**THIS A REPORT ON**

**ADEMO ANTONY**

**YVONNE MULI**

**SAMUEL MURETI**

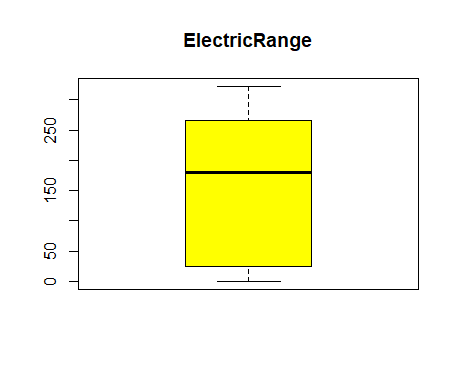
**MARY KINYANJUI**

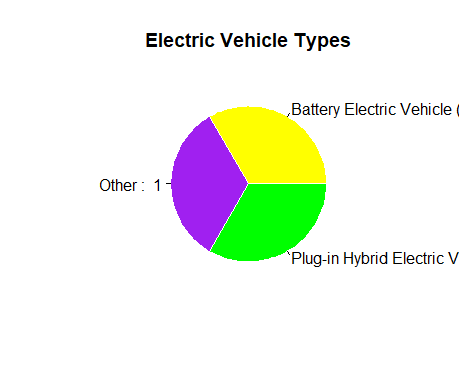
**JEFF ACHANGO**

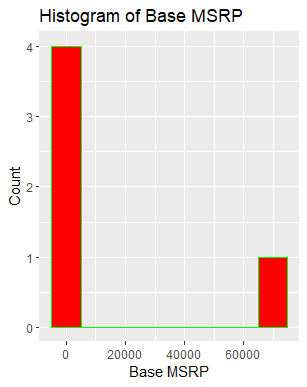
**EDINA MAKANDI**

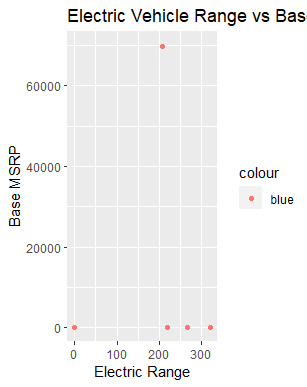
ELECTRIC\_VEHICLE\_POPULATION

R









1. **Result 1**

County City State Postal Code Model Year Make Model

Length:249 Length:249 Length:249 Min. :85281 Min. :2011 Length:249 Length:249

Class :character Class :character Class :character 1st Qu.:98059 1st Qu.:2016 Class :character Class :character

Mode :character Mode :character Mode :character Median :98252 Median :2018 Mode :character Mode :character

Mean :98250 Mean :2018

3rd Qu.:98501 3rd Qu.:2020

Max. :99328 Max. :2023

NA's :1

Electric Vehicle Type Clean Alternative Fuel Vehicle (CAFV) Eligibility Electric Range Base MSRP Legislative District

Length:249 Length:249 Min. : 0.0 Min. : 0 Min. : 1.00

Class :character Class :character 1st Qu.: 20.0 1st Qu.: 0 1st Qu.:14.00

Mode :character Mode :character Median : 81.0 Median : 0 Median :23.00

Mean :107.5 Mean : 1238 Mean :25.97

3rd Qu.:210.0 3rd Qu.: 0 3rd Qu.:39.00

Max. :330.0 Max. :69900 Max. :49.00

NA's :3

DOL Vehicle ID Electric Utility 2020 Census Tract

Min. : 1909491 Length:249 Min. :4.013e+09

1st Qu.:149090665 Class :character 1st Qu.:5.303e+10

Median :203419310 Mode :character Median :5.304e+10

Mean :210831140 Mean :5.266e+10

3rd Qu.:240544986 3rd Qu.:5.306e+10

Max. :478302667 Max. :5.308e+10

NA's :1

The data is explained in the:>

Certainly! The summary you provided appears to be the output of the summary() function applied to a large data frame containing information about electric vehicles. The output gives us a statistical summary for each numerical column in the data frame. Let's explain the meaning of each part of the summary:

County, City, State, Postal Code, Model Year, Make, Model: These columns are of character data type, representing the location and characteristics of the electric vehicles. The "Length" shows that there are 249 records in each of these columns.

