Business Analytics in Practice

Data Visualization Using R: ggplot2

Hakeem-Ur-Rehman, PhD

IQTM–PU

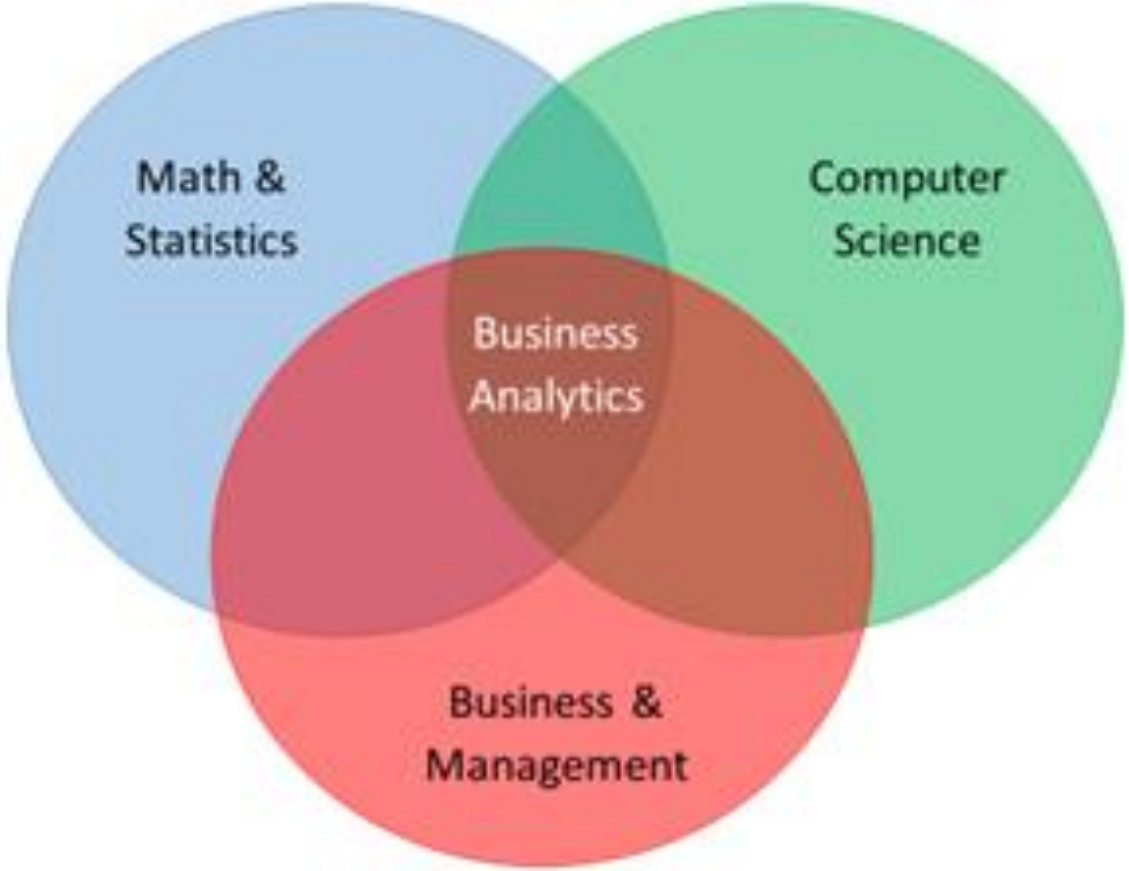
Outlines

- What is Business Analytics?
- Types of Business Analytics
- Business Analytics in Practice
  - HR Analytics
  - Marketing Analytics
  - Supply Chain Analytics
- Data Visualization Using R
  - R: Graphics Packages
  - Descriptive Analysis of Qualitative Data
  - Descriptive Analysis of Quantitative Data
  - Data Visualization for Different Data Stories
    - Data Visualization to Show Deviations, Correlations, Rankings, Distributions, Magnitudes
- Correlation & Regression Analysis Using R (YouTube Links)
- Parametric Testing of Hypothesis (YouTube Link)
- Business Analytics Using Python (YouTube Links)

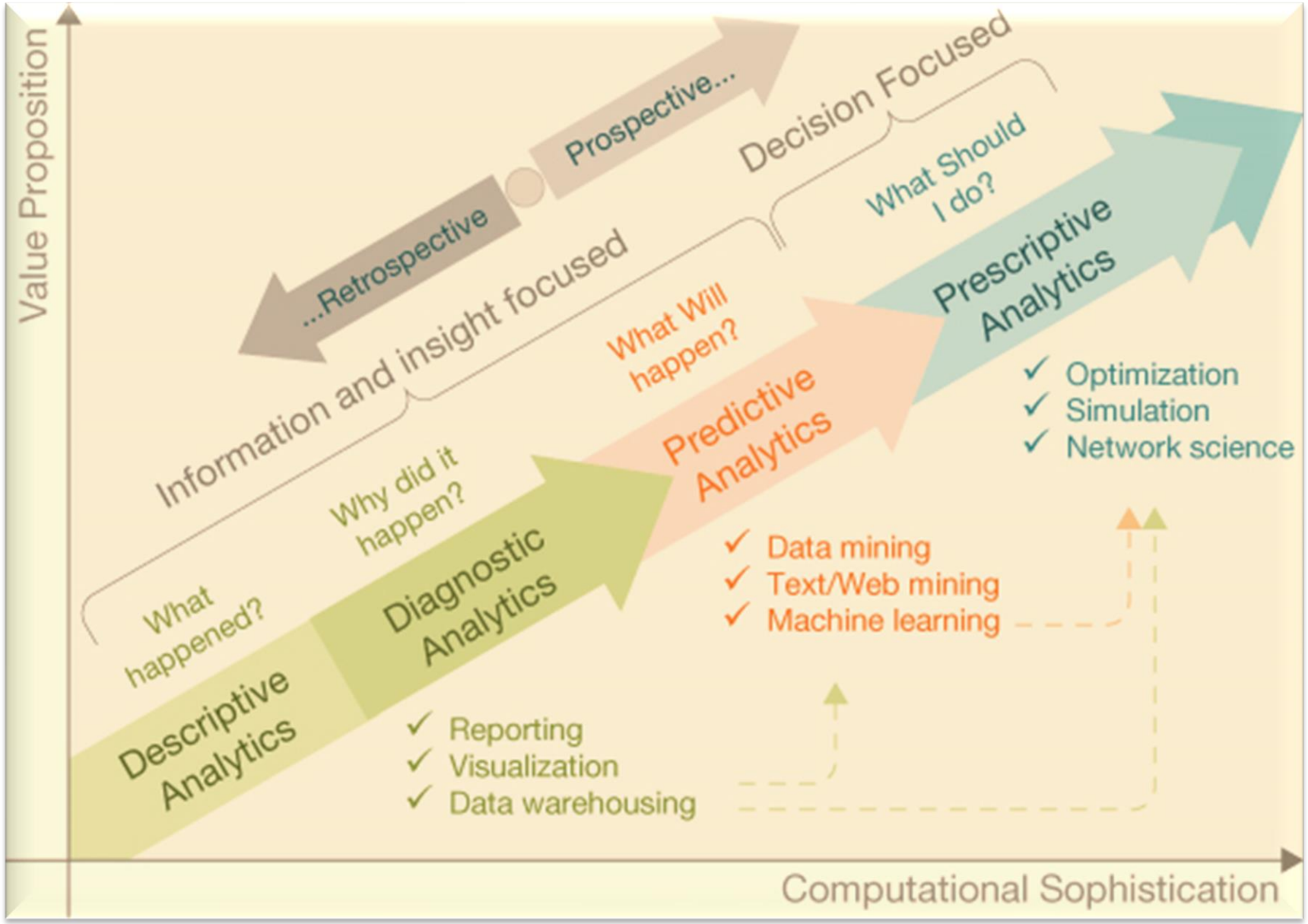
What is Business analytics?

- **Business analytics** is the scientific process of transforming data into insight for making better decisions (*fact-based management to drive decision making*)

A Visual Perspective of Business Analytics



Types of Business Analytics



Business Analytics in Practice: HR Analytics

Google refers to its HR Analytics function as "people analytics."

Google has analyzed substantial data on their own employees

- to determine the characteristics of great leaders,
- to assess factors that contribute to productivity, and
- to evaluate potential new hires

Google also uses predictive analytics to continually update their forecast of future employee turnover and retention

HR Data: *https://www.aihr.com/blog/hr-data-sources/*

Data Science / Machine Learning Data: *https://www.kaggle.com/*

Business Analytics in Practice: Marketing Analytics

Customer segmentation and a better understanding of *consumer behavior* through analytics leads to the better use of

- advertising budgets,
- more effective pricing strategies,
- improved forecasting of demand,
- improved product-line management, and
- increased customer satisfaction and loyalty

Marketing analytics tells you how your marketing programs are really performing by using important business metrics, such as ROI, marketing attribution and overall marketing effectiveness.

Business Analytics in Practice: Supply Chain Analytics

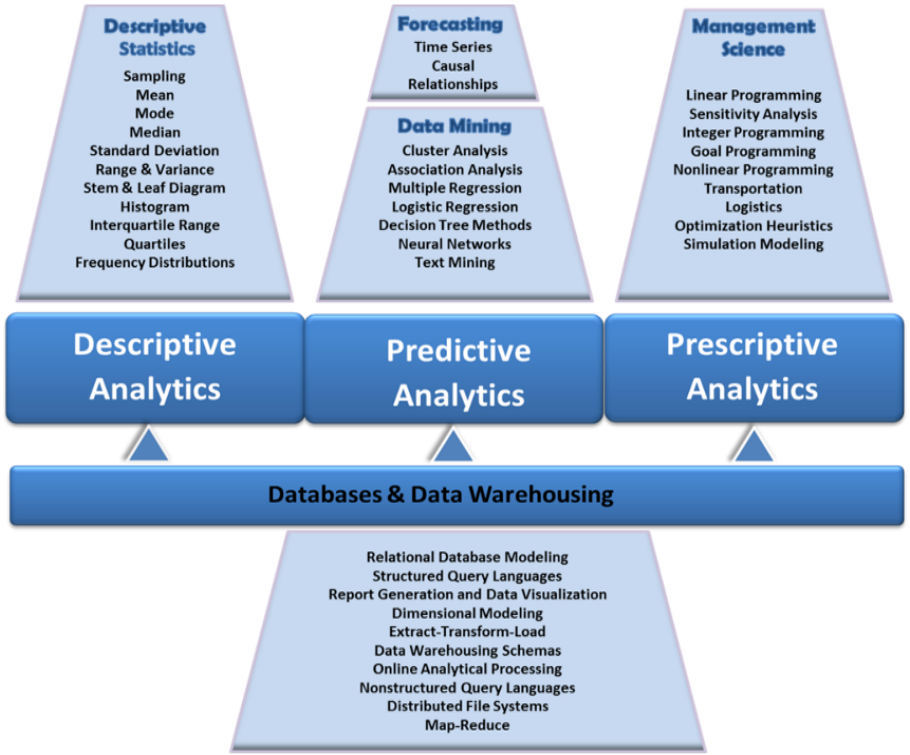
SCOR Domain	Source	Make	Deliver	Return
Activities	Order and receive materials and products	Schedule and manufacture, repair, remanufacture, or recycle materials and products	Receive, schedule, pick, pack, and ship orders	Request, approve, and determine disposal of products and assets
Strategic (time frame: years)	• Strategic sourcing • Supply chain mapping	• Location of plants • Product line mix at plants	• Location of distribution centers • Fleet planning	• Location of return centers
Tactical (time frame: months)	• Tactical sourcing • Supply chain contracts	• Product line rationalization • Sales and operations planning	• Transportation and distribution planning • Inventory policies at locations	• Reverse distribution plan
Operational (time frame: days)	• Materials requirement planning and inventory replenishment orders	• Workforce scheduling • Manufacturing, order tracking, and scheduling	• Vehicle routing (for deliveries)	• Vehicle routing (for returns collection)
Plan	Demand forecasting (long term, mid term, and short term)			

Analytics Techniques	Source	Make	Deliver	Return
Descriptive	• Supply chain mapping	• Supply chain visualization		
Predictive	• Time series methods (e.g., moving average, exponential smoothing, autoregressive models) • Linear, non-linear, and logistic regression • Data-mining techniques (e.g., cluster analysis, market basket analysis)			
Prescriptive	• Analytic hierarchy process • Game theory (e.g., auction design, contract design)	• Mixed-integer linear programming (MILP) • Non-linear programming	• Network flow algorithms • MILP • Stochastic dynamic programming	

Souza, G. C. (2014). Supply chain analytics. Business Horizons, 57(5), 595-605.

Business Analytics in Practice

The Process of Using techniques such as…



In domains such as…

To perform tasks such as…

- Measure Sales Force Effectiveness
- Campaign & Promotion Analysis
- Customer segmentation & Profiling
- Demand forecasting
- Warehouse planning
- Inventory Management
- Logistics Optimization
- Analysis of clinical trails
- Fraud Detection
- Many more

