ALY 6010 – Probability Theory and Introductory Statistics

A picture containing logo, text, emblem, circle

Description automatically generated

Northeastern University

Course : - ALY 6010 Section : - 3

Professor : - Sergiy Shevchenko

**Final Project**

**Electric Vehicle Population Data [ Report ]**

Group : - **HUSKY 4**

**Electric Vehicle Population Data**

**Introduction**

**Description : -**

The Electric Vehicle Population Data, sourced from a US Government website (data.gov), provides a detailed look at the world of electric cars. This dataset helps us understand how electric vehicles are growing and changing over time, which is important as we move towards more environmentally friendly transportation options. We're interested in this data because it helps us see how electric cars are becoming more popular and how they're helping to tackle environmental issues.

**Purpose : -**

In this exploratory data analysis, our goal is to delve into information about electric vehicles (EVs), including their types, manufacturers, geographical distribution, and eligibility for clean alternative fuel vehicle programs. By closely examining this data, we aim to uncover valuable insights into the factors that influence the adoption and usage patterns of EVs. This dataset provides a comprehensive view of the global or regional EV population, facilitating analysis of its growth, distribution, and market characteristics.

**Variables & Data Types : -**

**Text : -** Country , City , State , Make , Model , Electric vehicle Type , Clean Alternative Fuel ( CAFV ) Eligibility are in text type of data.

**Numeric : -** VIN , Postal Code , Model Year , Range , Base MSRP , legislative district and DOL are in numeric type of data.

**Data Summary : -**

**Rows : -** This Dataset contains 173534 Rows.

**Fields / Columns : -** This Dataset contains 16 variables.

**Data Source : -**

The data for the Electric Vehicle Population project comes from various places like government websites, reports from electric vehicle manufacturers, and international organizations, as well as government agencies. These sources are listed at the end of the report for reference.

**Data Analysis**

To make sure our analysis is trustworthy, we've carefully cleaned the data. This means we've removed any unusual characters in column names, dealt with missing information, standardized how data is presented, and made sure categories are organized consistently. These steps are important to make sure our analysis is solid and the conclusions we draw from the data are reliable.

**Data Visualization Graphs**

**A graph with different colored bars

Description automatically generated**

The Above Histogram of bar / column represents data of Average electric vehicle range according to EV\_Makers in which we analyze that JAGUAR has the highest average EV\_range among all whereas SUBARU has the least EV\_range. While some has negligible average range.

**A graph of different colored squares

Description automatically generated**

The Above histogram shows Electric vehicle types according to CAFV Eligibility with their count. We can Analyse that Not Eligible and Eligible type has the least count as compare to Unknown type of CAFV.

**A graph of models by top 8 car makers

Description automatically generated**The above graph / histogram shows the count of models by top 8 car makers of EV which contains makers such as BMW , CHERVOLET , FORD , KIA , NISSAN etc. This shows the top car makers among all the car makers along with their count

