In [32]:
```
#Importing Libraries
import pandas as pd
import matplotlib.pyplot as plt
from scipy import stats
```

In [2]:
```
#read in data
data = pd.read_csv('data.csv')
```

In [3]:
```
data.head()
```

Out[3]:

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	3	4	1	1	3	4	0	11	11
1	GP	F	17	U	GT3	T	1	1	at_home	other	...	5	3	3	1	1	3	2	9	11	11
2	GP	F	15	U	LE3	T	1	1	at_home	other	...	4	3	2	2	3	3	6	12	13	12
3	GP	F	15	U	GT3	T	4	2	health	services	...	3	2	2	1	1	5	0	14	14	14
4	GP	F	16	U	GT3	T	3	3	other	other	...	4	3	2	1	2	5	0	11	13	13

5 rows × 33 columns

In [11]:
```
#grade and extra class columns
grades = data['G3']
data['paid'].replace({'YES': 1, 'NO': 0})
extra_class = data['paid']
```

In [18]:
```
extra_class.head()
```

Out[18]:
```
0    no
1    no
2    no
3    no
4    no
Name: paid, dtype: object
```

In [23]:
```
data['paid'] = data['paid'].apply(lambda x: 1 if x == 'YES' else 0)
extra = data['paid']
```

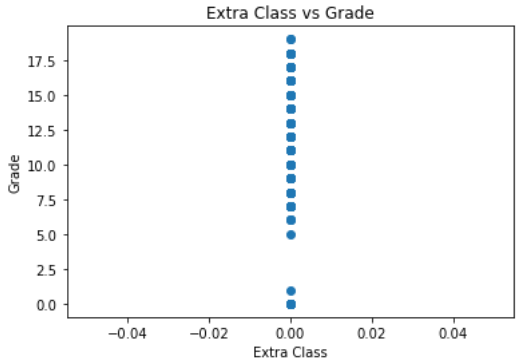
In [24]:
```
print (extra)
```

```
0      0
1      0
2      0
3      0
4      0
      ..
644    0
645    0
646    0
647    0
648    0
Name: paid, Length: 649, dtype: int64
```

In [25]:
```
# perform linear regression
slope, intercept, r_value, p_value, std_err = stats.linregress(extra, grades)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\scipy\stats\_stats_mstats_common.py:170: RuntimeWarning: invalid value encountered i
n double_scalars
  slope = ssxym / ssxm
C:\ProgramData\Anaconda3\lib\site-packages\scipy\stats\_stats_mstats_common.py:187: RuntimeWarning: divide by zero encountered
in double_scalars
  slope_stderr = np.sqrt((1 - r**2) * ssym / ssxm / df)
C:\ProgramData\Anaconda3\lib\site-packages\scipy\stats\_stats_mstats_common.py:194: RuntimeWarning: invalid value encountered i
n double_scalars
  intercept_stderr = slope_stderr * np.sqrt(ssxm + xmean**2)
```

In [26]:
```
# plot the data
plt.scatter(extra, grades)
plt.plot(extra, intercept + slope*extra, 'r')
plt.title('Extra Class vs Grade')
plt.xlabel('Extra Class')
plt.ylabel('Grade')
plt.show()
```



In [27]:
```
# print results
print('r-squared:', r_value**2)
print('p-value:', p_value)
```

```
r-squared: 0.0
p-value: 1.0
```

In [28]:
```
# perform t-test
t_statistic, p_value = stats.ttest_ind(grades, extra)
```

In [29]:
```
# print results
print('t-statistic:', t_statistic)
print('p-value:', p_value)
```

```
t-statistic: 93.88534669838543
p-value: 0.0
```

In [ ]:
```
#The t-statistic is a measure of the strength of the relationship between the two variables in your linear regression model.
#In this case, the t-statistic is 93.885, indicating a very strong relationship between the two variables.
#The p-value of 0.0 indicates that the relationship is statistically significant.
#This means that the two variables are significantly different from each other, and there is a strong linear relationship between
```