WEBVTT

NOTE duration:"01:06:19.1570000"

NOTE language:en-us

NOTE Confidence: 0.83166105

00:00:00.000 --> 00:00:03.552 Topic of today's is about speaking

NOTE Confidence: 0.83166105

00:00:03.552 --> 00:00:05.328 about data visualization.

NOTE Confidence: 0.83166105

00:00:05.330 --> 00:00:08.795 And so it will be very in general on

NOTE Confidence: 0.83166105

00:00:08.795 --> 00:00:11.838 ramps how to design some strategy,

NOTE Confidence: 0.83166105

00:00:11.840 --> 00:00:15.165 some issues, some principles to guide in

NOTE Confidence: 0.83166105

00:00:15.165 --> 00:00:17.917 the visualization of the of our data.

NOTE Confidence: 0.8231227

00:00:20.080 --> 00:00:22.414 So data visualization is important to

NOTE Confidence: 0.8231227

00:00:22.414 --> 00:00:25.369 explore the data and this is particularly

NOTE Confidence: 0.8231227

00:00:25.369 --> 00:00:28.027 crucial since nowadays data are becoming

NOTE Confidence: 0.8231227

00:00:28.027 --> 00:00:30.818 much more complex and much more bigger,

NOTE Confidence: 0.8231227

00:00:30.820 --> 00:00:32.680 and so in general there

NOTE Confidence: 0.8231227

00:00:32.680 --> 00:00:35.360 is a rise of data science.

NOTE Confidence: 0.8231227

00:00:35.360 --> 00:00:37.430 So not only in research,

NOTE Confidence: 0.8231227

00:00:37.430 --> 00:00:40.100 not only in biological research.

NOTE Confidence: 0.8231227

00:00:40.100 --> 00:00:42.680 The second function to data

NOTE Confidence: 0.8231227

00:00:42.680 --> 00:00:44.744 visualization for data visualization

NOTE Confidence: 0.8231227

00:00:44.744 --> 00:00:47.195 is to communicate the data and

NOTE Confidence: 0.8231227

00:00:47.195 --> 00:00:50.050 that may be the most traditional.

NOTE Confidence: 0.8231227

00:00:50.050 --> 00:00:53.620 I'm so that's what that.

NOTE Confidence: 0.8231227

00:00:53.620 --> 00:00:55.270 Creating publication figures,

NOTE Confidence: 0.8231227

00:00:55.270 --> 00:00:56.370 for example,

NOTE Confidence: 0.8231227

00:00:56.370 --> 00:01:01.025 is about to communicate data to others

NOTE Confidence: 0.8231227

00:01:01.025 --> 00:01:04.358 because communicating data visually is

NOTE Confidence: 0.8231227

00:01:04.358 --> 00:01:08.656 more efficient than than words in general.

NOTE Confidence: 0.8231227

00:01:08.660 --> 00:01:13.300 So in order to represent complex data here,

NOTE Confidence: 0.8231227

00:01:13.300 --> 00:01:16.219 I collected 3.

NOTE Confidence: 0.8231227

00:01:16.220 --> 00:01:18.132 General challenges and aims.

NOTE Confidence: 0.8231227

00:01:18.132 --> 00:01:21.570 So whenever you plot the data is

NOTE Confidence: 0.8231227

00:01:21.570 --> 00:01:24.486 important that the plots are and

NOTE Confidence: 0.8231227

00:01:24.486 --> 00:01:27.019 the representations are are precise,

NOTE Confidence: 0.8231227

00:01:27.020 --> 00:01:28.457 so they're truthful.

NOTE Confidence: 0.8231227

00:01:28.457 --> 00:01:30.852 That means that distortion has

NOTE Confidence: 0.8231227

00:01:30.852 --> 00:01:34.370 to be avoided as much as possible

NOTE Confidence: 0.8231227

00:01:34.370 --> 00:01:36.350 is not always achievable,

NOTE Confidence: 0.8231227

00:01:36.350 --> 00:01:38.810 so distortion sometimes is unavoidable.

NOTE Confidence: 0.8231227

00:01:38.810 --> 00:01:40.762 Think about for example,

NOTE Confidence: 0.8231227

00:01:40.762 --> 00:01:43.690 when you plot the 2D Maps

NOTE Confidence: 0.8231227

00:01:43.789 --> 00:01:46.229 for representing 3D data.

NOTE Confidence: 0.8231227

00:01:46.230 --> 00:01:47.850 But the point is that the

NOTE Confidence: 0.8231227

00:01:47.850 --> 00:01:49.277 distortion doesn't have to convey

NOTE Confidence: 0.8231227

00:01:49.277 --> 00:01:50.707 the message of the figure,

NOTE Confidence: 0.8231227

00:01:50.710 --> 00:01:53.626 so it has to be something that is not

NOTE Confidence: 0.8231227

00:01:53.626 --> 00:01:56.547 related to the main message of the feature.

NOTE Confidence: 0.8231227

00:01:56.550 --> 00:01:58.466 Otherwise it's a problem.

NOTE Confidence: 0.8231227

00:01:58.466 --> 00:02:01.340 Then the second point is clarity.

NOTE Confidence: 0.8231227

00:02:01.340 --> 00:02:05.570 So data the figure has not to be ambiguous,

NOTE Confidence: 0.8231227

00:02:05.570 --> 00:02:08.860 and the third one is the efficiency.

NOTE Confidence: 0.8231227

00:02:08.860 --> 00:02:10.970 So every.

NOTE Confidence: 0.8231227

00:02:10.970 --> 00:02:14.029 Inca every in every pixel is precious,

NOTE Confidence: 0.8231227

00:02:14.030 --> 00:02:17.089 so each decision in doing your plotter,

NOTE Confidence: 0.8231227

00:02:17.090 --> 00:02:20.338 each decision on the color on the size

NOTE Confidence: 0.8231227

00:02:20.338 --> 00:02:23.947 on the number of layers said that you

NOTE Confidence: 0.8231227

00:02:23.947 --> 00:02:28.009 that you plotter is important and it has an.

NOTE Confidence: 0.8231227

00:02:28.010 --> 00:02:31.943 Everything has to be has to have a purpose,

NOTE Confidence: 0.8231227

00:02:31.950 --> 00:02:34.130 so you should reduce the

NOTE Confidence: 0.8231227

00:02:34.130 --> 00:02:35.874 so called chartjunk here.

NOTE Confidence: 0.8231227

00:02:35.880 --> 00:02:38.065 Below the slide you see

NOTE Confidence: 0.8231227

00:02:38.065 --> 00:02:39.376 quotation from Edward.

NOTE Confidence: 0.8231227

00:02:39.380 --> 00:02:42.537 After that I discovered by the way.

NOTE Confidence: 0.8231227

00:02:42.540 --> 00:02:45.055 Only yesterday is that he

NOTE Confidence: 0.8231227

00:02:45.055 --> 00:02:48.190 never knew I'm I'm here from.

NOTE Confidence: 0.8231227

00:02:48.190 --> 00:02:50.225 Since three years and they

NOTE Confidence: 0.8231227

00:02:50.225 --> 00:02:52.260 never known that Edward Tufte,

NOTE Confidence: 0.8231227

00:02:52.260 --> 00:02:56.194 it is the most one of the

NOTE Confidence: 0.8231227

00:02:56.194 --> 00:02:57.880 most celebrated visualization.

NOTE Confidence: 0.8231227

00:02:57.880 --> 00:03:02.544 Antisa is is is there in new heaven.

NOTE Confidence: 0.8231227

00:03:02.550 --> 00:03:04.545 And the condition is that with an

NOTE Confidence: 0.8231227

00:03:04.545 --> 00:03:06.945 image you have to give to the viewer

NOTE Confidence: 0.8231227

00:03:06.945 --> 00:03:09.136 the greatest number of ideas in the

NOTE Confidence: 0.8231227

00:03:09.136 --> 00:03:11.397 shortest time and with the least possible,

NOTE Confidence: 0.8231227

00:03:11.400 --> 00:03:13.150 Inc.

NOTE Confidence: 0.8231227

00:03:13.150 --> 00:03:15.105 This is another general representation

NOTE Confidence: 0.8231227

00:03:15.105 --> 00:03:17.536 of his that you should consider

NOTE Confidence: 0.8231227

00:03:17.536 --> 00:03:19.566 to make a good visualization.

NOTE Confidence: 0.8231227

00:03:19.570 --> 00:03:22.671 Also, this is very general, so it's not.

NOTE Confidence: 0.8231227

00:03:22.671 --> 00:03:24.506 It's not only about science

NOTE Confidence: 0.8231227

00:03:24.506 --> 00:03:26.618 and basically they criteria are

NOTE Confidence: 0.8231227

00:03:26.618 --> 00:03:28.788 organized in four different sets,

NOTE Confidence: 0.8231227

00:03:28.790 --> 00:03:31.597 so you need to represent the information,

NOTE Confidence: 0.8231227

00:03:31.600 --> 00:03:34.113 but the fever also need to display

NOTE Confidence: 0.8231227

00:03:34.113 --> 00:03:37.006 to convey to communicate a story and

NOTE Confidence: 0.8231227

00:03:37.006 --> 00:03:39.616 that's the concept of the figure.

NOTE Confidence: 0.8231227

00:03:39.620 --> 00:03:43.328 This is connected also with the goal of the.

NOTE Confidence: 0.8231227

00:03:43.330 --> 00:03:46.690 All of the figures, so the function.

NOTE Confidence: 0.8231227

00:03:46.690 --> 00:03:49.735 So what is the message that you

NOTE Confidence: 0.8231227

00:03:49.735 --> 00:03:53.257 want to display and also the visual

NOTE Confidence: 0.8231227

00:03:53.257 --> 00:03:54.844 format is important.

NOTE Confidence: 0.8231227

00:03:54.850 --> 00:03:57.556 Obviously the weight of these four

NOTE Confidence: 0.8231227

00:03:57.556 --> 00:03:59.960 different layers is different in

NOTE Confidence: 0.8231227

00:03:59.960 --> 00:04:02.048 different applications for images,

NOTE Confidence: 0.8231227

00:04:02.050 --> 00:04:04.930 so visual form probably is more

NOTE Confidence: 0.8231227

00:04:04.930 --> 00:04:06.850 important for artistic display,

NOTE Confidence: 0.8231227

00:04:06.850 --> 00:04:09.270 while for scientific displays that

NOTE Confidence: 0.8231227

00:04:09.270 --> 00:04:11.690 probably in formation goal and

NOTE Confidence: 0.8589544

00:04:11.767 --> 00:04:13.627 story are more important.

NOTE Confidence: 0.8589544

00:04:13.630 --> 00:04:16.787 This doesn't mean that you should not

NOTE Confidence: 0.8589544

00:04:16.787 --> 00:04:19.789 consider also the visual visual part.

NOTE Confidence: 0.8589544

00:04:19.790 --> 00:04:22.110 I ideally the perfect visualization

NOTE Confidence: 0.8589544

00:04:22.110 --> 00:04:25.949 is at the center of these four steps.

NOTE Confidence: 0.89499336

00:04:28.090 --> 00:04:30.575 So this is for the introduction now.

NOTE Confidence: 0.89499336

00:04:30.580 --> 00:04:32.855 The rest of the presentation will be

NOTE Confidence: 0.89499336

00:04:32.855 --> 00:04:35.209 structured with some very concrete examples,

NOTE Confidence: 0.89499336

00:04:35.210 --> 00:04:36.990 and it's also organized in

NOTE Confidence: 0.89499336

00:04:36.990 --> 00:04:38.770 a way that is interactive,

NOTE Confidence: 0.89499336

00:04:38.770 --> 00:04:40.550 so I will show something.

NOTE Confidence: 0.89499336

00:04:40.550 --> 00:04:43.871 Some example of figures and I will try to

NOTE Confidence: 0.89499336

00:04:43.871 --> 00:04:47.685 ask you what could be wrong with this figure.

NOTE Confidence: 0.89499336

00:04:47.690 --> 00:04:48.794 Starting with this,

NOTE Confidence: 0.89499336

00:04:48.794 --> 00:04:52.064 but this is a figure that is very

NOTE Confidence: 0.89499336

00:04:52.064 --> 00:04:54.520 frequent in scientific publication.

NOTE Confidence: 0.89499336

00:04:54.520 --> 00:04:57.698 It's a barplot and it's the most.

NOTE Confidence: 0.89499336

00:04:57.700 --> 00:05:00.619 It's actually the most frequent disk image

NOTE Confidence: 0.89499336

00:05:00.619 --> 00:05:04.067 that you can find in biomedical journals.

NOTE Confidence: 0.92155665

00:05:07.530 --> 00:05:10.792 Do you have any ideas of what

NOTE Confidence: 0.92155665

00:05:10.792 --> 00:05:13.360 could be wrong with this?

NOTE Confidence: 0.92155665

00:05:13.360 --> 00:05:16.150 Not pretty, yes. Lots of it.

NOTE Confidence: 0.43369204

00:05:18.110 --> 00:05:21.133 It. It's lacking data.

NOTE Confidence: 0.43369204

00:05:21.133 --> 00:05:22.738 It's it's not showing you

NOTE Confidence: 0.43369204

00:05:22.738 --> 00:05:24.230 the data distribution.

NOTE Confidence: 0.43369204

00:05:24.230 --> 00:05:26.072 Yeah, yes, exactly there are.

NOTE Confidence: 0.43369204

00:05:26.072 --> 00:05:28.660 It's putting the treatment on the left,

NOTE Confidence: 0.43369204

00:05:28.660 --> 00:05:30.500 which I always don't like.

NOTE Confidence: 0.43369204

00:05:30.500 --> 00:05:33.090 I always want the control on the

NOTE Confidence: 0.7957011

00:05:33.090 --> 00:05:34.562 left. Oh yes, OK.

NOTE Confidence: 0.7957011

00:05:34.562 --> 00:05:36.436 Yes, that's true, yes,

NOTE Confidence: 0.7957011

00:05:36.436 --> 00:05:42.150 so it has a lot of like. Capital A.

NOTE Confidence: 0.7957011

00:05:42.150 --> 00:05:44.268 Visual problems, but the main yes.

NOTE Confidence: 0.7957011

00:05:44.270 --> 00:05:46.748 The main thing is that it doesn't

NOTE Confidence: 0.7957011

00:05:46.748 --> 00:05:49.182 show the data, so that's the main.

NOTE Confidence: 0.7957011

00:05:49.182 --> 00:05:51.710 That's the main drawback of this image,

NOTE Confidence: 0.7957011

00:05:51.710 --> 00:05:54.188 and so, particularly in the last year.

NOTE Confidence: 0.7957011

00:05:54.190 --> 00:05:57.097 So that's the trend that a lot of also

NOTE Confidence: 0.7957011

00:05:57.097 --> 00:05:59.138 publishers are requesting in images.

NOTE Confidence: 0.7957011

00:05:59.140 --> 00:06:02.119 So the principle is that you need ideally to

NOTE Confidence: 0.7957011

00:06:02.119 --> 00:06:05.154 show always the data points in every figure,

NOTE Confidence: 0.7957011

00:06:05.160 --> 00:06:07.652 because you should show the data that

NOTE Confidence: 0.7957011

00:06:07.652 --> 00:06:10.614 make up your fingers and these for a

NOTE Confidence: 0.7957011

00:06:10.614 --> 00:06:13.079 barplot means that you have to show.

NOTE Confidence: 0.7957011

00:06:13.080 --> 00:06:16.240 They did data points.

NOTE Confidence: 0.7957011

00:06:16.240 --> 00:06:19.291 So here you see an example of how these

NOTE Confidence: 0.7957011

00:06:19.291 --> 00:06:22.184 bar plot can be represented with the

NOTE Confidence: 0.7957011

00:06:22.184 --> 00:06:25.172 data points and you see here on the

NOTE Confidence: 0.7957011

00:06:25.172 --> 00:06:27.720 on the right you see the the single

NOTE Confidence: 0.7957011

00:06:27.720 --> 00:06:30.320 data points and also you see a summary

NOTE Confidence: 0.7957011

00:06:30.392 --> 00:06:32.846 statistics that could be for example,

NOTE Confidence: 0.7957011

00:06:32.850 --> 00:06:35.100 they mean plus minus standard deviation

NOTE Confidence: 0.7957011

00:06:35.100 --> 00:06:37.900 for the treatment and the end the control.

NOTE Confidence: 0.7957011

00:06:37.900 --> 00:06:39.730 In general this showing barplots

NOTE Confidence: 0.7957011

00:06:39.730 --> 00:06:42.294 with only the mean with the standard

NOTE Confidence: 0.7957011

00:06:42.294 --> 00:06:44.625 deviation is a problem and there was

NOTE Confidence: 0.7957011

00:06:44.625 --> 00:06:46.988 a publication of five years ago.

NOTE Confidence: 0.7957011

00:06:46.990 --> 00:06:49.348 The teacher, wife and one issues,

NOTE Confidence: 0.7957011

00:06:49.350 --> 00:06:50.180 for example,

NOTE Confidence: 0.7957011

00:06:50.180 --> 00:06:52.255 that the different data distribution

NOTE Confidence: 0.7957011

00:06:52.255 --> 00:06:54.867 can lead to the same bar block.

NOTE Confidence: 0.7957011

00:06:54.870 --> 00:06:56.840 You see an example here.

NOTE Confidence: 0.7957011

00:06:56.840 --> 00:06:59.198 So in a you should use.

NOTE Confidence: 0.7957011

00:06:59.200 --> 00:07:01.105 You see a barplot representation

NOTE Confidence: 0.7957011

00:07:01.105 --> 00:07:03.864 of distribution of data and all the

NOTE Confidence: 0.7957011

00:07:03.864 --> 00:07:06.228 distribution that you see from B2ER

NOTE Confidence: 0.7957011

00:07:06.228 --> 00:07:07.868 representing could be represented

NOTE Confidence: 0.7957011

00:07:07.868 --> 00:07:09.050 by that padlock.

NOTE Confidence: 0.7957011

00:07:09.050 --> 00:07:11.444 So the ideal situation would be

NOTE Confidence: 0.7957011

00:07:11.444 --> 00:07:13.780 what you see here in plot.

NOTE Confidence: 0.7957011

00:07:13.780 --> 00:07:18.008 Be where you have data that are.

NOTE Confidence: 0.7957011

00:07:18.010 --> 00:07:18.770 Symmetrically distributed,

NOTE Confidence: 0.7957011

00:07:18.770 --> 00:07:21.430 so this is the if the distribution

NOTE Confidence: 0.7957011

00:07:21.430 --> 00:07:22.947 of your real data is.

NOTE Confidence: 0.7957011

00:07:22.950 --> 00:07:25.774 These are the bar plot is less problematic,

NOTE Confidence: 0.7957011

00:07:25.780 --> 00:07:28.420 but for example in C use your situation

NOTE Confidence: 0.7957011

00:07:28.420 --> 00:07:30.838 where you have an outlier and so

NOTE Confidence: 0.7957011

00:07:30.838 --> 00:07:33.347 for example this would mean that the

NOTE Confidence: 0.7957011

00:07:33.347 --> 00:07:35.232 supposed difference that you are

NOTE Confidence: 0.7957011

00:07:35.232 --> 00:07:37.670 showing in the padlock is not real,

NOTE Confidence: 0.7957011

00:07:37.670 --> 00:07:39.295 but it's present only because

NOTE Confidence: 0.7957011

00:07:39.295 --> 00:07:41.307 you have these outlier pointer.

NOTE Confidence: 0.7957011

00:07:41.310 --> 00:07:43.795 But most of the other data are

NOTE Confidence: 0.7957011

00:07:43.795 --> 00:07:46.239 overlapping in the two distributions.

NOTE Confidence: 0.7957011

00:07:46.240 --> 00:07:48.718 Sometimes as you see in the,

NOTE Confidence: 0.7957011

00:07:48.720 --> 00:07:51.618 this plot could hide some patterns in

NOTE Confidence: 0.7957011

00:07:51.618 --> 00:07:54.928 the data, so that's what you see here.

NOTE Confidence: 0.7957011

00:07:54.930 --> 00:07:58.759 In the do you see my cursor?

NOTE Confidence: 0.7957011

00:07:58.760 --> 00:08:00.006 Yes, OK,

NOTE Confidence: 0.7957011

00:08:00.006 --> 00:08:04.990 so this for example shows that there are.

NOTE Confidence: 0.7957011

00:08:04.990 --> 00:08:06.840 The distributions that you see

NOTE Confidence: 0.7957011

00:08:06.840 --> 00:08:08.320 here are by model.

NOTE Confidence: 0.7957011

00:08:08.320 --> 00:08:10.540 This could be linked, for example,

NOTE Confidence: 0.7957011

00:08:10.540 --> 00:08:12.020 to replicate for example,

NOTE Confidence: 0.7957011

00:08:12.020 --> 00:08:13.130 technical replicates and

NOTE Confidence: 0.7957011

00:08:13.130 --> 00:08:13.870 biological replicates,

NOTE Confidence: 0.7957011

00:08:13.870 --> 00:08:16.090 or it could be an important

NOTE Confidence: 0.7957011

00:08:16.090 --> 00:08:17.570 property of the data.

NOTE Confidence: 0.7957011

00:08:17.570 --> 00:08:18.012 Nevertheless,

NOTE Confidence: 0.7957011

00:08:18.012 --> 00:08:20.222 it’s something that you cannot

NOTE Confidence: 0.7957011

00:08:20.222 --> 00:08:22.704 see if you represent with with

NOTE Confidence: 0.7957011

00:08:22.704 --> 00:08:24.944 a bar plot and the also Bartlett

NOTE Confidence: 0.7957011

00:08:24.944 --> 00:08:27.331 hide the number of data that are

NOTE Confidence: 0.7957011

00:08:27.331 --> 00:08:29.007 used to visualize the plot.

NOTE Confidence: 0.7957011

00:08:29.007 --> 00:08:30.692 The barplot themselves and so

NOTE Confidence: 0.7957011

00:08:30.692 --> 00:08:32.740 for example in EU situation,

NOTE Confidence: 0.7957011

00:08:32.740 --> 00:08:35.386 where you have an equal number of.

