Dealing with Materials Data: Collection, Analysis and Interpretation

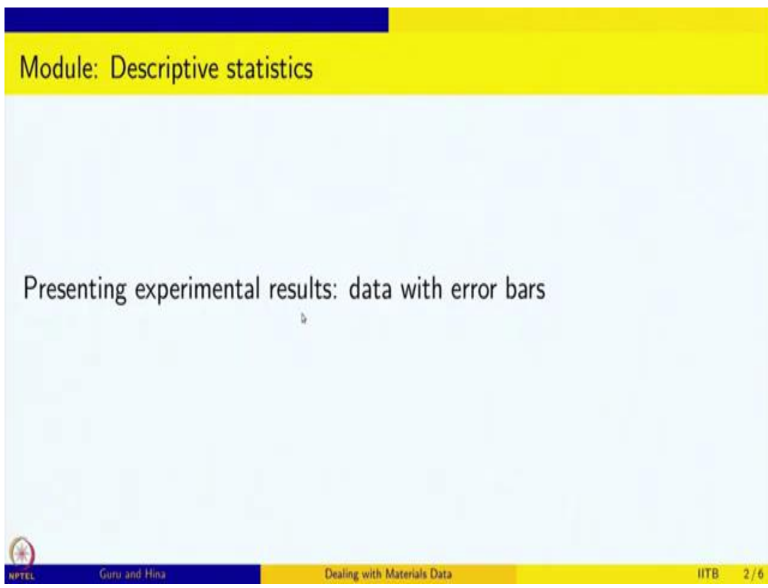
Professor. M P Gururajan
Professor. Hina A Gokhale
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay

Lecture 26

Course Introduction

Welcome to dealing with materials data. This course is about the Collection, Analysis and Interpretation of data from Material Science and Engineering. We are in the module on using or to do descriptive statistics, and we have already looked at ways of looking at data, visualizing data and preparing ranked based reports of the data and presenting them. As well as preparing summary based reports on data and presenting them.

(Refer Slide Time: 00:56)



We are going to continue with representation or presentation of data, and this session we are going to look at presenting experimental results in which the data is presented with error bars. Sometimes, we know that the data is obtained with given error bars. How do we incorporate this information when we present the experimental results as well as in interpreting results, without the error bars if you interpret you might make wrong interpretation? So, we want to look at these two aspects.

(Refer Slide Time: 01:23)

ETP copper: conductivity data with errors

% Deformation	Mean Conductivity (deformed)	Standard Deviation (deformed)	Mean Conductivity (annealed)	Standard Deviation (annealed)
2.9	101.3	0.1	102	0.1
8.7	101	0.1	102	0.1
12.1	100.8	0.1	102	0.1
15.9	100.5	0.1	101.9	0.1
22.5	100.2	0.1	101.9	0.1
30	100	0.2	102	0.2
40	99.7	0.1	101.9	0.2
60	103	0.2	106.1	0.3

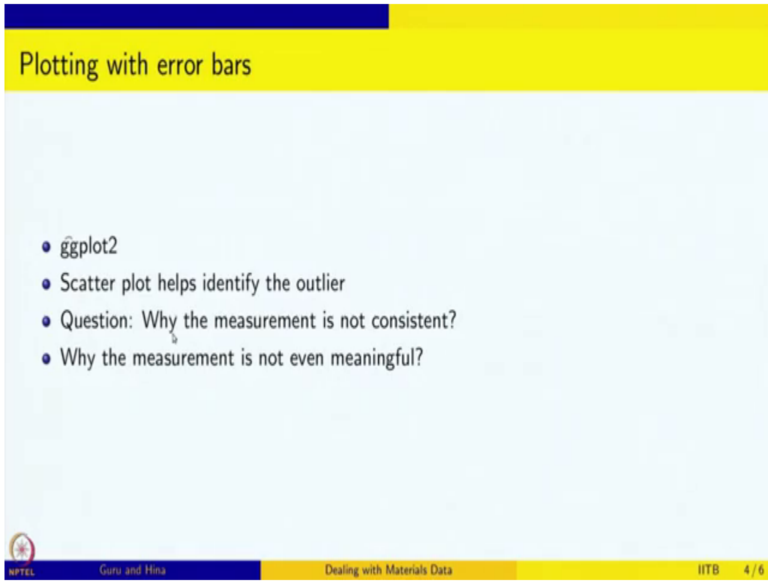
For doing this, we are going to go back to the conductivity data, from the electrolytic tough pitch copper from the thesis of Dr. Harshvardhan and this data consists of five columns. This is the percentage deformation that was given to the sample 2.9 percent, 8.7 percent, 12.1 percent etc. Up to 60 percent deformation and in the deformed state, the conductivity measurements were made and the mean conductivity and standard deviation is given.

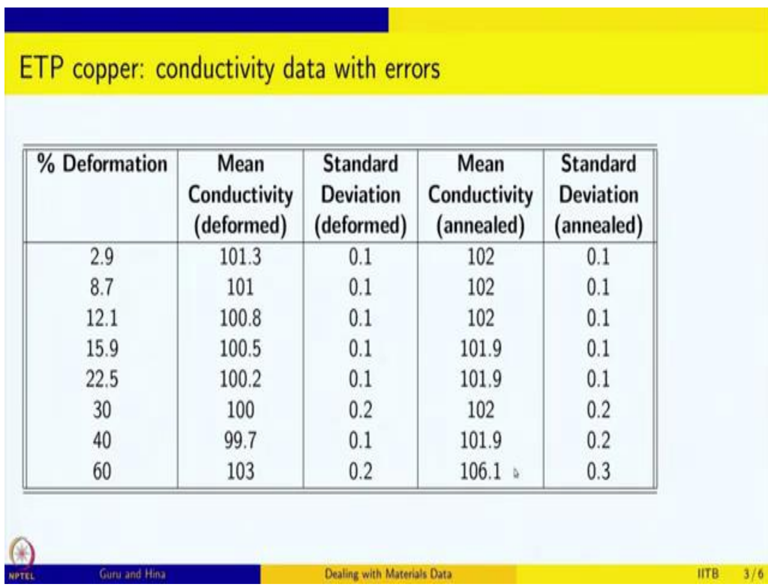
And then, all these deformed samples were annealed for the same amount of time and at the end of the annealing treatment these samples are taken, again the conductivity is measured and the mean conductivity and standard deviation is given, and in each 1 of these cases 20 measurements were taken to calculate this mean conductivity and standard deviation.

So, the experiment was repeated, the measurement was repeated, 20 times to make sure that you get the proper conductivity and standard deviation. So, each of these values is a result of 20 measurements, you deform by 2.9 percent, you make 20 measurements, get these quantities, annealed the sample, make another 20 measurements and get this values.

So, this is the data that is there and it is stored as a CSV file, which is what we are going to load and use, and as you can see, the mean conductivity of the deformed samples is decreasing and annealed sample remains more or less constant, and there is large conductivity that is reported for 60 percent deformed samples in both the deformed state and the annealed state as compared to the other values.

(Refer Slide Time: 03:31)



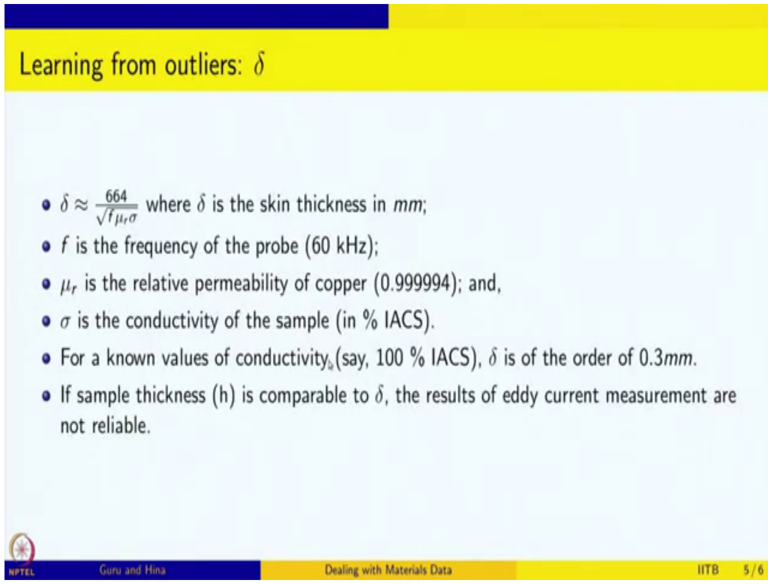

ETP copper: conductivity data with errors

% Deformation	Mean Conductivity (deformed)	Standard Deviation (deformed)	Mean Conductivity (annealed)	Standard Deviation (annealed)
2.9	101.3	0.1	102	0.1
8.7	101	0.1	102	0.1
12.1	100.8	0.1	102	0.1
15.9	100.5	0.1	101.9	0.1
22.5	100.2	0.1	101.9	0.1
30	100	0.2	102	0.2
40	99.7	0.1	101.9	0.2
60	103	0.2	106.1	0.3

Guru and Hina Dealing with Materials Data IITB 3/6

So, we will look at this aspect also as we move along. So, you can plot data with error bars and Scatter plots help identify the outlier, and you can also ask the question why the measurement is not consistent, why the measurement is not consistent to that as it is deformed this is going down and suddenly you find that the value has increased, and is this even meaningful? I mean, can you get in 60 percent deformed sample conductivity which is much higher than some sample in which you only had some about 3 percent deformation.

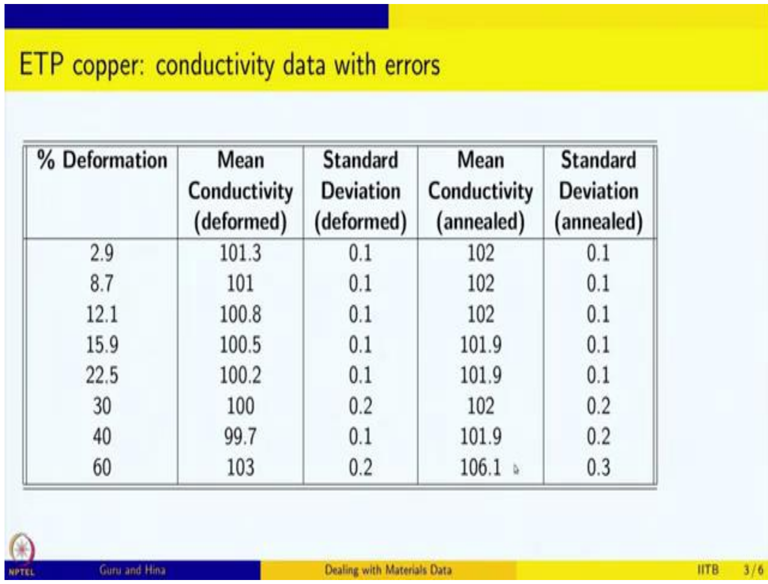
(Refer Slide Time: 04:26)



So, this is a question that we have to look at, and to answer this questions, so we are going to look at this aspect. There is some known as skin thickness and this skin thickness is given in millimetres and it is given by this formula, it is approximately equal to 664 divided by square root of f mu r sigma, where f is the frequency of the eddy current probe and in this case it was 60 kilo Hertz, mu r is the relative permeability of copper, so this you can read from tables this is some 5 9's and 4.

