Ch07_Data Visualization

September 5, 2018

1 Chapter 7: Data Visualization

```
In [3]: import pandas as pd

        dataset = pd.read_csv("./Data/Salaries.csv")

        rank = dataset['rank']
        discipline =  dataset['discipline']
        phd =  dataset['phd']
        service =  dataset['service']
        sex =  dataset['sex']
        salary =  dataset['salary']

        dataset.head()

Out[3]:    rank discipline  phd  service   sex  salary
        0  Prof          B   56       49  Male  186960
        1  Prof          A   12        6  Male   93000
        2  Prof          A   23       20  Male  110515
        3  Prof          A   40       31  Male  131205
        4  Prof          B   20       18  Male  104800
```
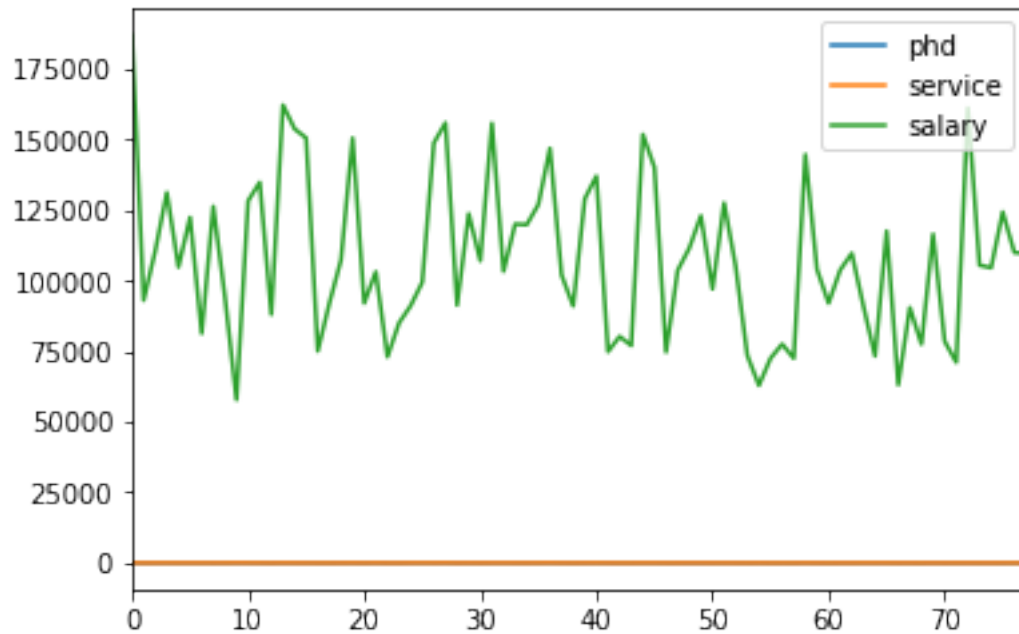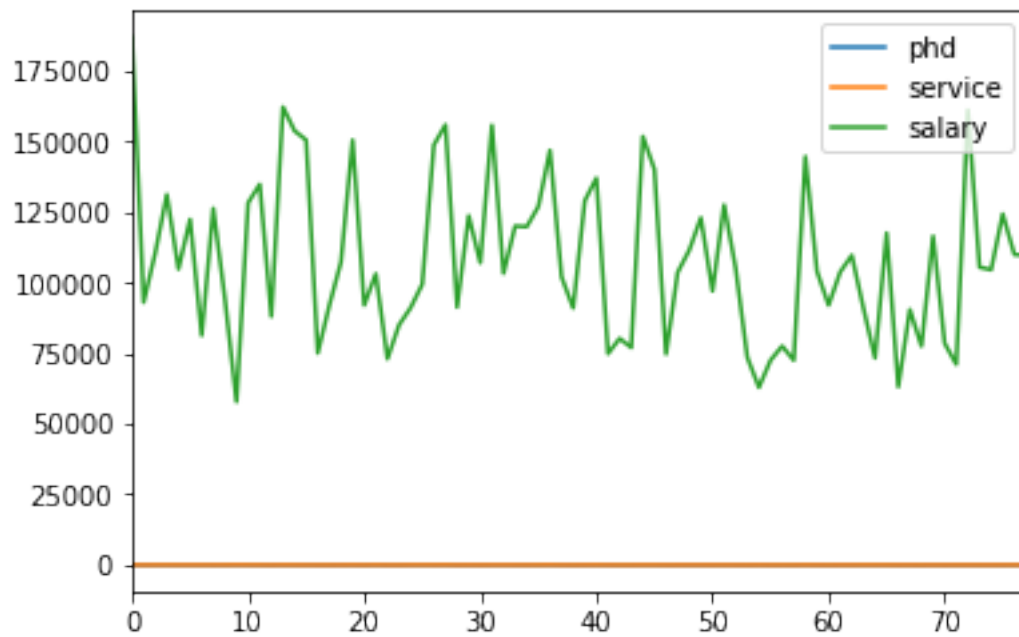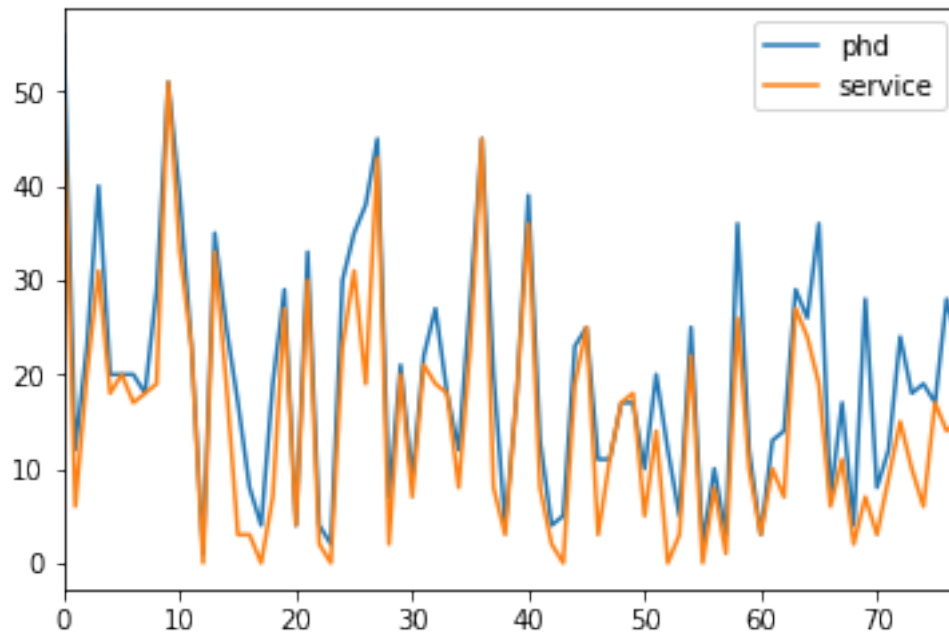
2 Line plotting

```
In [4]: dataset.plot()

Out[4]: <matplotlib.axes._subplots.AxesSubplot at 0x7f151e35a630>
```

In [5]: dataset[["rank", "discipline","phd","service", "sex", "salary"]].plot()

Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0x7f151e274240>

```
In [6]: dataset[["phd","service"]].plot()
```

```
Out[6]: <matplotlib.axes._subplots.AxesSubplot at 0x7f151dff5390>
```


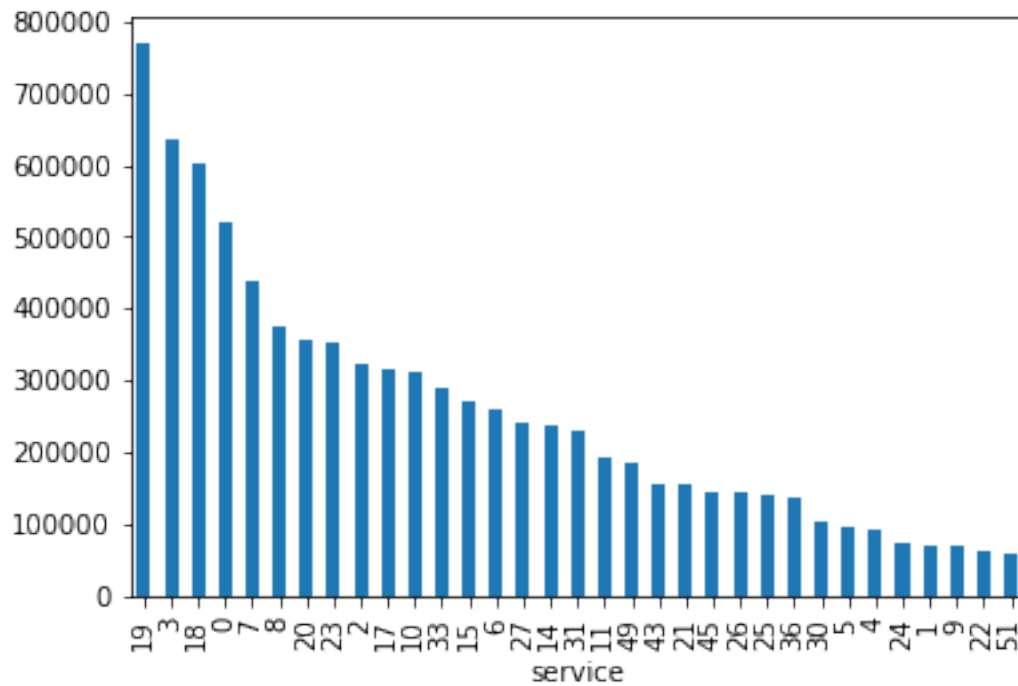# 3   Visualize grouped data

```
In [7]: dataset1 = dataset.groupby(['service']).sum()
        dataset1.sort_values("salary", ascending = False, inplace=True)
