w3_assessment

October 18, 2019

In this assignment we'll ask you to plot multiple variables.

You will use what you find in this assignment to answer the questions in the quiz that follows. It may be useful to keep this notebook side-by-side with this week's quiz on your screen.

In [2]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import scipy.stats as stats
%matplotlib inline
import matplotlib.pyplot as plt
pd.set_option('display.max_columns', 100)

path = "Cartwheeldata.csv"
```

In [3]:
```
# First, you must import the cartwheel data from the path given above
df = pd.read_csv(path)# using pandas, read in the csv data found at the url defined by
```

If you can't remember a function, open a previous notebook or video as a reference, or use your favorite search engine to look for a solution.

In [4]:
```
# Next, look at the 'head' of our DataFrame 'df'.
df.head()
```

Out[4]:

	ID	Age	Gender	GenderGroup	Glasses	GlassesGroup	Height	Wingspan	\
0	1	56	F	1	Y	1	62.0	61.0	
1	2	26	F	1	Y	1	62.0	60.0	
2	3	33	F	1	Y	1	66.0	64.0	
3	4	39	F	1	N	0	64.0	63.0	
4	5	27	M	2	N	0	73.0	75.0	

	CWDistance	Complete	CompleteGroup	Score
0	79	Y	1	7
1	70	Y	1	8
2	85	Y	1	7
3	87	Y	1	10
4	72	N	0	4

In [9]:
```
df.describe()
```

Out[9]:

	ID	Age	GenderGroup	GlassesGroup	Height	Wingspan
count	25.000000	25.000000	25.000000	25.000000	25.000000	25.000000
mean	13.000000	28.240000	1.520000	0.560000	67.650000	66.260000
std	7.359801	6.989754	0.509902	0.506623	4.431187	5.492647
min	1.000000	22.000000	1.000000	0.000000	61.500000	57.500000
25%	7.000000	24.000000	1.000000	0.000000	64.000000	62.000000
50%	13.000000	26.000000	2.000000	1.000000	68.000000	66.000000
75%	19.000000	29.000000	2.000000	1.000000	71.000000	71.000000
max	25.000000	56.000000	2.000000	1.000000	75.000000	76.000000

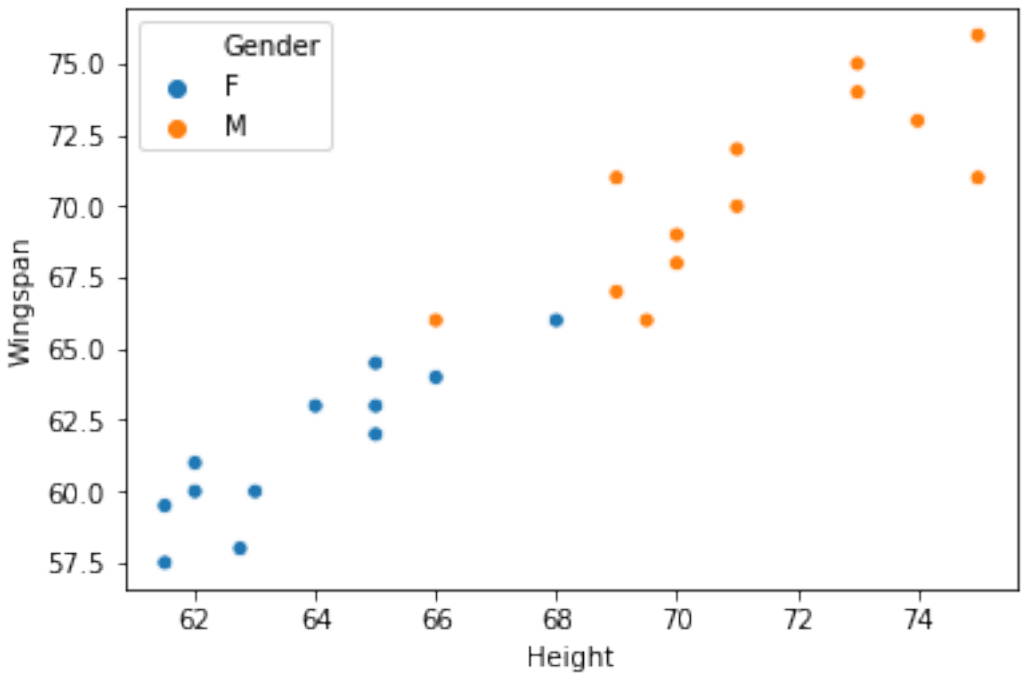
	CWDistance	CompleteGroup	Score
count	25.000000	25.00000	25.000000
mean	82.480000	0.76000	6.400000
std	15.058552	0.43589	2.533114
min	63.000000	0.00000	2.000000
25%	70.000000	1.00000	4.000000
50%	81.000000	1.00000	6.000000
75%	92.000000	1.00000	8.000000
max	115.000000	1.00000	10.000000

0.1 Scatter plots

First, let's looks at two variables that we expect to have a strong relationship, 'Height' and 'Wingspan'.

In [8]:
```
# Make a Seaborn scatter plot with x = height and y = wingspan using sns.scatterplot(x
sns.scatterplot(x='Height', y='Wingspan', data=df,hue=df['Gender'])
```

Out[8]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f8cfb394dd8>
```



