V506 R Introductory Lab 2 Exercise – Fall 2024 - Solution

Luis Navarro

August 2024

```
# run this code on your computer to install all the packages required for the code below
if(!require(dplyr)) {install.packages("dplyr")}
```

```
## Loading required package: 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
```

```
if(!require(pacman)) {install.packages("pacman")}
```

```
## Loading required package: pacman
```

```
if(!require(here)) {install.packages("here")}
```

```
## Loading required package: here
```

```
## here() starts at /Users/luisenriquenavarro/Library/CloudStorage/OneDrive-IndianaUniversity/V506/Fall1
```

```
if(!require(rio)) {install.packages("rio")}
```

```
## Loading required package: rio
```

```
## Warning: package 'rio' was built under R version 4.4.1
```

```
library(pacman)
p_load(dplyr, here, rio)

# Clean the environment
rm(list=ls())
```

- 1. Install tidyverse

```
install.packages("tidyverse", repos = "http://cran.us.r-project.org")
library(tidyverse)
```

- 2. Set your working directory using base R or here.

```
# Base R
setwd('/Users/luisenriquenavarro/Library/CloudStorage/OneDrive-IndianaUniversity/V506/Fall24')
# Verify it
getwd()
```

```
## [1] "/Users/luisenriquenavarro/Library/CloudStorage/OneDrive-IndianaUniversity/V506/Fall24"
```

```
# Here
v506_path <- here('/Users/luisenriquenavarro/Library/CloudStorage/OneDrive-IndianaUniversity/V506/Fall24
# You can create the paths for each folder with this object easily
lab2_path <- here(v506_path,'Lab2')
```

- 3. Load the autos data

```
# load the data using base R
filepath <- '/Users/luisenriquenavarro/Library/CloudStorage/OneDrive-IndianaUniversity/V506/Fall24/'
filename <- "Lab2/autodata.csv"
autos_path <- paste(filepath,filename, sep="")
autos <- read.csv(file = autos_path)
```

```
# using here and rio
autos <- import(file = here(lab2_path, 'autodata.csv'))
```

- 4. Structure of the data set

```
str(autos)
```

```
## 'data.frame':    128 obs. of  8 variables:
##  $ Car     : chr  "Acura Integra GS-R" "Acura Integra LS" "Acura NSX-T" "AM General Hummer" ...
##  $ Price   : int  20015 18560 86642 71760 36802 40960 22770 36015 54044 66837 ...
##  $ Time1   : num  7 7.5 5.2 18.1 8.3 9.7 7.8 7.7 5.7 8.4 ...
##  $ Time2   : num  15.6 15.9 13.8 21.2 16.5 17.2 16.1 16 14.3 16.6 ...
##  $ MaxSpeed: int  133 125 162 83 127 126 116 127 129 127 ...
##  $ Brake   : int  185 189 173 253 191 189 175 182 186 181 ...
##  $ MPG     : int  25 25 18 10 18 18 22 20 14 16 ...
##  $ Domestic: int  0 0 0 1 0 0 0 0 0 0 ...
```

```
summary(autos)
```

```
##      Car                Price            Time1            Time2
##  Length:128         Min.   : 12065   Min.   : 3.700   Min.   :12.30
##  Class :character   1st Qu.: 20811   1st Qu.: 6.675   1st Qu.:15.28
##  Mode  :character   Median : 27544   Median : 8.000   Median :16.30
##                     Mean   : 38636   Mean   : 7.990   Mean   :16.07
```

```
##                     3rd Qu.: 41128   3rd Qu.: 9.000   3rd Qu.:17.00
##                     Max.   :275056   Max.   :18.100   Max.   :21.20
##     MaxSpeed         Brake             MPG           Domestic
##  Min.   : 83.0   Min.   :150.0   Min.   : 9.00   Min.   :0.0000
##  1st Qu.:110.8   1st Qu.:174.8   1st Qu.:17.00   1st Qu.:0.0000
##  Median :126.5   Median :184.5   Median :18.00   Median :0.0000
##  Mean   :128.3   Mean   :184.5   Mean   :18.88   Mean   :0.3281
##  3rd Qu.:142.0   3rd Qu.:193.2   3rd Qu.:21.00   3rd Qu.:1.0000
##  Max.   :185.0   Max.   :253.0   Max.   :35.00   Max.   :1.0000
```

