L- 2: Descriptive Statistics

Today…..

- Recall the past for a while_ Simple tools
- Visualization of data
- Basics of probability
- Discussion & Problems on probability
- Conditional probability

Visualization

- Summary gives an idea about the data

  summary(income)

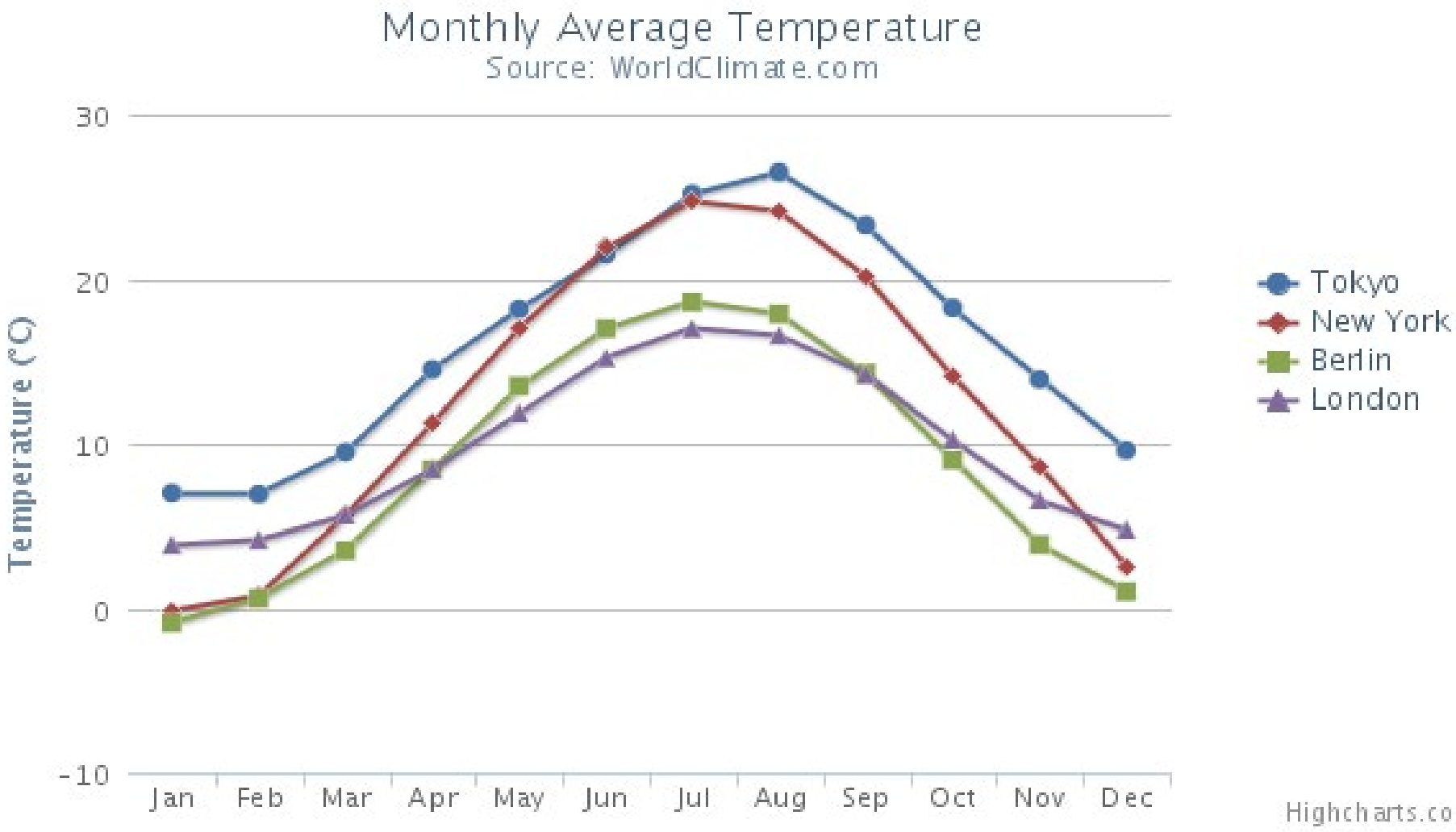
Min	*1st QU.*	*Median*	*Mean*	*3rd Qu*	*Max*
- 7.8	*12.5*	*32.0*	*52.03*	*67.2*	*585*

