

# w3\_assessment

August 30, 2020

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 [1]: 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 [2]: # 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
```

```
In [3]: # Next, look at the 'head' of our DataFrame 'df'.
df.head()
```

```
Out[3]:
```

	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

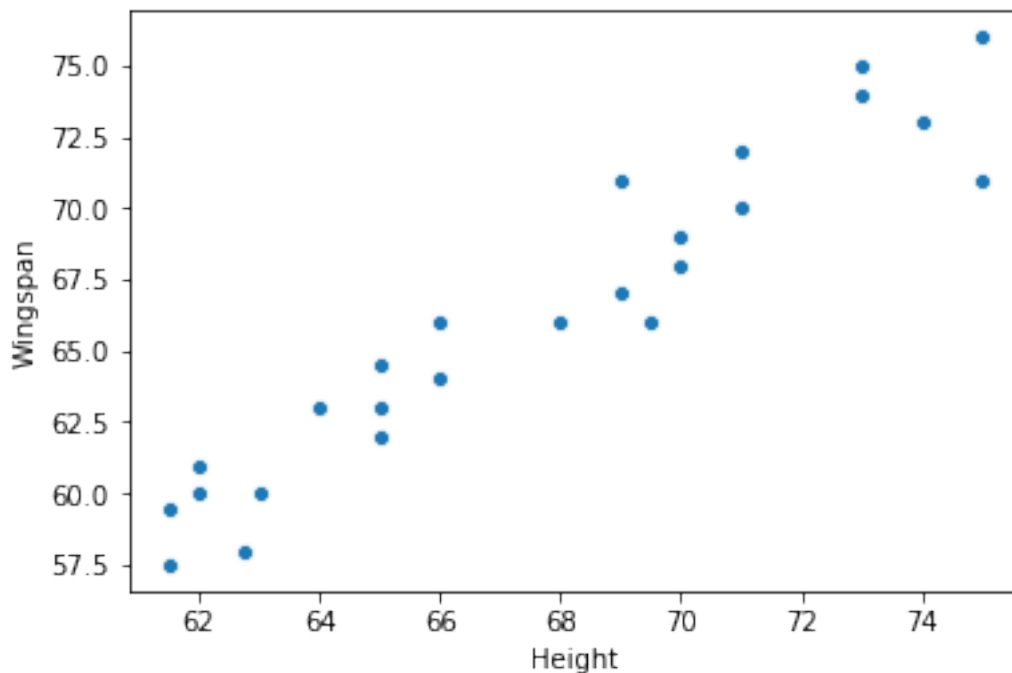
If you can't remember a function, open a previous notebook or video as a reference, or use your favorite search engine to look for a solution.

## 0.1 Scatter plots

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

```
In [5]: # Make a Seaborn scatter plot with x = height and y = wingspan using sns.scatterplot(x
x = df.Height
y = df.Wingspan
sns.scatterplot(x,y)
```

```
Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0x7fb449da1978>
```



How would you describe the relationship between 'Height' and 'Wingspan'?

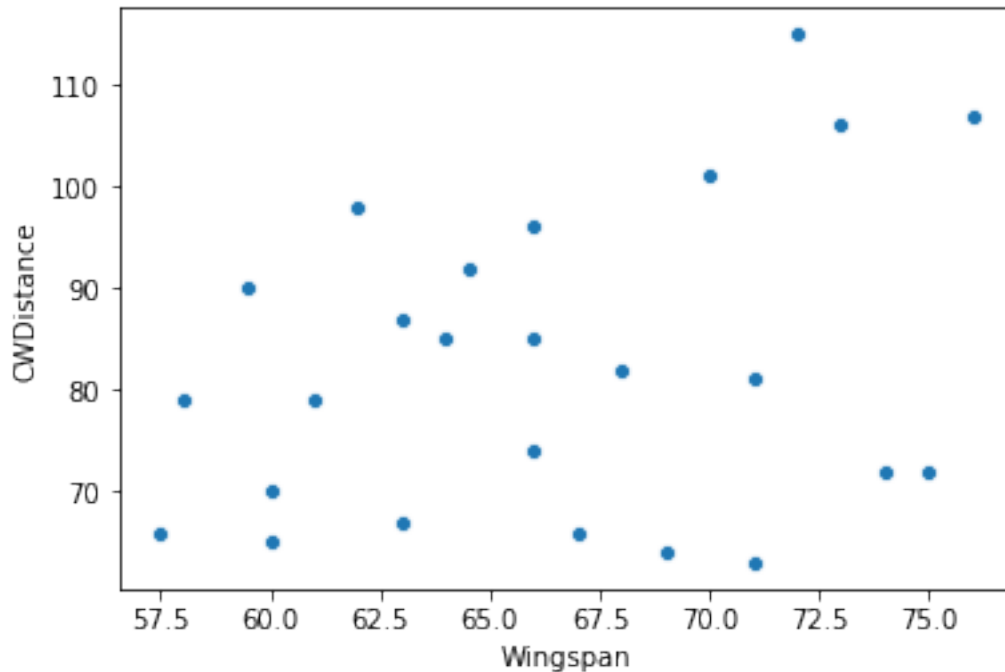
Questions you can ask: \* Is it linear? \* Are there outliers? \* Are their ranges similar or different?

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
x = df.Wingspan
y = df.CWDistance
sns.scatterplot(x, y)
```

```
Out[6]: <matplotlib.axes._subplots.AxesSubplot at 0x7fb449da1908>
```



