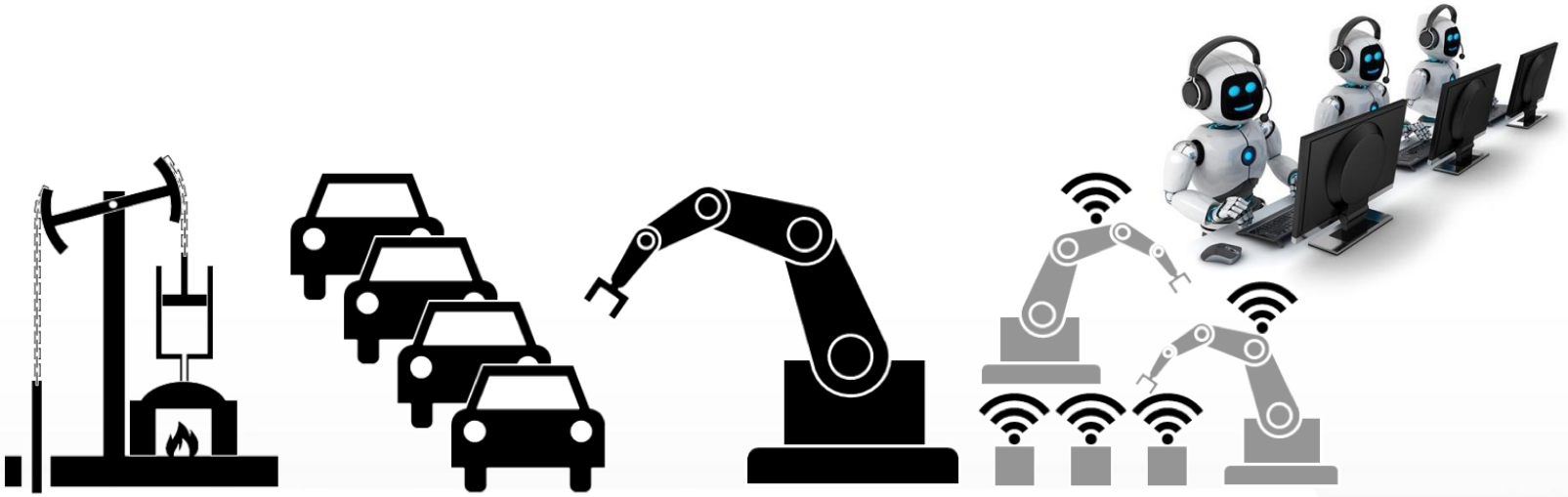
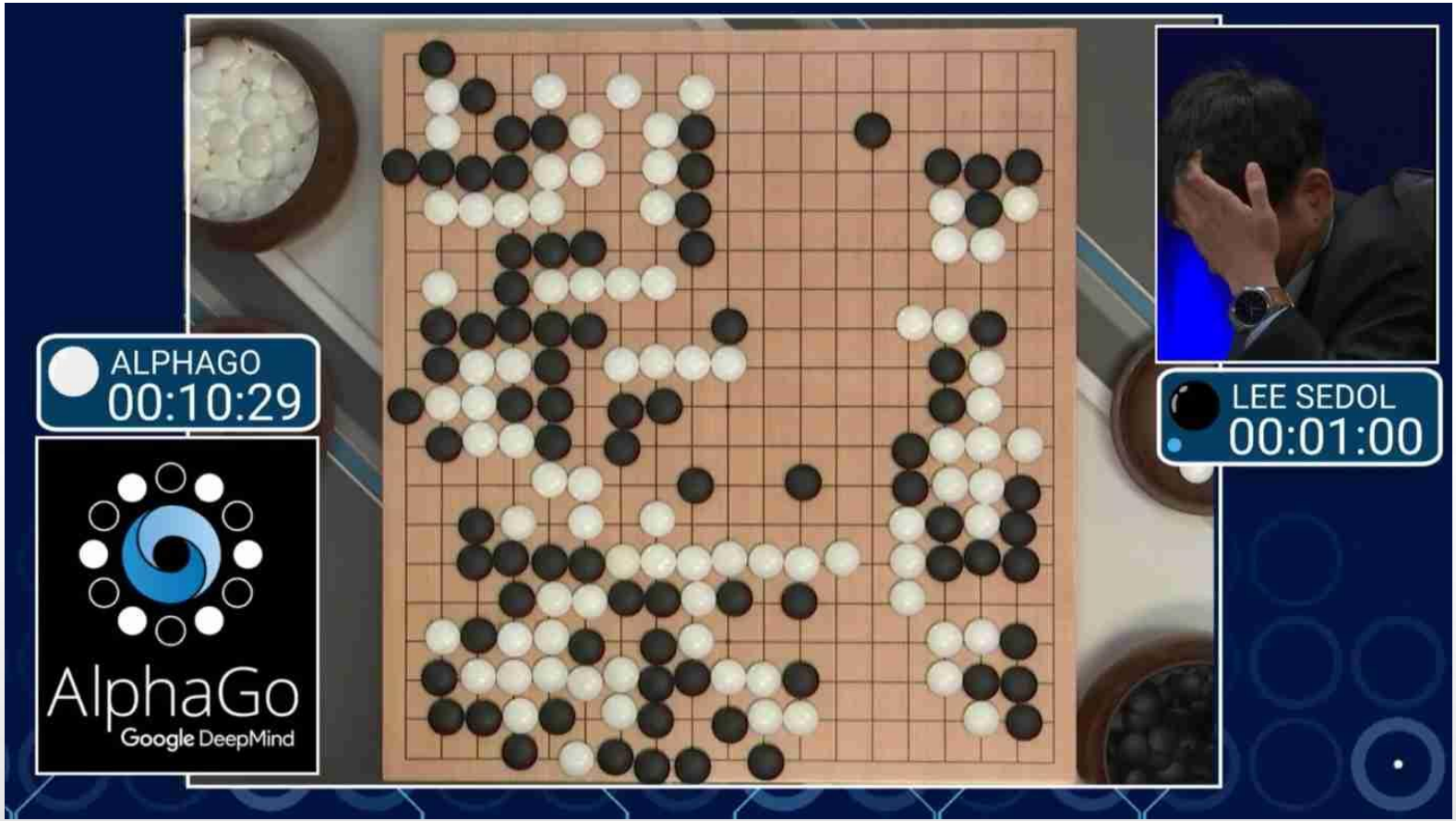
Practical Machine Learning With Python



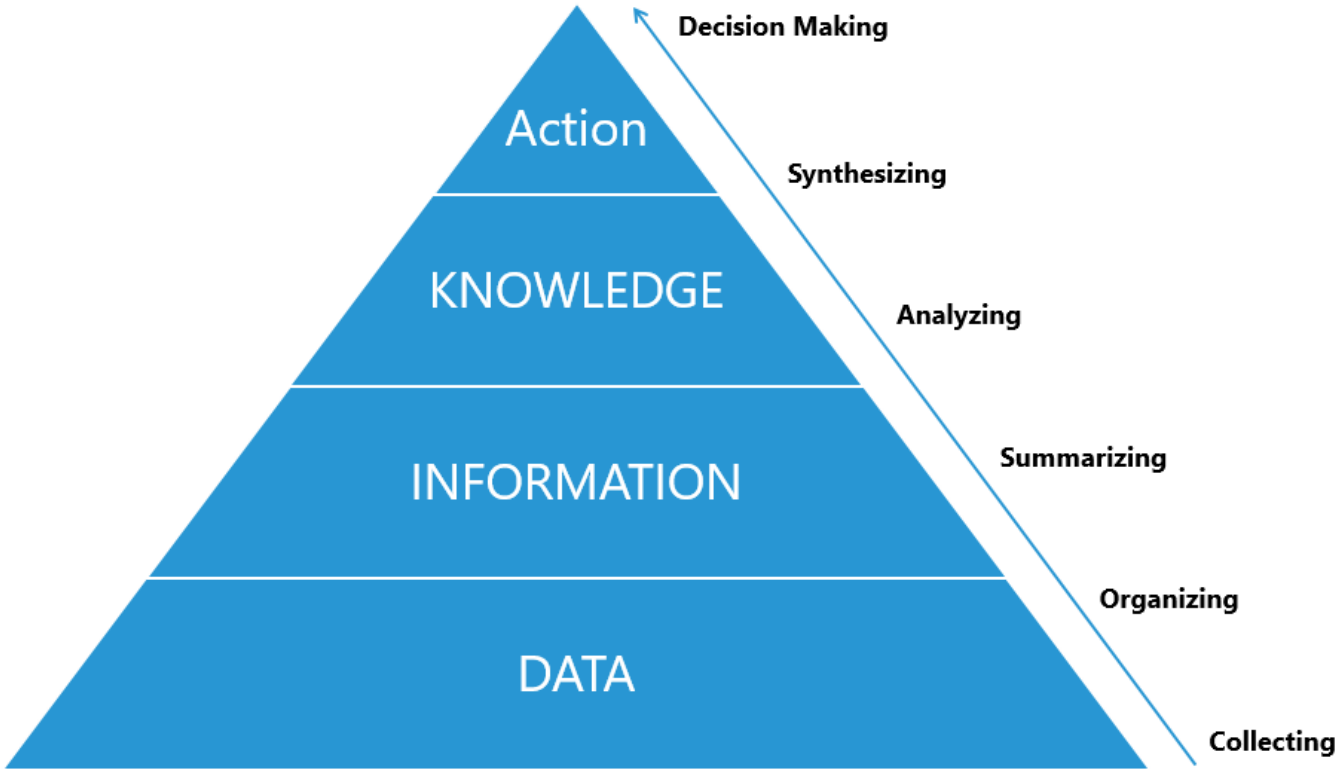
1st	2nd	3rd	4th
Mechanization, water power, steam power	Mass production, assembly line, electricity	Computer and automation	Cyber Physical Systems