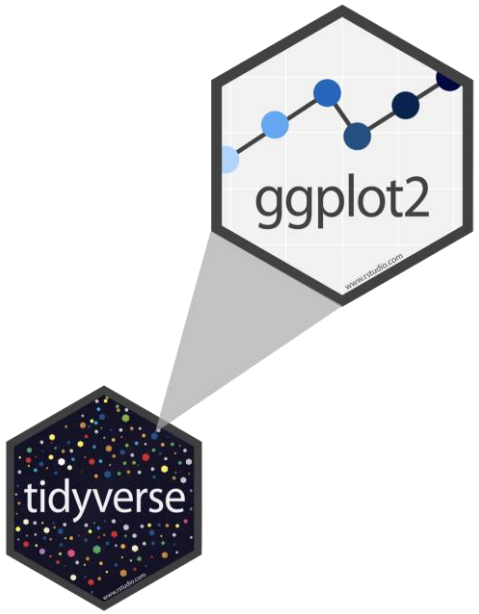
Data Visualization Using R

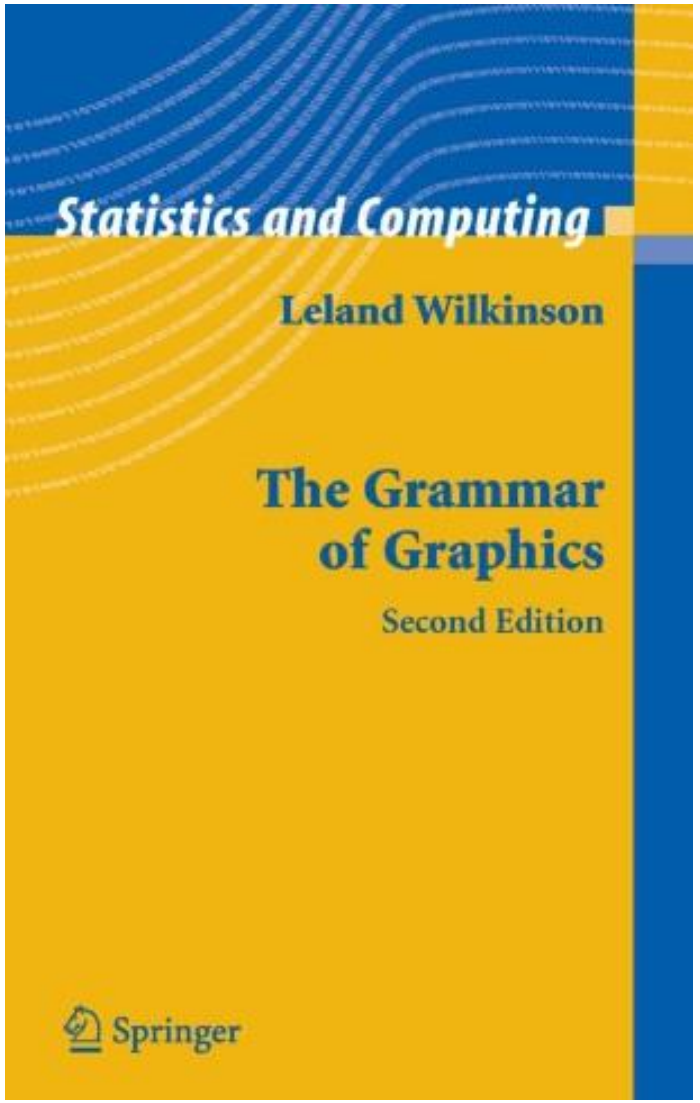
Data Visualization: Best practices for Business Users

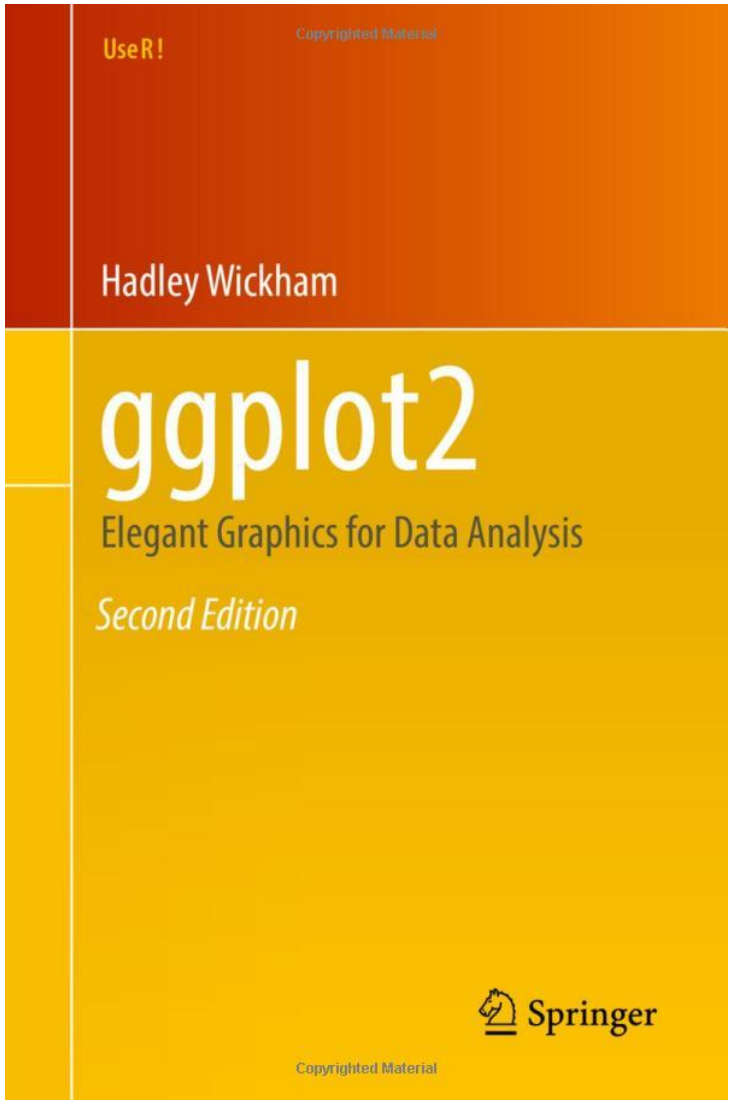
Data Visualization Techniques for Better Data Story Telling

R: Graphics Packages

1. ggplot2
2. ggvis
3. ggforce
4. Lattice
5. Plotly
6. patchwork
7. quantmod
8. RGL
9. Colourpicker
10. Esquisse







R Package: ggplot2

ggplot Syntax:

- Based on Grammar of Graphics book by Leland Wilkinson hence ‘gg’

- **Data** → must be stored as an R data frame
  - **ggplot**(data = df)

- **Aesthetic mapping (“aes”** → x-axis, y-axis, fill(“inside color”), color (“outside”), line type, size, shape “of points”)
  - **ggplot**(data = df, **aes**(x=categorical.var, fill=group.var)) + **geom_bar**()

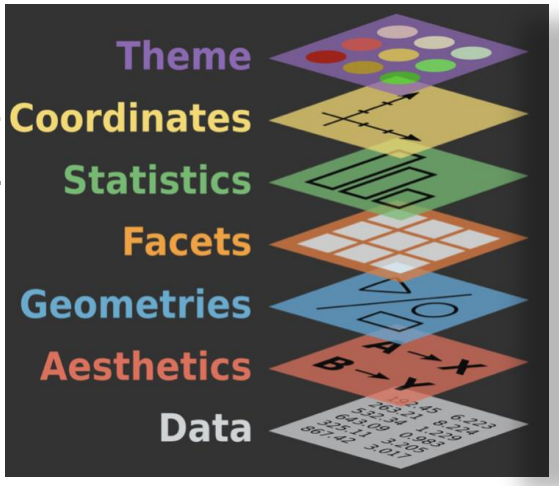
- **Geometric Object** → “geom_” + Plot Type → “*A plot must have at least one geom; there is no upper limit* ”
  - **ggplot**(data = df) + **geom_bar**()

All non-data ink.

The space on which the data will be plotted.

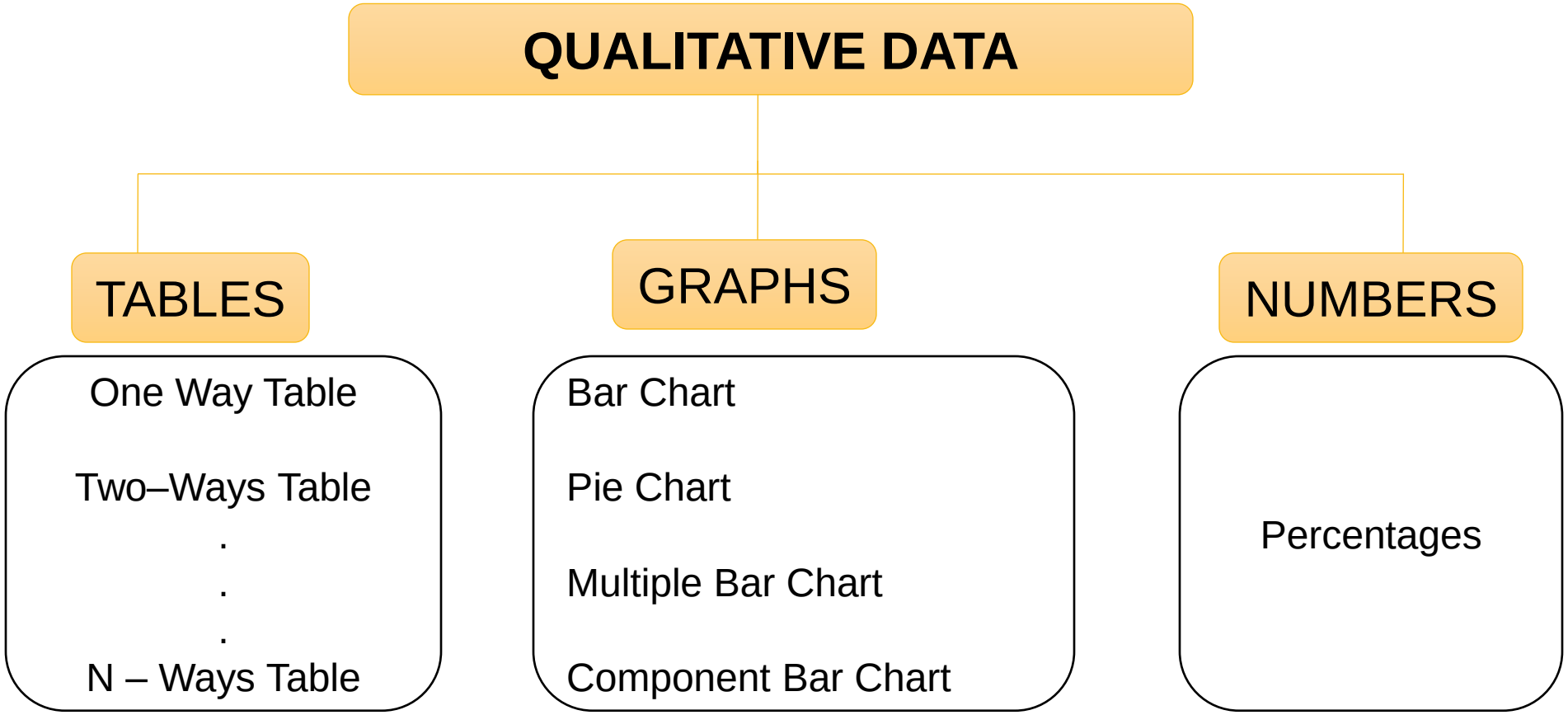
Representations of our data to aid understanding.

Plotting small multiples.



Data	*{variables of interest}*				
Aesthetics	*x-axis* *y-axis*	*colour* *fill*	*size* *labels*	*alpha* *shape*	*line width* *line type*
Geometries	*point*	*line*	*histogram*	*bar*	*boxplot*
Facets	*columns*	*rows*			
Statistics	*binning*	*smoothing*	*descriptive*	*inferential*	
Coordinates	*cartesian*	*fixed*	*polar*	*limits*	
Themes	*non-data ink*				

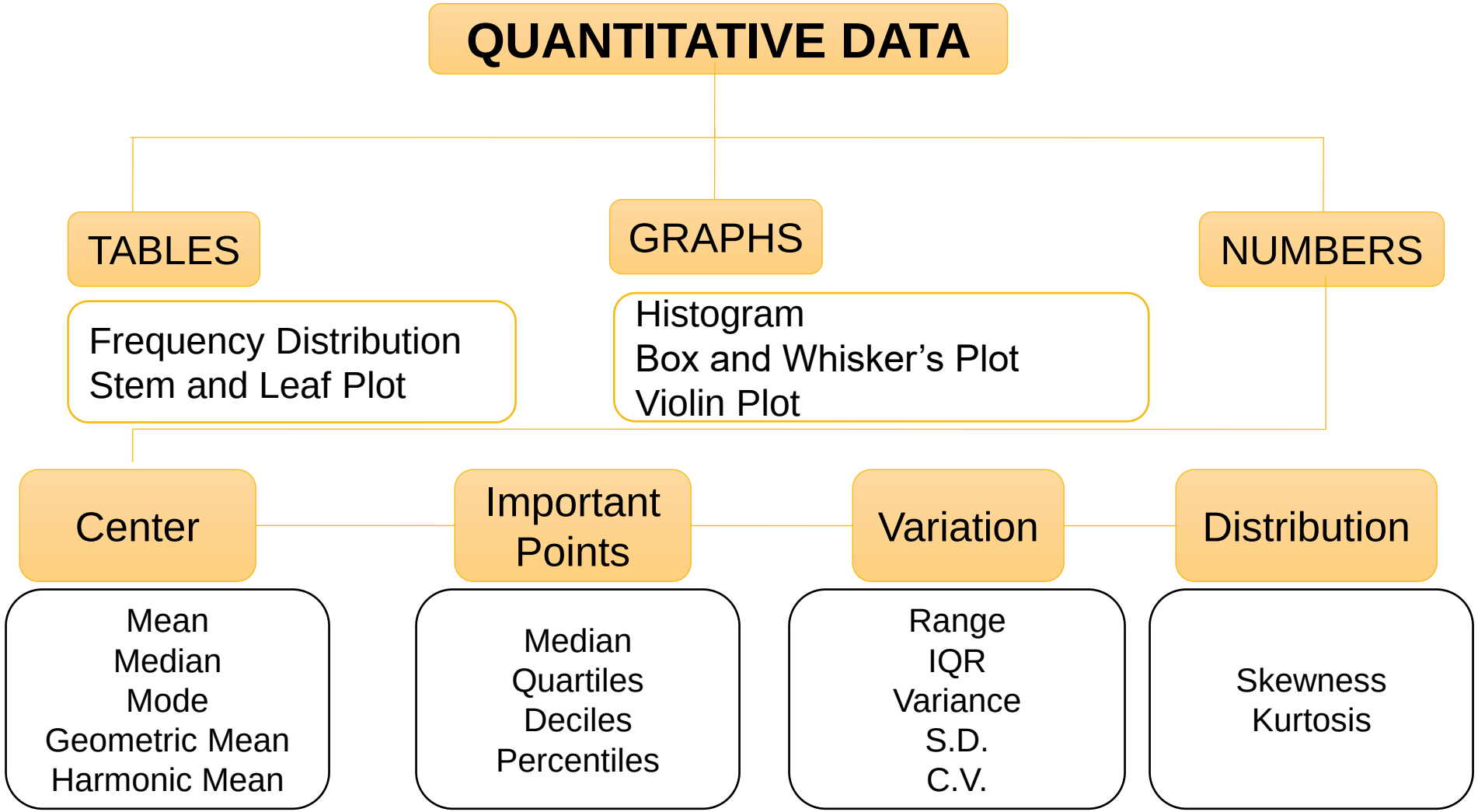
Descriptive Analysis of Qualitative Data



Line Plot:

- *Line graphs are used to track changes over short and long periods of time.*
- When smaller changes exist, line graphs are better to use than bar graphs.
- Line graphs can also be used to compare changes over the same period of time for more than one group.

