upx
Move up in life

Welcome to Session 6 of DS Foundation Course!

upx
Move up in life

What is the probability (chance) that this hamburger weighs exactly 0.25 pounds?

0

Recap of Session 5

- Importance of Central Tendency
- Calculation of Mean, Median and Mode
- Standard Deviation and Variance
- Confidence Interval
- Interquartile Range

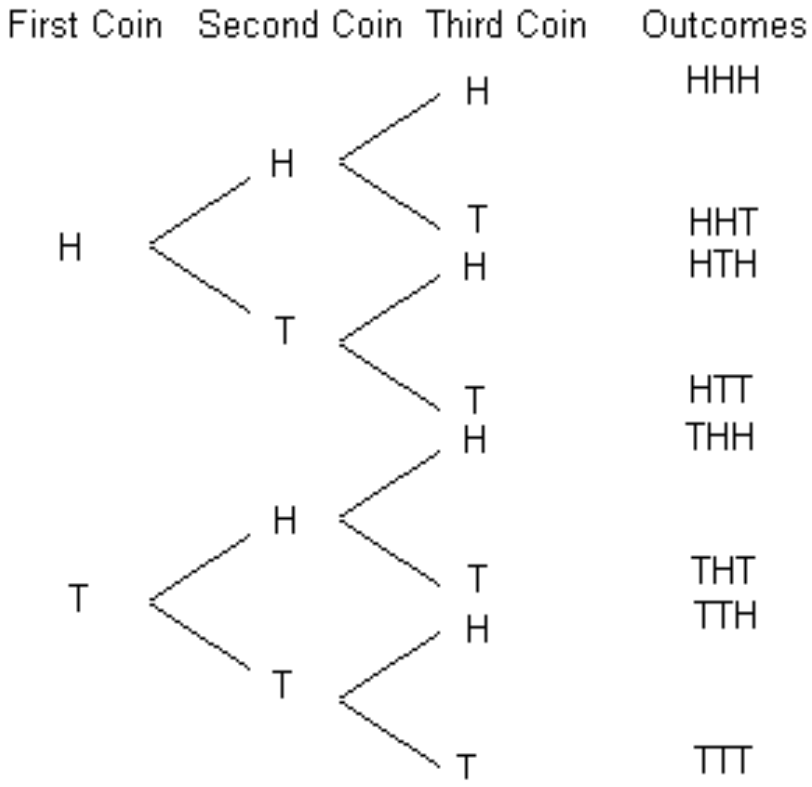
Probability, Normal Distribution and Pictorial Representations

Where does Sarah stand?

- Sarah is a student of English literature
- Class of 50 students complete an assignment
- Sarah scores 70
- Mean is 60; Standard deviation is 15
- Whilst Sarah has still scored much higher than the mean score, she has not necessarily achieved one of the best marks in her class.

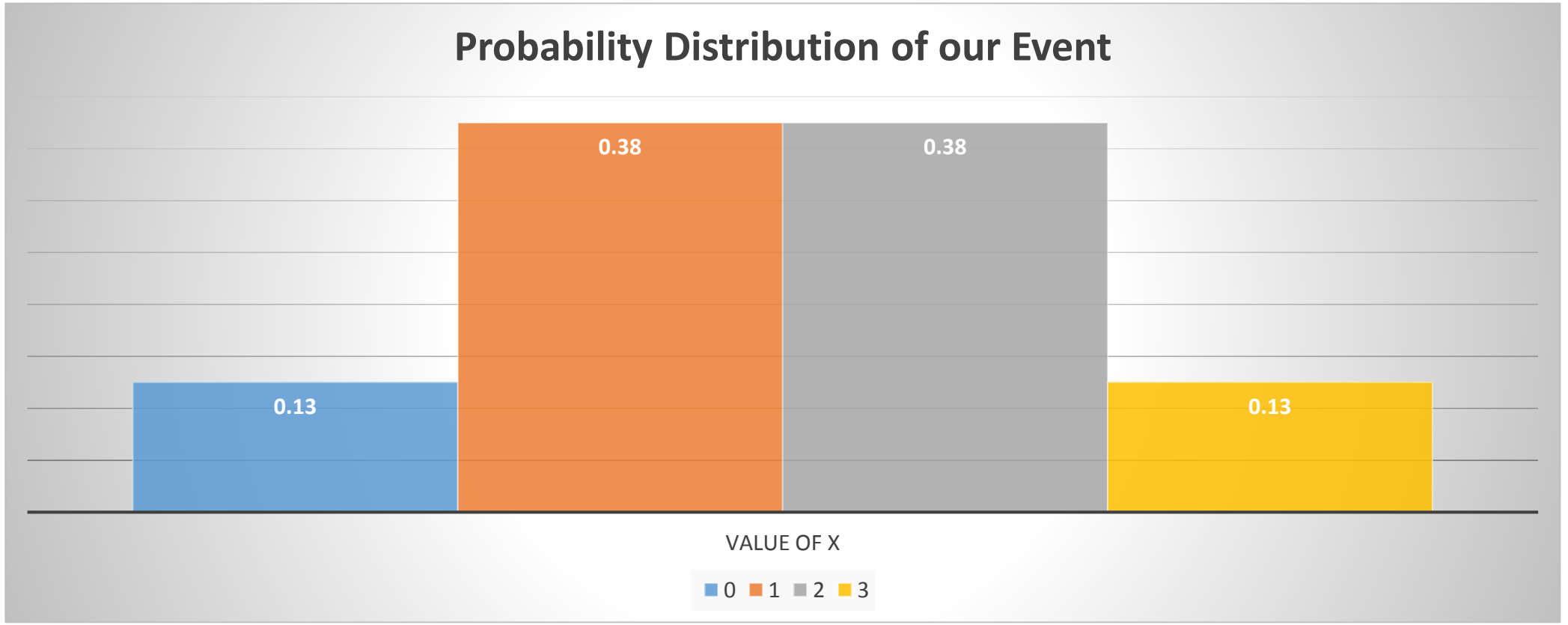
- How well did Sarah perform in her English literature work?

