

```
In [335... import pandas as pd
import matplotlib as mplib
import seaborn as sns
df = pd.read_csv('Salaries.csv')
```

```
In [337... df = df.rename(columns={'rank': 'Rank', 'yrs.service': 'Yrs_Service', 'yrs.since.phd': 'Yrs_Since_PHD'})
```

```
In [339... df.head(5)
```

```
Out[339]:
```

	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 [341... df.Yrs_Service.describe()
```

```
Out[341]: count    397.000000
mean      17.614610
std       13.006024
min        0.000000
25%        7.000000
50%       16.000000
75%       27.000000
max       60.000000
Name: Yrs_Service, dtype: float64
```

```
In [343... df.shape
```

```
Out[343]: (397, 6)
```

```
In [345... df.size
```

```
Out[345]: 2382
```

```
In [347... df['sex'].value_counts()
```

```
Out[347]: sex
Male      358
Female    39
Name: count, dtype: int64
```

```
In [349... df.salary.value_counts()
```

```
Out[349]: salary
92000      5
72500      4
74000      4
105000     3
101000     3
..
76840      1
83001      1
113278     1
155500     1
81035      1
Name: count, Length: 371, dtype: int64
```

```
In [351... df.groupby(['Rank'], sort=False)['salary'].mean()
```

```
Out[351]: Rank
Prof      126772.109023
AsstProf   80775.985075
AssocProf  93876.437500
Name: salary, dtype: float64
```

```
In [353... df.groupby(['Rank'])['salary'].mean()
```

```
Out[353]: Rank
AssocProf  93876.437500
AsstProf   80775.985075
Prof      126772.109023
Name: salary, dtype: float64
```

```
In [355... df.iloc[0:2, 0:6]
```

Out[355]:

	Rank	discipline	Yrs_Since_PHD	Yrs_Service	sex	salary
--	------	------------	---------------	-------------	-----	--------

0	Prof	B	19	18	Male	139750
---	------	---	----	----	------	--------

1	Prof	B	20	16	Male	173200
---	------	---	----	----	------	--------

In [357... `#df.iloc[0]`
`df.iloc[-1]`

Out[357]:

Rank	AsstProf
discipline	A
Yrs_Since_PHD	8
Yrs_Service	4
sex	Male
salary	81035

Name: 396, dtype: object

In [202... `df_sub = df[df['salary'] < 100000]`
`df_sub.head(2)`

Out[202]:

	Rank	discipline	Yrs_Since_PHD	Yrs_Service	sex	salary
--	------	------------	---------------	-------------	-----	--------

2	AsstProf	B	4	3	Male	79750
---	----------	---	---	---	------	-------

5	AssocProf	B	6	6	Male	97000
---	-----------	---	---	---	------	-------

In [214... `df[1:3]`

Out[214]:

	Rank	discipline	Yrs_Since_PHD	Yrs_Service	sex	salary
--	------	------------	---------------	-------------	-----	--------

1	Prof	B	20	16	Male	173200
---	------	---	----	----	------	--------

2	AsstProf	B	4	3	Male	79750
---	----------	---	---	---	------	-------

In [222... `df.loc[1:10,['sex','salary','Rank']]`

Out[222]:

	sex	salary	Rank
1	Male	173200	Prof
2	Male	79750	AsstProf
3	Male	115000	Prof
4	Male	141500	Prof
5	Male	97000	AssocProf
6	Male	175000	Prof
7	Male	147765	Prof
8	Male	119250	Prof
9	Female	129000	Prof
10	Male	119800	AssocProf

In [240... `df.sort_values(by = 'salary').head(2)`

Out[240]:

	Rank	discipline	Yrs_Since_PHD	Yrs_Service	sex	salary
282	Prof	A	51	51	Male	57800
123	AssocProf	A	25	22	Female	62884

In [359... `# sorting data using 2 or more columns`
`df_sorted = df.sort_values(by = ['Yrs_Service', 'salary'], ascending = [True, True])`
`df_sorted.head()`

Out[359]:

	Rank	discipline	Yrs_Since_PHD	Yrs_Service	sex	salary
127	AsstProf	A	2	0	Female	72500
308	AsstProf	A	5	0	Male	74000
28	AsstProf	B	11	0	Male	77000
35	AsstProf	B	5	0	Female	77000
13	AsstProf	B	2	0	Male	78000

In [283...]

