Data visualization !

Mohammad Ashour

- 2th year CS student – Medical Informatics.
- Data Science and Machine learning practitioner.
- Former AI intern at NTI.
- Former AI intern at IBM.
- Former Data Analyst intern at Google , FWD & DataCamp.
- Former Data Science Mentor @CIS & EGI.
- Former Team Leader @EGI Team.

What is data visualization ?

- Data visualization is the act of taking information (data) and placing it into a visual context , such as a map or graph.
- Data (or information) visualization is used to interpret and gain insight into large amounts of data. This is achieved through visual representations, often interactive, of raw data.

Why is it important?

- Data visualization gives us a clear idea of what the information means by giving it visual context.
- Data visualizations make big and small data easier for the human brain to understand
- Correlations in Relationships
- Trends Over Time.
- Find Patterns .
- Frequency.
- Risk and Reward.

Is data visualization a part of data science ?

Data science and data visualization are not two different entities . They are bound to each other . Data visualization is a subset of data science . Data science is not a single process or a method or any workflow .

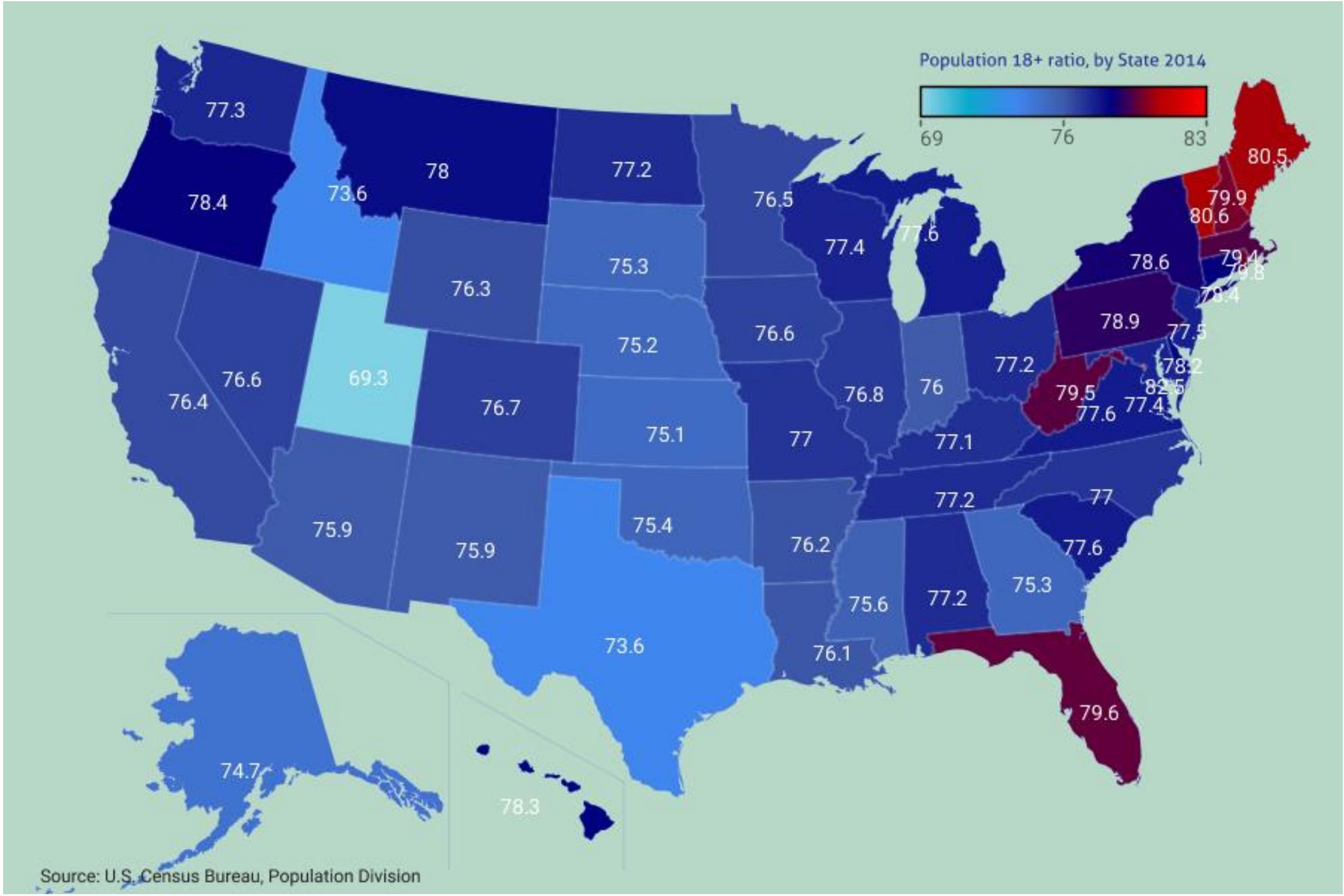
What are types of data ?

Numerical & Categorical

Examples of Data Visualizations

Maps

Help organize data geographically.


Population 18+ ratio, by State 2014
69
76
83
77.3
78.4
73.6
78
77.2
76.5
77.4
77.6
80.5
79.9
80.6
78.6
79.4
79.8
78.4
78.9
77.5
78.2
82.5
77.4
77.6
79.5
77.2
76
76.8
77
77.1
76.6
75.3
76.3
75.2
75.1
76.7
69.3
76.6
76.4
75.9
75.9
75.4
76.2
77.2
77
77.6
75.3
77.2
75.6
76.1
73.6
79.6
74.7
78.3
Source: U.S. Census Bureau, Population Division

How can i visualize Categorical data?