**Inferential Statistics & Hypothesis Testing**

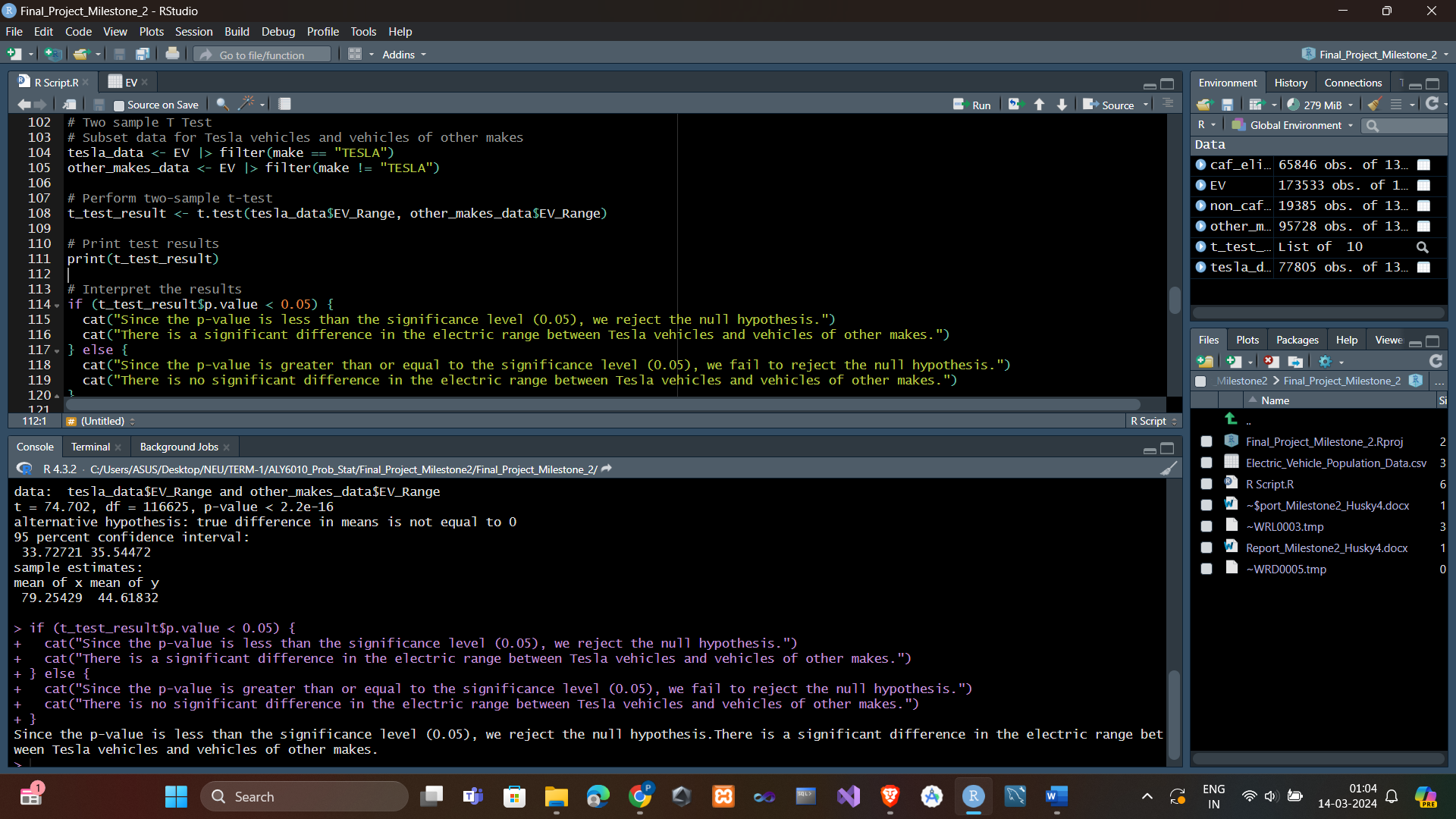
**Null & Alternative Hypothesis [ One / Two sample Test ]**

**Question 1 : -**  Is there a significant difference in the electric range between Tesla vehicles and vehicles of other makes?

**Hypothesis Testing : -**

**Null Hypothesis (H0) :** There is no significant difference in the electric range between Tesla vehicles and vehicles of other makes.

