

Data Visualization in Data Science

Understanding, Communicating, and Exploring Data

CS202 Lecture 27

What is Data Visualization?

- **Graphical Representation** of Data and Insights
- Identifies **Patterns, Trends, and Anomalies**
- Converts Raw Data into **Visual Understanding**



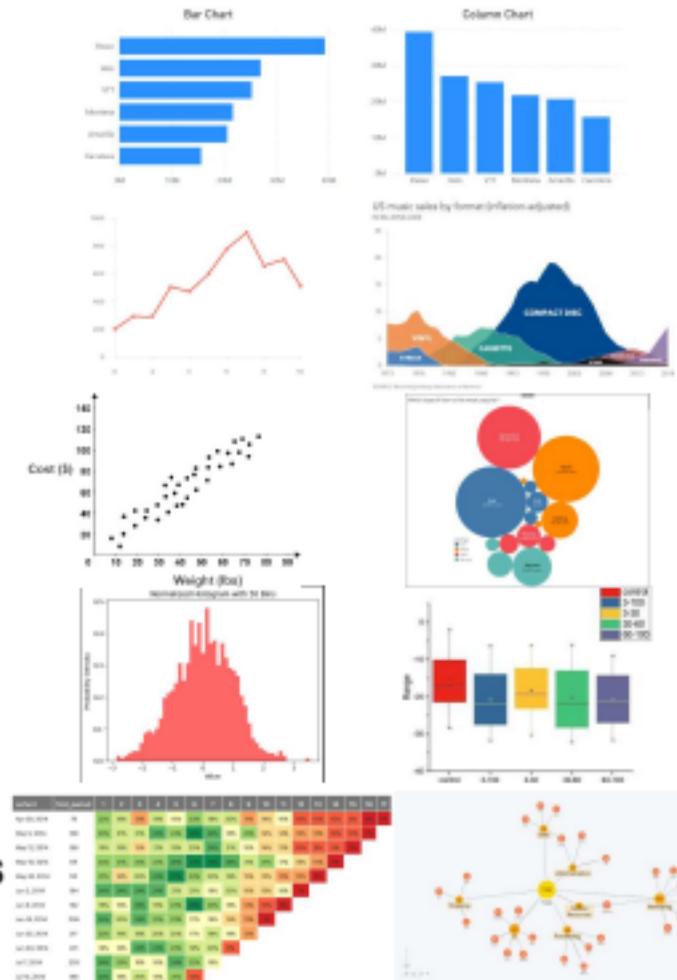
Why Data Visualization Matters

- Simplifies Complex **Datasets**
- Aids Faster **Decision-Making**
- Reveals Hidden **Insights**
- Communicates **Results** Effectively

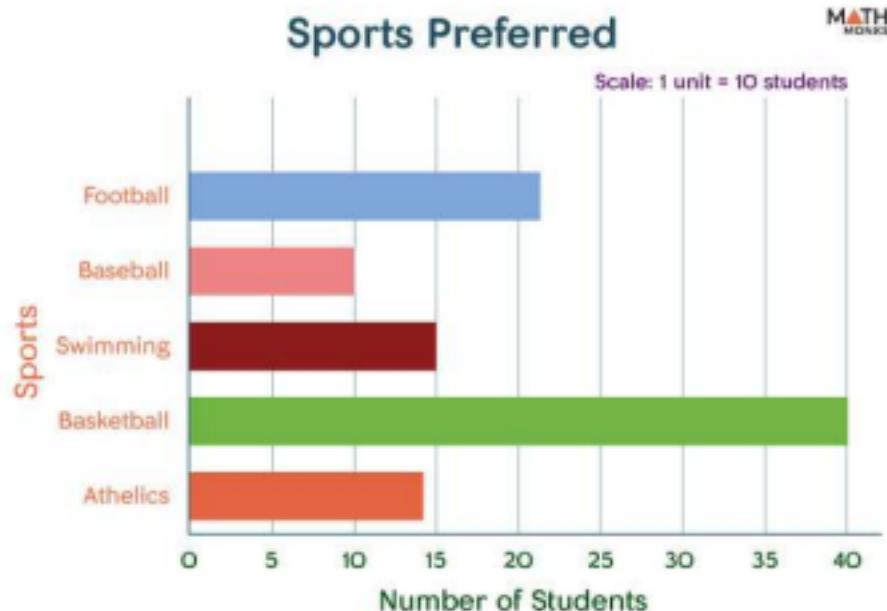


Types of Data Visualizations

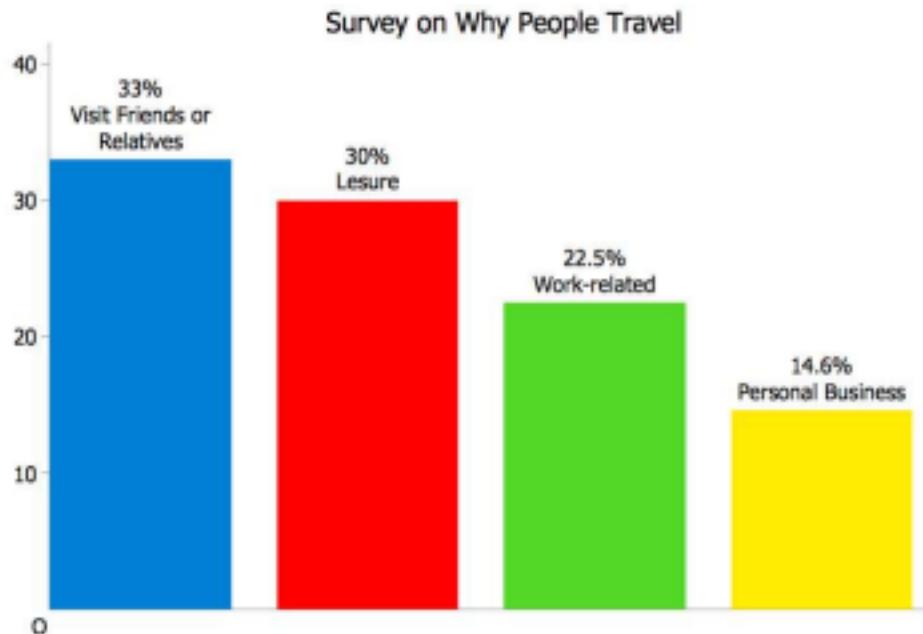
- **Comparisons:** Bar charts, column charts
- **Trends:** Line charts, area charts
- **Relationships:** Scatter plots, bubble charts
- **Distributions:** Histograms, box plots
- **Advanced:** Heatmaps, maps, network graphs



