w3_assessment

November 11, 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)
```

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

Out[3]:

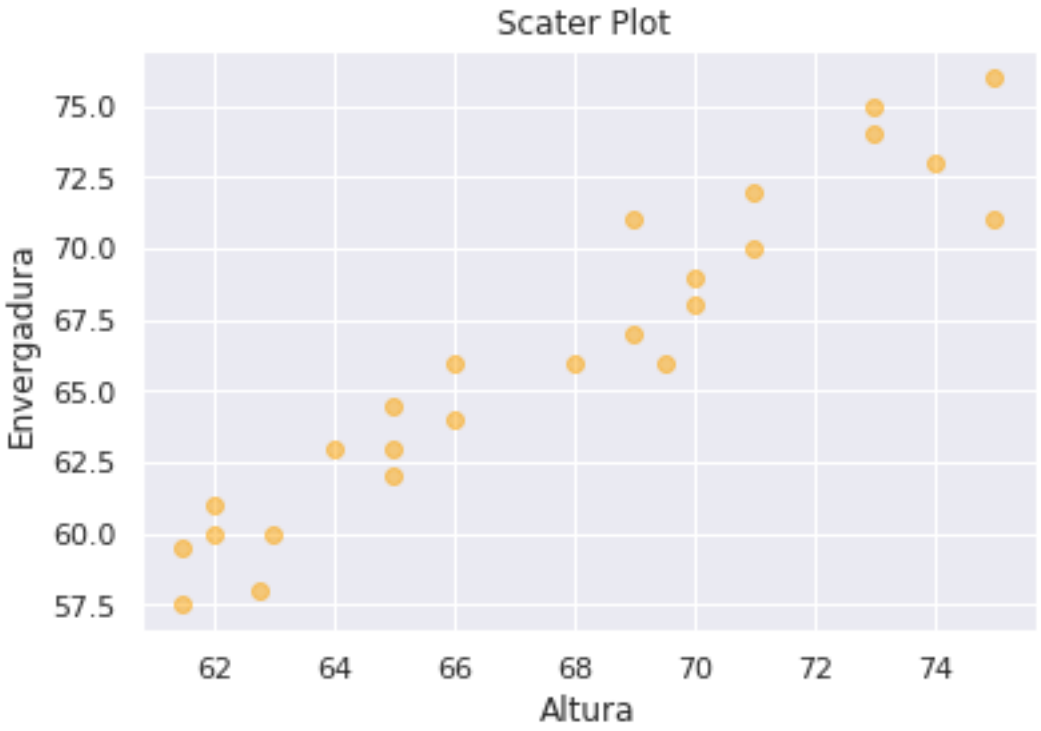
	ID	Age	Gender	GenderGroup	Glasses	GlassesGroup	Height	Wingspan	CWDistance	Complete	CompleteGroup	Score
0	1	56	F	1	Y	1	62.0	61.0	79	Y	1	7
1	2	26	F	1	Y	1	62.0	60.0	70	Y	1	8
2	3	33	F	1	Y	1	66.0	64.0	85	Y	1	7
3	4	39	F	1	N	0	64.0	63.0	87	Y	1	10
4	5	27	M	2	N	0	73.0	75.0	72	N	0	4

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.

0.1 Scatter plots

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

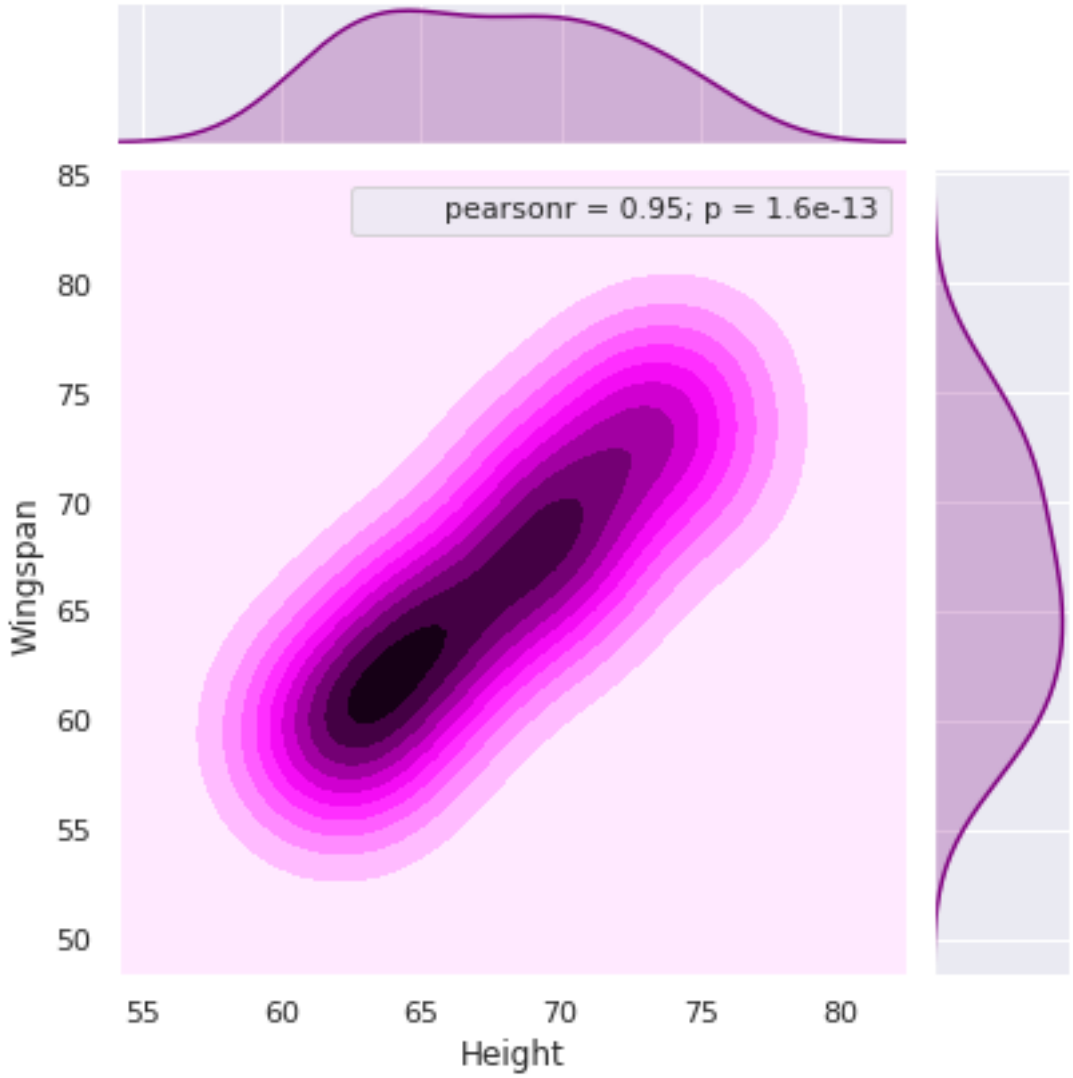
```
In [10]: # Haga un diagrama de dispersión de Seaborn con x = altura e y = envergadura usando s
         sns.set()
         sns.regplot(x="Height",y="Wingspan",data=df,fit_reg=False,scatter_kws={"alpha": 0.5},
         _ = plt.xlabel("Altura")
         _ = plt.ylabel("Envergadura")
         _ = plt.title("Scater Plot ")
         plt.show()
```




```
In [5]: from scipy import stats
```

```
In [9]: sns.set()
        _ = sns.jointplot(x="Height", y="Wingspan", kind='kde', data=df,color='purple').annota
        plt.show()
```

```
/opt/conda/lib/python3.6/site-packages/seaborn/axisgrid.py:1847: UserWarning: JointGrid annota
  warnings.warn(UserWarning(msg))
```

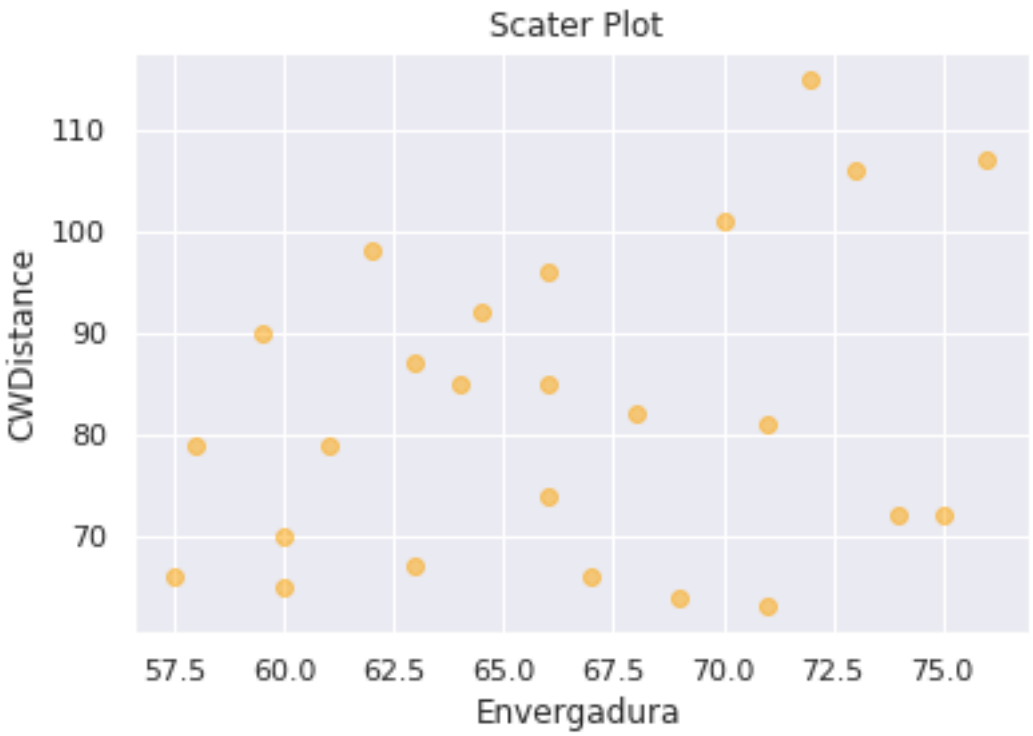


£Cómo describiría la relación entre 'Altura' y 'Envergadura'? Preguntas que puede hacer: * Es lineal? * £Hay valores atípicos? * £Son sus rangos similares o diferentes?

£De qué otra manera podrías describir la relación?

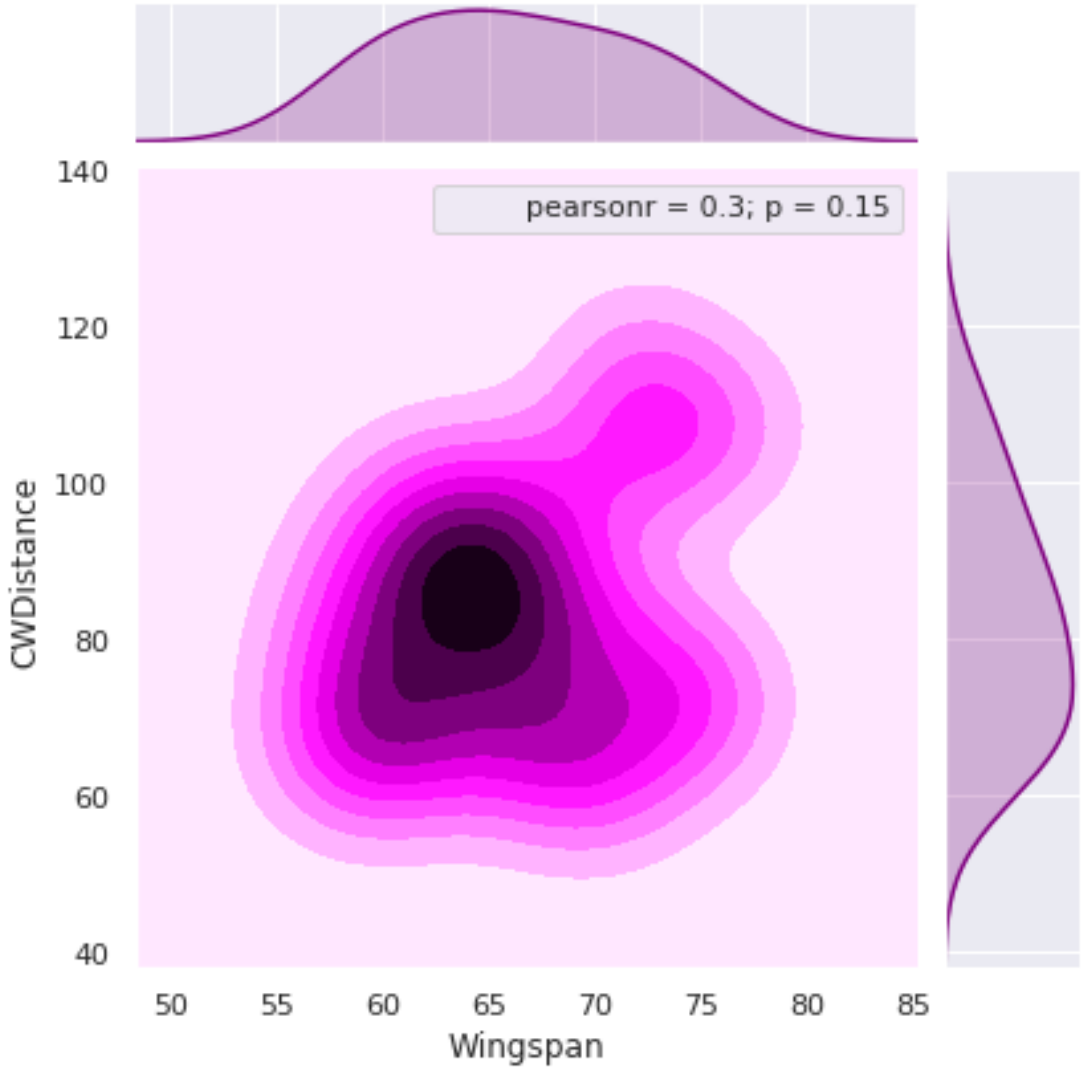
Ahora veamos dos variables que todavía no suponemos que tengan una relación sólida, 'Envergadura' y 'CWDistance'

