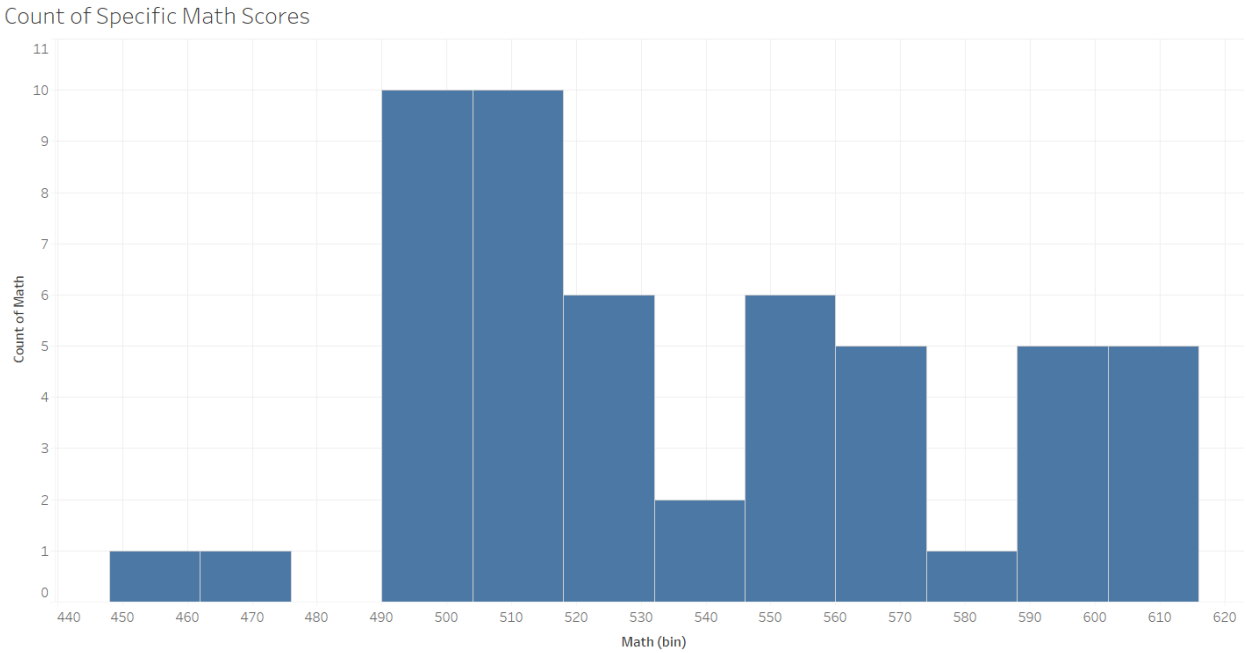
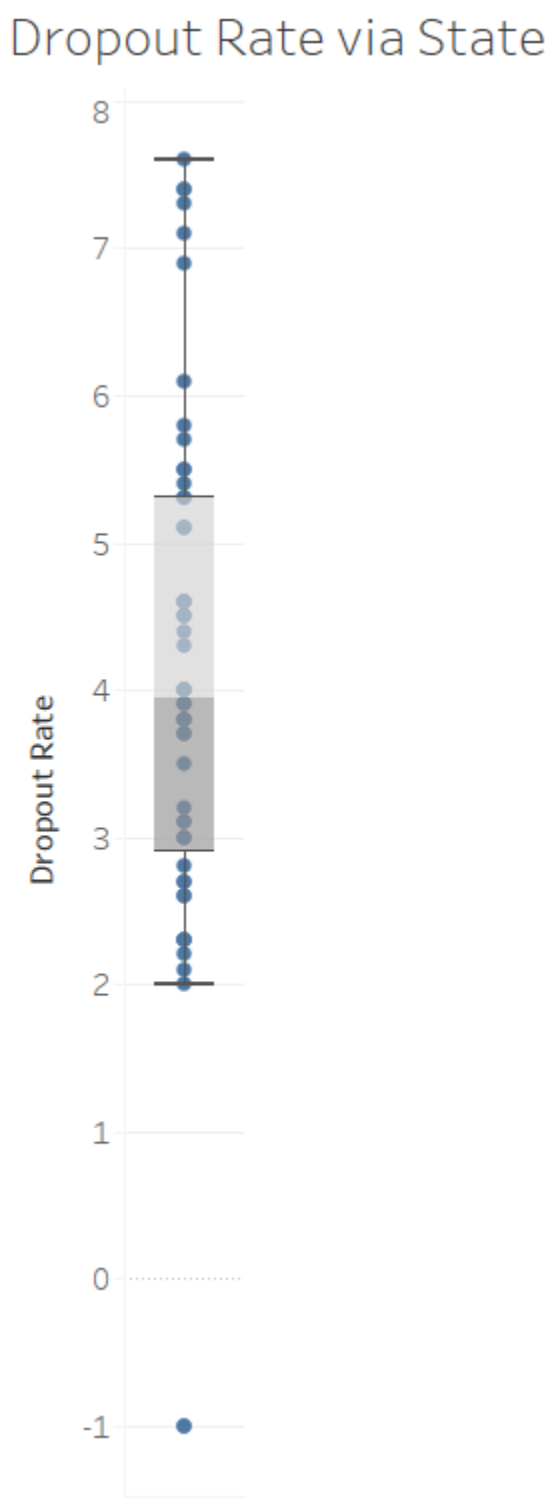
Tableau