“Deep Learning beats Go world Champion”



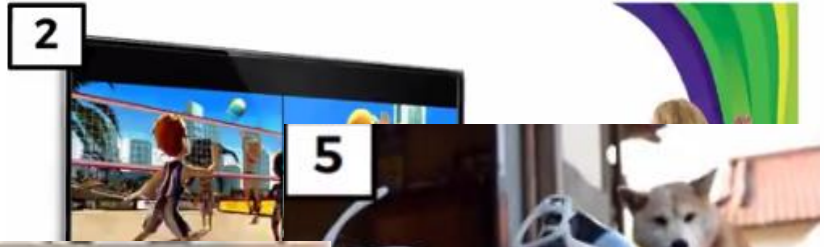
Learning by practicing over and over again

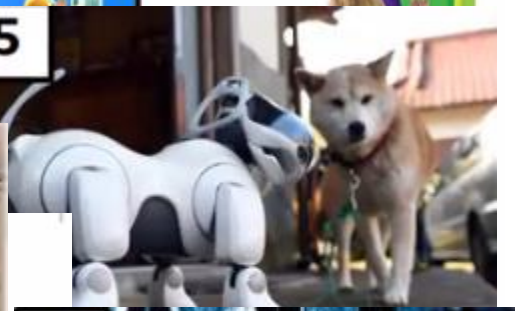
Learning from Data

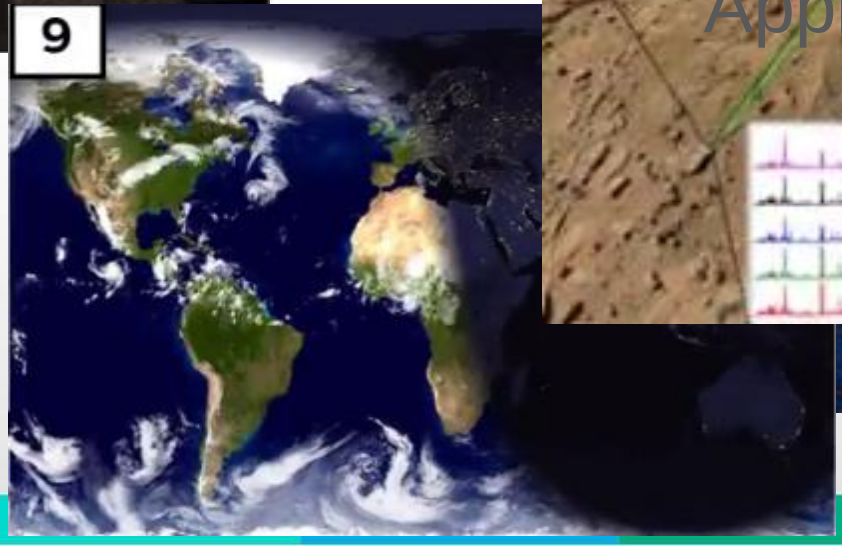


“Machine to learn that knowledge by themselves just as we do”

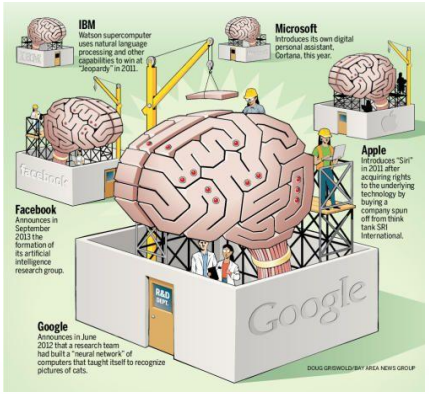
Applications of AI







A computer wins at the world's hardest boardgame

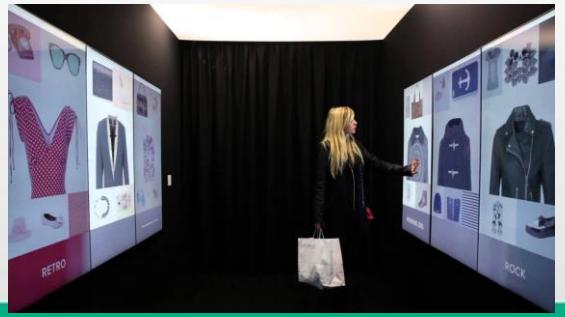


Read my lips, LipNet

Machines and humans pair up to fight fraud online

2016

Natural language processing gives life to a digital personal shopper



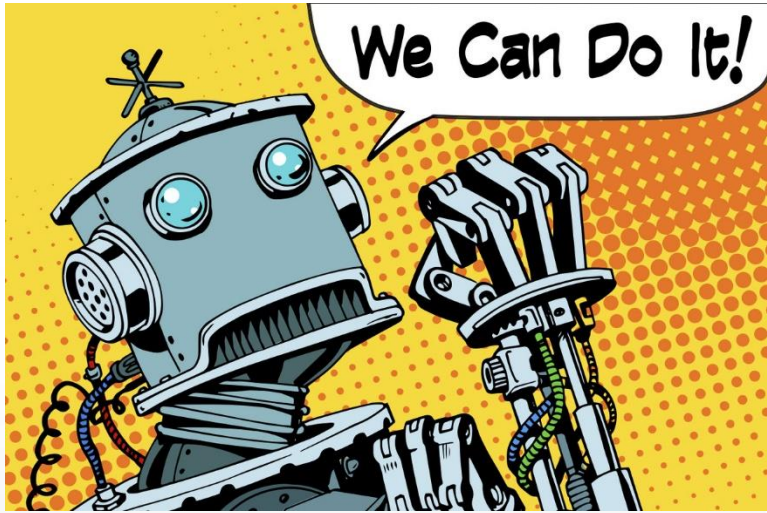
A machine learns how to stop online trolling

“Don’t be fooled we are still very very far from a machine that would be as able as humans to learn to master many aspects of our world” - Yoshua Bengio

Child learn thing that computer are not able to do right now

Computer Learning

- Application of deep learning is what we call supervised learning.
  - With Supervised learning the computer need to be taken by hand and humans have to tell the computer the answers to many questions



Humans Teaching Computers

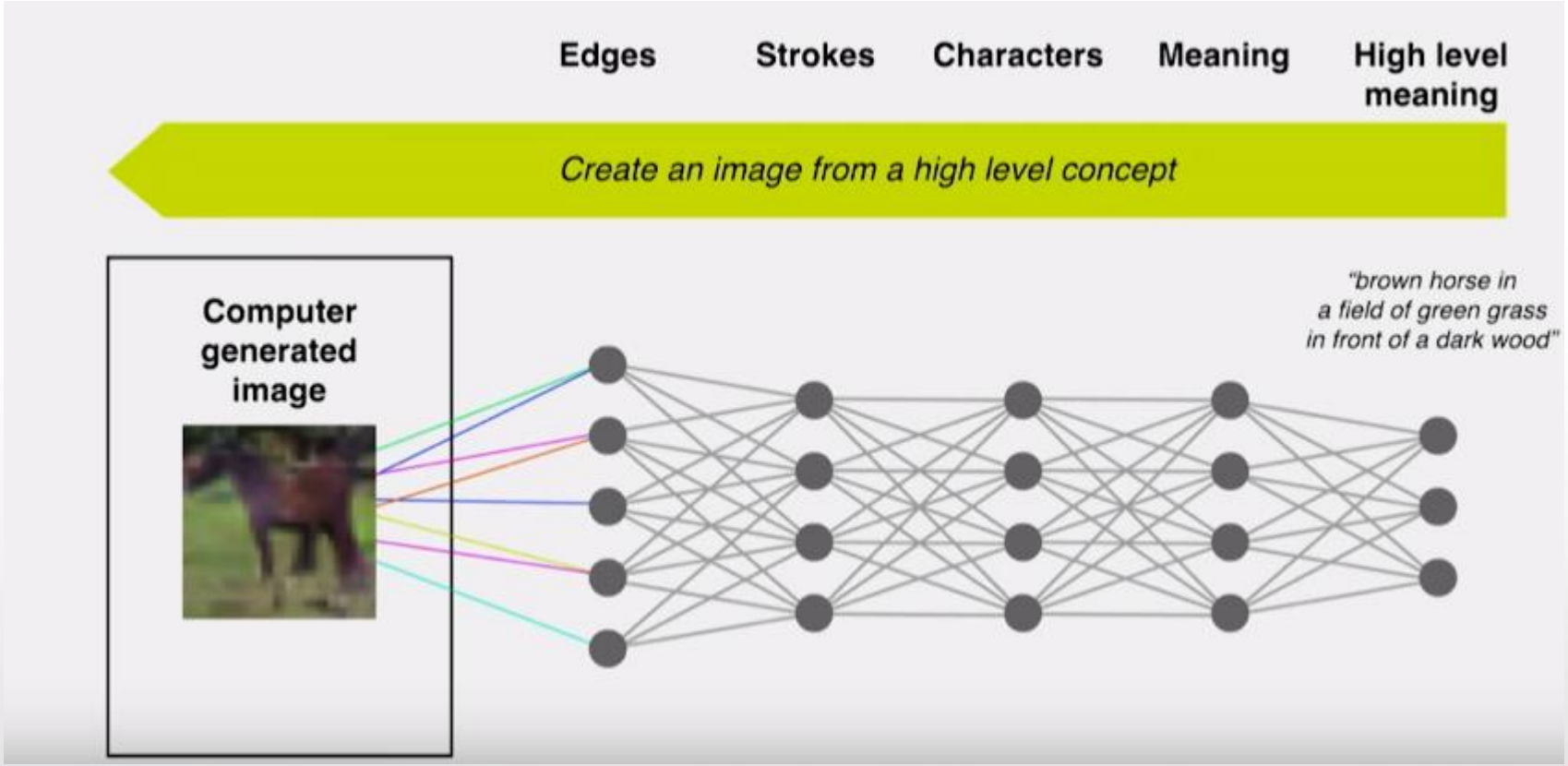
Humans teaching Computers

Although it is painful, it is very powerful which solves many interesting problems.

2 year old child is much stronger

- 2 year old child learning about unsupervised learning helps to deal with *“Self-Driving cars”*

Model to be Creative



Computer Generate Image

- Just like our dreams which is creative and generate new images, new abstracts

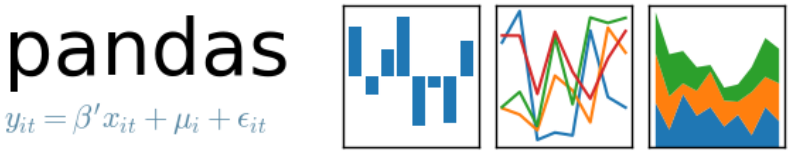
What's next in Machine Learning?

It will be part of your life

Let’s get give you a quick overview of the course!


spyder


jupyter



We start with a quick crash course in the essential Python data Science libraries!

Ready

Data

Data and Data Sets

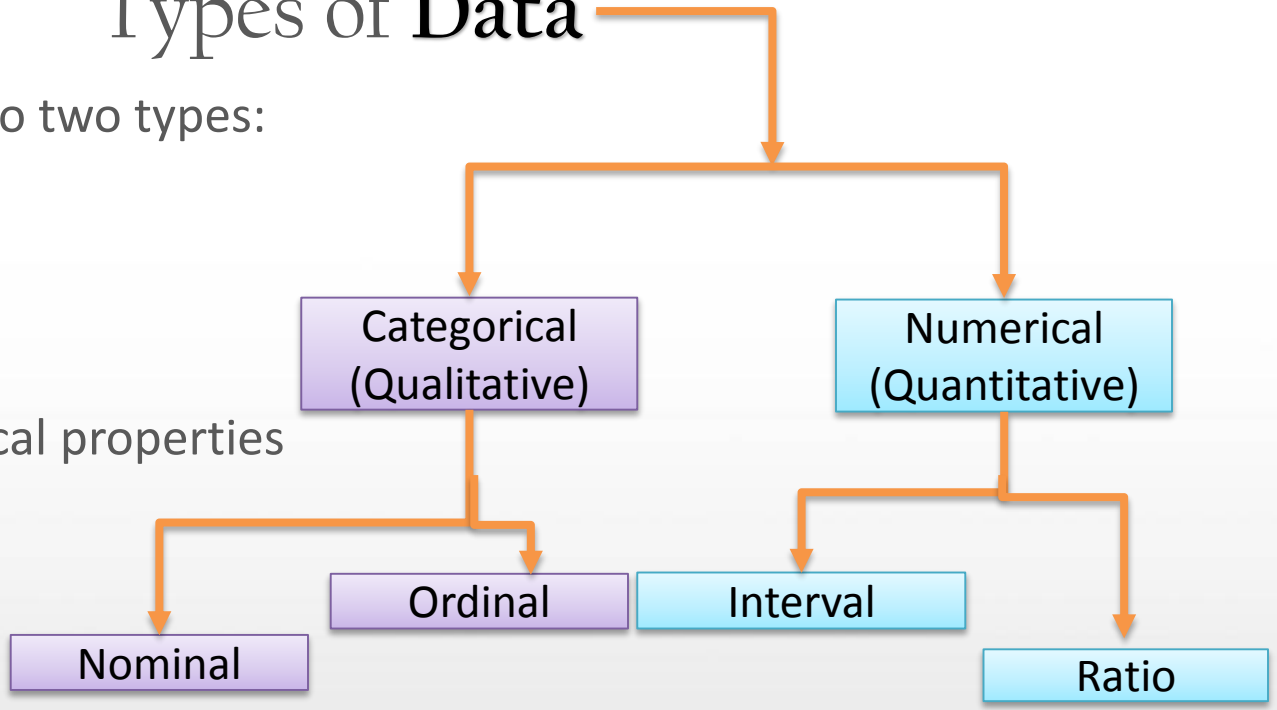
- Data are facts and figures collected summarized, analyzed and interpreted.

- The data collected in a particular study are referred to as dataset.

Types of **Data**

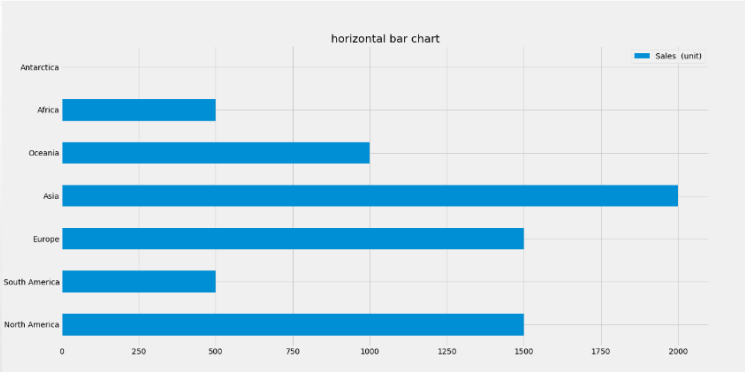
- Data are classified into two types:
  - Qualitative data
  - Quantitative data

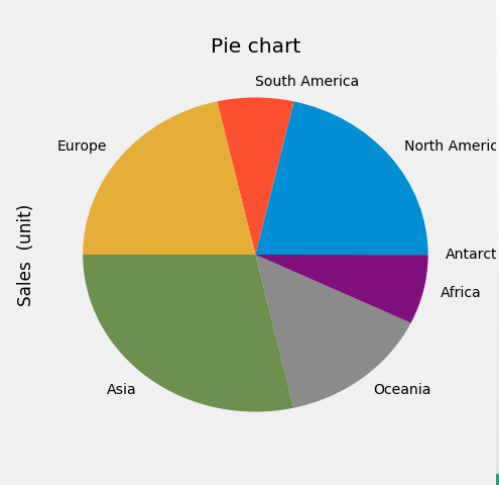
- Based on Mathematical properties
  - Nominal
  - Ordinal
  - Interval
  - Ratio



Nominal Data (Qualitative)

- Nominal means name and count
  - Data are alphabetic or numerical in name
- They are categories without order or direction
- They use to keep track of people, objects and events.

