### Dealing with Material Data: Collection, Analysis and Interpretation Professor. M. P Gururajan Professor Hina A Gokhale Department of Metallurgical Engineering and Material Science Indian Institute of Technology Bombay Lecture 24 Case Study: Grain size in two phase steel

Welcome to dealing with materials data, we are looking at the collection analysis and interpretation of data from the material science and engineering and we have done two modules, we have done an introduction to R module and this module is meant for doing descriptive data analysis using R and we have already done analysis on two sets of data, one is on the conductivity of electromagnetic, tough pitch copper and we also did an analysis on grain size and we found that the conductivity measurements and the grain size data are of two different types.

The case of conductivity measurement, the repeated measurements just gave errors about some mean value, because those are measurement errors or random errors or uncertainties associated with the experiment. But, on the other hand, when we measured something like a grain size, it has naturally a distribution. Not all grains are of the same size, they follow a distribution and the distribution is not normal or Gaussian or anything like that it is slightly more complicated.

So, it makes sense to represent the data in this case, not just with the mean and standard deviation, like we did in the case of conductivity, but also by plotting something like a histogram plot to indicate how the data looks like.

(Refer Slide Time: 01:49)



So let us now continue with the second data set. It is also data set which deals with distributions and this is the grain size data set to dot CSV and this is slightly complicated data set because it contains grain size of two phases and we are going to carry out all the rank based and property based analysis for both these phases and in all the cases we are going to do it, one next to the other.

So, we can have information about grain sizes of these two phases. But, we will also have a comparison between the grain sizes of phase one and phase two in terms of their rank based and property based summaries. So, that is what we are going to do.

(Refer Slide Time: 02:32)



One of the things that you have to carry from this session is that if you just looked at the grain size, likely reported conductivity just by reporting the mean and standard deviation. You will see that the grain size of grains are phase one is 24.1, plus or minus point 4 and grain size of grains of phase 2 is 23.4, plus or minus 2.

Sometimes students make a mistake of thinking that these two grain sizes are different and grain size of grains of phase two is smaller than one that is not true. Because you also have to take into account the fact that there is an uncertainty, when we say 23.4 plus or minus 2, it means that the number could be anywhere between 21.4 to 25.4 and so the 21.4 to 25.4 actually covers 24 also, when we say 24.1 plus or minus point 4 that means, it is 23.7 to 24.5.

So, within the error bars, all that you can say is that these two phases have the same grain size. However, if you look at the histogram as we will do or look at the quantile information, you will see that the mean and standard deviation are not the complete picture. These two grain size distributions are very different even though they might end up giving you more or less the same grain size and that is the part that has to come through the session. So, that is what we are going to see and we are going to understand.

(Refer Slide Time: 04:26)

Addether 🔮 Ritada v	The New 14, 1991 #Studie			1.4.5.
fie feit Cole Yew Pats Jessie Beld Debug Pathe Just 1949				2 Part de
County Terretories : John -	1	Revenues Mart	niy Commertians	
Type 'q()' to quit R.		· G Cone Branned		1.000
		Data		
> getwd()		οX	3664 obs. of 6 va	riables
[1] "/home/guru/Desktop/DealingWithMater	ialsData"			
> X <- read.csv("Data/GrainSizeDataSet2.	csv")			
> str(X)				
'data.frame': 3664 obs. of 6 variable	s:			
\$ Phase.Identity	: int 111111	Files Plats Pack	agan Noig Vesar	-
1111		Jan 1	Etuart - 0 🧹	S Ratio -
<pre>\$ Integer.identifying.grain</pre>	: int 9 78 117 11		1.000 000 1.0	1.1.1.1
8 131 149 154 181 184 234		Passing		1
<pre>\$ Number.of.measurement.points.in.the.g</pre>	rain: int 222223	11	B. 198	10 10 10 10 10 10 10 10 10 10 10 10 10 1
3 4 2 2		1	hannen an	
\$ Area.of.grain.in.square.microns	: Factor w/ 224 le			1 1 1
vels "0.00624","0.00935",: 1 1 1 1 1 2	2311			
<pre>\$ Diameter.of.grain.in.microns</pre>	: num 0.09 0.09			-1-1
0.09 0.09 0.09 0.11 0.11 0.13 0.09 0.09		-	1655.01 2001	
\$ ASTM.grain.size	: num 24.3 24.3 2		darks in	
4.3 24.3 24.3 23.7 23.7 23.3 24.3 24.3 .	••	1	194	
> plot(X)		24	P00007	
		100/000		
a 🚍 🚓 🔊 🔊 🔊 🔄 🔊 🚳 🙆				

So, let us do as usual we want to open R and so we have to check the version, we have to make sure that we are in the right directory and so we are all set to now start importing the data and in this case, the data is in CSV format. So, read dot CSV and it is from data directory and grain size two is the data set two. As you can see, there are 3664 observations and there are 6 variables 6 variables because in addition to the 5 variables that you saw in the other case, here the grain at the phase identity is also included before grain identity.

Of course, you can get more information by looking at the structure of this object x. It is a data frame, there are 300 and, 3,664 observations, there are 6 variables, the variables are phase identity, integer, identifying grade number of measurement points, the area of grain diameter of grain and ASTM grain size, like we did earlier, we are going to be worried only about integers identifying the grain and ASTM grain size. Of course, for the two phases, phase identity is the, is this is a 2 phase microstructure. So, there are 2 phase identities 1 and 2 and we are going to be working with these two.

First thing of course is to plot we can always try plot x so it is a 6 by 6. So, you should have 36 boxes 3, 4, 5, 6 and 3, 6. So 36 boxes are there and so against phase identity against integer identifying grain etc, all the parameters are plotted and so you can see how the data looks. So, this is the first step and this does not distinguish between the grain identities.

(Refer Slide Time: 06:58)





So, let us do the plotting of, what is this command? So, we want to plot the grain identity versus grain size and we want to color them according to the phase identity. So, phase identity, one should get one color and phase identity two should get another color.

So, this is how the plot looks. So, these are the sizes and these are the grain identities, you can also switch them, which is how it was the original one and you will get this I am showing this plot because this is closer to what you would see in the case of a dot chart.

But it is very difficult here we know to distinguish between these black points and red points. However, in dot chart, they will be separated. So, dot chart is always a nice way of visualizing the data instead of plain scatterplot that you can make. Of course, you can also play with the plain scatterplot yourself and come up with commands which will separate these data. But there is an easier way by using the existing libraries.

(Refer Slide Time: 08:12)



So, that is what and here it is much clearer. So, there are lots of red points and fewer black points and the black points are all clustered here and the red points are spread all over. So, black points are between 20 and 24. Whereas red points are between 12 and 24 and odd. So, this is an important point. So, we will come back to it.

(Refer Slide Time: 08:38)





So, let us do the other thing. Let us separate out the data. So, this tells R that we are going to make two plots, and one on top of another and those two plots we are going to plot by using phase one phase two. So, the grep command, so is going to get all the line numbers or rows, which have data of phase one and this grep command is going to get the row numbers of all the data points which has data for phase two, that is what I1 and I2 do. So, if I plot x I16 versus xI12.

So, this will be only data that is corresponding to phase one, because we have separated that those line numbers and I2 is for those data rows which have data about phase 2. So, that is what this is and of course we are going to label them as phase 1 and phase 2 and x label is ASTM grain size y label is grain ID.

(Refer Slide Time: 09:53)



So, let us do this plot and you can see that phase 1 has grain sizes, ranging somewhere between 20 and 24 and odd and phase 2 has between 12 and 24. So, this is about to 3, this is about 12 the spread is 4 times as much, like I said we later we are going to see that both of them are willing to give you a grain size which were somewhere which is somewhere about 24 both are going to show you the same grain size and of course, this is going to show more of spread it will show a spread of two, as compared to spread of 0.4 here.

So, 5 times however, looking at this data now, it is very clear that the phase one and phase two grain sizes are completely different in terms of their distribution, even though the overall grace properties like me might be the same.

So, that is the point behind making this broad. Like I said the, you do not have to make all this two plot separate, dot chart will do it automatically for us. So we are going to look at that. But before we do the dot chart.

# (Refer Slide Time: 11:15)

Ritidle		
i fall Gode. Yow Buts Jensin Bull Getug Pullis Just getu		1.00
mante familiari : juin -	S Revenues Malay Constitute	
a	· · · · · · · · · · · · · · · · · · ·	10
ab="Grain ID")	Data	
<pre>stem(X[i1,6])</pre>	•X 3664 obs. of 6 variables	
a construction and a construction	Values	
The decimal point is 1 digit(s) to the left of the	i1 int [1:457] 1 2 3 4 5 6 7 8 9 1	θ
	i2 int [1:3207] 458 459 460 461 46	2 46
208   0		
210	Files Parts Pectages Role Versar	
212	# Phone diffuset • 0 d	5.00
214	Phase 1	
216   0		- 1
218		
220 00	° • • • • • • • • • • • • • • • • • • •	
222 00	21.0 21.5 22.0 22.5 23.0 25.5 20	4.0
224   0	ASTM grain size	
226 0000	Dises 2	
228	Friday 2	
230 000000		Ш
232   000000000000000000		
234	12 14 18 18 20 22	24
236   00000000000000000000000000000000000	ASTM grain sitre	
brian 🔹 Aldada *		Y #1
nter Bitala - Terrer (sis 188 pår gen pas post båt påg gen bå påg på	*	1.4
niner ∰ 2024år + The Nex ( 10 100 jule jule jule jule jule jule jule jule	a fermant May, Gazdan	14
niter ∎ Stala - The Nex ( 1) Dis Data Spec Spec Species Bull Species	S Research Many Constitutes S State Second S (	14
nine €itida - Termert til pit pår gen pan pall pag palle pår pår på € 2010 – 2 distansen © 1 killet i konstruktionen som	S Antonional Matrix Constitute C S to Trans France - X S Total Sources	14
the month of the		1.4
the month of the		
Next € Statute * The movie (1) (All Solar Spece Species Build Desig Portin Spece Spece Resolution Resolution * Spece Species Build Desig Portin Spece Spece Resolution Resolution * Spece Species Build Desig Portin Spece Spece Resolution * Species Build Desig Portin Spece Species Resolution * Species Build Desig Portin Species Spec		8
Aller of Balance - The most 4 the The part of the period part of the period pe	Determine         Operations           Data         64           eX         3664 obs. of 6 variables           Values         1           int [1:457] 1 2 3 4 5 6 7 8 9 1           iz         int [1:457] 1 2 3 4 5 6 7 8 9 1	0
Normal         Normality	Determine         Operation           Bit and boards         Construction           Data         Construction           X         3664 obs. of 6 variables           Values         11           11         int [1:457] 1 2 3 4 5 6 7 8 9 1           12         int [1:3207] 458 459 460 461 46	0 2 46
Note:         Note: <t< td=""><td><pre>     browner many formation     f = 1</pre></td><td>.0 2 46</td></t<>	<pre>     browner many formation     f = 1</pre>	.0 2 46
None         None <th< td=""><td>Interment Make, Learning           Interment Make, Learning           Interment Make, Learning           Data           X           J664 obs. of 6 variables           Values           i1         int [1:457] 1 2 3 4 5 6 7 8 9 1           i2         int [1:3207] 458 459 460 461 46           The Makem Makement &amp; Makement           Abort State: 0</td><td>ιθ 2 46</td></th<>	Interment Make, Learning           Interment Make, Learning           Interment Make, Learning           Data           X           J664 obs. of 6 variables           Values           i1         int [1:457] 1 2 3 4 5 6 7 8 9 1           i2         int [1:3207] 458 459 460 461 46           The Makem Makement & Makement           Abort State: 0	ιθ 2 46
Note         District         Note were fit if the most if th	Marrier Many Communic      Prove Theorem (	ο 0 2 46
Control         Statute         Statute         Statute         Statute           101         Statute         Statute         Statute         Statute	Internet         Way Constant           If is a part base - (         If is a part base - (           If is a part base - (         If is a part base - (           If is a part base - (         If is a part base - (           If is a part base - (         If is a part base - (           If is a part base - (         If is a part base - (           If is a part base - (         If is a part base - (           If is a part base - (         If is a part base - (	0 2 46
Bit Bit Syme         Bit Bits         Bit Bits         Bit Bits         Bit Bits         Bit Bits	Internet Matty framework           If if much house           If if if much house           If	ιθ 2 46
Dirty         Diry         Dirty         Dirty <thd< td=""><td><pre></pre></td><td>8 mar 10 2 46</td></thd<>	<pre></pre>	8 mar 10 2 46
Bit Solution         Discrete Solution <t< td=""><td>Internet         Market Assessment           Image: Second Assessment         Image: Second Assessment           Data         3664 obs. of 6 variables           Values         Int [1:457] 1 2 3 4 5 6 7 8 9 1           I1         int [1:3207] 458 459 460 461 46           The met Network New Year         Image: Second Assessment Ass</td><td>θ 2 46</td></t<>	Internet         Market Assessment           Image: Second Assessment         Image: Second Assessment           Data         3664 obs. of 6 variables           Values         Int [1:457] 1 2 3 4 5 6 7 8 9 1           I1         int [1:3207] 458 459 460 461 46           The met Network New Year         Image: Second Assessment Ass	θ 2 46
Other (* 1)         Other (* 1) <thother (*="" 1)<="" th=""> <thother (*="" 1)<="" th=""></thother></thother>	Marriest Name, Communication           Image Name, Communication     <	0 2 46 40
Bit Def.         Def. <thdef.< th=""> <thdef.< th=""> <thdef.< th=""> <t< td=""><td><math display="block"> \begin{array}{c} \hline \\ \hline </math></td><td>.0 2 46</td></t<></thdef.<></thdef.<></thdef.<>	$ \begin{array}{c} \hline \\ \hline $	.0 2 46
Bit Det gene         Bit Det gene<	Internet         Martine           Image: strategy of strategy           Image: strategy<	ιθ 2 46
No. 00         District         District <thdistrict< th=""> <thdistrict< th=""> <th< td=""><td>Instant         Margin Learning           Data         3664 obs. of 6 variables           Values         1           11         int [1:457] 1 2 3 4 5 6 7 8 9 1           12         int [1:3207] 458 459 460 461 46           Verify the more than the new           Phase 1           0         0         0           0         0         0         0           10         215         225         220         215           All Margin etce         Phase 2</td><td>0 2 46</td></th<></thdistrict<></thdistrict<>	Instant         Margin Learning           Data         3664 obs. of 6 variables           Values         1           11         int [1:457] 1 2 3 4 5 6 7 8 9 1           12         int [1:3207] 458 459 460 461 46           Verify the more than the new           Phase 1           0         0         0           0         0         0         0           10         215         225         220         215           All Margin etce         Phase 2	0 2 46
Differ         Differ <thdifer< th=""> <thdifer< th="">         Differ</thdifer<></thdifer<>	Marrier         Marrier           Data         3664 obs. of 6 variables           Values         1           12         int [1:457] 1 2 3 4 5 6 7 8 9 1           12         int [1:3207] 458 459 460 461 46           Vertice         Present more           0         0         0         0           10         10         215         228         225         20         23         23           11         10         215         228         225         20         23         23         24           12         11         11/12/2007] 458 459 460 461 46         10	θ 2 46
Notes         Object         Object </td <td><math display="block"> \frac{1}{1000} = \frac{1}{10000} + \frac{1}{10000000000000000000000000000000000</math></td> <td>θ 2 46</td>	$ \frac{1}{1000} = \frac{1}{10000} + \frac{1}{10000000000000000000000000000000000$	θ 2 46
Notes         Observed         Description           10         Set Over Data         Description         Description           10         Set Over Data         Description         Description           11         Description         Description         Description           120         Description         Description         Description           121         Description         Description         Description           1210         Description         Description         Description           1211         Description         Description         Description           1212         Description         Description         Description           1213         Description         Description         Description           1226         Description         Description         Description           1230         Description         Description         Description           1231         Description         Description         Description           1232         Description         Description         Description           1233         Description         Description         Description           1234         Description         Description         Description           1242	$ \frac{1}{2} = \frac{1}{2} \frac$	θ θ 2 46       
Bit Market         Distance	$\frac{1}{1}$	θ 2 46 3 mm 2 46 3 mm 2 46 3 mm 2 46 3 mm



Let us do the stem plots and again, we are going to do two stem plots. First stem plot is for identity one, you can see. So, the decimal point is one digit to the left of the pipe symbol. So, it is 20.8 and 21.0, 21.2, etc, 21.6, and 22.0, 22.0, 22.0. So, these are the data points and you can see that it has a tail and then it slowly peaks and the peak is here. There are 275 data with this 24.2 number and that is why the average falls somewhere about 24.2 because the rest of the numbers are small compared to this value and the data goes like this and then it just speaks right.

So, it has a long tail on one side, but there is nothing on this side there is no tail at all I mean this is a peak and from the peak on 1 side, you see that the data has a tail. So we can now do this for 2 stem plot because it is much more skewed. So, you see these data points 11, 12, 13, etc and then you see that there are 52 more data points 616 more data points, thousand 924 more data points.

So, this is also the peak and from there, it just goes this way. So stem and leaf plots are nice to know about the structure of the data how it looks like, which will become apparent when we do the histogram plot, which we are going to do.

(Refer Slide Time: 13:14)



So, before that we are going to do the dot chart. Let us do the dot chart. So, we had x. Notice that. So, let us do the dot chart, notice that because we previously said that you have to do these two plots, two rows of plots, it continues.

If you want to change it, you have to give again, give the command to make sure that it starts plotting single, but in this case, we want to see the two plots for phase 1 and phase two. So let us continue. So, this is the second chart, so you can see that this is the same as the two plots we made previously and so dot chart again gives you this information.

### (Refer Slide Time: 14:32)

2				
tend Ober - 1 Sein dam - 1' Stampio - 2'				
(2 mas 1) 12 mas 12 We can also get a single dot chart:				
<pre>upen_ u par(mfrow=c(1,1))</pre>				
<pre>dotchart(X[,6],col*factor(X[,1]))</pre>				
6	, °			
	5			
	6			
				_
~				
aller Billiole * Piersen held false false laste and	**			V + 5
nensi ∰ stadu v De Nort C De Der Syne Des Sesses Juli Delle Delle Delle Delle Delle @ @ La Colomana i Male i	11			145
Mine: ● Klade + Pre Mont C. DR Dat your Dat Sealer gull gelag (tribs Jost pay @ @	Therement Holey C	1		-
Milet         ■ Kildela *         The Way (K)           DR         Dries         Sealers         Stational           DR         Dries         Sealers         Stational           DR         Dries         Sealers         Stational           DR         Dries         Sealers         Stational           DR         Delit         Stational         Stational           T         00000111111223333334444455555667778999999         18         B           18         00111112222333333444445555556666666677777777788888888999         19	and for the second state of the second state o	entertana C		354
Miles         ■ Elades         The West (4)           285         Date         System         System <td>11 Married Many 0 (* 1.) Organization (* 1.</td> <td>annetten ≠ 3664 obs. of</td> <td>6 variables</td> <td>3140</td>	11 Married Many 0 (* 1.) Organization (* 1.	annetten ≠ 3664 obs. of	6 variables	3140
Milet         ■ Exades*         Processor         P	9 Values	anetten € 3664 obs. of	6 variables	2.41 2.54 2.54
Noise         Billadas         Distance         Other State         Other	Image: second	3664 obs. of ( int [1:457] 1	6 variables 2 3 4 5 6 7 8 9 1	10
Note:         ■ Exadors         The Wey 14           101         Date:	<ul> <li>Brease Makes 0</li> <li>Bata</li> <li>X</li> <li>Values</li> <li>11</li> <li>12</li> </ul>	3664 obs. of int [1:457] 1 int [1:3207]	6 variables 2 3 4 5 6 7 8 9 1 458 459 460 461 46	10 52 463
Note         Distance         Distance <thdistance< th="">         Distance         <thd< td=""><td>Image: State of the s</td><td><pre>3664 obs. of int [1:457] 1 int [1:3207] / mm max</pre></td><td>6 variables 2 3 4 5 6 7 8 9 1 458 459 460 461 46</td><td>   </td></thd<></thdistance<>	Image: State of the s	<pre>3664 obs. of int [1:457] 1 int [1:3207] / mm max</pre>	6 variables 2 3 4 5 6 7 8 9 1 458 459 460 461 46	   
Mile         Bill Date         Date         Diff	Deressen Many D     D	<pre>// 3664 obs. of // int [1:457] 1 int [1:3207] // mme memory </pre>	6 variables 2 3 4 5 6 7 8 9 1 458 459 460 461 46	у н 8 в нал с н с н с н с н с н с н с н с н с н с н
Mint         Distance         Distance         Other Section	Deresse Many D     Deta     San times     Deta     San times     Deta     S     Deta     S     S     Deta     S     S     S     S     S     S     S     S     S     S     S	<pre>// 3664 obs. of // int [1:457] 1 int [1:3207] // mm mem // 0 // </pre>	6 variables 2 3 4 5 6 7 8 9 1 458 459 460 461 46	8 mar 8 mar 9 mar 10 2 463
Open Control         Open Control<	Image: Second	3664 obs. of f int [1:457] 1 int [1:3207] 4	6 variables 2 3 4 5 6 7 8 9 1 458 459 460 461 46	8 min 8 min 1 min 2 463 9 min 9
Open (1)	Normality         Normality <thnormality< th="">         Normality         <thnormality< th="">         Normality         <thnormality< th=""> <thnormality< th=""> <thnor< td=""><td>3664 obs. of f int [1:457] 1 int [1:3207]</td><td>6 variables 2 3 4 5 6 7 8 9 1 458 459 460 461 46</td><td>10 2 463</td></thnor<></thnormality<></thnormality<></thnormality<></thnormality<>	3664 obs. of f int [1:457] 1 int [1:3207]	6 variables 2 3 4 5 6 7 8 9 1 458 459 460 461 46	10 2 463
Open (************************************	Normality         Normality <thnormality< th="">         Normality         <thnormality< th="">         Normality         <thn< td=""><td>3664 obs. of 0 int [1:457] 1 int [1:3207] 0</td><td>6 variables 2 3 4 5 6 7 8 9 1 458 459 460 461 46</td><td>0 1 mar 1 mar 2 463 5 mar</td></thn<></thnormality<></thnormality<>	3664 obs. of 0 int [1:457] 1 int [1:3207] 0	6 variables 2 3 4 5 6 7 8 9 1 458 459 460 461 46	0 1 mar 1 mar 2 463 5 mar
Num         Num <td>Decension Many 0     To an and the second of the seco</td> <td>3664 obs. of int [1:457] 1 int [1:3207] . www.www</td> <td>6 variables 2 3 4 5 6 7 8 9 1 458 459 460 461 46</td> <td>0 1 mm 1 mm 2 463</td>	Decension Many 0     To an and the second of the seco	3664 obs. of int [1:457] 1 int [1:3207] . www.www	6 variables 2 3 4 5 6 7 8 9 1 458 459 460 461 46	0 1 mm 1 mm 2 463
Open (************************************	Notessel Naty 0	3664 obs. of ( int [1:457] 1 int [1:3207] .	6 variables 2 3 4 5 6 7 8 9 1 458 459 460 461 46	8 Parts 8 Parts 1 10 1 10
Num         District         District         Other	August Strange Control of Co	3664 obs. of f int [1:457] 1 int [1:3207] 4	6 variables 2 3 4 5 6 7 8 9 1 458 459 460 461 46	0
Open (*)	August Marge Company Comp	3664 obs. of int [1:457] 1 int [1:3207] .	6 variables 2 3 4 5 6 7 8 9 1 458 459 460 461 46	0
Description         Description           17         00000111111122333333444445555566666667777899999         18         00111112222333333444445555555666666677777777778888888999           19         00000000011222223333334444445555555566666667777777778888888999         19         000000000011122223333334444445555555566666667777777778888888999           20         00000111111111111111222222322223333334445555555566666667         777777888888999         21         000000000000000000000000000000000000	Determine Many of the second sec	3664 obs. of 0 int [1:457] 1 int [1:3207]	6 variables 2 3 4 5 6 7 8 9 1 458 459 460 461 46	8 mar

