Homework 4

Thu Nguyen

12 May, 2019

Libraries

```
library(ggplot2, lib='C:/temp')
library(tidyr, lib='C:/temp')
library(tidyverse, lib='C:/temp')
```

Problem 1

Write a function `boot.t.test(x, y, conf = 0.95, B = 9999)` modeled after the built-in function `t.test`, but instead based on a Studentized bootstrap approach rather than a Student t-distribution for the computation of quantiles (that go in the confidence interval) and the p-value. Apply your function to the following dataset.

`boot.t.test(x, y, conf = 0.95, B = 9999)` function

```
boot.t.test = function(x, y, conf = 0.95, B = 9999){
  # Extract lengths of samples
  m <- length(x)
  n <- length(y)

  # Standard Error from Original samples
  sehat <- sqrt((var(x)/m) + (var(y)/n))
  # Differeces from Original samples
  diff <- mean(x) - mean(y)

  # vector to store permuted t-ratio
  t_ratio <- numeric(B)
  for(b in 1:B){
    # Bootstrap the new samples from the Original samples with replacement
    x_bs <- sample(x, m, replace = TRUE)
    y_bs <- sample(y, n, replace = TRUE)
    # Current bootstrap's t-ratio
    t_ratio[b] = (mean(x_bs) - mean(y_bs) - diff)/sqrt( (var(x_bs)/m) + (var(y_bs)/n) )
  }

  # Re-order t-ratios, then take the (conf_level)^th quantile
  t_alpha_bs <- as.numeric(quantile(t_ratio, conf))

  cat("The one-sided ", conf, " confidence interval is: ", (diff - t_alpha_bs*sehat), " to infinity")
}
```

Apply on the dataset.

```r
# Prepare data
data <- read.table('memory.txt', header=TRUE, sep='\t', fill = FALSE)
gingko <- data$Gingko
placebo <- as.numeric(na.omit(data$Placebo))
# Apply function
boot.t.test(gingko, placebo, conf = .95, B = 9999)
```

```
The one-sided  0.95  confidence interval is:  -2.058273  to infinity
```

Problem 2

Problem 2. Examine the following dataset on various car models and some of their characteristics. Our goal here is to compare gas consumption between automatic and manual transmissions. Do the following for City MPG and for Highway MPG, separately.

A. Group cars according to their transmission type, either automatic or manual. (This requires some work, as the Transmission variable provides additional details.) Produce side-by-side violin plots of gas consumption according to this grouping. Then apply a test for comparing the two groups.
B. Refine the analysis by considering a finer classification in terms of transmission type. First produce side-by-side violin plots. Then apply a test for comparing the various groups.

`pertest(x, y, B)` function to hypothesis tests later

```
# Permutation test
pertest <- function(x,y,B) {
  nx <- length(x)
  ny <- length(y)
  testorig <- mean(x) - mean(y)
  z <- c(x,y)
  n <- length(z)
  count <- 0
  for (b in 1:B) {
    temp <- sample(z, n, replace = FALSE)
    xtemp <- temp[1:nx]
    ytemp <- temp[(nx+1):n]
    meantemp <- mean(xtemp) - mean(ytemp)
    if (meantemp >= testorig) {
      count <- count + 1
    }
  }
  return( round((count+1)/(B+1), 5) )
}
```

