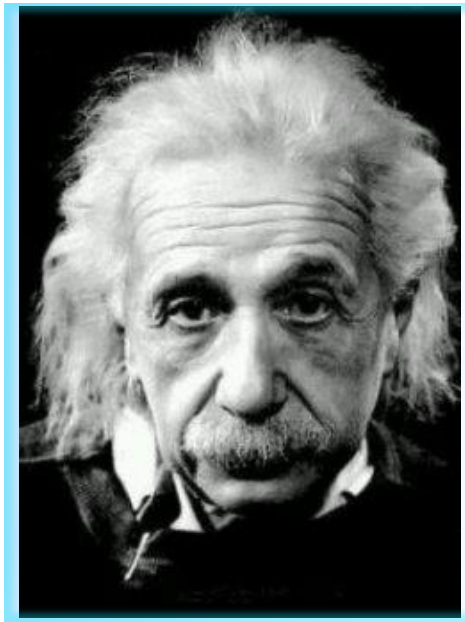
30 Hours Online Certificate Course On

"*Research & Data Analysis*"

- Organised By: **Accounting and Finance Lab, Department of Commerce, Ramanujan College, University of Delhi**
- Resource Person: **Dr. Arnav Kumar**
- Designation: Assistant Professor, Department of Management Studies, Ramanujan College
- Email: arnavkumardse@gmail.com
- Day & Date: **June 17, 2020.**

Let's Start with Three Thoughts…

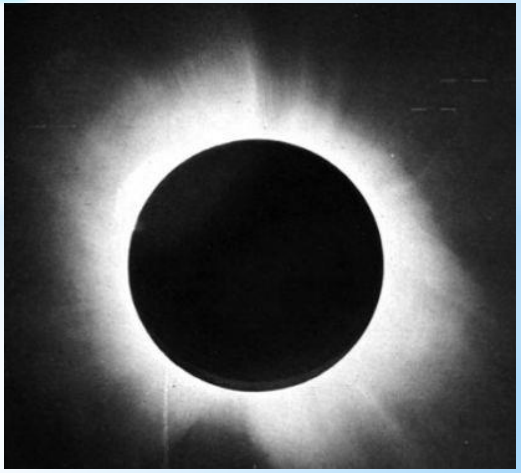
Thought #1:



If you can't explain it simply, you don't understand it well enough.

Albert Einstein

A great lesson for us!

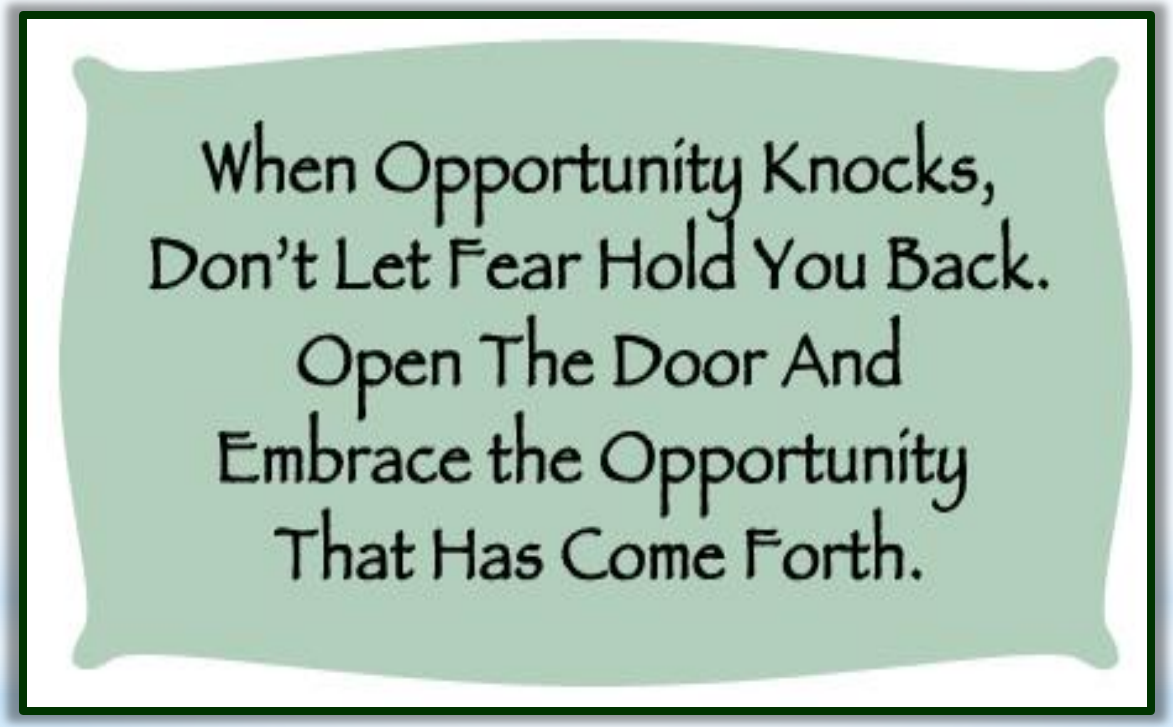


Thought #2:

