Data Scale

	Nominal	Ordinal	Interval	Ratio
Categorizes and labels variables	✓	✓	✓	✓
Ranks categories in order		✓	✓	✓
Has known, equal intervals			✓	✓
Has a true or meaningful zero				✓

Visualization

Importance of Data Visualization

1 Summary statistics can be dangerous

2 It is essential for exploratory data analysis

3 It make data easier for the people to understand
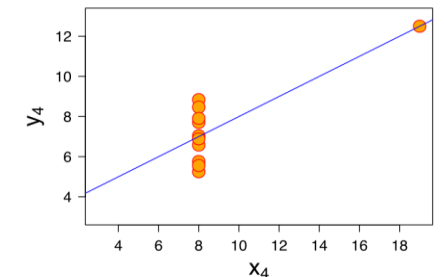
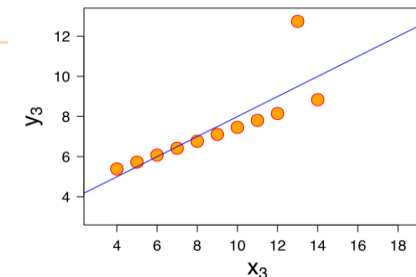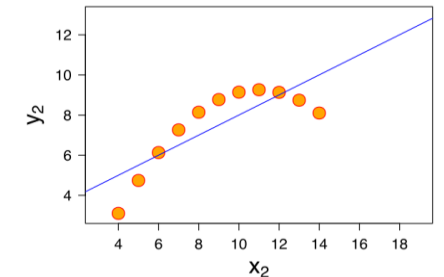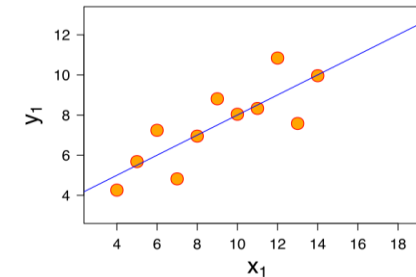# Importance of Data Visualization

1

Summary statistics can be dangerous

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8,04	10	9,14	10	7,46	8	6,58
8	6,95	8	8,14	8	6,77	8	5,76
13	7,58	13	8,74	13	12,74	8	7,71
9	8,81	9	8,77	9	7,11	8	8,84
11	8,33	11	9,26	11	7,81	8	8,47
14	9,96	14	8,1	14	8,84	8	7,04
6	7,24	6	6,13	6	6,08	8	5,25
4	4,26	4	3,1	4	5,39	19	12,5
12	10,84	12	9,13	12	8,15	8	5,56
7	4,82	7	7,26	7	6,42	8	7,91
5	5,68	5	4,74	5	5,73	8	6,89
SUM 99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG 9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV 3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

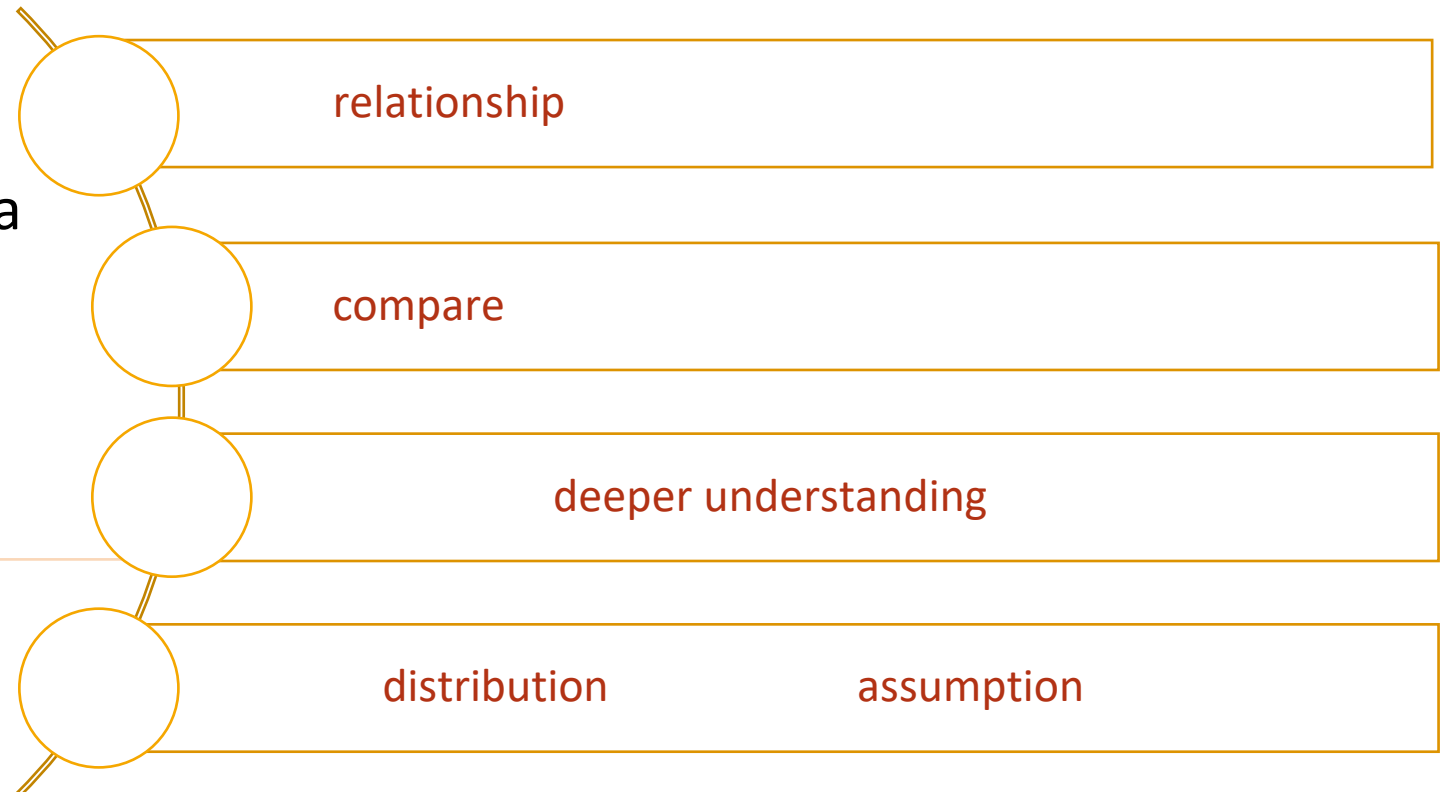**0.816** is the correlation coefficient for each dataset