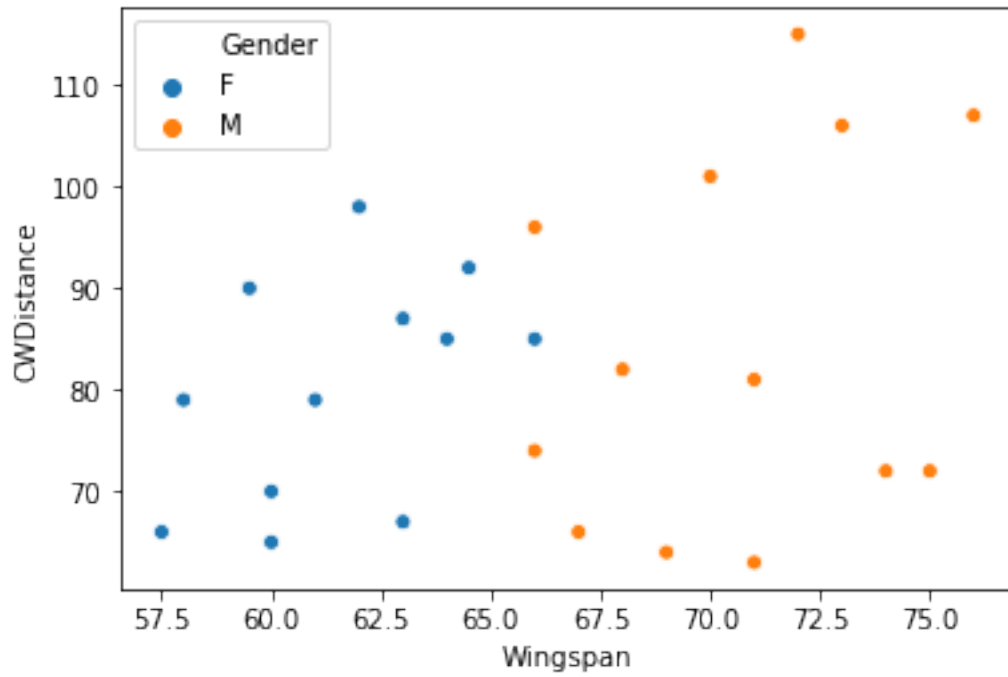
How would you describe the relationship between 'Wingspan' and 'CWDistance'?  
 \* Is it linear? \* Are there outliers? \* Are their ranges similar or different?  
 How else could you describe the relationship?  
 Let makes the same plot as above, but now include 'Gender' as the color scheme by including the argument