So, that is the value and sigma is the conductivity of the sample in percentage IACS, and we know that the conductivity of copper, the ETP copper that we are dealing with is somewhere around 100, 102 etc.

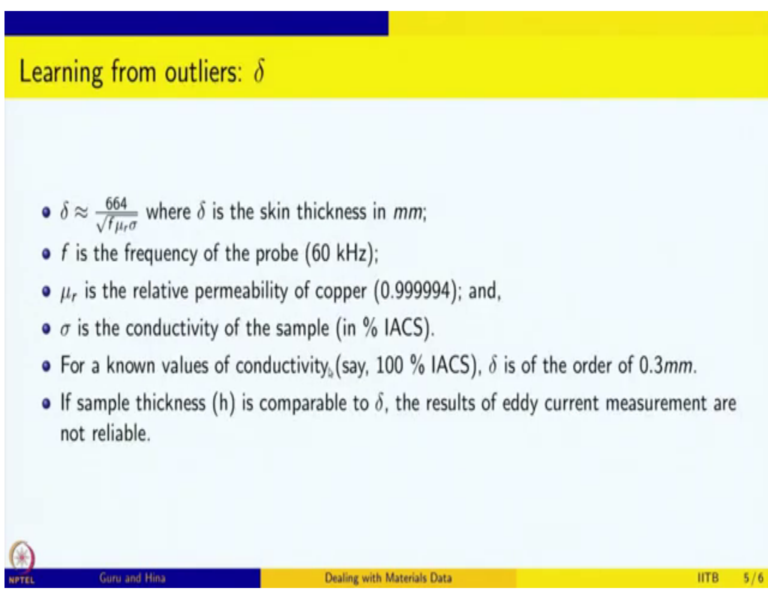
(Refer Slide Time: 05:24)



% Deformation	Mean Conductivity (deformed)	Standard Deviation (deformed)	Mean Conductivity (annealed)	Standard Deviation (annealed)
2.9	101.3	0.1	102	0.1
8.7	101	0.1	102	0.1
12.1	100.8	0.1	102	0.1
15.9	100.5	0.1	101.9	0.1
22.5	100.2	0.1	101.9	0.1
30	100	0.2	102	0.2
40	99.7	0.1	101.9	0.2
60	103	0.2	106.1	0.3

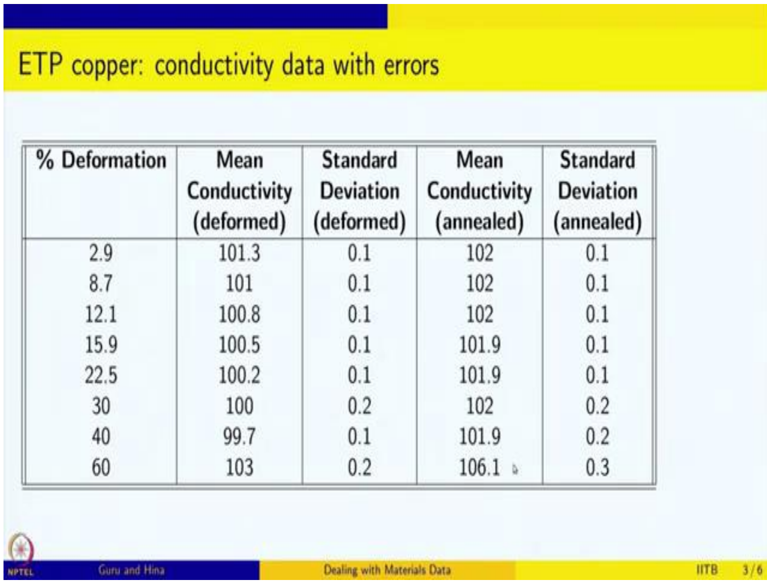
Let us say that it is about 100 IACS, then delta is of the order of 0.3 millimetre. If the sample thickness is comparable to this value, then the results of eddy current measurements are not reliable and the reason why in 60 percent deformed samples, we are getting meaningless numbers and inconsistent numbers is because, given our initial thickness, when we made 60 percent deformation on these samples, the thickness became comparable to this skin thickness and so this information.

(Refer Slide Time: 05:50)



I mean that we need to go look at this value and find out why it is not consistent came from looking at the data and once you pay attention to such outliers and try to understand why such outliers exist, of course, you learn a little bit more about the experiment and the data and the material and so on and so forth. So, it is very important to look at the outliers and try to understand them and try to make sense out of such results because everything else being the same.

