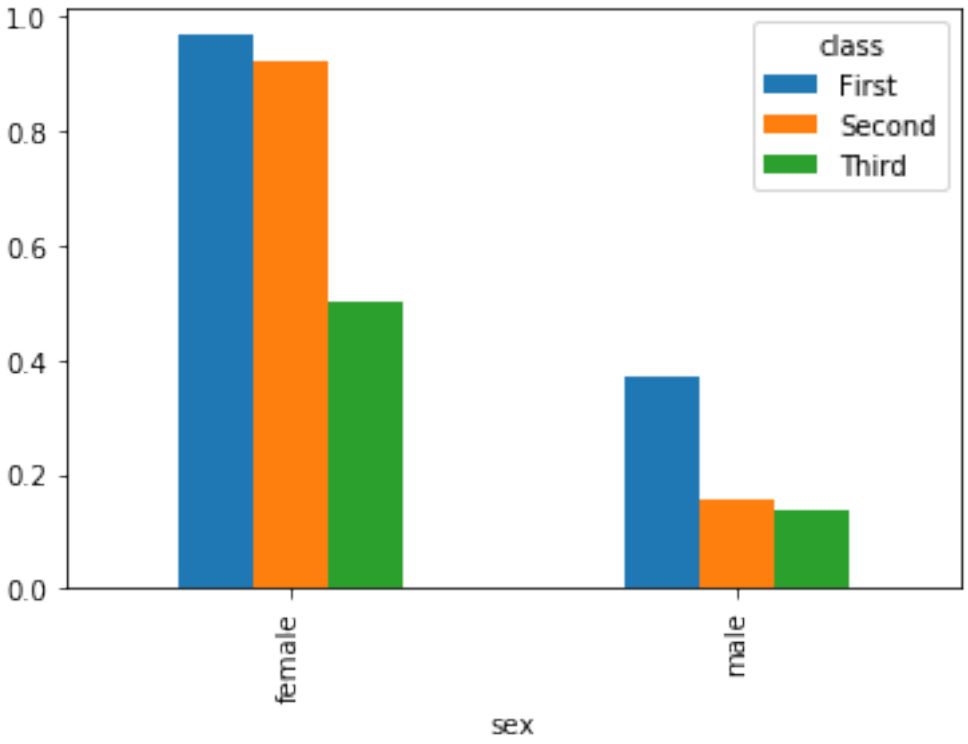
Assignment6

December 3, 2018

question 1

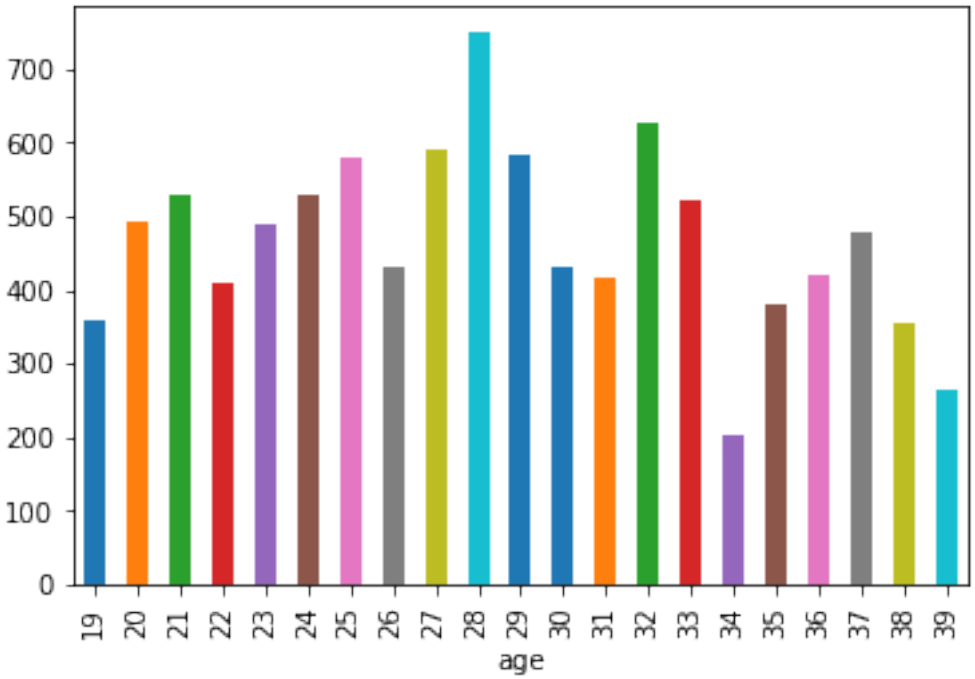
```
In [8]: import numpy as np
        import pandas as pd
        import seaborn as sns
        %matplotlib inline

        titanic = sns.load_dataset('titanic')
        titanic= titanic.pivot_table('survived',index='sex',columns='class')
        titanic.plot.bar()
```

```
Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x7f5cf1f831d0>
```



