**Video Script: Section 5 Video 1 – controlling axes with scales and limits**

|  |  |  |  |
| --- | --- | --- | --- |
| 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 section, we will see how to override the choices made by default by ggplot2 and customize our plots.  In this video, we’ll see how to control axes, for plotting on different scales or changing the area to be plotted. |
| 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 often useful to transform the data before plotting it, in order to make a relationship more apparent. A typical transformation is taking the log of the value, to produce a log or log-log plot if both axes are transformed. |
| 3 | Guidance (How to do it and how it works): | Switch to RStudio with activity05\_01.R open | Open activity\_05\_01.R and load ggplot2 |
| 4 |  |  | Using our usual diamonds dataset, we want to plot carat vs price on a log-log scale. |
| 5 |  |  | A simple way to achieve this is to directly transform the data and replacing price by log10(price) and carat by log10(carat). |
| 6 |  | Highlight and run:  ggplot(diamonds) +  geom\_point(aes(x=log10(carat),y=log10(price))) +  ggtitle('Manipulating data') | Run the first example: |
| 7 |  | A description... | We see a very clear linear relationship on the log-log scale, which suggests that the price is the carat with some exponentiation. |
| 8 |  |  | However, the axes are difficult to read: reading 3.5 on the y-axis, for example, means a price of 10^3.5 =3125  Using scale\_x\_log10() and scale\_y\_log10(), we can ask ggplot2 to transform the axes, without us applying the transformation manually. |
| 9 |  | Highlight and run:  ggplot(diamonds) +  geom\_point(aes(x=carat,y=price)) +  scale\_x\_log10() + scale\_y\_log10() +  ggtitle("Manipulating axes' scales") | Run the second example |
| 10 |  | A description... | This way, we haven’t manipulated the data and it’s much easier to read off values of the axes. |
| 11 |  | Highlight and run:  p <- ggplot(movies, aes( x = length)) + geom\_histogram()  p + ggtitle('Showing all movies') | You can also control the limits of your plot with xlim and ylim. Run the next example. |
| 12 |  | A description... | This is the distribution of movie lengths from the dataset movies. Because of some rare and very long movies, most of the data is squashed on the left hand side. |
| 13 |  | Highlight and run:  p + xlim(c(0,180)) + ggtitle('Showing only movies of duration between 0 and 3 hours.') | Use xlim()) to only plot the distribution between 0 and 3 hours. Run the last example. |
| 14 |  | A description... | We are now looking at a more interesting portion of the distribution. |
| 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 | We saw how to plot on different scales and to plot only parts of the graphs without manipulating the data, using scales and limits.  In the next video, we’ll see how to control the order in which categorical variables are plotted. |