CSC8631 Assignment Report

Marc Birkett

07/11/2021

```
library(ProjectTemplate);
```

```
## Loading required package: digest

## Loading required package: tibble
```

```
library(dplyr);
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr);
library(rmarkdown);
load.project();
```

```
## Project name: csc8631

## Loading project configuration

## Autoloading helper functions

##  Running helper script: globals.R

##  Running helper script: helpers.R

## Autoloading data

## Munging data
```

```
##  Running preprocessing script: 01-A.R

##  Running preprocessing script: 02-E.R

##  Running preprocessing script: 03-L.S.R

##  Running preprocessing script: 04-Q.R

##  Running preprocessing script: 05-S.A.R

##  Running preprocessing script: 06-T.M.R

##  Running preprocessing script: 07-V.S.R

##  Running preprocessing script: 08-S.S.R
```

CSC 8631 - Data Investigation with Student Data

Introduction

Report into investigation of Student Data using the CRISP-DM model. This report covers two iterations of the model and includes the processes of Business Understanding, Data Understanding, Data Preparation. The subprocesses I've chosen are to do the following steps:

- Import
- Tidy
- Visualise
- Understand
- Communicate

The project has been set up using ProjectTemplate to provide some structure and repeatability, which will be tested on a regular basis. Version control is provided by Git and this report created with R Markdown.

Libraries

- Readr - library to provide extra functionality to import the data from CSV. In this case it allows me to import that data and assign type.
- Dply - data management.

Iteration 1

Iteration 1 will be used to investigate the data and generate a hypothesis for further investigation. It will go through the entire list of sub-processes outlined above. Once a hypothesis has been identified this will be further investigated in iteration 2.

Iteration 2

Iteration 2 will further investigate the hypothesis identified in iteration 1 and will present the findings.

Findings

To answer the hypothesis XYZ the findings are that ABC

Conclusion

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.

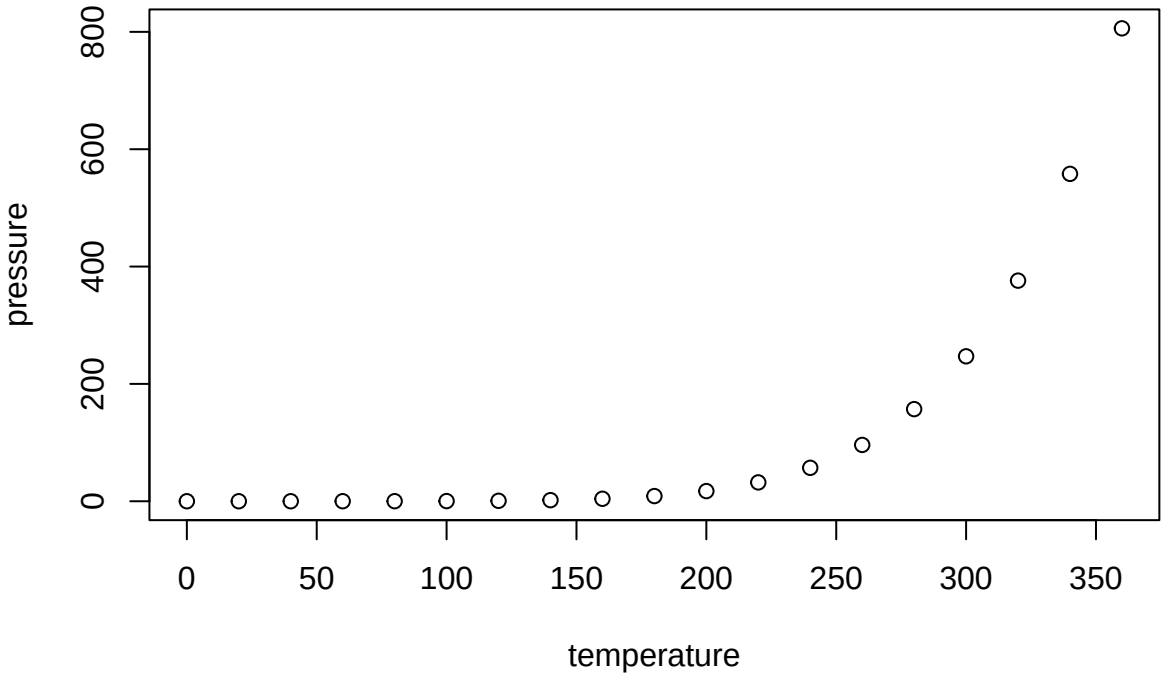
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
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.