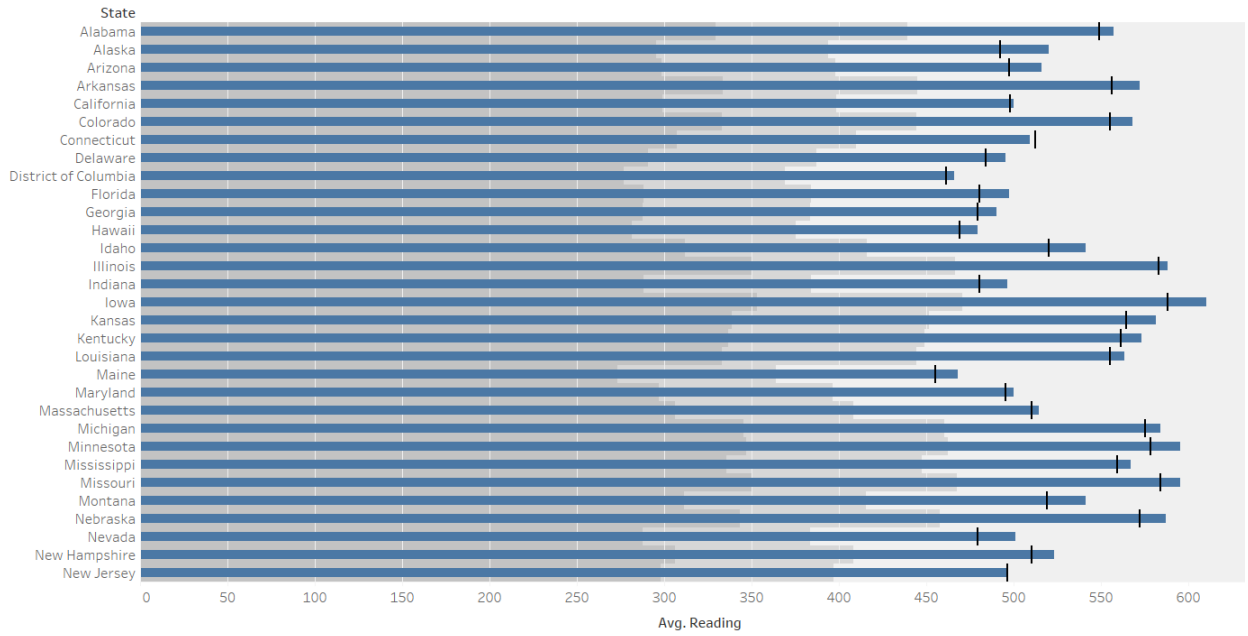
Histogram:



