INST0065 Practical work – Week 4

This week we have focussed on two things: working through and understanding an existing R script, and then looking at the use of RMarkdown files.

Candles data

The first part of the practical involves replicating the candles visualizations, and then altering them so that we can be more sure that we understand how they work.

1. If you have not already done so, copy and unzip the Candles data and code from github or Moodle.
2. Load the R file into RStudio; using File->Open file…
3. Try running the code. You could do either:
    a. Press 'run' repeatedly to run one line at a time (the cursor should advance each time), or
    b. Select the whole file, and then click 'Run'
4. Does this work? The answer is probably 'no', if R has not found and read in the data files. You have two choices
    a. Alter the file paths in the read_excel() statements, or add a setwd() command before you try to read the file.

       Note, for importing the data, it will be easiest to use a full filepath, e.g.:

       ```
       Scented_All <- read_excel("N:/modules/inst0065/Candles-main/Scented_all.xlsx")
       ```

       The path will obviously depend on where you have stored the data; note that if you are using Windows, we are using '/' in the filepath rather than the usual '\'.
    b. Use RStudio's File->Import dataset->From Excel tool

       Note: when I have tried this, it imports it as an object called Scented_all. R is case sensitive, so this is not the same as Scented_All. You will need to either change the subsequent code, or, add in an extra line of the form:

       ```
       Scented_All <- Scented_all
       ```

       in order to make sure that this is consistent.
5. Having corrected the data import stage, try running the rest of the file again. Does this work now?
6. The code as is may not produce the final outputs; why not? What do you need to add to see the visualizations?
7. Once you have produced the original visualizations, make sure you save your revised .R file. You should then modify it
    a. Comment out different layers; using '#' to comment out a line. Try producing the s1720 plot a number of times, with various layers removed. For example:

```
s1720 <- ggplot(s, aes(x = (as.Date(Date)), y = Rating)) +
#  geom_vline(xintercept = as.numeric(as.Date("2020-01-20")), colour =
"indianred1", linetype = "dashed")+
#  geom_smooth(method = "loess", size = 1.5, colour = "lightseagreen", fill
= "lightseagreen") +
  geom_point(alpha = 0.2, colour = "lightseagreen")+
  labs(x = "Date", y = "Average daily rating (1-5)", title = "Top 3 scented
candles Amazon reviews 2017-2020")+
  theme_light()+
  theme(plot.title = element_text(size=16))+
  scale_x_date(date_labels = "%m-%Y", date_breaks = "6 month")
```

Remember that you will need to select the code producing the plot and run it each time, and will then need to display the resulting object.

b. Which layers 'work' on their own? Which work in combination? Are there differences in the structure of the output when you only use one geom_ layer?
c. Try altering the settings for the various layers, and look at the results
    i. Alter colour settings
    ii. Alter the alpha setting in the point layer
    iii. Alter the labels

RMarkdown files

Note for help with RMarkdown options, see the RMarkdown cheatsheet and for help with ggplot see the ggplot cheatsheet.

8. Use the RStudio: File->New file->RMarkdown template to do the following. Remember that you should save the file after editing it, and should then use 'knit' to create output.
    a. Write an RMarkdown file that shows a table of summary data (using summary()) for the cars data set. The resulting knitted file should contain text explaining what you are doing
    b. Write an RMarkdown file that uses ggplot() to plot the cars data as a scatterplot showing (stopping) distance against speed. Adjust the appearance of the scatterplot as you see appropriate.
    c. Create a new version of the RMarkdown file used in (b), and add to the ggplot command another layer that uses geom_smooth() to plot a smooth line through the car stopping distances. Again, add text commentary to the markdown file explaining what you are showing. Consider whether you should add 'echo=false' to any of your chunks of R code.

9. Rework the R script that was used for the scented candles exercise so that it works as an RMarkdown file.

    You should add suitable titles etc. As with any academic work, you should cite your sources, in this case the code and data being used.

    As with the R script work above, you will need to set full pathnames in order to read in the Excel data.

    You should include suitable explanatory text, as well as the graphs.

    You should decide which R chunks should be echoed to the knitted output, and which do not need to be echoed.
10. As before, experiment with changes to the R code to alter the appearance of these visualizations. If they are bound into an output document, do you want to title the graphs in the same manner that you did before.