The Art and Science of Storytelling and Visualization

Introduction To Charts,Tableau

Pavan Kumar KV & Alok Tiwari

21/May/2022

Intro To Data Visualization

➢ **Data Visualization is used to communicate information clearly and efficiently**

➢ **It makes complex data more accessible, understandable, and usable.**

➢ **Tables are used where users need to see the pattern of a specific parameter, while charts are used to show patterns or relationships in the data for one or more parameters.**

Pro and Cons of Data Visualization

Pros :

    1. It can be accessed quickly by a wider audience.
    2. It conveys a lot of information in a small space.
    3. It makes your report more visually appealing.

Cons :

    1. It can misrepresent information – if an incorrect visual representation is made.
    2. It can be distracting – if the visual data is distorted or excessively used.

Charts for Data Scientists

Chart types based on data (Simpler view)

Categorical vs Numerical	Numerical Distributio	Numerical vs Numerical
• Bar • Line • Area • Pie • Geographical Map	• Box • Histogram	• Scatter

❖ Note
  ▪ A very high level distribution of most commonly used charts
  ▪ You may need to explore several beyond this

Class Exercise: Draw appropriate visualization

PC Name	Margin
Karimnagar	205077
Secundrabad	254735
Peddapalli	291158
Warangal	392574
Medak	397029

2014 General Elections data from Telangana Constituency
Source: India Votes

Bar/Column Charts
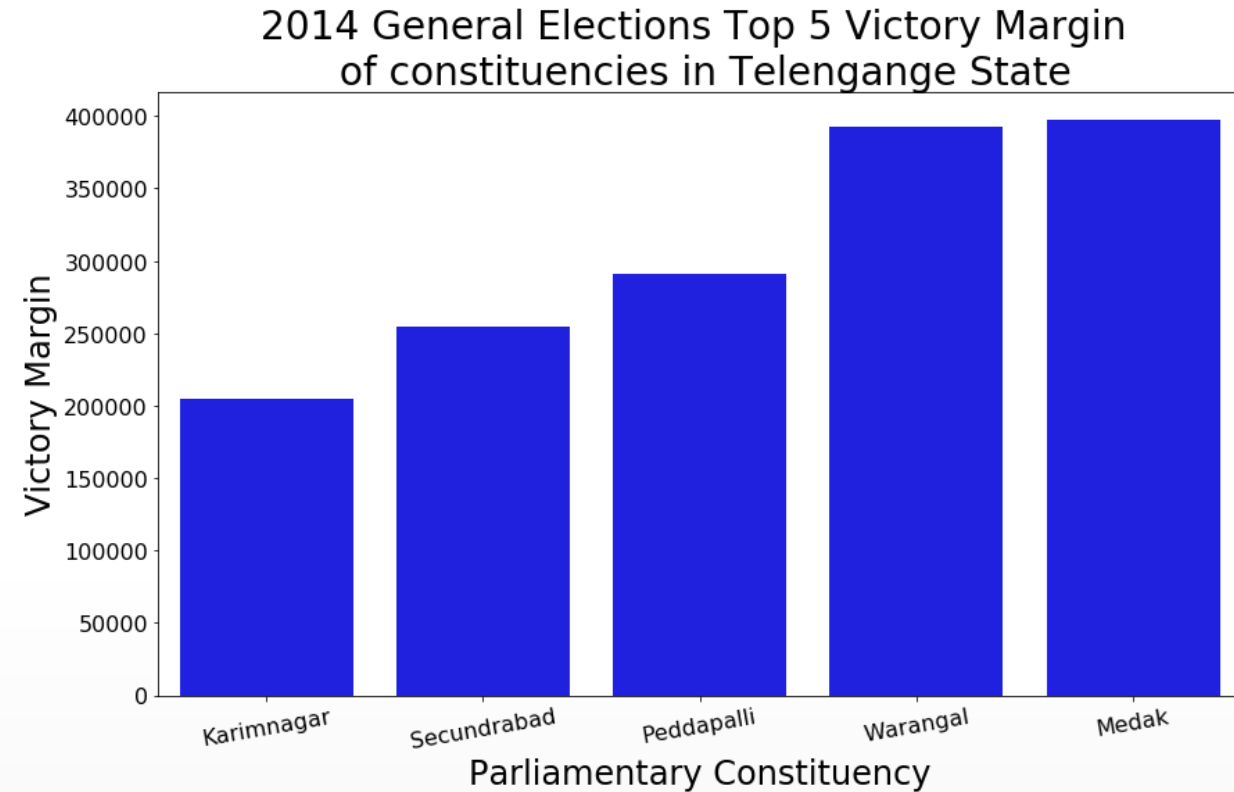
What are Bar Plots?

Categorical vs Numerical

Independent variable is categorical

Dependent variable is

numerical Rectangular bars

Heights proportional to values they present
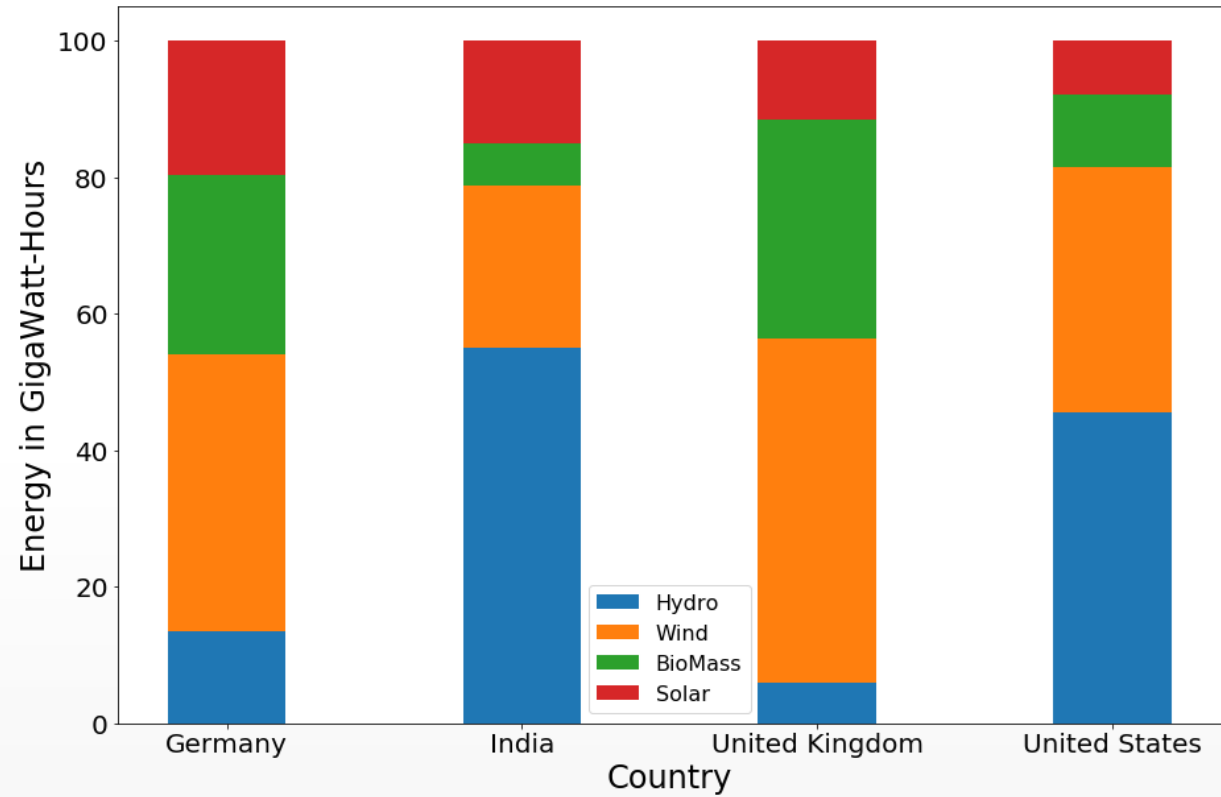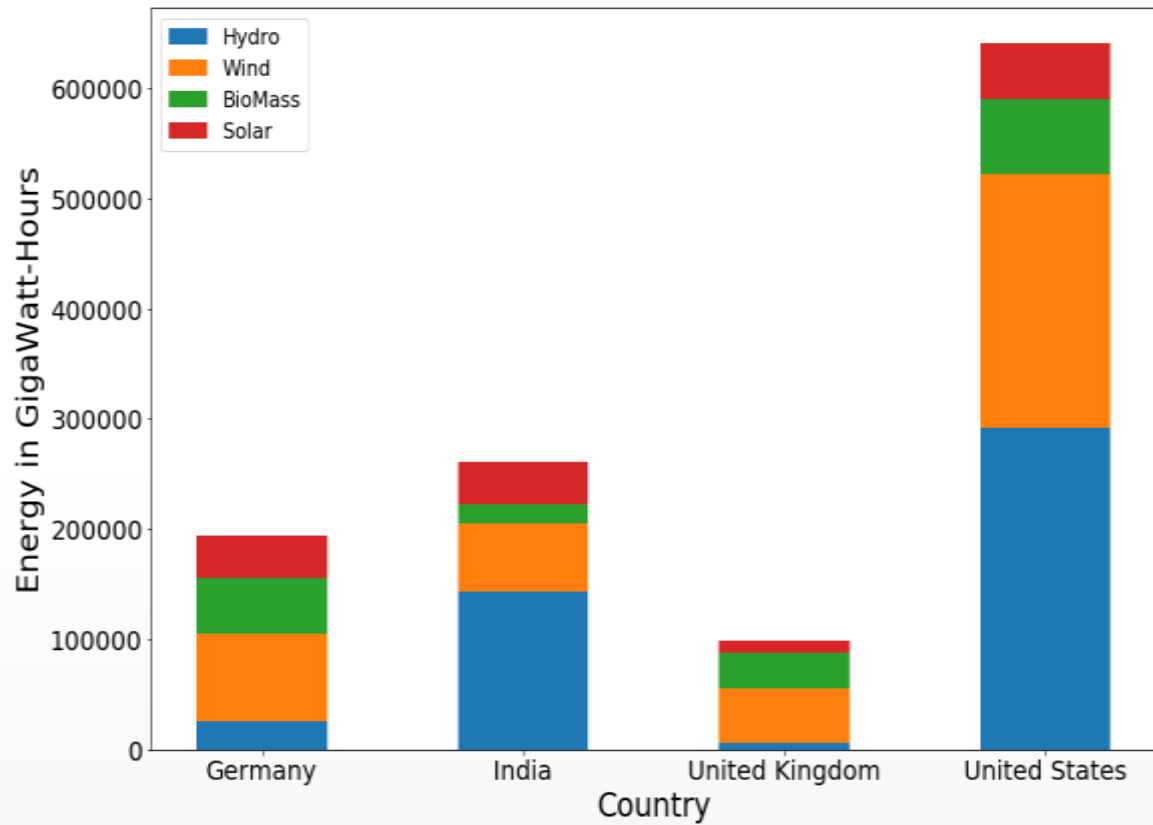
Width has no meaning (Usually)

2014 General Elections Top 5 Victory Margin of constituencies in Telengange State

Grouped bar chart

	Country	Source	Watts
0	Germany	Hydropower\n(GWh)	26135
1	Germany	Wind power\n(GWh)	78598
2	Germany	Biomass\nand waste\n(GWh)	50924
3	Germany	Solar power\n(GWh)	38098
4	India	Hydropower\n(GWh)	143743
5	India	Wind power\n(GWh)	62036
6	India	Biomass\nand waste\n(GWh)	16325
7	India	Solar power\n(GWh)	39268
8	United Kingdom	Hydropower\n(GWh)	5928
9	United Kingdom	Wind power\n(GWh)	50004
10	United Kingdom	Biomass\nand waste\n(GWh)	31869
11	United Kingdom	Solar power\n(GWh)	11525
12	United States	Hydropower\n(GWh)	292113
13	United States	Wind power\n(GWh)	229471
14	United States	Biomass\nand waste\n(GWh)	69017
15	United States	Solar power\n(GWh)	50334

Energy Distribution by Country
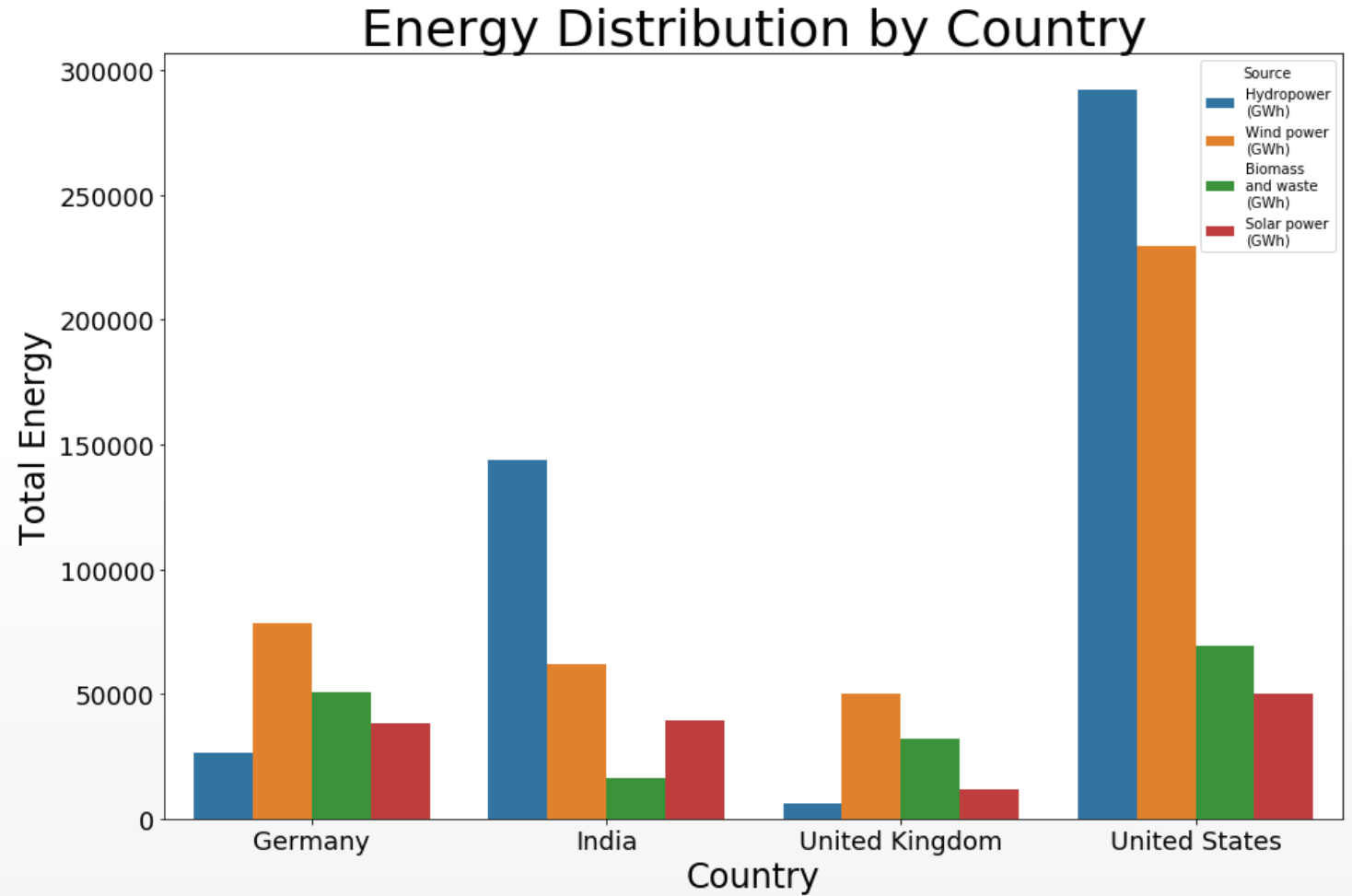
Stacked bar chart