        dataset1.head()
```

```
Out[7]:          phd   salary
        service
        19       178   769448
        3         56   635216
        18        91   603060
        0         26   519500
        7         70   440408
```
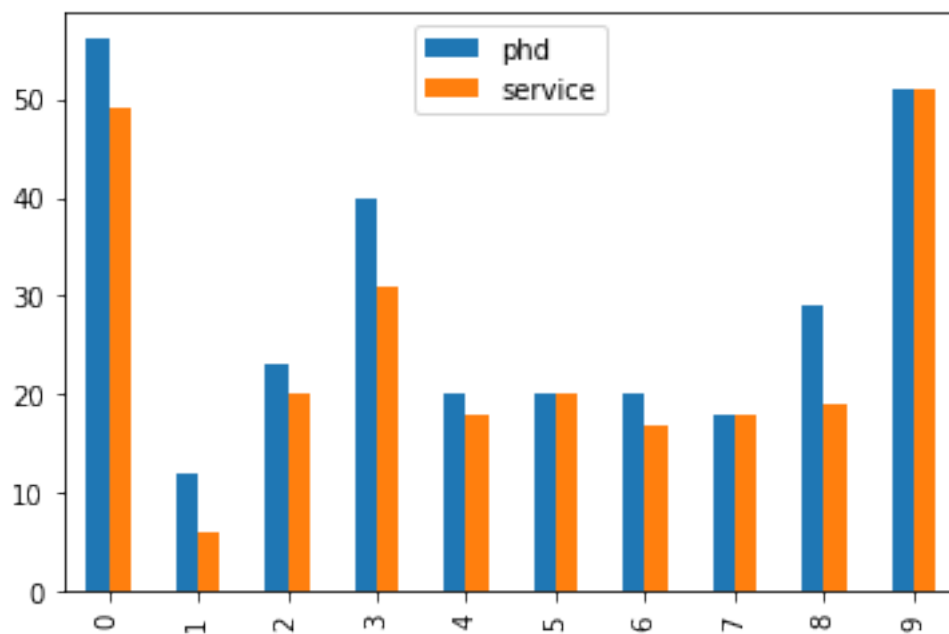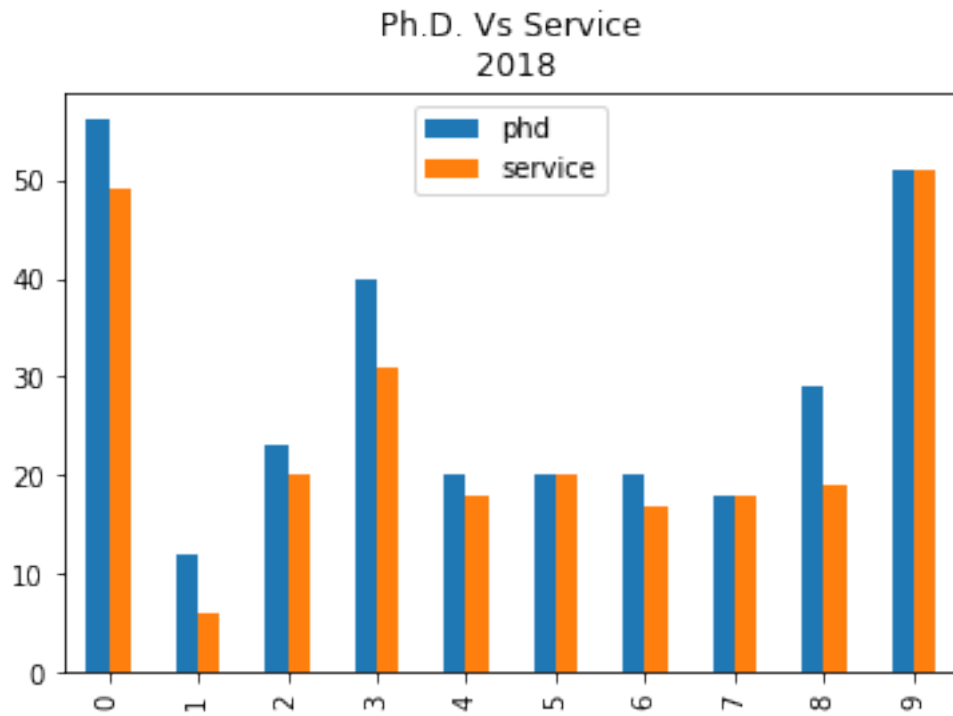
```
In [8]: dataset1["salary"].plot.bar()
```

```
Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x7f151df3c710>
```

4 Bar plotting

```
In [9]: dataset[[ 'phd', 'service' ]].head(10).plot.bar()

Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x7f151db8aef0>
```
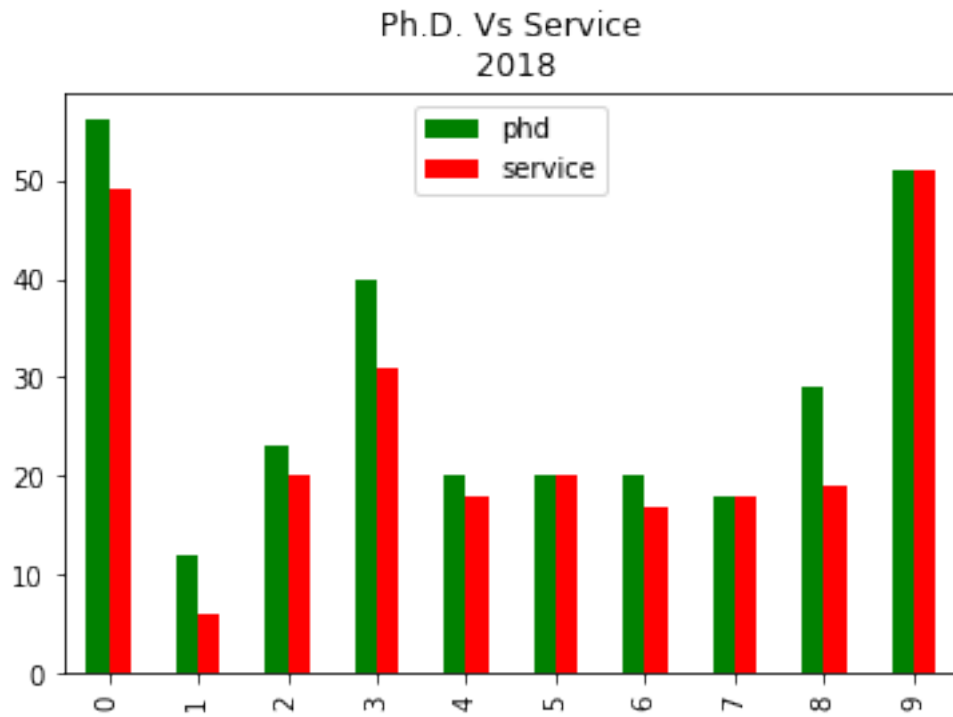
```
In [10]: dataset[['phd', 'service']].head(10).plot.bar(title="Ph.D. Vs Service\n 2018")
