

# Project 1: SAT Scores + Summary Statistics

Cheng Ji

# Contents

- Introduction
- Exploratory Data Analysis
- Visualization
- Conclusion

# Introduction

- **Data:** mean SAT math and verbal scores, and the participation rate for each state and the District of Columbia for the year 2001.

- **Data Dictionary**
- Since it's boring to do it in excel...

```
variable={'Description': ['States', 'Participant Rate', 'Average  
Verbal Scores', 'Average Math Scores'], 'Data Type':  
scores.dtypes, 'Mean': scores.mean()}  
datadict=pandas.DataFrame(variable)  
datadict.to_csv('data dictionary.csv')
```

	Data Type	Description	Mean
Math	int64	Average Verbal Scores	531.8
Rate	int64	Participant Rate	37
State	object	States	
Verbal	int64	Average Math Scores	532

# Exploratory Data Analysis

- **Load the Data to a list of lists:**

```
import csv
```

```
scorelist=[]
```

```
with open('sat_scores.csv', 'rb') as inputdata:
```

```
    filelist=csv.reader(inputdata)
```

```
    for row in filelist:
```

```
        scorelist.append(row)
```

- **Compute std using list comprehension:**

```
import numpy
```

```
lst=[rate, verbal, math]
```

```
def func():
```

```
    sd=[numpy.std([y for y in x]) for x in lst]
```