A good stacked bar chart

Revenue by User Geography
In Millions
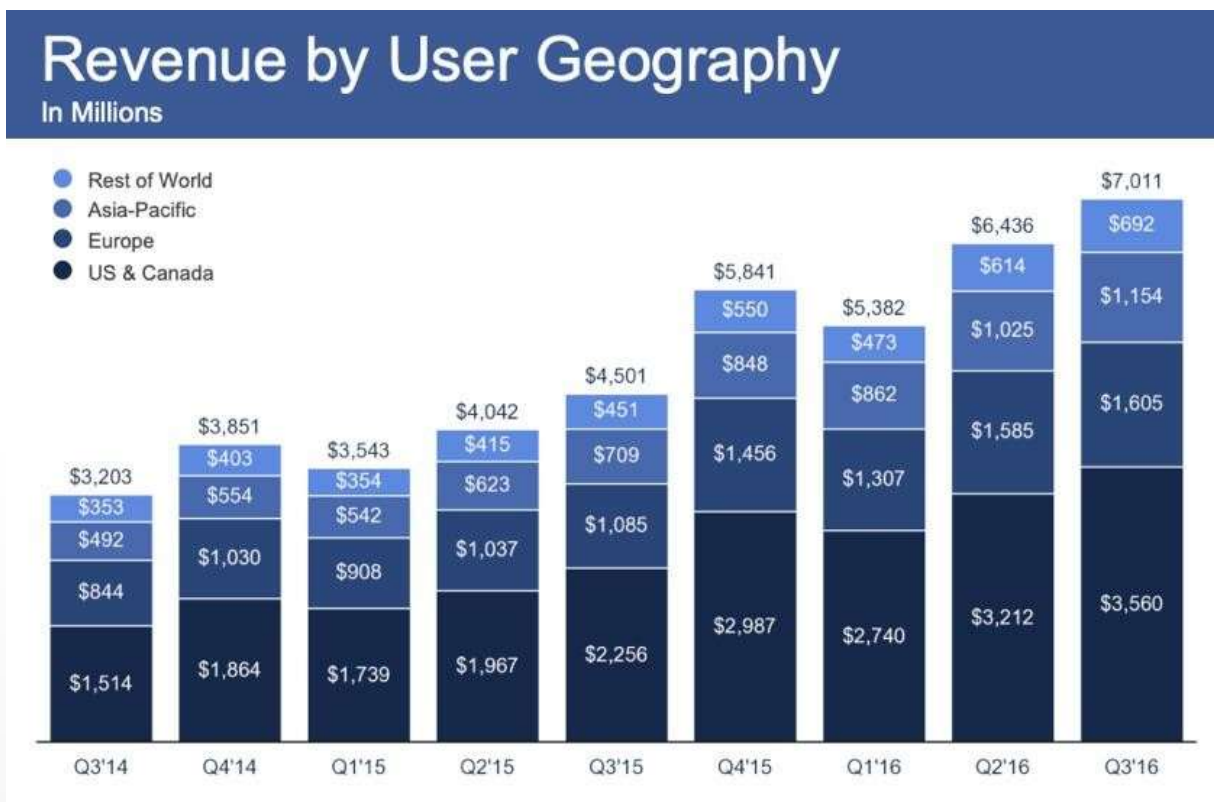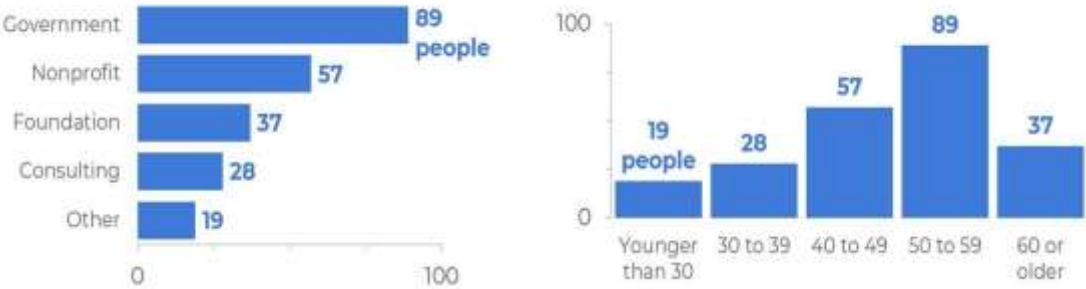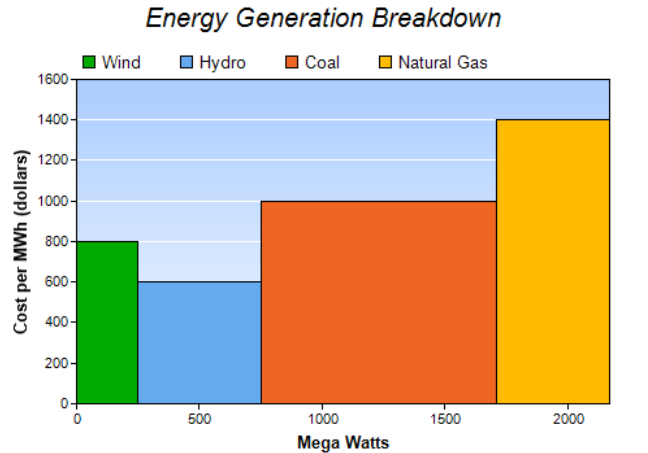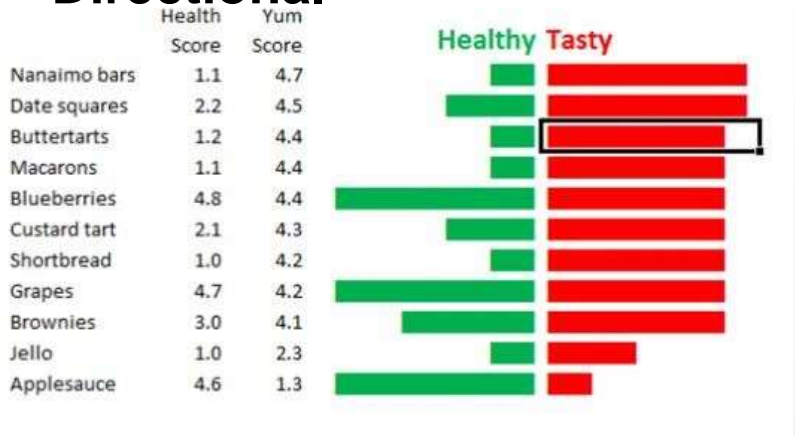
- Lot of data but still presented in compact design
- Numbers can be directly read from each stack and also visualized for relative change
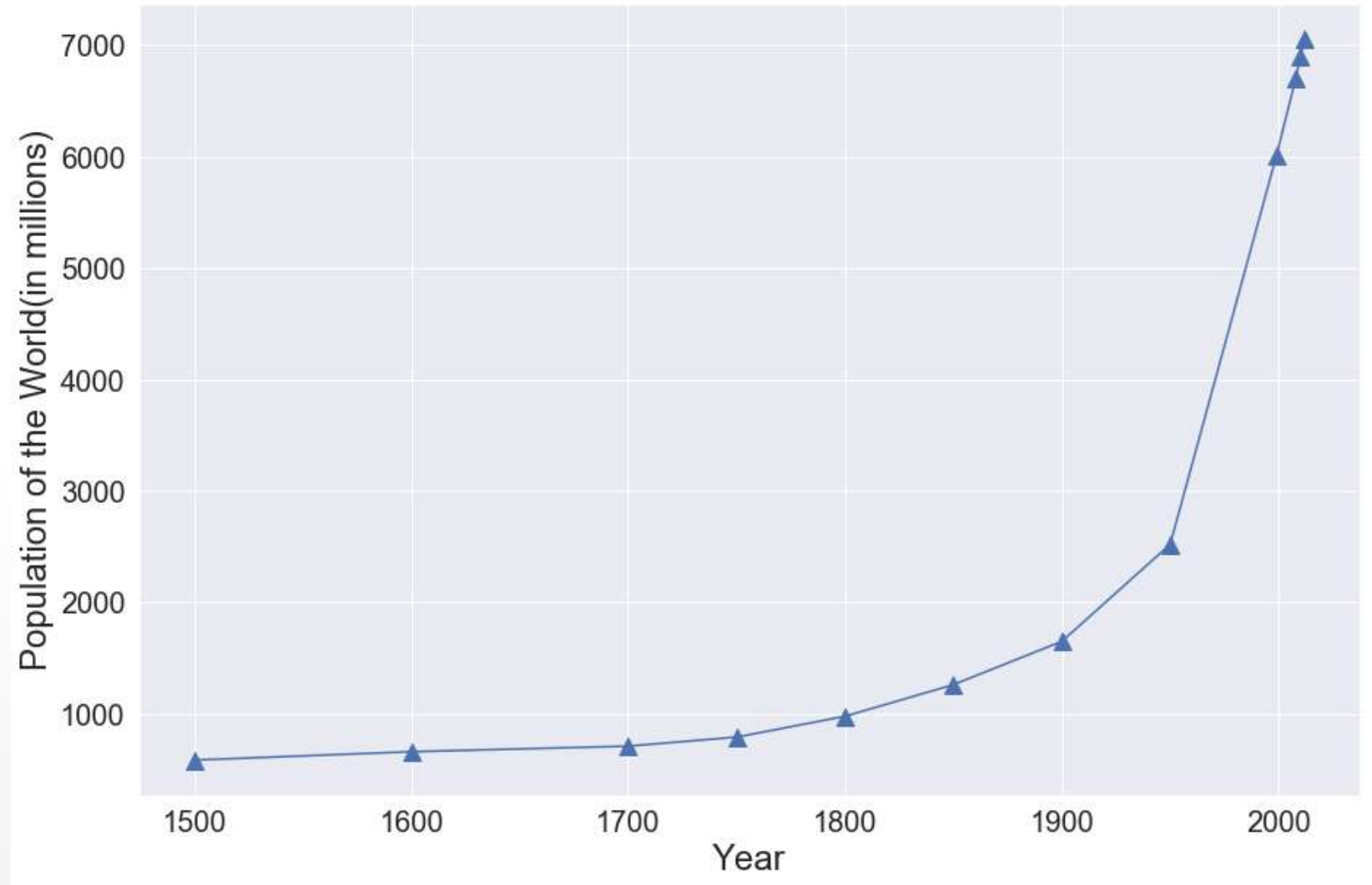
Bar Chart Types

INSOFE

Horizontal/Vertical

Grouped

Stacked
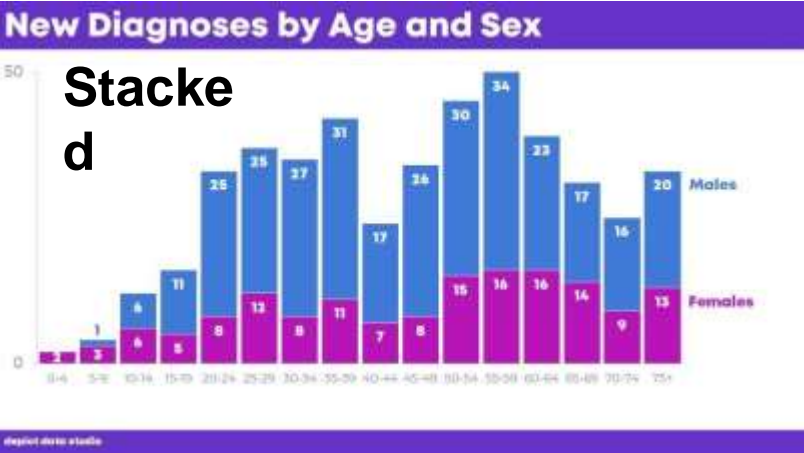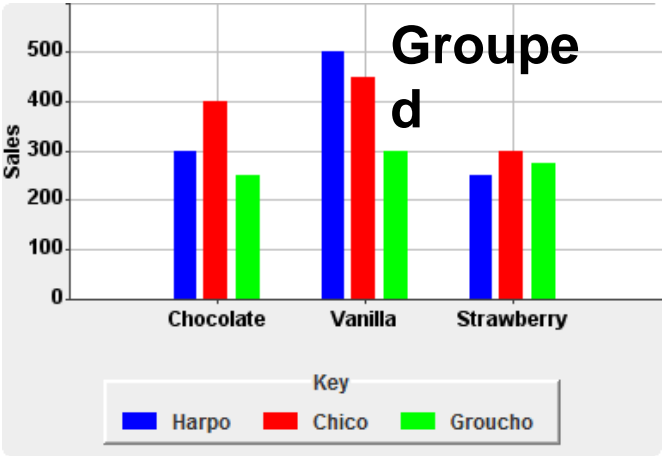
Bi-Directional

Line Chart

Class Exercise

	Year	World
0	1500	585
1	1600	660
2	1700	710
3	1750	791
4	1800	978
5	1850	1262
6	1900	1650
7	1950	2521
8	1999	6008
9	2008	6707
10	2010	6896
11	2012	7052