Probability



- X = Number of “heads” after 3 flips of a fair coin.
- P(X=0) = 1/8
- P(X=1) = 3/8
- P(X=2) = 3/8
- P(X=3) = 1/8

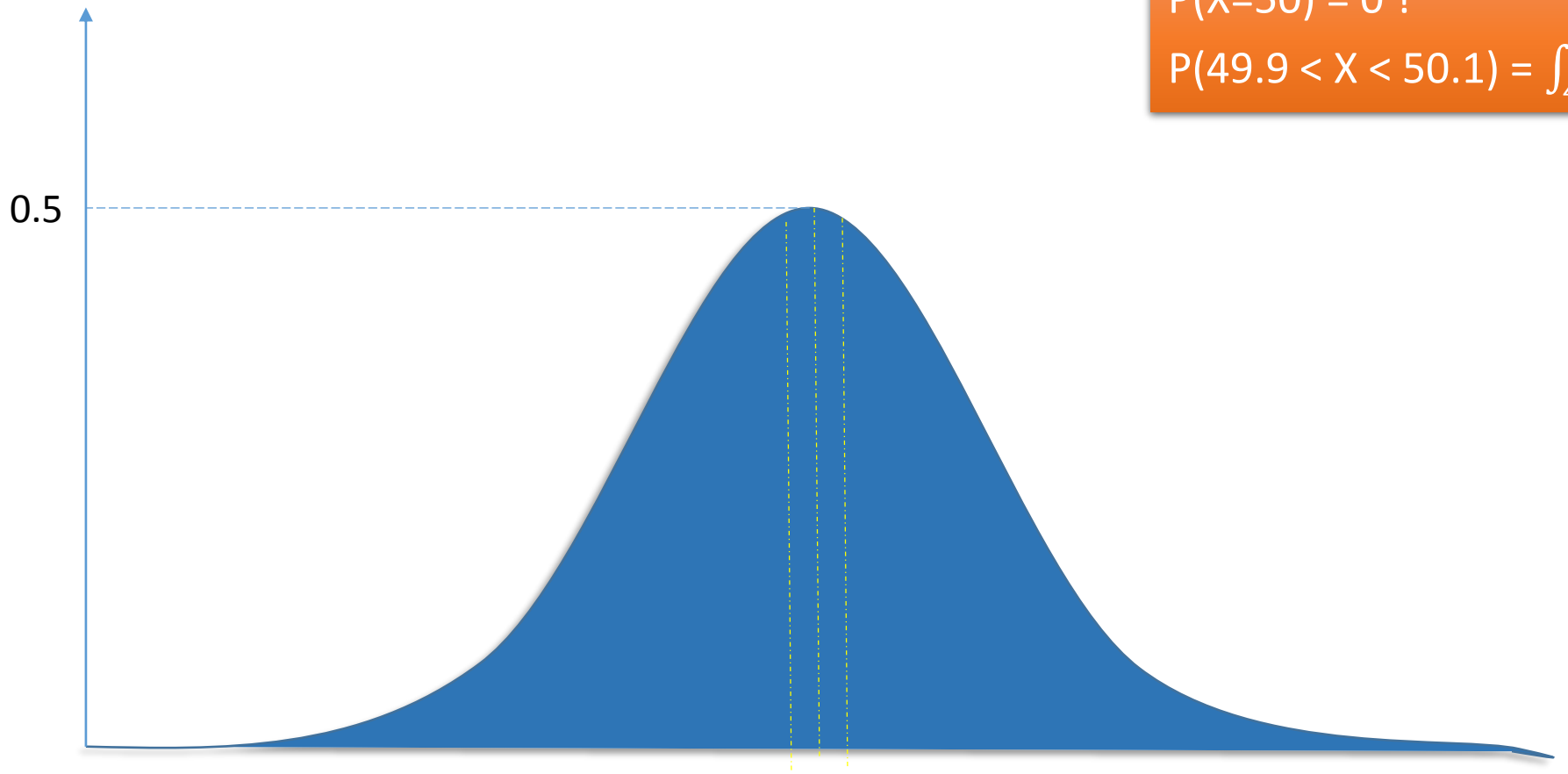
Probability Distribution



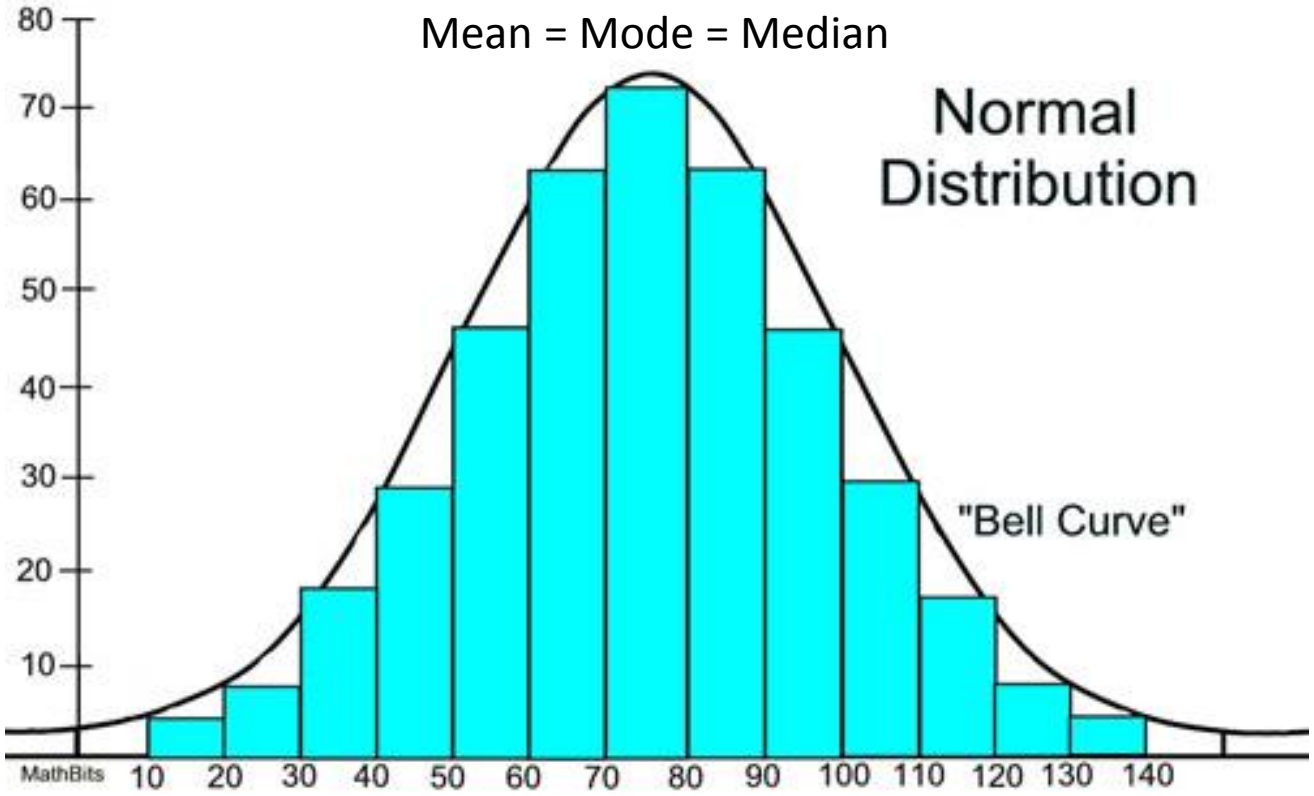
Probability Density Function

P(X=50) = 0 ?