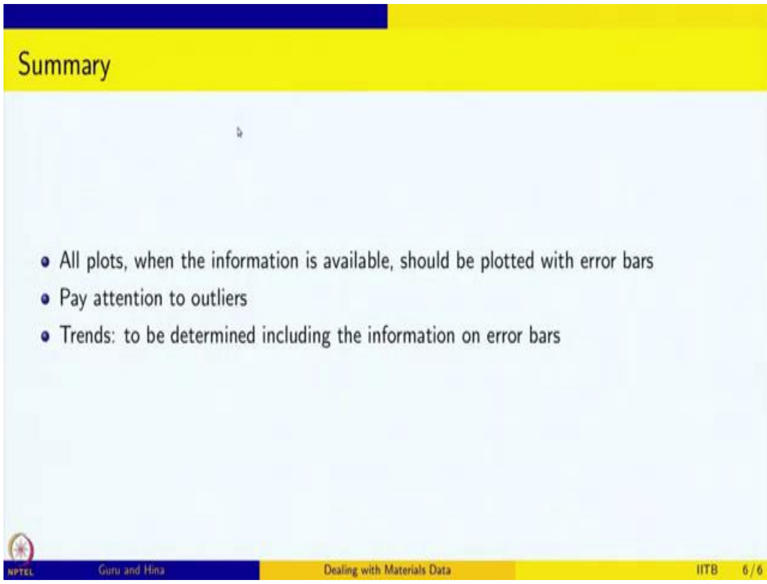
(Refer Slide Time: 06:12)



% Deformation	Mean Conductivity (deformed)	Standard Deviation (deformed)	Mean Conductivity (annealed)	Standard Deviation (annealed)
2.9	101.3	0.1	102	0.1
8.7	101	0.1	102	0.1
12.1	100.8	0.1	102	0.1
15.9	100.5	0.1	101.9	0.1
22.5	100.2	0.1	101.9	0.1
30	100	0.2	102	0.2
40	99.7	0.1	101.9	0.2
60	103	0.2	106.1	0.3

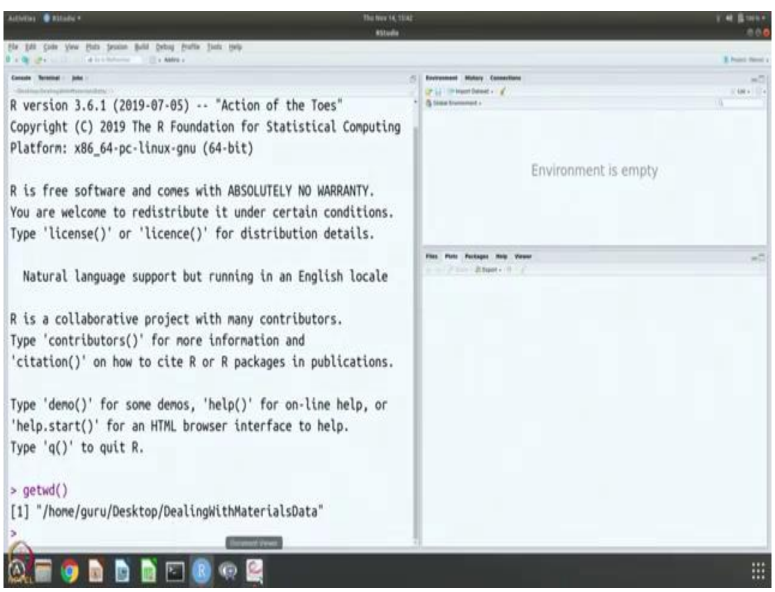
If you make 20 measurements and if you get a number like this, of course, you can also see the standard deviation also is relatively high but that is high for all 30, 40, 60. For example, even in annealed sample these give relatively higher standard deviations, but the numbers themselves are does not make sense and that is because of this reason. So, it is very important to pay attention to outliers and learn from them.

(Refer Slide Time: 06:43)



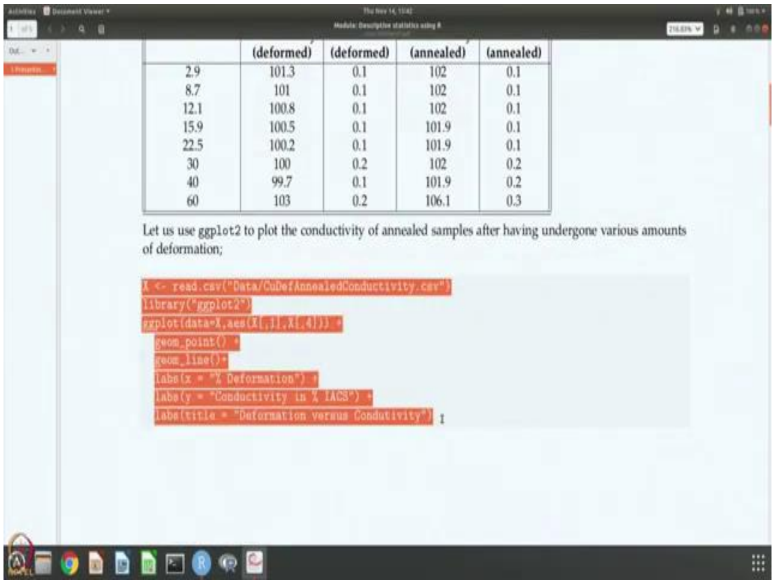
So, all plots when the information is available should be plotted with error bars. If you have information on standard deviation, which you would generally have. You should always put the error bars in the plots, it is incorrect to plot without error bars and I will show you why and how. And one should always pay attention to outliers, and when we look at trends in the plots, we should look at the information that is coming from error bars also to interpret the result. So, this is what we are going to learn from this session and we are going to learn that by doing the analysis in R.