Bar Chart



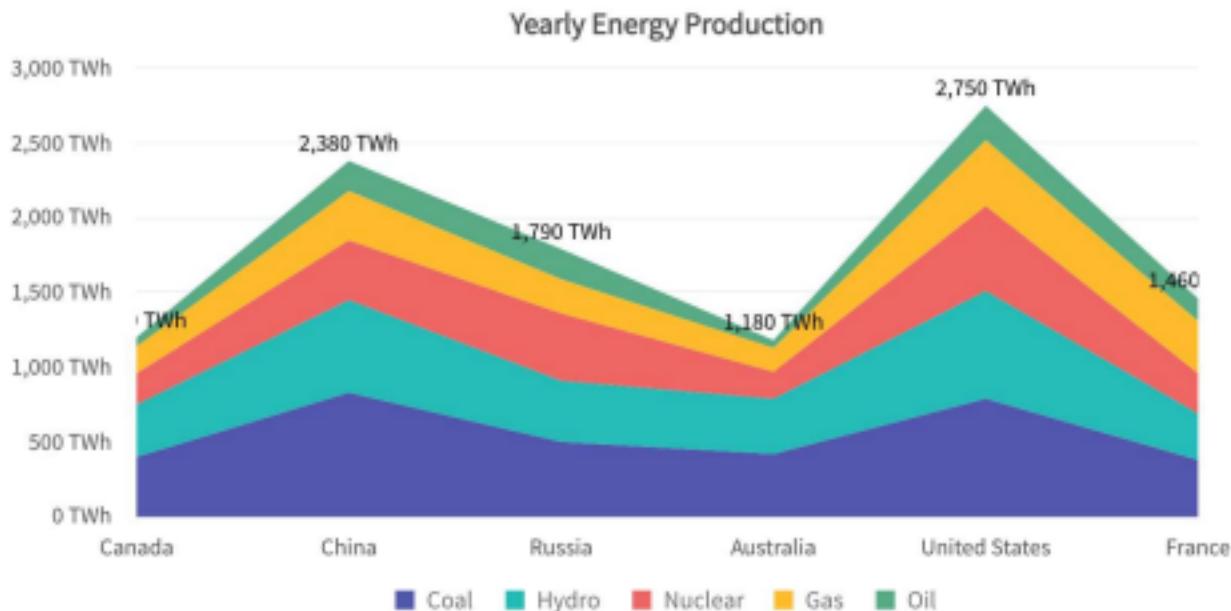
Column Chart



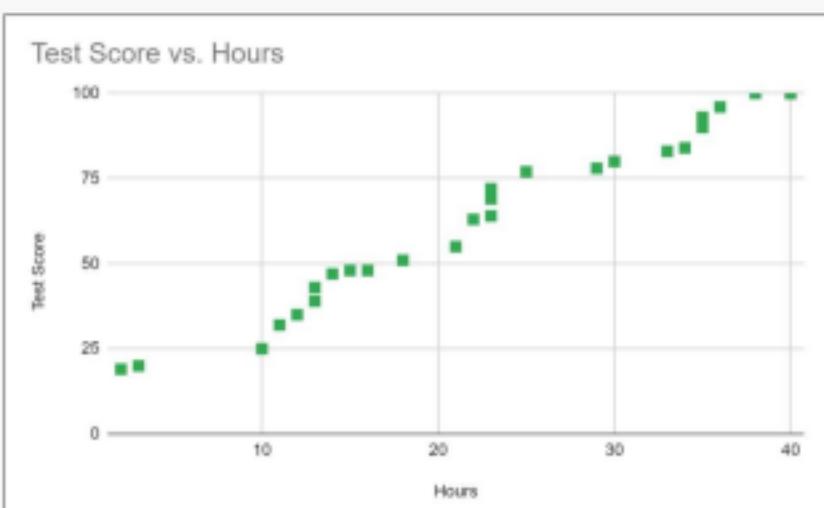
Line Chart



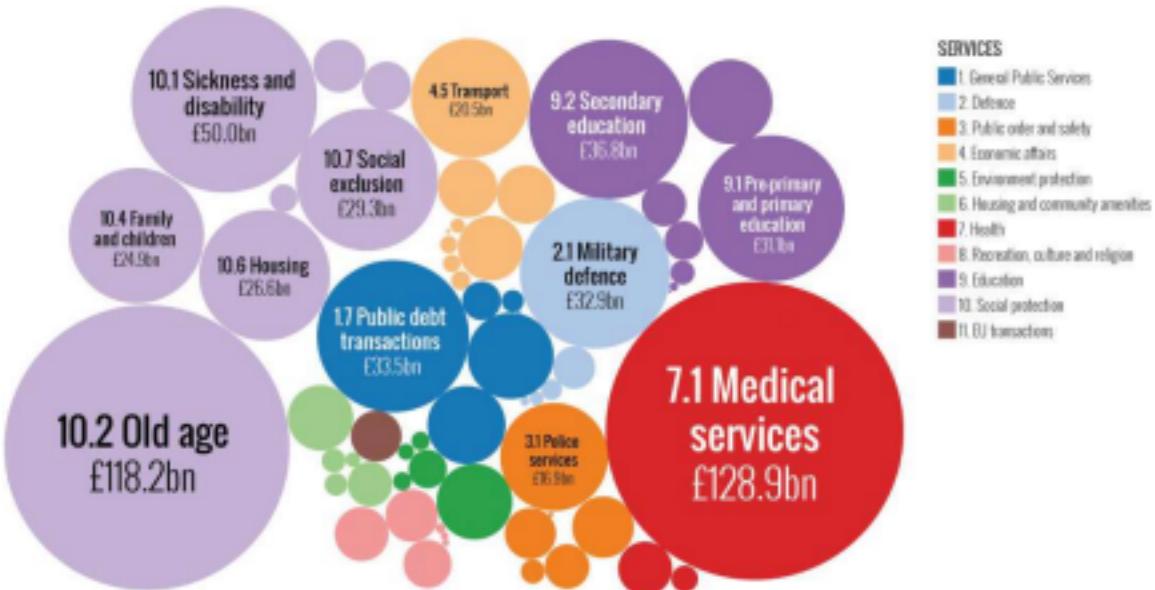
Area Chart



Scatter Plot

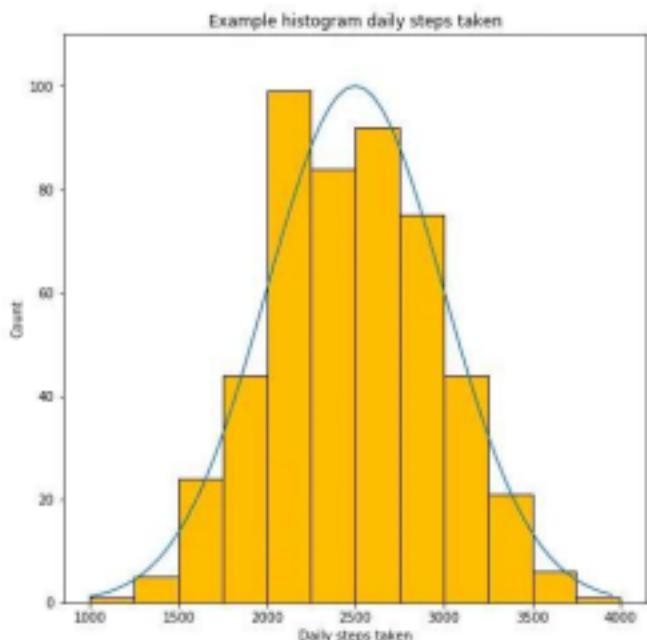


Bubble Chart

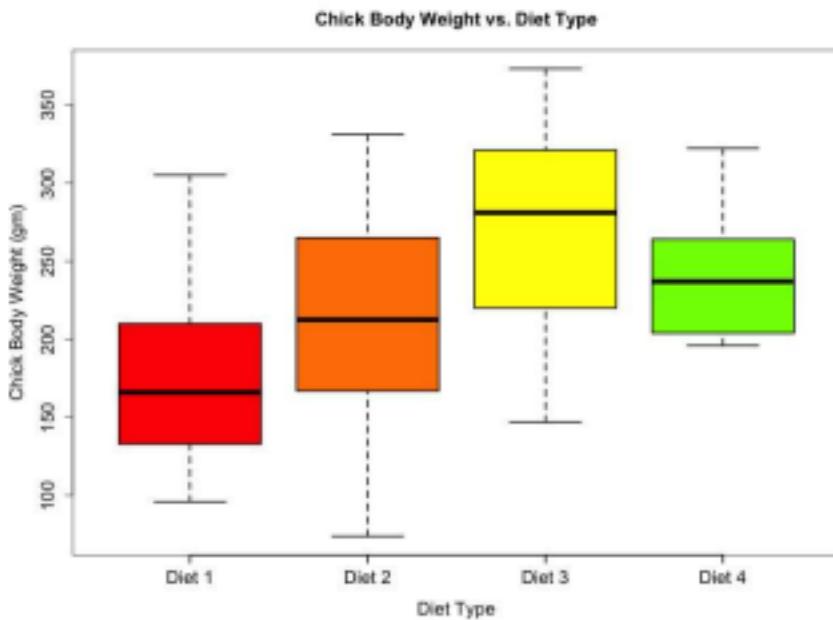


Analysis of Public sector capital expenditure (£ million) on services by function of the UK Government during 2014/15
Data from HM Treasury, 'Public Expenditure Statistical Analysis 2015' <http://www.gov.uk/government/statistics/public-expenditure-statistical-analysis-2015>