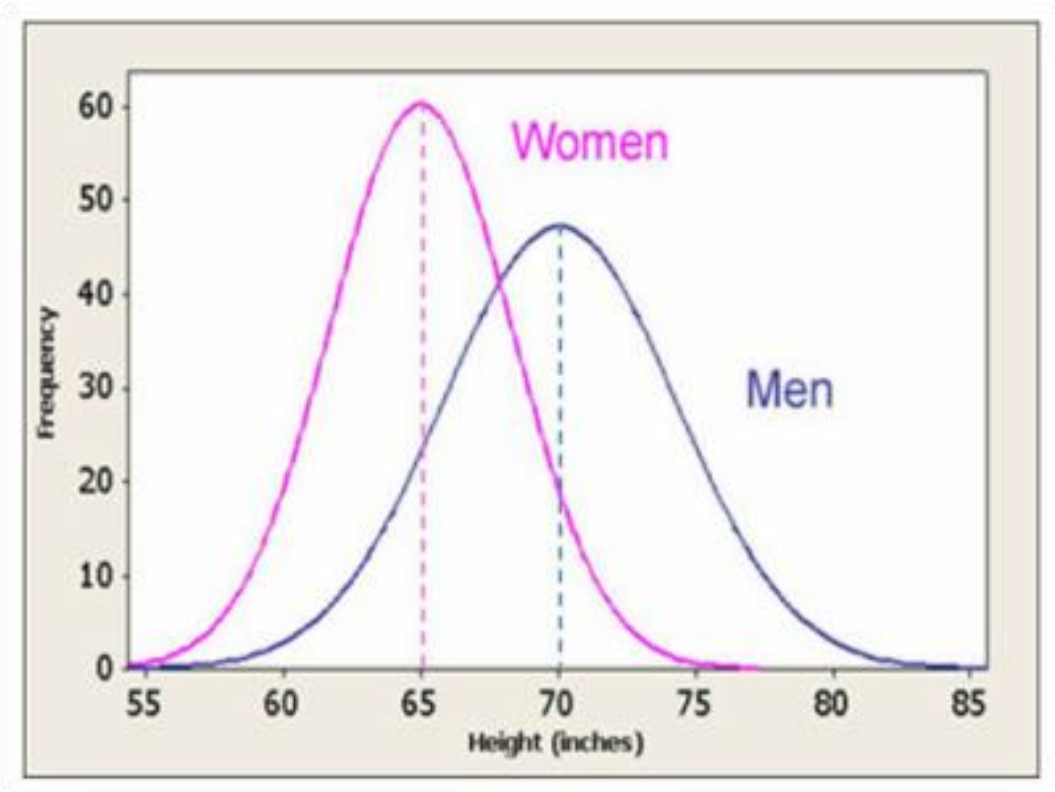
P(49.9 < X < 50.1) = $\int_{49.9}^{50.1} p(x)dx$



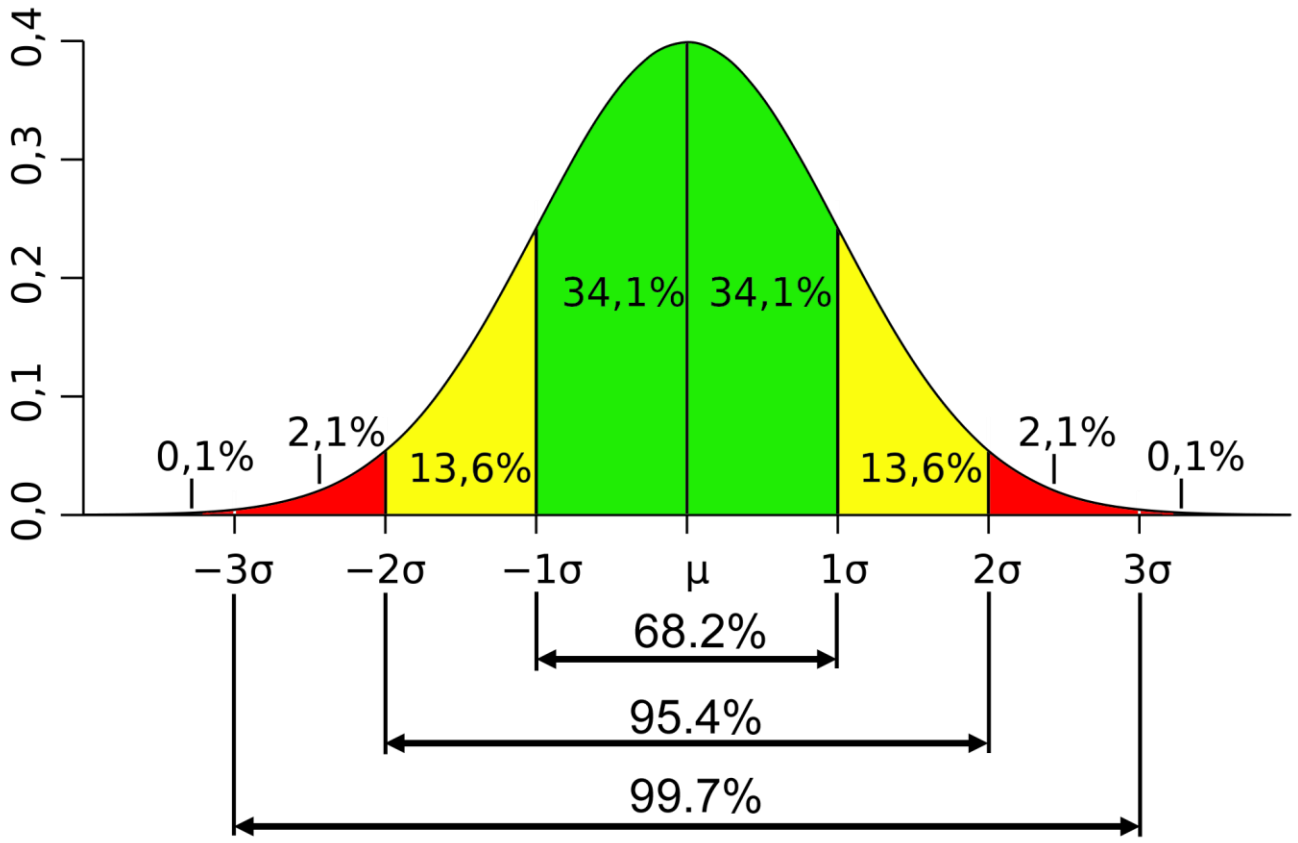
Normal Distribution



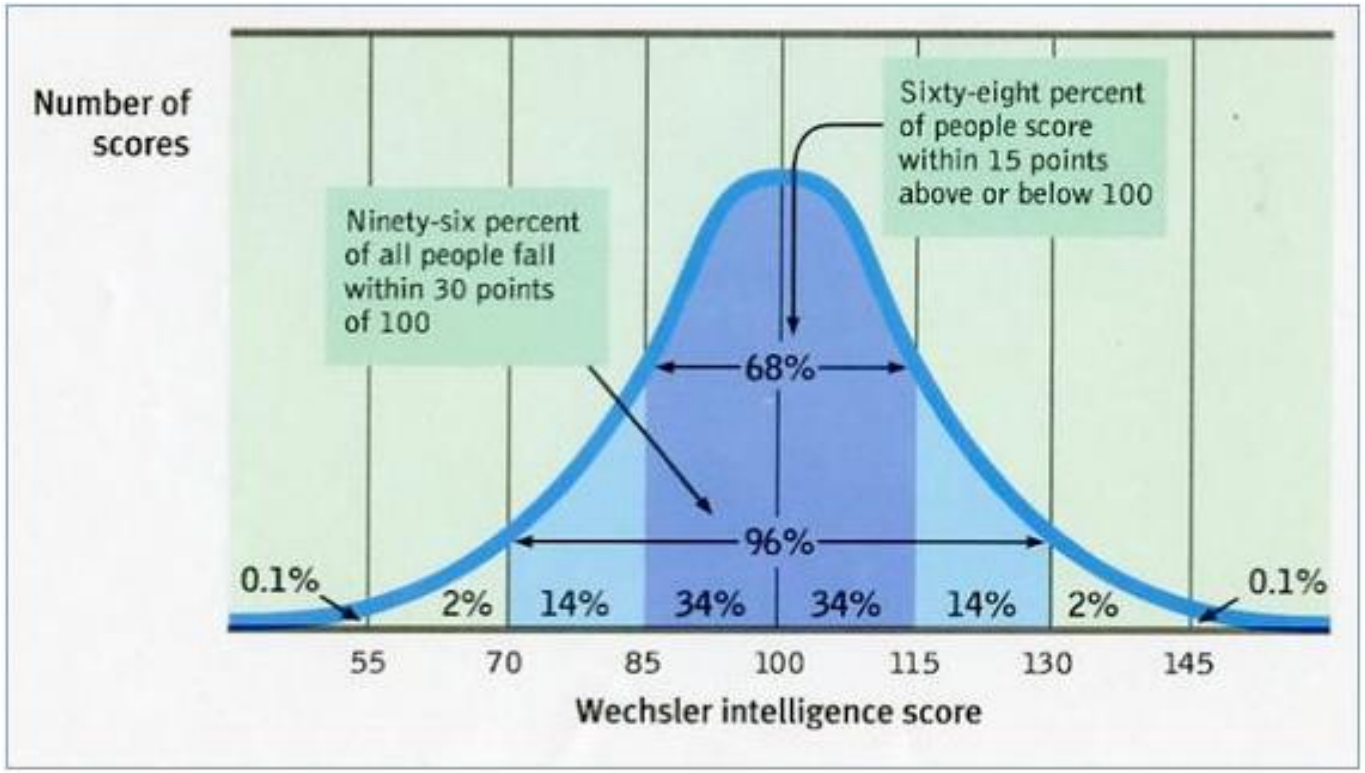
Examples


Women
Men
Frequency
Height (inches)
0
10
20
30
40
50
60
55
60
65
70
75
80
85

Normal Distribution & Std. Deviation



Example


Number of scores
Sixty-eight percent of people score within 15 points above or below 100
Ninety-six percent of all people fall within 30 points of 100
68%
96%
0.1%
2%
14%
34%
34%
14%
2%
0.1%
55
70
85
100
115
130
145
Wechsler intelligence score

