p8105_hw2_mc5698.Rmd

2024-09-27

#Question 1

```
#loading necessary packages
library(readr)
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(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(tidyr)
library(stringr)
```

```
nyc_t=
  read.csv(
    "/Users/nicolechen/Downloads/p8105_hw2_mc5698/NYC_Transit_Subway_Entrance_And_Exit_Data.csv",
    na = c("NA",",",".")) |>
  janitor:: clean_names() |>
  select(line, station_name, station_latitude, station_longitude, route1, route2,           route3, rou
  mutate(
    entry =
      case_match(
        entry,
        "Yes" ~ TRUE,
        "No" ~ FALSE),
    entry = as.logical(entry)
      )
```

The dataset contains line, station_name, station_latitude, station_longitude, route1, route2, route3, route4, route5, entry, vending, entrance_type, ada. For the data cleaning, I removed unnecessary columns and convert the entry variable from character to a logical variable by using `case_match` function. The dimension of the resulting dataset is 1868, 13. These data are mostly tidy but we could pivot different route columns into one variable.

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.

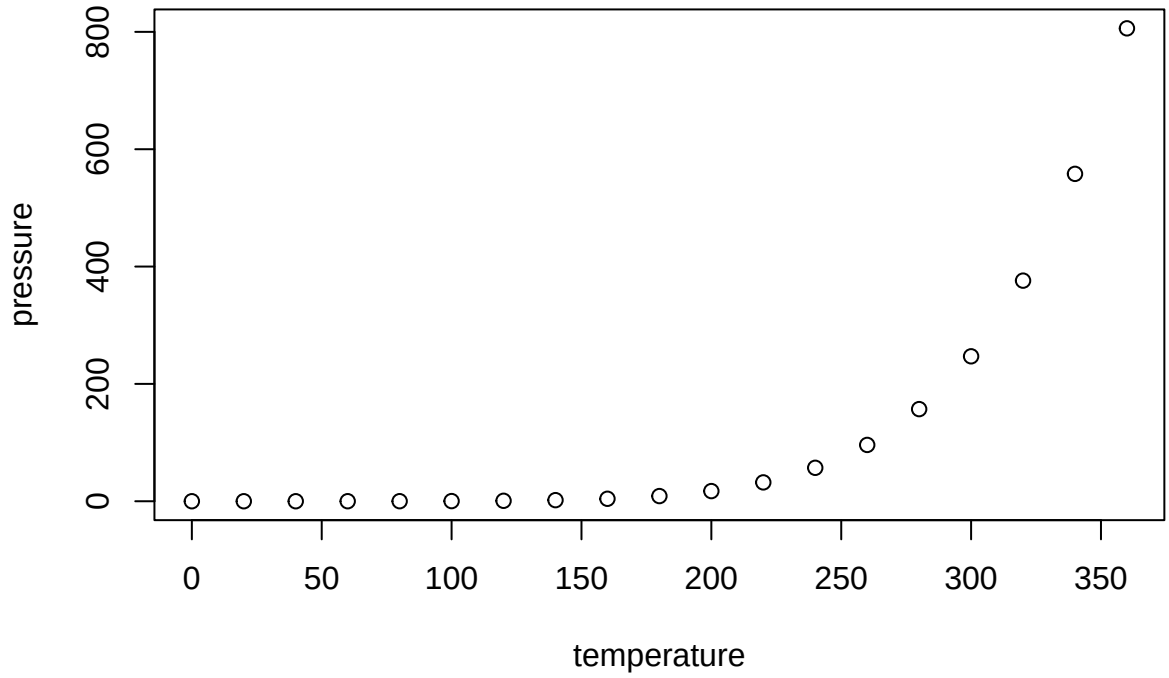
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
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.