problemset2_506

JUNYI ZHANG

Link to GitHub

https://github.com/juny1z/Problemset2.git

Probelm 1

(a). Four versions of the dice game:

```
#loop version
dice_loop <- function(n) {
  wins <- 0

  for (i in 1:n) {
    roll <- sample(1:6, 1)

    if (roll == 3 || roll == 5) {
      wins <- wins + (2 * roll) - 2
    } else {
      wins <- wins - 2
    }
  }

  return(wins)
}

#vectorized version
dice_vectorized <- function(n) {
  roll <- sample(1:6, n, replace = TRUE)
  wins <- ifelse(roll == 3 | roll == 5, (2 * roll)-2, -2)
  return(sum(wins))
```

```
}

#table version
dice_table <- function(n) {
  roll <- sample(1:6, n, replace = TRUE)
  roll_counts <- table(factor(roll, levels = 1:6))
  num_3 <- ifelse(is.na(roll_counts[as.character(3)]), 0, roll_counts[as.character(3)])
  num_5 <- ifelse(is.na(roll_counts[as.character(5)]), 0, roll_counts[as.character(5)])
  wins <- (num_3 * 6 + num_5 * 10) - 2 * n
  return(wins)
}

#apply version
dice_apply <- function(n) {
  rolls <- sample(1:6, n, replace = TRUE)
  wins <- sapply(rolls, function(roll) {
    if (roll == 3 || roll == 5) {
      return((2 * roll)-2)
    } else {
      return(-2)
    }
  })
  return(sum(wins))
}
```

(b) and (c). the results of 3 and 3000 rolls

```
set.seed(123)
print(dice_loop(3))
```

```
[1] 6
```

```
set.seed(123)
print(dice_vectorized(3))
```

```
[1] 6
```

```
set.seed(123)
print(dice_table(3))
```

```
3
6
```

```
set.seed(123)
print(dice_apply(3))
```

```
[1] 6
```

```
set.seed(123)
print(dice_loop(3000))
```

```
[1] 2174
```

```
set.seed(123)
print(dice_vectorized(3000))
```

```
[1] 2174
```

```
set.seed(123)
print(dice_table(3000))
```

```
   3
2174
```

```
set.seed(123)
print(dice_apply(3000))
```

```
[1] 2174
```

(d). the results of 1000 and 100000 rolls using *microbenchmark* package

```
#install.packages("microbenchmark")
library(microbenchmark)
set.seed(123)
microbenchmark(
  loop_1000 = dice_loop(1000),
  vectorized_1000 = dice_vectorized(1000),
  table_1000 = dice_table(1000),
```

```
  apply_1000 = dice_apply(1000),

  loop_100000 = dice_loop(100000),
  vectorized_100000 = dice_vectorized(100000),
  table_100000 = dice_table(100000),
  apply_100000 = dice_apply(100000),

  times = 10
)
```

```
Unit: microseconds
              expr        min         lq        mean      median         uq
         loop_1000   3829.793   3883.626   4063.1049   3975.6045   4039.209
   vectorized_1000    205.501    207.459    229.3630    224.7295    251.126
        table_1000    183.709    187.959    250.9673    271.4800    282.876
        apply_1000    625.084    641.042    666.0798    665.5840    688.959
       loop_100000 397358.709 409086.459 416658.9838 411373.7920 421483.959
 vectorized_100000  16632.793  16801.959  17420.8258  17295.5630  18130.792
      table_100000   7576.209   7706.626   8252.1464   8019.1670   8952.376
      apply_100000  62374.626  63573.334  65654.6591  64782.3965  66704.209
        max neval
   4904.459    10
    266.500    10
    329.584    10
    706.792    10
 452421.917    10
  18380.459    10
   9163.584    10
  71093.168    10
```

```
#The vectorized and table version seems that efficient than loop and apply version.
```

(e). Monte Carlo simulation

```
monte_carlo_simulation <- function(num_simulations, num_rolls) {
  results <- replicate(num_simulations, dice_vectorized(num_rolls))
  expectation <- mean(results)
  return(expectation)
}
set.seed(123)
expectation <- monte_carlo_simulation(10000, 10)
print(expectation)
```

```
[1] 6.7676
```

```
#Since the expected value is much higher than 2 (the cost of this game), although it's not fa
```

Probelm 2

```
#install.packages("dplyr")
#install.packages("ggplot2")
library(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
```

```
library(ggplot2)
cars <- read.csv("/Users/zjuny/Desktop/cars.csv", header = TRUE)
head(cars)
```

```
  Dimensions.Height Dimensions.Length Dimensions.Width
1               140               143              202
2               140               143              202
3               140               143              202
4               140               143              202
5               140               143              202
6                91                17               62
  Engine.Information.Driveline              Engine.Information.Engine.Type
1              All-wheel drive         Audi 3.2L 6 cylinder 250hp 236ft-lbs
2            Front-wheel drive Audi 2.0L 4 cylinder 200 hp 207 ft-lbs Turbo
3            Front-wheel drive Audi 2.0L 4 cylinder 200 hp 207 ft-lbs Turbo
4              All-wheel drive Audi 2.0L 4 cylinder 200 hp 207 ft-lbs Turbo
5              All-wheel drive Audi 2.0L 4 cylinder 200 hp 207 ft-lbs Turbo
```

```
6              All-wheel drive        Audi 3.2L 6 cylinder 265hp 243 ft-lbs
  Engine.Information.Hybrid Engine.Information.Number.of.Forward.Gears
1                      True                                          6
2                      True                                          6
3                      True                                          6
4                      True                                          6
5                      True                                          6
6                      True                                          6
  Engine.Information.Transmission Fuel.Information.City.mpg
1  6 Speed Automatic Select Shift                         18
2  6 Speed Automatic Select Shift                         22
3                  6 Speed Manual                         21
4  6 Speed Automatic Select Shift                         21
5  6 Speed Automatic Select Shift                         21
6                  6 Speed Manual                         16
  Fuel.Information.Fuel.Type Fuel.Information.Highway.mpg
1                   Gasoline                           25
2                   Gasoline                           28
3                   Gasoline                           30
4                   Gasoline                           28
5                   Gasoline                           28
6                   Gasoline                           27
  Identification.Classification          Identification.ID Identification.Make
1        Automatic transmission           2009 Audi A3 3.2                Audi
2        Automatic transmission      2009 Audi A3 2.0 T AT                Audi
3           Manual transmission         2009 Audi A3 2.0 T                Audi
4        Automatic transmission 2009 Audi A3 2.0 T Quattro                Audi
5        Automatic transmission 2009 Audi A3 2.0 T Quattro                Audi
6           Manual transmission           2009 Audi A5 3.2                Audi
  Identification.Model.Year Identification.Year
1              2009 Audi A3                2009
2              2009 Audi A3                2009
3              2009 Audi A3                2009
4              2009 Audi A3                2009
5              2009 Audi A3                2009
6              2009 Audi A5                2009
  Engine.Information.Engine.Statistics.Horsepower
1                                             250
2                                             200
3                                             200
4                                             200
5                                             200
6                                             265
```

```
  Engine.Information.Engine.Statistics.Torque
1                                         236
2                                         207
3                                         207
4                                         207
5                                         207
6                                         243
```

(a). Rename of variables

```
colnames(cars)
```

```
 [1] "Dimensions.Height"
 [2] "Dimensions.Length"
 [3] "Dimensions.Width"
 [4] "Engine.Information.Driveline"
 [5] "Engine.Information.Engine.Type"
 [6] "Engine.Information.Hybrid"
 [7] "Engine.Information.Number.of.Forward.Gears"
 [8] "Engine.Information.Transmission"
 [9] "Fuel.Information.City.mpg"
[10] "Fuel.Information.Fuel.Type"
[11] "Fuel.Information.Highway.mpg"
[12] "Identification.Classification"
[13] "Identification.ID"
[14] "Identification.Make"
[15] "Identification.Model.Year"
[16] "Identification.Year"
[17] "Engine.Information.Engine.Statistics.Horsepower"
[18] "Engine.Information.Engine.Statistics.Torque"
```

```
colnames(cars) <- c("Height", "Length", "Width", "Driveline", "Engine.Type", "Hybrid", "Num_
head(cars)
```

```
  Height Length Width         Driveline
1    140    143   202   All-wheel drive
2    140    143   202 Front-wheel drive
3    140    143   202 Front-wheel drive
4    140    143   202   All-wheel drive
5    140    143   202   All-wheel drive
6     91     17    62   All-wheel drive
```

```
                                  Engine.Type Hybrid Num_Gears
1         Audi 3.2L 6 cylinder 250hp 236ft-lbs   True         6
2 Audi 2.0L 4 cylinder 200 hp 207 ft-lbs Turbo   True         6
3 Audi 2.0L 4 cylinder 200 hp 207 ft-lbs Turbo   True         6
4 Audi 2.0L 4 cylinder 200 hp 207 ft-lbs Turbo   True         6
5 Audi 2.0L 4 cylinder 200 hp 207 ft-lbs Turbo   True         6
6        Audi 3.2L 6 cylinder 265hp 243 ft-lbs   True         6
                    Transmission City.mpg Fuel.Type Highway.mpg
1 6 Speed Automatic Select Shift       18  Gasoline          25
2 6 Speed Automatic Select Shift       22  Gasoline          28
3                6 Speed Manual       21  Gasoline          30
4 6 Speed Automatic Select Shift       21  Gasoline          28
5 6 Speed Automatic Select Shift       21  Gasoline          28
6                6 Speed Manual       16  Gasoline          27
          Classification                          ID Make   Model.Year Year
1 Automatic transmission           2009 Audi A3 3.2 Audi 2009 Audi A3 2009
2 Automatic transmission      2009 Audi A3 2.0 T AT Audi 2009 Audi A3 2009
3    Manual transmission         2009 Audi A3 2.0 T Audi 2009 Audi A3 2009
4 Automatic transmission 2009 Audi A3 2.0 T Quattro Audi 2009 Audi A3 2009
5 Automatic transmission 2009 Audi A3 2.0 T Quattro Audi 2009 Audi A3 2009
6    Manual transmission           2009 Audi A5 3.2 Audi 2009 Audi A5 2009
  Stat.Horsepower Stat.Torque
1             250         236
2             200         207
3             200         207
4             200         207
5             200         207
6             265         243
```