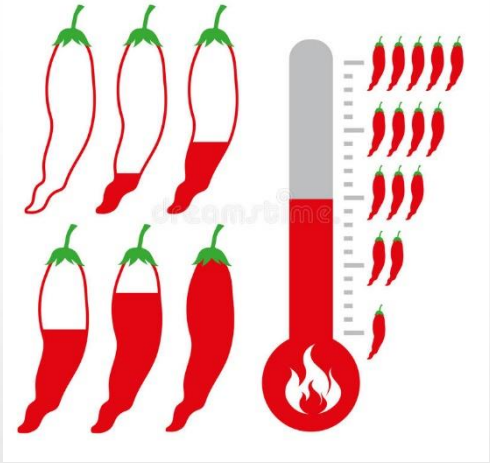




Continent	Sales (unit)
North America	1,500
South America	500
Europe	1,500
Asia	2,000
Oceania	1,000
Africa	500
Antarctica	1

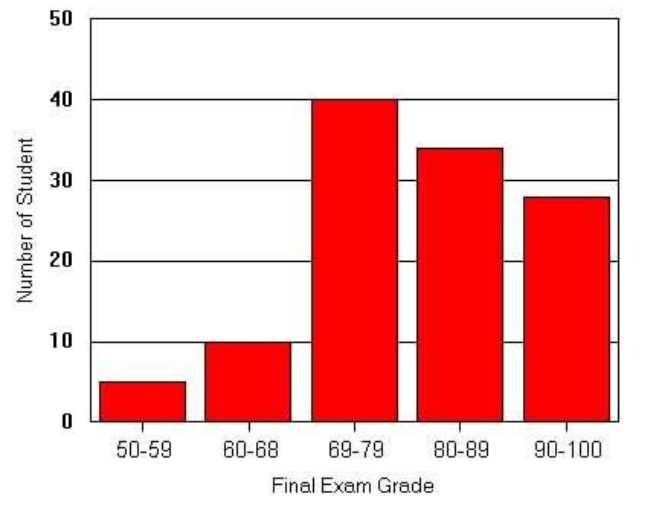
Ordinal Data (Qualitative)

- Ordinal means rank or order
- Data place in order. They are ordered categories like ranking or scaling.
- Has no absolute value
- More precise comparison are not possible



Interval Data (Quantitative)

- Interval data in addition to ranking further allow forming difference.
- For interval data there is no absolute zero. Unique origin does not exists.
- Interval data are more powerful than ordinal scale due to equality of interval.



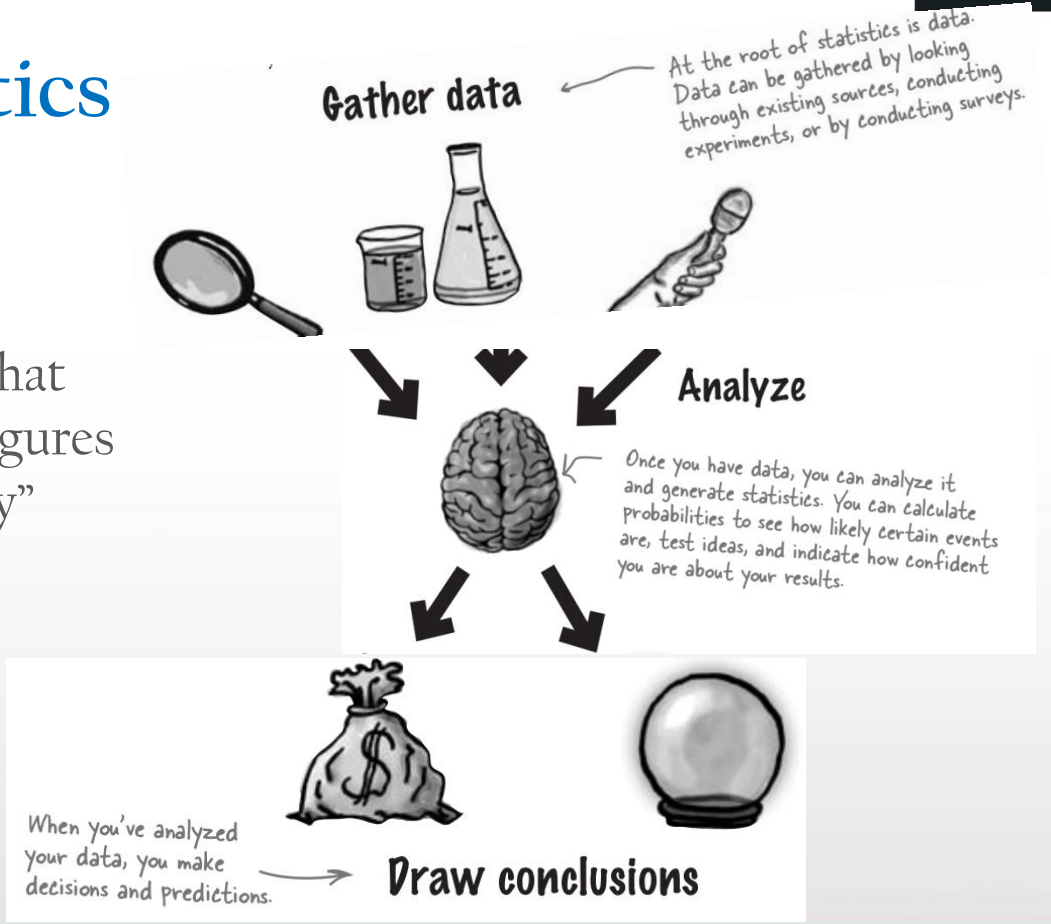
Ratio Data (Quantitative)

- Ratio data allow for forming quotients in addition to setting up inequalities and forming difference.
- All mathematical operations are possible on ratio data.
- The most precise data and allows for applications of all statistical techniques.
- Has absolute zero.

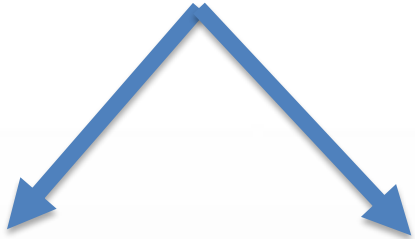
Statistics

Statistics

"Statistics are numbers that summarize raw facts and figures in some meaningful way"



Statistics



Descriptive Statistics

Inferential Statistics

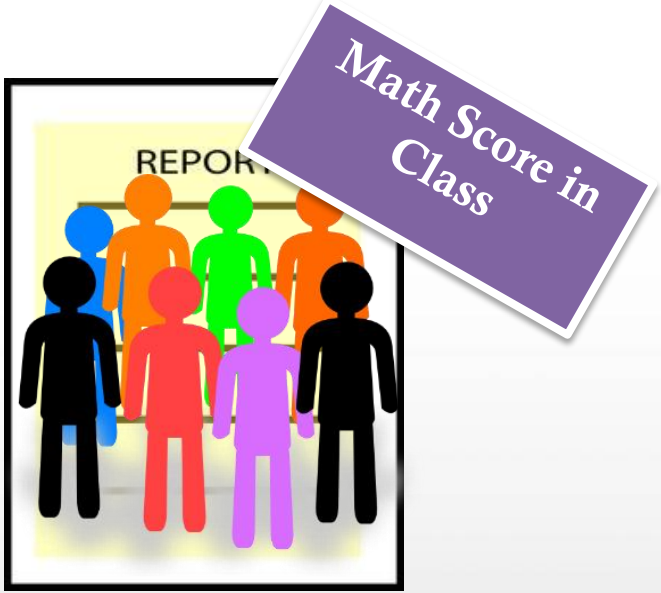
Descriptive Statistics

Descriptive Statistics are the tabular, graphical, and numerical methods used to summarize data

- It is term applied to Meaningful Data Analysis.
- Generally involves organizing summarizing large amount of data. So that it is easy to interpret.

