# Basics

July 23, 2019

## 0.1 Employment and Salary of Graduates between 2010-12 by Major (USA)

In [18]:
```
import pandas as pd
import matplotlib
import numpy as np

%matplotlib inline
```

In [19]:
```
recent_grads = pd.read_csv('recent-grads.csv')
recent_grads.iloc[0,:]
```

Out[19]:
```
Rank                                    1
Major_code                           2419
Major                 PETROLEUM ENGINEERING
Total                                2339
Men                                  2057
Women                                 282
Major_category                Engineering
ShareWomen                       0.120564
Sample_size                            36
Employed                             1976
Full_time                            1849
Part_time                             270
Full_time_year_round                 1207
Unemployed                             37
Unemployment_rate                0.0183805
Median                              110000
P25th                                95000
P75th                               125000
College_jobs                         1534
Non_college_jobs                      364
Low_wage_jobs                         193
Name: 0, dtype: object
```

In [20]:
```
recent_grads.head()
```

Out[20]:

| | Rank | Major_code | Major | Total | \ |
|---|---|---|---|---|---|
| 0 | 1 | 2419 | PETROLEUM ENGINEERING | 2339.0 | |
| 1 | 2 | 2416 | MINING AND MINERAL ENGINEERING | 756.0 | |

```
2     3        2415                    METALLURGICAL ENGINEERING    856.0
3     4        2417  NAVAL ARCHITECTURE AND MARINE ENGINEERING   1258.0
4     5        2405                        CHEMICAL ENGINEERING  32260.0

       Men    Women Major_category  ShareWomen  Sample_size  Employed  \
0   2057.0    282.0    Engineering    0.120564           36      1976
1    679.0     77.0    Engineering    0.101852            7       640
2    725.0    131.0    Engineering    0.153037            3       648
3   1123.0    135.0    Engineering    0.107313           16       758
4  21239.0  11021.0    Engineering    0.341631          289     25694

       ...        Part_time  Full_time_year_round  Unemployed  \
0      ...              270                  1207          37
1      ...              170                   388          85
2      ...              133                   340          16
3      ...              150                   692          40
4      ...             5180                 16697        1672

   Unemployment_rate  Median  P25th   P75th  College_jobs  Non_college_jobs  \
0           0.018381  110000  95000  125000          1534               364
1           0.117241   75000  55000   90000           350               257
2           0.024096   73000  50000  105000           456               176
3           0.050125   70000  43000   80000           529               102
4           0.061098   65000  50000   75000         18314              4440

   Low_wage_jobs
0            193
1             50
2              0
3              0
4            972

[5 rows x 21 columns]
```

```
In [21]: recent_grads.tail()
```

```
Out[21]:      Rank  Major_code                  Major   Total     Men   Women  \
168   169        3609                 ZOOLOGY  8409.0  3050.0  5359.0
169   170        5201  EDUCATIONAL PSYCHOLOGY  2854.0   522.0  2332.0
170   171        5202     CLINICAL PSYCHOLOGY  2838.0   568.0  2270.0
171   172        5203   COUNSELING PSYCHOLOGY  4626.0   931.0  3695.0
172   173        3501         LIBRARY SCIENCE  1098.0   134.0   964.0

                Major_category  ShareWomen  Sample_size  Employed  \
168    Biology & Life Science    0.637293           47      6259
169  Psychology & Social Work    0.817099            7      2125
170  Psychology & Social Work    0.799859           13      2101
171  Psychology & Social Work    0.798746           21      3777
```

```
172                Education    0.877960           2       742

         ...        Part_time  Full_time_year_round  Unemployed  \
168      ...             2190                  3602         304
169      ...              572                  1211         148
170      ...              648                  1293         368
171      ...              965                  2738         214
172      ...              237                   410          87

     Unemployment_rate  Median  P25th  P75th  College_jobs  Non_college_jobs  \
168           0.046320   26000  20000  39000          2771              2947
169           0.065112   25000  24000  34000          1488               615
170           0.149048   25000  25000  40000           986               870
171           0.053621   23400  19200  26000          2403              1245
172           0.104946   22000  20000  22000           288               338

     Low_wage_jobs
168            743
169             82
170            622
171            308
172            192

[5 rows x 21 columns]
```