NOTE Confidence: 0.7957011

00:08:35.390 --> 00:08:38.870 Points for the black and the white are

NOTE Confidence: 0.7957011

00:08:38.870 --> 00:08:41.840 Bartlett on the left and the right.

NOTE Confidence: 0.7957011

00:08:41.840 --> 00:08:45.280 At this is a problem also when you

NOTE Confidence: 0.7957011

00:08:45.280 --> 00:08:49.217 want to show paired data in barplots.

NOTE Confidence: 0.7957011

00:08:49.220 --> 00:08:52.832 So again here you see a situation

NOTE Confidence: 0.7957011

00:08:52.832 --> 00:08:54.380 where a barplot

NOTE Confidence: 0.7462741

00:08:54.490 --> 00:08:57.230 displays some is the same.

NOTE Confidence: 0.7462741

00:08:57.230 --> 00:08:59.010 For situations that you see

NOTE Confidence: 0.7462741

00:08:59.010 --> 00:09:01.148 displayed in BC&D, so be Cmdr.

NOTE Confidence: 0.7462741

00:09:01.148 --> 00:09:02.572 Very different situations here

NOTE Confidence: 0.7462741

00:09:02.572 --> 00:09:04.342 you could imagine, for example,

NOTE Confidence: 0.7462741

00:09:04.342 --> 00:09:06.448 that this data obtained from single

NOTE Confidence: 0.7462741

00:09:06.448 --> 00:09:08.618 patients at treated with the dragon,

NOTE Confidence: 0.7462741

00:09:08.620 --> 00:09:11.468 and you measure a parameter of the patients,

NOTE Confidence: 0.7462741

00:09:11.470 --> 00:09:14.088 and so the information related to each

NOTE Confidence: 0.7462741

00:09:14.088 --> 00:09:16.903 patient has to be connected so that the

NOTE Confidence: 0.7462741

00:09:16.903 --> 00:09:19.660 meaning of the of the pair that plot.

NOTE Confidence: 0.7462741

00:09:19.660 --> 00:09:22.396 So the situation in B shows that the

NOTE Confidence: 0.7462741

00:09:22.396 --> 00:09:25.012 Dragon has a consistent effect on all

NOTE Confidence: 0.7462741

00:09:25.012 --> 00:09:27.600 the patients and you can see that.

NOTE Confidence: 0.7462741

00:09:27.600 --> 00:09:29.524 Calculating for each patient,

NOTE Confidence: 0.7462741

00:09:29.524 --> 00:09:32.410 the difference between the dots on

NOTE Confidence: 0.7462741

00:09:32.487 --> 00:09:35.071 the left and on the right give rise

NOTE Confidence: 0.7462741

00:09:35.071 --> 00:09:37.600 to this to this plot here below,

NOTE Confidence: 0.7462741

00:09:37.600 --> 00:09:39.665 where all the differences are

NOTE Confidence: 0.7462741

00:09:39.665 --> 00:09:42.000 positive and are also consistent in.

NOTE Confidence: 0.7462741

00:09:42.000 --> 00:09:44.382 See you see a situation where

NOTE Confidence: 0.7462741

00:09:44.382 --> 00:09:46.400 the drug has very big,

NOTE Confidence: 0.7462741

00:09:46.400 --> 00:09:48.416 very different effects depending

NOTE Confidence: 0.7462741

00:09:48.416 --> 00:09:49.928 on the patient.

NOTE Confidence: 0.7462741

00:09:49.930 --> 00:09:51.735 So that the distribution of

NOTE Confidence: 0.7462741

00:09:51.735 --> 00:09:53.179 the differences is skewed.

NOTE Confidence: 0.7462741

00:09:53.180 --> 00:09:54.832 And by the way,

NOTE Confidence: 0.7462741

00:09:54.832 --> 00:09:56.897 this line represents the median

NOTE Confidence: 0.7462741

00:09:56.897 --> 00:09:58.938 difference that you see for

NOTE Confidence: 0.7462741

00:09:58.938 --> 00:10:00.878 each patients for the treatment.

NOTE Confidence: 0.7462741

00:10:00.880 --> 00:10:03.088 And the third plot indeed that you see

NOTE Confidence: 0.7462741

00:10:03.088 --> 00:10:05.469 has a composition of effects that.

NOTE Confidence: 0.7462741

00:10:05.470 --> 00:10:07.752 So here you see that the again

NOTE Confidence: 0.7462741

00:10:07.752 --> 00:10:09.409 the difference is by model.

NOTE Confidence: 0.7462741

00:10:09.410 --> 00:10:11.408 That means that there are patients

NOTE Confidence: 0.7462741

00:10:11.408 --> 00:10:13.670 that do not respond to the dragon,

NOTE Confidence: 0.7462741

00:10:13.670 --> 00:10:15.452 and you see here with the

NOTE Confidence: 0.7462741

00:10:15.452 --> 00:10:17.072 horizontal lines and some patients

NOTE Confidence: 0.7462741

00:10:17.072 --> 00:10:18.917 that responded to the dragon.

NOTE Confidence: 0.7462741

00:10:18.920 --> 00:10:21.020 So the resulting distribution of the

NOTE Confidence: 0.7462741

00:10:21.020 --> 00:10:23.507 difference as you see here is by model.

NOTE Confidence: 0.7462741

00:10:23.510 --> 00:10:25.145 The problem with her plots

NOTE Confidence: 0.7462741

00:10:25.145 --> 00:10:26.453 are and the problem.

NOTE Confidence: 0.7462741

00:10:26.460 --> 00:10:26.802 Also,

NOTE Confidence: 0.7462741

00:10:26.802 --> 00:10:28.854 if you use barplots with paired

NOTE Confidence: 0.7462741

00:10:28.854 --> 00:10:31.188 data is that you don't see any.

NOTE Confidence: 0.7462741

00:10:31.190 --> 00:10:32.700 Any of this structure so

NOTE Confidence: 0.7462741

00:10:32.700 --> 00:10:34.210 you you are losing it.

NOTE Confidence: 0.7462741

00:10:34.210 --> 00:10:36.298 So the best way is always to show

NOTE Confidence: 0.7462741

00:10:36.298 --> 00:10:38.437 the dots are of your distribution,

NOTE Confidence: 0.7462741

00:10:38.440 --> 00:10:40.150 maybe together with the bar plots

NOTE Confidence: 0.7462741

00:10:40.150 --> 00:10:42.247 and if the data are paid also

NOTE Confidence: 0.7462741

00:10:42.247 --> 00:10:43.732 to show the single connection

NOTE Confidence: 0.7462741

00:10:43.732 --> 00:10:45.379 with in between the dots.

NOTE Confidence: 0.7590634

00:10:49.080 --> 00:10:52.400 There is also an issue about about the

NOTE Confidence: 0.7590634

00:10:52.400 --> 00:10:56.167 choice of displaying the meme of your data,

NOTE Confidence: 0.7590634

00:10:56.170 --> 00:10:58.380 for example versus the media,

NOTE Confidence: 0.7590634

00:10:58.380 --> 00:11:01.212 or to show the standard deviation

NOTE Confidence: 0.7590634

00:11:01.212 --> 00:11:04.140 versus the standard error of the mean,

NOTE Confidence: 0.7590634

00:11:04.140 --> 00:11:07.404 so mean versus median are ways to represent

NOTE Confidence: 0.7590634

00:11:07.404 --> 00:11:10.786 summary of the centrality of a distribution.

NOTE Confidence: 0.7590634

00:11:10.790 --> 00:11:13.919 An the mean is preferable if you

NOTE Confidence: 0.7590634

00:11:13.919 --> 00:11:16.788 suppose your data are for example.

NOTE Confidence: 0.7590634

00:11:16.790 --> 00:11:17.628 Symmetrically distributed.

NOTE Confidence: 0.7590634

00:11:17.628 --> 00:11:20.980 For example, if you assume that the data

NOTE Confidence: 0.7590634

00:11:21.050 --> 00:11:23.426 has a normal or Gaussian distribution,

NOTE Confidence: 0.7590634

00:11:23.430 --> 00:11:25.405 while the median represents the

NOTE Confidence: 0.7590634

00:11:25.405 --> 00:11:27.894 mid is the point that represents

NOTE Confidence: 0.7590634

00:11:27.894 --> 00:11:30.069 the middle of your data.

NOTE Confidence: 0.7590634

00:11:30.070 --> 00:11:32.150 The middle of your distribution,

NOTE Confidence: 0.7590634

00:11:32.150 --> 00:11:34.265 and it’s more generally applied

NOTE Confidence: 0.7590634

00:11:34.265 --> 00:11:35.957 independently from the shaper

NOTE Confidence: 0.7590634

00:11:35.957 --> 00:11:38.366 of the distribution of the data.

NOTE Confidence: 0.7590634

00:11:38.370 --> 00:11:41.402 So here you see an example where you

NOTE Confidence: 0.7590634

00:11:41.402 --> 00:11:43.713 have four different samples population

NOTE Confidence: 0.7590634

00:11:43.713 --> 00:11:47.950 and you plot the mean plus the standard.

NOTE Confidence: 0.7590634

00:11:47.950 --> 00:11:49.620 And that’s the most conventional

NOTE Confidence: 0.7590634

00:11:49.620 --> 00:11:52.130 way that you see in publication.

NOTE Confidence: 0.7590634

00:11:52.130 --> 00:11:54.970 They mean plus standard deviation.

NOTE Confidence: 0.7590634

00:11:54.970 --> 00:11:57.686 And the median of the population will

NOTE Confidence: 0.7590634

00:11:57.686 --> 00:12:00.249 receive the single point and the

NOTE Confidence: 0.7590634

00:12:00.249 --> 00:12:02.459 horizontal bar represents the median.

NOTE Confidence: 0.7590634

00:12:02.460 --> 00:12:05.750 So an important point about.

NOTE Confidence: 0.7590634

00:12:05.750 --> 00:12:07.886 Mean versus median is that the

NOTE Confidence: 0.7590634

00:12:07.886 --> 00:12:11.289 mean and can be used only with

NOTE Confidence: 0.7590634

00:12:11.289 --> 00:12:12.557 symmetrical distributions.

NOTE Confidence: 0.7590634

00:12:12.560 --> 00:12:14.830 Otherwise it can be misleading.

NOTE Confidence: 0.7590634

00:12:14.830 --> 00:12:17.100 While the median is more

NOTE Confidence: 0.7590634

00:12:17.100 --> 00:12:18.008 generally appropriate.

NOTE Confidence: 0.80607957

00:12:19.220 --> 00:12:21.138 When you have an outlier like that,

NOTE Confidence: 0.80607957

00:12:21.140 --> 00:12:23.786 you would always recommend the meat being.

NOTE Confidence: 0.80607957

00:12:23.790 --> 00:12:26.792 Honey. When you have an outlier

NOTE Confidence: 0.80607957

00:12:26.792 --> 00:12:28.938 like in the third group there,

NOTE Confidence: 0.80607957

00:12:28.940 --> 00:12:31.439 yeah, then it makes more sense to

NOTE Confidence: 0.7963572

00:12:31.440 --> 00:12:32.514 use the median.

NOTE Confidence: 0.7963572

00:12:32.514 --> 00:12:34.304 Yeah, nobody showed them young.

NOTE Confidence: 0.7963572

00:12:34.310 --> 00:12:36.627 Is that the median is more robust

NOTE Confidence: 0.7963572

00:12:36.627 --> 00:12:38.468 data with outliers is totally

NOTE Confidence: 0.7963572

00:12:38.468 --> 00:12:40.036 more robust with outliers,

NOTE Confidence: 0.7963572

00:12:40.040 --> 00:12:41.830 and the median is not,

NOTE Confidence: 0.7963572

00:12:41.830 --> 00:12:44.399 so the presence of over outlier as

NOTE Confidence: 0.7963572

00:12:44.399 --> 00:12:47.706 you see here in C can shift a lot the

NOTE Confidence: 0.7963572

00:12:47.706 --> 00:12:50.418 mean while the median is is affected,

NOTE Confidence: 0.7963572

00:12:50.420 --> 00:12:53.260 but not so much.

NOTE Confidence: 0.7963572

00:12:53.260 --> 00:12:54.816 Especially from the magnitude

NOTE Confidence: 0.7963572

00:12:54.816 --> 00:12:55.980 of the outlier,

NOTE Confidence: 0.7963572

00:12:55.980 --> 00:12:58.240 I would say. So

NOTE Confidence: 0.80480194

00:12:58.240 --> 00:12:59.338 tomorrow question right?

NOTE Confidence: 0.80480194

00:12:59.338 --> 00:13:01.568 So so also, being you know,

NOTE Confidence: 0.80480194

00:13:01.568 --> 00:13:03.072 knowing there's a difference

NOTE Confidence: 0.80480194

00:13:03.072 --> 00:13:04.840 between me and a medium,

NOTE Confidence: 0.80480194

00:13:04.840 --> 00:13:07.409 but one of the things I heard,

NOTE Confidence: 0.80480194

00:13:07.410 --> 00:13:08.878 of course, haven't looked

NOTE Confidence: 0.80480194

00:13:08.878 --> 00:13:10.713 myself into this deeply enough.

NOTE Confidence: 0.80480194

00:13:10.720 --> 00:13:13.275 Is that for the meeting the distribution,

NOTE Confidence: 0.80480194

00:13:13.280 --> 00:13:15.115 unlike mean not necessarily follows

NOTE Confidence: 0.80480194

00:13:15.115 --> 00:13:16.950 a Gaussian or normal distribution,

NOTE Confidence: 0.80480194

00:13:16.950 --> 00:13:19.894 so that from a statistical point of view

NOTE Confidence: 0.80480194

00:13:19.894 --> 00:13:23.187 is going to be a little hard to calculate,

NOTE Confidence: 0.80480194

00:13:23.190 --> 00:13:24.291 certain significance etc.

NOTE Confidence: 0.80480194

00:13:24.291 --> 00:13:25.759 Based on medium data.

NOTE Confidence: 0.80480194

00:13:25.760 --> 00:13:26.861 Is that true?

NOTE Confidence: 0.80480194

00:13:26.861 --> 00:13:28.696 Or it's simply a misnomer?

NOTE Confidence: 0.89396536

00:13:29.690 --> 00:13:32.330 How to calculate the

NOTE Confidence: 0.89396536

00:13:32.330 --> 00:13:34.310 significance of differences?

NOTE Confidence: 0.89396536

00:13:34.310 --> 00:13:35.609 That's a different.

NOTE Confidence: 0.89396536

00:13:35.609 --> 00:13:38.640 So that's the difference of the approach.

NOTE Confidence: 0.89396536

00:13:38.640 --> 00:13:40.810 If you choose parametric test,

NOTE Confidence: 0.89396536

00:13:40.810 --> 00:13:43.652 such as the tester or the ANOVA

NOTE Confidence: 0.89396536

00:13:43.652 --> 00:13:46.544 and those tests assume that the

NOTE Confidence: 0.89396536

00:13:46.544 --> 00:13:49.189 distribution is Goshen is normal.

NOTE Confidence: 0.89396536

00:13:49.190 --> 00:13:52.030 Yeah, so you need to be careful so he is

NOTE Confidence: 0.89396536

00:13:52.102 --> 00:13:54.727 usually if it is a repeated measures.

NOTE Confidence: 0.89396536

00:13:54.730 --> 00:13:56.475 So if you’re testing repeated

NOTE Confidence: 0.89396536

00:13:56.475 --> 00:13:58.865 measure yes soon the error is is

NOTE Confidence: 0.89396536

00:13:58.865 --> 00:14:00.593 is distributed in a Goshen way,

NOTE Confidence: 0.89396536

00:14:00.600 --> 00:14:02.875 but that is not always the case.

NOTE Confidence: 0.89396536

00:14:02.880 --> 00:14:03.806 For example,

NOTE Confidence: 0.89396536

00:14:03.806 --> 00:14:06.584 if you’re comparing two population of

NOTE Confidence: 0.89396536

00:14:06.584 --> 00:14:09.337 jeans with a signal for each gene.

NOTE Confidence: 0.89396536

00:14:09.340 --> 00:14:10.170 Just have to check it.

NOTE Confidence: 0.81003946

00:14:11.400 --> 00:14:13.216 Well, so so this is something that I

NOTE Confidence: 0.81003946

00:14:13.216 --> 00:14:14.997 think will be particularly important

NOTE Confidence: 0.81003946

00:14:14.997 --> 00:14:16.572 for experimental scientist, right?

NOTE Confidence: 0.81003946

00:14:16.572 --> 00:14:18.324 Because you know, as an experiment

NOTE Confidence: 0.81003946

00:14:18.324 --> 00:14:20.855 is when we are trained, we know OK,

NOTE Confidence: 0.81003946

00:14:20.855 --> 00:14:22.380 when we did design experiment,

NOTE Confidence: 0.81003946

00:14:22.380 --> 00:14:24.452 we do service replica so we can

NOTE Confidence: 0.81003946

00:14:24.452 --> 00:14:26.023 join error bar without thinking

NOTE Confidence: 0.81003946

00:14:26.023 --> 00:14:28.172 Y and how to deal with it.

NOTE Confidence: 0.81003946

00:14:28.180 --> 00:14:30.357 And if you go to a statistician

NOTE Confidence: 0.81003946

00:14:30.357 --> 00:14:32.450 that will tell you say oh look,

NOTE Confidence: 0.81003946

00:14:32.450 --> 00:14:34.578 if you're going to use the test,

NOTE Confidence: 0.81003946

00:14:34.580 --> 00:14:37.118 you have to show me first that this is

NOTE Confidence: 0.81003946

00:14:37.118 --> 00:14:38.928 actually largely a normal distribution

NOTE Confidence: 0.81003946

00:14:38.928 --> 00:14:41.609 before you can actually use the T test.

NOTE Confidence: 0.81003946

00:14:41.610 --> 00:14:43.010 Whereas the vast majority

NOTE Confidence: 0.81003946

00:14:43.010 --> 00:14:44.760 of people in the lab,

NOTE Confidence: 0.81003946

00:14:44.760 --> 00:14:46.435 that's not how they will

NOTE Confidence: 0.81003946

00:14:46.435 --> 00:14:48.610 think about in the 1st place,

NOTE Confidence: 0.81003946

00:14:48.610 --> 00:14:50.644 and they also not trendy enough

NOTE Confidence: 0.81003946

00:14:50.644 --> 00:14:53.134 to think you know how to prove

NOTE Confidence: 0.81003946

00:14:53.134 --> 00:14:54.904 or disprove that's the case.

NOTE Confidence: 0.81003946

00:14:54.910 --> 00:14:56.660 So what would you suggest,

NOTE Confidence: 0.81003946

00:14:56.660 --> 00:14:57.972 especially when we're doing

NOTE Confidence: 0.81003946

00:14:57.972 --> 00:14:59.612 experiment that you cannot do

NOTE Confidence: 0.81003946

00:14:59.612 --> 00:15:01.558 200 replicas for each experiment.

NOTE Confidence: 0.81003946

00:15:01.560 --> 00:15:03.660 So what would be a good

NOTE Confidence: 0.81003946

00:15:03.660 --> 00:15:05.060 approach in that regard?

NOTE Confidence: 0.83941036

00:15:06.130 --> 00:15:08.909 Yeah, so there is a tradeoff between

NOTE Confidence: 0.83941036

00:15:08.909 --> 00:15:11.363 the ideal situation where the ideal

NOTE Confidence: 0.83941036

00:15:11.363 --> 00:15:13.745 situation would be always to have

NOTE Confidence: 0.83941036

00:15:13.745 --> 00:15:16.380 enough data points so that you can

NOTE Confidence: 0.83941036

00:15:16.380 --> 00:15:18.654 understand the shape of the distribution

NOTE Confidence: 0.83941036

00:15:18.654 --> 00:15:21.538 and the real case scenario with you

NOTE Confidence: 0.83941036

00:15:21.538 --> 00:15:24.458 can do as many replicates as you can,

NOTE Confidence: 0.83941036

00:15:24.460 --> 00:15:27.190 and so usually you have to assume

NOTE Confidence: 0.83941036

00:15:27.190 --> 00:15:29.530 that the distribution is normal, so.

NOTE Confidence: 0.88484627

00:15:32.700 --> 00:15:35.886 Ideally, you should always check her.

NOTE Confidence: 0.88484627

00:15:35.890 --> 00:15:38.515 And again, if we are repeating measures

NOTE Confidence: 0.88484627

00:15:38.515 --> 00:15:40.869 and you are collecting a measure

NOTE Confidence: 0.88484627

00:15:40.869 --> 00:15:43.550 of the same data in a repeated,

NOTE Confidence: 0.88484627

00:15:43.550 --> 00:15:45.530 that way you can assume that

NOTE Confidence: 0.88484627

00:15:45.530 --> 00:15:47.760 if the error is stochastic,

NOTE Confidence: 0.88484627

00:15:47.760 --> 00:15:49.680 it should be normally distributed.

NOTE Confidence: 0.88484627

00:15:49.680 --> 00:15:52.221 So you assume that the distribution of

NOTE Confidence: 0.88484627

00:15:52.221 --> 00:15:55.278 the error is Goshen, and it makes sense.

NOTE Confidence: 0.88484627

00:15:55.278 --> 00:15:57.360 But for example in other situation

NOTE Confidence: 0.88484627

00:15:57.429 --> 00:15:59.949 where you have a lot of measurements

NOTE Confidence: 0.88484627

00:15:59.949 --> 00:16:02.319 and measurements of different entities,

NOTE Confidence: 0.88484627

00:16:02.320 --> 00:16:04.410 for example, the expression of

NOTE Confidence: 0.88484627

00:16:04.410 --> 00:16:06.500 different genes we’re doing like.

NOTE Confidence: 0.88484627

00:16:06.500 --> 00:16:08.236 A compilation of Jesus.