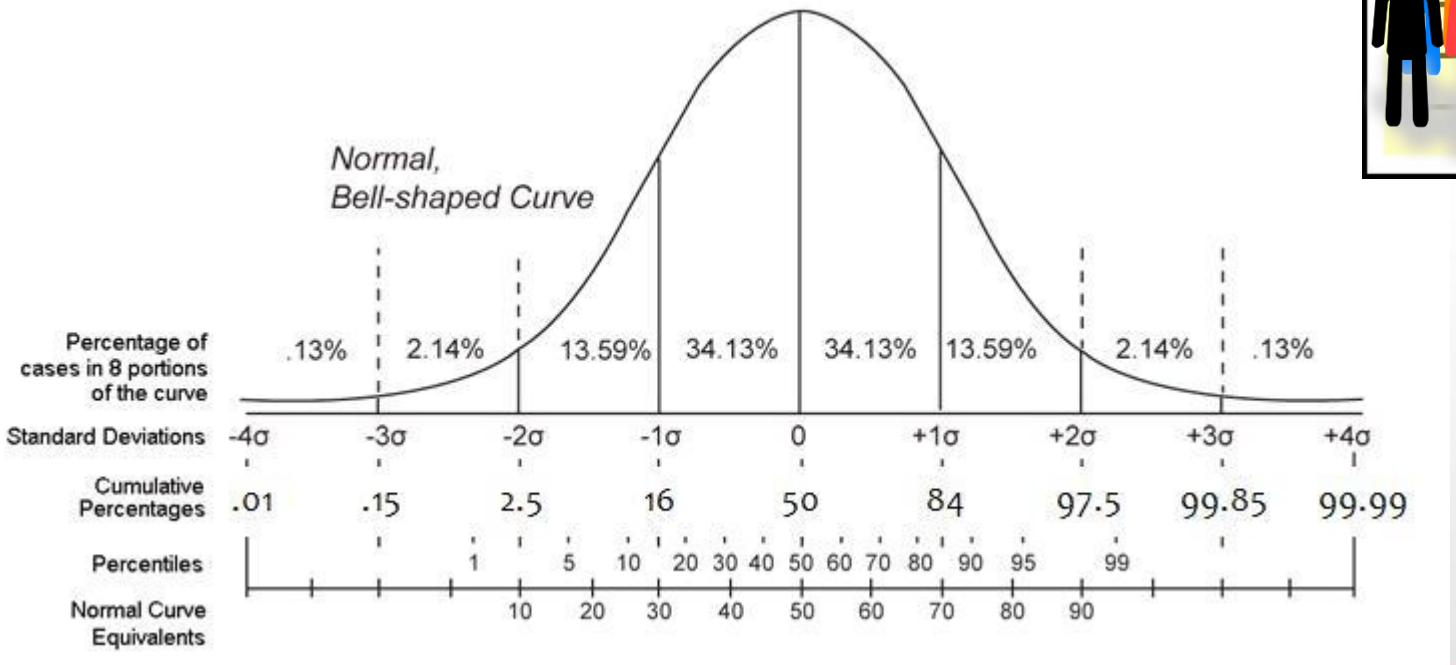
Descriptive Statistics

- It will describe **central tendencies** of those scores
  - Mean
  - Median
  - Mode
  - Variance
  - Standard Deviation