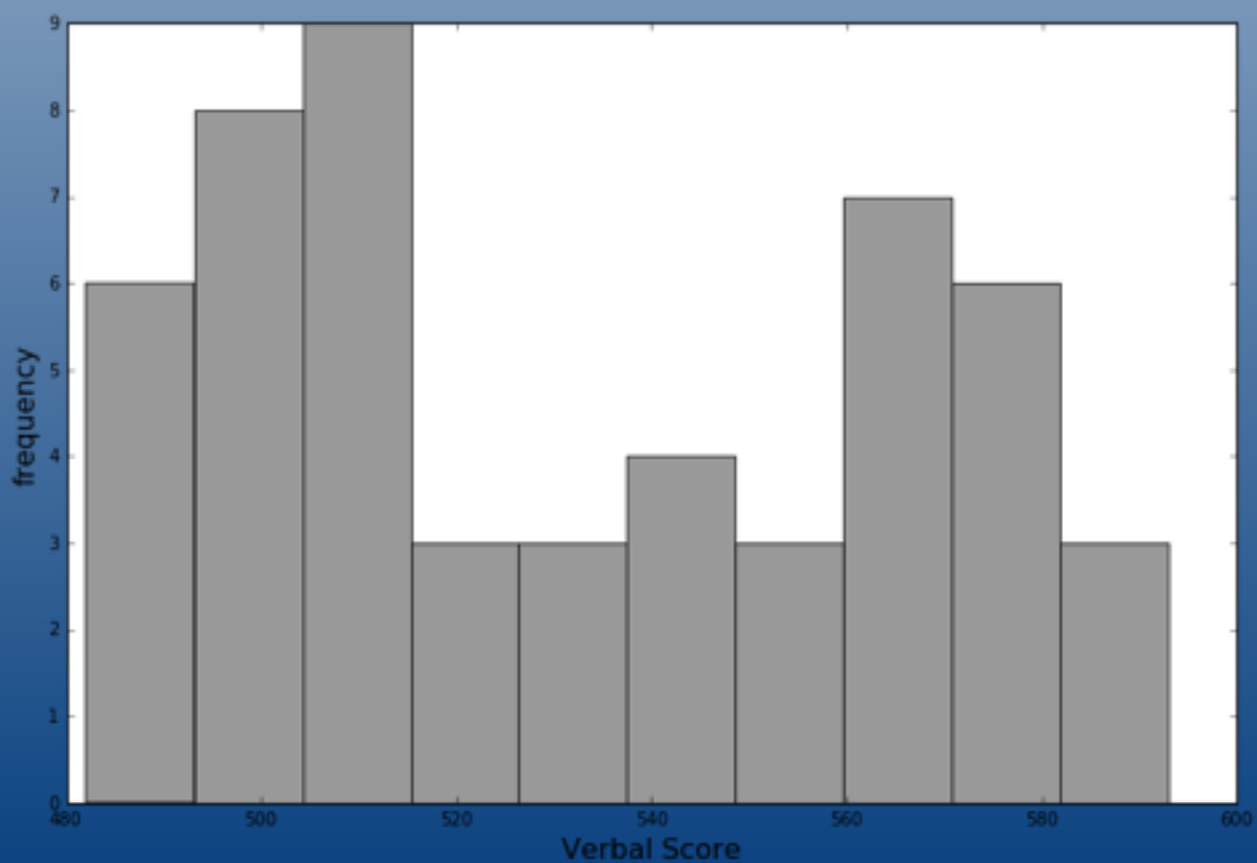
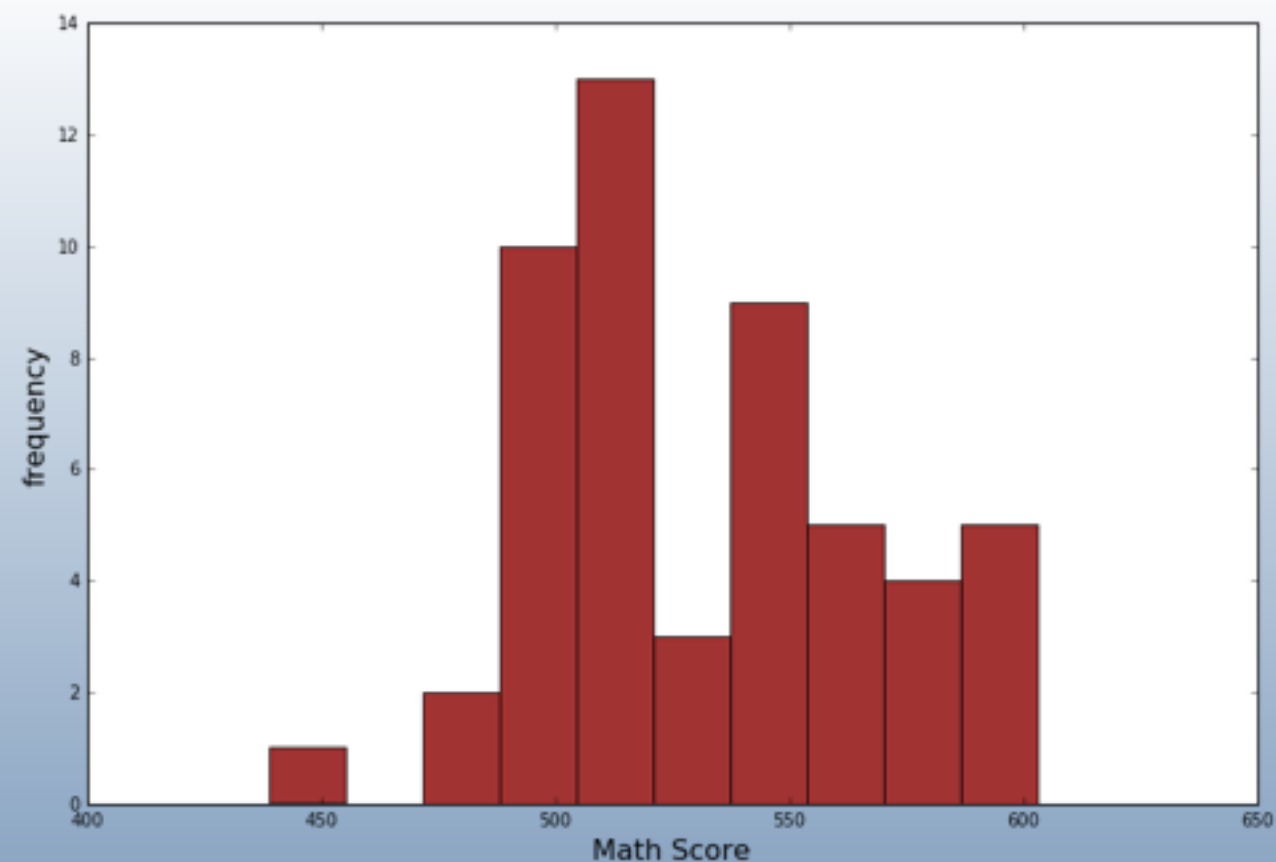