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

in the Seaborn function

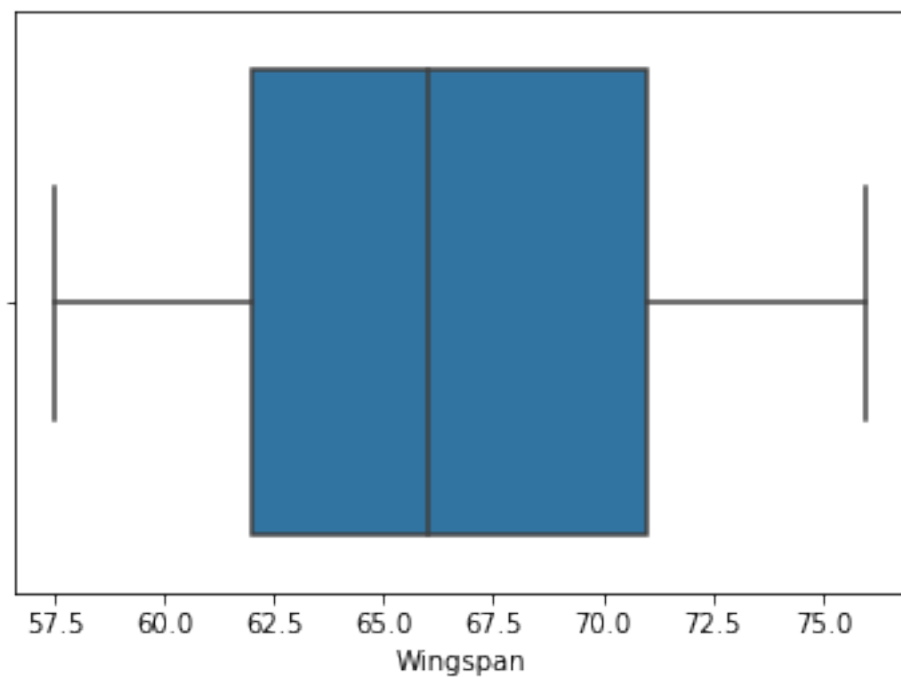
```
In [16]: # Make a Seaborn scatter plot with x = wingspan and y = cartwheel distance, and hue =
x = df.Wingspan
y = df.CWDistance
sns.scatterplot(x, y, hue=df['Gender'])
```

```
Out[16]: <matplotlib.axes._subplots.AxesSubplot at 0x7fb447930d30>
```



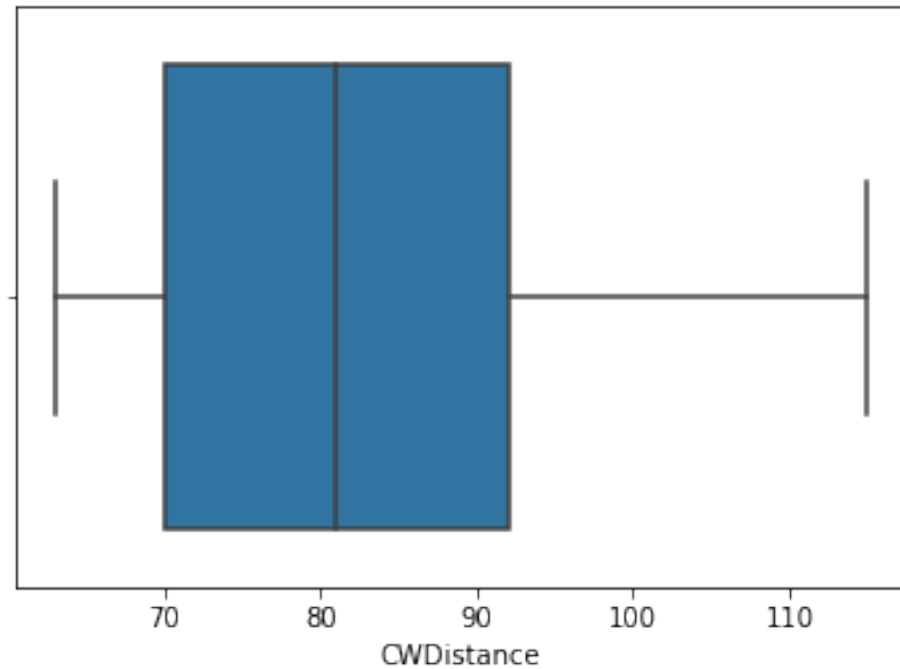
```
In [15]: sns.boxplot(x)
```

```
Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x7fb4479659b0>
```



```
In [14]: sns.boxplot(y)
```

```
Out[14]: <matplotlib.axes._subplots.AxesSubplot at 0x7fb447ac1588>
```



