Features Importance for the price of car

Final Project

Jiahao Li

MSDS 597 DATA WRANGLING

4/20/2021

Library(Kaggler)
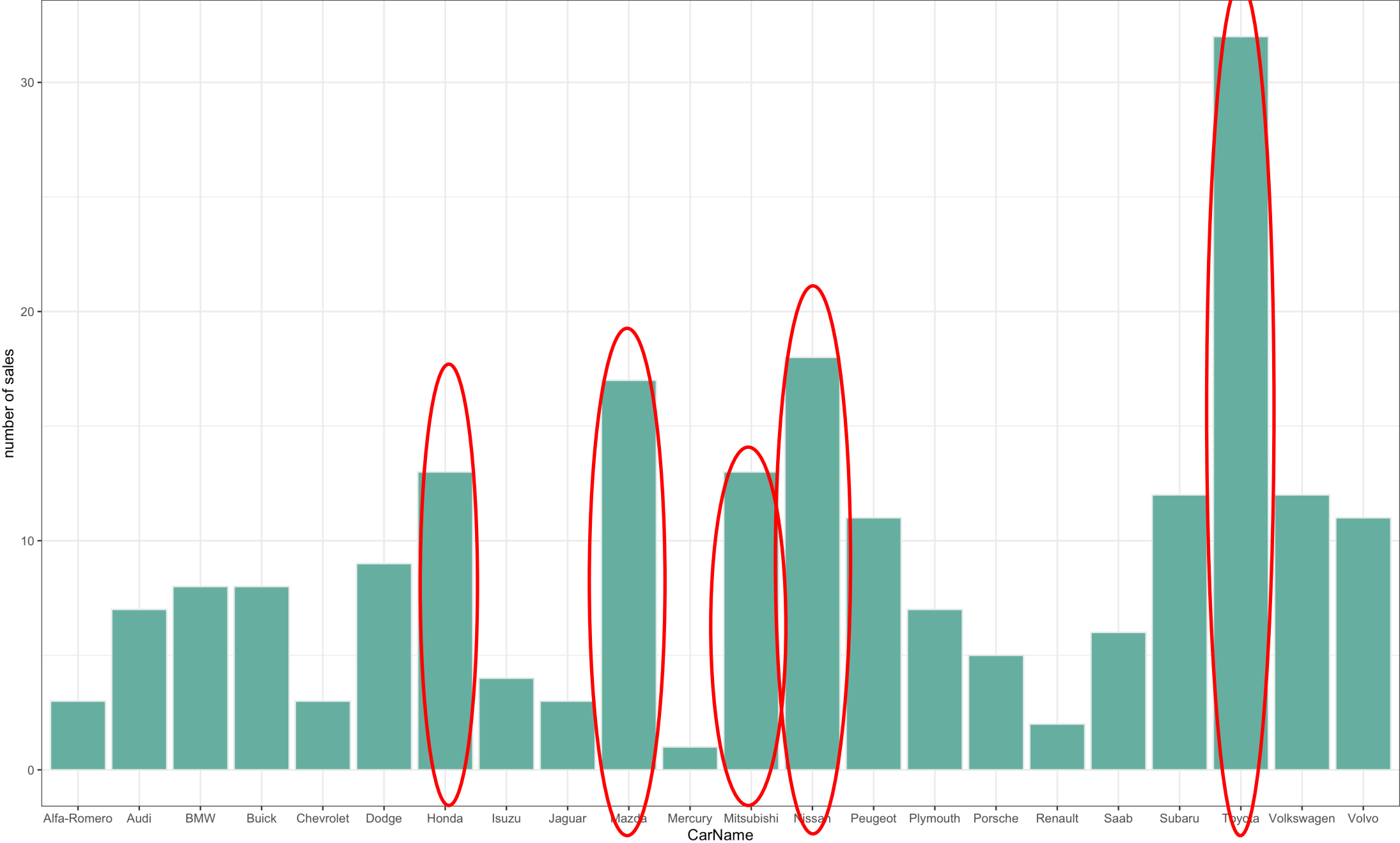
Original Dataset

car_ID	symboling	CarName	fueltype	aspiration	doornumber	carbody	drivewheel	enginelocation	wheelbase
1	3	alfa-romero giulia	gas	std	two	convertible	rwd	front	88.6
2	3	alfa-romero stelvio	gas	std	two	convertible	rwd	front	88.6
3	1	alfa-romero Quadrifoglio	gas	std	two	hatchback	rwd	front	94.5
4	2	audi 100 ls	gas	std	four	sedan	fwd	front	99.8
5	2	audi 100ls	gas	std	four	sedan	4wd	front	99.4

carlength	carwidth	carheight	curbweight	enginetype	cylindernumber	enginesize	fuelsystem	boreratio	stroke	compressionratio
