Python Project

SDS348 Spring 2021

Daeun Paeng, dp29339

In [14]:
```
# Import packages
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

melanoma = pd.read_excel(r'C:\Users\okcij\Documents\R\melanoma.xlsx')
print(melanoma)
```

```
    status     sex  age  year  thickness    ulcer
0    Other    Male   76  1972       6.76  Present
1    Other    Male   56  1968       0.65   Absent
2    Alive    Male   41  1977       1.34   Absent
3    Other  Female   71  1968       2.90   Absent
4     Dead    Male   52  1965      12.08  Present
..     ...     ...  ...   ...        ...      ...
200  Alive    Male   29  1965       7.06  Present
201  Alive  Female   40  1965       6.12   Absent
202  Alive  Female   42  1965       0.48   Absent
203  Alive  Female   50  1964       2.26   Absent
204  Alive  Female   41  1962       2.90   Absent

[205 rows x 6 columns]
```

The dataset melanoma has 205 rows and 6 columns. The columns are status(status of the patients), sex(sex of the patients), age(age of the patients), year(year of the operation), thickness(thickness of the tumor in mm), and ulcer(presence of ulceration).

EDA

Statistics summary

In [12]:
```
(melanoma.filter(['thickness','sex'])
.groupby(['sex'])
.agg(['mean','min','median','max','std']))
```

Out[12]:

	thickness				
	mean	min	median	max	std
sex					
Female	2.486429	0.10	1.62	17.42	2.754616
Male	3.611139	0.16	2.58	14.66	3.155724

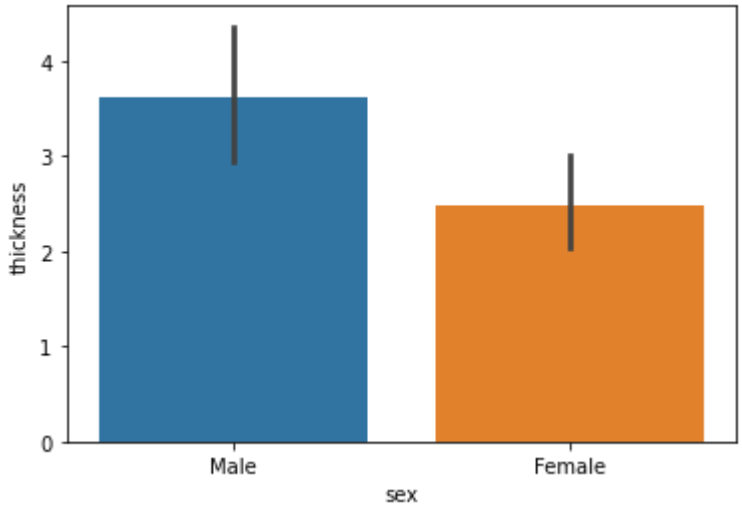
The average thickness of tumor for females and males were 2.486 and 3.611 mm. The lowest thickness of tumor found in females was 0.10 mm while in males it was 0.16 mm. The median thickness of tumor found in females were 1.62 mm while in males it was 2.58 mm. The highest thickness of tumor found in females was 17.42 mm while in males it was 14.66. The standard

deviation for females was 2.754 while for males it was 3.155. From the summary statistics, we can see that males tend to have thicker tumors than females.

Visualization

In [20]:
```
sns.barplot(x="sex", y="thickness", data=melanoma)
```

Out[20]:
```
<AxesSubplot:xlabel='sex', ylabel='thickness'>
```



This bar graph shows that males have thicker tumors than females.