```

```
Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x7f151db1be80>
```



```
In [11]: dataset[['phd', 'service']].head(10).plot.bar(title="Ph.D. Vs Service\n 2018" , color=[
```
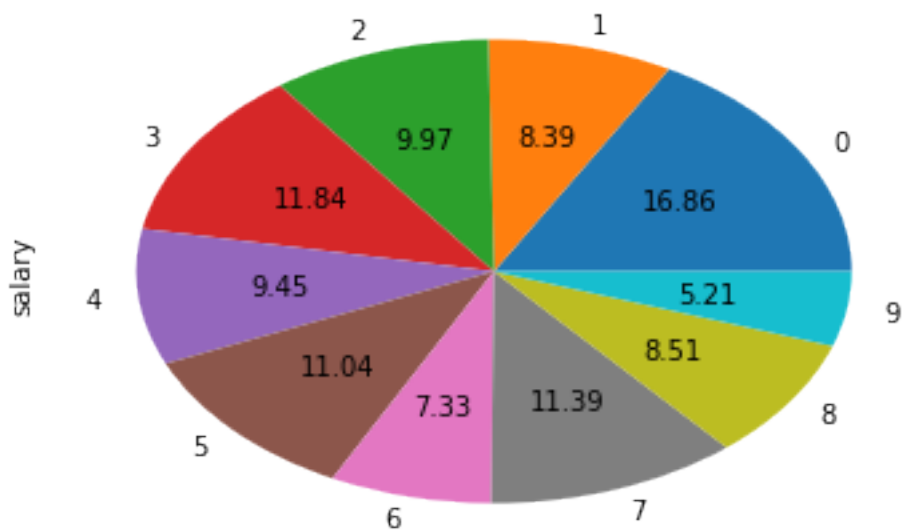
```
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x7f151dbae828>
```

Ph.D. Vs Service
2018

5 Pie Chart

```
In [12]: dataset["salary"].head(10).plot.pie(autopct='%.2f')

Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x7f151d7fefd0>
```
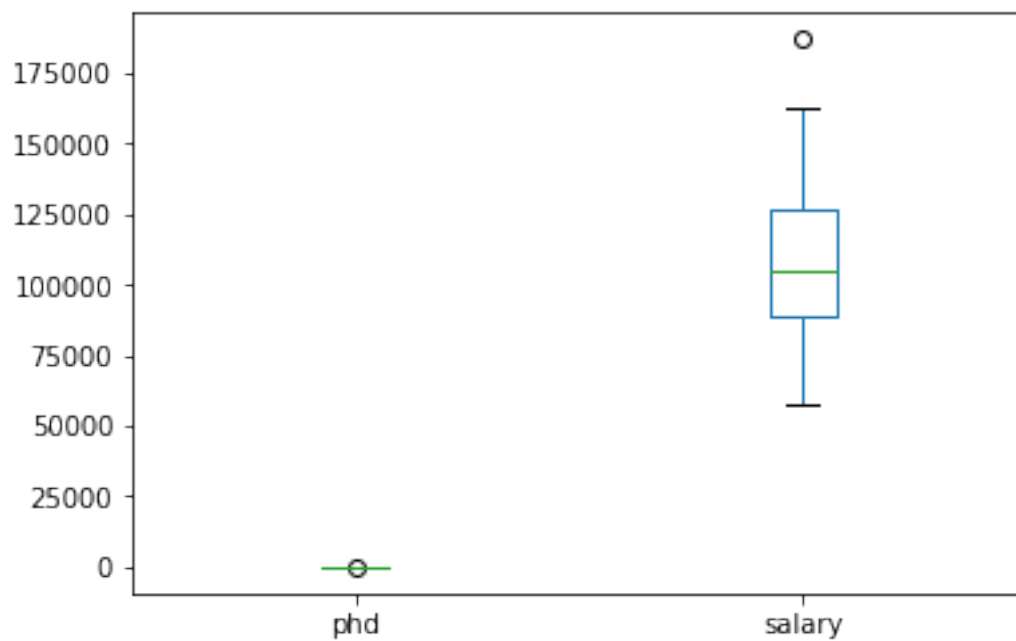
6 Box Plotting

```
In [13]: dataset[["phd","salary"]].head(100).plot.box()
```
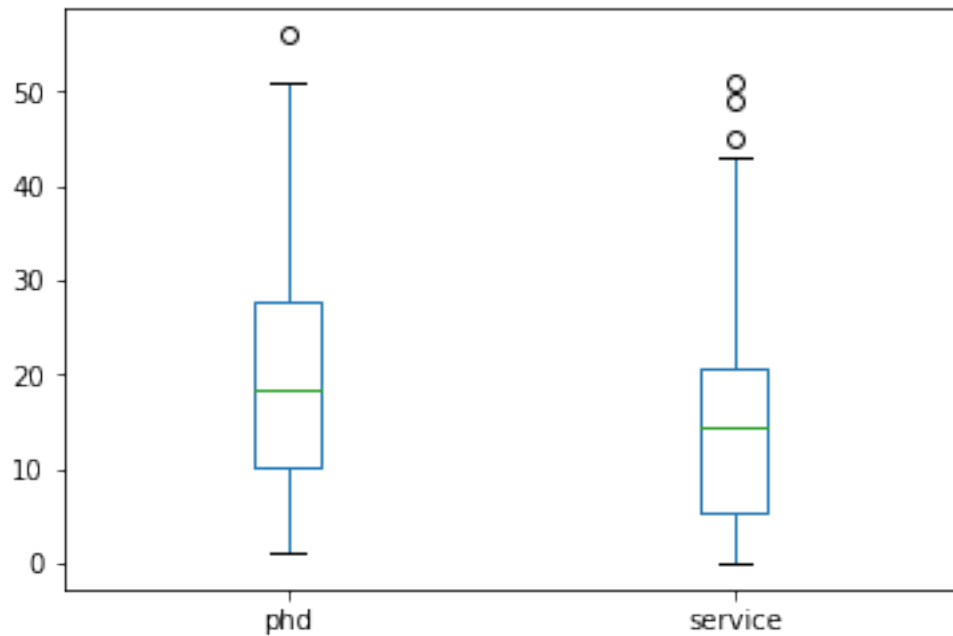
```
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x7f151d682b70>
```



```
In [14]: dataset[["phd","service"]].plot.box()
```

```
Out[14]: <matplotlib.axes._subplots.AxesSubplot at 0x7f151d3b21d0>
```
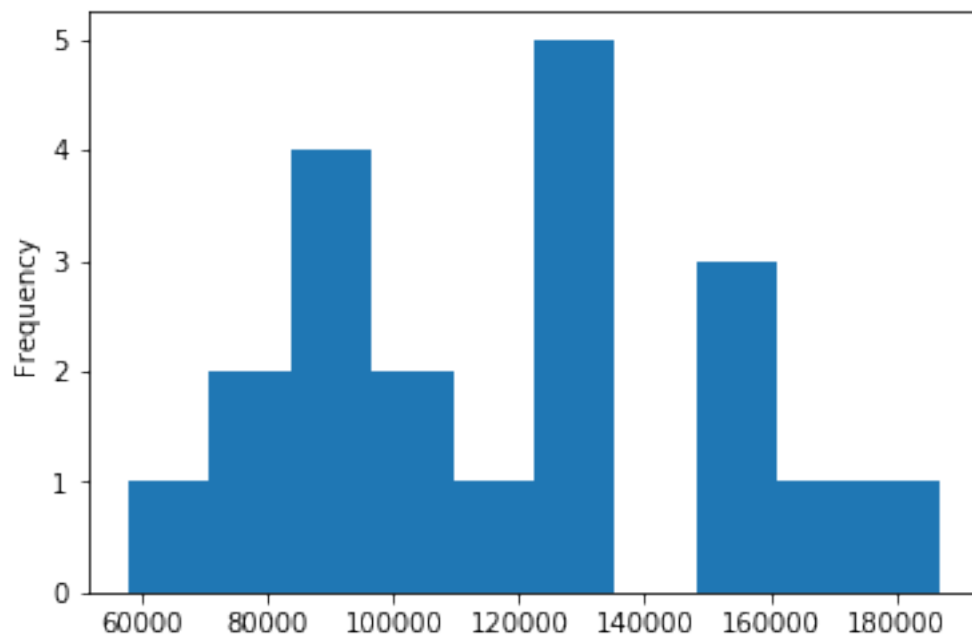
7 Histogram

```
In [15]: dataset["salary"].head(20).plot.hist()
```
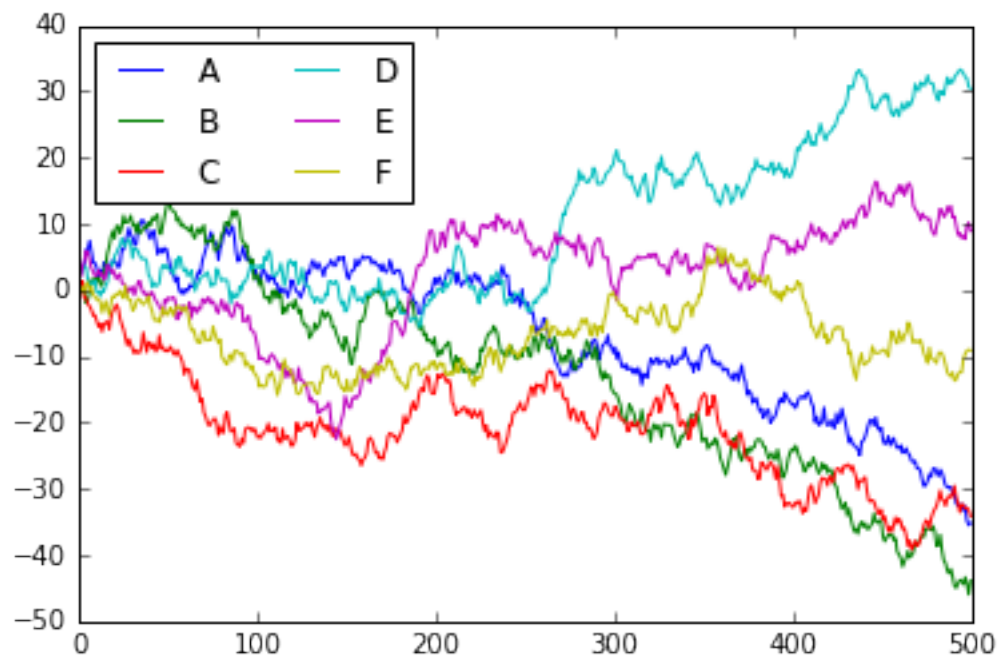
```
Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x7f151d32a7f0>
```

```
In [ ]: # A scatterplot

In [3]: # Exercises

In [4]: import matplotlib.pyplot as plt
        plt.style.use('classic')
        %matplotlib inline
        import numpy as np
        import pandas as pd

In [30]: # Create temprature data
         rng = np.random.RandomState(0)
         season1 = np.cumsum(rng.randn(500, 6), 0)

In [32]: # Plot the data with Matplotlib defaults
         plt.plot(season1)
         plt.legend('ABCDEF', ncol=2, loc='upper left');
```
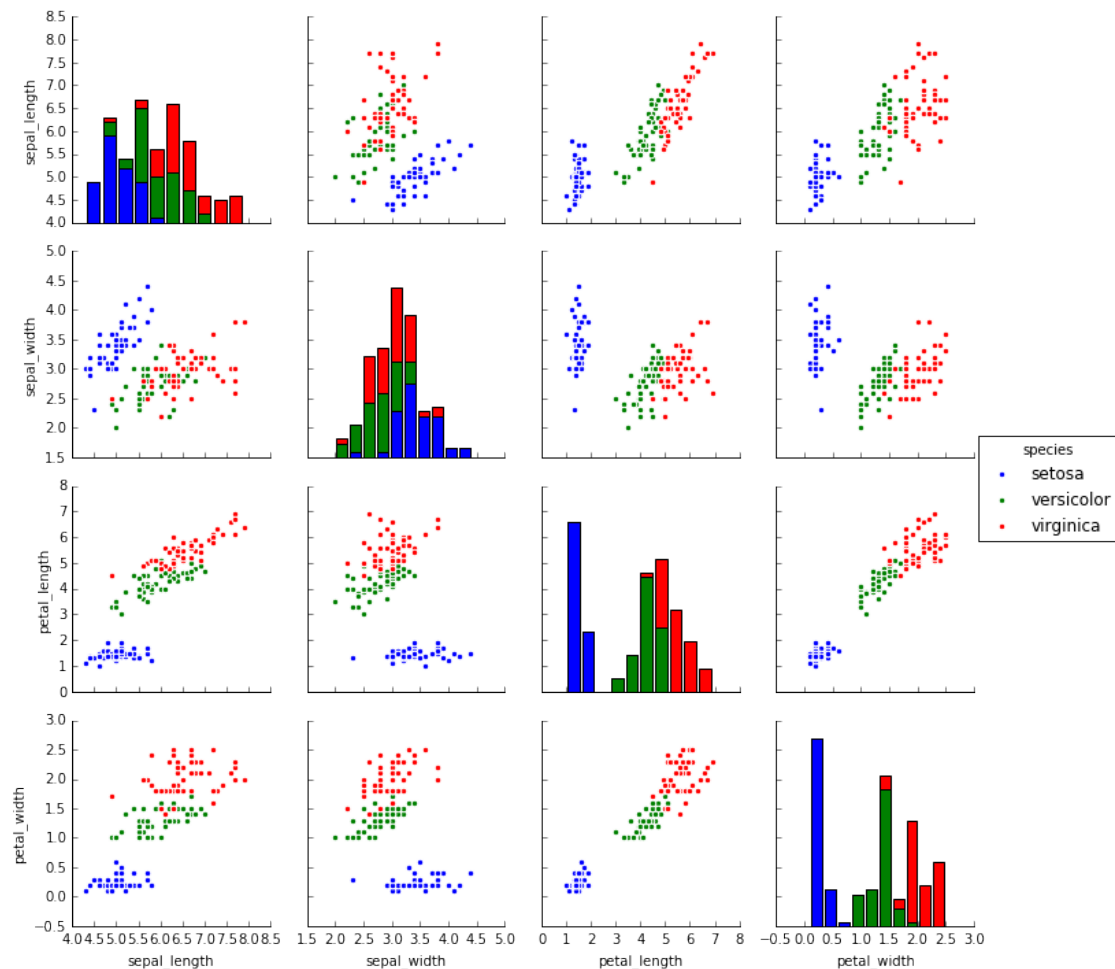


```
In [33]: import seaborn as sns
         iris = sns.load_dataset("iris")
         iris.head()
         sns.pairplot(iris, hue='species', size=2.5);
```

9

In [36]:
```python
import seaborn as sns
tips = sns.load_dataset('tips')
tips.head()
```

Out[36]:
```
   total_bill   tip     sex smoker  day    time  size
0       16.99  1.01  Female     No  Sun  Dinner     2
1       10.34  1.66    Male     No  Sun  Dinner     3
2       21.01  3.50    Male     No  Sun  Dinner     3
3       23.68  3.31    Male     No  Sun  Dinner     2
4       24.59  3.61  Female     No  Sun  Dinner     4
```

In [37]:
```python
tips['Tips Percentage'] = 100 * tips['tip'] / tips['total_bill']
grid = sns.FacetGrid(tips, row="sex", col="time", margin_titles=True)
grid.map(plt.hist, "Tips Percentage", bins=np.linspace(0, 40, 15));
```
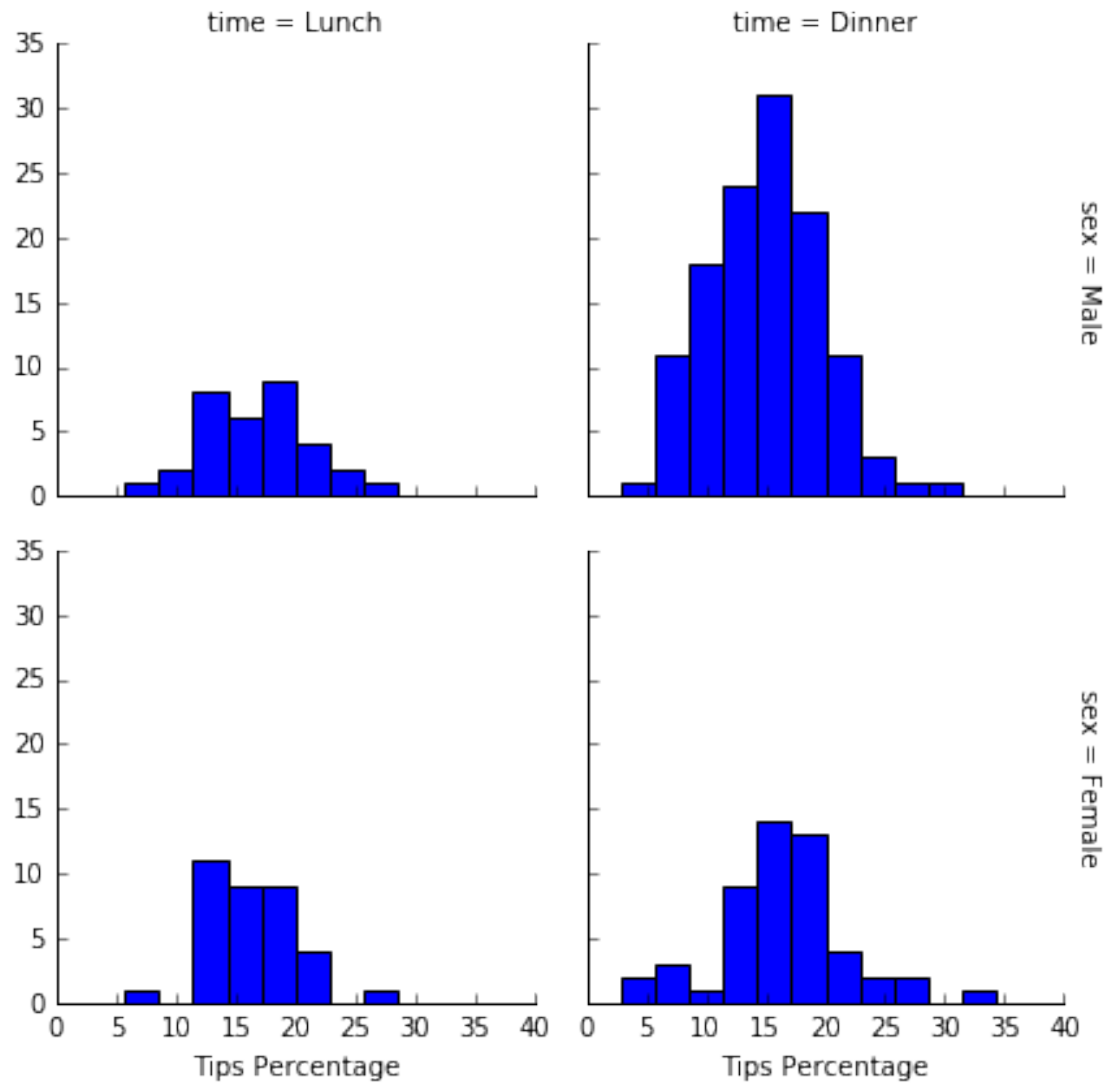
```
In [39]: import seaborn as sns
         tips = sns.load_dataset('tips')
         with sns.axes_style(style='ticks'):
             g = sns.factorplot("day", "total_bill", "sex", data=tips, kind="box")
             g.set_axis_labels("Bill Day", "Total Bill Amount");
```
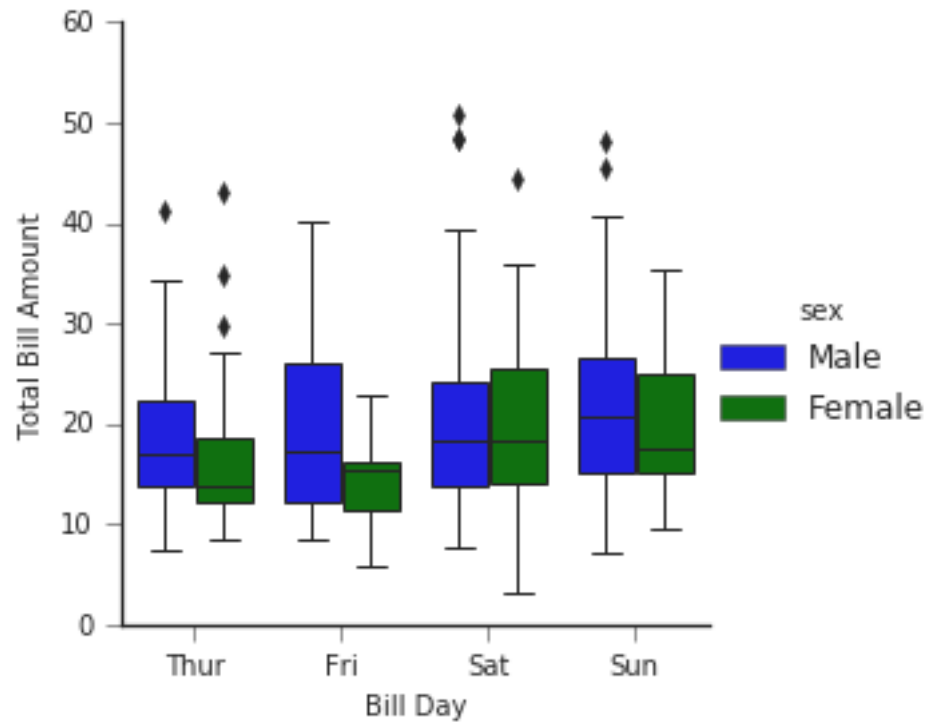
```
In [43]: import seaborn as sns
         tips = sns.load_dataset('tips')
