

Assignment 11

March 7, 2021

0.0.1 Portfolio assignment 11

20 min: Do a Numerical VS Categorical bivariate analysis on the penguins dataset. - Choose one of the categorical columns: species, island or sex - use `.groupby('').mean()` to look at the means of the numerical columns. Does it look like there is a difference between categories? - Use the seaborn barplot to plot the mean and confidence. Create this plot for each of the numerical columns (`bill_length_mm`, `bill_depth_mm`, `flipper_length_mm`, `body_mass_g`) - For each of the plots, write a conclusion: Is there a statistically significant difference for this numerical column for each category? - Optional: Repeat this process for the other two categorical columns

```
[1]: import pandas as pd
import seaborn as sns
```

```
[4]: penguins = sns.load_dataset('penguins')
penguins.head()
```

```
[4]:  species      island  bill_length_mm  bill_depth_mm  flipper_length_mm  \
0  Adelie  Torgersen         39.1           18.7           181.0
1  Adelie  Torgersen         39.5           17.4           186.0
2  Adelie  Torgersen         40.3           18.0           195.0
3  Adelie  Torgersen          NaN           NaN           NaN
4  Adelie  Torgersen         36.7           19.3           193.0

      body_mass_g      sex
0         3750.0    Male
1         3800.0  Female
2         3250.0  Female
3            NaN     NaN
4         3450.0  Female
```

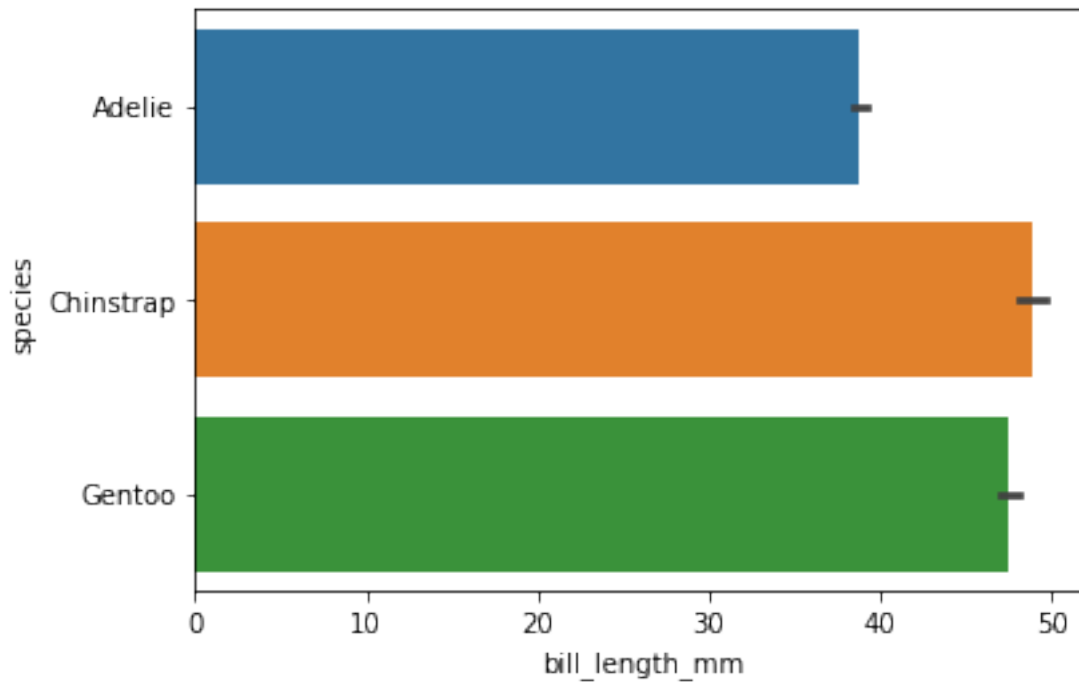
```
[6]: species = penguins.groupby('species').mean()
species
```

```
[6]:          bill_length_mm  bill_depth_mm  flipper_length_mm  body_mass_g
species
Adelie          38.791391      18.346358      189.953642    3700.662252
Chinstrap       48.833824      18.420588      195.823529    3733.088235
Gentoo          47.504878      14.982114      217.186992    5076.016260
```