(Refer Slide Time: 07:24)



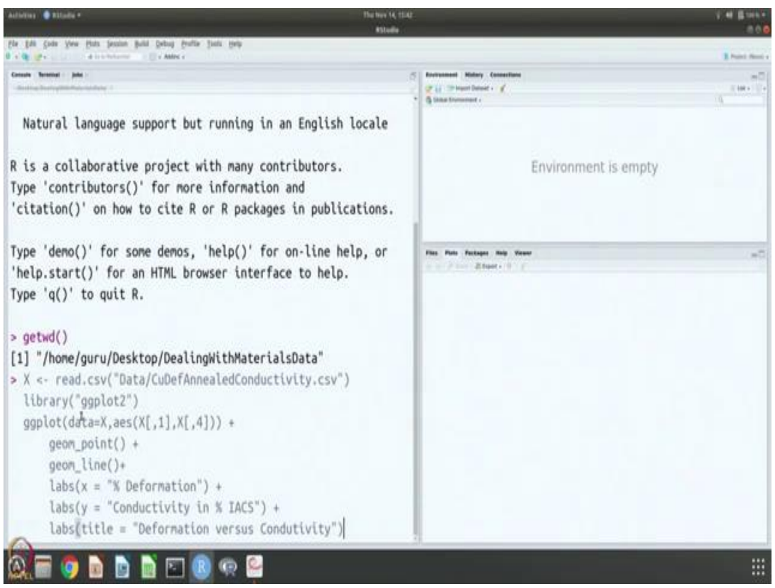
So, let us do that, so it is version 3.6.1 and we are in the, dealing with materials data directory.

(Refer Slide Time: 07:39)



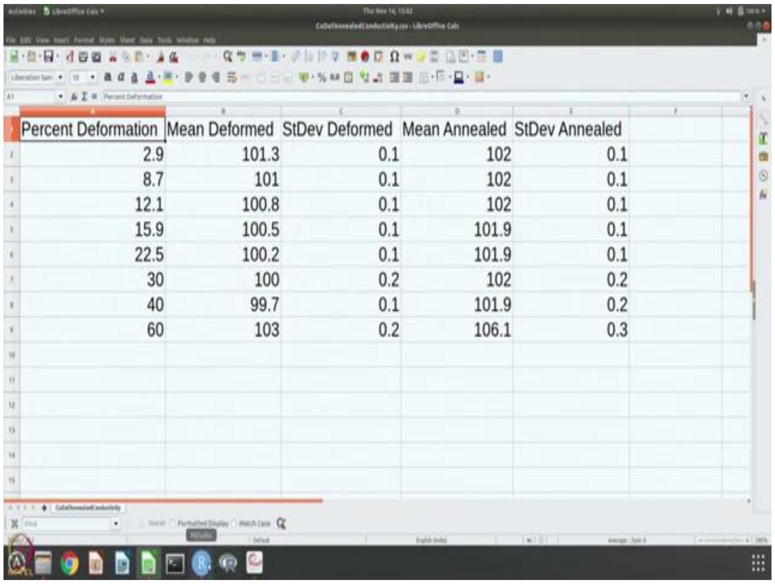
So, we are ready to start. So, let us get the data read and let us plot it.

(Refer Slide Time: 07:46)



Import the data. So, this imports the data, this is copper deformed and annealed copper conductivity and we are you going to use ggplot and as usual to ggplot you have to tell which is the data, what is it that you are trying to plot in the data, so let us open the data file and see to understand what is it that we are plotting.

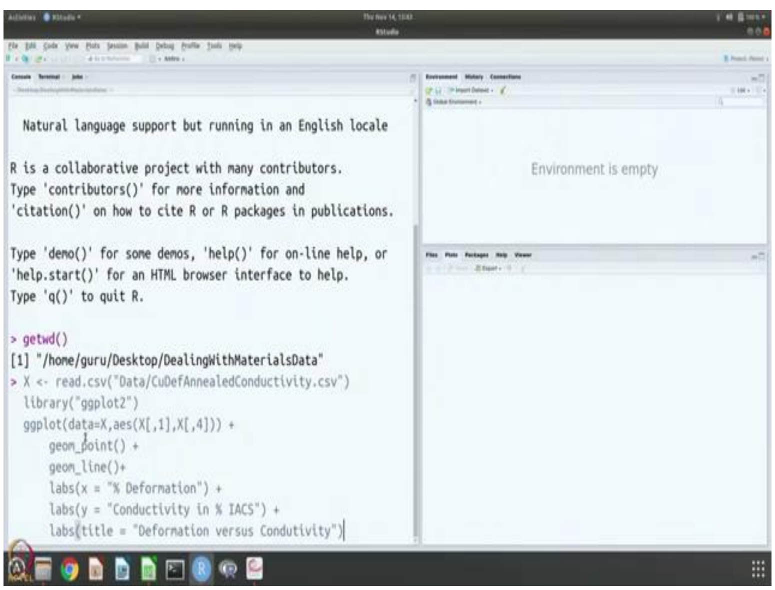
(Refer Slide Time: 08:16)

