**Video Script: Section 4 Video 1 linear and polynomial trends**

|  |  |  |  |
| --- | --- | --- | --- |
| 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 uncover trends in the data from scatter plots using built-in R functions. |
| 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. |  | A scatter plot can give us a good idea of the relation between two variables. You can easily add trends calculated from the data to the scatter plot and make the trends stand out. |
| 3 | Guidance (How to do it and how it works): |  | ggplot supports two types of smoothing: linear regression , and a LOESS, which is a non-parametric model which can accommodate more complicated trends. |
| 4 |  | Highlight and run:  library(ggplot2)  # plot all data points, with some jittering  p <- ggplot(ChickWeight, aes(x=Time, y=weight, colour=Diet)) +  geom\_point(  # Connected lines  p + geom\_line(aes(group = Chick)) + ggtitle("connected lines")  A description... | Open activity\_04\_01.R  And run the first 9 lines |
| 5 |  |  | The plot shows the Chickweight data we came across before: the weight of 50 chicks under 4 different diets was monitored for a period of time. |
| 6 |  |  | The plot is a bit confusing because all data is there. Using geom\_smooth(), we can summarise the effect of each diet and thus get 4 curves instead of 50.  Each curve will capture the underlying trend for each diet. |
| 7 |  | Highlight and run:  # Default smoothing (loess) with small dataset  p + geom\_smooth(alpha=.2, size=1) + ggtitle("Default smoothing (loess)")  A description... | Run the next ggplot command. |
| 8 |  |  | Geom\_smooth() built a polynomial fit to each of the diet.  The gray area represents the standard error on the best-fit line, so the better the fit the thinner the gray ribbon.  The optional parameters alpha and size control the ribbon’s transparency and the line’s thickness, respectively.  geom\_smooth uses loess by default if there are fewer than 1000 points, as it is the case here. |
| 9 |  |  | By using the parameter ‘method’ in geom\_smooth(), you can get a linear model instead. |
| 10 |  | Highlight and run:  # Linear model  p + geom\_smooth(alpha=.2, size=1, method="lm") + ggtitle("Smoothing with 'lm'")  A description... | Run the next ggplot command. |
| 11 |  |  | This time the trend is forced to be linear, i.e. a straight line. |
| 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 | Back to PPT | We have seen how to use geom\_smooth to quickly view trends in the data, where many individual lines are too confusing, and also visually appreciate the fit of a model.  In the next video, we’ll see how to fit more sophisticated curves to the data. |