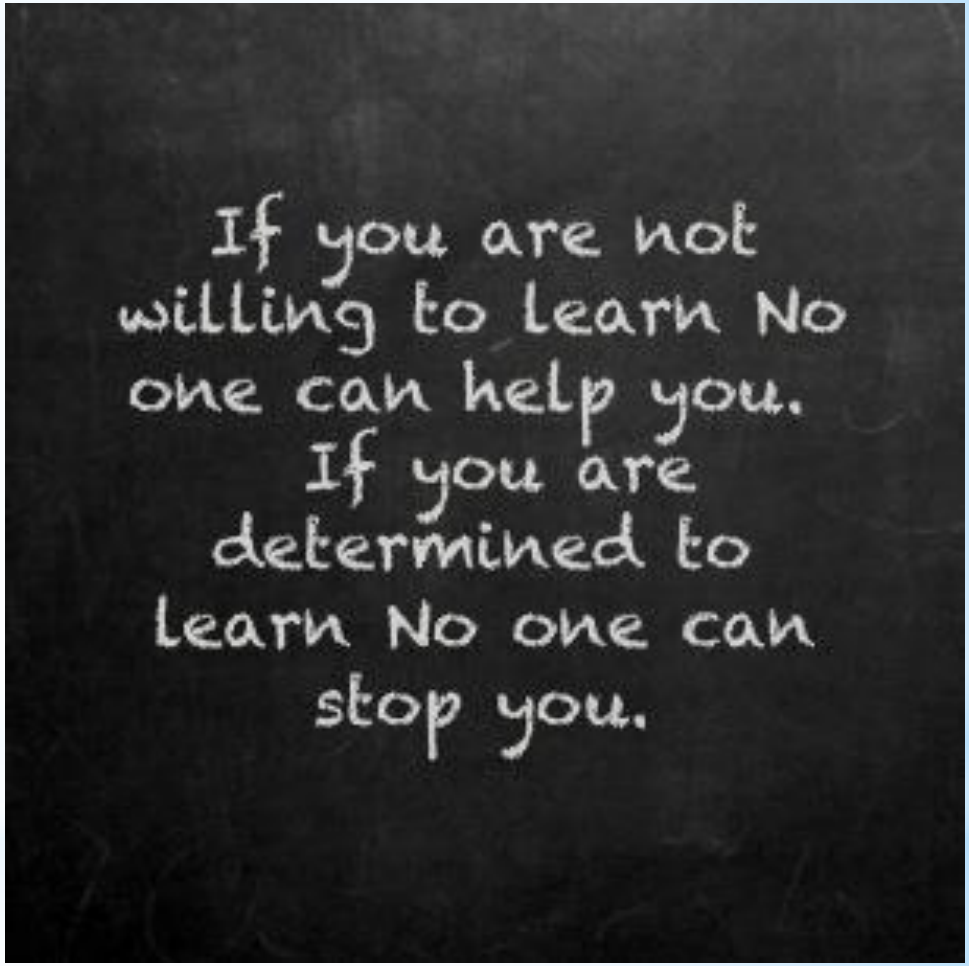
You are missing the excitement of life!


STATISTICS is a journey from data to WISDOM!

This Course is a GREAT OPPORTUNITY to learn!!!



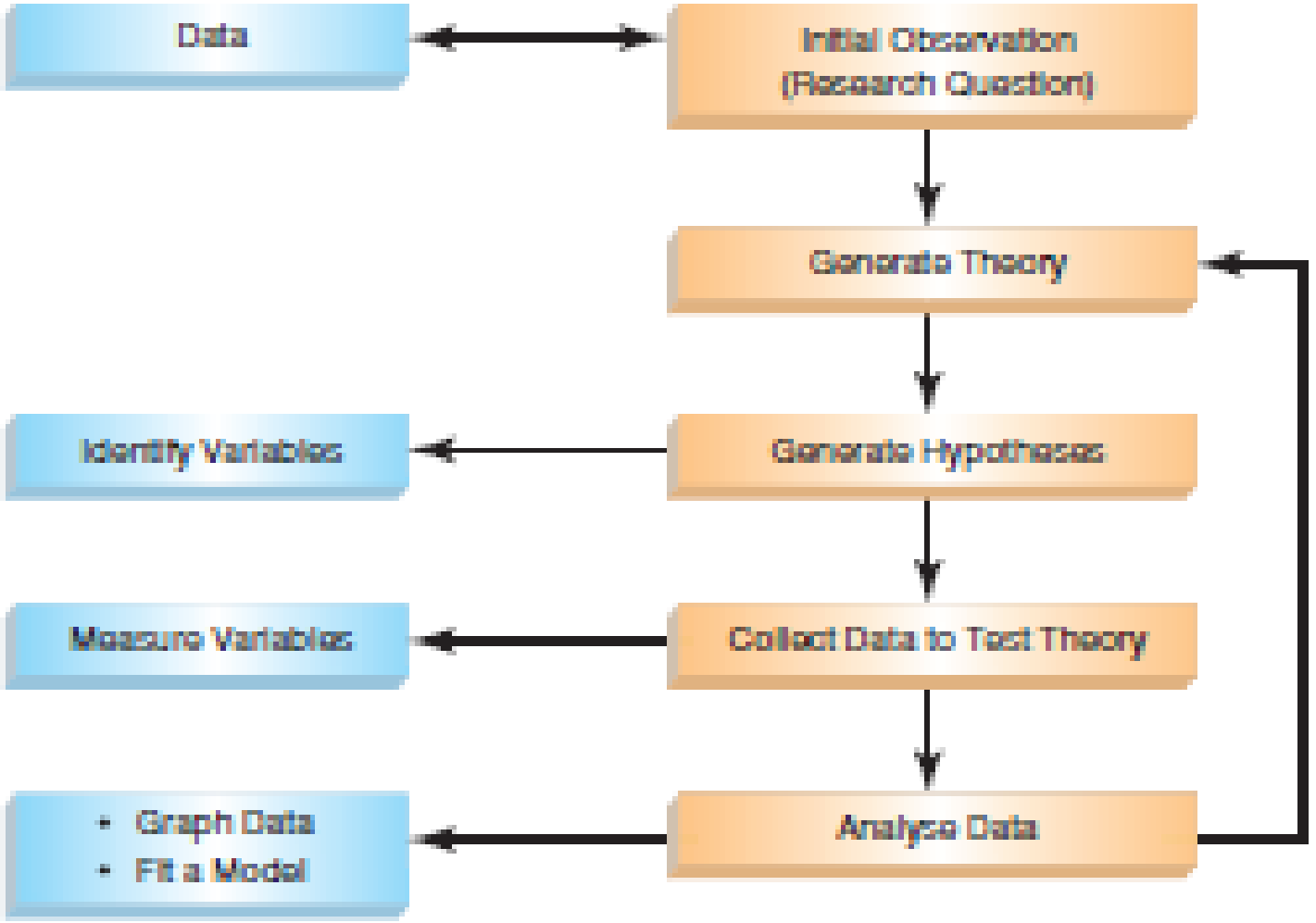
Thought #3: Please remember…



Choice is YOURS!!!

Introduction to Data Analysis

Research Process



Why to do an Empirical Research?

- Conducting empirical research is one of the best ways to get to **grips with technical material**, and find out what **practical difficulties** that we encounter when conducting research
- Conducting research gives you the opportunity to solve an **puzzle or unexplored topic or real life problem**
- Doing empirical work is usually **less risky** than trying to develop a **new Theory.**
- Your **report-writing** and **time-management** skills will improve
- A project will provide something to discuss in **Job interviews.**
- A dissertation may be a route into **M.Phil. or Ph.D.** research.

Examples of Applications of Data Analysis in Different Business Areas

S. No.	Sample Data	Possible Applications	Area
1	Current Customers Database	Identifying characteristics of High Value & Repeat Purchasers	Marketing
2	Prospective Customers Market Survey	Understanding Customer's Needs, Preferences	Marketing
3	Past and Current Accounting and Financial Data of Own firm & Rivals	Finding Significant factors affecting Profitability, Share Prices	Finance
4	Current and Previous Employees Database	Investigating reasons for High Employee Turnover or Poor Performance	Human Resource (HR)

Some Real Life Examples of Companies using Data Analysis

- Banks, such as Capital One, use data analysis (or analytics), to differentiate among customers based on credit risk, usage and other characteristics and then to match customer characteristics with appropriate product offerings.
- Harrah's, the gaming firm, uses analytics in its customer loyalty programs.
- E & J Gallo Winery quantitatively analyses and predicts the appeal of its wines.

Successful Data Analysis

Successful Data Analysis depends on:

- Data Availability & Quality;
- Skilled Analysts who understand the technologies (Software Packages and Statistical Methods); and
- Organizational Commitment to data-driven decision making.

Descriptive Statistics

Basic Concepts

- **Population** – The entire/complete/whole set of entities.
- **Sample** – A part/subset of Population.
- **Parameter -** Any measure related to the Population.
- **Statistics -** Any measure related to the Sample.
- **Frequency** – How many times a value of a variable occurs or is repeated.
- **Frequency Distribution** – List out all the values of a variable along with their frequencies.
- **Probability** – Chance of occurrence, Range [0 <= P(X) <=1].
- **Probability Distribution** - List out all the values of a variable along with their probabilities.

❖ **How we make Probability Distribution from Frequency Distribution?**

Measures of Central Tendency

- Mean
- Median
- Mode
- Quartiles
- Deciles
- Percentiles
- Positional Averages
- How many Quartiles, Deciles and Percentiles are their?
- Which is the best Measure & Why?
- Which is most affected by Extreme Values?

Measures of Variation or Dispersion

Deviations or Differences from Mean

Sum of Deviations

Sum of Squared Deviations

Mean Sum of Squared Deviations (Variance)

+ Square Root of Variance [Standard Deviation(SD)]

Coefficient of Variation (CV = SD/Mean)

Measures of Variation or Dispersion

- **Simplest Measure is Range** (Highest Value – Lowest Value).
- **Absolute Measures** (in same unit as the variables) and **Relative Measures** (unit free).
- **What is Unit of Variance, S.D. & C.V.?** – Variance (units squared), S.D. (units of the Variable) & C.V. (unit free – Relative measure).
- **What does Large S.D. & C.V. values imply?** – High Variation/Volatility/Dispersion or Deviation from the Central Value.
- **What if S.D. is O?** – No Deviation from Mean, i.e., All observations are exactly same!!!

Why “IBM SPSS Statistics” for Data Analysis?

- IBM SPSS Statistics is an integrated family of products that addresses the entire analytical process, from planning to data collection to analysis, and reporting.
- Helps in quickly gaining understanding and insights from datasets in any format using advanced statistical procedures.
- Significantly increases the Analytical Power, and Flexibility (Choice of Methods etc.).
- Ensures high accuracy.
- Easily communicate results via presentation ready output and high quality graphs.

Basic Data Analysis (Frequency & Descriptive Statistics) using IBM SPSS Statistics

- **Variables:**
  - ✓ Grade & Percent;
  - ✓ Gender & Ethnicity.
- **Frequency Distribution:**

Analyze > Descriptive Statistics > Frequencies

- **Descriptive Statistics:**

Analyze > Descriptive Statistics > Frequencies > Statistics

Graphical Analysis using IBM SPSS Statistics

What makes a Good Graph?

- Tufte (2001) points out that graphs should:
  - Show the data.
  - Induce the reader to think about the data being presented (rather than some other aspect of the graph, like how pink it is).
  - Avoid distorting the data.
  - Present many numbers with minimum ink.
  - Make large data sets coherent.
  - Encourage the reader to compare different pieces of data.
  - Reveal the underlying message of the data.