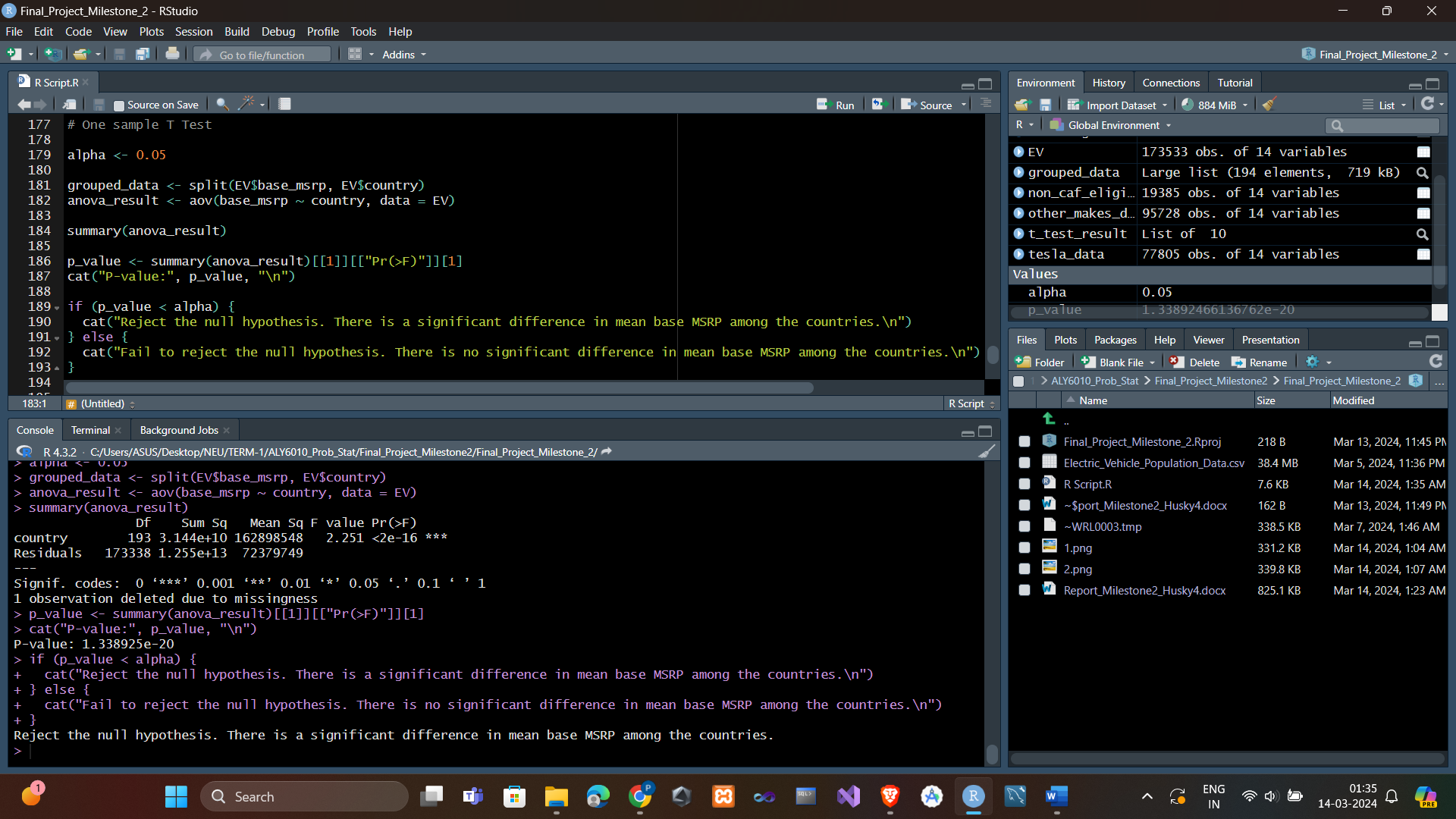
**Alternative Hypothesis (H1) :** There is a significant difference in the electric range between Tesla vehicles and vehicles of other makes.

Select a significance level (α). Let's choose α = 0.05.  


**# OUTPUT :** Because the p-value is lower than our predetermined significance level of 0.05, we can reject the idea of no difference (null hypothesis). This means there's a notable contrast in the electric range between Tesla cars and those made by other companies.

**# Interpretation :** In simpler terms, when we tested whether there's a big difference in how far Tesla cars can go on electric power compared to cars from other brands, the results show that yes, there is indeed a significant gap.

Based on the two-sample t-test conducted with a significance level of 0.05:  
Result: The p-value obtained from the test is less than the significance level (0.05).  
Interpretation: Therefore, we reject the null hypothesis.  
Conclusion: There is a significant difference in the electric range between Tesla vehicles and vehicles of other makes.

**Question 2 : -**  Is there a significant difference in the mean base MSRP of Electric Vehicles between different countries?  
 **Hypothesis Testing : -**   
**Null Hypothesis (H0) :** The mean base MSRP for Electric Vehicles is the same across all countries.  
**Alternative Hypothesis (H1) :** There is a significant difference in the mean base MSRP for Electric Vehicles across different countries.  
  
Select a significance level (αlpha). Let's choose αlpha = 0.05.  
  
  
**# OUTPUT :** ANOVA F-statistic: This number tells us if the averages of the groups are really different from each other. P-value: This tells us the likelihood of getting results as unusual as the ones we observed, if the idea of no difference (null hypothesis) were true.

**# Interpretation :**

If the p-value is less than the significance level (α = 0.05), we reject the null hypothesis.

If the p-value is greater than or equal to 0.05, we fail to reject the null hypothesis.

**# Conclusion :**

If the p-value is less than 0.05, it means we can confidently say that there's a meaningful difference in the average base MSRP (Manufacturer's Suggested Retail Price) across different countries.

If the p-value is greater than or equal to 0.05, it means we can't make a strong conclusion about there being a significant difference in the average base MSRP among the countries; there's no clear evidence supporting such a difference.