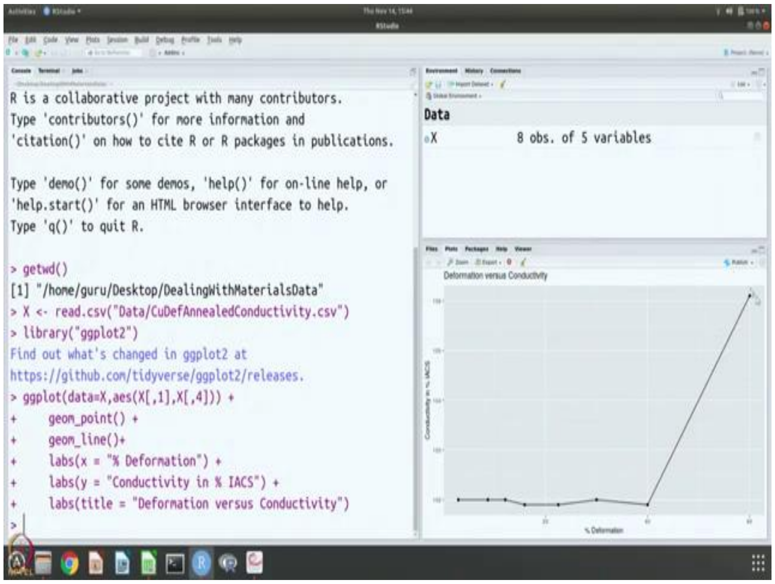


Percent Deformation	Mean Deformed	StDev Deformed	Mean Annealed	StDev Annealed
2.9	101.3	0.1	102	0.1
8.7	101	0.1	102	0.1
12.1	100.8	0.1	102	0.1
15.9	100.5	0.1	101.9	0.1
22.5	100.2	0.1	101.9	0.1
30	100	0.2	102	0.2
40	99.7	0.1	101.9	0.2
60	103	0.2	106.1	0.3

So, it is percentage deformation, mean conductivity of deformed sample, standard deviation of deformed sample, mean conductivity of annealed sample and standard deviation for annealed copper. So, this is the data. So, what we want to plot this percentage deformation and the conductivity, mean conductivity.

(Refer Slide Time: 08:35)

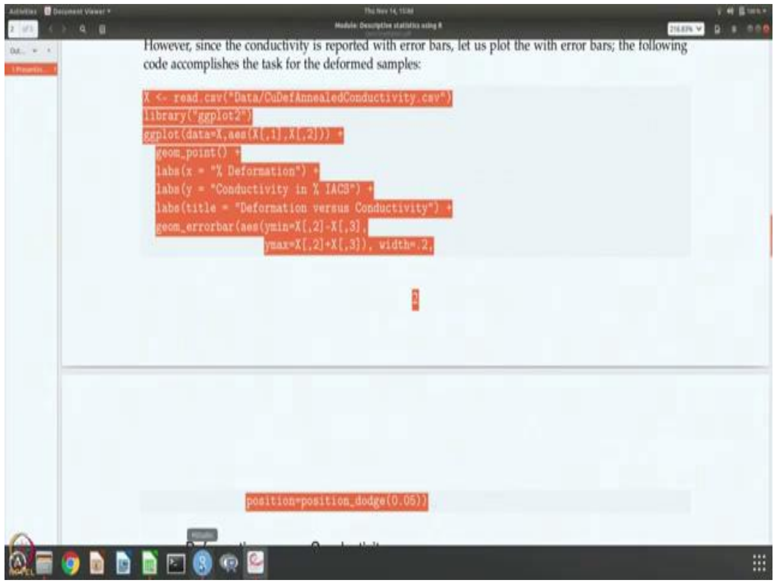




So, this is for annealed sample, so it is percentage deformation and after annealing what is the mean value, and we are going to put points and typically you will see that students also join the data points with a line. So, I have also added a line and then x axis is percentage deformation, y axis is conductivity in percentage, IACS and it is titled deformation versus conductivity. Let us do this. So, if you see then you get a plot like this and I have seen sometimes this kind of plot being interpreted as follows.

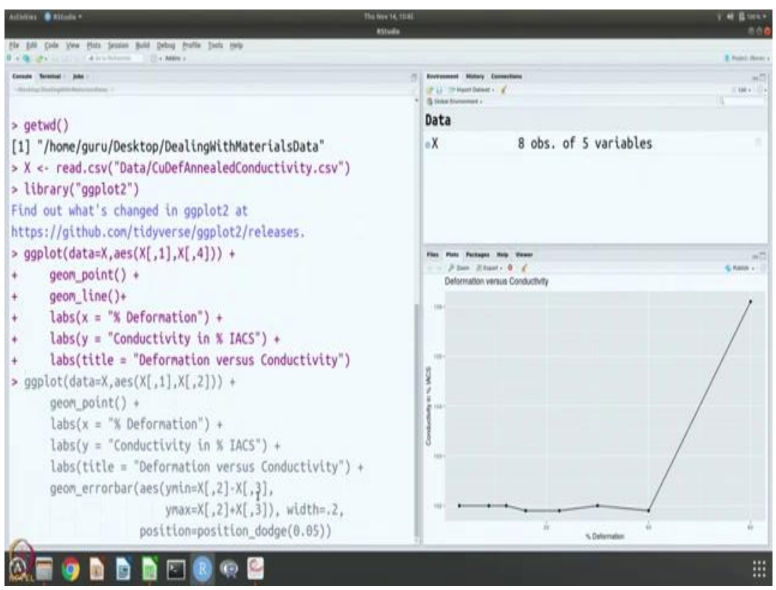
So, initially the conductivity remains a constant with deformation, these are annealed samples and then it decreases then it increases then it decreases and then it increases a lot. If you just look at the data and if you just join them by points, and that seems to be the way the data is behaving. However, these interpretations are wrong, because we know that irrespective of what the deformation was if the sample is well annealed, the conductivity should remain more or less constant.