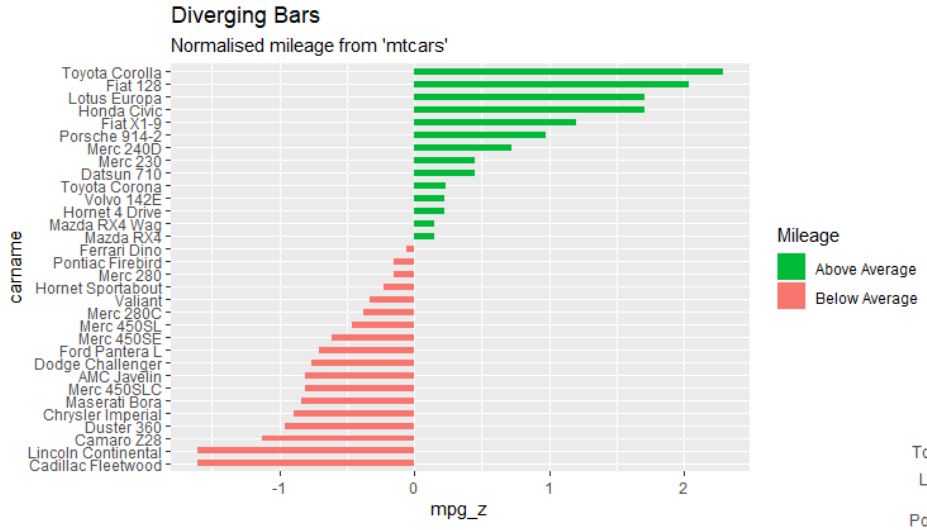
Descriptive Analysis of Quantitative Data

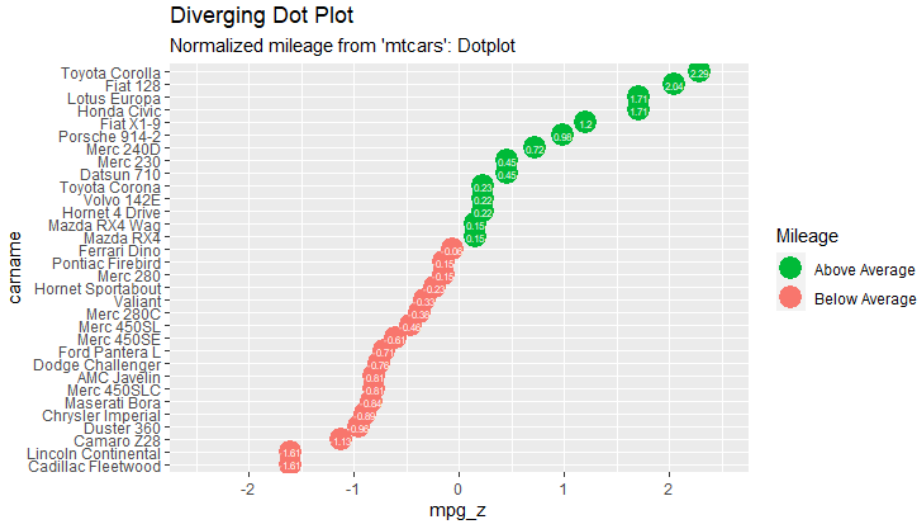


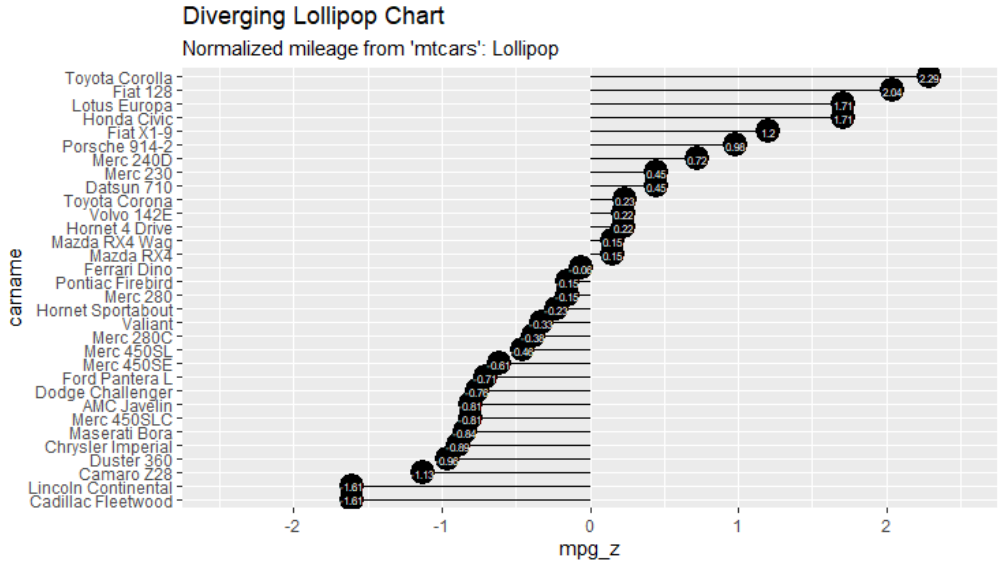
Data Visualization for Different Data Stories

Data Visualization to Show Deviations

- Variations, be it positive or negative, are compared with a reference point, which is usually zero. However, reference points can also be a target or a long-term average.
  - Diverging Bar Chart
  - Diverging Dot Plot
  - Diverging Lollipop Chart



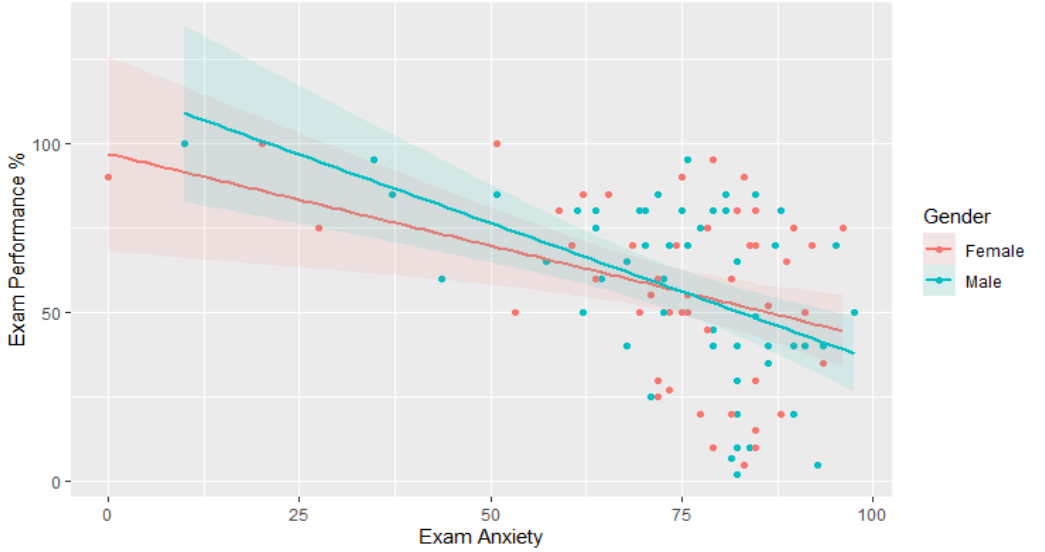


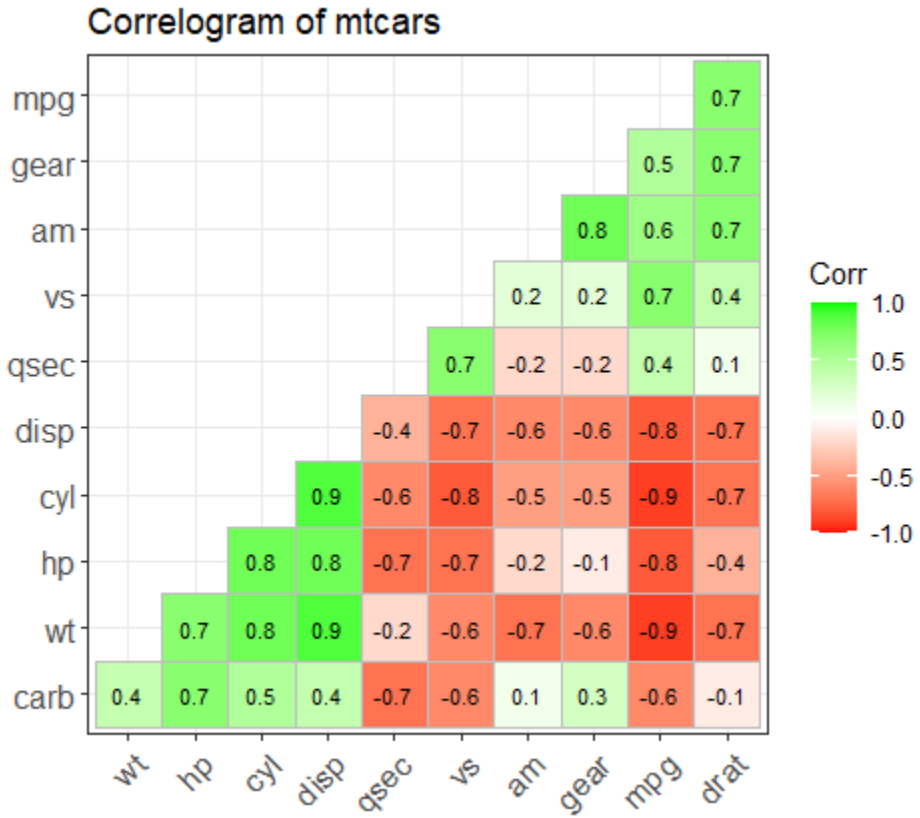


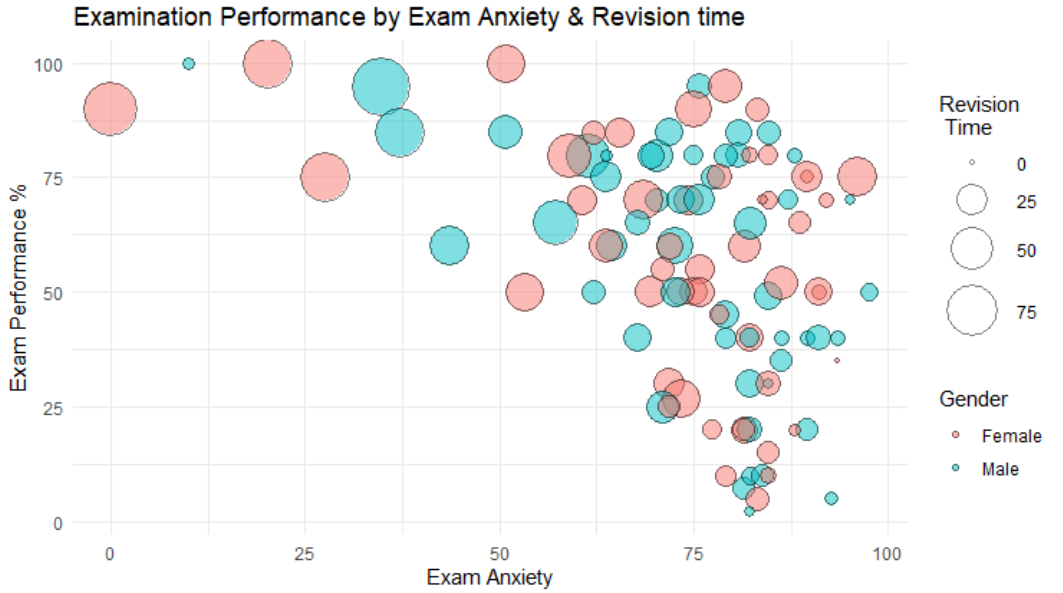
Data Visualization for Different Data Stories

Data Visualization to Show Correlations

- To show the relationship between two or more variables.
  - Scatter Plot
  - Bubble Plot
  - Correlogram



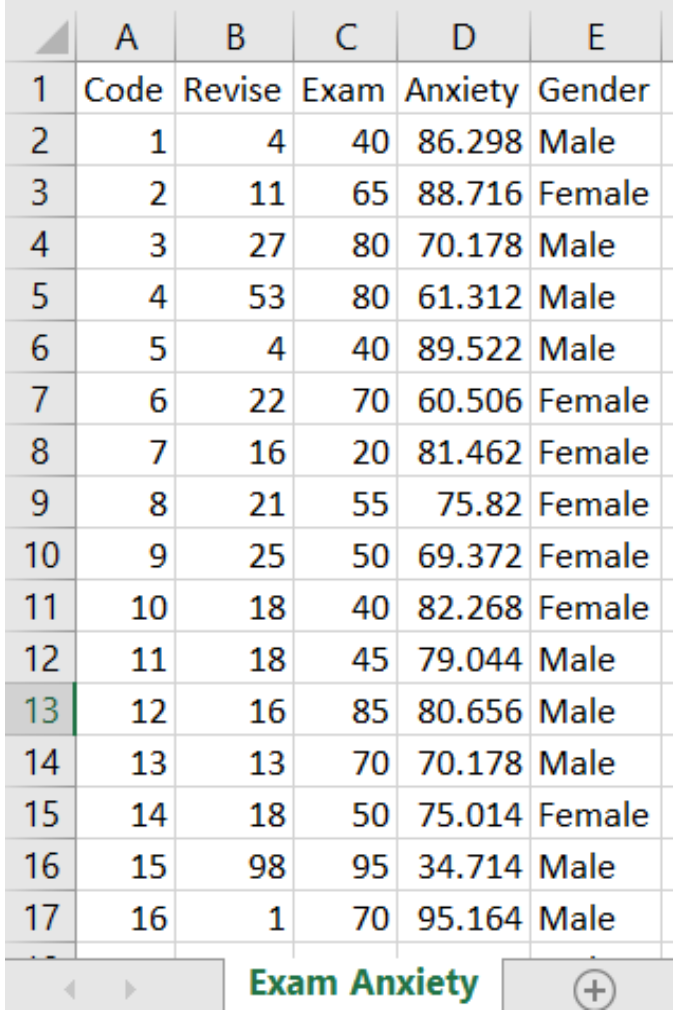




Exam Anxiety Data

Data File: Exam Anxiety.dat

- **Code**: a number indicating from which participant the scores came.
- **Revise**: the total hours spent revising.
- **Exam**: mark on the exam as a percentage.
- **Anxiety**: the score on the Exam Anxiety Questionnaire (EAQ).
- **Gender**: whether the participant was male or female (stored as strings of text).