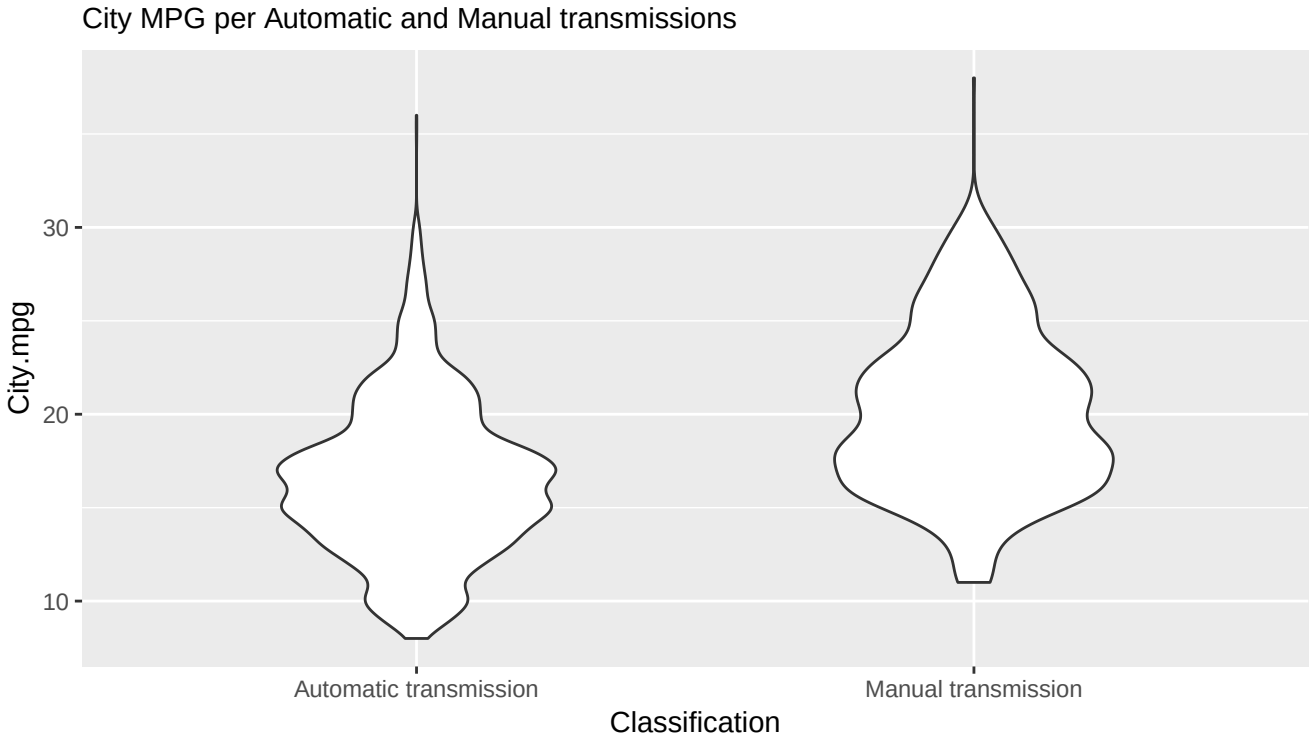
```
# Load data
cars = read.csv('cars.csv', header=TRUE, sep=",", fill=FALSE)
```

Part A

A.1. City MPG

Violin plot

```
cars %>%
  group_by(Classification) %>%
  select(City.mpg, Classification) %>%
  ggplot(aes(x = Classification, y = City.mpg)) +
  geom_violin(scale = 'width', adjust = 1, width = .5) +
  labs(subtitle = 'City MPG per Automatic and Manual transmissions')
```



Prepare Data for Hypothesis tests

```
auto <- cars %>%
  filter(Classification == "Automatic transmission") %>%
  select(City.mpg, Highway.mpg)
manual <- cars %>%
  filter(Classification == "Manual transmission") %>%
  select(City.mpg, Highway.mpg)
```

Hypothesis Test via Permutation Test

$$H_0 : A_C \sim M_C$$
$$H_1 : A_C \geq^{sto} M_C$$

where A_C is the City MPG for automatic transmission, and M_C is for manual transmission.

```
pertest(auto[,1], manual[,1], B = 2000)
```

```
[1] 1
```

$$H_0 : M_C \sim A_C$$
$$H_1 : M_C \geq^{sto} A_C$$

```
pertest(manual[,1], auto[,1], B = 2000)
```

```
[1] 5e-04
```

