ECON 690 Training 1 Fall 2018

Mark Banghart

September 14, 2018

Set up a folder for the class

Create an econ690 folder.

Basics of using Rstudio

- RStudio IDE
- R Projects
  - Create an R project named "HW1" in the econ690 folder.
  - Close RStudio and then open your project again.
  - Move the DB1B_MARKET.csv file into this folder. This is not done in RStudio or R.

RMarkdown

- RMarkdown
  - Open sample pdf RMarkdown file and knit it. Pandoc may need to be installed on your laptop.
  - Paragraphs
  - Lists
  - Code chunks

Basics of R console, commands, data, and packages

- R Console
  - Prompts
  - Escape
  - Recalling commands
- Commands / Syntax
  - Assignment

    ```
    X <- 3.14
    ```

  - Comments

    ```
    # comments should state what you are doing and not how you are doing it
    ```

  - Functions and their parameters

    ```
    seq(1, 10)
    ```

    ```
    ##  [1]  1  2  3  4  5  6  7  8  9 10
    ```

```r
seq(to = 10, from = 1)
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10
```

```r
seq(1, 10, by = 2)
```

```
## [1] 1 3 5 7 9
```

- Importing data

  Create a .csv file and .RData file to import

```r
df <- data.frame(student = c("Maya", "Omar", "Emma", "Roy", "Aki", "Sara", "Jabari"),
                 math = c(21, 11, 27, 25, 23, 27, 25),
                 english = c(31, 38, 33, 26, 19, 10, 26)
                 )
write.table(df, file = "students.csv", sep = ",", row.names = FALSE)

save(mtcars, file = "mtcars.RData")
```

  - read.table()
    - View first few lines of a data file
    - Import the mtcars data set

```r
students <- read.table("students.csv", sep = ",", header = TRUE)

str(students)
```

```
## 'data.frame':    7 obs. of  3 variables:
##  $ student: Factor w/ 7 levels "Aki","Emma","Jabari",..: 4 5 2 6 1 7 3
##  $ math   : int  21 11 27 25 23 27 25
##  $ english: int  31 38 33 26 19 10 26
```

  - load()

```r
load("mtcars.RData")

str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

  - Data types and their structure
    - Subset by $

```r
mtcars$mpg
```

```
##  [1] 21.0 21.0 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4 17.3 15.2
```

```
## [15] 10.4 10.4 14.7 32.4 30.4 33.9 21.5 15.5 15.2 13.3 19.2 27.3 26.0 30.4
## [29] 15.8 19.7 15.0 21.4
```

* Subset by []

```
mtcars[1:5, c("mpg", "hp", "cyl")]
```

```
##                   mpg  hp cyl
## Mazda RX4         21.0 110   6
## Mazda RX4 Wag     21.0 110   6
## Datsun 710        22.8  93   4
## Hornet 4 Drive    21.4 110   6
## Hornet Sportabout 18.7 175   8
```

```
mtcars$mpg[1:5]
```

```
## [1] 21.0 21.0 22.8 21.4 18.7
```

– Loading and installing

*library()

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------- t
## v ggplot2 3.0.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.6
## v tidyr   0.8.1     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0
## -- Conflicts ------------------------------------------------------------------------- tidyvers
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(knitr)
```

Tidyverse introduction

```
* Tibble - another kind of data.frame

* Make new variables

    mutate()


    ```r
    mtcars <- mutate(mtcars, log_mpg = log(mpg))

    str(mtcars)
    ```

    ```
    ## 'data.frame':    32 obs. of  12 variables:
    ##  $ mpg    : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
    ##  $ cyl    : num  6 6 4 6 8 6 8 4 4 6 ...
    ##  $ disp   : num  160 160 108 258 360 ...
    ##  $ hp     : num  110 110 93 110 175 105 245 62 95 123 ...
```
```