168.8	64.1	48.8	2548	dohc	four	130	mpfi	3.47	2.68	9.00
168.8	64.1	48.8	2548	dohc	four	130	mpfi	3.47	2.68	9.00
171.2	65.5	52.4	2823	ohcv	six	152	mpfi	2.68	3.47	9.00
176.6	66.2	54.3	2337	ohc	four	109	mpfi	3.19	3.40	10.00
176.6	66.4	54.3	2824	ohc	five	136	mpfi	3.19	3.40	8.00

horsepower	peakrpm	citympg	highwaympg	price
111	5000	21	27	13495.00
111	5000	21	27	16500.00
154	5000	19	26	16500.00
102	5500	24	30	13950.00
115	5500	18	22	17450.00

There are 205 observations and 26 variables without missing value

Histogram of the sales number of each CarName

Histogram of the mean price of each CarName
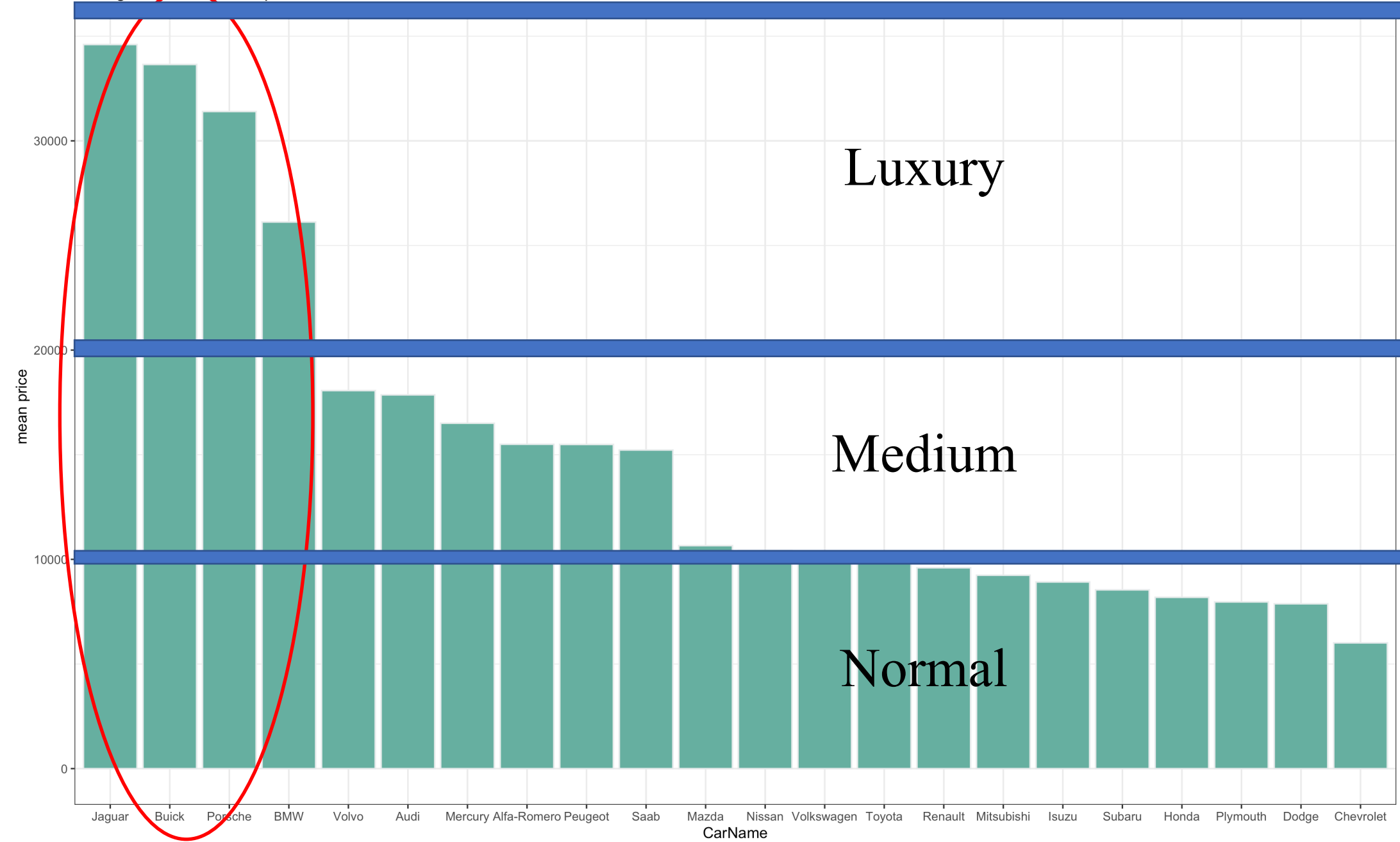
Cleaned Dataset