- 5. Create a New Variable

```
autos2 = autos$Price/autos$MPG
```

- 6. Combine the data sets

```
autos = cbind(autos,autos2)
names(autos)
```

```
## [1] "Car"      "Price"    "Time1"    "Time2"    "MaxSpeed" "Brake"    "MPG"
## [8] "Domestic" "autos2"
```

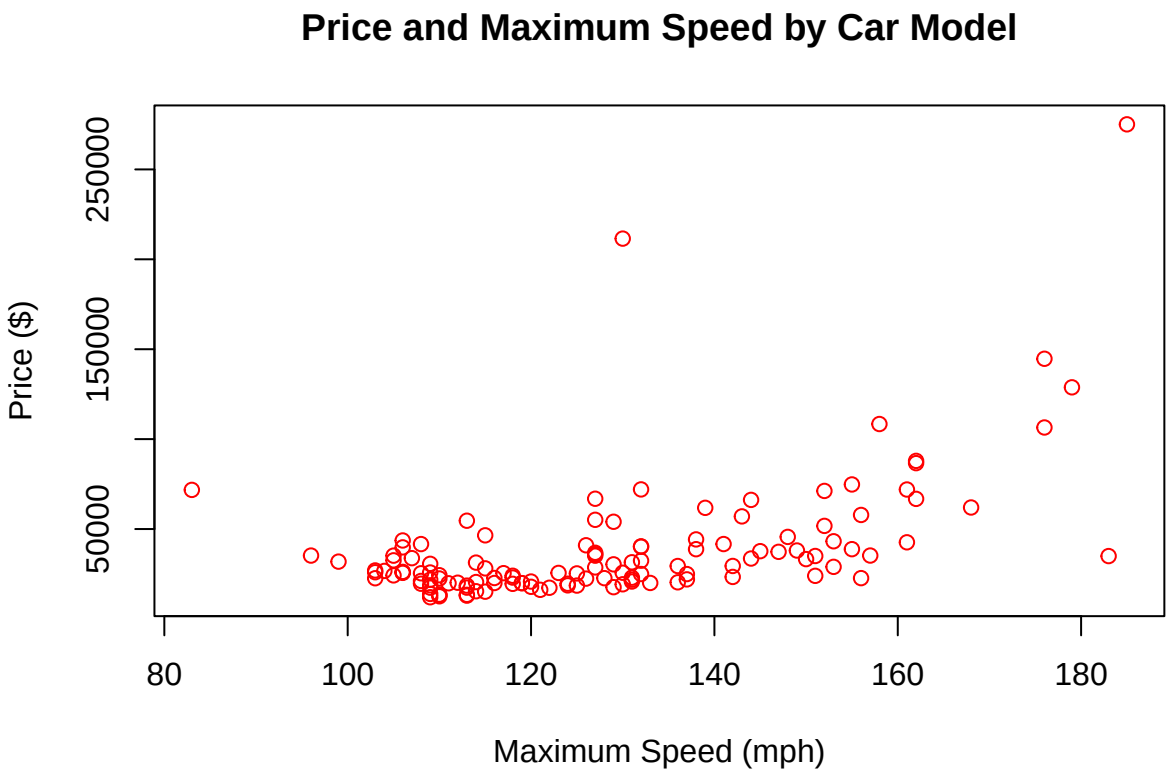
- 7. Filter Data sets

```
filter(autos, Price >= 100000)
```

```
subset(autos, Price >= 100000)
```

- 8. Scatter Plot

```
plot(autos$MaxSpeed, autos$Price, main = "Price and Maximum Speed by Car Model",
     ylab = "Price ($)", xlab="Maximum Speed (mph)", type="p", col = "red")
```



- 9. Summary of Statistics

```
mean(autos$MaxSpeed)
```

```
## [1] 128.3125
```

```
median(autos$MaxSpeed)
```

```
## [1] 126.5
```

```
var(autos$MaxSpeed)
```

```
## [1] 420.9252
```

```
sd(autos$MaxSpeed)
```

```
## [1] 20.51646
```

- 10. Export the data set

```r
autos2_df <- data.frame(price_mpg = autos2)

# use base R: note this code will save it on the address you set with setwd()
write.csv(autos2_df, file = "autos2.csv")

# Use rio
rio::export(autos2_df, file = here(lab2_path, 'autos2.csv'))
```