In [22]:
```
recent_grads.describe(include=np.number)
```

Out[22]:

| | Rank | Major_code | Total | Men | Women |
|---|---|---|---|---|---|
| count | 173.000000 | 173.000000 | 172.000000 | 172.000000 | 172.000000 |
| mean | 87.000000 | 3879.815029 | 39370.081395 | 16723.406977 | 22646.674419 |
| std | 50.084928 | 1687.753140 | 63483.491009 | 28122.433474 | 41057.330740 |
| min | 1.000000 | 1100.000000 | 124.000000 | 119.000000 | 0.000000 |
| 25% | 44.000000 | 2403.000000 | 4549.750000 | 2177.500000 | 1778.250000 |
| 50% | 87.000000 | 3608.000000 | 15104.000000 | 5434.000000 | 8386.500000 |
| 75% | 130.000000 | 5503.000000 | 38909.750000 | 14631.000000 | 22553.750000 |
| max | 173.000000 | 6403.000000 | 393735.000000 | 173809.000000 | 307087.000000 |

| | ShareWomen | Sample_size | Employed | Full_time | Part_time |
|---|---|---|---|---|---|
| count | 172.000000 | 173.000000 | 173.000000 | 173.000000 | 173.000000 |
| mean | 0.522223 | 356.080925 | 31192.763006 | 26029.306358 | 8832.398844 |
| std | 0.231205 | 618.361022 | 50675.002241 | 42869.655092 | 14648.179473 |
| min | 0.000000 | 2.000000 | 0.000000 | 111.000000 | 0.000000 |
| 25% | 0.336026 | 39.000000 | 3608.000000 | 3154.000000 | 1030.000000 |
| 50% | 0.534024 | 130.000000 | 11797.000000 | 10048.000000 | 3299.000000 |
| 75% | 0.703299 | 338.000000 | 31433.000000 | 25147.000000 | 9948.000000 |
| max | 0.968954 | 4212.000000 | 307933.000000 | 251540.000000 | 115172.000000 |

| | Full_time_year_round | Unemployed | Unemployment_rate | Median |
|---|---|---|---|---|

```
count            173.000000    173.000000         173.000000     173.000000
mean           19694.427746   2416.329480           0.068191   40151.445087
std            33160.941514   4112.803148           0.030331   11470.181802
min              111.000000      0.000000           0.000000   22000.000000
25%             2453.000000    304.000000           0.050306   33000.000000
50%             7413.000000    893.000000           0.067961   36000.000000
75%            16891.000000   2393.000000           0.087557   45000.000000
max           199897.000000  28169.000000           0.177226  110000.000000

                P25th          P75th   College_jobs  Non_college_jobs  \
count      173.000000     173.000000     173.000000        173.000000
mean     29501.445087   51494.219653   12322.635838      13284.497110
std       9166.005235   14906.279740   21299.868863      23789.655363
min      18500.000000   22000.000000       0.000000          0.000000
25%      24000.000000   42000.000000    1675.000000       1591.000000
50%      27000.000000   47000.000000    4390.000000       4595.000000
75%      33000.000000   60000.000000   14444.000000      11783.000000
max      95000.000000  125000.000000  151643.000000     148395.000000

       Low_wage_jobs
count     173.000000
mean     3859.017341
std      6944.998579
min         0.000000
25%       340.000000
50%      1231.000000
75%      3466.000000
max     48207.000000
```