         with sns.axes_style('white'):
             sns.jointplot( "total_bill", "tip", data=tips, kind='hex')
```
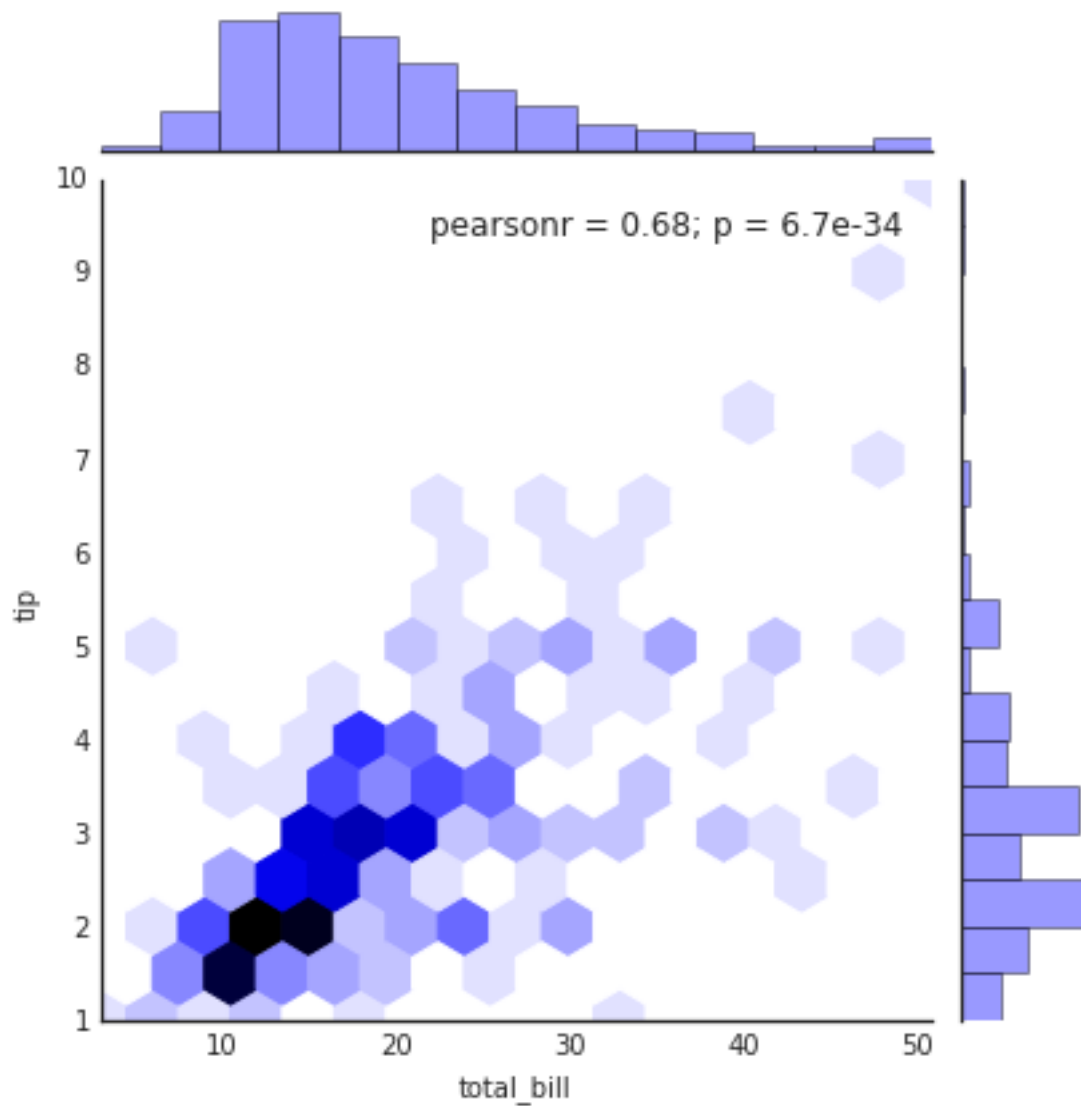
```
In [25]: import seaborn as sns
         planets = sns.load_dataset('planets')
         planets.head()

Out[25]:          method  number  orbital_period   mass  distance  year
         0  Radial Velocity       1         269.300   7.10     77.40  2006
         1  Radial Velocity       1         874.774   2.21     56.95  2008
         2  Radial Velocity       1         763.000   2.60     19.84  2011
         3  Radial Velocity       1         326.030  19.40    110.62  2007
         4  Radial Velocity       1         516.220  10.50    119.47  2009
```