

ID Barbe : 403715  
ID Gaucher : 403783

Barbe Victor – International Student  
Pierre-louis Gaucher – International Student

## Assignment 1: Data visualization techniques

For each type of the graphs, present a small summary indicating the following parameters:

1. What type of input data or table is required?
2. How do they process the input
3. What is/are the results shown in the graph
4. What do we learn from it?
5. Provide an example of input data, resulting data, ...

### A. Line plot

To make a line plot, the function used is to make a graph of a function. The input used must be a list of points which are array likes or scalar. Then many other parameters can be addressed like color, shape, ect.

This graph gives us the relation between two points of a function. We can thus get the resulting value of Y with an input of X, and the link between the input of a function and its output (is it linear, exponential, is there any link, ...).

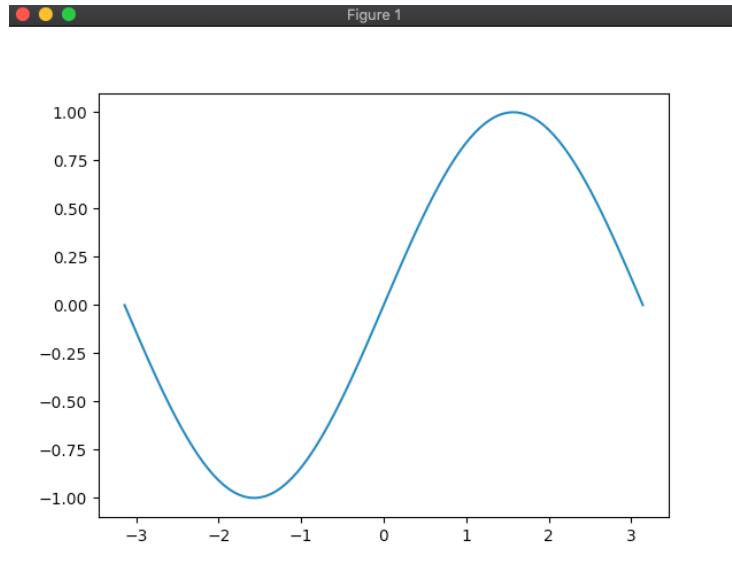
Here as an example, we could plot a sinusoidal function. X would be a value between -1 and 1, and Y would be the corresponding value of  $\sin(x)$ .

```
import matplotlib.pyplot as plt
import numpy as np

x = np.linspace(-np.pi,np.pi,100)
y = np.sin(x)

plt.plot(x,y)
plt.show()
```

ID Barbe : 403715  
ID Gaucher : 403783



Then, we can make the graph more user friendly by adding labels, names, colors, ...

## B. Bar chart

A bar chart is a graph that presents data ranked by categories, using rectangular bars that are proportional to the value of this category. They can be used to compare the importance of different parameters, with one axis presenting the different categories and another their importance.

This graph needs at least two inputs, the categories and their values. The values must be numbers, the input can be a vector or a matrix. Depending on that, there will be simple or stacked bars.

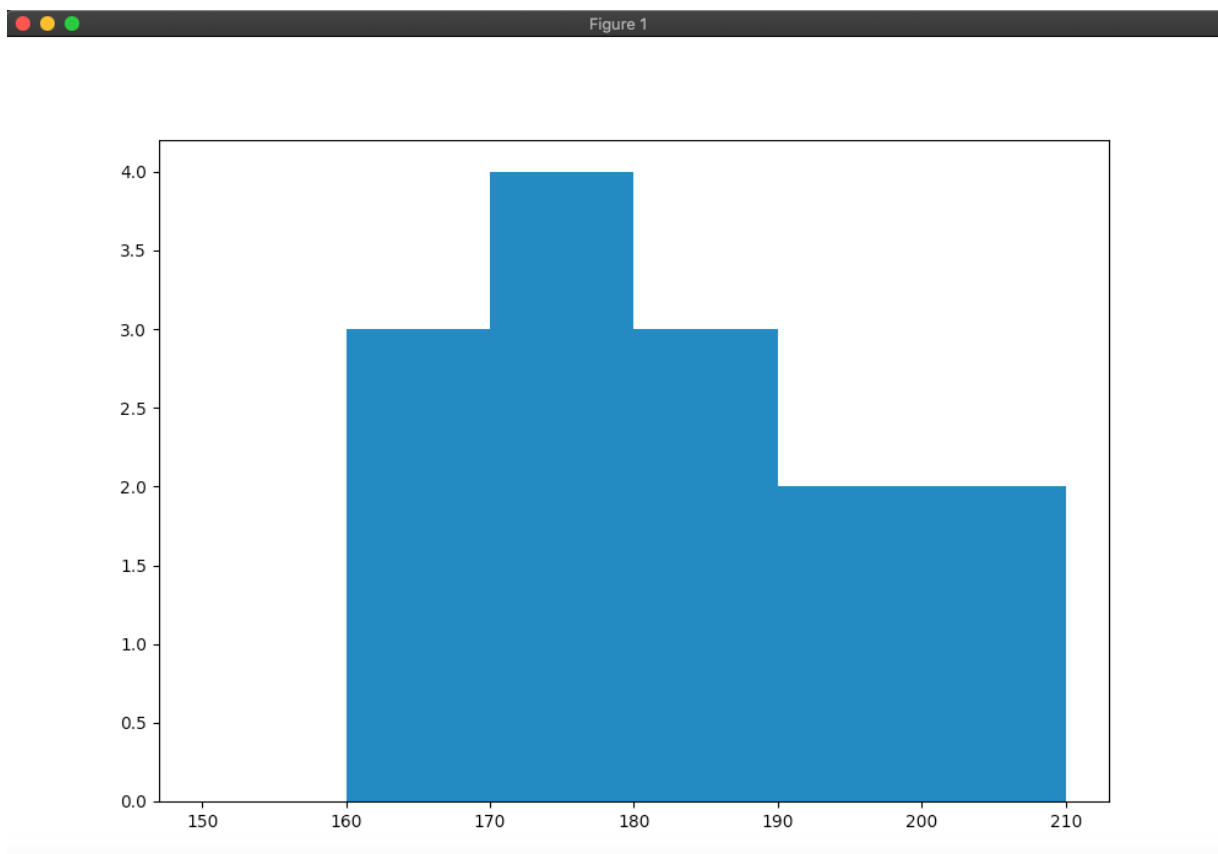
Here is an example of this graph with stacked bars, representing the survivals of the fall of the titanic depending on the class of the passengers:



