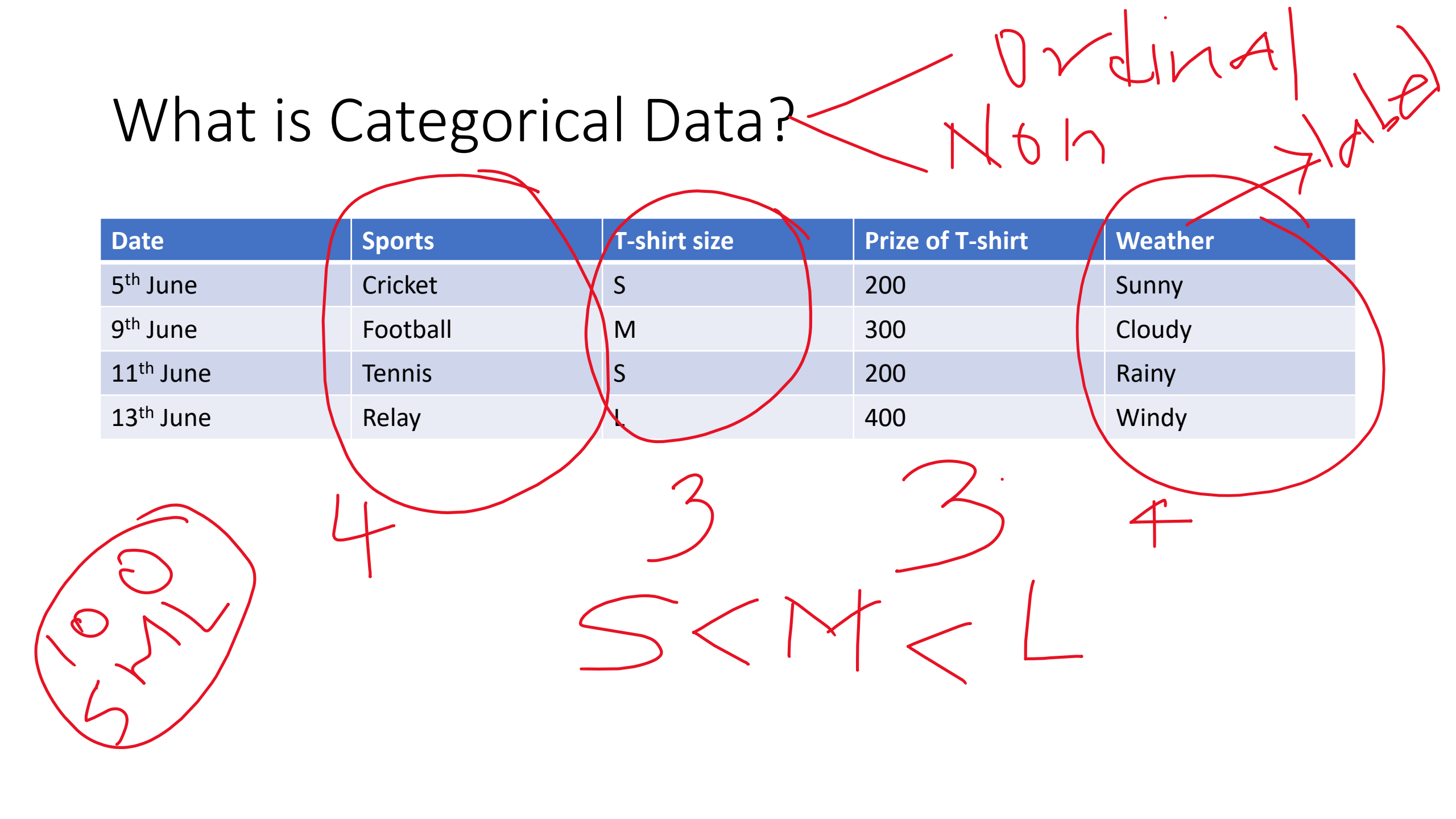
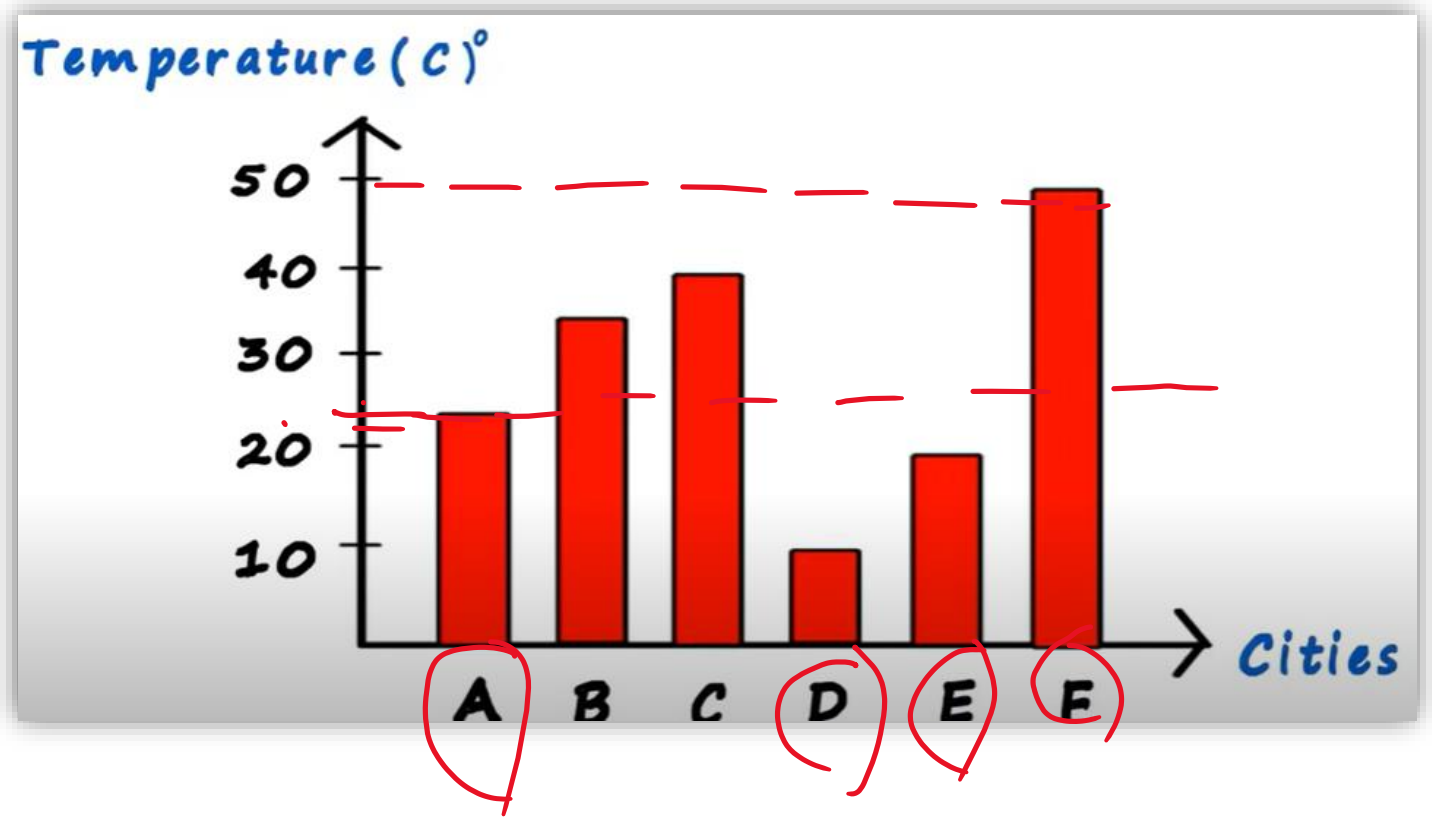
Exploring Categorical Data