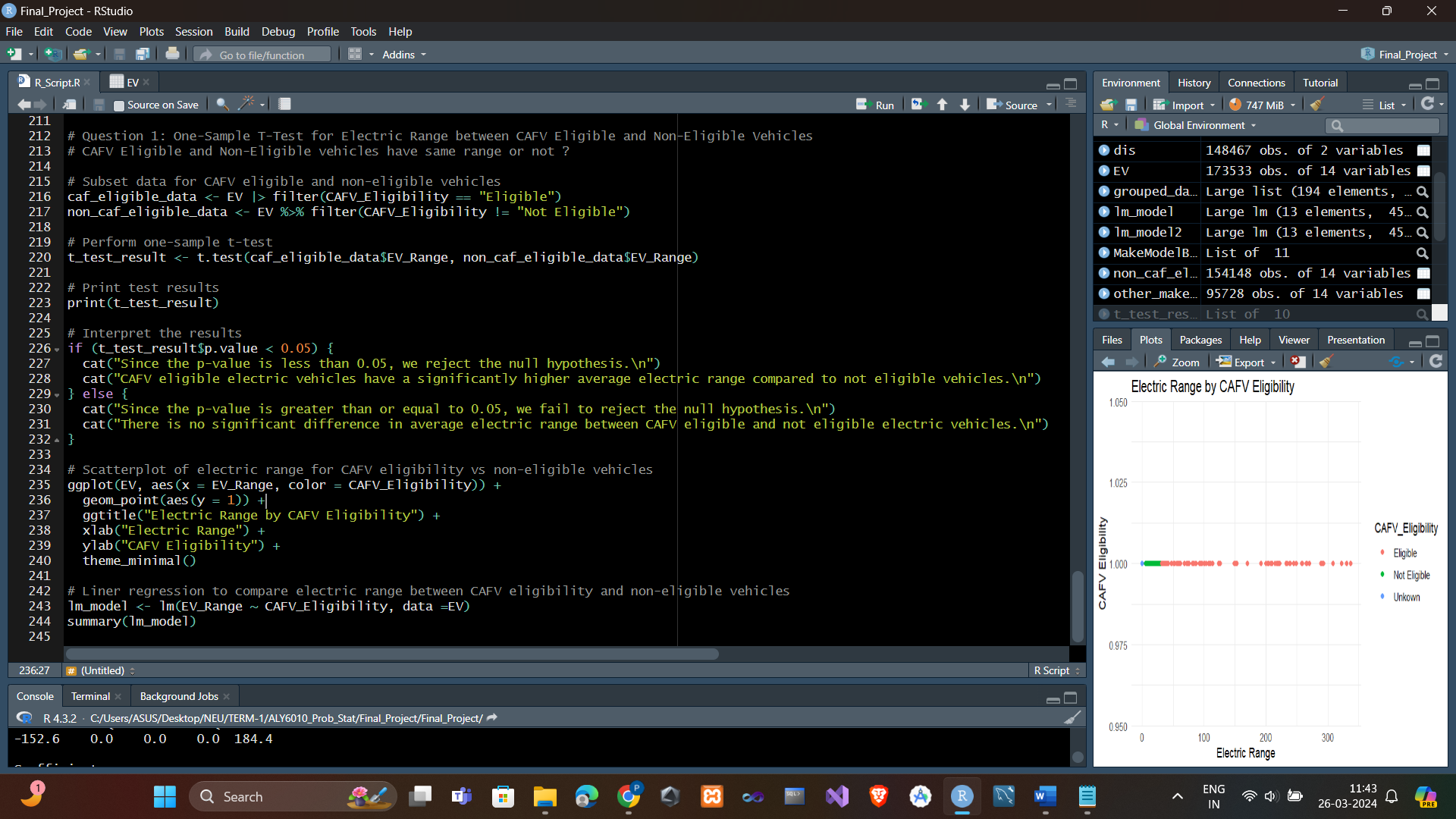
**Relationship Between Variables in EV Dataset  
Hypothesis & Regression Testing , Inferential Statistics , Scatterplots**

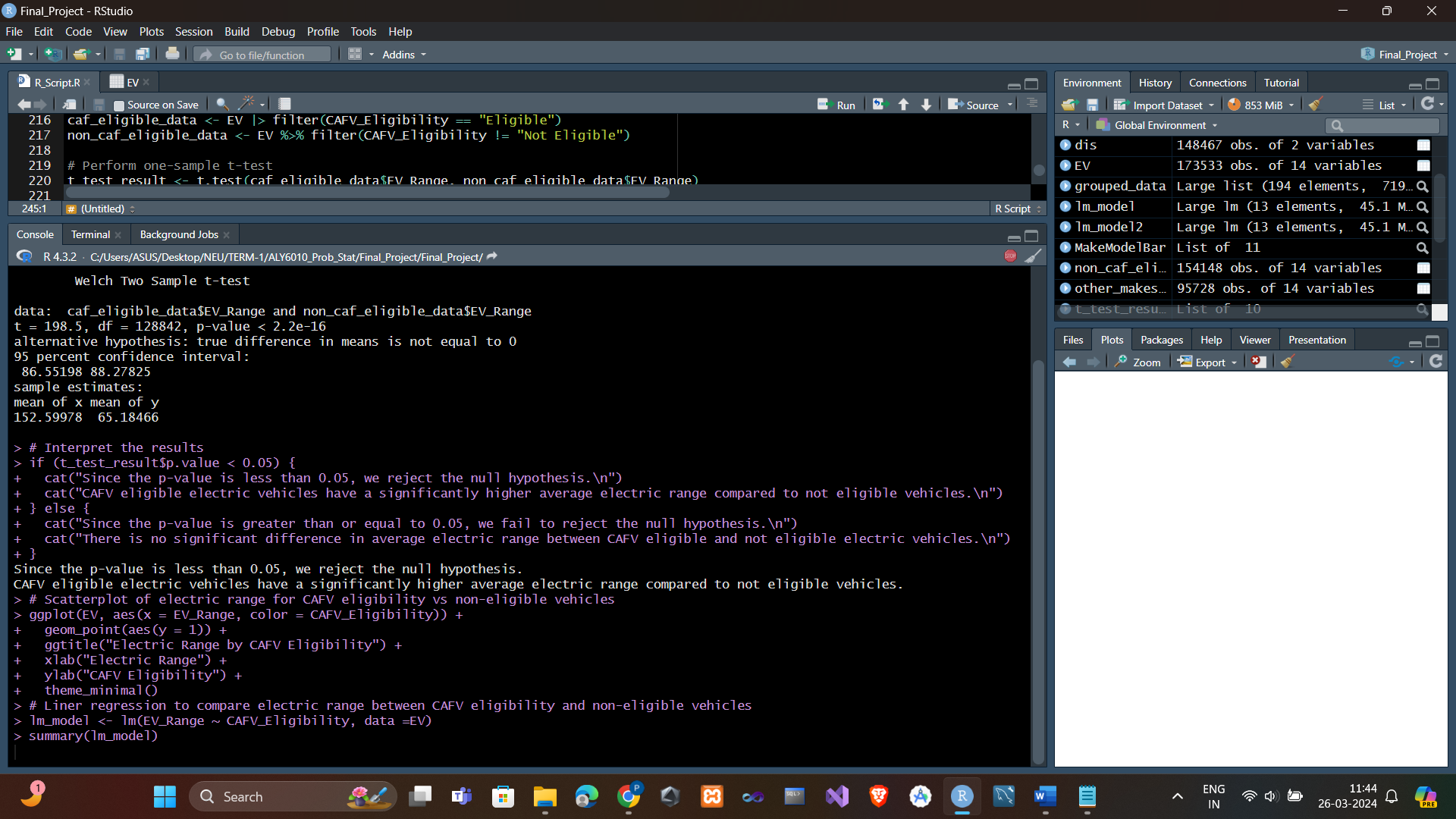
**Question 1 : -**  One-Sample T-Test for Electric Range between CAFV Eligible and Non-Eligible Vehicles. Is the CAFV Eligible and Non-Eligible vehicles have same range or not ?

