Understanding the cartwheel data set

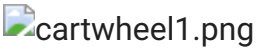
The notebook aims to undertand the content of the cartwheel data set.

Acknowledgments

- Data from https://www.coursera.org/ from the course "Understanding and Visualizing Data with Python" by University of Michigan

Cartwheel data set

1. A cartwheel



2. The dataset description
    - The dataset used here is an extension from the original cartwheel dataset from cursera
    - Total numer of observations: 28
    - Many observations/measurements/recordings of the characteristics/attributes/variables of cartwheel executions
    - Variables: Age, Gender, GenderGroup, Glasses, GlassesGroup, Height, Wingspan, CWDistance, ... (X variables)

Importing and inspecting the data

```
# Define where you are running the code: colab or local
RunInColab          = True     # (False: no  | True: yes)

# If running in colab:
if RunInColab:
    # Mount your google drive in google colab
    from google.colab import drive
    drive.mount('/content/drive')

    # Find location
    #!pwd
    #!ls
    #!ls "/content/drive/My Drive/Colab Notebooks/a01637205/NotebooksProfessor/datas

    # Define path del proyecto
    Ruta            = "/content/drive/MyDrive/Colab Notebooks/a01637205/NotebooksPro

else:
    # Define path del proyecto
    Ruta            = ""
```

```
Mounted at /content/drive
```

```
# Import the packages that we will be using
import matplotlib.pyplot as plt
import pandas as pd

# Dataset url
url = Ruta + "cartwheel/cartwheel.csv"

# Load the dataset
dataset = pd.read_csv(url)


# Print the dataset
dataset
```

	ID	Age	Gender	GenderGroup	Glasses	GlassesGroup	Height	Wingspan	CWDista
0	1	56.0	F	1	Y	1	62.00	61.0	
1	2	26.0	F	1	Y	1	62.00	60.0	
2	3	33.0	F	1	Y	1	66.00	64.0	
3	4	39.0	F	1	N	0	64.00	63.0	
4	5	27.0	M	2	N	0	73.00	75.0	
5	6	24.0	M	2	N	0	75.00	71.0	
6	7	28.0	M	2	N	0	75.00	76.0	
7	8	22.0	F	1	N	0	65.00	62.0	
8	9	29.0	M	2	Y	1	74.00	73.0	
9	10	33.0	F	1	Y	1	63.00	60.0	
10	11	30.0	M	2	Y	1	69.50	66.0	
11	12	28.0	F	1	Y	1	62.75	58.0	
12	13	25.0	F	1	Y	1	65.00	64.5	
13	14	23.0	F	1	N	0	61.50	57.5	
14	15	31.0	M	2	Y	1	73.00	74.0	
15	16	26.0	M	2	Y	1	71.00	72.0	
16	17	26.0	F	1	N	0	61.50	59.5	
17	18	27.0	M	2	N	0	66.00	66.0	
18	19	23.0	M	2	Y	1	70.00	69.0	
19	20	24.0	F	1	Y	1	68.00	66.0	
20	21	23.0	M	2	Y	1	69.00	67.0	
21	22	29.0	M	2	N	0	71.00	70.0	
22	23	25.0	M	2	N	0	70.00	68.0	
23	24	26.0	M	2	N	0	69.00	71.0	
24	25	23.0	F	1	Y	1	65.00	63.0	
25	26	28.0	M	2	N	0	75.00	76.0	
26	27	24.0	M	2	N	0	78.40	71.0	
27	28	25.0	M	2	Y	1	76.00	73.0	
28	29	32.0	F	1	Y	1	63.00	60.0	
29	30	38.0	F	1	Y	1	61.50	61.0	

30	31	27.0	F	1	Y	1	62.00	60.0
31	32	33.0	F	1	Y	1	65.30	64.0
32	33	38.0	F	1	N	0	64.00	63.0
33	34	27.0	M	2	N	0	77.00	75.0
34	35	24.0	F	1	N	0	67.80	62.0
35	36	27.0	M	2	N	0	68.00	66.0
36	37	25.0	F	1	Y	1	65.00	64.5
37	38	26.0	F	1	N	0	61.50	59.5
38	39	31.0	M	2	Y	1	73.00	74.0
39	40	30.0	M	2	Y	1	69.50	66.0
40	41	23.0	F	1	N	0	70.40	71.0
41	42	26.0	M	2	Y	1	73.50	72.0
42	43	28.0	F	1	Y	1	72.50	72.0
43	44	26.0	F	1	Y	1	72.00	72.0
44	45	30.0	F	1	Y	1	66.00	64.0
45	46	39.0	F	1	N	0	64.00	63.0
46	47	27.0	M	2	N	0	78.00	75.0
47	48	24.0	M	2	N	0	79.50	75.0
48	49	28.0	M	2	N	0	77.80	76.0
49	50	30.0	F	1	N	0	74.60	NaN
50	51	NaN	M	2	N	0	71.00	70.0
51	52	27.0	M	2	N	0	NaN	71.5

Next steps: View recommended plots

```
# Print the number of rows
Nrows = dataset.shape[0]
Nrows
```

```
52
```

```
# Print the number of columns
Ncols = dataset.shape[1]
Ncols
```

```
12
```

Data types

```
types = dataset.dtypes

print (types)
```

```
ID                int64
Age             float64
Gender           object
GenderGroup       int64
Glasses          object
GlassesGroup      int64
Height          float64
Wingspan        float64
CWDistance        int64
Complete         object
CompleteGroup   float64
Score             int64
dtype: object
```

Activity: work with the iris dataset

1. Load the iris.csv file in your computer and understand the dataset
2. How many observations (rows) are in total?
3. How many variables (columns) are in total? What do they represent?
4. How many observations are for each type of flower?
5. What is the type of data for each variable?
6. What are the units of each variable?

Parte 1

```
# Dataset url
url2 = Ruta + "iris/iris.csv"

# Load the dataset
dataset2 = pd.read_csv(url2, header = None, names = ["sepal_length", "sepal_width",


dataset2
```

	sepal_length	sepal_width	petal_length	petal_width	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 5 columns

Next steps: View recommended plots

Parte 2

```
# Print the number of rows
Nrows2 = dataset2.shape[0]
print("There are", Nrows2, "observations/rows in total")
```

```
There are 150 observations/rows in total
```

Parte 3

```
# Print the number of columns
Ncols2 = dataset2.shape[1]
print("There are", Ncols2, "variables/columns in total")
print("They represent the different data there is for each flower: ")
for col in dataset2.columns:
    print(col)
```

```
There are 5 variables/columns in total
They represent the different data there is for each flower:
sepal_length
sepal_width
petal_length
petal_width
```

Parte 4

```
nObservations = dataset2.groupby('class').size()
print("There are 50 observations for each type of flower: ")
nObservations
```

```
There are 50 observations for each type of flower:
class
Iris-setosa        50
Iris-versicolor    50
Iris-virginica     50
dtype: int64
```

Parte 5

```
typeData = dataset2.dtypes
print("The type of data for each variable is: ")
print (typeData)
```

```
The type of data for each variable is:
sepal_length    float64
sepal_width     float64
```