(b). Restrict Fuel type into Gasoline

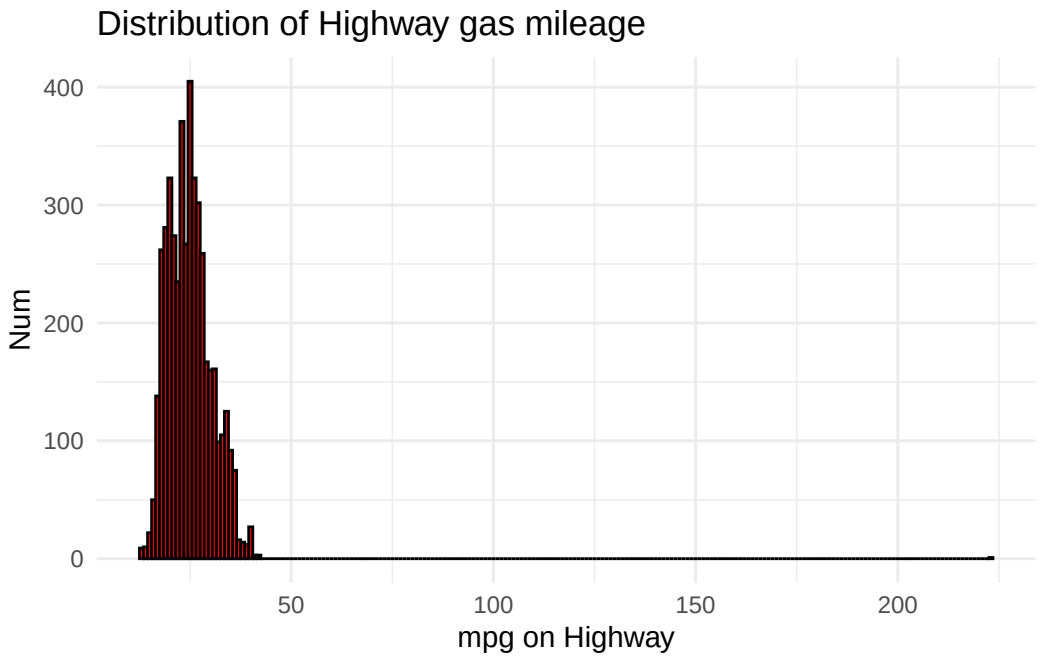
```
cars_Gasoline <- cars %>% filter(Fuel.Type == "Gasoline")
head(cars_Gasoline)
```

```
  Height Length Width         Driveline
1    140    143   202   All-wheel drive
2    140    143   202 Front-wheel drive
3    140    143   202 Front-wheel drive
4    140    143   202   All-wheel drive
5    140    143   202   All-wheel drive
6     91     17    62   All-wheel drive
                                  Engine.Type Hybrid Num_Gears
```

```
1         Audi 3.2L 6 cylinder 250hp 236ft-lbs   True         6
2 Audi 2.0L 4 cylinder 200 hp 207 ft-lbs Turbo   True         6
3 Audi 2.0L 4 cylinder 200 hp 207 ft-lbs Turbo   True         6
4 Audi 2.0L 4 cylinder 200 hp 207 ft-lbs Turbo   True         6
5 Audi 2.0L 4 cylinder 200 hp 207 ft-lbs Turbo   True         6
6        Audi 3.2L 6 cylinder 265hp 243 ft-lbs   True         6
                    Transmission City.mpg Fuel.Type Highway.mpg
1 6 Speed Automatic Select Shift       18  Gasoline          25
2 6 Speed Automatic Select Shift       22  Gasoline          28
3                6 Speed Manual       21  Gasoline          30
4 6 Speed Automatic Select Shift       21  Gasoline          28
5 6 Speed Automatic Select Shift       21  Gasoline          28
6                6 Speed Manual       16  Gasoline          27
          Classification                          ID Make   Model.Year Year
1 Automatic transmission           2009 Audi A3 3.2 Audi 2009 Audi A3 2009
2 Automatic transmission      2009 Audi A3 2.0 T AT Audi 2009 Audi A3 2009
3    Manual transmission         2009 Audi A3 2.0 T Audi 2009 Audi A3 2009
4 Automatic transmission 2009 Audi A3 2.0 T Quattro Audi 2009 Audi A3 2009
5 Automatic transmission 2009 Audi A3 2.0 T Quattro Audi 2009 Audi A3 2009
6    Manual transmission           2009 Audi A5 3.2 Audi 2009 Audi A5 2009
  Stat.Horsepower Stat.Torque
1             250         236
2             200         207
3             200         207
4             200         207
5             200         207
6             265         243
```