Histogram



Box Plot



Heatmap

Business Heatmap Analysis

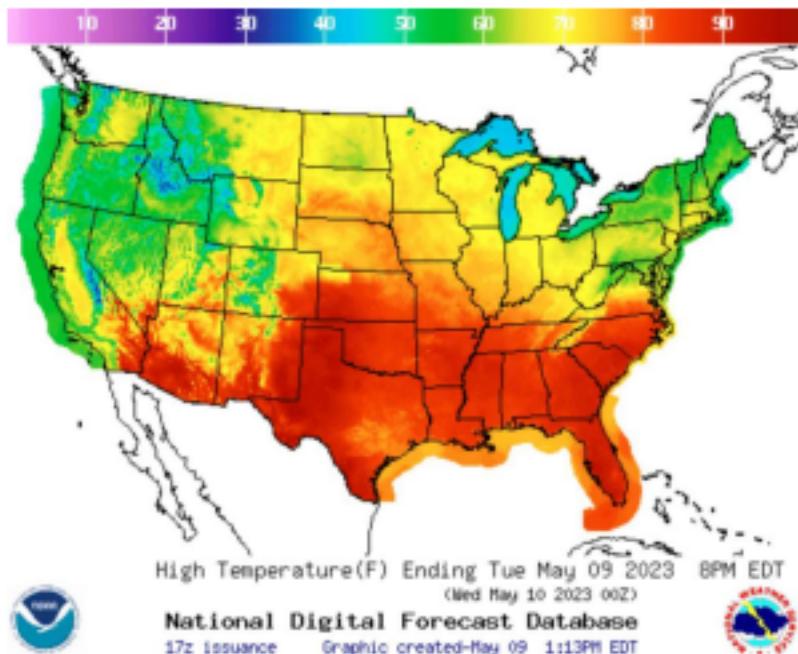
Questions	Product 1	Product 2	Product 3	Product 4	Product 5	Product 6	Product 7	Product 8	Product 9	Product 10	Total Average Per Product
Question 1	10.00	8.76	4.33	5.00	7.40	4.80	7.00	4.00	7.20	9.20	6.89
Question 2	8.50	4.00	7.00	7.50	6.50	3.50	3.00	6.50	7.00	7.00	6.20
Question 3	6.00	10.00	3.67	9.00	4.50	8.33	1.00	2.20	8.40	8.00	4.19
Question 4	6.00	7.00	9.00	4.67	4.00	3.67	1.00	6.50	1.50	4.00	4.93
Question 5	3.00	4.00	8.50	6.50	3.00	8.80	6.00	7.00	3.60	6.60	5.73
Question 6	3.80	4.20	8.20	6.00	4.20	2.80	3.80	4.20	9.20	8.60	4.40
Total Average Per Question	6.47	6.16	9.95	4.08	4.43	4.81	4.80	1.25	5.45	5.37	6.36

Rankings



This slide is 100% editable. Adapt it to your needs and capture your audience's attention.

Map



Network Graph



Principles of Good Visualization

- **Clarity** and Simplicity
- **Accuracy** and Honesty in Representation
- Right Chart for the Right Data
- Avoid Visual **Clutter** and Distortion



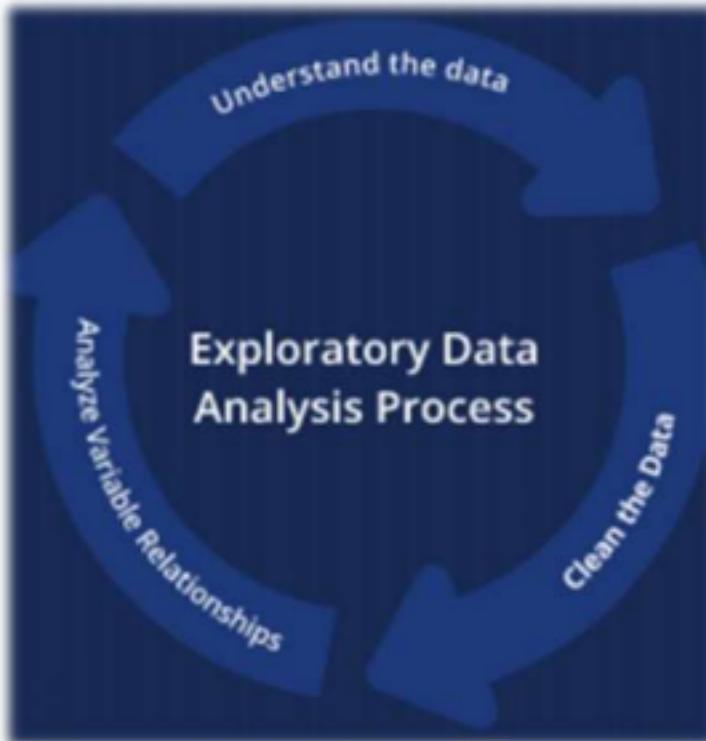
Tools for Data Visualization

- **Python:** Matplotlib, Seaborn, Plotly
- **R:** ggplot2, Shiny
- **BI Tools:** Tableau, Power BI
- **Web-based:** D3.js, Dash



Exploratory Data Analysis (EDA)

- Understanding **Distributions & Structure**
- Detecting **Patterns, Correlations, and Outliers**
- Guiding **Feature Engineering**
- Hypothesis Generation



Storytelling with Data

- Turning Insights into **Narratives**
- Highlighting Key **Findings**
- Supporting **Decisions** with Visuals
- Designing Visuals for **Audience Needs**



