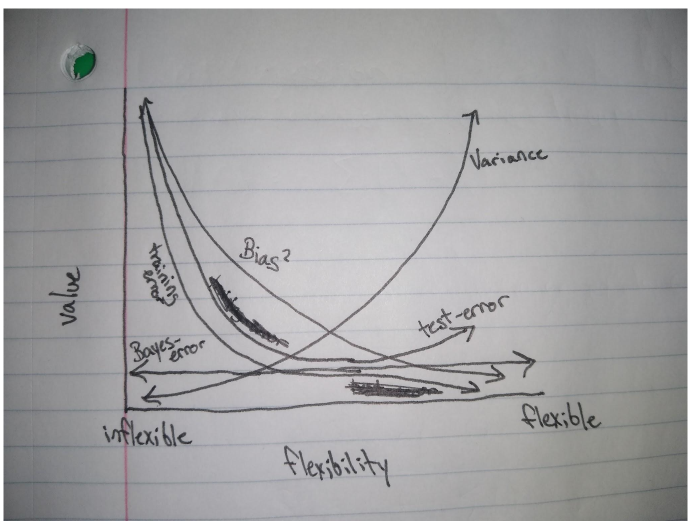
1.
    a. This situation would be better suited to a flexible model because n is large, making overfitting less likely, and there are few predictors, there is less need to narrow them down for time efficiency.
    b. This would be better for an inflexible model because there are many parameters, some of which are unlikely to have a large influence on the result and n is small, making overfitting more likely with a more flexible model.
    c. Flexible - if the relationship is highly non-linear, a more flexible model (higher order polynomial) will be better at approximating it.
    d. Inflexible - a more flexible model is likely to fit to the noise (variance) so an inflexible model is safer.
2.
    a. Regression, inference, n = 500, p = 3 (profit, number of employees, industry)
    b. Classification, prediction, n = 20, p = 13 (price charged for the product, marketing budget, competition price, and ten other variables)
    c. Regression, prediction, n = 52 (weeks in 2012), p = 3 (the % change in the US market, the % change in the British market, and the % change in the German market)
3.
    a. 

    b.
        i. Squared bias: decreases with increase in flexibility.
        ii. Variance: increases with flexibility due to overfitting
        iii. Training error decreases with flexibility because the model can fit more exactly to the data that it is trained on
        iv. Test error decreases with flexibility until the model is optimal, and then increases as overfitting occurs
        v. Bayes error is constant due to the natural noise in the data
4.
    a. Classification:
        i. Image classification: response = what the image is of; predictors = image/array of pixels; This is prediction, because you are training a

machine to recognize images and decide what they are, but are not trying to learn more about this process itself.

      ii. Spam email detection: response = spam or not; predictors = email, sender; this is prediction because it is on a case by case basis for each email and again is not to describe the process/function.
      iii. Deciding if a customer will buy a product or not: response = purchase or not; predictors = customer data(history, location, hobbies, etc.), product data(price, usage, size, shipping, specs, etc…); This is also prediction because you are trying to predict what a person will do.
   b. Regression:
      i. Price of house: response = price of the house; predictors = size, location, year built, num bedrooms, bathrooms, etc.; This could be prediction if you are trying to put a price on a given house, or it could be inference if your goal is to decide what to change about your house to increase its value the most.
      ii. Temperature: response = ex. the temperature in 1 hour; predictors = current temp, weather, month, etc.; This is prediction because you are trying to predict the future based on current measurements.
      iii. Exam grades: response = a students grade on an exam; predictors = study time, previous grades, previous knowledge; This is prediction because you are trying to predict the future based on current measurements.
   c. Cluster:
      i. Recommending content
      ii. Fraud detection
      iii. Credit scores
5. Flexible model
   a. Advantages: fits more complex data, lower bias
   b. Disadvantages: prone to overfitting, higher variance
6. Parametric: can easily change the flexibility by adding and removing parameters, but you need to find what parameters matter
7.

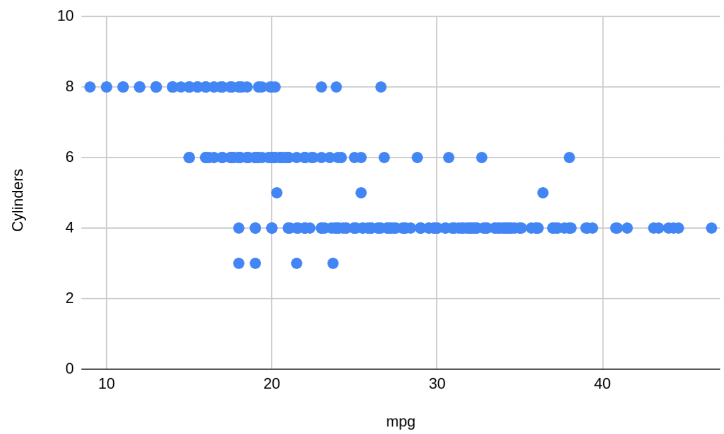
obs	x1	x2	x3	y	distance
1	0	3	0	Red	3
2	2	0	0	Red	2
3	0	1	3	Red	3.162
4	0	1	2	Green	2.236
5	-1	0	1	Green	1.414
6	1	1	1	Red	1.732