Graphs - Basics

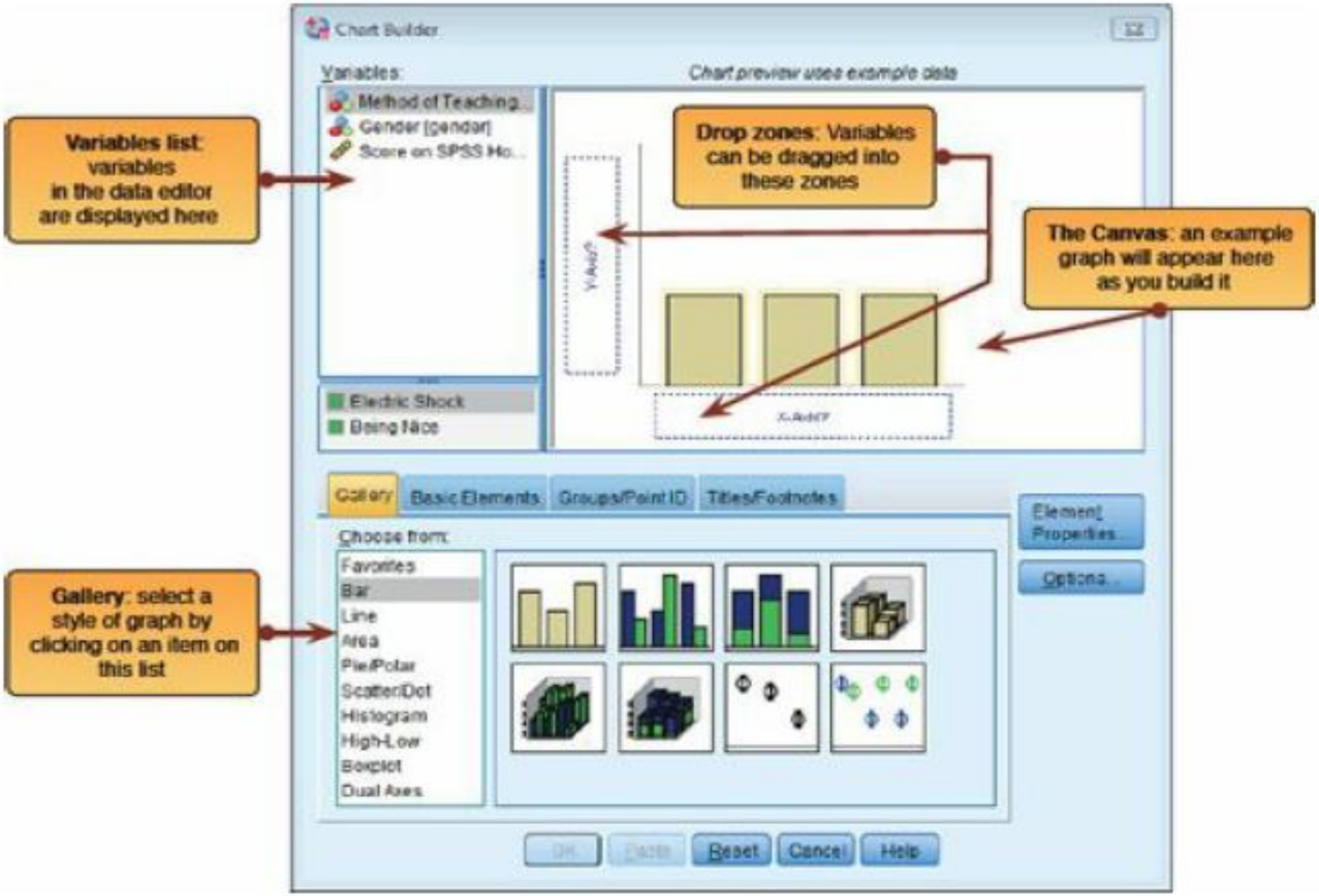
- Vertical Axis of Graph = Y-axis (or Ordinate).
- Horizontal Axis of Graph = X-axis (or Abscissa).
- Some Useful Graphing Tips:

✓ Don't create false impressions of what the data actually show (likewise, don't hide effects) by scaling the y-axis in some weird way.

✓ Abolish Chart-junk: Don't use patterns, 3-D effects, shadows, pictures/photos of anything else.

✓ Avoid excess ink: If you don't need the axes/lines, then get rid of them.

The SPSS Chart Builder



Common Types of Graphs

1. Line Charts
2. Pie Chart
3. Bar Charts
4. Histograms
5. Box Plot – It is commonly used to Check for Outliers.
6. Scatter Plot – It is used to study relationship between two Variables. We will learn more about it in “Correlation Analysis”.