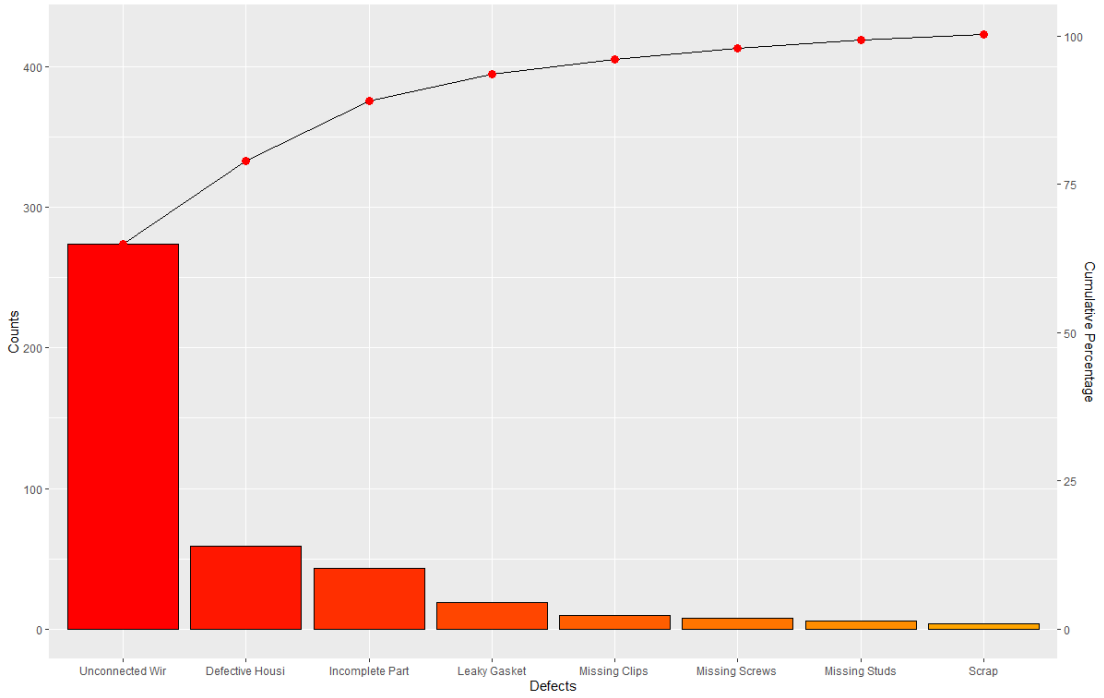
	A	B	C	D	E
1	Code	Revise	Exam	Anxiety	Gender
2	1	4	40	86.298	Male
3	2	11	65	88.716	Female
4	3	27	80	70.178	Male
5	4	53	80	61.312	Male
6	5	4	40	89.522	Male
7	6	22	70	60.506	Female
8	7	16	20	81.462	Female
9	8	21	55	75.82	Female
10	9	25	50	69.372	Female
11	10	18	40	82.268	Female
12	11	18	45	79.044	Male
13	12	16	85	80.656	Male
14	13	13	70	70.178	Male
15	14	18	50	75.014	Female
16	15	98	95	34.714	Male
17	16	1	70	95.164	Male

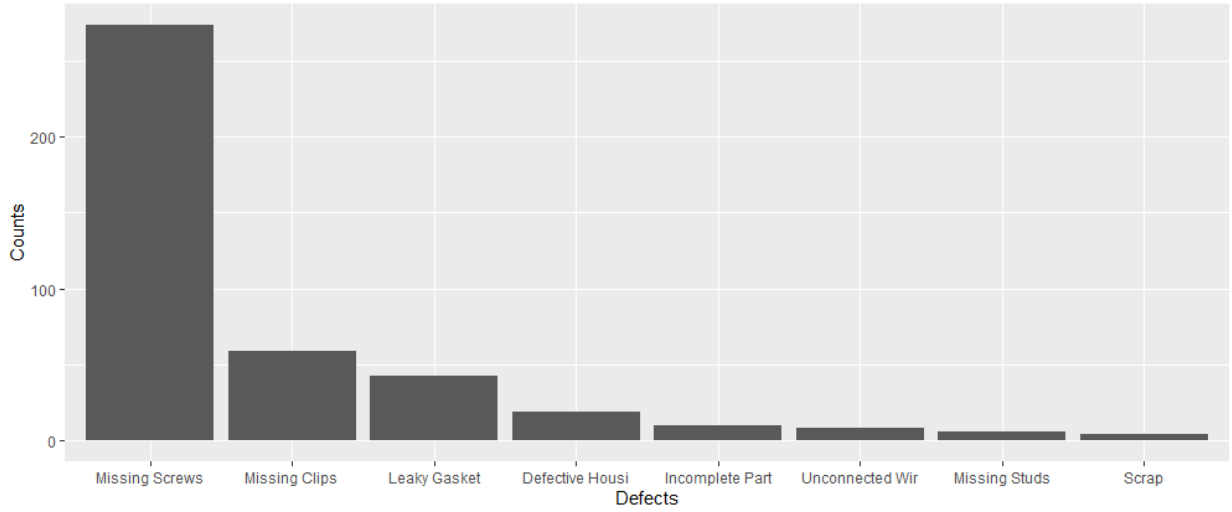
Exam Anxiety

Data Visualization for Different Data Stories

Data Visualization to Show Rankings

- Ordered lists or rankings are useful to quickly identify top or bottom performers.
  - Ordered Bar
  - Pareto Chart

