Data Analysis Sample Project

Exploring a classic ‘canned’ data set to determine which car manufacture has the ‘best’ efficiency. The study breaks the analysis into two parts – city and highway – before combining the results into a final conclusion.

1. Manufacturer and Highway MPG

Starting by looking at the median highway MPG per manufacturer reveals that Honda and Volkswagen have the best highway MPG, while Jeep, Ford, Mercury, Dodge, Lincoln and Land Rover have the worst.

## Manufacturer Highway\_MPG  
## honda 32.0  
## volkswagen 29.0  
## hyundai 26.5  
## audi 26.0  
## nissan 26.0  
## pontiac 26.0  
## subaru 26.0  
## toyota 26.0  
## chevrolet 23.0  
## jeep 18.5  
## ford 18.0  
## mercury 18.0  
## dodge 17.0  
## lincoln 17.0  
## land rover 16.5

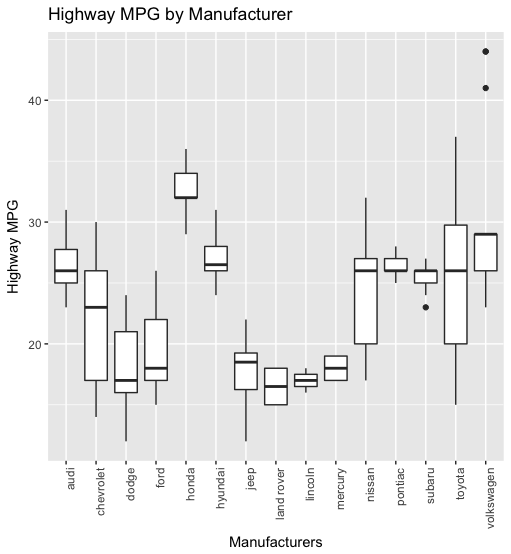
In order to determine the “worst” and “best”, the study calculated a range of ‘significant’ values – those more than one SD away from the median (24 MPG) for all manufacturers (18.04, 29.95).

Visualizing the results reaffirms that Honda has the most fuel-efficient cars on average. while Volkswagen has several models that surpass the higher end of the Honda range (represented by the dots on the plot below). Moreover, with a median is in the upper range of its IQR, it appears that a majority of the Volkswagen models are high efficiency – even though the median MPG is barely within the study’s ‘significance zone’.

While Land Rover has the lowest median, Dodge and Jeep offer the least fuel-efficient cars – denoted by the long whiskers of the plot.

Toyota and Chevrolet have the widest spread of MPG -- suggesting they offer the widest mix of car-type offerings (truck, compact, sports, etc.).

Conversely, Subaru, Lincoln and Pontiac have the narrowest IQR – signaling a narrow offering of car-type.



Turning to a simple regression, Honda (6.1) and Volkswagen (2.7) have a positive relationship with fuel-efficiency – increasing highway MPG compared to the overall average.

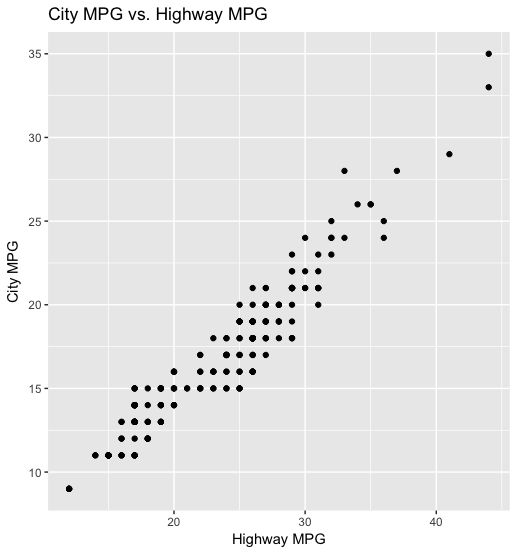
Hyundai, Nissan, Pontiac, Subaru, and Toyota are do not have significant coefficients – meaning that they offer average fuel-efficiency.

Mercury, Lincoln, Land rover, Jeep, Ford, Dodge and Chevrolet have a negative relationship with highway MPG, with Jeep (-8.8), Lincoln (-9.4) and Land Rover (-9.9) being the worst of the group.

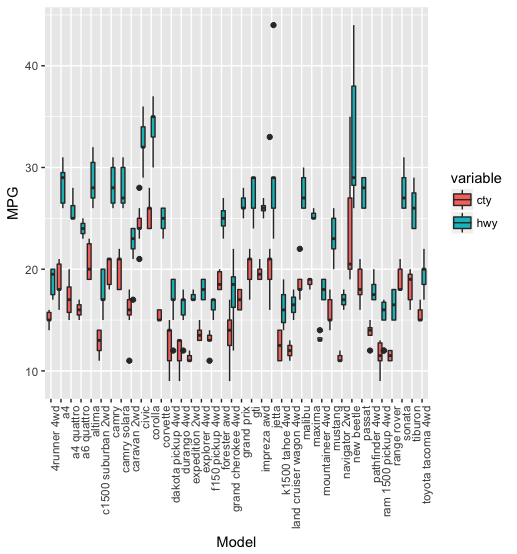
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 26.44444 0.98206 26.928 < 2e-16 \*\*\*  
## manufacturerchevrolet -4.54971 1.37044 -3.320 0.001055 \*\*   
## manufacturerdodge -8.49850 1.19734 -7.098 1.74e-11 \*\*\*  
## manufacturerford -7.08444 1.28796 -5.501 1.05e-07 \*\*\*  
## manufacturerhonda 6.11111 1.70097 3.593 0.000404 \*\*\*  
## manufacturerhyundai 0.41270 1.48473 0.278 0.781304   
## manufacturerjeep -8.81944 1.77043 -4.982 1.28e-06 \*\*\*  
## manufacturerland rover -9.94444 2.30313 -4.318 2.39e-05 \*\*\*  
## manufacturerlincoln -9.44444 2.59828 -3.635 0.000347 \*\*\*  
## manufacturermercury -8.44444 2.30313 -3.667 0.000309 \*\*\*  
## manufacturernissan -1.82906 1.51651 -1.206 0.229082   
## manufacturerpontiac -0.04444 2.10628 -0.021 0.983184   
## manufacturersubaru -0.87302 1.48473 -0.588 0.557141   
## manufacturertoyota -1.53268 1.21450 -1.262 0.208298   
## manufacturervolkswagen 2.77778 1.26783 2.191 0.029509 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

