In [ ]:

```
# Project 3

## SDS348 Spring 2021
```

In [ ]:

```
### Hannah LeBlanc (hkl362)
```

In [66]:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
#import packages
```

```
In [24]: studentp = pd.read_csv("student.csv", index_col = 0) #import the Student Preferences dataset
         print(studentp) #show first few rows of dataset
```

```
       race/ethnicity parental_level_of_education         lunch  \
gender
female        group B           bachelor's degree      standard
female        group C                some college      standard
female        group B             master's degree      standard
male          group A          associate's degree  free/reduced
male          group C                some college      standard
female        group B          associate's degree      standard
female        group B                some college      standard
male          group B                some college  free/reduced
male          group D                 high school  free/reduced
female        group B                 high school  free/reduced
male          group C          associate's degree      standard
male          group D          associate's degree      standard
female        group B                 high school      standard
male          group A                some college      standard
female        group A             master's degree      standard
female        group C            some high school      standard
male          group C                 high school      standard
female        group B            some high school  free/reduced
male          group C             master's degree  free/reduced
female        group C          associate's degree  free/reduced
male          group D                 high school      standard
female        group B                some college  free/reduced
male          group D                some college      standard
female        group C            some high school      standard
male          group D           bachelor's degree  free/reduced
male          group A             master's degree  free/reduced
male          group B                some college      standard
female        group C           bachelor's degree      standard
male          group C                 high school      standard
female        group D             master's degree      standard
...               ...                         ...           ...
female        group D           bachelor's degree      standard
male          group C            some high school      standard
female        group A                 high school  free/reduced
female        group D                some college  free/reduced
female        group A                some college      standard
female        group C                some college      standard
male          group B                some college  free/reduced
male          group C          associate's degree      standard
male          group D                 high school      standard
female        group C          associate's degree      standard
female        group B                 high school  free/reduced
male          group D            some high school      standard
male          group B            some high school      standard
female        group A                some college      standard
female        group C            some high school      standard
male          group A                 high school      standard
female        group C          associate's degree      standard
male          group E            some high school      standard
female        group A            some high school  free/reduced
female        group D                some college  free/reduced
male          group E                 high school  free/reduced
female        group B            some high school      standard
female        group D          associate's degree  free/reduced
female        group D           bachelor's degree  free/reduced
```

```
male          group A                high school      standard
female        group E            master's degree      standard
male          group C                high school  free/reduced
female        group C                high school  free/reduced
female        group D               some college      standard
female        group D               some college  free/reduced

       test_preparation course  math_score  reading_score  writing_scor
e
gender
female                    none          72             72             7
4
female               completed          69             90             8
8
female                    none          90             95             9
3
male                      none          47             57             4
4
male                      none          76             78             7
5
female                    none          71             83             7
8
female               completed          88             95             9
2
male                      none          40             43             3
9
male                 completed          64             64             6
7
female                    none          38             60             5
0
male                      none          58             54             5
2
male                      none          40             52             4
3
female                    none          65             81             7
3
male                 completed          78             72             7
0
female                    none          50             53             5
8
female                    none          69             75             7
8
male                      none          88             89             8
6
female                    none          18             32             2
8
male                 completed          46             42             4
6
female                    none          54             58             6
1
male                      none          66             69             6
3
female               completed          65             75             7
0
male                      none          44             54             5
3
female                    none          69             73             7
```

```
3
male                 completed          74             71             8
0
male                      none          73             74             7
2
male                      none          69             54             5
5
female                    none          67             69             7
5
male                      none          70             70             6
5
female                    none          62             70             7
5
...                        ...         ...            ...
...
female                    none          89            100            10
0
male                 completed          78             72             6
9
female               completed          53             50             6
0
female                    none          49             65             6
1
female                    none          54             63             6
7
female               completed          64             82             7
7
male                 completed          60             62             6
0
male                      none          62             65             5
8
male                 completed          55             41             4
8
female                    none          91             95             9
4
female                    none           8             24             2
3
male                      none          81             78             7
8
male                 completed          79             85             8
6
female               completed          78             87             9
1
female                    none          74             75             8
2
male                      none          57             51             5
4
female                    none          40             59             5
1
male                 completed          81             75             7
6
female                    none          44             45             4
5
female               completed          67             86             8
3
male                 completed          86             81             7
5
```

```
female                completed          65             82             7
8
female                     none          55             76             7
6
female                     none          62             72             7
4
male                       none          63             63             6
2
female                completed          88             99             9
5
male                       none          62             55             5
5
female                completed          59             71             6
5
female                completed          68             78             7
7
female                     none          77             86             8
6

[1000 rows x 7 columns]
```

In [27]:

```
studentp.info() #show information about dataset 'studentp'
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1000 entries, female to female
Data columns (total 7 columns):
race/ethnicity                 1000 non-null object
parental_level_of_education    1000 non-null object
lunch                          1000 non-null object
test_preparation course        1000 non-null object
math_score                     1000 non-null int64
reading_score                  1000 non-null int64
writing_score                  1000 non-null int64
dtypes: int64(3), object(4)
memory usage: 62.5+ KB
```

In [ ]:

```
This dataset is known as Student Performances, which I renamed 'student
p' for ease of carrying out data analysis. The
dataset has 8 variables all together (the information above says 7, but
it neglects the column known as 'gender'). The categorical variables inc
luded in this dataset are as follows: gender, race/
ethnicity, parental level of education, lunch and test preparation cours
e (whether completed or not). Numerical
variables included math scores, reading scores, and writing scores. Ther
e were 1000 observations recorded per variable
in the dataset.
```

In [30]:
```
studentp.describe() #gives statistics for numeric variables in dataset s
tudentp
```

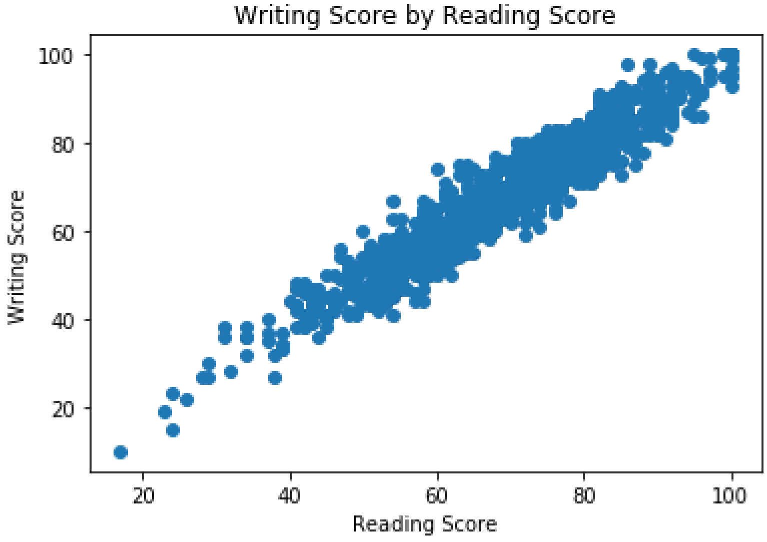
Out[30]:

	math_score	reading_score	writing_score
count	1000.00000	1000.000000	1000.000000
mean	66.08900	69.169000	68.054000
std	15.16308	14.600192	15.195657
min	0.00000	17.000000	10.000000
25%	57.00000	59.000000	57.750000
50%	66.00000	70.000000	69.000000
75%	77.00000	79.000000	79.000000
max	100.00000	100.000000	100.000000

In [ ]:
```
The summary statistic table above includes only the statistics for the n
umeric variables in the dataset. All three
numeric variables have counts of 1000 for the 1000 indiviuduals who were
observed across the variables. The summary
table, among many things, tells us the minimum and maximum values for ea
ch numerical variable, allowing for a calcula-
tion of range for each variable. The range of math scores was 100, the r
ange of reading scores was 83 and the range of
writing scores was 90. The summary statistic table gave a value for mean
, which represents the average or middle of
the data if it is normally distributed. The mean of math scores was 66.0
9, the mean of reading scores was 69.17 and
the mean of writing scores was 68.05.
```

In [31]:

```
plt.scatter(studentp.reading_score, studentp.writing_score)
plt.title("Writing Score by Reading Score")
plt.xlabel("Reading Score")
plt.ylabel("Writing Score")
plt.show()
#create a scatterplot to display the relationship between reading scores
and writing scores
```



In [ ]:

```
Anytime I am looking to improve my writing scores, the suggestion I rece
ive the most is that I need to pick up some
books and get reading. It was for this reason that I decided to explore
the relationship between the numerical variab-
les known as reading scores and writing scores. I expected to see a posi
tve correlation, the higher the reading score
the higher the writing score I expected to see. The graph shown above do
es in fact show the positive correlation I
expected to see based on my knowledge of the relationship between readin
g more books and being a stronger writer.
```

In [ ]:

```
The statistics that were not included in the summary statistics table ab
ove were the categorical variables. In order
to interpret categorical variables, we use counts. The following shows t
he counts for two of the categorical variables
in the dataset, race/ethnicity and parental level of education.
```

In [32]:

```
studentp['race/ethnicity'].value_counts() #decribe categorical variable
 'race/ethnicity'
```

Out[32]:

```
group C    319
group D    262
group B    190
group E    140
group A     89
Name: race/ethnicity, dtype: int64
```

In [ ]:

```
The counts above show that the largest number of individuals included in
the dataset, 319, belong in the group C race/
ethnicity category, while the smallest number of individuals, 89, are co
unted in the race/ethnicity group A. Group D
encompassed 262 individuals, group B had 190 and group E included 140.
```

In [33]:

```
studentp['parental_level_of_education'].value_counts() #decribe categori
cal variable 'parental level of education'
```

Out[33]:

```
some college          226
associate's degree    222
high school           196
some high school      179
bachelor's degree     118
master's degree        59
Name: parental_level_of_education, dtype: int64
```

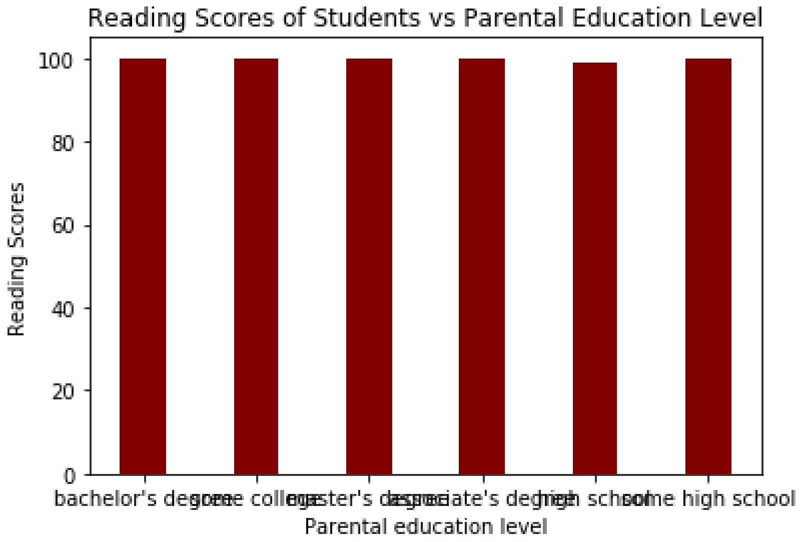
In [ ]:

```
The counts above give the level of education for parents of the students
involved in the dataset. The group with the
greatest parent count was the 'some college' group, which encompassed 22
6 parents. The smallest group, with 59 people,
represented the parents with a master's degree. Parents with an associat
e's degree numbered 222, the high school educ-
ation group has 196 parents, the some high school group included 179 par
ents, and the parents with a bachelor's degree
encompassed 118 parents.
```

In [75]:

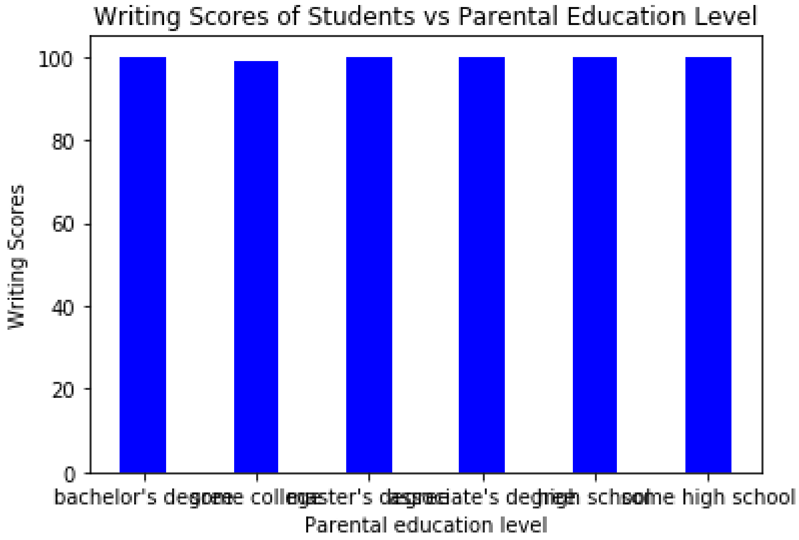
```
plt.bar(studentp.parental_level_of_education, studentp.reading_score, co
lor ='maroon',
        width = 0.4)

plt.xlabel("Parental education level")
plt.ylabel("Reading Scores")
plt.title("Reading Scores of Students vs Parental Education Level")
plt.show()
#create bar plot of the relationship between reading scores and parental
education level
```



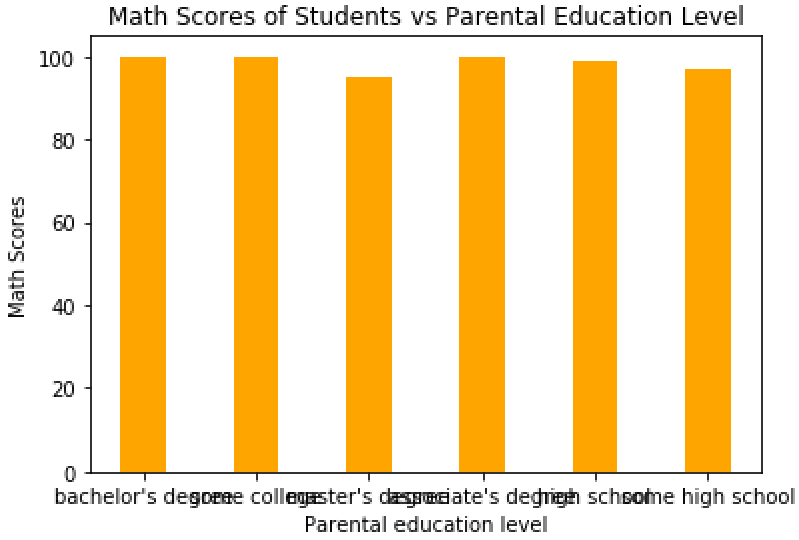
In [78]:
```
plt.bar(studentp.parental_level_of_education, studentp.writing_score, color ='blue',
        width = 0.4)

plt.xlabel("Parental education level")
plt.ylabel("Writing Scores")
plt.title("Writing Scores of Students vs Parental Education Level")
plt.show()
#create bar plot of the relationship between writing scores and parental education level
```



In [77]:
```
plt.bar(studentp.parental_level_of_education, studentp.math_score, color ='orange',
        width = 0.4)

plt.xlabel("Parental education level")
plt.ylabel("Math Scores")
plt.title("Math Scores of Students vs Parental Education Level")
plt.show()
#create bar plot of the relationship between reading scores and parental education level
```



In [ ]:

```
The three bar plots I included represent the categorical variable, paren
tal education level, in comparison to the scoresof students on the three
subjects taken into consideration during this study. I though perhaps th
ere may be an obviousdifference in the scores that could be seen on the
bar plots, however, that was not the case. Further research should becon
ducted in order to determine if there are significant differences in sco
res amongst students with parents of differenteducational backgrounds.
```