Bar Graphs

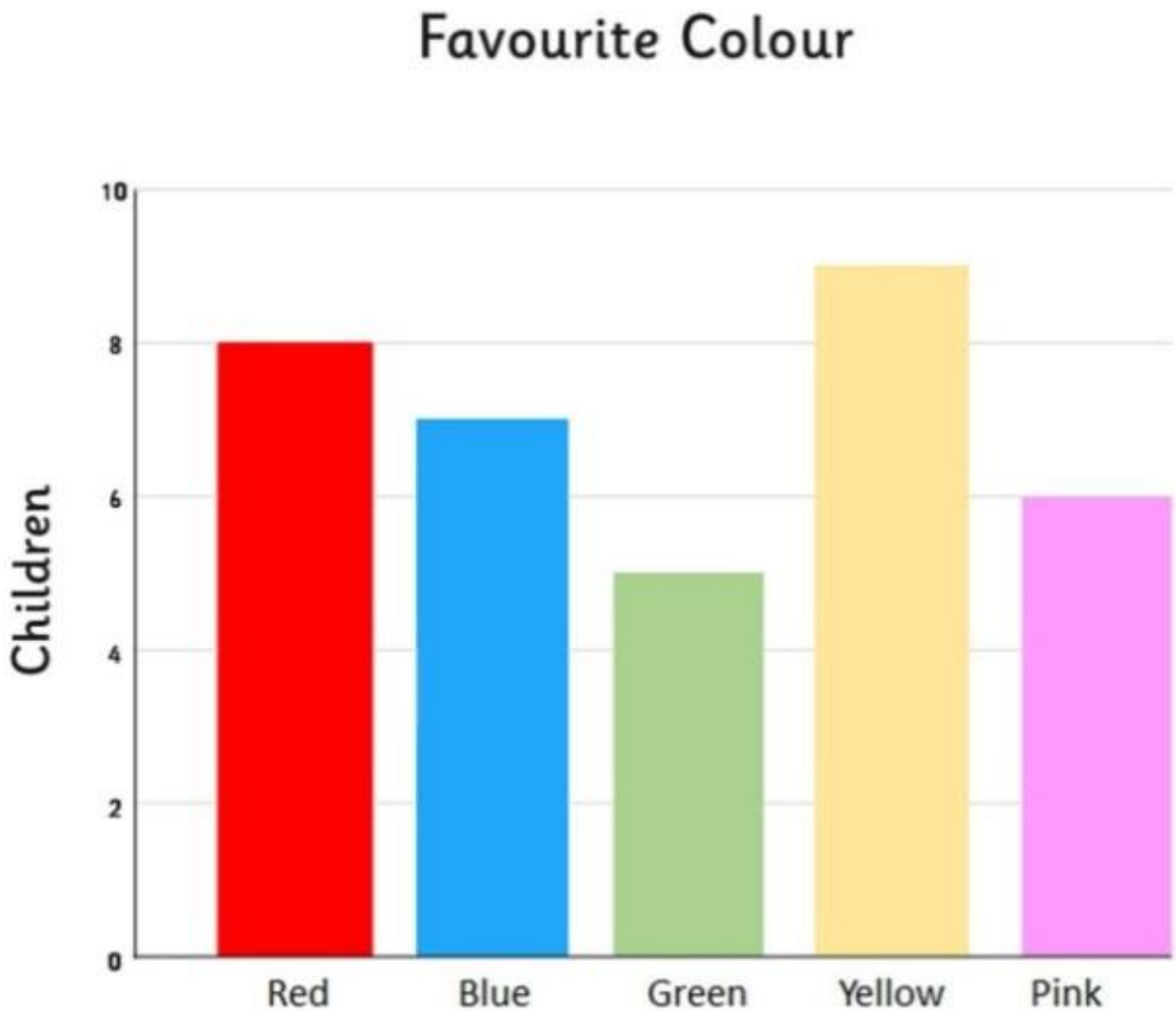
- Use size contrast to compare two or more variables.
- Bar charts, seen often in the popular press, are a common visual tool for displaying a single categorical variable.

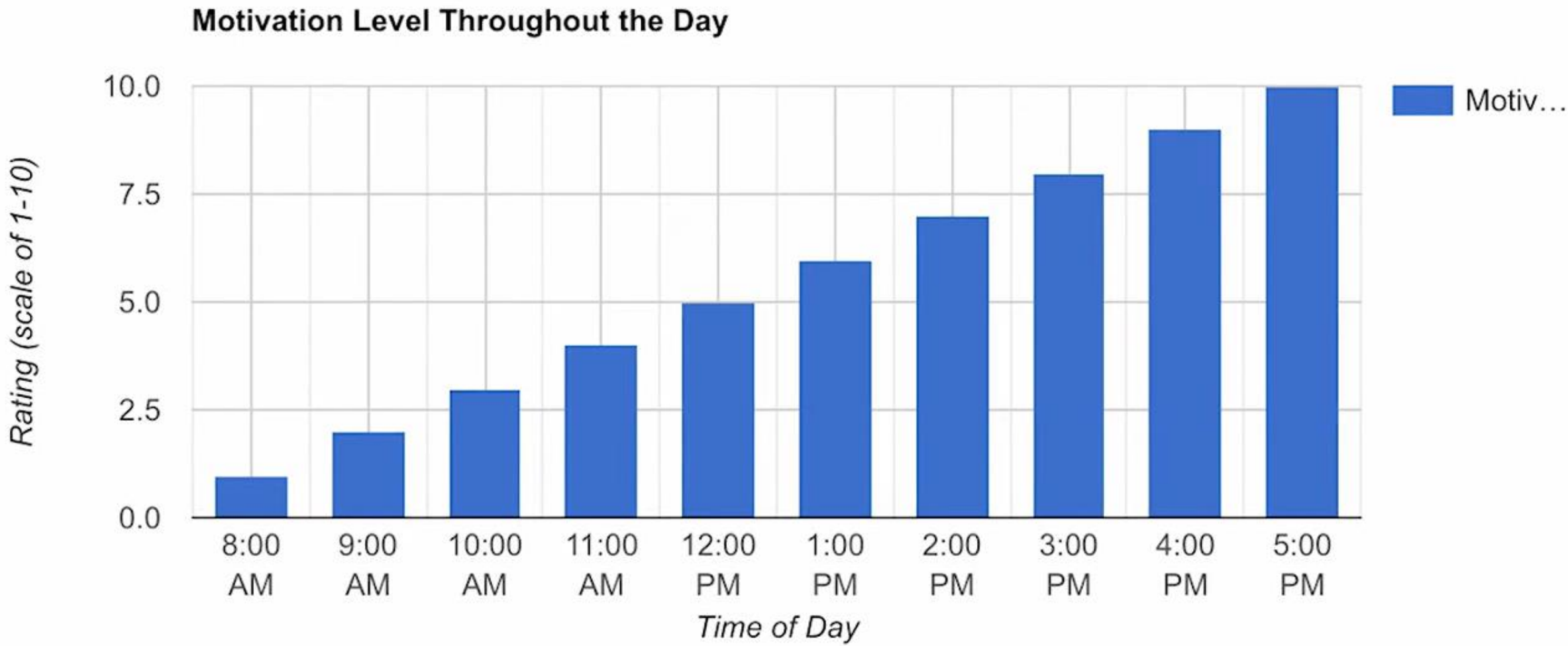
x-axis

- The horizontal line of a bar graph , usually placed at the bottom.
- The x-axis is used to present categories data.

y-axis

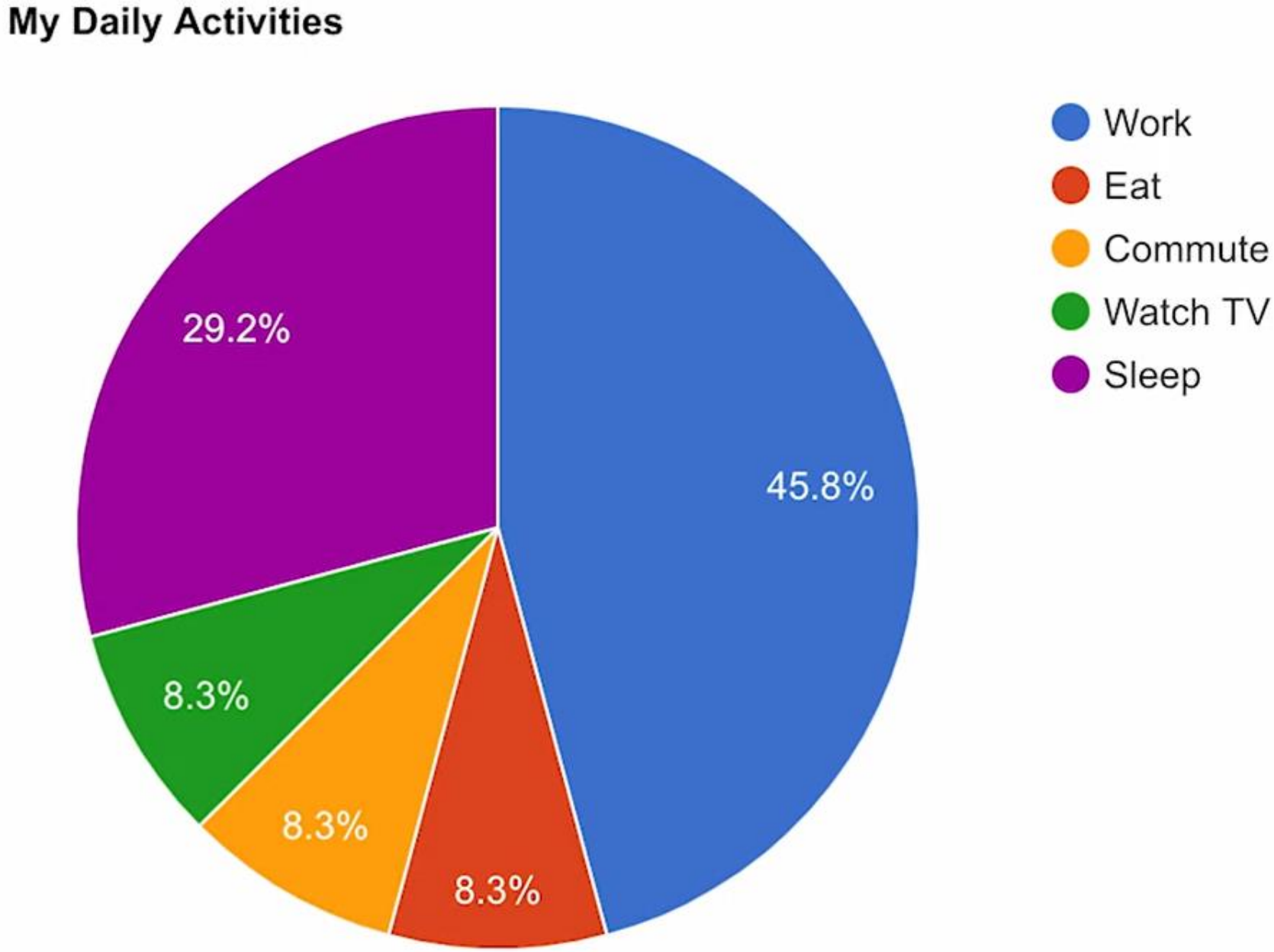
- The vertical line of a bar graph , usually placed at the left.
- The y-axis is used to present numerical data.


Favourite Colour
Children
10
8
6
4
2
0
Red
Blue
Green
Yellow
Pink


Motivation Level Throughout the Day
Rating (scale of 1-10)
10.0
7.5
5.0
2.5
0.0
8:00 AM
9:00 AM
10:00 AM
11:00 AM
12:00 PM
1:00 PM
2:00 PM
3:00 PM
4:00 PM
5:00 PM
Time of Day
Motiv…

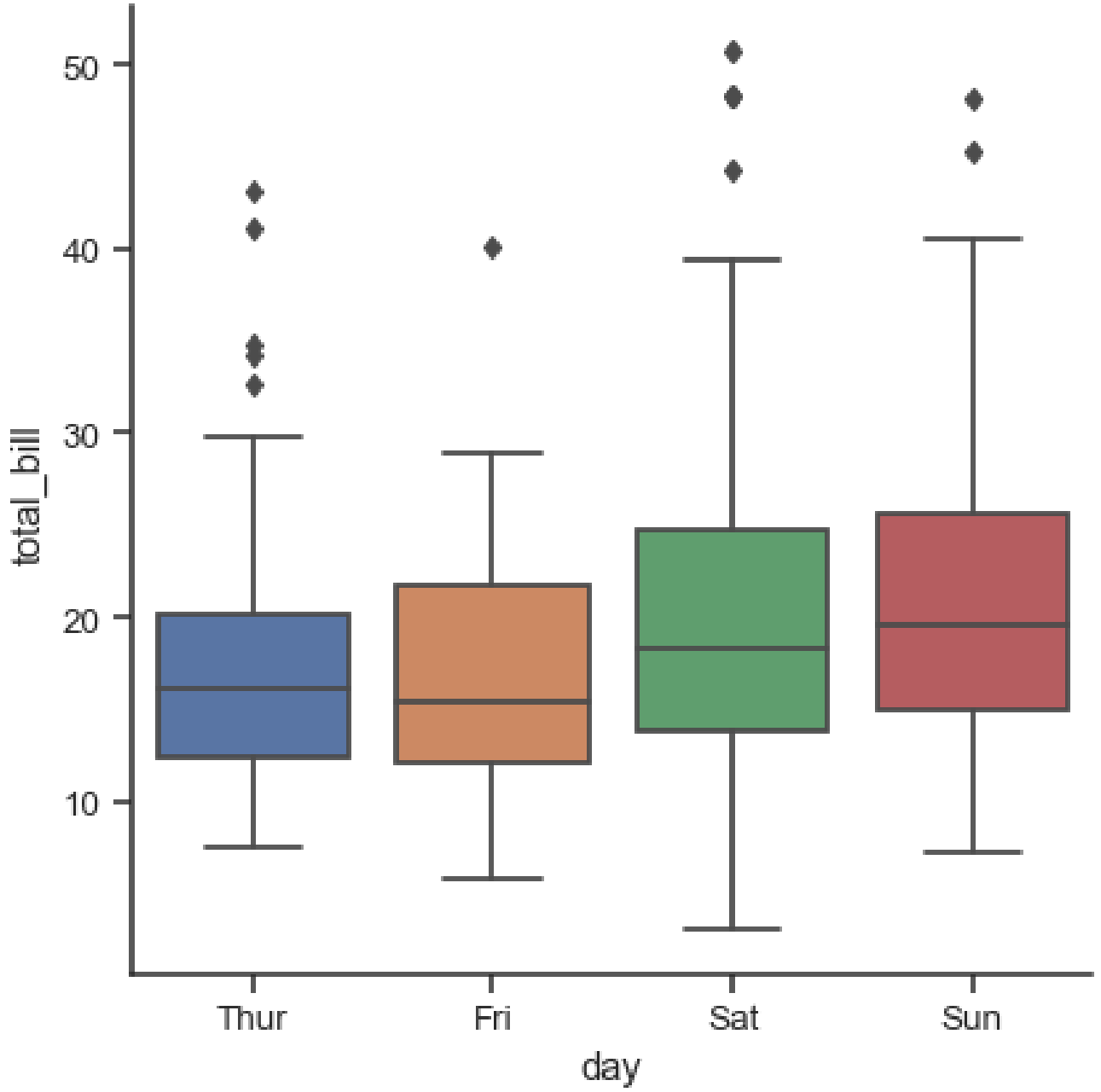
Pie Graphs

- Show how much each part of something makes up the whole.
- Represent categorical data.


My Daily Activities
45.8%
8.3%
8.3%
8.3%
29.2%
Work
Eat
Commute
Watch TV
Sleep

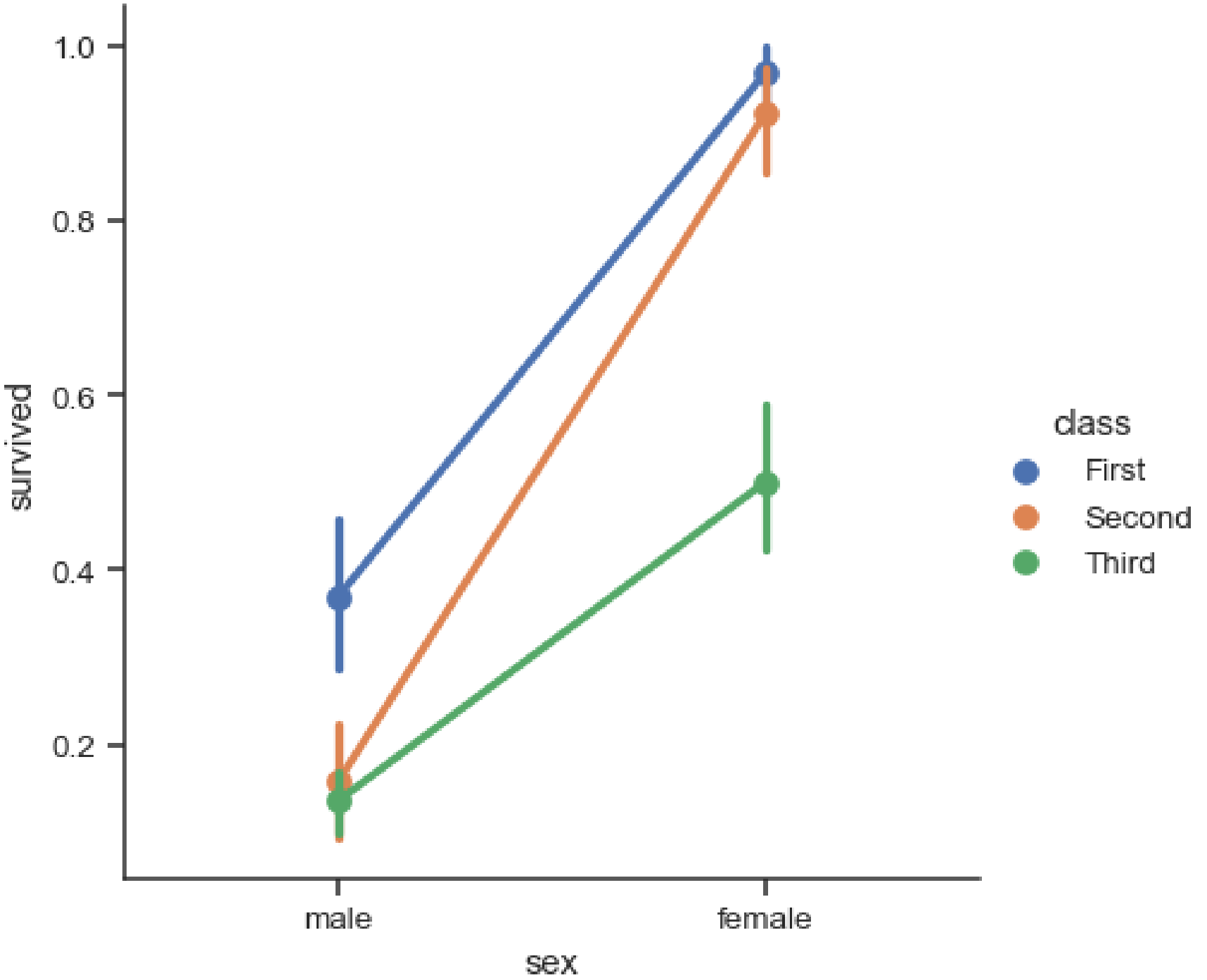
Box plots

This kind of plot shows the three quartile values of the distribution along with extreme values.


50
40
30
20
10
total_bill
Thur
Fri
Sat
Sun
day

Point plots

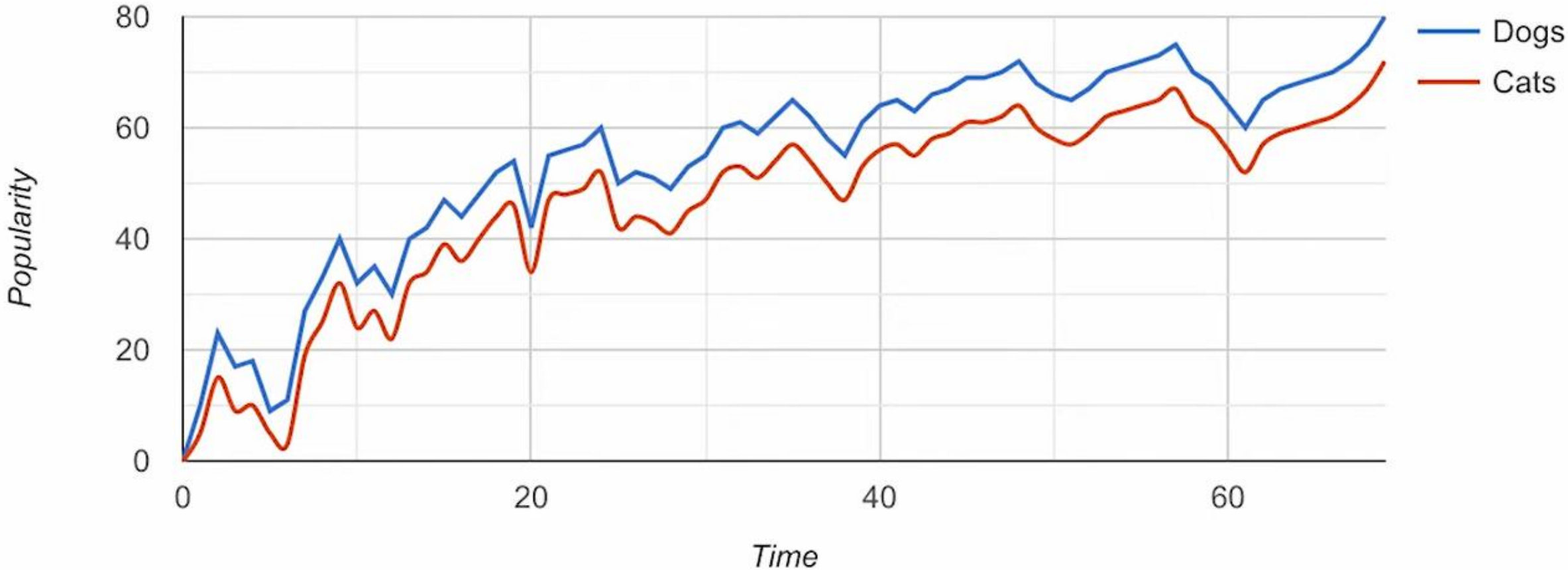
- This function also encodes the value of the estimate with height on the other axis.
- Additionally, point plot connects points from the same hue category.


1.0
0.8
0.6
0.4
0.2
survived
male
female
sex
class
First
Second
Third

How can i visualize Numerical data?

Line Graphs

- Help your audience to understand shifts or changes in your data.


80
60
40
20
0
Popularity
0
20
40
60
Time
Dogs
Cats

Scatter Plot

- Scatter plots are a great tool for visualizing relationships between two numerical variables. We can use it to better understand how one variable impacts another if it does.

Scatter Plot

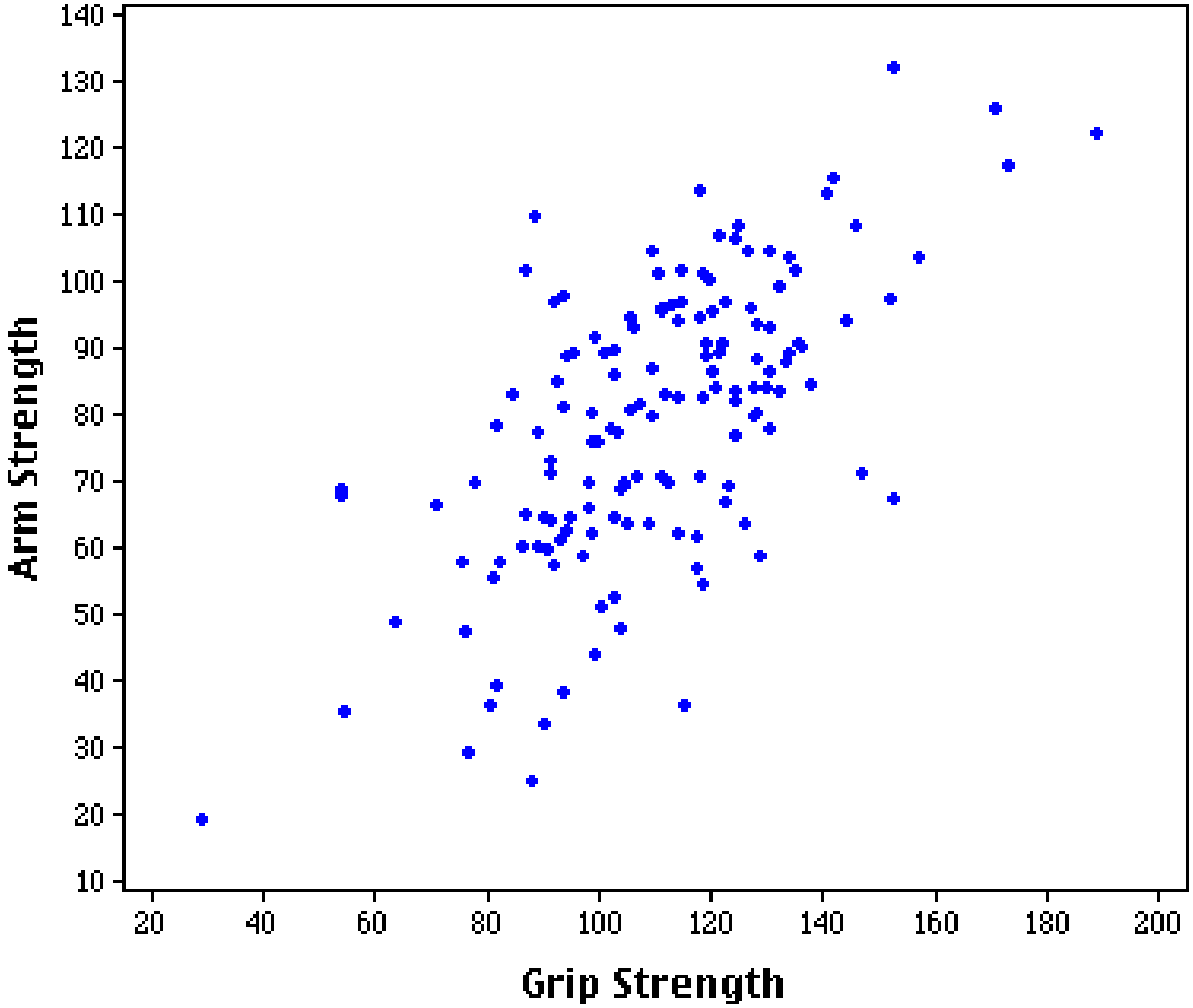
- Scatter plots are a great tool for visualizing relationships between two numerical variables. We can use it to better understand how one variable impacts another if it does.

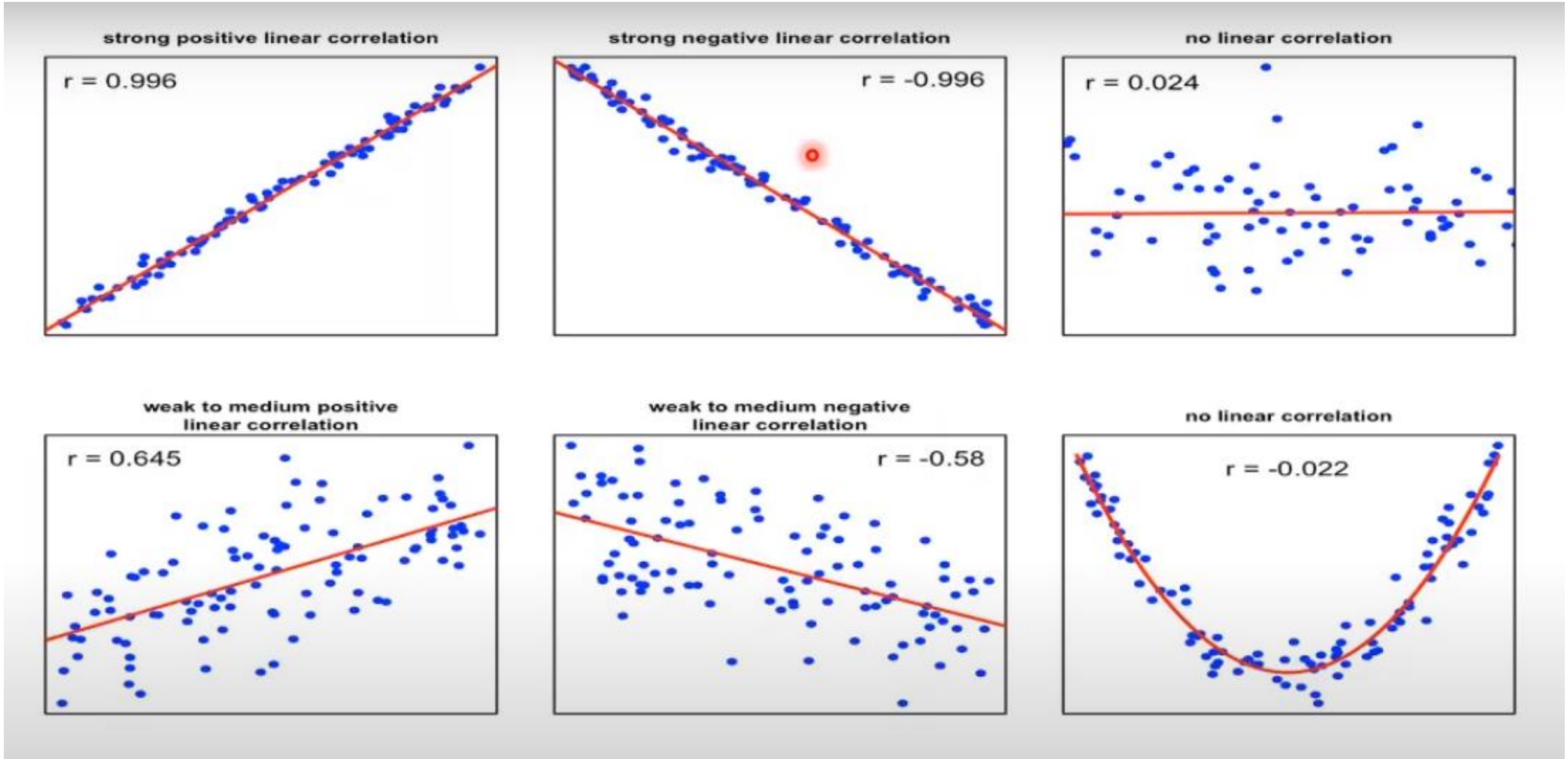
x-axis

- The horizontal line of a Scatter graph , usually placed at the bottom.
- the x-axis is the value of one variable.

y-axis

- The vertical line of a bar graph , usually placed at the left.
- The y-axis has the value of another.


Arm Strength
Grip Strength
140
130
120
110
100
90
80
70
60
50
40
30
20
10
20
40
60
80
100
120
140
160
180
200


strong positive linear correlation
r = 0.996
strong negative linear correlation
r = -0.996
no linear correlation
r = 0.024
weak to medium positive linear correlation
r = 0.645
weak to medium negative linear correlation
r = -0.58
no linear correlation
r = -0.022

Exploring the Distribution

Histogram

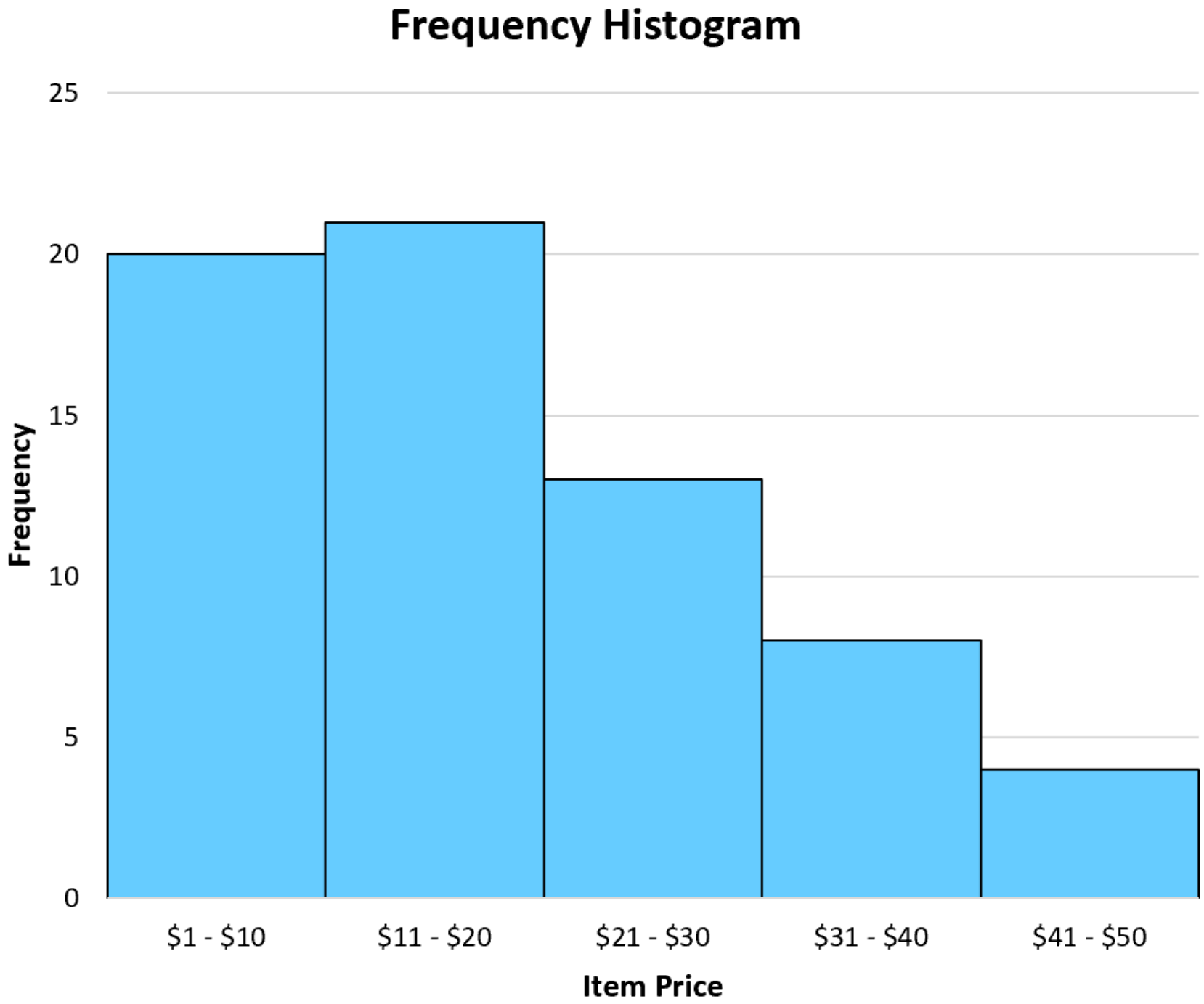
- A histogram is the graphical representation of data where data is grouped into continuous number ranges and each range corresponds to a vertical bar

x-axis

- The horizontal line of a Scatter graph , usually placed at the bottom.
- bins on the x-axis.

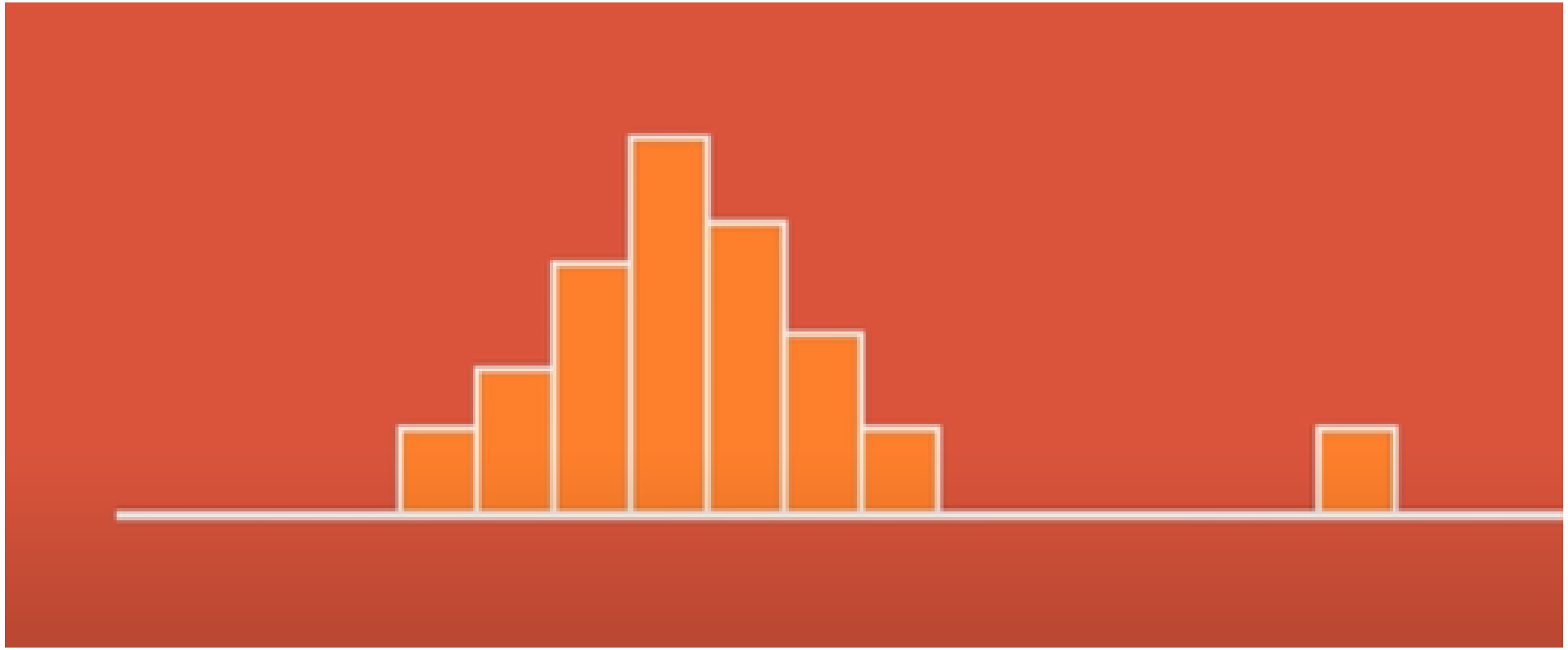
y-axis

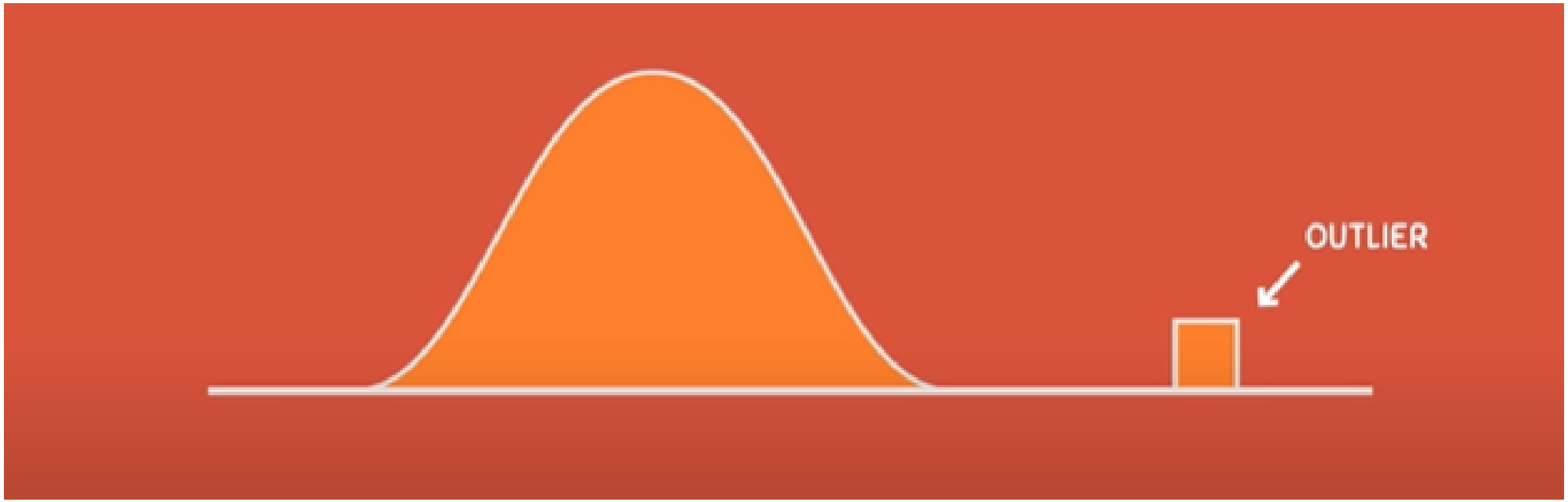
- The vertical line of a bar graph , usually placed at the left.
- Data count on the y-axis