1. Model, Highway MPG and City MPG

Highway MPG and City MPG are measures of fuel-efficiency that rely on many of the same factors. As such, the two terms are highly positively correlated (0.95). This suggests that the best performing models on the highway should also be the best performing models on city streets.



Looking at a side-by-side boxplot of city MPG and highway MPG by model seems to reaffirm this – with those models with the higher highway fuel-efficiency having higher city MPG.



However, this plot is overly-crowded and again leaves the question of which models are quantitatively the best and worst. To solve this issue, the study again created a ‘significant range’ of one SD above the median value for city MPG (12.7, 21.2) and highway MPG (18.04, 29.95). Those models deemed the best are ones that have MPGs above the upper bounds of both range, with the worst models possessing MPGs that fall beneath the lower bounds of both ranges.

With this criteria, Civic and Corolla are the best models while the Expedition, Land Cruiser, Navigator, Ram 1500 Pickup and Range Rover are the worst performing models.

Overall, Corolla is the most efficient and the Ram 1500 Pickup 4wd being the least efficient.

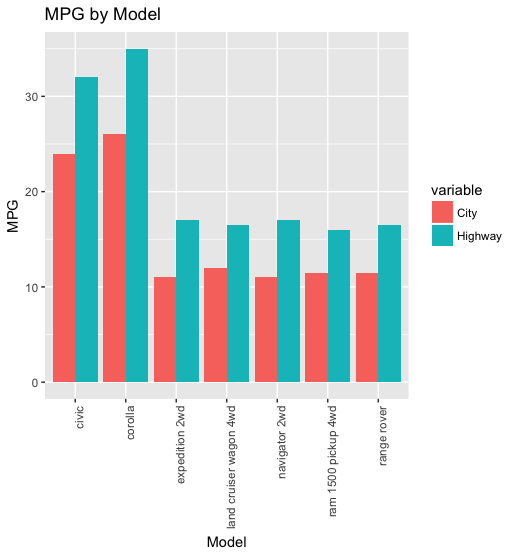
best

## Model City Highway Hwy - Cty  
## 10 civic 24 32 8  
## 11 corolla 26 35 9

worst

## Model City Highway Hwy - Cty  
## 15 expedition 2wd 11.0 17.0 6.0  
## 25 land cruiser wagon 4wd 12.0 16.5 4.5  
## 30 navigator 2wd 11.0 17.0 6.0  
## 34 ram 1500 pickup 4wd 11.5 16.0 4.5  
## 35 range rover 11.5 16.5 5.0

The best cars when only considering Highway MPG or only considering City MPG remained Civic and Corolla – reaffirming the high correlation between the two variables.



The worst cars when only considering City MPG expands to:

worst\_highway

## Model City Highway Hwy - Cty  
## 6 c1500 suburban 2wd 13.0 17.0 4.0  
## 13 dakota pickup 4wd 14.0 17.0 3.0  
## 14 durango 4wd 13.0 17.0 4.0  
## 15 expedition 2wd 11.0 17.0 6.0  
## 17 f150 pickup 4wd 13.0 17.0 4.0  
## 24 k1500 tahoe 4wd 12.5 16.0 3.5  
## 25 land cruiser wagon 4wd 12.0 16.5 4.5  
## 30 navigator 2wd 11.0 17.0 6.0  
## 33 pathfinder 4wd 14.0 17.5 3.5  
## 34 ram 1500 pickup 4wd 11.5 16.0 4.5  
## 35 range rover 11.5 16.5 5.0

Most of the additions are trucks or heavier models – suggesting that the stop-and-go of city driving is draining while the vehicles can use their weight to “cruse” and conserve fuel.

Running two regressions[[1]](#footnote-1) (city MPG ~ model, highway MPG ~ model) confirms the findings with significant positive relationship for Civic (9.2, 13.7) and Corolla (10.4, 15.1). While several other models have significantly positive high efficiency in individual MPG categories (a4, Altima), none possess the overall efficiency of those two models.

The Expedition (-3.8, -1.5), Land Cruiser (-3.1, -2.3), Navigator (-3.8, -1.8), Ram 1500 Pickup (-3.8, -3.5) and Range Rover (-3.7, -2.3) are the statistical significantly worse than average.

1. As the MPG variables are highly correlated, they cannot be included in a linear regression model. [↑](#footnote-ref-1)