In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [12]:
```
%matplotlib inline
```

In [13]:
```
df = pd.read_csv("w11-12/education.csv")
df
```

Out[13]:

	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
5	California	500	513	498	49	10.9	5.5
6	Colorado	568	575	555	20	8.1	6.9
7	Connecticut	509	513	512	83	6.6	2.1
8	Delaware	495	498	484	71	7.9	5.5
9	District of Columbia	466	451	461	79	6.3	7.1
10	Florida	497	498	480	59	8.1	3.8
11	Georgia	490	491	479	71	7.0	4.6
12	Hawaii	479	502	469	58	8.3	5.4
13	Idaho	541	540	520	18	10.0	2.6
14	Illinois	588	604	583	6	9.9	4.0
15	Indiana	496	507	480	63	7.5	2.7
16	Iowa	610	615	588	3	6.8	2.3
17	Kansas	581	589	564	7	8.6	2.7
18	Kentucky	573	573	561	7	6.6	3.0
19	Louisiana	563	558	555	7	6.8	7.4
20	Maine	468	467	455	90	4.9	5.3
21	Maryland	500	502	495	69	7.2	3.8

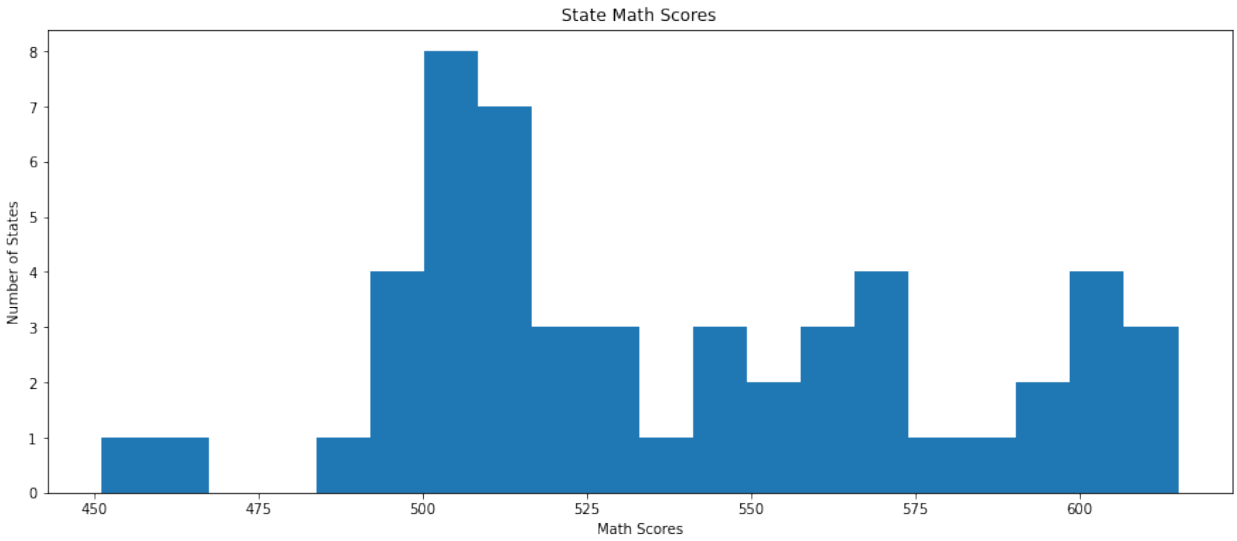
22	Massachusetts	514	526	510	84	7.8	3.8
23	Michigan	584	603	575	5	8.1	7.4
24	Minnesota	595	609	578	7	7.7	3.0
25	Mississippi	567	554	559	4	6.9	4.3
26	Missouri	595	600	584	5	6.9	3.7
27	Montana	541	542	519	22	7.4	3.7
28	Nebraska	587	594	572	4	6.7	2.8
29	Nevada	501	505	479	42	12.1	4.5
30	New Hampshire	523	523	510	75	6.2	3.2
31	New Jersey	496	513	496	76	6.9	2.0
32	New Mexico	553	546	534	11	7.0	6.1
33	New York	485	502	478	85	7.4	5.3
34	North Carolina	495	511	480	63	7.3	5.7
35	North Dakota	590	593	566	3	6.2	2.3
36	Ohio	537	546	523	22	7.5	4.5
37	Oklahoma	575	571	557	5	7.4	3.5
38	Oregon	523	525	499	52	8.8	4.6
39	Pennsylvania	493	501	483	71	7.1	-1.0
40	Rhode Island	498	496	494	66	8.4	5.8
41	South Carolina	486	496	470	67	10.8	3.9
42	South Dakota	589	600	569	3	7.1	3.9
43	Tennessee	571	565	565	10	7.6	3.1
44	Texas	486	506	475	51	7.4	4.0
45	Utah	559	558	540	6	11.9	3.1
46	Vermont	518	518	506	64	4.9	-1.0
47	Virginia	511	512	498	68	6.0	2.6
48	Washington	524	531	507	53	9.9	5.1
49	West Virginia	511	501	499	18	7.4	4.0
50	Wisconsin	594	608	582	5	8.3	2.2
51	Wyoming	567	568	550	5	5.6	5.1

In [14]:
```
math=df['math']
math
```

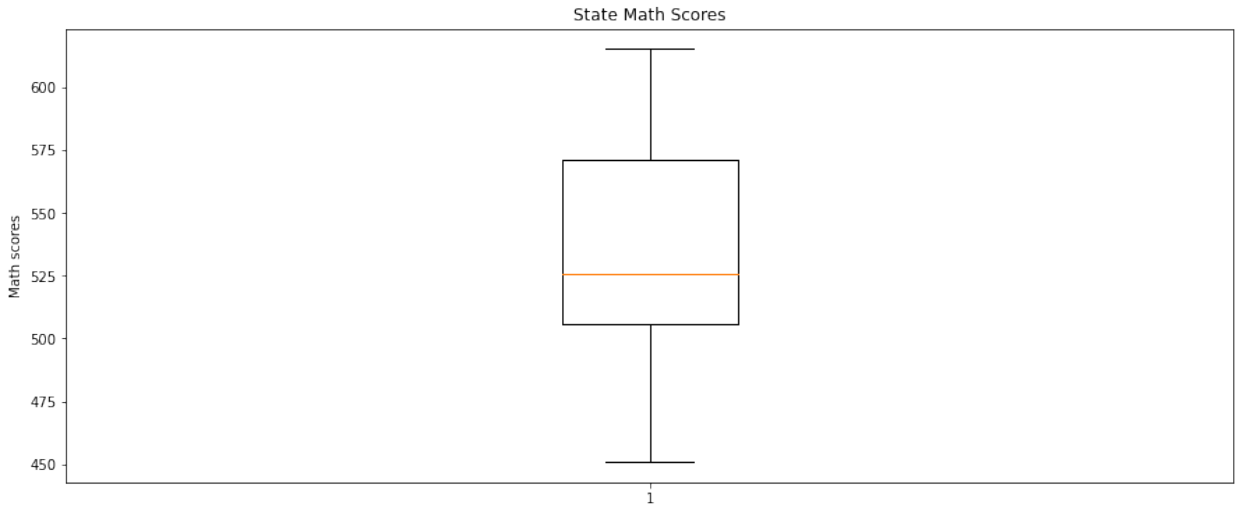
Out[14]:
```
0     515
1     552
2     516
3     521
4     572
5     513
6     575
7     513
8     498
9     451
10    498
11    491
12    502
13    540
14    604
15    507
16    615
17    589
18    573
19    558
20    467
21    502
22    526
23    603
24    609
25    554
26    600
27    542
28    594
29    505
30    523
31    513
32    546
33    502
34    511
35    593
36    546
37    571
38    525
39    501
40    496
41    496
42    600
43    565
44    506
45    558
46    518
```

```
47    512
48    531
49    501
50    608
51    568
Name: math, dtype: int64
```

In [15]:
```
plt.figure(figsize=(15, 6))
plt.hist(math, bins=20)
plt.title('State Math Scores')
plt.xlabel('Math Scores')
plt.ylabel('Number of States')
plt.show()
```



In [16]:
```
plt.figure(figsize=(15, 6))
plt.boxplot(math)
plt.title('State Math Scores')
plt.ylabel('Math scores')
plt.show()
```



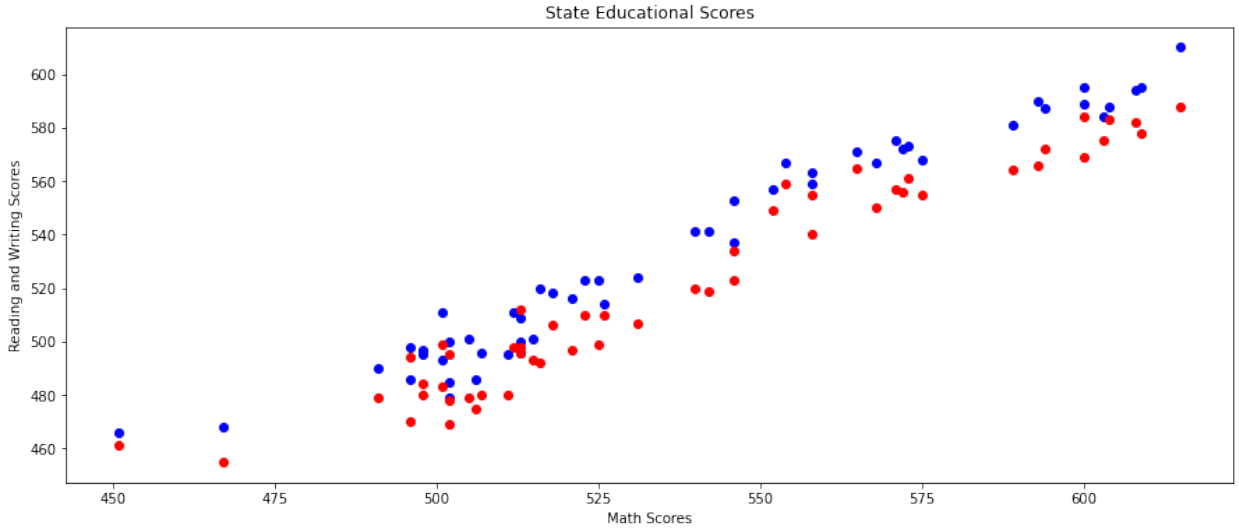
In [19]:
```
palette = sns.color_palette("Blues_r", len(math))
```

In [24]:
```
fig, ax = plt.subplots()
ax.set_aspect('equal')
ax.set_yticks([1])
ax.set_yticklabels(['Math Scores'])

prev_limit = 0
for idx, lim in enumerate(math):
    ax.barh([1], lim-prev_limit, left=prev_limit, height=15, color=pal
    prev_limit = lim
```

In [25]:
```
read= df['reading']
write=df['writing']
```

In [26]:
```
plt.figure(figsize=(15, 6))
plt.scatter(math,read,c='b',label='Math Scores Versus Reading Scores')
plt.scatter(math,write,c='r',label='Math Scores Versus Writing Scores'
plt.title('State Educational Scores')
plt.xlabel('Math Scores')
plt.ylabel('Reading and Writing Scores')
plt.show()
```



In [ ]: