Exploring EV Data

Samuel Adebayo, Jack Kester

Introduction

In this analysis, we will use a CSV file from data.gov to explore certain properties of electric vehicles (EV's). The main things that we will be looking at include Make, Model, and Electric Range (in Miles).

It is important to note that some of the vehicles on this list are hybrid vehicles, meaning that they run on both electricity and gas. This detail will come up later on in the text.

Question 1

Research Question Q1: Is there any significant difference in vehicle range across manufacturers?

To do this, I am simply going to make a histogram that shows each manufacturer along with their average range on their electric vehicles.
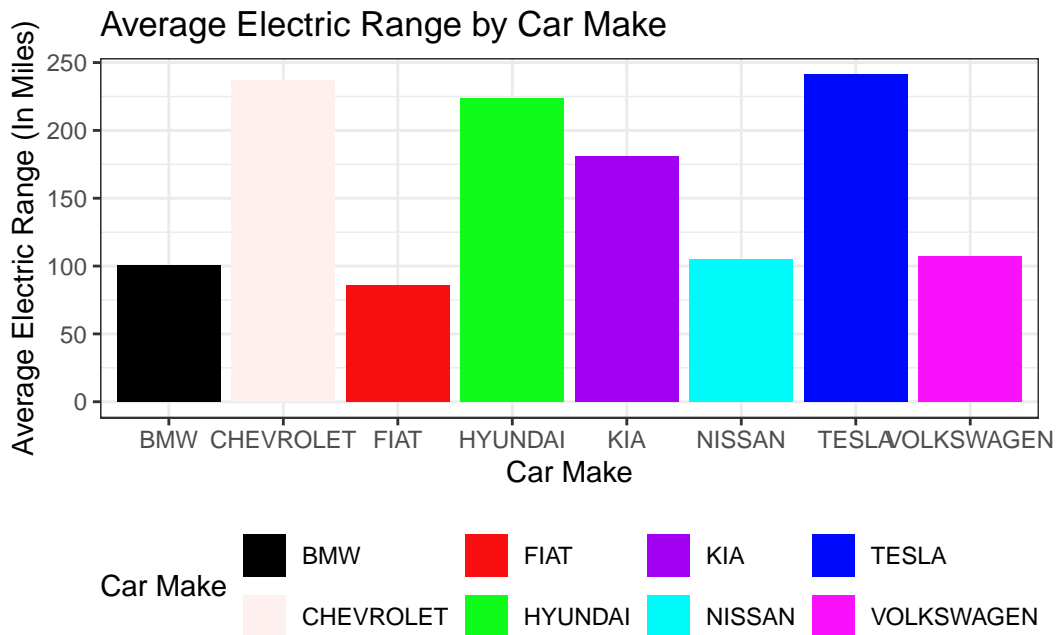

Response:

Q1: Based on the histogram made for question 1, it is safe to assume that there is a vast difference between an EV's make and its range in miles.

As seen in the graph, companies like Hyundai, Chevy, Tesla and KIA all make EV's that can comfortably drive over 150 miles on one charge, with the first 3 being capable of driving around 200 miles on one charge.

As noted in the introduction there are hybrid vehicles in the data set, meaning that some of the values for manufacturers like BMW and Nissan have their averages dragged down due to that fact.