Importance of Data Visualization

2

It is essential for exploratory data analysis

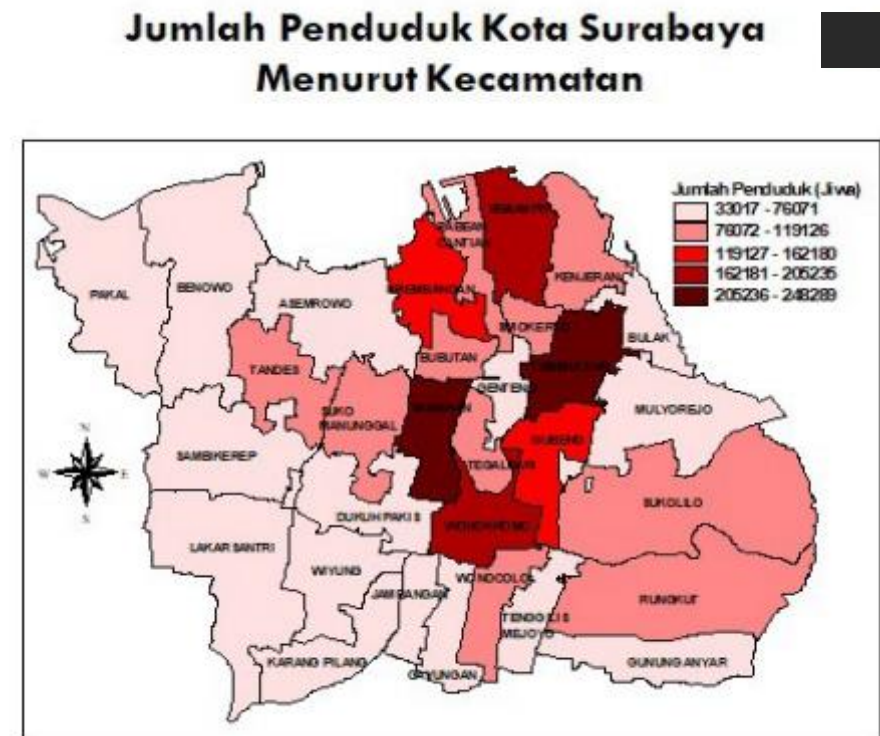EDA is a process by which a data scientist/data analyst seeks **to understand data** after gathering or preprocessing
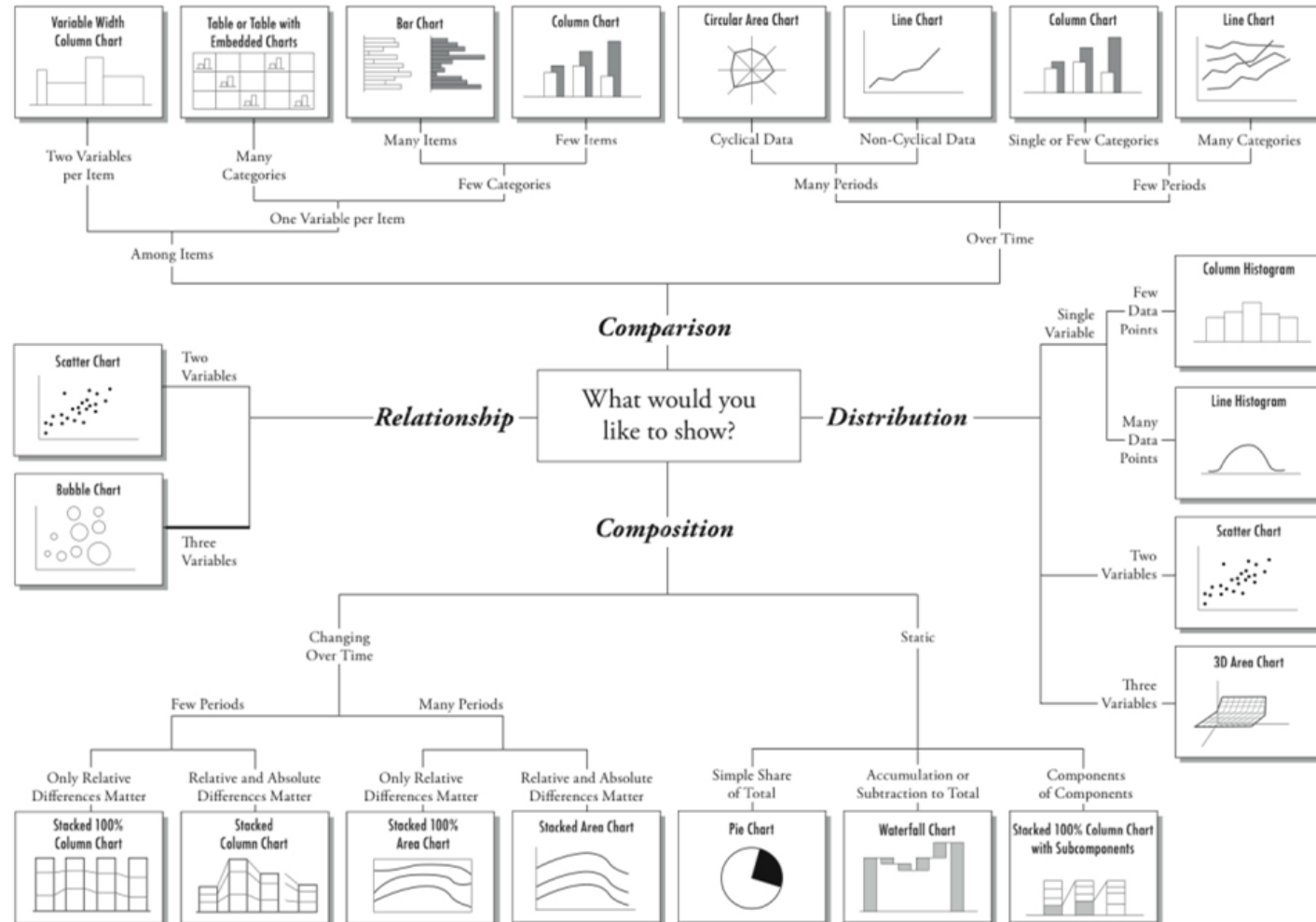
- relationship
- compare
- deeper understanding
- distribution   assumption

