

Processing and Analysis of Data

Prof. Dr. Ayad Mohammed Fadhil
(*Ph.D. Eng. Remote Sensing & GIS*)

Petroleum & Mining Eng. Dept., Faculty of
Engineering,
Tishk International University (TIU)
(ayad.alquraishi@tiu.edu.iq)
2023/2024

PTR 411_Lect. 6

Outlines:

- The role of data processing and analysis in research.
- Importance of data in the fields of petroleum and mining engineering.



Objectives:

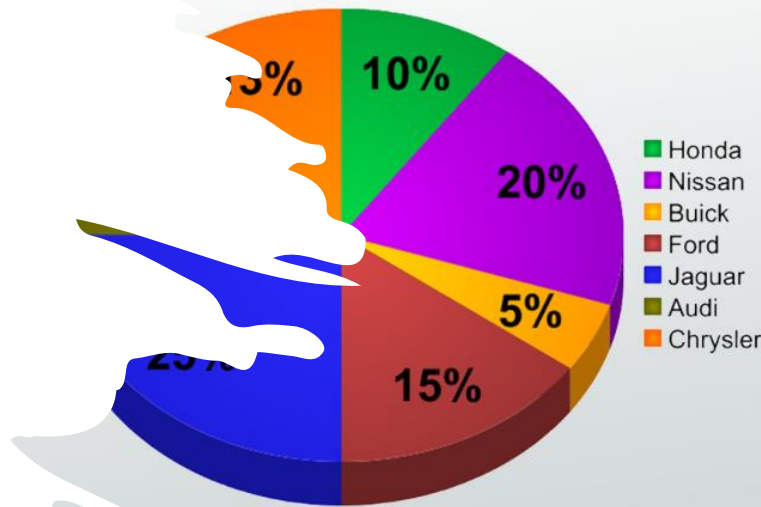
- Understand the data processing and analysis stages.
- Learn about data types and sources.
- Familiarize with data cleaning and preparation techniques.
- Explore data analysis methods.



Types of Data:

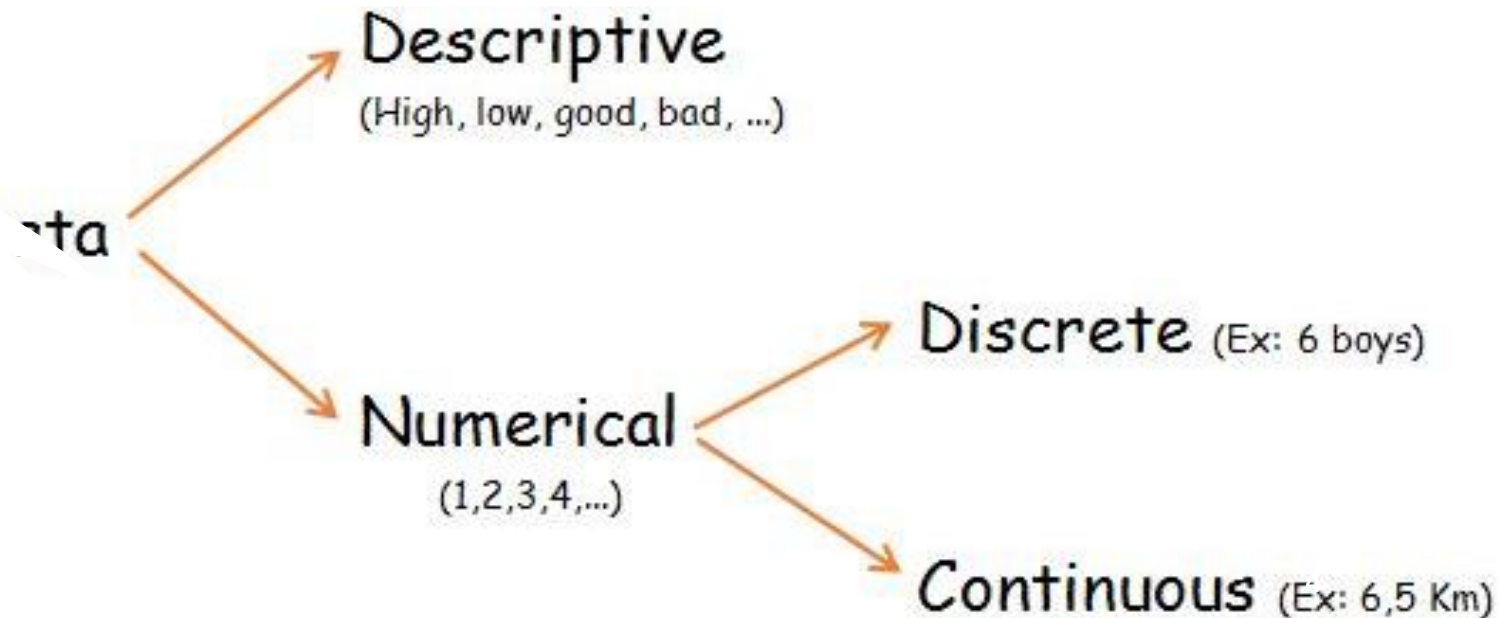
- Categorical data.
- Numerical data (continuous and discrete).
- Time-series data.
- Spatial data.
- Multivariate data.

Categorical Data



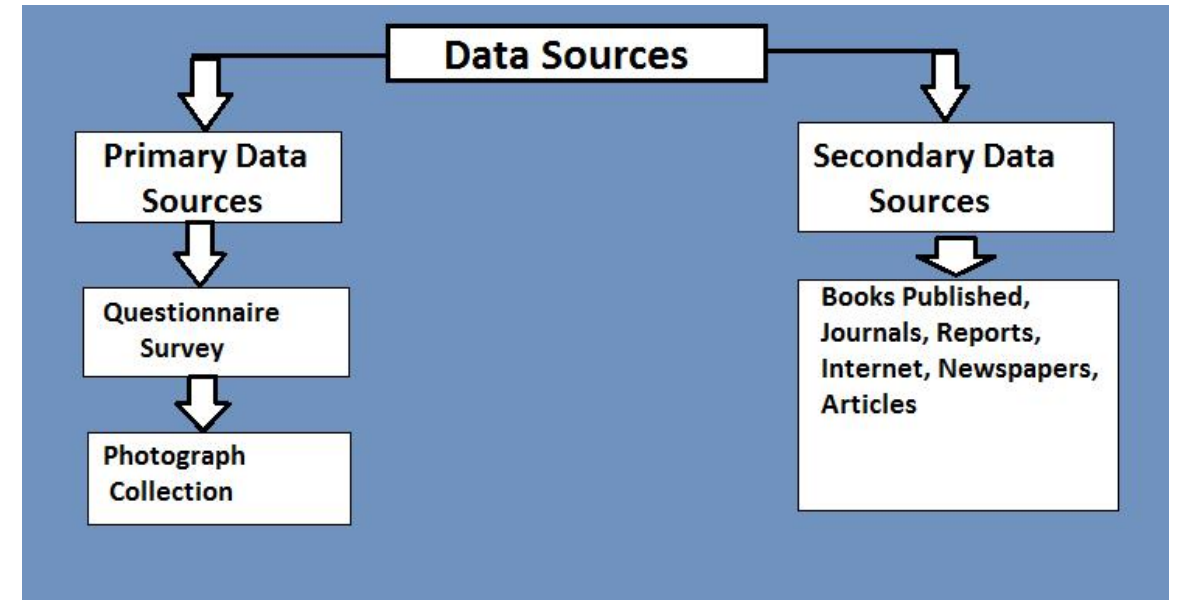
when numbers are collected in groups or categories

© Study.com



Sources of Data:

- Primary data.
- Secondary data.
- Public datasets.
- Surveys.
- Experiments.



Data Processing:

- Data collection.
- Data validation.
- Data cleaning:
- Handling missing values.
- Outlier detection and treatment.
- Data transformation.

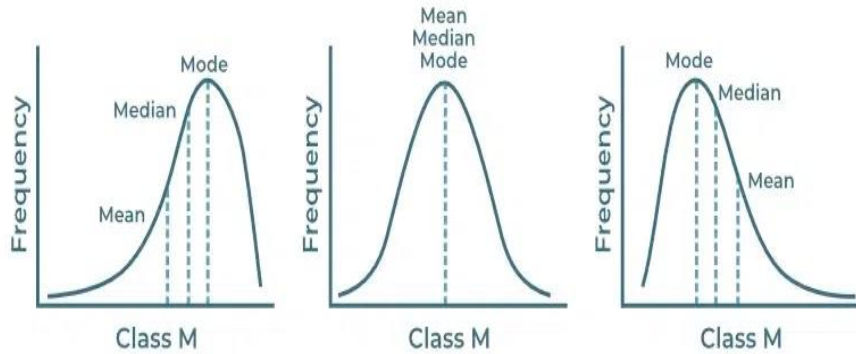


Data Analysis

Descriptive Statistics

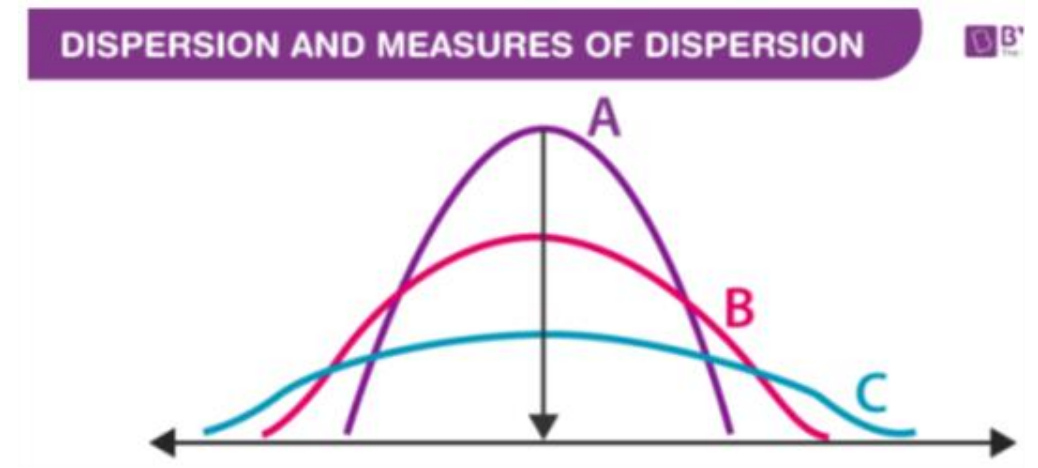
1. Measures of central tendency.

- It represents the single value of the entire distribution or a dataset.
- It aims to provide an accurate description of all the data in the distribution.
- The central tendency of the dataset can be found out using the three important measures namely mean, median and mode.



Measures of dispersion.

- Dispersion is the state of getting dispersed or spread.
- It means the extent to which numerical data will likely vary about an average value.
- In other words, dispersion helps to understand the data distribution.



Data visualization:

1. Histograms.
2. Box plots.
3. Scatter plots.



- Correlation and regression analysis.
- Time-series analysis.
- Spatial data analysis.

- **Software Tools for Data Analysis:**

- Introduction to software tools (e.g., Python, R, Excel, etc.).
- Data import and manipulation.
- Basic data analysis in the chosen software.

• **Case Studies:**

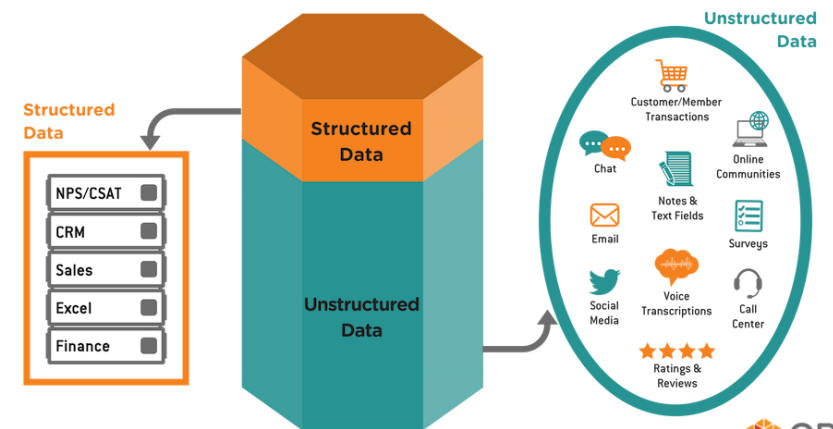
- Real-world examples of data processing and analysis in petroleum and mining engineering.
- How data analysis informs decision-making in these industries.
- considerations.
 - Regulatory compliance.

Challenges in Data Analysis for Petroleum and Mining Engineering:

- Dealing with large datasets.
- Handling unstructured data.
- Safety and ethical



What's Hiding in Your Unstructured Data?



Source: Graphic adapted from January 2018 CXPA Presentation "The Why Behind the What," Jim Kitterman

Software Tools for Data Analysis:

- Introduction to software tools (e.g., Python, R, Excel, etc.).
- Data import and manipulation.
- Basic data analysis in the chosen software.



Case Studies:

- Real-world examples of data processing and analysis in petroleum and mining engineering.
- How data analysis informs decision-making in these industries.



Challenges in Data Analysis for Petroleum and Mining Engineering:

- Dealing with large datasets.
- Handling unstructured data.
- Safety and ethical considerations.
- Regulatory compliance.



Data Collection:

- Collecting data is the foundation of any research project.
- Importance of careful planning and design in data collection.
- Data collection methods, such as surveys, experiments, field observations, and remote sensing.



Data Validation:

- Ensuring the accuracy and reliability of collected data.
- Verification of data sources.
- Cross-checking data with field measurements or known standards.



Data Cleaning:

- Handling Missing Values:
 - Methods for imputing missing data.
- Outlier Detection and Treatment:
 - Identifying outliers that may skew the analysis.
 - Strategies for handling outliers (e.g., removal or transformation).



Descriptive Statistics:

- Measures of Central Tendency:
 - Mean, median, and mode.
- Measures of Dispersion:
 - Variance, standard deviation, and range.
- Interpreting descriptive statistics in the context of your data.



Data Visualization:

- Importance of visualization in data analysis.
- Creating effective visualizations using appropriate tools.
- Interpretation of common plots and charts.



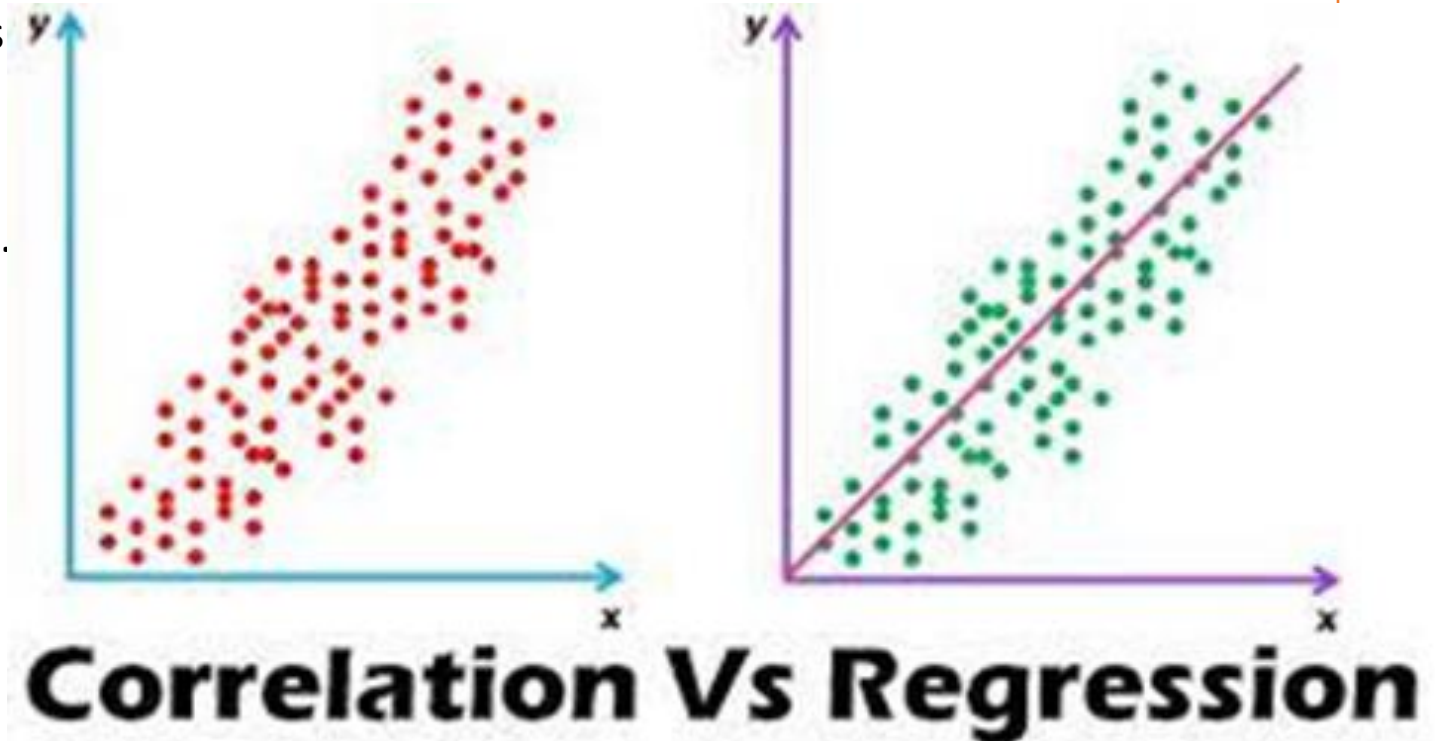
Inferential Statistics:

- Hypothesis Testing:
- Null and alternative hypotheses.
- Significance level and p-values.
- Types of statistical tests (t-tests, ANOVA, chi-square, etc.).
- Confidence Intervals:
- Calculating confidence intervals.
- Interpreting confidence levels.
- Sample size determination and power analysis.



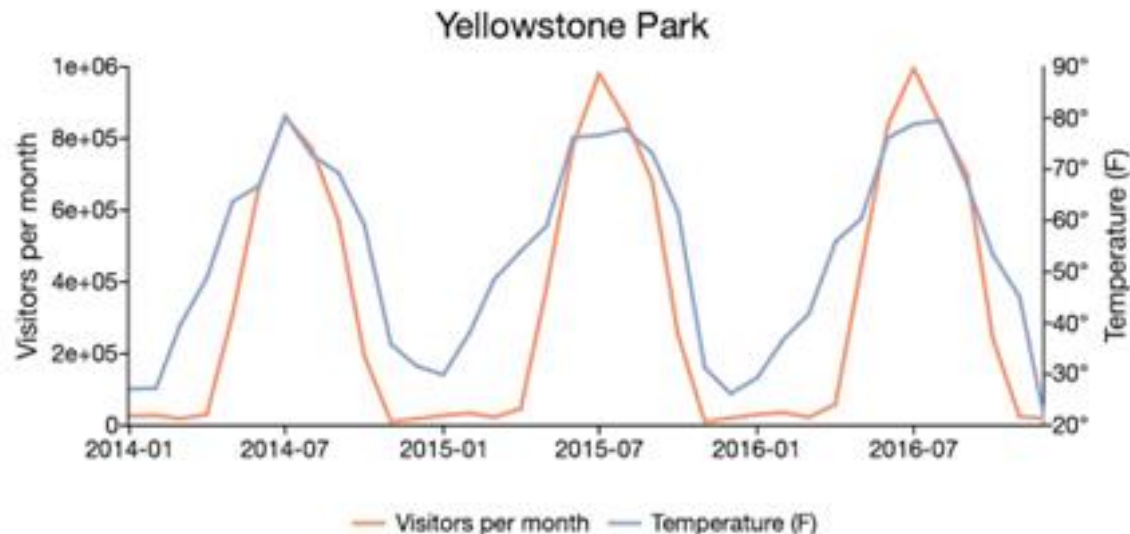
Correlation and Regression Analysis:

- Measuring relationships between variables.
- Interpretation of correlation coefficients
- Simple and multiple linear regression.
- Regression assumptions and diagnostics.



Time-Series Analysis:

- Handling time-series data in petroleum and mining engineering.
- Time-series components (trend, seasonality, noise).
- Time-series modeling (ARIMA, exponential smoothing).



Spatial Data Analysis:

- Introduction to geographic information systems (GIS) in geospatial data analysis.
- Spatial data visualization and analysis techniques.
- Applications in geological mapping and resource exploration.



Software Tools for Data Analysis:

- Deeper dive into data analysis software (e.g., Python libraries like NumPy, Pandas, Matplotlib, and Scikit-Learn).
- Hands-on exercises for data manipulation and analysis using software.



Challenges in Data Analysis for Petroleum and Mining Engineering:

- Handling large volumes of data from sensors and geological surveys.
- Unstructured data sources like well logs and rock samples.
- Safety considerations in data collection and analysis.
- Compliance with industry regulations and environmental standards.





















Questions???