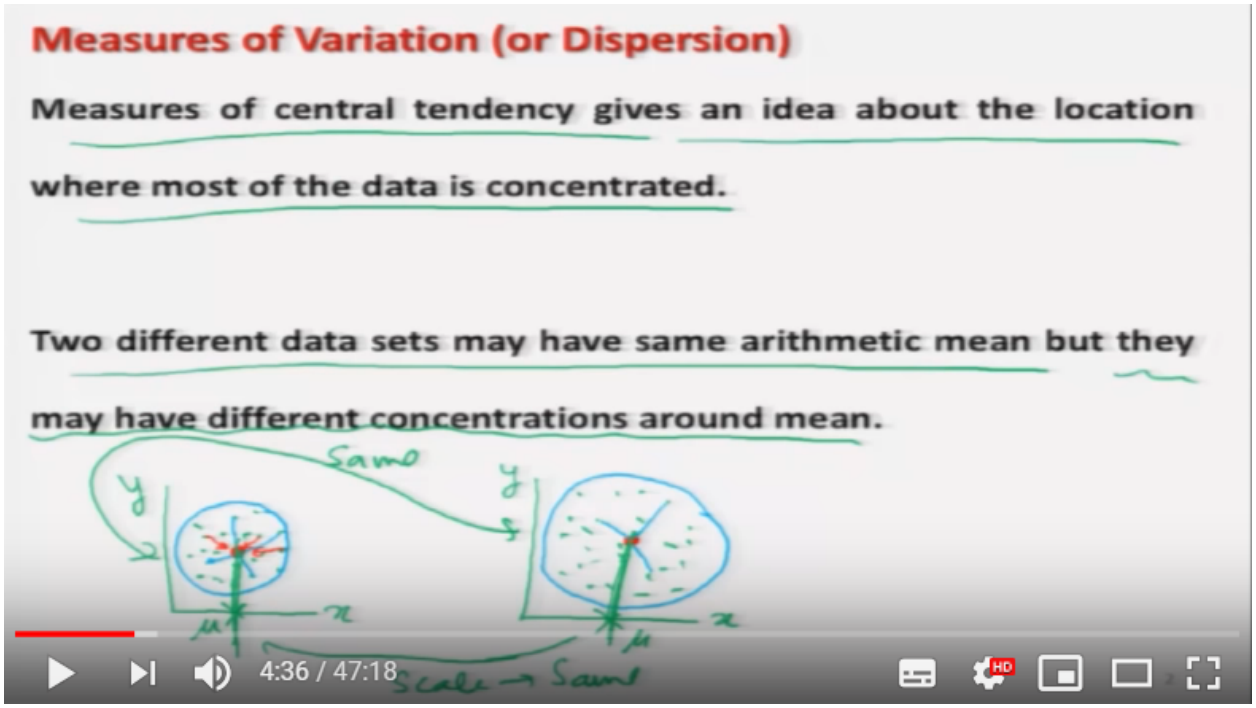
Lecture – 18
Variation in Data – Range, Interquartile Range and Quartile Deviation

Welcome to the lecture on the course, 'Descriptive Statistics with R Software.' Now, you may recall that, when we started the topics of descriptive statistics, we have taken several aspects. One option was the central tendency of the data, which we have discussed in the last couple of lectures. Now, we will aim to discuss the topic of variation in data. So, now the first question comes, what is this variation? Why it is important? How it is useful? What type of information it is going to give us and what are the different quantitative measures of such variations? So, in this lecture, you will try to develop the concept, need, requirement, of having the measures of variation. And, we will discuss three possible measures in this lecture; range, inter quartile range and quartile deviation.

So, let us start our discussion. You have seen that whenever we have the data. We simply want to dig out the information contained inside the data. And, as we had to discuss that, data itself cannot tell you that I

have these properties. So, in the last couple of lectures, we have concentrated on the central tendency of the data.

Refer Slide Time: (01:56)



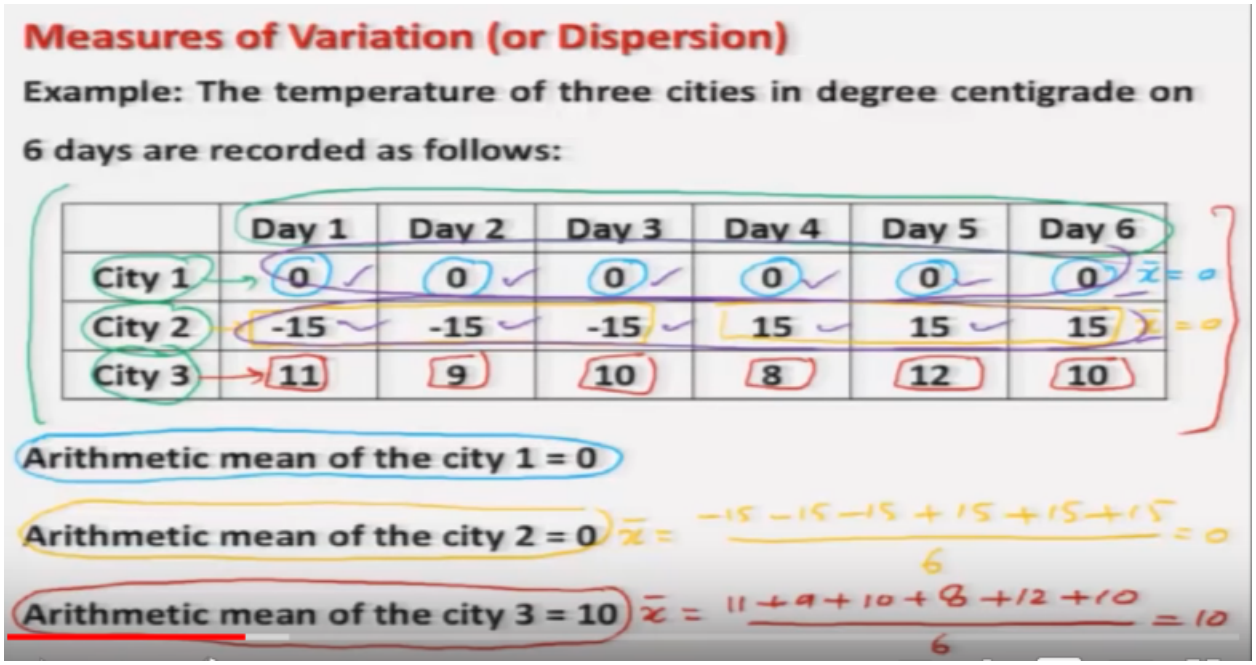
And, we have seen that those measures of central tendency gives us an idea about the location, where most of the data is concentrated. What does this mean? Means, if I have suppose this data, which I'm trying to plot here through a graphical measure. And, suppose if I say here like this is my x-axis, y-axis. So, I can see here that this data is concentrated somewhere here. So, this is trying to give us the information about, that where this where all this data is concentrated. But, there is another thing; you can see here that there is a deviation between the center point and those individual points. And, some points are close to the center and some points are away from the center. So, if I say here that suppose I have here, these two types of data sets, and suppose their scales are the same on y-axis and so, scales on the x and y axis they are the same. So, there is no issue.

Now, one data is here like this and another data here is like this. So, you can see here, in this case the most of the data is again concentrated over here. But, these deviations that means the difference between point and this center point or where is the mean is located, this is changing. And, you can see here, that in the first figure this region is of this type and then it's another figure this region is of bigger shape. So, what I can see here, that there can be two different data sets, which may have the same mean, for example, here in this case the mean is here and this of the mean is here. So, I'm assuming that this point

on the excess is suppose here is mu and here mu. Which is the generating the mean suppose. But, the spread of the values around the mean is different. And, similarly I can take any other point instead of a mean also.

So, now what we can see from these two figures that there can be two different data sets, which have got the same automatic mean. But, they may have different concentrations around mean. Now, the question is this, from the graphics I can show you that there are two deserts data sets in which the individual observations are scattered around the central point, in a different way. Graphically I can view it, but, now the question is how to quantify it? For example, I can take here a simple example to explain you, that what type of information is conveyed by the measures of variations. Suppose, there are three cities and we have measured our temperature, with the weather temperature in those cities on six days.

Refer Slide Time: (05:29)

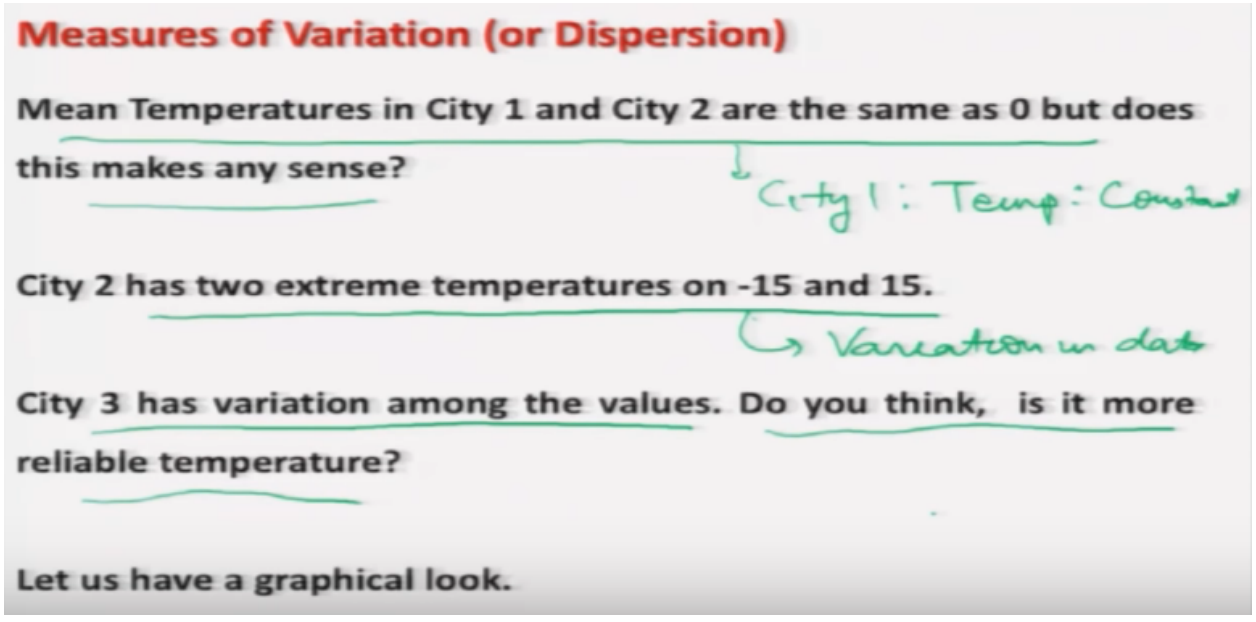


And, you can see here, those temperatures in degrees centigrade are recorded here in this table. So, now please try to have a look on the data that is given inside this table. So, here I am taking here three cities, city one, city two, city three. And, in the rows we have here different days, days one, two, three, four, five, six. So, if you try to observe in the city number here one. You can see here, this temperature here is zero, on that day one, temperature on that day two is zero, the temperature on the day three is zero and the same temperature continues for all the six days. So, now in this case if you try to find out the mean of these observations, mean of these temperatures this will come out to be zero. So, what we can see here, that the arithmetic mean of the temperatures of the city one is zero.

Now, similarly in case if you try to observe in the city number two. First three values that mean the temperature on first three days minus fifteen. And, the temperature in the day four, five and six, this is plus fifteen. Once again, in case if you try to find out here the average, this will come out to be some of minus fifteen, minus fifteen, minus fifteen, plus fifteen, plus fifteen, plus fifteen divided by six. And, this will again come out to be zero. So, in this case also this x bar is coming out to be zero. So, now you can see here, there are two cities; city one and city two. In which the automatic mean is coming out to be the same, zero, zero. But, in case if you try to look into the data, see here, here, here, here… Do you think that the data in the city one, which is all zero and the data in the city two are the same, the answer is no. They are different, but somehow their automatic mean is coming out to be zero.

