**Video Script: Section 3 Video 5 Faceting with several variables**

|  |  |  |  |
| --- | --- | --- | --- |
| 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’re going to see how to generate plots for each value of several variables. |
| 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. |  | With a large enough dataset, it’s often interesting to look at subsets defined by several variables. |
| 3 | Guidance (How to do it and how it works): |  | We’re going to see two ways to make conditional plots over 2 variables. |
| 4 |  | Highlight and run library(ggplot2)  # distribution of diamond price given color and cut.  # note the the use of scale\_x\_log10 for plotting on a log scale, which is more appropriate here.  ggplot(diamonds) +  geom\_histogram(aes(x=price)) +  scale\_x\_log10() +  facet\_wrap( color ~ cut) +  ggtitle("with facet\_wrap")  A description... | Open activity\_03\_05.R  Run the first 11 lines. |
| 5 |  |  | We end up with a graph similar to what we had in the previous video: the distribution of the price for different types of diamonds.  By using facet\_wrap(color ~ cut), a plot is produced for each of the combinations of color and cut, which is what we want here. |
| 6 |  |  | However, as it is, it is difficult to appreciate what is the importance of cut on the price distribution, because the plots are simply strung one after the other. |
| 7 |  | Highlight and run:  # with facet\_grid  ggplot(diamonds) +  geom\_histogram(aes(x=price)) +  scale\_x\_log10() +  facet\_grid( color ~ cut) +  ggtitle("with facet\_grid")  A description... | Run the second ggplot command |
| 8 |  | Highlight  color ~ cut + clarity  But don’t run, too slow (~10 seconds) | We have used facet\_grid instead of facet\_wrap, which produces a more orderly table of graphs that makes for easier reading.  The 2 variables ‘cut’ and ‘color’ now run on the x and y axes.  It’s possible to use more than 2 variables by using the formula in the usual way, for example:  color ~ cut + clarity |
| 9 |  |  | You can also add extra plots with all the data by setting the option ‘margin’ to TRUE. |
| 10 |  | Highlight and run  ?facet\_grid  (note that code also points to ggplot2 webpage for facet\_grid). | See the help file for facet grid to see the complete list of options. |
| 11 |  |  |  |
| 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 |  | From this video, you have learned that faceting with several variables is possible with either facet\_wrap or facet\_grid but the plots resulting from facet\_grid are easier to interpret.  In the next section, we will see how to use a statistical summary in our plots. |