1. Line Charts

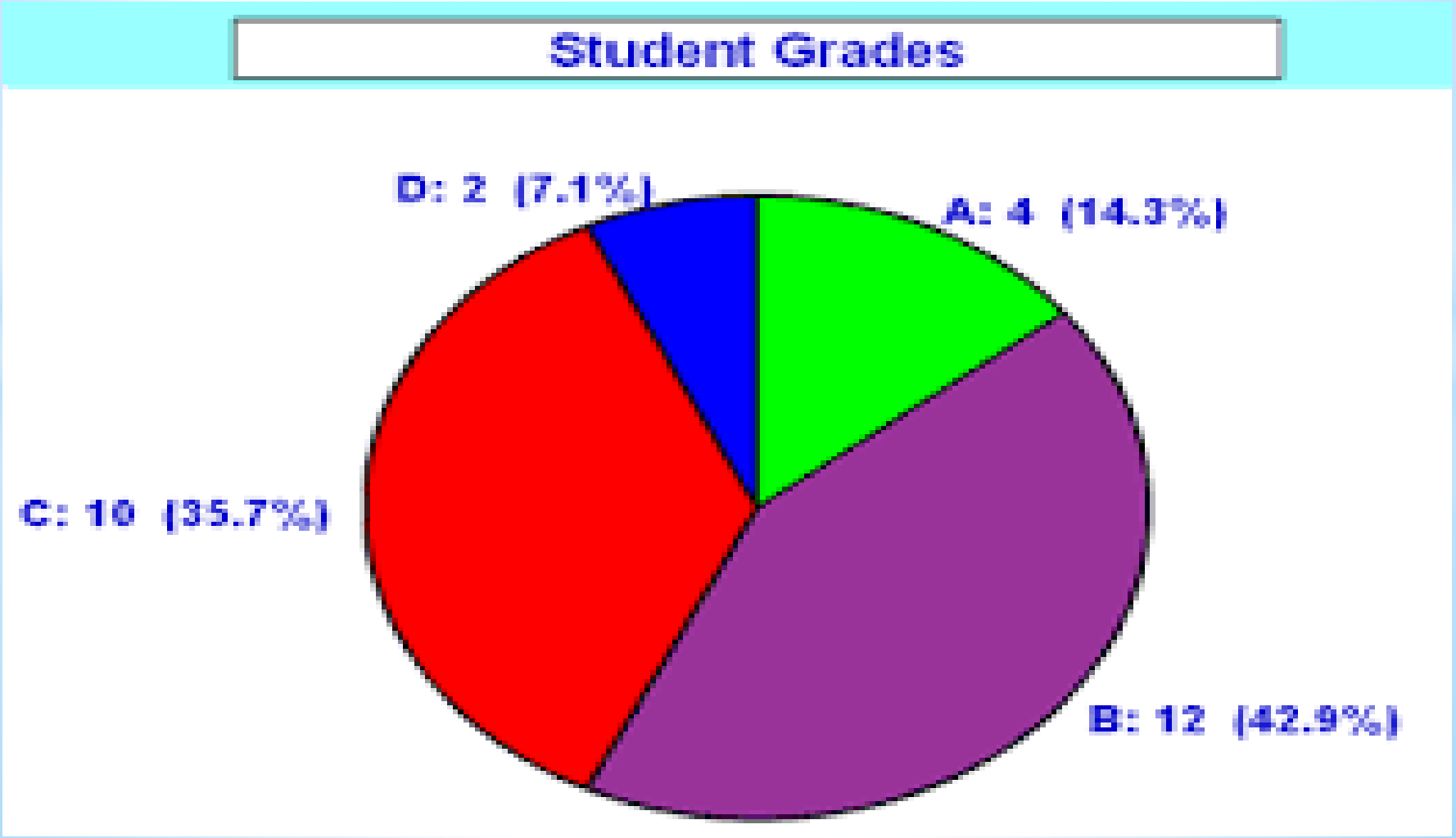
- We can use a Line Chart to summarize categorical variables, in which case it is similar to a bar chart.
- Line charts are also useful for Time Series/Case Wise/Category Wise analysis data.
- How to create a Simple Line Chart:
- ✓ In Chart Builder, click Gallery tab and select Line.
- ✓ Drag the Simple Line icon onto the canvas.
- ✓ Drag a Date/Case/Categorical Variable (*ID*/*Gender*) to the X-axis drop zone.
- ✓ Drag a Scale Variable (*GPA*/*Percentage*) to the Y-axis drop zone. This is the variable whose values were recorded over Categories/Cases/Time.

1. Line Charts

- We can also simultaneously draw multiple line charts.
- How to create a Multiple Line Chart:
- ✓ In Chart Builder, click Gallery tab and select Line.
- ✓ Drag the Multiple Line icon onto the canvas.
- ✓ Drag a Date/Case/Categorical Variable (*Gender/ Ethnicity*) to the X-axis drop zone.
- ✓ Drag a Scale Variable (*Quiz 1, Quiz 2, Quiz 3, Quiz 4 & Quiz 5*) to the Y-axis drop zone.

2. Pie Charts

- Each slice of a pie chart displays the proportion of parts to a whole.



2. Pie Charts

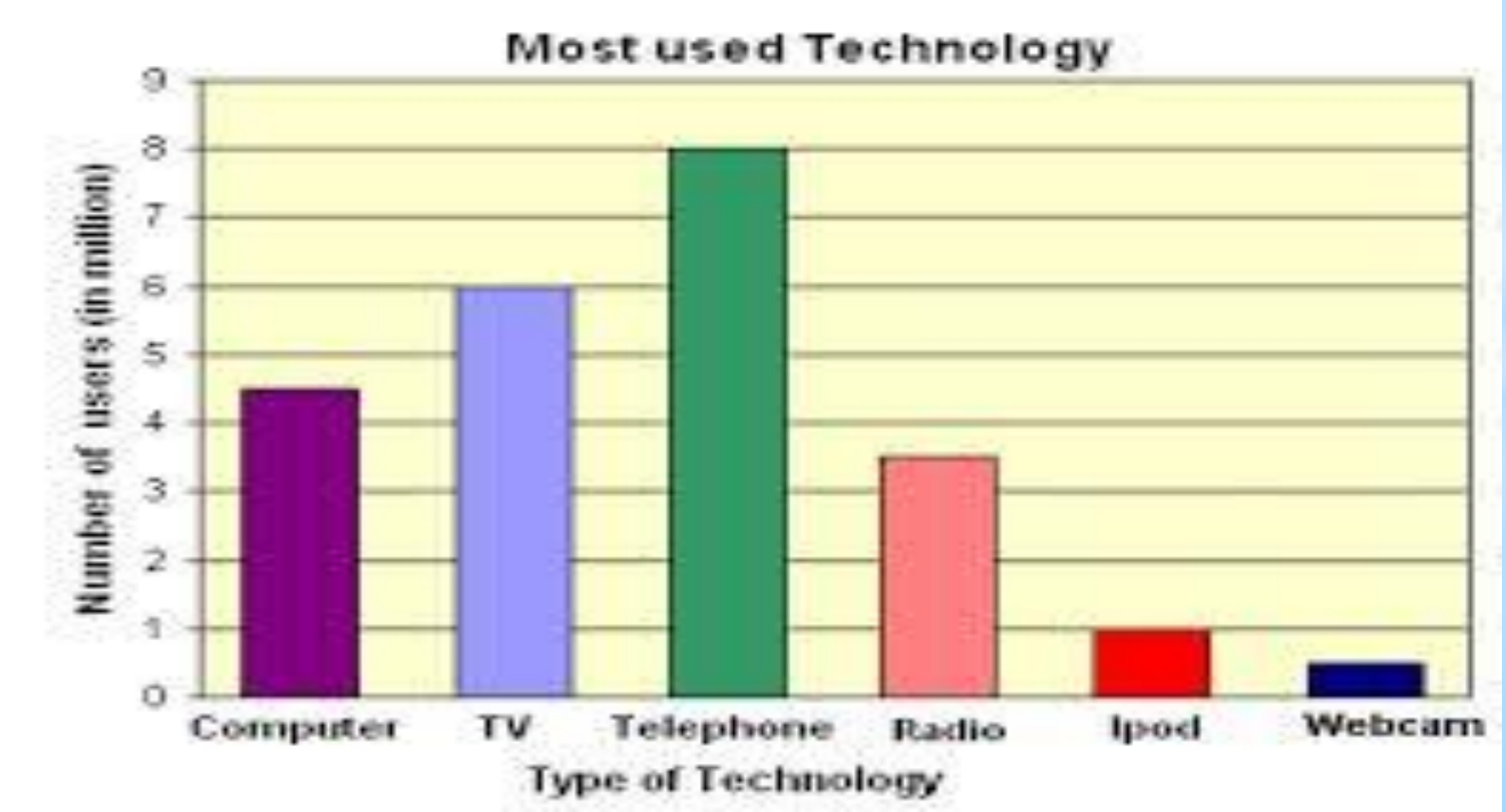
- A pie chart is useful for comparing proportions. Ex: To demonstrate that a greater proportion of women are enrolled in a certain course.
- How to create a Simple Pie Chart:
- ✓ In Chart Builder, click Gallery & select Pie/Polar.
- ✓ Drag the Pie Chart icon onto the canvas.
- ✓ Drag a categorical (nominal or ordinal) variable to the Slice By drop zone.
- ✓ The number of categories in this variable determine the number of slices in the pie chart.

