#1

The data describes what is likely the average Math and Verbal SAT scores per state. (I say average because I can't be certain from the data where the values we're derived)

#2

The data has an extra row 'ALL' that was a problem throughout my code. I realized I should have removed it earlier. I had a similar problem where I was forced to use .pop which I didn't like because if the code was ran again it would throw off the data moving forward (I made a note on #6). Additionally, Nebraska's state abbreviation was wrong which mean's nothing now until you get to the bonus.

#3

I wasn't sure what this was asking...

#4

```
import csv

allData = []
with open ('/Users/Paul/Desktop/General_Assembly/DSI_SM_01/projects/01-projects-weekly/
project-01/assets/sat_scores.csv') as f:
    reader = csv.reader(f)
    for row in reader:
        allData.append(row[0:4])
```

#5

```
print allData
```

#6

```
###can only run this once....###

lables = allData.pop(0)
print lables
```

#7

```
states = [s[0] for s in allData]
print states
```

#8

```
for index, val in enumerate(allData[2]):
    print lables[index], type(allData[2][index])
```

```
#9

for h in allData:
    h[1] = int(h[1])
    h[2] = int(h[2])
    h[3] = int(h[3])

print allData


#10

rate = [s[1] for s in allData]
verbal = [s[2] for s in allData]
mathL = [s[3] for s in allData]

dicRate = {}
for index, value in enumerate(allData):
    dicRate[states[index]] = allData[index][1]

dicMath = {}
for index, value in enumerate(allData):
    dicMath[states[index]] = allData[index][2]

dicVerbal = {}
for index, value in enumerate(allData):
    dicVerbal[states[index]] = allData[index][3]


print 'Rate', dicRate
print 'Math', dicMath
print 'Verbal', dicVerbal

#11

scores = {states[i]: allData[i] for i, state in enumerate(states)}
print scores

#12

import numpy as np

minRate = np.min(rate)
maxRate = np.max(rate)

minVerbal = np.min(verbal)
maxVerbal = np.max(verbal)

minMath = np.min(mathL)
```

```
maxMath = np.max(mathL)

print "min: Rate" , minRate
print "max: Rate" , maxRate
print "min: Verbal" , minVerbal
print "max: Verbal" , maxVerbal
print "min: Math" , minMath
print "max: Math" , maxMath


#13

import math as mt

rateAverage = sum(rate)/51
rateNewList = []

for i in rate[0:50]:
    result = (rateAverage - i) ** 2
    rateNewList.append(result)

rateStd = mt.sqrt(sum(rateNewList)/51)

verbalAverage = sum(verbal)/51
verbalNewList = []

for x in verbal[0:50]:
    result2 = (verbalAverage - x) ** 2
    verbalNewList.append(result2)

verbalStd = mt.sqrt(sum(verbalNewList)/51)
mathAverage = sum(mathL)/ 51

mathNewList = []

for y in mathL[0:50]:
    result3 = (mathAverage - x) ** 2
    mathNewList.append(result3)

mathStd = mt.sqrt(sum(mathNewList)/51)

print "rate: std" , rateStd
print "verbal: std" , verbalStd
print "math: std " , mathStd



#14
```

```
import matplotlib.pyplot as plt
import numpy as np
%matplotlib inline

plt.hist(rate[0:50])
plt.show()
```

#15

```
plt.hist(mathL)
plt.show()
```

#16

```
plt.hist(verbal)
plt.show()
```

#17

That it is a Normal Distribution

#18

No

#19

```
plt.scatter(verbal, mathL)
plt.show()

plt.scatter(mathL, verbal)
plt.show()

plt.scatter(rate, mathL)
plt.show()

plt.scatter(rate, verbal)
plt.show()
```

#20

```
plt.scatter(rate, mathL)
plt.show()

plt.scatter(rate, verbal)
```

plt.show()

Math and Verbal scores show a pattern of decreasing in relation to rate. In other words, the more students that take the SATS per state, the more likely the scores will represent a normal distribution. However, in some states, the SATS are not taken except for students wishing to attend universities requiring an SAT scores for admission.

#21

```
plt.boxplot(rate)
plt.show()

plt.boxplot(verbal)
plt.show()

plt.boxplot(mathL)
plt.show()
```

#22

I created 2 views in project1FINAL tableu file.

1st - Using information from data.gov on 2014 State education budgets, I hypothesized that the more a state would invest in education the greater the state SAT scores would be.

2nd - Comparing Math and Verbal total SAT scores by State