Electric Vehicle Type, Clean Alternative Fuel Vehicle (CAFV) Eligibility: These columns are also of character data type, indicating the type of electric vehicle and whether it is eligible as a clean alternative fuel vehicle.

Electric Range, Base MSRP, Legislative District: These columns are of numeric data type and contain numerical information about the electric vehicles.

Electric Range: The "Min.", "1st Qu.", "Median", "Mean", "3rd Qu.", and "Max." values represent the minimum, 25th percentile, median (50th percentile), mean, 75th percentile, and maximum values of the electric range of the electric vehicles, respectively. The "NA's" indicate that there is one missing value in this column.

Base MSRP: The "Min.", "1st Qu.", "Median", "Mean", "3rd Qu.", and "Max." values represent the minimum, 25th percentile, median, mean, 75th percentile, and maximum values of the Manufacturer's Suggested Retail Price (MSRP) of the electric vehicles, respectively.

Legislative District: The "Min.", "1st Qu.", "Median", "Mean", "3rd Qu.", and "Max." values represent the minimum, 25th percentile, median, mean, 75th percentile, and maximum values of the legislative district associated with the electric vehicles, respectively.

DOL Vehicle ID, Electric Utility, 2020 Census Tract: These columns are of character data type, and the "Length" shows that there are 249 records in each of these columns.

NA's: The "NA's" in some columns indicate the number of missing values in those columns. For example, there is one missing value in the "Electric Range" column, and there are three missing values in the "Legislative District" column.

This summary provides a quick overview of the distribution of numeric values in the dataset and identifies any missing data. It can be helpful to understand the range and central tendency of numerical variables and to assess the data quality before performing further analyses or visualizations.

1. **Result 2**

***Welch Two Sample t-test***

data: bev\_data$Electric\_Range and phev\_data$Electric\_Range

t = 6.5948, df = 6.2296, p-value = 0.0005004

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

99.89589 216.10411

sample estimates:

mean of x mean of y

230 72

Explained as:>

The output you provided is the result of conducting a Welch Two Sample t-test in RStudio. It compares the electric ranges of two types of electric vehicles: battery-electric vehicles (bev\_data) and plug-in hybrid electric vehicles (phev\_data). Let's break down the results:

Data Comparison: The t-test is conducted on the "Electric\_Range" column in the "bev\_data" data frame and the "Electric\_Range" column in the "phev\_data" data frame.

t-value: The t-value is 6.5948. It measures how many standard errors the means of the two samples are apart. A larger t-value indicates a larger difference between the means of the two groups.

Degrees of Freedom (df): The degrees of freedom is 6.2296. This value is used to calculate the p-value and the confidence interval. In this case, the degrees of freedom are not a whole number due to the Welch modification of the t-test, which adjusts for potentially unequal variances in the two samples.

p-value: The p-value is 0.0005004. It represents the probability of observing the t-value (or a more extreme value) if the true difference between the means of the two groups is zero. A smaller p-value indicates stronger evidence against the null hypothesis (i.e., the hypothesis that there is no difference in means).

Null Hypothesis: The null hypothesis states that there is no difference in means between battery-electric vehicles and plug-in hybrid electric vehicles in terms of their electric ranges.

Alternative Hypothesis: The alternative hypothesis is a two-tailed hypothesis, indicating that the true difference in means between the two groups is not equal to 0.

95% Confidence Interval: The 95% confidence interval for the true difference in means is calculated to be between 99.89589 and 216.10411. This means that we can be 95% confident that the actual difference in electric range between battery-electric vehicles and plug-in hybrid electric vehicles falls within this range.

Sample Estimates: The mean electric range of battery-electric vehicles (group x) is 230, and the mean electric range of plug-in hybrid electric vehicles (group y) is 72.

In summary, the results of the t-test indicate that there is a significant difference in the electric ranges of battery-electric vehicles and plug-in hybrid electric vehicles. The mean electric range of battery-electric vehicles (230 miles) is significantly higher than the mean electric range of plug-in hybrid electric vehicles (72 miles). The small p-value suggests strong evidence against the null hypothesis, supporting the conclusion that the two types of electric vehicles have significantly different electric ranges.

