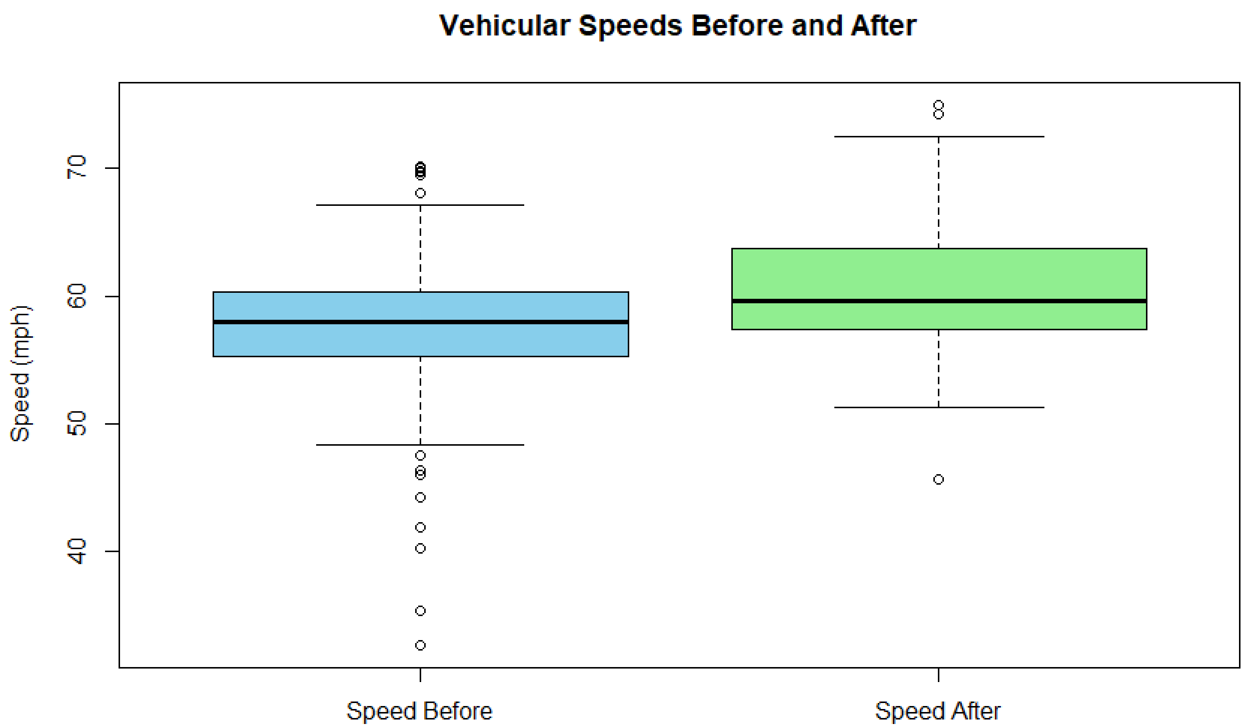
1. Generate box plots for vehicular before and after speeds data and compare the results.



From the observation, we can see the position of the median line of speed after is higher than that of speed before, indicating a potentially higher speed range for speed after compared to speed before. In addition, the length of the box for speed after is longer than the speed before which implies a larger spread of the data. It also means it has more variability in the speed range. In addition, from the graph, we can see that the median lines of the box plot of speed before and speed after are not centered in the box, which suggests skewness in the data distribution. Finally, there are more outliers in the box plot of speed before compared to the box plot of speed after.

2. Generate 95% confidence intervals for the variance of before-speed data. Explain each step and interpret the results.

```r
# Question 2

# Calculate the variance
var.sbf = var(speed$speed_before)

# Get the sample size
n = length(speed$speed_before)

# Confidence level
conf.level = 0.95

# Significance level
alpha = 1 - conf.level

# Chi-squared values for upper and lower confidence limits
chi.1 = qchisq(1 - alpha/2, df=n-1)
chi.2 = qchisq(alpha/2, df=n-1)

# Calculate confidence interval for the variance of before-speed data.
var.interval = c((n-1)*var.sbf/chi.1, (n-1)*var.sbf/chi.2)

# Print results
cat("95% Confidence Interval for for the variance of before-speed data:", var.interval[1], "-", var.interval[2], "\n")
```

```
> cat("95% Confidence Interval for for the variance of before-speed data:", var.interval[1], "-", var.interval[2], "\n")
95% Confidence Interval for for the variance of before-speed data: 16.7784 - 20.51003
```

We are 95% confident that the true variance of the before-speed data falls within this interval. In addition, if the population variance were to be repeatedly estimated using this method, we would expect the true variance to fall within the calculated interval 95% of the time.

3. Generate 90% confidence intervals for mean vehicular after-speed data, assuming the population variance is unknown. Explain each step and interpret the results.

```r
# Question 3

# Remove NA values from speed_after
speed_after <- speed$speed_after
speed_after_clean <- speed_after[!is.na(speed_after)]

# Calculate sample mean and sample standard deviation
sample_mean <- mean(speed_after_clean)
sample_sd <- sd(speed_after_clean)

# Confidence level
conf_level <- 0.90

# Get the sample size
length_speed_after = length(speed$speed_after)

# Calculate standard error of the mean
standard_error <- sample_sd / sqrt(length_speed_after)

# Find t-value for given confidence level and degrees of freedom
t_value <- qt((1 + conf_level) / 2, df = length_speed_after  - 1)

# Calculate margin of error
margin_of_error <- t_value * standard_error

# Calculate confidence interval
confidence_interval <- c(sample_mean - margin_of_error, sample_mean + margin_of_error)

# Print results
cat("90% Confidence Interval for Mean Vehicular After-Speed Data:", confidence_interval[1], "-", confidence_interval[2], "\n")
```

```
> cat("90% Confidence Interval for Mean Vehicular After-Speed Data:", confidence_interval[1], "-", confidence_interval[2]
"\n")
90% Confidence Interval for Mean Vehicular After-Speed Data: 60.41016 - 60.94886
```

We are 90% confident that the true mean of the after-speed data falls within this interval. In addition, if the population mean were to be repeatedly estimated using this method, we would expect the true mean to fall within the calculated interval 90% of the time.

4. Test whether the mean speed is greater than 65 mph after repealing the speed limit at the α=1% significance level. Explain each step and interpret the results.

```r
# Problem 4

# Set the hypothesized mean
test_mean <- 65

# Perform one-sample t-test
test_result <- t.test(speed_after_clean, mu = test_mean, alternative = "greater", conf.level = 0.99)

# Get the p-value
p_value <- test_result$p.value

# Set significance level
alpha <- 0.01

# Print the test result
cat("p-value:", p_value, "\n")

# Check if the p-value is less than alpha
if (p_value < alpha) {
  cat("The mean speed is greater than 65 mph at the 1% significance level.\n")
} else {
  cat("There is not enough evidence to conclude that the mean speed is greater than 65 mph at the 1% significance level.\n")
}
```

```
> cat("p-value:", p_value, "\n")
p-value: 1
```

```
+   cat("There is not enough evidence to conclude that the mean speed is greater than 65 mph at the 1% significance level.\n")
+ }
There is not enough evidence to conclude that the mean speed is greater than 65 mph at the 1% significance level.
```

To test whether the mean speed is greater than 65 mph after repealing the speed limit at the α=1% significance level, I use a one-sample t-test. The null hypothesis is that the mean speed is less than or equal to 65 mph, and the alternative hypothesis is that the mean speed is greater than 65 mph. If the p-value is less than alpha, reject the null hypothesis and conclude that the mean speed is greater than 65 mph. Otherwise, it fails to reject the null hypothesis. In this case, the p-value is greater than alpha (1 > 0.01), so we fail to conclude that the mean speed is greater than 65 mph at the 1% significance level.

5. Test whether the variance of after-speed data is less than 20 mph2 at the α=5% significance level. Explain each step and interpret the results

```r
# Question 5

# Variance to test
var_to_test <- 20

# Degrees of freedom for the numerator
df_numerator <- length_speed_after - 1

# Degrees of freedom for the denominator (fixed at 1 for testing variance)
df_denominator <- 1

# Calculate F statistic
F_statistic <- var(speed_after_clean) / var_to_test

# Calculate p-value
p_value <- pf(F_statistic, df_numerator, df_denominator, lower.tail = FALSE)

# Print the test result
cat("p-value:", p_value, "\n")

# Set significance level
alpha <- 0.05


# Check if the p-value is less than alpha
if (p_value < alpha) {
  cat("The variance of the after-speed data is less than 20 mph^2 at the 5% significance level.\n")
} else {
  cat("There is not enough evidence to conclude that the variance of the after-speed data is less than 20 mph^2 at the 5% significance level.\n")
}
```

```
> cat("p-value:", p_value, "\n")
p-value: 0.6771526
```

```
+   cat("There is not enough evidence to conclude that the variance of the after-speed data is less than 20 mph^2 at the 5% significance level.\n")
+ }
There is not enough evidence to conclude that the variance of the after-speed data is less than 20 mph^2 at the 5% significance level.
```

To test whether the variance of after-speed data is less than 20 mph^2 at the α=5% significance level, I use the F-distribution. The null hypothesis is that the variance of the after-speed data is equal to or greater than 20 mph^2, and the alternative hypothesis is that the variance is less than 20 mph^2. If the p-value is less than alpha, reject the null hypothesis and conclude that the variance of the after-speed data is less than 20 mph^2. Otherwise, it fails to reject the null hypothesis. In this case, the p-value is greater than alpha (0.6771 > 0,05), so we fail to conclude that the variance of the after-speed data is less than 20 mph^2 at the 5% significance level.

6. Test that the vehicular speed variances before and after are equal at the α=10% significance level. Explain each step and interpret the results.

```r
# Question 6

# Perform an F-test for equality of variances
alpha <- 0.10
test_result <- var.test(speed_after_clean,speed$speed_before, conf.level = 1 - alpha)
print(test_result)
```

```
> print(test_result)

        F test to compare two variances

data:  speed_after_clean and speed$speed_before
F = 1.1045, num df = 570, denom df = 763, p-value = 0.2022
alternative hypothesis: true ratio of variances is not equal to 1
90 percent confidence interval:
 0.9715999 1.2572732
sample estimates:
ratio of variances
          1.104475
```

To check that the vehicular speed variances before and after are equal at the α=10% significance level, I performed an F-test for equality of variances. The null hypothesis for this test is that the variances are equal. In this test, if the resulting p-value is less than the chosen significance level, you reject the null hypothesis and conclude that the variances are significantly different. Otherwise, you fail to reject the null hypothesis. In other words, they are considered equal. In our case, the resulting p-value is 0.2022 higher than our significance level of 0.1 which means we fail to reject the null hypothesis. This means the vehicular speed variances before and after are equal at the α=10% significance level.

7. Test that the vehicular after-speed mean is greater than before-speed at the α=1% significance level. Explain each step and interpret the results.

```r
# Question 7

alpha_2 <- 0.01
# Perform the one-tailed t-test
test_result <- t.test(speed_after_clean,speed$speed_before, alternative = "greater", conf.level = 0.99)
print(test_result)
```

```
> test_result <- t.test(speed_after_clean,speed$speed_before, alternative = "greater", conf.level = 0.99)
> print(test_result)

        Welch Two Sample t-test

data:  speed_after_clean and speed$speed_before
t = 11.867, df = 1194, p-value < 2.2e-16
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 2.336348      Inf
sample estimates:
mean of x mean of y 
 60.67951  57.77251 
```