- Visualization – why

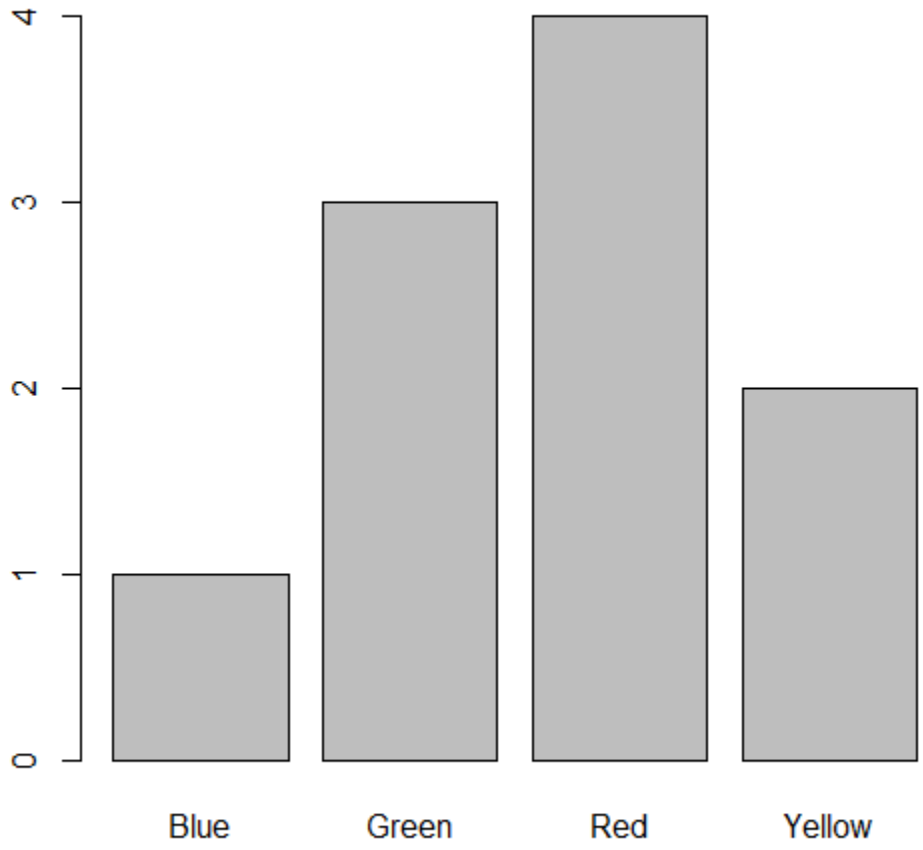
Data Visualisation

- Line chart
- Bar chart
- Histogram
- Pie chart
- Scatter plot
- Box plot

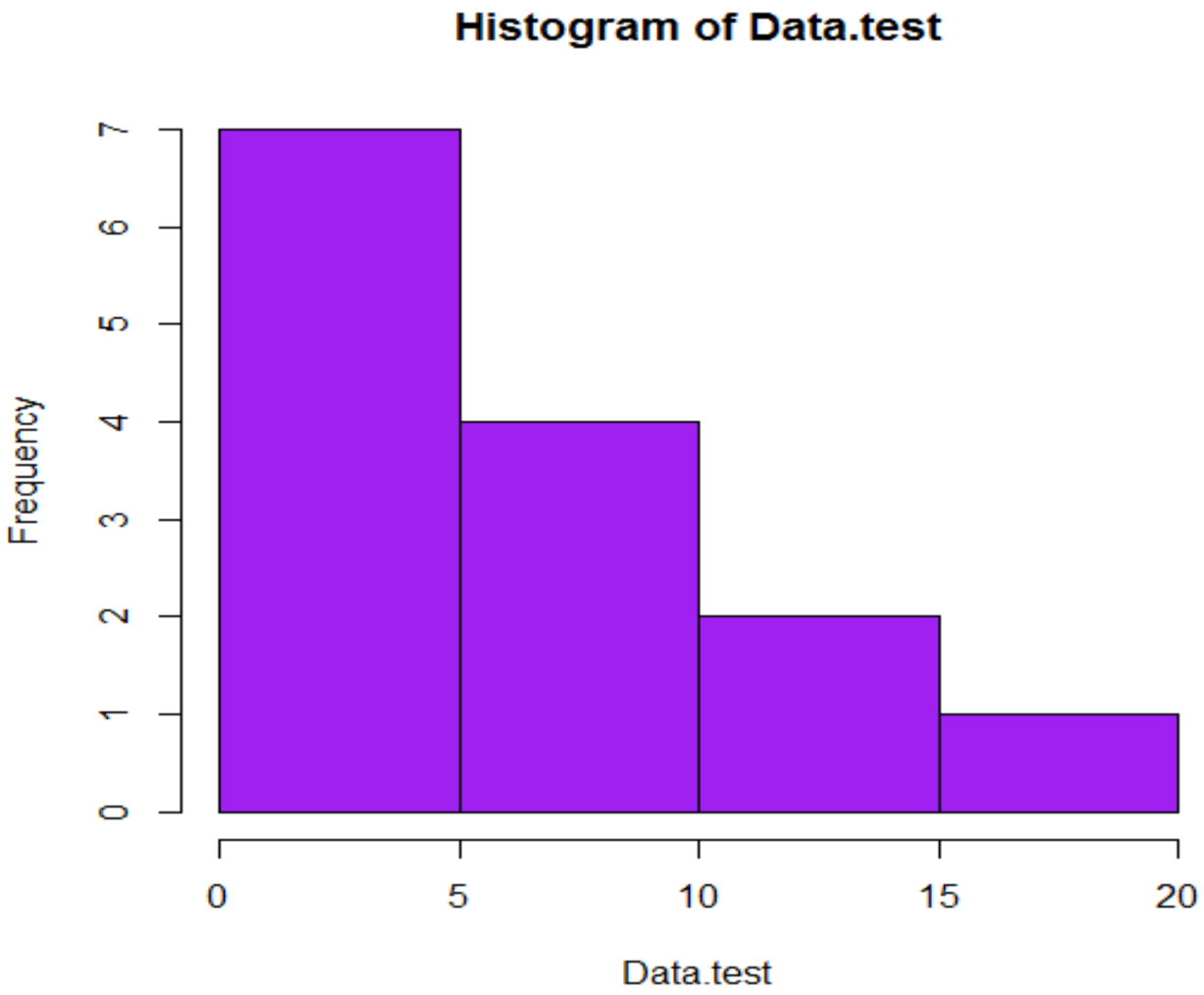
Line Chart


Monthly Average Temperature
Source: WorldClimate.com
Temperature (°C)
30
20
10
0
-10
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Tokyo
New York
Berlin
London
Highcharts.com