Let us make a single dot chart. So how do we do? We say, and then we do a chart of x and then we color. We color by the next factor X 1. So, we are going to say dot chart of the 6 column which is the sizes and color by factor of 1, which means phase 1 and phase 2 should be plotted with different colors and here is a plot.

Now you can see that dot chart automatically separates out the black points from red points previously, just our scatterplot actually overlap them, but dot chart automatically separate them out and puts them. So you do not have to make two different plots, just by calling dot chart with

factor. Now you can look at how the data looks like. So, which is very good, which is a nice way of looking at the data. Let us now move to the rank based properties.

(Refer Slide Time: 16:17)



So, we will do the cumulative distribution function of course. So, you can say ecdf, plot dot ecdf of X, 1, 6. So, this is the cumulative distribution function and you can see it just goes like that, because we know that there is a peak somewhere here and it has a tail, and this is true for two also. It is also but it is a much longer and fatter tail as compared to the other one right.

(Refer Slide Time: 16:48)



So, if you actually do this two rows then you will immediately see because see both of them go from 0 to 1, and in this case you have so both are skewed. Both have long tails and the tail is only on one side. That is what I mean by skew. But, but the tail is much longer here and because they are on the same scale, you can also see that this tail is much fatter than this.

So, so these information about these skewedness and how much information is there in detail, but peak at this value somewhere around this value is where they peak. But, their tails have different distributions, different characteristics. So, let us go back to single plot.

(Refer Slide Time: 17:38)



So, we have looked at the empirical cumulative distribution function and we will do. So, can we do that using ggplot? Let us do that also. So, we use the library ggplot and library scales. The reason why we are doing this is because we want to see whether the data will show any normal behavior. Obviously, it is not well to show because it is peaking at one end and it is highly skewed. Normal Distribution should have nice symmetric tails on either side, which is not the case. So, we are obviously not expecting, but if you see some kind of this kind of skew, what kind of plots do you get, for the probability scale

(Refer Slide Time: 18:56)



So, let us take a look at it, let us. So, I have plotted the both this on the same plot. So, usually ggplot he has to say data is x1 and you brought the grain size and that is ecdf and then you have to plot the second data. Again the ASTM grain size and you have to plot the empirical cumulative distribution function. So that is what we have done and they are on the same plot, and then what do we do?