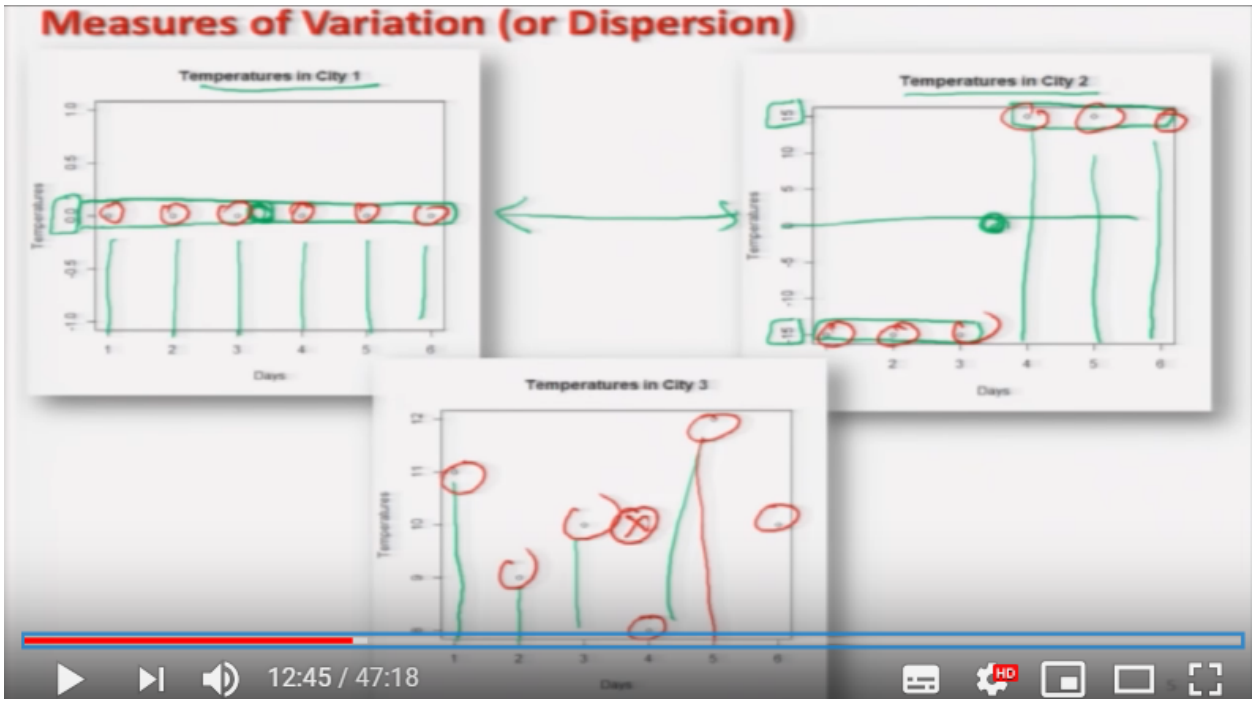
Now, in case if you try to observe in the data in the city number three. You can see here, on day one, city three has temperature eleven, day two-nine, day three-ten, day four-eight, day five-twelve, day six-ten. And, now in case if you try to find out the arithmetic mean of these temperatures, like eleven, plus nine, plus ten, plus eight, plus twelve, plus ten divided by six, this will come out to be ten. So, the arithmetic mean of the temperature in city three is coming out to be ten. So, now you can see here, in this case the temperatures are little bit different, then in compared to the city one and city two. So, you can see here, in these three cities, I have artificially taken three different types of data sets. And, I'm trying to find out their arithmetic mean. What you have to notice, that the arithmetic means in the city one and city two, they are the same. But, their data values are trying to show us a different information.

Refer Slide Time: (09:15)



Now, can I say that since the mean temperature in city one and city two are the same as zero? So, the pattern of the temperature in both the cities are the same. Why? Because, city two has an, has a peculiar characteristic that it has that temperature on two extreme, minus fifteen degree centigrade and fifteen degree centigrade. Whereas in the city one, the temperature is, city one the temperature is actually constant, it is always zero. Where is in the city two? Yes, there is some variation in the data. Now, in case if you try to look into the data of city three, this also has some variation in the data. But, now incase if you try to look at the temperature patterns of these three cities, what do you think? Can I say that the information provided by the city three temperatures, it's more reliable? No.

Slide Refer Time: (10:41)

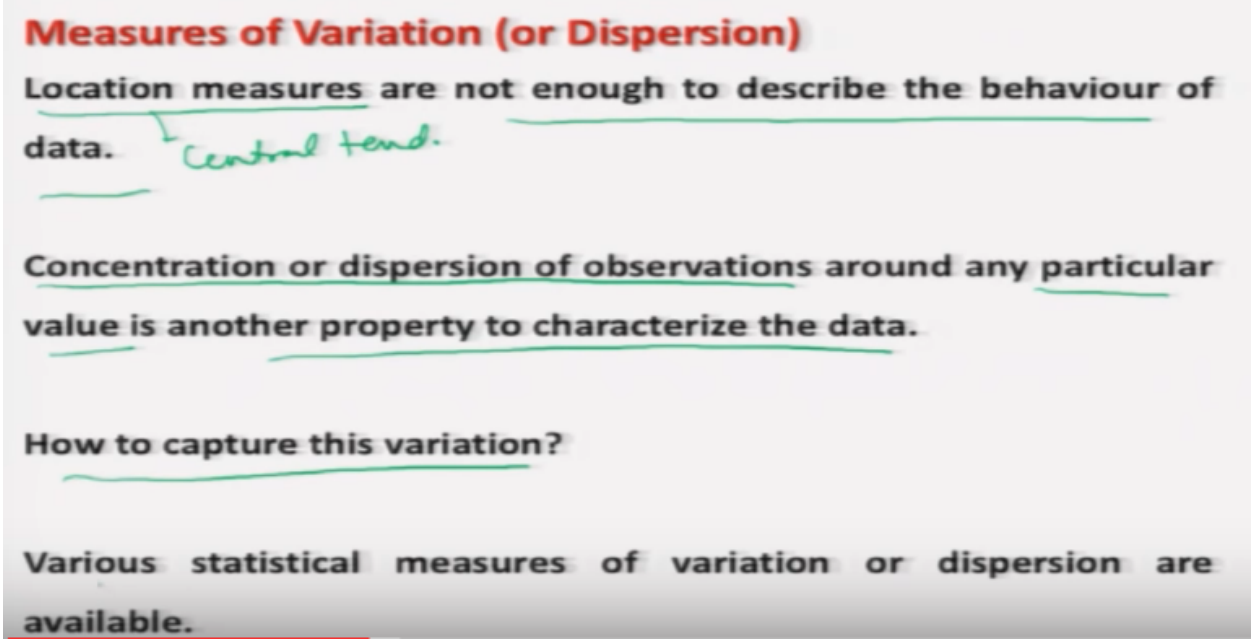


Let us have to see, whether this statement is right or wrong. So, first let me try to plot, this data on a simple graph. And, let us try to see what a type of information I am going to get. Well, that's a very simple plot and I will show you later on that how to plot this type of data in our software. But, if you try to see here in that city number one, this temperature is constant, zero means all these dots are denoting the temperature on, day one, day two, day three, day four, day five and day six, and it is here zero. Similarly, in the city two, you can see here, there are two values here minus fifteen and plus fifteen. And, there are three observations on days one, two and three, which are the same, minus 15. And, there are three temperatures on day four, day five and day six they are also plus fifteen and they are the same.

But, can you see the pattern in Figure one and figure two. And, you can see here that the average value will be somewhere here at about zero, in both the cases. Yeah! But, if you try to see such a diagram for

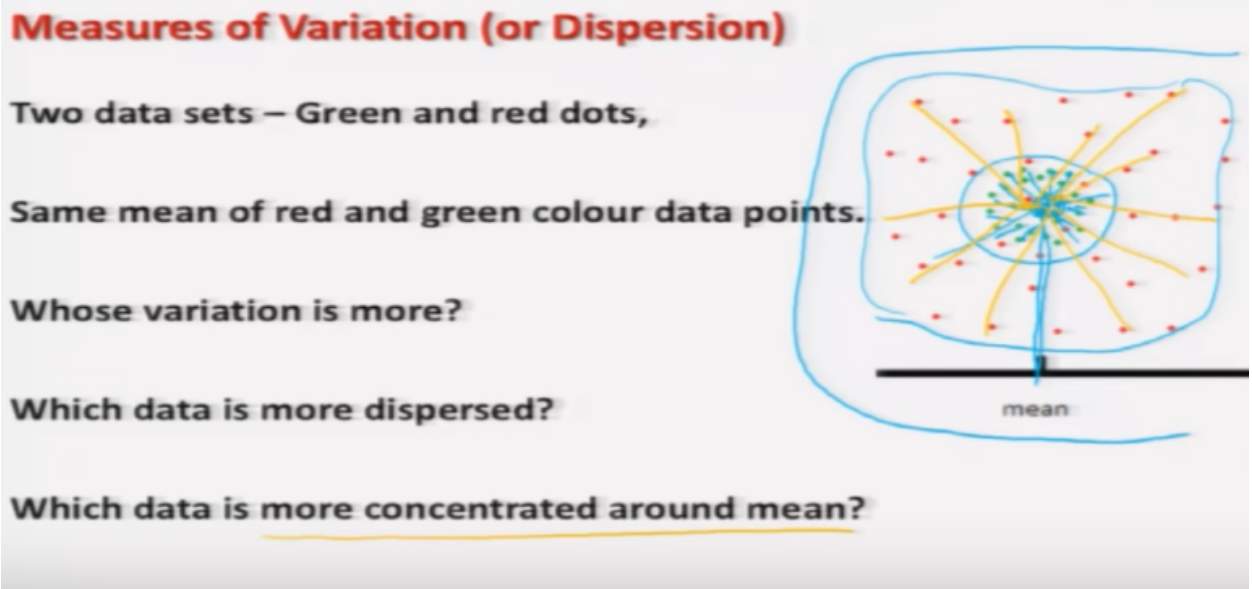
city three so on, day one this is trying to show that the temperature is eleven, day two it is nine, day three it is ten and so on. So, if you try to see here this pattern don't you see that, the points here are scattered in different places. For example, you can see here, this is the figure number one, figure number two and figure number three. Right? What this point should be here. Right? So, you can see here, and the mean value is somewhere here. So, now our objective is very simple, we have understood that the mean value is giving us some information and the variation in the values is giving us different type of information. So, from graphics I can see, but I would like to quantify it.

Refer Slide Time: (13:03)



So, now the next question is how to get it done? So, I can see here, that the location measures, location means the central tendency. I will say in simple language. The measures of central tendency are not enough to describe the behavior of the data. The, there is another aspect the concentration of data or the dispersion of data around any particular value. This is another characteristic of the data that we would like to study. And, now the question is, how to capture this variation? And, in order to capture this variation, various statistical measures of variation or dispersions are available.

Refer Slide Time: (13:47)

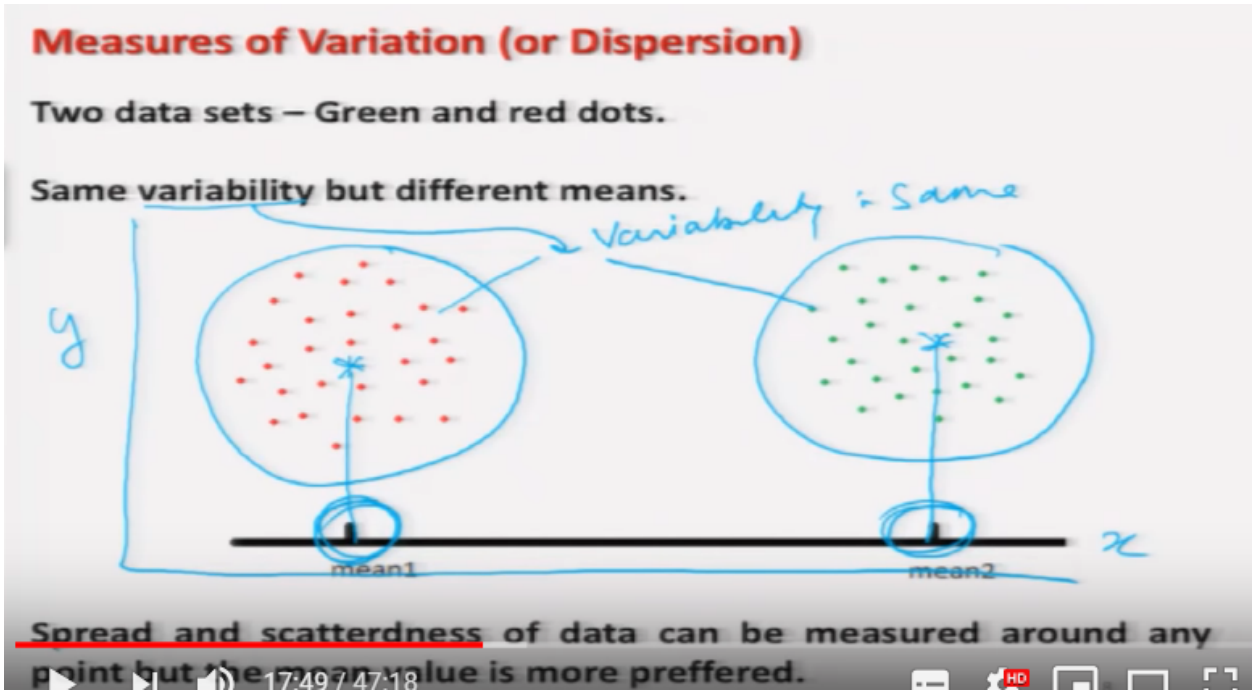


Now, we are going to say that, what is the means, how to use it, what is the interpretation and how to implement these tools on the R software? These are the three objectives in this lecture. Okay. So, now let we try to show you here a simple graph. So, here you can see here I have made here two types of dots in this picture here. One here is in green color like as here and another here is in red colors, which are concentrated somewhere around this. So, these are two data sets and I have simply plotted the scatter plot. You can see here, in both the cases the arithmetic mean is going to be somewhere here. But, you can see here, that the data set in green color, that is more close to the central value, which is here. And, whereas the data in the red color, that is more scattered from the center of the data. So, I can see here, that in the case of green dots the variation is only up to this point. And, whereas in case of, see here red dots this variation is going up to this point. So, this orange color pen and then this blue color pen, they are trying to give us the idea of the scatterdness of the data. So, I would try to now devise some tools, which can measure this scatterdness or this constant. That which data is more concentrated around the central mean or which data is more scattered around the mean or say. That mean is the general consideration otherwise I can measure it around any particular value also.

Now, I'm going to address one simple issue, before going further. Sometimes people do ask me that they have got two different data sets, and they have got the same variation, is it possible that they also have the same mean? So, I'm just trying to create here two different data set hyper means hypothetical graphics to show you the answer. The answer is this by looking at the variation of the data, you cannot really comment on the central tendency of the data and vice-versa. For example, in this in the last slide we have seen, that there are two data set, which have got the same mean, but they have got different variation.

Now, I will try to show you that, I have got two data set which have got the same variation, but they have got different means. Okay.

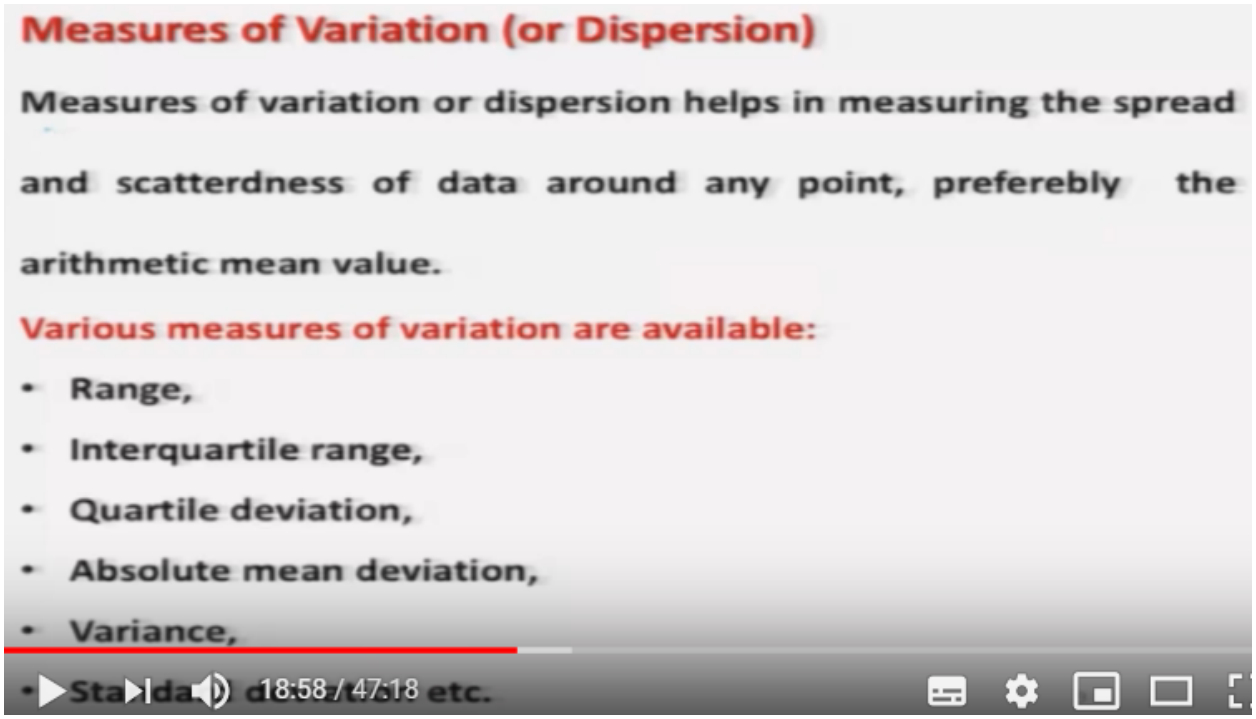
Refer Slide Time: (16:45)



Now, if you try to see here, I have to pick in here two data sets, one in if denoted by red dots and another by green dots. Well! I have prepared it by hand. So, essentially I have tried to make it very similar. So, you can see here, the mean in the data set one is somewhere here, and the mean in the data set is somewhere here. And, here is our x-axis and here is our y-axis. So, you can see here, that the variability in both the cases, this is the same; nearly the same means I have tried to make it as close as possible graphically. But, they have got different mean, the mean of first data set is here, calling as mean one and the data set two has a mean two. So, you can see here, even of the two the data sets have got the same variability, but they can have different means.

Usually, you will see that the spread or scatterdness or concentration that can be measured around any particular point. But, we will see that measuring this concentration or spread around the mean value. And, particularly the arithmetic mean this is more preferable. And, there is a statistical reason actually, that when we try to measure, some measures like is variance, standard deviation around the arithmetic mean, then they have got certain statistical advantages. Definitely, this is not really the platform or this is not really the course, where I can really explain you the advantages of using arithmetic mean. But, as I go further in the lectures, I will try to show you that, what is the most preferable location with respect to the given tool, around which we should measure the variability. Okay?

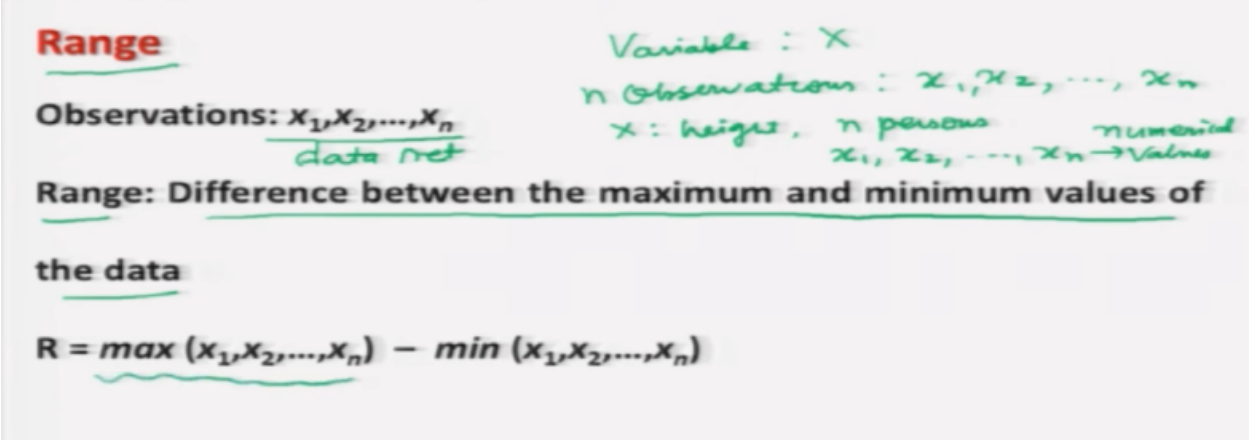
Refer Slide Time: (18:58)



So, now we have understood that different measures of variations or different measures of dispersions, they help in measuring the spread and scatterdness of the data around any point, and preferably the arithmetic mean value. Now, there are different types of measures which are available in statistics. There is a long list, but I will try to take here some of them in the given time, and those things with which I can show you easily on the R software. So, I would write, so there are different major like as range, interquartile range, quartile deviation, absolute mean deviation, variance, standard deviation and there are some graphical things also. And, there is a long list. So, I will try to consider here, these many measures in this lecture and the forthcoming lectures, and I will try to show you that how to measure them on the R software also.

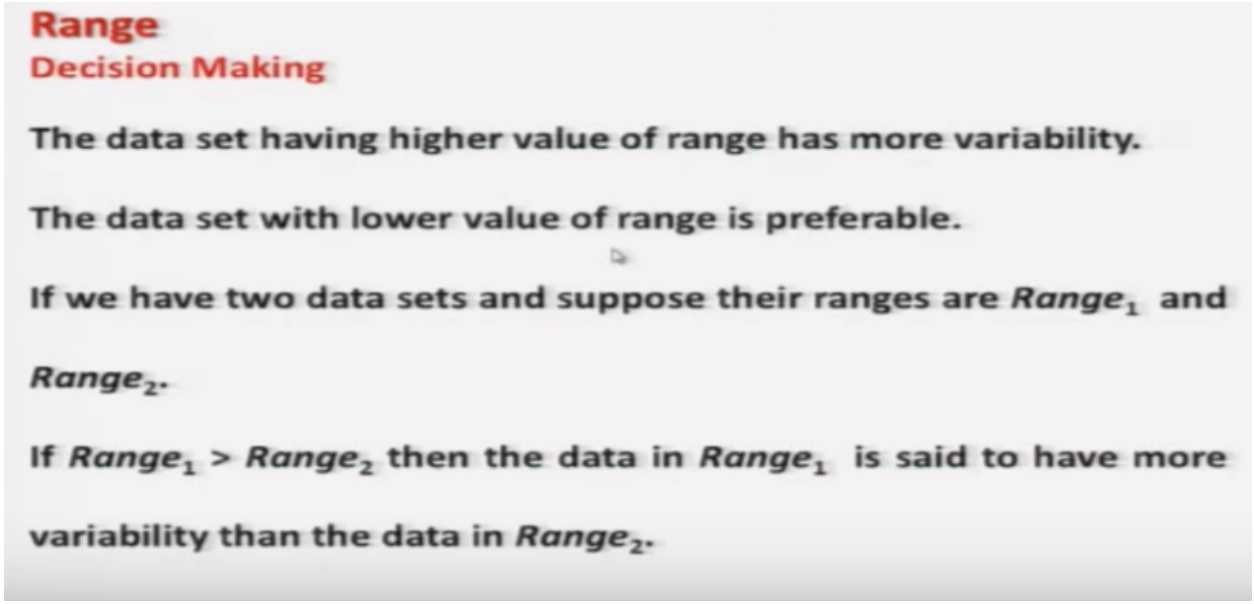
So, now what I will do, I will try to take these measures one by one. And, I will have several questions to, to answer depending on the tool. First thing will be what is the definition of the measure, second thing will be how to compute it on the R software, third thing will be how to interpret it? And, in some of the cases I will try to show you that in case if some missing values are present in the data set, then how to handle them, how to compute the tool in the presence of some missing observation. And, lastly wherever possible I will try to show you the tools for group and ungroup data. So, this idea or these things will continue during the entire topics of variation in data. Okay?

Refer Slide Time: (20:56)



So, now let us take the first topic here, which is here the range. So, first I will assume here and that will be valid forever all other lectures, that I have here a variable x. On which we are collecting the n observations. And, I am denoting it by small x1, is small x2 and says small xn. For example, in case if I say x is here height, then I am trying to collect the data on the heights of say here n persons, and the height of first person is denoted by small x1, height of second person is denoted by small x2 and the height of n it person is rooted by a small xn. So, this small x1, small x2, small xn they are going to be some numerical values, Right? So, I will now say in simple words that we have a set of observation x1, x2,….,xn which is our data set. And, our objective is this how to define the tools, and how to compute them using this data? The range is defined as a difference between the maximum and minimum values of the data. So, it is pretty simple. Just try to find out the maximum value out of x1,x2,….,xn that is the given data. Try to find out the minimum value from the given data set among x1, x2,., xn. And, just try to find out the difference between that two. And, this will give us the value of the range.

Refer Slide Time: (22:46)

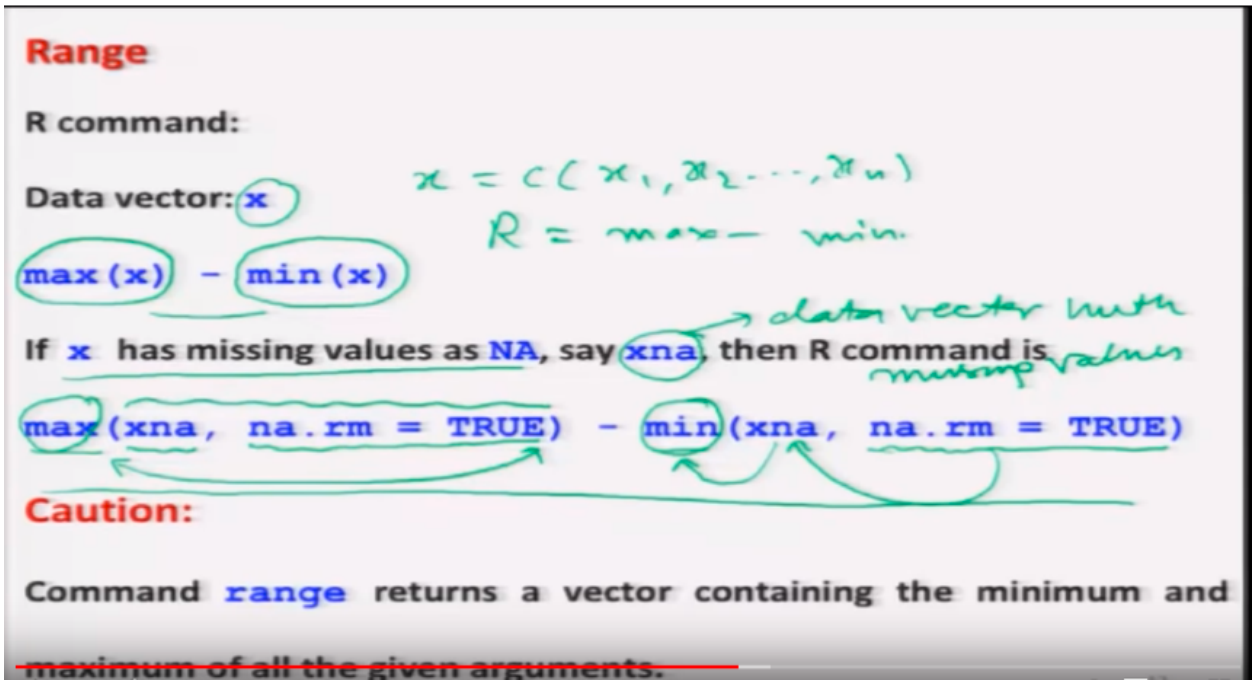


So, this is pretty simple actually. Now, the question is this once you get the range then how do you interpret it? So, the rule is pretty simple, the data set having higher value of range has more variability. So, I can say one thing that if you have got more than one data sets, and if you want to measure the variability in terms of range. Then, what we have to do? Just try to find out the ranges of all the data sets. Try to compare them and whose server range is coming out smallest, the corresponding data set will be thought to have a smaller variability. So, I can say nowhere that the data set which is having the lower value of range is preferable. And, in case if we have two data set and suppose their ranges are represented by range one and say here range two. Then if range one is greater than range two, then we say that the data set of range one is having more variability than the data in the data set of range two.

One thing I would like to make it clear. That we are going to discuss different types of tools range, interquartile deviation, quartile deviation, absolute mean deviation, variance, standard deviation and so on. So, whenever we are trying to measure the variability. Then, we are trying to make such decision only with respect to that measure. Now, if I say that suppose you have two data sets, and you try to find out the range of one data sets and say a standard deviation of say and in the data set. Now, if you try to compare the range of first data set and the variance or the standard deviation of the second data set, that may not be appropriate.

So, my advice is that whenever you want to compare the variability try to use the same tool, and then you try to make an inter comparison.

Refer Slide Time: (25:07)



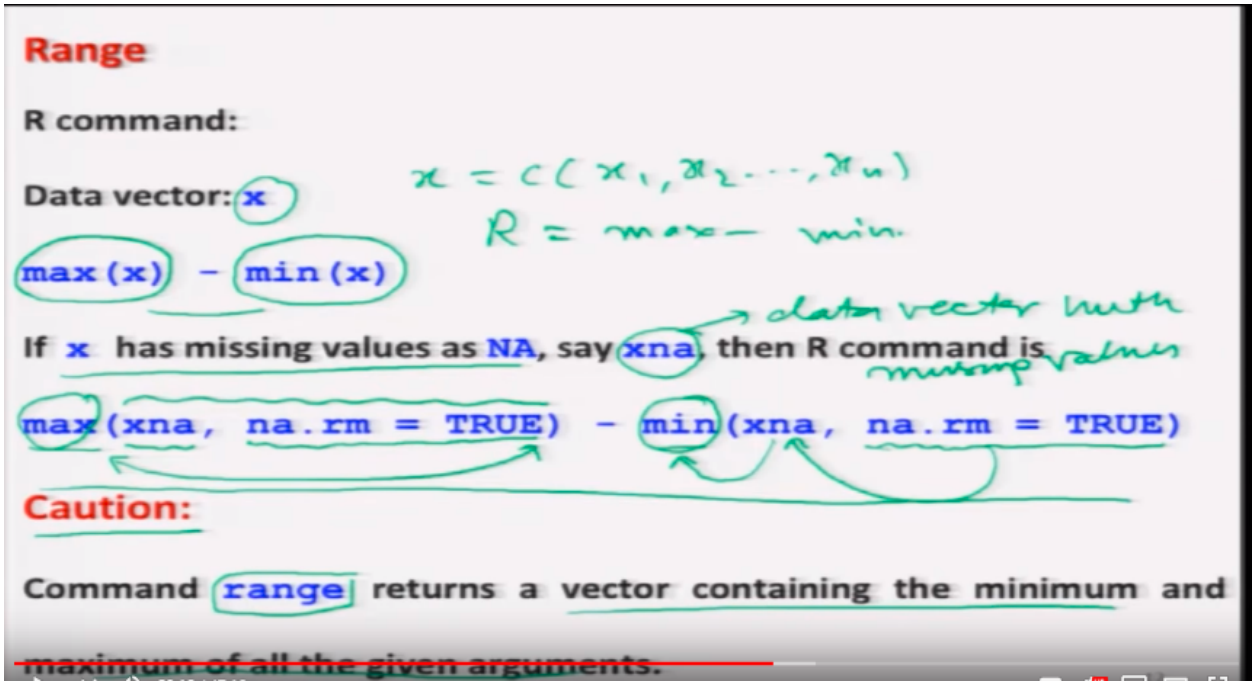
Now, I am coming on the aspect that how to compute the range in the R software. So, I will denote here by this x the data vector that whatever are the data values they are contained here like c ($x_1, x_2 \ldots\ldots x_n$). Now, you know as we have defined the range here the range has been defined as maximum value minus the minimum value. So, what I can do here, that I can use the built-in commands in the R software to find out the maximum value and the minimum value. That we had discussed in the earlier lectures, so the maximum value of x_1, x_2, x_n is going to be computed by max of (x), max and inside the argument you have to give the data vector and the minimum value of x_1, x_2, x_n is going to be computed by the command min and inside the argument you have to give the data and then you try to find out the difference between the two. So, max of (x) minus min of (x).

Now, suppose it happens that x has some missing values and they are denoted by capital N capital A. And, suppose I try to store this data into another data vector say xna, so xna is my another data vector which has got some missing values. So, in case if you want to compute the range, of such a data vector where the values are missing, you simply have to use the same command max and min, but inside the argument you have to give the data, in this format xna the data vector in which you have got some missing values and you have to use the command na dot rm is equal to logically TRUE. That is capital T, capital R, capital U, capital E. So, what will happen? That once you try to operate the max command on this vector, this data set, then it will try to remove the missing value and then it will try to compare how to compute the maximum value. And, similarly when we try to use the min minimum command on this data vector, then when we specify na dot rm is equal to TRUE, then this operator will remove the missing values from

this vector x and a which are denoted by capital N, capital A and then after that it will try to compute the minimum value. So, this is how you have to compute the range in case some data is missing.

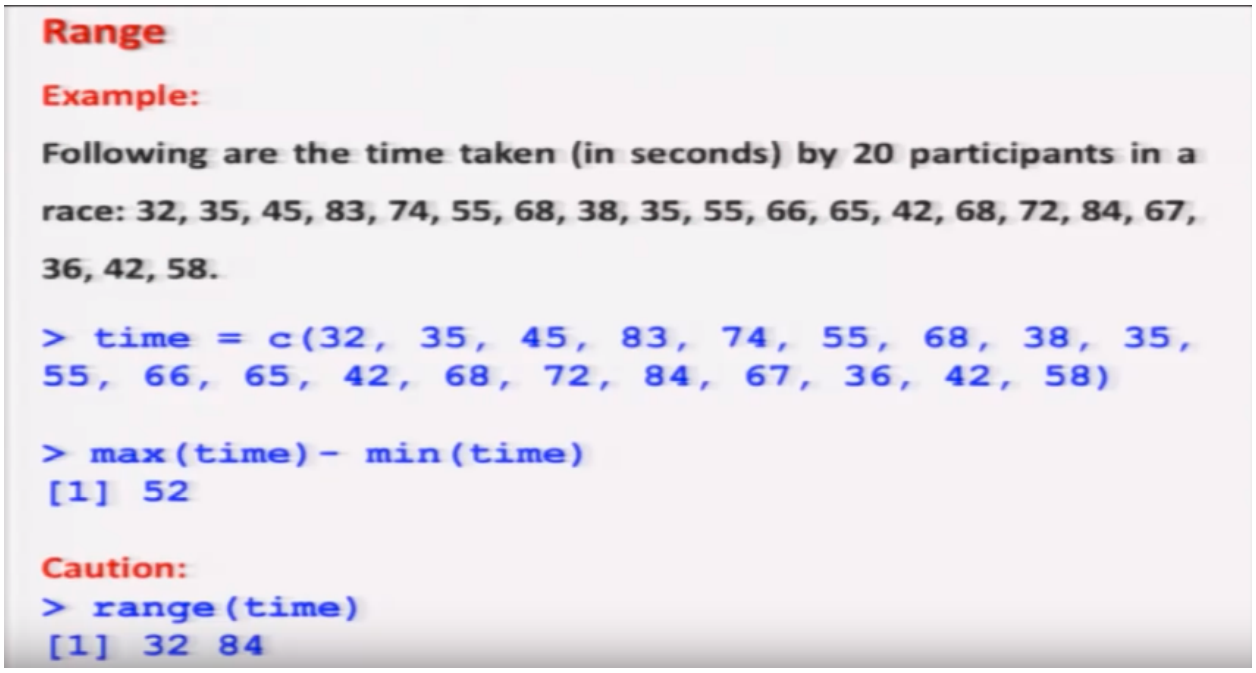
One thing which I would like to caution you here, as you have seen in the R software, we have the names of the function which are giving a value which is easily understood by the name, like as mean. Mean, means the arithmetic mean of the data vector median. Which is trying to give the median of the data vector. Similarly, when you try to use here the command range, 'range'. Then it appears that as if this is going to give me the value of the range, that is maximum value minus minimum value, but it does not happen. Range will try to give you the two values. The range command will give you two values, one is the maximum value of the data vector and another is the minimum value out of those data inside the data vector, so be careful.

Refer Slide Time: (29:04)



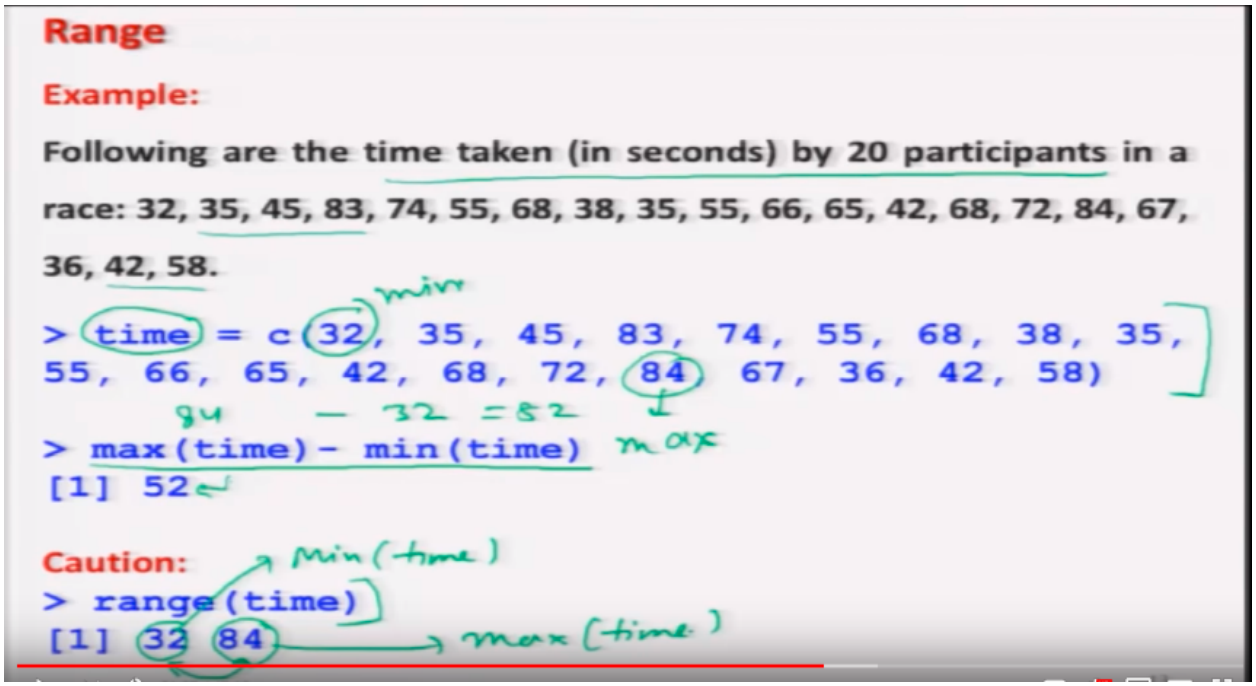
So, here I would like to make here a note of caution, that if you try to use the command range, then this will return a vector containing the minimum and maximum values of the given argument.

Refer Slide Time: (29:22)



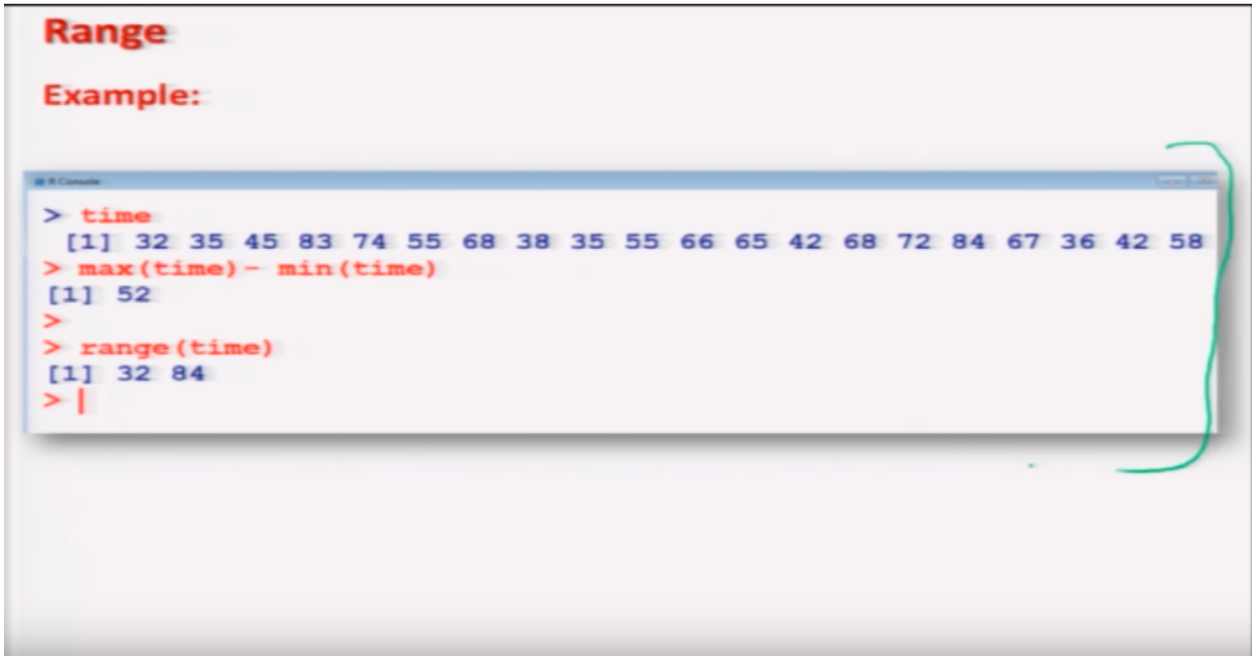
So, just be careful and if you recall the same thing happened in the case of mode also, mode that was trying to give some other information but by name it appears that as if this is going to give me the value corresponding to the maximum frequency. So, similar is the case with range, so you need to be careful. Now, after this I will try to take an example and I will try to show you that how to compute the range on the given set of data. So, I will take the same example which I have given or which I have used in the earlier lectures.

Refer Slide Time: (29:55)



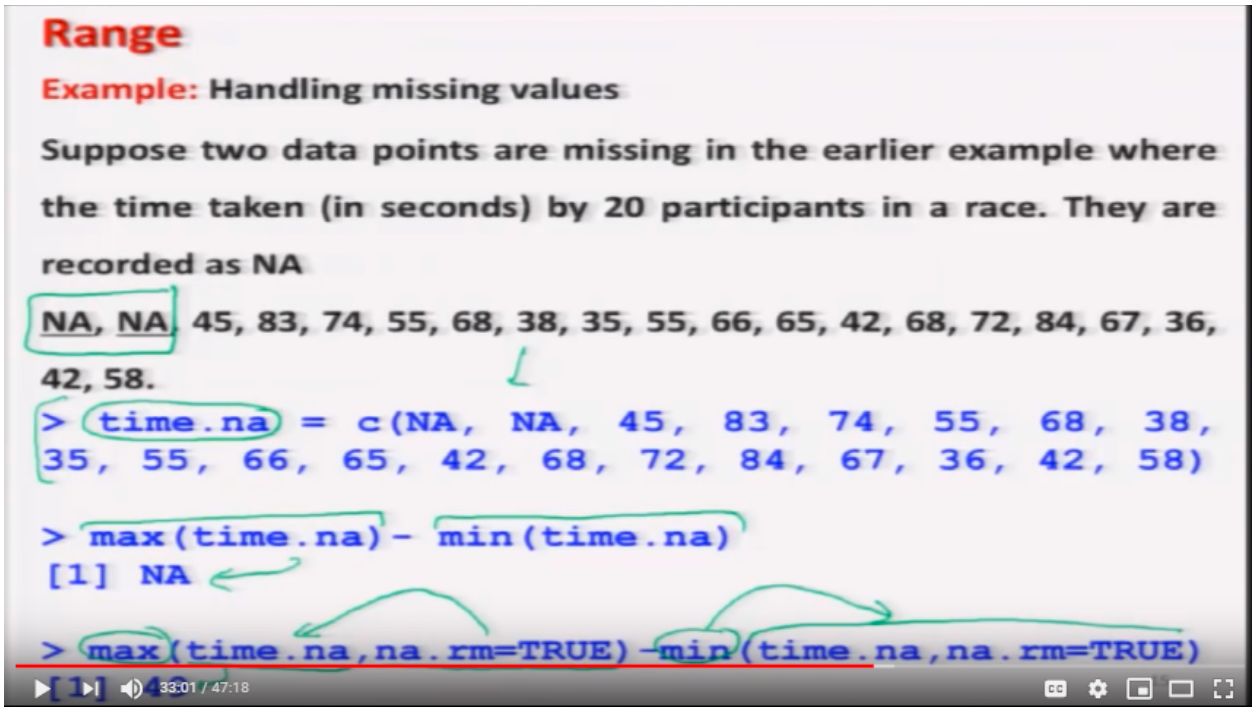
So, I have computed up or I have observed the time taken by 20 participants in a race and they are given in seconds over here like this, and this data is recorded inside a variable here time. So, this is my here the data vector and now I'm simply trying to execute the r command on maximum time minus minimum time, and I can see here that this is giving me the value here 52. And, you can also verify it from the given set of data, for example here you can see here this is here the maximum value and this is here the minimum value. So, if you try to subtract it here 84 minus 32 you get the value equal to 52. Just to show you what will be the command, or what will be the outcome of the command range, then you can see here this is giving me two different values. This is here the minimum value of time and this is 84 is the maximum value in the data given inside the time in theta vector, Right.

Refer Slide Time: (31:12)



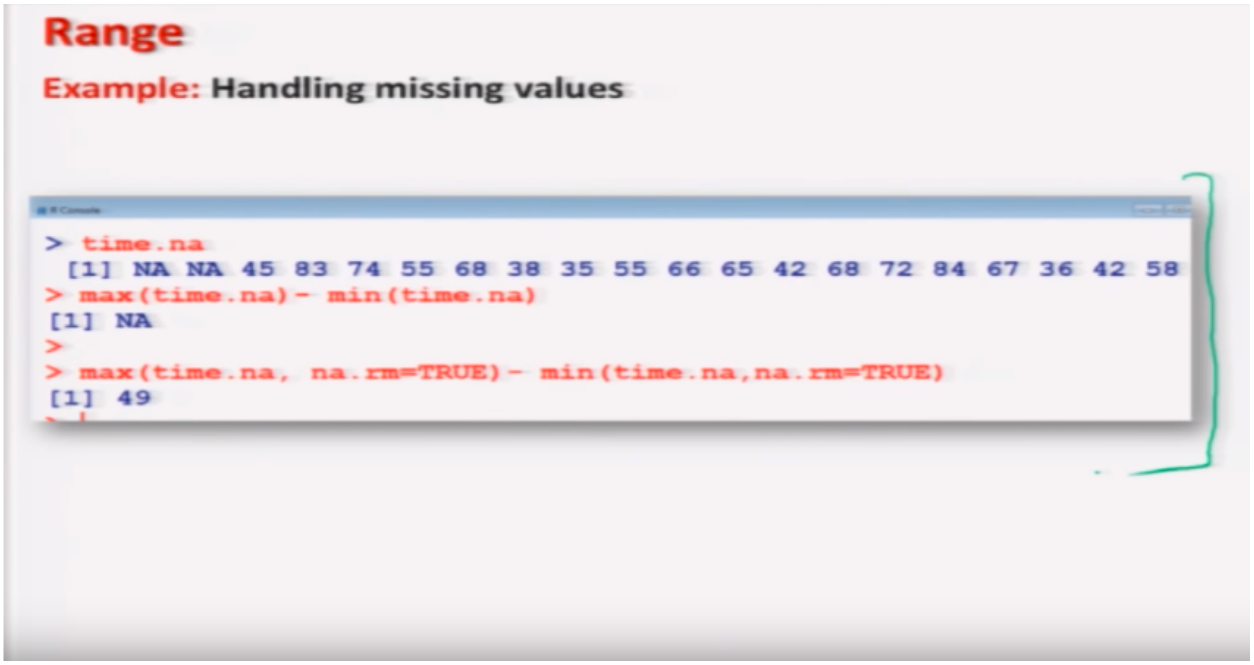
So, before I try to show you on the r console, let me try to give you one more example and then I will come back and here, in this slide you can see here the screen shot on the r console and now what I am trying to show you in the same example if that data values are missing then how you are going to handle it, so in the same example where we have recorded the time taken by twenty participants in a race.

Refer Slide Time: (31:22)



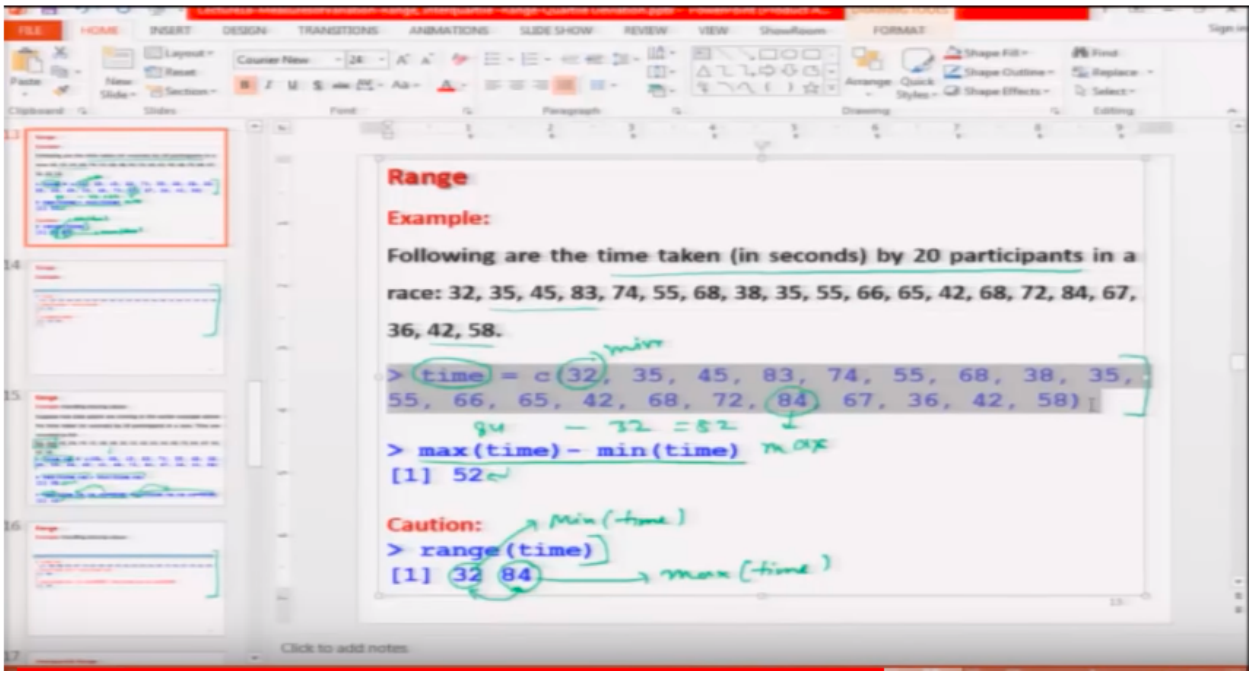
I have made this first two values to be a NA, that means they are not available, and all other values are the same. So, now I'm trying to record this data inside a new variable, this is time dot na. so, time dot na is simply indicating that, the data is not available inside the time vector, Right. And, there is no rule that the, that you always have to use here the point or dot, that is your choice. So, I have stored this data and now I'm trying to find out the maximum value of this time na, minus minimum value of the time na. But now if you see this will give me a output like na. Why? Because I have not used here that command na dot rm is equal to TRUE. So, I try to correct myself and I try to find out the maximum and minimum on the data set time dot na using the command that na dot rm is equal to TRUE. So, as soon as we give na dot rm is equal to TRUE, this will understand, the maximum command will understand that there are some missing values in the time dot na, which have to be removed before computing the maximum, and the same thing will happen here. This minimum command will understand that the first the any values or the missing value we have to be removed and this value comes out to be here 49. Right?

Refer Slide Time: (33:07)



So, now I will try to show you here on the r console and you can see here this is the data this is the screenshot of the operation on the r console.

Video Start Time: (33:22)



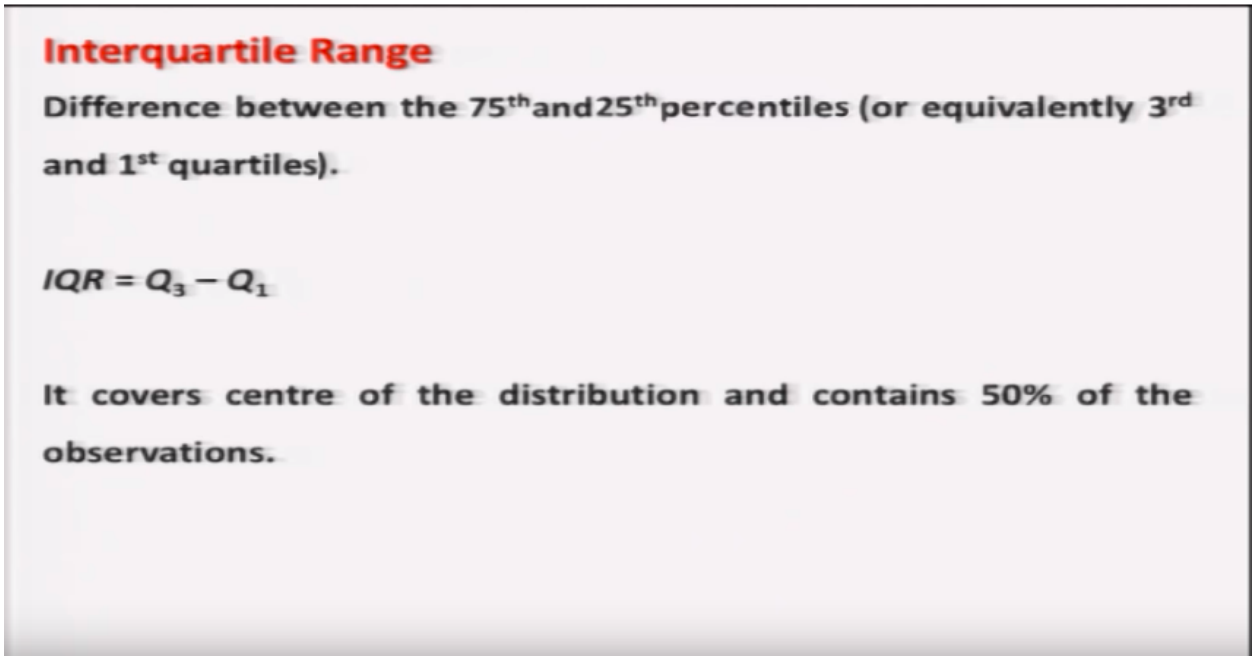
So, let me try to first copy, this data set, say her time and I try to put it here time. So, I can see if this is my here that data time. So, now suppose if you say by mistake if you try to operate the command range, you can see here this will come out to be 32, 84 that is the same outcome that we had received, Right. But if you try to find out the maximum of time minus minimum of time you can see here, you are getting the this value which is the value of the range, and similarly if you try to remove the essing data Then how to operate it? So, I try to create here say here another, the data vector time dot na, so you can see here I have created this data vector here time dot na in which the first two values are missing and suppose if I try to find out here the range of time dot na. This will give me there is something wrong, so I try to use here the command na dot rm is equal to TRUE. I mean now it will give me the minimum and maximum values.