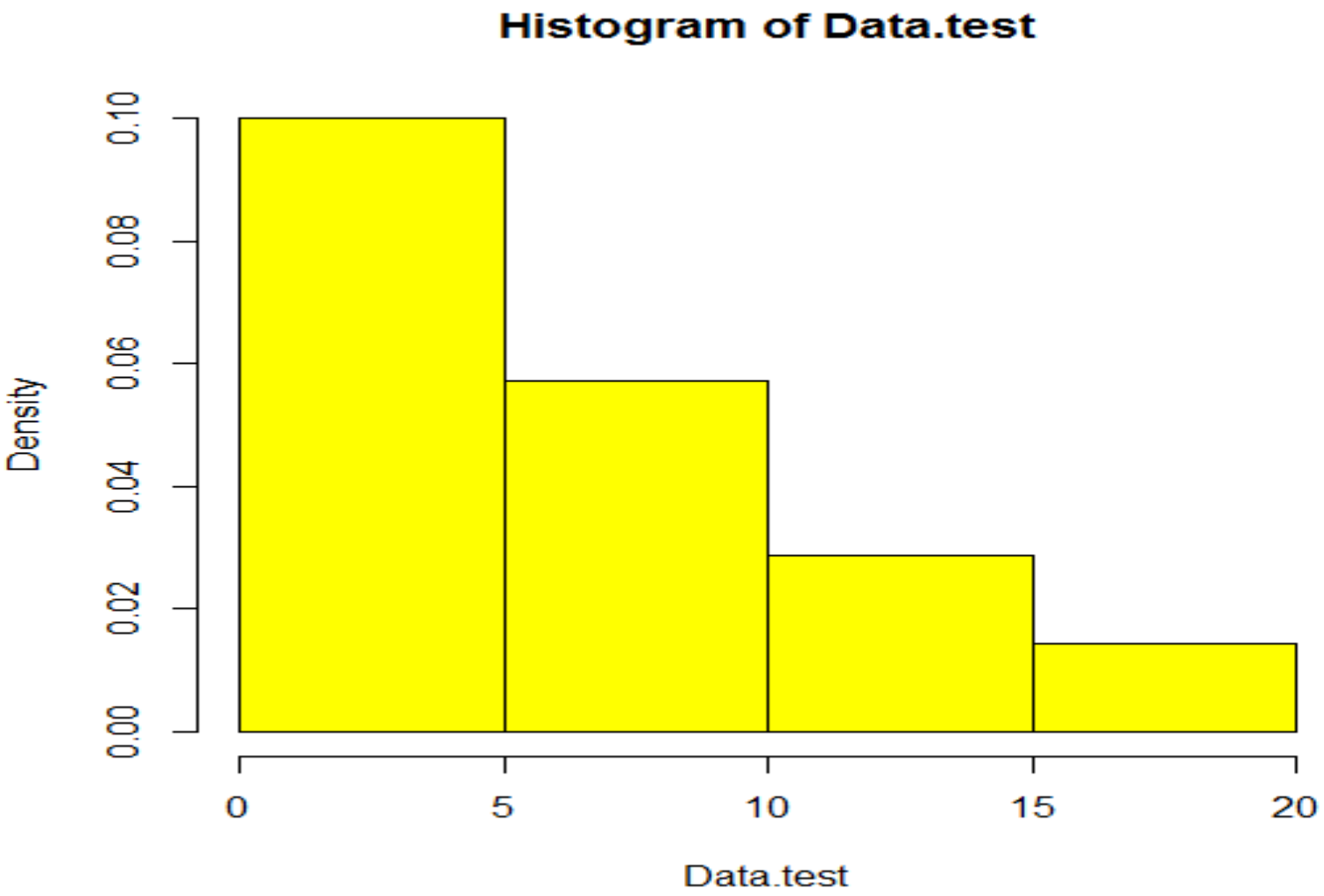
Bar Chart



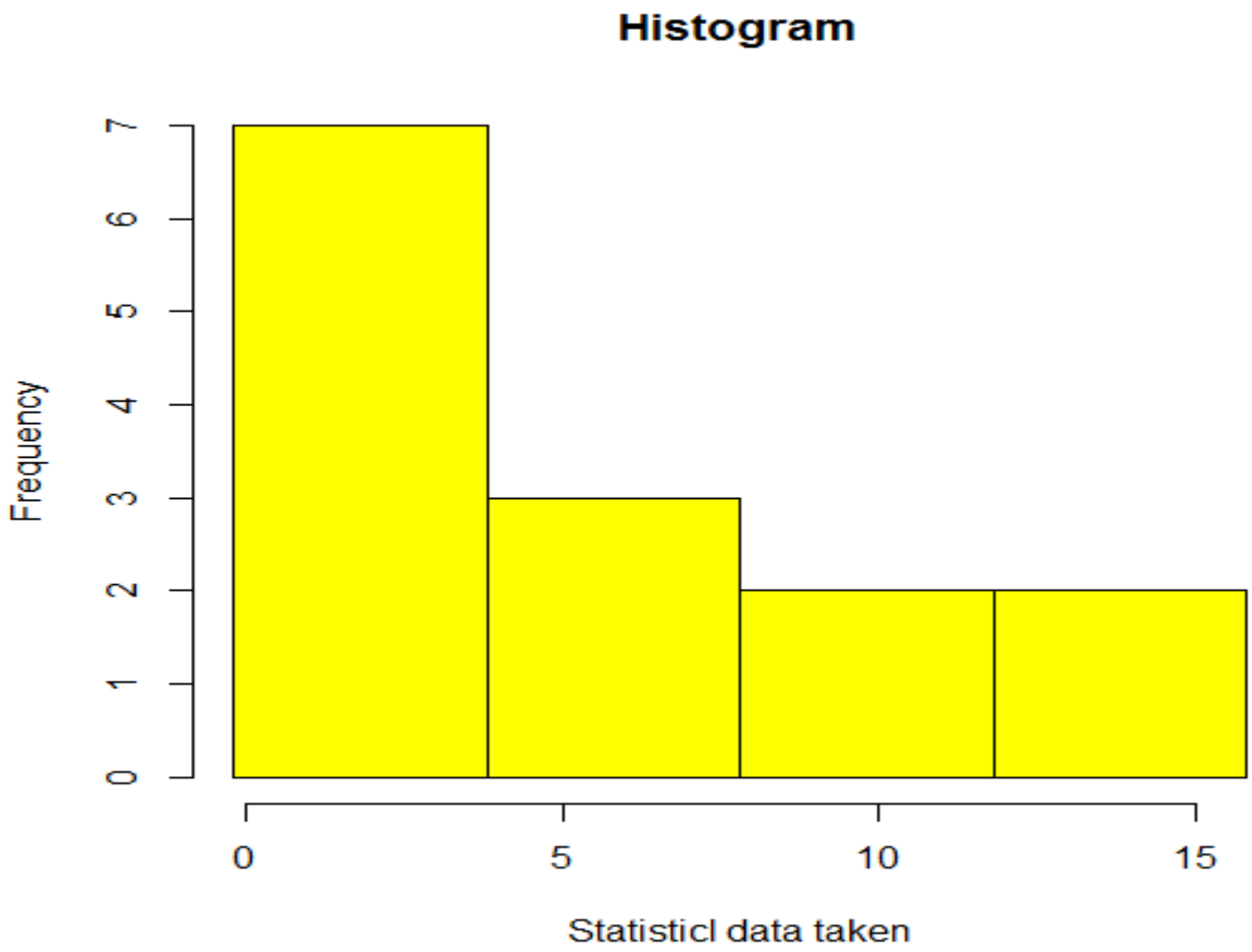
Histograms


Histogram of Data.test
Frequency
0
1
2
3
4
5
6
7
0