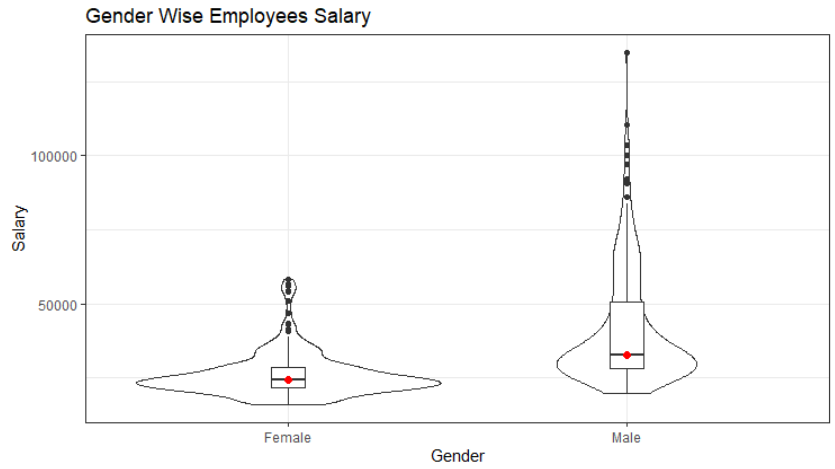


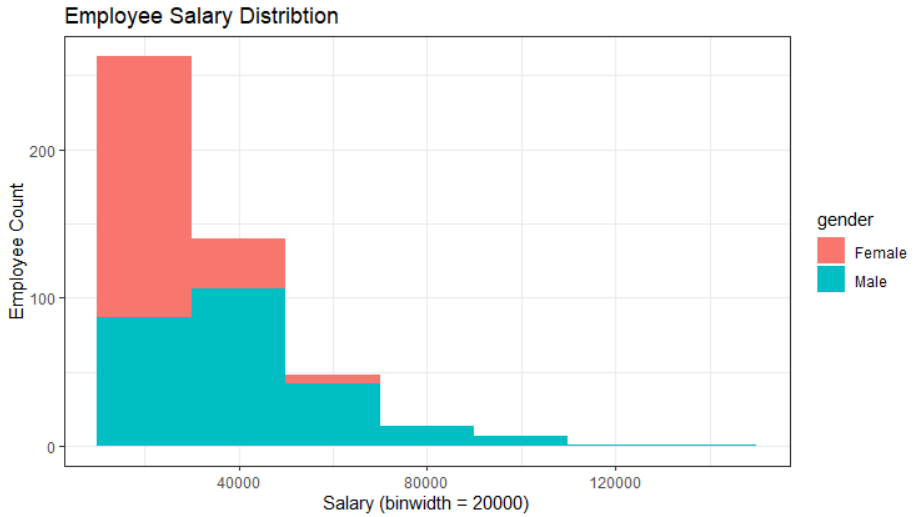


Data Visualization for Different Data Stories…

Data Visualization to Show Distributions

- Histogram
- Box plot
- Violin Plot
- Frequency Polygons

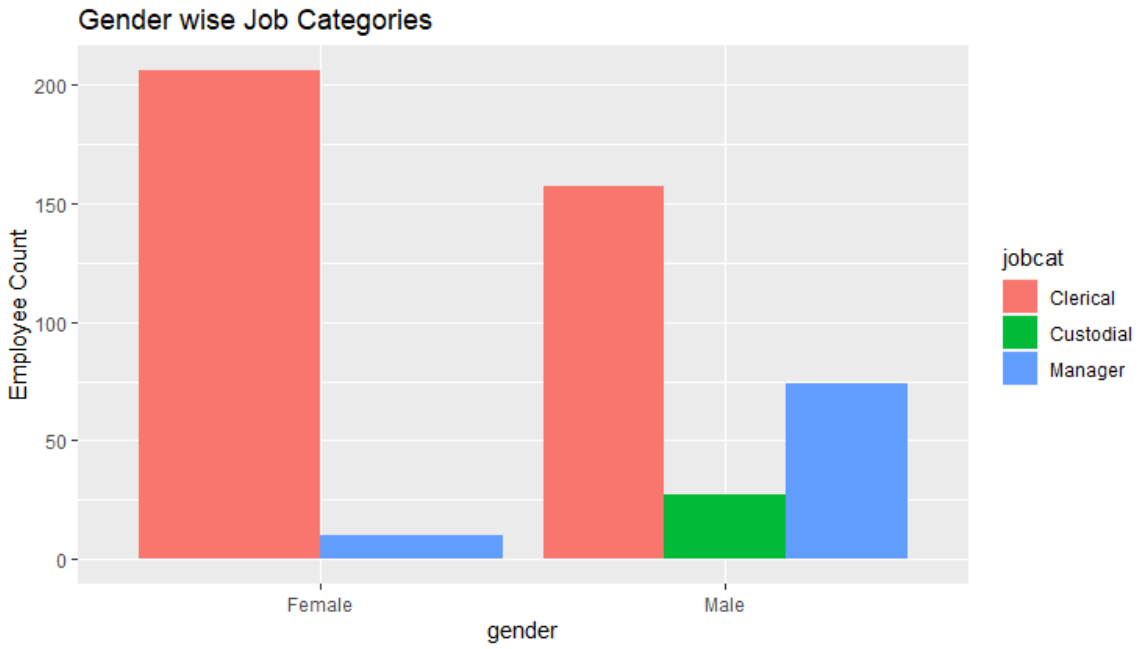




Data Visualization for Different Data Stories…

Data Visualization to Show Huge Magnitudes

- Set of charts to show size comparisons in data.
- These charts are good to show counted numbers rather than a value such as changing rate or percentage.
  - Simple Bar Chart
  - Multiple Bar Chart



Correlation & Regression Analysis Using R

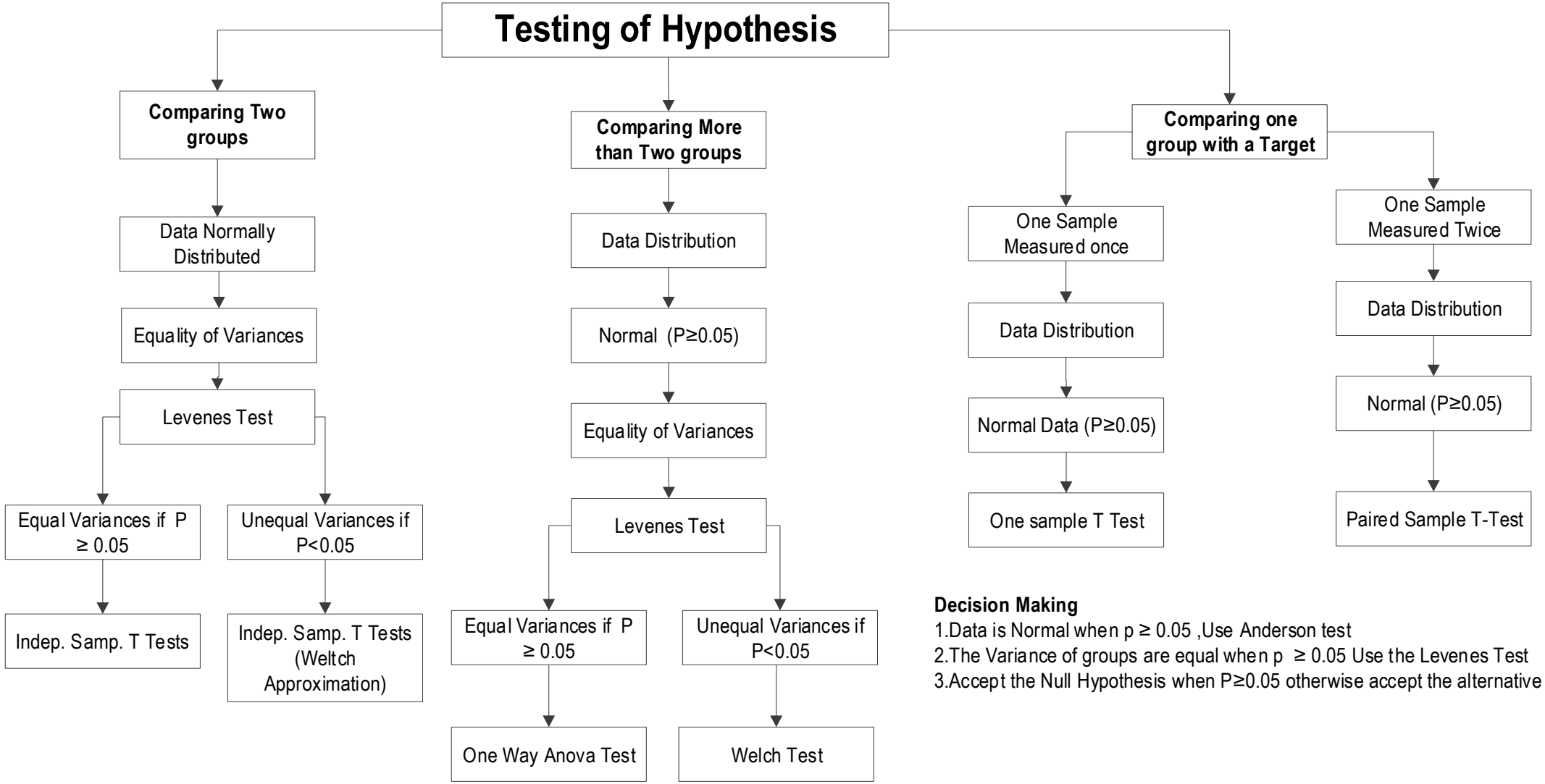
Correlation Analysis (Part-1 to 3)

- https://www.youtube.com/watch?v=ZLikejwTtIA
- https://www.youtube.com/watch?v=aEL-bUsOOSk
- https://www.youtube.com/watch?v=qmpWjtrQGdk&t=161s

Regression Analysis

- Simple Regression Analysis
  - https://www.youtube.com/watch?v=RH5NyPAdwuA&t=2619s
- Multiple Regression Analysis
  - https://www.youtube.com/watch?v=FAKS3vbve0Y
- Multiple Regression Analysis - Robust Regression: Bootstrapping Using R
  - https://www.youtube.com/watch?v=iCofhLpZc9o&t=171s
- Regression Analysis with Dummy Variables Using R
  - https://www.youtube.com/watch?v=mIp3o8AhOHM&t=2s

Parametric Testing of Hypothesis



Decision Making

1.Data is Normal when p ≥ 0.05 ,Use Anderson test
2.The Variance of groups are equal when p ≥ 0.05 Use the Levenes Test
3.Accept the Null Hypothesis when P≥0.05 otherwise accept the alternative hypothesis

https://www.youtube.com/watch?v=cIe6mvIgn0Q

Business Analytics Using Python

Data Manipulation & Analysis Using Python

- https://www.youtube.com/watch?v=FYvwgsKZkXo

Predictive Analytics

- Time Series Forecasting Using Python
  - https://www.youtube.com/watch?v=xSn9aEtPlDs

Perspective analytics – Operations Research / Optimization

- Linear & Integer Programming Using Python (Playlist)
  - https://www.youtube.com/watch?v=NNmRQEgupuo&list=PLW39o_Nls7NwOmVjYprSySpeHjexDHt3i

THANK YOU