(c). Examination of distribution

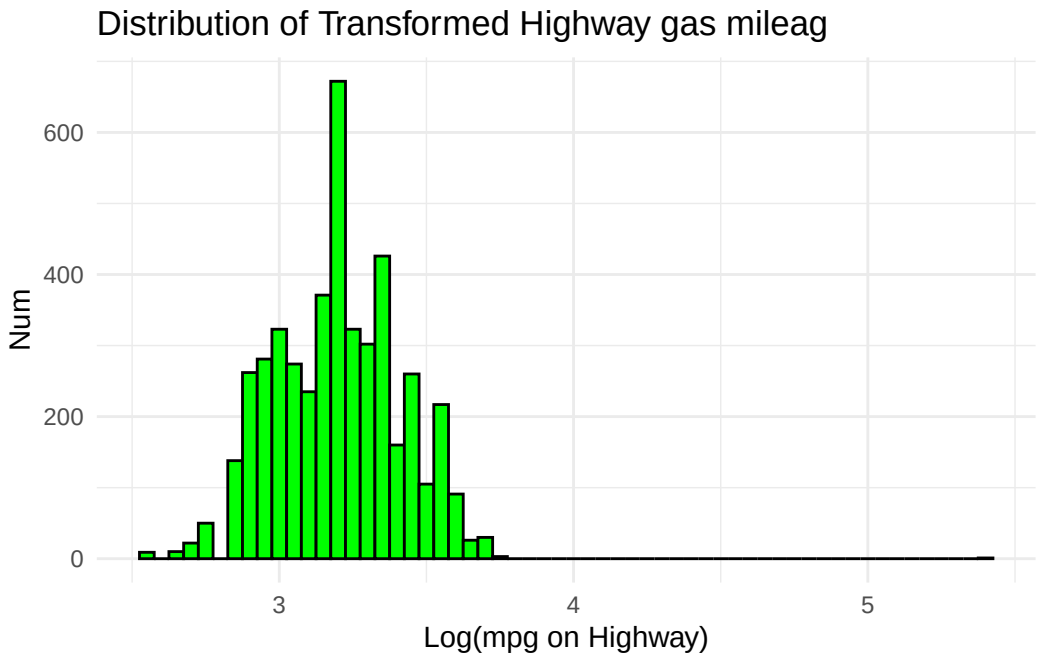
```r
#distribution of highway gas mileage
ggplot(cars_Gasoline, aes(x = Highway.mpg)) +
  geom_histogram(binwidth = 1, fill = "red", color = "black") +
  labs(title = "Distribution of Highway gas mileage",
       x = "mpg on Highway",
       y = "Num")+
  theme_minimal()
```




```
summary(cars_Gasoline$Highway.mpg)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  13.00   21.00   25.00   24.97   28.00  223.00 
```

```
#distribution of transformed highway gas mileage
cars_Gasoline <- cars_Gasoline %>%
  mutate(log_Highway.mpg = log(Highway.mpg))
ggplot(cars_Gasoline, aes(x = log_Highway.mpg)) +
  geom_histogram(binwidth = 0.05, fill = "green", color = "black") +
  labs(title = "Distribution of Transformed Highway gas mileag",
       x = "Log(mpg on Highway)",
       y = "Num")+
  theme_minimal()
```




```
summary(cars_Gasoline$log_Highway.mpg)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.565   3.045   3.219   3.194   3.332   5.407
```

(d). Fitted linear regression model

```
model <- lm(Highway.mpg ~ Stat.Torque + Stat.Horsepower + Height + Length + Width + as.factor
summary(model)
```

```
Call:
lm(formula = Highway.mpg ~ Stat.Torque + Stat.Horsepower + Height +
    Length + Width + as.factor(Year), data = cars_Gasoline)

Residuals:
    Min      1Q  Median      3Q     Max
-10.824  -2.550  -0.452   2.372 202.639

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         32.2926630  0.7225982  44.690  < 2e-16 ***
```

```
Stat.Torque          -0.0507425  0.0022030 -23.034  < 2e-16 ***
Stat.Horsepower       0.0163556  0.0022772   7.182 7.96e-13 ***
Height                0.0099079  0.0011267   8.794  < 2e-16 ***
Length                0.0017290  0.0008836   1.957   0.0504 .
Width                -0.0003343  0.0009045  -0.370   0.7117
as.factor(Year)2010 -0.4539681  0.6768246  -0.671   0.5024
as.factor(Year)2011  0.1711016  0.6757043   0.253   0.8001
as.factor(Year)2012  1.3029279  0.6810076   1.913   0.0558 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.602 on 4582 degrees of freedom
Multiple R-squared:  0.4192,    Adjusted R-squared:  0.4182
F-statistic: 413.3 on 8 and 4582 DF,  p-value: < 2.2e-16
```

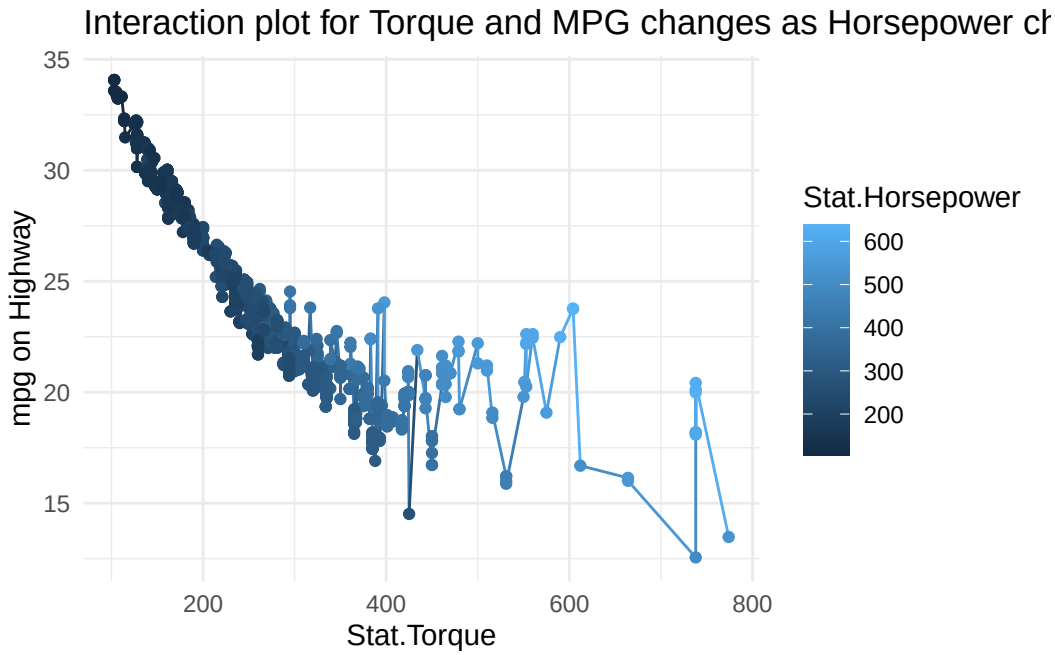
(e). Interaction plot

```
model2 <- lm(Highway.mpg ~ Stat.Torque * Stat.Horsepower + Height + Length + Width + as.facto
summary(model2)
```

```
Call:
lm(formula = Highway.mpg ~ Stat.Torque * Stat.Horsepower + Height +
    Length + Width + as.factor(Year), data = cars_Gasoline)

Residuals:
    Min      1Q  Median      3Q     Max
-11.109  -2.313  -0.258   2.062 203.540

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                  4.219e+01  7.930e-01  53.199  < 2e-16 ***
Stat.Torque                 -8.606e-02  2.533e-03 -33.972  < 2e-16 ***
Stat.Horsepower             -1.666e-02  2.539e-03  -6.563 5.84e-11 ***
Height                       6.560e-03  1.070e-03   6.133 9.32e-10 ***
Length                       1.777e-03  8.318e-04   2.136   0.0327 *
Width                       -1.169e-03  8.521e-04  -1.372   0.1700
as.factor(Year)2010         -5.628e-01  6.372e-01  -0.883   0.3771
as.factor(Year)2011          7.254e-02  6.361e-01   0.114   0.9092
as.factor(Year)2012          1.197e+00  6.411e-01   1.867   0.0619 .
Stat.Torque:Stat.Horsepower  1.124e-04  4.628e-06  24.276  < 2e-16 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.333 on 4581 degrees of freedom
Multiple R-squared:  0.4854,    Adjusted R-squared:  0.4844
F-statistic: 480.1 on 9 and 4581 DF,  p-value: < 2.2e-16
```

```r
interaction_plot <- cars_Gasoline %>%
  filter(Year == 2010) %>%
  mutate(mpg_new = predict(model2, newdata = .))

ggplot(interaction_plot, aes(x = Stat.Torque, y = mpg_new, color = Stat.Horsepower))+
  geom_line()+
  geom_point()+
  labs(title = "Interaction plot for Torque and MPG changes as Horsepower changes",
       x = "Stat.Torque",
       y = "mpg on Highway")+
  theme_minimal()
```



(f). Calculation of beta

```r
x <- model.matrix(~Stat.Torque * Stat.Horsepower + Height + Length + Width + as.factor(Year)
y <- cars_Gasoline$Highway.mpg
betahat <- solve(t(x) %*% x) %*% t(x) %*% y
```

```r
coeff <- as.vector(betahat)
names(coeff) <- colnames(x)
print(coeff)
```

```
                (Intercept)                 Stat.Torque
              42.1879478687               -0.0860592704
            Stat.Horsepower                       Height
              -0.0166633227                0.0065603903
                    Length                        Width
               0.0017767232               -0.0011694485
         as.factor(Year)2010          as.factor(Year)2011
              -0.5627857770                0.0725356431
         as.factor(Year)2012 Stat.Torque:Stat.Horsepower
               1.1970329986                0.0001123567
```