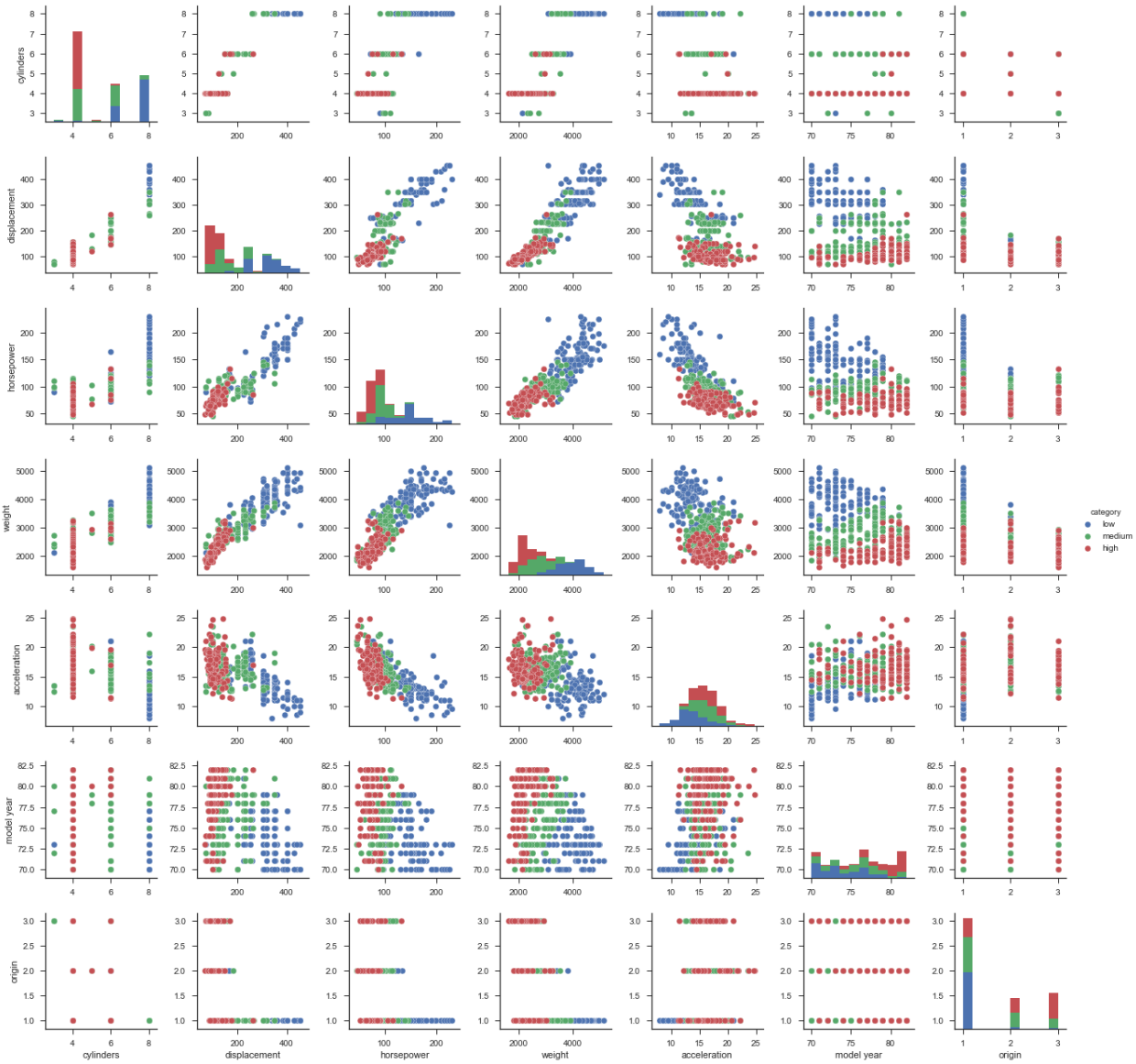
ECS 171 – Homework 1

Instructions: The code for all questions is contained in homework1.py script. I have indicated in the comments the part of the script that is for each question.

1. Output:

```
Low-Medium Threshold: 18.503
Medium-High Threshold: 26.959400000000006
The size of each bin for (low, medium, high):
130 131 131
```

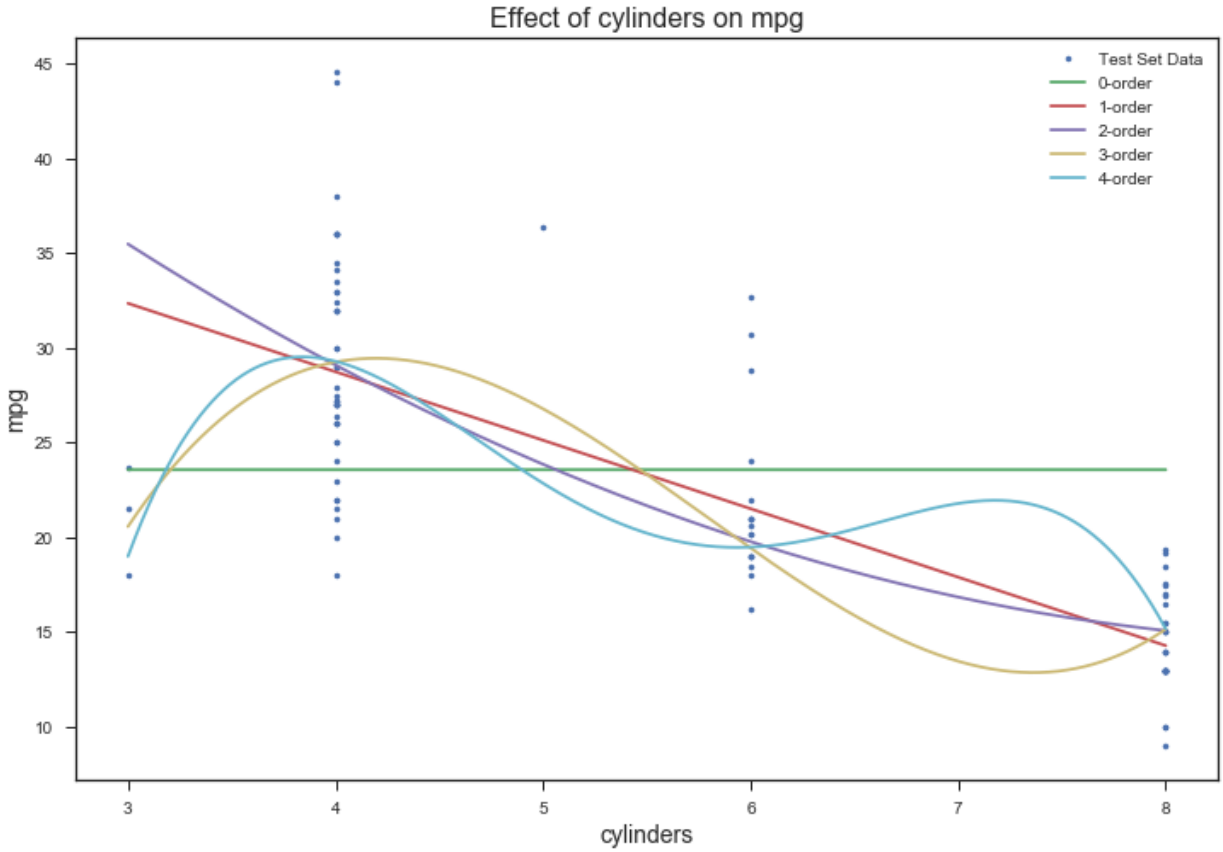
2. Output:




```
Best feature combinations: horsepower-weight, horsepower-acceleration, and
 weight-acceleration
```

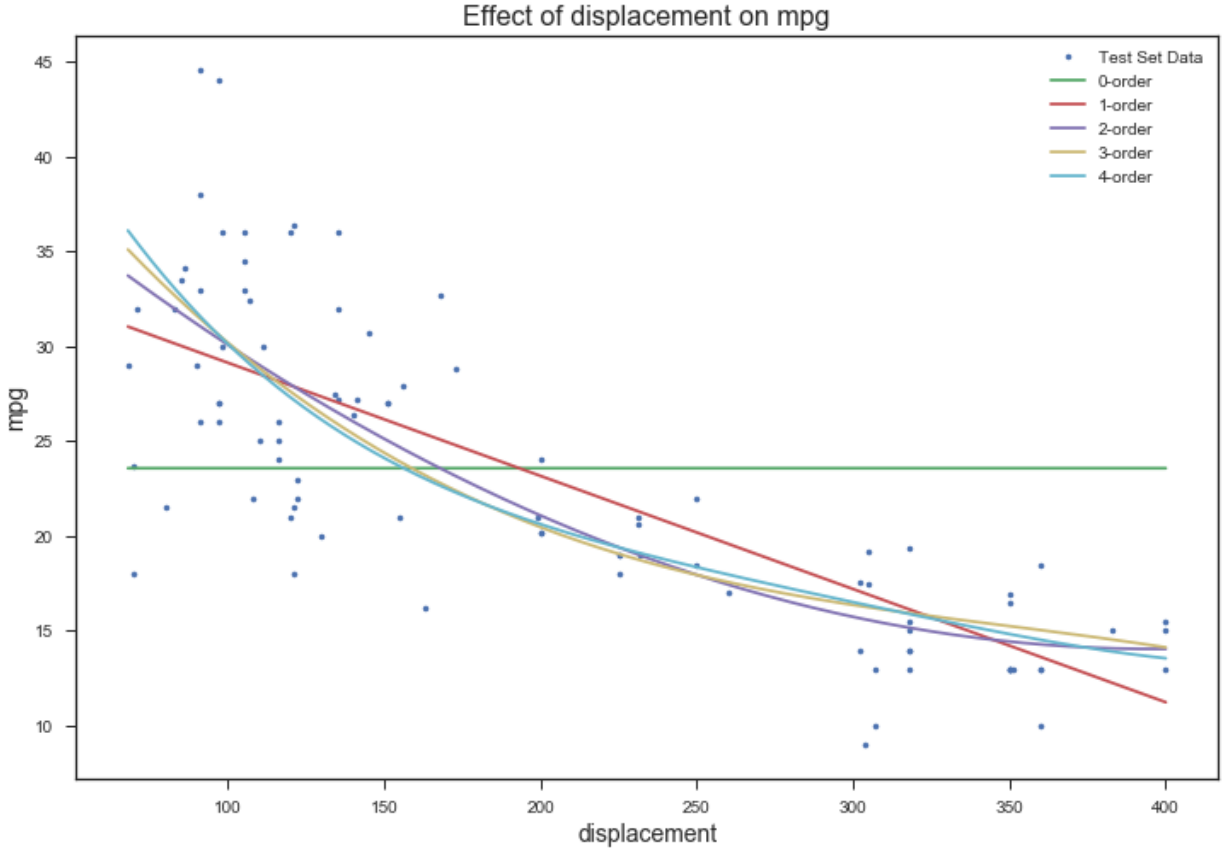
3. No output. Look at script for function.
4. Output:

```
Feature: cylinders
0-order polynomial:
Training set MSE: 59.3911132222
Testing set MSE: 65.3006788068
1-order polynomial:
Training set MSE: 22.9362058909
Testing set MSE: 27.5962715864
2-order polynomial:
Training set MSE: 22.008379162
Testing set MSE: 31.3358556707
3-order polynomial:
Training set MSE: 21.1677234371
Testing set MSE: 24.1798546121
4-order polynomial:
Training set MSE: 21.0550193474
Testing set MSE: 25.2457158742
```



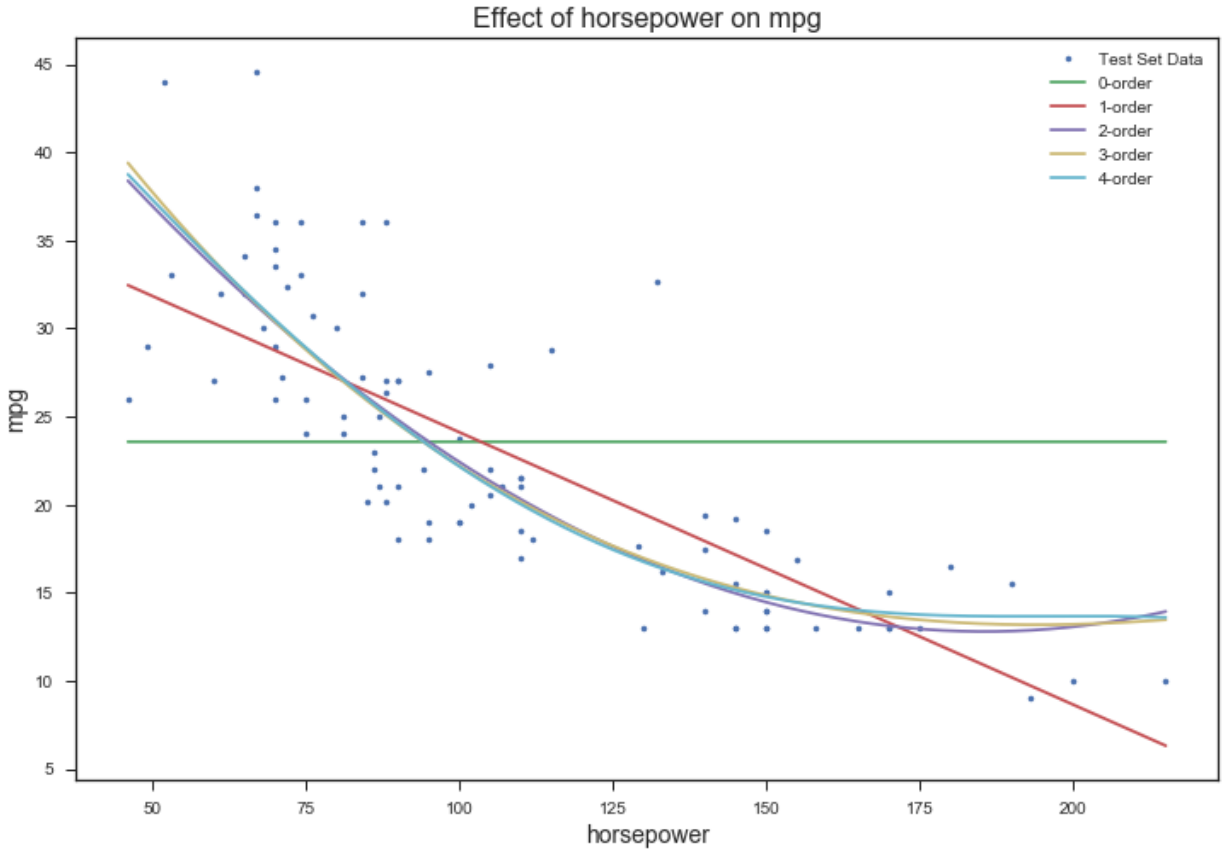

```
Feature: displacement
0-order polynomial:
Training set MSE: 59.3911132222
Testing set MSE: 65.3006788068
1-order polynomial:
```

```
Training set MSE: 20.5064436137
Testing set MSE: 24.2240126697
2-order polynomial:
Training set MSE: 17.3279726394
Testing set MSE: 24.1479733719
3-order polynomial:
Training set MSE: 16.9653827297
Testing set MSE: 25.7875877965
4-order polynomial:
Training set MSE: 16.8499368684
Testing set MSE: 26.6089135427
```

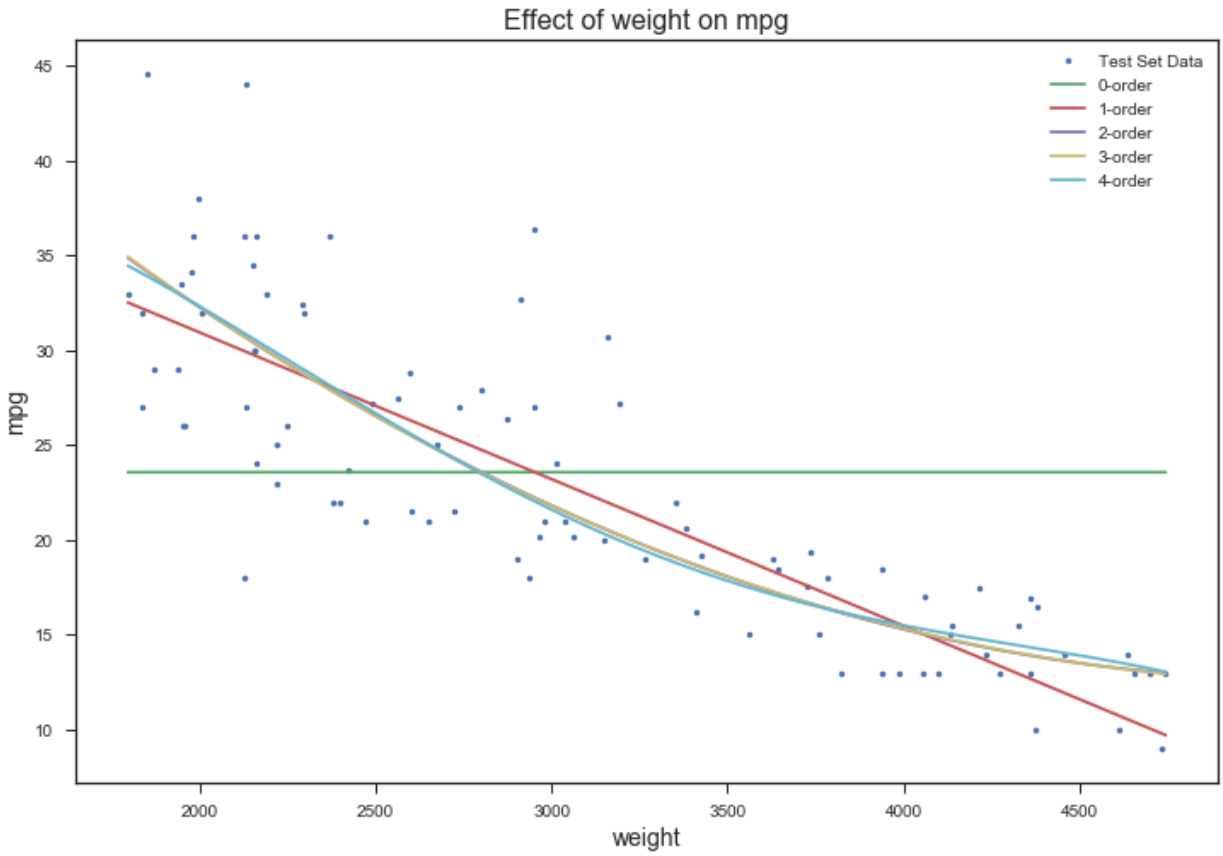



```
Feature: horsepower
0-order polynomial:
Training set MSE: 59.3911132222
Testing set MSE: 65.3006788068
1-order polynomial:
Training set MSE: 24.4653662978
Testing set MSE: 22.3103952801
2-order polynomial:
Training set MSE: 18.5256671609
Testing set MSE: 20.6017387696
3-order polynomial:
Training set MSE: 18.4325222408
Testing set MSE: 20.7956106036
4-order polynomial:
Training set MSE: 18.4064825098
```

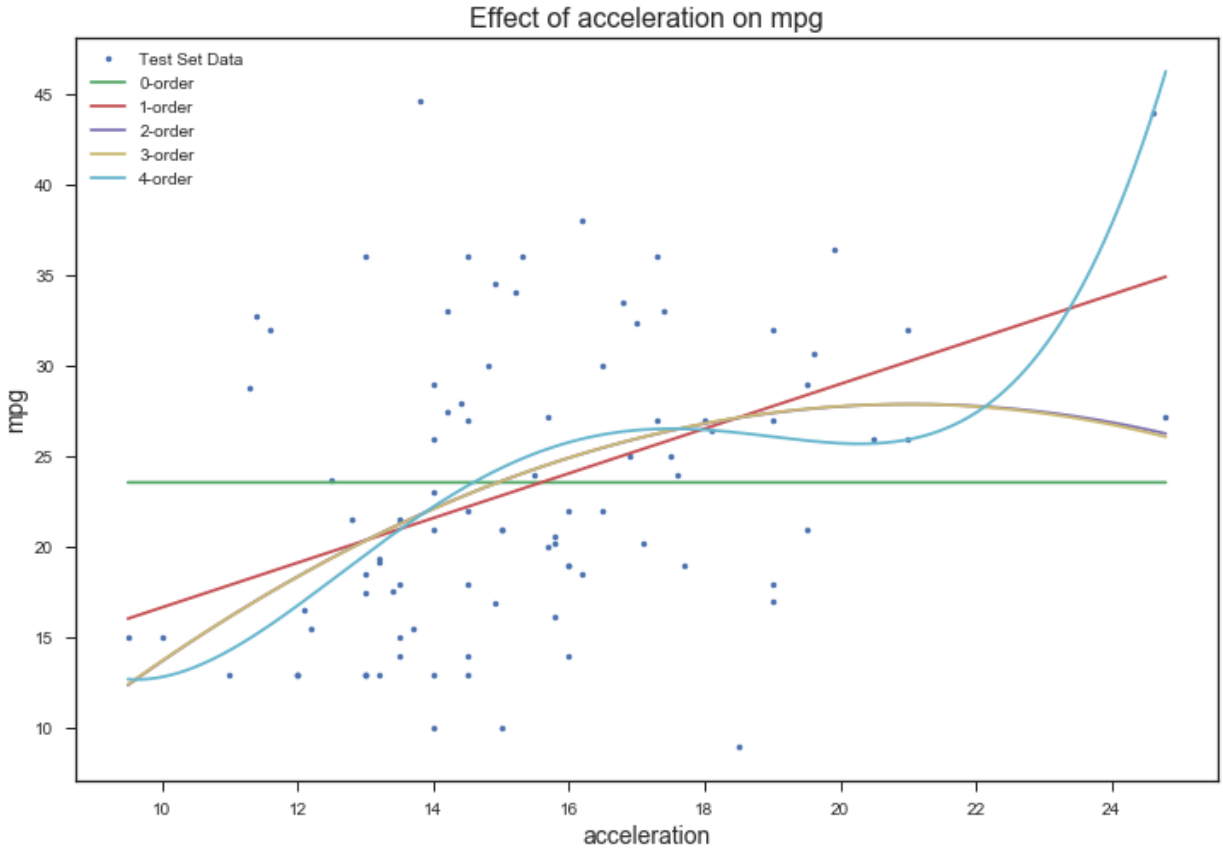
```
Testing set MSE: 20.6330883772
```



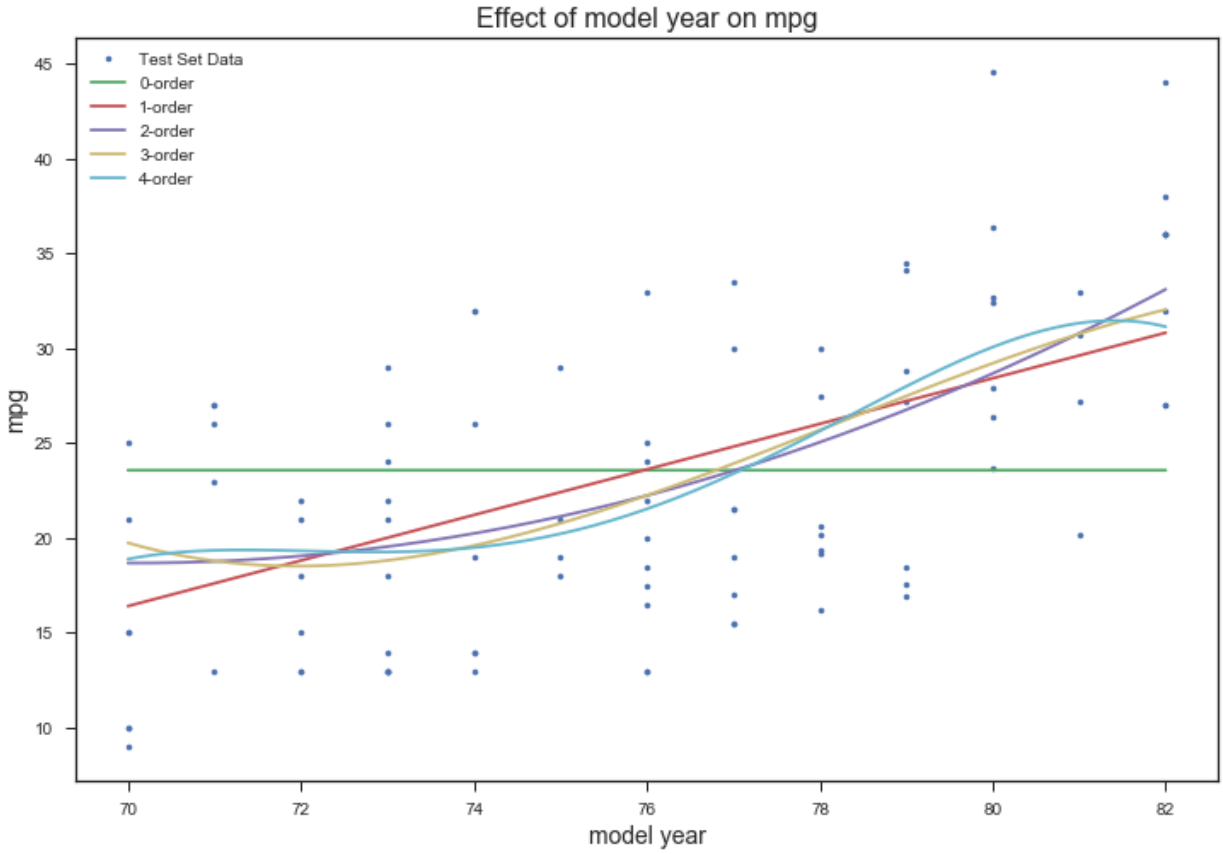

```
Feature: weight
0-order polynomial:
Training set MSE: 59.3911132222
Testing set MSE: 65.3006788068
1-order polynomial:
Training set MSE: 18.2420124621
Testing set MSE: 20.1360001033
2-order polynomial:
Training set MSE: 16.3802326653
Testing set MSE: 20.4921998723
3-order polynomial:
Training set MSE: 16.3794991821
Testing set MSE: 20.5009409692
4-order polynomial:
Training set MSE: 16.3063357169
Testing set MSE: 20.7368563202
```



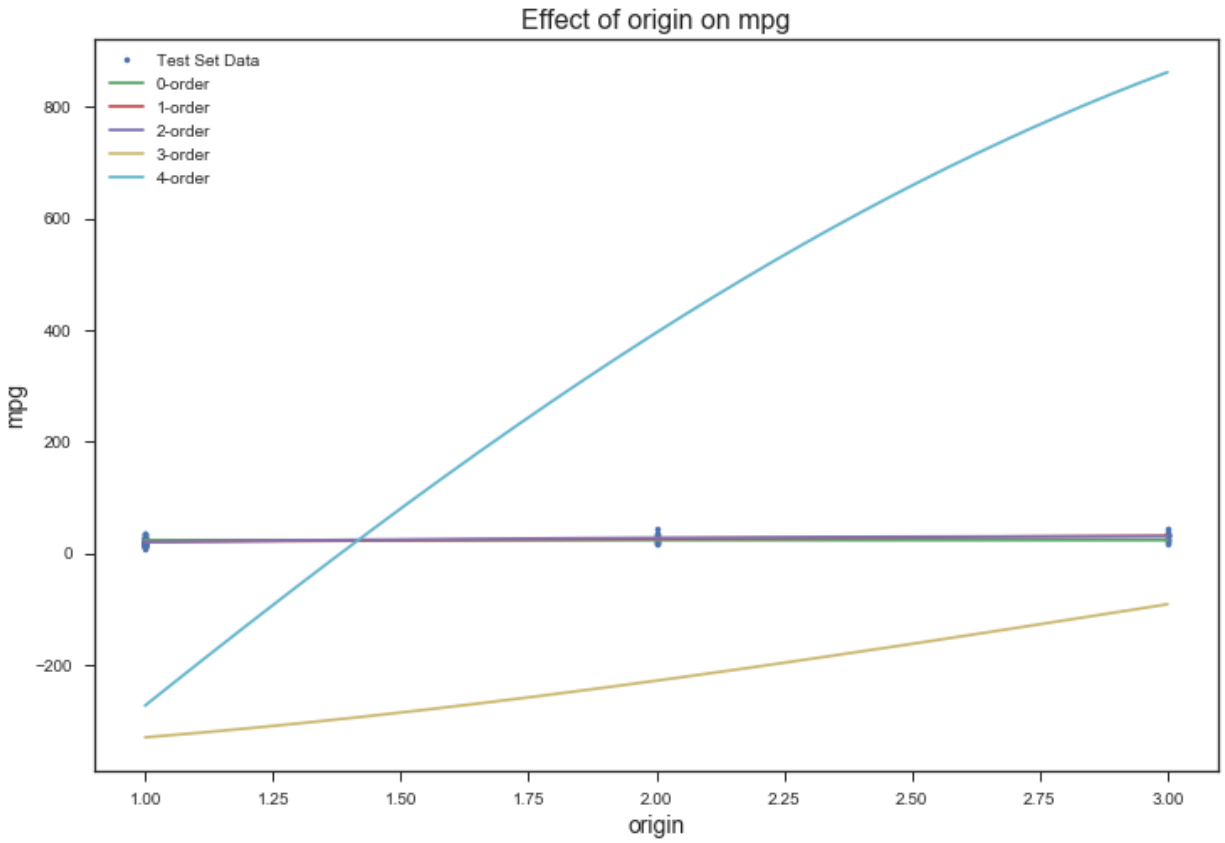

```
Feature: acceleration
0-order polynomial:
Training set MSE: 59.3911132222
Testing set MSE: 65.3006788068
1-order polynomial:
Training set MSE: 48.121251842
Testing set MSE: 55.6418723963
2-order polynomial:
Training set MSE: 46.4549442486
Testing set MSE: 57.8961199869
3-order polynomial:
Training set MSE: 46.4546624976
Testing set MSE: 57.9655038038
4-order polynomial:
Training set MSE: 44.810373809
Testing set MSE: 59.6789357221
```




```
Feature: model year
0-order polynomial:
Training set MSE: 59.3911132222
Testing set MSE: 65.3006788068
1-order polynomial:
Training set MSE: 39.8334280588
Testing set MSE: 41.8888085899
2-order polynomial:
Training set MSE: 38.2826742161
Testing set MSE: 39.0812646935
3-order polynomial:
Training set MSE: 37.9009646241
Testing set MSE: 40.5710641524
4-order polynomial:
Training set MSE: 37.5210601908
Testing set MSE: 40.199241013
```




```
Feature: origin
0-order polynomial:
Training set MSE: 59.3911132222
Testing set MSE: 65.3006788068
1-order polynomial:
Training set MSE: 40.1154680598
Testing set MSE: 45.4338181017
2-order polynomial:
Training set MSE: 39.3091554603
Testing set MSE: 44.8786483855
3-order polynomial:
Training set MSE: 90618.3722949
Testing set MSE: 91733.6680159
4-order polynomial:
Training set MSE: 221163.068342
Testing set MSE: 201315.083603
```




```
Over a few trials, it seems like 2-order, 3-order, and 4-order polynomials
 have the best performances on the test set.
The best features for predicting mpg consumption are cylinders, displaceme
nt, horsepower, and weight.
The results from each trial can vary because of randomization of test set
```

Note: I have no idea why 3rd order and 4th order polynomials for origin vs. mpg are incorrect. Everything else seems to make sense: Error decreases for higher order polynomials and error increases for test set from training set.

5. Output:

```
0-order polynomial:
Training set MSE: 59.7067265556
Testing set MSE: 64.2239350386
1-order polynomial:
Training set MSE: 10.0302979095
Testing set MSE: 13.9590032718
2-order polynomial:
Training set MSE: 6.44373029529
Testing set MSE: 10.9999449251
```

6. Output:

```
Training set accuracy: 0.833333333333
Test set accuracy: 0.79347826087
```

7. Output:

```
Second-order, multi-variate polynomial regression:
Predicted MPG rating - 19.5330541074
Predicted category - medium
Logistic regression:
Predicted category - low
```

Note: While the two regression methods predicted two different categories, this is understandable beca use the threshold is 18.503, which is really close to the predicted MPG rating.

8. Output:

```
Assumptions: cylinders: 0, displacement: 0, horsepower: 500, weight: 2000,
 acceleration: 1, model year: 50, origin: 1
Second-order, multi-variate polynomial regression:
Predicted MPG rating - 98.2198987178
Predicted category – high
Logistic regression:
Predicted category - low
```