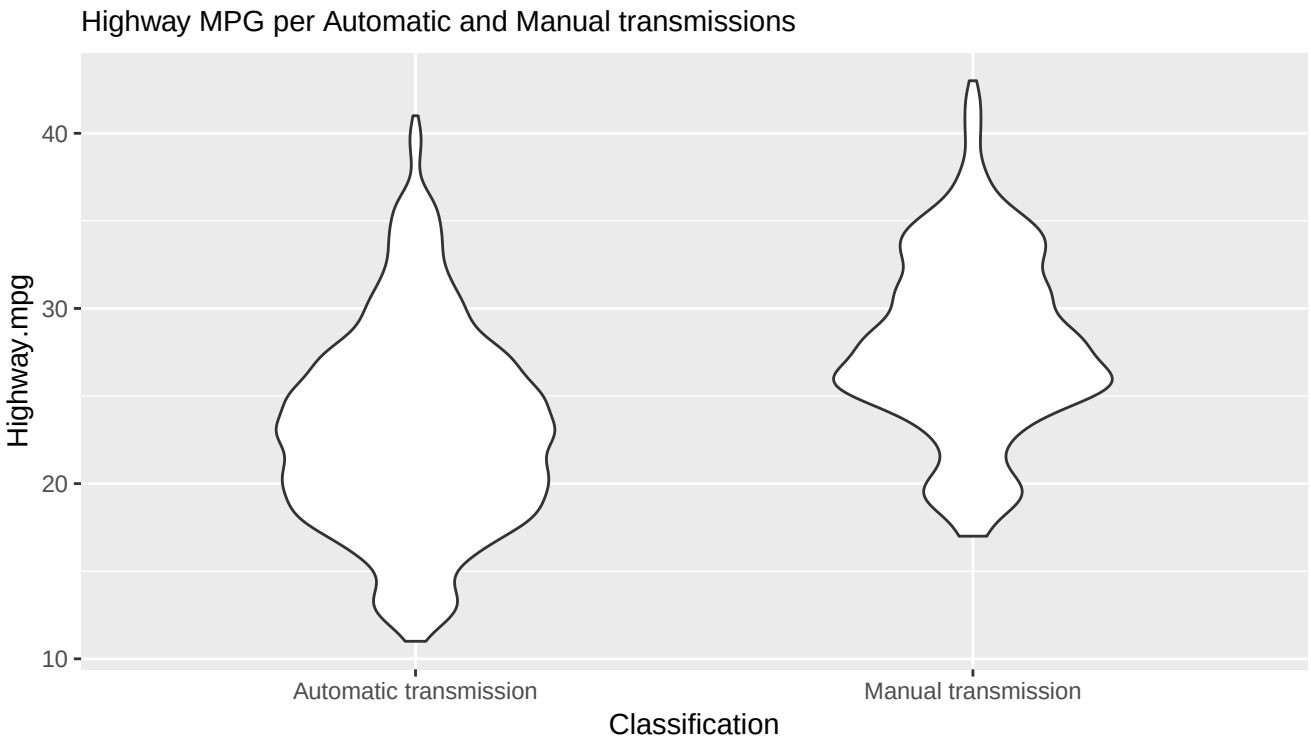
A.2. Highway MPG

In the data for Highway MPG, there is an entry of value 223, which is likely to be an outlier, and might have been entered by mistake. As such, I have decided to replace it with a more reasonable value of 23.

```
cars$Highway.mpg[which.max(cars$Highway.mpg)] <- 23
```

Violin plot

```
cars %>%
  group_by(Classification) %>%
  select(Highway.mpg, Classification) %>%
  ggplot(aes(x = Classification, y = Highway.mpg)) +
  geom_violin(scale = 'width', adject = 1, width = .5) +
  labs(subtitle = 'Highway MPG per Automatic and Manual transmissions')
```



Hypothesis Test via Permutation Test

$$H_0 : A_H \sim M_H$$
$$H_1 : A_H \geq^{sto} M_H$$

where A_H is the Highway MPG for automatic transmission, and M_H is for manual transmission.

```
pertest(auto[,2], manual[,2], B = 2000)
```

```
[1] 1
```

$$H_0 : M_H \sim A_H$$
$$H_1 : M_H \geq^{sto} A_H$$

```
pertest(manual[,2], auto[,2], B = 2000)
```