### (Refer Slide Time: 19:36)

<pre></pre>	10 C 1 Q		Medule Descriptive statistics using 8		SARY D'S O
<pre>10 10 10 ASTM.gain.size In order to check if the gain size distributions are normal, we have to make the scales of the y-axes of the cumulative distributions greeobability scales. Let us do that: 11 &lt; gree(0.21(1)) 12 &lt; gree(0.21(1)) 13 &lt; gree(0.21(1)) 14 &lt; gree(0.21(1)) 15 &lt; gree(0.21(1)) 10 &lt; cold and find function "dotchat" 10 &lt; main find function "dotchat" 11 &lt; main find function "dotchat" 12 &lt; main find function "dotchat" 12 &lt; main find function (dotchat" 12 &lt; main find function (dotchat") 11 &lt; main find function (dotchat") 11 &lt; main find function (dotchat") 11 &lt; main find function (dotchat") 12 &lt; main find function (dotchat") 12 &lt; main find function (dotchat") 12 &lt; main find function (dotchat") 13 &lt; main find function (dotchat") 14 &lt; main find function (dotchat") 15 &lt; main find function (dotchat") 15 &lt; main find function (dotchat") 16 &lt; main find function (dotchat") 17 &lt; main find function (dotchat") 18 &lt; main find function (dotchat") 18 &lt; main find function (dotchat") 19 <!-- main find function (dotchat")<br-->19 <!-- main find</th--><th>M., w. 1</th><th></th><th></th><th>لتحميه</th><th></th></pre>	M., w. 1			لتحميه	
ASTM.grain.siz	Head the 1	0.00			
ASTM.grain.size ASTM.grain.size ASTM.grain.size Astm.grain.size Astm.grain.size Astm.grain.size)	Alfred		4	20	
In order to check if the grain size distributions are normal, we have to make the scales of the y-axes of the canualities distributions probability scales. Let us do that:	3.2 Hole		ASTM.grain.s	ize	
In order to check if the grain size distributions are normal, we have to make the scales of the y-axes of the comulative significance of the y-axes of the y-axes of the y-axes of the comulative significance of the y-axes of	Properti 13		•		
I bit correct (2,1)         I correct (		In order to check if the grain s cumulative distributions proba	ize distributions are normal, we ability scales. Let us do that:	have to make the scales of the y-axes of t	the
9		library(*ggplot2*) library(*scales*) 11 <- grep(1,X[,1]) 12 <- grep(2,X[,1]) X1 <- X[11,] X2 <- X[12,]			
Image: Solution of the second of the seco		par(mfrow=c(2,1))	9		
Nume       Nume       Nume       Nume       Nume         # Del per		-			
B) Die Vermin Weiner Beiter					
Image Service       Image Service       Image Service       Image Service         Image Service		b 🖻 🗉 🕄 🛛 😫	That have \$4, 1214		
<pre>trow is the second second</pre>	2. (m) (0) (0) (0) (0) (0) (0) (0) (0) (0) (0	Senie Bell Stag Roth Stati Brig	The New 14, 1514 \$31edia		Y 41 B 4
<pre>Fror in dotchat(X[11, 6]) : could not find function "dotchat"</pre>	Alarias Dels You Pas a lat pes You Pas a lat pes You Pas	Senan gold Ophie Both Taki Ipie	The layer 54, 1514 Kitada		7 4 8 4 0 2 here 1
<pre>• dotchart(X[13,6]) ************************************</pre>	a Tal Day Ann	peer juli peg pela juli juli juli	Malayi A, 1994 Abbala Statula	Nitry Constant	Y H St
<pre>• otchart(X[12,6]) • par(mfrow=c(1,1)) • otchart(X[13,6]) • plot.ecdf(X[i1,6]) • plot.ecdf(X[i2,6]) • plot.ecdf(X[i2,6]) • plot.ecdf(X[i2,6]) • plot.ecdf(X[i2,6]) • plot.ecdf(X[i2,6]) • plot.ecdf(X[i2,6]) • plot.ecdf(X[i1,6]) • plot.ecdf(X[i1,6]) • tit = grep(1,X[,1]) • tit = grep(1,X[,1]) • X1 &lt;- X[i1,] • X2 &lt;- X[i2,] g gplot(data=X1,aes(ASTM.grain.size)) • par(mfrow=c(2,1)) • par(mfrow=c(2,1)) • tit = grep(1,X[,1]) • X1 &lt;- X[i1,] • X2 &lt;- X[i2,] g gplot(data=X1,aes(ASTM.grain.size)) • par(mfrow=c(2,1)) • tit = grep(1,X[,1]) • X1 &lt;- X[i2,] • gaplot(data=X1,aes(ASTM.grain.size)) • par(mfrow=c(2,1)) • tit = grep(1,X[,1]) • tit = grep(1,X[,1]) • X1 &lt;- X[i2,] • gaplot(data=X1,aes(ASTM.grain.size)) • par(mfrow=c(2,1)) • tit = grep(1,X[,1]) • tit = grep(1,X[,</pre>	Annual Debugan	pain pair (pag port juk pe () Mar () and port juk pe () Mar () and () () () () () () () () () () () () ()	nalwey (1914 asole unction "dotchat"	may Comutana un thua. ∉ mar. 3664 obs. of 6 variablas	V H BS
<pre>&gt; par(mfrow=c(2,1)) &gt; dotchart(X[6],col=factor(X[,1])) &gt; plot.ecdf(X[i1,6]) &gt; plot.ecdf(X[i2,6]) &gt; plot.ecdf(X[i2,6]) &gt; plot.ecdf(X[i2,6]) &gt; plot.ecdf(X[i2,6]) &gt; plot.ecdf(X[i2,6]) &gt; plot.ecdf(X[i2,6]) &gt; plot.ecdf(X[i2,6]) &gt; plot.ecdf(X[i1,6]) &gt; plot.ecdf(X[i1,6]) &gt; plot.ecdf(X[i2,6]) &gt; plot.ecdf(X[i1,6]) &gt; library("sgalex") = library("sgale</pre>	Annual Description	<pre>peak peak peak peak peak peak peak peak</pre>	The law Y (1514 stude) unction "dotchat"	mmo commune enterna more discusse of 6 variables 457 obs. of 6 variables	2 4 2 3 2 bart 2 2 (10
obtic.edf(X[i1,6])       ii (it [1:457] 1 2 3 4 5 6 7 8 9 10         i2 int [1:457] 1 2 3 4 5 6 7 8 9 10       ii (it [1:457] 1 2 3 4 5 6 7 8 9 10         i2 int [1:3207] 458 459 460 461 462 461       iii (it [1:457] 1 2 3 4 5 6 7 8 9 10         i2 int [1:3207] 458 459 460 461 462 461       iii (it [1:457] 1 2 3 4 5 6 7 8 9 10         i2 int [1:3207] 458 459 460 461 462 461       iii (it [1:457] 1 2 3 4 5 6 7 8 9 10         polt.ecdf(X[i2,6])       iii (it [1:457] 1 2 3 4 5 6 7 8 9 10         polt.ecdf(X[i2,6])       iii (it [1:457] 1 2 3 4 5 6 7 8 9 10         polt.ecdf(X[i2,6])       iii (it [1:457] 1 2 3 4 5 6 7 8 9 10         polt.ecdf(X[i2,6])       iii (it [1:457] 1 2 3 4 5 6 7 8 9 10         polt.ecdf(X[i2,6])       iii (it [1:457] 1 2 3 4 5 6 7 8 9 10         polt.ecdf(X[i2,6])       iii (it [1:457] 1 2 3 4 5 6 7 8 9 10         par(mfrow=c(2,1))       iii (it [1:457] 1 2 3 4 5 6 7 8 9 10         iii (it [1:457] 1 2 3 4 5 6 7 8 9 10       iii (it [1:457] 1 2 3 4 5 6 7 8 9 10         iii (it [1:457] 1 2 3 4 5 6 7 8 9 10       iii (it [1:457] 1 2 3 4 5 6 7 8 9 10         par(mfrow=c(2,1))       iii (it [1:457] 1 2 3 4 5 6 7 8 9 10         par(mfrow=c(2,1))       iii (it [1:457] 1 2 3 4 5 6 7 8 9 10         par(mfrow=c(2,1))       iii (it [1:457] 1 2 3 4 5 6 7 8 9 10         par(mfrow=c(2,1))       iii (it [1:457] 1 2 3 4 5 6	Annu Control C	<pre>pair pair pair pair pair pair pair pair</pre>	The law 15 (1) 14 Ethods unction "dotchat" (X eX eX	Many Constant and Data - √ 3664 obs. of 6 variables 3707 obs. of 6 variables	V 4 S
plot.ccdf(X[i2,6])       it         ibrary("scales")       it         ilbrary("scales")       it         ilt       it         x1 << X[i1,1]	Control of the second sec	<pre>puis pair pair pair pair pair pair pair pair</pre>	Industry (1914 Elicity unction "dotchat"	Many Commune and Data - 3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables	(7.48) 8 had 9
<pre>prof(nfrow=(2,1)) par(nfrow=(2,1)) par(nfrow=(2,1))</pre>	Anne () Control () () Control () Control () () Co	<pre>peim pail peim peim peim peim peim peim peim peim</pre>	unction "dotchat"	Mony Commune and Data - 3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 6	2 4 6 10
<pre>plot.ecdf(X[i1,6]) plot.ecdf(X[i2,6]) plot.ecdf(X[i2,6]) plot.ecdf(X[i2,6]) plot.ecdf(X[i2,6]) plot.ecdf(X[i2,6]) plot.ecdf(X[i2,6]) plot.ecdf(X[i1,1]) ilibrary("spolt2") ilibrary("spolt2") ilitary("spolt2") ilitary("spolt2</pre>	Annu Control C	<pre>peak pair per per per per per per per per per pe</pre>	The law Y (12) 4 Resolution unction "dotchat"	Mony Commune and Dawa - ( 3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 f int [1:3207] 456 450 450	7 8 9 10