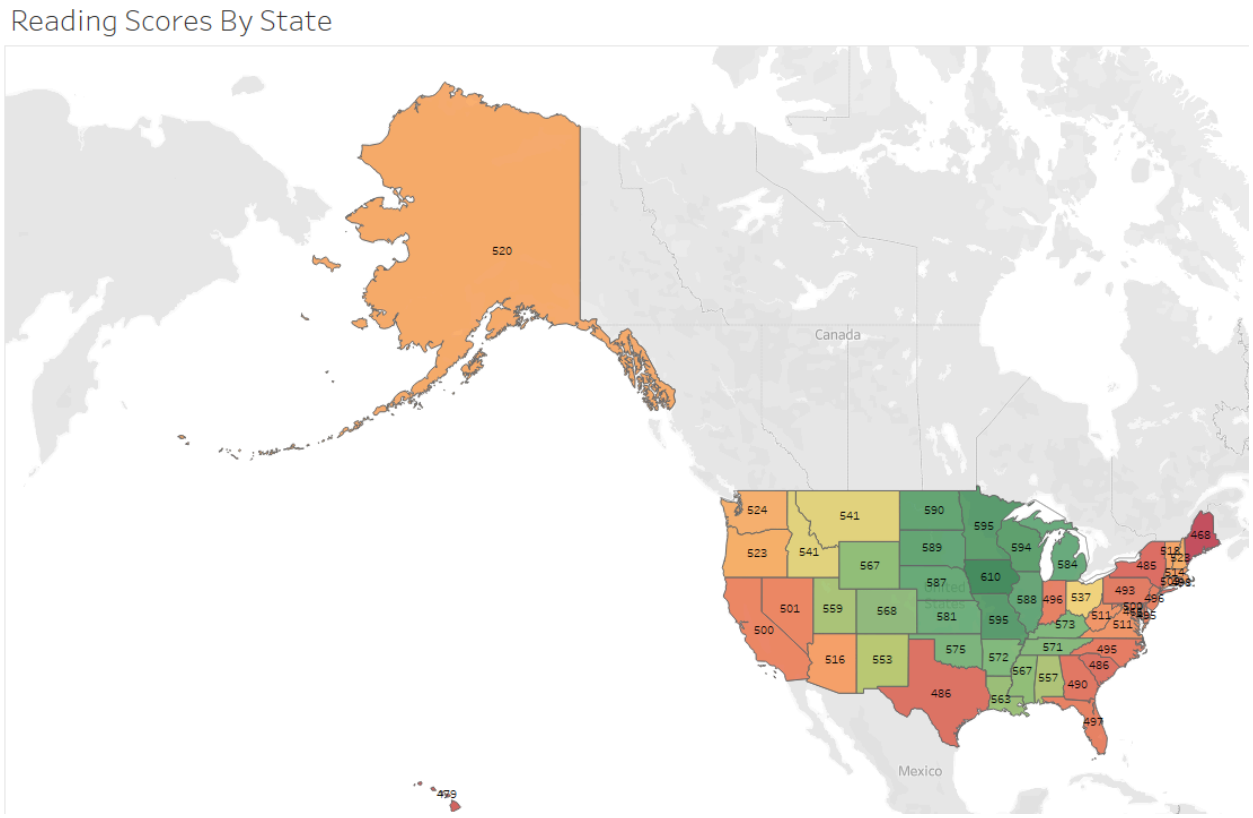
Box Plot



Bullet Chart

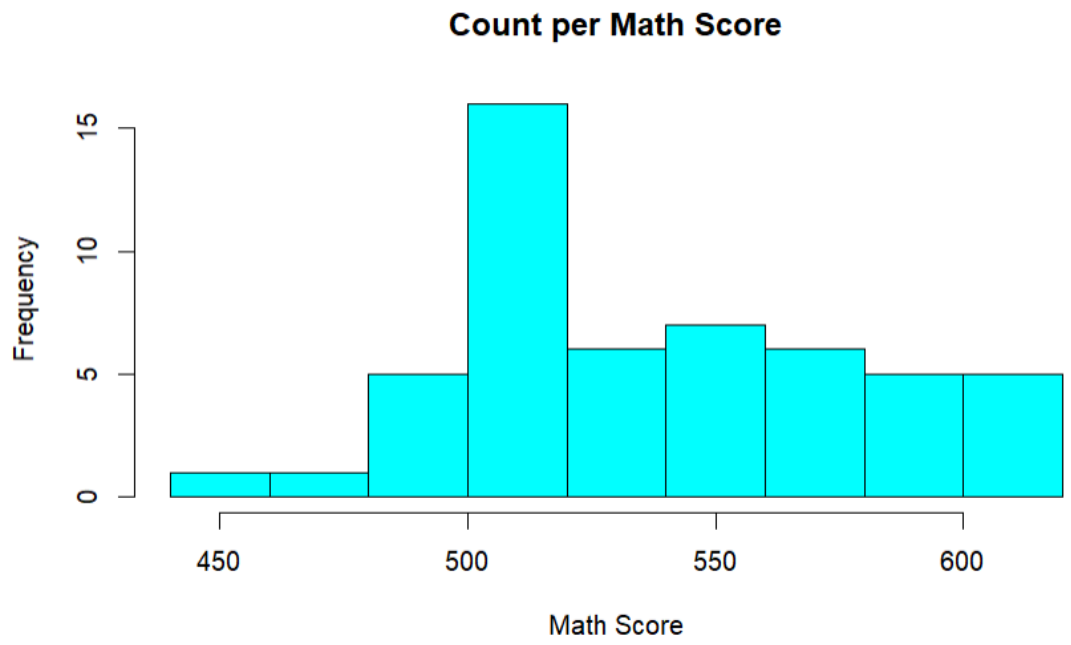


GeoSpatial Map

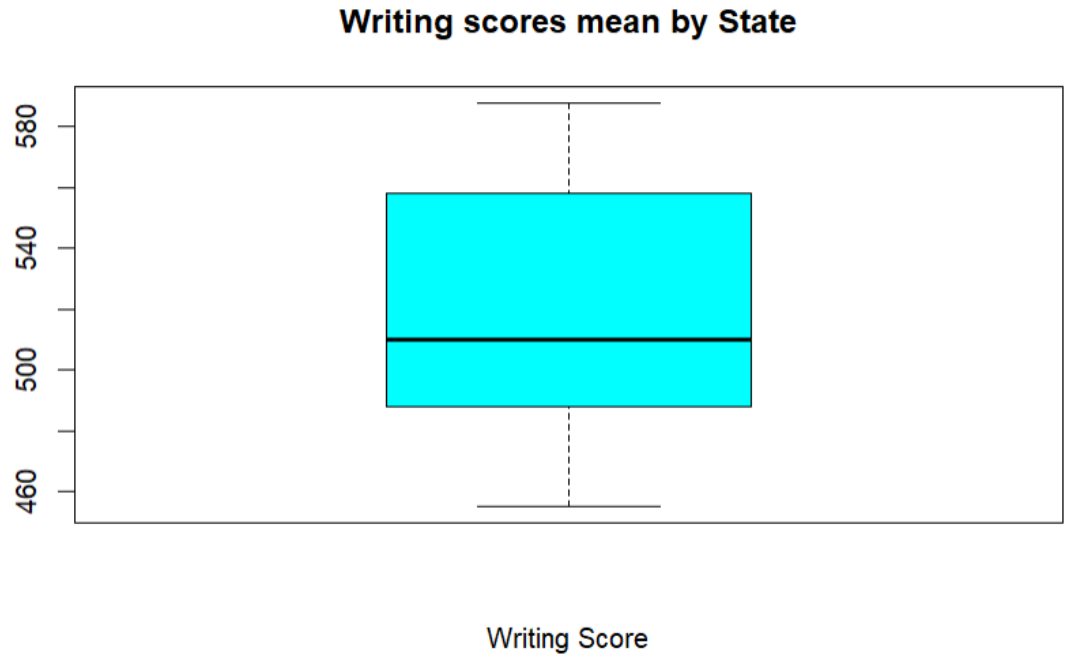


Rstudio

Histrogram



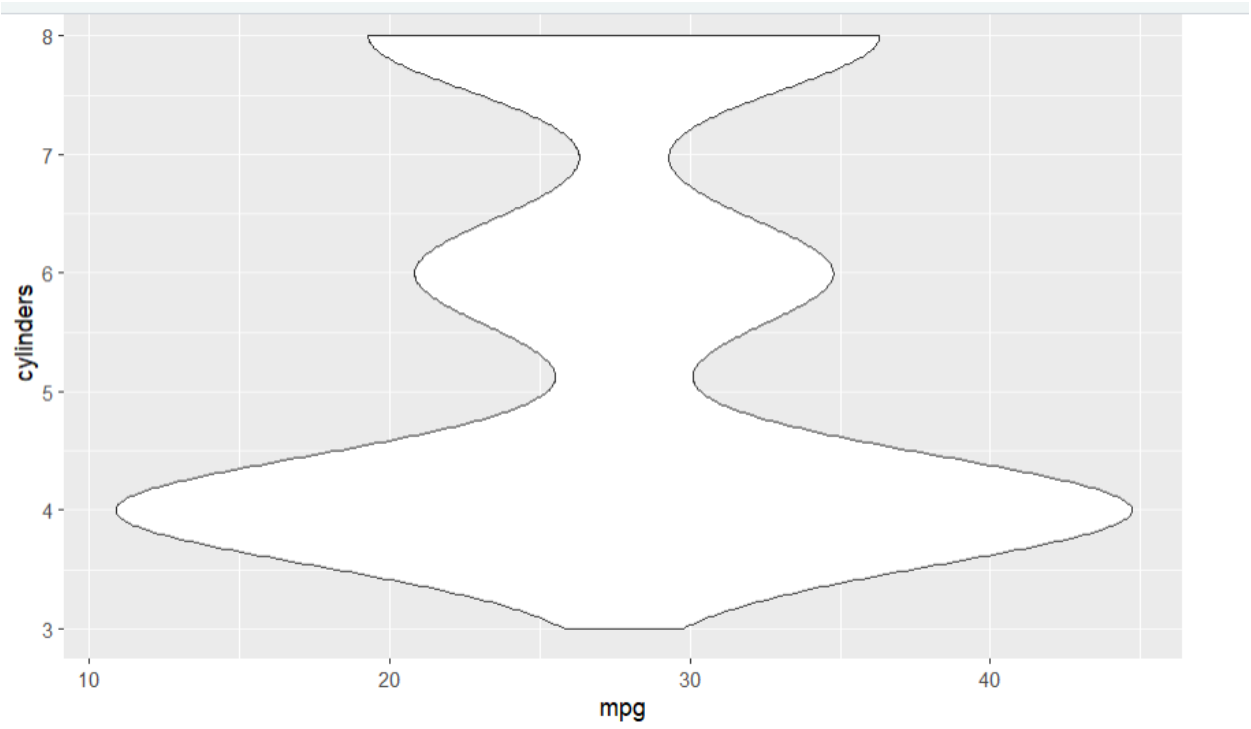
Box Plot



Bullet Chart

#got some code but couldn't figure out how to make it.

Violin Chart!



In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import chart_studio.plotly as py
import plotly.figure_factory as ff
```

In [2]:
```
#import data to dataframe, number 1
df1=pd.read_csv('birth-rate.csv')
df1.head()
```

Out[2]:

Country	1960	1961	1962	1963	1964	1965	1966	1967	1968	...	1999	2000	2001	2002	2003	2004	2005	
Aruba	36.400	35.179	33.863	32.459	30.994	29.513	28.069	26.721	25.518	...	15.024	14.528	14.041	13.579	13.153	12.772	12.441	1
anistan	52.201	52.206	52.208	52.204	52.192	52.168	52.130	52.076	52.006	...	51.229	50.903	50.486	49.984	49.416	48.803	48.177	4
Angola	54.432	54.394	54.317	54.199	54.040	53.836	53.585	53.296	52.984	...	48.662	48.355	48.005	47.545	46.936	46.184	45.330	4
Albania	40.886	40.312	39.604	38.792	37.913	37.008	36.112	35.245	34.421	...	17.713	16.850	16.081	15.444	14.962	14.644	14.485	1
erlands Antilles	32.321	30.987	29.618	28.229	26.849	25.518	24.280	23.173	22.230	...	15.809	15.412	15.096	14.824	14.565	14.309	14.051	1

× 50 columns

In [3]:
```
#import data to dataframe, number 1
df2=pd.read_csv('education.csv')
df2.head()
```

Out[3]:

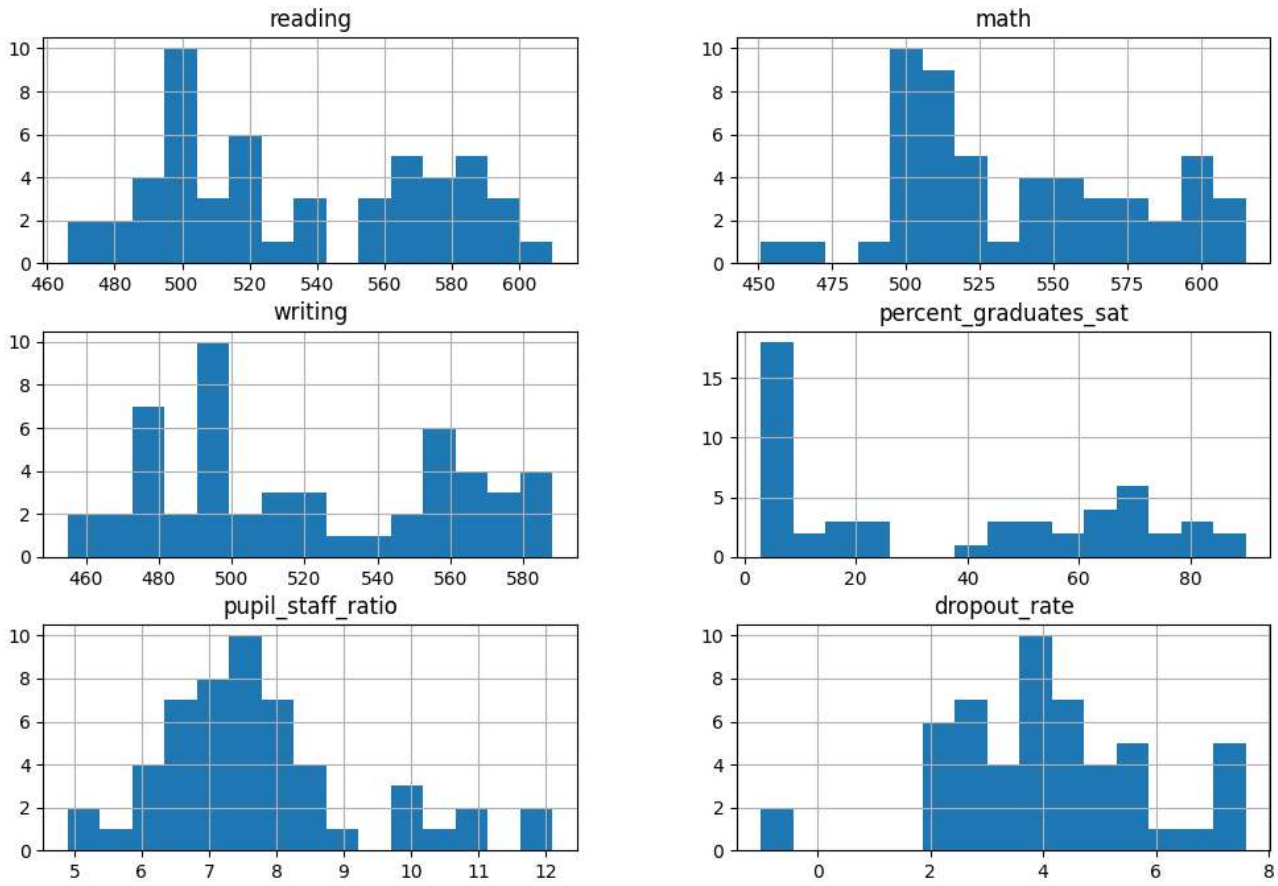
	state	reading	math	writing	percent_graduates_sat	pupil_staff_ratio	dropout_rate
0	United States	501	515	493	46	7.9	4.4
1	Alabama	557	552	549	7	6.7	2.3
2	Alaska	520	516	492	46	7.9	7.3
3	Arizona	516	521	497	26	10.4	7.6
4	Arkansas	572	572	556	5	6.8	4.6

In [4]:
```
#import data to dataframe, number 1
df3=pd.read_csv('crimeratesbystate-formatted.csv')
df3.head()
```

Out[4]:

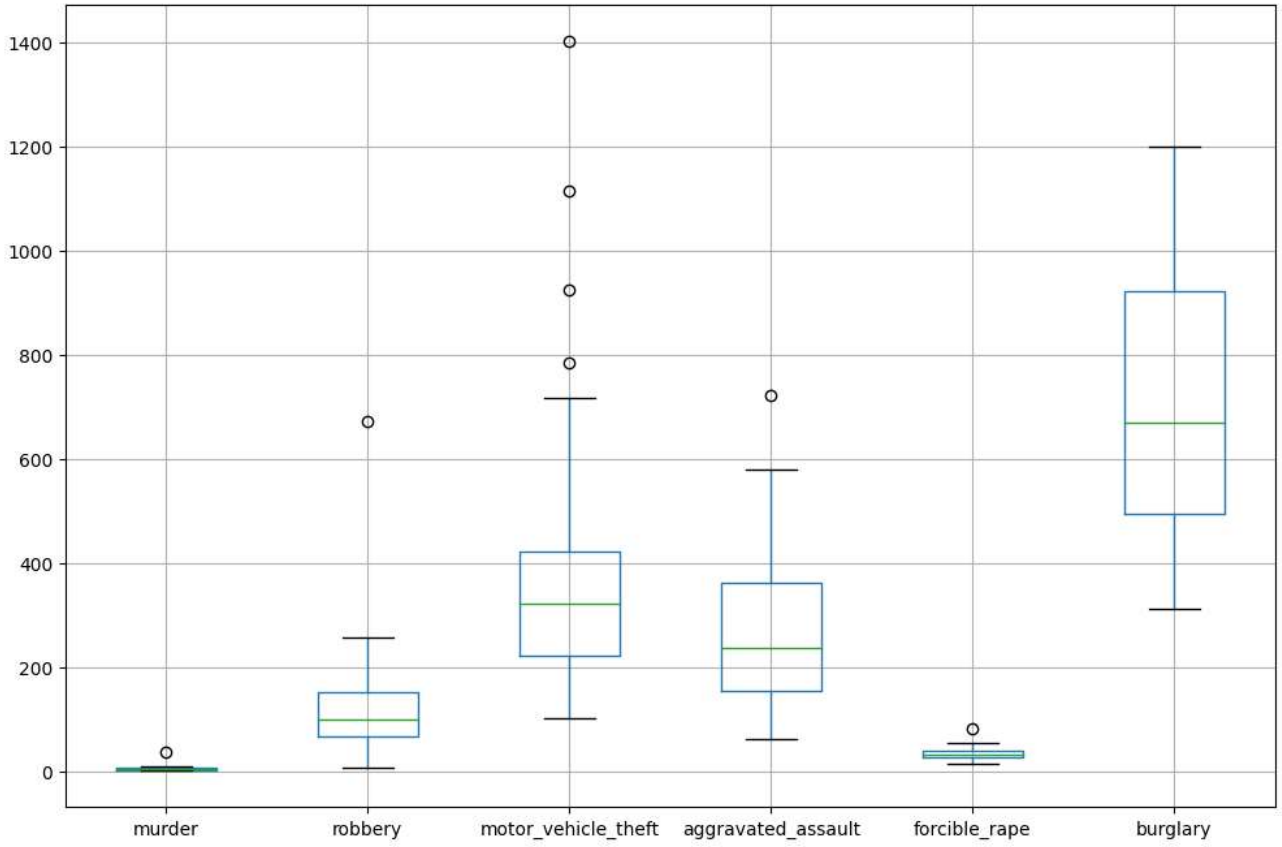
	state	murder	forcible_rape	robbery	aggravated_assault	burglary	larceny_theft	motor_vehicle_theft
0	United States	5.6	31.7	140.7	291.1	726.7	2286.3	416.7
1	Alabama	8.2	34.3	141.4	247.8	953.8	2650.0	288.3
2	Alaska	4.8	81.1	80.9	465.1	622.5	2599.1	391.0
3	Arizona	7.5	33.8	144.4	327.4	948.4	2965.2	924.4
4	Arkansas	6.7	42.9	91.1	386.8	1084.6	2711.2	262.1

In [7]:
```
#histogram using dataset 2. showing the number of instances per column
df2.hist(figsize=[12, 8], bins=15)
plt.show()
```



In [8]:
```
#boxplot using dataset 3. the outliers of multiple columns.
df3.boxplot(column=['murder', 'robbery', 'motor_vehicle_theft' ,'aggravated_assault', 'forcible_rape', 'burglary
```

Out[8]: <AxesSubplot: >



In [16]:
```
df4 = df2.head()
df4


  #add legend and format borders
  plt.legend(loc=(0.35,1.0))
  ax.spines['left'].set_linewidth(0.2)
  ax.spines['bottom'].set_linewidth(0)
  ax.spines['right'].set_linewidth(0)
  ax.spines['top'].set_linewidth(0)
  fig.subplots_adjust(left=0.2, top=0.8)
```

Out[16]:

	state	reading	math	writing	percent_graduates_sat	pupil_staff_ratio	dropout_rate	col
0	United States	501	515	493	46	7.9	4.4	gold
1	Alabama	557	552	549	7	6.7	2.3	gold
2	Alaska	520	516	492	46	7.9	7.3	gold
3	Arizona	516	521	497	26	10.4	7.6	gold
4	Arkansas	572	572	556	5	6.8	4.6	gold

Used this website for the below code. Not sure how to source code used? Had a very hard time figuring this out.
https://community.sisense.com/t5/knowledge/python-bullet-charts/ta-p/9381

In [31]:
```
def bullet_chart(df4,color_code=False):
  y_pos = np.arange(len(df4.index))

  #assign coloring
  df4["col"]="indigo"
  if (color_code==True):
    for i in y_pos:
      if(df4["reading"][i]>=800):
        df4["col"][i]="green"
      elif(df4["reading"][i]>=400):
        df4["col"][i]="gold"
      else:
        df4["col"][i]="lightcoral"

  #Initialize plot
  fig, ax = plt.subplots()
  ax.barh(y_pos, 800, height=0.5, align='center', color='thistle', label="stretch goal")
  ax.barh(y_pos, 400, height=0.5, align='center', color='mediumorchid', label = "base goal")
  ax.barh(y_pos, df4["reading"], height=0.2, align='center',color=df4["col"])
  ax.set_yticklabels(df4["state"])
  ax.set_yticks(y_pos)
  ax.invert_yaxis()
  ax.set_title("Math SAT Scores")

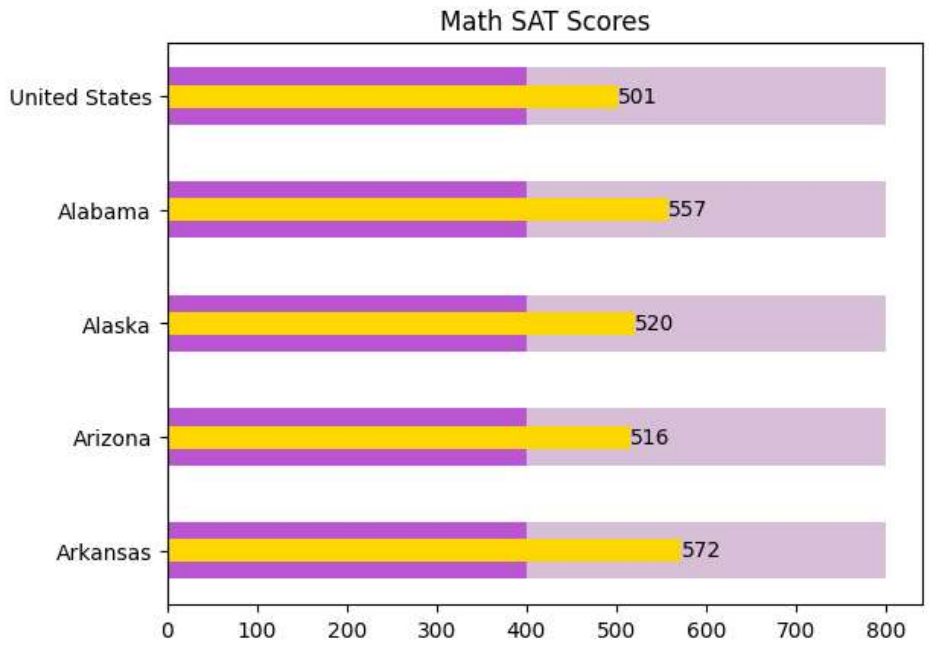

  #add data labels
  for i in y_pos:
    ax.text(df4["reading"][i], i+0.05, df4["reading"][i])