But my objective is not to find the minimum and maximum value. I want to find out the maximum and minimum values and I want to find out their difference, so I will say hey time dot na, and when na dot rm is equal to true, minus the minimum of time dot na, na dot rm is equal to TRUE. Now, if you try to see here you will get here the when you 49. But in case if you by mistake if you do not give the command na dot rm is equal to TRUE, then you can see the outcome will be na.

Video End Time: (35:27)

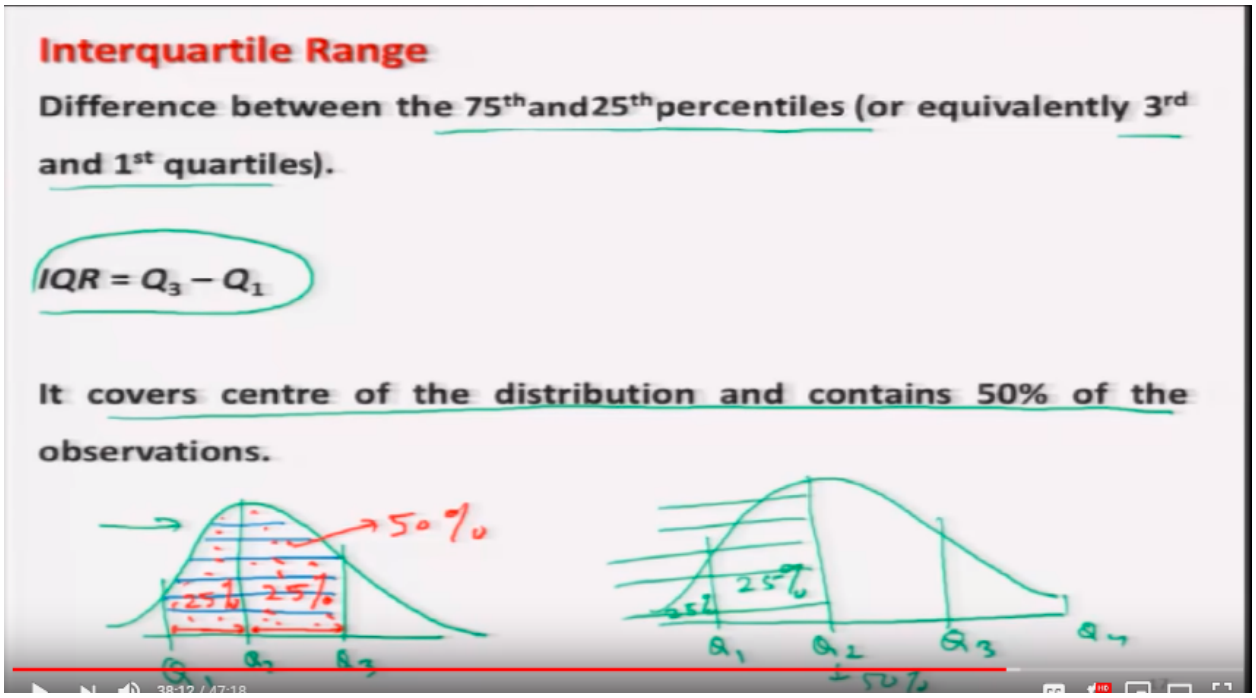
```
> time.na = c(NA, NA, 45, 83, 74, 55, 68, 38, 35, 55, 66,$
> time.na
 [1] NA NA 45 83 74 55 68 38 35 55 66 65 42 68 72 84 67 36
[19] 42 58
> range(time.na)
[1] NA NA
> range(time.na, na.rm=TRUE)
[1] 35 84
>
> max(time.na, na.rm=TRUE) - min(time.na, na.rm=TRUE)
[1] 49
> max(time.na) - min(time.na)
[1] NA
>
```

Refer Slide Time: (35:28)



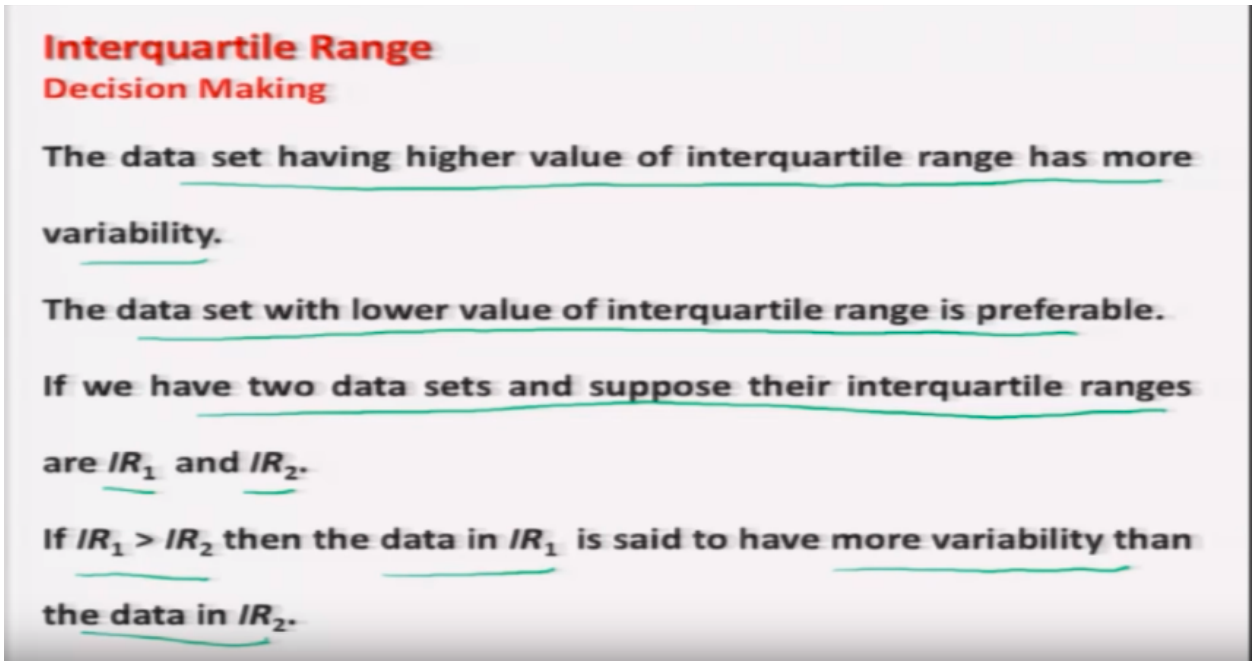
Now, after this I come on at this another topic, which the interquartile range. Just like range is trying to measure the difference between the maximum and minimum values. Similarly, if we have another measure what is called as interquartile range, and interquartile range simply tries to measure the difference between the third and first quartiles. Now, you may recall, what was the quartile?

Refer Slide Time: (36:00)



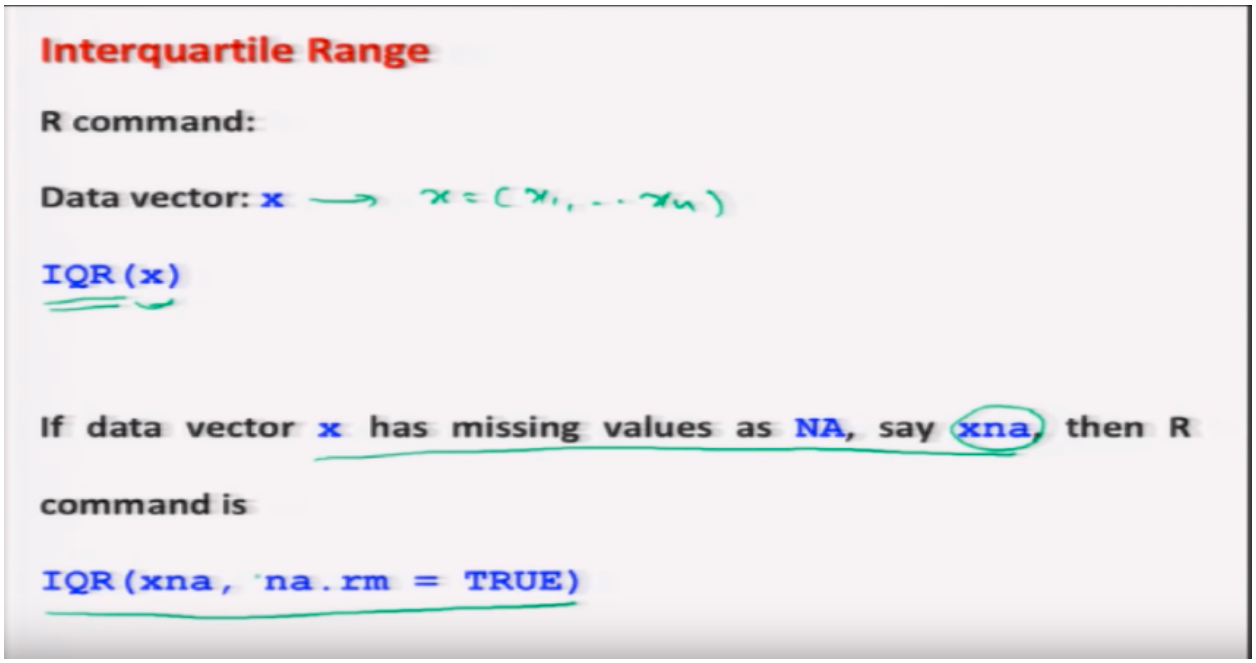
If you try to recall, then we had discussed that in case if we have the frequency distribution like this one, then this frequency distribution is divided into four equal parts and the first 25 percent of the frequency is covered in the first quartile to denote it as Q_1 , next 25 percent of the frequency is contained between Q_1 and Q_2 .So, essentially Q_2 is trying to consider the total fifty percentage of the frequency, so this is the median and similarly we had here Q_3 and finally here Q_4 . So, now what I try to do here, that I try to take here Q_1 ,Q_2 and Q_3 . Q_2 is the median, and I try to consider this area so you can see here that this area is consisting of 25 percent of the total frequency and this area is consisting of the another 25 percent of the frequency. So, the area between Q_1 and Q_2 is 25 percent of the total frequency and the area between Q_2 and Q_3 is 25 percentage of the total frequency. So, altogether if you try to add it together then this entire area, which I am denoting by here dots this is going to take care of the 50 percent of the total frequency. So, the interquartile range is defined as the difference between the 75^{th} and 25^{th} percentiles or equivalently, this is nothing but the third and first quartile, so this is denoted by here IQR is equal to Q_3 minus Q_1 ,Right. And, as I have shown you in this figure here, that this IQR or the interquartile range covers the center of the distribution and contains 50 percent of the observations.

Refer Slide Time: (38:16)



So, now how to make the decision making. Once again the rule is the same the data set having the higher value of inter quartile range has more variability that will be the interpretation. So, obviously if we would always like to have a data set which has got the smaller variability, so the data set with lower value of interquartile range is more variable. So, suppose if I have got two data sets and suppose their interquartile ranges are computed as IR_1 and IR_2 . So, if IR_1 is greater than IR_2 , then we say that the data in IR_1 is more variable or has more variability than the data in IR_2 . So, this is our interpretation, so now with the these two examples you can see here that range and interquartile range there both of them are trying to measure the same aspect of the data that is the variation, but they are doing in a different way, Right. Now how to compute it on the R software this is pretty simple. There are two ways that you can write your own program or just use the command to compute the quartiles and try to find out their difference or there is a built-in command in R software to find out the interquartile range.