plot.ecdf(X[i2,6]) par(mFrow=c(1,1)) library("spolet2") library("scales") il <- grep(1,X[,1]) X1 <- X[i1,] X2 <- X[i2,] ggplot(data=X1,aes(ASTM.grain.size))+stat_ecdf()+ stat_ecdf(data=X2,aes(ASTM.grain.size)) par(mFrow=c(2,1)) a f	Compared and a contract of the contract on the contract o	<pre>peak paid paig paid paid paid paid paid paid paid paid</pre>	The law Y (12) if a strategy unction "dotchat" (X oX oX vAlues i1 i2 i2 i2 i1 i2 i2 i2 i1 i2 i2 i2 i2 i2 i2 i2 i2 i2 i2 i2 i2 i2	meny commune mentar	7 8 9 10 461 462 463
<pre>pr(mfrow=c(1,1)) Library("sgplot2") Library("scales") Library("scales") Litrary("scales") Litrary</pre>	Annual Control () Annual Contro	<pre>&gt;</pre>	The law Y (2) I strate unction "dotchat" (V) Values 11 12 12 12 12 12 12 12 12 12	Miney Construe motions- ℓ 3664 obs. of 6 variables 3207 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 1 int [1:3207] 458 459 460 Miney Mee money Mee construct ℓ ℓ	7 8 9 10 461 462 463.
<pre>library("ggplot2") library("scales") i1 &lt;- grep(1,X[,1]) x1 &lt;- X[1,] x2 &lt;- X[1,] ggplot(data=X2,aes(ASTM.grain.size))+stat_ecdf()+ par(mfrow=c(2,1)) i </pre>	Annual Control of the second s	<pre>pair put put put put put put put put put put</pre>	The term 14 (1514 Bit day unction "dotchat"	Misoy Commune anterna 3664 obs. of 6 variables 3207 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 1 int [1:3207] 458 459 460 Misager Miso Wear a Start 9 6	7 8 9 10 461 462 463.
<pre>library("scales") i1 &lt;- grep(1,X[,1]) i2 &lt;- grep(2,X[,1]) X1 &lt;- X[i1,] ygplot(data=X1,aes(ASTM.grain.size))+stat_ecdf()+ stat_ecdf(data=X2,aes(ASTM.grain.size)) par(mfrow=c(2,1)) i </pre>	Areas Control of the second se	<pre>pair pair pair pair pair pair pair pair</pre>	The term 1 ( 12) is the term 1 ( 12) is unction "dotchat" () Values 11 12 No. Non 16 16 12 16 16 16 16 16 16 16 16 16 16	Many Committee and Tables - 3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 1 int [1:3207] 458 459 460 Names New Wee m 2000+2 0 €	7 8 9 10 461 462 463.
<pre>i1 &lt;- grep(1,X[,1]) i2 &lt;- grep(2,X[,1]) X1 &lt;- X[i1,] ygplot(data=X1,aes(ASTM.grain.size))+stat_ecdf()+ stat_ecdf(data=X2,aes(ASTM.grain.size)) par(mfrow=c(2,1)) i </pre>	Annu Control C	<pre>public public publ</pre>	The tensor of (1914 Ethering) unction "dotchat"	Many Commune and Heart - 3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460 Anagen Men Menu a diment 0 €	r 4 5 I han 1 T 8 9 10 461 462 463.
i2 <- grep(2,X[,1]) X1 <- X[i1,] y2 <- X[12,] ggplot(data=X1,aes(ASTM.grain.size))+stat_ecdf()+ stat_ecdf(data=X2,aes(ASTM.grain.size)) par(nfrow=c(2,1)) i i	A second se	<pre>press paid para point just pet press paid para point just pet tt(X[i1, 6]) : could not find f ,6]) i,col=factor(X[,1])) i,col=factor(X[,1])) i,6]) 2,c6]) ,1)) ot2" ,es")</pre>	The law V (12) I Relation unction "dotchat" Values in value in value va	Many Commune mark mark 457 obs. of 6 variables 3207 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460 Manger May Mark m. Blant. 0 €	7 8 9 10 461 462 463.
<pre>x1 &lt;- X[i1,] x2 &lt;- X[i2,] ggplot(data=X1,aes(ASTM.grain.size))+stat_ecdf()+ stat_ecdf(data=X2,aes(ASTM.grain.size)) par(mfrow=c(2,1)) i i ATMpener AtMpener i AtMpene i AtMpener i AtMpener i AtMpener i AtMpener i AtMpener i AtM</pre>	Annue i annue	Image         Image         Image         Image           years         years         years         years           it(X[i1, 6])         years         years         years           j:col=factor(X[,1]))         i.6])         years         years           j:f(j)         years         years         years           j:f(j)         years         years         years           j:f(j)         years         years         years           j:f(j)         years         years         years	The law V(1)14 atoms unction "dotchat" (A ) (A ) (A ) (A ) (A ) (A ) (A ) (A )	Жину Сонитин натана- ∉ 3664 obs. of 6 variables 3207 obs. of 6 variables 3207 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 1 int [1:3207] 458 459 460 Минирание а диниг € ∉	7 8 9 10 461 462 463.
<pre>x2 &lt;- X[i2,] ggplot(data=X1,aes(ASTM.grain.size))+stat_ecdf()+ stat_ecdf(data=X2,aes(ASTM.grain.size)) par(mfrow=c(2,1)) </pre>	Arease in the second se	<pre>&gt;</pre>	The term V (12) a state unction "dotchat" (X) (X) (X) (X) (X) (X) (X) (X)	Nony commune anterna 3664 obs. of 6 variables 3207 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 1 int [1:3207] 458 459 460 Nonger Non Your m Shout 0 ¢	7 8 9 10 461 462 463.
<pre>ggplot(data=X1,aes(ASTM.grain.size))+stat_ecdf()+ stat_ecdf(data=X2,aes(ASTM.grain.size)) par(mfrow=c(2,1)) </pre>	Arrent Control of the second s	Seale         pedge pedge pedge pedge pedge pedge         Seale	The law V (12) a Unction "dotchat"	Notion         Communities           3664 obs. of 6 variables         3267 obs. of 6 variables           3207 obs. of 6 variables         3207 obs. of 6 variables           int [1:457] 1 2 3 4 5 6 1         int [1:3207] 458 459 460           Notagen New Weak         # 2000 0000	7 8 9 10 461 462 463.
stat_ecdf(data=X2,aes(ASTM.grain.size)) par(mfrow=c(2,1))	Anter a second s	<pre>paid peig peig peig peig peig peig peig peig</pre>	Unction "dotchat" With the second s	Warry Consumm           and Data-           3664 obs. of 6 variables           457 obs. of 6 variables           3207 obs. of 6 variables           int [1:457] 1 2 3 4 5 6 7           int [1:3207] 458 459 460           Moment Wey Wey           # Board P (	7 8 9 10 461 462 463.
par(wfrow=c(2,1)) ; i	And the second s	Image: Second	The law (1)(1) a train unction "dotchat" () () () () () () () () () ()	Many Commun and Dana. ∉ 457 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460 Manyan Many Many an Johant. 0 ∉	7 8 9 10 461 462 463. 5 mm
All light de	And the set of the set	<pre>peake paid pang park jusk pre it(X[i1, 6]) : could not find f (,6]) (,6]) (,6]) (,6]) (,6]) (,6]) (,6]) (,6]) (,6]) (,6]) (,6]) (,6]) (,6]) (,7)) (,6]) (,6]) (,7)) (,6]) (,7)) (,6]) (,7)) (,6]) (,7)) (,6]) (,7))</pre>	the law V(12)4 attain unction "dotchat"	Mony Construe montain 457 obs. of 6 variables 3207 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460 Mongar Mon Kee m 2 mart. 0 Ke	7 8 9 10 9 Mar 10 9 Mar 10 9 Mar 10 9 Mar 10 9 Mar 10 9 Mar 10
	And the set of the set	Section         Section <t< td=""><td>The law (4,25)4           Bitlaw           unction *dotchat*           0X           0X2           Values           11           12           Non           0%</td><td>Nony Commune and Dense 3664 obs. of 6 variables 3207 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 1 int [1:3207] 458 459 460 Nonger Non Yoor a Shout 0 ¢</td><td>7 8 9 10 461 462 463.</td></t<>	The law (4,25)4           Bitlaw           unction *dotchat*           0X           0X2           Values           11           12           Non           0%	Nony Commune and Dense 3664 obs. of 6 variables 3207 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 1 int [1:3207] 458 459 460 Nonger Non Yoor a Shout 0 ¢	7 8 9 10 461 462 463.