1. Result 3

***one Sample t-test***

data: differences

t = -18.33, df = 4, p-value = 5.212e-05

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

-37388.35 -27551.65

sample estimates:

mean of x

-32470

Explained as:>

The output you provided is the result of conducting a One Sample t-test in R. The test is performed on a dataset called "differences." Let's break down the results:

Data: The t-test is conducted on the dataset called "differences."

t-value: The t-value is -18.33. It measures how many standard errors the sample mean is away from the hypothesized population mean (in this case, the hypothesized mean is 0). A large absolute t-value indicates a significant difference between the sample mean and the hypothesized mean.

Degrees of Freedom (df): The degrees of freedom is 4. It represents the number of independent pieces of information available to estimate a statistic. In this case, the degrees of freedom is relatively small, suggesting a small sample size.

p-value: The p-value is 5.212e-05 (0.00005212). It represents the probability of observing the t-value (or a more extreme value) if the true mean of the population is equal to the hypothesized mean (0 in this case). A smaller p-value indicates strong evidence against the null hypothesis (i.e., the hypothesis that the true mean is equal to 0).

Null Hypothesis: The null hypothesis states that the true mean of the population is equal to 0.

Alternative Hypothesis: The alternative hypothesis is a two-tailed hypothesis, indicating that the true mean of the population is not equal to 0.

95% Confidence Interval: The 95% confidence interval for the true mean is calculated to be between -37388.35 and -27551.65. This means that we can be 95% confident that the true mean of the population falls within this range.

Sample Estimate: The mean of the dataset "differences" is -32470.

In summary, the results of the One Sample t-test indicate that there is strong evidence to reject the null hypothesis. The data suggests that the true mean of the population is not equal to 0. The large negative t-value and small p-value support the conclusion that the sample mean is significantly different from 0. The 95% confidence interval also shows that the true mean likely lies within the range of -37388.35 to -27551.65.

1. Result on correlation

Correlation between ModelYear and ElectricRange: -0.2102131

Correlation between ModelYear and BaseMSRP: -0.860663

Correlation between ElectricRange and BaseMSRP: 0.1342485

Correlation between ModelYear and LegislativeDistrict: 0.0204586

**5.Result on p-value**

P-value for correlation between ModelYear and ElectricRange: 0.5599459

P-value for correlation between ModelYear and BaseMSRP: 0.001389044

P-value for correlation between ElectricRange and BaseMSRP: 0.7115672

P-value for correlation between ModelYear and LegislativeDistrict: 0.9552655

This is the script we used to do this work.

library(readr)

ELECTRIC\_VEHICLE\_POPULATION\_ <- read\_csv("C:/Users/Prince Ademo/Desktop/ELECTRIC VEHICLE POPULATION .CSV")

View(ELECTRIC\_VEHICLE\_POPULATION\_)

data1<-c(ELECTRIC\_VEHICLE\_POPULATION\_)

summary(ELECTRIC\_VEHICLE\_POPULATION\_)

library(ggplot2)

# Load required libraries

library(ggplot2)

# Create a data frame with the provided data

data1 <- data.frame(County = c("Thurston", "Island", "Snohomish", "King", "Snohomish", "Chelan", "Snohomish", "Snohomish", "Thurston", "Thurston","Snohomish", "King", "Kitsap", "Snohomish", "Snohomish", "Yakima", "Chelan", "Snohomish", "Thurston", "Snohomish", "Thurston", "Snohomish", "Kitsap", "Snohomish", "Kitsap", "Kitsap", "Kitsap", "Kitsap", "Kitsap", "Snohomish","Kitsap", "Snohomish", "Marin", "Snohomish", "Snohomish", "Whitman", "Island", "Snohomish", "Thurston", "Snohomish", "Snohomish", "Kitsap", "Thurston", "Snohomish", "Kitsap", "Snohomish", "Kitsap", "King", "Thurston", "Snohomish","Kitsap", "Thurston", "Thurston", "Island", "Snohomish", "Snohomish", "Kitsap", "Kitsap", "Thurston", "Snohomish","Snohomish", "Kitsap", "Snohomish", "Snohomish", "Kitsap", "Kitsap", "King", "Snohomish", "Chelan", "Thurston","King", "Snohomish", "Yakima", "Snohomish", "Snohomish", "Kitsap", "Kitsap", "Thurston", "King", "Snohomish","Whitman", "Kitsap", "Snohomish", "Snohomish", "Thurston", "Snohomish", "King", "Snohomish", "Snohomish","Snohomish", "Snohomish", "Douglas", "Snohomish", "Snohomish", "Kitsap", "Kitsap", "Yakima", "Snohomish","Thurston", "King", "Snohomish", "Whitman", "Kitsap", "Snohomish", "Douglas", "Thurston", "Snohomish","Snohomish", "Kitsap", "Kitsap", "Yakima", "Snohomish", "Snohomish", "Snohomish", "Snohomish", "Douglas", "Thurston", "King", "Snohomish", "Snohomish", "Thurston", "King", "Snohomish", "Whitman", "Kitsap", "Kitsap","Thurston", "King", "Snohomish", "Kitsap", "Snohomish", "Kitsap", "Kitsap", "Snohomish", "Kitsap", "Thurston","Snohomish", "Snohomish", "Snohomish", "Thurston", "Snohomish", "Snohomish", "Thurston", "Snohomish", "Snohomish","Snohomish", "Snohomish", "King", "Thurston", "Snohomish", "Kitsap", "Thurston", "Thurston", "Island","Snohomish", "Snohomish", "Kitsap", "Kitsap", "Thurston", "Snohomish", "Snohomish", "Kitsap", "Snohomish","Snohomish", "Kitsap", "Kitsap", "King", "Snohomish", "Chelan", "Thurston", "King", "Snohomish", "Yakima","Snohomish", "Snohomish", "Kitsap", "Thurston", "Snohomish", "Kitsap", "Snohomish", "Kitsap", "King","Thurston", "Snohomish", "Kitsap", "Thurston", "Thurston"))

# Generate the box plot

ggplot(data1, aes(x = "", y = County)) +geom\_boxplot() +labs(x = "", y = "County")+ggtitle("Bar Plot of County")

data1 <- data.frame(County = c("Thurston", "Island", "Snohomish", "King", "Snohomish", "Chelan", "Snohomish", "Snohomish", "Thurston", "Thurston"),City = c("Tumwater", "Clinton", "Snohomish", "Seattle", "Edmonds", "Manson", "Marysville", "Edmonds", "Olympia", "Lacey"),State = c("WA", "WA", "WA", "WA", "WA", "WA", "WA", "WA", "WA", "WA"),PostalCode = c(98512, 98236, 98290, 98134, 98020, 98831, 98271, 98026, 98501, 98503),ModelYear = c(2019, 2022, 2020, 2020, 2013, 2018, 2020, 2021, 2019, 2018),Make = c("TESLA", "NISSAN", "TESLA", "TESLA", "TESLA", "BMW", "TESLA", "AUDI", "TOYOTA", "NISSAN"),Model = c("MODEL 3", "LEAF", "MODEL 3", "MODEL 3", "MODEL S", "I3", "MODEL 3", "Q5 E", "PRIUS PRIME", "LEAF"),ElectricVehicleType = c("Battery Electric Vehicle (BEV)", "Battery Electric Vehicle (BEV)", "Battery Electric Vehicle (BEV)", "Battery Electric Vehicle (BEV)", "Battery Electric Vehicle (BEV)", "Plug-in Hybrid Electric Vehicle (PHEV)", "Battery Electric Vehicle (BEV)", "Plug-in Hybrid Electric Vehicle (PHEV)", "Plug-in Hybrid Electric Vehicle (PHEV)", "Battery Electric Vehicle (BEV)"),CAFVEligibility = c("Clean Alternative Fuel Vehicle Eligible", "Eligibility unknown as battery range has not been researched", "Clean Alternative Fuel Vehicle Eligible", "Clean Alternative Fuel Vehicle Eligible", "Clean Alternative Fuel Vehicle Eligible", "Clean Alternative Fuel Vehicle Eligible", "Clean Alternative Fuel Vehicle Eligible", "Not eligible due to low battery range", "Not eligible due to low battery range", "Clean Alternative Fuel Vehicle Eligible"),ElectricRange = c(220, 0, 266, 322, 208, 97, 308, 18, 25, 151),BaseMSRP = c(0, 0, 0, 0, 69900, 0, 0, 0, 0, 0),LegislativeDistrict = c(22, 10, 44, 11, 21, 12, 38, 21, 22, 22),DOLVehicleID = c(242565116, 183272785, 112552366, 6336319, 186212960, 215122904, 110992472, 138909032, 272310279, 235573929),ElectricUtility = c("PUGET SOUND ENERGY INC", "PUGET SOUND ENERGY INC", "PUGET SOUND ENERGY INC", "CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)", "PUGET SOUND ENERGY INC", "PUD NO 1 OF CHELAN COUNTY", "PUGET SOUND ENERGY INC", "PUGET SOUND ENERGY INC", "PUGET SOUND ENERGY INC", "PUGET SOUND ENERGY INC"),CensusTract2020 = c(53067010910, 53029972000, 53061052502, 53033009300, 53061050403, 53007960400, 53061053102, 53061050300, 53067010700, 53067011422))