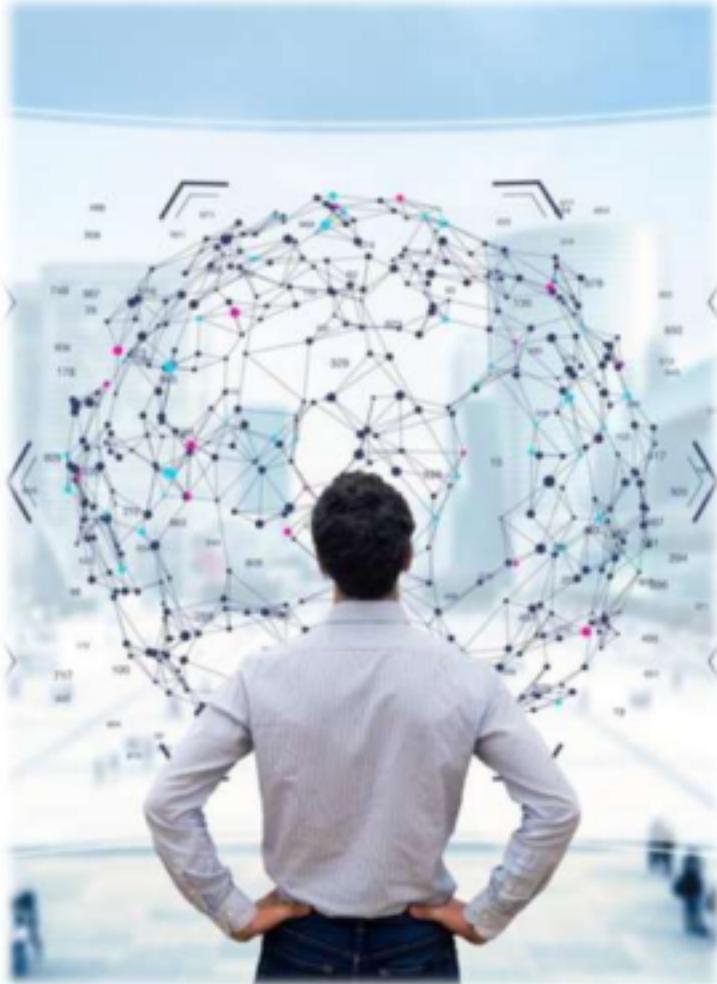
Common Visualization Pitfalls

- Misleading Axes and Scales
- Overplotting and Clutter
- Incorrect Color Schemes
- Using Inappropriate Chart Types



Applications in Data Science

- Business Intelligence **Dashboards**
- Scientific **Research and Analytics**
- Machine Learning **Model Explanation**
- Smart City and IoT **Analytics**





Power BI



CS202 Lecture 28

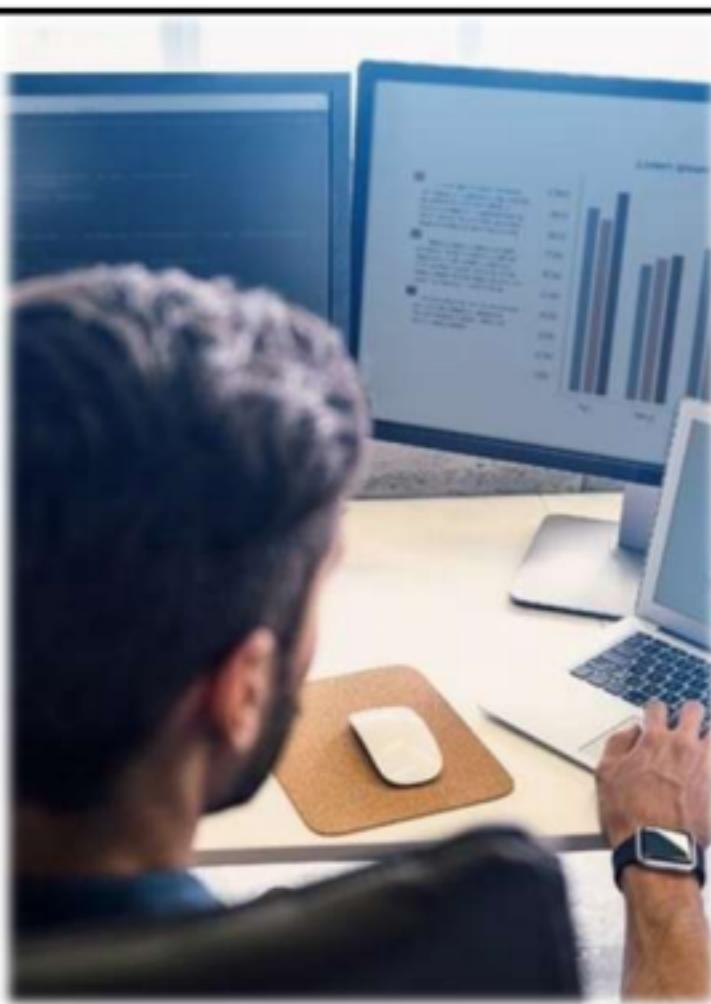
Topics To Cover

- **Importance of Data Analytics** in Decision-Making
- **Popular Tools** for Data Analysis
- **Key Features** of Power BI
- **Installation and Interface** of Power BI
- **Data Loading and Data Transformation** in Power BI
- **Creation of Basic Visuals** in Power BI



Data Analytics for Decision-Making

- Imagine a University Planning to **Open a New Campus**
- You are Part of the **Strategic Team** at the University
- Critical **Data Insights** are Required
- To Make **Informed Decisions**



Data Analytics for Decision-Making

Required Decisions:

- Location:** Where to Open (Rawalpindi, Islamabad, Lahore, Karachi)?
- Programs:** Which Programs to Start (Graduate, Postgraduate, PhD)?
- Disciplines:** Which Disciplines to Offer (IT, Management, Engineering)?
- Fees:** What should be the Tuition Fees?