Obviously, we are going to change the scale the chain the scale these are not needed we already have this information. We also have this information. So, this is also not needed. So, we are going to make two plots,

# (Refer Slide Time: 20:04)

a series of the	B Module: Descriptive stat	father writing #	8	1116 V B B B
d				
Dealing F		9		
Alfrei Z				
8.2 Hills 17				
Properts				
	<pre>ggplot(data=X1,aes(ASTH.grain.size))+stat_e</pre>	cdf()+	(1997)	
	scale_y_continuous(trans-scales::pr	obability_trans( n	iorn //	
	## Warning: Transformation introduced infi	nite values in con	tinuous y-axis	
	ggplot(data=X2,aes(ASTM.grain.size))+stat_e	cdf()*		
	scale_y_continuous(trans=scales::pr	obability_trans("n	iorm"))	
	## Warning: Transformation introduced infi	nite values in con	tinuous v-axis	
	0.25+			
~	0.00-			
	0.20*			_
24 <b>- 9</b> 1				
eletter 🔹 Köladla +	nteres (	m .		148
Listen ( and the second s	s point phil phil phil phil phil			1 4 8- 0 1 had 3
alation a standar -	The same of the part of the pa	S Referent Hore, Const	-	1 H B) 2 Hard
Alertan Collector a par pole your Par a par pole your Par a part collector b plot.ecdf(X[	The server of th	الله المحمد	-	V 40 S B hand 1 C 40 C
<pre>bill color you put bill color you put pill color you put p</pre>	The served of th	11. Statement Many Grant Statement - ℓ Statement - ℓ A X		t e g
<pre>characterize in the second secon</pre>	Image: Section (1)         Image: Section (1)         Image: Section (1)         Image: Section (1)           Image: Section (1)         Image: Section (1)         Image: Section (1)         Image: Section (1)           Image: Section (1)         Image: Section (1)         Image: Section (1)         Image: Section (1)           Image: Section (1)         Image: Section (1)         Image: Section (1)         Image: Section (1)           Image: Section (1)         Image: Section (1)         Image: Section (1)         Image: Section (1)           Image: Section (1)         Image: Section (1)         Image: Section (1)         Image: Section (1)           Image: Section (1)         Image: Section (1)         Image: Section (1)         Image: Section (1)           Image: Section (1)         Image: Section (1)         Image: Section (1)         Image: Section (1)           Image: Section (1)         Image: Section (1)         Image: Section (1)         Image: Section (1)           Image: Section (1)         Image: Section (1)         Image: Section (1)         Image: Section (1)         Image: Section (1)           Image: Section (1)         Image: Section (1)         Image: Section (1)         Image: Section (1)         Image: Section (1)           Image: Section (1)         Image: Section (1)         Image: Section (1)         Image: Section (1)         Image: Section	۲۱۱. ۲۰۰۰ کی محمد الفتین (میرود) ۲۰۰۰ کی محمد الفتین (میروی) (میروی) (میرود) (میروی) (میروی) (میروی) (میروی) (میر	3664 obs. of 6 variables 457 obs. of 6 variables	a de B Brend C de C
<pre>Addition = Bit Code = Bit Code yee Paid a Bit Code yee Paid a Dist.ecdf(X[     par(mfrow=c(     plot.ecdf(X[     plot.ec</pre>	Image: Section (1)         Image:	31 32 ■ Волжник Валау Санал 32 ■ 32 = 32 = 32 = 4 34 ■ 52 = 32 = 12 = 2 4 ■ 54 = 32 = 12 = 2 4 ■ 54 = 12 = 12 = 12 = 12 = 12 = 12 = 12 = 1	3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables	t e g
<pre>control control c</pre>	Image: Second	S Parameter Mann, Comp S C Parameter Mann, Comp S The International	3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables	(† 4 8 18 mai - 1 16
<pre>bit control of the second second</pre>	Image: Second	<ul> <li>Mercure Many General State Construction of State</li></ul>	3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 7	7 8 9 10
<pre>bit control contr</pre>	Image: Section (Section (	Marrier Many, Const Subscriptions - C Subscriptions -	3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460	7 8 9 10 461 462 463.
<pre>plot.ecdf(X) plot.ecdf(X) plot.ecdf(X)</pre>	Image: Solution         Image: Sol	Descence Name Const Descence Name Const Descence Name Const Descence Name Const Name Const Name Const Descence Name Const Descence	3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460	1 4 8 4 2 mar 2 3 mar 2 4 5 10 4 6 1 4 6 2 4 6 3.
<pre>plot.ecdf(X) par(mfrowec) library("speet libra</pre>	Image: Solution         Image: Sol	S Research Marky General S Hard House A (1) S Hard House - ( S	3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460	* 44 84 ********************************
<pre>plot.ecdf(X[ par(mfrowcc) plot.ecdf(X[ par(mfrowcc) library("ggp it - grep() it - gre</pre>	Image: Second	Parameter Marky Connel     Parameter Marky Connel     Parameter Marky Connel     Parameter Market A     Market Market A	3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460 ver	1 4 8 1 mm 1 (101) 7 8 9 18 461 462 463.
<pre>&gt; III pay yee the plot.ecdf(X[ plot.ecdf(X[ plot.ecdf(X[ plot.ecdf(X[ plot.ecdf(X[ library("spe library(</pre>	Image: Second	Internet Water, Count         Internet Water, Count           Image: Count of the state of	3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460	* 4 8 8 bes 2 * 8 9 10 \$ 461 462 463. \$ 400
<pre>bill of the second second</pre>	Image: Second	Anticipal State Count State (Count Stat	3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460	7 8 9 10 461 462 463.
Discrete and a second	The second         The sec	Martine Many Count State States State States	3664 ob5. of 6 variables 457 ob5. of 6 variables 3207 ob5. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460	7 8 9 10 461 462 463.
<pre>plot.ecdf(X) plot.ecdf(X) plot.ecdf(X)</pre>	Image: Solution         Image: Sol	Martiner Many, Const Martiner Martiner Many, Const Martiner Martin	3664 ob5. of 6 variables 457 ob5. of 6 variables 3207 ob5. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460	8 9 10 461 462 463.
Image: Second	Image: Solution         Image: Sol	Martiner Many Genet     Martiner Many Genet     Martiner Many     Martiner	3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460	8 9 10 461 462 463.
<pre>&gt; ID out you find the second sec</pre>	Image: Second	Internet         Martin           Image: State State State         Image: State State           Image: State State         Image: State State           Image: State State         Image: State State           Image: State State         Image: State           Image: State State         Image: State           Image: State         Image: State           Imag	3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460	7 8 9 10 461 462 463.
<pre>&gt; It is a set of the set of</pre>	Image: Second	Annual Maria Count Count of the second market of the Count of the second market of the Count of the second market Count of the second	3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460	<ul> <li>4 8 9</li> <li>7 8 9 10</li> <li>7 8 9 10</li> <li>461 462 463.</li> </ul>
<pre>&gt; Display &amp; Display &amp;</pre>	Image:	A contrast many count of the contrast of the count of the contrast of the count of the count of the count of the count of the count of the count of the count of the of the count of the count of the count of the of the count of the count of the count of the of the count of the count of the count of the of the count of the count of the count of the of the count of the count of the count of the count of the of the count of the count of the count of the count of the of the count of the count of the count of the count of the of the count of the count of the count of the count of the of the count of the of the count of the co	3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460	7 8 9 10 461 462 463.
<pre>Place of the second secon</pre>	The second sec	A constant and con	3664 obs. of 6 variables 457 obs. of 6 variables 3207 obs. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460	7 8 9 10 461 462 463.
<pre>plot.ecdf(X) plot.ecdf(X) plot.ecdf(X)</pre>	Image: State State         Image: State<	A constraint Manage Constraint of Section 2010 Constraints of Cons	3664 ob5. of 6 variables 457 ob5. of 6 variables 3207 ob5. of 6 variables int [1:457] 1 2 3 4 5 6 7 int [1:3207] 458 459 460	* 44 8 * 1000 * 8 9 10 461 462 463. * 8000

Consta Reserved - John -	-	Connections	
The second se	The sectors	1. C	1000
<pre>&gt; plot.ecdf(X[12,6])</pre>	X	3664 obs of 6 variables	
<pre>par(mfrow=c(2,1))</pre>	V1	457 also of 6 unsighter	
<pre>plot.ecdf(X[i1,6])</pre>	evi Na	457 dbs. of 6 variables	
<pre>plot.ecdf(X[i2,6])</pre>	0X2	3207 ODS. OF 6 VARIABLES	
<pre>&gt; par(mfrow=c(1,1))</pre>	Values		
<pre>&gt; library("ggplot2")</pre>	i1	int [1:457] 1 2 3 4 5 6 7 8 9	) 10
<pre>library("scales")</pre>	i2	int [1:3207] 458 459 460 461	462 463_
<pre>i1 &lt;- grep(1,X[,1])</pre>	Fies Plats Pectager	Relp Verser	
i2 <- grep(2,X[,1])	0.1 (#3an 23) 12-	ur. 0 2	S Anto -
• X1 <- X[i1,]	10-		
> X2 <- X[i2,]	10		
ggplot(data=X1,aes(ASTM.grain.size))+stat_ecdf()+	100		
<pre>+ stat_ecdf(data=X2,aes(ASTM.grain.size))</pre>	a nexe		l.
<pre>par(mfrow=c(2,1))</pre>	10		
<pre>ggplot(data=X1,aes(ASTM.grain.size))+stat_ecdf()+</pre>	×		
<pre>scale y continuous(trans=scales::probability trans("nor</pre>			
r"))			
larning message:			
Transformation introduced infinite values in continuous v-axis			
	_		
	14.	à	÷.

And what are those plots going to be? We are going to do the probability distribution. The scale, we are going to change it to probability scale normal probability scale. Obviously, we do not expect it to be normal.

(Refer Slide Time: 20:29)

			V 4 845
al fai yes yes years held being built into into into	_		I Passi dan
neete foreine : jans :	g bernest the	y Connections	-
library("scales")	· G Sinter Brunnenger -	ee - 1	1.686+110
i1 <- greg(1,X[,1])	٥X	3664 obs. of 6 variables	
12 <- grep(2,X[,1])	•X1	457 obs. of 6 variables	
X1 <- X[i1.]	•X2	3207 obs. of 6 variables	
X2 <- X[12,]	Values		
<pre>ggplot(data=X1.aes(ASTM.grain.size))+stat ecdf()+</pre>	i1	int [1:457] 1 2 3 4 5 6 7 8 9 10	9
<pre>stat ecdf(data=X2.aes(ASTM.grain.size))</pre>	12	int [1:3207] 458 459 460 461 462	2 463
par(mfrow=c(2,1))	Files Plate Packet	per Melly Wester	-
<pre>ggplot(data=X1,aes(ASTM.grain.size))+stat ecdf()+</pre>	+ / 2m 2	tuar - 9 🖌	5 Anio - 1
<pre>scale v continuous(trans=scales::probability trans("nor</pre>			-
"))			1
arning message:			r'
ransformation introduced infinite values in continuous y-axis	22		
<pre>ggplot(data=X2,aes(ASTM.grain.size))+stat_ecdf()+ 1</pre>			
<pre>scale y continuous(trans=scales::probability_trans("nor</pre>			
"))		1	
arning message:		1	
ransformation introduced infinite values in continuous y-axis	, r	F	
	1		
	2	ACTM grain size	

So, we do not expect this curves to turn out to be a straight line. We are just confirming and so you can see, it is not clear here that these two figures are one on top of each other, but you will see it here.

(Refer Slide Time: 20:51)





So, in both the cases, so this is also not a straight line, sort of some amount of deviation. And here it is very clearly seen that it is not a straight line at all and you can see that this does not go up to it is only point 4 here and it is only point 25 here, the scale is not going up to 1, as you would have seen in other cases.

That is because the data is not symmetric about the mean. It is only 1 side that you have the tail. So, that is what is seen in this also. So we will come back and we will do more of the other analysis, histograms and box plots, etc for the same data set, thank you.