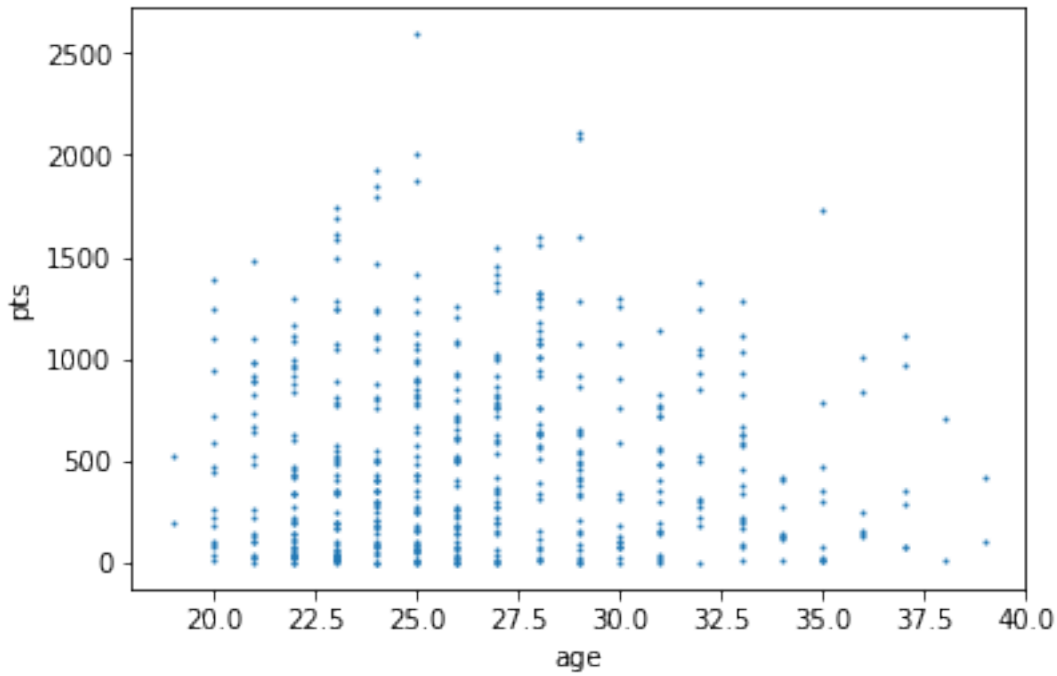
question 2

```
In [15]: nba = pd.read_csv("/home/lubo/data301/nba_2013.csv")
         nba= nba.groupby('age')['pts'].mean()
         nba.plot.bar()

Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x7f5cf1c73da0>
```



question 3

```
In [22]: nba = pd.read_csv("/home/lubo/data301/nba_2013.csv")
         nba.plot.scatter(x="age", y="pts",s=1)

Out[22]: <matplotlib.axes._subplots.AxesSubplot at 0x7f5cf1ac9160>
```



the first(bar) graph is the better representation of the data because it gives you a better ideas of how the data is distributed over the graph as a whole. That is because each "group" is aggregated. That being said, the second graph does a better job of showing how the vertices are distributed within each group.