Importance of Data Visualization

3

It make data easier for the people to understand

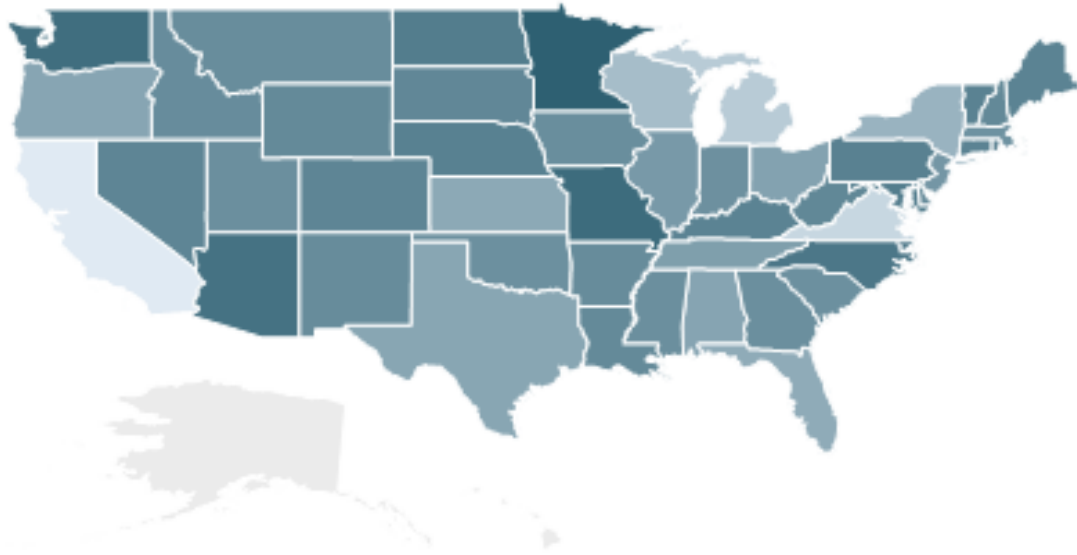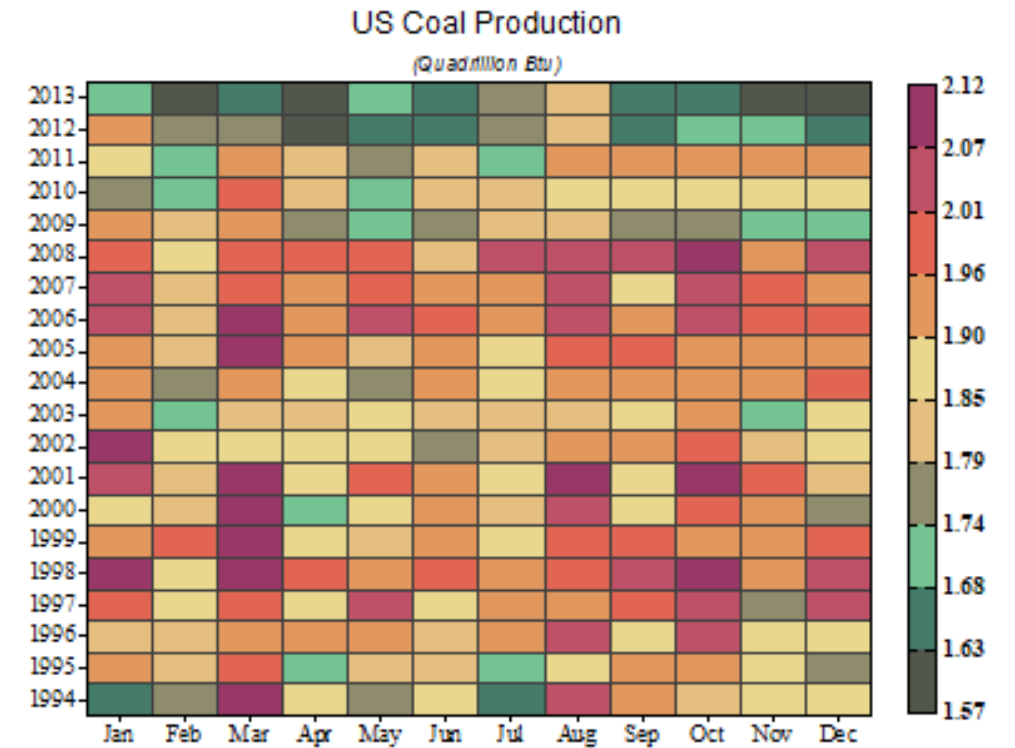| Kecamatan	Penduduk	KK	Rata-rata Anggota Keluarga
010. Karangpilang	77,557	24,264	3.20
020. Jambangan	54,105	16,813	3.22
030. Gayungan	47,827	15,156	3.16
040. Wonocolo	85,284	26,139	3.26
050. Tenggilis Mejoyo	60,274	19,070	3.16
060. Gunung Anyar	60,505	18,617	3.25
070. Rungkut	121,247	37,319	3.25
080. Sukolilo	116,915	36,091	3.24
090. Mulyorejo	91,339	28,897	3.16
100. Gubeng	143,874	47,197	3.05
110. Wonokromo	169,994	53,749	3.16
120. Dukuh Pakis	62,846	19,821	3.17
130. Wiyung	74,024	22,911	3.23
140. Lakarsantri	61,907	19,058	3.25
141. Sambikerep	66,833	20,638	3.24
150. Tandes	96,590	29,816	3.24

Jumlah Penduduk Kota Surabaya Menurut Kecamatan

Cheat Sheet

www.ExtremePresentation.com
© 2009 A. Abela — a.v.abela@gmail.com

Other Graph

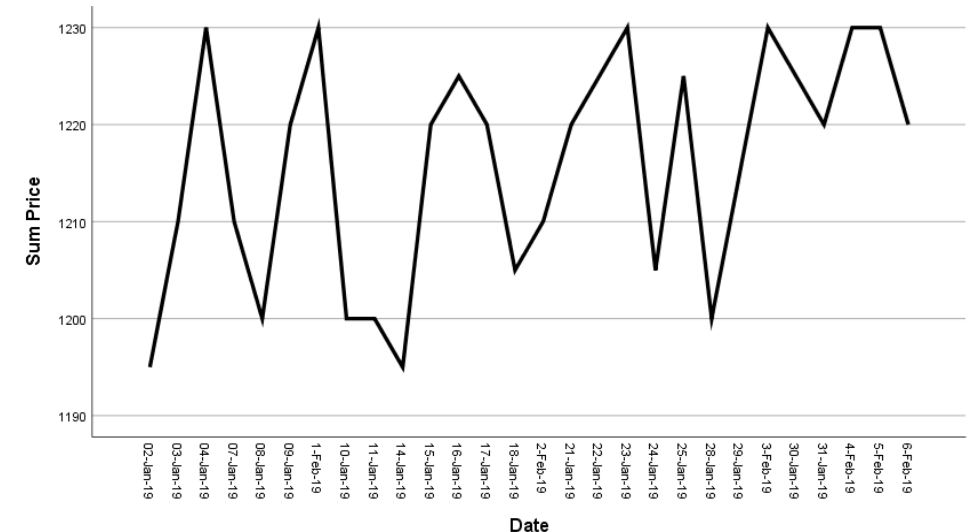- Map Chart

- Heat Map

US Coal Production
(Quadrillion Btu)

Type of Chart

Line Chart

- **To show trends.** For example, how house prices have increased over time.

- **to make predictions** based on a data history over time.

- **To compare** two or more different variables, situations, and information over a given period of time.
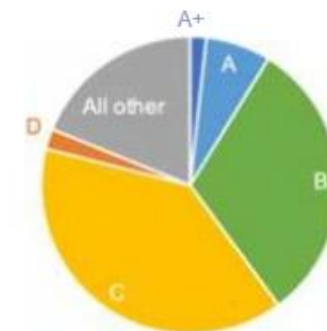
Type of Chart

Pie Chart

- To display data that are grouped into **nominal or ordinal categories**

- To **show percentage** or proportional data.

- **To compare data** among different categories.

- Pie charts work best for displaying data for **3 to 7** categories

Pie Chart of Sex

Category
F
M

187; 47,3%   208; 52,7%
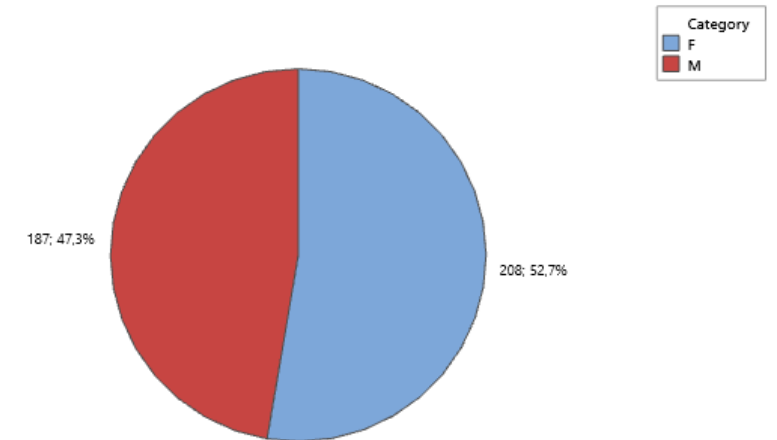
New client tier share

% of Total Accounts   % of Total Revenue

Type of Chart

Box Plot

- To **shows** the **median** of the data, the **upper and lower quartiles**, and any data points that possibly are **outside value**
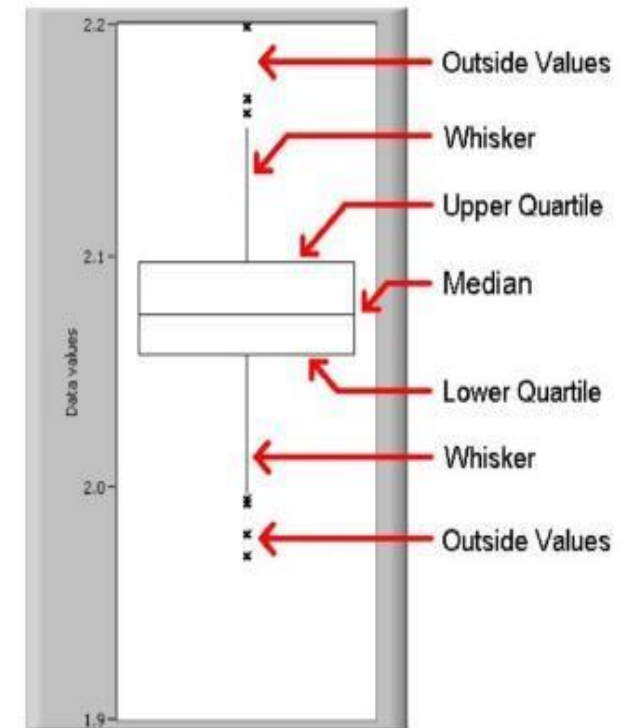
Type of Chart

Histogram

- For **continuous** data.

- To **represent the shape** of the data's **distribution**.

- To **summarize large data** sets graphically.
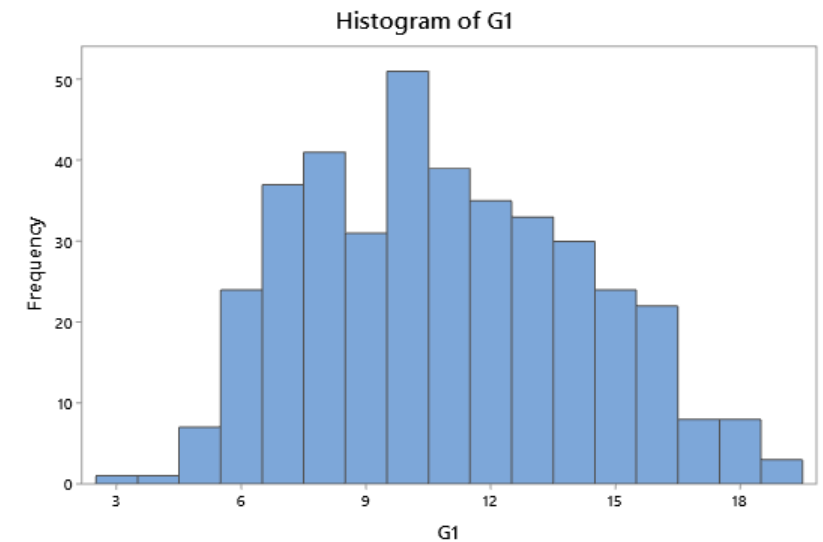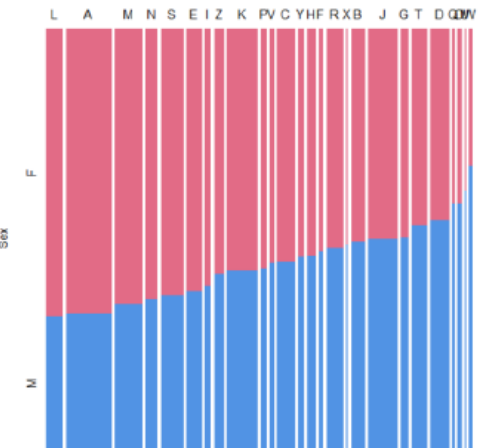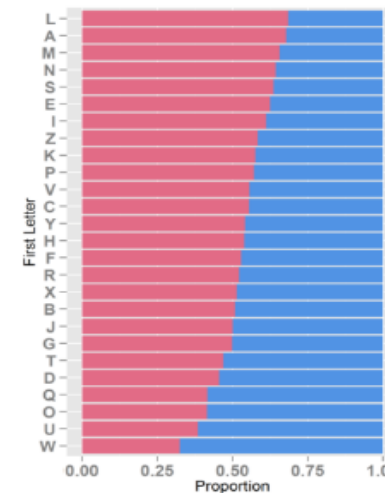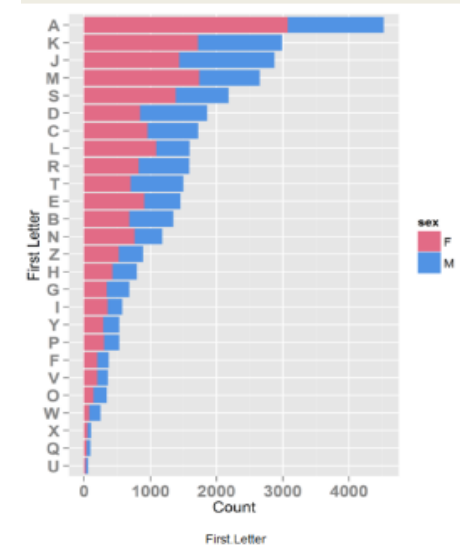

Histogram of G1

Type of Chart

Bar Chart

- For **categorical** data and **continuous** data.

- To **present the frequency** of the categorical data
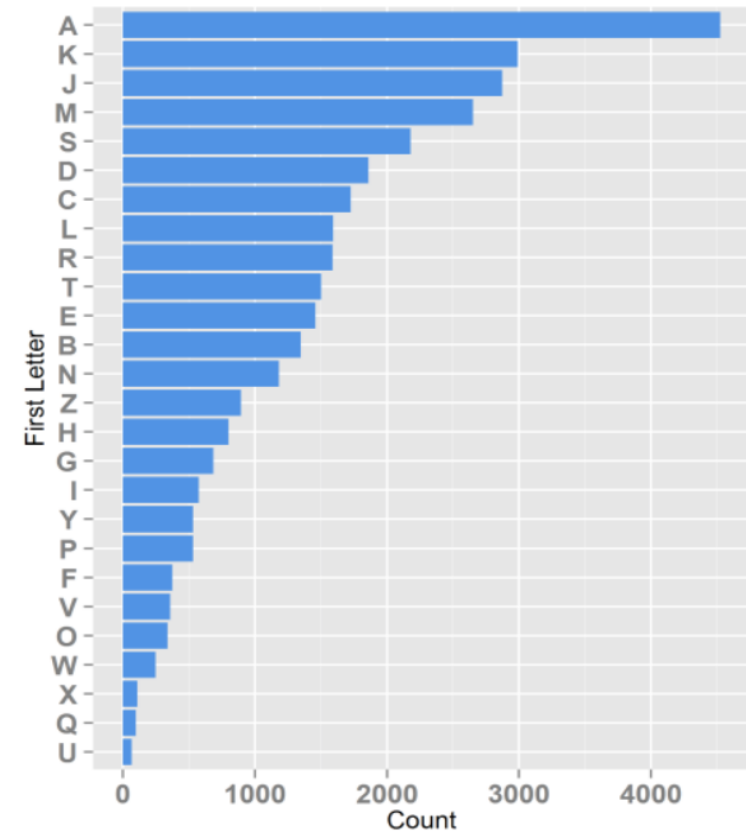
Type of Chart

Bar Chart

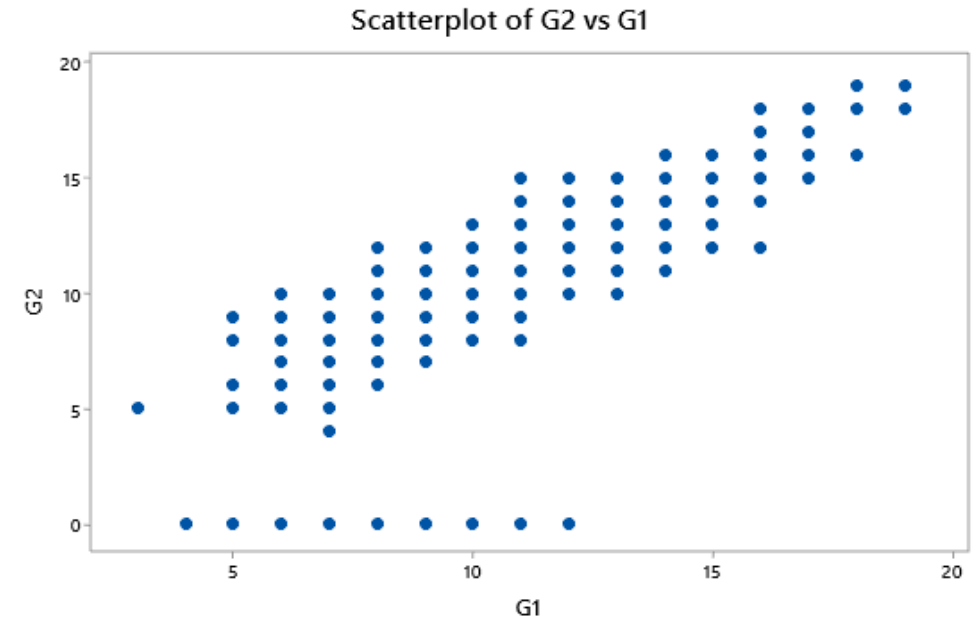- For **categorical** data and **continuous** data.

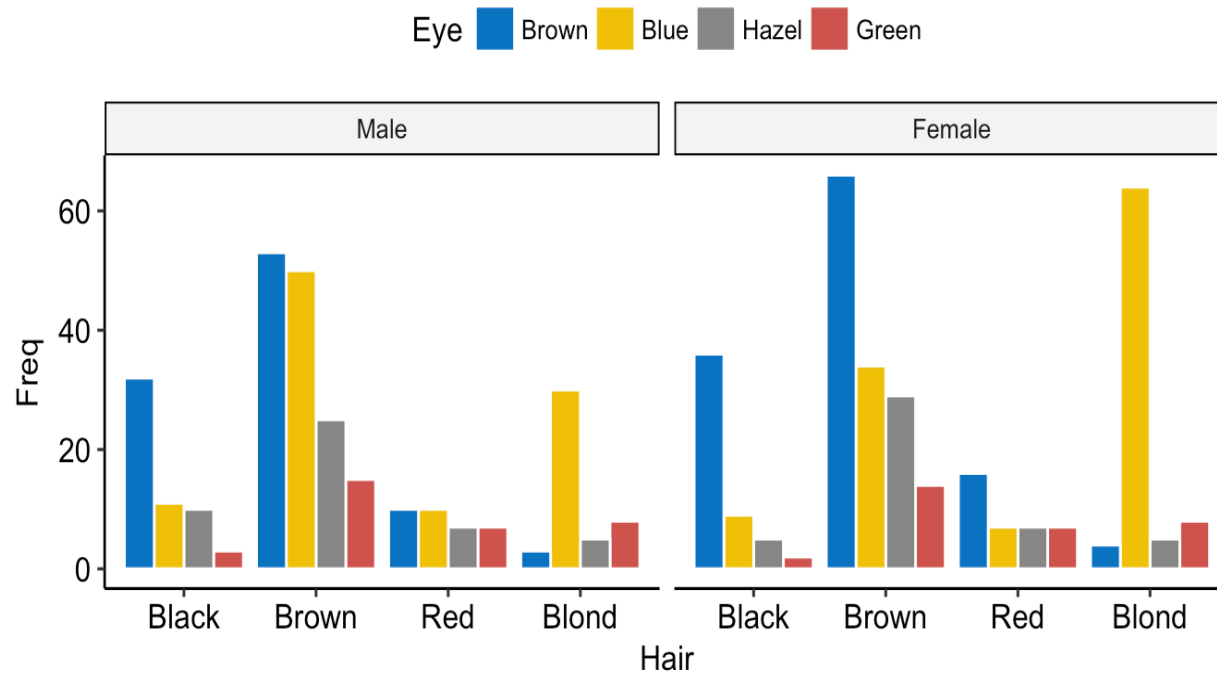- To **present the frequency** of the categorical data
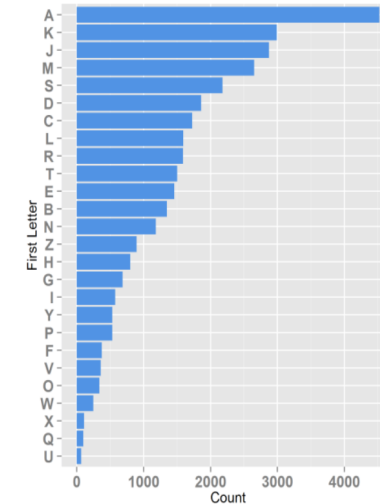
Type of Chart

Scatterplot

- To **find out a relationship** between 2 variables.