5
10
15
20
Data.test

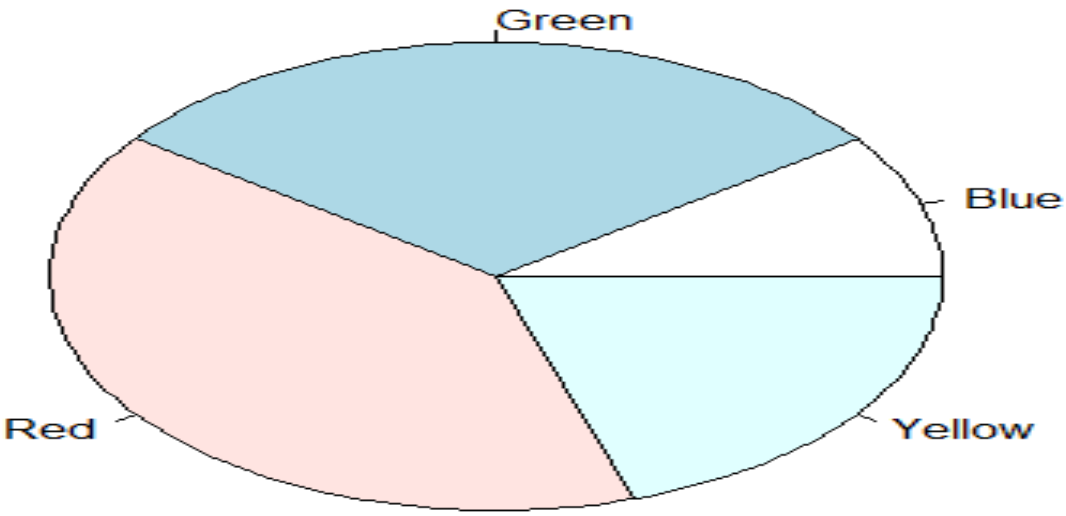
Histograms


Histogram of Data.test
Density
0.00
0.02
0.04
0.06
0.08
0.10
0
5
10
15
20
Data.test

