Week 5 Activity

Manas Kore

2024-06-23

knitr::opts\_chunk$set(echo = TRUE)  
  
# Reading the data into R  
library(readr) # loading the readr package  
library(dplyr) # loading the dplyr package

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2) # loading the ggplot2 package for visualizations  
library(knitr) # loading the knitr package  
library(scales) # loading the scales package for formatting

##   
## Attaching package: 'scales'

## The following object is masked from 'package:readr':  
##   
## col\_factor

knitr::opts\_chunk$set(echo = TRUE)  
  
# Reading the data into R  
library(readr) # loading the readr package

knitr::opts\_chunk$set(echo = TRUE)  
# Use read\_csv() function to read the data  
ev\_data <- read\_csv("/Users/mkore/Downloads/Electric\_Vehicle\_Population\_Data.csv")

## Rows: 186879 Columns: 17  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (10): VIN (1-10), County, City, State, Make, Model, Electric Vehicle Typ...  
## dbl (7): Postal Code, Model Year, Electric Range, Base MSRP, Legislative Di...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

knitr::opts\_chunk$set(echo = TRUE)  
# Cleaning the data  
library(dplyr) # loading the dplyr package

knitr::opts\_chunk$set(echo = TRUE)  
# 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`  
 )

knitr::opts\_chunk$set(echo = TRUE)  
# Getting the number of rows and columns  
num\_rows <- nrow(ev\_data)  
num\_cols <- ncol(ev\_data)

We have 186879 rows and 17columns in the data frame

knitr::opts\_chunk$set(echo = TRUE)  
# 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"  
 )  
)

knitr::opts\_chunk$set(echo = TRUE)  
# Display the table using kable  
library(knitr)  
kable(column\_descriptions, col.names = c("Column Name", "Description"))

| Column Name | Description |
| --- | --- |
| VIN | Vehicle Identification Number (first 10 characters) |
| County | County of vehicle registration |
| City | City of vehicle registration |
| State | State of vehicle registration |
| Postal\_Code | Postal code of vehicle registration |
| Model\_Year | Year of the vehicle model |
| Make | Manufacturer of the vehicle |
| Model | Model of the vehicle |
| EV\_Type | Type of electric vehicle (BEV or PHEV) |
| CAFV\_Eligibility | Eligibility for Clean Alternative Fuel Vehicle |
| Electric\_Range | Electric range of the vehicle in miles |
| Base\_MSRP | Manufacturer’s Suggested Retail Price |
| Legislative\_District | Legislative district of vehicle registration |
| DOL\_Vehicle\_ID | Unique vehicle ID assigned by the Department of Licensing |
| Vehicle\_Location | Location of the vehicle in geographical coordinates |
| Electric\_Utility | Electric utility provider |
| Census\_Tract | 2020 Census tract code |

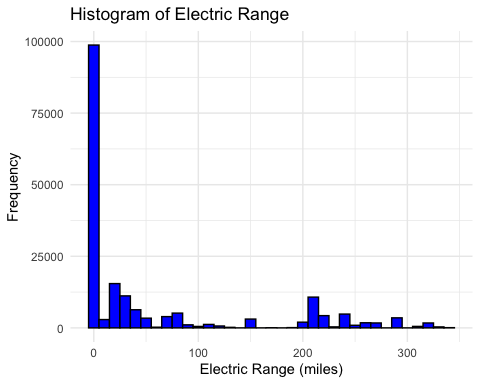
knitr::opts\_chunk$set(echo = TRUE)  
# Selecting three columns for summary statistics  
selected\_columns <- ev\_data %>%  
 select(Electric\_Range, Base\_MSRP, Model\_Year)

knitr::opts\_chunk$set(echo = TRUE)  
  
  
# 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))  
 )

knitr::opts\_chunk$set(echo = TRUE)  
# Displaying summary statistics in a table  
kable(summary\_stats, col.names = c("Min Range", "Max Range", "Mean Range", "Missing Range", "Min MSRP", "Max MSRP", "Mean MSRP", "Missing MSRP", "Min Year", "Max Year", "Mean Year", "Missing Year"))

| Min Range | Max Range | Mean Range | Missing Range | Min MSRP | Max MSRP | Mean MSRP | Missing MSRP | Min Year | Max Year | Mean Year | Missing Year |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 337 | 56.70779 | 0 | 0 | 845000 | 1011.918 | 0 | 1997 | 2024 | 2020.66 | 0 |

knitr::opts\_chunk$set(echo = TRUE)  
  
ggplot(ev\_data, aes(x = Electric\_Range)) +  
 geom\_histogram(binwidth = 10, fill = "blue", color = "black") +  
 labs(title = "Histogram of Electric Range", x = "Electric Range (miles)", y = "Frequency") +  
 theme\_minimal()



knitr::opts\_chunk$set(echo = TRUE)  
  
ggplot(ev\_data, aes(x = Base\_MSRP, y = Electric\_Range)) +  
 geom\_point(alpha = 0.6) +  
 labs(title = "Scatterplot of Base MSRP vs. Electric Range", x = "Base MSRP (USD)", y = "Electric Range (miles)") +  
 scale\_x\_continuous(labels = scales::label\_dollar(), breaks = seq(0, 800000, by = 150000)) +  
 theme\_minimal()

