**Video Script: Section 3 Video 4 Automatically plotting subsets of a data set.**

|  |  |  |  |
| --- | --- | --- | --- |
| 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 will see how to automatically produce individual plots for each value of one variable. |
| 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. |  | It is always possible to draw a whole dataset on a single graph but it’s often more informative to look at subsets of it individually. |
| 3 | Guidance (How to do it and how it works): |  | With faceting, it is very easy to produce a similar plot for multiple subsets of a dataset. |
| 4 |  | Highlight and run:  library(ggplot2)  # one single plot  ggplot(diamonds) +  geom\_histogram(aes(x=price, fill=cut)) +  ggtitle("One single plot") | Open activity\_03\_04.R  Run the first 9 lines. |
| 5 |  |  | We have plotted the distributions of the price of 50’000 diamonds for each possible cut, from fair to ideal. |
| 6 |  |  | This can be useful to compare one with another but the plot is rather difficult to read if you’re interested in each distribution individually. |
| 7 |  |  | What we want is a plot of the price distribution for each value of cut. |
| 8 |  |  | We can do this with the facet() function |
| 9 |  | Highlight and run:  ggplot(diamonds) +  geom\_histogram(aes(x=price)) +  facet\_wrap( ~ cut) +  ggtitle("One plot per cut") | Run the second ggplot command. |
| 10 |  |  | With very little change to the code, we produced 5 individual graphs, one for each value of cut. |
| 11 |  |  | Each facet panel shares axes. This makes it easier to compare values across panels than just drawing five separate plots. It's also much more space-efficient. |
| 12 |  |  | You can also control the plots layout and axes with options like ‘ncol’, ‘nrow’ and scales. |
| 13 |  | Higlight and run:  ?facet\_wrap  Help appears on the right hand side. | Use the ?facet\_wrap for the complete list of options. |
| 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 | Back to PPT | Now, we can automatically produce individual plots for each value of a particular column of the data frame. This makes for less crowded and more legible graphs.  In the next video, we’ll see how to make subplots for each values of several variables. |