In [9]:
```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
import seaborn as sns
```

In [10]:
```
data = pd.read_csv("Salaries.csv")
```

In [11]:
```
data.head()
```

Out[11]:

	rank	discipline	yrs.since.phd	yrs.service	sex	salary
0	Prof	B	19	18	Male	139750
1	Prof	B	20	16	Male	173200
2	AsstProf	B	4	3	Male	79750
3	Prof	B	45	39	Male	115000
4	Prof	B	40	41	Male	141500

In [12]:
```
data.columns
```

Out[12]:
```
Index(['rank', 'discipline', 'yrs.since.phd', 'yrs.service', 'sex', 'salary'], dtype='object')
```

In [13]:
```
data.describe()
```

Out[13]:

	yrs.since.phd	yrs.service	salary
count	397.000000	397.000000	397.000000
mean	22.314861	17.614610	113706.458438
std	12.887003	13.006024	30289.038695
min	1.000000	0.000000	57800.000000
25%	12.000000	7.000000	91000.000000
50%	21.000000	16.000000	107300.000000
75%	32.000000	27.000000	134185.000000
max	56.000000	60.000000	231545.000000

In [14]:
```
data.isnull()
```

Out[14]:

	rank	discipline	yrs.since.phd	yrs.service	sex	salary
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...	...	...	...	...	...	...
392	False	False	False	False	False	False
393	False	False	False	False	False	False
394	False	False	False	False	False	False
395	False	False	False	False	False	False
396	False	False	False	False	False	False

397 rows × 6 columns

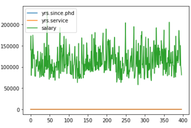
In [15]:
```
data.head()
```

Out[15]:

	rank	discipline	yrs.since.phd	yrs.service	sex	salary
0	Prof	B	19	18	Male	139750
1	Prof	B	20	16	Male	173200
2	AsstProf	B	4	3	Male	79750
3	Prof	B	45	39	Male	115000
4	Prof	B	40	41	Male	141500

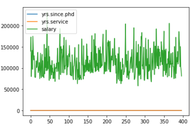
In [27]:
```
data.plot()
```

Out[27]:
```
<AxesSubplot:>
```



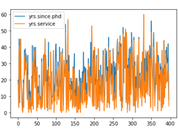
In [32]:
```
data[["rank","discipline","yrs.since.phd","yrs.service","sex","salary"]].plot()
```

Out[32]:
```
<AxesSubplot:>
```



In [33]:
```
data[["yrs.since.phd","yrs.service"]].plot()
```

Out[33]:
```
<AxesSubplot:>
```



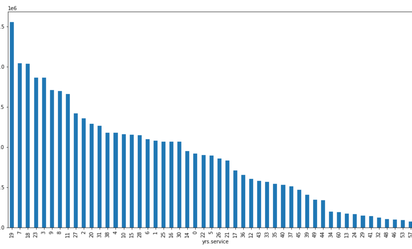
In [37]:
```
data1 = data.groupby(["yrs.service"]).sum()
data1.sort_values("salary",ascending = False,inplace=True)
data1.head()
```

Out[37]:

	yrs.since.phd	salary
yrs.service		
19	518	2554779
7	285	2044623
18	327	2037195
23	443	1865521
3	145	1861760

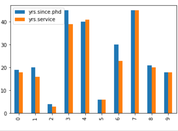
In [43]:
```
plt.figure(figsize=(15,8))
data["salary"].plot.bar()
```

Out[43]:
```
<AxesSubplot:xlabel='yrs.service'>
```



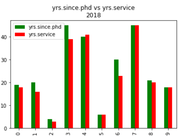
In [45]:
```
data[["yrs.since.phd","yrs.service"]].head(10).plot.bar()
```

Out[45]:
```
<AxesSubplot:>
```



In [47]:
```
data[["yrs.since.phd","yrs.service"]].head(10).plot.bar(title="yrs.since.phd vs yrs.service\n2018",color=["green","red"])
```

Out[47]:
```
<AxesSubplot:title={'center':'yrs.since.phd vs yrs.service\n2018'}>
```

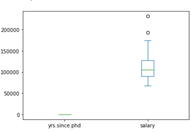


In [48]:
```
data["salary"].head(10).plot.pie(autopct="%.2f")
```

Out[48]:
```
<AxesSubplot:ylabel='salary'>
```

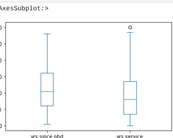
In [49]:
```
data[["yrs.since.phd","salary"]].head(100).plot.box()
```

Out[49]:
```
<AxesSubplot:>
```



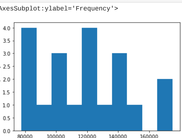
In [50]:
```
data[["yrs.since.phd","yrs.service"]].plot.box()
```

Out[50]:
```
<AxesSubplot:>
```



In [52]:
```
data['salary'].head(20).plot.hist()
```

Out[52]:
```
<AxesSubplot:ylabel='Frequency'>
```



In [ ]: