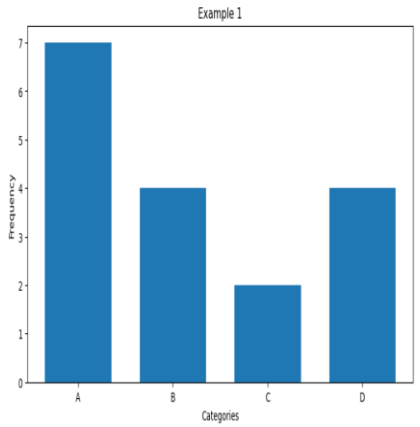
FREQUENCY TABLES AND HISTOGRAMS

We can tabulate both categorical and numeric data

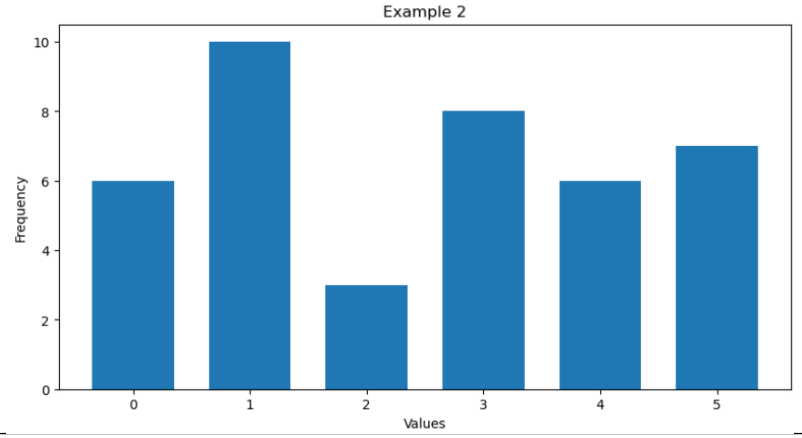
raw data			
A	B	A	A
A	E	A	D
B	B	D	A
A	C	E	D
D	C	B	

Category	Frequency
A	7
B	4
C	2
D	4



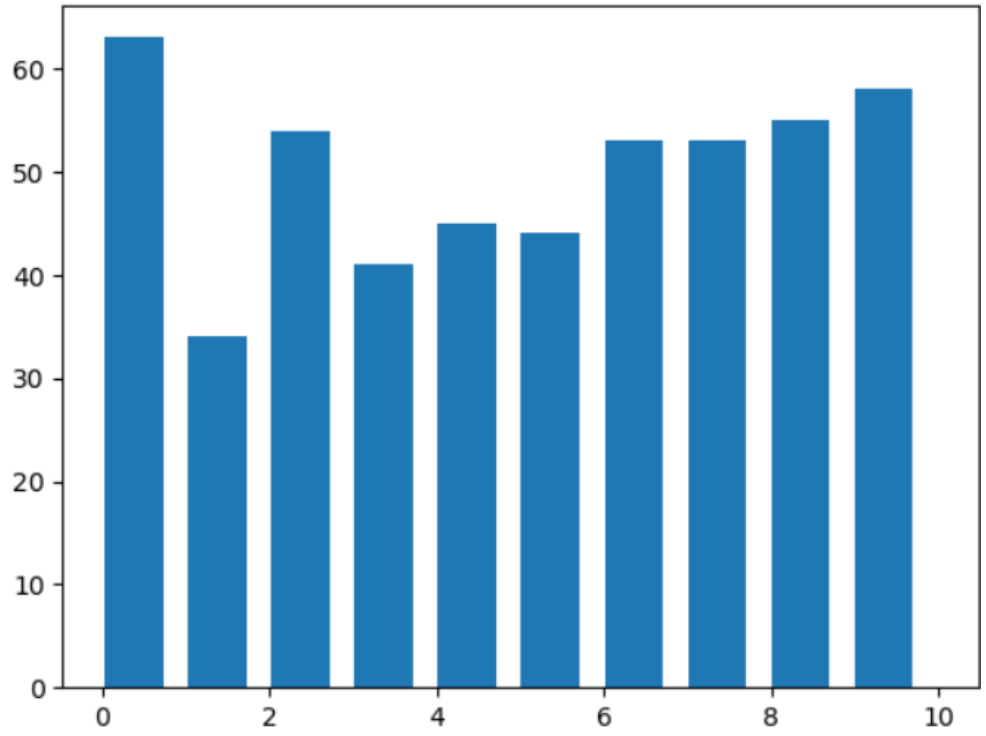
raw data (discrete)			
1	4	3	4
0	2	1	0
4	1	0	5
5	5	2	1
5	1	1	5
5	4	3	4
2	3	0	3
0	4	0	3
1	1	3	1
3	5	1	3

Value	Frequency
0	6
1	10
2	3
3	8
4	6
5	7



For continuous data (or discrete with many values) we need to create ranges, or intervals, and count how many numbers belong to each of the intervals.

The graph (bar chart type) of such frequency table is called a **Histogram**



Frequency table with tabulated data

Range (exclude upper limit)	**frequency**
between 0 and 1	58
between 1 and 2	48
between 2 and 3	44
between 3 and 4	57
between 4 and 5	56
between 5 and 6	48
between 6 and 7	44
between 7 and 8	37
between 8 and 9	59
between 9 and 10	49

Total: 500 values

Practice in Python

Now, we will complete the following activity together.

1. We will create a code that takes some categorical data and create a frequency table.

2. We will create a distribution given some continuous data, and the number of bins that we want to use.

- You will need the file **data_lecture3.csv**