Treat this is as categorical

Class Exercise

|	Year	World	Africa	Asia	Europe	Latin America[Note 1]	North America[Note 1]	Oceania
0	1500	585	86	282	168	40	6	3
1	1600	660	114	350	170	20	3	3
2	1700	710	106	411	178	10	2	3
3	1750	791	106	502	190	16	2	2
4	1800	978	107	635	203	24	7	2
5	1850	1262	111	809	276	38	26	2
6	1900	1650	133	947	408	74	82	6
7	1950	2521	221	1402	547	167	172	13
8	1999	6008	783	3700	675	508	312	30
9	2008	6707	973	4054	732	577	337	34
10	2010	6896	1022	4164	738	590	345	37
11	2012	7052	1052	4250	740	603	351	38

Line Chart
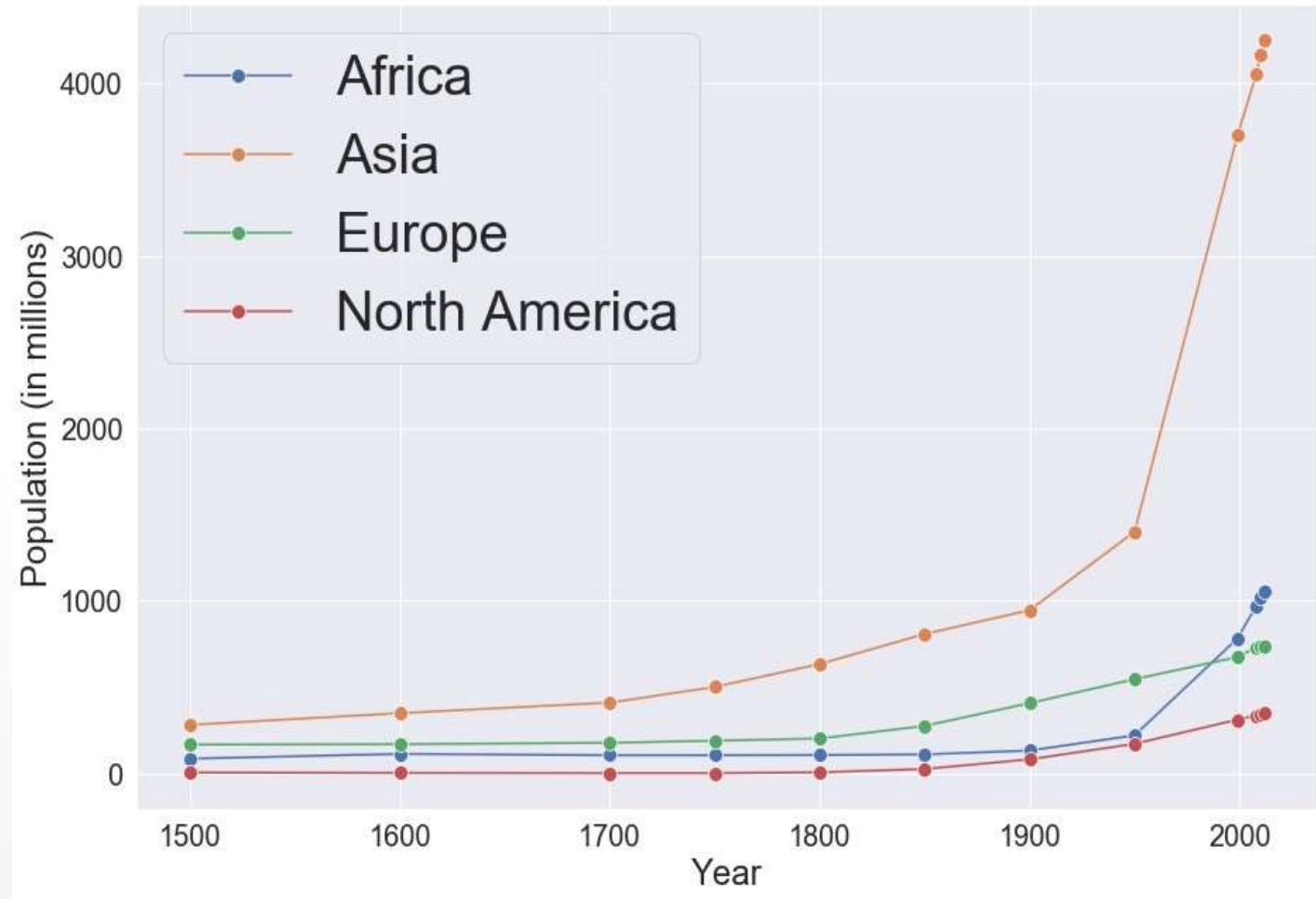
Q: Can a stacked bar chart replace this?

Same data as bar chart

Same data with engineered feature

Different data (on different axis)
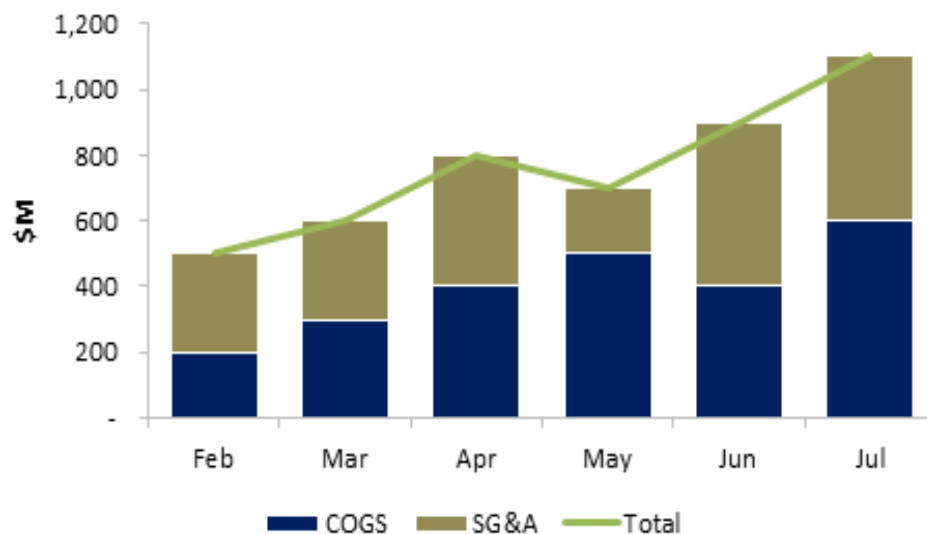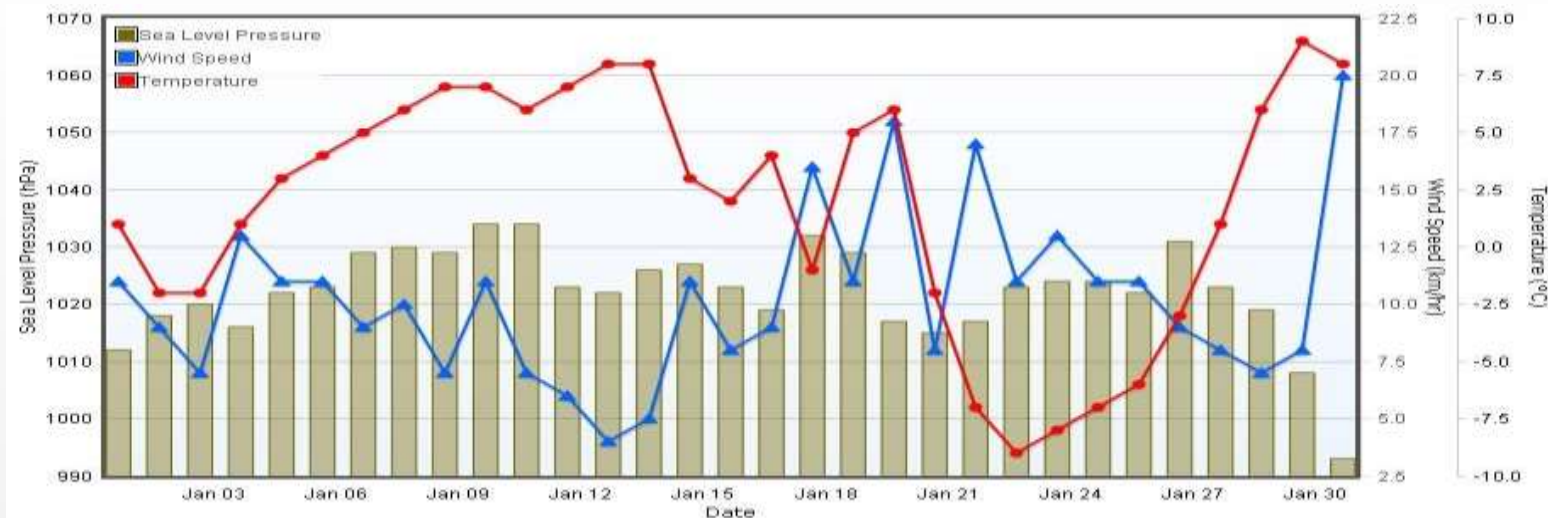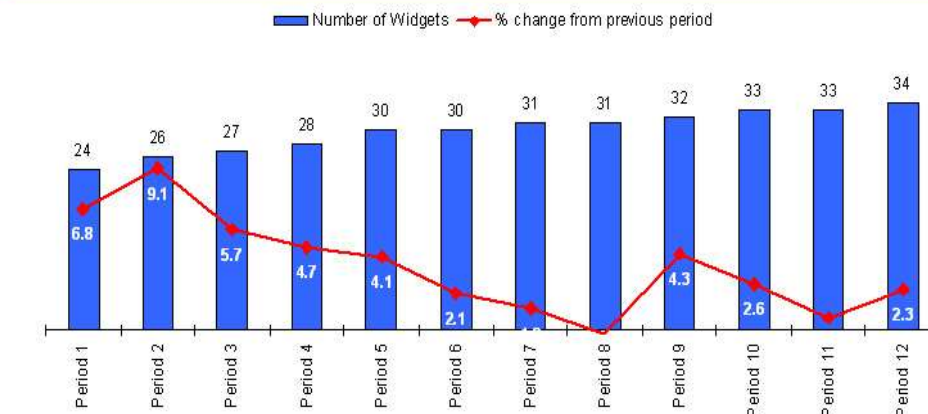
Area Chart

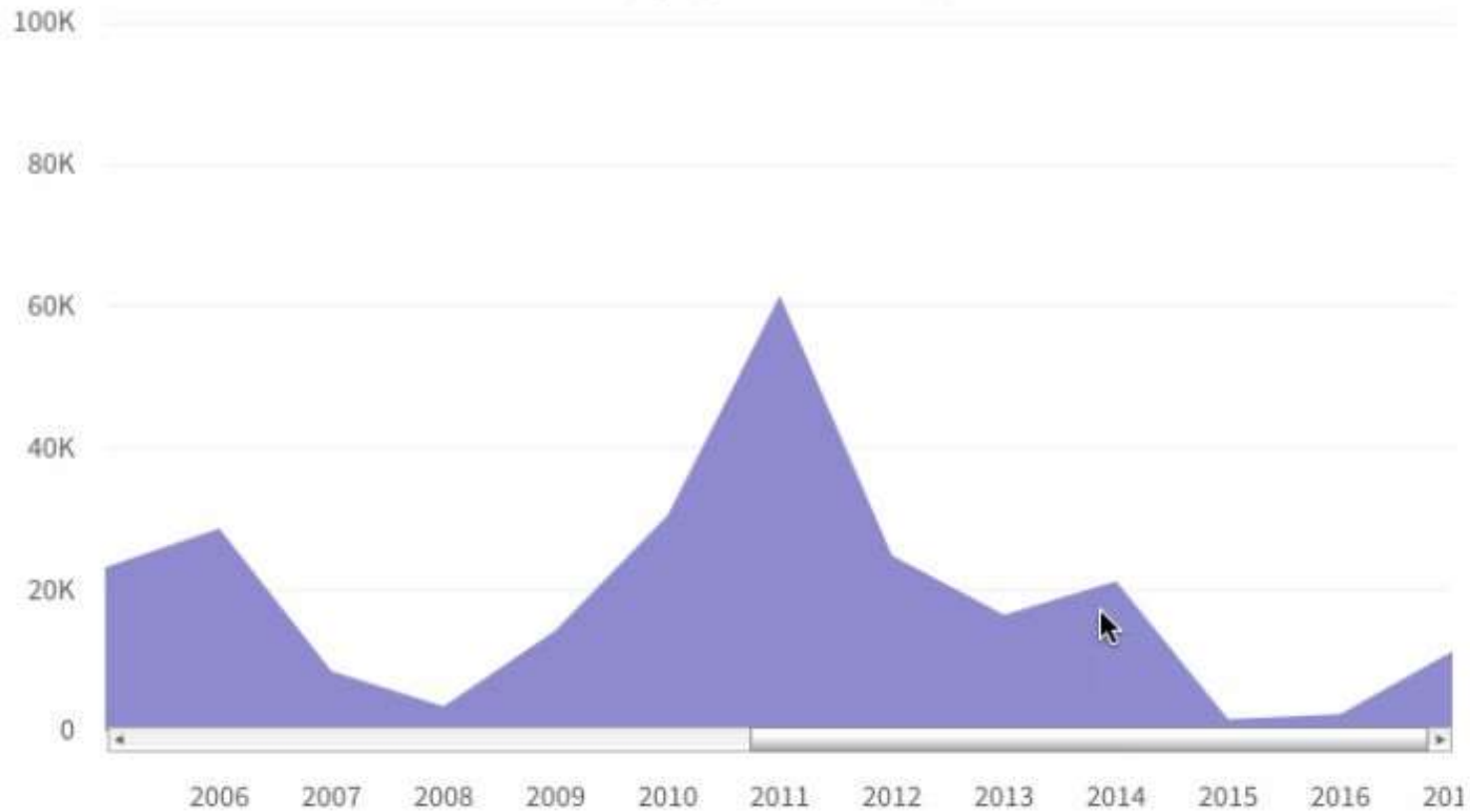# Area Plot

Line chart shaded with area between the line and X-axis shaded

Histogram
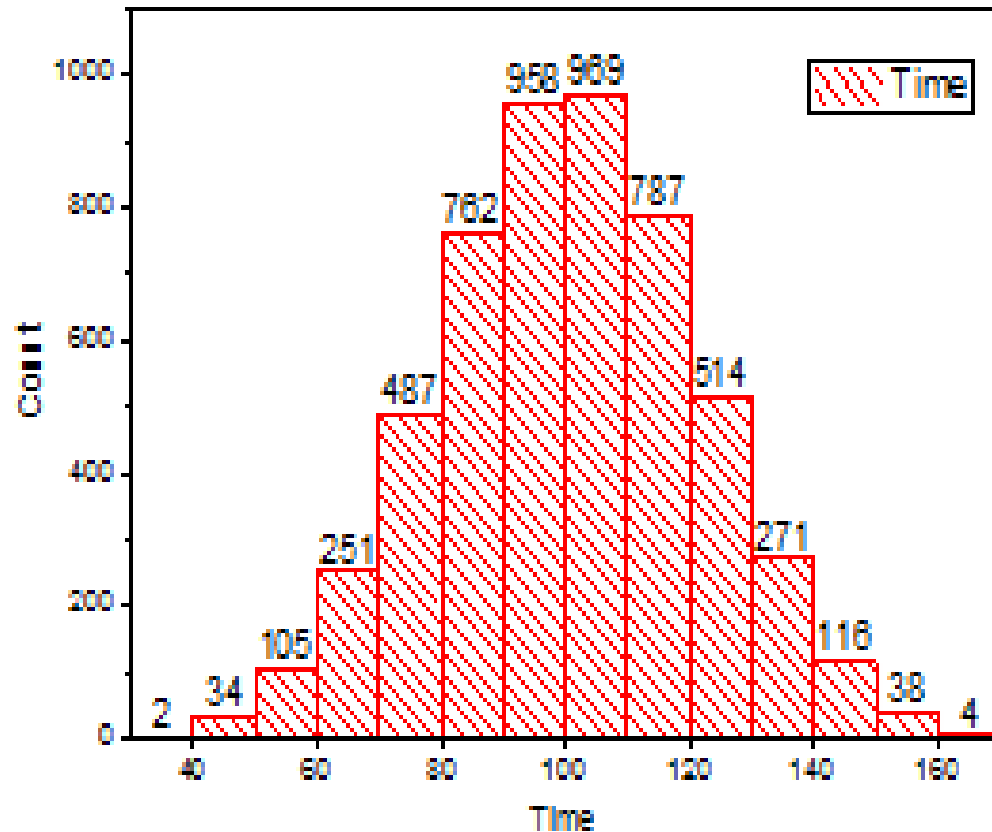
Single dimensional frequency chart. Y axis
(height) denoting counts of X axis
variable

Primary use

1.Centre
2.Spread
3.Skewness
4.Outliers
5.Modes

No spacing between the bins (unless to
denote zero frequency)

Skewed Histogram

Age at Admission to Hospital
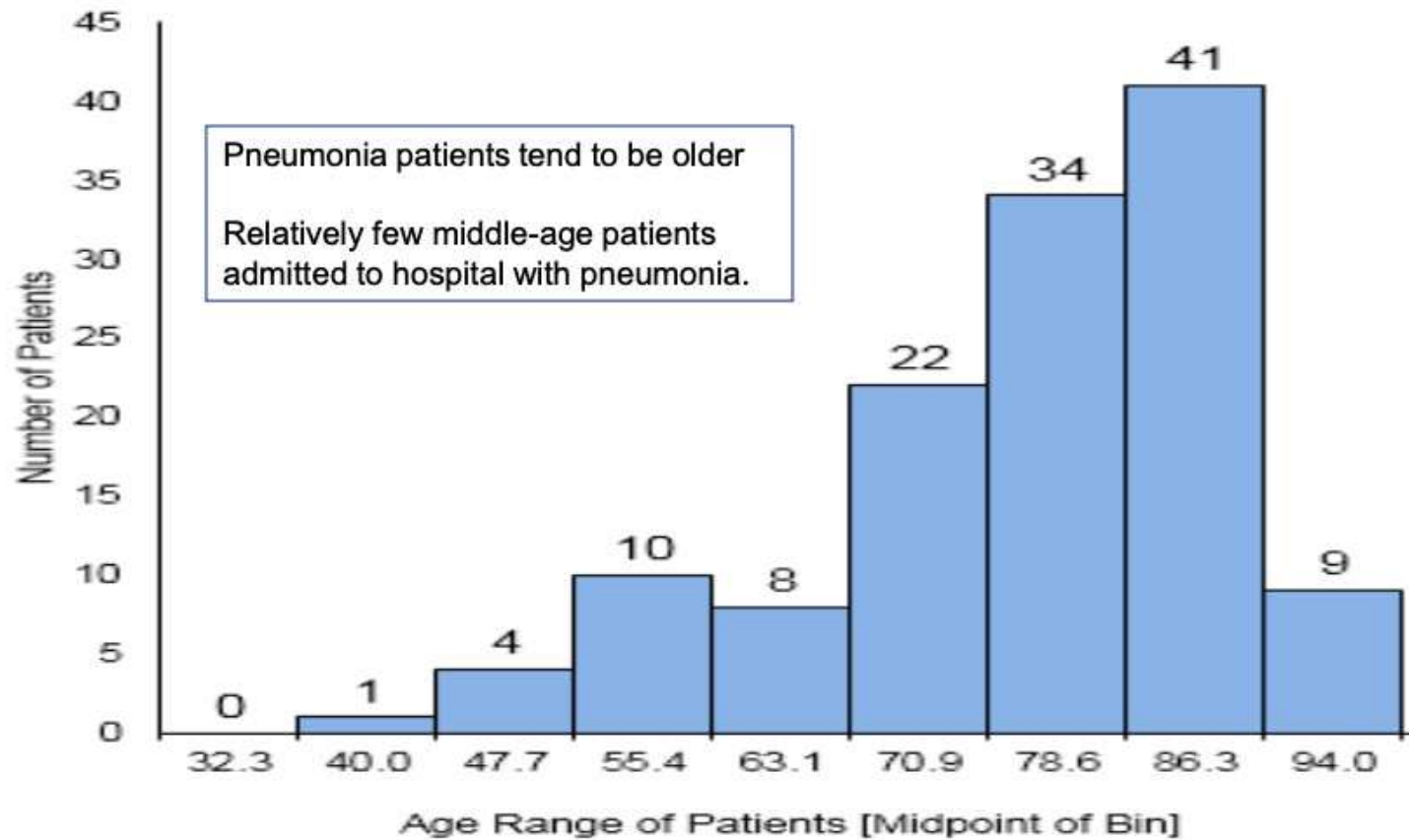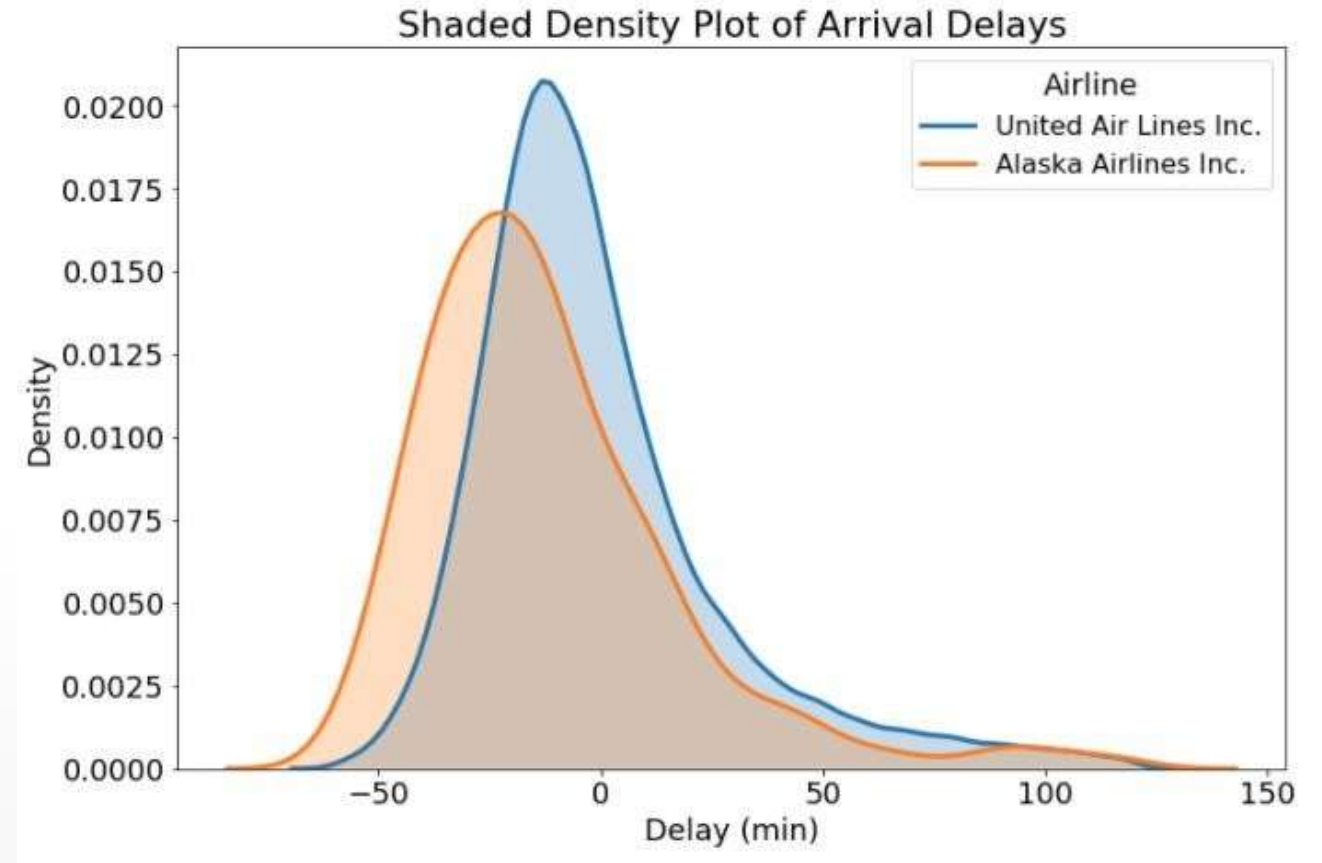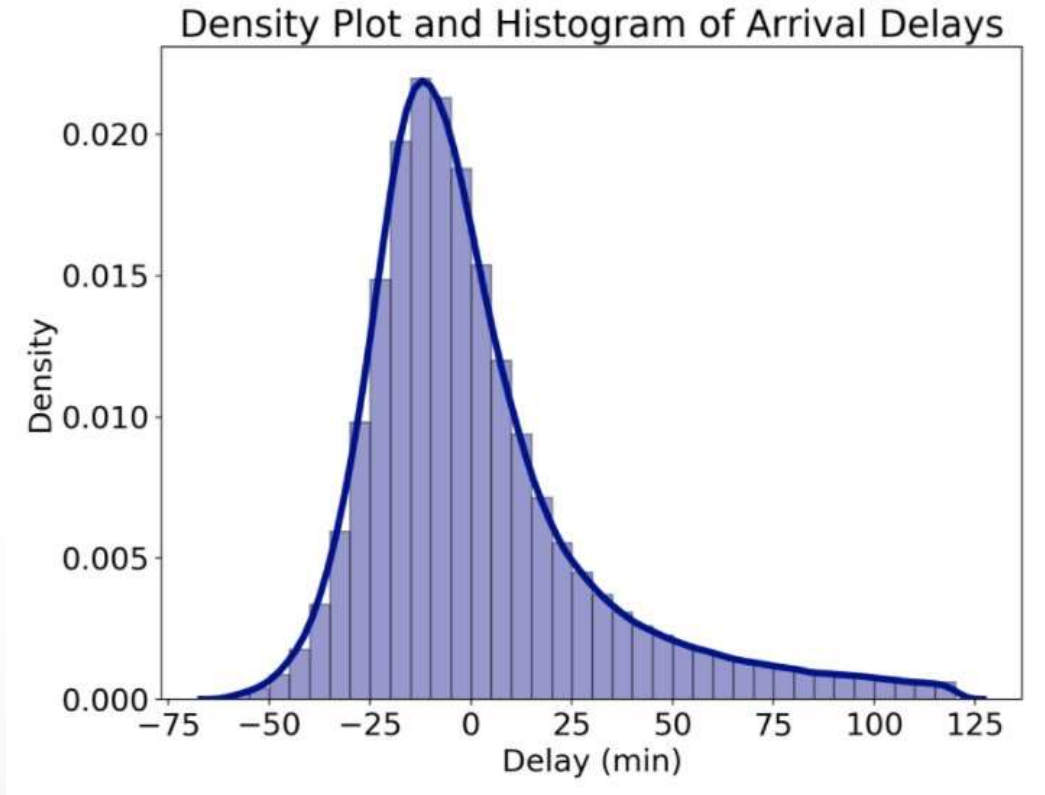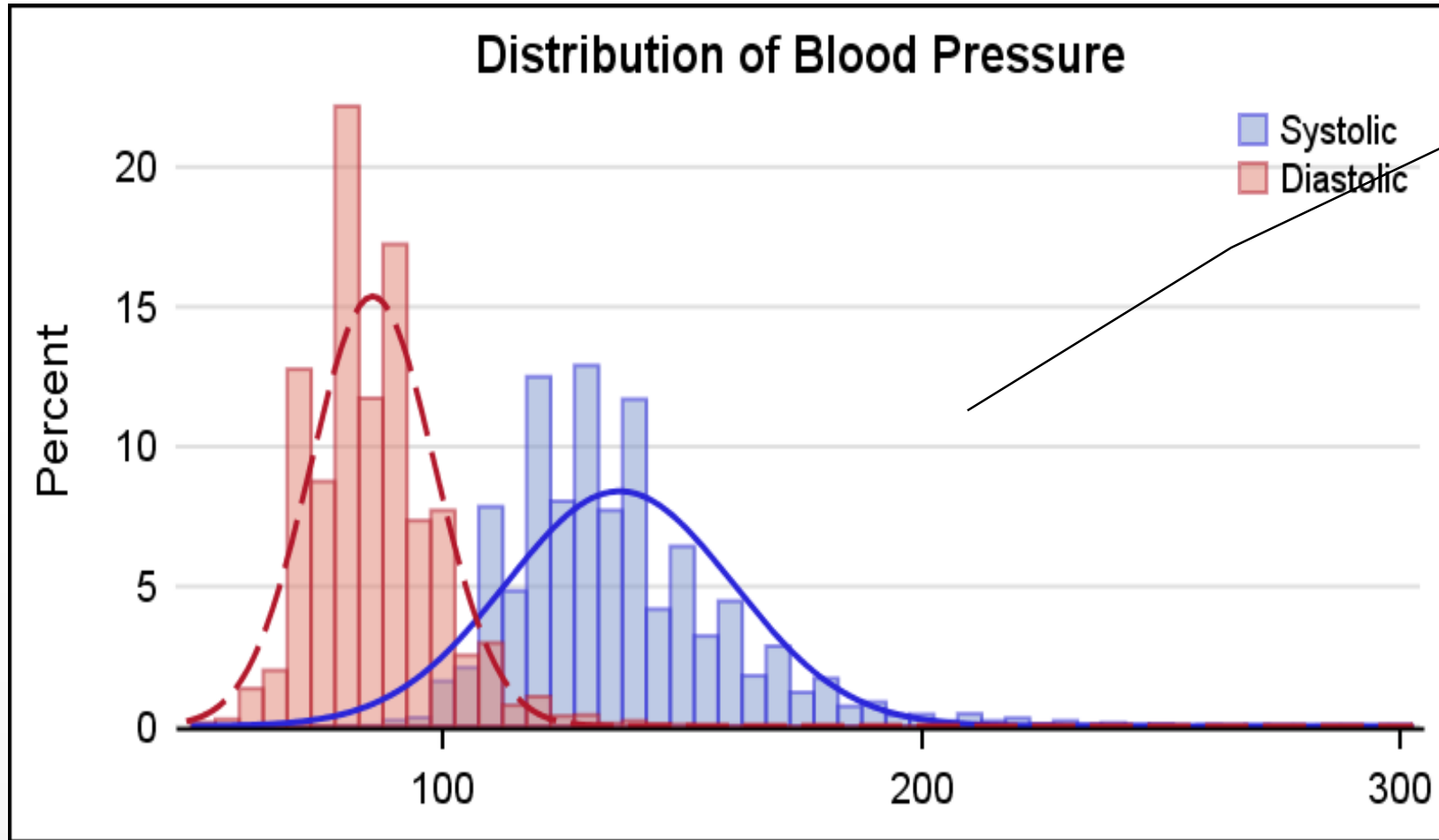Pneumonia Primary Diagnosis - December 2014

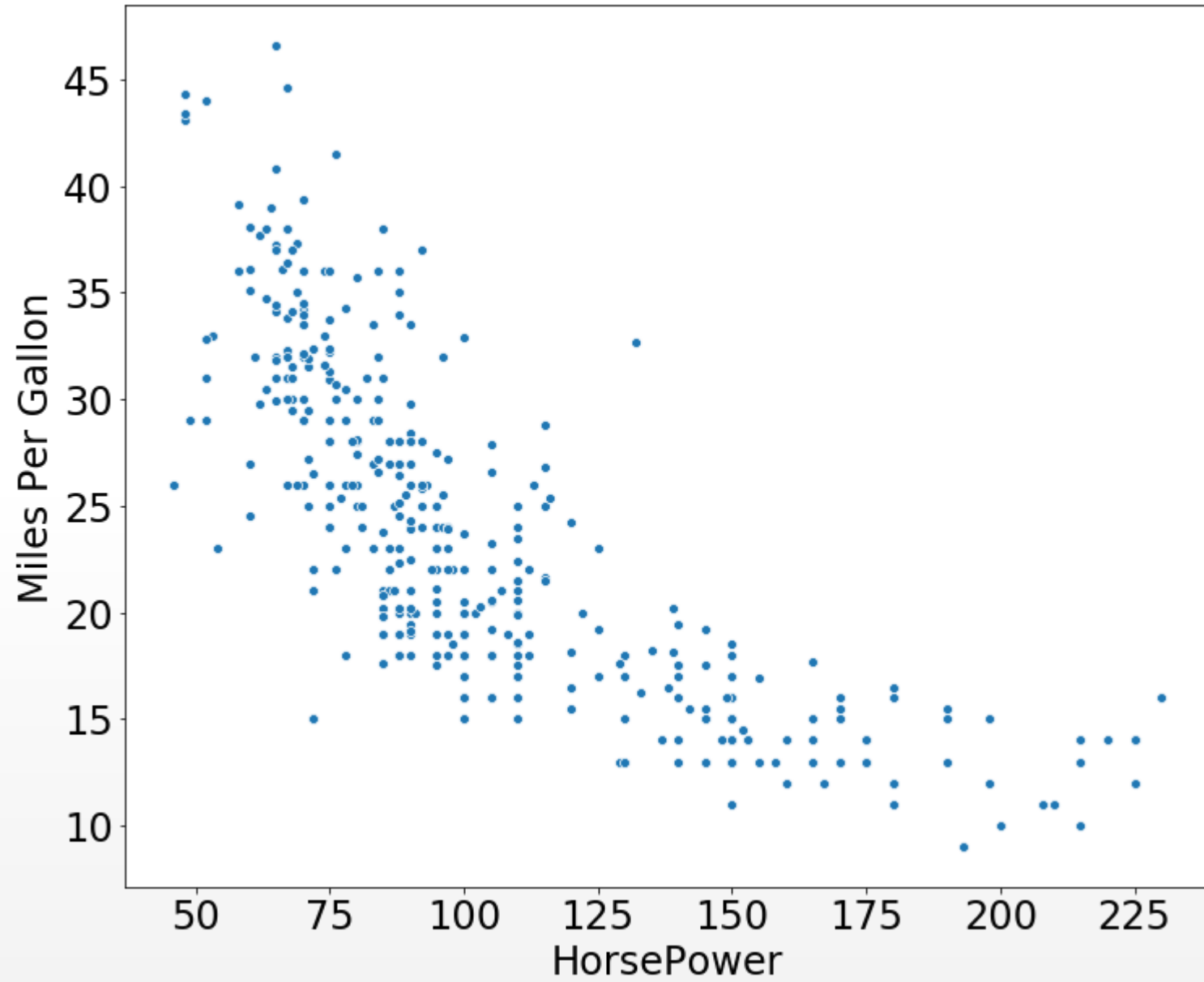# Histogram: Density plots

Histogram Example

Different subcategory of same variable.

Scatter Plots

Class Exercise

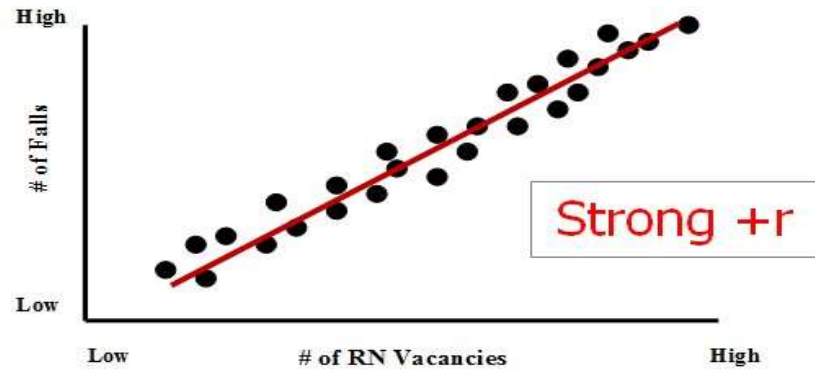|	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	name
0	18.0	8	307.0	130.0	3504	12.0	70	usa	chevrolet chevelle malibu
1	15.0	8	350.0	165.0	3693	11.5	70	usa	buick skylark 320
2	18.0	8	318.0	150.0	3436	11.0	70	usa	plymouth satellite
3	16.0	8	304.0	150.0	3433	12.0	70	usa	amc rebel sst
4	17.0	8	302.0	140.0	3449	10.5	70	usa	ford torino
5	15.0	8	429.0	198.0	4341	10.0	70	usa	ford galaxie 500
6	14.0	8	454.0	220.0	4354	9.0	70	usa	chevrolet impala
7	14.0	8	440.0	215.0	4312	8.5	70	usa	plymouth fury iii
8	14.0	8	455.0	225.0	4425	10.0	70	usa	pontiac catalina
9	15.0	8	390.0	190.0	3850	8.5	70	usa	amc ambassador dpl
10	15.0	8	383.0	170.0	3563	10.0	70	usa	dodge challenger se
11	14.0	8	340.0	160.0	3609	8.0	70	usa	plymouth 'cuda 340
12	15.0	8	400.0	150.0	3761	9.5	70	usa	chevrolet monte carlo
13	14.0	8	455.0	225.0	3086	10.0	70	usa	buick estate wagon (sw)
14	24.0	4	113.0	95.0	2372	15.0	70	japan	toyota corona mark ii
15	22.0	6	198.0	95.0	2833	15.5	70	usa	plymouth duster
16	18.0	6	199.0	97.0	2774	15.5	70	usa	amc hornet
17	21.0	6	200.0	85.0	2587	16.0	70	usa	ford maverick
18	27.0	4	97.0	88.0	2130	14.5	70	japan	datsun pl510
19	26.0	4	97.0	46.0	1835	20.5	70	europe	volkswagen 1131 deluxe sedan
20	25.0	4	110.0	87.0	2672	17.5	70	europe	peugeot 504
21	24.0	4	107.0	90.0	2430	14.5	70	europe	audi 100 ls

Scatter

Scatter plot
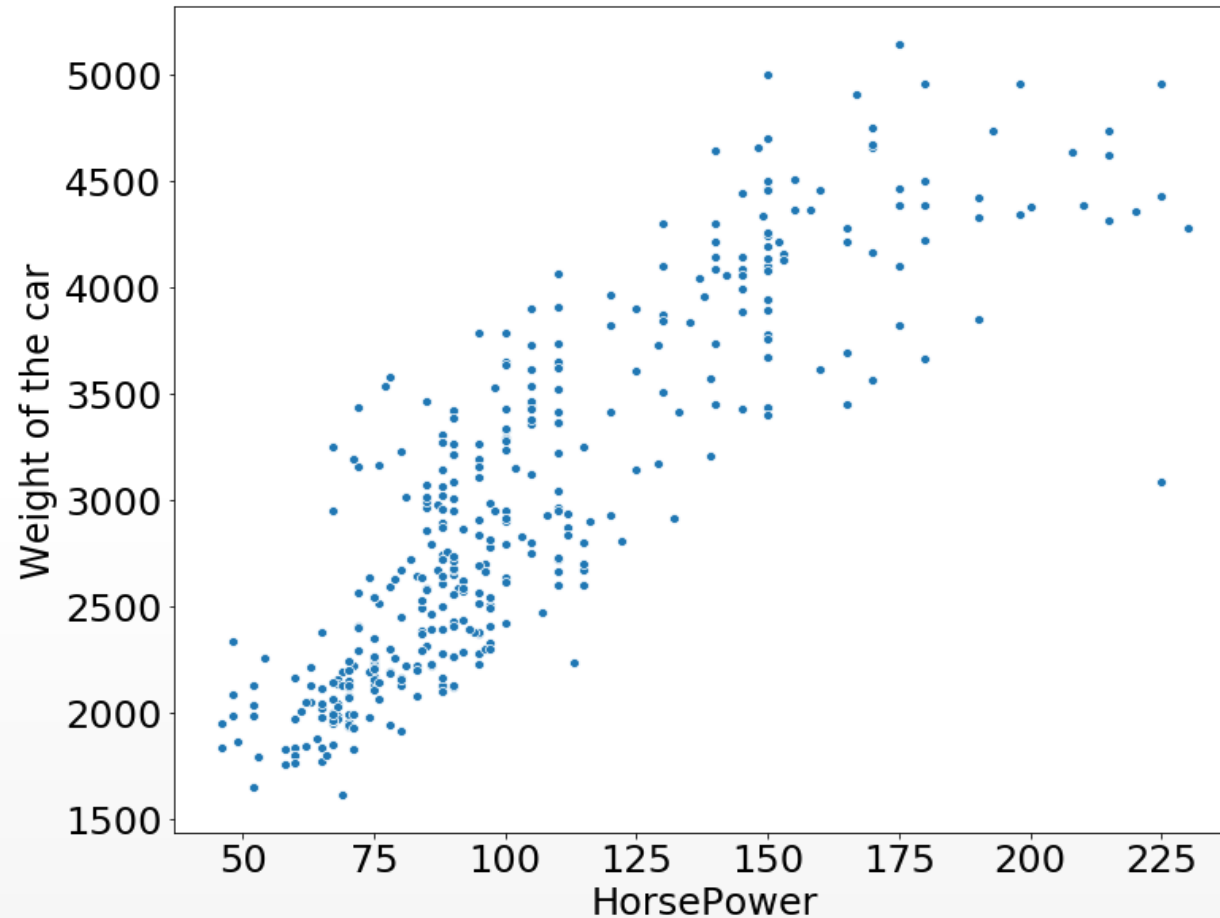
Numerical variable on both axis
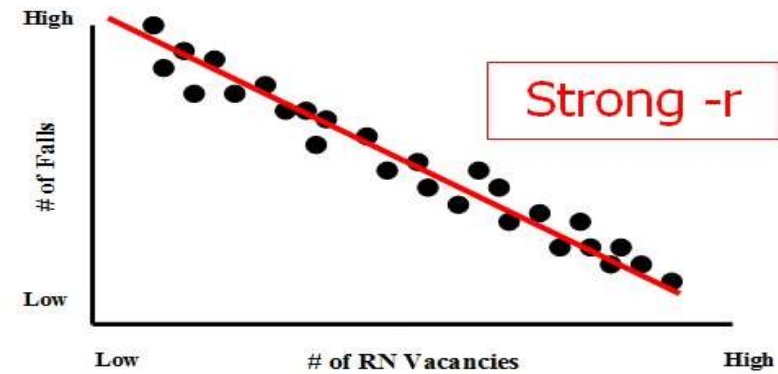
Y is dependent variable of X

Plot shows correlation between X and Y

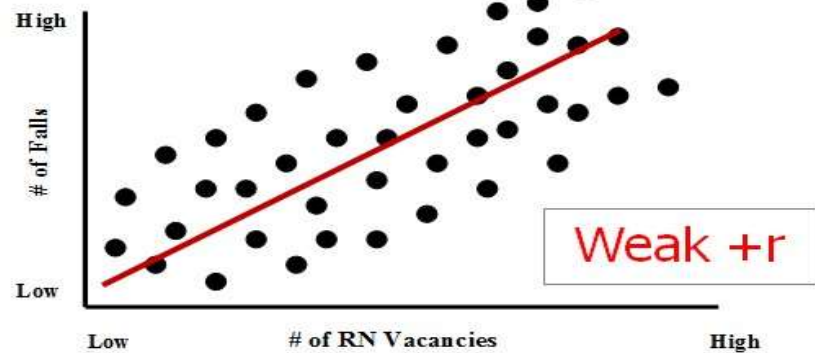# Scatter plots – correlation interpretation
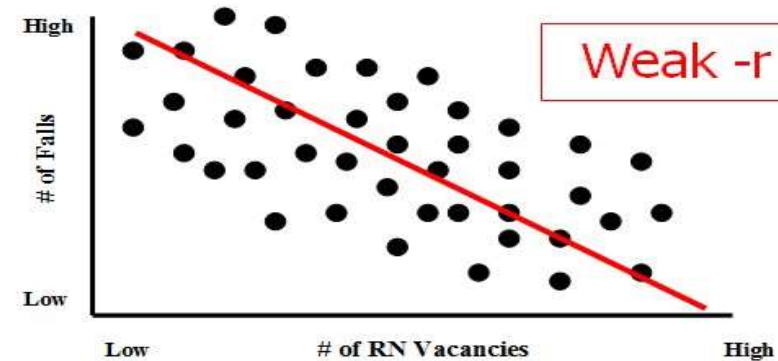
New Dimension to the plot

|	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	name
0	18.0	8	307.0	130.0	3504	12.0	70	usa	chevrolet chevelle malibu
1	15.0	8	350.0	165.0	3693	11.5	70	usa	buick skylark 320
2	18.0	8	318.0	150.0	3436	11.0	70	usa	plymouth satellite
3	16.0	8	304.0	150.0	3433	12.0	70	usa	amc rebel sst
4	17.0	8	302.0	140.0	3449	10.5	70	usa	ford torino
5	15.0	8	429.0	198.0	4341	10.0	70	usa	ford galaxie 500
6	14.0	8	454.0	220.0	4354	9.0	70	usa	chevrolet impala
7	14.0	8	440.0	215.0	4312	8.5	70	usa	plymouth fury iii
8	14.0	8	455.0	225.0	4425	10.0	70	usa	pontiac catalina
9	15.0	8	390.0	190.0	3850	8.5	70	usa	amc ambassador dpl
10	15.0	8	383.0	170.0	3563	10.0	70	usa	dodge challenger se
11	14.0	8	340.0	160.0	3609	8.0	70	usa	plymouth 'cuda 340
12	15.0	8	400.0	150.0	3761	9.5	70	usa	chevrolet monte carlo
13	14.0	8	455.0	225.0	3086	10.0	70	usa	buick estate wagon (sw)
14	24.0	4	113.0	95.0	2372	15.0	70	japan	toyota corona mark ii
15	22.0	6	198.0	95.0	2833	15.5	70	usa	plymouth duster
16	18.0	6	199.0	97.0	2774	15.5	70	usa	amc hornet
17	21.0	6	200.0	85.0	2587	16.0	70	usa	ford maverick
18	27.0	4	97.0	88.0	2130	14.5	70	japan	datsun pl510
19	26.0	4	97.0	46.0	1835	20.5	70	europe	volkswagen 1131 deluxe sedan
20	25.0	4	110.0	87.0	2672	17.5	70	europe	peugeot 504
21	24.0	4	107.0	90.0	2430	14.5	70	europe	audi 100 ls

Bubble Size
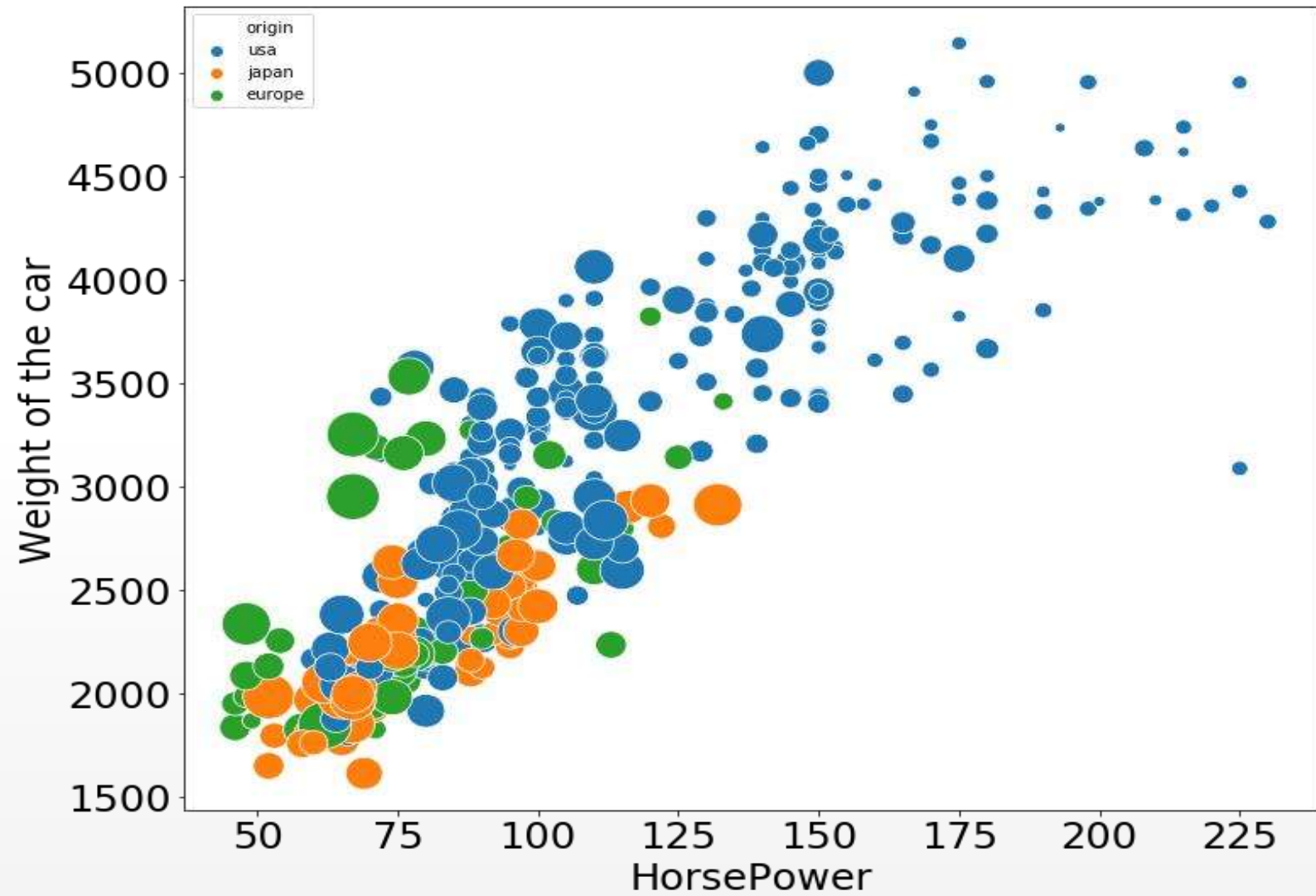
Additional Dimensions: Shape instead of colors
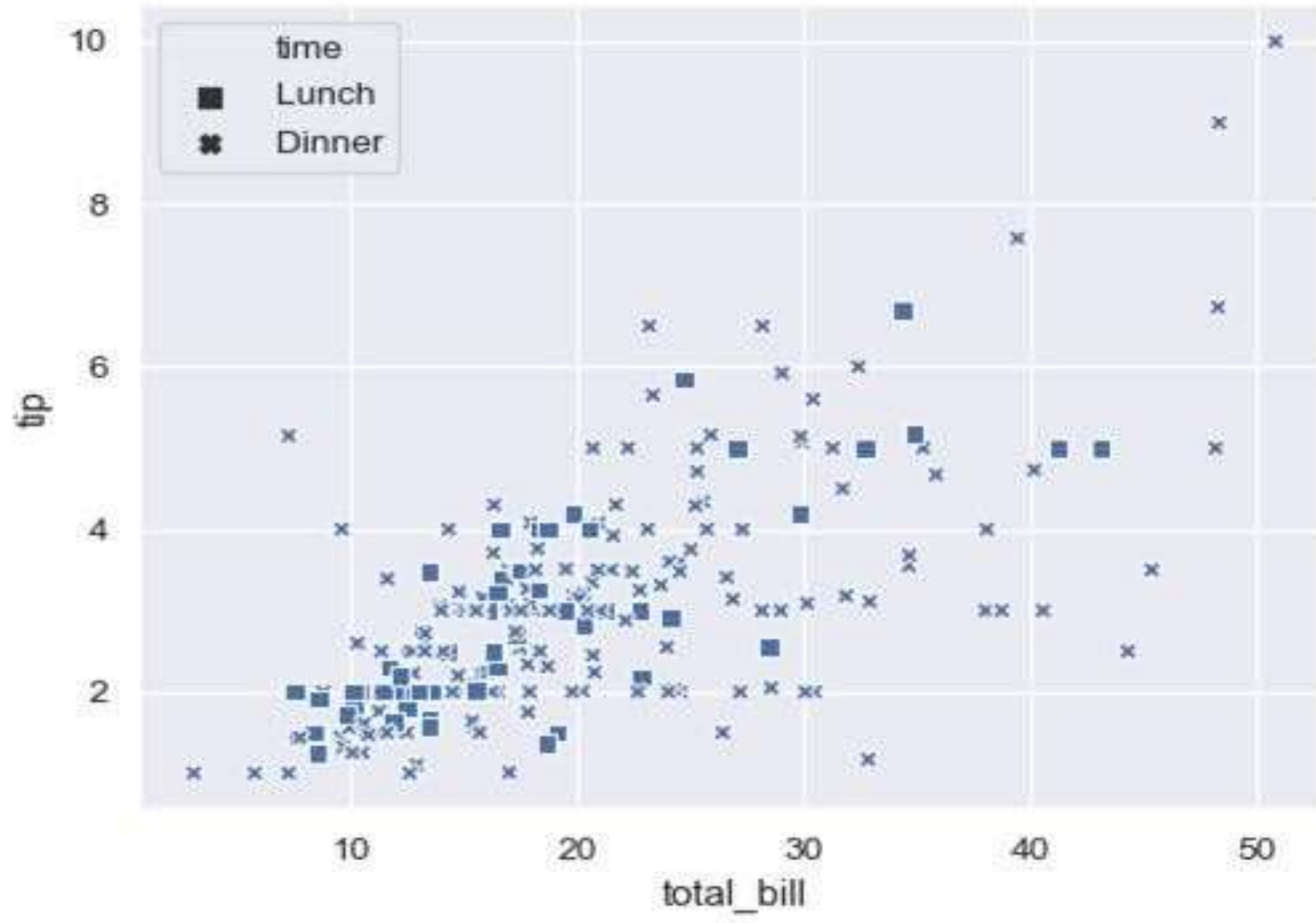
Pie charts

Pie Chart/Count plat
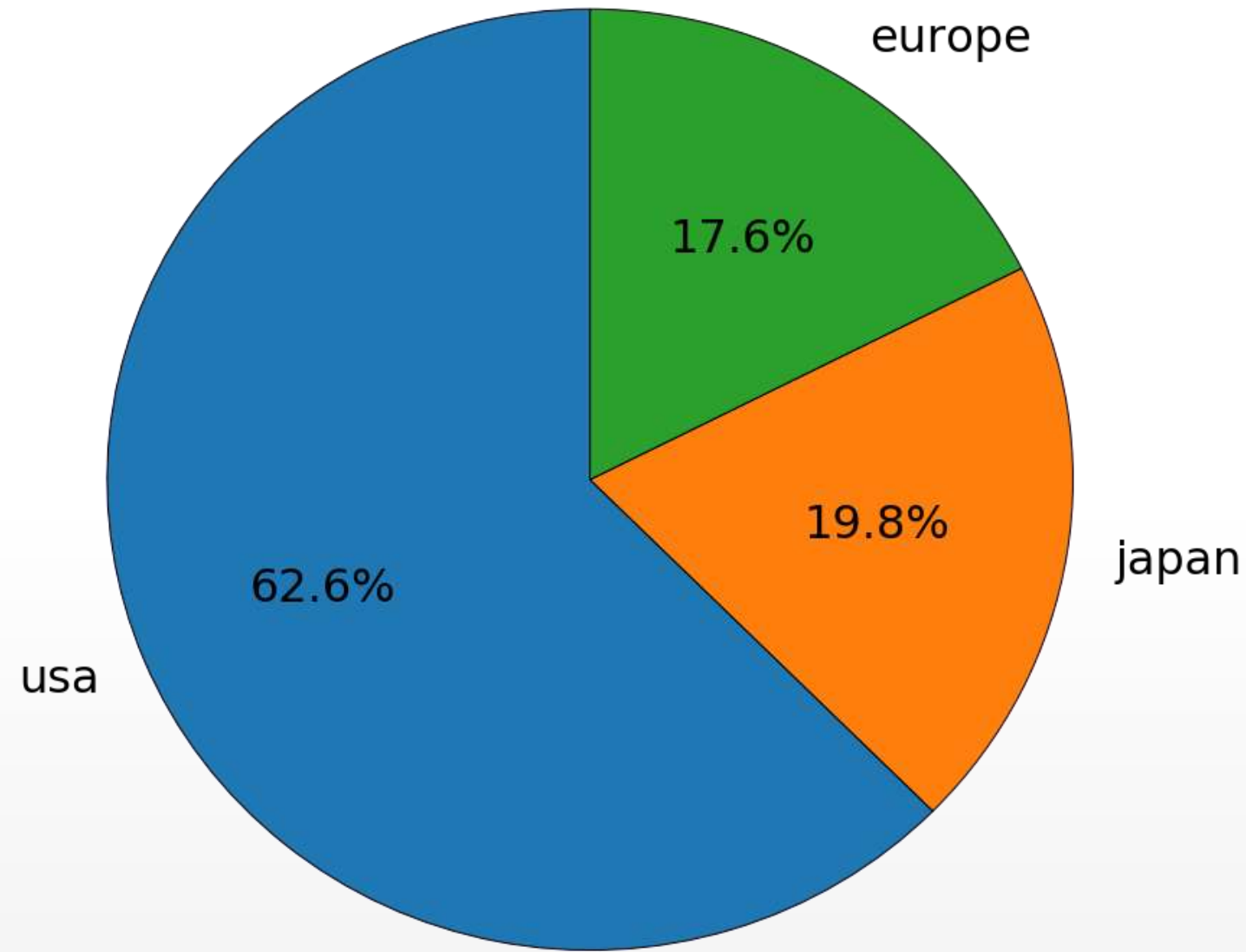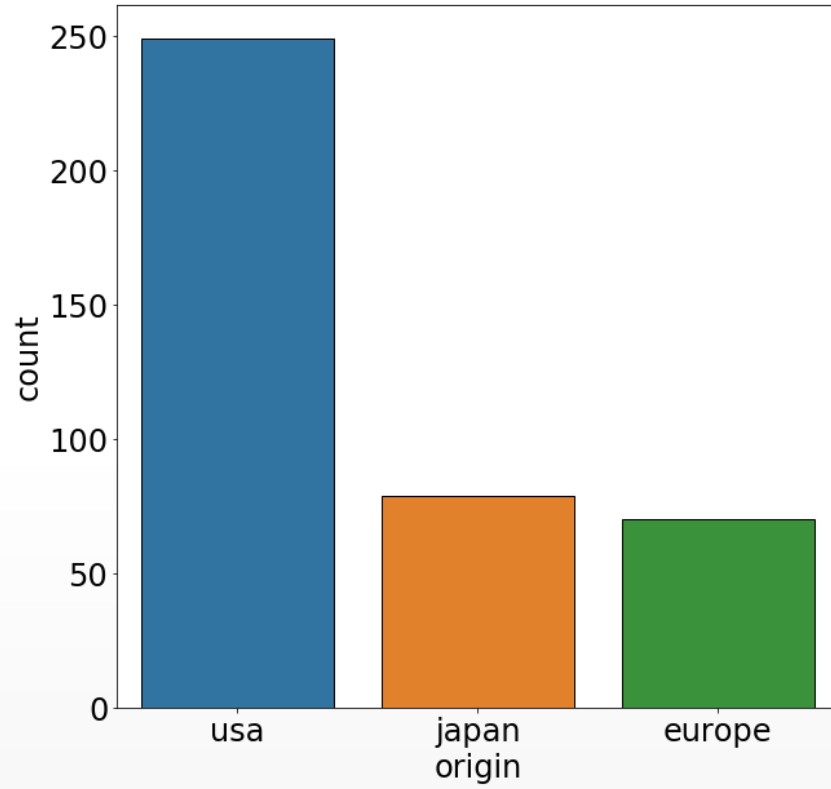
Pie charts

Facebook Ad Spending (Top 100 Buyers)

Right Leaning
17.8%

$2,651,983

State Props/Other
18.7%

$2,778,235

$9,431,551
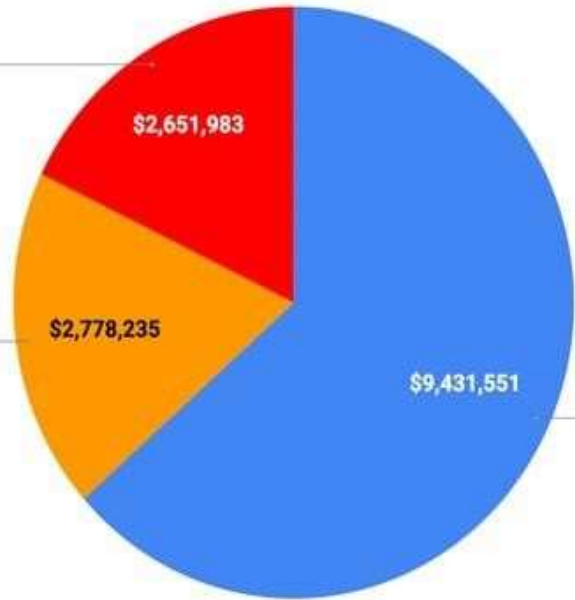
Left Leaning
63.5%

Desktop Browser Market Share 2016

IE
12%

Firefox
15%

Other
11%

Chrome
61%

Edge
3%

Opera
2%

Other
2%

Safari
5%
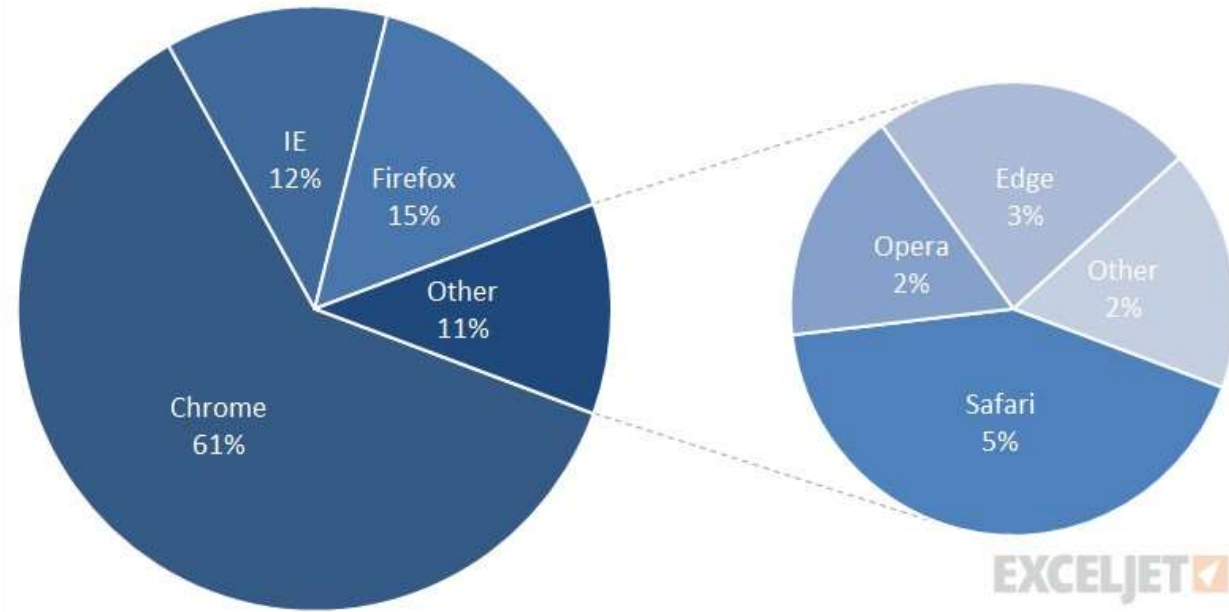
A Close Alternative: Donut Chart

INS◉FE

Borough
- Brooklyn
- Queens
- Manhattan
- Bronx
- Staten Island

Total Population
7,858,035
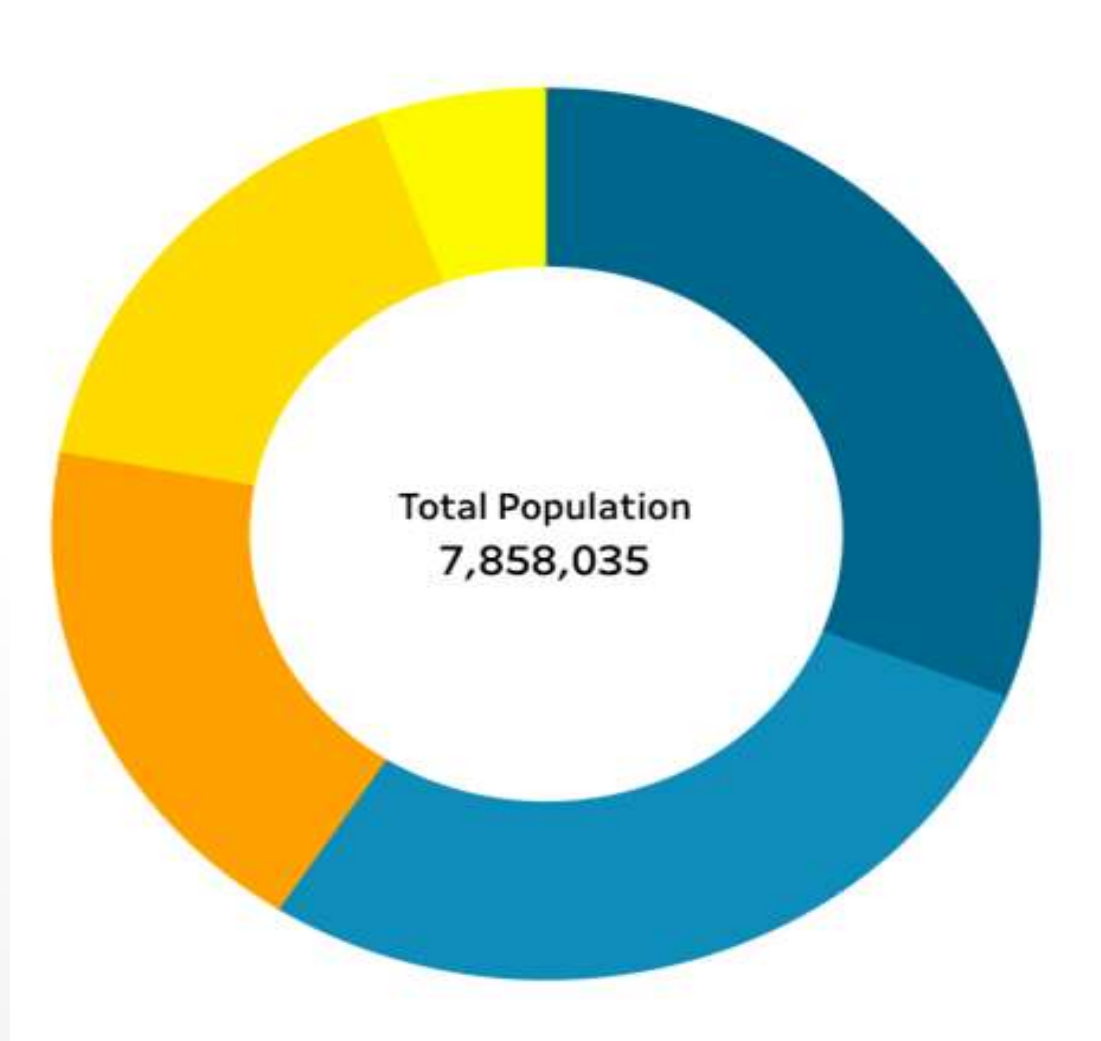
Identical information as pie chart except
1. Hollowed portion can contain additional information
2. Human eye perception is better on Dount than on pie!

Pie vs Bar Chart
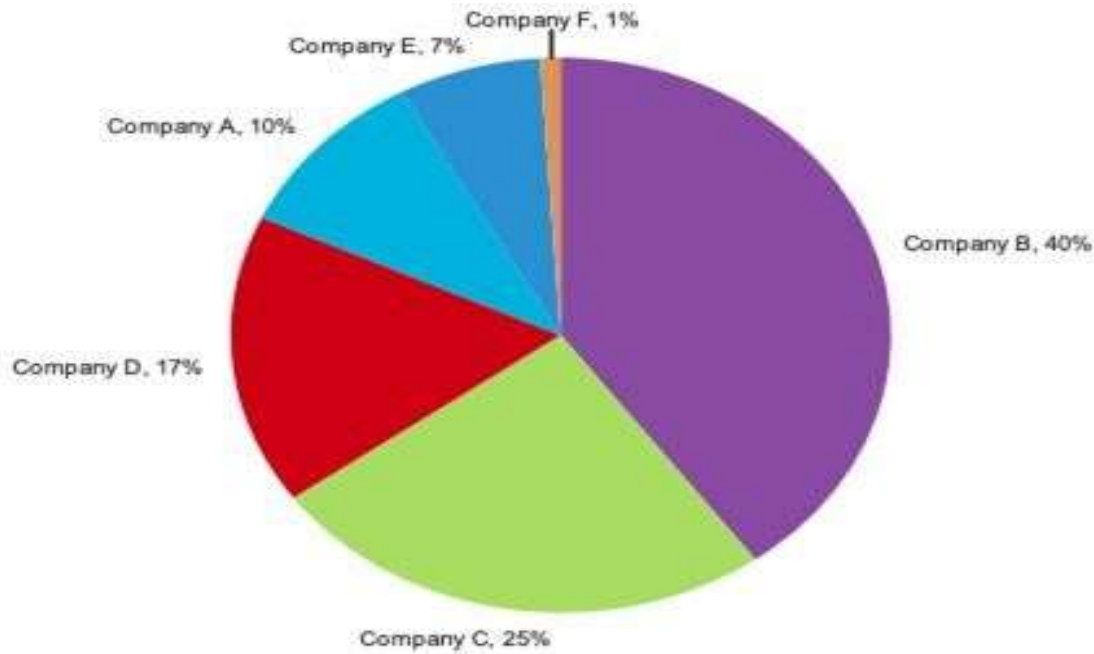
Companies	Percentage
Company B	40%
Company C	25%
Company D	17%
Company A	10%
Company E	7%
Company F	1%
Total	100%
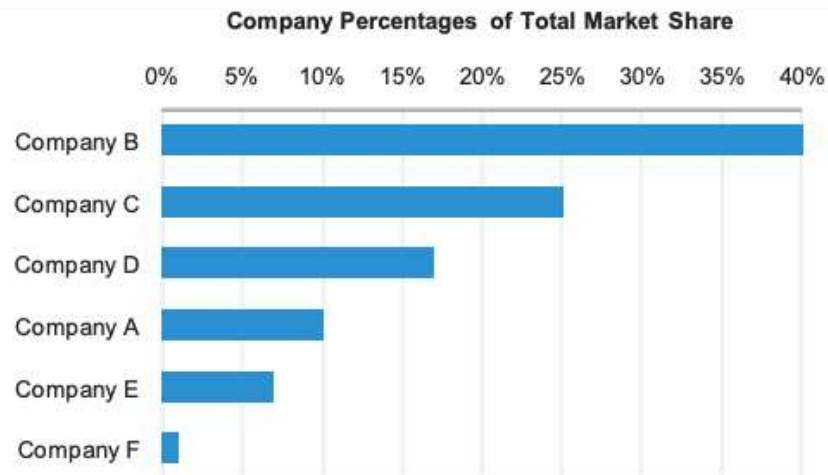
In most cases, bar charts are better alternatives

Other Plots

Map Chart

CO₂ emissions per capita, 2016
Average carbon dioxide (CO₂) emissions per capita measured in tonnes per year.

Our World in Data
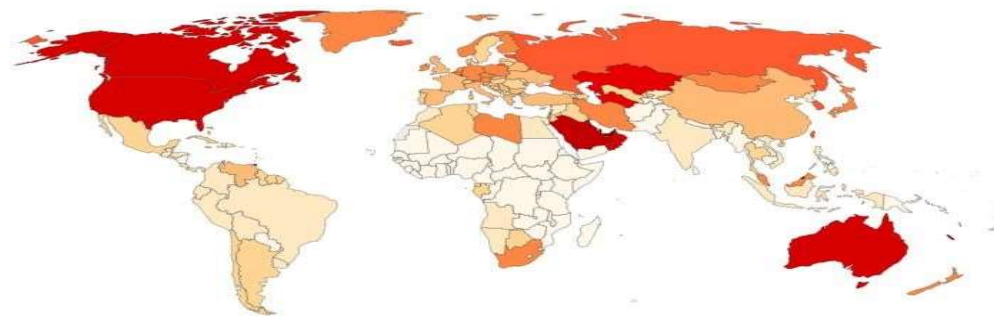
No data   0 t   1 t   2.5 t   5 t   7.5 t   10 t   12.5 t   15 t   17.5 t   20 t   25 t   >50 t

Source: OWID based on Global Carbon Project; Gapminder & UN   CC BY

▶ 1785   2016   CHART   MAP   DATA   SOURCES

Same data: Several Alternatives
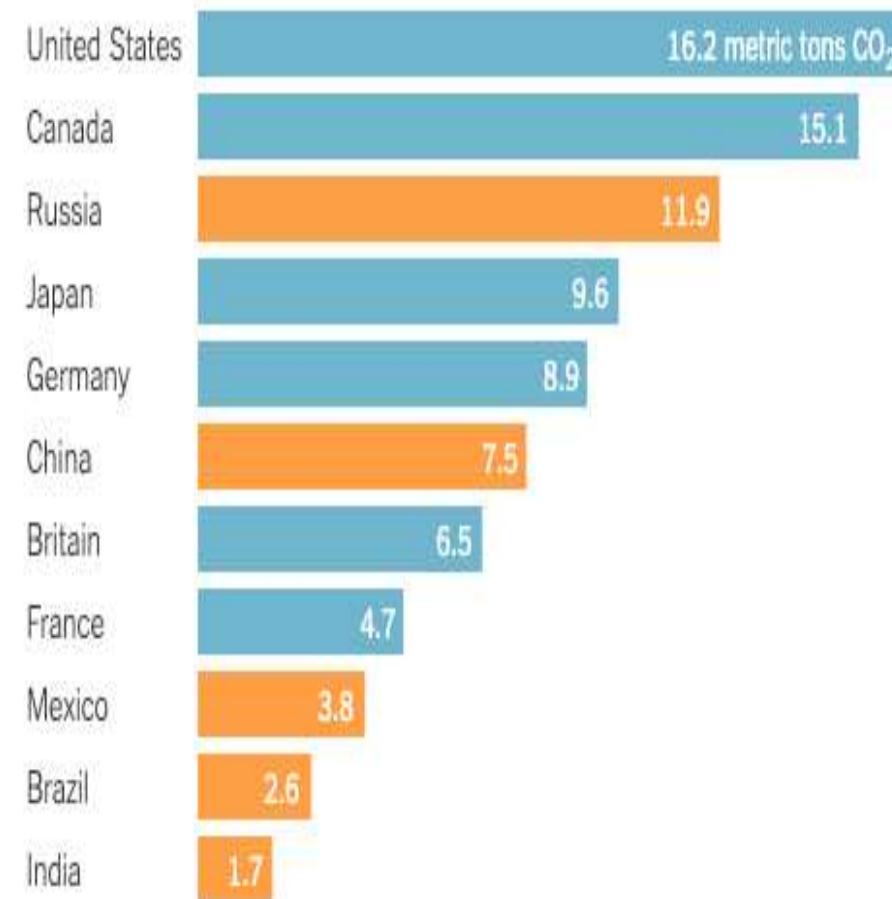
Summary and Review

Graph and Usage Summary

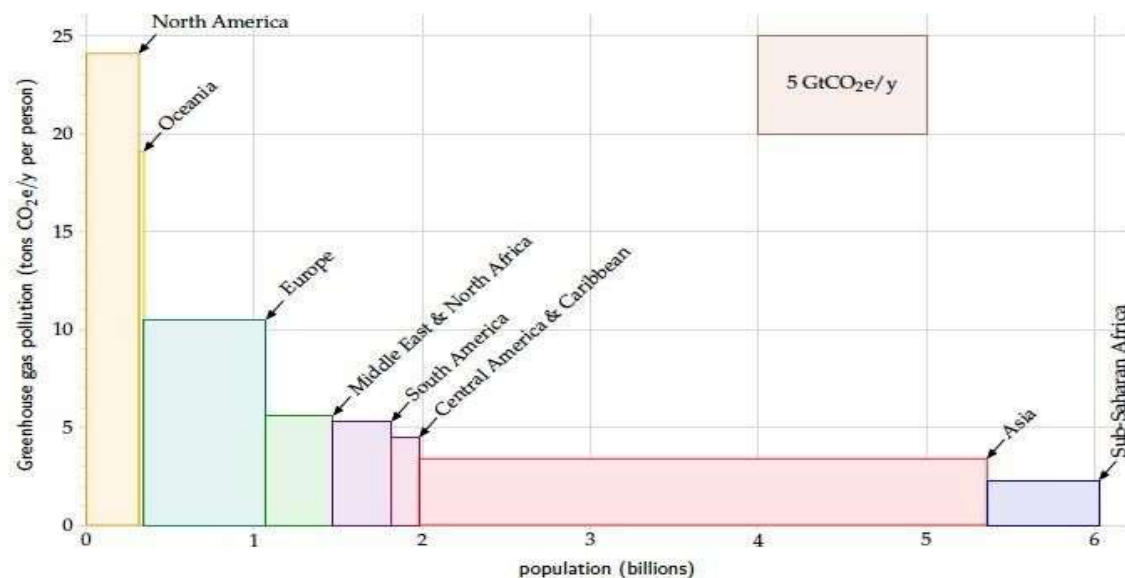| Graph Type	Usage	Additional Comments
Bar graphs	Comparing data values within or across categories; Discrete data	Consider Line graphs for Continuous data
Histograms	Distribution of values across a possible range	
Line graphs	Continuous data; Display trends	
Time series	Data with a time dimension	
Pie graphs	Comparing fractions of a whole; Very few fractions and precision is not important	Avoid to the extent possible
Gauge charts	Comparing values between a small number of variables	
Scatter plots	Understanding correlations between two quantitative dimensions of data	3 or 4 dimensions possible by encoding data points as bubbles, etc.
Heat maps	Area graphs that use colour or brightness to indicate values (or changes in value) of large data sets; Show relationships between 2 factors	
Box plots	Understanding distribution of a numerical data; Comparing distributions across categories; Identify outliers	
Tree maps	Display hierarchical data in rectangles	

Chart types based on data (Simpler view)

INSOFE

Categorical vs Numerical	Numerical Distributions	Numerical vs Numerical
• Bar • Line • Area • Pie • Geographical Map	• Box • Histogram	• Scatter

❖ Note
- A very high level distribution of most commonly used charts
- You may need to explore several beyond this

TABLEAU INTRODUCTION

- A tableau is a business intelligence and visualisation tool.

- It was designed to help the user to make sense of their data through interactive charts and graphics.

- Without the help of any program or any prior knowledge of programming.

Why Tableau ?

- User Friendly

- Easily establish a secure connection

- It is a easy drag and drop functionality

- Advanced visualisations

- Trend lines and predictive analysis

- Interactive dashboards

- Collaboration and sharing

- Mobile view

Tableau (By Salesforce)

- Salesforce announced that it has closed the $15.7 billion Tableau deal in June 2019. The deal is by far the biggest acquisition in the Salesforce history, a company known for being highly acquisitive

Source:

- https://techcrunch.com/2019/08/01/salesforce-closes-15-7b-tableau-deal/

Tableau Products

Tableau has 5 main products :
1. Desktop
2. Public
3. Online
4. Server
5. Reader

Desktop : Tableau desktop is to analyse data, create workbooks, visualisation, dashboards and stories.

Public : Tableau is a free service that lets anyone publish interactive dashboard on the web they can interact with the data, download it, or create their own visualisations of it out of it.

Online : Tableau Online, is a secure, cloud-based solution for sharing, distributing, and collaborating on content created in Tableau desktop.

Server : Tableau server is an online solution for sharing, distributing and collaborating on content created in tableau desktop. Tableau users can create workbooks, views and dashboards in tableau desktop and then publish this content to the server.

Reader : Tableau reader is a free desktop application that you can use to open and interact with data visualisation built in tableau desktop.

Types Of Charts

- Bar chart
- Stacked bar chart
- Pie chart
- Line chart
- Scatter plot
- Bubble chart
- Word cloud
- Geo maps
- Funnel chart
- Tree maps

Pie   Bar   Column   Line   Area   Doughnut

Bubble Chart   Spider and Radar   Scatter   Comparison Chart   Stacked bar chart   Gauges

Data types
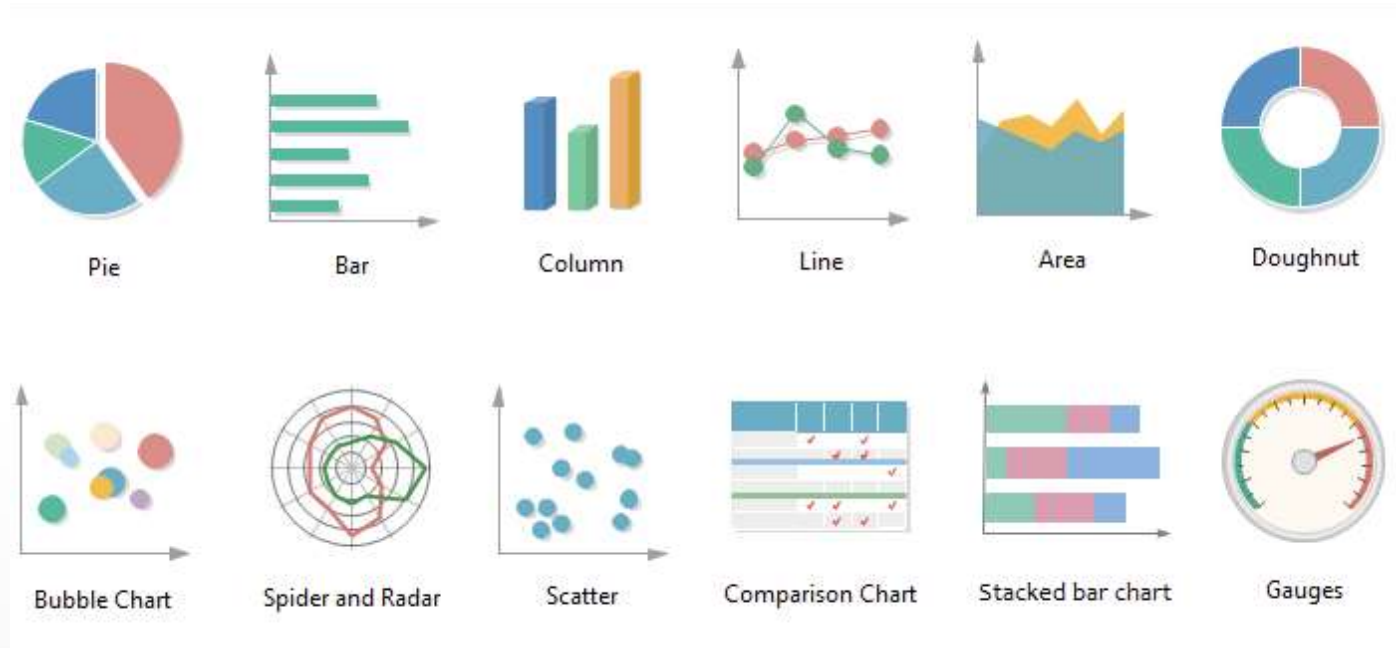
1. Abc   - Text / String / Character value number
2. Calendar - Date values
3. Calendar - Date and time values
4. #   - Numeric / Integer values / Float
5. T|F   - Boolean values
6. Globe   - Geographic

Save your work

Two different file format we can save the work books. They are:
.twb
.twbx

.twb : Tableau Workbook:
➢ The Tableau Workbook is the basic file created by Tableau Desktop. This includes all of the worksheets, dashboards, story points, etc. Everything that you've done within the workbook .twb is only opened in desktop.

.twbx: Tableau Packaged Workbook:
➢ twbx file is a Tableau Packaged Workbook, meaning it is the original .twb file grouped together with the datasource(s) in one package.
➢ The primary advantage to using .twbx files is that analysis can be performed.
➢ A twbx is a " zipped" archive containing a twb + any external files associated with that workbook, such as extracts and background images.

INSOFE's Vision

The BEST GLOBAL DESTINATION for individuals and organizations to learn and adopt disruptive technologies for solving business and society's challenges.

INSOFE

Map locations: GLASGOW, OTTAWA, CLEVELAND, RENNES, ROPAR, MUMBAI, HYDERABAD, BENGALURU

INSOFE - HYDERABAD
2nd Floor, Jyothi Imperial, Vamsiram Builders
Janardana Hills, Gachibowli
Hyderabad – 500032
+91 93199 77257

INSOFE - BENGALURU
Floors 1-3, L77, 15th Cross Road
Sector 6, HSR Layout
Bengaluru - 560102
+91 93199 77267

INSOFE - MUMBAI
4th Floor - A Wing, Spaces - Kanaki
Andheri-Kurla Road, Chakala
Andheri East, Mumbai - 400093
+91 93199 77269

Email:
info@insofe.edu.in

Website:
www.insofe.edu.in

Follow us on Social Media: /insofeglobal/   /school/insofe/   /INSOFEedu   /insofe_global/   /InsofeVideos