

UNIT 1

Graphs and Charts

Presented By :

Aditya Joshi

Assistant Professor

Graphic Era Deemed to be University, Dehradun

Graphs and Charts

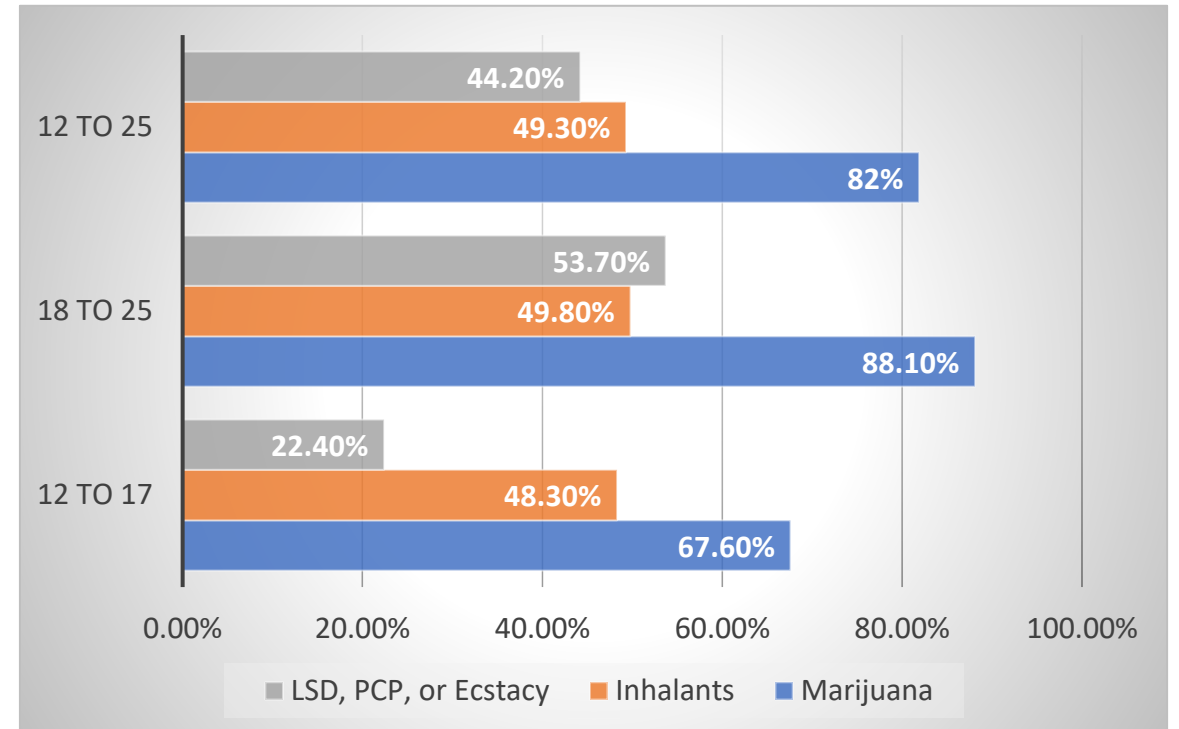
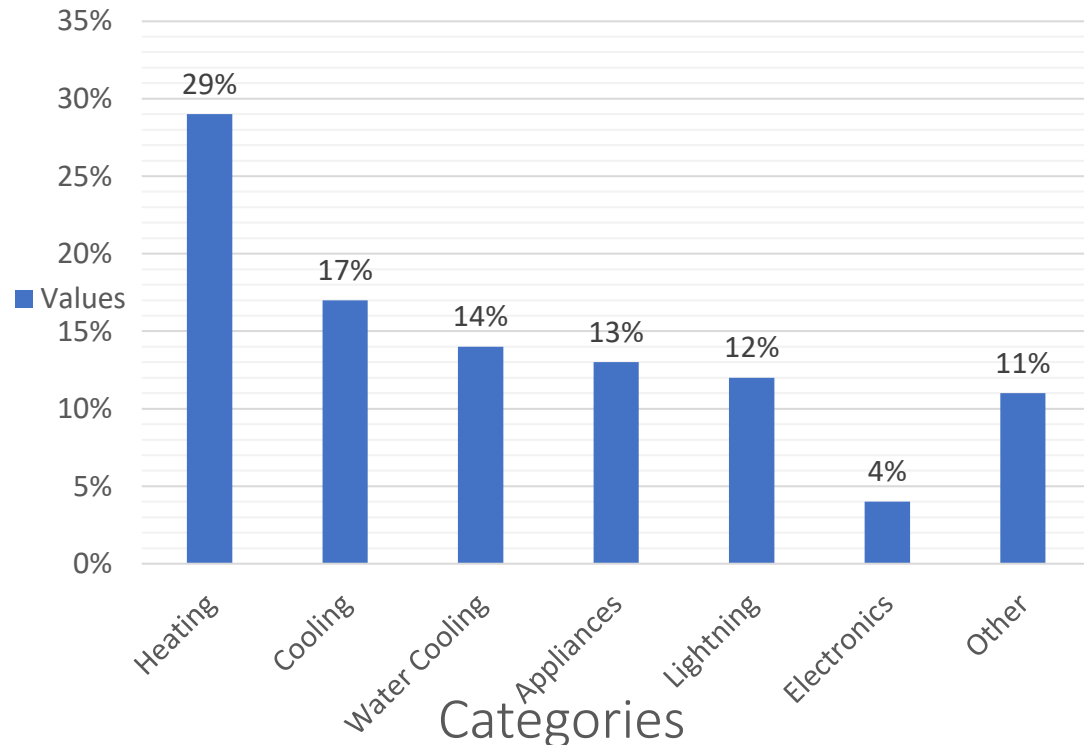
- There are number of ways to display summary data using graphs and charts some of the most common ones are
- Bar charts
- Histogram
- Frequency distribution table
- Boxplot
- Scatter plot
- Pie chart

