**Video Script: Section 4 Video 3 using stat\_function**

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Description | Action on screen | Narration |
| 1 | Introduction  (Outcome and why it is desirable)  This should give the viewer an idea of the outcome of the task at the beginning of the videos and set the stage and expectations of the viewer. | Refer to PPT | In this video, we are going to see how to add a user-defined function to a plot. |
| 2 | Context(Problem/Solution)  Present the viewer with a real-world solution and how the situation would pose as a challenge. It always helps to draw the viewer's attention using a use-case. Metadata template can be used here. |  | Rather than fitting a model, you might want to plot a particular function to compare your data against, for example to check that the data is behaving in the manner you’re expecting. |
| 3 | Guidance (How to do it and how it works): |  | With stat\_function(), you can draw any function you want, as well as passing parameters to it. |
| 4 |  | Highlight and run:  library(ggplot2)  # checking the empirical distribution against the theoretical distribution  ggplot(data.frame( value = rnorm(1E3))) +  geom\_density(aes(x=value)) +  stat\_function(fun = dnorm, colour='red') +  ggtitle("Normally distributed data")  A description... | Open activity\_04\_03.R  Run the first ggplot command. |
| 5 |  |  | We took a large sample of 1000 normally distributed variables and plotted its distribution. |
| 6 |  |  | We then added a plot of the theoretical normal distribution (dnorm()) using stat\_function. |
| 7 |  |  | We can visually check that the sample is large enough and that its distribution is very close to its theoretical value, in red. |
| 8 |  | Highlight and run:  someAverages <- data.frame( averages = replicate(1E3, mean(runif(10)))) | Run the next command |
| 9 |  |  | The data frame someAverages contains 1000 averages of samples of 10 uniformly distributed values.  An important theorem in statistics ( the central limit theorem) tells us that the averages should be normally distributed with a predictable mean and standard deviation. |
| 10 |  | Highlight and run:  predictedMean <- 0.5  predictedStandardDeviation <- sqrt(1/12)/sqrt(10)    ggplot(someAverages) +  geom\_density(aes( x = averages)) +  stat\_function(fun = dnorm, colour='red', args = list( mean = predictedMean, sd = predictedStandardDeviation )) +  ggtitle("The central limit theorem holds.")  A description... | The next ggplot command helps us verify that this is indeed the case. |
| 11 |  |  | We’re adding dnorm to the plot again but this time with some extra parameters (mean and sd) specified as a list in the parameter args.  The theorem gave us the predicted values for those. You can find their derivation in the source code.  You can see a good fit between our data and its theoretical distribution. |
| 12 |  |  |  |
| 13 |  |  |  |
| 14 |  |  |  |
| 15 |  |  |  |
| 16 | Conclusion:The video concludes by showing the viewer that the goal has been achieved, and reminding them why they should be happy about that. A PowerPoint summary slide with the key points emphasized would make it easier for the viewer to remember what was covered in the video |  | Now you know that it is very easy to draw a function on top of a plot with stat\_function. This is very useful to check an assumption about the data for example.  In the next video, we’ll see how to plot very large datasets efficiently. |