**Video Script: Section 1 Video 3 – mapping data to graphical elements with aesthetics**

|  |  |  |  |
| --- | --- | --- | --- |
| 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. | Opening slides. | **In this video**, we are going to look at how graphical elements can be controlled by parameters called aesthetics. |
| 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. |  | In ggplot2, the data is always passed on in the form of a data.frame. The idea is to map the columns of the data.frame to these parameters and use their type (e.g. continuous/categorical) to automatically produce the correct presentation. |
| 3 | Guidance (How to do it and how it works): | Activity01\_03.R appears in the editor. | Let's see an example. Open activity01\_03.R. (CTRL+O) in Rstudio. Don't run the code yet. |
| 4 | Consider a simple scatter plot. | Highlight and run:  library(ggplot2)  set.seed(42)  small<-diamonds[sample(nrow(diamonds),1000),]  head(small)  The first 6 elements of 'small' are displayed in the console.  carat cut color clarity depth table price x y z  49345 0.71 Very Good H SI1 62.5 60 2096 5.68 5.75 3.57  50545 0.79 Premium H SI1 61.8 59 2275 5.97 5.91 3.67  15434 1.03 Ideal F SI1 62.4 57 6178 6.48 6.44 4.03  44792 0.50 Ideal E VS2 62.2 54 1624 5.08 5.11 3.17  34614 0.27 Ideal E VS1 61.6 56 470 4.14 4.17 2.56  27998 0.30 Premium E VS2 61.7 58 658 4.32 4.34 2.67 | Run the first 4 lines of the code (select , CTRL+ENTER).  By doing this, we load the ggplot2 library and prepare a small dataset for future plotting.  The dataset is a collection of diamonds' properties: carat, cut, color, clarity, depth, table, price and 3 values (x,y,z) related to the size of the diamonds.  Some of these properties are continuous (e.g. carat, price, depth), other are categorical (e.g. cut, color, clarity). |
| 5 |  | Highlight and run:  ggplot(small) + geom\_point(aes(x=carat,y=price))  A black and white plot appears.  01_03_example01.png | Suppose we want to investigate the relation between carat and price with a scatter plot.  For this, we use geom\_point() for plotting points on the graph.  Run the first example. |
| 6 |  | Highlight ggplot(small) | Here is how we did it:  ggplot(small) creates a ggplot object and tells it to use some data (always a data frame). |
| 7 |  | Highlight geom\_point(aes(x=carat,y=price)) | We then add a layer with geom\_point():  geom\_point(aes(x=carat,y=price)) |
| 8 |  | Highlight aes(x=carat, y=price) | The code aes(x=carat, y=price) maps the data to some of the properties of the geom, or geometrical object. A point on a graph needs to have at least two properties: its x and y co-ordinates. Here small$carat will be used for x and small$price will be used for y. |
| 9 |  |  | This is all that is needed for producing the plot. Note how the labels are automatically produced and the plot has nice default colours.. |
| 10 | Other graphical properties | Highlight and run:  # example 02  ggplot(small) + geom\_point(aes(x = carat, y = price, colour = cut))  A colourful plot appears.  01_03_example02.png | Run example 02. This similar to the previous plot but this time each point has a color that depends on the cut of the diamond.  Note that since 'cut' is a categorical variable (I.e a factor in R terminology), ggplot has automatically decided to use a discrete scale for the colour scheme.  Also , you’ll notice that the colour scale appears automatically on the graph. As an example, in case the data or the number of cuts change, the graph will change accordingly, but you won't need to update the code, since all graphical elements are a function of the data. |
| 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 | We’ve seen how to link the data to graphical elements through the use of aesthetics and how the type of the data (continuous / categorical) is automatically used by ggplot2.  In the next video, we will clarify some subtleties about aesthetics which can be puzzling at first sight. |