 01:14:28.254 And so again, I'm not going to

NOTE Confidence: 0.898670131

01:14:28.254 --> 01:14:29.958 go through all the details here,

NOTE Confidence: 0.898670131

01:14:29.960 --> 01:14:32.032 but I just wanted to show you an

NOTE Confidence: 0.898670131

01:14:32.032 --> 01:14:33.718 an application of this in practice.

NOTE Confidence: 0.898670131

01:14:33.720 --> 01:14:35.799 So this is a paper which was

NOTE Confidence: 0.898670131

01:14:35.799 --> 01:14:37.120 on archive in 2018.

NOTE Confidence: 0.898670131

01:14:37.120 --> 01:14:38.640 It might be out already.

NOTE Confidence: 0.898670131

01:14:38.640 --> 01:14:40.440 I hope it's out by now.

NOTE Confidence: 0.898670131

01:14:40.440 --> 01:14:42.239 And in the, in this data set,

NOTE Confidence: 0.898670131

01:14:42.240 --> 01:14:43.662 they were looking,

NOTE Confidence: 0.898670131

01:14:43.662 --> 01:14:46.032 they were imaging the vocal

NOTE Confidence: 0.898670131

01:14:46.032 --> 01:14:48.798 cords of humans as they were,

NOTE Confidence: 0.898670131

01:14:48.800 --> 01:14:51.520 they were making some sounds.

NOTE Confidence: 0.898670131

01:14:51.520 --> 01:14:53.680 And so when you're making like

NOTE Confidence: 0.898670131

01:14:53.680 --> 01:14:55.680 a rhythmic pattern of sounds,

NOTE Confidence: 0.898670131

01:14:55.680 --> 01:14:58.380 your vocal cords, they open, they close,

NOTE Confidence: 0.898670131

01:14:58.380 --> 01:15:00.480 they open and then they close.

NOTE Confidence: 0.898670131

01:15:00.480 --> 01:15:02.520 And so in this data set,

NOTE Confidence: 0.898670131

01:15:02.520 --> 01:15:04.458 obviously there's a periodic nature to

NOTE Confidence: 0.898670131

01:15:04.458 --> 01:15:06.799 the type of sound you're producing.

NOTE Confidence: 0.898670131

01:15:06.800 --> 01:15:08.571 And so if you look at these

NOTE Confidence: 0.898670131

01:15:08.571 --> 01:15:10.933 images of the vocal cords and you

NOTE Confidence: 0.898670131

01:15:10.933 --> 01:15:12.473 compute image self similarity,

NOTE Confidence: 0.898670131

01:15:12.480 --> 01:15:14.461 you can kind of guess that there

NOTE Confidence: 0.898670131

01:15:14.461 --> 01:15:16.544 is a period after which the

NOTE Confidence: 0.898670131

01:15:16.544 --> 01:15:18.479 image becomes similar to itself.

NOTE Confidence: 0.898670131

01:15:18.480 --> 01:15:20.856 And so you can kind of quantify the

NOTE Confidence: 0.898670131

01:15:20.856 --> 01:15:22.757 periodicity of the data in this way.

NOTE Confidence: 0.898670131

01:15:22.760 --> 01:15:23.424 But also,

NOTE Confidence: 0.898670131

01:15:23.424 --> 01:15:25.748 if you take the sequence of images

NOTE Confidence: 0.898670131

01:15:25.748 --> 01:15:28.724 and you do this time delay embedding

NOTE Confidence: 0.898670131

01:15:28.724 --> 01:15:30.440 and compute cubicle homology,

NOTE Confidence: 0.898670131

01:15:30.440 --> 01:15:32.799 you can end up with a persistence

NOTE Confidence: 0.898670131

01:15:32.799 --> 01:15:34.587 diagram where very clearly you

NOTE Confidence: 0.898670131

01:15:34.587 --> 01:15:36.723 see this H1 feature which tells

NOTE Confidence: 0.898670131

01:15:36.723 --> 01:15:38.799 you there's a loop in your data,

NOTE Confidence: 0.898670131

01:15:38.800 --> 01:15:42.034 which means that the data is periodic.