	symboling	CarName	fueltype	aspiration	doornumber	carbody	drivewheel	enginelocation	wheelbase	carlength	carwidth	carheight	curbweight
1	3	Alfa-Romero	gas	std	two	convertible	rwd	front	88.6	168.8	64.1	48.8	2548
2	3	Alfa-Romero	gas	std	two	convertible	rwd	front	88.6	168.8	64.1	48.8	2548
3	1	Alfa-Romero	gas	std	two	hatchback	rwd	front	94.5	171.2	65.5	52.4	2823
4	2	Audi	gas	std	four	sedan	fwd	front	99.8	176.6	66.2	54.3	2337
5	2	Audi	gas	std	four	sedan	4wd	front	99.4	176.6	66.4	54.3	2824

| curbweight	enginetype	cylindernumber	enginesize	fuelsystem	boreratio	stroke	compressionratio	horsepower	peakrpm	citympg	highwaympg	price	categorical_price
2548	dohc	four	130	mpfi	3.47	2.68	9.00	111	5000	21	27	13495.00	medium
2548	dohc	four	130	mpfi	3.47	2.68	9.00	111	5000	21	27	16500.00	medium
2823	ohcv	six	152	mpfi	2.68	3.47	9.00	154	5000	19	26	16500.00	medium
2337	ohc	four	109	mpfi	3.19	3.40	10.00	102	5500	24	30	13950.00	medium
2824	ohc	five	136	mpfi	3.19	3.40	8.00	115	5500	18	22	17450.00	medium
2507	ohc	five	136	mpfi	3.19	3.40	8.50	110	5500	19	25	15250.00	medium