NOTE Confidence: 0.88484627

00:16:08.236 --> 00:16:10.840 Then these assumption is less probable,

NOTE Confidence: 0.88484627

00:16:10.840 --> 00:16:12.088 is less likely,

NOTE Confidence: 0.88484627

00:16:12.088 --> 00:16:15.000 and you should have enough data points

NOTE Confidence: 0.88484627

00:16:15.086 --> 00:16:17.837 so that you can switch from parametric

NOTE Confidence: 0.88484627

00:16:17.837 --> 00:16:22.809 tests one on parametric, so we're not.

NOTE Confidence: 0.88484627

00:16:22.810 --> 00:16:24.540 That doesn't make assumption of

NOTE Confidence: 0.88484627

00:16:24.540 --> 00:16:25.924 on the underlying distribution.

NOTE Confidence: 0.88484627

00:16:25.930 --> 00:16:27.234 It is, for example,

NOTE Confidence: 0.88484627

00:16:27.234 --> 00:16:29.644 they will cook some test or the

NOTE Confidence: 0.88484627

00:16:29.644 --> 00:16:30.787 Mann Whitney test.

NOTE Confidence: 0.88484627

00:16:30.790 --> 00:16:33.734 And the problem is that you need them

NOTE Confidence: 0.88484627

00:16:33.734 --> 00:16:36.093 or replicates because if the end is

NOTE Confidence: 0.88484627

00:16:36.093 --> 00:16:39.384 the size is less than five, you don't.

NOTE Confidence: 0.88484627

00:16:39.384 --> 00:16:42.644 You cannot reach the statistical

NOTE Confidence: 0.88484627

00:16:42.644 --> 00:16:47.130 significance as it is accepted below 0.05,

NOTE Confidence: 0.88484627

00:16:47.130 --> 00:16:52.576 but it's usually the more correct way.

NOTE Confidence: 0.88484627

00:16:52.580 --> 00:16:55.109 Then they the standard is not to use that,

NOTE Confidence: 0.88484627

00:16:55.110 --> 00:16:57.238 and so I remember there was a case

NOTE Confidence: 0.88484627

00:16:57.238 --> 00:16:59.039 where the paper was in review.

NOTE Confidence: 0.88484627

00:16:59.040 --> 00:17:01.839 It was from.

NOTE Confidence: 0.88484627

00:17:01.840 --> 00:17:04.668 Young bean and I remember we performed

NOTE Confidence: 0.88484627

00:17:04.668 --> 00:17:06.906 the Wilcoxon test and the reviewers

NOTE Confidence: 0.88484627

00:17:06.906 --> 00:17:09.596 as to why we didn't do the parameter

NOTE Confidence: 0.88484627

00:17:09.596 --> 00:17:12.740 test so so they asked for the opposite.

NOTE Confidence: 0.88484627

00:17:12.740 --> 00:17:15.372 They asked us to go against the

NOTE Confidence: 0.88484627

00:17:15.372 --> 00:17:16.124 ideal situation.

NOTE Confidence: 0.84749943

00:17:18.000 --> 00:17:19.698 I think this is very helpful.

NOTE Confidence: 0.84749943

00:17:19.700 --> 00:17:20.832 I think it's really,

NOTE Confidence: 0.84749943

00:17:20.832 --> 00:17:22.247 you know telling about me,

NOTE Confidence: 0.84749943

00:17:22.250 --> 00:17:24.066 especially for people who

NOTE Confidence: 0.84749943

00:17:24.066 --> 00:17:26.336 are not familiar with with.

NOTE Confidence: 0.84749943

00:17:26.340 --> 00:17:28.820 Test and also the the World Cup test.

NOTE Confidence: 0.84749943

00:17:28.820 --> 00:17:30.370 I think it’s really suggest

NOTE Confidence: 0.84749943

00:17:30.370 --> 00:17:31.920 you to look into that.

NOTE Confidence: 0.8276753

00:17:31.920 --> 00:17:33.470 Things can be very helpful.

NOTE Confidence: 0.8276753

00:17:33.470 --> 00:17:35.010 Yeah, and and obviously you’re

NOTE Confidence: 0.8276753

00:17:35.010 --> 00:17:36.880 so it’s important when you plan.

NOTE Confidence: 0.8276753

00:17:36.880 --> 00:17:39.814 If you if you can to have enough data

NOTE Confidence: 0.8276753

00:17:39.814 --> 00:17:42.587 points to perform a nonparametric test.

NOTE Confidence: 0.8276753

00:17:42.590 --> 00:17:43.526 In high throughput

NOTE Confidence: 0.8276753

00:17:43.526 --> 00:17:45.086 experiments that they see now,

NOTE Confidence: 0.8276753

00:17:45.090 --> 00:17:47.055 for example single cell that’s

NOTE Confidence: 0.8276753

00:17:47.055 --> 00:17:49.020 not there anymore problem because

NOTE Confidence: 0.8276753

00:17:49.085 --> 00:17:50.975 you have usually a lot of data

NOTE Confidence: 0.8276753

00:17:50.975 --> 00:17:53.158 points and so that’s less of a

NOTE Confidence: 0.8276753

00:17:53.158 --> 00:17:54.793 problem that sometimes we work.

NOTE Confidence: 0.8276753

00:17:54.800 --> 00:17:57.509 Is it after him because they thought

NOTE Confidence: 0.8276753

00:17:57.509 --> 00:18:00.299 the number of data are increasing?

NOTE Confidence: 0.8276753

00:18:00.300 --> 00:18:01.310 And not

NOTE Confidence: 0.8346292

00:18:01.310 --> 00:18:04.360 that generic comment comma. These people

NOTE Confidence: 0.8346292

00:18:04.360 --> 00:18:09.930 are a lot of these lot of our group is blood

NOTE Confidence: 0.8346292

00:18:09.930 --> 00:18:11.448 hematology researchers. Yeah,

NOTE Confidence: 0.8346292

00:18:11.450 --> 00:18:15.739 and neither blood nor blood advances require.

NOTE Confidence: 0.8346292

00:18:15.740 --> 00:18:17.430 The investigator in their papers

NOTE Confidence: 0.9140715

00:18:17.430 --> 00:18:18.770 to show all the

NOTE Confidence: 0.890373675

00:18:18.770 --> 00:18:20.958 data points. And now

NOTE Confidence: 0.86667585

00:18:20.960 --> 00:18:22.420 I'm on the publication committee.

NOTE Confidence: 0.86667585

00:18:22.420 --> 00:18:23.874 We've actually talked about this,

NOTE Confidence: 0.86667585

00:18:23.874 --> 00:18:27.510 but we go by the Journal of Cell Bio.

NOTE Confidence: 0.86667585

00:18:27.510 --> 00:18:30.048 Instructions to authors and prep for figures,

NOTE Confidence: 0.8307862

00:18:30.050 --> 00:18:31.870 and there are Rockefeller Press

NOTE Confidence: 0.822556846666667

00:18:31.870 --> 00:18:33.538 publication. And they

NOTE Confidence: 0.81844217

00:18:33.540 --> 00:18:35.790 haven't. So they have genome research

NOTE Confidence: 0.81844217

00:18:35.790 --> 00:18:38.416 and germ cell bio Med and stuff,

NOTE Confidence: 0.81844217

00:18:38.416 --> 00:18:41.042 so they haven't come around to making

NOTE Confidence: 0.81844217

00:18:41.042 --> 00:18:43.295 people show all their dots etc.

NOTE Confidence: 0.81844217

00:18:43.295 --> 00:18:46.293 But a number of journals, as you know,

NOTE Confidence: 0.81844217

00:18:46.293 --> 00:18:49.698 half like JC I you know JC AI

NOTE Confidence: 0.81844217

00:18:49.698 --> 00:18:51.370 advances etc. They might not

NOTE Confidence: 0.8858033

00:18:51.370 --> 00:18:52.670 even review your paper if

NOTE Confidence: 0.8858033

00:18:52.670 --> 00:18:53.710 you show, for instance,

NOTE Confidence: 0.8858033

00:18:53.710 --> 00:18:55.930 your plots on the left here.

NOTE Confidence: 0.8858033

00:18:55.930 --> 00:18:57.472 Well, they might, you know,

NOTE Confidence: 0.8858033

00:18:57.472 --> 00:18:59.008 might not even go out

NOTE Confidence: 0.87476665

00:18:59.010 --> 00:19:01.170 for review. The pre reviewer's will say

NOTE Confidence: 0.87476665

00:19:01.170 --> 00:19:03.012 you know your figures are inadequate

NOTE Confidence: 0.87476665

00:19:03.012 --> 00:19:04.861 for our instructions, authors etc etc.

NOTE Confidence: 0.87476665

00:19:04.861 --> 00:19:06.710 So I think some journals are

NOTE Confidence: 0.87476665

00:19:06.710 --> 00:19:08.864 coming around to this is the way

NOTE Confidence: 0.87476665

00:19:08.864 --> 00:19:11.020 we really want to see the data.

NOTE Confidence: 0.8233126

00:19:11.910 --> 00:19:14.584 Yeah, I think there is a shift

NOTE Confidence: 0.8233126

00:19:14.584 --> 00:19:16.492 in the paradigm, let's say,

NOTE Confidence: 0.8233126

00:19:16.492 --> 00:19:18.397 and it will take years.

NOTE Confidence: 0.8233126

00:19:18.400 --> 00:19:21.224 But for example, I have a slide here

NOTE Confidence: 0.8233126

00:19:21.224 --> 00:19:24.126 where so this is from my experience,

NOTE Confidence: 0.8233126

00:19:24.130 --> 00:19:25.478 so that for example,

NOTE Confidence: 0.8233126

00:19:25.478 --> 00:19:28.000 all the family of the network journals

NOTE Confidence: 0.8233126

00:19:28.000 --> 00:19:31.010 have already this policies for the figure.

NOTE Confidence: 0.8233126

00:19:31.010 --> 00:19:34.041 So this is something I received after

NOTE Confidence: 0.8233126

00:19:34.041 --> 00:19:37.593 the review of a paper as an editorial

NOTE Confidence: 0.8233126

00:19:37.593 --> 00:19:40.779 guidelines and the food for these like.

NOTE Confidence: 0.8233126

00:19:40.780 --> 00:19:44.028 Policies that I had to change a

NOTE Confidence: 0.8233126

00:19:44.028 --> 00:19:47.240 lot of figures and you see that.

NOTE Confidence: 0.8233126

00:19:47.240 --> 00:19:48.983 And so that the one of the

NOTE Confidence: 0.8233126

00:19:48.983 --> 00:19:50.360 policy as you see here,

NOTE Confidence: 0.8233126

00:19:50.360 --> 00:19:52.495 the last one is that for sample

NOTE Confidence: 0.8233126

00:19:52.495 --> 00:19:54.368 size that are less than 10.

NOTE Confidence: 0.8233126

00:19:54.370 --> 00:19:56.714 And they want you to get to plot

NOTE Confidence: 0.8233126

00:19:56.714 --> 00:19:58.361 the individual data points and

NOTE Confidence: 0.8233126

00:19:58.361 --> 00:20:00.353 so they don't accept bar graphs.

NOTE Confidence: 0.8233126

00:20:00.360 --> 00:20:03.240 Got bargraphs anymore.

NOTE Confidence: 0.8233126

00:20:03.240 --> 00:20:04.980 And then, for example,

NOTE Confidence: 0.8233126

00:20:04.980 --> 00:20:08.094 if you have some statistics such as

NOTE Confidence: 0.8233126

00:20:08.094 --> 00:20:11.069 error bars with the lesson 3 replicates,

NOTE Confidence: 0.8233126

00:20:11.070 --> 00:20:12.618 you have to remove,

NOTE Confidence: 0.8233126

00:20:12.618 --> 00:20:15.483 remove them and you have to show

NOTE Confidence: 0.8233126

00:20:15.483 --> 00:20:18.249 to show the data without the

NOTE Confidence: 0.8233126

00:20:18.249 --> 00:20:20.210 statistics without the error.

NOTE Confidence: 0.8233126

00:20:20.210 --> 00:20:22.947 Then this also is a point that

NOTE Confidence: 0.8233126

00:20:22.947 --> 00:20:25.430 you usually is not satisfied.

NOTE Confidence: 0.8233126

00:20:25.430 --> 00:20:28.040 So when you plot some statistical

NOTE Confidence: 0.8233126

00:20:28.040 --> 00:20:28.910 significance values,

NOTE Confidence: 0.8233126

00:20:28.910 --> 00:20:31.062 they don't accept anymore,

NOTE Confidence: 0.8233126

00:20:31.062 --> 00:20:33.214 they start the stars.

NOTE Confidence: 0.8233126

00:20:33.220 --> 00:20:35.266 But you have to provide the

NOTE Confidence: 0.8233126

00:20:35.266 --> 00:20:37.529 precise P value in the figure.

NOTE Confidence: 0.8233126

00:20:37.530 --> 00:20:40.036 It means that you have some stars.

NOTE Confidence: 0.8233126

00:20:40.040 --> 00:20:42.455 You have to change the stars that

NOTE Confidence: 0.8233126

00:20:42.455 --> 00:20:44.699 converting start to the precise P

NOTE Confidence: 0.8233126

00:20:44.699 --> 00:20:46.589 value before before publishing and

NOTE Confidence: 0.8233126

00:20:46.589 --> 00:20:49.111 then also you have to provide the

NOTE Confidence: 0.8233126

00:20:49.111 --> 00:20:51.890 precise number size for each of your bars.

NOTE Confidence: 0.8233126

00:20:51.890 --> 00:20:52.610 For example,

NOTE Confidence: 0.8233126

00:20:52.610 --> 00:20:53.447 I mean I,

NOTE Confidence: 0.8233126

00:20:53.447 --> 00:20:55.859 I think in the past it was enough

NOTE Confidence: 0.8233126

00:20:55.859 --> 00:20:58.097 to provide a range like from

NOTE Confidence: 0.8233126

00:20:58.097 --> 00:21:00.150 three to six replicates,

NOTE Confidence: 0.8233126

00:21:00.150 --> 00:21:03.750 but now they really want the number for each.

NOTE Confidence: 0.8233126

00:21:03.750 --> 00:21:05.460 For each app and population,

NOTE Confidence: 0.8233126

00:21:05.460 --> 00:21:08.388 for each sample that you have.

NOTE Confidence: 0.8233126

00:21:08.390 --> 00:21:09.734 So these are,

NOTE Confidence: 0.8233126

00:21:09.734 --> 00:21:11.974 in my experience were something

NOTE Confidence: 0.8233126

00:21:11.974 --> 00:21:14.688 that I had to provide that,

NOTE Confidence: 0.8233126

00:21:14.690 --> 00:21:18.290 but after the radio so it was not.

NOTE Confidence: 0.8233126

00:21:18.290 --> 00:21:21.640 It was the editorial like.

NOTE Confidence: 0.8233126

00:21:21.640 --> 00:21:24.426 At stage of acceptance of the paper,

NOTE Confidence: 0.8233126

00:21:24.430 --> 00:21:27.985 and I think this is true now for all

NOTE Confidence: 0.8233126

00:21:27.985 --> 00:21:31.539 the families of the of the natural.

NOTE Confidence: 0.8233126

00:21:31.540 --> 00:21:32.210 Jordans

NOTE Confidence: 0.902883

00:21:33.780 --> 00:21:35.680 it can I add something?

NOTE Confidence: 0.902883

00:21:35.680 --> 00:21:37.580 Although this is only for

NOTE Confidence: 0.902883

00:21:37.580 --> 00:21:39.480 publication that goal of publication,

NOTE Confidence: 0.902883

00:21:39.480 --> 00:21:42.066 but it’s important that we start

NOTE Confidence: 0.902883

00:21:42.066 --> 00:21:44.230 practicing all these rules in

NOTE Confidence: 0.902883

00:21:44.230 --> 00:21:46.420 our daily life because it’s so

NOTE Confidence: 0.902883

00:21:46.420 --> 00:21:48.904 painful that you have to do this

NOTE Confidence: 0.902883

00:21:48.904 --> 00:21:50.872 when you you’re trying to get

NOTE Confidence: 0.902883

00:21:50.880 --> 00:21:52.780 the figures into the Journal.

NOTE Confidence: 0.902883

00:21:52.780 --> 00:21:56.056 It’s a lot easier to do it while you’re

NOTE Confidence: 0.902883

00:21:56.056 --> 00:21:58.476 making the figures in real life.

NOTE Confidence: 0.8273201

00:21:59.090 --> 00:22:00.510 Yeah, so obviously it

NOTE Confidence: 0.8273201

00:22:00.510 --> 00:22:01.930 says worker before there.

NOTE Confidence: 0.8273201

00:22:01.930 --> 00:22:04.630 Yeah it says work because otherwise

NOTE Confidence: 0.8273201

00:22:04.630 --> 00:22:09.700 you have to repeat all day. Fevers so

NOTE Confidence: 0.79503936

00:22:09.700 --> 00:22:11.004 yeah also echo that,

NOTE Confidence: 0.79503936

00:22:11.004 --> 00:22:13.903 and also just want to say that you know

NOTE Confidence: 0.79503936

00:22:13.903 --> 00:22:16.650 I used to just use Excel to placings.

NOTE Confidence: 0.79503936

00:22:16.650 --> 00:22:19.074 But since my many of my lab members

NOTE Confidence: 0.79503936

00:22:19.074 --> 00:22:21.616 start to use Graphpad prism to plot,

NOTE Confidence: 0.79503936

00:22:21.620 --> 00:22:23.888 that makes a huge difference in

NOTE Confidence: 0.79503936

00:22:23.888 --> 00:22:25.400 converting between different types

NOTE Confidence: 0.79503936

00:22:25.460 --> 00:22:27.580 of parts such as this kind of things.

NOTE Confidence: 0.79503936

00:22:27.580 --> 00:22:29.890 If you had a bar bar graph,

NOTE Confidence: 0.79503936

00:22:29.890 --> 00:22:31.190 Indiana in that software,

NOTE Confidence: 0.79503936

00:22:31.190 --> 00:22:34.197 then you can very easily change that to a

NOTE Confidence: 0.79503936

00:22:34.197 --> 00:22:36.177 bar graph with different dots distributed.

NOTE Confidence: 0.79503936

00:22:36.180 --> 00:22:38.497 So it's very easy to work with.

NOTE Confidence: 0.7834576

00:22:39.850 --> 00:22:41.512 Yeah, that's also I have something

NOTE Confidence: 0.7834576

00:22:41.512 --> 00:22:43.670 at the end of the presentation.

NOTE Confidence: 0.7834576

00:22:43.670 --> 00:22:46.253 So basically there are a lot of tools now

NOTE Confidence: 0.7834576

00:22:46.253 --> 00:22:48.916 more or less commercial, but tequila.

NOTE Confidence: 0.7834576

00:22:48.916 --> 00:22:51.148 They aren't really available.

NOTE Confidence: 0.7834576

00:22:51.150 --> 00:22:54.718 U as which are too many different formats

NOTE Confidence: 0.7834576

00:22:54.718 --> 00:22:57.928 and starting with the same initial data,

NOTE Confidence: 0.7834576

00:22:57.930 --> 00:23:01.010 basically formatted as a table.

NOTE Confidence: 0.7834576

00:23:01.010 --> 00:23:03.404 So that from the same table you can switch

NOTE Confidence: 0.7834576

00:23:03.404 --> 00:23:06.108 to there too many different visualizations.

NOTE Confidence: 0.7834576

00:23:06.110 --> 00:23:07.834 So that's that's true,

NOTE Confidence: 0.7834576

00:23:07.834 --> 00:23:11.210 and it's probably easier also to plot these

NOTE Confidence: 0.7834576

00:23:11.210 --> 00:23:14.820 dots with single dots as it was in the past.

NOTE Confidence: 0.7834576

00:23:14.820 --> 00:23:18.140 Without respect.

NOTE Confidence: 0.7834576

00:23:18.140 --> 00:23:21.880 OK, so that was the main point of this part.

NOTE Confidence: 0.7834576

00:23:21.880 --> 00:23:24.498 I had a part on the standard

NOTE Confidence: 0.7834576

00:23:24.498 --> 00:23:25.620 deviation standard error.

NOTE Confidence: 0.7834576

00:23:25.620 --> 00:23:27.500 That's another issue because the

NOTE Confidence: 0.7834576

00:23:27.500 --> 00:23:29.380 standard error is basically the

NOTE Confidence: 0.7834576

00:23:29.441 --> 00:23:31.517 standard deviation divided by the square

NOTE Confidence: 0.7834576

00:23:31.517 --> 00:23:33.850 root of the number of experiments,

NOTE Confidence: 0.7834576

00:23:33.850 --> 00:23:35.715 and so usually the standard

NOTE Confidence: 0.7834576

00:23:35.715 --> 00:23:36.834 error is displayed.

NOTE Confidence: 0.7834576

00:23:36.840 --> 00:23:39.680 But you have just be careful that it's

NOTE Confidence: 0.7834576

00:23:39.680 --> 00:23:42.741 a measure that tends to go to zero

NOTE Confidence: 0.7834576

00:23:42.741 --> 00:23:45.133 just because they increase the number

NOTE Confidence: 0.7834576

00:23:45.133 --> 00:23:47.954 of replicates or the number of points.

NOTE Confidence: 0.7834576

00:23:47.960 --> 00:23:50.186 So you see an example here where

NOTE Confidence: 0.7834576

00:23:50.186 --> 00:23:52.383 it seems by plotting the standard

NOTE Confidence: 0.7834576

00:23:52.383 --> 00:23:55.113 error that the black bar and the

NOTE Confidence: 0.7834576

00:23:55.188 --> 00:23:57.666 white bar have the same like measure

NOTE Confidence: 0.7834576

00:23:57.666 --> 00:23:59.484 of spread of the data.

NOTE Confidence: 0.7834576

00:23:59.484 --> 00:24:02.123 But if you look at the standard

NOTE Confidence: 0.7834576

00:24:02.123 --> 00:24:04.244 deviation you see that this is

NOTE Confidence: 0.7834576

00:24:04.244 --> 00:24:05.909 an effect of the factor.

NOTE Confidence: 0.7834576

00:24:05.910 --> 00:24:08.416 Today the Black bar has higher spread,

NOTE Confidence: 0.7834576

00:24:08.420 --> 00:24:09.856 but also more points,

NOTE Confidence: 0.7834576

00:24:09.856 --> 00:24:11.651 and that’s why the standard

NOTE Confidence: 0.7834576

00:24:11.651 --> 00:24:13.450 error seems seems the same.

NOTE Confidence: 0.8393723

00:24:16.020 --> 00:24:18.010 So that’s another another issue.

NOTE Confidence: 0.8393723

00:24:18.010 --> 00:24:20.332 So obviously for publication at the

NOTE Confidence: 0.8393723

00:24:20.332 --> 00:24:23.169 standard error of the mean is preferred,

NOTE Confidence: 0.8393723

00:24:23.170 --> 00:24:26.308 because it usually gives an impression

NOTE Confidence: 0.8393723

00:24:26.308 --> 00:24:29.620 of the data being less sparse.

NOTE Confidence: 0.8393723

00:24:29.620 --> 00:24:31.180 But especially with different

NOTE Confidence: 0.8393723

00:24:31.180 --> 00:24:33.130 number of samples in different

NOTE Confidence: 0.8393723

00:24:33.130 --> 00:24:35.379 in different bars that it could.

NOTE Confidence: 0.8393723

00:24:35.380 --> 00:24:36.920 This could be misleading.

NOTE Confidence: 0.85438687

00:24:39.890 --> 00:24:41.822 And all these issues were presented

NOTE Confidence: 0.85438687

00:24:41.822 --> 00:24:43.861 in these in this paper published

NOTE Confidence: 0.85438687

00:24:43.861 --> 00:24:46.269 five years ago in in plus biology.

NOTE Confidence: 0.8659432

00:24:48.760 --> 00:24:50.044 I would skip this,

NOTE Confidence: 0.8659432

00:24:50.044 --> 00:24:52.450 just that we will touch this later,

NOTE Confidence: 0.8659432

00:24:52.450 --> 00:24:54.125 but an alternative solution if

NOTE Confidence: 0.8659432

00:24:54.125 --> 00:24:55.800 you have enough data points.

NOTE Confidence: 0.8659432

00:24:55.800 --> 00:24:58.789 So I would say more than 10.

NOTE Confidence: 0.8659432

00:24:58.790 --> 00:25:00.514 An alternative solution instead

NOTE Confidence: 0.8659432

00:25:00.514 --> 00:25:03.100 of showing like but lotsa Ann

NOTE Confidence: 0.8659432

00:25:03.176 --> 00:25:05.126 is to show the distribution of

NOTE Confidence: 0.8659432

00:25:05.126 --> 00:25:07.420 the data is box whisker plot.

NOTE Confidence: 0.8659432

00:25:07.420 --> 00:25:10.764 As you see here they have some light

NOTE Confidence: 0.8659432

00:25:10.764 --> 00:25:13.559 model with more details on this.

NOTE Confidence: 0.8659432

00:25:13.560 --> 00:25:17.088 OK, so this is the next example.

NOTE Confidence: 0.8659432

00:25:17.090 --> 00:25:19.484 I think it's a biplot with the

NOTE Confidence: 0.8659432

00:25:19.484 --> 00:25:22.110 usage of the different browsers,

NOTE Confidence: 0.8659432

00:25:22.110 --> 00:25:25.876 so this is extra science image so.

NOTE Confidence: 0.8659432

00:25:25.880 --> 00:25:28.664 So this is a classic example

NOTE Confidence: 0.8659432

00:25:28.664 --> 00:25:30.520 in like visualization lessons.

NOTE Confidence: 0.8659432

00:25:30.520 --> 00:25:33.768 So what could we run with this?

NOTE Confidence: 0.86750424

00:25:39.120 --> 00:25:44.000 There's no end. Yeah, so that's a yes,

NOTE Confidence: 0.86750424

00:25:44.000 --> 00:25:47.456 so there is no endless so that you cannot.

NOTE Confidence: 0.86750424

00:25:47.460 --> 00:25:49.758 You don't know of how many,

NOTE Confidence: 0.86750424

00:25:49.760 --> 00:25:52.376 how many data points you use that in

NOTE Confidence: 0.86750424

00:25:52.376 --> 00:25:55.140 order to build the other frequencies.

NOTE Confidence: 0.86750424

00:25:55.140 --> 00:25:59.256 Obviously pie charts are used to display

NOTE Confidence: 0.86750424

00:25:59.256 --> 00:26:01.800 frequencies and proportions of some

NOTE Confidence: 0.86750424

00:26:01.800 --> 00:26:05.259 classes that sum up to 100 or or to one.

NOTE Confidence: 0.86750424

00:26:05.260 --> 00:26:07.156 The main problem is that so

NOTE Confidence: 0.86750424

00:26:07.156 --> 00:26:09.500 the idea is that you shouldn't.

NOTE Confidence: 0.86750424

00:26:09.500 --> 00:26:12.370 You should avoid by chance.

NOTE Confidence: 0.86750424

00:26:12.370 --> 00:26:14.806 So the idea for displaying an

NOTE Confidence: 0.86750424

00:26:14.806 --> 00:26:17.069 information of our proportion or of

NOTE Confidence: 0.86750424

00:26:17.069 --> 00:26:19.325 a percentage as a pie chart are is.

NOTE Confidence: 0.816895016470588

00:26:21.470 --> 00:26:23.318 Not the best choice.

NOTE Confidence: 0.816895016470588

00:26:23.318 --> 00:26:26.677 Because that it was shown that humans

NOTE Confidence: 0.816895016470588

00:26:26.677 --> 00:26:29.629 are very bad at reading angles,

NOTE Confidence: 0.816895016470588

00:26:29.630 --> 00:26:31.870 so we're not very precise,

NOTE Confidence: 0.816895016470588

00:26:31.870 --> 00:26:34.870 precise in understanding differences between

NOTE Confidence: 0.816895016470588

00:26:34.870 --> 00:26:39.250 angles and so between the designs of the.

NOTE Confidence: 0.816895016470588

00:26:39.250 --> 00:26:43.543 Slices of the pie and so usually if you

NOTE Confidence: 0.816895016470588

00:26:43.543 --> 00:26:47.175 convert the pie chart into a bar plot.

NOTE Confidence: 0.816895016470588

00:26:47.180 --> 00:26:48.690 Information is much more clear.

NOTE Confidence: 0.816895016470588

00:26:48.690 --> 00:26:50.805 It's true that the pie

NOTE Confidence: 0.816895016470588

00:26:50.805 --> 00:26:52.497 chart is more aesthetic.

NOTE Confidence: 0.816895016470588

00:26:52.500 --> 00:26:54.540 Appeared, but the bar plotter

NOTE Confidence: 0.816895016470588

00:26:54.540 --> 00:26:56.580 is in in any circumstances,

NOTE Confidence: 0.816895016470588

00:26:56.580 --> 00:26:59.028 usually more affecting in displaying girl.

NOTE Confidence: 0.816895016470588

00:26:59.030 --> 00:27:01.555 For example, differences in the

NOTE Confidence: 0.816895016470588

00:27:01.555 --> 00:27:04.080 usage of this genome browsers.

NOTE Confidence: 0.816895016470588

00:27:04.080 --> 00:27:06.910 So this has been a long issue and if you

NOTE Confidence: 0.816895016470588

00:27:06.989 --> 00:27:09.866 in many presentation so there is always

NOTE Confidence: 0.816895016470588

00:27:09.866 --> 00:27:13.017 this suggestion to avoid at all a pie charts.

NOTE Confidence: 0.816895016470588

00:27:13.020 --> 00:27:15.428 There are also some example of these.

NOTE Confidence: 0.816895016470588

00:27:15.430 --> 00:27:18.780 So these are three pie charts and you can see

NOTE Confidence: 0.816895016470588

00:27:18.857 --> 00:27:22.209 that it's they are different from each other.

NOTE Confidence: 0.816895016470588

00:27:22.210 --> 00:27:23.995 But it's very difficult to

NOTE Confidence: 0.816895016470588

00:27:23.995 --> 00:27:25.423 understand that the difference,

NOTE Confidence: 0.816895016470588

00:27:25.430 --> 00:27:28.118 so the difference is is in the size

NOTE Confidence: 0.816895016470588

00:27:28.118 --> 00:27:30.796 of the slice of the three pies,

NOTE Confidence: 0.816895016470588

00:27:30.800 --> 00:27:32.232 but it's very different.

NOTE Confidence: 0.816895016470588

00:27:32.232 --> 00:27:34.810 For example, to understand in each

NOTE Confidence: 0.816895016470588

00:27:34.810 --> 00:27:38.350 pie which one is the largest slides.

NOTE Confidence: 0.816895016470588

00:27:38.350 --> 00:27:40.975 And to draw comparison it much more

NOTE Confidence: 0.816895016470588

00:27:40.975 --> 00:27:43.978 more easier to understand these issues.

NOTE Confidence: 0.816895016470588

00:27:43.980 --> 00:27:47.508 So which pie is larger if the information is

NOTE Confidence: 0.816895016470588

00:27:47.508 --> 00:27:51.340 not displaced is not displayed as pie charts,

NOTE Confidence: 0.816895016470588

00:27:51.340 --> 00:27:54.379 but as market.

NOTE Confidence: 0.816895016470588

00:27:54.380 --> 00:27:56.010 So that's on the web.

NOTE Confidence: 0.816895016470588

00:27:56.010 --> 00:27:58.520 I also found these provocative.

NOTE Confidence: 0.816895016470588

00:27:58.520 --> 00:28:00.608 Label of pie charts as lighters.

NOTE Confidence: 0.8758027

00:28:02.790 --> 00:28:05.814 So in general it would be better to avoid

NOTE Confidence: 0.8758027

00:28:05.814 --> 00:28:07.827 displaying information as pie chart.

NOTE Confidence: 0.8758027

00:28:07.830 --> 00:28:09.804 And prefer a bar chart instead

NOTE Confidence: 0.8758027

00:28:09.804 --> 00:28:12.030 to show the same information.

NOTE Confidence: 0.89527786

00:28:14.400 --> 00:28:16.010 OK, so that was faster.

NOTE Confidence: 0.89527786

00:28:16.010 --> 00:28:17.506 This is another example.

NOTE Confidence: 0.89527786

00:28:17.506 --> 00:28:20.170 What could be wrong with this plot?

NOTE Confidence: 0.89527786

00:28:20.170 --> 00:28:21.420 Again, we have a treatment.

NOTE Confidence: 0.89527786

00:28:21.420 --> 00:28:22.360 We have a control.

NOTE Confidence: 0.89527786

00:28:22.360 --> 00:28:24.150 This time we see the data point.

NOTE Confidence: 0.8554094

00:28:26.330 --> 00:28:28.692 Scale is so wrong, so it covers

NOTE Confidence: 0.8554094

00:28:28.692 --> 00:28:30.800 the distribution of the lower end.

NOTE Confidence: 0.8554094

00:28:31.390 --> 00:28:34.214 Yes, exactly so this is a case where

NOTE Confidence: 0.8554094

00:28:34.214 --> 00:28:36.927 most of the data are compressed,

NOTE Confidence: 0.8554094

00:28:36.930 --> 00:28:39.306 since they have very different magnitude.

NOTE Confidence: 0.8554094

00:28:39.310 --> 00:28:41.932 Most of the data are compressed

NOTE Confidence: 0.8554094

00:28:41.932 --> 00:28:45.850 air in a very small part of the

NOTE Confidence: 0.8554094

00:28:45.850 --> 00:28:48.440 plot and we cannot understand.

NOTE Confidence: 0.8554094

00:28:48.440 --> 00:28:50.684 Very much how they are distributed

NOTE Confidence: 0.8554094

00:28:50.684 --> 00:28:53.392 because most of the plotter is related

NOTE Confidence: 0.8554094

00:28:53.392 --> 00:28:55.660 to these kind of two outliers.

NOTE Confidence: 0.8554094

00:28:55.660 --> 00:28:58.140 So this is an issue with the measures

NOTE Confidence: 0.8554094

00:28:58.140 --> 00:29:00.220 that have different magnitudes,

NOTE Confidence: 0.8554094

00:29:00.220 --> 00:29:03.260 so it could in my experience it happens.

NOTE Confidence: 0.8554094

00:29:03.260 --> 00:29:07.100 For example in gene expression measurements.

NOTE Confidence: 0.8554094

00:29:07.100 --> 00:29:08.328 Because they can vary,

NOTE Confidence: 0.8554094

00:29:08.328 --> 00:29:09.556 especially with the sequencing.

NOTE Confidence: 0.8554094

00:29:09.560 --> 00:29:12.955 They can value of four to five.

NOTE Confidence: 0.8554094

00:29:12.960 --> 00:29:16.136 Magnitude and the main way to solve this

NOTE Confidence: 0.8554094

00:29:16.136 --> 00:29:19.129 issue is to log transform the data.

NOTE Confidence: 0.8554094

00:29:19.130 --> 00:29:21.254 So instead of plotting in a

NOTE Confidence: 0.8554094

00:29:21.254 --> 00:29:23.650 linear scale to log normalizing,

NOTE Confidence: 0.8554094

00:29:23.650 --> 00:29:26.205 the scale of the data and this

NOTE Confidence: 0.8554094

00:29:26.205 --> 00:29:28.327 allows to restrict the distance

NOTE Confidence: 0.8554094

00:29:28.327 --> 00:29:30.215 between these two points,

NOTE Confidence: 0.8554094

00:29:30.220 --> 00:29:31.064 the outliers,

NOTE Confidence: 0.8554094

00:29:31.064 --> 00:29:34.018 and allow you to see also the

NOTE Confidence: 0.8554094

00:29:34.018 --> 00:29:35.568 distribution of the points.

NOTE Confidence: 0.8554094

00:29:35.570 --> 00:29:37.620 That here seems all compressed.

NOTE Confidence: 0.866823

00:29:39.770 --> 00:29:42.885 So usually log transformation allow you to

NOTE Confidence: 0.866823

00:29:42.885 --> 00:29:45.344 capture some information on the difference

NOTE Confidence: 0.866823

00:29:45.344 --> 00:29:47.808 of your points that are more clear.

NOTE Confidence: 0.866823

00:29:47.810 --> 00:29:50.225 Not in all cases, but in some

NOTE Confidence: 0.866823

00:29:50.225 --> 00:29:52.279 cases rather than displaying the

NOTE Confidence: 0.866823

00:29:52.279 --> 00:29:54.639 information in a linear scale,

NOTE Confidence: 0.866823

00:29:54.640 --> 00:29:57.832 especially when you have a lot of range

NOTE Confidence: 0.866823

00:29:57.832 --> 00:30:00.249 between your minimal and maximal.

NOTE Confidence: 0.866823

00:30:00.250 --> 00:30:03.913 Measurements. An alternative way.

NOTE Confidence: 0.866823

00:30:03.913 --> 00:30:07.279 Is not also a panel breaks,

NOTE Confidence: 0.866823

00:30:07.280 --> 00:30:10.454 so personally I prefer log log

NOTE Confidence: 0.866823

00:30:10.454 --> 00:30:13.259 transformation over panel breaker because

NOTE Confidence: 0.866823

00:30:13.259 --> 00:30:16.289 there is mathematically more likely.

NOTE Confidence: 0.866823

00:30:16.290 --> 00:30:17.505 Linear or elegant,

NOTE Confidence: 0.866823

00:30:17.505 --> 00:30:20.340 but there are situations where you can.

NOTE Confidence: 0.866823

00:30:20.340 --> 00:30:23.580 You can choose so this is an example.

NOTE Confidence: 0.866823

00:30:23.580 --> 00:30:25.610 You have a bar chart.

NOTE Confidence: 0.866823

00:30:25.610 --> 00:30:28.208 You have a huge difference between

NOTE Confidence: 0.866823

00:30:28.208 --> 00:30:30.868 the measurements of a 2D and E&F.

NOTE Confidence: 0.866823

00:30:30.870 --> 00:30:34.965 So this is how you solve the problem by

NOTE Confidence: 0.866823

00:30:34.965 --> 00:30:37.649 introducing a breaker in your panel.

NOTE Confidence: 0.866823

00:30:37.650 --> 00:30:41.293 So from 25 to 200 to 210 and this is the

NOTE Confidence: 0.866823

00:30:41.293 --> 00:30:44.418 equivalent solution by log transformation.

NOTE Confidence: 0.866823

00:30:44.420 --> 00:30:45.824 As you see,

NOTE Confidence: 0.866823

00:30:45.824 --> 00:30:48.632 the solution that the two solutions

NOTE Confidence: 0.866823

00:30:48.632 --> 00:30:51.440 give a fight a similar result.

NOTE Confidence: 0.866823

00:30:51.440 --> 00:30:54.344 But here you insert the manual break of

NOTE Confidence: 0.866823

00:30:54.344 --> 00:30:57.348 the data and this could be misleading.

NOTE Confidence: 0.866823

00:30:57.350 --> 00:30:59.716 Here you saw the issue by log

NOTE Confidence: 0.866823

00:30:59.716 --> 00:31:01.680 transforming all the measurements.

NOTE Confidence: 0.866823

00:31:01.680 --> 00:31:04.448 So this is for example is an advantage

NOTE Confidence: 0.866823

00:31:04.448 --> 00:31:07.198 because it affects all the measurement.

NOTE Confidence: 0.866823

00:31:07.200 --> 00:31:09.170 And while this panel breaker

NOTE Confidence: 0.866823

00:31:09.170 --> 00:31:10.746 affects only for example,

NOTE Confidence: 0.866823

00:31:10.750 --> 00:31:13.894 these two bars and could distort the data.

NOTE Confidence: 0.798156

00:31:18.080 --> 00:31:22.238 Another another scenario where you should

NOTE Confidence: 0.798156

00:31:22.238 --> 00:31:25.690 consider log transformation is these.

NOTE Confidence: 0.798156

00:31:25.690 --> 00:31:30.163 This could be a plotter that shows for gene

NOTE Confidence: 0.798156

00:31:30.163 --> 00:31:33.339 expression levels from Aaron Isike for.

NOTE Confidence: 0.798156

00:31:33.340 --> 00:31:34.546 Population of jeans.

NOTE Confidence: 0.798156

00:31:34.546 --> 00:31:37.360 So each gene could be a doctor

NOTE Confidence: 0.798156

00:31:37.441 --> 00:31:38.937 and he received it.

NOTE Confidence: 0.798156

00:31:38.940 --> 00:31:41.106 There is a different year age,

NOTE Confidence: 0.798156

00:31:41.110 --> 00:31:43.702 but you see that there are outliers like

NOTE Confidence: 0.798156

00:31:43.702 --> 00:31:46.158 for example genes of ribosomal proteins.

NOTE Confidence: 0.798156

00:31:46.160 --> 00:31:49.440 Histones usually are in these.

NOTE Confidence: 0.798156

00:31:49.440 --> 00:31:51.638 Are in this part of the plot,

NOTE Confidence: 0.798156

00:31:51.640 --> 00:31:54.126 but most of the gene are 90% of

NOTE Confidence: 0.798156

00:31:54.126 --> 00:31:56.414 your jeans are in this part of the

NOTE Confidence: 0.798156

00:31:56.414 --> 00:31:58.550 plot and you cannot really see.

NOTE Confidence: 0.8560322

00:32:01.170 --> 00:32:02.960 You cannot really inspect them

NOTE Confidence: 0.8560322

00:32:02.960 --> 00:32:05.214 because most of the plot is

NOTE Confidence: 0.8560322

00:32:05.214 --> 00:32:06.778 dedicated to some outliers.

NOTE Confidence: 0.8560322

00:32:06.780 --> 00:32:08.778 So again, here is a situation

NOTE Confidence: 0.8560322

00:32:08.778 --> 00:32:10.890 where you can log transform.

NOTE Confidence: 0.8560322

00:32:10.890 --> 00:32:12.374 Both are the coordinates.

NOTE Confidence: 0.8560322

00:32:12.374 --> 00:32:15.050 So let's say that here is the

NOTE Confidence: 0.8560322

00:32:15.050 --> 00:32:17.276 control and this is the treatment

NOTE Confidence: 0.8560322

00:32:17.276 --> 00:32:20.287 and this will allow you to see more

NOTE Confidence: 0.8560322

00:32:20.287 --> 00:32:22.485 in detail the differences in the

NOTE Confidence: 0.8560322

00:32:22.485 --> 00:32:25.110 expression of the bug of your jeans.

NOTE Confidence: 0.8636312

00:32:28.250 --> 00:32:30.420 In a situation like Visa,

NOTE Confidence: 0.8636312

00:32:30.420 --> 00:32:32.148 you should also consider

NOTE Confidence: 0.8636312

00:32:32.148 --> 00:32:33.876 issue if you're interested,

NOTE Confidence: 0.8636312

00:32:33.880 --> 00:32:36.550 for example in showing differences in

NOTE Confidence: 0.8636312

00:32:36.550 --> 00:32:39.079 expression between 3 between a control.

NOTE Confidence: 0.8636312

00:32:39.080 --> 00:32:41.240 For example, are one and

NOTE Confidence: 0.8636312

00:32:41.240 --> 00:32:42.968 the treatment are two.

NOTE Confidence: 0.8636312

00:32:42.970 --> 00:32:47.989 You have also the possibility to show.

NOTE Confidence: 0.8636312

00:32:47.990 --> 00:32:49.622 As the Y axis,

NOTE Confidence: 0.8636312

00:32:49.622 --> 00:32:52.070 the differences in the log values.

NOTE Confidence: 0.8636312

00:32:52.070 --> 00:32:53.291 So this representation

NOTE Confidence: 0.8636312

00:32:53.291 --> 00:32:55.733 here is the same as this,

NOTE Confidence: 0.8636312

00:32:55.740 --> 00:32:57.685 but it maximizes the visualization

NOTE Confidence: 0.8636312

00:32:57.685 --> 00:32:59.630 of the differences in the

NOTE Confidence: 0.8636312

00:32:59.698 --> 00:33:01.450 expression levels of genes.