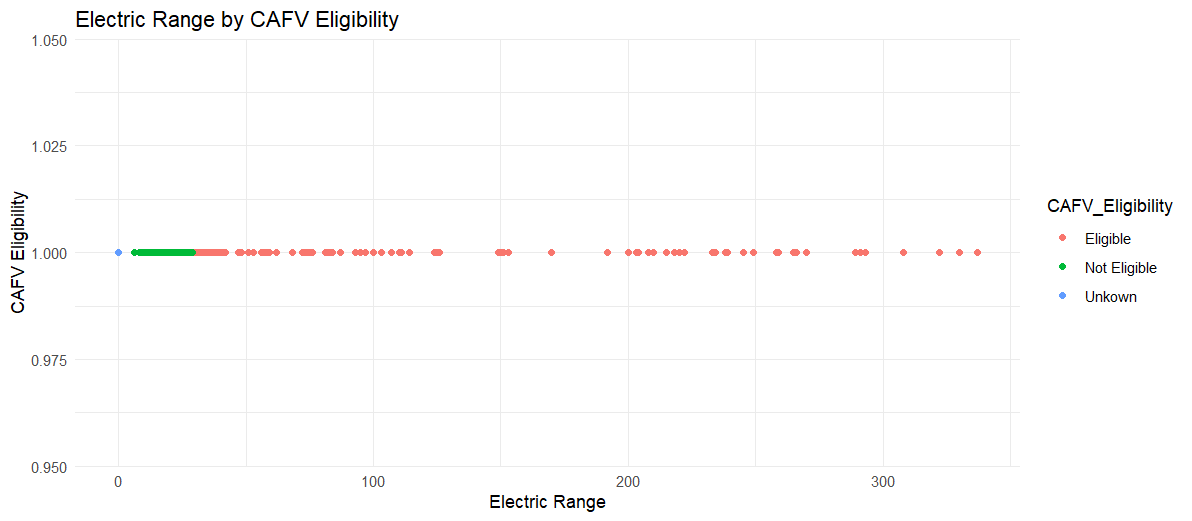
**Hypothesis Testing : -**

**Null Hypothesis (H0) :** There is no significant difference in average electric range between CAFV eligible and not eligible electric vehicles.

**Alternative Hypothesis (H1) :** CAFV eligible electric vehicles have a significantly higher average electric range compared to not eligible vehicles.



