**Video Script: Section 2 Video 1 Drawing lines**

|  |  |  |  |
| --- | --- | --- | --- |
| 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 PPT slide | Hi, in this section of the course, we will take a look at basic plots that often come up during data analysis:   * Lines * Paths * Bar charts * Histograms and density plots * Boxplots   **In this video**, we will learn how to use geom\_line() to make lines very easily. |
| 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. |  | So far, we only have plotted individual data points on a scatter plot, using geom\_points().  It often makes sense to connect the data points, ~~e.g. body temperature over time.~~ Lines can also be used for visualizing two continuous variables. For this, we are going to use geom\_line(). |
| 3 | Guidance (How to do it and how it works): | Switch to RStudio.  Open activity\_02\_01.R  Run the first few lines. | In RStudio, open ‘activity\_02\_01.R’.  Run the first few lines in the console, up until ‘head(X)’.  As usual, select the text and press ctrl+enter. |
| 4 |  | In the console:  index Stock value  6015 435 FTSE 2868.00  3424 1564 SMI 5669.90  6325 745 FTSE 3106.00  1100 1100 DAX 2305.58  3787 67 CAC 1887.40  2490 630 SMI 2741.90 | On the screen, you’ll notice that we have a sample data set created from a recording of stock market values (EuStockMarkets). This is a data frame with 3 columns:   * index: an index which increases with time. * stock: the name of the stock * value: the value of the stock |
| 5 |  | Highlight  # shuffle the rows  X <- X[sample(1:nrow(X)),] | We have also removed 200 rows at random and shuffled the index. |
|  |  | Run ggplot(X) +  geom\_line( aes( x = index, y = value, colour = Stock)) +  ggtitle("geom\_lines, using colour")  A description... | We use the aesthetics ‘colour’ to differentiate the stocks:  ggplot(X) +  geom\_line( aes( x = index, y = value, colour = Stock)) +  ggtitle("geom\_lines, using colour")  Geom\_line() requires at least 2 aesthetics: x and y. It will connect the points along those coordinates with a line, but will also make sure to sort the values for x in increasing order.  We see that despite the index having been shuffled, the points are arranged in the correct time order. Lines are always drawn from left to right. They never zigzag across the plot.  It automatically produces a legend on the graph. |
|  |  | Run the second example:  # plotting without grouping  ggplot(X) +  geom\_line( aes( x = index, y = value)) +  ggtitle("geom\_lines, without defining groups") | Run the second example.  In this case, all the datapoints are joined up since we didn’t give ggplot any information as to how to differentiate between them. This is obviously not terribly useful. |
| 6 |  | Run the third example:  ggplot(X) +  geom\_line( aes( x = index, y = value, group = Stock)) +  ggtitle("geom\_lines, using group ") | Run the third example:  ggplot(X) +  geom\_line( aes( x = index, y = value, group = Stock)) +  ggtitle("geom\_lines, using group only") |
| 7 |  | Highlight ‘group = Stock’ | In this case, we used the aesthetics ‘group’ in order to separate the four different stocks,. |
| 8 |  |  |  |
| 9 |  |  |  |
| 10 |  |  |  |
| 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 | Last slide of the PPT | In this video, we’ve seen Geom\_line(), which allows you to connect data points in a graph, which is especially useful for time series or for visualizing the relation between two variables.  In the next video, we will see how to draw paths. |