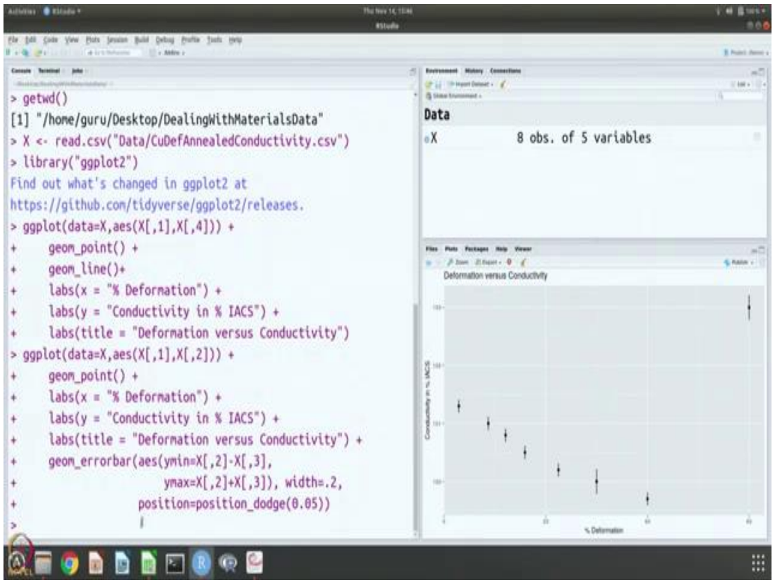
(Refer Slide Time: 10:12)



It is not clear why suddenly there is an increase but these all should be the same value that is the expectation from knowing the physics and material science of this system and that information you can also see if you plot with the error bars. So, that is what we are going to do.

(Refer Slide Time: 10:24)

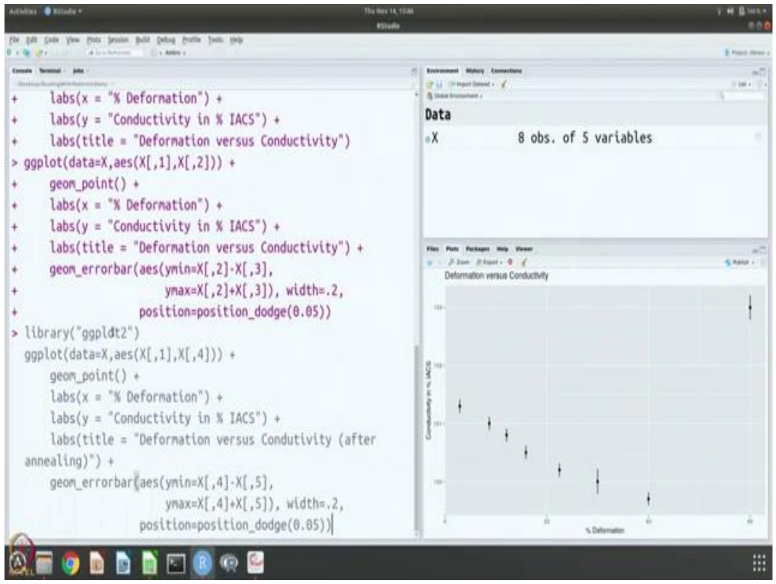


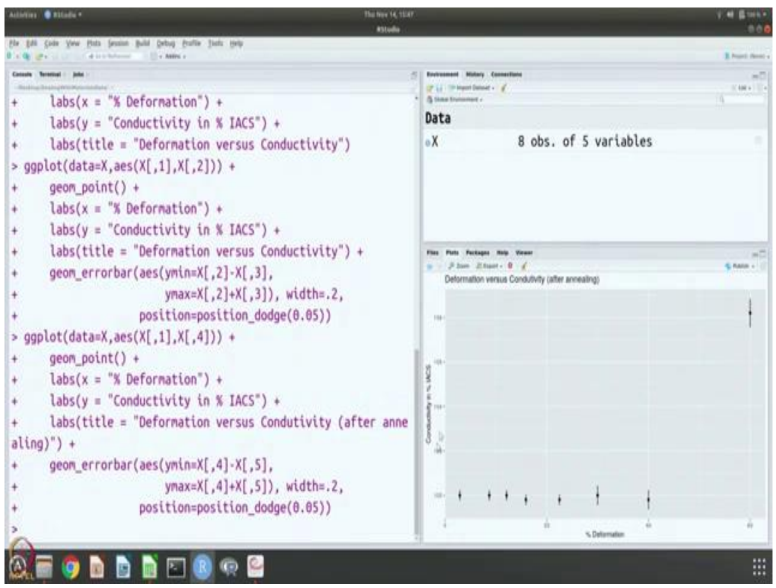


So, let us do that. So, we have read the data, so it is not needed. So, we are going to use ggplot data is X and we are going to look at the deformed sample first. So, it is X 1 versus X 2 and we are going to put points and we are going to label the X, Y axis and we are going to give a title, it is deformation versus conductivity and the data points are plotted with error bars, and what is the error bar? So, we are going to take the standard deviation plus or minus about the mean and so that is what we have done here and let us look at the plot.

So, you can see that the conductivity with deformation decreases, and these are the error bars and you know, this decrease is, for example, these two there is some overlap but this and this for example or this and this for example or this and this for example, have certainly decrease so you can see a trend that the conductivity is decreasing with increasing percentage deformation in deformed samples. Except that 60 percent shows really large value more than even something that is deformed only a few percent, so it does not make sense.