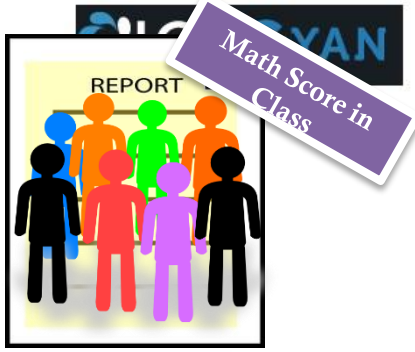


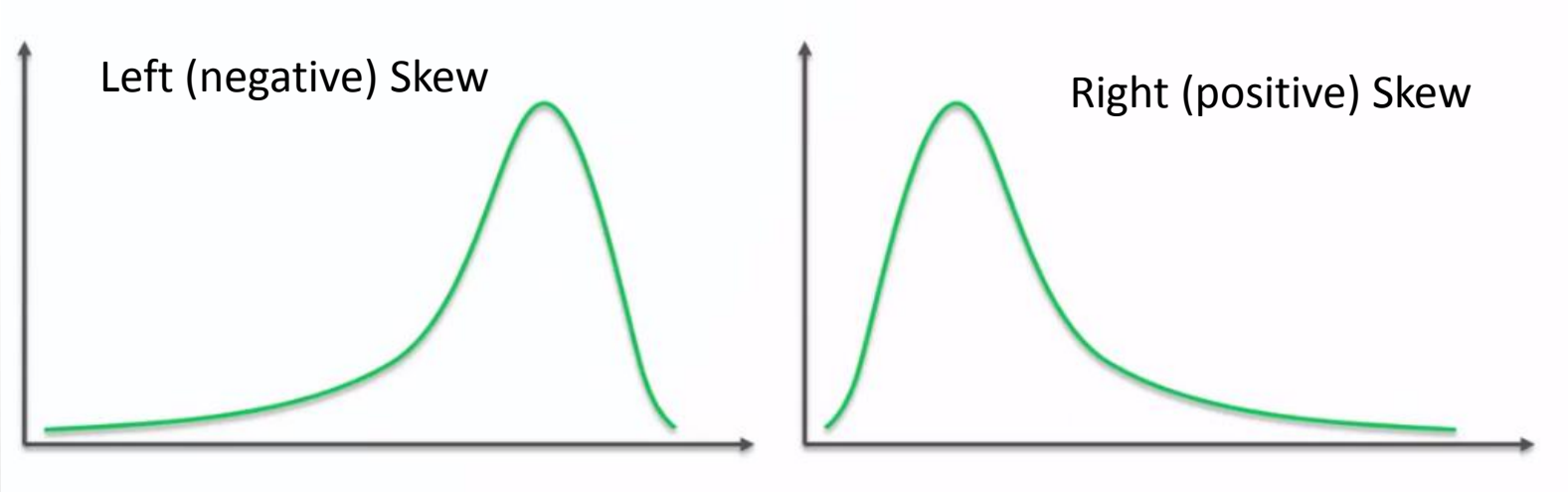
Mean

Math Score in Class

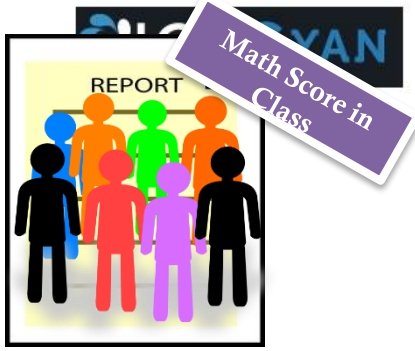


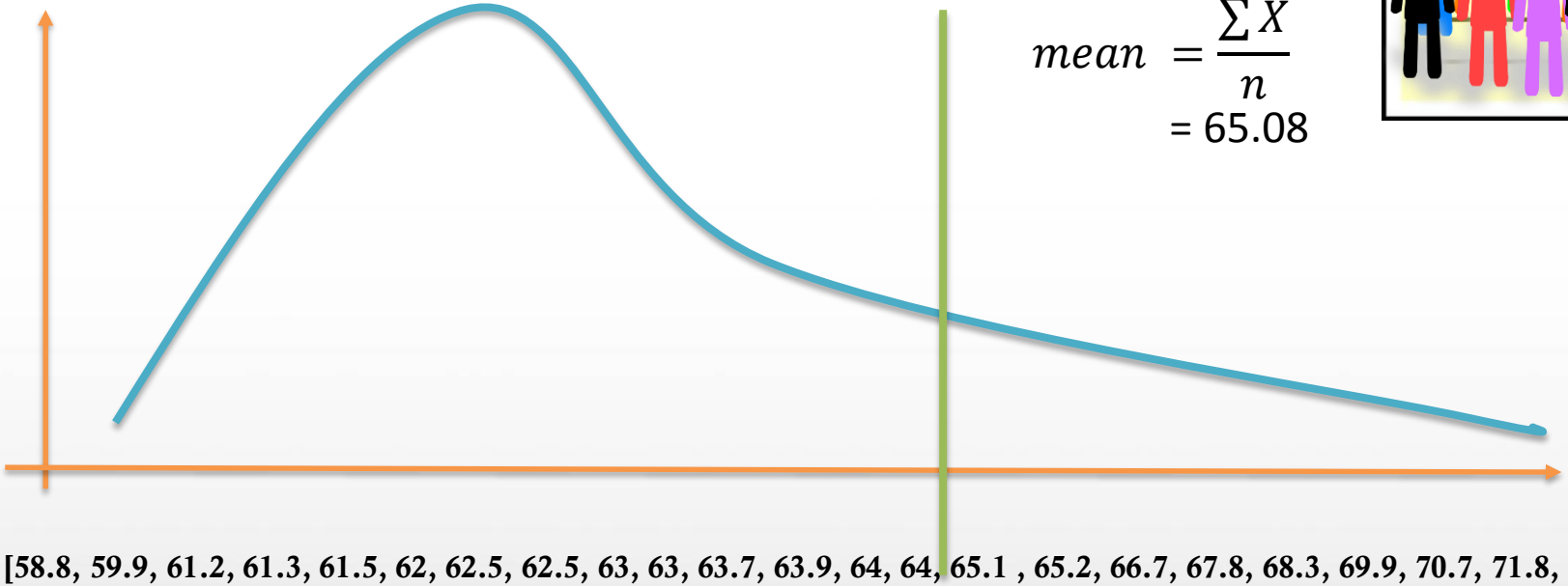
Mean


REPORT
Math Score in Class


Left (negative) Skew
Right (positive) Skew

Mean

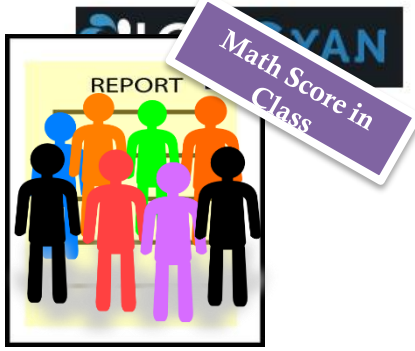


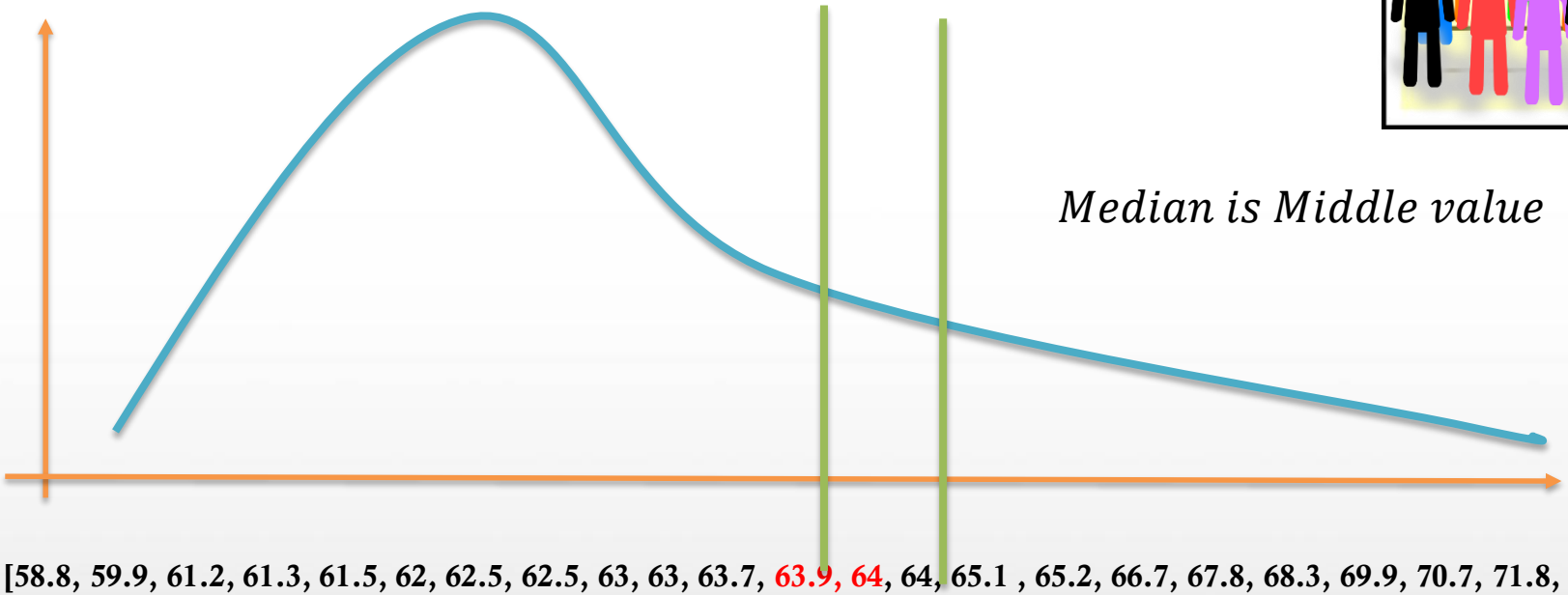


$$mean = \frac{\sum X}{n}$$
$$= 65.08$$

[58.8, 59.9, 61.2, 61.3, 61.5, 62, 62.5, 62.5, 63, 63, 63.7, 63.9, 64, 64, 65.1 , 65.2, 66.7, 67.8, 68.3, 69.9, 70.7, 71.8, 72.2, 73]

Median

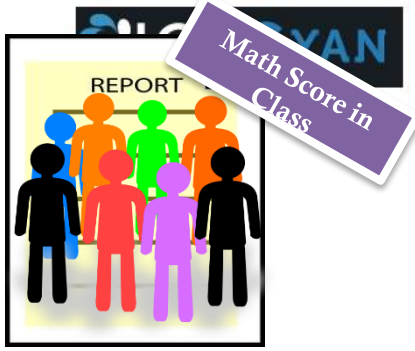


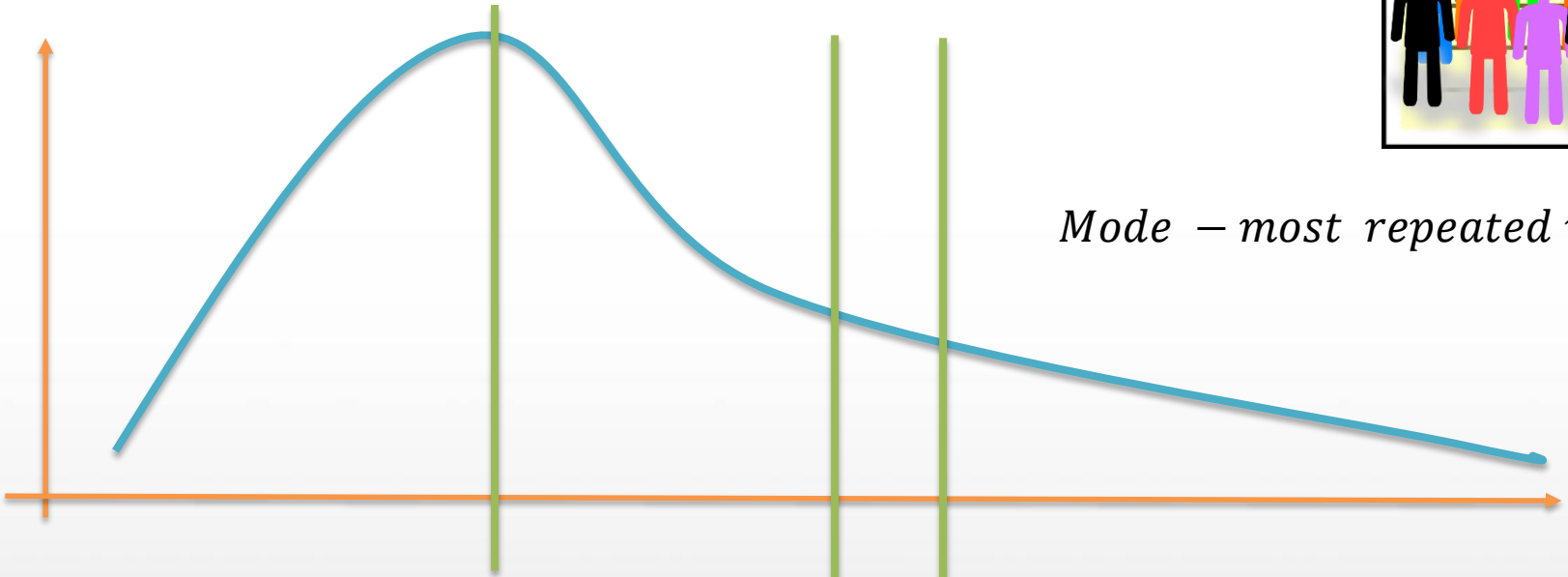


Median is Middle value

[58.8, 59.9, 61.2, 61.3, 61.5, 62, 62.5, 62.5, 63, 63, 63.7, 63.9, 64, 64, 65.1 , 65.2, 66.7, 67.8, 68.3, 69.9, 70.7, 71.8, 72.2, 73]

Mode

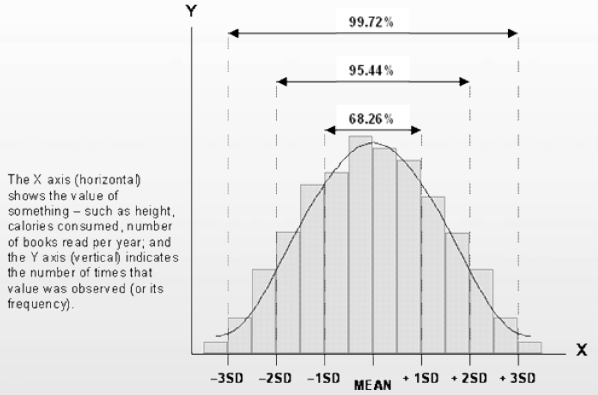




Mode – most repeated value

[58.8, 59.9, 61.2, 61.3, 61.5, 62, 62.5, 62.5, 63, 63, 63.7, 63.9, 64, 64, 65.1 , 65.2, 66.7, 67.8, 68.3, 69.9, 70.7, 71.8, 72.2, 73]

Measuring Variability and spread



The X axis (horizontal) shows the value of something – such as height, calories consumed, number of books read per year; and the Y axis (vertical) indicates the number of times that value was observed (or its frequency).

Range to differentiate between dataset

- It is quite often, the average only gives part of the picture.
- Averages give us a way of determining where the center of a set of data is, but they don't tell us how the data varies.

"The range tells us over how many numbers the data extends, a bit like measuring its width."

Range

The range is a way of measuring how spread out a set of values are. It's given by Upper bound - Lower bound where the upper bound is the highest value, and the lower bound the lowest.

Range = upper bound - lower bound

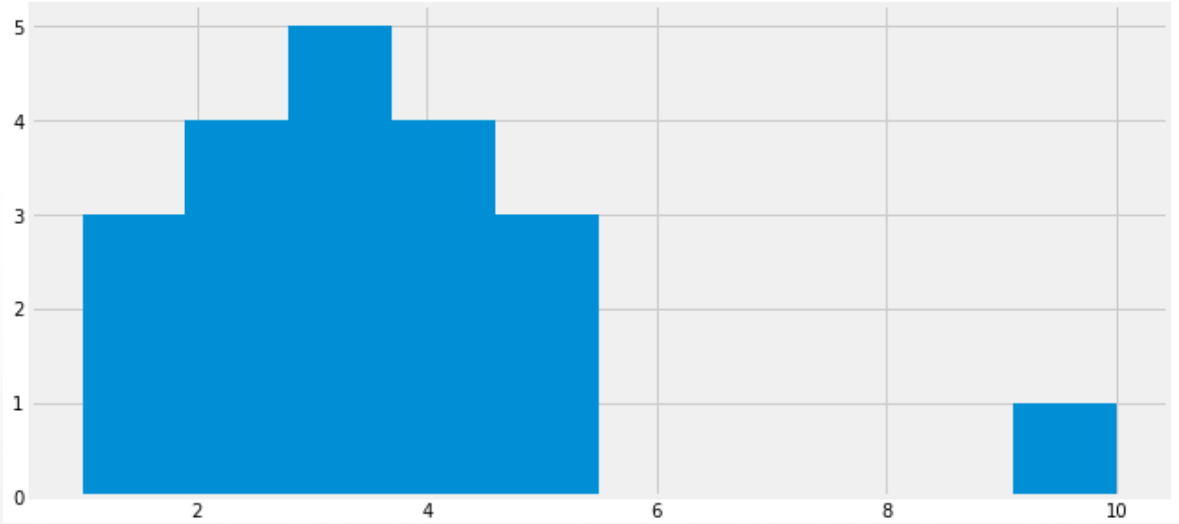
= 10 -1

= 9

so, the range is 9

Range

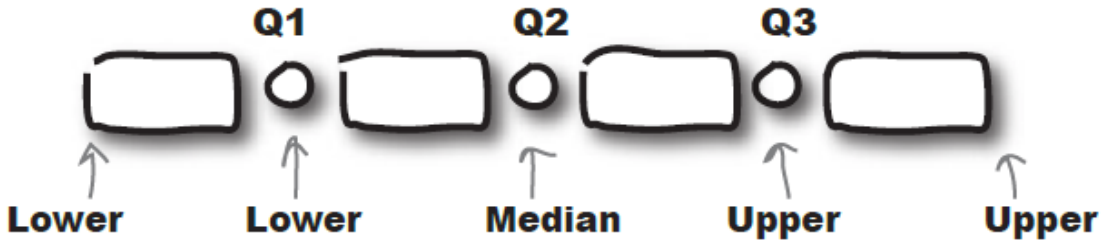
Limitation

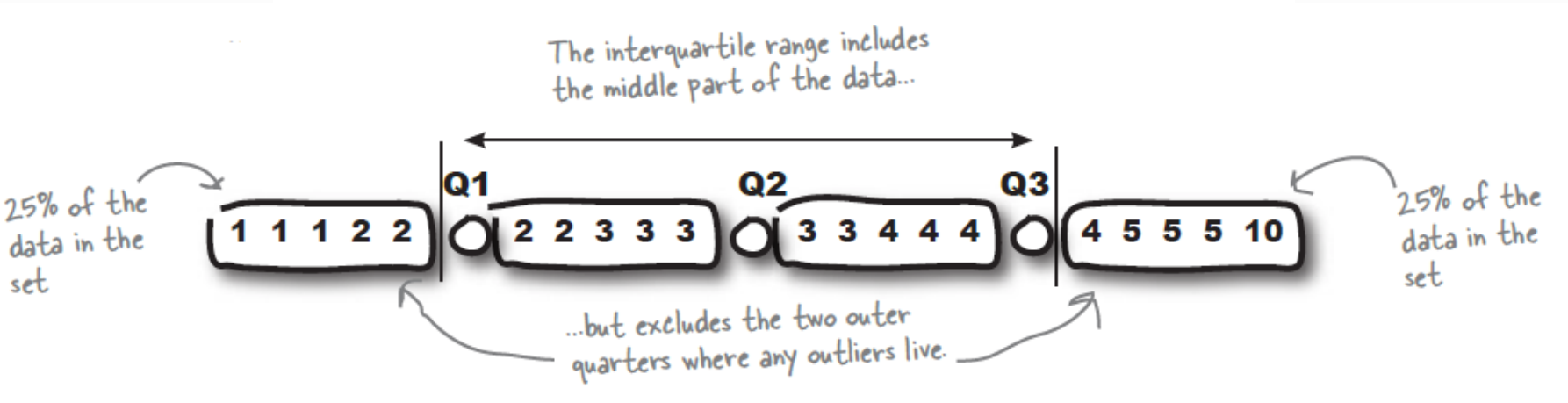


"If your data has outliers, using the range to describe how your values are dispersed can be very misleading because of its sensitivity to outliers"