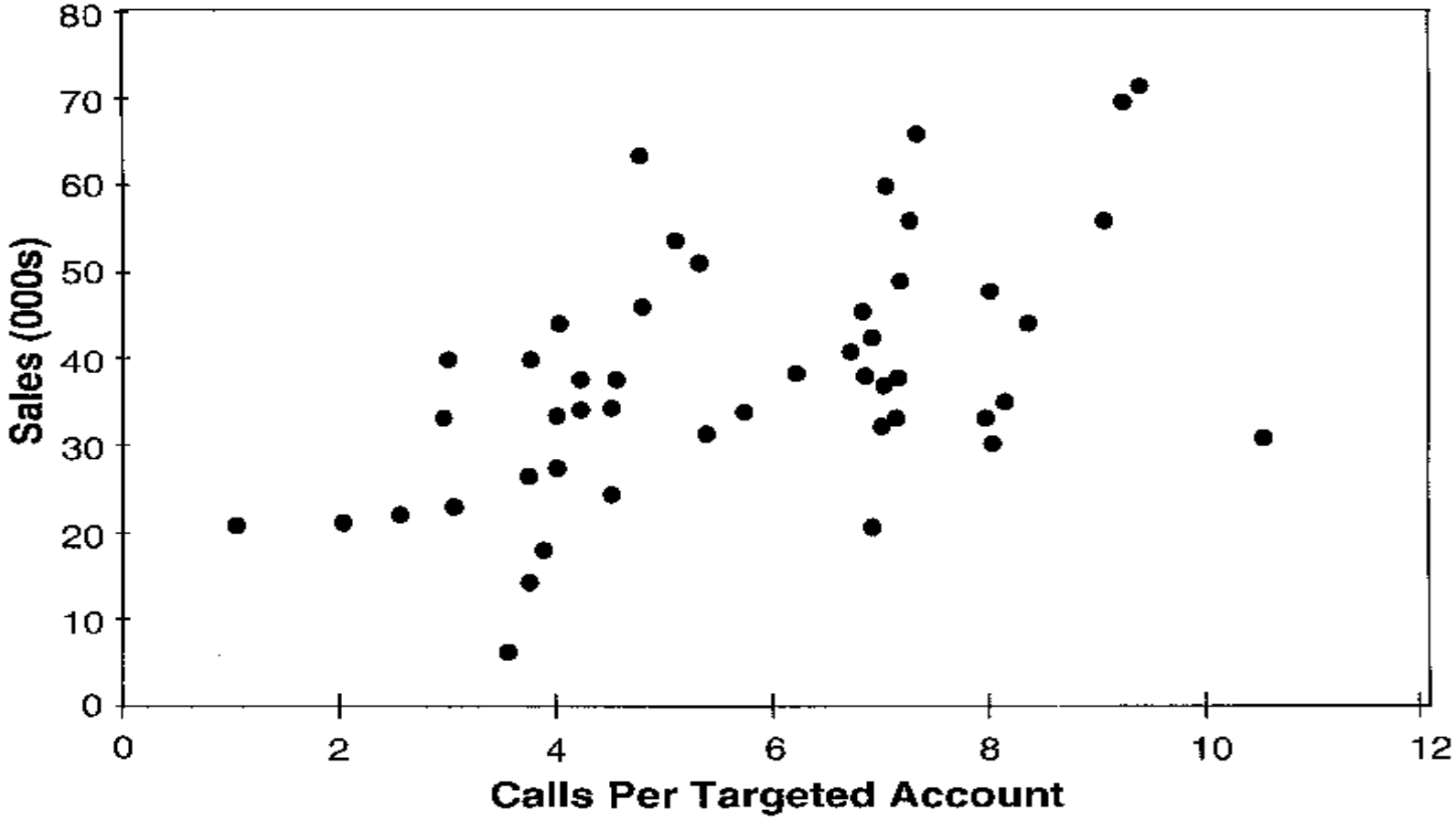
Histograms


Histogram
Frequency
0
1
2
3
4
5
6
7
0
5
10
15
Statisticl data taken

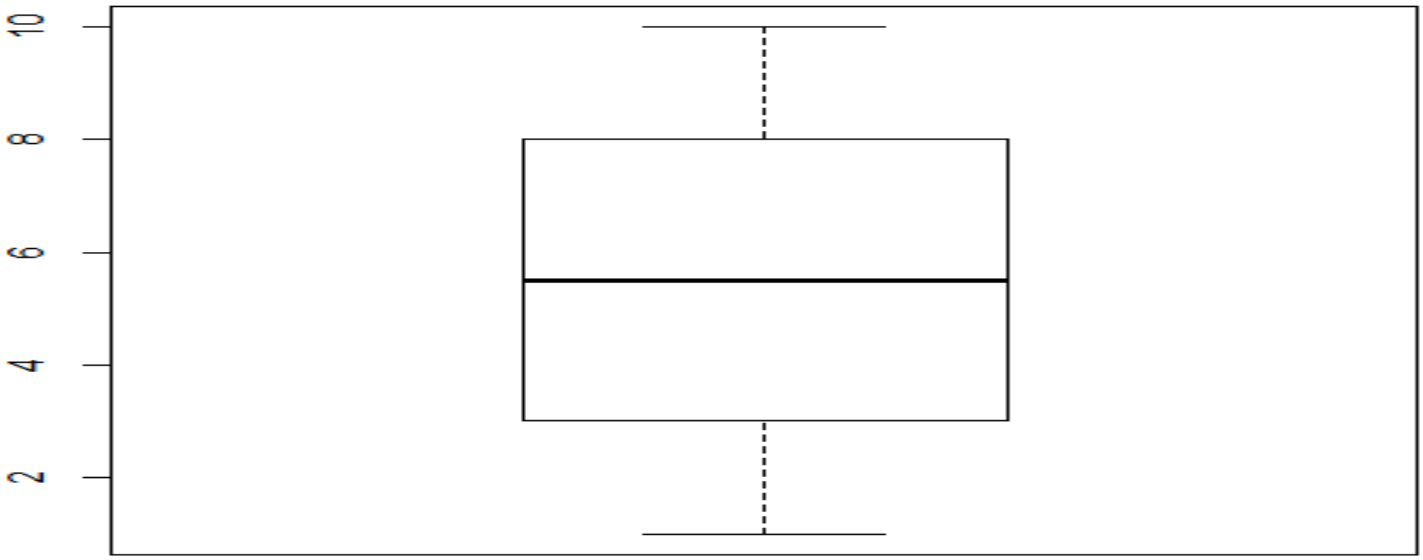
Pie charts



Scatter Plot



Box plot



To conclude _ Visualization

- Visualization gives a sense of data distribution and relationship among variables

- Visualization is an iterative process and helps answer questions about the data. Time spent is not wasted during the modelling process and helps to find the optimal model to fit the data

!!!!

A famous statistician would never travel by airplane, because she had studied air travel and estimated the probability of there being a bomb on any given flight was 1 in a million, and she was not prepared to accept these odds.

One day a colleague met her at a conference far from home.

"How did you get here, by train?"

"No, I flew"

"What about the possibility of a bomb?"

"Well, I began thinking that if the odds of one bomb are 1:million, then the odds of TWO bombs are (1/1,000,000) x (1/1,000,000) = 10^{-12}. This is a very, very small probability, which I can accept. So, now I bring my own bomb along!"

Random Experiment

Term "**random experiment**" is used to describe any action whose outcome is not known in advance. Here are some examples of experiments dealing with statistical data:

- Tossing a coin
- Counting how many times a certain word or a combination of words appears in the text of the “King Lear” or in a text of Confucius
- counting occurrences of a certain combination of amino acids in a protein database.
- pulling a card from the deck

Sample Space

- Discrete sample spaces.
- Continuous sample spaces

- Event

- Independent events
- Dependent events

Probability

Axioms of Probability

Probability is a number that is assigned to each member of a collection of events from a random experiment that satisfies the following properties:

If S is the sample space and E is any event in a random experiment,

(1) $P(S) = 1$

(2) $0 \leq P(E) \leq 1$

(3) For two events E_1 and E_2 with $E_1 \cap E_2 = \emptyset$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

1

The sales manager of an e commerce company says that 80% of those who visit their website for the first time do not buy any mobile. If a new customer visits the website, what is the probability that the customer would buy mobile

2

	Blue	Black	Brown	Total
Software prog	35	25	20	80
Project Mgrs	7	8	5	20
Total	42	33	25	100

If an employee is selected at random , what is the probability that he is a software prog?

……..,what is the probability that he is wearing a blue trouser

3

A Survey conducted by a bank revealed that 40% of the accounts are savings accounts and 35% of the accounts are current accounts and the balance are loan accounts.

- What is the probability that an account taken at random is a loan account ?
- What is the probability that an account taken at random is NOT savings account ?
- What is the probability that an account taken at random is NOT a current account
- What is the probability that an account taken at random is a current account or a loan account?

Thanks