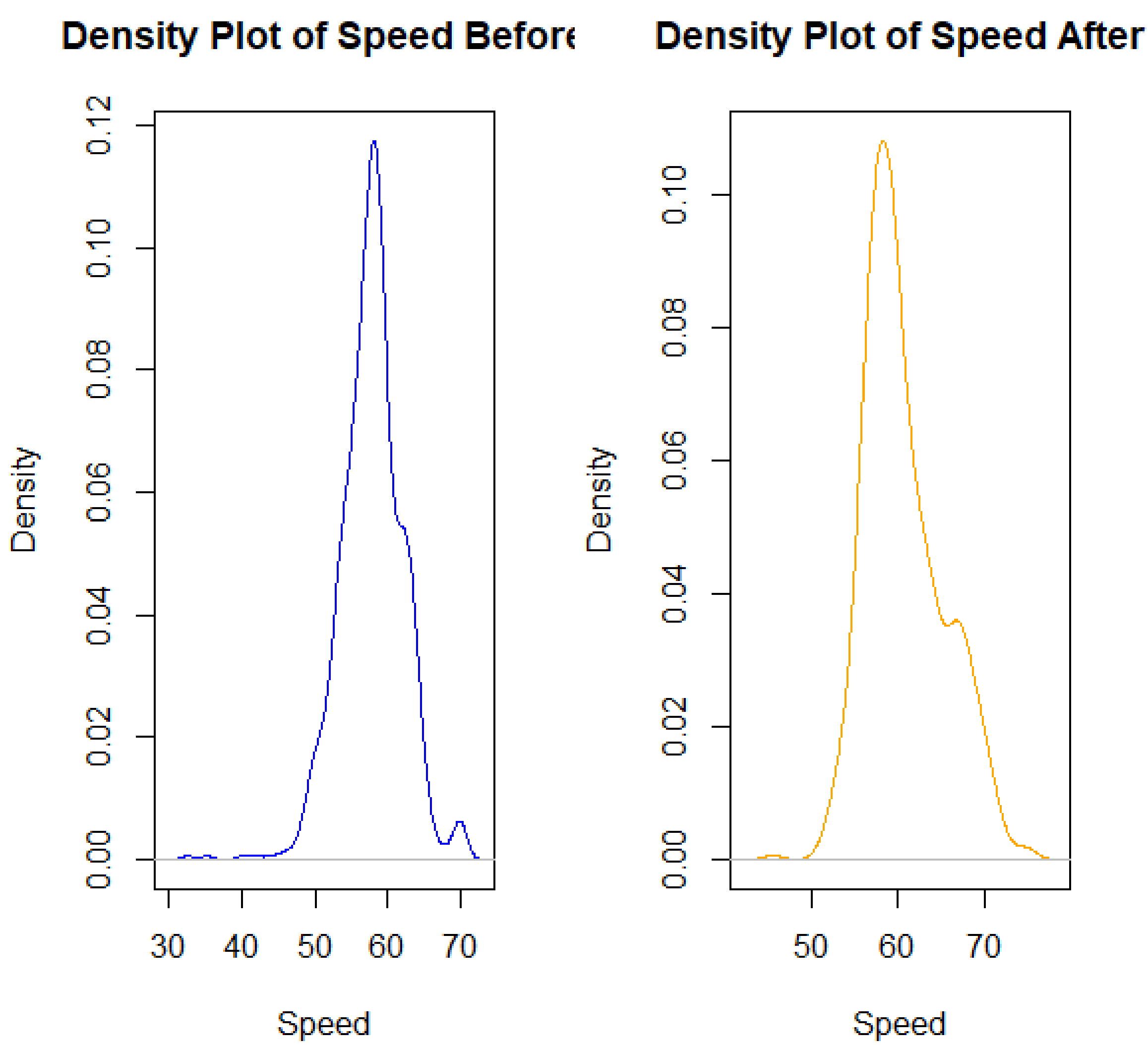
To test that the vehicular after-speed mean is greater than before-speed at the α=1% significance level, I performed a one-tailed t-test. The null hypothesis for this test is that the mean vehicular after-speed is not greater than the before-speed mean. In this test, if the resulting p-value is less than the chosen significance level, you reject the null hypothesis and conclude that the mean vehicular after-speed is greater than the before-speed mean . Otherwise, you fail to reject the null hypothesis. In our case, the resulting p-value is less than 2.2e-16 lower than our significance level of 0.01 which means that we reject the null hypothesis. This means that the vehicular after-speed mean is greater than before-speed at the α=1% significance level

8. Use a Mann-Whitney-Wilcoxon test to assess whether the distributions of speeds before and after are equal. Also, draw density plots using before and after speed data. Interpret the results based on both the test and drawing

```
> print(test_result)

        Wilcoxon rank sum test with continuity correction

data:  speed$speed_before and speed_after_clean
W = 146050, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```



After performing a Mann-Whitney-Wilcoxon test to speed before and speed after, we get the p-value of 2.2e-16, which is extremely small so that we can assess that the distributions of speeds before and after are not equal. With such a small p-value, we reject the null hypothesis, suggesting that there is a significant difference between the distributions of speeds before and after. In addition, looking at the density plot, this result suggests that something has changed (repealing the national maximum speed limit in this case) between the two time periods, affecting the distribution of speeds.