```
# Aggregation func in pandas
df[['Yrs_Service', 'salary']].agg(['min', 'mean', 'max'])
```

Out[283]:

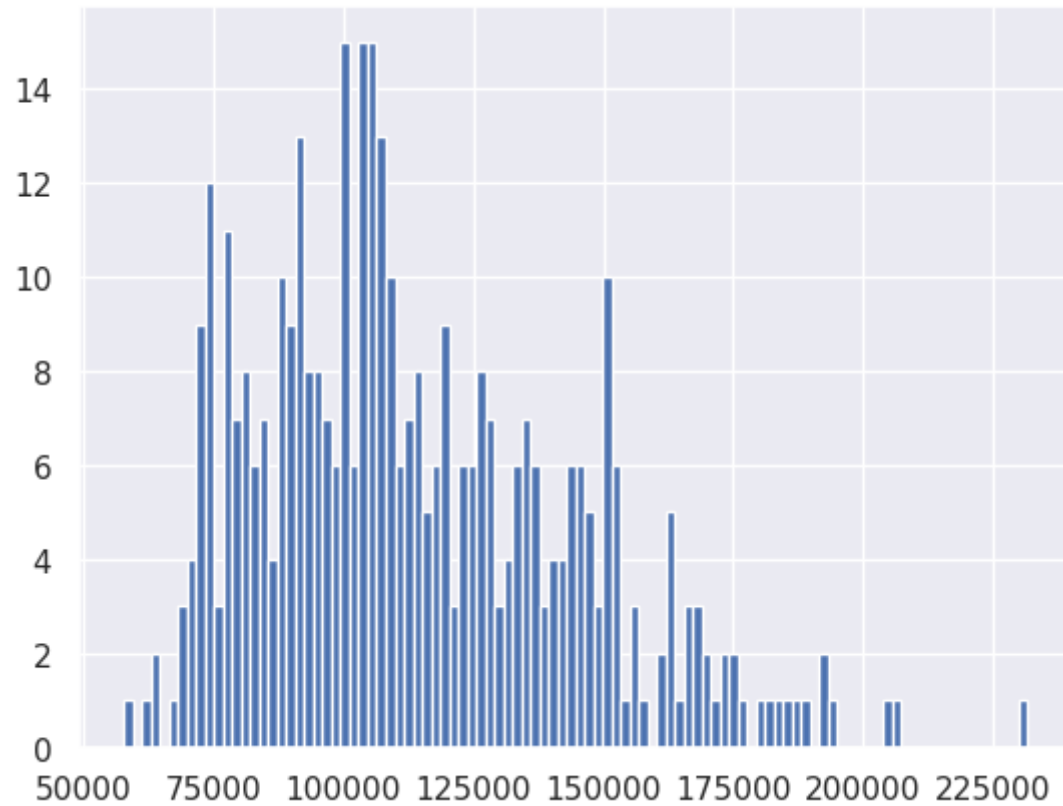
	Yrs_Service	salary
min	0.00000	57800.000000
mean	17.61461	113706.458438
max	60.00000	231545.000000

In [395...]

```
df['salary'].hist(bins=100, density=0)
```

Out[395]:

<Axes: >

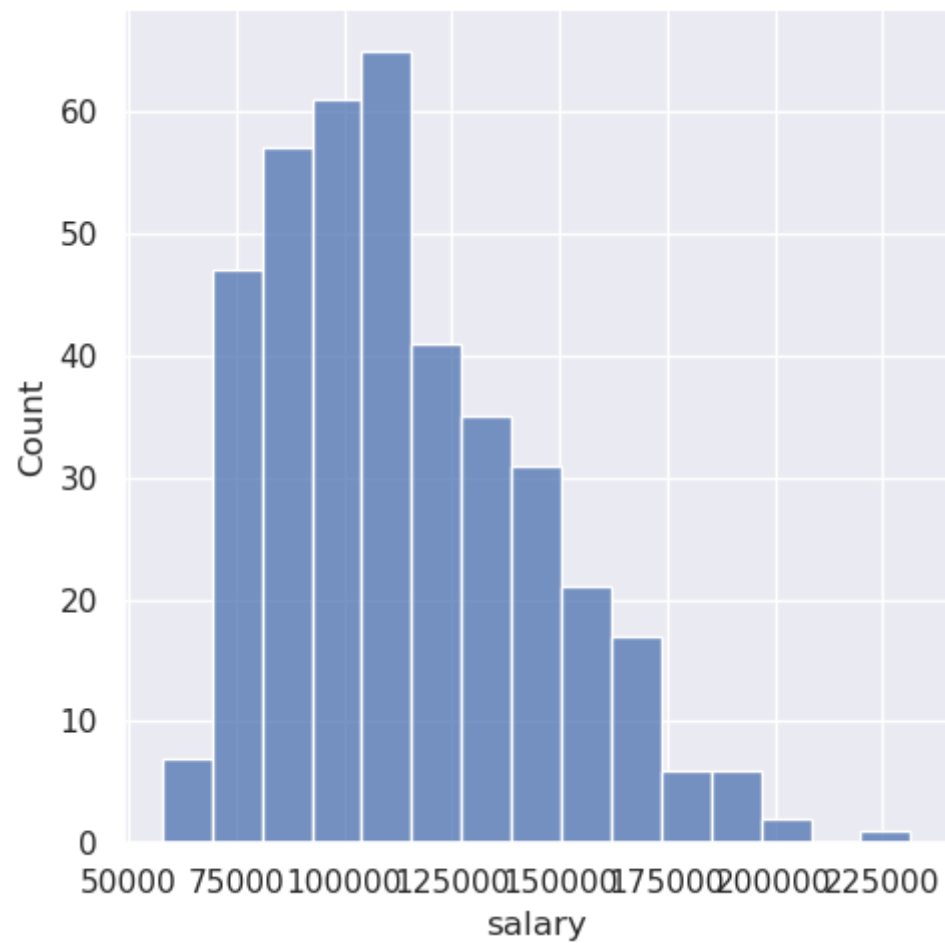


```
In [393...] sns.displot(df[['salary']])
```

```
/opt/conda/envs/anaconda-panel-2023.05-py310/lib/python3.11/site-packages/seaborn/axisgrid.py:118: UserWarning: The figure layout has changed to tight
```

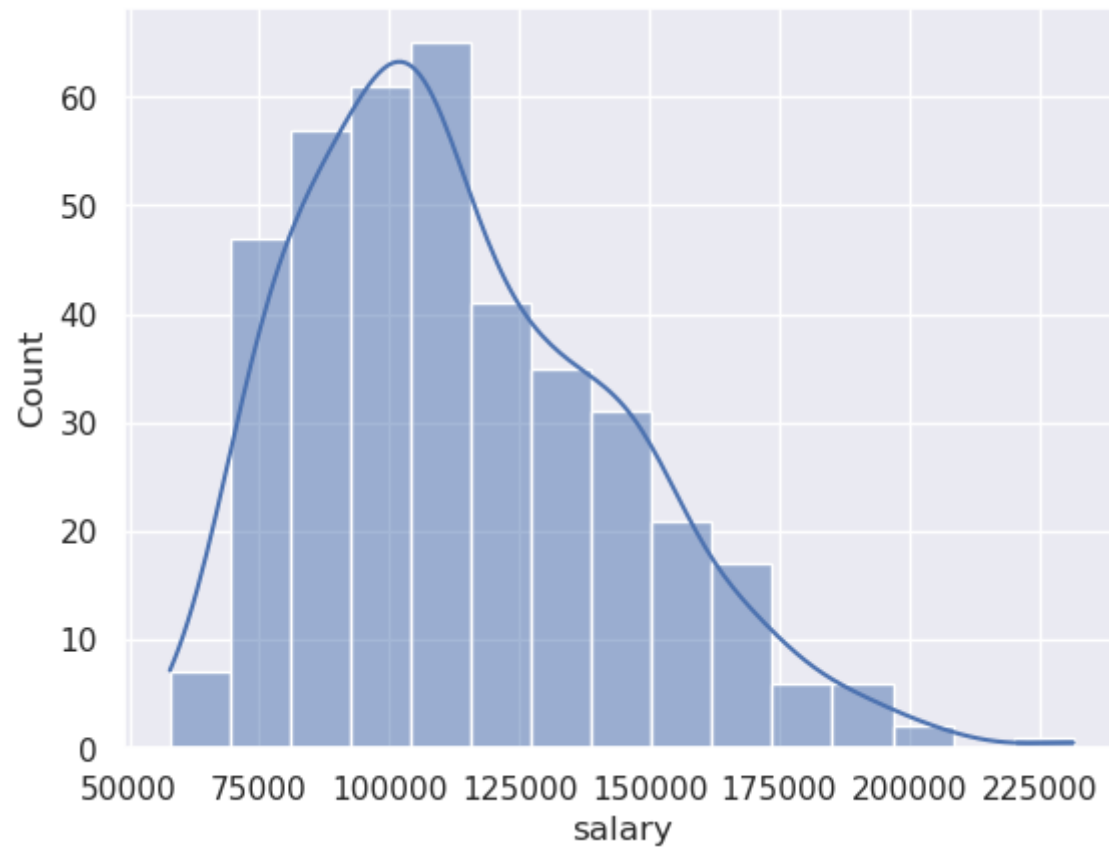
```
self._figure.tight_layout(*args, **kwargs)
```

```
Out[393]: <seaborn.axisgrid.FacetGrid at 0x7ff478552490>
```



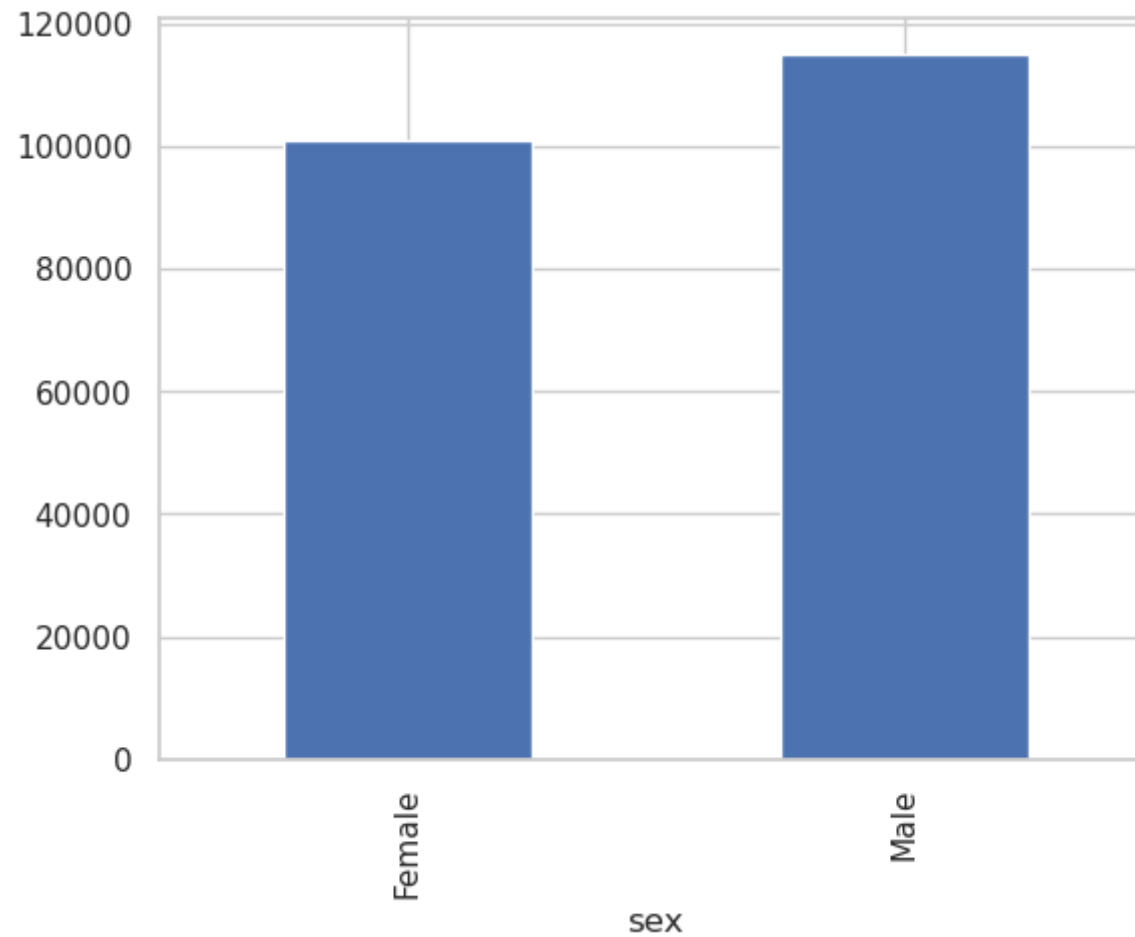
```
In [ ]: sns.histplot(df["salary"], kde=False)
```

```
Out[ ]: <Axes: xlabel='salary', ylabel='Count'>
```

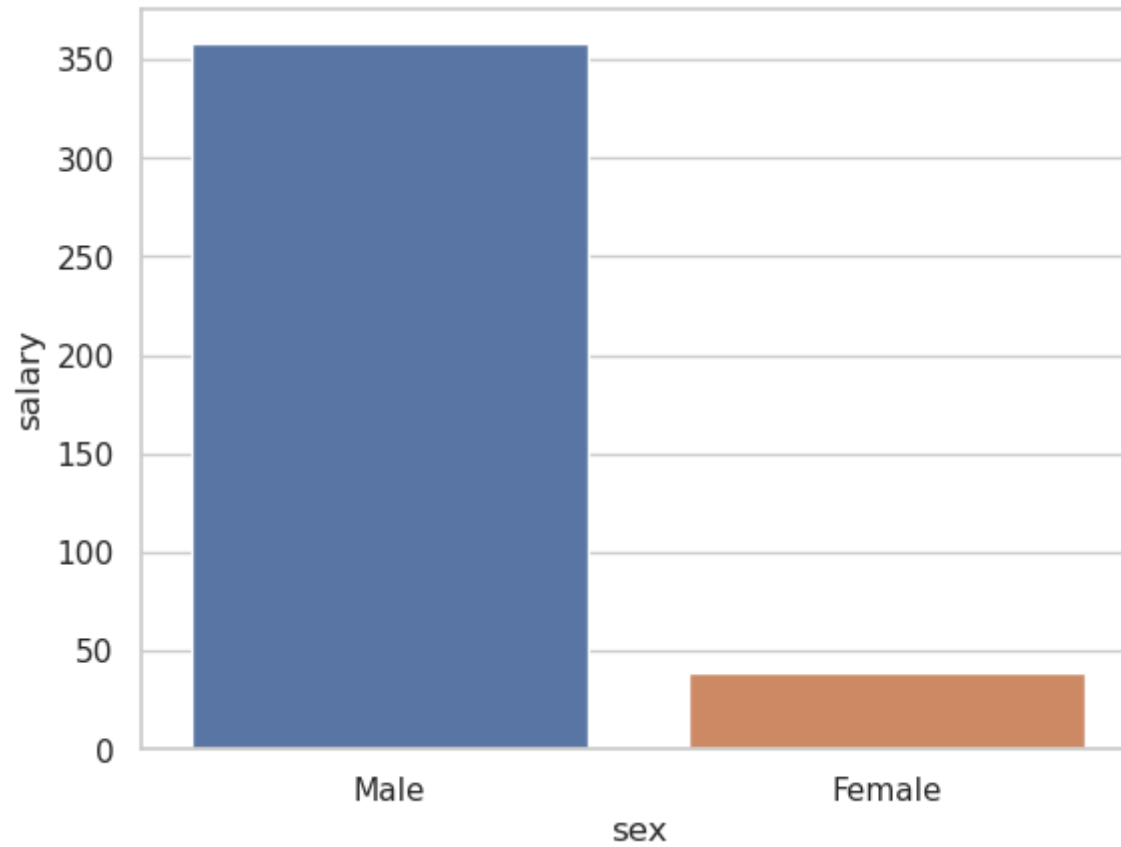