```
In [11]: # Make a Seaborn scatter plot with x = wingspan and y = cartwheel distance
         sns.set()
         sns.regplot(x="Wingspan",y="CWDistance",data=df,fit_reg=False,scatter_kws={"alpha": 0
         _ = plt.xlabel("Envergadura")
         _ = plt.ylabel("CWDistance")
         _ = plt.title("Scater Plot ")
         plt.show()
```



In [13]:
```
sns.set()
_ = sns.jointplot(x="Wingspan", y="CWDistance", kind='kde', data=df,color='purple').a
plt.show()
```

```
/opt/conda/lib/python3.6/site-packages/seaborn/axisgrid.py:1847: UserWarning: JointGrid annota
  warnings.warn(UserWarning(msg))
```



£Cómo describiría la relación entre 'Wingspan' y 'CWDistance'?
* £Es lineal? * £Hay valores atípicos? * £Son sus rangos similares o diferentes?

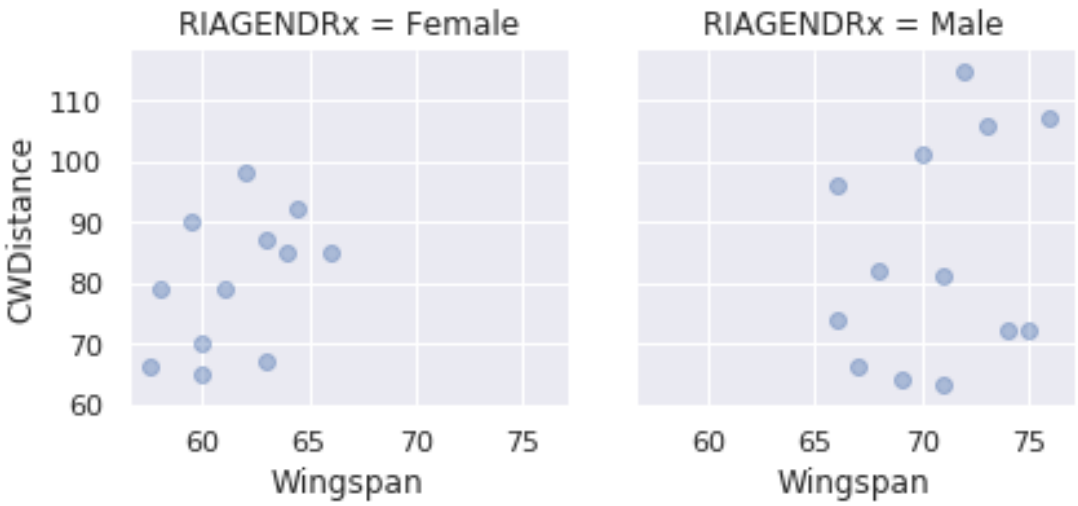
£De qué otra manera podrías describir la relación?

Hagamos el mismo diagrama que el anterior, pero ahora incluya 'Género' como esquema de color al incluir el argumento

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

en la función Seaborn

```
In [23]: df["RIAGENDRx"] = df.Gender.replace({"M": "Male", "F": "Female"})
         sns.FacetGrid(df, col="RIAGENDRx").map(plt.scatter, "Wingspan", "CWDistance", alpha=0
         plt.show()
```



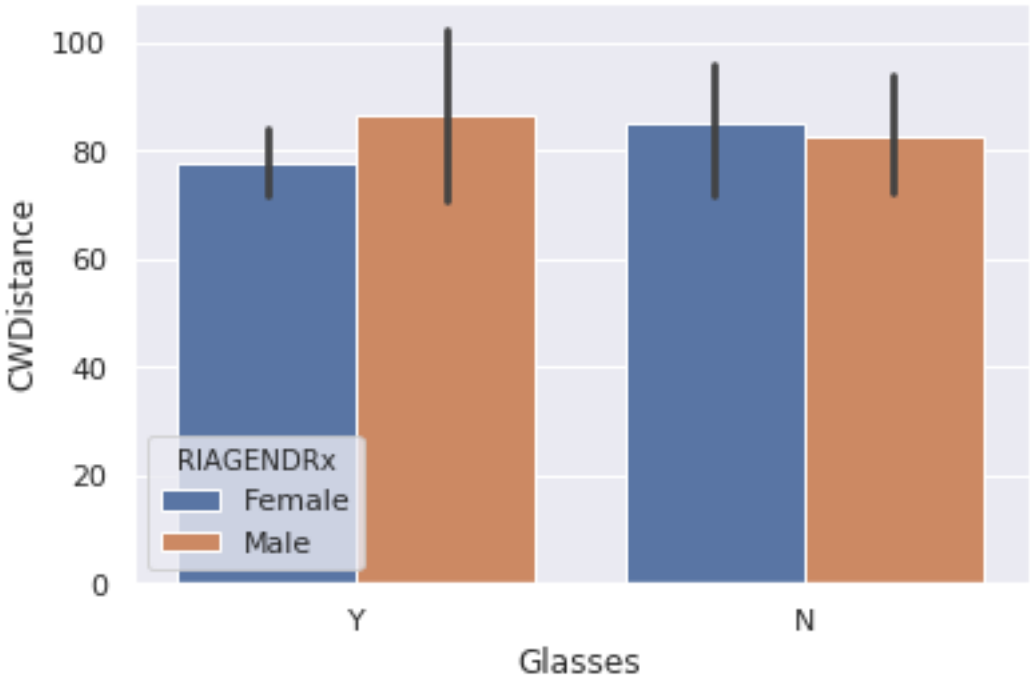

```
In [29]: def outlier_treatment(datacolumn):
             sorted(datacolumn)
             Q1,Q3 = np.percentile(datacolumn , [25,75])
             IQR = Q3 - Q1
             lower_range = Q1 - (1.5 * IQR)
             upper_range = Q3 + (1.5 * IQR)
             return lower_range,upper_range

In [33]: IQR_CWD = df["CWDistance"]
         IQR_Wing = df["Wingspan"]

         print(outlier_treatment(IQR_CWD))
         print(outlier_treatment(IQR_Wing))

(37.0, 125.0)
(48.5, 84.5)
```

£Esta nueva información sobre la trama cambia su interpretación de la relación entre 'Wingspan' y 'CWDistance'?

0.2 Barcharts

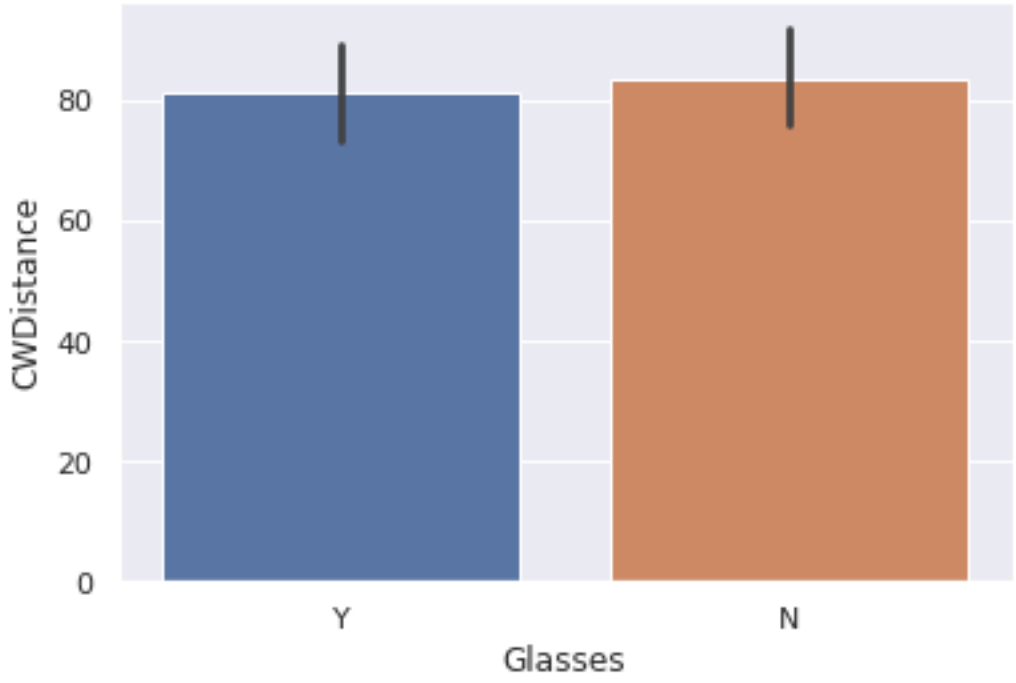
Now lets plot barplots of 'Glasses'

```
In [24]: # Make a Seaborn barplot with x = glasses and y = cartwheel distance
         ax = sns.barplot(x="Glasses", y="CWDistance", hue="RIAGENDRx", data=df)
```



£Qué puedes decir sobre la relación de 'Gafas' y 'CWDistance'?

```
In [34]: # Make the same Seaborn boxplot as above, but include gender for the hue argument
         ax = sns.barplot(x="Glasses", y="CWDistance", data=df)
```



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