Question 2

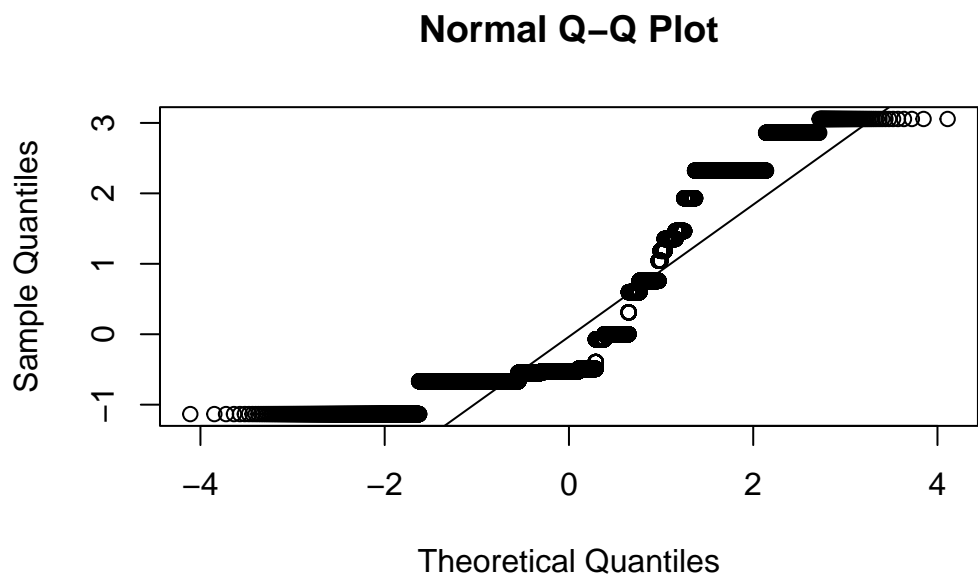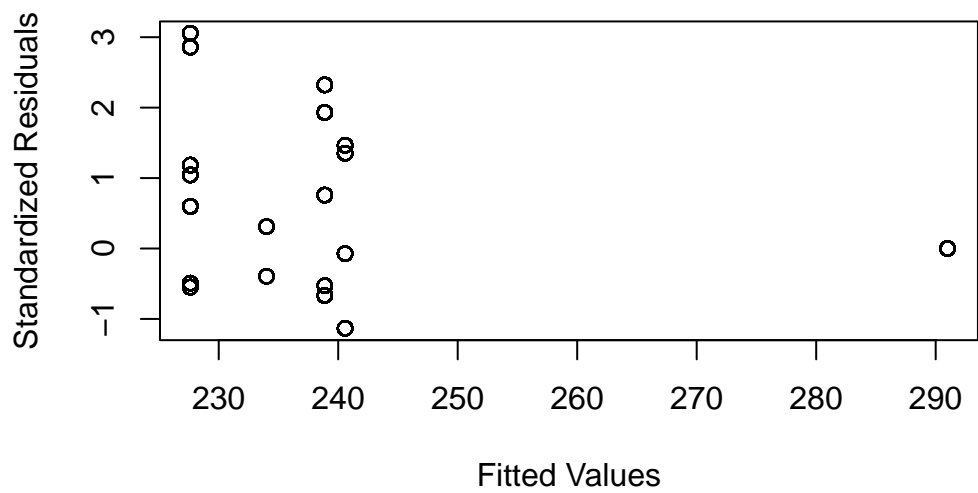Research Questions: Q2: Is there a difference in average electric range between the Tesla models?
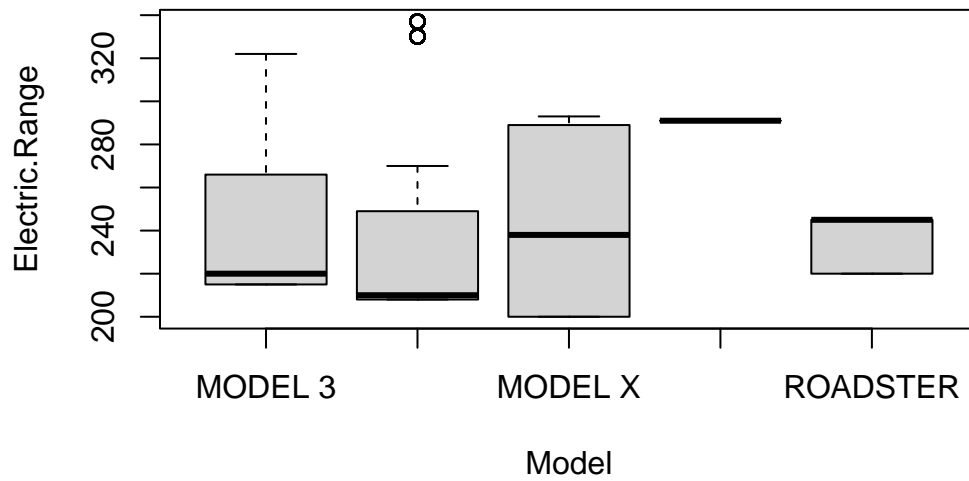

```
             Df   Sum Sq Mean Sq F value Pr(>F)
Model         4  6903021 1725755    1347 <2e-16 ***
Residuals 25299 32408761    1281
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Normal Q–Q Plot

Research Questions: Is there a difference in average electric range between the Tesla models?

Q2: Anova Hypothesis Test (At 5% sig level) (1) $H_0 : \mu_1 = \mu_2 = \mu_3$ $H_1 : Some \ \mu_i \neq \mu_k$

(2)

```
               Df    Sum Sq Mean Sq F value Pr(>F)
Model           4   6903021 1725755    1347 <2e-16 ***
Residuals   25299 32408761    1281
```

(3) Validate Assumptions Independence: Cars aren't randomly assigned Models because they are manufactured specifically as that model, however we can assume cars were randomly chosen to be tested so we can validate independence

Normality: Normality tails off on the sides, however slightly follows a normal distribution

Constant Variance: Variance does not randomly disperse around 0 so cannot really validate variance.

Assumptions cannot be fully validated, however, to respond properly to the question we will continue

(4) P-Value = $P(F >= 1347) = [(<\text{2e-16}) < 0.05$ (significance level)$)]$

Because our p-value is small ($<5\%$), we can conclude that at least one mean electric range is different than the others.

Appendix

Code sourced from: https://catalog.data.gov/dataset/electric-vehicle-population-data