Introduction to data manipulation

Homework week 3

Solution

15 September 2020

pacman::p\_load(tidyverse)

## Data structure

str(mpg)

## Classes 'tbl\_df', 'tbl' and 'data.frame': 234 obs. of 11 variables:  
## $ manufacturer: chr "audi" "audi" "audi" "audi" ...  
## $ model : chr "a4" "a4" "a4" "a4" ...  
## $ displ : num 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...  
## $ year : int 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...  
## $ cyl : int 4 4 4 4 6 6 6 4 4 4 ...  
## $ trans : chr "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...  
## $ drv : chr "f" "f" "f" "f" ...  
## $ cty : int 18 21 20 21 16 18 18 18 16 20 ...  
## $ hwy : int 29 29 31 30 26 26 27 26 25 28 ...  
## $ fl : chr "p" "p" "p" "p" ...  
## $ class : chr "compact" "compact" "compact" "compact" ...

## Summary table

### Summarization script

mpg2 <-   
 mpg %>%  
 rename(drive = drv) %>%   
 mutate(drive= recode(drive, "4"="Four-wheel",   
 "f"="Front-wheel",   
 "r"="Rear-wheel"))   
means <-   
 mpg2 %>%   
 group\_by(drive) %>%  
 summarize(mean\_hwy = mean(hwy))

### Print the table

Either simply call *means* to see what the object contains:

means

## # A tibble: 3 x 2  
## drive mean\_hwy  
## <chr> <dbl>  
## 1 Four-wheel 19.2  
## 2 Front-wheel 28.2  
## 3 Rear-wheel 21

Or give it some formatting with the *kable* function in **knitr** package:

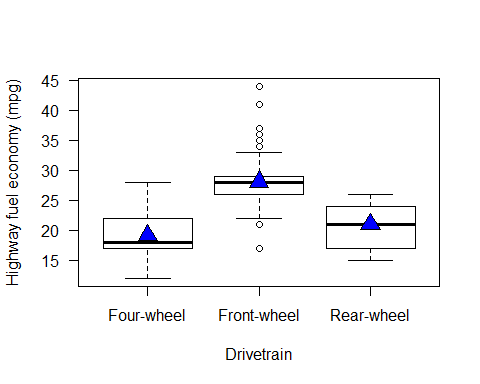
knitr::kable(means, caption = "Easy to make a nice table.")

Easy to make a nice table.

|  |  |
| --- | --- |
| drive | mean\_hwy |
| Four-wheel | 19.17476 |
| Front-wheel | 28.16038 |
| Rear-wheel | 21.00000 |

## Boxplot

boxplot(hwy ~ drive, mpg2, las = 1,   
 xlab = "Drivetrain",   
 ylab = "Highway fuel economy (mpg)")  
 points(mean\_hwy ~ as.factor(drive), means,   
 pch=24, bg="blue", cex=2)



Boxplot with means added.