Mustapha Brima DACSS 601 - Assignment 4

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

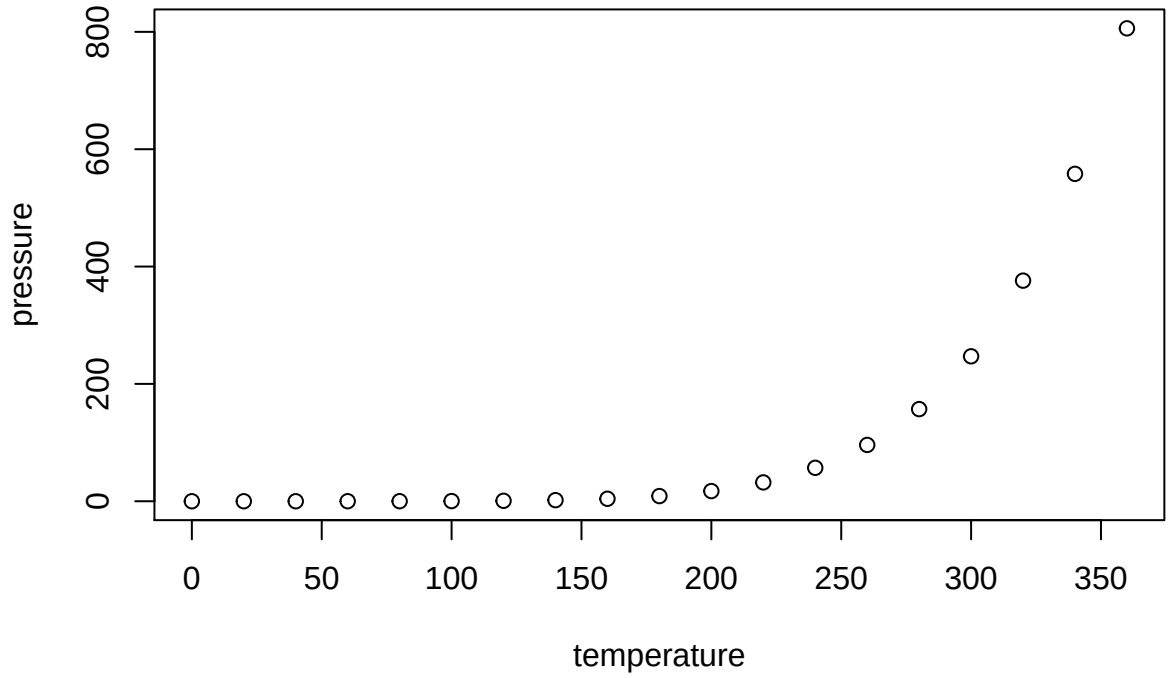
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



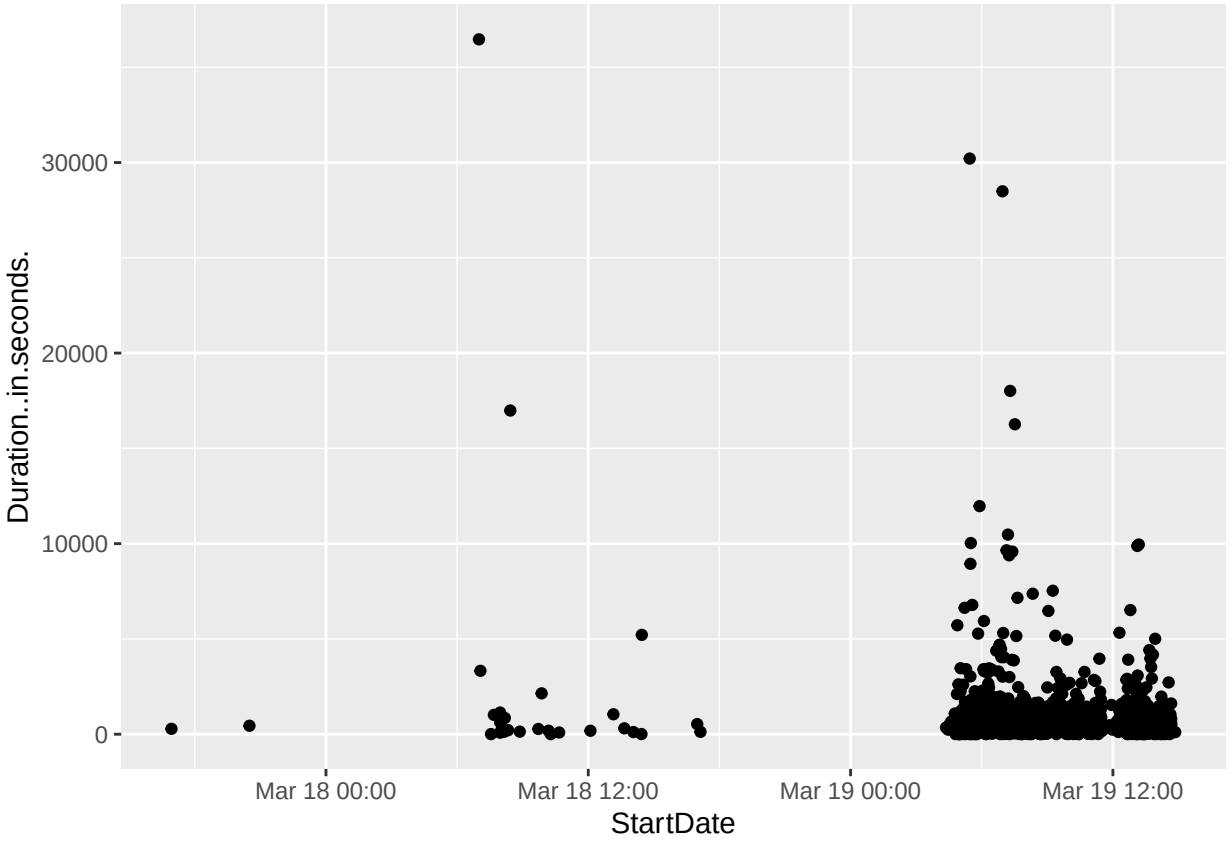
Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

covid_responses_tab %>% select(Duration..in.seconds., StartDate, EndDate)%>% arrange(StartDate)

Plot

##{r plot, include=FALSE} ggplot(covid_responses_tab, mapping = aes(x = StartDate, y = Duration..in.seconds.)) + geom_point() ##

```
plot(x,y)
```



Study:

I used the dataset 'covid_responses_tab' for my assignment. The dataset was a result of a survey conducted by the italian government for Italian citizens. It would be used to record how the citizens had were affected by and responded to the COVID19 Pandemic. I organized the entries in the dataset by the object 'StartDate'. StartDate represents the date and time that each participant in the survey began to take the survey. I then formed the plot for the dataset by writing code that included the StartDate or the x-axis, as well as the 'Duration..in.seconds.' by the y-axis. I had doe this in order to rey to see if thee was a correlation between Starttimes and Duration of the surveys, to which I had found that there was none. I was originally going to plot the entries in order from smallest to highest Duration, but the data would need two variables to be shown graphically based on the examples offered online. Where I felt limited was in the ability to show many of the different entries for 'Mar 19 12:00'. It seems a great amount of the participants of the survey took it around the later part of this date or from the morning until noon. Becaus of this, many of the data points on the scatter plot seem to be on top of each other. If I could have found a way to change the scaling of the plot to make more of the points known while also keeping all of the points be accounted for on the plot, then it would have been fairly easier to represent the data.