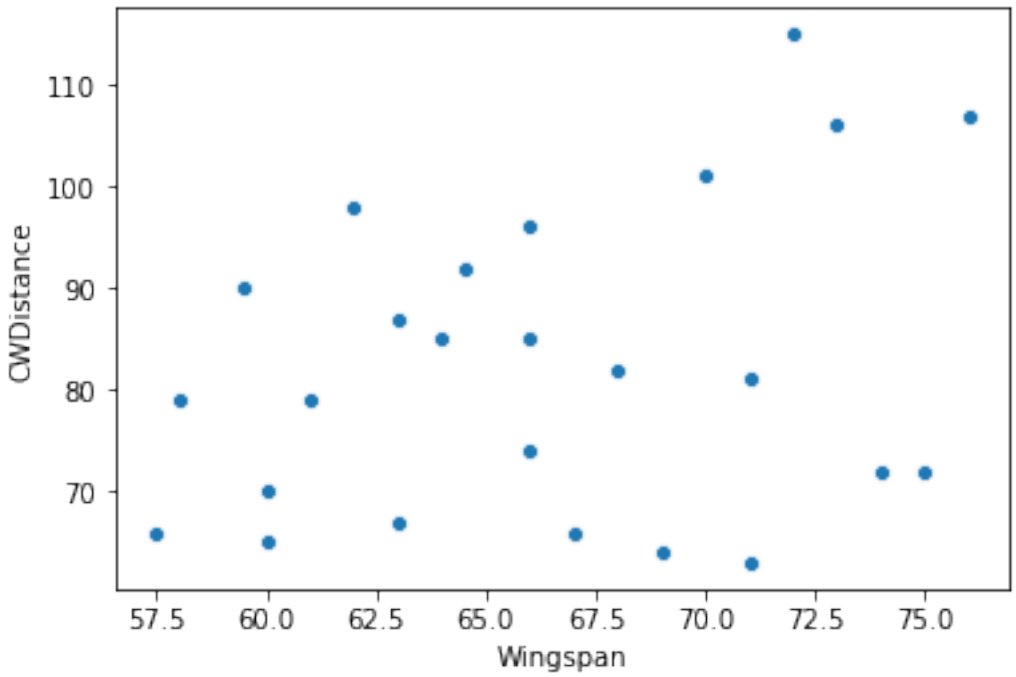
How would you describe the relationship between 'Height' and 'Wingspan'?
Questions you can ask: * Is it linear? * Are there outliers? * Are their ranges similar or different?

How else could you describe the relationship?

Now let's look at two variables that we don't yet assume have a strong relationship, 'Wingspan' and 'CWDistance'

```
In [6]: # Make a Seaborn scatter plot with x = wingspan and y = cartwheel distance
        sns.scatterplot(x='Wingspan', y='CWDistance', data=df)
```

```
Out[6]: <matplotlib.axes._subplots.AxesSubplot at 0x7f8cfb4b6320>
```



How would you describe the relationship between 'Wingspan' and 'CWDistance'?
* Is it linear? * Are there outliers? * Are their ranges similar or different?