(Refer Slide Time: 12:00)

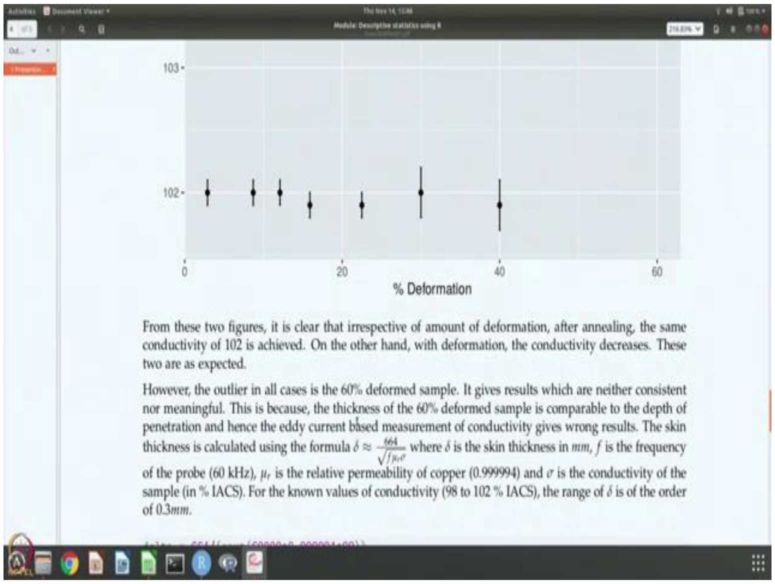




Anyway, so that is 1 and let us do now for the annealed sample and it is the same and of plot. So, we have the data and we are saying 1 to 4 and the error bar is given by the standard deviation, which is in the fifth column and we are plotting with geometry. Now with error bars you can see because the error bars are spanning all these points except these two were here also it is touching.

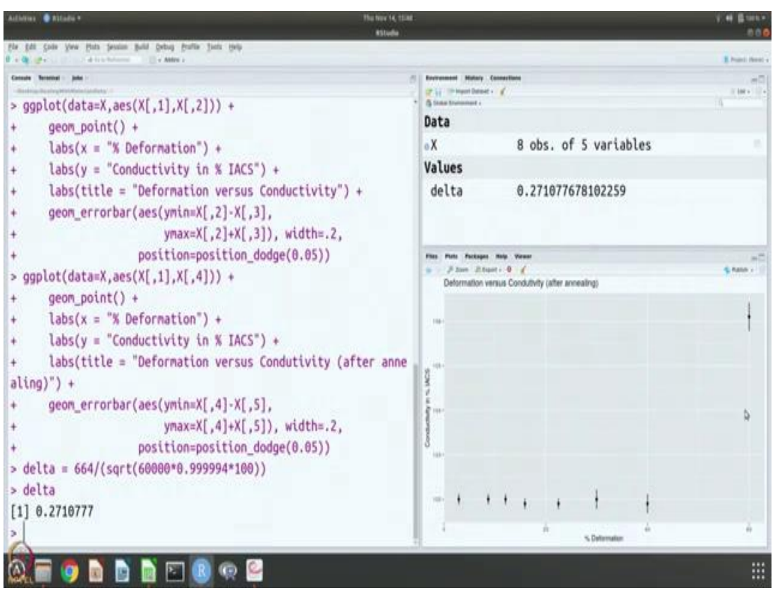
So, except for this point, which is just touching you can see that more or less the conductivity remains a constant, which makes sense because that is what the expectation is. So, after annealing we see that irrespective of what percentage deformation you gave, annealing got rid of all the effects of that and you have this conductivity to be here more or less a constant and of course 60 percent is giving you something that is not meaningful.

(Refer Slide Time: 13:04)



So, like we mention and one has to, so, here this is clearer. So, the data is somewhere here and one should consider this to be a straight line, this little bit of deviation here maybe it is like this, so it is a straight line with small deviation here. So, why is the 60 percent deformed sample showing different value, like I mentioned it is because of this.

(Refer Slide Time: 13:35)



So, let us calculate that delta and that delta was 664 divided by square root of 60000 that is a 60 kilo Hertz and the mu r and the conductivity is about a 100, and the delta value is about 0.3 millimetre. So, it is about 0.3 millimetres. So, it so happened that in this case, in the 60 percent deformed case the thickness of the sample is about 2 or 2 and a half times this 0.3 and that is the reason why we got this outlier.

So, we now understand why this happens and of course the conductivity is not really 100. We know that because if you have deformed sample, let us say 60 percent deformed, maybe it is conductivity will go down 2, 3 percentage IACS points and very well annealed samples can also give you above 100 like we had a mean of about 101 but it sometimes you can also get 102 etc. So, this leads us to the next question.

So, we found an outlier, we found an explanation for that, but it so happens that one of the quantities that goes into that formula is not quite correct. It is not just a number; it has some uncertainty or error and what is the effect of that error on this quantity that we are calculating, because we just used 100, we could have used 102, we could have used 98 and if somebody has some slightly impure copper and they have let us say the, the conductivity of that copper alloy to be some 40 percent IACS then what is the skin depth?

So, if there is a large variation in this quantity what happens to this quantity, or if there is some small error or uncertainty in the measurement what happens to this quantity? So, this is the next question that we want to address and that we will do in the next session. Thank you.