   a. See distance column above

b. Green because its closest neighbor is obs 5 which is green.
c. Red because the closest are obs 5, 6, and 2. Two of which are red
d. Small would be better because it produces a more flexible model.

Data set analysis

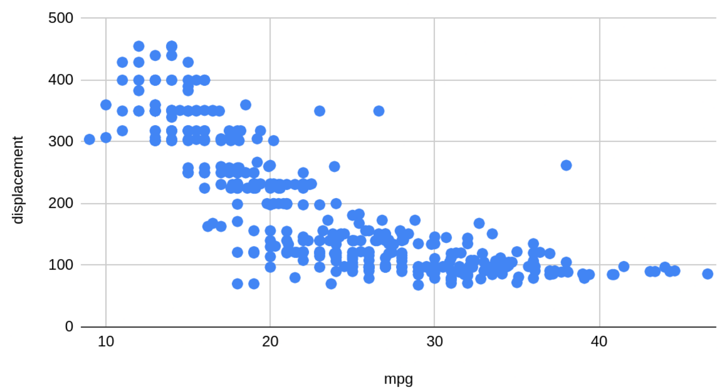
1. Auto data set analysis
    a. Predictors
        i. Quantitative: mpg, cylinders, displacement, horsepower, weight, acceleration, year
        ii. Qualitative: origin, name
    b. Ranges:
        i. Mpg: 37.6
        ii. Cylinders: 5
        iii. Displacement: 387
        iv. Horsepower: 184
        v. Weight: 3527
        vi. Acceleration: 16.8
        vii. Year: 12
    c. Mean, stdev (separated by commas in this order)
        i. Mpg: 23.551256281407, 7.84776025290344
        ii. Cylinders: 5.4572864321608, 1.69958794039033
        iii. Displacement: 194.018844221106, 104.698128000081
        iv. Horsepower: 104.671755725191, 38.6507998335987
        v. Weight: 2971.6608040201, 847.295253956189
        vi. Acceleration: 15.5587939698492, 2.74723777973473
        vii. Year: 75.8341708542714, 4.88586489366503
    d. Range, mean, stdev (separated by commas in this order)
        i. Mpg: 35.6, 24.4732919254658, 7.92031779597766
        ii. Cylinders: 5, 5.3695652173913, 1.65103737721684
        iii. Displacement: 387, 187.670807453416, 100.102176858483
        iv. Horsepower: 184, 101.216981132075, 36.1401944370692
        v. Weight: 3348, 2935.24844720497, 809.708085786237
        vi. Acceleration: 16.3, 15.7248447204969, 2.67652844052648
        vii. Year: 12, 77, 4.77832755610901
    e.



i.

This shows a general negative correlation between numbers of cylinders and mpg which makes sense as engines with more cylinders tend to have more displacement and, while more powerful, have worse mpg.

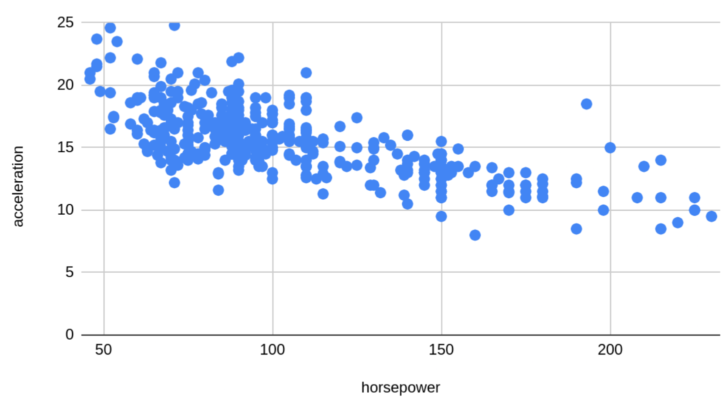
displacement vs. mpg



ii.

