Assignment 2

Manas Kore

2024-06-16

# Reading the data into R

library(readr) # loading the readr package

# Use read\_csv() function to read the data

ev\_data <- read\_csv(“/Users/mkore/Downloads/Electric\_Vehicle\_Population\_Data.csv”)

# Cleaning the data

library(dplyr) # loading the dplyr package

# Renaming columns for easier reference

ev\_data <- ev\_data %>% rename( VIN = VIN (1-10), County = County, City = City, State = State, Postal\_Code = Postal Code, Model\_Year = Model Year, Make = Make, Model = Model, EV\_Type = Electric Vehicle Type, CAFV\_Eligibility = Clean Alternative Fuel Vehicle (CAFV) Eligibility, Electric\_Range = Electric Range, Base\_MSRP = Base MSRP, Legislative\_District = Legislative District, DOL\_Vehicle\_ID = DOL Vehicle ID, Vehicle\_Location = Vehicle Location, Electric\_Utility = Electric Utility, Census\_Tract = 2020 Census Tract )

# Getting the number of rows and columns

num\_rows <- nrow(ev\_data) num\_cols <- ncol(ev\_data)

# Create a data frame for column descriptions

column\_descriptions <- data.frame( Column\_Name = colnames(ev\_data), Description = c( “Vehicle Identification Number (first 10 characters)”, “County of vehicle registration”, “City of vehicle registration”, “State of vehicle registration”, “Postal code of vehicle registration”, “Year of the vehicle model”, “Manufacturer of the vehicle”, “Model of the vehicle”, “Type of electric vehicle (BEV or PHEV)”, “Eligibility for Clean Alternative Fuel Vehicle”, “Electric range of the vehicle in miles”, “Manufacturer’s Suggested Retail Price”, “Legislative district of vehicle registration”, “Unique vehicle ID assigned by the Department of Licensing”, “Location of the vehicle in geographical coordinates”, “Electric utility provider”, “2020 Census tract code” ) )

# Display the table using kable

library(knitr) kable(column\_descriptions, col.names = c(“Column Name”, “Description”))

# Selecting three columns for summary statistics

selected\_columns <- ev\_data %>% select(Electric\_Range, Base\_MSRP, Model\_Year)

# Calculating summary statistics

summary\_stats <- selected\_columns %>% summarise( Min\_Range = min(Electric\_Range, na.rm = TRUE), Max\_Range = max(Electric\_Range, na.rm = TRUE), Mean\_Range = mean(Electric\_Range, na.rm = TRUE), Missing\_Range = sum(is.na(Electric\_Range)),

Min\_MSRP = min(Base\_MSRP, na.rm = TRUE),  
Max\_MSRP = max(Base\_MSRP, na.rm = TRUE),  
Mean\_MSRP = mean(Base\_MSRP, na.rm = TRUE),  
Missing\_MSRP = sum(is.na(Base\_MSRP)),  
  
Min\_Year = min(Model\_Year, na.rm = TRUE),  
Max\_Year = max(Model\_Year, na.rm = TRUE),  
Mean\_Year = mean(Model\_Year, na.rm = TRUE),  
Missing\_Year = sum(is.na(Model\_Year))

)

summary\_stats