- To **predict** the behavior of dependent variable based on the measure of the independent variable

Scatterplot of G2 vs G1

Type of Chart

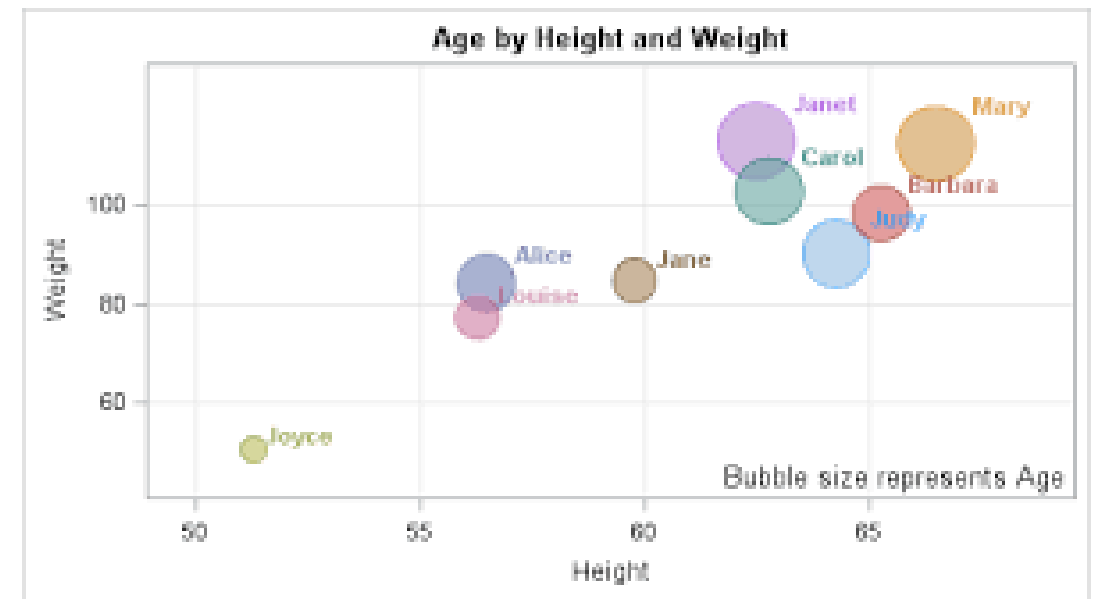## Some Examples of Multivariate Chart
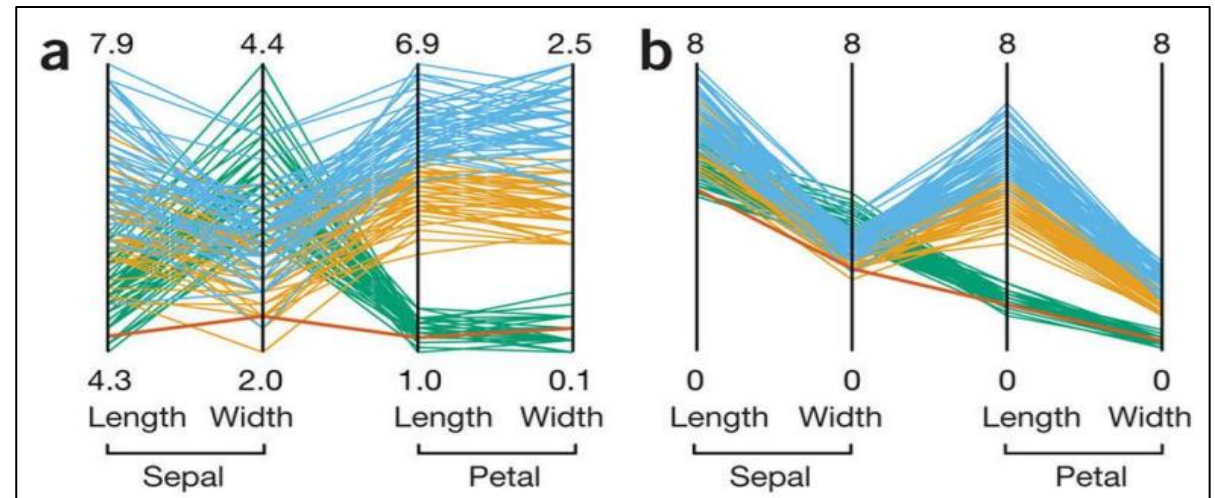
Type of Chart