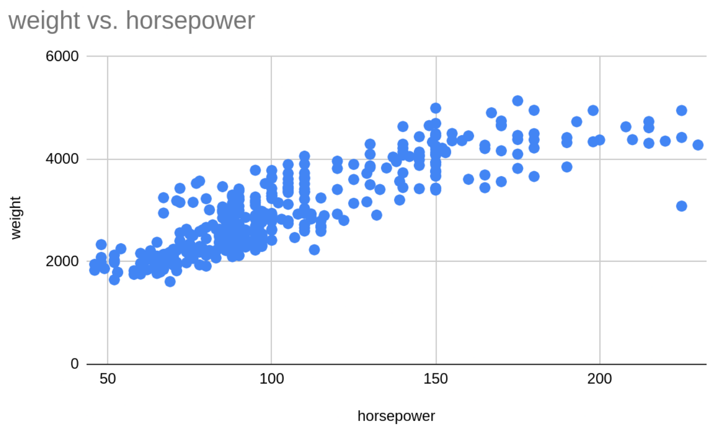
This shows a general negative correlation between displacement and mpg which makes sense for the same reason as above.

acceleration vs. horsepower



iii.

This shows a general negative correlation between acceleration and horsepower. (this one is unintuitive)

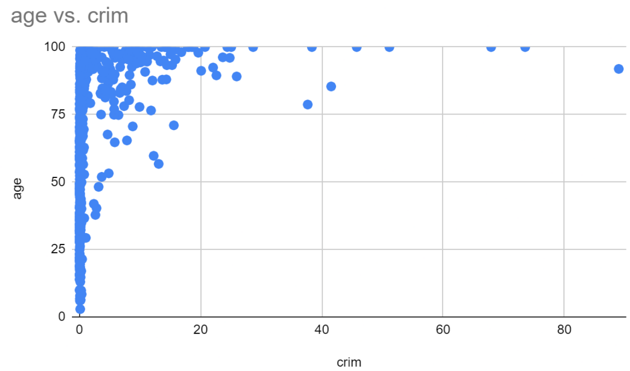


iv.

This shows a positive correlation between horsepower and weight which makes sense due to needing more power to move a heavier vehicle (also is likely the reason for the unintuitive graph above)

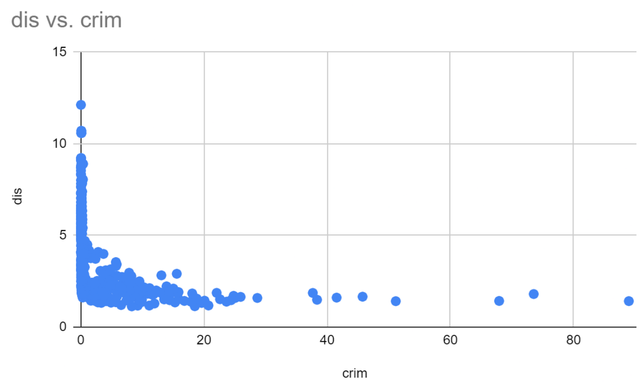
f. Based on the observations above (and prior knowledge), I believe that cylinders and displacement would likely be related (although since these are correlated as well maybe only one would suffice). In addition, weight would certainly be relevant and horsepower would likely be related as well.

2. Boston data set
    a.
        i. 506 rows representing suburbs in boston
        ii. 14 rows:
            1. Index
            2. per capita crime rate by town
            3. proportion of residential land zoned for lots over 25,000 sq.ft.
            4. proportion of non-retail business acres per town.
            5. Charles River dummy variable (= 1 if tract bounds river; 0 otherwise).
            6. nitrogen oxides concentration (parts per 10 million).
            7. average number of rooms per dwelling.
            8. proportion of owner-occupied units built prior to 1940.
            9. weighted mean of distances to five Boston employment centres.
            10. index of accessibility to radial highways.
            11. full-value property-tax rate per $10,000.
            12. pupil-teacher ratio by town.
            13. lower status of the population (percent).
            14. median value of owner-occupied homes in $1000s.
    b. Three notable scatter plots that show some correlation:



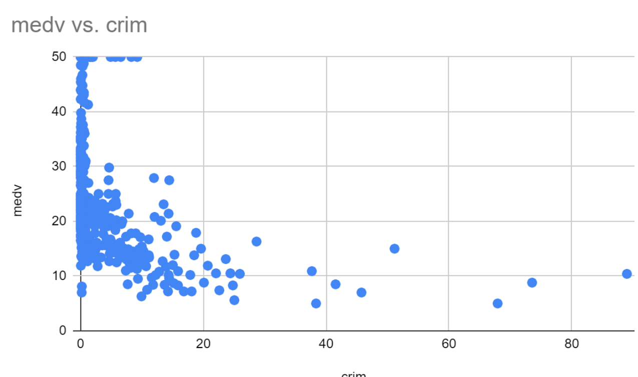
i.

Crime rates seem to increase with age of the suburbs.



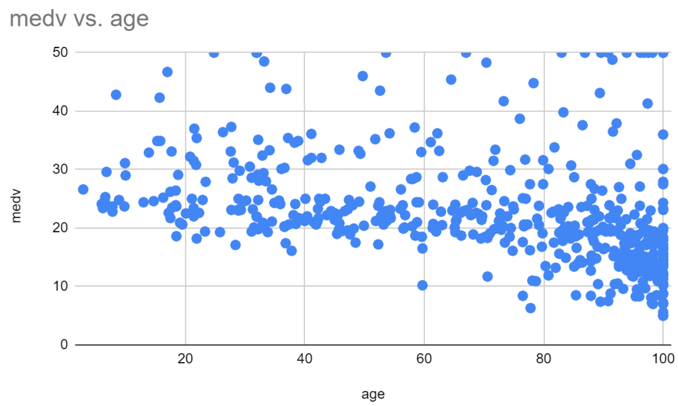
ii.

Crime seems to increase in suburbs with closer employment centers.



iii.

Crime seems to increase as the value of houses decreases.



iv. Based on findings i and iii above, I made the assumption that in areas with older buildings, the buildings might be of less value. This, however, is not backed up by the data. There are many more old, low/med value buildings but there does not seem to be a strong correlation here.

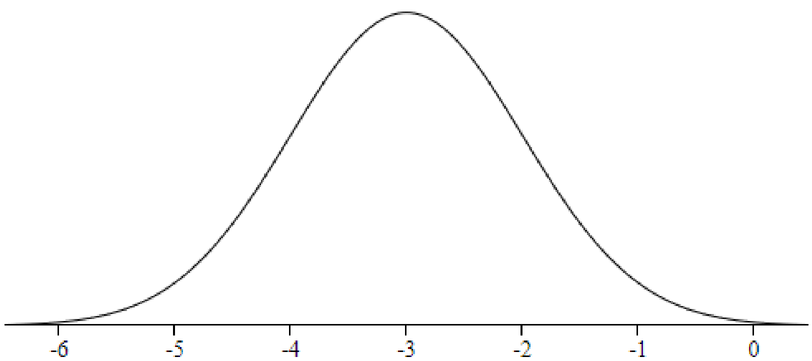
c. Older suburbs, suburbs with closer employment centers, and suburbs with less valuable buildings tend to have higher crime rates. I would hypothesize that this is because all of these likely correlate to lower income/higher poverty rates, which tend to correlate with crime rates.
d. Ranges:
   i. Crim: 88.96988 - this is notable as most suburbs have low rates, but a few are around 3 orders of magnitude larger
   ii. Zn: 100 - this is a proportion, so this covers the entire possible range (0-100%)
   iii. Indus: 27.28 - not that notable
   iv. Chas: 1 - not that notable (boolean)
   v. Nox: 0.486 - not that notable
   vi. Rm: 5.219 - not that notable
   vii. Age: 97.1 - again, proportion covers almost the entire range, very diverse
   viii. Dis: 10.9969 - not that notable
   ix. Rad: 23 - most suburbs have below 5, but a few have >20
   x. Tax: 524 - not that notable
   xi. Ptratio: 9.4 - not that notable
   xii. Lstat: 36.24 - not that notable
   xiii. Medv: 45 - not that notable
e. 35
f. 19.05
g. Suburbs 399 and 406 are tied at a value of 5. They are both very old, near more highways than average, have fairly high crime rates, a high pupil/teacher ratio, fairly high

lstat, and low distance to employment centers. As mentioned above, all of these are often factors of lower income/high poverty rate areas.

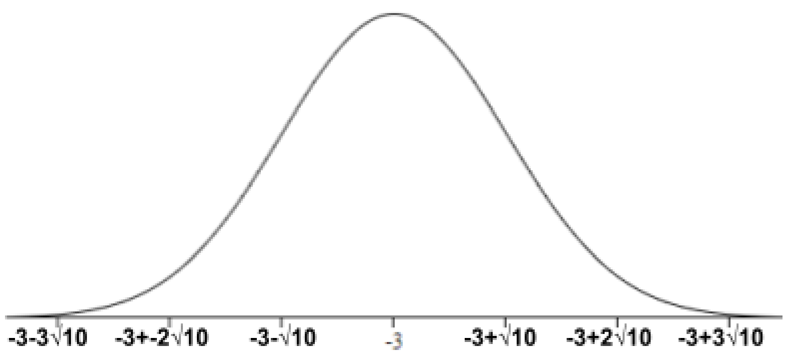
h. Rooms
   i. >7: 64
   ii. >8: 13
      1. All but one has a significantly higher than average median value, all but two have significantly low industry value(more retail and housing focused), all but one have low zn values (smaller buildings likely due to the high value of land), and finally they all have fairly low crime rates.

Mathematics and Probability Questions

1.
   a. 
   b.
      i. $\frac{\delta f}{\delta x} = 6x + 2y - 4$
      ii. $\frac{\delta f}{\delta x} = 2x - 4y + 6$
   c. The minimum is -inf.
2. MLE

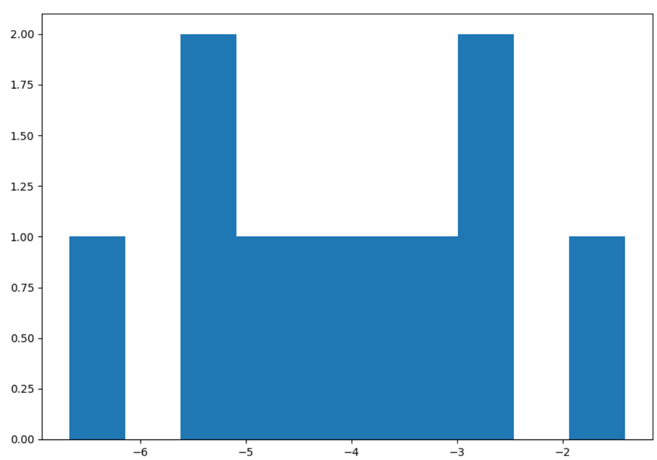


a.



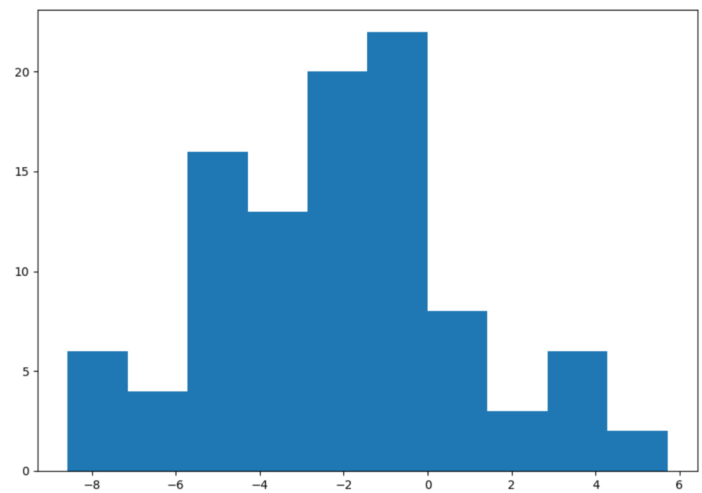
b.

c.

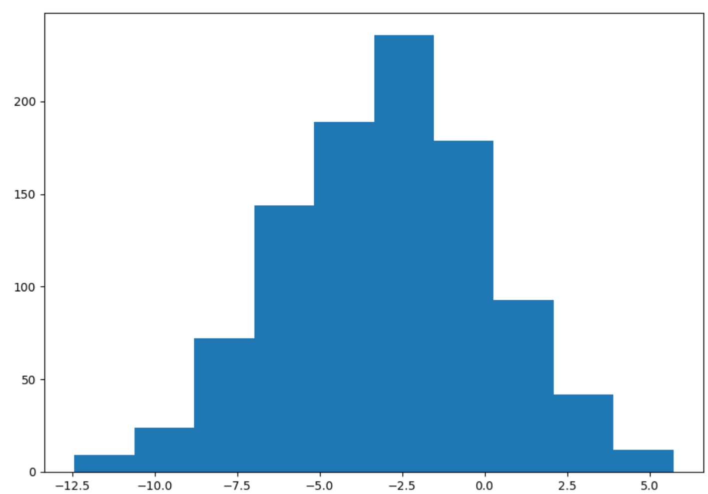


d.

i. n = 10, mean = -4.056661970123534, var = 2.0863223688096695

e. 

i. n = 100, mean = -2.1855296426882145, var = 9.128179102582312

f. 

i. n = 1000, mean = -2.955990926640081, var = 9.838943972503994

3. Bayes Rule and Conditional Distribution

	phd	master	bachelor	Totals
Acccepted	20	30	40	90
Rejected	80	125	60	265
Totals	100	155	100	355

4.

a. 100/355 = 0.2817
b. 50/255 = 0.1961
c. 90/355 = 0.2535
d. 20/90 = 0.2222