Z score

$$z = \frac{\text{raw score} - \text{mean}}{\text{standard deviation}}$$

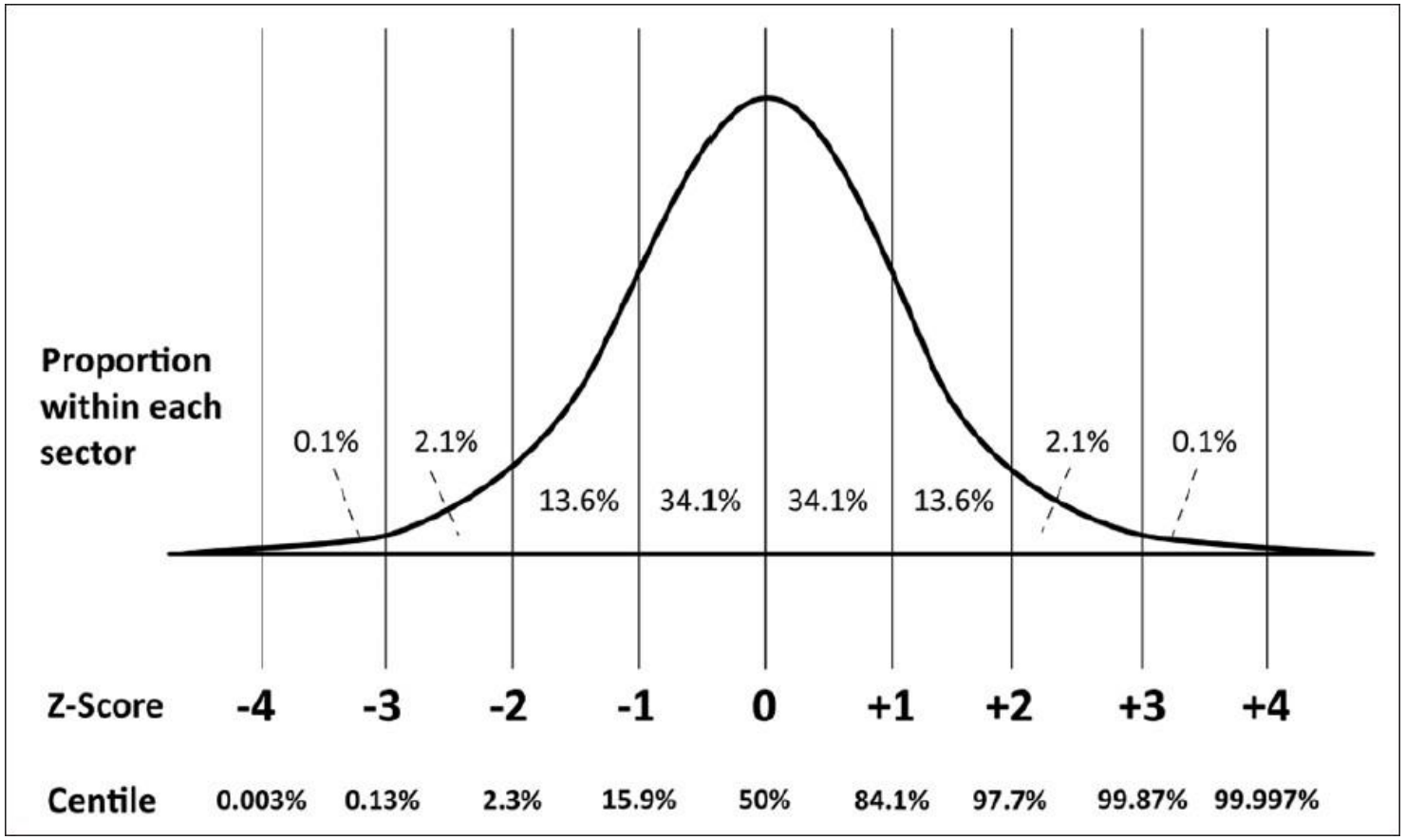
or

$$z = \frac{X - \mu}{\sigma}$$

or

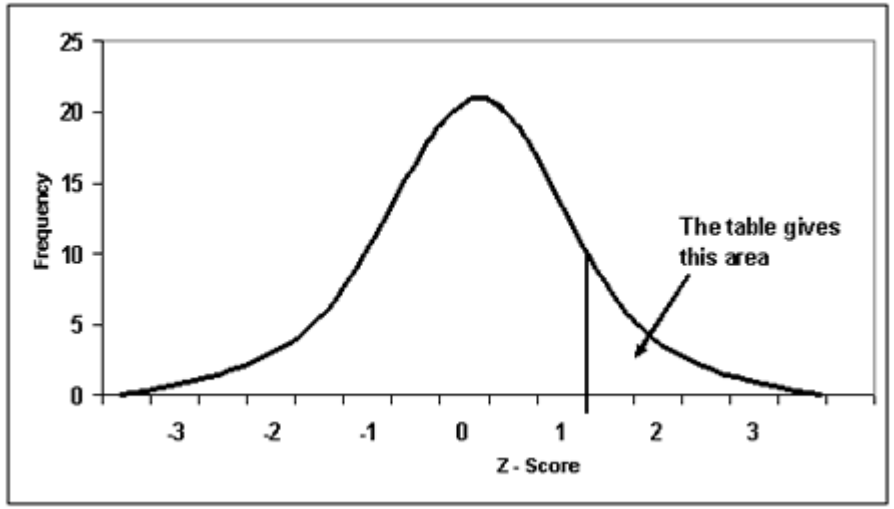
$$z = \frac{X - \bar{X}}{s}$$

where X = raw score
μ = population mean
σ = population standard deviation
$\bar{X}$ = sample mean
s = sample standard deviation



Sarah's position unravelled

- Sarah's Z-score is 0.67
- Looking at the Z-table, we get the value for 0.67 as 0.2514
- This means that the probability of a score being greater than 0.677 is 0.2514.
- If we look at this as a percentage, we simply times the score by 100; hence 0.2514 x 100 = 25.14%.
- In other words, around 25% of the class got a better mark than Sarah (roughly 13 students since there is no such thing as part of a student!).