There is definitely a difference between species, Gentoo seems to be bigger in body mass by a lot.

```
[13]: sns.barplot(data=penguins,x='bill_length_mm', y='species')
```

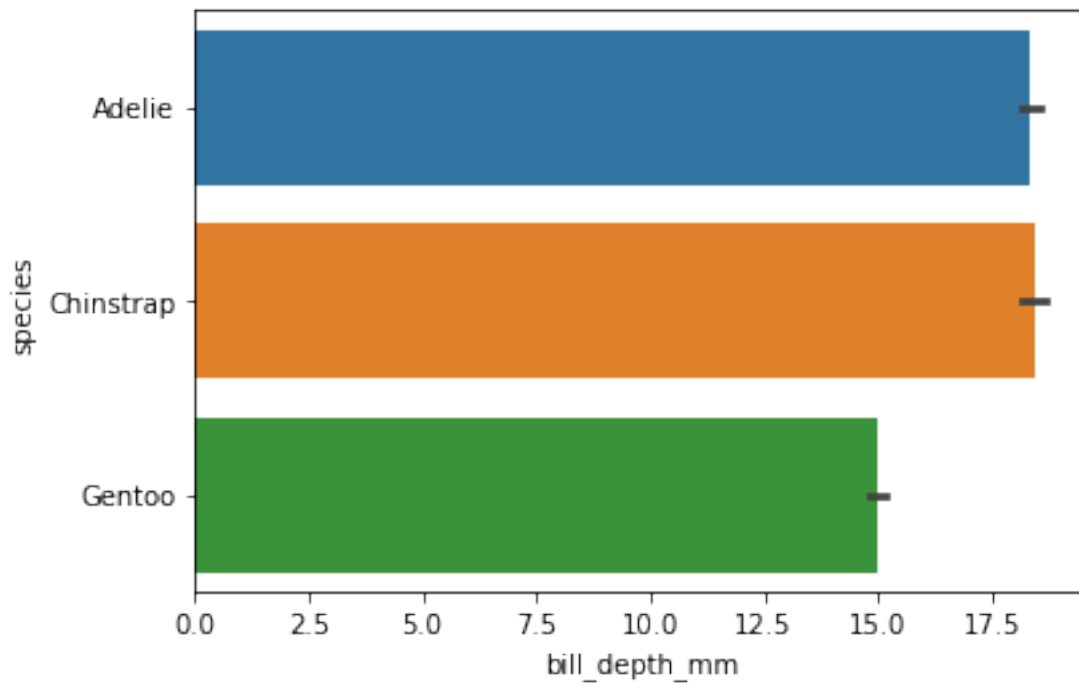
```
[13]: <AxesSubplot:xlabel='bill_length_mm', ylabel='species'>
```



Gentoo on average has a shorter bill depth. Chinstrap on average has the longest bill length.

```
[14]: sns.barplot(data=penguins,x='bill_depth_mm', y='species')
```

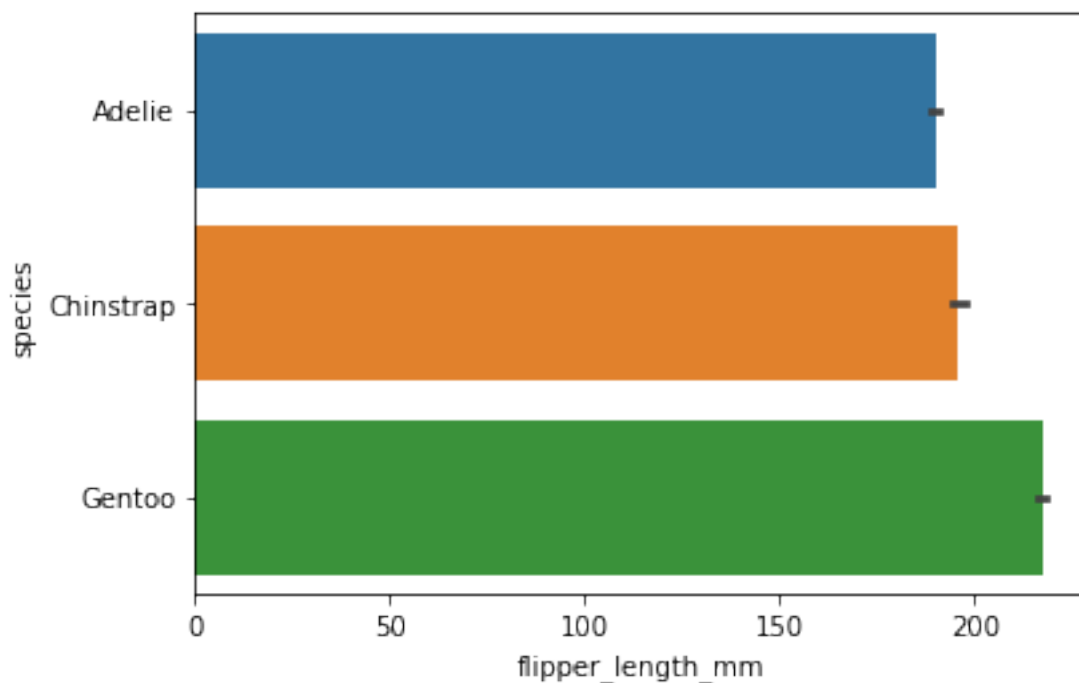
```
[14]: <AxesSubplot:xlabel='bill_depth_mm', ylabel='species'>
```