boxplot(data1$ElectricRange,col ="yellow",main ="ElectricRange")

# Example data

data <- c("Battery Electric Vehicle (BEV)", "Plug-in Hybrid Electric Vehicle (PHEV)", "Other")

# Create a pie chart

pie(table(data),main = "Electric Vehicle Types",col = c("blue", "green", "violet"),labels = paste(names(table(data)), ": ", table(data)),border = "white")

library(ggplot2)

# Create a data frame with the given data

data <- data.frame(County = c("Thurston", "Island", "Snohomish", "King", "Snohomish"),City = c("Tumwater", "Clinton", "Snohomish", "Seattle", "Edmonds"),State = c("WA", "WA", "WA", "WA", "WA"),PostalCode = c(98512, 98236, 98290, 98134, 98020),ModelYear = c(2019, 2022, 2020, 2020, 2013),Make = c("TESLA", "NISSAN", "TESLA", "TESLA", "TESLA"),Model = c("MODEL 3", "LEAF", "MODEL 3", "MODEL 3", "MODEL S"),ElectricVehicleType = c("Battery Electric Vehicle (BEV)", "Battery Electric Vehicle (BEV)", "Battery Electric Vehicle (BEV)", "Battery Electric Vehicle (BEV)", "Battery Electric Vehicle (BEV)"),CleanAlternativeFuelVehicleCAFVEligibility = c("Clean Alternative Fuel Vehicle Eligible", "Eligibility unknown as battery range has not been researched", "Clean Alternative Fuel Vehicle Eligible", "Clean Alternative Fuel Vehicle Eligible", "Clean Alternative Fuel Vehicle Eligible"),ElectricRange = c(220, 0, 266, 322, 208),BaseMSRP = c(0, 0, 0, 0, 69900),LegislativeDistrict = c(22, 10, 44, 11, 21),DOLVehicleID = c(242565116, 183272785, 112552366, 6336319, 186212960),ElectricUtility = c("PUGET SOUND ENERGY INC", "PUGET SOUND ENERGY INC", "PUGET SOUND ENERGY INC", "CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)", "PUGET SOUND ENERGY INC"),CensusTract2020 = c(53067010910, 53029972000, 53061052502, 53033009300, 53061050403))

# Plot a histogram of the "Base MSRP" column

ggplot(data, aes(x = BaseMSRP)) +geom\_histogram(binwidth = 10000, fill = "skyblue", color = "black") +labs(title = "Histogram of Base MSRP", x = "Base MSRP", y = "Count")

library(ggplot2)

# Create a data frame with the given data

data <- data.frame(County = c("Thurston", "Island", "Snohomish", "King", "Snohomish"),City = c("Tumwater", "Clinton", "Snohomish", "Seattle", "Edmonds"),Electric\_Range = c(220, 0, 266, 322, 208),Base\_MSRP = c(0, 0, 0, 0, 69900))

# Create the scatterplot

ggplot(data, aes(x = Electric\_Range, y = Base\_MSRP)) +geom\_point() +labs(x = "Electric Range", y = "Base MSRP") +ggtitle("Electric Vehicle Range vs Base MSRP")