Bar charts

A bar chart is a graph with **rectangular bars**, the graph normally compares different categories.

Although the graphs can be plotted **vertically** (bar standing up) or **horizontally** (bar laying from left to right) usual type is vertical.

The horizontal **x axis** represents categories and vertical **y axis** represents value for those categories.



- Make a bar graph that represent exotic pet ownership in India there are
8000000 fish
1500000 rabbits
1300000 dogs
1000000 poultry
900000 sheep

Number the y axis : you know how many pets

x axis : label each bar in x axis

Histogram

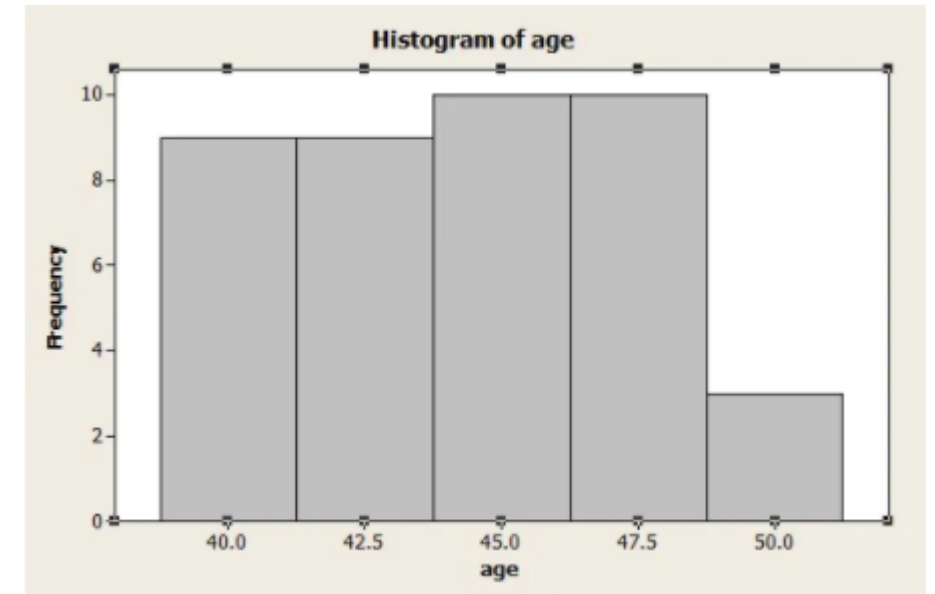
It is similar to bar charts they are a way to display counts of data. A bar graph charts actual counts against categories. The height of bar indicates the number of items in that category. A histogram displays the same categorical values in “bins”

A bin shows how many data points are with in range (an interval) you choose the range that best fits your data.

There are no set rules about how many bins. **The rule of thumb is 5-20 bins.**

More than 20 bins hard to read

If less then 5 bins graph will have little meaning



Height of a bar: unlike a bar chart, the area of a bar in histogram represents the frequency, not a height. The frequency is calculated by multiplying the width of the bin by the height.

Properties:

Quantitative data

No gaps

Bar width (bin size or class size)

Y axis corresponds to frequency or counts

Q. how are grades of 15 students

88, 48, 60, 51, 57, 85, 69, 75, 97, 72, 71, 79, 65, 63, 73 build a histogram from the data

1. Break the range of values into intervals called bins or classes
2. Find the lowest and highest value

48 and 97

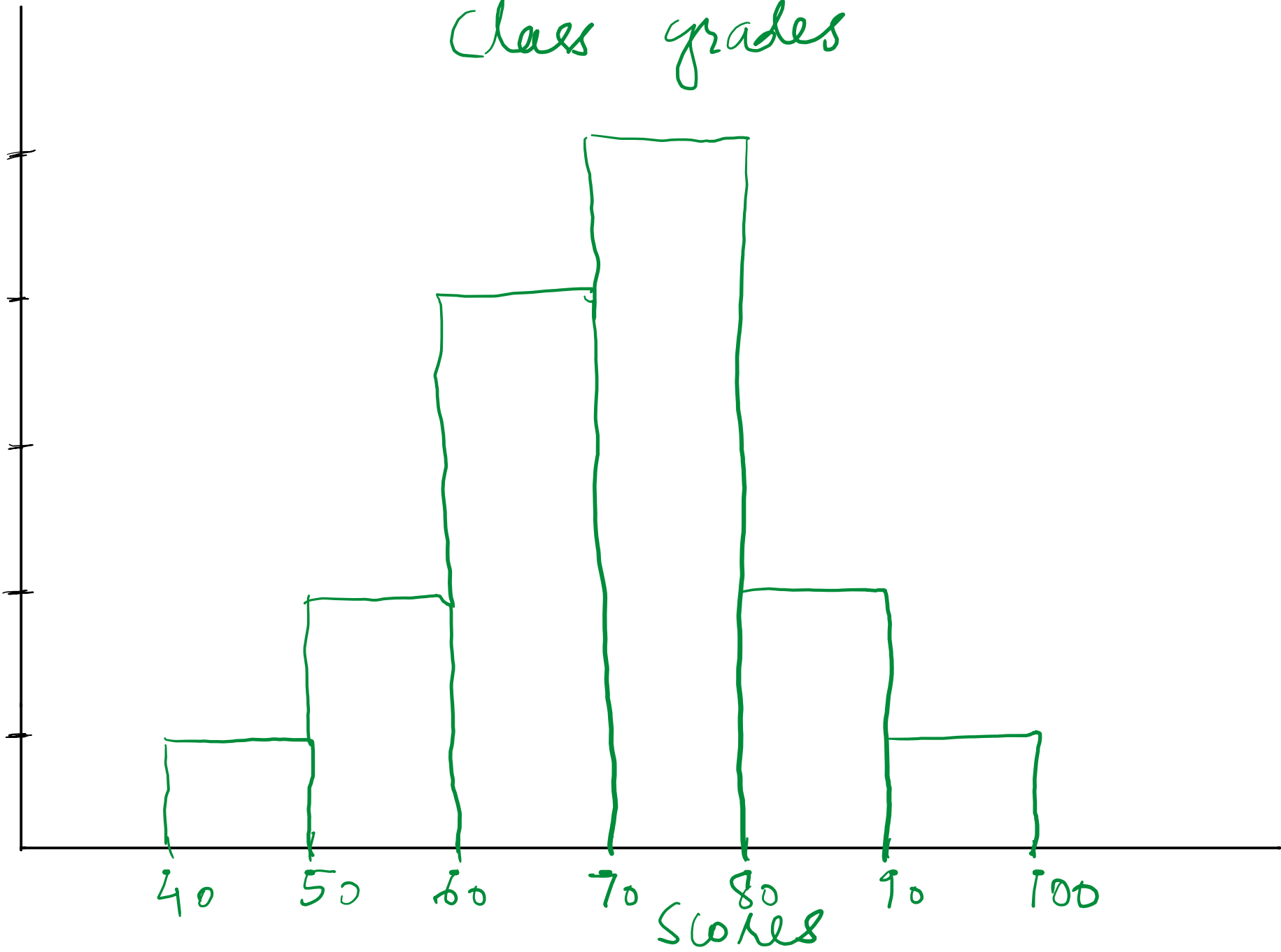
$$\text{Bins} = \frac{97-48}{5} = 9.8 \quad \text{size of bin may be 10}$$