NOTE Confidence: 0.898670131

01:15:42.040 --> 01:15:44.812 What's cooler is when you do

NOTE Confidence: 0.898670131

01:15:44.812 --> 01:15:46.198 what's called bifonation.

NOTE Confidence: 0.898670131

01:15:46.200 --> 01:15:47.193 So in bifonation,

NOTE Confidence: 0.898670131

01:15:47.193 --> 01:15:49.510 the vocal cords move in a way

NOTE Confidence: 0.898670131

01:15:49.582 --> 01:15:51.727 that they produce two different

NOTE Confidence: 0.898670131

01:15:51.727 --> 01:15:53.872 frequencies at the same time.

NOTE Confidence: 0.967520644285714

01:15:53.880 --> 01:15:57.156 So you have like a high frequency,

NOTE Confidence: 0.967520644285714

01:15:57.160 --> 01:15:58.344 a sound coming out,

NOTE Confidence: 0.967520644285714

01:15:58.344 --> 01:16:00.909 and at the same time you have like

NOTE Confidence: 0.967520644285714

01:16:00.909 --> 01:16:03.075 a low frequency whine coming out.

NOTE Confidence: 0.967520644285714

01:16:03.080 --> 01:16:04.970 And I was thinking if I can do a

NOTE Confidence: 0.967520644285714

01:16:04.970 --> 01:16:06.237 demonstration of this kind of voice,

NOTE Confidence: 0.967520644285714

01:16:06.240 --> 01:16:07.956 but I really cannot do it.

NOTE Confidence: 0.967520644285714

01:16:07.960 --> 01:16:08.940 So you have to.

NOTE Confidence: 0.967520644285714

01:16:08.940 --> 01:16:10.840 If you search online for Bifor Nation,

NOTE Confidence: 0.967520644285714

01:16:10.840 --> 01:16:12.954 you will find examples of people who

NOTE Confidence: 0.967520644285714

01:16:12.954 --> 01:16:14.604 can produce both high frequencies

NOTE Confidence: 0.967520644285714

01:16:14.604 --> 01:16:16.998 and low frequencies at the same time.

NOTE Confidence: 0.967520644285714

01:16:17.000 --> 01:16:19.664 And if you look at the vocal cord

NOTE Confidence: 0.967520644285714

01:16:19.664 --> 01:16:21.998 images of producing this kind of sound,

NOTE Confidence: 0.967520644285714

01:16:22.000 --> 01:16:23.672 this is kind of what it looks like

NOTE Confidence: 0.967520644285714

01:16:23.672 --> 01:16:25.320 when you look at self similarity.

NOTE Confidence: 0.967520644285714

01:16:25.320 --> 01:16:27.720 You do observe a pattern here.

NOTE Confidence: 0.967520644285714

01:16:27.720 --> 01:16:28.743 But very importantly,

NOTE Confidence: 0.967520644285714

01:16:28.743 --> 01:16:32.017 when you take this data and you plug it

NOTE Confidence: 0.967520644285714

01:16:32.017 --> 01:16:34.477 through the techniques that I've described,

NOTE Confidence: 0.967520644285714

01:16:34.480 --> 01:16:35.900 you get a persistence

NOTE Confidence: 0.967520644285714

01:16:35.900 --> 01:16:37.675 diagram that looks like this,

NOTE Confidence: 0.967520644285714

01:16:37.680 --> 01:16:42.360 which has 2H1 features and one H2 feature.

NOTE Confidence: 0.967520644285714

01:16:42.360 --> 01:16:43.786 So remember,

NOTE Confidence: 0.967520644285714

01:16:43.786 --> 01:16:48.272 H1 is is the dimension 1 hole and

NOTE Confidence: 0.967520644285714

01:16:48.272 --> 01:16:51.440 H2 is dimension 2 hole or a void.

NOTE Confidence: 0.967520644285714

01:16:51.440 --> 01:16:53.120 And so from our previous quiz,

NOTE Confidence: 0.967520644285714

01:16:53.120 --> 01:16:56.032 hopefully you recall that 2H1 holes and

NOTE Confidence: 0.967520644285714

01:16:56.032 --> 01:16:59.317 one H2 hole means it's like a torus,

NOTE Confidence: 0.967520644285714

01:16:59.320 --> 01:17:01.567 which means there's empty hole in the

NOTE Confidence: 0.967520644285714

01:17:01.567 --> 01:17:04.195 middle and there are two loops in the torus.

NOTE Confidence: 0.967520644285714

01:17:04.200 --> 01:17:06.186 And that makes perfect sense for

NOTE Confidence: 0.967520644285714

01:17:06.186 --> 01:17:08.547 this data set because you have a

NOTE Confidence: 0.967520644285714

01:17:08.547 --> 01:17:10.497 high frequency and a low frequency

NOTE Confidence: 0.967520644285714

01:17:10.497 --> 01:17:11.678 forming 2 loops here.

NOTE Confidence: 0.967520644285714

01:17:11.680 --> 01:17:13.312 And then you have because it's

NOTE Confidence: 0.967520644285714

01:17:13.312 --> 01:17:14.400 arranged like a Taurus,

NOTE Confidence: 0.967520644285714

01:17:14.400 --> 01:17:15.816 both of those things are happening

NOTE Confidence: 0.967520644285714

01:17:15.816 --> 01:17:16.760 at the same time.

NOTE Confidence: 0.967520644285714

01:17:16.760 --> 01:17:18.992 You get a dimension to all in the

NOTE Confidence: 0.967520644285714

01:17:18.992 --> 01:17:21.076 data set or as it says in here,

NOTE Confidence: 0.967520644285714

01:17:21.080 --> 01:17:24.720 to a two cycle in the data set.

NOTE Confidence: 0.967520644285714

01:17:24.720 --> 01:17:27.320 And this again is from the same paper.

NOTE Confidence: 0.967520644285714

01:17:27.320 --> 01:17:31.246 This is an example where the person is

NOTE Confidence: 0.967520644285714

01:17:31.246 --> 01:17:34.076 showing irregular vocal fold vibrations.

NOTE Confidence: 0.967520644285714

01:17:34.080 --> 01:17:36.080 So there is no periodicity,

NOTE Confidence: 0.967520644285714

01:17:36.080 --> 01:17:37.880 no quasi periodicity.

NOTE Confidence: 0.967520644285714

01:17:37.880 --> 01:17:39.680 It appears random.

NOTE Confidence: 0.967520644285714

01:17:39.680 --> 01:17:42.074 When you look at image self similarity,

NOTE Confidence: 0.967520644285714

01:17:42.080 --> 01:17:44.156 it just goes along the diagonal.

NOTE Confidence: 0.967520644285714

01:17:44.160 --> 01:17:46.580 You don't see a lot of like important

NOTE Confidence: 0.967520644285714

01:17:46.580 --> 01:17:49.080 self similarity off the diagonal,

NOTE Confidence: 0.967520644285714

01:17:49.080 --> 01:17:50.739 which means that all of these images

NOTE Confidence: 0.967520644285714

01:17:50.739 --> 01:17:52.757 look kind of different from each other.

NOTE Confidence: 0.967520644285714

01:17:52.760 --> 01:17:55.032 And if you throw this into TDA,

NOTE Confidence: 0.967520644285714

01:17:55.032 --> 01:17:57.192 you get topological features that

NOTE Confidence: 0.967520644285714

01:17:57.192 --> 01:17:59.999 are very close to the diagonal.

NOTE Confidence: 0.967520644285714

01:18:00.000 --> 01:18:00.584 Again,

NOTE Confidence: 0.967520644285714

01:18:00.584 --> 01:18:03.285 you can compute like confidence

NOTE Confidence: 0.967520644285714

01:18:03.285 --> 01:18:05.840 intervals and so forth for these things.

NOTE Confidence: 0.967520644285714

01:18:05.840 --> 01:18:06.642 But again,

NOTE Confidence: 0.967520644285714

01:18:06.642 --> 01:18:08.647 it shows there's no interesting

NOTE Confidence: 0.967520644285714

01:18:08.647 --> 01:18:10.692 topology happening in this data

NOTE Confidence: 0.967520644285714

01:18:10.692 --> 01:18:12.240 set because it's irregular.

NOTE Confidence: 0.967520644285714

01:18:12.240 --> 01:18:13.768 There's no quasi periodicity

NOTE Confidence: 0.967520644285714

01:18:13.768 --> 01:18:15.678 or periodicity in this data.

NOTE Confidence: 0.9470462

01:18:18.520 --> 01:18:21.695 Lastly, topology is invertible

NOTE Confidence: 0.9470462

01:18:21.695 --> 01:18:23.475 to a certain extent.

NOTE Confidence: 0.9470462

01:18:23.480 --> 01:18:26.033 So folks often ask like, OK,

NOTE Confidence: 0.9470462

01:18:26.033 --> 01:18:28.398 I created this persistence diagram.

NOTE Confidence: 0.9470462

01:18:28.400 --> 01:18:31.459 I want to interpret the where these

NOTE Confidence: 0.9470462

01:18:31.459 --> 01:18:33.200 topological features come from.

NOTE Confidence: 0.9470462

01:18:33.200 --> 01:18:36.084 And you can do that using something

NOTE Confidence: 0.9470462

01:18:36.084 --> 01:18:37.320 called cycle representatives.

NOTE Confidence: 0.9470462

01:18:37.320 --> 01:18:38.668 And what cycle representatives

NOTE Confidence: 0.9470462

01:18:38.668 --> 01:18:41.932 allow you to do is they allow you to

NOTE Confidence: 0.9470462

01:18:41.932 --> 01:18:43.504 interrogate a specific topological

NOTE Confidence: 0.9470462

01:18:43.504 --> 01:18:45.560 feature and ask the question,

NOTE Confidence: 0.9470462

01:18:45.560 --> 01:18:47.420 where does that feature come

NOTE Confidence: 0.9470462

01:18:47.420 --> 01:18:49.640 from in your input data set?

NOTE Confidence: 0.9470462

01:18:49.640 --> 01:18:50.669 So for example,

NOTE Confidence: 0.9470462

01:18:50.669 --> 01:18:53.070 if we have this persistence diagram that's

NOTE Confidence: 0.9470462

01:18:53.132 --> 01:18:55.435 derived from this point cloud data set,

NOTE Confidence: 0.9470462

01:18:55.440 --> 01:18:57.495 you can then interrogate this

NOTE Confidence: 0.9470462

01:18:57.495 --> 01:18:59.550 topological feature and this it

NOTE Confidence: 0.9470462

01:18:59.621 --> 01:19:01.793 will tell you that this dimension

NOTE Confidence: 0.9470462

01:19:01.793 --> 01:19:04.029 0 feature appears here because

NOTE Confidence: 0.9470462

01:19:04.029 --> 01:19:07.047 these two clusters of data became

NOTE Confidence: 0.9470462

01:19:07.047 --> 01:19:09.358 connected at that epsilon value.

NOTE Confidence: 0.9470462

01:19:09.360 --> 01:19:11.415 So the these two connected

NOTE Confidence: 0.9470462

01:19:11.415 --> 01:19:12.237 components disappeared,

NOTE Confidence: 0.9470462

01:19:12.240 --> 01:19:15.796 they merged together at that epsilon value.

NOTE Confidence: 0.9470462

01:19:15.800 --> 01:19:17.880 And likewise for dimension one,

NOTE Confidence: 0.9470462

01:19:17.880 --> 01:19:19.890 you can interrogate that topological

NOTE Confidence: 0.9470462

01:19:19.890 --> 01:19:22.322 feature and it will tell you

NOTE Confidence: 0.9470462

01:19:22.322 --> 01:19:24.067 that that particular loop is

NOTE Confidence: 0.9470462

01:19:24.067 --> 01:19:26.000 formed by these four points.

NOTE Confidence: 0.9470462

01:19:26.000 --> 01:19:26.999 This is very,

NOTE Confidence: 0.9470462

01:19:26.999 --> 01:19:29.330 very important for us because when we

NOTE Confidence: 0.9470462

01:19:29.401 --> 01:19:32.005 are dealing with the state space of

NOTE Confidence: 0.9470462

01:19:32.005 --> 01:19:34.240 cellular activity and neural activity,

NOTE Confidence: 0.9470462

01:19:34.240 --> 01:19:36.540 having access to these cycle

NOTE Confidence: 0.9470462

01:19:36.540 --> 01:19:38.840 representatives will allow us to,

NOTE Confidence: 0.9470462

01:19:38.840 --> 01:19:41.332 will will give us the ability to

NOTE Confidence: 0.9470462

01:19:41.332 --> 01:19:43.655 say which time points and which

NOTE Confidence: 0.9470462

01:19:43.655 --> 01:19:46.013 parts of the brain precisely led

NOTE Confidence: 0.9470462

01:19:46.013 --> 01:19:49.467 to the formation of a cycle which

NOTE Confidence: 0.9470462

01:19:49.467 --> 01:19:50.955 indicates periodic activity.

NOTE Confidence: 0.9470462

01:19:50.960 --> 01:19:54.376 So we can indeed go back in reverse

NOTE Confidence: 0.9470462

01:19:54.376 --> 01:19:56.627 from topological features back to

NOTE Confidence: 0.9470462

01:19:56.627 --> 01:19:59.243 our original data set and figure

NOTE Confidence: 0.9470462

01:19:59.243 --> 01:20:01.697 out why a certain topological

NOTE Confidence: 0.9470462

01:20:01.697 --> 01:20:04.117 feature exists in our data.

NOTE Confidence: 0.9470462

01:20:04.120 --> 01:20:06.730 And so I think we are at a stage here

NOTE Confidence: 0.9470462

01:20:06.805 --> 01:20:09.559 where we don't have a lot of time left.

NOTE Confidence: 0.9470462

01:20:09.560 --> 01:20:11.720 We're supposed to end at 5:30.

NOTE Confidence: 0.9470462

01:20:11.720 --> 01:20:13.960 So I'm not going to go through the ML parts.

NOTE Confidence: 0.9470462

01:20:13.960 --> 01:20:15.598 I had a few ML slides,

NOTE Confidence: 0.9470462

01:20:15.600 --> 01:20:17.760 but I think we can punt that to

NOTE Confidence: 0.9470462

01:20:17.760 --> 01:20:18.960 our third workshop.

NOTE Confidence: 0.9470462

01:20:18.960 --> 01:20:20.856 I think it would be a good time

NOTE Confidence: 0.9470462

01:20:20.856 --> 01:20:22.798 to take questions and end here.

NOTE Confidence: 0.9470462

01:20:22.800 --> 01:20:24.200 I just wanted to mention if you're,

NOTE Confidence: 0.9470462

01:20:24.200 --> 01:20:25.280 if you're about to leave,

NOTE Confidence: 0.9470462

01:20:25.280 --> 01:20:27.620 we're going to have another workshop

NOTE Confidence: 0.9470462

01:20:27.620 --> 01:20:30.222 next week that will be given by

NOTE Confidence: 0.9470462

01:20:30.222 --> 01:20:32.469 Rahul and he'll be telling you how

NOTE Confidence: 0.9470462

01:20:32.540 --> 01:20:34.934 we can take a graph consisting of

NOTE Confidence: 0.9470462

01:20:34.934 --> 01:20:37.888 nodes and edges and use graph signal

NOTE Confidence: 0.9470462

01:20:37.888 --> 01:20:40.298 processing to quantify how some

NOTE Confidence: 0.9470462

01:20:40.298 --> 01:20:43.238 signal is distributed on that graph.

NOTE Confidence: 0.9470462

01:20:43.240 --> 01:20:45.120 And then the following week,

NOTE Confidence: 0.9470462

01:20:45.120 --> 01:20:45.999 me and Brian,

NOTE Confidence: 0.9470462

01:20:45.999 --> 01:20:47.757 we're going to put all of

NOTE Confidence: 0.9470462

01:20:47.757 --> 01:20:49.159 these things together,

NOTE Confidence: 0.9470462

01:20:49.160 --> 01:20:52.040 talk about GSTH as a technique

NOTE Confidence: 0.9470462

01:20:52.040 --> 01:20:53.825 all combined together and we'll

NOTE Confidence: 0.9470462

01:20:53.825 --> 01:20:56.404 show you how we have used this

NOTE Confidence: 0.9470462

01:20:56.404 --> 01:20:58.917 technique with some of our data sets.

NOTE Confidence: 0.9470462

01:20:58.920 --> 01:21:00.078 Thanks for listening.

NOTE Confidence: 0.9470462

01:21:00.078 --> 01:21:01.236 Thanks for coming.

NOTE Confidence: 0.9470462

01:21:01.240 --> 01:21:01.520 Jay,

NOTE Confidence: 0.835997186363636

01:21:01.520 --> 01:21:02.592 I have a question.

NOTE Confidence: 0.835997186363636

01:21:02.592 --> 01:21:04.920 It may be a little bit premature,

NOTE Confidence: 0.835997186363636

01:21:04.920 --> 01:21:07.495 but is my intuition just

NOTE Confidence: 0.835997186363636

01:21:07.495 --> 01:21:09.040 correct my intuition?

NOTE Confidence: 0.835997186363636

01:21:09.040 --> 01:21:11.000 Do you understand correctly that

NOTE Confidence: 0.835997186363636

01:21:11.000 --> 01:21:12.960 if you have more organised,

NOTE Confidence: 0.835997186363636

01:21:12.960 --> 01:21:16.372 more complexly organized systems,

NOTE Confidence: 0.835997186363636

01:21:16.372 --> 01:21:21.137 you should see higher level holes?

NOTE Confidence: 0.835997186363636

01:21:21.137 --> 01:21:24.119 And if it's mostly random noise,

NOTE Confidence: 0.835997186363636

01:21:24.120 --> 01:21:27.320 you kind of don't really see much? Yeah.

NOTE Confidence: 0.965181258333333

01:21:27.320 --> 01:21:30.390 So if you have random noise, then in space,

NOTE Confidence: 0.965181258333333

01:21:30.390 --> 01:21:32.792 everything will get filled in, right?

NOTE Confidence: 0.965181258333333

01:21:32.792 --> 01:21:34.200 It's all just noise.

NOTE Confidence: 0.965181258333333

01:21:34.200 --> 01:21:36.972 So there'll be no structure to the data set

NOTE Confidence: 0.965181258333333

01:21:36.972 --> 01:21:39.714 and you won't see any holes in the data.

NOTE Confidence: 0.965181258333333

01:21:39.720 --> 01:21:40.878 The connectivity pattern

NOTE Confidence: 0.965181258333333

01:21:40.878 --> 01:21:42.036 also looks different.

NOTE Confidence: 0.965181258333333

01:21:42.040 --> 01:21:45.144 So you can do statistical tests where you

NOTE Confidence: 0.965181258333333

01:21:45.144 --> 01:21:48.585 can take real data from an experiment and

NOTE Confidence: 0.965181258333333

01:21:48.585 --> 01:21:50.954 compare that with topological features

NOTE Confidence: 0.965181258333333

01:21:50.954 --> 01:21:53.639 derived from like standard distributions,

NOTE Confidence: 0.965181258333333

01:21:53.640 --> 01:21:54.840 like uniform distribution

NOTE Confidence: 0.965181258333333

01:21:54.840 --> 01:21:56.040 and Gaussian distribution,

NOTE Confidence: 0.965181258333333

01:21:56.040 --> 01:21:58.200 and it will tell you that in your

NOTE Confidence: 0.965181258333333

01:21:58.200 --> 01:21:59.718 experiment in the state space,

NOTE Confidence: 0.965181258333333

01:21:59.720 --> 01:22:01.680 it kind of looks like a uniform distribution.

NOTE Confidence: 0.965181258333333

01:22:01.680 --> 01:22:04.160 There's no structure to it. Yeah.

NOTE Confidence: 0.965181258333333

01:22:04.160 --> 01:22:06.160 So that's that's one aspect.

NOTE Confidence: 0.965181258333333

01:22:06.160 --> 01:22:08.218 Also like dimension zero will tell you

NOTE Confidence: 0.965181258333333

01:22:08.218 --> 01:22:11.410 if in your trajectory here you have two

NOTE Confidence: 0.965181258333333

01:22:11.410 --> 01:22:12.999 different connected components, right.

NOTE Confidence: 0.965181258333333

01:22:12.999 --> 01:22:14.911 So if you have like one like set

NOTE Confidence: 0.965181258333333

01:22:14.911 --> 01:22:16.797 of States and then a completely

NOTE Confidence: 0.965181258333333

01:22:16.797 --> 01:22:18.748 different set of States and they're

NOTE Confidence: 0.965181258333333

01:22:18.748 --> 01:22:20.834 kind of far apart from each other.

NOTE Confidence: 0.965181258333333

01:22:20.840 --> 01:22:22.688 That's what we learned from like

NOTE Confidence: 0.965181258333333

01:22:22.688 --> 01:22:24.919 dimension 0 homology in addition to like,

NOTE Confidence: 0.965181258333333

01:22:24.920 --> 01:22:25.506 you know,

NOTE Confidence: 0.965181258333333

01:22:25.506 --> 01:22:27.557 noise and how the data is distributed.

NOTE Confidence: 0.965181258333333

01:22:27.560 --> 01:22:29.975 And dimension one will tell us these

NOTE Confidence: 0.965181258333333

01:22:29.975 --> 01:22:31.724 periodic loop like structures that

NOTE Confidence: 0.965181258333333

01:22:31.724 --> 01:22:34.332 might exist in our data and also the

NOTE Confidence: 0.965181258333333

01:22:34.397 --> 01:22:36.653 empty spaces being states that cannot

NOTE Confidence: 0.965181258333333

01:22:36.653 --> 01:22:39.165 really exist based off of like this

NOTE Confidence: 0.965181258333333

01:22:39.165 --> 01:22:40.840 this experimental data again being

NOTE Confidence: 0.965181258333333

01:22:40.840 --> 01:22:42.772 cognizant of the earlier question where

NOTE Confidence: 0.965181258333333

01:22:42.772 --> 01:22:45.200 if your data is not sampled correctly,

NOTE Confidence: 0.965181258333333

01:22:45.200 --> 01:22:46.720 it might be telling you the wrong thing.

NOTE Confidence: 0.911909383333333

01:22:49.200 --> 01:22:54.205 OK, thank you. Do you have any other

NOTE Confidence: 0.911909383333333

01:22:54.205 --> 01:22:56.292 questions in the chat or if anybody

NOTE Confidence: 0.911909383333333

01:22:56.292 --> 01:22:58.838 wants to ask any follow up questions?

NOTE Confidence: 0.878974094285714

01:23:01.160 --> 01:23:02.959 All right, I don't see any questions.

NOTE Confidence: 0.878974094285714

01:23:02.960 --> 01:23:04.672 Thank you so much.

NOTE Confidence: 0.878974094285714

01:23:04.672 --> 01:23:07.500 Just a note, all these papers that

NOTE Confidence: 0.878974094285714

01:23:07.500 --> 01:23:09.425 you mentioned in the presentation

NOTE Confidence: 0.878974094285714

01:23:09.425 --> 01:23:11.848 I'm going to add to your website.

NOTE Confidence: 0.878974094285714

01:23:11.848 --> 01:23:13.658 So if people are interested

NOTE Confidence: 0.878974094285714

01:23:13.658 --> 01:23:15.799 in looking at those papers,

NOTE Confidence: 0.878974094285714

01:23:15.800 --> 01:23:20.560 there will be links on your maps site.

NOTE Confidence: 0.878974094285714

01:23:20.560 --> 01:23:21.568 And thank you again.

NOTE Confidence: 0.878974094285714

01:23:21.568 --> 01:23:23.160 And I'll see you next week. See

NOTE Confidence: 0.892415235

01:23:24.040 --> 01:23:27.000 you next week. See you. Bye.