26 variables

Scattle of the price of mile per gallon
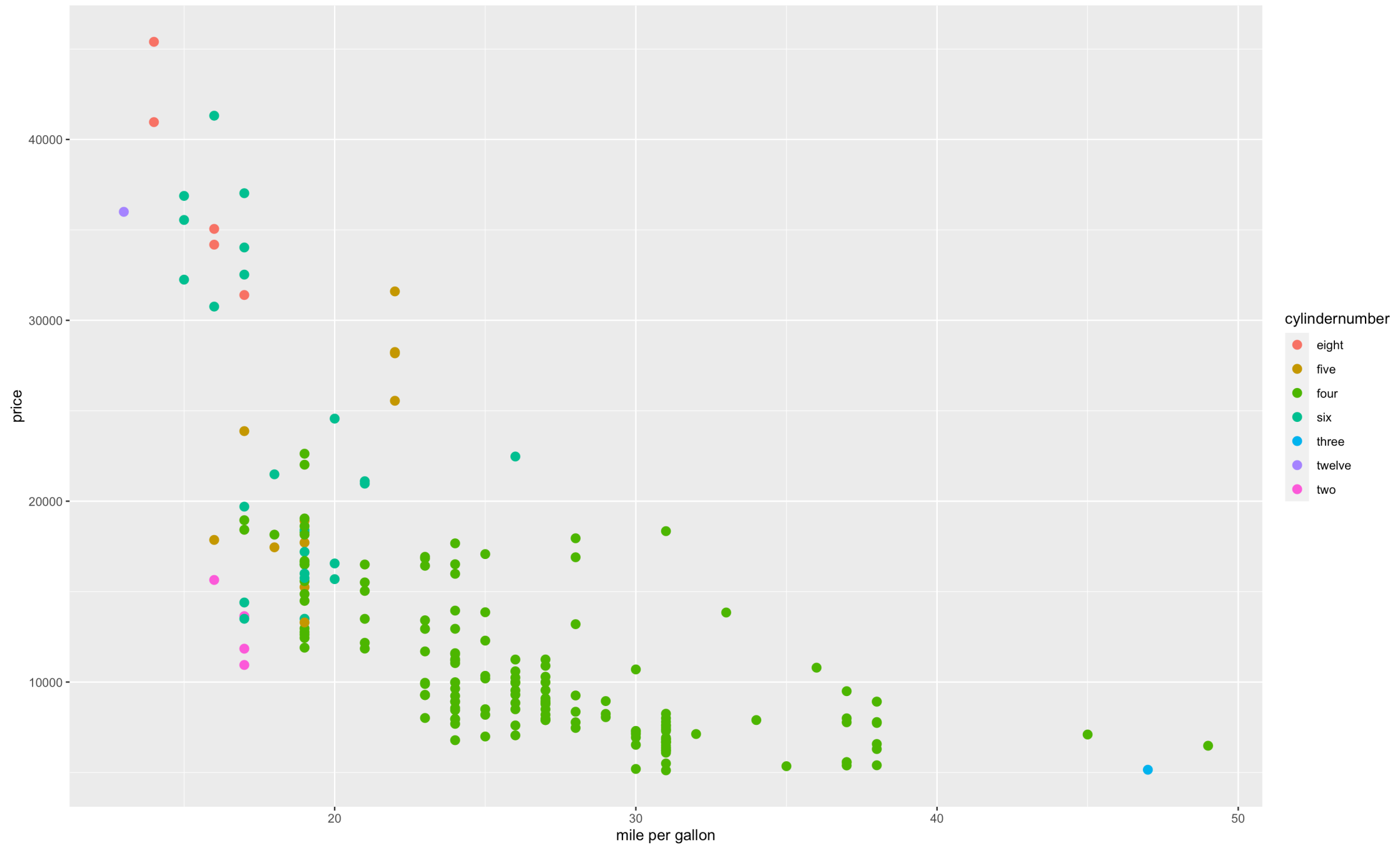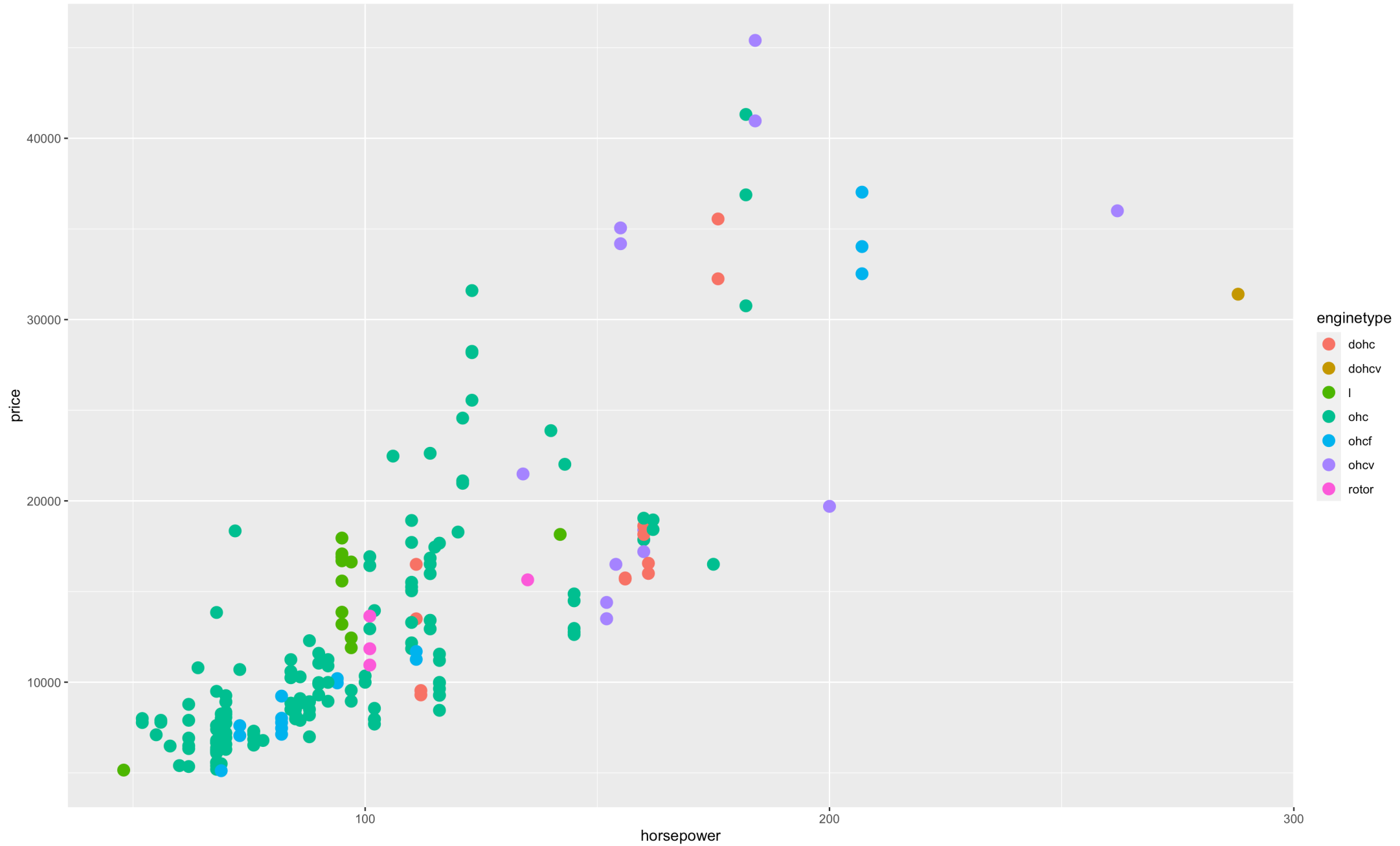
Scattle of the price of engine size
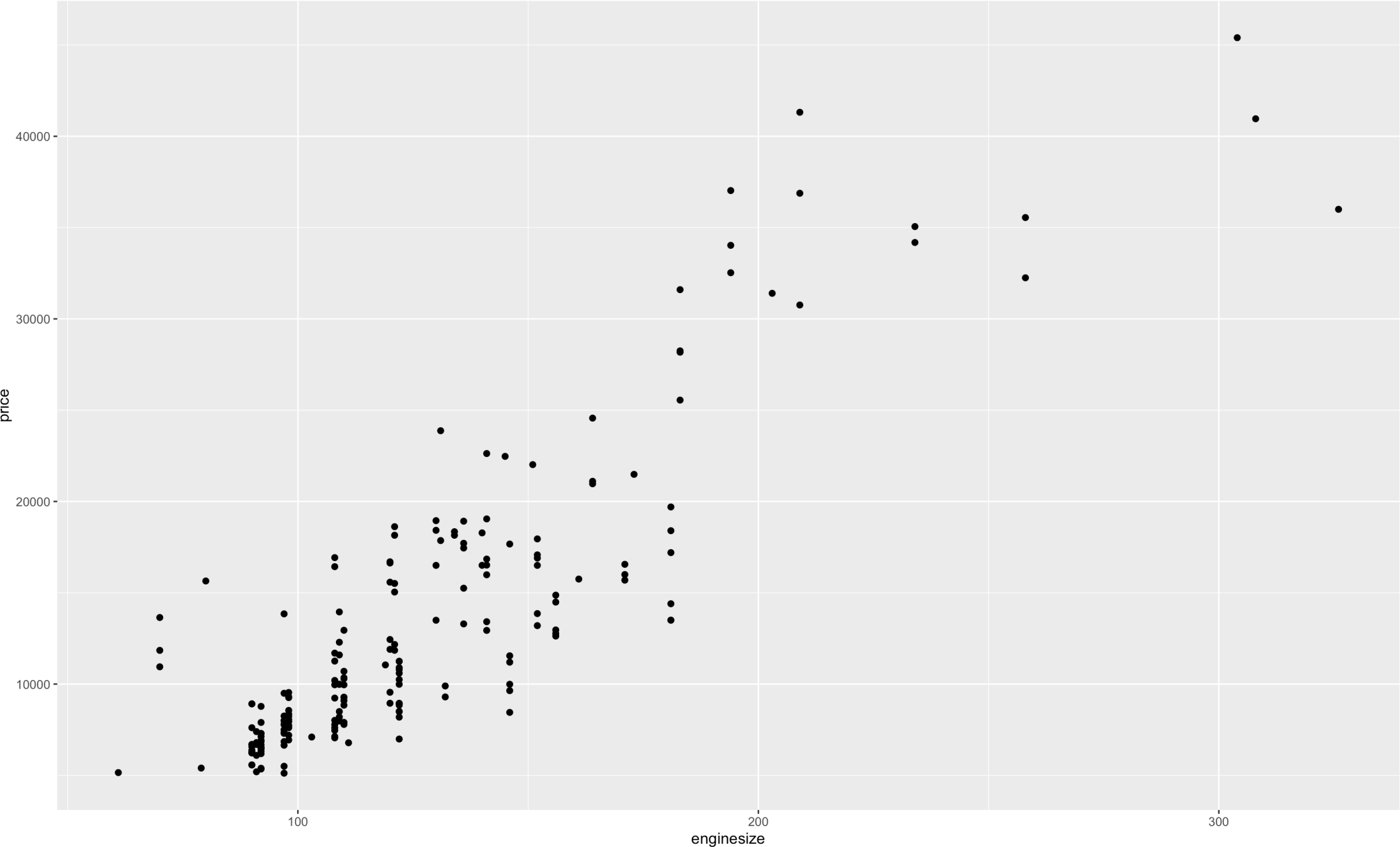
Heat Map for the correlation of numeric variables
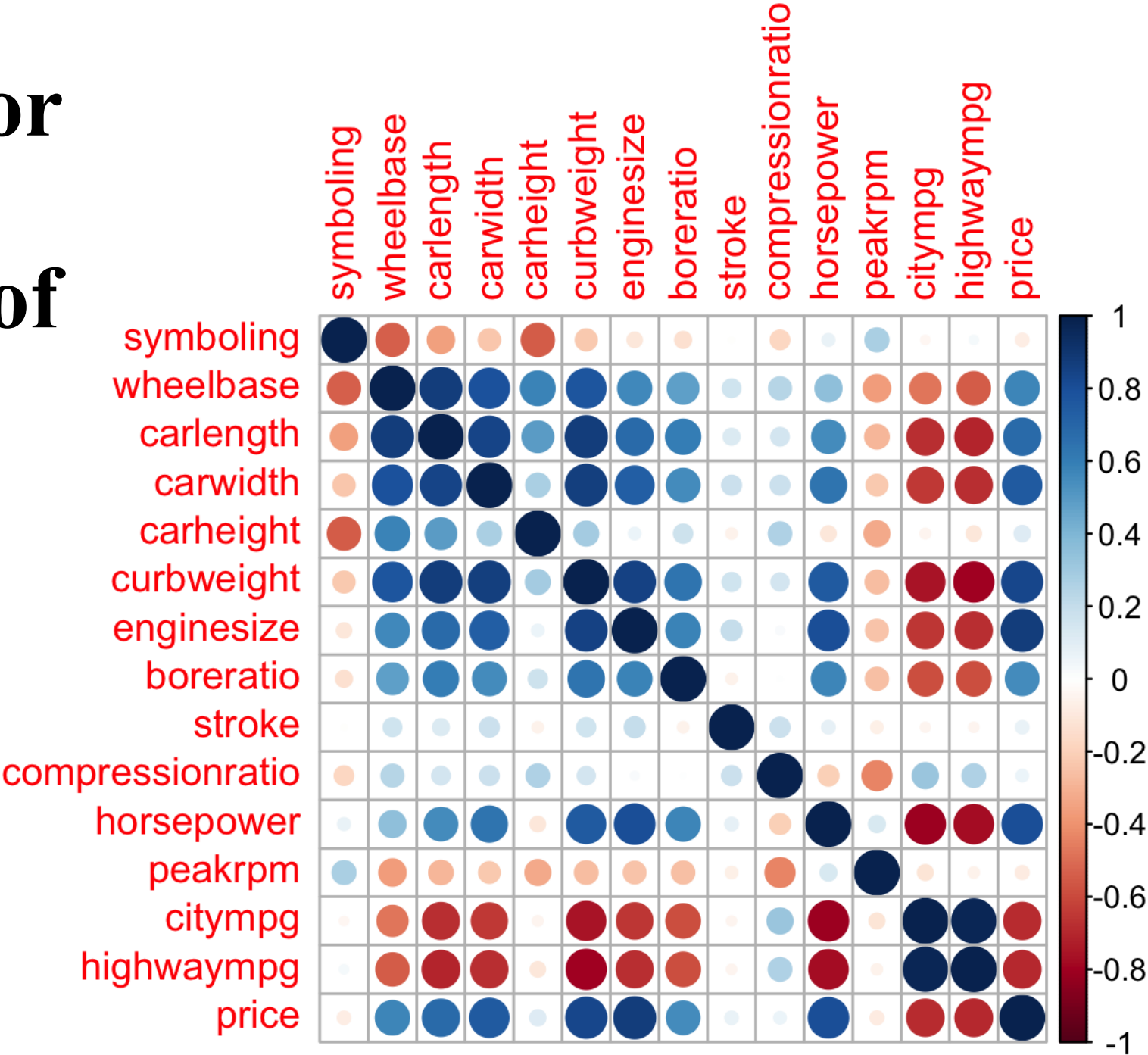
Price ~ .

```
Linear Regression

205 samples
 49 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 185, 184, 185, 185, 183, 185, ...
Resampling results:

  RMSE        Rsquared    MAE
  0.3039155   0.9210799   0.2195451
```
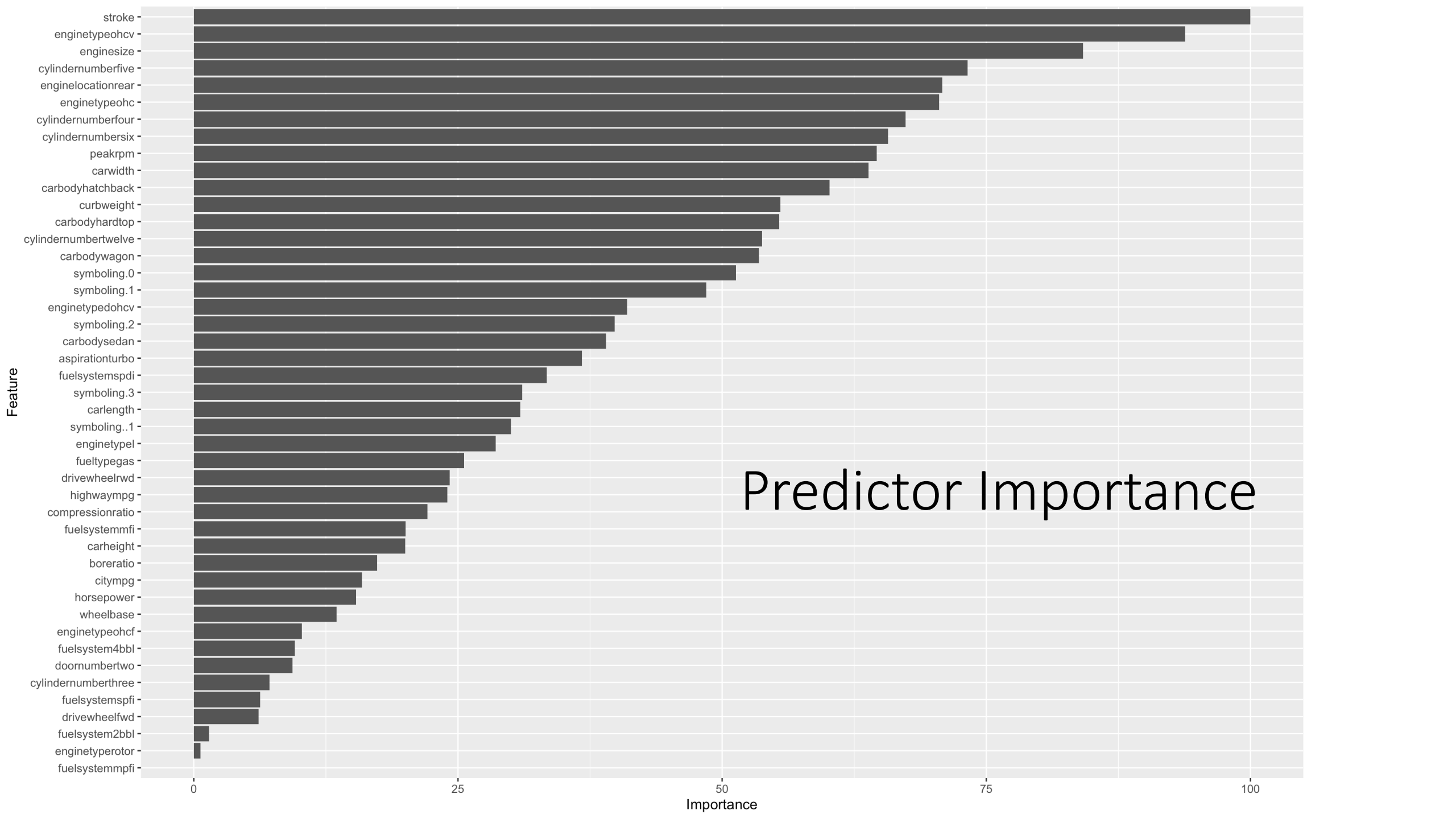
Predictor Importance

Thank you