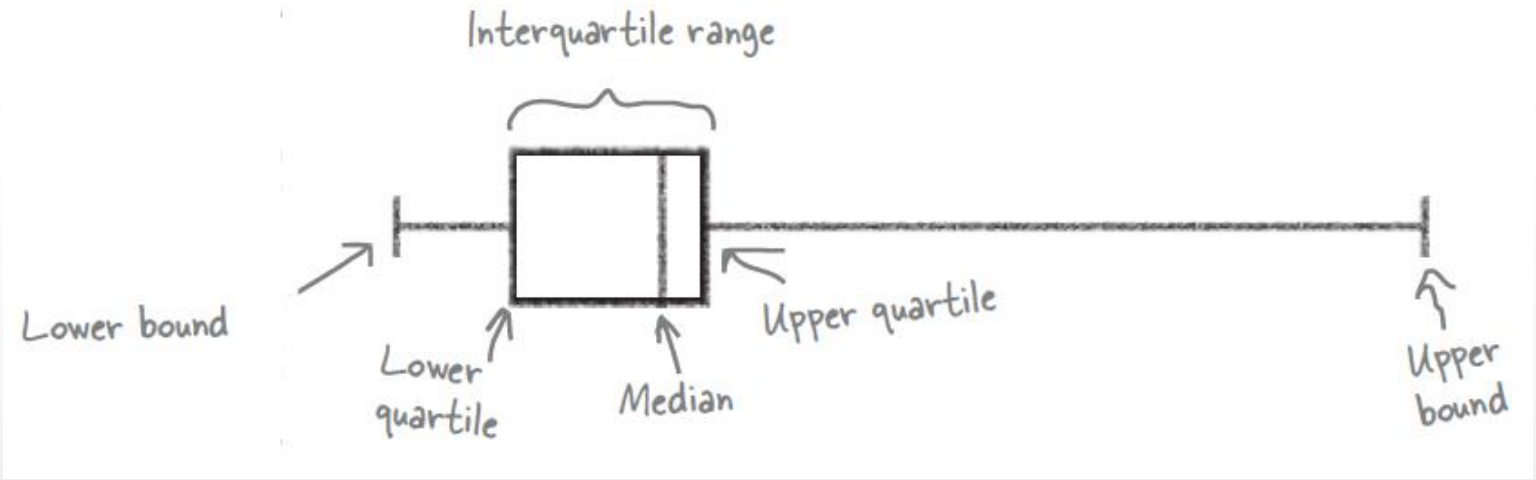
Quartiles will rescue the problem

- Quartiles of a set of data is a very similar process to finding the median.





Box plot → Quartiles



Variance

- The variance is a way of measuring spread, and it's the average of the distance of values from the mean squared.

$$\sigma^2 = \frac{\sum(x - \mu)^2}{n}$$

- This is a method of measuring spread

Standard Deviation

- Standard deviation is a way of saying how far typical values are from the mean.
- The smaller the standard deviation, the closer values are to the mean.
- The smallest value the standard deviation can take is 0.

$$\sigma = \sqrt{Variance}$$

$$\sigma = \sqrt{\frac{\sum(x - \mu)^2}{n}}$$

- This is a method of measuring spread

Z - Score

- How far is any given data point from the mean ? (Distance)
  - Z – score can help us answer
- How many standard deviation away (above and below) from the mean is a data point ?
- Units for Z- score is “standard deviation”
- Z – score is measure of distance from mean.

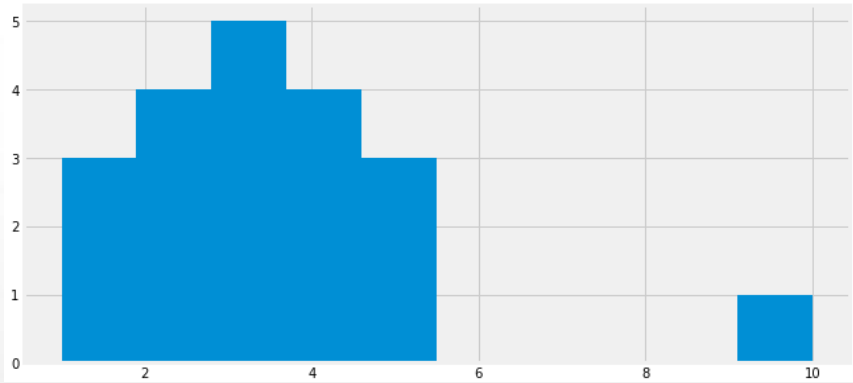
$$Z = \frac{x - \mu}{\sigma}$$

Value	Mean	value- Mean	Z	Outlier ?
1	3.26	-2.26	-1.1037	Not outlier
1	3.26	-2.26	-1.1037	Not outlier
1	3.26	-2.26	-1.1037	Not outlier
2	3.26	-1.26	-0.6160	Not outlier
2	3.26	-1.26	-0.6160	Not outlier
2	3.26	-1.26	-0.6160	Not outlier
2	3.26	-1.26	-0.6160	Not outlier
3	3.26	-0.26	-0.6160	Not outlier
3	3.26	-0.26	-0.1283	Not outlier
3	3.26	-0.26	-0.1283	Not outlier
3	3.26	-0.26	-0.1283	Not outlier
3	3.26	-0.26	-0.1283	Not outlier
4	3.26	1.74	0.3593	Not outlier
4	3.26	1.74	0.3593	Not outlier
4	3.26	1.74	0.3593	Not outlier
4	3.26	1.74	0.3593	Not outlier
5	3.26	2.74	0.8470	Not outlier
5	3.26	2.74	0.8470	Not outlier
10	3.26	7.74	3.28	outlier

The lower bound is still 1.

But the upper bound has increased to 10.

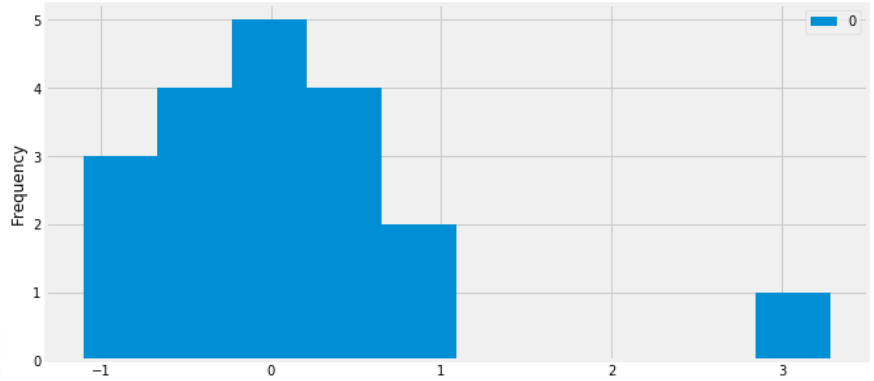
1 1 1 2 2 2 2 3 3 3 3 3 4 4 4 4 5 5 5 10

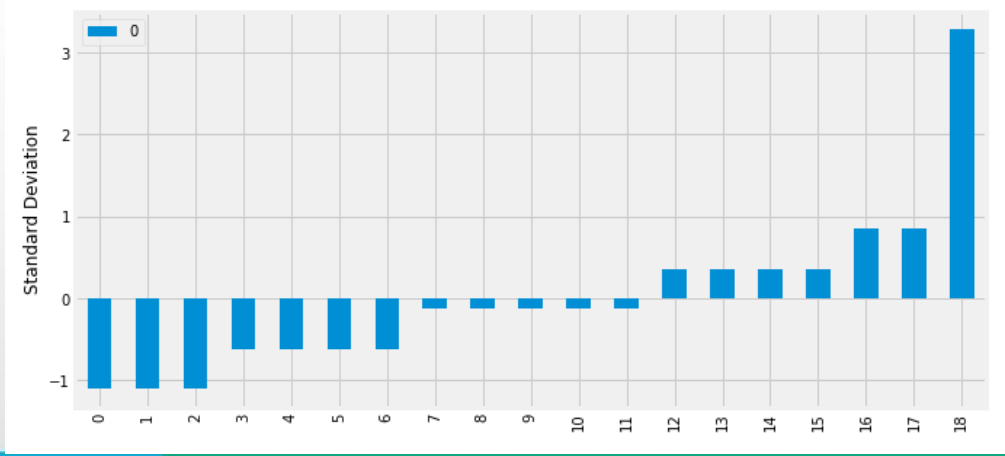


The lower bound is still 1.

But the upper bound has increased to 10.

1 1 1 2 2 2 2 3 3 3 3 3 4 4 4 4 5 5 5 10





IOTGYAN