How else could you describe the relationship?

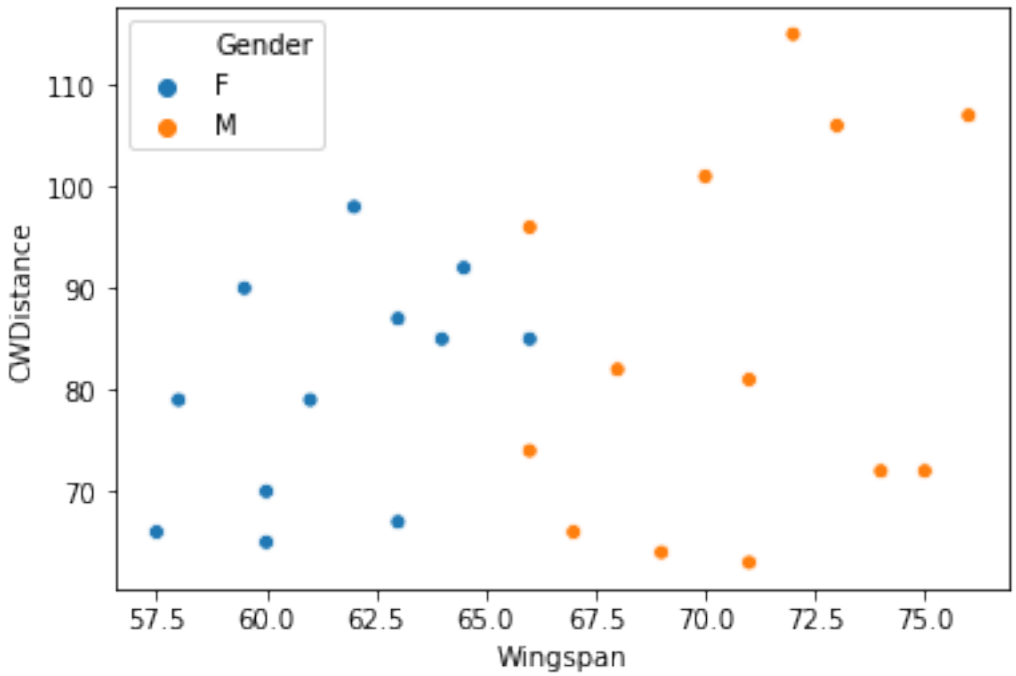
Let makes the same plot as above, but now include 'Gender' as the color scheme by including the argument

```
hue=df['Gender']
```

in the Seaborn function

```
In [22]: # Make a Seaborn scatter plot with x = wingspan and y = cartwheel distance, and hue =
         sns.scatterplot(x='Wingspan', y='CWDistance', data=df,hue=df['Gender'])
```

```
Out[22]: <matplotlib.axes._subplots.AxesSubplot at 0x7f8cfb09cb38>
```



Does does this new information on the plot change your interpretation of the relationship between 'Wingspan' and 'CWDistance'?

0.2 Barcharts

Now lets plot barplots of 'Glasses'

```
In [1]: # Make a Seaborn barplot with x = glasses and y = cartwheel distance
        sns.barplot(x='Glasses',hue='CWDistance',data=df)

        ---------------------------------------------------------------------------

        NameError                                 Traceback (most recent call last)

        <ipython-input-1-1cc9e02b7e7f> in <module>()
          1 # Make a Seaborn barplot with x = glasses and y = cartwheel distance
        ----> 2 sns.barplot(x='Glasses',hue='CWDistance',data=df)


        NameError: name 'sns' is not defined
```

What can you say about the relationship of 'Glasses' and 'CWDistance'?

```
In [0]: # Make the same Seaborn boxplot as above, but include gender for the hue argument
```

How does this new plot change your interpretation about the relationship of 'Glasses' and 'CWDistance'?