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 'path'
```

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

Out[3]:

	ID	Age	Gender	GenderGroup	Glasses	GlassesGroup	Height	Wingspan	CWDistance	Complete	CompleteGroup	Score
0	1	56	F	1	Y	1	62.0	61.0	79	Y	1	7
1	2	26	F	1	Y	1	62.0	60.0	70	Y	1	8
2	3	33	F	1	Y	1	66.0	64.0	85	Y	1	7
3	4	39	F	1	N	0	64.0	63.0	87	Y	1	10
4	5	27	M	2	N	0	73.0	75.0	72	N	0	4

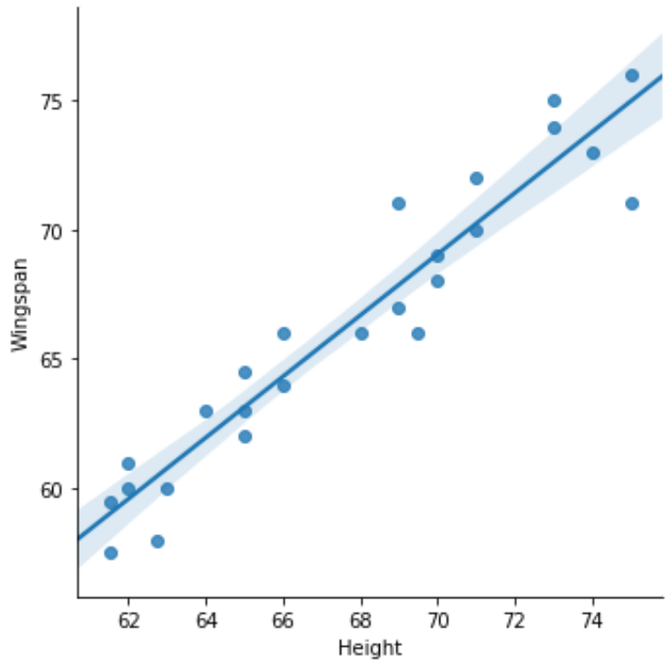
If you can't remember a function, open a previous notebook or video as a reference, or use your favorite search engine to look for a solution.

Scatter plots

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

In [4]:
```
# Make a Seaborn scatter plot with x = height and y = wingspan using sns.scatterplot(x, y)
sns.lmplot(x='Height', y='Wingspan', data=df)
```

Out[4]: <seaborn.axisgrid.FacetGrid at 0x7fab482f10f0>



How would you describe the relationship between 'Height' and 'Wingspan'?
Questions you can ask:

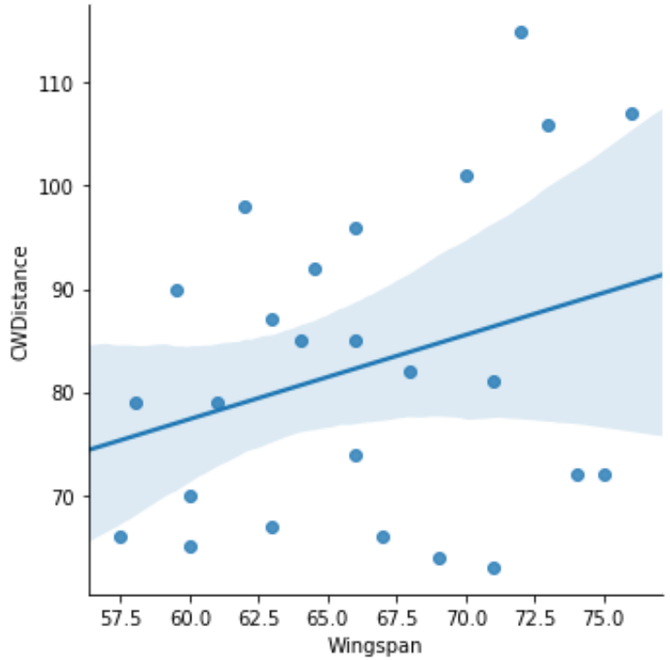
- Is it linear?
- Are there outliers?
- Are their ranges similar or different?

How else could you describe the relationship? ==> About as direct as you can get

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

In [5]:
```
# Make a Seaborn scatter plot with x = wingspan and y = cartwheel distance
sns.lmplot(x='Wingspan', y='CWDistance', data=df)
```

Out[5]: <seaborn.axisgrid.FacetGrid at 0x7fac187fe3c8>



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? ==> very slightly positive; lots of outliers; not really linear

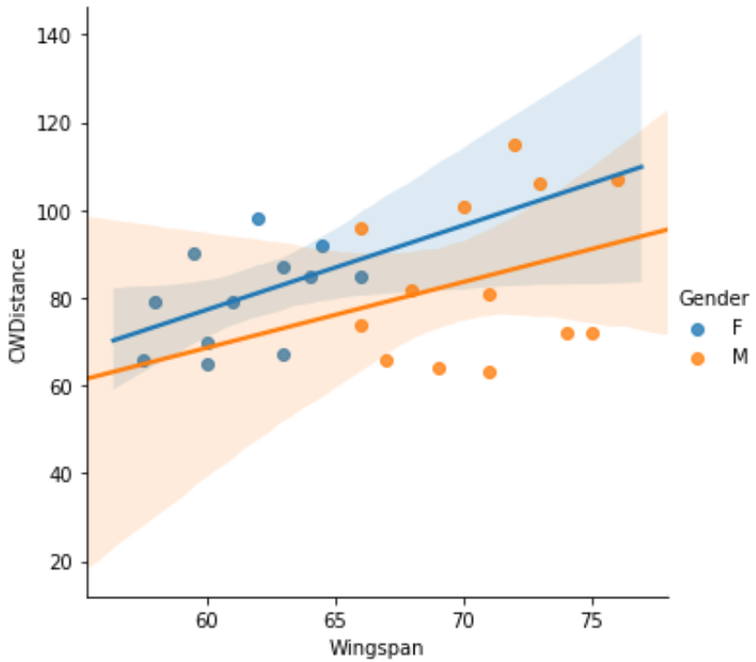
Let makes the same plot as above, but now include 'Gender' as the color scheme by including the argument

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

in the Seaborn function

In [12]:
```
# Make a Seaborn scatter plot with x = wingspan and y = cartwheel distance, and hue = gender
sns.lmplot(x='Wingspan', y='CWDistance', data=df, hue='Gender')
```

Out[12]: <seaborn.axisgrid.FacetGrid at 0x7fab45e7fba8>



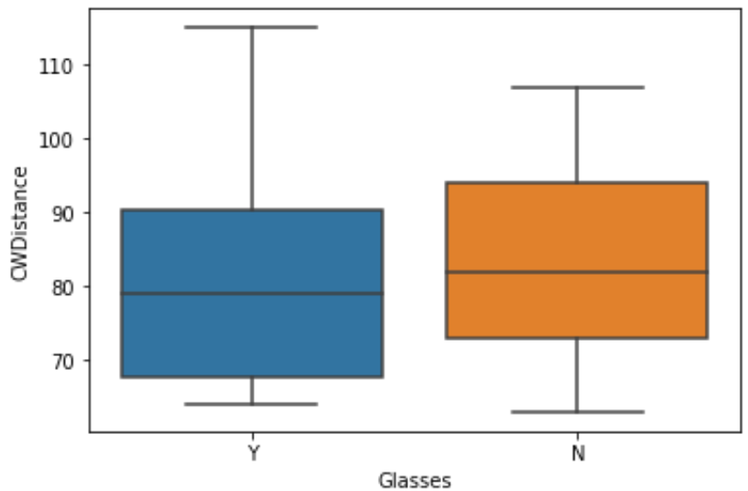
Does does this new information on the plot change your interpretation of the relationship between 'Wingspan' and 'CWDistance'? ==> Not really

Barcharts

Now lets plot barplots of 'Glasses'

In [9]:
```
# Make a Seaborn barplot with x = glasses and y = cartwheel distance
sns.boxplot(x='Glasses', y='CWDistance', data=df)
```

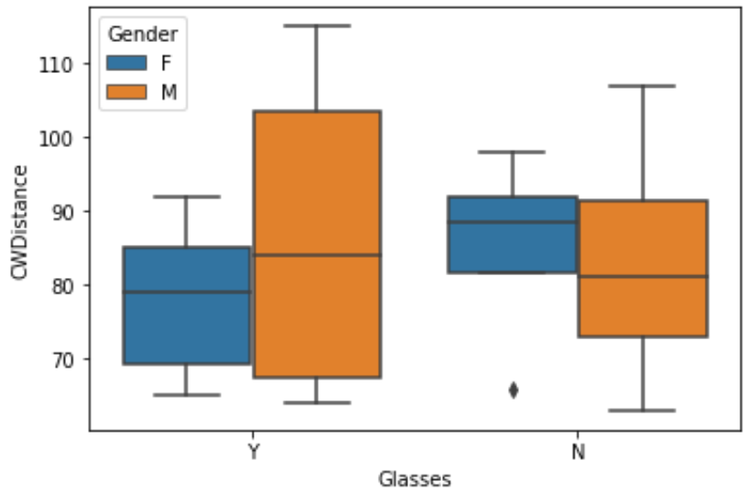
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x7fab46070828>



What can you say about the relationship of 'Glasses' and 'CWDistance'? ==> CWDistance roughly 10% greater for no Glasses

In [14]:
```
sns.boxplot(x='Glasses', y='CWDistance', data=df, hue='Gender')
```

Out[14]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fab45dd5780>
```



How does this new plot change your interpretation about the relationship of 'Glasses' and 'CWDistance'? ==> difference much more pronounced for females; looks to be a bit lower for males