What is Categorical Data?

Date	Sports	T-shirt size	Prize of T-shirt	Weather
5th June	Cricket	S	200	Sunny
9th June	Football	M	300	Cloudy
11th June	Tennis	S	200	Rainy
13th June	Relay	L	400	Windy



Bar Graph



1) Hottest — F

temp < 25°

→ warm

(3) H — W

50 - 10

= 40

Expected Value

When the categories can be associated with a numeric value, this gives an average value based on a category’s probability of occurrence.

Example

1. Lisa plays a game in which there are only two outcomes. The cost to play the game is $100. If she wins, she receives $500. The probability of winning is 20%. What is the expected value for winning a single game on average?

$E(X) = X_1P_1 + X_2P_2$

$= 500(.2) + (-100)(.8)$

$= 100 - 80$

$E(X) = \$20$

Out	Win	Loss
V	$500	-$100
P	20%	80%
	0.2	0.8

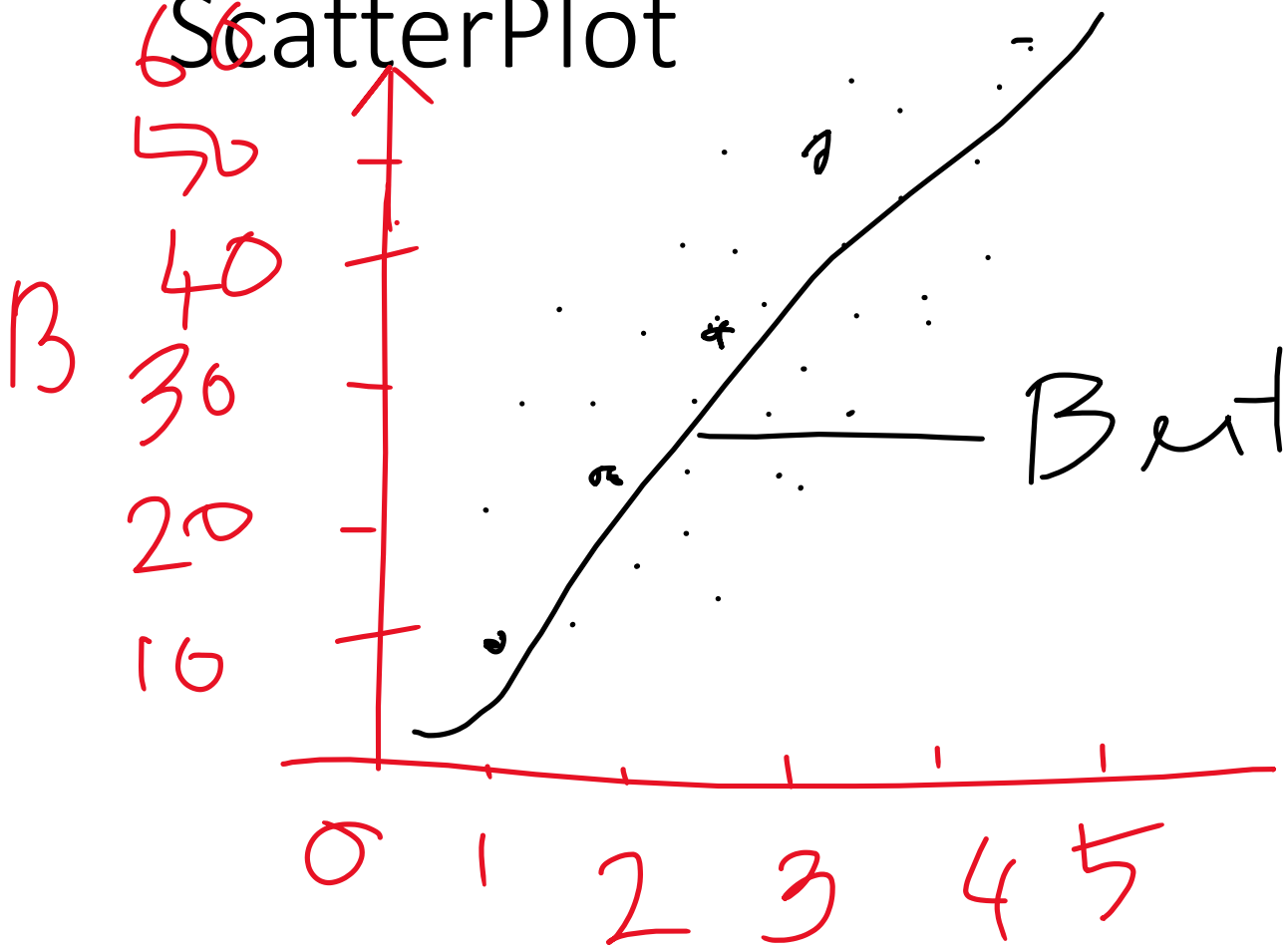
10 → 200

100 → $2000

Correlation

- Correlation is relation between two variables.
- When X increases then Y also increases – positive correlation
- When X decreases then Y also decreases – negative correlation
- When X increases and Y decreases – No correlation

ScatterPlot



A	1	2	3	4	5	6
B	10	25	35	50	60	46

positively.

Positive Correlation, Negative Correlation and No Correlation