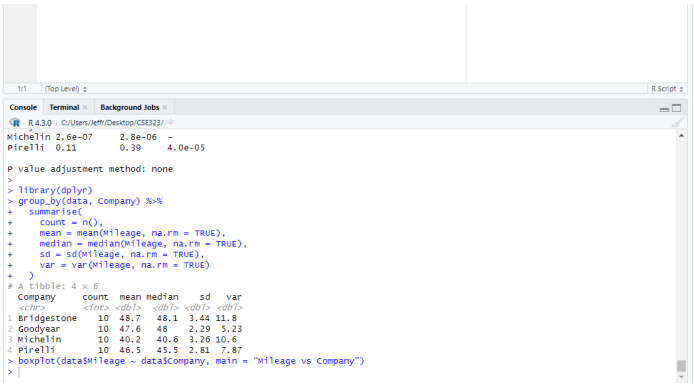
Jeffrey Chin
113257669
jefchin

For dataset1:

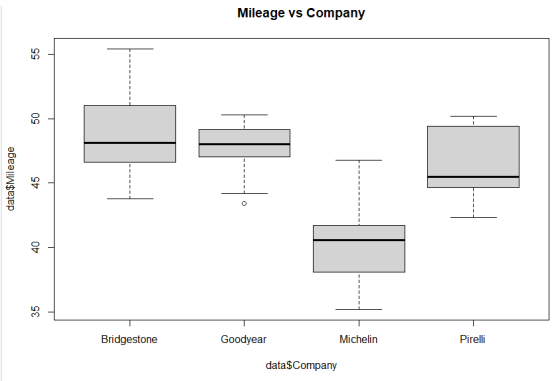
```
> summary(model)
            Df Sum Sq Mean Sq F value   Pr(>F)
Company      3  428.6  142.87   16.09 8.47e-07 ***
Residuals   36  319.7    8.88
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> pairwise_results

        Pairwise comparisons using t tests with pooled SD

data:  data$Mileage and data$Company

          Bridgestone Goodyear Michelin
Goodyear  0.44        -        -
Michelin  2.6e-07     2.8e-06  -
Pirelli   0.11        0.39     4.0e-05

P value adjustment method: none
```

Based on the ANOVA test results, we can conclude that from the F (16.09) and p (8.47e-07) values that there are significant differences in mileage between the 4 tire companies. Based on the pairwise t-tests, the results indicate that there is a huge difference in mean mileage between Michelin and both Goodyear and Pirelli, as well as between Bridgestone and Michelin. However, there seems to be no significant difference in mean mileage between Bridgestone and Goodyear or Goodyear and Pirelli.

BONUS:



```
Michelin 2.6e-07     2.8e-06  -
Pirelli  0.11        0.39     4.0e-05

P value adjustment method: none
>
> library(dplyr)
> group_by(data, Company) %>%
+   summarise(
+     count = n(),
+     mean = mean(Mileage, na.rm = TRUE),
+     median = median(Mileage, na.rm = TRUE),
+     sd = sd(Mileage, na.rm = TRUE),
+     var = var(Mileage, na.rm = TRUE)
+   )
# A tibble: 4 x 6
  Company     count  mean median    sd   var
  <chr>       <int> <dbl>  <dbl> <dbl> <dbl>
1 Bridgestone    10  48.7   48.1  3.44 11.8
2 Goodyear       10  47.6   48    2.29  5.23
3 Michelin       10  40.2   40.6  3.26 10.6
4 Pirelli        10  46.5   45.5  2.81  7.87
> boxplot(data$Mileage ~ data$Company, main = "Mileage vs Company")
>
```



To provide more insight on the data of the csv files, I calculated the statistics for each company. These statistics include the count, mean, median, standard deviation, and variance of the mileage values.

Looking at the statistics from the left in the console, we see that Bridgestone has the highest mean mileage (48.7), followed by Goodyear (47.6), Pirelli (46.5), and Michelin (40.2). We can also see that Bridgestone has the largest variance (11.8) in mileage and Goodyear has the lowest (5.23). The graph on the right is my box and whisker plot for the first dataset. It shows the minimum, first quartile, median, third quartile, and maximum mileage values for each of the companies. From the boxplot, we can see that Bridgestone has the highest median mileage, while Michelin has the most spread in the distribution. Overall, both the statistics and boxplot make it seem that Bridgestone provides the best overall performance.

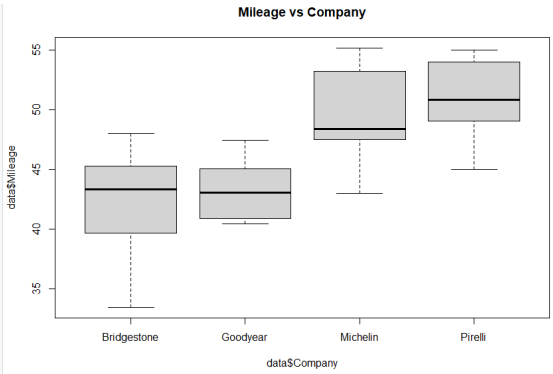
For dataset2:

```
> summary(model)
            Df Sum Sq Mean Sq F value   Pr(>F)
Company      3  568.3  189.43   12.84 7.39e-06 ***
Residuals   36  531.1   14.75
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> pairwise_results

        Pairwise comparisons using t tests with pooled SD

data:  data$Mileage and data$Company

          Bridgestone Goodyear Michelin
Goodyear  0.45442     -        -
Michelin  0.00011     0.00097  -
Pirelli   1.2e-05     0.00012  0.48034

P value adjustment method: none
```

Based on the ANOVA test results, we can conclude that from the F (12.84) and p (7.39e-06) values that there are significant differences between the tire company and the performance of mileage just like in dataset1. To further this claim, the pairwise t-tests show that there’s significant difference in mean mileage between Bridgestone and Michelin, Michelin and Goodyear, and Pirelli and Michelin. The pairwise comparisons also show that there is no significant difference between Goodyear and Bridgestone, Pirelli and Bridgestone, and between Pirelli and Goodyear. The tire company has huge impacts on mileage performance of the tires. Michelin and Bridgestone have the highest and lowest average mileage performances respectively. Goodyear and Pirelli have no significant differences between their mileage performances.

BONUS:

```
Michelin 0.00011      0.00097  -
Pirelli  1.2e-05      0.00012  0.48034

P value adjustment method: none
>
> library(dplyr)
> group_by(data, Company) %>%
+   summarise(
+     count = n(),
+     mean = mean(Mileage, na.rm = TRUE),
+     median = median(Mileage, na.rm = TRUE),
+     sd = sd(Mileage, na.rm = TRUE),
+     var = var(Mileage, na.rm = TRUE)
+   )
# A tibble: 4 x 6
  Company     count  mean median    sd   var
  <chr>       <int> <dbl>  <dbl> <dbl> <dbl>
1 Bridgestone    10  42.0   43.3  4.87 23.8
2 Goodyear       10  43.3   43.1  2.42  5.87
3 Michelin       10  49.5   48.4  4.09 16.7
4 Pirelli        10  50.7   50.8  3.56 12.6
> boxplot(data$Mileage ~ data$Company, main = "Mileage vs Company")
>
```



To provide more insight on the data of the csv files, I calculated the statistics for each company. These statistics include the count, mean, median, standard deviation, and variance of the mileage values. Based on the statistics, we see that Pirelli has the highest mean mileage (50.7), followed by Michelin (49.5), Goodyear (43.3), and Bridgestone (42.0). The graph on the right is my box and whisker plot for the second dataset. It shows the minimum, first quartile, median, third quartile, and maximum values for each of the companies. From the boxplot, we can see that Pirelli has the highest median mileage, while Bridgestone has the most spread out distribution. Overall, we can conclude that Pirelli and Michelin have somewhat better mileage performance than Bridgestone and Goodyear.