2. Pie Charts

- ✓ Specify a statistic (Count/Sum/Percentage) in the Element Properties dialog box.
- ✓ For pie charts, you typically want a Count-based Statistic or Sum (for Nominal or Ordinal variable).
- ✓ The result of the statistic determines the relative size of each slice out of the total pie.
- ✓ To see Values/Data Labels: Double left click graph > Chart Editor > Elements > Show Data Labels.
- ✓ Create a Pie Chart of *Grades* (Categorical Variable) based on *Count & Percentages* (Statistics).
- ✓ In Data Label, simultaneously Show *Grade*, *Count & Percentage*.

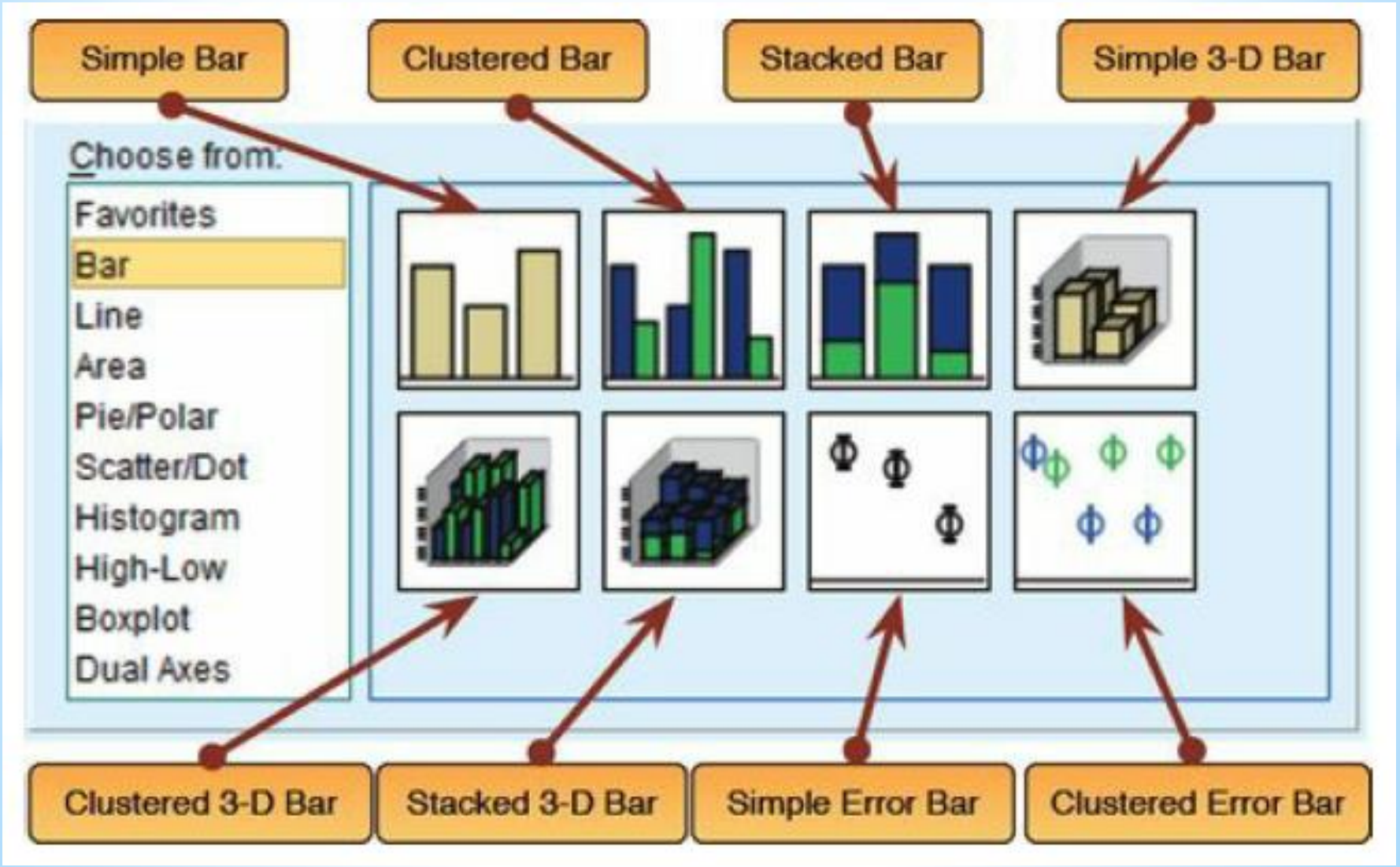
3. Bar Charts

- Displays the count for each distinct value or category as a separate bar.