Data Analytics for Decision-Making

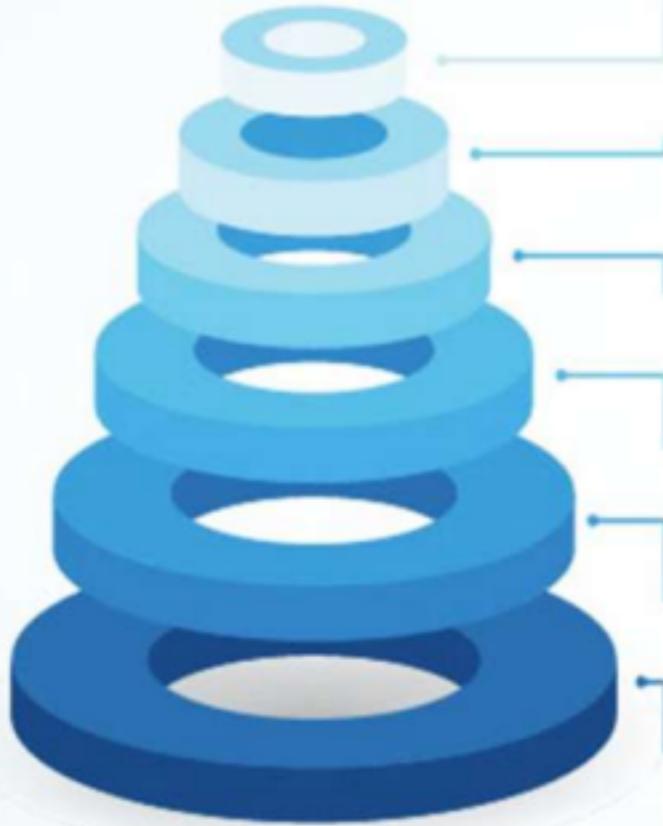
Key Insights for Data Analysis:

1. Location Analysis:

- Demographics of Potential Students
- Economic Conditions and Affordability
- Competition from Other Universities
- Accessibility and Transportation Infrastructure

2. Program Demand:

- Popularity of Graduate, Postgraduate, and PhD Programs
- Market Demand for Specific Disciplines
- Employment Trends and Industry Needs



Data Analytics for Decision-Making

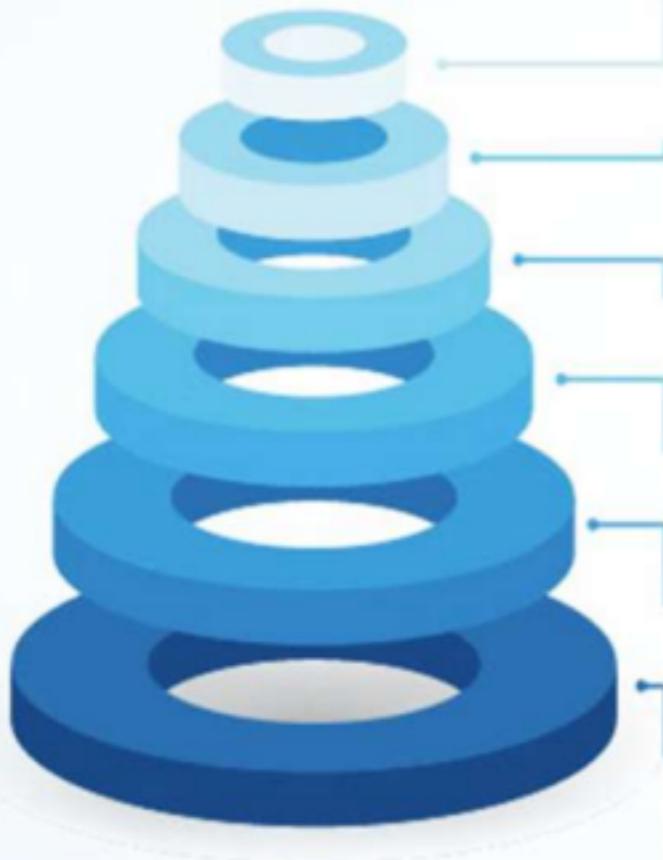
Key Insights for Data Analysis:

3. Discipline Demand:

- Local Industry Requirements (IT, Management, Engineering)
- Student Interest and Enrolment Trends

4. Fee Structure:

- Average Tuition Fees in the Region
- Cost of the Program



Data Analytics for Decision-Making

Consequences of Wrong Analytics and Decisions:

- **Wrong Location:** Poor Student Enrolment, High Dropout Rates, Financial Losses
- **Unsuitable Programs:** Low Demand Leading to Underutilized Resources
- **Misaligned Disciplines:** Graduates Struggling to Find Employment, Damaging the University's Reputation
- **Incorrect Fee Structure:** Making Education Unaffordable or Undervaluing the University's Offerings



Popular Tools for Data Analysis

Method	Introduction
Manual	Analyzing data by hand using pen and paper
Excel	Spreadsheet software for data organization, analysis, and visualization
Power BI	Business analytics tool for interactive data visualization and sharing insights
SPSS	Statistical software for in-depth statistical analysis and data management
Tableau	Data visualization tool for transforming data into interactive, shareable dashboards
Database with SQL	Using Structured Query Language to query and manage databases for data analysis

Popular Tools for Data Analysis

Method	Pros	Cons
Manual	Simple, no technology required	Time-consuming, error-prone, not scalable
Excel	Widely used, easy to learn, good for small data	Limited advanced analytics, not ideal for large datasets
Power BI	User-friendly, powerful visualizations, real-time insights	Requires learning, can be difficult without guidance
SPSS	Robust statistical analysis, used in research	Steep learning curve, expensive
Tableau	Excellent visualizations, user-friendly	Costly, may require training
Database with SQL	Powerful, handles large datasets, precise	Requires SQL knowledge, not instinctive

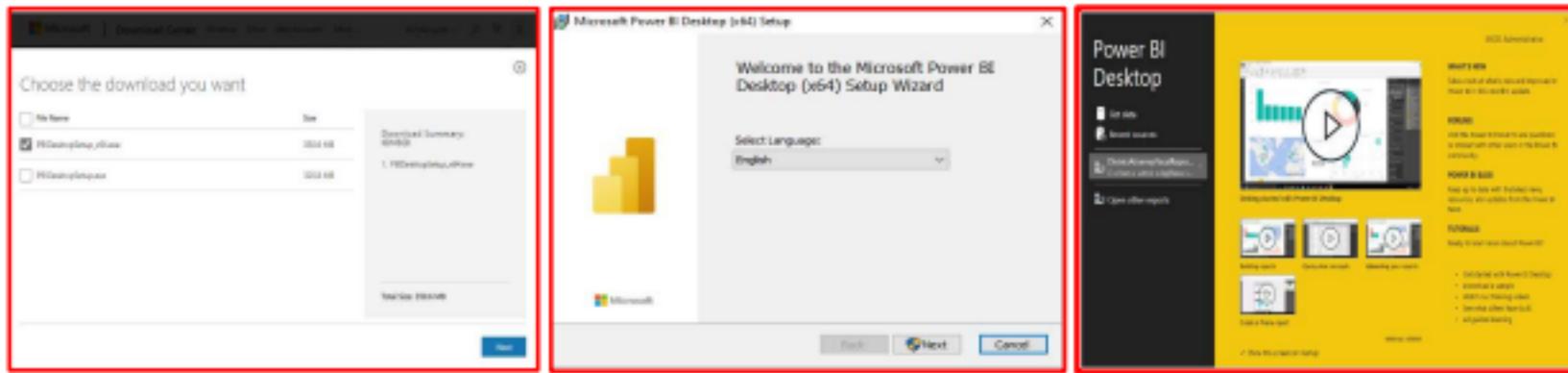
Why Power BI

- Power BI offers a **balance of user-friendliness and powerful capabilities** for data visualization and analysis
- It enables us to **easily transform data into actionable insights**, making it a preferred choice for this task
- **Key Features:**
 - Bring all: Unified data
 - Turn insights: Actionable impact
 - Uncover insights: AI-driven
 - Empower every: Team collaboration



Installation of Power BI

- 1. Download the Installation File from the Microsoft Website:**
www.microsoft.com/download/details.aspx?id=58494
- 2. Run the Installation File**
- 3. Launch Power BI**



Power BI Interface

The screenshot displays the Microsoft Power BI Desktop application window. The top navigation bar includes File, Home (selected), Insert, Modeling, View, Help, and External Tools. The ribbon below the bar contains sections for Data, Queries, and Insert.

The main workspace shows a dashboard titled "Store Sales Overview" containing several visualizations:

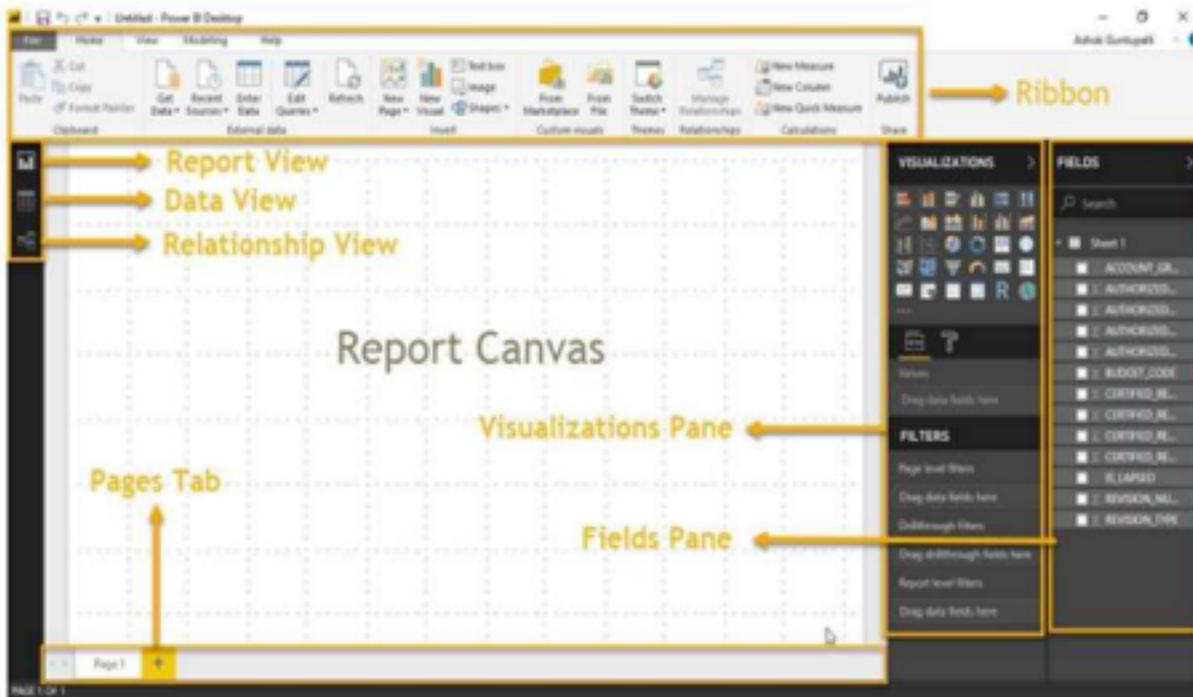
- A pie chart showing the distribution of sales by category.
- A bar chart titled "District Monthly Sales" with values 10 and 104.
- A map of the United States where states are colored red or blue.
- A bubble chart showing data points across a grid.

On the right side of the interface, there are three panels:

- Filters:** A search bar and a list of filters applied to the current page, including Chain, City, District, Name, Open Month, Store Type, and more.
- Visualizations:** A large library of visualization icons categorized under Build visual, such as bar charts, line graphs, maps, and tables.
- Fields:** A search bar and a list of fields available for selection, including Sales, District, Item, Store, and Time.

At the bottom of the workspace, tabs include Info, Overview (selected), District Monthly Sales, New Stores, and a plus sign icon. The status bar at the bottom right shows "Page 2 of 4" and "51%".

Power BI Interface



Data Loading in Power BI

- 1. Open Power BI Desktop:** Launch Application
- 2. Get Data:** Click “Get Data”
- 3. Choose Data Source:** Select Source
- 4. Connect to Data Source:** Enter Details
- 5. Load Data:** Preview, Load



Sample Data

City	Program	Discipline	Annual Fee	Population	Compete.	Unemp. Rate	Avg. Salary
RWP	Postgrad	Mgmt	183000	5000000	7	5	70000
KHI	PhD	Mgmt	152000	30000000	4	7	80000
ISB	Grad	Engg	337000	2000000	3	5.5	70000
RWP	Grad	Mgmt	285000	5000000	8	5	70000
KHI	PhD	IT	219000	30000000	1	6	100000
KHI	Grad	IT	120000	30000000	4	6	100000
KHI	Grad	Mgmt	233000	30000000	1	7	80000
KHI	Grad	Engg	253000	30000000	4	8	90000
ISB	Grad	Mgmt	256000	2000000	7	4.5	60000
KHI	Grad	Mgmt	309000	30000000	2	7	80000
ISB	Grad	Mgmt	455000	2000000	3	4.5	60000
LHE	PhD	IT	193000	20000000	5	5.5	110000
RWP	Grad	IT	111000	5000000	2	4	90000
KHI	PhD	IT	460000	30000000	4	6	100000

Meta Data

- **City:** The city where the new campus should be considered
- **Program:** Type of program to be offered
- **Discipline:** Academic disciplines to be considered
- **Annual Fee:** Tuition fee per annum, rounded to nearest 1000
- **Population:** Population of city, rounded to the nearest 100,000
- **Competitors:** Number of competing universities in the city
- **Unemployment Rate:** City unemployment % rate for the discipline
- **Average Salary:** Average salary for graduates in the specific discipline in the city, rounded to the nearest 100,000

Data Transformation in Power BI

- Click on “Transform Data” in the “Home” tab of Power BI to open the Query Editor
- Apply any transformation such as remove column, sort data, merge queries and pivot/unpivot
- Remove Unnecessary Columns (If any)
- Rename Columns to more meaningful names
 - Program -> Program Level
 - Population -> City Population
 - Competitors -> No of Competitors
- Apply Changes

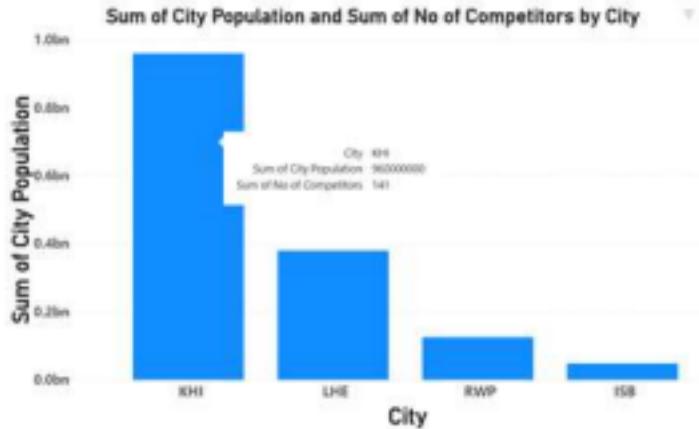


Data Visualization in Power BI

1. Location: Where to Open the Campus?

Column Plot:

- X-axis: City
- Y-axis: City Population
- Tooltip: No of Competitors



2. Programs: Which Programs to Start?

Cluster Column Chart:

- X-axis: City
- Y-axis: Program Level
- Legend: Program Level



Data Visualization in Power BI

3. Disciplines: Which Disciplines to Offer?

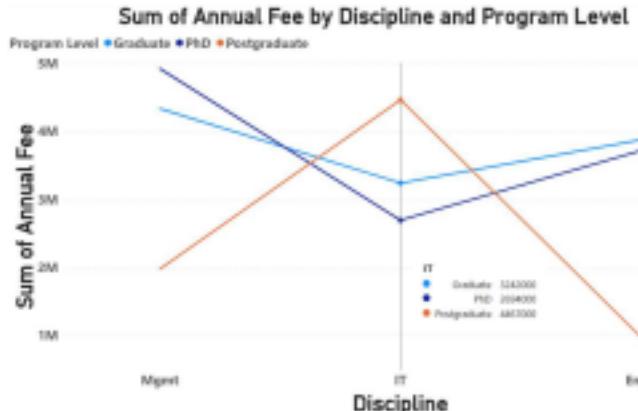
Cluster Column Chart:

- X-axis: City
- Y-axis: Count of Discipline
- Legend: Discipline

4. Fees: What Should Be the Tuition Fees?

Line Plot:

- X-axis: Discipline
- Y-axis: Annual Fee
- Legend: Program Level



Conclusion

- Microsoft Power BI is an **Interactive Data Visualization Software Product** developed by Microsoft with a primary focus on business intelligence.
- Power BI stands for **Power Business Intelligence** and refers to a collection of software services, tools, and connectors that help transform data from multiple sources into actionable insights.
- By using Power BI, one can make **Data-Informed Decisions** across organization.

Thank You !