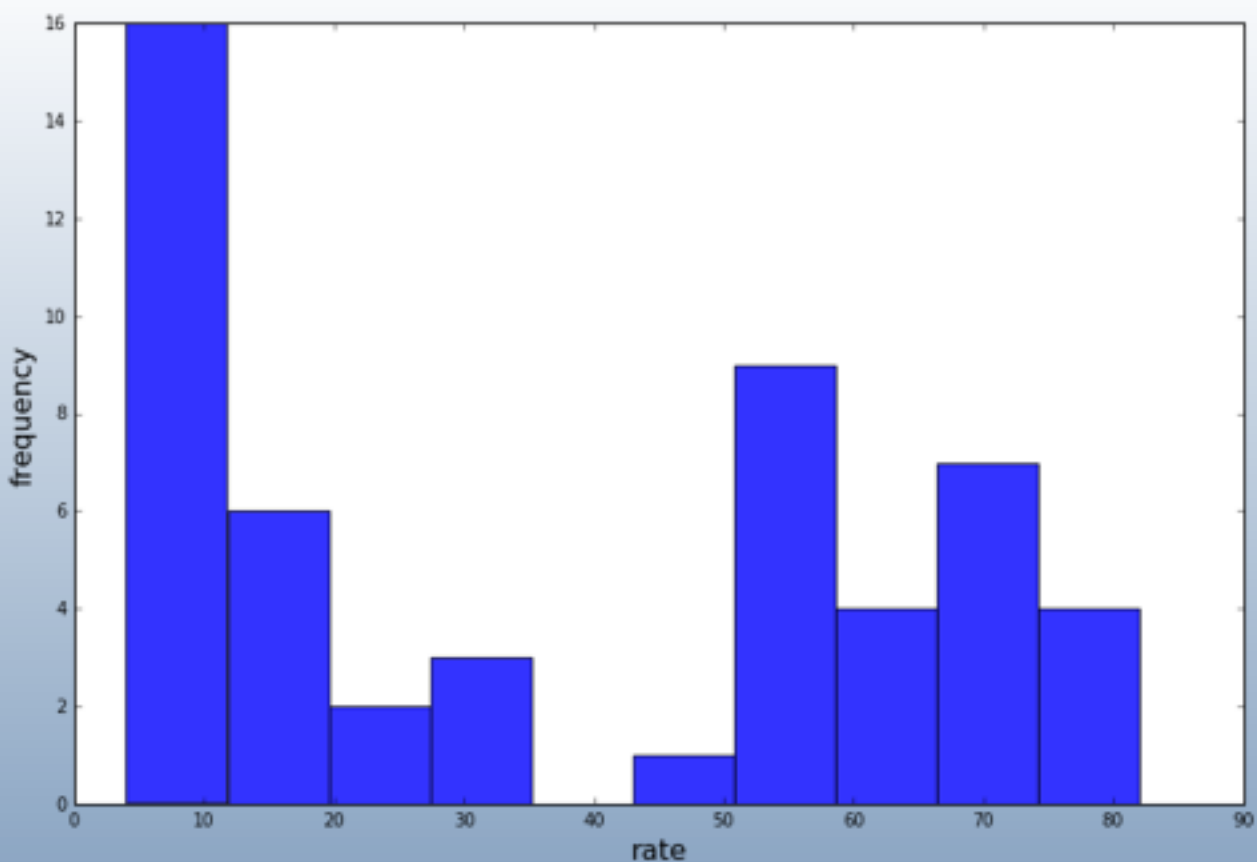
```
    return sd
```

```
print func()
```

- Result: [27.037996494523586, 32.915094961603266, 35.666996164305949]

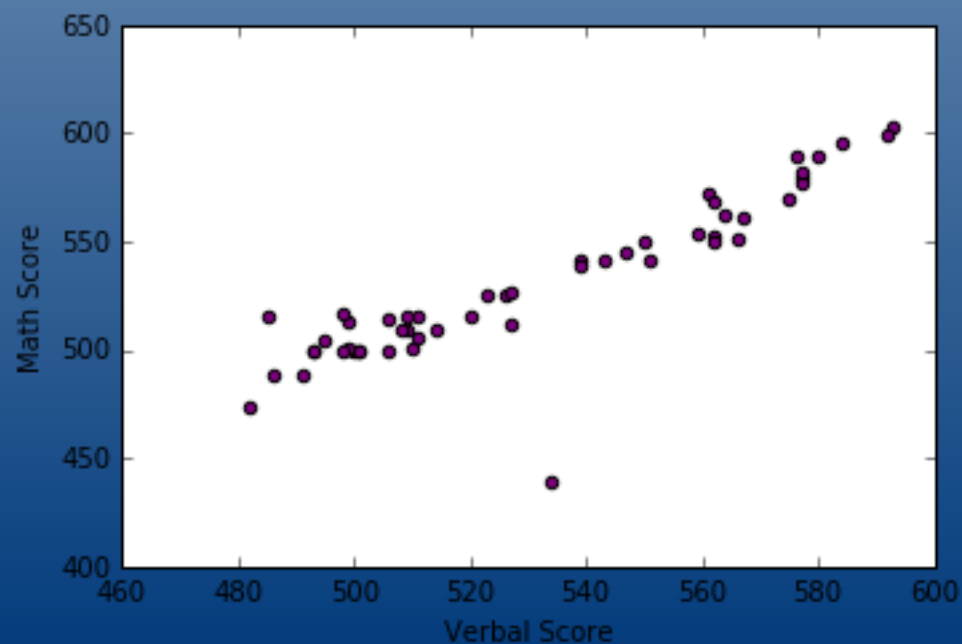
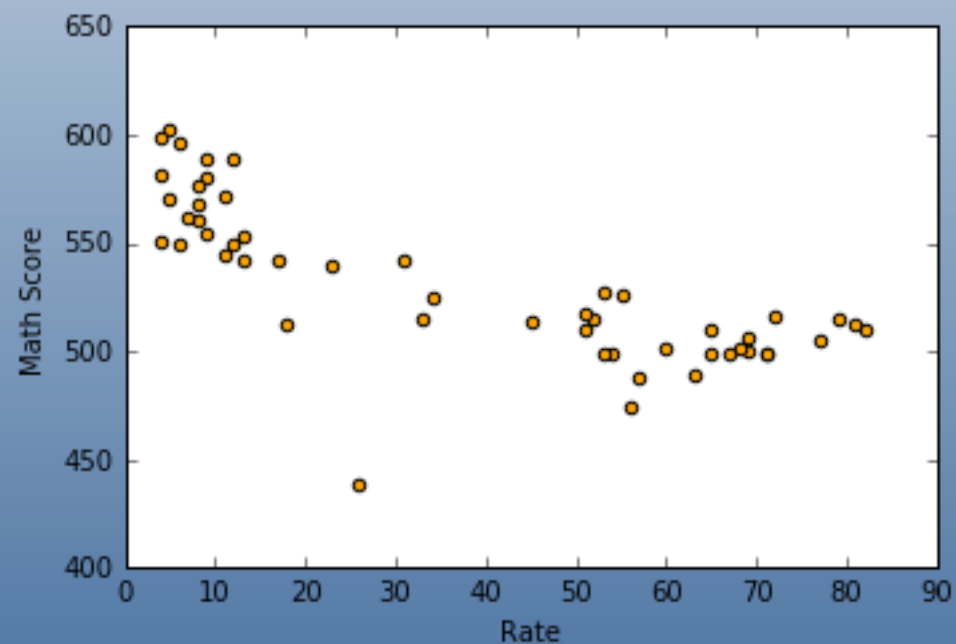
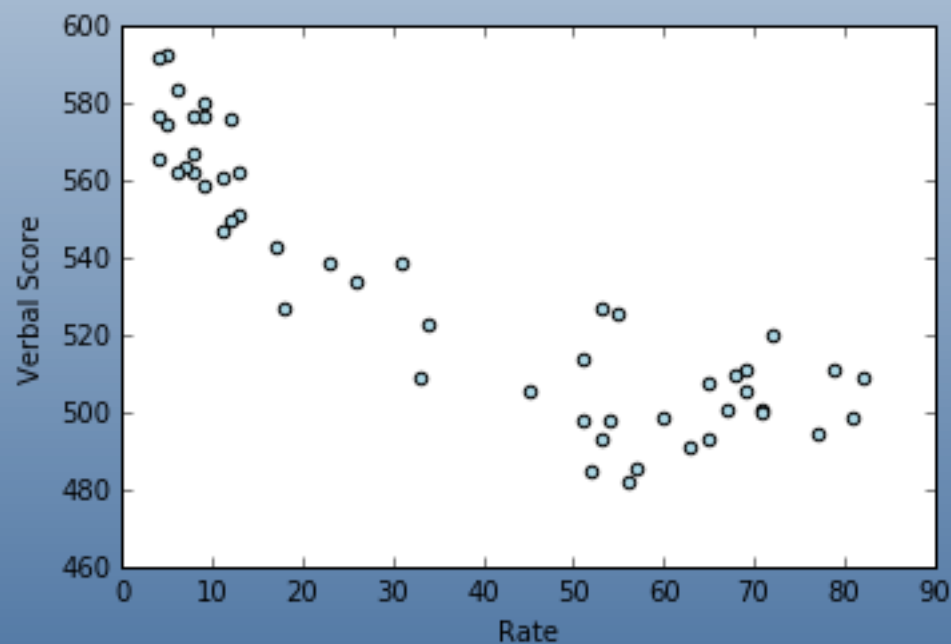
# Visualization