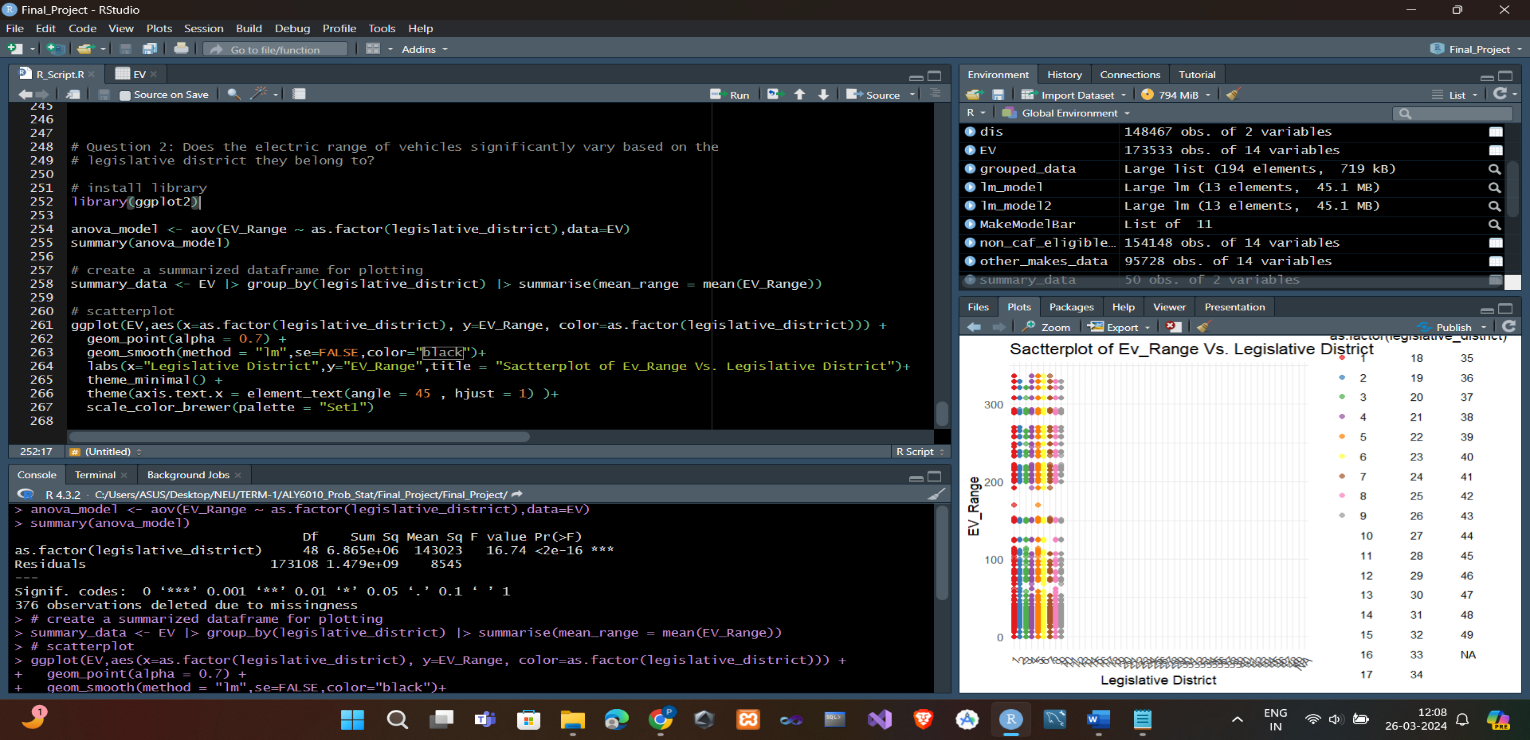




The p-value from the one-sample t-test is compared to the significance level (α = 0.05) to determine whether to reject the null hypothesis.

**Scatterplot and Linear Regression :-** Another scatterplot is created to visualize the electric range distribution for CAFV eligible vehicles compared to non-eligible vehicles.Linear regression is performed to quantify the relationship between CAFV eligibility and electric range.

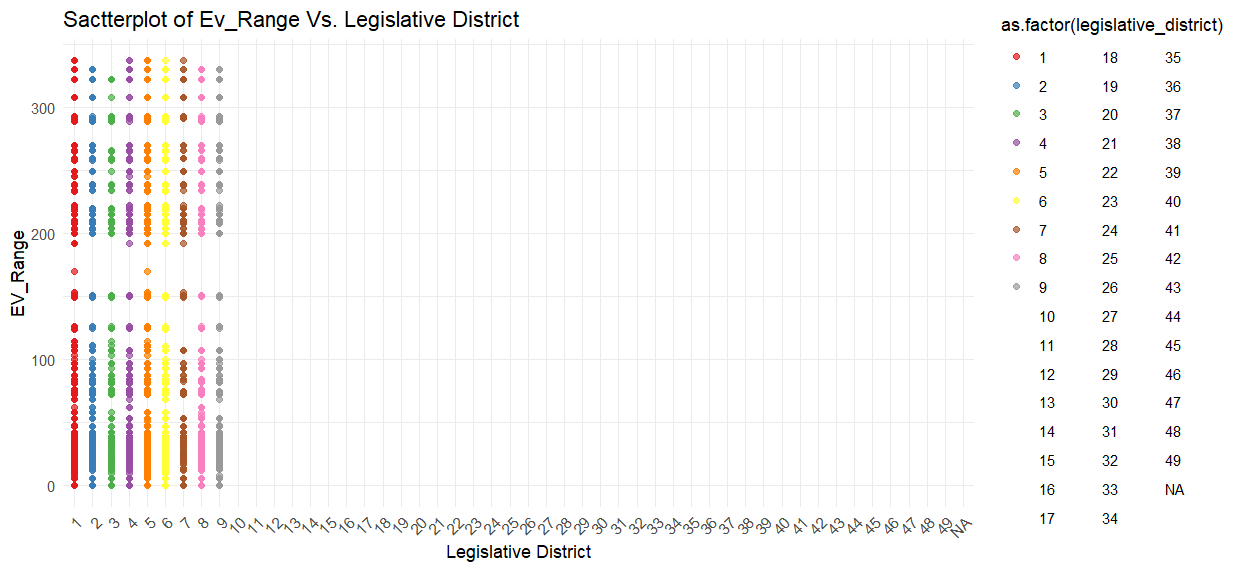
**Question 2: -**  Does the electric range of vehicles significantly vary based on the legislative district they belong to? **Hypothesis Testing : -**   
**Null Hypothesis (H0) :** There is no significant difference in electric range based on legislative district.  
**Alternative Hypothesis (H1) :** There is a significant difference in electric range based on legislative district.