# Import the data

ELECTRIC\_VEHICLE\_POPULATION <- data.frame(ELECTRIC\_VEHICLE\_POPULATION\_)

ELECTRIC\_VEHICLE\_POPULATION <- ELECTRIC\_VEHICLE\_POPULATION %>% summarize(Mean\_Range = mean(Electric.Range), Median\_Range = median(Electric.Range), Min\_Range = min(Electric.Range), Max\_Range = max(Electric.Range), Mean\_MSRP = mean(Base.MSRP), Median\_MSRP = median(Base.MSRP), Min\_MSRP = min(Base.MSRP), Max\_MSRP = max(Base.MSRP))

print(ELECTRIC\_VEHICLE\_POPULATION)

# Load the necessary library

library(tidyverse)

# Create a sample dataset

electric\_vehicle\_data <- data.frame(Electric\_Vehicle\_Type = c("BEV", "BEV", "PHEV", "PHEV", "BEV", "PHEV", "PHEV", "BEV", "PHEV", "BEV"),Electric\_Range = c(250, 300, 40, 50, 220, 80, 100, 180, 90, 200))

# Split the dataset into two groups: BEVs and PHEVs

bev\_data <- electric\_vehicle\_data %>% filter(Electric\_Vehicle\_Type == "BEV")

phev\_data <- electric\_vehicle\_data %>% filter(Electric\_Vehicle\_Type == "PHEV")

# Perform an independent t-test

t\_test\_result <- t.test(bev\_data$Electric\_Range, phev\_data$Electric\_Range)

# Print the t-test result

print(t\_test\_result)

# Load the necessary library

library(stats)

# Create a data frame with the paired observations of Electric Range and Base MSRP

data <- data.frame(Electric\_Range = c(120, 130, 110, 140, 150),Base\_MSRP = c(30000, 35000, 28000, 32000, 38000))

# Calculate the differences between Electric Range before and after Base MSRP increase

differences <- data$Electric\_Range - data$Base\_MSRP

# Perform the paired t-test

result <- t.test(differences)

# Print the result

print(result)

# Create a data frame with the provided dataset

data <- data.frame(County = c("Thurston", "Island", "Snohomish", "King", "Snohomish", "Chelan", "Snohomish", "Snohomish", "Thurston", "Thurston"),City = c("Tumwater", "Clinton", "Snohomish", "Seattle", "Edmonds", "Manson", "Marysville", "Edmonds", "Olympia", "Lacey"),State = c("WA", "WA", "WA", "WA", "WA", "WA", "WA", "WA", "WA", "WA"),PostalCode = c(98512, 98236, 98290, 98134, 98020, 98831, 98271, 98026, 98501, 98503),ModelYear = c(2019, 2022, 2020, 2020, 2013, 2018, 2020, 2021, 2019, 2018),Make = c("TESLA", "NISSAN", "TESLA", "TESLA", "TESLA", "BMW", "TESLA", "AUDI", "TOYOTA", "NISSAN"),Model = c("MODEL 3", "LEAF", "MODEL 3", "MODEL 3", "MODEL S", "I3", "MODEL 3", "Q5 E", "PRIUS PRIME", "LEAF"),ElectricVehicleType = c("Battery Electric Vehicle (BEV)", "Battery Electric Vehicle (BEV)", "Battery Electric Vehicle (BEV)", "Battery Electric Vehicle (BEV)", "Battery Electric Vehicle (BEV)", "Plug-in Hybrid Electric Vehicle (PHEV)", "Battery Electric Vehicle (BEV)", "Plug-in Hybrid Electric Vehicle (PHEV)", "Plug-in Hybrid Electric Vehicle (PHEV)", "Battery Electric Vehicle (BEV)"),CAFVEligibility = c("Clean Alternative Fuel Vehicle Eligible", "Eligibility unknown as battery range has not been researched", "Clean Alternative Fuel Vehicle Eligible", "Clean Alternative Fuel Vehicle Eligible", "Clean Alternative Fuel Vehicle Eligible", "Clean Alternative Fuel Vehicle Eligible", "Clean Alternative Fuel Vehicle Eligible", "Not eligible due to low battery range", "Not eligible due to low battery range", "Clean Alternative Fuel Vehicle Eligible"),ElectricRange = c(220, 0, 266, 322, 208, 97, 308, 18, 25, 151),BaseMSRP = c(0, 0, 0, 0, 69900, 0, 0, 0, 0, 0),LegislativeDistrict = c(22, 10, 44, 11, 21, 12, 38, 21, 22, 22),DOLVehicleID = c(242565116, 183272785, 112552366, 6336319, 186212960, 215122904, 110992472, 138909032, 272310279, 235573929),ElectricUtility = c("PUGET SOUND ENERGY INC", "PUGET SOUND ENERGY INC", "PUGET SOUND ENERGY INC", "CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)", "PUGET SOUND ENERGY INC", "PUD NO 1 OF CHELAN COUNTY", "PUGET SOUND ENERGY INC", "PUGET SOUND ENERGY INC", "PUGET SOUND ENERGY INC", "PUGET SOUND ENERGY INC"),CensusTract = c(53067010910, 53029972000, 53061052502, 53033009300, 53061050403, 53007960400, 53061053102, 53061050300, 53067010700, 53067011422))