3. Bar Charts

- Bar Charts are the usual way to display means.



3. Bar Charts

- Simple Bar: To see the Means of Values of Scale Variables across different groups/categories of Categorical Variable.
- Clustered Bar: If we have a second Grouping or Categorical Variable, we could produce a simple bar chart (as above) but with bars produced in different colours for levels of the second grouping variable.
- Stacked Bar: Same as the clustered bar, except that the different coloured bars are stacked on top of each other rather than placed side by side.

3. Bar Charts

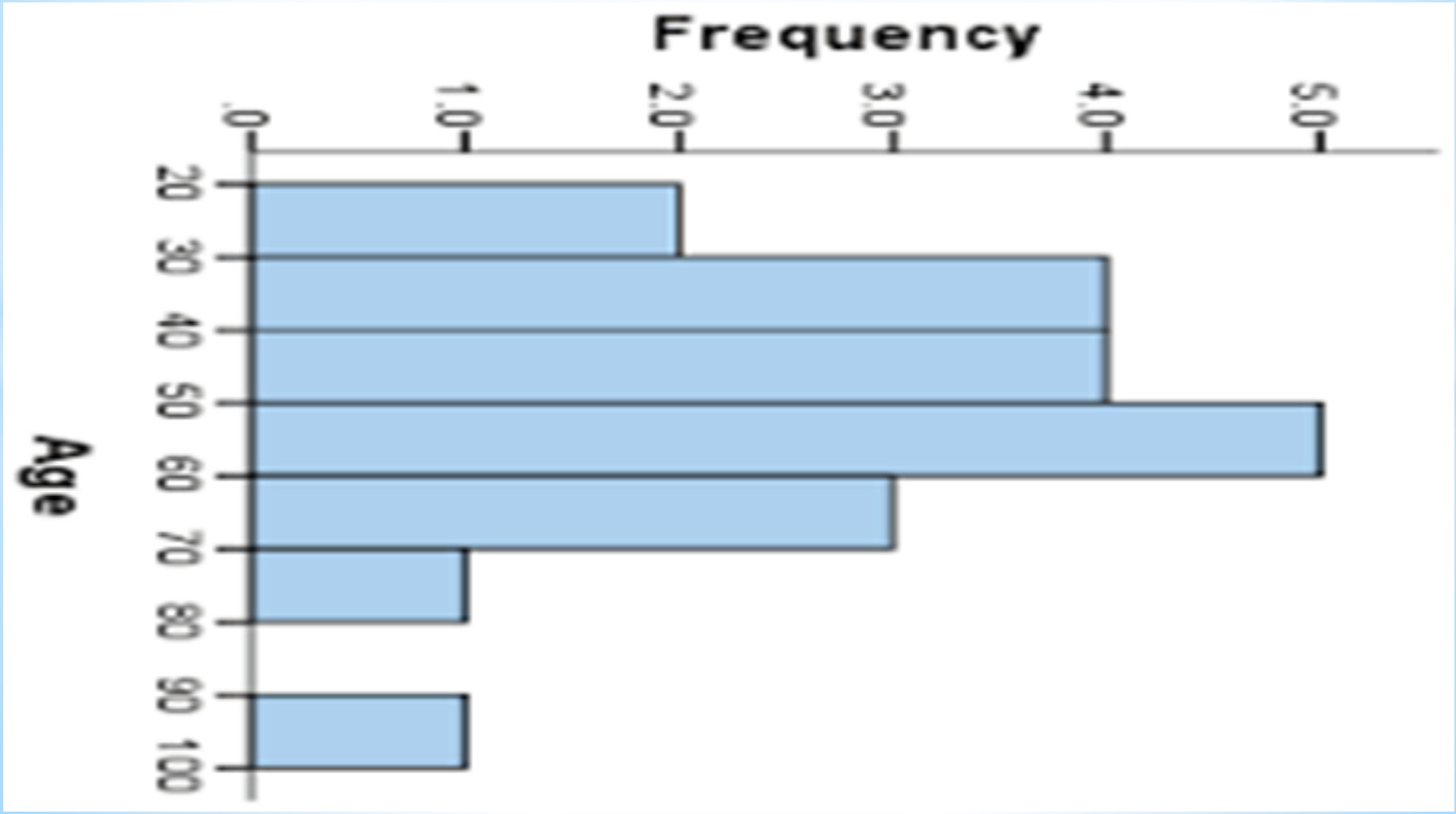
- How to create a Simple Bar Chart:
- ✓ In Chart Builder, click Gallery tab and select Bar.
- ✓ Drag the Simple Bar icon onto the canvas.
- ✓ Drag a categorical (nominal or ordinal) variable (*Grade*) to the x-axis drop zone. You can use a scale variable, but the results will be useful in only a few special cases. A bar chart looks best with a limited number of distinct values.
- ✓ Specify a Statistic (*Cumulative Count*/*Cumulative Percentage*) in the Element Properties dialog box. The result of any statistic determines the height of the bars.

3. Bar Charts

- How to create a Clustered Bar Chart:
- ✓ In Chart Builder, click Gallery tab and select Bar.
- ✓ Drag the Clustered Bar icon onto the canvas.
- ✓ Drag a categorical (nominal or ordinal) variable (*Grade*) to the X-axis drop zone.
- ✓ Drag & drop the second Categorical Variable (*Gender*) to the "Cluster on X: Set Color" box.
- ✓ Specify a Statistic (*Count*/*Percentage*) in the Element Properties dialog box.
- ✓ With all the same settings, just drag the "Stacked Bar" icon on the canvas to get a Stacked Bar Chart.

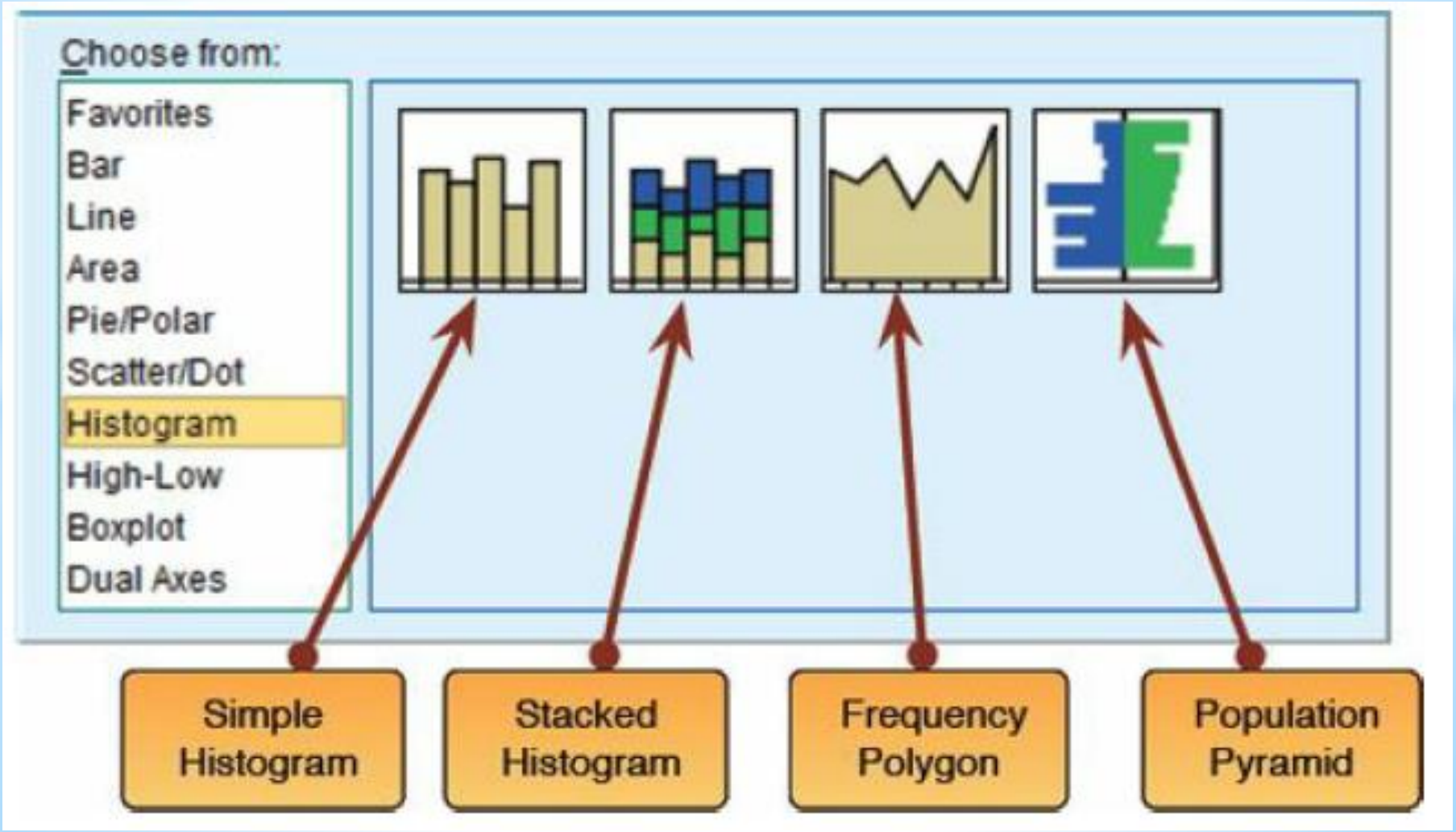
4. Histograms

- Also has bars, but along an equal interval scale.
- Height is Frequency of values.



4. Histograms

- Histograms are useful for showing the distribution or shape of data of a single scale variable.



4. Histograms

- Simple Histogram: Use this option to see the frequencies of scores for a single variable.
- Stacked Histogram: If you had a grouping variable you could produce a histogram in which each bar is split by group. This is a good way to compare the relative frequency of values of the Scale variable across groups of the Categorical Variable.
- Frequency Polygon: This option displays the same data as the simple histogram, except that it uses a line instead of bars to show the frequency, and the area below the line is shaded.

4. Histograms

- Population Pyramid:
- Like a stacked histogram, this shows the relative frequency of scores in two populations.
- It plots the variable on the vertical axis and the frequencies for each population on the horizontal: the populations appear back to back on the graph.
- If the bars either side of the dividing line are equally long then the distributions have equal frequencies.

4. Histograms

- How to create a Simple Histogram:
- ✓ In Chart Builder, click Gallery and select Histogram.
- ✓ Drag the Simple Histogram icon onto the canvas.
- ✓ Drag a scale variable (*GPA*) to the X-axis drop zone.

- How to create a Stacked Histogram:
- ✓ In Chart Builder, click Gallery and select Histogram.
- ✓ Drag the Stacked Histogram icon onto the canvas.
- ✓ Drag a scale variable (*GPA*) to the X-axis drop zone.
- ✓ Drag a Categorical Variable (*Gender*) to the "Stack: Set Color" box.

4. Histograms

- How to create a Frequency Polygon:
- ✓ In Chart Builder, click Gallery and select Histogram.
- ✓ Drag the Frequency Polygon icon onto the canvas.
- ✓ Drag a scale variable (*Percent*) to the X-axis drop zone.

4. Histograms

- How to create a Population Pyramid:
- ✓ In Chart Builder, click Gallery and select Histogram.
- ✓ Drag the Population Pyramid icon onto the canvas.
- ✓ Drag a Scale Variable (*Percentage*) to the Distribution Variable drop zone.
- ✓ Drag a Categorical Variable (*Gender*) to the Split Variable drop zone. Although it is possible to use a split variable with many categories, it is recommend to use a variable that has only two categories.
- ✓ The split variable acts as a paneling variable in that it creates multiple graphs.

The end!
Thank you for viewing and listening!