

Review II of DS 1: Extraction, Manipulation, Visualization (EMV)

1. Extraction

1a. Read Excel File

```
library(readxl)
setwd("C:/Users/james/Desktop/DS_Summer_2020")
amazon_world_df = read_xlsx('games.xlsx', sheet = 'amazon_world')
```

1b. Read Wikipedia Tables

```
library(htmllab)
billboard_df = htmllab("https://en.wikipedia.org/wiki/List_of_Billboard_Hot_100_number-one_singles_of_2020",1)
```

2. Manipulation

2a. Select

```
library(tidyverse)
mtcars %>%
  select(vs,mpg)
```

2b. Mutate

```
mtcars %>%
  mutate(categ_mpg = case_when(
    mpg <= 15 ~ 'low_mpg',
    mpg > 15 & mpg < 21 ~ 'med_mpg',
    mpg > 21 ~ 'hi_mpg'
  ))
```

2c. Filter

```
mtcars %>%
  filter(vs == 0)
```

3. Visualization

3a. Bar Plot

```
mtcars %>%
  ggplot(aes(x = as.character(vs))) +
  geom_bar(stat = 'count')
```

3b. Boxplot

```
mtcars %>%
  ggplot(aes(y = mpg)) +
  geom_boxplot()
```

3c. Scatter Plot

```
mtcars %>%  
  ggplot(aes(x = mpg, y = wt)) +  
  geom_point()
```