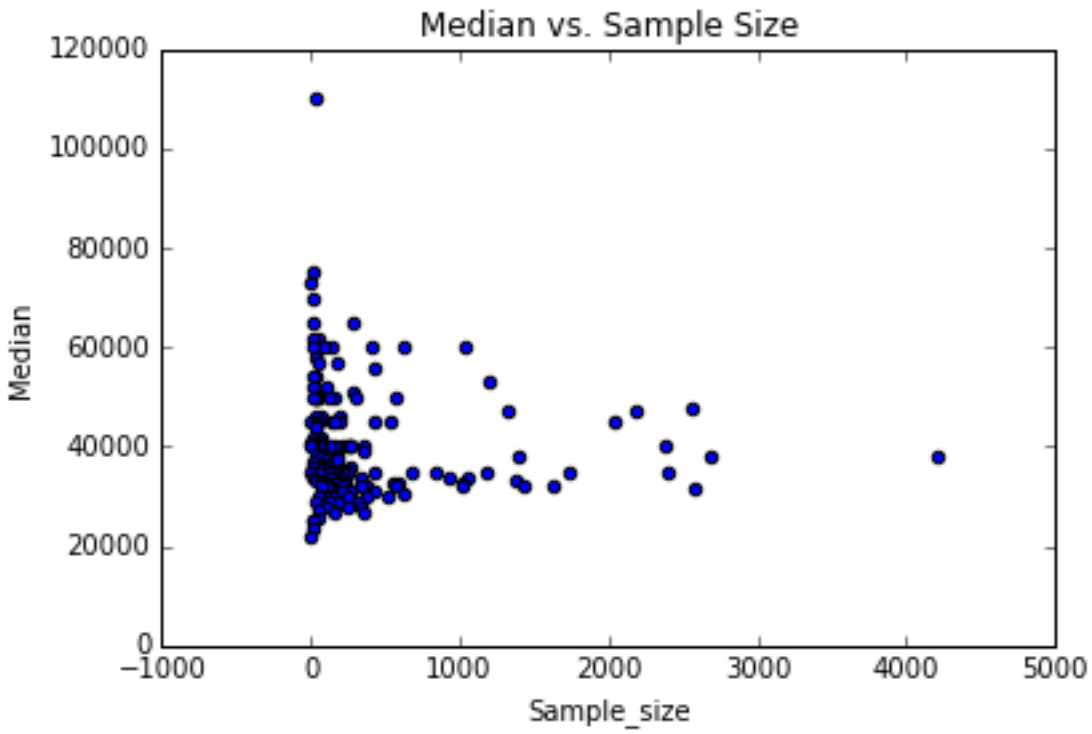
In [23]:
```
raw_data_count = len(recent_grads.index) + 1
print(raw_data_count)
recent_grads = recent_grads.dropna(axis=0, how='any')
```

```
174
```

In [24]:
```
cleaned_data_count = len(recent_grads.index) + 1
```

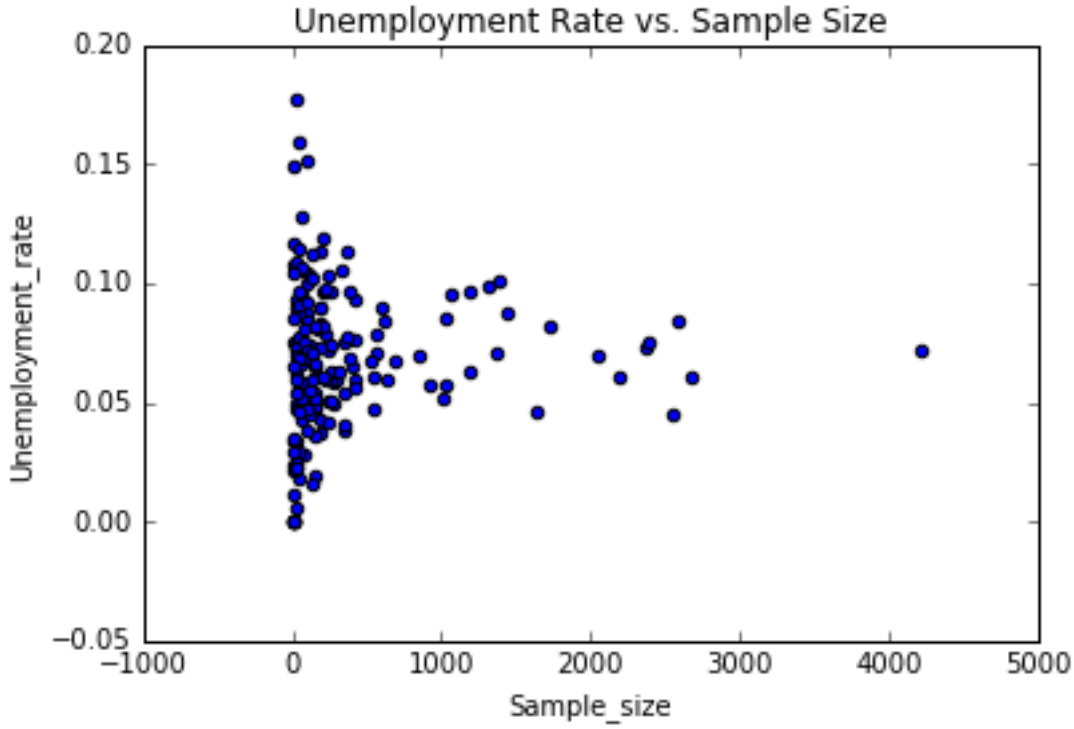
In [25]:
```
recent_grads.plot(x='Sample_size', y='Median', kind='scatter', title= 'Median vs. Samp
```