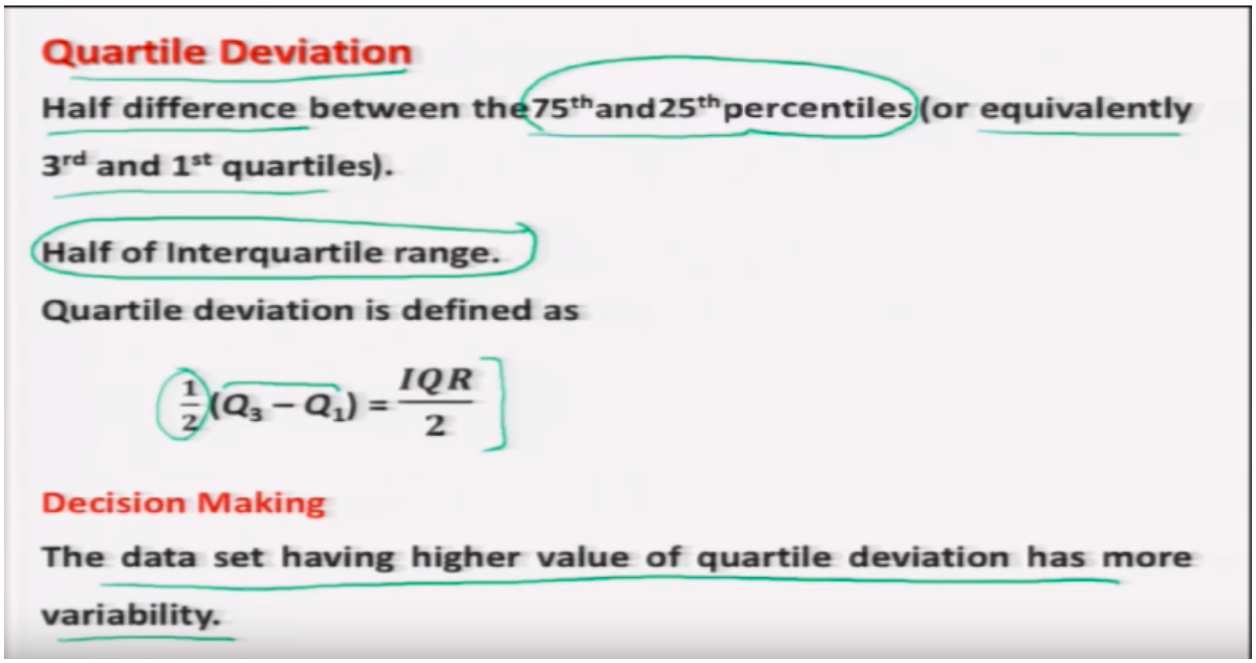
Refer Slide Time: (39:36)



So, if I say that my data vector here is x which is consisting of here say observation ($x_1, x_2, ..x_n$). What we had assumed? Then the interquartile range is computed by the command IQR and inside the argument you have to give the data vector, and in case if the data vector has some missing values which are denoted by here xna, then the command will be modified as IQR of xna in the command na dot rm is equal to TRUE.

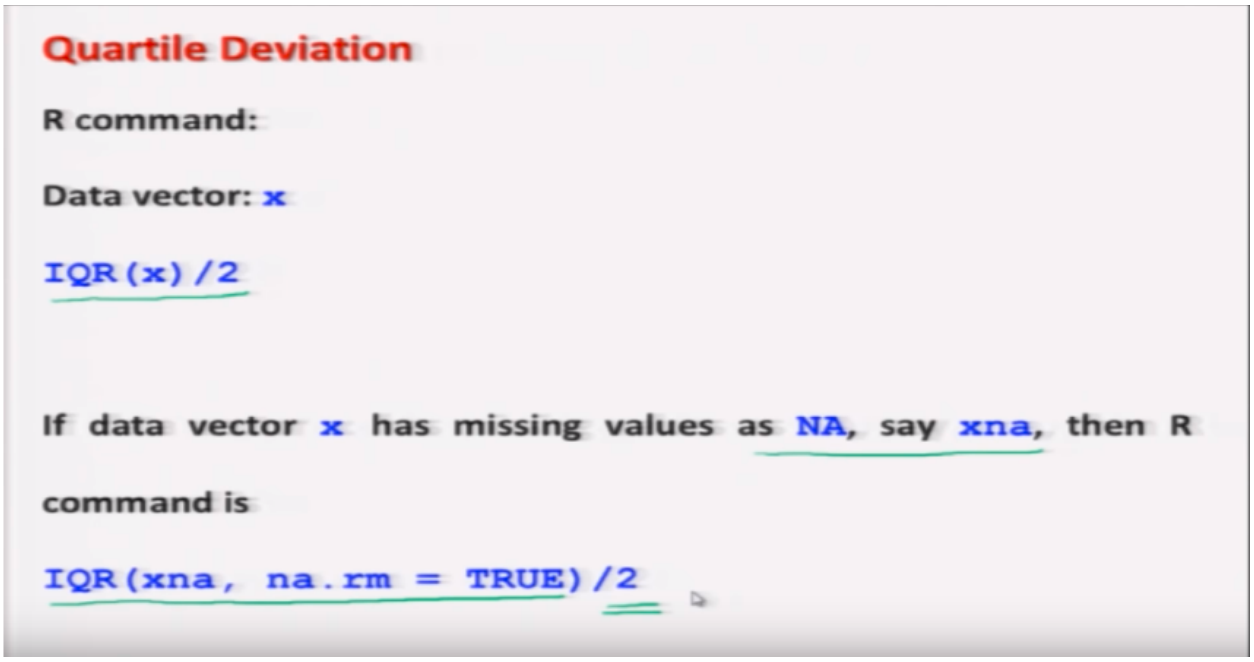
Now, with this thing I would like to introduce one more measure that is called as quartile deviation. This quartile deviation and interquartile range both are very closely related to each other, and after this I will try to show you how to compute it on the R software. Okay?

Refer Slide Time: (40:32)



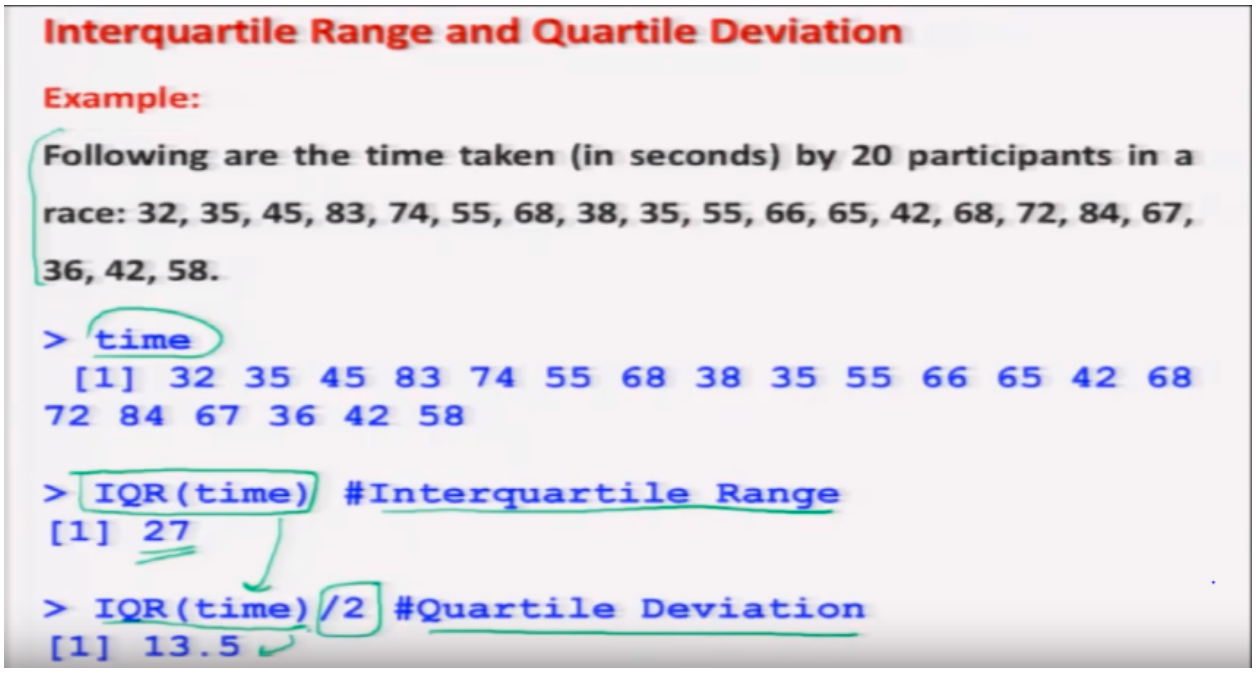
So, this quarter deviation is another measure to find out the variability in the data and this is defined as the half difference between the 75th and 25th percentiles are the half difference between the third and first quartile. So, this is essentially half of the value of the interquartile range, so half of the interquartile range is called as quartile deviation. So, it is not really difficult to give the definition of the quartile deviation, we simply have to take the difference between Q_3 minus Q_1. Which is nothing but your interquartile range and you have to divide it by 2, which I am doing here, so this is nothing but the half of the interquartile range and the decision making in this case is the same as in the case of interquartile range the data set having a higher value of quartile deviation is said to have more variability. Right?

Refer Slide Time: (41:36)



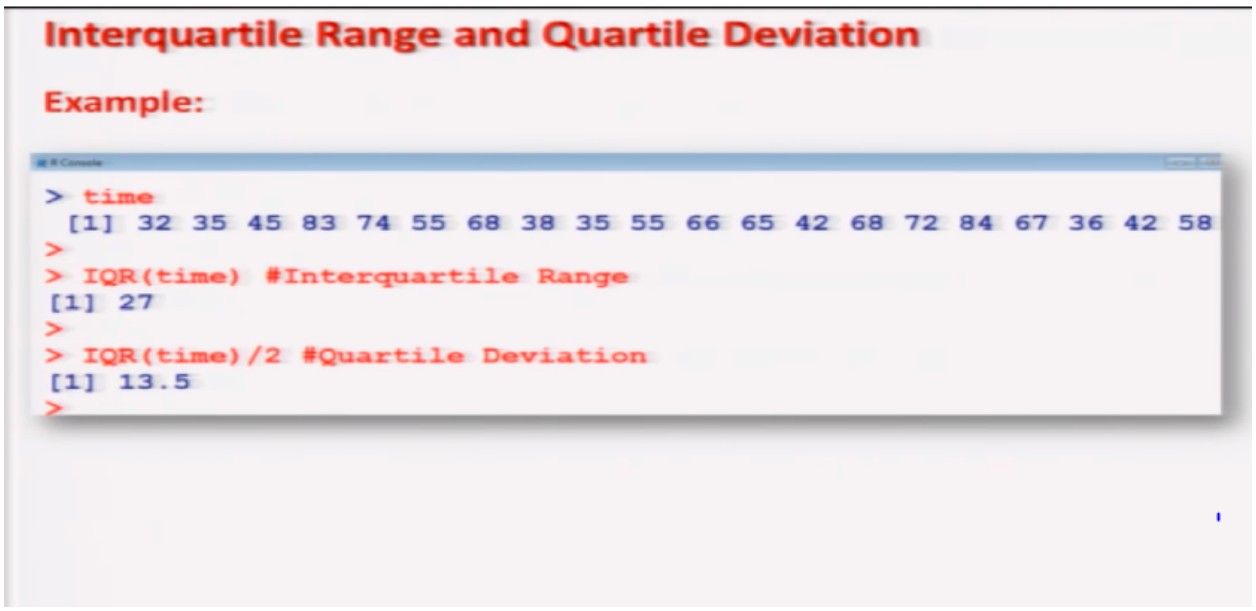
Now in case if you want to compute the quartile deviation on the R software it is pretty simple, you already have learned how to compute the interquartile range so you simply have to write down the same command and divide it by 2, and suppose if the data vector has some guessing values, then again just write the same command define for the interquartile range in the case of missing data and divide it by 2. So, this command will give you the value of the quartile deviation. Now I will try to take an example to show you how to compute these things and then I will to show you on the r console also. So, again I'm going to take the same example.

Refer Slide Time: (42:17)



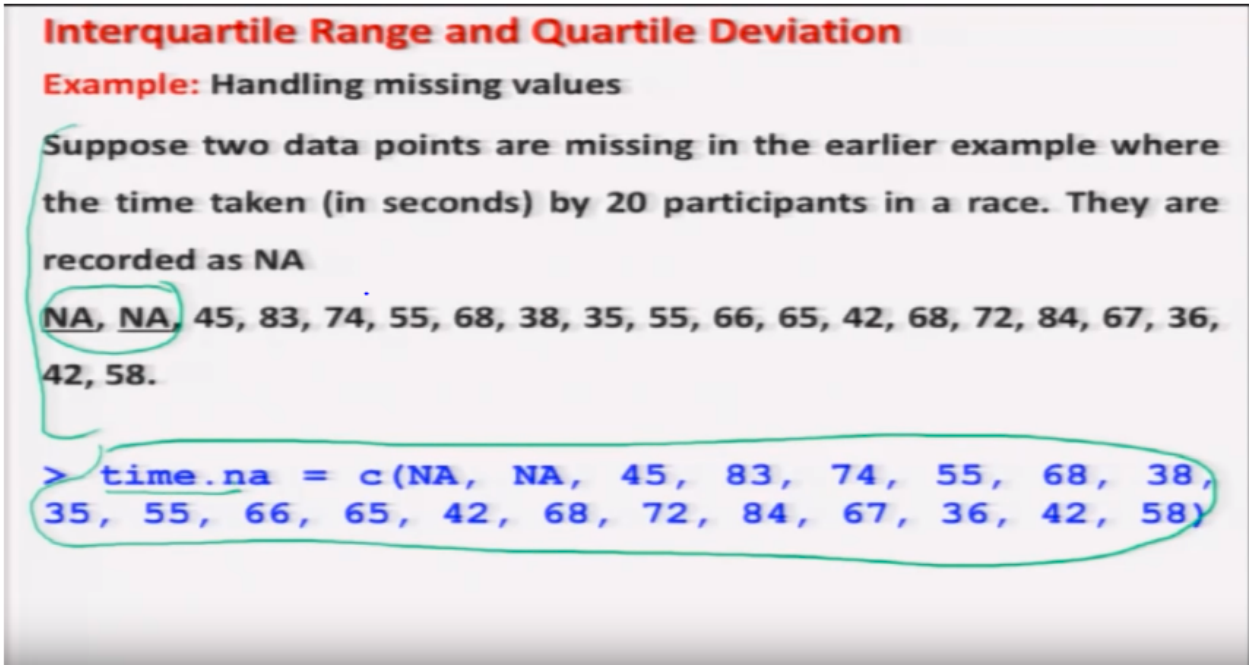
Where, I have restored the data on the time of 20 participants, now if I simply tried to operate here the command IQR inside the argument time. This will give me the value of the interquartile range and this value comes out to be here 27, and similarly, if I want to find out the quartile deviation. I simply have to write down IQR of time the same command which I have used here and just divided by 2. So, this value will come out to be thirteen point five.