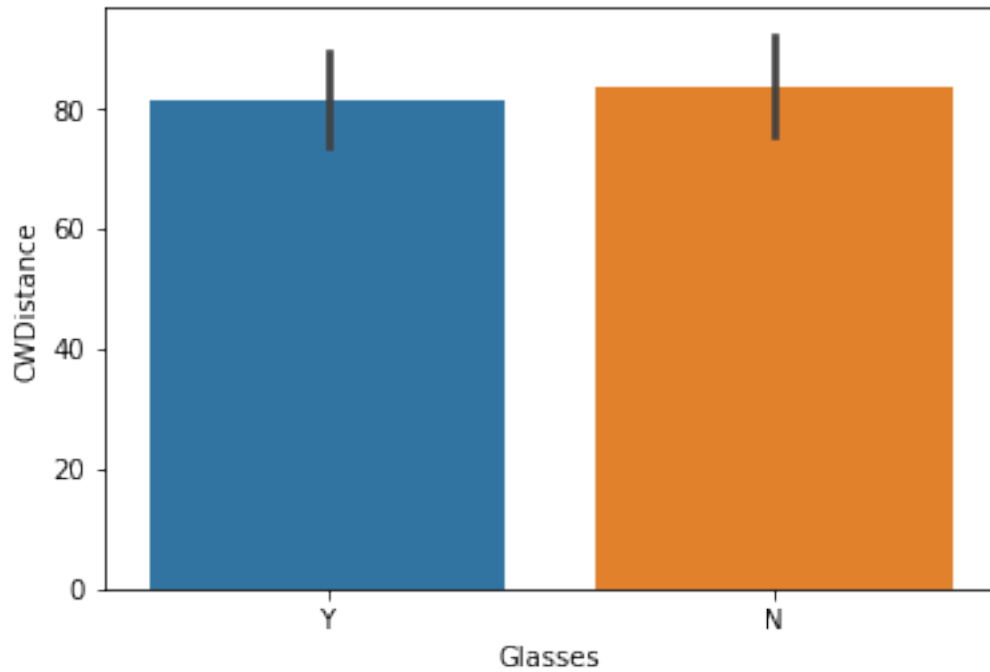
Does this new information on the plot change your interpretation of the relationship between 'Wingspan' and 'CWDistance'?

## 0.2 Barcharts

Now lets plot barplots of 'Glasses'

```
In [8]: # Make a Seaborn barplot with x = glasses and y = cartwheel distance
sns.barplot(x=df['Glasses'], y=df['CWDistance'])
```

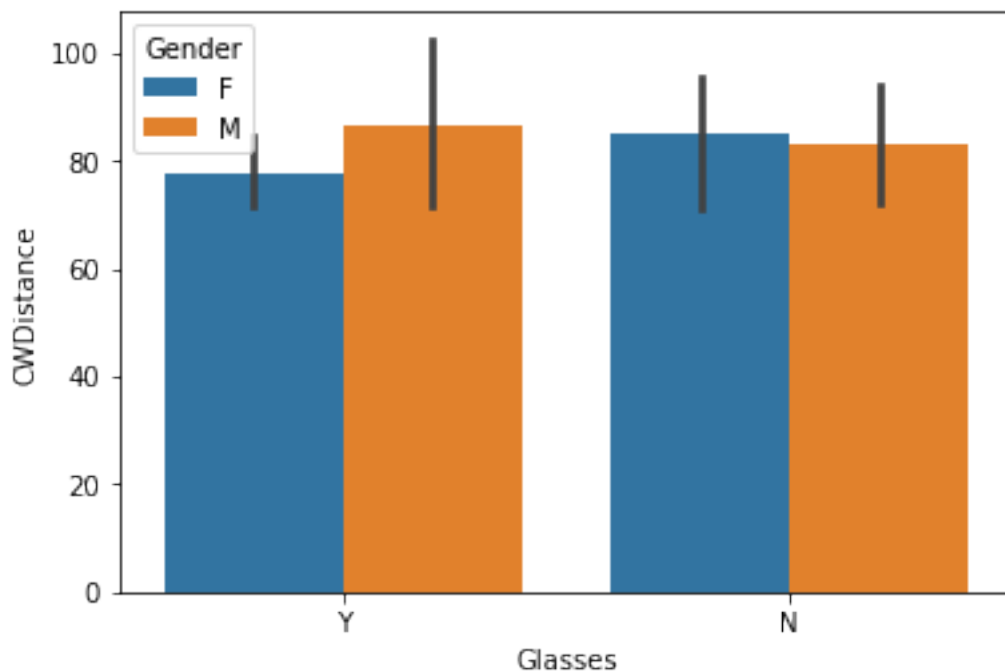
```
Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x7fb447cd3c50>
```



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

In [10]: *# Make the same Seaborn boxplot as above, but include gender for the hue argument*  
`sns.barplot(x=df['Glasses'], y=df['CWDistance'], hue=df['Gender'])`

Out[10]: `<matplotlib.axes._subplots.AxesSubplot at 0x7fb447ba59e8>`



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