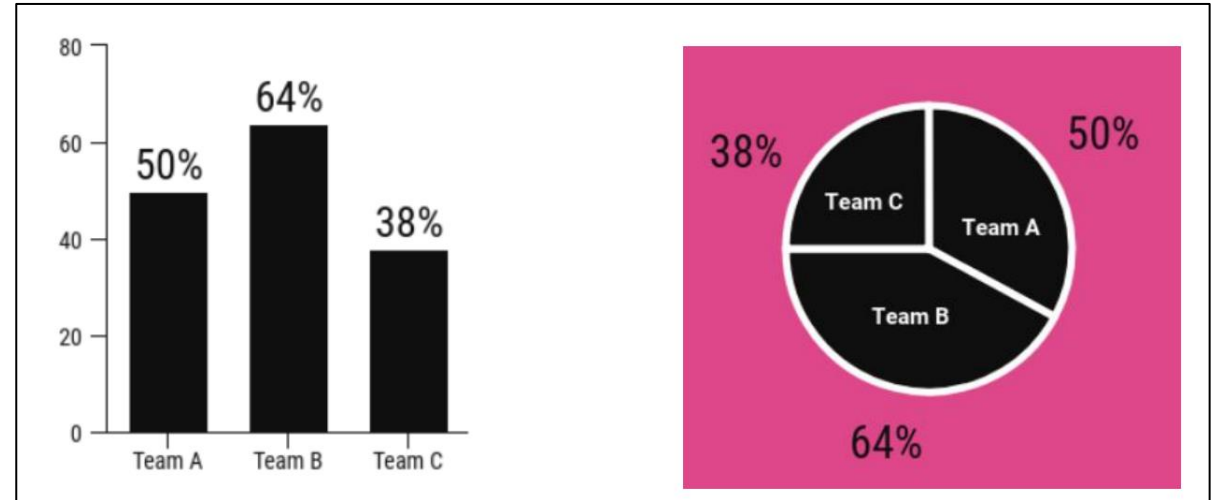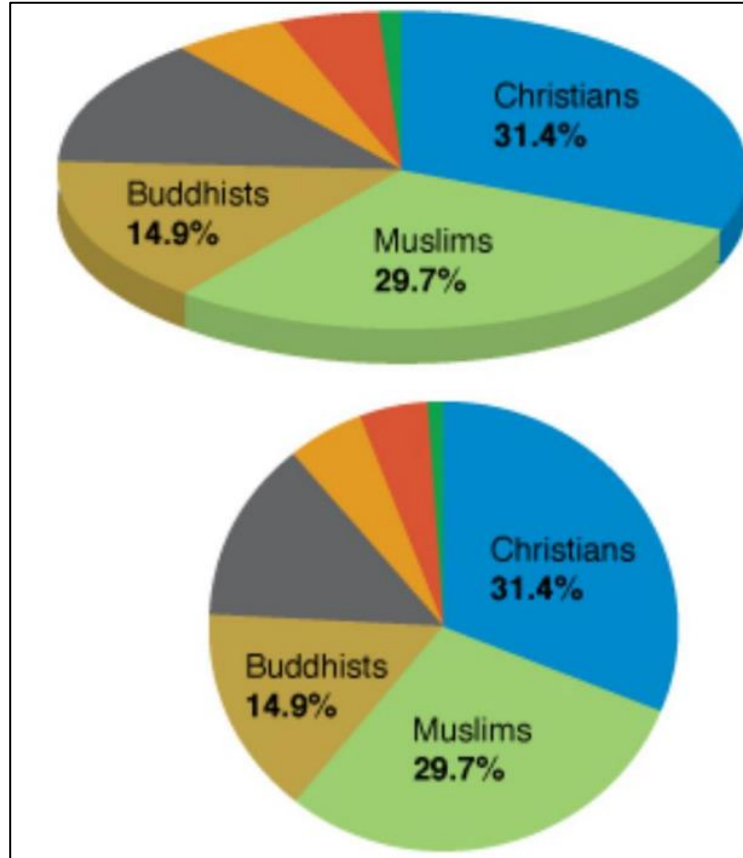## Some Examples of Multivariate Chart

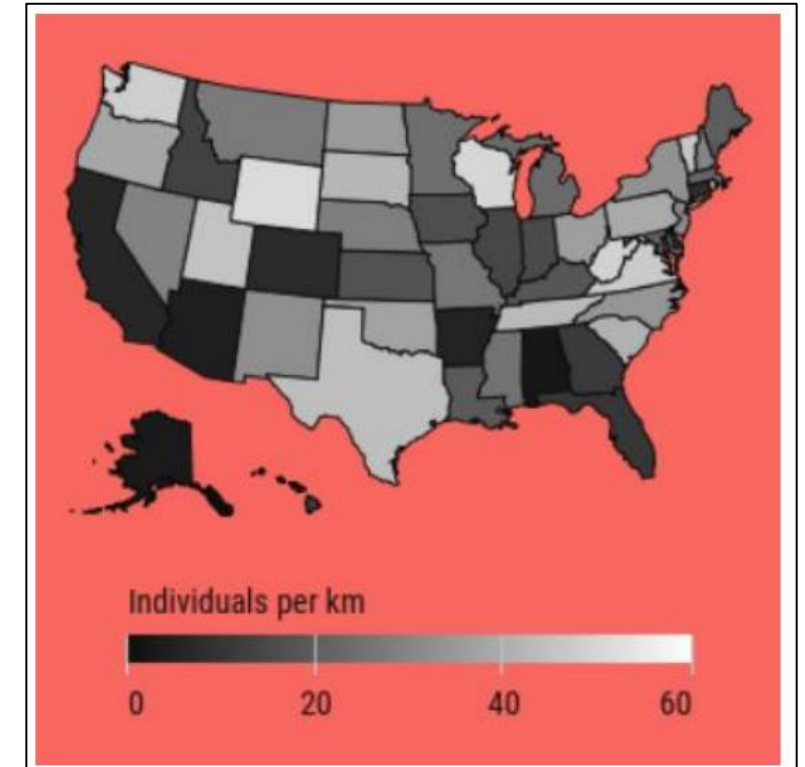Heat map of school grades within a fictional class
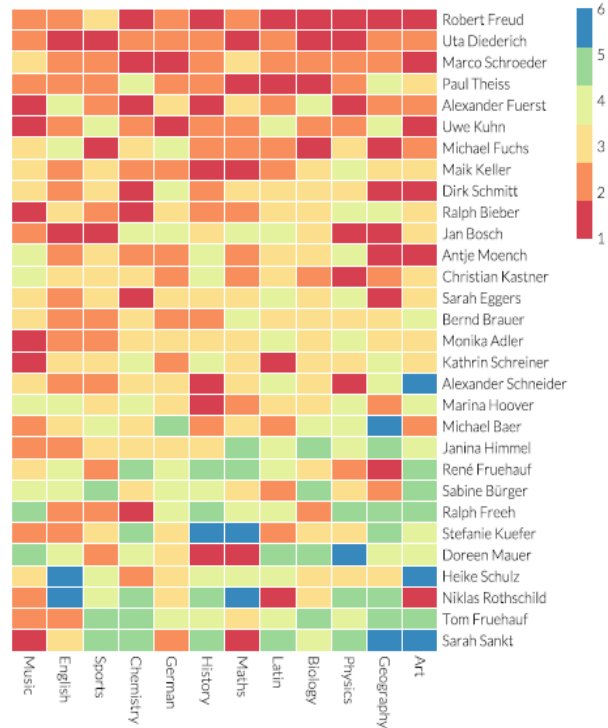
Fictional data, names generated with de.fakenamegenerator.com

Age by Height and Weight

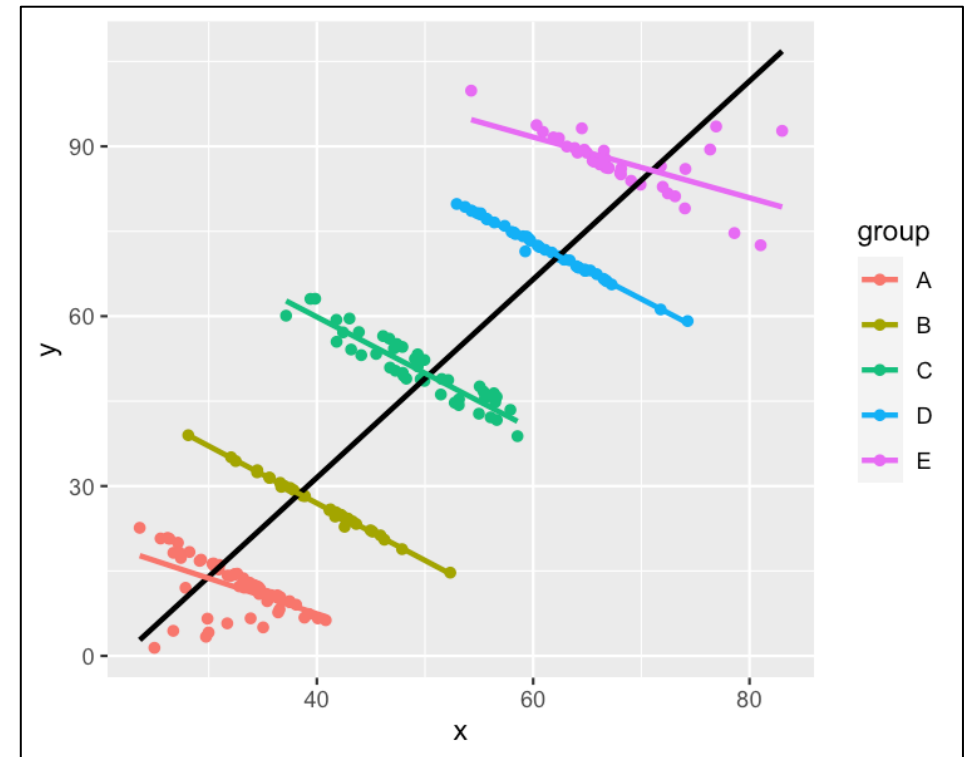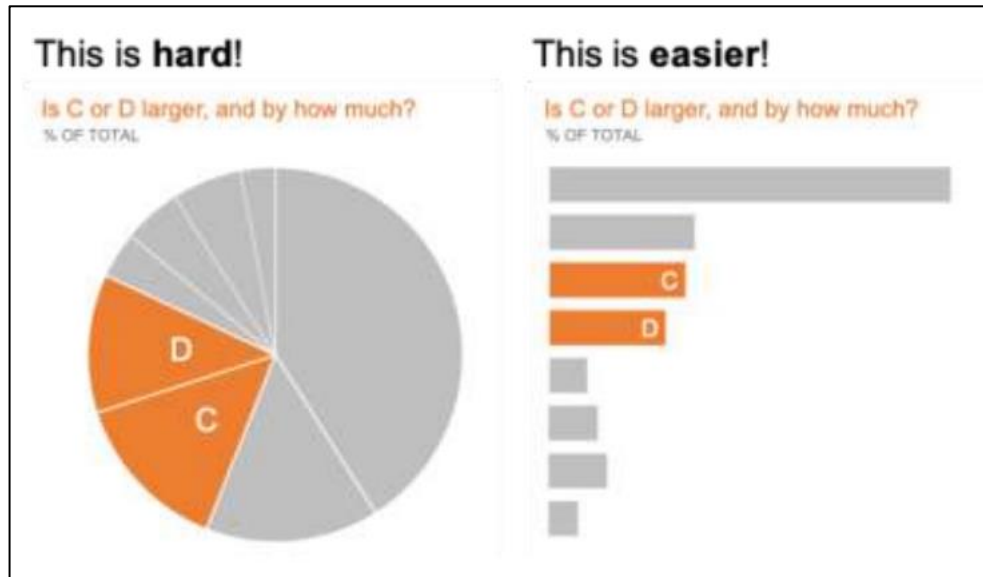Bubble size represents Age

Misleading Graph

Data Analysis 2023

Misleading Graph
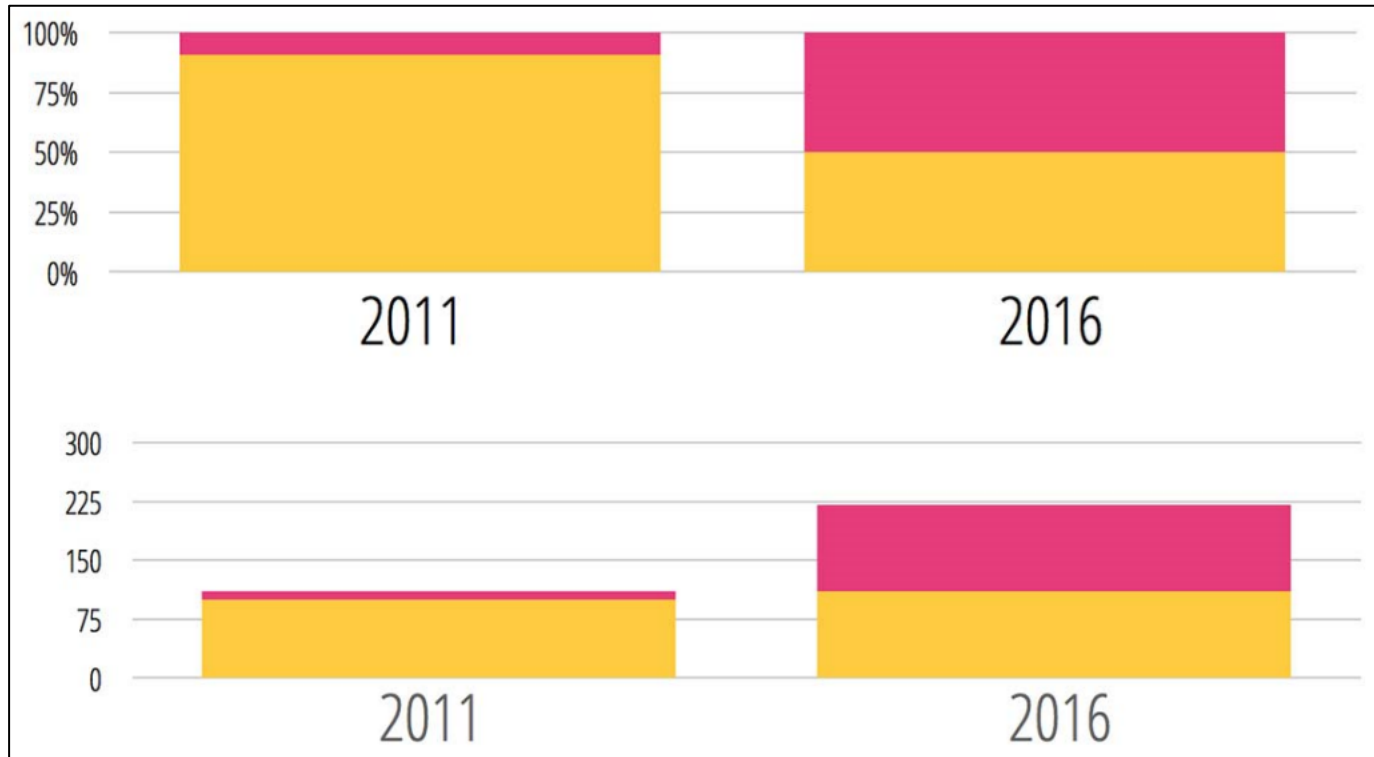
Misleading Graph

More tips on building a better chart: https://medium.com/responsibleml/data-visualization-cheat-sheets-1c12ba8a7671

Code Cheat Sheet

https://raw.githubusercontent.com/rstudio/cheatsheets/main/data-visualization.pdf

https://matplotlib.org/cheatsheets/