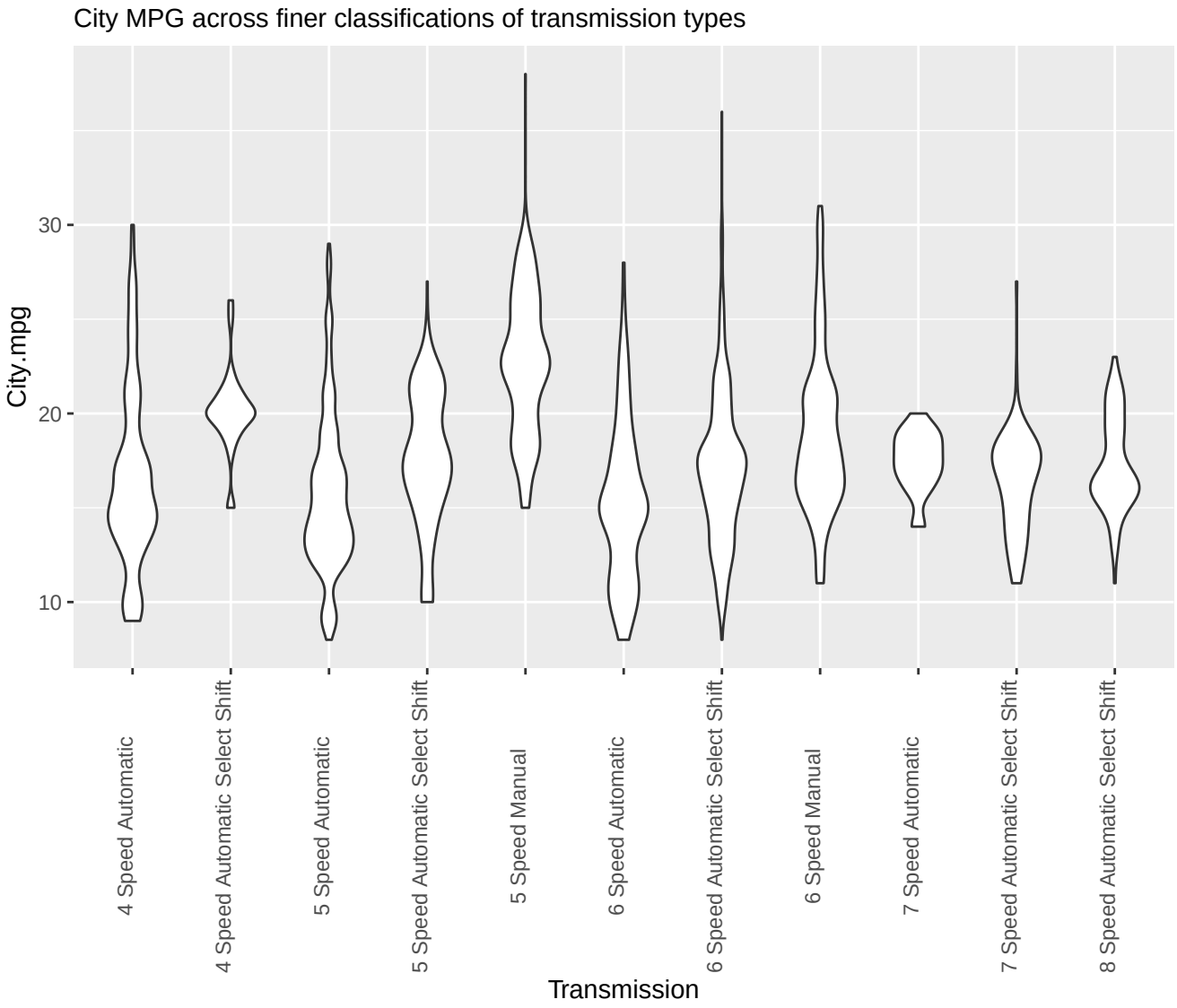
```
[1] 5e-04
```

Part B: Finer Classifications

B.1. City MPG

Violin plot

```
cars %>%
  group_by(Transmission) %>%
  select(City.mpg, Transmission) %>%
  ggplot(aes(x = Transmission, y = City.mpg)) +
  geom_violin(scale = 'width', adjust = 1, width = .5) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0)) +
  labs(subtitle = 'City MPG across finer classifications of transmission types')
```



Prepare Data for Hypothesis tests for finer classifications

```
auto_4s <- cars %>%
  filter(Transmission == '4 Speed Automatic') %>%
  select(City.mpg, Highway.mpg)
auto_5s <- cars %>%
```

```
  filter(Transmission == '5 Speed Automatic') %>%
  select(City.mpg, Highway.mpg)
auto_6s <- cars %>%
  filter(Transmission == '6 Speed Automatic') %>%
  select(City.mpg, Highway.mpg)
auto_7s <- cars %>%
  filter(Transmission == '7 Speed Automatic') %>%
  select(City.mpg, Highway.mpg)
auto_4s_ss <- cars %>%
  filter(Transmission == '4 Speed Automatic Select Shift') %>%
  select(City.mpg, Highway.mpg)
auto_5s_ss <- cars %>%
  filter(Transmission == '5 Speed Automatic Select Shift') %>%
  select(City.mpg, Highway.mpg)
auto_6s_ss <- cars %>%
  filter(Transmission == '6 Speed Automatic Select Shift') %>%
  select(City.mpg, Highway.mpg)
auto_7s_ss <- cars %>%
  filter(Transmission == '7 Speed Automatic Select Shift') %>%
  select(City.mpg, Highway.mpg)
auto_8s_ss <- cars %>%
  filter(Transmission == '8 Speed Automatic Select Shift') %>%
  select(City.mpg, Highway.mpg)
manual_5s <- cars %>%
  filter(Transmission == '5 Speed Manual') %>%
  select(City.mpg, Highway.mpg)
manual_6s <- cars %>%
  filter(Transmission == '6 Speed Manual') %>%
  select(City.mpg, Highway.mpg)
```

Hypothesis Test via Permutation Test

Let:

- $A_C^{(i)}$ for $i \in \{4, 5, 6, 7\}$ be the City MPG for i Speed Automatic transmission,
- $A_C^{(i_{ss})}$ for $i \in \{4, 5, 6, 7, 8\}$ be the City MPG for i Speed Automatic Select Shift transmission, and
- $M_C^{(i)}$ for $i \in \{5, 6\}$ be the City MPG for i Speed Manual transmission.