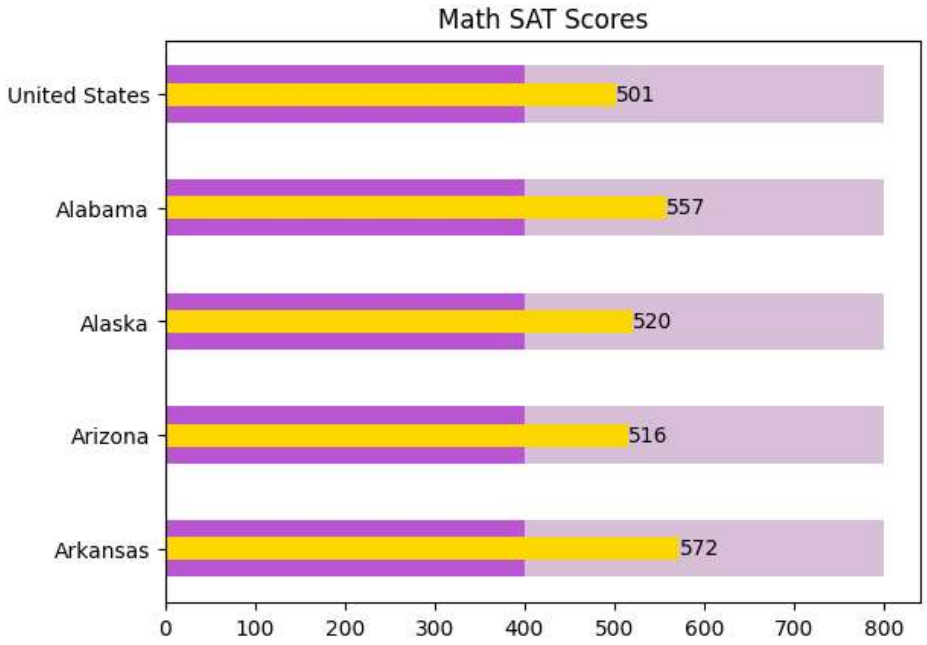
  return fig

bullet_chart(df4, color_code=True)
```

```
C:\Users\brean\AppData\Local\Temp\ipykernel_24276\675034634.py:20: UserWarning: FixedFormatter should only be
used together with FixedLocator
  ax.set_yticklabels(df4["state"])
```

Out[31]:

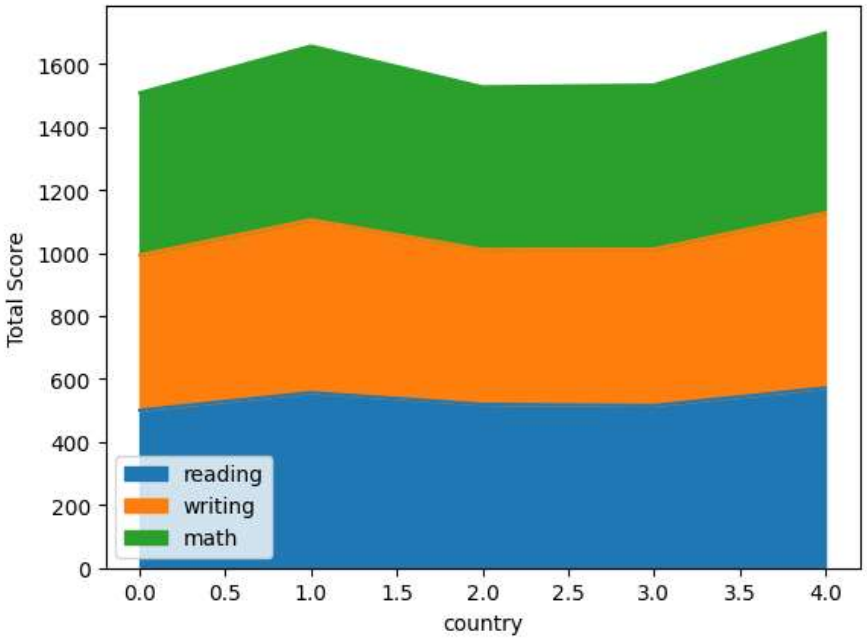




In [38]:
```
df4 = df4[['reading', 'writing', 'math']]
```

In [42]:
```
df4.plot.area()
plt.xlabel("country")
plt.ylabel("Total Score")
```

Out[42]:
```
Text(0, 0.5, 'Total Score')
```



````
---
title: "final"
output: html_document
date: "2024-02-24"
---


```{r setup, include=FALSE}
setwd("C:/Users/brean/OneDrive/Desktop/640/final/breannaparkerdscc640final")
knitr::opts_chunk$set(echo = TRUE)


```



```{r}
#load dataset
library(readr)

data <- read.csv("education.csv")
data

```

```{r}
#histogram
graph1 <- data$math
hist(graph1, main="Count per Math Score",
xlab="Math Score", col="cyan")


```

```{r}
#boxplot
boxplot(data$writing,
main = "Writing scores mean by State",
xlab = "Writing Score",
col = "cyan",
border = "black"
)

```
```{r}
#couldnot figure out the bullet chart for rstudio. this was the farthest that I got.
library(dplyr)


ggplot() +
  geom_bar(data = data,
    aes(x = data$state, y = data$reading, fill = math), stat = "identity",
    position = "stack") +
  geom_bar(data = data,
    aes(x = data$state, y = data$reading), fill = "black", width = .2,
    stat = "identity") +
  scale_fill_manual(values = bullet_colors) +
  coord_flip(expand = FALSE)

```

```{r}
#load second dataset
````
````

```
data2 <- read.csv("Automobile.csv")
data2
```

```{r}
library(ggplot2)
#violin chart!
ggplot(data2, aes(x=mpg, y=cylinders)) +
  geom_violin()
```