Refer Slide Time: (42:53)



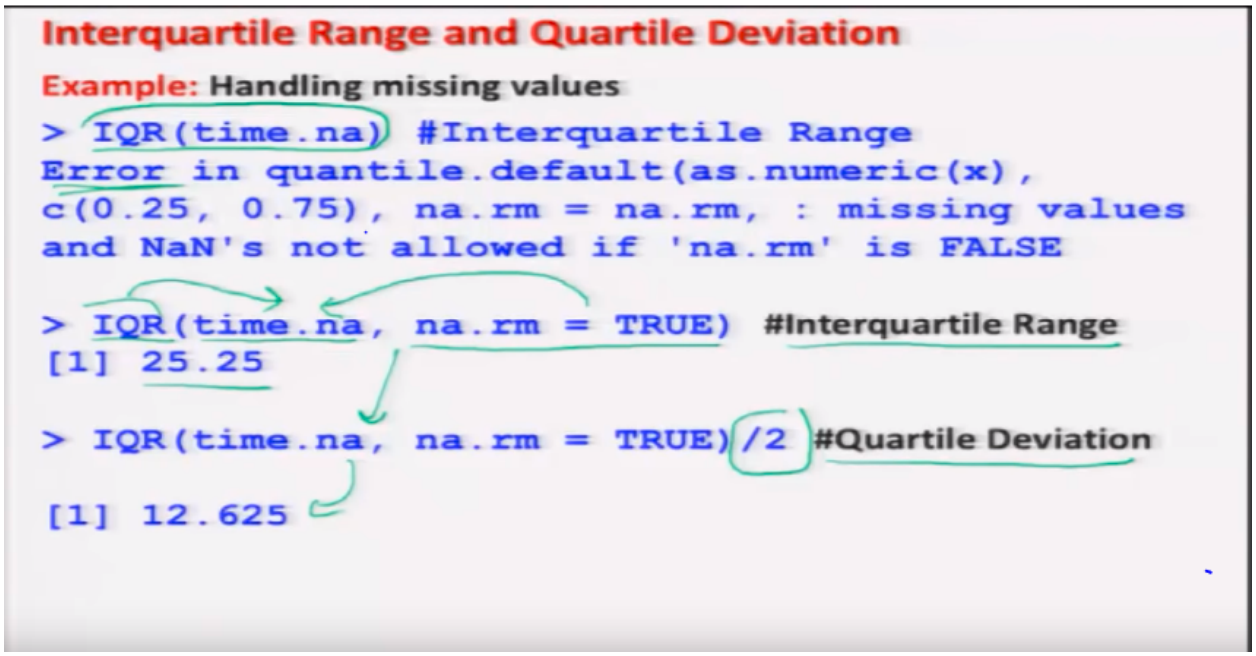
And, this is here the screenshot of the operation.

Refer Slide Time: (42:55)



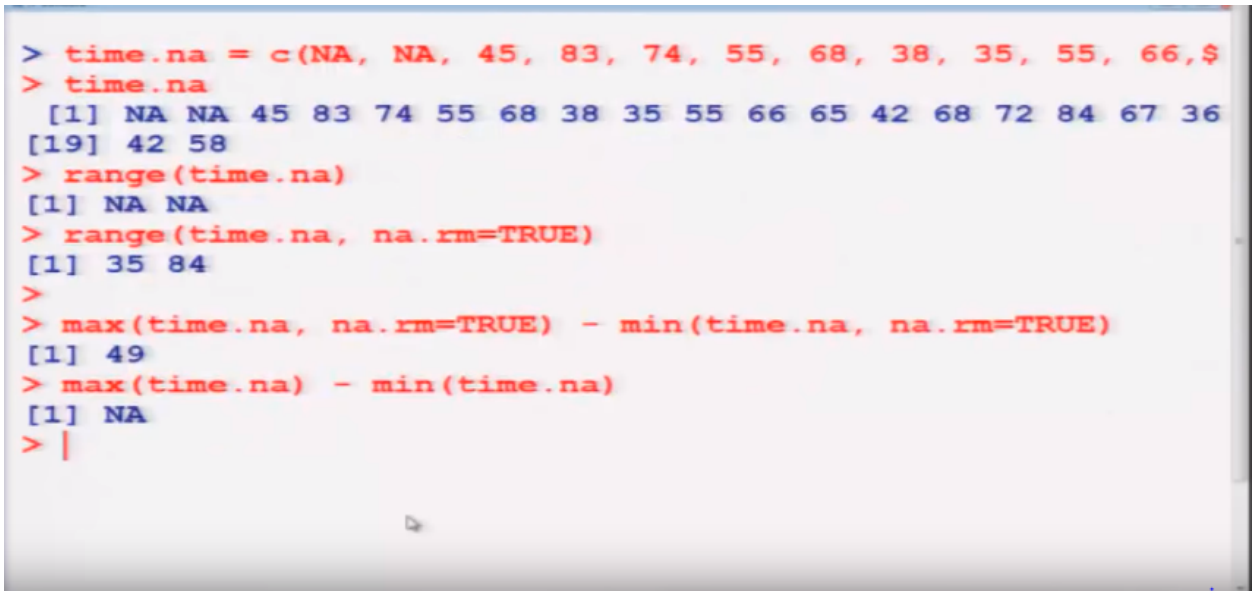
And, now I will try to show you first that how you are going to handle the essing value. So, once again I will try to take the same example, which I have taken earlier that in the time data the first two values have been replaced by na, so they are representing the missing values and these missing values have been given inside the data vector time dot na.

Refer Slide Time: (43:17)



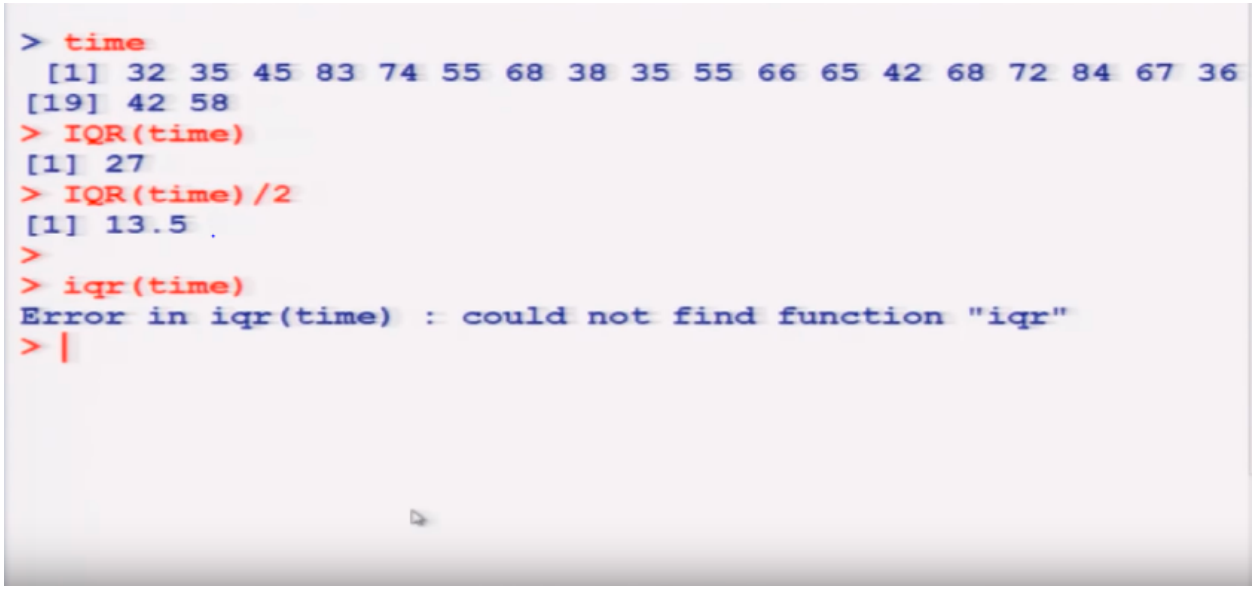
Now after this in case if you simply try to put here IQR of time dot na, so obviously this is going to give you an error because there are missing values. So, what you have to do in case if you want to compute the interquartile range, then give IQR as the command the data containing the missing value time dot na and write the command na dot rm is equal to TRUE. So, this is going to tell this time dot na that when we are trying to compute IQR then first the missing values have to be removed, so this value comes out to be a 25 point two five, and if you try to find out the quartile deviation, just try to use the same command here and divide it by 2. So, this will give you the value of interquartile range.

Video Start Time: (44:09)

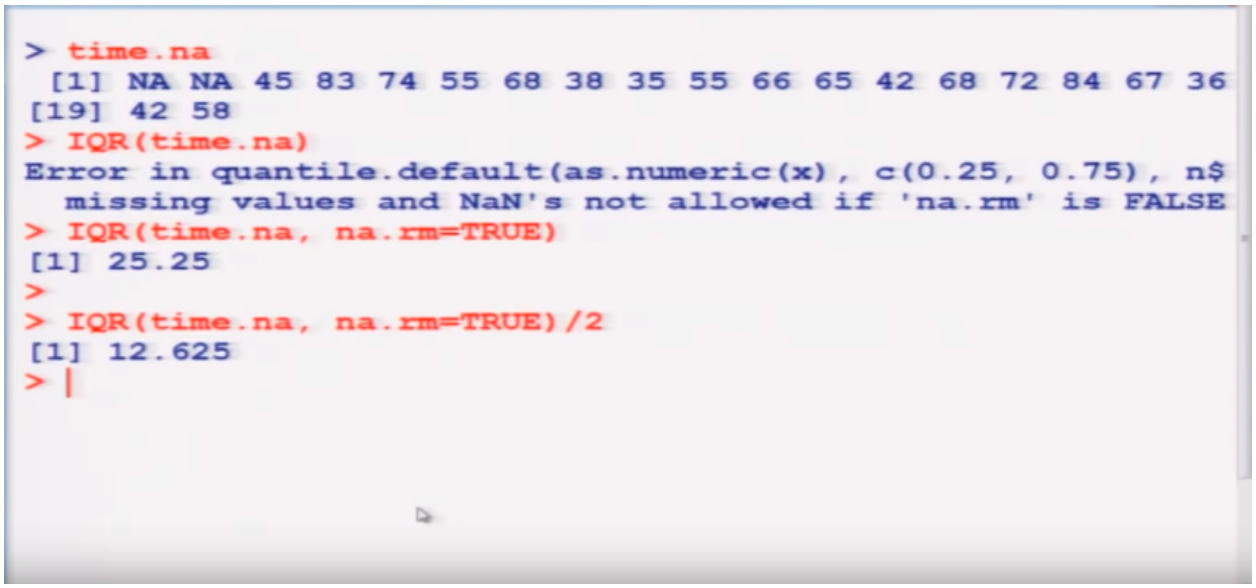


So, I will try to now show you this thing on the r console. So, you can see here you already have the data entered here time, so I will say here IQR of time. This is coming out to be twenty-seven and if you want to find out the inter quartile deviation you just divided by 2, this will come to be like this. You see what happens if I try to use the small quick address i, iqr this will give me a mistake. So, what you have to keep in mind that IQR command is case sensitive. They're small iqr and say capital IQR these are different thing, Right.

Video End Time: (44:47)



Refer Slide Time: (44:48)



And, similarly if I try to take the data on the a single use time dot na. So, this is my data vector now if I want to find out the IQR of this time dot na, so I try to do here without giving the argument na dot rm is equal to TRUE, and you can see here this gives me a mistake. So, I try to add here the command that na dot rm is equal to TRUE, and you get here this value and if you want to find out the, the IQR divided by two, the quadrille deviation. So, you have to simply divide the interquartile range by two, and this gives me the value here twelve point six two five. Okay?

So, in this lecture I have given you concept of variation in data and I have introduced two measures which are based on certain values, for example range is based on two values minimum and maximum, and quartile deviations they are or interquartile range. Both are based on two values first quartile and third quartile. So, at this moment, I would request you, you please try to go through with the lecture take some example and practice them on the R software try to make different experiment try to get the values and try to see that how the values which you have obtained for range and interquartile range they are going to give what type of information which is contained inside the data so you practice it and I will see you in the next lecture. Till then, goodbye.