Assignment 3

March 7, 2021

0.0.1 Portfolio assignment 3

15 min: Perform a univariate analysis on all the categorical data of the penguins dataset. Commit the notebook to your portfolio when you're finished. Optional: Start working on portfolio assignment 4

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

[2]:
```
penguins = sns.load_dataset("penguins")
```

[3]:
```
penguins
```

[3]:
```
     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
..       ...        ...             ...            ...                ...
339   Gentoo     Biscoe             NaN            NaN                NaN
340   Gentoo     Biscoe            46.8           14.3              215.0
341   Gentoo     Biscoe            50.4           15.7              222.0
342   Gentoo     Biscoe            45.2           14.8              212.0
343   Gentoo     Biscoe            49.9           16.1              213.0

     body_mass_g     sex
0         3750.0    Male
1         3800.0  Female
2         3250.0  Female
3            NaN     NaN
4         3450.0  Female
..           ...     ...
339          NaN     NaN
340       4850.0  Female
341       5750.0    Male
342       5200.0  Female
343       5400.0    Male
```

```
[344 rows x 7 columns]
```

1 Split values

These values are usually 2 options that won't change anytime soon. They don't really need any visualization.

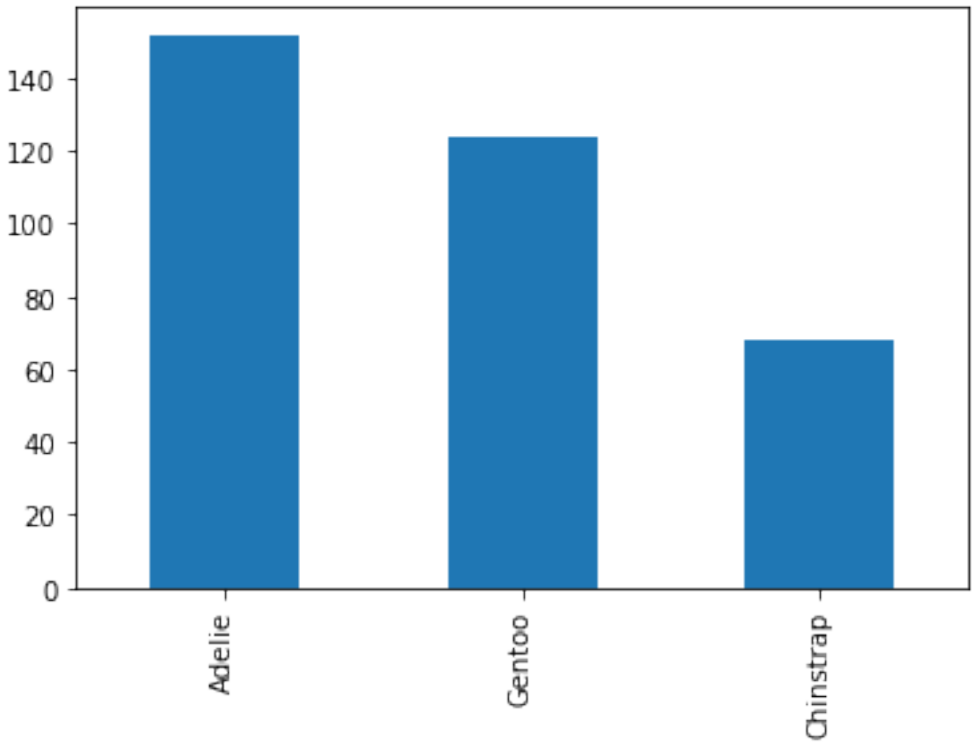
[4]:
```
penguins['sex'].value_counts()
```

[4]:
```
Male      168
Female    165
Name: sex, dtype: int64
```

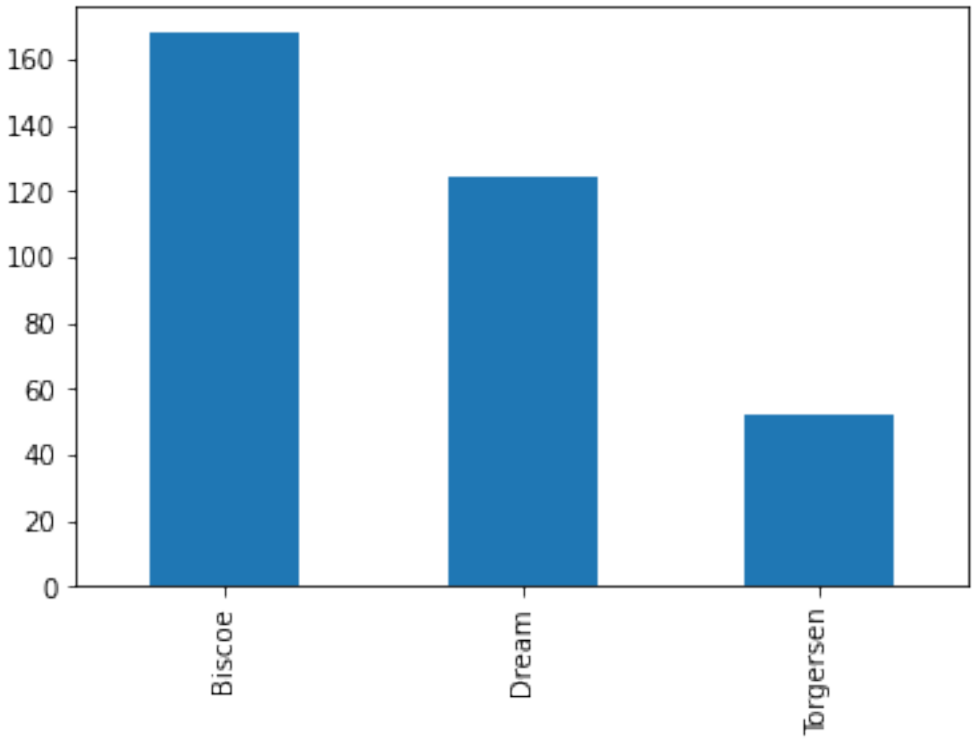
2 Division

Here we can see the division between different species and islands. Might be fun to see some correlation between them.

[5]:
```
penguins['species'].value_counts().plot.bar();
```



[6]:
```
penguins['island'].value_counts().plot.bar();
```

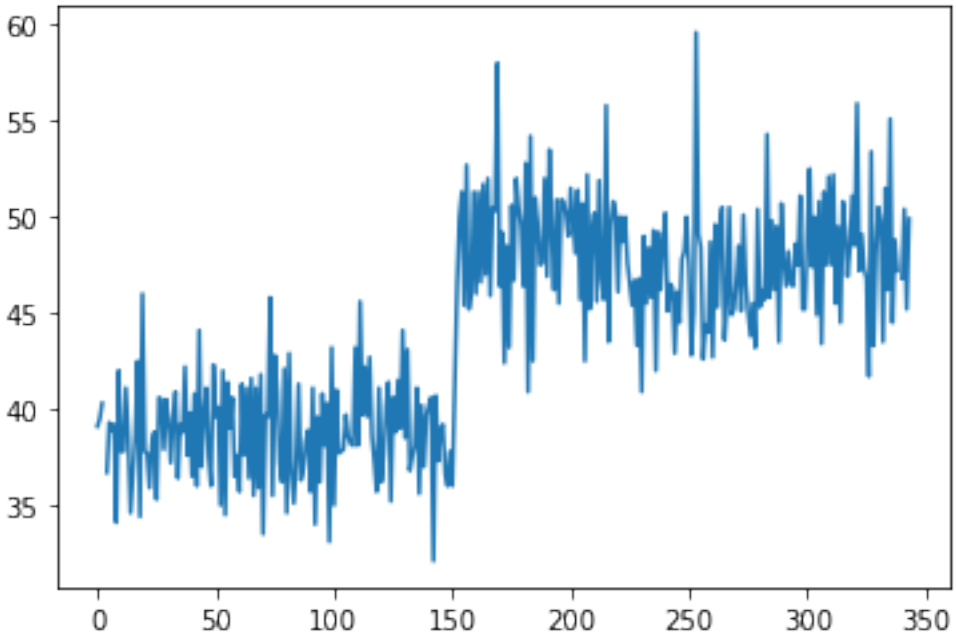


2.1 Lengths

You can see the amount of penguins that has a certain bill length. I'm not so sure on the usefulness here.

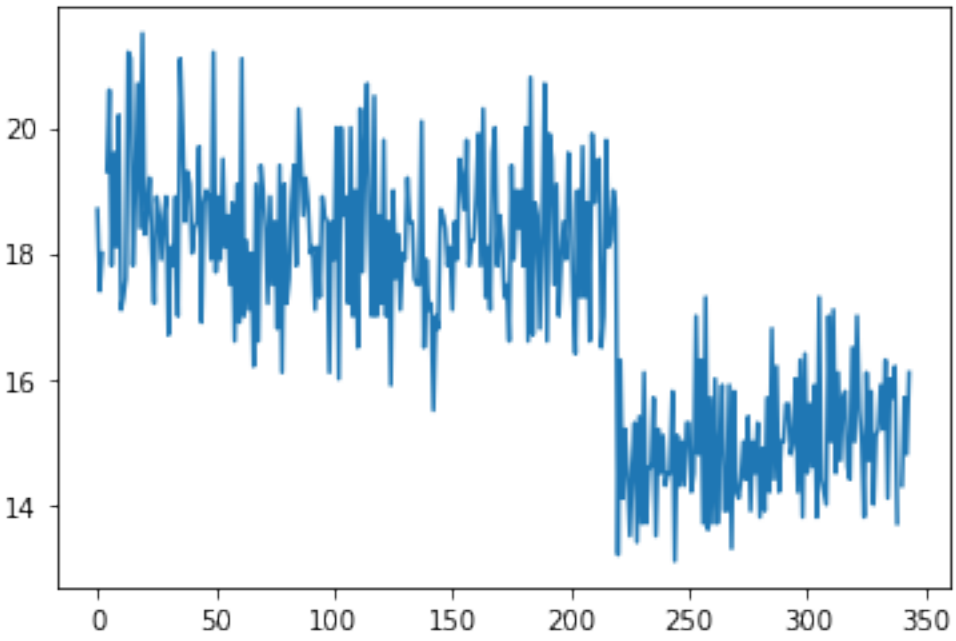
[7]:
```
penguins['bill_length_mm'].plot()
```

[7]:
```
<AxesSubplot:>
```



[8]:
```
penguins['bill_depth_mm'].plot()
```

[8]: <AxesSubplot:>



[9]:
```
penguins['flipper_length_mm'].plot()
```

[9]: <AxesSubplot:>