```
In [401... df.groupby(['sex'])['salary'].mean().plot(kind='bar')
```

```
Out[401]: <Axes: xlabel='sex'>
```

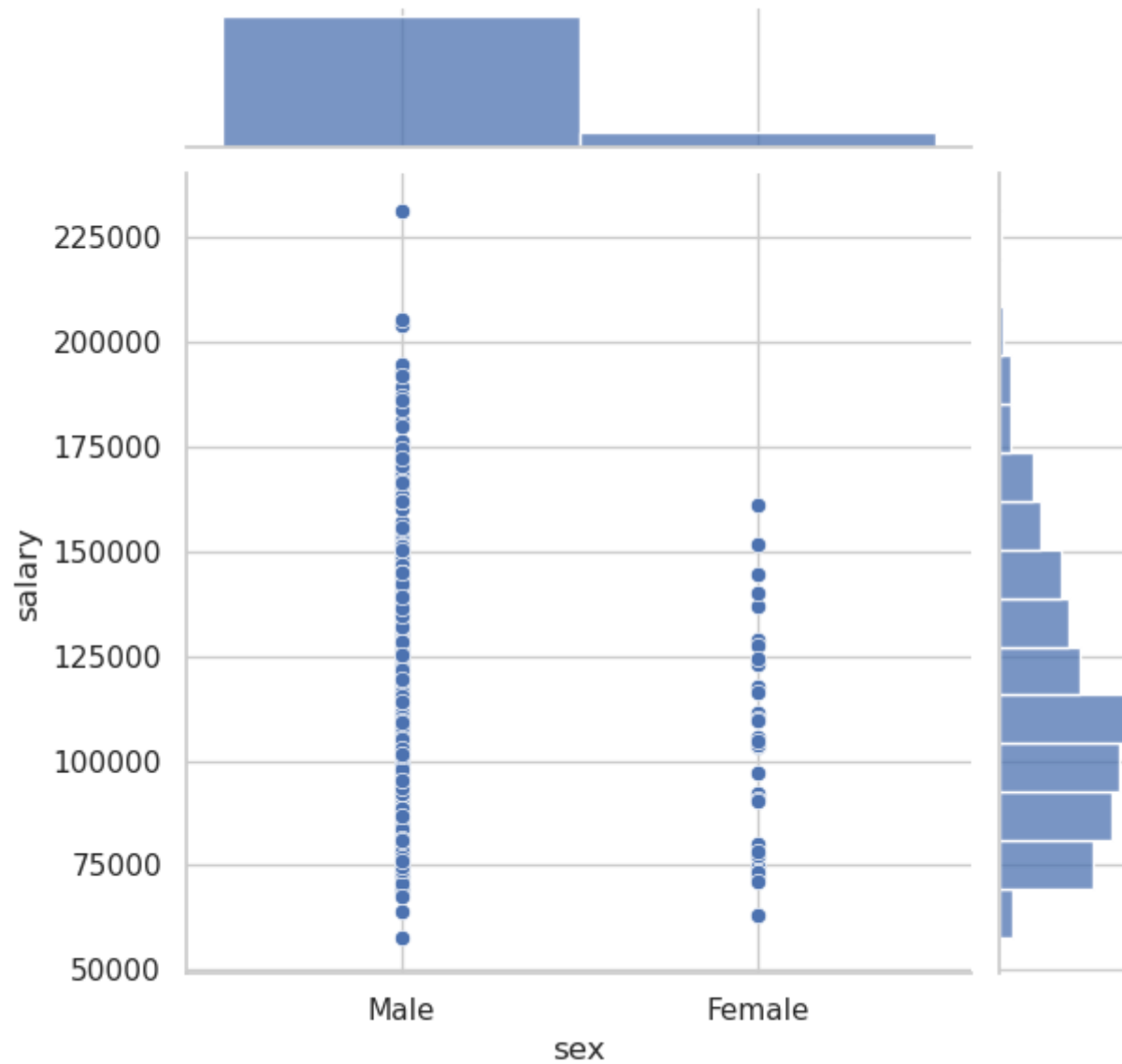



```
In [403... sns.set_style("whitegrid")  
ax = sns.barplot(x='sex',y='salary',data=df,estimator=len)
```



```
In [405]: sns.jointplot(x='sex',y='salary', data=df)
```

```
Out[405]: <seaborn.axisgrid.JointGrid at 0x7ff478461150>
```



```
In [407]: sns.boxplot(x='sex',y='salary', data=df)
```

```
Out[407]: <Axes: xlabel='sex', ylabel='salary'>
```