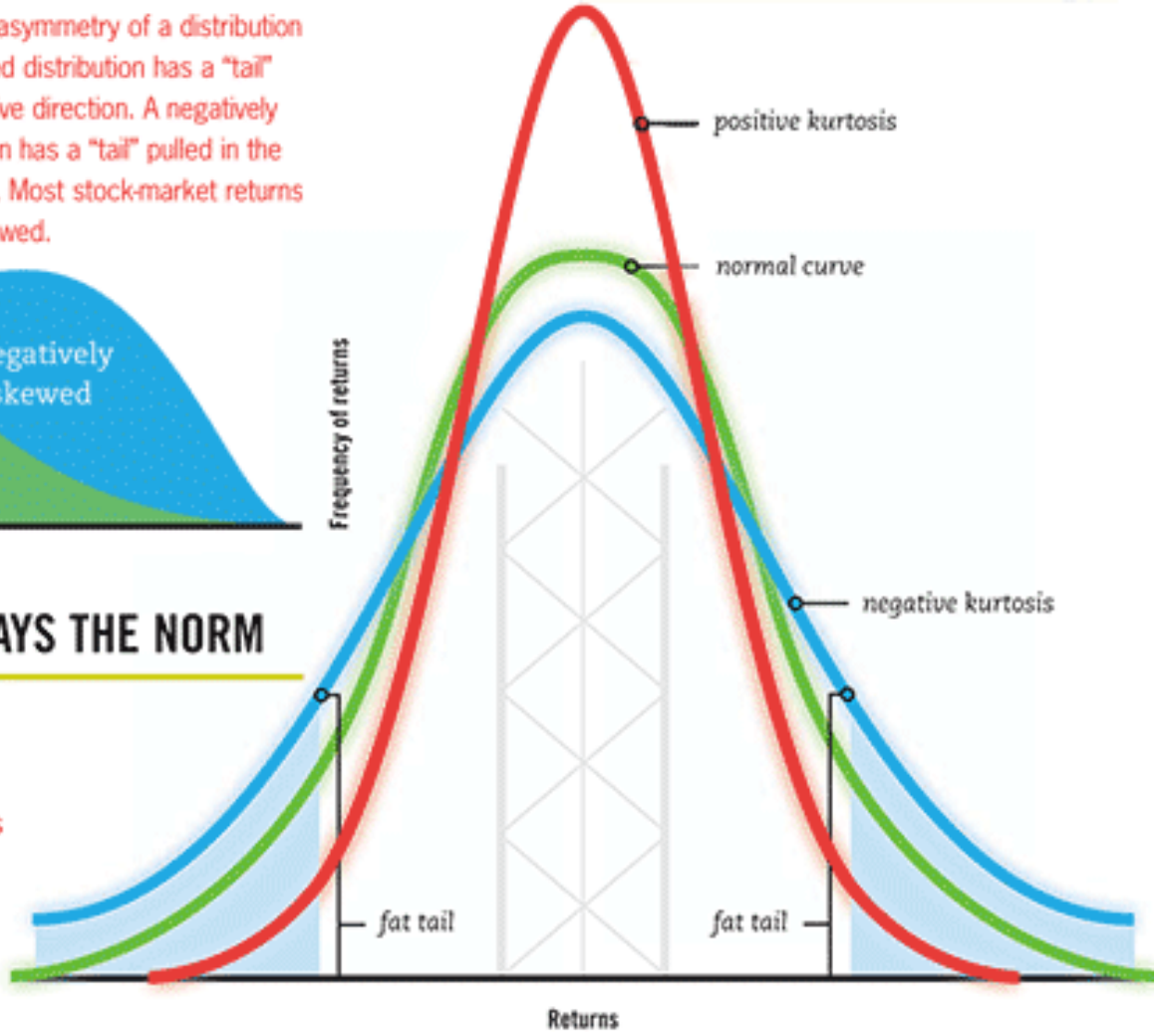


Skewness and Kurtosis

Skewness is the asymmetry of a distribution A positively skewed distribution has a "tail" pulled in the positive direction. A negatively skewed distribution has a "tail" pulled in the negative direction. Most stock-market returns are negatively skewed.

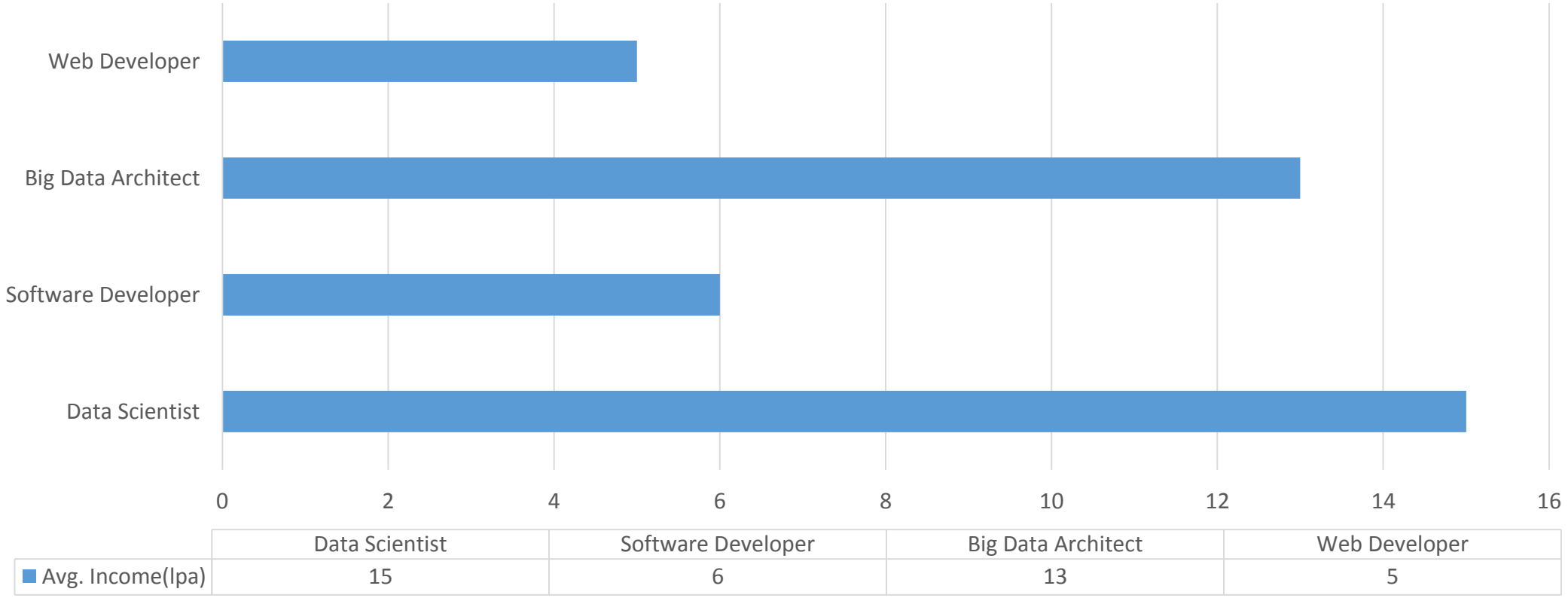
NORMAL NOT ALWAYS THE NORM

Kurtosis refers to how peaked the curve is: steeper means positive kurtosis and flatter means negative kurtosis. Fat tails occur when there are more outsize returns on the downside or upside, or both, than the normal curve suggests.