Test for various groups.

Test 1: City MPG: 4 Speed Automatic vs. 5 Speed Automatic

$$H_0 : A_C^{(4)} \sim A_C^{(5)}$$

$$H_1 : A_C^{(4)} \geq^{sto} A_C^{(5)}$$

```
p <- pertest(auto_4s[,1], auto_5s[,1], B = 2000)
print(paste0('p-value for 4 Speed Automatic vs. 5 Speed Automatic: ', p))
```

```
[1] "p-value for 4 Speed Automatic vs. 5 Speed Automatic: 5e-04"
```

Test 2: City MPG: 5 Speed Automatic vs. 6 Speed Automatic

$$H_0 : A_C^{(5)} \sim A_C^{(6)}$$

$$H_1 : A_C^{(5)} \geq^{sto} A_C^{(6)}$$

```
p <- pertest(auto_5s[,1], auto_6s[,1], B = 2000)
print(paste0('p-value for 5 Speed Automatic vs. 6 Speed Automatic: ', p))
```

```
[1] "p-value for 5 Speed Automatic vs. 6 Speed Automatic: 0.01799"
```

Test 3: City MPG: 6 Speed Automatic vs. 7 Speed Automatic

$$H_0 : A_C^{(6)} \sim A_C^{(7)}$$

$$H_1 : A_C^{(6)} \geq^{sto} A_C^{(7)}$$

```
p <- pertest(auto_6s[,1], auto_7s[,1], B = 2000)
print(paste0('p-value for 6 Speed Automatic vs. 7 Speed Automatic: ', p))
```

```
[1] "p-value for 6 Speed Automatic vs. 7 Speed Automatic: 0.999"
```

Test 4: City MPG: 4 Speed Automatic Select Shift vs. 4 Speed Automatic

$$H_0 : A_C^{(4_{ss})} \sim A_C^{(4)}$$

$$H_1 : A_C^{(4_{ss})} \geq^{sto} A_C^{(4)}$$

```
p <- pertest(auto_4s_ss[,1], auto_4s[,1], B = 2000)
print(paste0('p-value for 4 Speed Automatic Select Shift vs. 4 Speed Automatic: ', p))
```

```
[1] "p-value for 4 Speed Automatic Select Shift vs. 4 Speed Automatic: 5e-04"
```

Test 5: City MPG: 5 Speed Automatic Select Shift vs. 5 Speed Automatic

$$H_0 : A_C^{(5_{ss})} \sim A_C^{(5)}$$

$$H_1 : A_C^{(5_{ss})} \geq^{sto} A_C^{(5)}$$

```
p <- pertest(auto_5s_ss[,1], auto_5s[,1], B = 2000)
print(paste0('p-value for 5 Speed Automatic Select Shift vs. 5 Speed Automatic: ', p))
```

```
[1] "p-value for 5 Speed Automatic Select Shift vs. 5 Speed Automatic: 5e-04"
```

Test 6: City MPG: 6 Speed Automatic Select Shift vs. 6 Speed Automatic

$$H_0 : A_C^{(6_{ss})} \sim A_C^{(6)}$$

$$H_1 : A_C^{(6_{ss})} \geq^{sto} A_C^{(6)}$$

```
p <- pertest(auto_6s_ss[,1], auto_6s[,1], B = 2000)
print(paste0('p-value for 6 Speed Automatic Select Shift vs. 6 Speed Automatic: ', p))
```

```
[1] "p-value for 6 Speed Automatic Select Shift vs. 6 Speed Automatic: 5e-04"
```

Test 7: City MPG: 7 Speed Automatic Select Shift vs. 7 Speed Automatic

$$H_0 : A_C^{(7_{ss})} \sim A_C^{(7)}$$

$$H_1 : A_C^{(7_{ss})} \geq^{sto} A_C^{(7)}$$

```
p <- pertest(auto_7s_ss[,1], auto_7s[,1], B = 2000)
print(paste0('p-value for 7 Speed Automatic Select Shift vs. 7 Speed Automatic: ', p))
```

```
[1] "p-value for 7 Speed Automatic Select Shift vs. 7 Speed Automatic: 0.997"
```

Test 8: City MPG: 5 Speed Automatic vs. 5 Speed Manual

$$H_0 : A_C^{(5)} \sim A_M^{(5)}$$

$$H_1 : A_C^{(5)} \geq^{sto} A_M^{(5)}$$

```
p <- pertest(auto_5s[,1], manual_5s[,1], B = 2000)
print(paste0('p-value for 5 Speed Automatic vs. 5 Speed Manual: ', p))
```

```
[1] "p-value for 5 Speed Automatic vs. 5 Speed Manual: 1"
```

Test 9: City MPG: 6 Speed Automatic vs. 6 Speed Manual

$$H_0 : A_C^{(6)} \sim A_M^{(6)}$$

$$H_1 : A_C^{(6)} \geq^{sto} A_M^{(6)}$$

```
p <- pertest(auto_6s[,1], manual_6s[,1], B = 2000)
print(paste0('p-value for 6 Speed Automatic vs. 6 Speed Manual: ', p))
```

```
[1] "p-value for 6 Speed Automatic vs. 6 Speed Manual: 1"
```

Test 10: City MPG: 5 Speed Automatic Select Shift vs. 5 Speed Manual

$$H_0 : A_C^{(5_{ss})} \sim A_M^{(5)}$$

$$H_1 : A_C^{(5_{ss})} \geq^{sto} A_M^{(5)}$$

```
p <- pertest(auto_5s_ss[,1], manual_5s[,1], B = 2000)
print(paste0('p-value for 5 Speed Automatic Select Shift vs. 5 Speed Manual: ', p))
```

```
[1] "p-value for 5 Speed Automatic Select Shift vs. 5 Speed Manual: 1"
```

Test 11: City MPG: 6 Speed Automatic Select Shift vs. 6 Speed Manual

$$H_0 : A_C^{(6_{ss})} \sim A_M^{(6)}$$

$$H_1 : A_C^{(6_{ss})} \geq^{sto} A_M^{(6)}$$

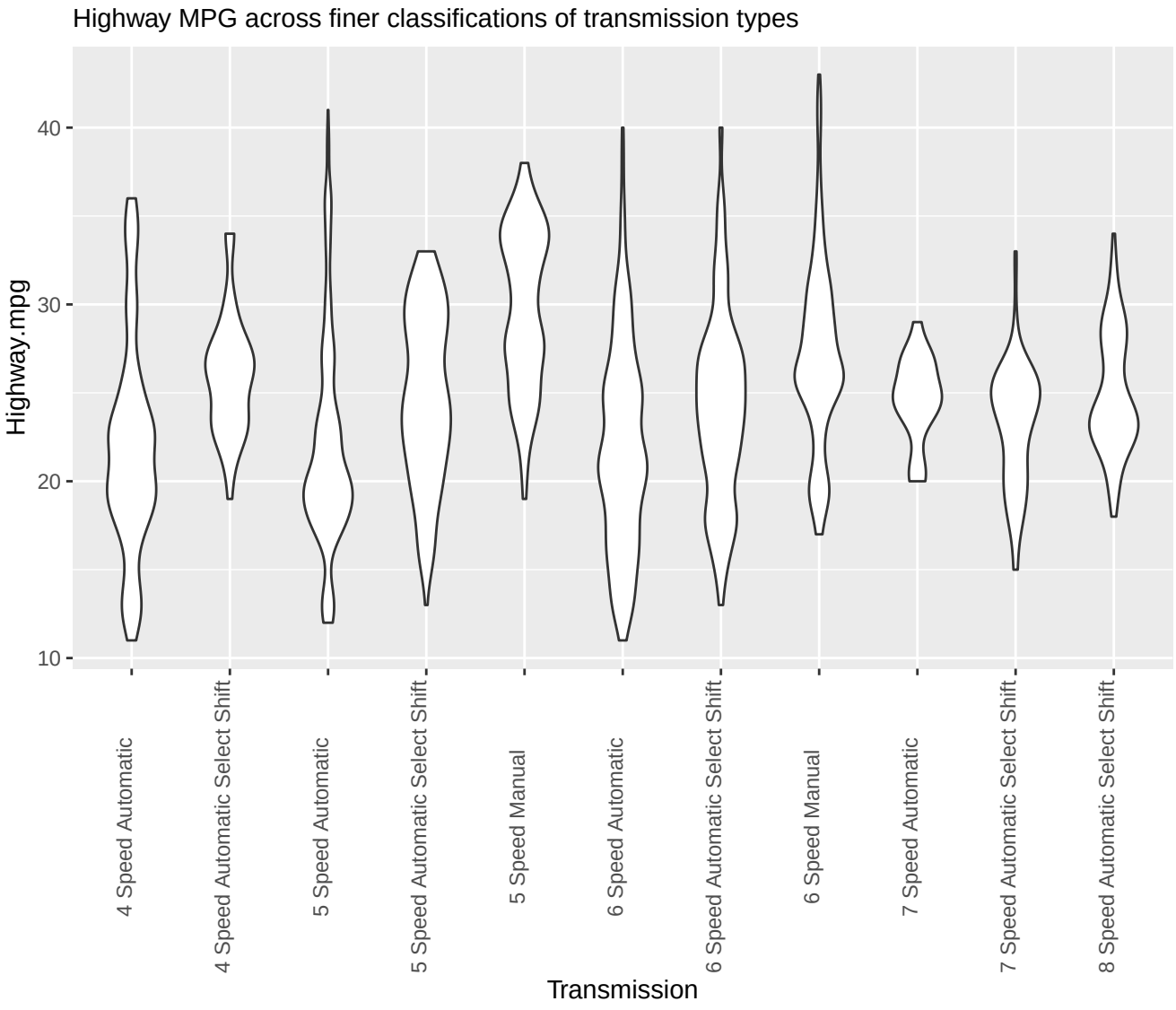
```
p <- pertest(auto_6s[,1], manual_6s[,1], B = 2000)
print(paste0('p-value for 6 Speed Automatic Select Shift vs. 6 Speed Manual: ', p))
```

```
[1] "p-value for 6 Speed Automatic Select Shift vs. 6 Speed Manual: 1"
```

B.2. Highway MPG

Violin plot

```
cars %>%
  group_by(Transmission) %>%
  select(Highway.mpg, Transmission) %>%
  ggplot(aes(x = Transmission, y = Highway.mpg)) +
  geom_violin(scale = 'width', adjust = 1, width = .5) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0)) +
  labs(subtitle = 'Highway MPG across finer classifications of transmission types')
```



Hypothesis Test via Permutation Test

Let:

- $A_H^{(i)}$ for $i \in \{4, 5, 6, 7\}$ be the Highway MPG for i Speed Automatic transmission,
- $A_H^{(i_{ss})}$ for $i \in \{4, 5, 6, 7, 8\}$ be the Highway MPG for i Speed Automatic Select Shift transmission, and
- $M_H^{(i)}$ for $i \in \{5, 6\}$ be the Highway MPG for i Speed Manual transmission.

Test for various groups.

Test 1: Highway MPG: 4 Speed Automatic vs. 5 Speed Automatic

$$H_0 : A_H^{(4)} \sim A_H^{(5)}$$

$$H_1 : A_H^{(4)} \geq^{sto} A_H^{(5)}$$

```
p <- pertest(auto_4s[,2], auto_5s[,2], B = 2000)
print(paste0('p-value for 4 Speed Automatic vs. 5 Speed Automatic: ', p))
```

```
[1] "p-value for 4 Speed Automatic vs. 5 Speed Automatic: 0.34283"
```

Test 2: Highway MPG: 5 Speed Automatic vs. 6 Speed Automatic

$$H_0 : A_H^{(5)} \sim A_H^{(6)}$$

$$H_1 : A_H^{(5)} \geq^{sto} A_H^{(6)}$$

```
p <- pertest(auto_5s[,2], auto_6s[,2], B = 2000)
print(paste0('p-value for 5 Speed Automatic vs. 6 Speed Automatic: ', p))
```

```
[1] "p-value for 5 Speed Automatic vs. 6 Speed Automatic: 0.58171"
```

Test 3: Highway MPG: 6 Speed Automatic vs. 7 Speed Automatic

$$H_0 : A_H^{(6)} \sim A_H^{(7)}$$

$$H_1 : A_H^{(6)} \geq^{sto} A_H^{(7)}$$

```
p <- pertest(auto_6s[,2], auto_7s[,2], B = 2000)
print(paste0('p-value for 6 Speed Automatic vs. 7 Speed Automatic: ', p))
```

```
[1] "p-value for 6 Speed Automatic vs. 7 Speed Automatic: 0.9995"
```

Test 4: Highway MPG: 4 Speed Automatic Select Shift vs. 4 Speed Automatic

$$H_0 : A_H^{(4_{ss})} \sim A_H^{(4)}$$

$$H_1 : A_H^{(4_{ss})} \geq^{sto} A_H^{(4)}$$

```
p <- pertest(auto_4s_ss[,2], auto_4s[,2], B = 2000)
print(paste0('p-value for 4 Speed Automatic Select Shift vs. 4 Speed Automatic: ', p))
```

```
[1] "p-value for 4 Speed Automatic Select Shift vs. 4 Speed Automatic: 5e-04"
```

Test 5: Highway MPG: 5 Speed Automatic Select Shift vs. 5 Speed Automatic

$$H_0 : A_H^{(5_{ss})} \sim A_H^{(5)}$$

$$H_1 : A_H^{(5_{ss})} \geq^{sto} A_H^{(5)}$$

```
p <- pertest(auto_5s_ss[,2], auto_5s[,2], B = 2000)
print(paste0('p-value for 5 Speed Automatic Select Shift vs. 5 Speed Automatic: ', p))
```

```
[1] "p-value for 5 Speed Automatic Select Shift vs. 5 Speed Automatic: 5e-04"
```

Test 6: Highway MPG: 6 Speed Automatic Select Shift vs. 6 Speed Automatic

$$H_0 : A_H^{(6_{ss})} \sim A_H^{(6)}$$

$$H_1 : A_H^{(6_{ss})} \geq^{sto} A_H^{(6)}$$

```
p <- pertest(auto_6s_ss[,2], auto_6s[,2], B = 2000)
print(paste0('p-value for 6 Speed Automatic Select Shift vs. 6 Speed Automatic: ', p))
```

```
[1] "p-value for 6 Speed Automatic Select Shift vs. 6 Speed Automatic: 5e-04"
```

Test 7: Highway MPG: 7 Speed Automatic Select Shift vs. 7 Speed Automatic

$$H_0 : A_H^{(7_{ss})} \sim A_H^{(7)}$$

$$H_1 : A_H^{(7_{ss})} \geq^{sto} A_H^{(7)}$$

```
p <- pertest(auto_7s_ss[,2], auto_7s[,2], B = 2000)
print(paste0('p-value for 7 Speed Automatic Select Shift vs. 7 Speed Automatic: ', p))
```

```
[1] "p-value for 7 Speed Automatic Select Shift vs. 7 Speed Automatic: 0.9975"
```

Test 8: Highway MPG: 5 Speed Automatic vs. 5 Speed Manual

$$H_0 : A_H^{(5)} \sim M_H^{(5)}$$

$$H_1 : A_H^{(5)} \geq^{sto} M_H^{(5)}$$

```
p <- pertest(auto_5s[,2], manual_5s[,2], B = 2000)
print(paste0('p-value for 5 Speed Automatic vs. 5 Speed Manual: ', p))
```

```
[1] "p-value for 5 Speed Automatic vs. 5 Speed Manual: 1"
```

Test 9: Highway MPG: 6 Speed Automatic vs. 6 Speed Manual

$$H_0 : A_H^{(6)} \sim M_H^{(6)}$$

$$H_1 : A_H^{(6)} \geq^{sto} M_H^{(6)}$$

```
p <- pertest(auto_6s[,2], manual_6s[,2], B = 2000)
print(paste0('p-value for 6 Speed Automatic vs. 6 Speed Manual: ', p))
```

```
[1] "p-value for 6 Speed Automatic vs. 6 Speed Manual: 1"
```

Test 10: Highway MPG: 5 Speed Automatic Select Shift vs. 5 Speed Manual

$$H_0 : A_H^{(5_{ss})} \sim M_H^{(5)}$$

$$H_1 : A_H^{(5_{ss})} \geq^{sto} M_H^{(5)}$$

```
p <- pertest(auto_5s_ss[,2], manual_5s[,2], B = 2000)
print(paste0('p-value for 5 Speed Automatic Select Shift vs. 5 Speed Manual: ', p))
```

```
[1] "p-value for 5 Speed Automatic Select Shift vs. 5 Speed Manual: 1"
```

Test 11: Highway MPG: 6 Speed Automatic Select Shift vs. 6 Speed Manual

$$H_0 : A_H^{(6_{ss})} \sim M_H^{(6)}$$

$$H_1 : A_H^{(6_{ss})} \geq^{sto} M_H^{(6)}$$

```
p <- pertest(auto_6s[,2], manual_6s[,2], B = 2000)
print(paste0('p-value for 6 Speed Automatic Select Shift vs. 6 Speed Manual: ', p))
```

```
[1] "p-value for 6 Speed Automatic Select Shift vs. 6 Speed Manual: 1"
```