# Perform correlational analysis on numeric variables

correlation <- cor(data[, c("ModelYear", "ElectricRange", "BaseMSRP", "LegislativeDistrict")], use = "complete.obs")

print(correlation)

library(readxl)

data1 <- read\_csv("ELECTRIC\_VEHICLE\_POPULATION")

view(data1)

subset\_data <- data1[, c("ModelYear", "ElectricRange", "BaseMSRP")]

# Import the data

data3<- data.frame(ELECTRIC\_VEHICLE\_POPULATION)

# Perform correlation analysis

correlations <- cor(data3[, c("ModelYear", "ElectricRange", "BaseMSRP")])

# Print the correlation matrix

print(correlations)

colnames(data2)

correlation\_matrix <- cor(data1[c("ModelYear", "ElectricRange", "BaseMSRP","LegislativeDistrict")])

library(readxl)

# Read the Excel file

data1 <- read\_excel("ELECTRIC VEHICLE POPULATION.xlsx")

# Calculate correlation coefficients

correlation\_matrix <- cor(data1[c("ModelYear", "ElectricRange", "BaseMSRP", )])

# Extract the correlation coefficients

cor\_model\_year\_range <- correlation\_matrix["ModelYear", "ElectricRange"]

cor\_model\_year\_msrp <- correlation\_matrix["ModelYear", "BaseMSRP"]

cor\_range\_msrp <- correlation\_matrix["ElectricRange", "BaseMSRP"]

cor\_model\_year\_district <- correlation\_matrix["ModelYear", "LegislativeDistrict"]

# Print the correlation coefficients

cat("Correlation between ModelYear and ElectricRange:", cor\_model\_year\_range, "\n")

cat("Correlation between ModelYear and BaseMSRP:", cor\_model\_year\_msrp, "\n")

cat("Correlation between ElectricRange and BaseMSRP:", cor\_range\_msrp, "\n")

cat("Correlation between ModelYear and LegislativeDistrict:", cor\_model\_year\_district, "\n")

# Test the significance of the correlations

p\_value\_model\_year\_range <- cor.test(data1$ModelYear, data1$ElectricRange)$p.value

p\_value\_model\_year\_msrp <- cor.test(data1$ModelYear, data1$BaseMSRP)$p.value

p\_value\_range\_msrp <- cor.test(data1$ElectricRange, data1$BaseMSRP)$p.value

p\_value\_model\_year\_district <- cor.test(data1$ModelYear, data1$LegislativeDistrict)$p.value

# Print the p-values

cat("P-value for correlation between ModelYear and ElectricRange:", p\_value\_model\_year\_range, "\n")

cat("P-value for correlation between ModelYear and BaseMSRP:", p\_value\_model\_year\_msrp, "\n")

cat("P-value for correlation between ElectricRange and BaseMSRP:", p\_value\_range\_msrp, "\n")

cat("P-value for correlation between ModelYear and LegislativeDistrict:", p\_value\_model\_year\_district, "\n")