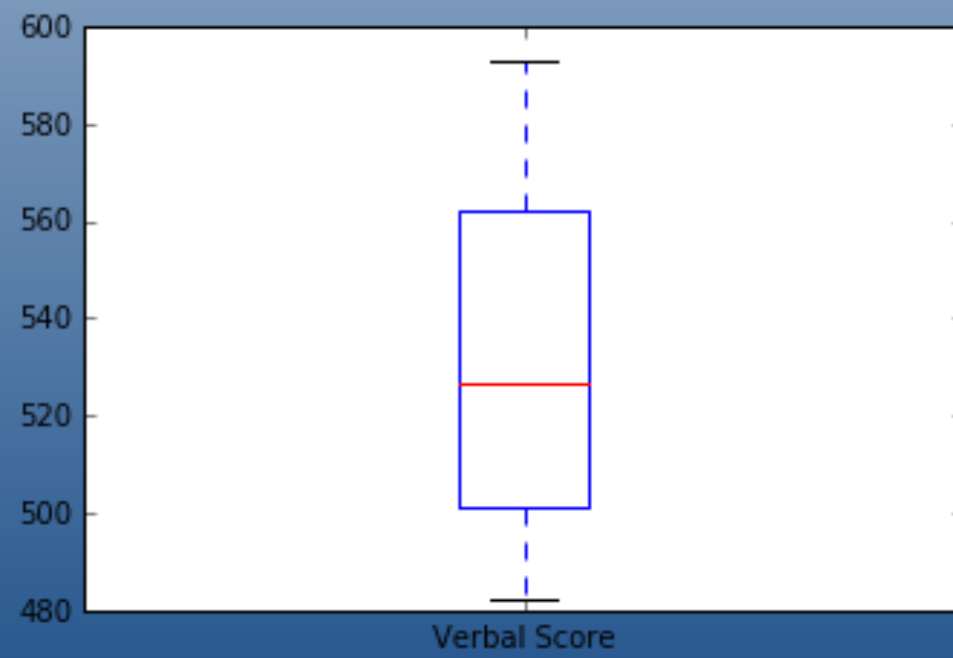
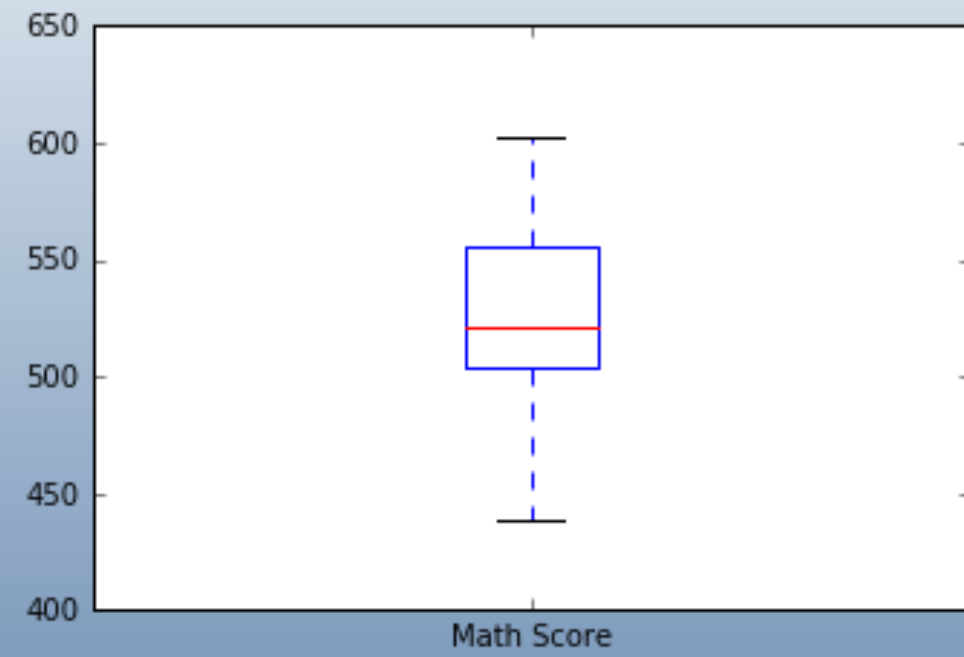
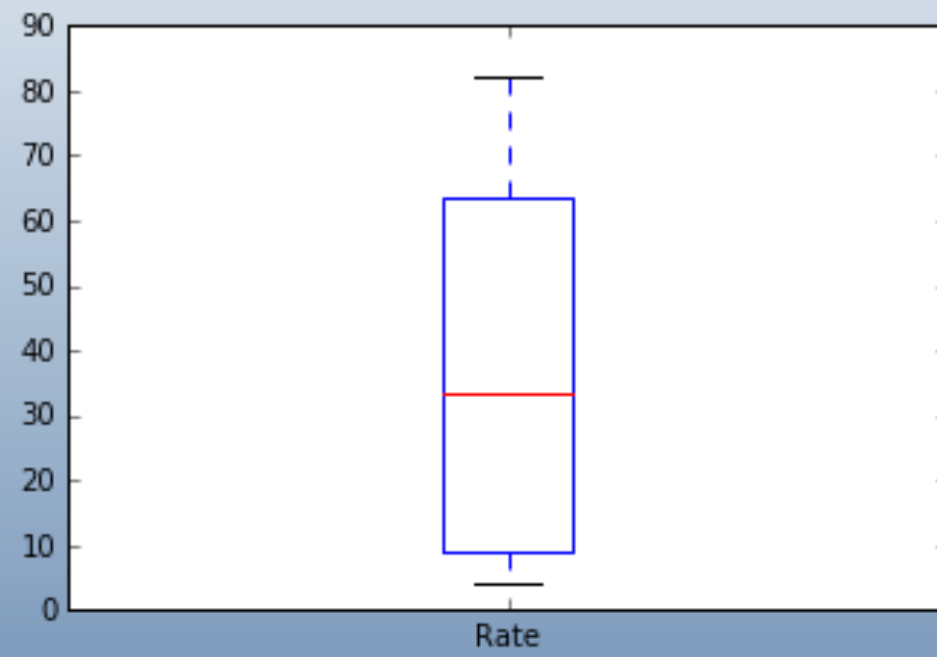




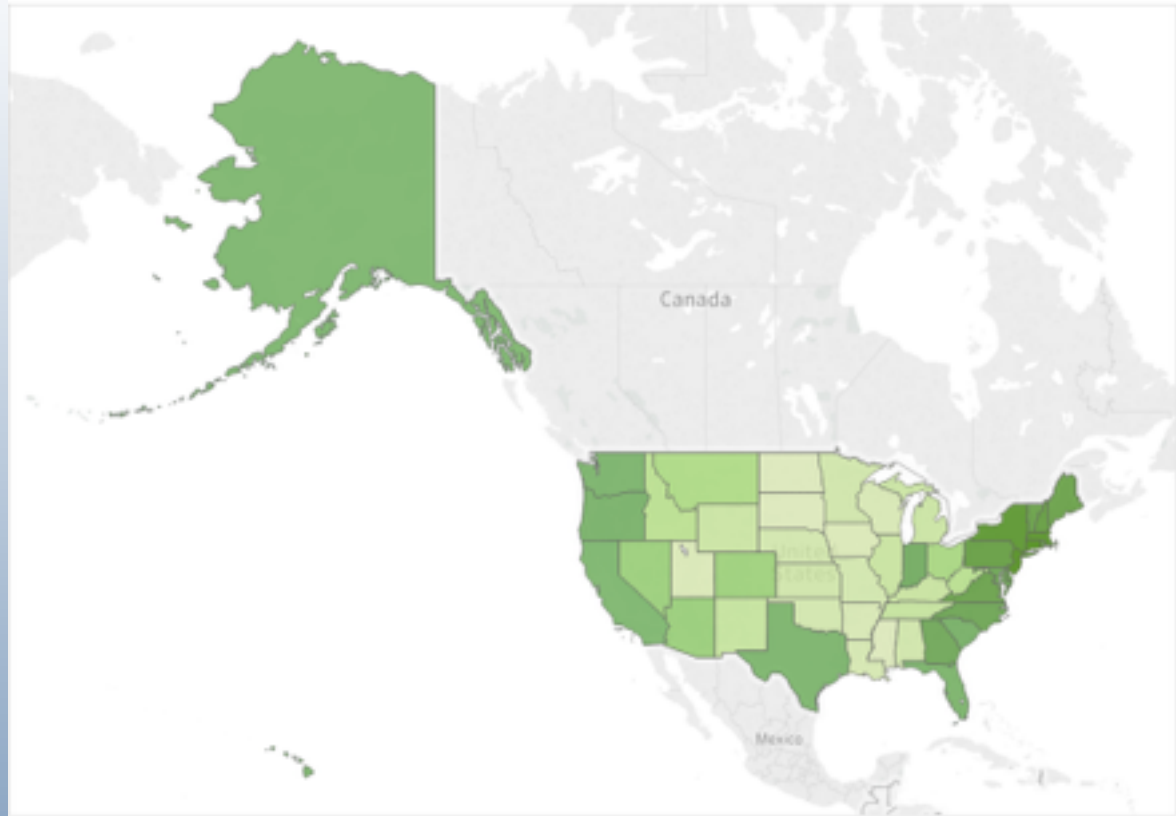
Rate and Verbal Score dist  
don't seem like normal.  
Possible Reasons:  
1. small sample size.  
2. wrong data

- Negative correlation between rate and test scores.
- Positive correlation between verbal and math scores.
- Both make sense.



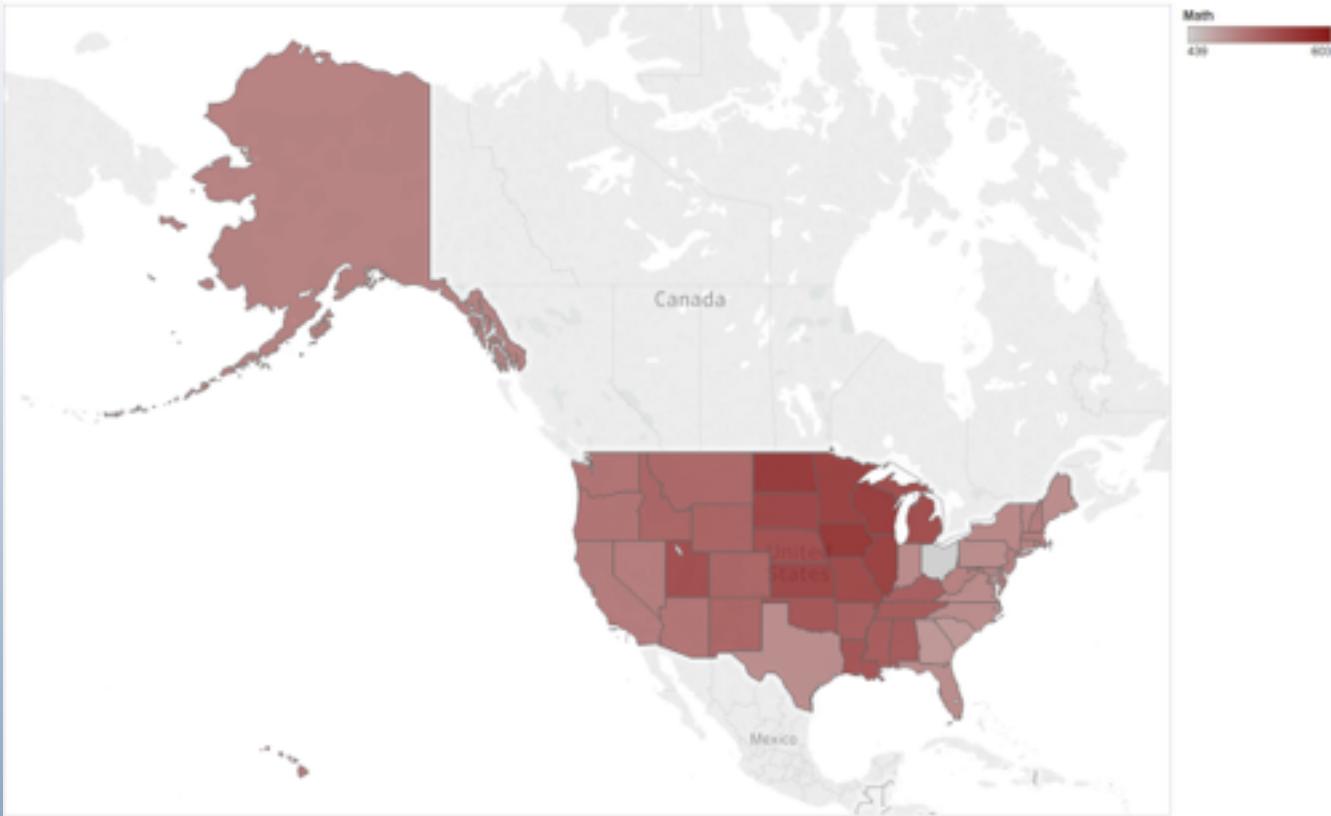


Sheet 1



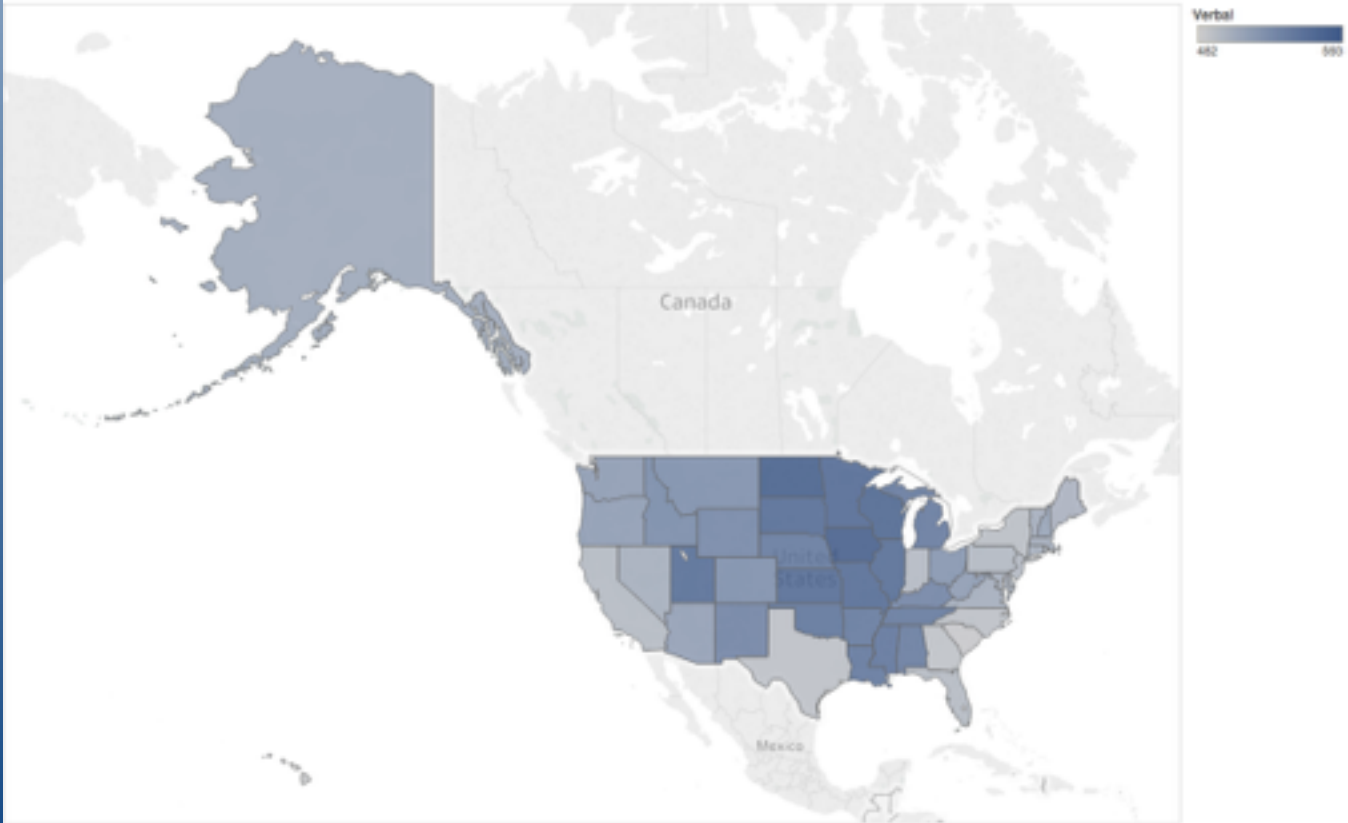
Map based on Longitude (generated) and Latitude (generated). Color shows sum of Rate. Details are shown for State.

Sheet 1



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Math. Details are shown for State.

Sheet 1



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Verbal. Details are shown for State.

# Conclusion

- Next steps: linear regression?
- Need more information and data to draw useful conclusion