Gentoo on average has a shorter bill depth, the other two seems to be equal.

```
[15]: sns.barplot(data=penguins,x='flipper_length_mm', y='species')
```

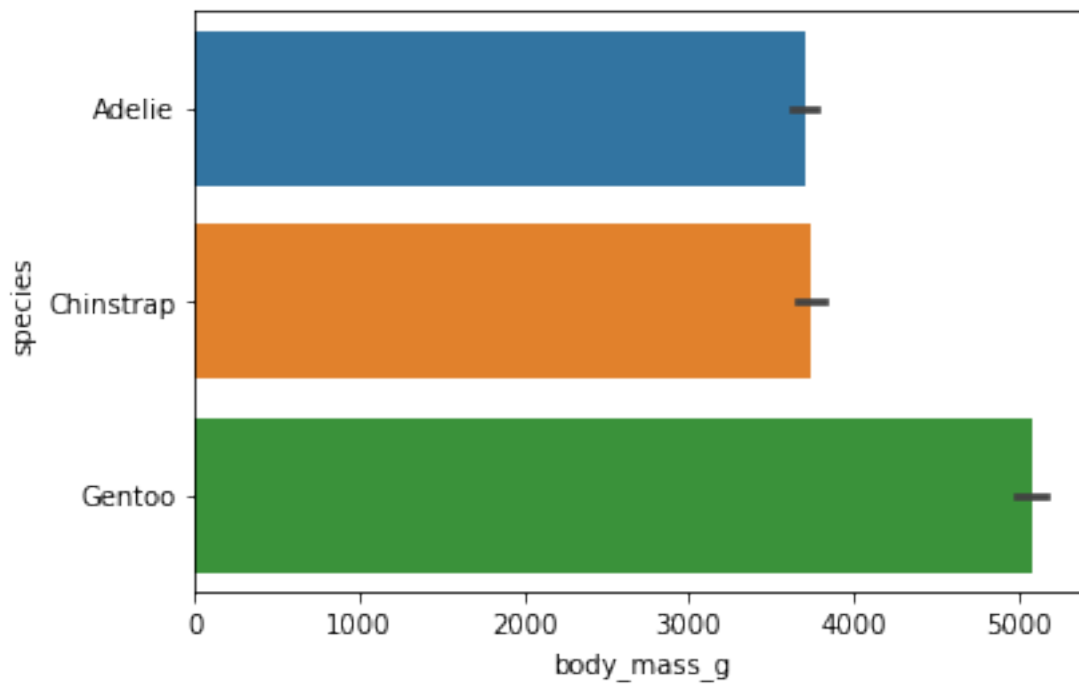
```
[15]: <AxesSubplot:xlabel='flipper_length_mm', ylabel='species'>
```



Gentoo also seems to have a longer flipper length on average, while the Adelie has the shortest.

```
[17]: sns.barplot(data=penguins,x='body_mass_g', y='species')
```

```
[17]: <AxesSubplot:xlabel='body_mass_g', ylabel='species'>
```



As you can see, the gentoo is on average significantly bigger.