```
## $ drat   : num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
## $ wt     : num  2.62 2.88 2.32 3.21 3.44 ...
## $ qsec   : num  16.5 17 18.6 19.4 17 ...
## $ vs     : num  0 0 1 1 0 1 0 1 1 1 ...
## $ am     : num  1 1 1 0 0 0 0 0 0 0 ...
## $ gear   : num  4 4 4 3 3 3 3 4 4 4 ...
## $ carb   : num  4 4 1 1 2 1 4 2 2 4 ...
## $ log_mpg: num  3.04 3.04 3.13 3.06 2.93 ...
```

* Subsetting

    filter()

    ```r
    gasHogs <- filter(mtcars, mpg > mean(mpg))

    str(gasHogs)
    ```

    ```
    ## 'data.frame':    14 obs. of  12 variables:
    ## $ mpg    : num  21 21 22.8 21.4 24.4 22.8 32.4 30.4 33.9 21.5 ...
    ## $ cyl    : num  6 6 4 6 4 4 4 4 4 4 ...
    ## $ disp   : num  160 160 108 258 147 ...
    ## $ hp     : num  110 110 93 110 62 95 66 52 65 97 ...
    ## $ drat   : num  3.9 3.9 3.85 3.08 3.69 3.92 4.08 4.93 4.22 3.7 ...
    ## $ wt     : num  2.62 2.88 2.32 3.21 3.19 ...
    ## $ qsec   : num  16.5 17 18.6 19.4 20 ...
    ## $ vs     : num  0 0 1 1 1 1 1 1 1 1 ...
    ## $ am     : num  1 1 1 0 0 0 1 1 1 0 ...
    ## $ gear   : num  4 4 4 3 4 4 4 4 4 3 ...
    ## $ carb   : num  4 4 1 1 2 2 1 2 1 1 ...
    ## $ log_mpg: num  3.04 3.04 3.13 3.06 3.19 ...
    ```

    select()

    ```r
    mtcars <- select(mtcars, -log_mpg)

    str(mtcars)
    ```

    ```
    ## 'data.frame':    32 obs. of  11 variables:
    ## $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
    ## $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
    ## $ disp: num  160 160 108 258 360 ...
    ## $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
    ## $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
    ## $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
    ```

```
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

* Pipe operator

```r
mtcars <-
    mtcars %>%
    mutate(gasHog = mpg > mean(mpg))

str(mtcars)
```

```
## 'data.frame':    32 obs. of  12 variables:
##  $ mpg   : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl   : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp  : num  160 160 108 258 360 ...
##  $ hp    : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat  : num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt    : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec  : num  16.5 17 18.6 19.4 17 ...
##  $ vs    : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am    : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear  : num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb  : num  4 4 1 1 2 1 4 2 2 4 ...
##  $ gasHog: logi  TRUE TRUE TRUE TRUE FALSE FALSE ...
```

* grouping data

    group_by() and ungroup()

```r
mtcars <-
    mtcars %>%
    group_by(cyl, gear) %>%
    mutate(groupSize = n()) %>%
    ungroup()

str(mtcars)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    32 obs. of  13 variables:
##  $ mpg      : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl      : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp     : num  160 160 108 258 360 ...
##  $ hp       : num  110 110 93 110 175 105 245 62 95 123 ...
```

```
##  $ drat     : num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt       : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec     : num  16.5 17 18.6 19.4 17 ...
##  $ vs       : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am       : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear     : num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb     : num  4 4 1 1 2 1 4 2 2 4 ...
##  $ gasHog   : logi  TRUE TRUE TRUE TRUE FALSE FALSE ...
##  $ groupSize: int  4 4 8 2 12 2 12 8 8 4 ...
```

```r
summary(mtcars$groupSize)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   3.500   8.000   7.438  12.000  12.000
```

* summarise

```r
gasHogSum <-
    mtcars %>%
    group_by(gasHog) %>%
    summarise(
      cyl = mean(cyl),
      disp = mean(disp),
      hp = mean(hp),
      wt = mean(wt)
    )

str(gasHogSum)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    2 obs. of  5 variables:
##  $ gasHog: logi  FALSE TRUE
##  $ cyl   : num  7.56 4.43
##  $ disp  : num  314 124
##  $ hp    : num  191.9 88.5
##  $ wt    : num  3.84 2.42
```

```r
kable(gasHogSum)
```

gasHog | cyl | disp | hp | wt

```
-------  ---------  ---------  ---------  ---------
FALSE      7.555556   313.8111   191.9444   3.838833
TRUE       4.428571   123.8929    88.5000   2.418071
```

Graphs

- ggplot()
  - Geometries
  - Aesthetics
  - Themes

```
ggplot(data = mtcars, aes(x = hp, y = mpg)) +
  geom_point() +
  ggtitle("Plot of car data") +
  xlab("Horse Power") +
  ylab("Miles Per Gallon") +
  theme_bw()
```