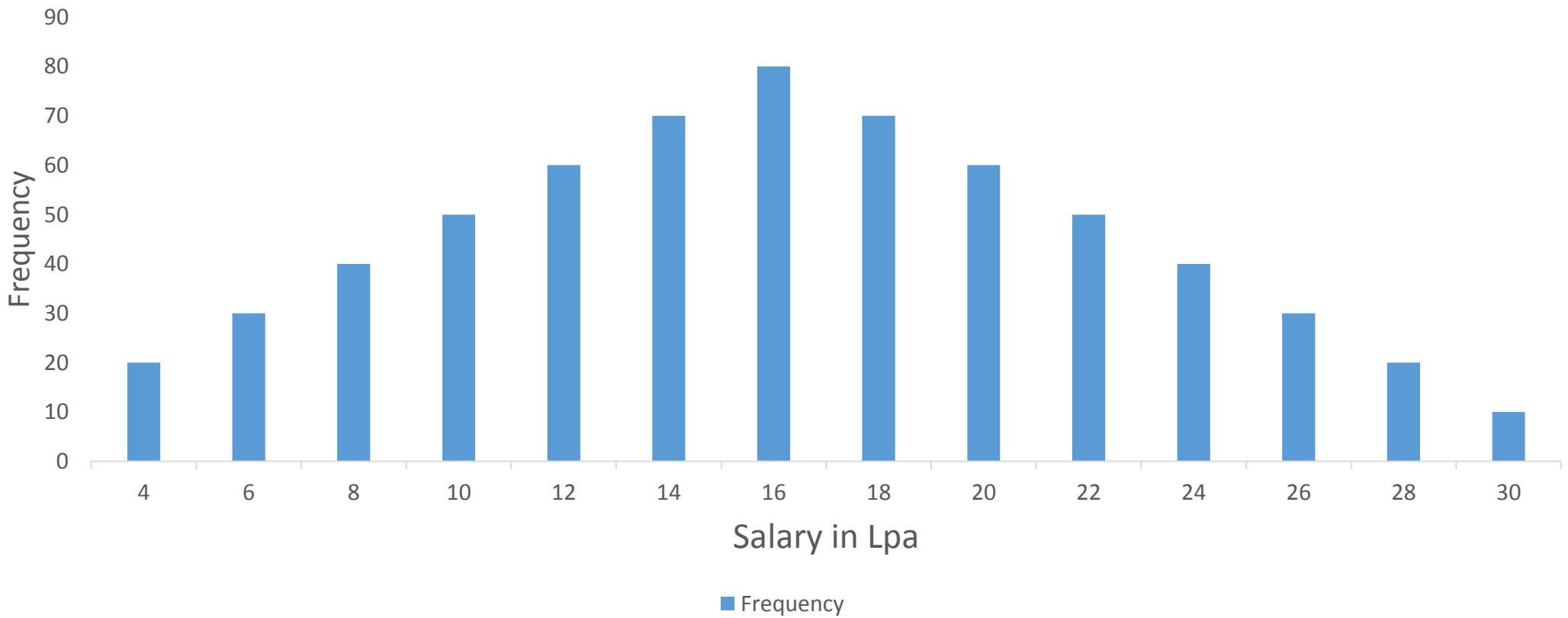
Pictorial Representations

Bar Graph

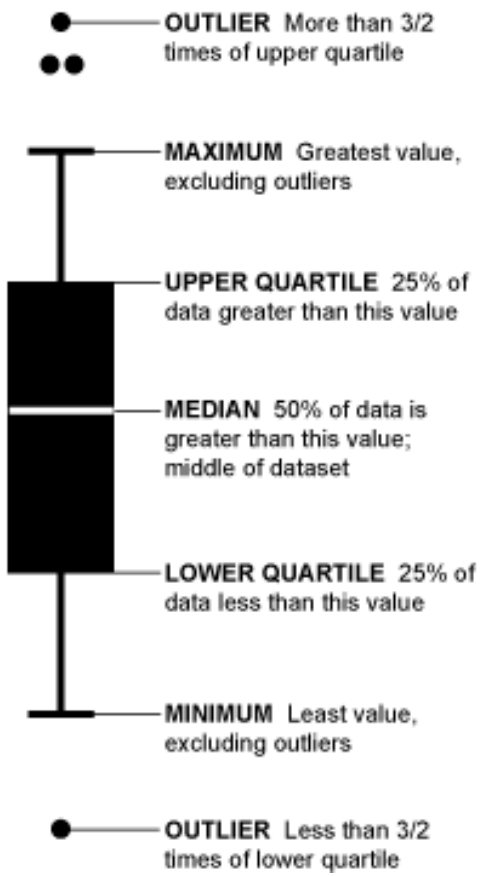


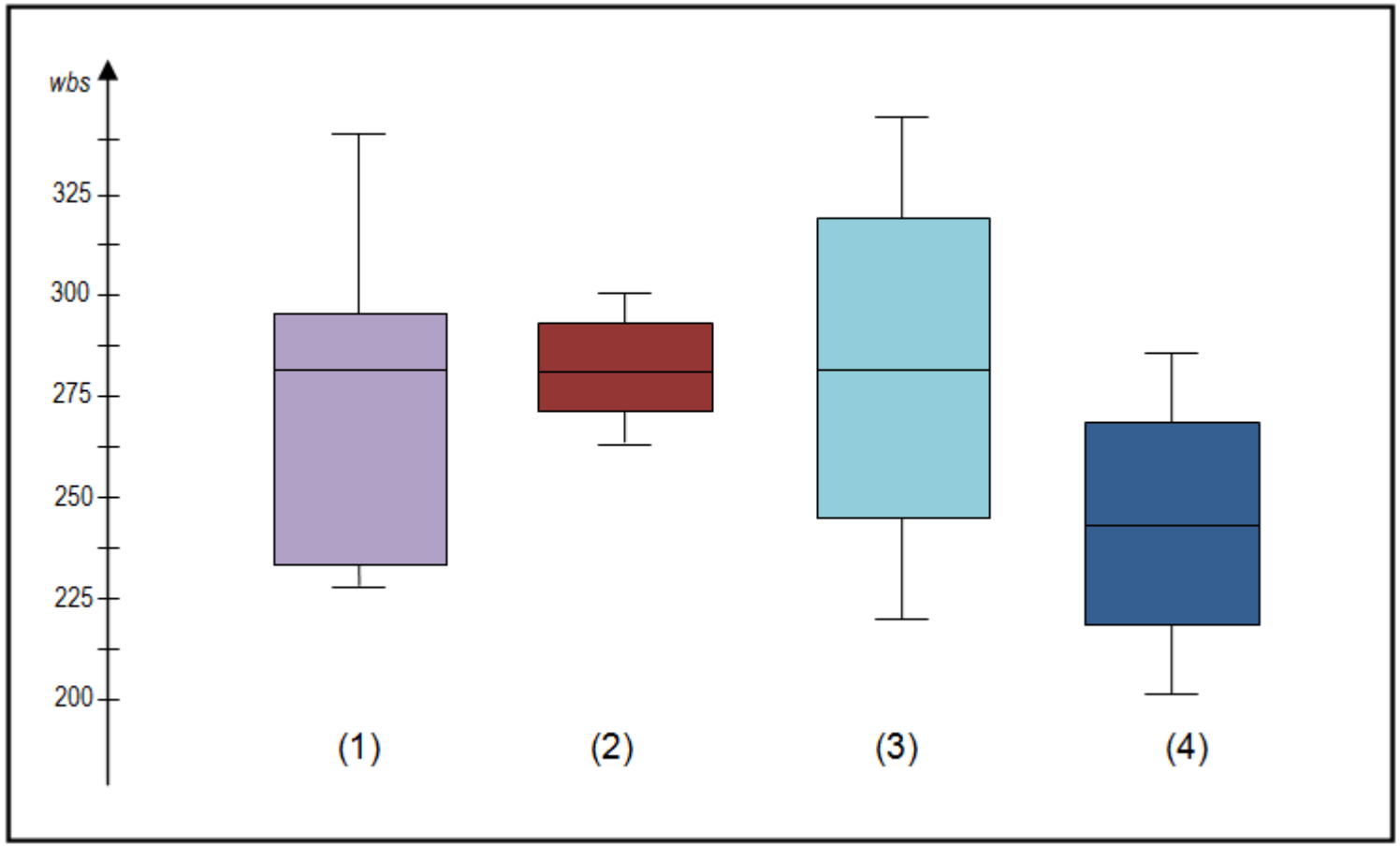
	Data Scientist	Software Developer	Big Data Architect	Web Developer
Avg. Income(lpa)	15	6	13	5

Histogram


90
80
70
60
50
40
30
20
10
0
Frequency
4
6
8
10
12
14
16
18
20
22
24
26
28
30
Salary in Lpa
Frequency

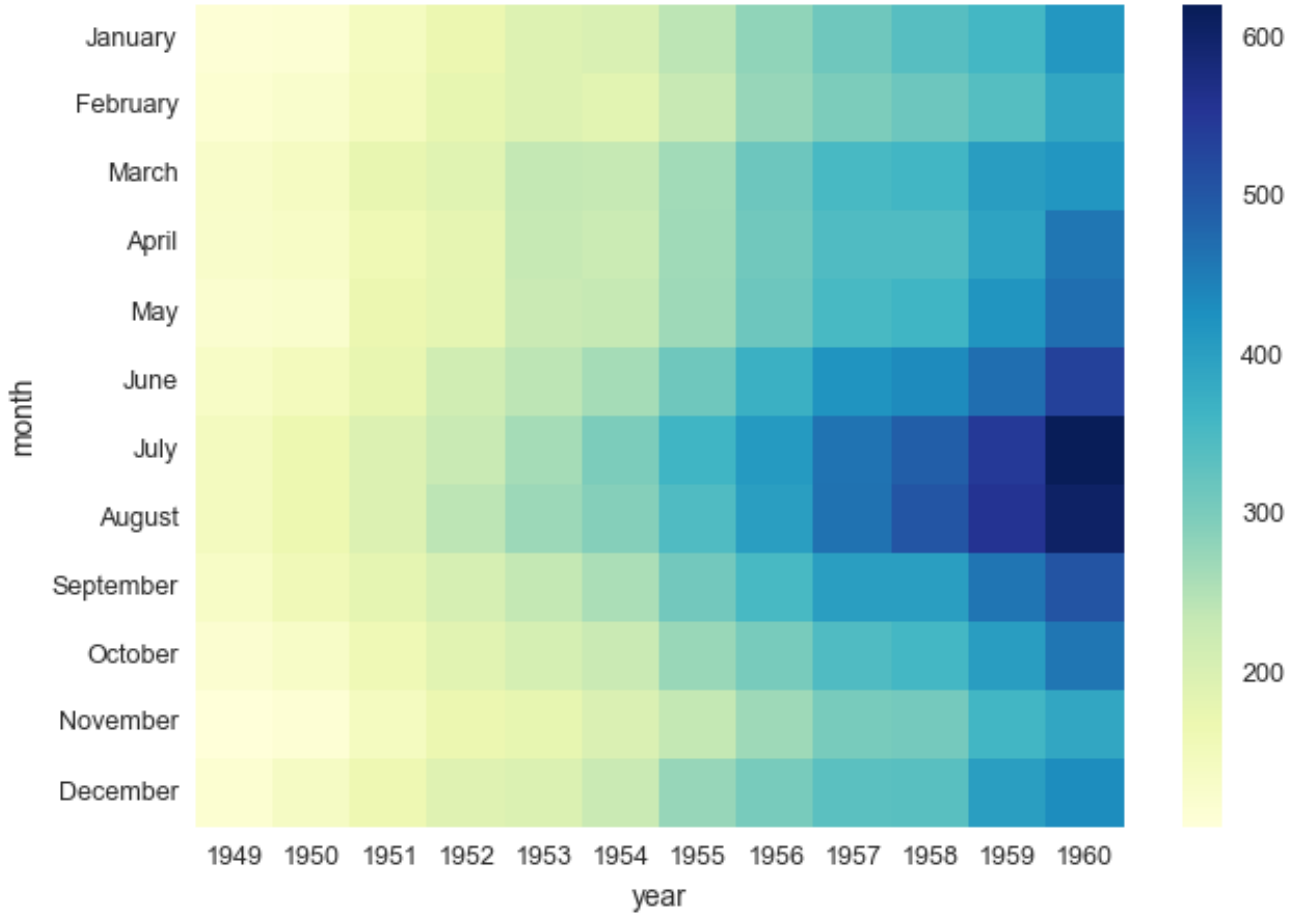
Box Plot


OUTLIER More than 3/2 times of upper quartile
MAXIMUM Greatest value, excluding outliers
UPPER QUARTILE 25% of data greater than this value
MEDIAN 50% of data is greater than this value; middle of dataset
LOWER QUARTILE 25% of data less than this value
MINIMUM Least value, excluding outliers
OUTLIER Less than 3/2 times of lower quartile


wbs
325
300
275
250
225
200
(1)
(2)
(3)
(4)

Time-series Plot

Heat Maps


January
February
March
April
May
June
July
August
September
October
November
December
month
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
year
600
500
400
300
200