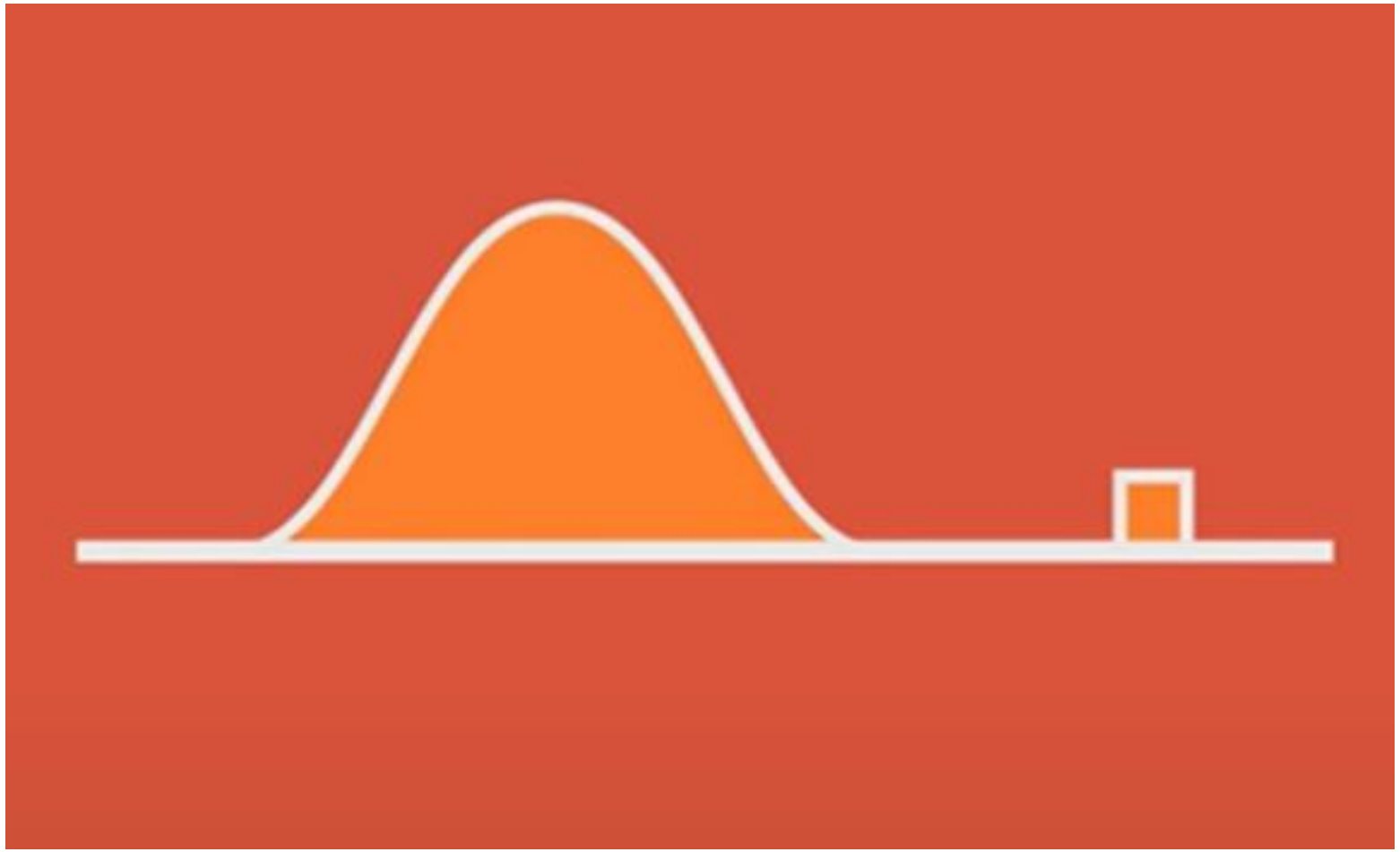

Frequency Histogram
25
20
15
10
5
0
Frequency
$1 - $10
$11 - $20
$21 - $30
$31 - $40
$41 - $50
Item Price

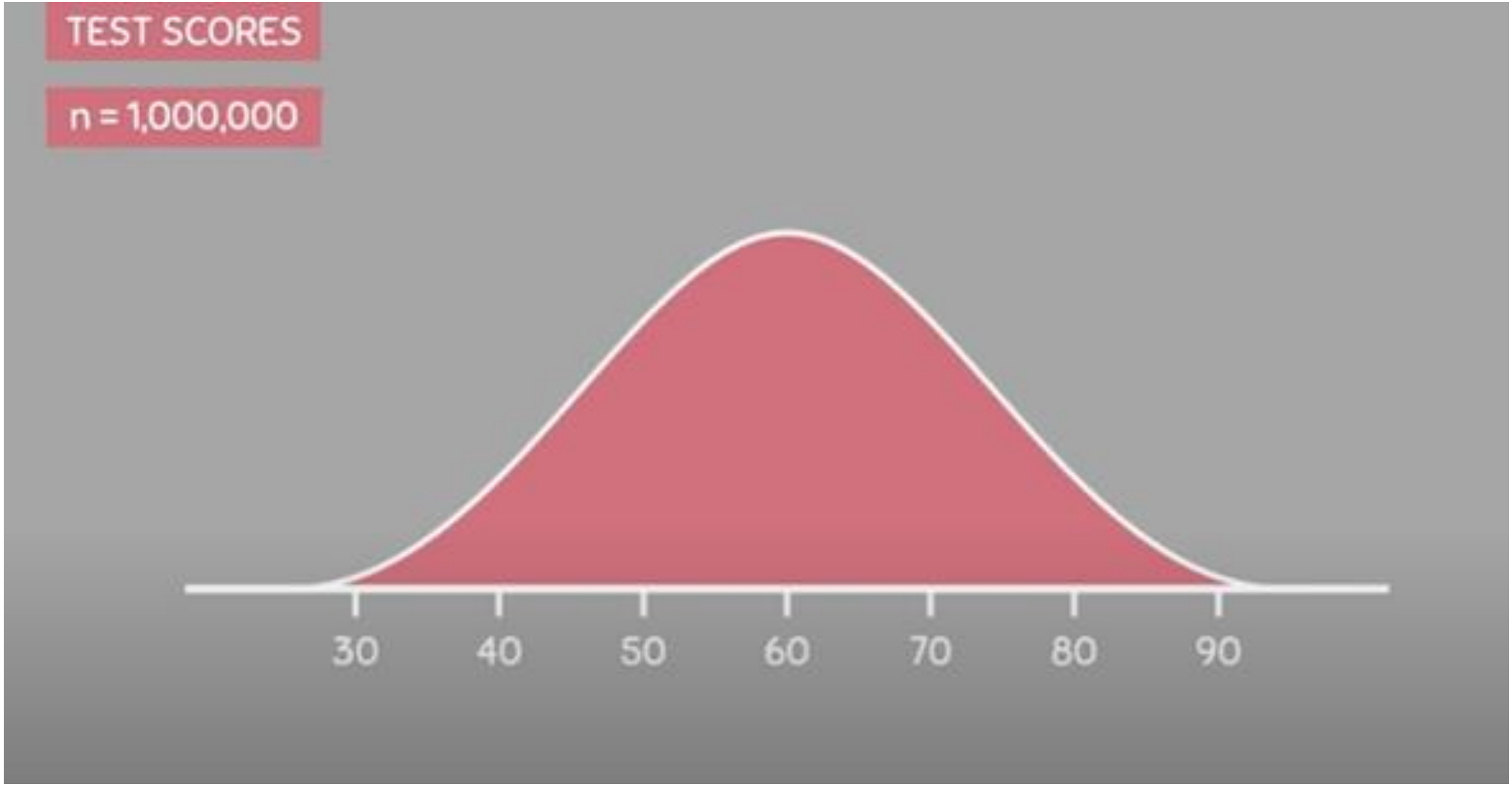
Density plot

- A density curve is a curve on a graph that represents the distribution of values in a dataset .
- A smoothed version of the histogram .




OUTLIER




TEST SCORES
n = 1,000,000
30
40
50
60
70
80
90

Any Questions ?

Now let us moving into practice !!