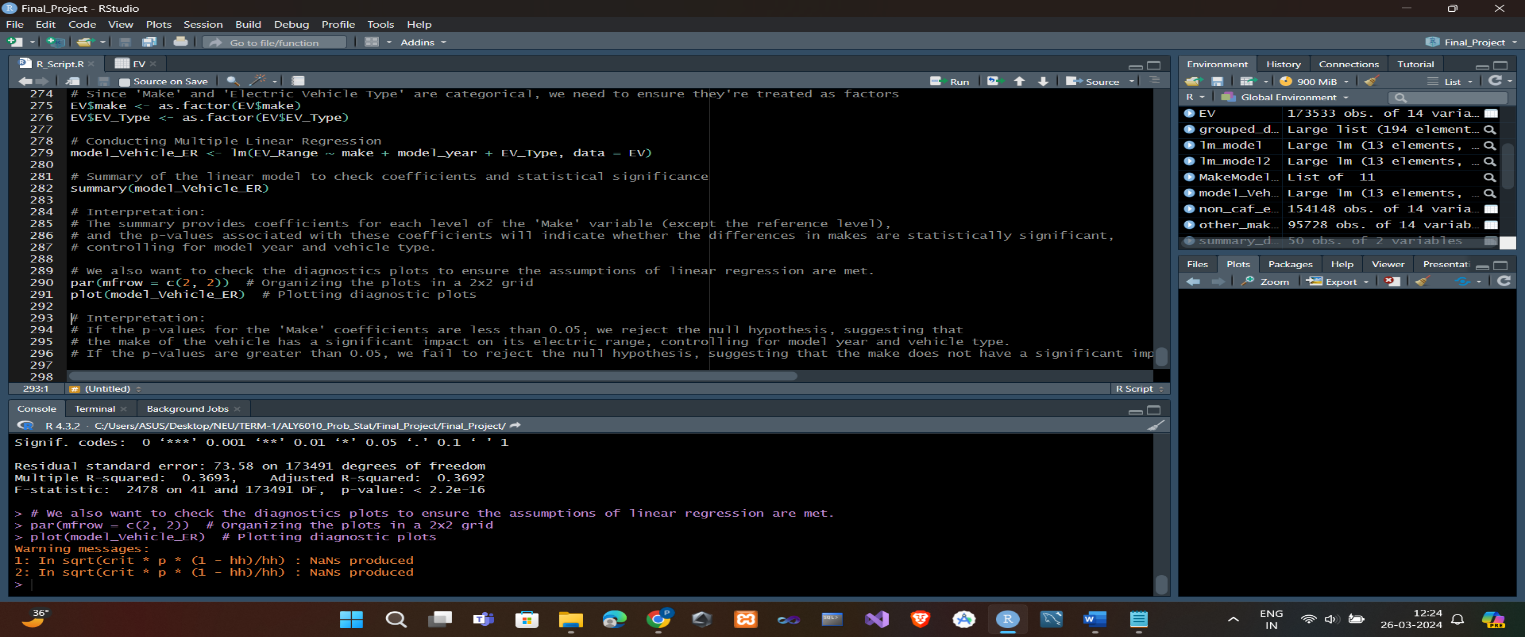
The F value is 16.74 with a very small p-value (< 2e-16), suggesting that there is a significant difference in electric range based on legislative district.

Given these results, we reject the null hypothesis and accept the alternative hypothesis. This indicates that there is indeed a significant difference in electric range based on legislative district.

In summary, we have sufficient evidence to conclude that legislative district has a significant effect on electric range.



**Question 3: -**  : Does the make of the vehicle significantly impact its electric range, controlling for other factors like model year and vehicle type? **Hypothesis Testing : -**   
**Null Hypothesis (H0) :** The make of the vehicle does not have a significant impact on its electric range, controlling for model year and vehicle type.

**Alternative Hypothesis (H1) :** The make of the vehicle has a significant impact on its electric range, controlling for model year and vehicle type.

A computer screen shot of a black screen

Description automatically generated

A computer screen with text on it

Description automatically generated

# Interpretation :-

If the p-values for the 'Make' coefficients are less than 0.05, we reject the null hypothesis, suggesting that the make of the vehicle has a significant impact on its electric range, controlling for model year and vehicle type.

If the p-values are greater than 0.05, we fail to reject the null hypothesis, suggesting that the make does not have a significant impact.

**Summary**

In summary, looking closely at the Electric Vehicle Population Data gives us some really useful information about the electric car market. We found out things like how far electric cars can go on a charge, which cars are eligible for clean fuel programs, and the different models made by various companies. To learn even more, we could dig into how these things change over time, what makes people decide to get electric cars, and how electric car use differs in different parts of the country.

**References : -**

Data Set : - “ Electric vehicle population Data “ , Published by data.gov platform / site , retrieved on 1ST March , 2024 Online from , <https://catalog.data.gov/dataset/electric-vehicle-population-data>