We can start bins with 40 and stop at 100

Scores	Counts
[40,50)	1
[50,60)	2
[60,70)	4
[70,80)	5
[80,90)	2
[90,100)	1

Counts / frequency

class grades



Frequency distribution table

Frequency tells you how often something happened the frequency of an observation tells you the number of times the observation occurs in the data.

For example

The frequency of number 9 is 5 in

1, 2, 3, 4, 6, 9, 9, 8, 5, 1, 1, 9, 9, 0, 6, 9

How to make frequency distribution table: **using tally marks** are used to make frequency distribution table

Q. You survey a number of house holds and find out how many pets they own the results are

3, 0, 1, 4, 4, 1, 2, 0, 2, 2, 0, 2, 0, 1, 3, 1, 2, 1, 1, 3

To make a frequency distribution table write the categories in one column

No of pets(x)	Tally Marks	Frequency(f)
0		4
1		6
2		5
3		3
4		2

Including classes

118, 123, 124, 125, 127, 128, 129, 130, 130, 133, 136, 138, 141, 142, 149, 150, 154

Choosing classes

1. figure out how many classes categories you need
 - Pick between 5 and 20 classes for the list of above IQ scores we pick 5 classes
 - we have 20 items choose 5 (4 items per category) not 20 which would give you 1 items per category

Sorting the Data

2. Subtract the min value with max value $154 - 118 = 36$
3. Divide answer by class selected in step 1
 $36 / 5 = 7.2$
4. Round the number to whole number i.e. 8

5. Write down the lowest value 118

6. Add class width from step 4 to step 5 to get next lower class limit

$$118+8=126$$

7. Repeat the step for other min data values until we reach the 5 minimum data values because in step 1 we choose the class size is 5

118

$$126(118+8)$$

$$134(126+8)$$

$$142(134+8)$$

$$150(142+8)$$

8. Write down the upper class limits. These are the highest values that can be in the category, so in most cases you can subtract 1 from the class width and add to the minimum data value

$$118+(8-1)=125$$

$$126+(8-1)=133$$

$$134+(8-1)=141$$

$$142+(8-1)=149$$

$$150+(8-1)=157$$

Finishing up the table

9. Add a second column for number of items in each class

IQ	Numbers
[118-125]	4
[126-133]	6
[134-141]	3
[142-149]	2
[150-157]	2

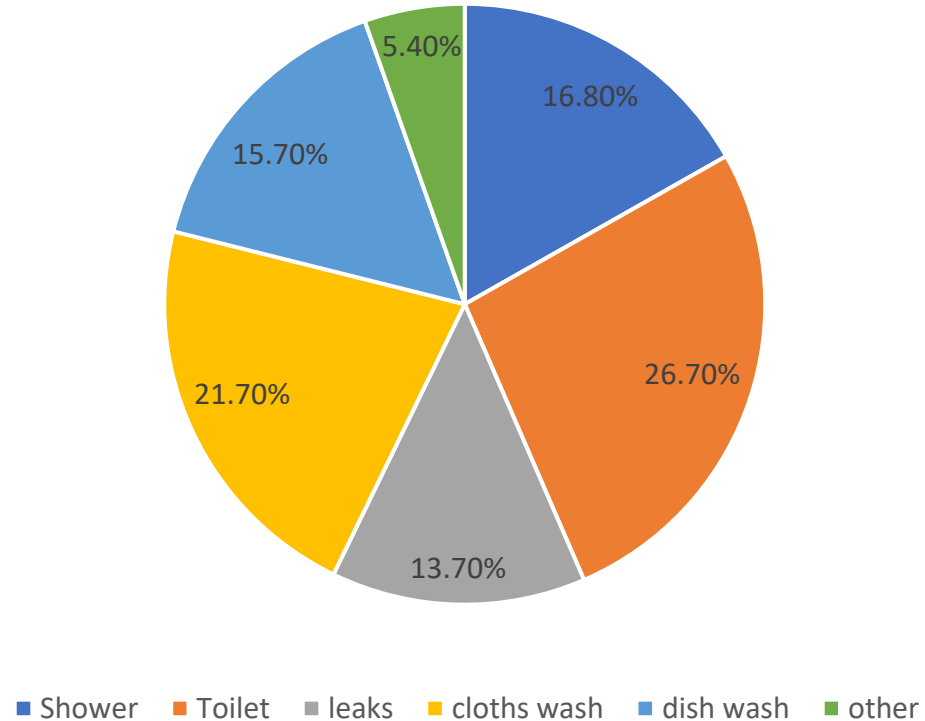
Pie chart

A pie chart is a type of graph that displays data in a circular graph. The pieces of the graph are proportional to the fraction of whole in each category.

The entire “pie” represent 100% of a whole.

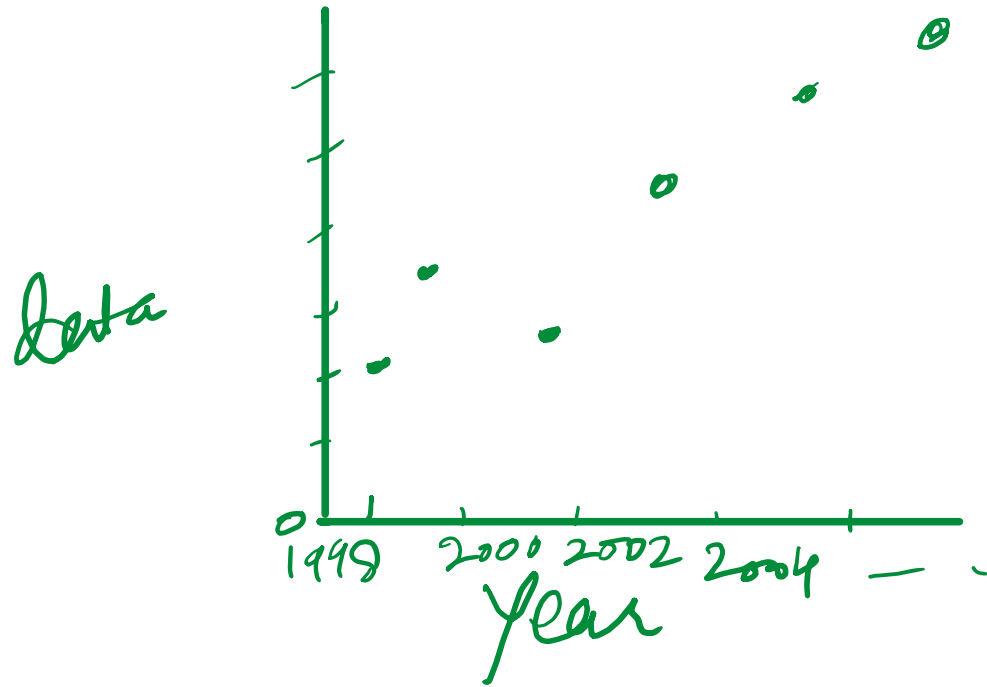
While the pie “slices” represent portions of the whole.

How much water do we use



Scatter plot

Similar to line graphs a line graph uses a line on a x-y axis to plot a continuous function. While scatter plot use dots to represent individual pieces of data



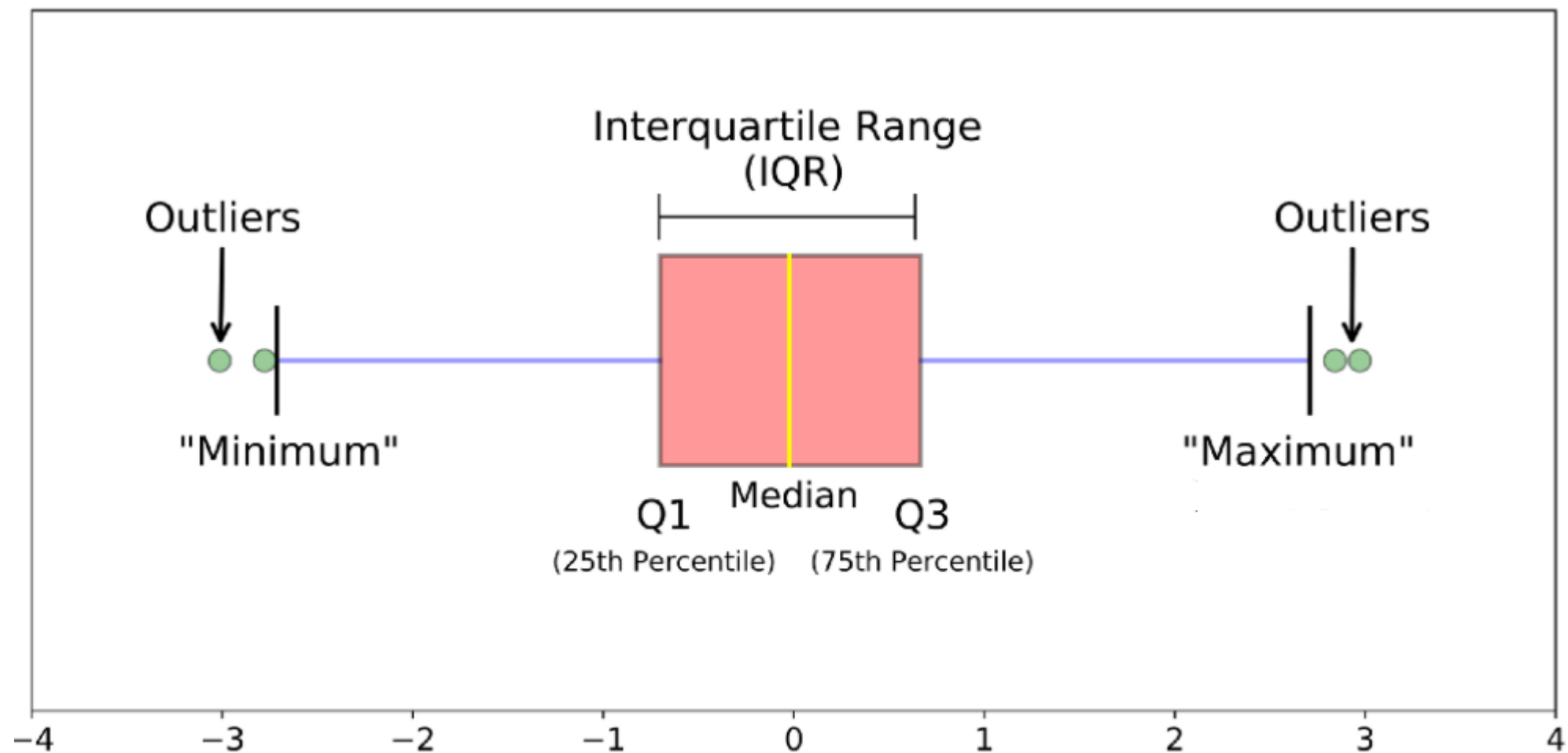
Box Plot

A box plot also called **whisker plot** is a way to show the spread and centers of a dataset.

Measures of a spread include interquartile range and the mean of dataset

Measures of center include the mean or average and median (the middle of dataset)

A boxplot is a standardized way of displaying the distribution of data based on a five number summary (**“minimum”, first quartile (Q1), median, third quartile (Q3), and “maximum”**). It can tell you about your outliers and what their values are. It can also tell you if your data is symmetrical, how tightly your data is grouped, and if and how your data is skewed.



Different parts of a boxplot

- **median (Q2/50th Percentile)**: the middle value of the dataset.
- **first quartile (Q1/25th Percentile)**
- **third quartile (Q3/75th Percentile)**
- **interquartile range (IQR)**: 25th to the 75th percentile.
- **whiskers (shown in blue)**
- **outliers (shown as green circles)**
- **“maximum”** (the largest number in the data set), shown at the far right of the box.
- **“minimum”** (the smallest number in the data set). The minimum is shown at the far left of the chart, at the end of the left “whisker.”