Out[25]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f452180fc88>
```



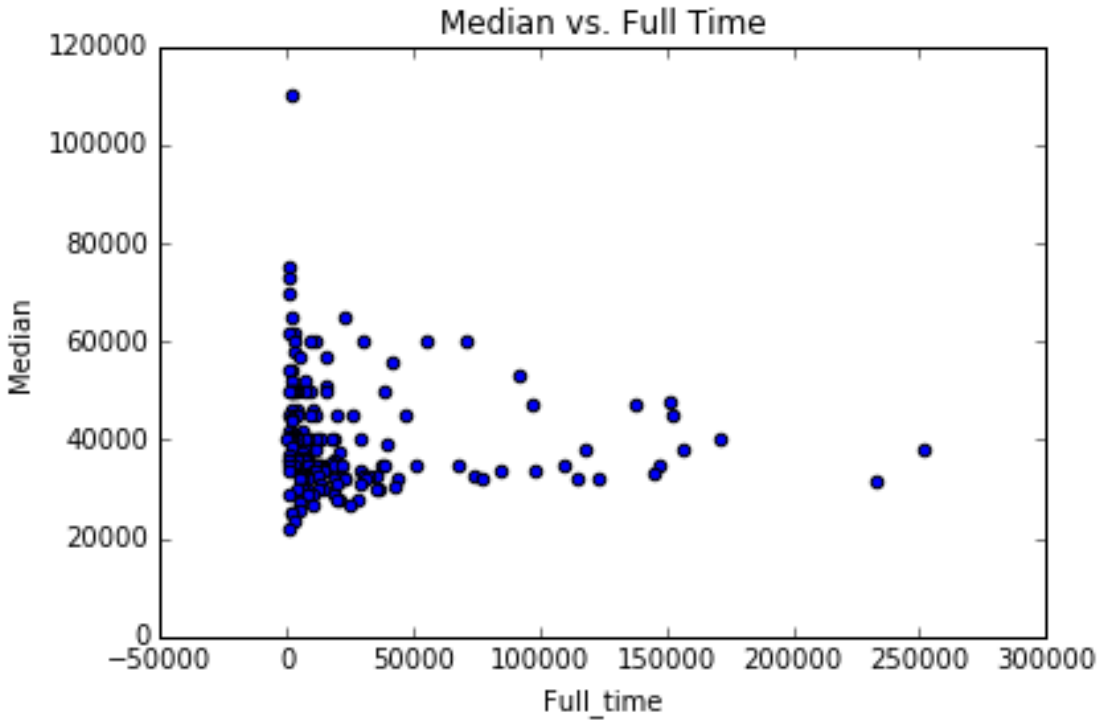

In [26]: recent_grads.plot(x='Sample_size', y='Unemployment_rate', kind='scatter', title='Unemp

Out[26]: <matplotlib.axes._subplots.AxesSubplot at 0x7f45217a07b8>