NOTE Confidence: 0.8636312

00:33:01.450 --> 00:33:04.006 So this is something that you

NOTE Confidence: 0.8636312

00:33:04.006 --> 00:33:07.169 find a cold as as an MA plot.

NOTE Confidence: 0.8636312

00:33:07.170 --> 00:33:09.205 It was introduced with the

NOTE Confidence: 0.8636312

00:33:09.205 --> 00:33:10.833 analysis of microarray data,

NOTE Confidence: 0.8636312

00:33:10.840 --> 00:33:13.282 but you can find it also

NOTE Confidence: 0.8636312

00:33:13.282 --> 00:33:14.503 with sequencing data.

NOTE Confidence: 0.8636312

00:33:14.510 --> 00:33:16.190 Sometimes these two different

NOTE Confidence: 0.8636312

00:33:16.190 --> 00:33:17.450 visualization are used.

NOTE Confidence: 0.8636312

00:33:17.450 --> 00:33:19.585 Depending on the aim of the figure,

NOTE Confidence: 0.8636312

00:33:19.590 --> 00:33:21.420 so sometimes you will find these,

NOTE Confidence: 0.8636312

00:33:21.420 --> 00:33:23.046 especially when the message of the

NOTE Confidence: 0.8636312

00:33:23.046 --> 00:33:25.208 figure is that you don't see big

NOTE Confidence: 0.8636312

00:33:25.208 --> 00:33:26.903 differences between the two conditions,

NOTE Confidence: 0.8636312

00:33:26.910 --> 00:33:29.448 while if the message is that you find big

NOTE Confidence: 0.8636312

00:33:29.448 --> 00:33:31.180 differences between the two condition,

NOTE Confidence: 0.8636312

00:33:31.180 --> 00:33:34.840 you will find mostly these visualization.

NOTE Confidence: 0.8636312

00:33:34.840 --> 00:33:37.416 So here I would just point out that

NOTE Confidence: 0.8636312

00:33:37.416 --> 00:33:40.439 in any at any sequencing experiments,

NOTE Confidence: 0.8636312

00:33:40.440 --> 00:33:43.772 you will probably never find any gene

NOTE Confidence: 0.8636312

00:33:43.772 --> 00:33:47.590 that is in this area because they.

NOTE Confidence: 0.8636312

00:33:47.590 --> 00:33:49.520 But most of the genes,

NOTE Confidence: 0.8636312

00:33:49.520 --> 00:33:51.440 the main difference they make

NOTE Confidence: 0.8636312

00:33:51.440 --> 00:33:52.976 the main like determinant,

NOTE Confidence: 0.8636312

00:33:52.980 --> 00:33:54.910 is the basil expression levels.

NOTE Confidence: 0.8636312

00:33:54.910 --> 00:33:57.220 So usually your perturbations do not

NOTE Confidence: 0.8636312

00:33:57.220 --> 00:34:00.298 affect so much the expression of a gene,

NOTE Confidence: 0.8636312

00:34:00.300 --> 00:34:02.771 so that the gene is in these

NOTE Confidence: 0.8636312

00:34:02.771 --> 00:34:05.687 area of the plot of the oranges.

NOTE Confidence: 0.8636312

00:34:05.690 --> 00:34:07.610 And that's why this visualization

NOTE Confidence: 0.8636312

00:34:07.610 --> 00:34:10.201 is much more efficient in capturing

NOTE Confidence: 0.8636312

00:34:10.201 --> 00:34:12.007 the expression differences.

NOTE Confidence: 0.8636312

00:34:12.010 --> 00:34:14.596 Because they scale on on the

NOTE Confidence: 0.8636312

00:34:14.596 --> 00:34:15.889 expression at baseline.

NOTE Confidence: 0.91227716

00:34:19.000 --> 00:34:21.385 OK, so now I have a section I don't

NOTE Confidence: 0.91227716

00:34:21.385 --> 00:34:23.555 know that I'm I have a section

NOTE Confidence: 0.91227716

00:34:23.555 --> 00:34:25.780 about how to display distributions.

NOTE Confidence: 0.8261615

00:34:28.500 --> 00:34:29.428 So let's say that

NOTE Confidence: 0.8261615

00:34:29.430 --> 00:34:31.626 we have a display. One time you had 15

NOTE Confidence: 0.8261615

00:34:31.626 --> 00:34:33.600 minutes and if we go a little over, that's

NOTE Confidence: 0.855771083333333

00:34:33.600 --> 00:34:36.673 OK, OK? So when you have to

NOTE Confidence: 0.855771083333333

00:34:36.673 --> 00:34:38.729 represent the distribution of data,

NOTE Confidence: 0.855771083333333

00:34:38.730 --> 00:34:40.134 you have many choices.

NOTE Confidence: 0.855771083333333

00:34:40.134 --> 00:34:43.509 The histogram is one of the most used choice.

NOTE Confidence: 0.855771083333333

00:34:43.510 --> 00:34:48.030 It has the advantage that it can present.

NOTE Confidence: 0.855771083333333

00:34:48.030 --> 00:34:51.156 With detail, the shape of the

NOTE Confidence: 0.855771083333333

00:34:51.156 --> 00:34:53.240 distribution of your data.

NOTE Confidence: 0.855771083333333

00:34:53.240 --> 00:34:56.112 And so basically you have a variable of

NOTE Confidence: 0.855771083333333

00:34:56.112 --> 00:34:58.395 interest that usually is a continuous

NOTE Confidence: 0.855771083333333

00:34:58.395 --> 00:35:00.633 variable and you wanted to show

NOTE Confidence: 0.855771083333333

00:35:00.702 --> 00:35:02.937 how this variable is distributed.

NOTE Confidence: 0.855771083333333

00:35:02.940 --> 00:35:06.081 So you divide the range of the values in

NOTE Confidence: 0.855771083333333

00:35:06.081 --> 00:35:09.218 some beans and then you count the number

NOTE Confidence: 0.855771083333333

00:35:09.218 --> 00:35:12.250 of points that fall inside each being.

NOTE Confidence: 0.855771083333333

00:35:12.250 --> 00:35:14.236 The issue with the histograms is

NOTE Confidence: 0.855771083333333

00:35:14.236 --> 00:35:16.818 that you should be careful when when

NOTE Confidence: 0.855771083333333

00:35:16.818 --> 00:35:19.194 building the histograms and when looking

NOTE Confidence: 0.855771083333333

00:35:19.194 --> 00:35:21.805 at the histograms that there are

NOTE Confidence: 0.855771083333333

00:35:21.805 --> 00:35:23.950 some are being arbitrary parameters.

NOTE Confidence: 0.855771083333333

00:35:23.950 --> 00:35:25.870 In building up his histogram,

NOTE Confidence: 0.855771083333333

00:35:25.870 --> 00:35:29.566 mainly the choice of the bin size.

NOTE Confidence: 0.855771083333333

00:35:29.570 --> 00:35:32.794 So this is an example where the same

NOTE Confidence: 0.855771083333333

00:35:32.794 --> 00:35:35.270 distribution of data that is the

NOTE Confidence: 0.855771083333333

00:35:35.270 --> 00:35:37.646 distribution of the price of abedy

NOTE Confidence: 0.855771083333333

00:35:37.724 --> 00:35:40.190 apartments in French City has been

NOTE Confidence: 0.855771083333333

00:35:40.190 --> 00:35:42.642 being there in two different ways.

NOTE Confidence: 0.855771083333333

00:35:42.642 --> 00:35:46.750 So here is the price and hear the bin sizes.

NOTE Confidence: 0.855771083333333

00:35:46.750 --> 00:35:52.080 So the size of each of the bin is 10.

NOTE Confidence: 0.855771083333333

00:35:52.080 --> 00:35:52.516 Dollars.

NOTE Confidence: 0.855771083333333

00:35:52.516 --> 00:35:56.440 While it in here on the writer it is

NOTE Confidence: 0.855771083333333

00:35:56.538 --> 00:36:00.462 of $2 so you can see that using more

NOTE Confidence: 0.855771083333333

00:36:00.462 --> 00:36:03.725 granular bins allow you to see some

NOTE Confidence: 0.855771083333333

00:36:03.725 --> 00:36:05.982 the presence of some accumulations

NOTE Confidence: 0.855771083333333

00:36:05.982 --> 00:36:09.006 in your data that you cannot really

NOTE Confidence: 0.855771083333333

00:36:09.006 --> 00:36:11.437 see with the larger bin size,

NOTE Confidence: 0.855771083333333

00:36:11.440 --> 00:36:14.398 and this could be important because

NOTE Confidence: 0.855771083333333

00:36:14.398 --> 00:36:16.370 these accumulation this probably

NOTE Confidence: 0.855771083333333

00:36:16.451 --> 00:36:18.629 are accumulation of price that are

NOTE Confidence: 0.855771083333333

00:36:18.629 --> 00:36:21.751 due to the fact that they are prices

NOTE Confidence: 0.855771083333333

00:36:21.751 --> 00:36:23.347 that are commonly used.

NOTE Confidence: 0.855771083333333

00:36:23.350 --> 00:36:25.738 By many different Airbnbs, for example,

NOTE Confidence: 0.855771083333333

00:36:25.740 --> 00:36:28.916 because they are multipliers of 50 or 100,

NOTE Confidence: 0.855771083333333

00:36:28.920 --> 00:36:29.726 for example.

NOTE Confidence: 0.855771083333333

00:36:29.726 --> 00:36:32.547 But the fact is that depending on

NOTE Confidence: 0.855771083333333

00:36:32.547 --> 00:36:34.489 the choice of the bin,

NOTE Confidence: 0.855771083333333

00:36:34.490 --> 00:36:37.920 you see a different story.

NOTE Confidence: 0.855771083333333

00:36:37.920 --> 00:36:41.502 And then you should be always

NOTE Confidence: 0.855771083333333

00:36:41.502 --> 00:36:44.600 careful to select bin size.

NOTE Confidence: 0.855771083333333

00:36:44.600 --> 00:36:48.878 That doesn't affect too much data.

NOTE Confidence: 0.855771083333333

00:36:48.880 --> 00:36:51.220 There are also software tools

NOTE Confidence: 0.855771083333333

00:36:51.220 --> 00:36:53.092 that calculates depending on

NOTE Confidence: 0.855771083333333

00:36:53.092 --> 00:36:55.569 your data depending on squared,

NOTE Confidence: 0.855771083333333

00:36:55.570 --> 00:36:58.264 your points are placed the best

NOTE Confidence: 0.855771083333333

00:36:58.264 --> 00:37:02.147 and size of the bins so that you

NOTE Confidence: 0.855771083333333

00:37:02.147 --> 00:37:05.129 reduce the distortion of your data.

NOTE Confidence: 0.79675037

00:37:08.710 --> 00:37:10.290 An alternative way to represent

NOTE Confidence: 0.79675037

00:37:10.290 --> 00:37:12.669 distribution is to use a density plot.

NOTE Confidence: 0.79675037

00:37:12.670 --> 00:37:15.214 So a density plot is basically

NOTE Confidence: 0.79675037

00:37:15.214 --> 00:37:17.910 a smoothing of a histogram.

NOTE Confidence: 0.79675037

00:37:17.910 --> 00:37:20.528 Here you collect being said and here

NOTE Confidence: 0.79675037

00:37:20.528 --> 00:37:23.379 use motor the shape of the distribution

NOTE Confidence: 0.79675037

00:37:23.379 --> 00:37:26.600 so that you have a continuous function.

NOTE Confidence: 0.79675037

00:37:26.600 --> 00:37:29.668 This is graphically nice.

NOTE Confidence: 0.79675037

00:37:29.670 --> 00:37:31.450 And it allows to compare,

NOTE Confidence: 0.79675037

00:37:31.450 --> 00:37:33.010 for example distribution of

NOTE Confidence: 0.79675037

00:37:33.010 --> 00:37:35.350 two variables as you see here

NOTE Confidence: 0.79675037

00:37:35.424 --> 00:37:37.475 in green and in and in Violet,

NOTE Confidence: 0.79675037

00:37:37.480 --> 00:37:40.152 and the advantages that you can see also

NOTE Confidence: 0.79675037

00:37:40.152 --> 00:37:42.099 complex shapes of the distribution.

NOTE Confidence: 0.79675037

00:37:42.100 --> 00:37:43.860 For example here the bimodality

NOTE Confidence: 0.79675037

00:37:43.860 --> 00:37:46.388 or hear the presence of this show

NOTE Confidence: 0.79675037

00:37:46.388 --> 00:37:48.128 is that of the distribution.

NOTE Confidence: 0.79675037

00:37:48.130 --> 00:37:50.615 The pitfall is similar to the histogram,

NOTE Confidence: 0.79675037

00:37:50.620 --> 00:37:52.744 so you should always be careful

NOTE Confidence: 0.79675037

00:37:52.744 --> 00:37:53.806 in selecting the.

NOTE Confidence: 0.79675037

00:37:53.810 --> 00:37:55.940 How much is Martha the distribution?

NOTE Confidence: 0.79675037

00:37:55.940 --> 00:37:58.070 And here you see an example.

NOTE Confidence: 0.79675037

00:37:58.070 --> 00:37:59.342 So these are the.

NOTE Confidence: 0.79675037

00:37:59.342 --> 00:38:01.675 Points that were used at the single

NOTE Confidence: 0.79675037

00:38:01.675 --> 00:38:04.146 points that were the that were used

NOTE Confidence: 0.79675037

00:38:04.146 --> 00:38:06.717 in order to build the distribution.

NOTE Confidence: 0.79675037

00:38:06.720 --> 00:38:09.919 They were randomly chosen from a normal

NOTE Confidence: 0.79675037

00:38:09.919 --> 00:38:12.359 distribution and you can see that.

NOTE Confidence: 0.79675037

00:38:12.360 --> 00:38:14.397 Problem is similar to the bin size,

NOTE Confidence: 0.79675037

00:38:14.400 --> 00:38:16.444 so here you have to select basically.

NOTE Confidence: 0.83262765

00:38:18.510 --> 00:38:20.958 A wavelength in order to approximate

NOTE Confidence: 0.83262765

00:38:20.958 --> 00:38:23.483 that the function to a curve

NOTE Confidence: 0.83262765

00:38:23.483 --> 00:38:25.528 and depending on the wavelength,

NOTE Confidence: 0.83262765

00:38:25.530 --> 00:38:27.182 the resolution of the

NOTE Confidence: 0.83262765

00:38:27.182 --> 00:38:28.834 wavelength that you choose.

NOTE Confidence: 0.83262765

00:38:28.840 --> 00:38:30.624 The result is different,

NOTE Confidence: 0.83262765

00:38:30.624 --> 00:38:34.092 so you could have this kind of plot

NOTE Confidence: 0.83262765

00:38:34.092 --> 00:38:37.510 that seems to show a lot of local pixel,

NOTE Confidence: 0.83262765

00:38:37.510 --> 00:38:40.888 but by smoothing more you have

NOTE Confidence: 0.83262765

00:38:40.888 --> 00:38:43.140 instead the normal distribution

NOTE Confidence: 0.83262765

00:38:43.240 --> 00:38:46.145 from which you draw the data so.

NOTE Confidence: 0.83262765

00:38:46.150 --> 00:38:48.328 There is a balance which appear

NOTE Confidence: 0.83262765

00:38:48.328 --> 00:38:50.667 in choosing beings that are two

NOTE Confidence: 0.83262765

00:38:50.667 --> 00:38:52.697 larger or hear excessive smoothing.

NOTE Confidence: 0.83262765

00:38:52.700 --> 00:38:54.236 Because these over simplifies

NOTE Confidence: 0.83262765

00:38:54.236 --> 00:38:55.388 the original distribution,

NOTE Confidence: 0.83262765

00:38:55.390 --> 00:38:57.320 but on the other side,

NOTE Confidence: 0.83262765

00:38:57.320 --> 00:39:00.776 if you take a resolution that is too small,

NOTE Confidence: 0.83262765

00:39:00.780 --> 00:39:01.550 too granular,

NOTE Confidence: 0.83262765

00:39:01.550 --> 00:39:03.860 you can obtain that strange effects.

NOTE Confidence: 0.83262765

00:39:03.860 --> 00:39:06.170 So you could see for example,

NOTE Confidence: 0.83262765

00:39:06.170 --> 00:39:08.876 pics that are depending on the

NOTE Confidence: 0.83262765

00:39:08.876 --> 00:39:10.680 extraction of random numbers.

NOTE Confidence: 0.83262765

00:39:10.680 --> 00:39:11.179 Again,

NOTE Confidence: 0.83262765

00:39:11.179 --> 00:39:14.672 also in this case there are softwares

NOTE Confidence: 0.83262765

00:39:14.672 --> 00:39:17.759 that given the the original data,

NOTE Confidence: 0.83262765

00:39:17.760 --> 00:39:22.464 your original vote data can calculate the

NOTE Confidence: 0.83262765

00:39:22.464 --> 00:39:25.620 optimal smoothing wavelength in order

NOTE Confidence: 0.83262765

00:39:25.620 --> 00:39:29.274 to avoid distortions based on your data.

NOTE Confidence: 0.83262765

00:39:29.280 --> 00:39:31.386 A compact way to represent the

NOTE Confidence: 0.83262765

00:39:31.386 --> 00:39:33.709 distribution is the box whisker plot,

NOTE Confidence: 0.83262765

00:39:33.710 --> 00:39:36.270 and here you can see how a box

NOTE Confidence: 0.83262765

00:39:36.270 --> 00:39:38.630 whisker plot they can be obtained

NOTE Confidence: 0.83262765

00:39:38.630 --> 00:39:41.090 by this distribution of 20 points.

NOTE Confidence: 0.83262765

00:39:41.090 --> 00:39:43.352 So basically the box whisker plot

NOTE Confidence: 0.83262765

00:39:43.352 --> 00:39:46.254 represents as a box 50% of the data

NOTE Confidence: 0.83262765

00:39:46.254 --> 00:39:47.726 of the distribution to.

NOTE Confidence: 0.83262765

00:39:47.730 --> 00:39:49.944 Usually you have a central line

NOTE Confidence: 0.83262765

00:39:49.944 --> 00:39:51.420 that is the media.

NOTE Confidence: 0.83262765

00:39:51.420 --> 00:39:52.130 It’s important,

NOTE Confidence: 0.83262765

00:39:52.130 --> 00:39:52.840 not laminar,

NOTE Confidence: 0.83262765

00:39:52.840 --> 00:39:56.218 but in the box whisker is always the medium.

NOTE Confidence: 0.83262765

00:39:56.220 --> 00:39:58.120 This is the first quartile

NOTE Confidence: 0.83262765

00:39:58.120 --> 00:39:59.640 and the third quartile.

NOTE Confidence: 0.83262765

00:39:59.640 --> 00:40:02.629 420 Percent 25th percentile of the data.

NOTE Confidence: 0.83262765

00:40:02.630 --> 00:40:04.510 75th percentile of the data.

NOTE Confidence: 0.83262765

00:40:04.510 --> 00:40:07.128 So in the box you have 50%

NOTE Confidence: 0.83262765

00:40:07.130 --> 00:40:09.380 of your day to the central.

NOTE Confidence: 0.83262765

00:40:09.380 --> 00:40:11.260 Here 50% of your data.

NOTE Confidence: 0.83262765

00:40:11.260 --> 00:40:13.730 Then you have the whiskers.

NOTE Confidence: 0.83262765

00:40:13.730 --> 00:40:16.610 They are standard definition of the

NOTE Confidence: 0.83262765

00:40:16.610 --> 00:40:20.004 Whisker Lanka is that they are as

NOTE Confidence: 0.83262765

00:40:20.004 --> 00:40:22.319 long as the interquartile range.

NOTE Confidence: 0.83262765

00:40:22.320 --> 00:40:27.068 That's the distance between Q1 and Q 3 * 1.5.

NOTE Confidence: 0.83262765

00:40:27.068 --> 00:40:30.274 And you see these as the whisker

NOTE Confidence: 0.83262765

00:40:30.274 --> 00:40:31.810 of your plot.

NOTE Confidence: 0.83262765

00:40:31.810 --> 00:40:34.096 So these collect most of the

NOTE Confidence: 0.83262765

00:40:34.096 --> 00:40:35.620 distribution of your data.

NOTE Confidence: 0.83262765

00:40:35.620 --> 00:40:38.210 The data that are outside the whiskers

NOTE Confidence: 0.83262765

00:40:38.210 --> 00:40:40.189 are considered to be outliers.

NOTE Confidence: 0.83262765

00:40:40.190 --> 00:40:40.952 For example,

NOTE Confidence: 0.83262765

00:40:40.952 --> 00:40:43.619 here you see there these three points.

NOTE Confidence: 0.83262765

00:40:43.620 --> 00:40:45.906 They are outside the whisker size,

NOTE Confidence: 0.83262765

00:40:45.910 --> 00:40:48.730 and so these usually are individually

NOTE Confidence: 0.83262765

00:40:48.730 --> 00:40:51.891 displayed in the whisker plot and are

NOTE Confidence: 0.83262765

00:40:51.891 --> 00:40:54.297 considered to be an outlier according

NOTE Confidence: 0.83262765

00:40:54.297 --> 00:40:57.208 to this definition of the whiskers.

NOTE Confidence: 0.83262765

00:40:57.210 --> 00:40:57.532 Yes,

NOTE Confidence: 0.83262765

00:40:57.532 --> 00:40:58.498 if you wanted

NOTE Confidence: 0.8877263

00:40:58.500 --> 00:41:00.110 to make these plots, yeah,

NOTE Confidence: 0.8877263

00:41:00.110 --> 00:41:02.686 is there an easy way to do it

NOTE Confidence: 0.8877263

00:41:02.686 --> 00:41:04.618 or do you like you personally,

NOTE Confidence: 0.8877263

00:41:04.620 --> 00:41:07.188 just do it by in R or something?

NOTE Confidence: 0.8273683

00:41:08.360 --> 00:41:10.448 Well, box plot. I don't think

NOTE Confidence: 0.8273683

00:41:10.448 --> 00:41:12.779 you can do them with Excel,

NOTE Confidence: 0.8273683

00:41:12.780 --> 00:41:14.615 but for example with Prisma

NOTE Confidence: 0.8273683

00:41:14.615 --> 00:41:16.460 or Origin you can totally.

NOTE Confidence: 0.84994245

00:41:19.850 --> 00:41:22.104 I think the only limitation is is

NOTE Confidence: 0.84994245

00:41:22.104 --> 00:41:24.356 Excel, but I be honest, I didn't

NOTE Confidence: 0.84994245

00:41:24.356 --> 00:41:26.310 check the last version of Excel.

NOTE Confidence: 0.8775666

00:41:27.070 --> 00:41:28.486 Right for us to think about,

NOTE Confidence: 0.8775666

00:41:28.490 --> 00:41:30.540 you know we can we have our data and there

NOTE Confidence: 0.8775666

00:41:30.593 --> 00:41:32.497 are many different ways of plotting it,

NOTE Confidence: 0.8775666

00:41:32.500 --> 00:41:34.620 but it sounds like prison might be the

NOTE Confidence: 0.8775666

00:41:34.620 --> 00:41:36.976 way to go in to try to do it in less.

NOTE Confidence: 0.8775666

00:41:36.980 --> 00:41:37.920 You're somebody like you.

NOTE Confidence: 0.8775666

00:41:37.920 --> 00:41:40.930 Who knows how to put it into our.

NOTE Confidence: 0.72675556

00:41:40.930 --> 00:41:44.878 Yes, probably, so please MA is it?

NOTE Confidence: 0.72675556

00:41:44.880 --> 00:41:46.826 Give you an option that is much.

NOTE Confidence: 0.72675556

00:41:46.830 --> 00:41:48.438 Use that if usually use them

NOTE Confidence: 0.72675556

00:41:48.438 --> 00:41:49.900 originally with respect to Prisma.

NOTE Confidence: 0.72675556

00:41:49.900 --> 00:41:52.440 I think it has more.

NOTE Confidence: 0.72675556

00:41:52.440 --> 00:41:53.586 I'm more power,

NOTE Confidence: 0.72675556

00:41:53.586 --> 00:41:56.260 so there are more things that you

NOTE Confidence: 0.72675556

00:41:56.339 --> 00:41:58.985 can do with origin then please MA.

NOTE Confidence: 0.72675556

00:41:58.990 --> 00:42:00.964 I think because it was designed

NOTE Confidence: 0.72675556

00:42:00.964 --> 00:42:03.610 for the for the physics community,

NOTE Confidence: 0.72675556

00:42:03.610 --> 00:42:05.920 but the tradeoff is always complexity,

NOTE Confidence: 0.72675556

00:42:05.920 --> 00:42:08.608 so please May is has less power,

NOTE Confidence: 0.72675556

00:42:08.610 --> 00:42:10.450 less choices, but it's easier

NOTE Confidence: 0.72675556

00:42:10.450 --> 00:42:12.850 to use rather than than origin,

NOTE Confidence: 0.72675556

00:42:12.850 --> 00:42:15.022 but both share the same philosophy

NOTE Confidence: 0.72675556

00:42:15.022 --> 00:42:18.031 so that you need to provide the data

NOTE Confidence: 0.72675556

00:42:18.031 --> 00:42:20.738 is a spreadsheet format and they are

NOTE Confidence: 0.72675556

00:42:20.738 --> 00:42:23.234 available in the software library at.

NOTE Confidence: 0.7910341

00:42:24.340 --> 00:42:26.852 OK, thank you to my can you say

NOTE Confidence: 0.7910341

00:42:26.852 --> 00:42:29.286 the name of the other not prism

NOTE Confidence: 0.7910341

00:42:29.290 --> 00:42:30.566 but the other programming?

NOTE Confidence: 0.7910341

00:42:30.566 --> 00:42:33.580 Or I have a slide after whether you show

NOTE Confidence: 0.7910341

00:42:33.580 --> 00:42:35.230 its origin? OK, thanks yeah.

NOTE Confidence: 0.76106936

00:42:36.750 --> 00:42:37.878 Ava question so,

NOTE Confidence: 0.76106936

00:42:37.878 --> 00:42:40.134 so my initial understanding is that

NOTE Confidence: 0.76106936

00:42:40.134 --> 00:42:42.305 the whisker lenses representing the

NOTE Confidence: 0.76106936

00:42:42.305 --> 00:42:44.885 95 percentile of the data range.

NOTE Confidence: 0.76106936

00:42:44.890 --> 00:42:47.308 But here it says the whisker

NOTE Confidence: 0.76106936

00:42:47.308 --> 00:42:50.180 length is 1.5 times this IQR lens.

NOTE Confidence: 0.76106936

00:42:50.180 --> 00:42:52.220 But if that's the case,

NOTE Confidence: 0.76106936

00:42:52.220 --> 00:42:54.476 why would the left side of

NOTE Confidence: 0.76106936

00:42:54.476 --> 00:42:56.951 the screen right side of risk

NOTE Confidence: 0.76106936

00:42:56.951 --> 00:42:58.727 are having different lens?

NOTE Confidence: 0.08699137

00:43:01.710 --> 00:43:06.384 Um? So that could be for example

NOTE Confidence: 0.08699137

00:43:06.384 --> 00:43:09.106 because here you have the, so that's

NOTE Confidence: 0.08699137

00:43:09.106 --> 00:43:11.822 the the maximal length of the whisker.

NOTE Confidence: 0.08699137

00:43:11.830 --> 00:43:14.152 But if the minimum of your

NOTE Confidence: 0.08699137

00:43:14.152 --> 00:43:16.499 data that is here is here,

NOTE Confidence: 0.08699137

00:43:16.500 --> 00:43:19.416 the whisker stops. So that's why.

NOTE Confidence: 0.08699137

00:43:19.420 --> 00:43:22.100 So I see here you have outliers and

NOTE Confidence: 0.08699137

00:43:22.100 --> 00:43:24.847 so that we can extend to the maximum

NOTE Confidence: 0.08699137

00:43:24.847 --> 00:43:27.374 point that is 1.5 at this measure.

NOTE Confidence: 0.08699137

00:43:27.380 --> 00:43:29.940 But if you before the the maximal distance

NOTE Confidence: 0.08699137

00:43:29.940 --> 00:43:32.217 here you meet the minimal pointer,

NOTE Confidence: 0.08699137

00:43:32.220 --> 00:43:34.255 the whisker and there and

NOTE Confidence: 0.08699137

00:43:34.255 --> 00:43:35.883 there so that's why.

NOTE Confidence: 0.08699137

00:43:35.890 --> 00:43:38.754 OK, I see it's also true that these

NOTE Confidence: 0.08699137

00:43:38.754 --> 00:43:40.738 whisker definition can be customized,

NOTE Confidence: 0.08699137

00:43:40.740 --> 00:43:42.978 so this is the default interpretation.

NOTE Confidence: 0.08699137

00:43:42.980 --> 00:43:45.584 I don't know who who decided this.

NOTE Confidence: 0.08699137

00:43:45.590 --> 00:43:47.828 I don't have the original publication,

NOTE Confidence: 0.08699137

00:43:47.830 --> 00:43:50.062 but you can choose whiskers to

NOTE Confidence: 0.08699137

00:43:50.062 --> 00:43:52.046 be differently, so that's why.

NOTE Confidence: 0.08699137

00:43:52.046 --> 00:43:54.106 Also in the Network Journal

NOTE Confidence: 0.08699137

00:43:54.106 --> 00:43:56.410 paper when you do a box plot,

NOTE Confidence: 0.08699137

00:43:56.410 --> 00:43:59.474 you have always to specify in the statistical

NOTE Confidence: 0.08699137

00:43:59.474 --> 00:44:01.996 methods how you design your box plot.

NOTE Confidence: 0.08699137

00:44:02.000 --> 00:44:04.238 So you have to provide how,

NOTE Confidence: 0.08699137

00:44:04.240 --> 00:44:06.616 for example, the skirts were defined.

NOTE Confidence: 0.08699137

00:44:06.620 --> 00:44:08.450 Because sometimes it's true that,

NOTE Confidence: 0.08699137

00:44:08.450 --> 00:44:09.180 for example,

NOTE Confidence: 0.08699137

00:44:09.180 --> 00:44:11.005 the whisker can represent like

NOTE Confidence: 0.08699137

00:44:11.005 --> 00:44:12.597 95% of the distribution.

NOTE Confidence: 0.08699137

00:44:12.597 --> 00:44:15.460 Right, so this is just the default,

NOTE Confidence: 0.08699137

00:44:15.460 --> 00:44:17.090 but it can be customized,

NOTE Confidence: 0.08699137

00:44:17.090 --> 00:44:18.730 so there are different choices.

NOTE Confidence: 0.87523544

00:44:20.740 --> 00:44:23.530 I have a question regarding the

NOTE Confidence: 0.87523544

00:44:23.530 --> 00:44:26.320 distribution again, maybe it's in

NOTE Confidence: 0.87523544

00:44:26.320 --> 00:44:30.040 continuation to what you just said.

NOTE Confidence: 0.87523544

00:44:30.040 --> 00:44:32.422 Some softwares allow a default value

NOTE Confidence: 0.87523544

00:44:32.422 --> 00:44:35.650 for the bin size and for the smoothening

NOTE Confidence: 0.87523544

00:44:35.650 --> 00:44:38.430 and all that say like Matlab that

NOTE Confidence: 0.87523544

00:44:38.430 --> 00:44:40.985 I've been trying to put this into.

NOTE Confidence: 0.87523544

00:44:40.990 --> 00:44:43.727 How reliable do you think that is?

NOTE Confidence: 0.87523544

00:44:43.730 --> 00:44:46.850 The default values and how would you suggest?

NOTE Confidence: 0.7996838

00:44:48.200 --> 00:44:49.408 Most of the time,

NOTE Confidence: 0.7996838

00:44:49.408 --> 00:44:51.529 most of the time, so I don't.

NOTE Confidence: 0.7996838

00:44:51.529 --> 00:44:53.347 I don't have experience with matter,

NOTE Confidence: 0.7996838

00:44:53.350 --> 00:44:55.702 but probably it will be that it's the

NOTE Confidence: 0.7996838

00:44:55.702 --> 00:44:58.057 same in our so so most of the time

NOTE Confidence: 0.7996838

00:44:58.057 --> 00:45:00.620 there is a sort of optimization there,

NOTE Confidence: 0.7996838

00:45:00.620 --> 00:45:04.660 so most of the time is fine. Uh, but.

NOTE Confidence: 0.87064826

00:45:06.980 --> 00:45:08.916 Sometimes, especially if you

NOTE Confidence: 0.87064826

00:45:08.916 --> 00:45:11.336 have a distribution of data,

NOTE Confidence: 0.87064826

00:45:11.340 --> 00:45:14.244 but you also have a pointer

NOTE Confidence: 0.87064826

00:45:14.244 --> 00:45:16.180 with cumulation of data.

NOTE Confidence: 0.87064826

00:45:16.180 --> 00:45:20.230 You could have problems in the.

NOTE Confidence: 0.87064826

00:45:20.230 --> 00:45:22.010 In the blocker so.

NOTE Confidence: 0.87064826

00:45:22.010 --> 00:45:24.843 But I don't have an example.

NOTE Confidence: 0.87064826

00:45:24.843 --> 00:45:29.530 OK, so like in 95% of the time I'm OK

NOTE Confidence: 0.87064826

00:45:29.530 --> 00:45:33.361 with the with the solution that is

NOTE Confidence: 0.87064826

00:45:33.361 --> 00:45:38.218 provided by the MATLAB or RA building tool.

NOTE Confidence: 0.87064826

00:45:38.220 --> 00:45:40.206 For example, sometimes when you compare

NOTE Confidence: 0.87064826

00:45:40.206 --> 00:45:42.238 to distribution with a different size

NOTE Confidence: 0.87064826

00:45:42.238 --> 00:45:44.200 with a different number of points,

NOTE Confidence: 0.87064826

00:45:44.200 --> 00:45:48.052 that could be that that can be a problem.

NOTE Confidence: 0.87064826

00:45:48.060 --> 00:45:49.548 Because sometimes there.

NOTE Confidence: 0.87064826

00:45:49.548 --> 00:45:52.028 If you're comparing for example

NOTE Confidence: 0.87064826

00:45:52.028 --> 00:45:54.222 distribution with 10 points with

NOTE Confidence: 0.87064826

00:45:54.222 --> 00:45:56.107 a distribution of 1000 points.

NOTE Confidence: 0.87064826

00:45:56.110 --> 00:45:57.630 Adopting the same wavelength

NOTE Confidence: 0.87064826

00:45:57.630 --> 00:45:59.150 could be a problem,

NOTE Confidence: 0.87064826

00:45:59.150 --> 00:46:03.174 and so you need to manually change it.

NOTE Confidence: 0.87064826

00:46:03.180 --> 00:46:04.832 So that's the yes,

NOTE Confidence: 0.87064826

00:46:04.832 --> 00:46:07.550 but that's that probably could be a.

NOTE Confidence: 0.87064826

00:46:07.550 --> 00:46:11.550 A practical example on when it's not ideal.

NOTE Confidence: 0.87064826

00:46:11.550 --> 00:46:12.825 Because the software,

NOTE Confidence: 0.87064826

00:46:12.825 --> 00:46:15.800 if you are trying to compare a

NOTE Confidence: 0.87064826

00:46:15.887 --> 00:46:18.257 10 points versus 1000 points,

NOTE Confidence: 0.87064826

00:46:18.260 --> 00:46:20.936 tries to define a common wavelength.

NOTE Confidence: 0.87064826

00:46:20.940 --> 00:46:22.724 But sometimes this leads

NOTE Confidence: 0.87064826

00:46:22.724 --> 00:46:24.508 to like distorted images.

NOTE Confidence: 0.87064826

00:46:24.510 --> 00:46:27.639 I don't have an example to show.

NOTE Confidence: 0.882226

00:46:28.510 --> 00:46:30.040 That's good enough, thank you.

NOTE Confidence: 0.7991695

00:46:31.720 --> 00:46:34.590 And well, I can leave the note.

NOTE Confidence: 0.7991695

00:46:34.590 --> 00:46:36.912 Sometimes you can see also the

NOTE Confidence: 0.7991695

00:46:36.912 --> 00:46:39.100 nutshack inside your box whisker,

NOTE Confidence: 0.7991695

00:46:39.100 --> 00:46:41.770 so they’re not sure is diesel

NOTE Confidence: 0.7991695

00:46:41.770 --> 00:46:44.492 feature that it represents a measure

NOTE Confidence: 0.7991695

00:46:44.492 --> 00:46:46.647 of certainty for the medium.

NOTE Confidence: 0.7991695

00:46:46.650 --> 00:46:48.736 So sometimes it is useful to have.

NOTE Confidence: 0.7991695

00:46:48.740 --> 00:46:50.805 These are ’cause if you are comparing

NOTE Confidence: 0.7991695

00:46:50.805 --> 00:46:53.731 a lot of box whisker plots a you can

NOTE Confidence: 0.7991695

00:46:53.731 --> 00:46:56.392 look at the uncertainty as if it was a

NOTE Confidence: 0.7991695

00:46:56.392 --> 00:46:58.566 sort of standard error of the media.

NOTE Confidence: 0.7991695

00:46:58.566 --> 00:47:01.667 And so if two box whisker overlapping,

NOTE Confidence: 0.7991695

00:47:01.670 --> 00:47:02.584 they’re not.

NOTE Confidence: 0.7991695

00:47:02.584 --> 00:47:05.326 She’s probably it means that the

NOTE Confidence: 0.7991695

00:47:05.326 --> 00:47:07.769 medians are not statistically.

NOTE Confidence: 0.7991695

00:47:07.770 --> 00:47:08.946 Inefficiently different.

NOTE Confidence: 0.7991695

00:47:08.946 --> 00:47:12.474 This could be a way to.

NOTE Confidence: 0.7991695

00:47:12.480 --> 00:47:14.696 The use of the notch or there is

NOTE Confidence: 0.7991695

00:47:14.696 --> 00:47:16.287 the interpretation of the data

NOTE Confidence: 0.7991695

00:47:16.287 --> 00:47:17.917 and the comparison of different

NOTE Confidence: 0.7991695

00:47:17.917 --> 00:47:19.498 distribution and that's why the

NOTE Confidence: 0.7991695

00:47:19.498 --> 00:47:21.238 box whisker plots are so popular,

NOTE Confidence: 0.7991695

00:47:21.240 --> 00:47:23.514 because they allow you to represent

NOTE Confidence: 0.7991695

00:47:23.514 --> 00:47:25.406 that distribution of data in

NOTE Confidence: 0.7991695

00:47:25.406 --> 00:47:26.650 a very compact format.

NOTE Confidence: 0.7991695

00:47:26.650 --> 00:47:28.990 This is another display of the

NOTE Confidence: 0.7991695

00:47:28.990 --> 00:47:31.270 anatomy of Big box whisker,

NOTE Confidence: 0.7991695

00:47:31.270 --> 00:47:33.370 but it doesn't add anything

NOTE Confidence: 0.7991695

00:47:33.370 --> 00:47:35.470 that I had also before.

NOTE Confidence: 0.7991695

00:47:35.470 --> 00:47:38.366 So here is an example where box whisker

NOTE Confidence: 0.7991695

00:47:38.366 --> 00:47:41.544 plots are used in order to compare

NOTE Confidence: 0.7991695

00:47:41.544 --> 00:47:43.448 the four different distributions.

NOTE Confidence: 0.7991695

00:47:43.450 --> 00:47:46.439 So the advantage is that they allow

NOTE Confidence: 0.7991695

00:47:46.439 --> 00:47:49.020 easy comparison so it's easy to

NOTE Confidence: 0.7991695

00:47:49.020 --> 00:47:51.848 compare the distribution of ABC and D.

NOTE Confidence: 0.7991695

00:47:51.850 --> 00:47:55.066 The problem they can have is that they

NOTE Confidence: 0.7991695

00:47:55.066 --> 00:47:58.180 hide the shape of the distribution.

NOTE Confidence: 0.7991695

00:47:58.180 --> 00:48:00.742 And also usually they hide the

NOTE Confidence: 0.7991695

00:48:00.742 --> 00:48:03.362 number of points that were used

NOTE Confidence: 0.7991695

00:48:03.362 --> 00:48:05.417 to build the box whisker.

NOTE Confidence: 0.7991695

00:48:05.420 --> 00:48:07.790 Sometimes you can code the number

NOTE Confidence: 0.7991695

00:48:07.790 --> 00:48:10.335 of points so the cardinality the

NOTE Confidence: 0.7991695

00:48:10.335 --> 00:48:12.987 size of the distribution as the

NOTE Confidence: 0.7991695

00:48:12.987 --> 00:48:15.218 width of the box whisker,

NOTE Confidence: 0.7991695

00:48:15.220 --> 00:48:18.202 but it’s rarely used because it’s not

NOTE Confidence: 0.7991695

00:48:18.202 --> 00:48:20.620 very visually beautiful, I would say.

NOTE Confidence: 0.7991695

00:48:20.620 --> 00:48:23.280 So one solution it could be to

NOTE Confidence: 0.7991695

00:48:23.374 --> 00:48:25.939 overlay over the box whisker,

NOTE Confidence: 0.7991695

00:48:25.940 --> 00:48:27.436 plot the jitter plot,

NOTE Confidence: 0.7991695

00:48:27.436 --> 00:48:29.680 so jitter plot represents the single

NOTE Confidence: 0.7991695

00:48:29.746 --> 00:48:31.720 points that were used to build

NOTE Confidence: 0.7991695

00:48:31.720 --> 00:48:33.938 the box whisker plot and they are.

NOTE Confidence: 0.7991695

00:48:33.940 --> 00:48:36.292 So while on the Y axis that there

NOTE Confidence: 0.7991695

00:48:36.292 --> 00:48:38.985 is the precise values on the X

NOTE Confidence: 0.7991695

00:48:38.985 --> 00:48:40.200 axis there randomly.

NOTE Confidence: 0.810526

00:48:42.740 --> 00:48:44.600 Place that let's say there are

NOTE Confidence: 0.810526

00:48:44.600 --> 00:48:46.685 also methods that do not display

NOTE Confidence: 0.810526

00:48:46.685 --> 00:48:48.640 these points randomly butting up.

NOTE Confidence: 0.810526

00:48:48.640 --> 00:48:50.428 Sell the random way that captures

NOTE Confidence: 0.810526

00:48:50.428 --> 00:48:52.460 the shape of the distribution,

NOTE Confidence: 0.810526

00:48:52.460 --> 00:48:55.020 and I think that that kind of plot

NOTE Confidence: 0.810526

00:48:55.020 --> 00:48:57.676 is also present in in graph for

NOTE Confidence: 0.810526

00:48:57.676 --> 00:49:00.305 the Prisma so the advantage of this

NOTE Confidence: 0.810526

00:49:00.305 --> 00:49:02.764 is that you can see, for example,

NOTE Confidence: 0.810526

00:49:02.764 --> 00:49:06.033 that would be the distribution is bimodal.

NOTE Confidence: 0.810526

00:49:06.040 --> 00:49:08.378 So because you see that there are

NOTE Confidence: 0.810526

00:49:08.378 --> 00:49:10.447 these high densities of points and

NOTE Confidence: 0.810526

00:49:10.447 --> 00:49:12.451 the box whisker plot cannot capture

NOTE Confidence: 0.810526

00:49:12.451 --> 00:49:14.458 that you cannot see from a box,

NOTE Confidence: 0.810526

00:49:14.460 --> 00:49:16.640 whisker plot data distribution is

NOTE Confidence: 0.810526

00:49:16.640 --> 00:49:19.242 bimodal and for example here you

NOTE Confidence: 0.810526

00:49:19.242 --> 00:49:21.312 can see that these box whisker

NOTE Confidence: 0.810526

00:49:21.312 --> 00:49:23.771 plot there has been is based on

NOTE Confidence: 0.810526

00:49:23.771 --> 00:49:25.799 much less data than the others.

NOTE Confidence: 0.810526

00:49:25.800 --> 00:49:28.768 So, uh, and a solution for these

NOTE Confidence: 0.810526

00:49:28.768 --> 00:49:32.303 are is to enclose the box whisker

NOTE Confidence: 0.810526

00:49:32.303 --> 00:49:34.998 plot into a violin plot.

NOTE Confidence: 0.810526

00:49:35.000 --> 00:49:36.952 So violin plot representation

NOTE Confidence: 0.810526

00:49:36.952 --> 00:49:40.430 like these allow you to see the

NOTE Confidence: 0.810526

00:49:40.430 --> 00:49:43.214 same information of a box whisker,

NOTE Confidence: 0.810526

00:49:43.220 --> 00:49:46.136 but also information on this shape

NOTE Confidence: 0.810526

00:49:46.136 --> 00:49:48.635 of the distribution is basically

NOTE Confidence: 0.810526

00:49:48.635 --> 00:49:50.479 in a violin plot.

NOTE Confidence: 0.810526

00:49:50.480 --> 00:49:54.323 You add a density plot that is

NOTE Confidence: 0.810526

00:49:54.323 --> 00:49:57.759 parallel to the vertical axis.

NOTE Confidence: 0.810526

00:49:57.760 --> 00:50:00.390 And here, by using a violin plot you can see.

NOTE Confidence: 0.810526

00:50:00.390 --> 00:50:01.016 That this,

NOTE Confidence: 0.810526

00:50:01.016 --> 00:50:02.894 that this distribution is one pick

NOTE Confidence: 0.810526

00:50:02.894 --> 00:50:05.460 has one pick. This one is by model.

NOTE Confidence: 0.83657587

00:50:08.230 --> 00:50:10.806 And you can add also the number here.

NOTE Confidence: 0.83657587

00:50:10.810 --> 00:50:12.672 He said of coding the number as

NOTE Confidence: 0.83657587

00:50:12.672 --> 00:50:14.695 the size of the distribution as

NOTE Confidence: 0.83657587

00:50:14.695 --> 00:50:16.630 the width of the distribution.

NOTE Confidence: 0.8613901

00:50:18.710 --> 00:50:21.265 So this is an example of compare

NOTE Confidence: 0.8613901

00:50:21.265 --> 00:50:22.907 of comparisons between different

NOTE Confidence: 0.8613901

00:50:22.907 --> 00:50:24.819 ways to show distribution.

NOTE Confidence: 0.8613901

00:50:24.820 --> 00:50:28.068 Here you see the histogram with the density,

NOTE Confidence: 0.8613901

00:50:28.070 --> 00:50:29.306 corresponding density plot,

NOTE Confidence: 0.8613901

00:50:29.306 --> 00:50:30.542 the same distribution

NOTE Confidence: 0.8613901

00:50:30.542 --> 00:50:32.550 visualized as a box plot,

NOTE Confidence: 0.8613901

00:50:32.550 --> 00:50:35.770 and visualized as a violin plot that

NOTE Confidence: 0.8613901

00:50:35.770 --> 00:50:39.362 captures both the features of a box

NOTE Confidence: 0.8613901

00:50:39.362 --> 00:50:42.012 plot cluster the density distribution.

NOTE Confidence: 0.8613901

00:50:42.020 --> 00:50:44.456 And this is for a normal distribution.

NOTE Confidence: 0.8613901

00:50:44.460 --> 00:50:47.132 This is for a bimodal distribution where you

NOTE Confidence: 0.8613901

00:50:47.132 --> 00:50:50.049 can see that the box plot doesn't capture,

NOTE Confidence: 0.8613901

00:50:50.050 --> 00:50:52.610 so the box plot can capture the fact

NOTE Confidence: 0.8613901

00:50:52.610 --> 00:50:55.176 that the data are not symmetrical and

NOTE Confidence: 0.8613901

00:50:55.176 --> 00:50:58.124 you see the for example the distance from

NOTE Confidence: 0.8613901

00:50:58.124 --> 00:51:01.218 the from the from the point of the box

NOTE Confidence: 0.8613901

00:51:01.218 --> 00:51:04.010 and the medium is much more than these.

NOTE Confidence: 0.8613901

00:51:04.010 --> 00:51:06.714 So the box whisker is good in capturing

NOTE Confidence: 0.8613901

00:51:06.714 --> 00:51:08.567 a symmetrical distributions but not

NOTE Confidence: 0.8613901

00:51:08.567 --> 00:51:11.157 the presence of more than one piece.

NOTE Confidence: 0.8613901

00:51:11.160 --> 00:51:14.720 So not the complex shape of the distribution.

NOTE Confidence: 0.8613901

00:51:14.720 --> 00:51:17.247 And there is a website here where

NOTE Confidence: 0.8613901

00:51:17.247 --> 00:51:20.663 you can where you can see a lot of

NOTE Confidence: 0.8613901

00:51:20.663 --> 00:51:23.255 examples where the different choice of

NOTE Confidence: 0.8613901

00:51:23.255 --> 00:51:25.775 visualization can lead to different.

NOTE Confidence: 0.8613901

00:51:25.780 --> 00:51:29.209 Conclusion as here.

NOTE Confidence: 0.8613901

00:51:29.210 --> 00:51:32.381 It's true also that the violin Plata

NOTE Confidence: 0.8613901

00:51:32.381 --> 00:51:34.999 is not efficient because you're

NOTE Confidence: 0.8613901

00:51:34.999 --> 00:51:37.335 sure you're showing twice.

NOTE Confidence: 0.8613901

00:51:37.340 --> 00:51:38.906 The same information,

NOTE Confidence: 0.8613901

00:51:38.906 --> 00:51:41.516 so this is aesthetically pleasant,

NOTE Confidence: 0.8613901

00:51:41.520 --> 00:51:44.245 but is not efficient because

NOTE Confidence: 0.8613901

00:51:44.245 --> 00:51:46.425 you're repeating basically this

NOTE Confidence: 0.8613901

00:51:46.425 --> 00:51:48.819 density twice above and below,

NOTE Confidence: 0.8613901

00:51:48.820 --> 00:51:52.040 and so that’s why there are two

NOTE Confidence: 0.8613901

00:51:52.040 --> 00:51:54.570 saver for efficiency sufficiency.

NOTE Confidence: 0.8613901

00:51:54.570 --> 00:51:56.918 There are recent visualization

NOTE Confidence: 0.8613901

00:51:56.918 --> 00:52:00.440 strategies as the rain cloud plotter.

NOTE Confidence: 0.8613901

00:52:00.440 --> 00:52:03.216 So the Raincloud plot that shows a box

NOTE Confidence: 0.8613901

00:52:03.216 --> 00:52:05.644 whisker plot in the middle half violin

NOTE Confidence: 0.8613901

00:52:05.644 --> 00:52:08.538 plot here and then also the single point.

NOTE Confidence: 0.8613901

00:52:08.540 --> 00:52:11.634 So that’s probably the one of the

NOTE Confidence: 0.8613901

00:52:11.634 --> 00:52:13.934 most complete exhaustive ways to

NOTE Confidence: 0.8613901

00:52:13.934 --> 00:52:16.164 represent a distribution of data.

NOTE Confidence: 0.8613901

00:52:16.170 --> 00:52:18.235 And they’re called the rain cloud because

NOTE Confidence: 0.8613901

00:52:18.235 --> 00:52:20.387 of this effect is should be the cloud.

NOTE Confidence: 0.8613901

00:52:20.390 --> 00:52:22.266 And this is the rain that falls

NOTE Confidence: 0.8613901

00:52:22.266 --> 00:52:23.480 on the proposed below.

NOTE Confidence: 0.8613901

00:52:23.480 --> 00:52:25.376 So you can find information on

NOTE Confidence: 0.8613901

00:52:25.376 --> 00:52:27.543 how to block these are following

NOTE Confidence: 0.8613901

00:52:27.543 --> 00:52:29.219 the following these link.

NOTE Confidence: 0.8613901

00:52:29.220 --> 00:52:30.530 Another yeah.

NOTE Confidence: 0.83115506

00:52:31.370 --> 00:52:35.240 Quick question, is there a?

NOTE Confidence: 0.83115506

00:52:35.240 --> 00:52:38.411 How to say the restriction or limitation

NOTE Confidence: 0.83115506

00:52:38.411 --> 00:52:42.864 as to how many data points are required

NOTE Confidence: 0.83115506

00:52:42.864 --> 00:52:45.859 for generating reliable violin plot?

NOTE Confidence: 0.8652727

00:52:49.810 --> 00:52:54.620 Generally not so. Probably more than 10,

NOTE Confidence: 0.8652727

00:52:54.620 --> 00:52:57.722 I would say because otherwise so you can

NOTE Confidence: 0.8652727

00:52:57.722 --> 00:53:00.144 see that you can see it empirically,

NOTE Confidence: 0.8652727

00:53:00.150 --> 00:53:03.237 because if the data are too few you can

NOTE Confidence: 0.8652727

00:53:03.237 --> 00:53:06.348 see that the violin basically have sort

NOTE Confidence: 0.8652727

00:53:06.348 --> 00:53:09.738 of waves around each point of your data.

NOTE Confidence: 0.8652727

00:53:09.740 --> 00:53:14.622 So as a general. Is a general threshold.

NOTE Confidence: 0.8652727

00:53:14.622 --> 00:53:18.552 I would say 10 points would be the

NOTE Confidence: 0.8652727

00:53:18.552 --> 00:53:21.190 like the minimum number. And asking

NOTE Confidence: 0.83410805

00:53:21.190 --> 00:53:23.188 that question is of course if

NOTE Confidence: 0.83410805

00:53:23.188 --> 00:53:25.739 you have a lot to data points,

NOTE Confidence: 0.83410805

00:53:25.740 --> 00:53:27.140 these would be informative.

NOTE Confidence: 0.83410805

00:53:27.140 --> 00:53:29.240 But if you have, let's say

NOTE Confidence: 0.83410805

00:53:29.240 --> 00:53:31.340 less than 10 or small number,

NOTE Confidence: 0.83410805

00:53:31.340 --> 00:53:33.090 this could be really distorting

NOTE Confidence: 0.83410805

00:53:33.090 --> 00:53:35.190 or faking the. Yeah, yeah, that's

NOTE Confidence: 0.83410805

00:53:35.190 --> 00:53:36.965 true. That's why I would

NOTE Confidence: 0.83410805

00:53:36.965 --> 00:53:39.040 say 10 because it below 10.

NOTE Confidence: 0.83410805

00:53:39.040 --> 00:53:41.427 Probably the best strategy is to show

NOTE Confidence: 0.83410805

00:53:41.427 --> 00:53:43.631 the single points and then a summary

NOTE Confidence: 0.83410805

00:53:43.631 --> 00:53:45.854 such as the mean or median plus

NOTE Confidence: 0.83410805

00:53:45.854 --> 00:53:47.790 some validation standard dialogue,

NOTE Confidence: 0.83410805

00:53:47.790 --> 00:53:50.004 but not the not the distribution

NOTE Confidence: 0.83410805

00:53:50.004 --> 00:53:51.480 as a violin plot.

NOTE Confidence: 0.83410805

00:53:51.480 --> 00:53:54.540 So that's for a like less than 10 data.

NOTE Confidence: 0.8752581

00:53:56.890 --> 00:53:58.666 Alright, when data are too much,

NOTE Confidence: 0.8752581

00:53:58.670 --> 00:54:00.330 for example, it doesn't make

NOTE Confidence: 0.8752581

00:54:00.330 --> 00:54:02.500 sense to show the single points.

NOTE Confidence: 0.8752581

00:54:02.500 --> 00:54:04.090 Because that they are overlap,

NOTE Confidence: 0.8752581

00:54:04.090 --> 00:54:06.253 they overlap each other and so you

NOTE Confidence: 0.8752581

00:54:06.253 --> 00:54:07.882 don't see anything that happens

NOTE Confidence: 0.8752581

00:54:07.882 --> 00:54:10.108 when you have more than 1000 points,

NOTE Confidence: 0.8752581

00:54:10.110 --> 00:54:12.422 and so the best solution in that case

NOTE Confidence: 0.8752581

00:54:12.422 --> 00:54:15.179 is for example to show only the violin.

NOTE Confidence: 0.86036754

00:54:18.190 --> 00:54:21.620 So there is a Ranger for which.

NOTE Confidence: 0.86036754

00:54:21.620 --> 00:54:24.329 The best solution is to show the

NOTE Confidence: 0.86036754

00:54:24.329 --> 00:54:27.019 single data points with the cross bar,

NOTE Confidence: 0.86036754

00:54:27.020 --> 00:54:30.037 so an element with captures mean or

NOTE Confidence: 0.86036754

00:54:30.037 --> 00:54:32.650 median plus standard deviation order.

NOTE Confidence: 0.86036754

00:54:32.650 --> 00:54:34.545 Confidence interval there is a

NOTE Confidence: 0.86036754

00:54:34.545 --> 00:54:37.683 Ranger that is in the middle from 10

NOTE Confidence: 0.86036754

00:54:37.683 --> 00:54:40.005 to some hundreds where the violin

NOTE Confidence: 0.86036754

00:54:40.005 --> 00:54:42.525 plot and the box whisker plot are

NOTE Confidence: 0.86036754

00:54:42.525 --> 00:54:44.353 the best option to visualize.

NOTE Confidence: 0.86036754

00:54:44.353 --> 00:54:46.618 And when you have many,

NOTE Confidence: 0.86036754

00:54:46.620 --> 00:54:48.474 many data more than 1000, probably.

NOTE Confidence: 0.86036754

00:54:48.474 --> 00:54:50.742 If you want to capture the distribution

NOTE Confidence: 0.86036754

00:54:50.742 --> 00:54:52.935 then only there the violin plot rather

NOTE Confidence: 0.86036754

00:54:52.935 --> 00:54:55.189 than the single points is the best way.

NOTE Confidence: 0.85021085

00:55:01.390 --> 00:55:05.640 Did did he? Did it answer?

NOTE Confidence: 0.85021085

00:55:05.640 --> 00:55:06.620 Yeah, that was awesome.

NOTE Confidence: 0.8187417

00:55:08.060 --> 00:55:09.080 That is a great explanation.

NOTE Confidence: 0.82946837

00:55:10.990 --> 00:55:12.826 OK, another another alternative

NOTE Confidence: 0.82946837

00:55:12.826 --> 00:55:15.580 way to maximize efficiency of the

NOTE Confidence: 0.82946837

00:55:15.659 --> 00:55:18.147 violence that I saw a lot in the

NOTE Confidence: 0.82946837

00:55:18.147 --> 00:55:20.328 with single cell data, for example,

NOTE Confidence: 0.82946837

00:55:20.328 --> 00:55:23.440 is the the user split violin plots are,

NOTE Confidence: 0.82946837

00:55:23.440 --> 00:55:26.329 so you use the violin plot to show a

NOTE Confidence: 0.82946837

00:55:26.329 --> 00:55:28.498 comparison between two distributions.

NOTE Confidence: 0.82946837

00:55:28.500 --> 00:55:31.929 So you see here are this plot shows the

NOTE Confidence: 0.82946837

00:55:31.929 --> 00:55:34.335 representation of Asia or female and

NOTE Confidence: 0.82946837

00:55:34.335 --> 00:55:36.668 males are using different social, social,

NOTE Confidence: 0.82946837

00:55:36.668 --> 00:55:38.220 media, Instagram, Facebook, Twitter.

NOTE Confidence: 0.82946837

00:55:38.220 --> 00:55:41.020 So it's a way to show using.

NOTE Confidence: 0.82946837

00:55:41.020 --> 00:55:43.981 Half of a violin plot are differences

NOTE Confidence: 0.82946837

00:55:43.981 --> 00:55:46.314 in the distributions and this can

NOTE Confidence: 0.82946837

00:55:46.314 --> 00:55:48.610 be used when you have a contrast

NOTE Confidence: 0.82946837

00:55:48.690 --> 00:55:51.707 of two conditions or you want to

NOTE Confidence: 0.82946837

00:55:51.707 --> 00:55:53.000 compare two distributions.

NOTE Confidence: 0.82946837

00:55:53.000 --> 00:55:55.538 I'm also in the single cell.

NOTE Confidence: 0.8256293

00:55:56.960 --> 00:55:58.268 About the violin. Plots,

NOTE Confidence: 0.8256293

00:55:58.268 --> 00:56:00.237 like in the such cases, yeah,

NOTE Confidence: 0.8256293

00:56:00.237 --> 00:56:02.853 So what determines the height of the peaks?

NOTE Confidence: 0.8256293

00:56:02.860 --> 00:56:04.852 Or is that everything is normalized

NOTE Confidence: 0.8256293

00:56:04.852 --> 00:56:07.130 so that the total area the same,

NOTE Confidence: 0.8256293

00:56:07.130 --> 00:56:09.419 or the maximum height is the same?

NOTE Confidence: 0.8786113

00:56:10.220 --> 00:56:12.320 So most of the time,

NOTE Confidence: 0.8786113

00:56:12.320 --> 00:56:16.136 so you have choices usually so you can

NOTE Confidence: 0.8786113

00:56:16.136 --> 00:56:19.879 choose to have the same maximum hate.

NOTE Confidence: 0.8786113

00:56:19.880 --> 00:56:21.930 And that's usually the then.

NOTE Confidence: 0.8786113

00:56:21.930 --> 00:56:23.970 That's usually what you find,

NOTE Confidence: 0.8786113

00:56:23.970 --> 00:56:26.922 so you you plot there in a way

NOTE Confidence: 0.8786113

00:56:26.922 --> 00:56:29.934 that the Ranger is the same from

NOTE Confidence: 0.8786113

00:56:29.934 --> 00:56:32.970 here to here from here to here,

NOTE Confidence: 0.8786113

00:56:32.970 --> 00:56:35.412 the alternative is to use the

NOTE Confidence: 0.8786113

00:56:35.412 --> 00:56:38.290 real criteria for a for a density,

NOTE Confidence: 0.8786113

00:56:38.290 --> 00:56:41.350 and that should be that the

NOTE Confidence: 0.8786113

00:56:41.350 --> 00:56:44.730 area under visa is equal to 1.

NOTE Confidence: 0.8786113

00:56:44.730 --> 00:56:47.979 And so that the two have the same area.

NOTE Confidence: 0.8786113

00:56:47.980 --> 00:56:50.182 An alternative is to have an

NOTE Confidence: 0.8786113

00:56:50.182 --> 00:56:52.134 area that is proportional to

NOTE Confidence: 0.8786113

00:56:52.134 --> 00:56:53.766 the number of observations,

NOTE Confidence: 0.8786113

00:56:53.770 --> 00:56:56.554 but I think that visually most

NOTE Confidence: 0.8786113

00:56:56.554 --> 00:56:59.529 of the time you find that.

NOTE Confidence: 0.8786113

00:56:59.530 --> 00:57:01.938 The criteria is that you have in order

NOTE Confidence: 0.8786113

00:57:01.938 --> 00:57:04.707 to have balanced plots are the criteria,

NOTE Confidence: 0.8786113

00:57:04.710 --> 00:57:06.780 is to have the same Ranger.

NOTE Confidence: 0.8786113

00:57:06.780 --> 00:57:08.628 Meaning from here to the maximum

NOTE Confidence: 0.8786113

00:57:08.628 --> 00:57:10.666 for all their pull the plot

NOTE Confidence: 0.8786113

00:57:10.666 --> 00:57:12.511 independently from the area and

NOTE Confidence: 0.8786113

00:57:12.511 --> 00:57:14.710 dependently from the number of points.

NOTE Confidence: 0.87719923

00:57:17.880 --> 00:57:19.637 It’s not probably the best solution from

NOTE Confidence: 0.87719923

00:57:19.637 --> 00:57:21.489 the point of view of communication,

NOTE Confidence: 0.87719923

00:57:21.490 --> 00:57:24.180 but it’s most used. OK, thank you.

NOTE Confidence: 0.7779449

00:57:26.570 --> 00:57:29.636 I variation of this is also the

NOTE Confidence: 0.7779449

00:57:29.636 --> 00:57:32.409 use of ridgeline plots are that.

NOTE Confidence: 0.7779449

00:57:32.410 --> 00:57:35.158 They allow you to compare a

NOTE Confidence: 0.7779449

00:57:35.158 --> 00:57:36.990 lot of different densities.

NOTE Confidence: 0.7779449

00:57:36.990 --> 00:57:39.713 For example, here you see a comparison

NOTE Confidence: 0.7779449

00:57:39.713 --> 00:57:42.562 of the density of temperatures in

NOTE Confidence: 0.7779449

00:57:42.562 --> 00:57:45.227 different month in allocation metadata.

NOTE Confidence: 0.7779449

00:57:45.230 --> 00:57:48.040 Remember Lincoln NE and this

NOTE Confidence: 0.7779449

00:57:48.040 --> 00:57:51.560 is used in a single cell.

NOTE Confidence: 0.7779449

00:57:51.560 --> 00:57:53.762 Is Alotta now in these years

NOTE Confidence: 0.7779449

00:57:53.762 --> 00:57:55.230 with single cell data?

NOTE Confidence: 0.7779449

00:57:55.230 --> 00:57:55.858 For example,

NOTE Confidence: 0.7779449

00:57:55.858 --> 00:57:58.370 here you see that it is used to

NOTE Confidence: 0.7779449

00:57:58.447 --> 00:58:00.517 compare the distribution of the

NOTE Confidence: 0.7779449

00:58:00.517 --> 00:58:03.022 expression of 1 gene leads A

NOTE Confidence: 0.7779449

00:58:03.022 --> 00:58:04.807 or CL5 in different population

NOTE Confidence: 0.7779449

00:58:04.807 --> 00:58:06.836 of cells that are probability

NOTE Confidence: 0.7779449

00:58:06.836 --> 00:58:09.816 can from some blood sample.

NOTE Confidence: 0.7779449

00:58:09.820 --> 00:58:10.906 Different population and

NOTE Confidence: 0.7779449

00:58:10.906 --> 00:58:12.716 these allow you to see.

NOTE Confidence: 0.7779449

00:58:12.720 --> 00:58:15.720 Sorry to see an marker genes or to

NOTE Confidence: 0.7779449

00:58:15.720 --> 00:58:19.410 see how the expression of a gene is

NOTE Confidence: 0.7779449

00:58:19.410 --> 00:58:22.199 specific for a population of cells.

NOTE Confidence: 0.7779449

00:58:22.200 --> 00:58:24.909 So that's why I included because I

NOTE Confidence: 0.7779449

00:58:24.909 --> 00:58:28.047 see that the frequency of this plot,

NOTE Confidence: 0.7779449

00:58:28.050 --> 00:58:31.050 specially in the single cell

NOTE Confidence: 0.7779449

00:58:31.050 --> 00:58:34.050 visualization field is quite increasing.

NOTE Confidence: 0.7779449

00:58:34.050 --> 00:58:36.276 I have visa section of the

NOTE Confidence: 0.7779449

00:58:36.276 --> 00:58:38.180 presentation that we could skip.

NOTE Confidence: 0.7779449

00:58:38.180 --> 00:58:40.472 In general the message about Visa

NOTE Confidence: 0.7779449

00:58:40.472 --> 00:58:42.789 is that Venn diagrams are good

NOTE Confidence: 0.7779449

00:58:42.789 --> 00:58:44.925 when you have two Venn diagrams,

NOTE Confidence: 0.7779449

00:58:44.930 --> 00:58:46.114 but if they are,

NOTE Confidence: 0.7779449

00:58:46.114 --> 00:58:48.952 if they're more there a bad way to

NOTE Confidence: 0.7779449

00:58:48.952 --> 00:58:50.928 represent intersections between sets.

NOTE Confidence: 0.7779449

00:58:50.930 --> 00:58:53.569 And this actually is a plot that

NOTE Confidence: 0.7779449

00:58:53.569 --> 00:58:56.357 was published in Nature and it it's

NOTE Confidence: 0.7779449

00:58:56.357 --> 00:58:58.715 about a comparison of the genome

NOTE Confidence: 0.7779449

00:58:58.794 --> 00:59:01.029 of banana with other species.

NOTE Confidence: 0.7779449

00:59:01.030 --> 00:59:03.662 So the problem in general is that

NOTE Confidence: 0.7779449

00:59:03.662 --> 00:59:06.381 when you have more than two, 3,

NOTE Confidence: 0.7779449

00:59:06.381 --> 00:59:08.667 four but also two Venn diagrams,

NOTE Confidence: 0.7779449

00:59:08.670 --> 00:59:11.281 it's it's not the best way to

NOTE Confidence: 0.7779449

00:59:11.281 --> 00:59:13.240 visualize intersection with the use

NOTE Confidence: 0.7779449

00:59:13.240 --> 00:59:15.160 of the traditional Venn diagrams.

NOTE Confidence: 0.7779449

00:59:15.160 --> 00:59:18.079 So a table is probably more effective

NOTE Confidence: 0.7779449

00:59:18.079 --> 00:59:20.629 than this because the areas are

NOTE Confidence: 0.7779449

00:59:20.629 --> 00:59:22.664 not proportional to the size.

NOTE Confidence: 0.7779449

00:59:22.670 --> 00:59:25.724 And it's quite confusing to see

NOTE Confidence: 0.7779449

00:59:25.724 --> 00:59:27.760 the specific intersection and

NOTE Confidence: 0.7779449

00:59:27.844 --> 00:59:29.680 so on alternative way.

NOTE Confidence: 0.7779449

00:59:29.680 --> 00:59:32.137 That was developed in the recent year

NOTE Confidence: 0.7779449

00:59:32.137 --> 00:59:34.693 was the user the concept of this

NOTE Confidence: 0.7779449

00:59:34.693 --> 00:59:37.345 upset plots are so to represent the

NOTE Confidence: 0.7779449

00:59:37.345 --> 00:59:39.510 intersections in a matrix format.

NOTE Confidence: 0.7779449

00:59:39.510 --> 00:59:41.960 So represent these are as a member

NOTE Confidence: 0.7779449

00:59:41.960 --> 00:59:43.670 as a sum object.

NOTE Confidence: 0.7779449

00:59:43.670 --> 00:59:43.989 Example,

NOTE Confidence: 0.7779449

00:59:43.989 --> 00:59:46.222 a gene that is present on only

NOTE Confidence: 0.7779449

00:59:46.222 --> 00:59:49.228 List A only list D only list C

NOTE Confidence: 0.7779449

00:59:49.228 --> 00:59:50.756 intersection between AMD origin

NOTE Confidence: 0.7779449

00:59:50.827 --> 00:59:53.117 present in all the intersections.

NOTE Confidence: 0.7779449

00:59:53.120 --> 00:59:55.304 So you can use these matrix format

NOTE Confidence: 0.7779449

00:59:55.304 --> 00:59:57.714 to show the intersections and then

NOTE Confidence: 0.7779449

00:59:57.714 --> 01:00:00.049 you can display the cardinality.

NOTE Confidence: 0.7779449

01:00:00.050 --> 01:00:00.714 Of each.

NOTE Confidence: 0.7779449

01:00:00.714 --> 01:00:02.706 Intersection so the number of genes,

NOTE Confidence: 0.7779449

01:00:02.710 --> 01:00:05.742 for example that are only in the PDF

NOTE Confidence: 0.7779449

01:00:05.742 --> 01:00:07.947 error pathway that you see here.

NOTE Confidence: 0.7779449

01:00:07.950 --> 01:00:10.982 The number of genes that are in the

NOTE Confidence: 0.7779449

01:00:10.982 --> 01:00:13.556 common between the EGFR and P-10 path.

NOTE Confidence: 0.7779449

01:00:13.560 --> 01:00:15.430 With that you see here.

NOTE Confidence: 0.7779449

01:00:15.430 --> 01:00:18.580 So this is a way to show the cardinality

NOTE Confidence: 0.7779449

01:00:18.580 --> 01:00:21.790 of the global list that you see here.

NOTE Confidence: 0.7779449

01:00:21.790 --> 01:00:24.352 And also you can rank the intersections

NOTE Confidence: 0.7779449

01:00:24.352 --> 01:00:26.266 between the different sets according

NOTE Confidence: 0.7779449

01:00:26.266 --> 01:00:28.516 to their size to their personality.

NOTE Confidence: 0.7779449

01:00:28.520 --> 01:00:30.480 So it’s much more clearer.

NOTE Confidence: 0.7779449

01:00:30.480 --> 01:00:33.588 To show the structure of the intersection.

NOTE Confidence: 0.7779449

01:00:33.590 --> 01:00:38.385 Rather than using the. A Venn diagram.

NOTE Confidence: 0.7779449

01:00:38.390 --> 01:00:40.898 I skip this because they are.

NOTE Confidence: 0.7779449

01:00:40.900 --> 01:00:43.936 There were some examples of bad

NOTE Confidence: 0.7779449

01:00:43.936 --> 01:00:46.960 usage of graphic in politics.

NOTE Confidence: 0.7779449

01:00:46.960 --> 01:00:49.627 And a lot are looking at online

NOTE Confidence: 0.7779449

01:00:49.627 --> 01:00:51.600 and related to Fox News.

NOTE Confidence: 0.7779449

01:00:51.600 --> 01:00:53.520 Of bad usage of klasa display.

NOTE Confidence: 0.7097202

01:00:53.520 --> 01:00:56.080 So the final part could be how to

NOTE Confidence: 0.7097202

01:00:56.080 --> 01:00:58.030 draw this pad. There is

NOTE Confidence: 0.7097202

01:00:58.030 --> 01:00:59.980 relative they were trying to.

NOTE Confidence: 0.7097202

01:00:59.980 --> 01:01:01.660 Not make the point. Yeah,

NOTE Confidence: 0.82830507

01:01:01.660 --> 01:01:03.676 well they were trying to make him

NOTE Confidence: 0.82830507

01:01:03.676 --> 01:01:06.348 to give a message by distorting the.

NOTE Confidence: 0.8635316

01:01:08.890 --> 01:01:10.762 Yeah, this for example is an

NOTE Confidence: 0.8635316

01:01:10.762 --> 01:01:12.701 issue if you always need to

NOTE Confidence: 0.8635316

01:01:12.701 --> 01:01:14.555 include the zero in your plot.

NOTE Confidence: 0.8635316

01:01:14.560 --> 01:01:16.712 Sir, this is controversial.

NOTE Confidence: 0.8635316

01:01:16.712 --> 01:01:19.034 Let's say that. In general,

NOTE Confidence: 0.8635316

01:01:19.034 --> 01:01:21.146 in Barplots it's a bad idea,

NOTE Confidence: 0.8635316

01:01:21.150 --> 01:01:23.100 but for example is a good

NOTE Confidence: 0.8635316

01:01:23.100 --> 01:01:25.040 idea in in time series,

NOTE Confidence: 0.8635316

01:01:25.040 --> 01:01:27.236 and that's because there in barplots

NOTE Confidence: 0.8635316

01:01:27.236 --> 01:01:30.102 the height of the bar plot is that

NOTE Confidence: 0.8635316

01:01:30.102 --> 01:01:32.094 your main message of the figure,

NOTE Confidence: 0.8635316

01:01:32.100 --> 01:01:34.172 while for example here in in a

NOTE Confidence: 0.8635316

01:01:34.172 --> 01:01:36.169 time series that the main message

NOTE Confidence: 0.8635316

01:01:36.169 --> 01:01:37.904 is how the two trajectories

NOTE Confidence: 0.8635316

01:01:37.904 --> 01:01:39.859 evolve and are interconnected.

NOTE Confidence: 0.8635316

01:01:39.860 --> 01:01:42.982 So the main issue is the horizontal

NOTE Confidence: 0.8635316

01:01:42.982 --> 01:01:46.510 axis and so you can skip the zero.

NOTE Confidence: 0.8635316

01:01:46.510 --> 01:01:47.288 So again,

NOTE Confidence: 0.8635316

01:01:47.288 --> 01:01:50.011 it depends on how much these inclusion

NOTE Confidence: 0.8635316

01:01:50.011 --> 01:01:52.585 or exclusion of the zero distort

NOTE Confidence: 0.8635316

01:01:52.585 --> 01:01:55.530 your your main message of the fever.

NOTE Confidence: 0.8635316

01:01:55.530 --> 01:01:57.580 So how to draw plots?

NOTE Confidence: 0.8635316

01:01:57.580 --> 01:02:00.408 So here are there is an outline

NOTE Confidence: 0.8635316

01:02:00.408 --> 01:02:02.909 of the software that you have,

NOTE Confidence: 0.8635316

01:02:02.910 --> 01:02:04.905 so this is some commercial

NOTE Confidence: 0.8635316

01:02:04.905 --> 01:02:07.420 software from the most from Excel.

NOTE Confidence: 0.8635316

01:02:07.420 --> 01:02:10.290 It's probably the most used or available,

NOTE Confidence: 0.8635316

01:02:10.290 --> 01:02:13.498 but it doesn't allow to plot all the

NOTE Confidence: 0.8635316

01:02:13.498 --> 01:02:16.423 solutions that they did show before, but.

NOTE Confidence: 0.8635316

01:02:16.423 --> 01:02:17.209 For example,

NOTE Confidence: 0.8635316

01:02:17.209 --> 01:02:17.602 Grandpa,

NOTE Confidence: 0.8635316

01:02:17.602 --> 01:02:18.388 Graphpad prism,

NOTE Confidence: 0.8635316

01:02:18.390 --> 01:02:20.790 or Origin Pro are through software

NOTE Confidence: 0.8635316

01:02:20.790 --> 01:02:23.195 that are available and with those

NOTE Confidence: 0.8635316

01:02:23.195 --> 01:02:25.750 that you should be able in an

NOTE Confidence: 0.8635316

01:02:25.750 --> 01:02:27.838 environment that is similar to Excel

NOTE Confidence: 0.8635316

01:02:27.838 --> 01:02:30.486 to produce most of the plots that

NOTE Confidence: 0.8635316

01:02:30.486 --> 01:02:33.474 you saw in the presentation today.

NOTE Confidence: 0.8635316

01:02:33.480 --> 01:02:35.340 So this is commercial software,

NOTE Confidence: 0.8635316

01:02:35.340 --> 01:02:36.450 doesn’t require programming

NOTE Confidence: 0.8635316

01:02:36.450 --> 01:02:37.930 skill on these sides.

NOTE Confidence: 0.8635316

01:02:37.930 --> 01:02:39.886 Are you see the main solutions

NOTE Confidence: 0.8635316

01:02:39.886 --> 01:02:42.380 that are used by data scientists,

NOTE Confidence: 0.8635316

01:02:42.380 --> 01:02:43.764 but that require programming

NOTE Confidence: 0.8635316

01:02:43.764 --> 01:02:45.840 skills so that the two most

NOTE Confidence: 0.8635316

01:02:45.908 --> 01:02:47.948 common languages in data science,

NOTE Confidence: 0.8635316

01:02:47.950 --> 01:02:52.630 RR and Python so far are you have is GG plot.

NOTE Confidence: 0.8635316

01:02:52.630 --> 01:02:54.490 Library for Python.

NOTE Confidence: 0.8635316

01:02:54.490 --> 01:02:57.590 You have matplotlib or Seaborn.

NOTE Confidence: 0.8635316

01:02:57.590 --> 01:02:59.270 At these require programming so,

NOTE Confidence: 0.8635316

01:02:59.270 --> 01:03:02.245 but I would say that the advantage

NOTE Confidence: 0.8635316

01:03:02.245 --> 01:03:05.283 nowadays of using visa is that you can

NOTE Confidence: 0.8635316

01:03:05.283 --> 01:03:08.519 find a lot of really a lot of examples.

NOTE Confidence: 0.8635316

01:03:08.520 --> 01:03:11.593 Because there are a lot of website

NOTE Confidence: 0.8635316

01:03:11.593 --> 01:03:14.596 that where you can choose that

NOTE Confidence: 0.8635316

01:03:14.596 --> 01:03:16.712 you're like data visualization

NOTE Confidence: 0.8635316

01:03:16.712 --> 01:03:20.279 type and you see already the code.

NOTE Confidence: 0.8635316

01:03:20.280 --> 01:03:22.182 That you can use in order

NOTE Confidence: 0.8635316

01:03:22.182 --> 01:03:23.450 to produce the blocked.

NOTE Confidence: 0.8635316

01:03:23.450 --> 01:03:25.778 So I would say that you just need

NOTE Confidence: 0.8635316

01:03:25.778 --> 01:03:28.469 to know how to insert that or how to

NOTE Confidence: 0.8635316

01:03:28.469 --> 01:03:30.841 load that in the this programming

NOTE Confidence: 0.8635316

01:03:30.841 --> 01:03:32.745 environment table of data.

NOTE Confidence: 0.8635316

01:03:32.750 --> 01:03:34.988 And then most of the difficulties

NOTE Confidence: 0.8635316

01:03:34.988 --> 01:03:37.080 are probably in fixing details,

NOTE Confidence: 0.8635316

01:03:37.080 --> 01:03:40.240 so it's very easy to realize the plot,

NOTE Confidence: 0.8635316

01:03:40.240 --> 01:03:41.266 different plot.

NOTE Confidence: 0.8635316

01:03:41.266 --> 01:03:44.344 It's more complicated to adapt the

NOTE Confidence: 0.8635316

01:03:44.344 --> 01:03:47.578 small things that we are to your taste.

NOTE Confidence: 0.8635316

01:03:47.580 --> 01:03:48.492 But so so,

NOTE Confidence: 0.8635316

01:03:48.492 --> 01:03:50.316 this suggestion is that if you

NOTE Confidence: 0.8635316

01:03:50.316 --> 01:03:52.278 do a lot of visualization,

NOTE Confidence: 0.8635316

01:03:52.280 --> 01:03:53.960 it's worth investing in this.

NOTE Confidence: 0.82958233

01:03:57.390 --> 01:04:00.614 Here you see a maybe a future perspective

NOTE Confidence: 0.82958233

01:04:00.614 --> 01:04:03.746 that could be their own online solution.

NOTE Confidence: 0.82958233

01:04:03.750 --> 01:04:05.870 They're already available so summer,

NOTE Confidence: 0.82958233

01:04:05.870 --> 01:04:08.838 for example. You can produce upset plots,

NOTE Confidence: 0.82958233

01:04:08.840 --> 01:04:10.352 Aurora rain plot, Sir,

NOTE Confidence: 0.82958233

01:04:10.352 --> 01:04:12.620 or some other like exotic type

NOTE Confidence: 0.82958233

01:04:12.698 --> 01:04:14.770 of data visualization online.

NOTE Confidence: 0.82958233

01:04:14.770 --> 01:04:16.562 So there are websites,

NOTE Confidence: 0.82958233

01:04:16.562 --> 01:04:19.776 web web servers where you can insert

NOTE Confidence: 0.82958233

01:04:19.776 --> 01:04:23.224 your data as tables and they produce at

NOTE Confidence: 0.82958233

01:04:23.224 --> 01:04:26.506 the data that you want and you have.

NOTE Confidence: 0.82958233

01:04:26.510 --> 01:04:28.390 Some sort of interactivity,

NOTE Confidence: 0.82958233

01:04:28.390 --> 01:04:31.544 so that could be the future. Sure.

NOTE Confidence: 0.82958233

01:04:31.544 --> 01:04:33.882 Where are web servers provide you with

NOTE Confidence: 0.82958233

01:04:33.882 --> 01:04:36.130 the main programming environment?

NOTE Confidence: 0.82958233

01:04:36.130 --> 01:04:38.260 You need just to interfere data

NOTE Confidence: 0.82958233

01:04:38.260 --> 01:04:41.049 and you can see by interactively.

NOTE Confidence: 0.78084993

01:04:43.580 --> 01:04:46.387 By the interaction with the web server.

NOTE Confidence: 0.78084993

01:04:46.390 --> 01:04:48.390 How to customize the data?

NOTE Confidence: 0.78084993

01:04:48.390 --> 01:04:52.856 Most of the solutions right now are.

NOTE Confidence: 0.78084993

01:04:52.860 --> 01:04:55.604 Commercial, and so you need to pay,

NOTE Confidence: 0.78084993

01:04:55.610 --> 01:04:57.968 and that's the drawback of this.

NOTE Confidence: 0.78084993

01:04:57.970 --> 01:05:00.040 But it could be probably the

NOTE Confidence: 0.78084993

01:05:00.040 --> 01:05:01.964 future of matching the programming

NOTE Confidence: 0.78084993

01:05:01.964 --> 01:05:03.860 with easiness of usage.

NOTE Confidence: 0.8579263

01:05:06.440 --> 01:05:08.426 This is a useful resource that

NOTE Confidence: 0.8579263

01:05:08.426 --> 01:05:10.882 you can use also to decide which

NOTE Confidence: 0.8579263

01:05:10.882 --> 01:05:12.994 kind of blocked are you want.

NOTE Confidence: 0.8579263

01:05:13.000 --> 01:05:15.960 So there are a lot of these trees are that

NOTE Confidence: 0.8579263

01:05:16.036 --> 01:05:18.857 depending on what you want to represent,

NOTE Confidence: 0.8579263

01:05:18.860 --> 01:05:20.585 one numeric variable to numeric

NOTE Confidence: 0.8579263

01:05:20.585 --> 01:05:21.965 variables or categorical variables,

NOTE Confidence: 0.8579263

01:05:21.970 --> 01:05:24.007 you can follow the tree and arrive

NOTE Confidence: 0.8579263

01:05:24.007 --> 01:05:26.158 to your to the best graphical

NOTE Confidence: 0.8579263

01:05:26.158 --> 01:05:28.178 solutions to display your data.

NOTE Confidence: 0.8579263

01:05:28.180 --> 01:05:30.126 So I suggest you to visit it

NOTE Confidence: 0.8579263

01:05:30.126 --> 01:05:32.721 also to look at what are the

NOTE Confidence: 0.8579263

01:05:32.721 --> 01:05:34.826 kind of possibilities for data

NOTE Confidence: 0.8579263

01:05:34.826 --> 01:05:36.750 representations that you have online.

NOTE Confidence: 0.8579263

01:05:36.750 --> 01:05:39.130 There are many of these sites and

NOTE Confidence: 0.8579263

01:05:39.130 --> 01:05:41.671 now and that's why it's easy to

NOTE Confidence: 0.8579263

01:05:41.671 --> 01:05:43.837 look at the documentation and also

NOTE Confidence: 0.8579263

01:05:43.917 --> 01:05:46.395 to retrieve and reproduce the code.

NOTE Confidence: 0.8579263

01:05:46.400 --> 01:05:49.616 This is another example I closed with Visa.

NOTE Confidence: 0.8365897

01:05:51.720 --> 01:05:54.420 Patricia, that I find particularly

NOTE Confidence: 0.8365897

01:05:54.420 --> 01:05:57.120 related to data visualization and

NOTE Confidence: 0.8365897

01:05:57.200 --> 01:05:59.864 science is not natural itself about

NOTE Confidence: 0.8365897

01:05:59.864 --> 01:06:02.680 its nature under our observation.

NOTE Confidence: 0.8365897

01:06:02.680 --> 01:06:05.459 And so the science of data visualization

NOTE Confidence: 0.8365897

01:06:05.459 --> 01:06:10.072 is a way to allow more adherence between

NOTE Confidence: 0.8365897

01:06:10.072 --> 01:06:12.076 observation science visualization.

NOTE Confidence: 0.84495366

01:06:16.840 --> 01:06:19.156 Thank you come on.