## C. Histogram

A histogram is a graph which is showing the frequency distributions. The histograms need as an input an array of data and can also take a second input that will define the different number of cuts that should be made.

The goal of this graph is to understand the distribution of the values of a function. As an example, we could take 100 persons and make a histogram of their height to understand the repartition.



```
# Creating dataset
a = np.array([175,180,160,161,190,200,205,182,173,174,186,169,191,176])

# Creating histogram
fig, ax = plt.subplots(figsize=(10, 7))
ax.hist(a, bins = [150, 160, 170, 180, 190, 200, 210])

# Show plot
plt.show()
```

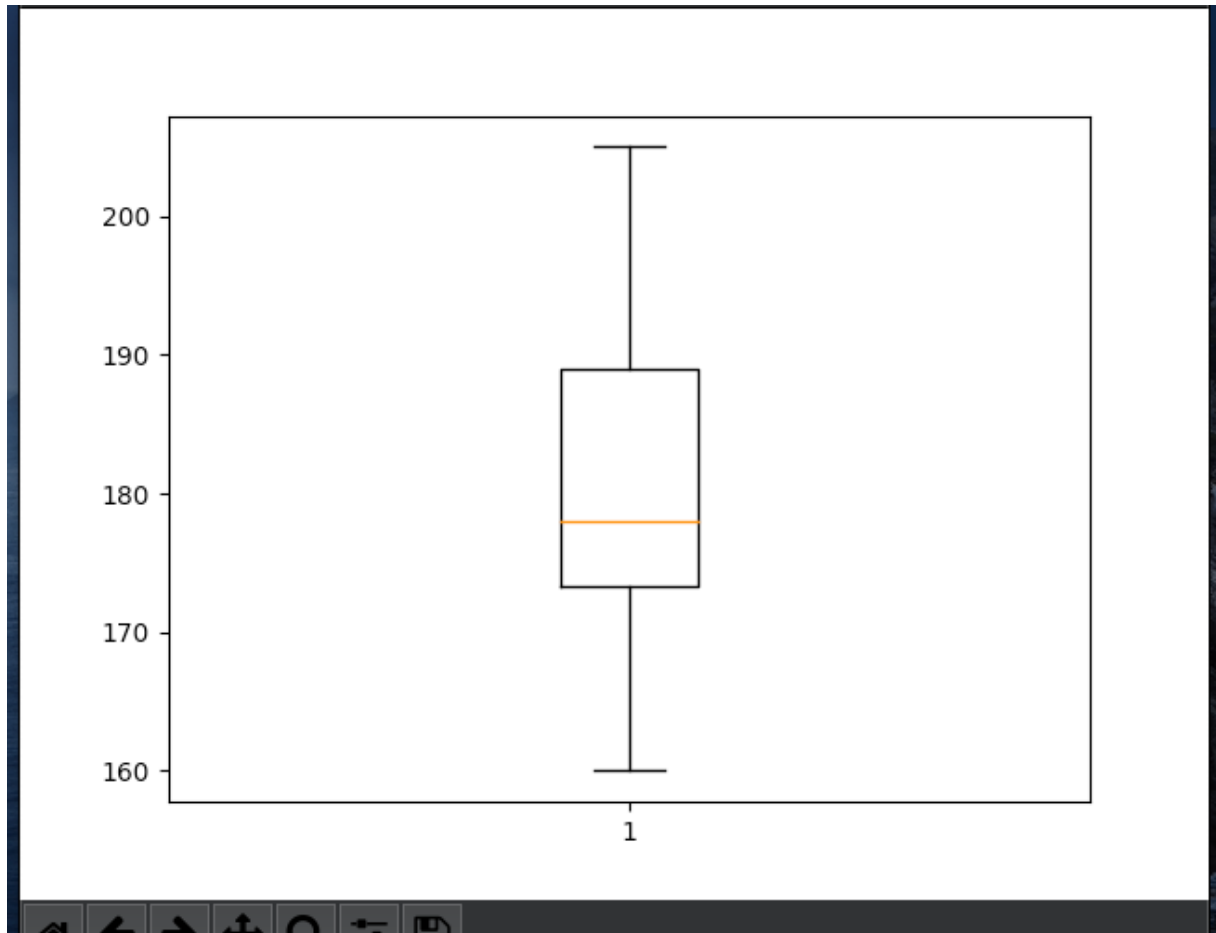
## D. Box and whiskers

ID Barbe : 403715

ID Gaucher : 403783

The box Plot is used to show a “summary of a set of values. It will show the minimum, maximum and the different quartiles. A box goes from the 1<sup>st</sup> to the 3<sup>rd</sup> quartile. There is also inside the box a line that shows the median. This plot can also be used to understand the frequency distribution of a data set, where the X-axis shows the data and the Y axis the frequency distribution.

Here we can use the same data set for the different heights of people and plot it:



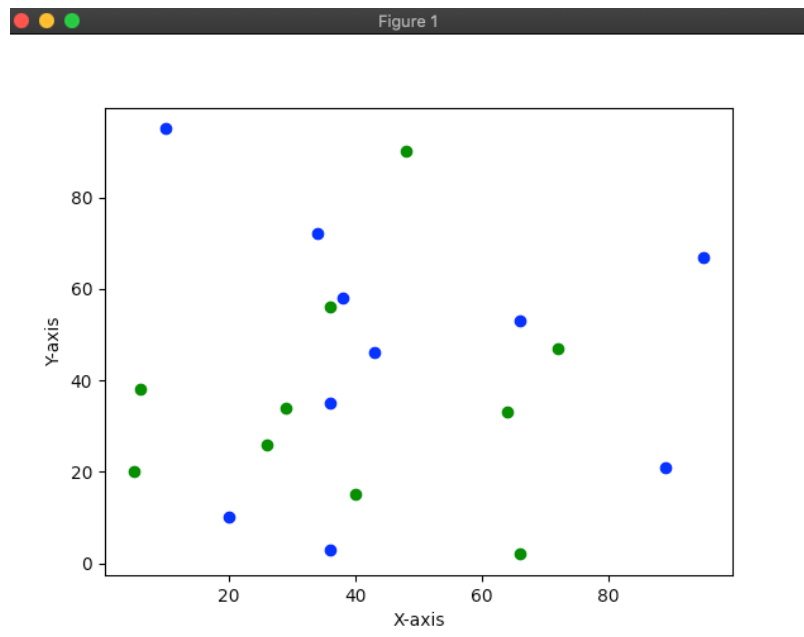
The lowest height is 160cm, the median is at 177cm, and the 3<sup>rd</sup> quartile at 190cm.

## E. Scatter plot

The scatter plot is used to observe the relation between two variables and represents them using dot. The inputs must be number which are linked by pair (X and Y). It is however possible to represent more than one dataset on one scatter plot. Again, it is then possible to change the aesthetics by adding different colors for each dataset, ect...

ID Barbe : 403715

ID Gaucher : 403783



```
import matplotlib.pyplot as plt
import numpy as np

x_set1 = [89, 43, 36, 36, 95, 10, 66, 34, 38, 20]
y_set1 = [21, 46, 3, 35, 67, 95, 53, 72, 58, 10]

# dataset2
x_set2 = [26, 29, 48, 64, 6, 5, 36, 66, 72, 40]
y_set2 = [26, 34, 90, 33, 38, 20, 56, 2, 47, 15]

plt.scatter(x_set1, y_set1, c="BLUE")

plt.scatter(x_set2, y_set2, c="GREEN")

plt.xlabel("X-axis")
plt.ylabel("Y-axis")
plt.show()
```

## F. Scatter plot matrix

A scatter plot matrix is used to see the bivariate relation between a combination of variable, in comparison with the scatter plot that shows a simple relation. Scatter plot matrix are often used to determine if there is a linear correlation between two continuous variables.

Here we can see an example of a scatter matrix presenting the relation between different attributes. We can see as an example that attribute 2 and 1 seem to be correlated linearly, but attribute 4 and 3 don't.

ID Barbe : 403715

ID Gaucher : 403783