```
In [27]: recent_grads.plot(x='Full_time', y = 'Median', kind='scatter', title='Median vs. Full
```

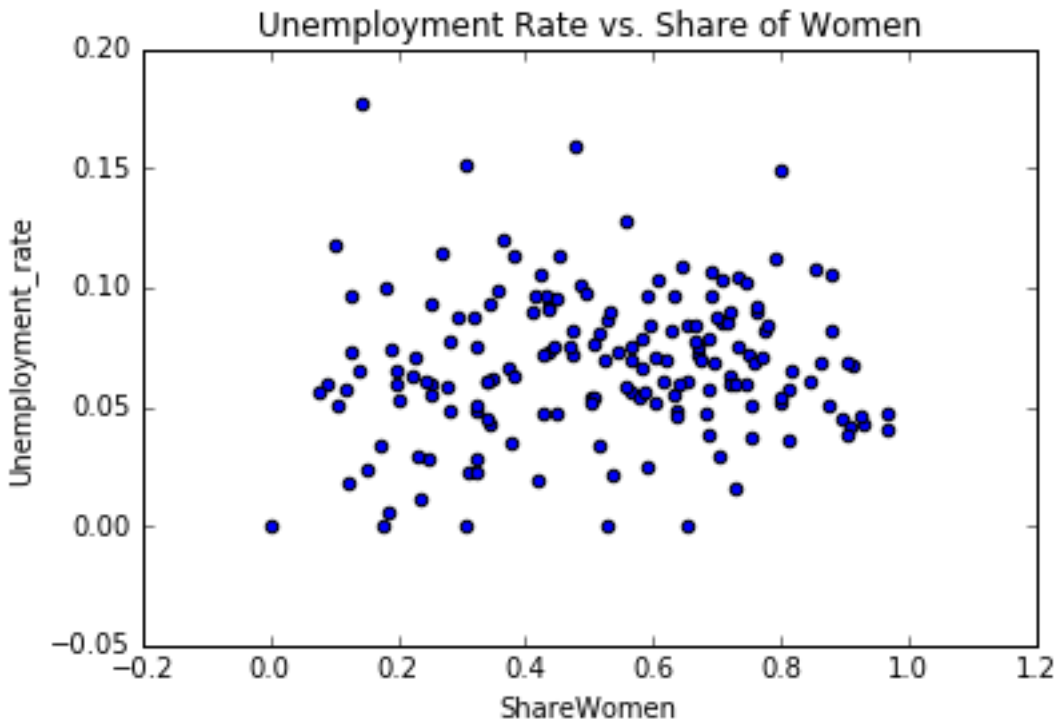
```
Out[27]: <matplotlib.axes._subplots.AxesSubplot at 0x7f45218275c0>
```




**Median vs. Full time** Higher numbers of full time employees tend be in lower salaried roles. While there is clustering between 40k and 80k these roles have fewer people in them while there is a long tail around 30k with high numbers of employees. >This suggests that popular majors tend to earn less.

```
In [28]: recent_grads.plot(x='ShareWomen', y='Unemployment_rate', kind='scatter', title='Unempl
```

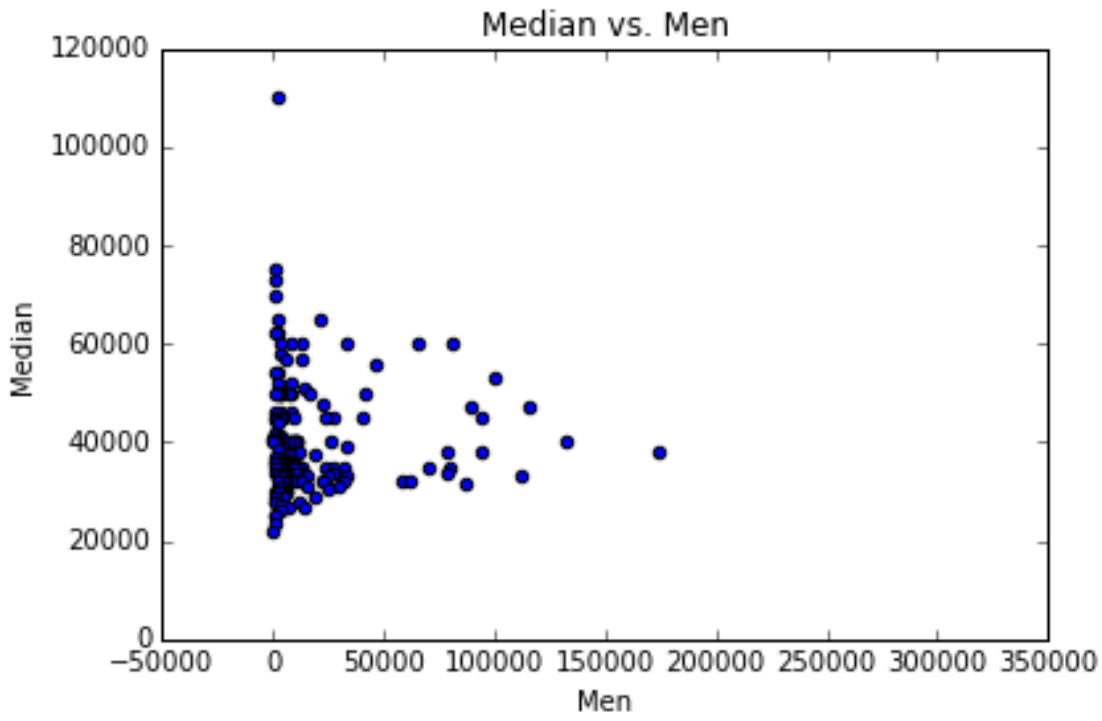
```
Out[28]: <matplotlib.axes._subplots.AxesSubplot at 0x7f45217489b0>
```




**Unemployment Rate vs. Share of Women** Possible weak connection between unemployment and share of women. Needs further analysis

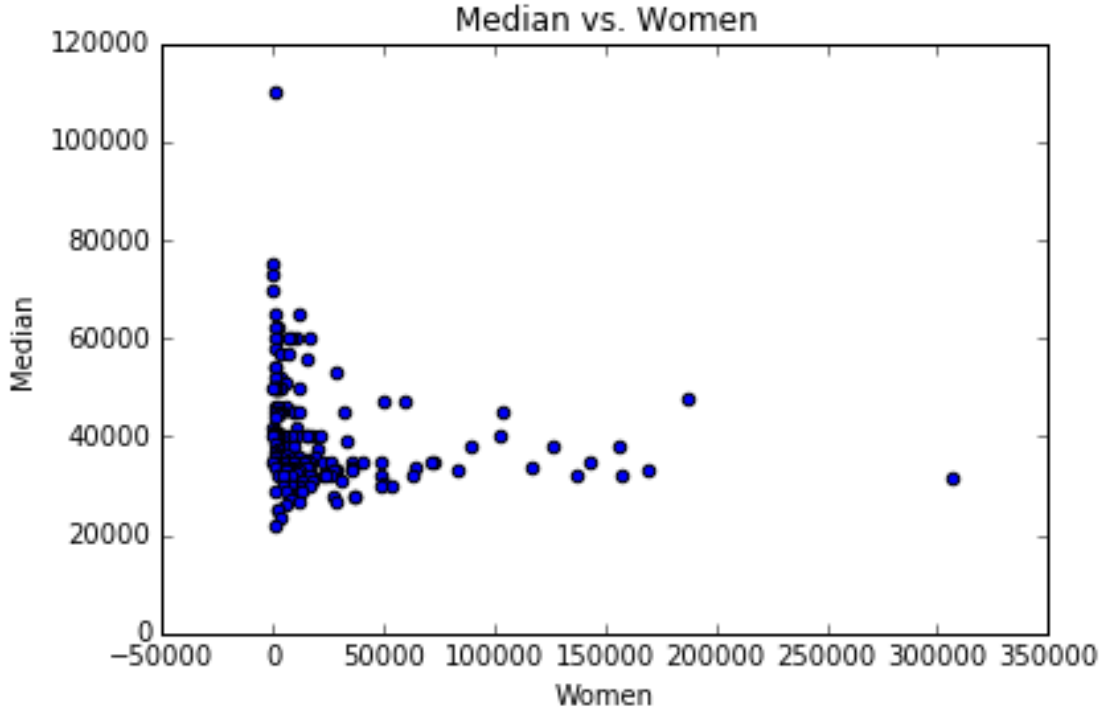
```
In [31]: ax1 = recent_grads.plot(x='Men', y='Median', kind='scatter', title='Median vs. Men')
         ax1.set(xlim=(-50000, 350000))
```

```
Out[31]: [(-50000, 350000)]
```




```
In [30]: ax2 = recent_grads.plot(x='Women', y='Median', kind='scatter', title='Median vs. Women
```

```
Out[30]: <matplotlib.axes._subplots.AxesSubplot at 0x7f45216916d8>
```



**Do students that majored in subjects that were majority female make more money?** The clustering in the less popular majors suggests no difference. However, majors that are very popular with females earn less. While it holds that popular majors earn less, it is clear that very high numbers of women graduated with majors that earn less.

In [ ]: