KAUTILYA SCHOOL OF PUBLIC POLICY
GITAM (Deemed to be University)
Rudraram, Patancheru Mandal
Hyderabad, Telangana 502329

Course Code: PPOL7121	Course Title: Data Visualization and Dashboards	
Trimester: 4	Course Type: Core	Credits: 3
Home Program(s): MPP	Batch/Academic Year: 2023-25	
Course Instructors: KPMG	Assigned T/RA:	

Course Description

This course dwells in data analytics, data visualization and dashboard design towards extracting and visualizing usable information from public data sets. It explores prominent government sources of data, data releases, and types of public data. Students will build foundations in the different stages of working with data. Students will learn how to understand complex statistics, and compile, clean, and visualize data. Students will use R to learn the basics of data analytics and visualization. In the present day where data is the biggest asset any individual or organization can have, the students will be learning ways to make data actionable. Students will also learn how to design dashboards to present data effectively and communicate insights to the relevant stakeholders through iterative designing and storytelling. The course culminates in a group project where students showcase their learning through analyzing a dataset, creating visualizations and dashboards, and presenting the same to stakeholders in a simulated environment

Learning Objectives

1. To learn how to access, understand, and use public data sources.
2. To understand the role of data analytics and data visualization in decision making and public policy
3. To apply different stages of working with data to real life data sets.
4. To compile, clean, standardize, analyze, and visualize data using R.
5. To learn how to create effective dashboards to communicate insights to relevant stakeholders.
6. To transition from using plain numbers to using maps, graphs, infographics, dashboards etc.

Course Outcomes

On successful completion of this course, students will be able to:

1. Understand how to use R for data analytics and advanced data visualization
2. Understand the public sources of data and ways to access and use public data.
3. Use different stages of working with data using basic and advanced tools.
4. Analyze techniques of visualization and converting data sets into visual form.
5. Create and use dashboards to enhance data visualization and presentation.

Course Schedule

Unit I	4.5 hours	Introduction to Data Analytics and Visualization

- Introduction to Public Data Sources including different government sources- Lok Sabha, Rajya Sabha, government data portals, sector specific and project specific sources (eg - UDISE, NDEAR, MGNREGA, MoHFW, Swachh Bharat) and different data releases (eg- Census, Surveys, Annual Reports)
- Introduction to databases like CMIE and different types of data and data stages
- Overview of Data Analytics in Public Policy: Discussing the role of data analytics in decision-making, policy formulation, and evaluation with examples
- Fundamentals of Data Visualization: best practices such as choosing the right chart type; color theory, clarity and accuracy in presenting data, common pitfalls like misleading axes. Example: Comparing bar charts vs. pie charts for demographic data presentation.
- Introduction to R: Installing R and RStudio, basic syntax, data types (R objects and attributes, vectors, lists, matrices, factors, missing values), sub setting, vectorized operations. Example: Writing simple R scripts to manipulate data frames. applicability in a real-world context.

Unit II	5.5 hours	Programming with R

- Programming Fundamentals (control structures, R functions, scoping rules, basic coding standards, dates and times in R)
- Loop Functions and Debugging
- Simulation (random sampling, linear models) and Profiling
- Basic familiarization to Tidyverse – 'ggplot2' for visualization, 'dplyr' for data manipulation, 'readr' for data import and 'tibble' for data frames
- **Case Study**: Analyzing Crime rates across major cities
1. Objective: Students learn how to load, clean and visualize crime data using R
2. Task: Importing crime data from different cities, using basic 'dplyr' commands to filter and summarize the data and create initial visualization using 'ggplot2' to identify trends.
3. Discussion: the varying crime rates and distribution to inform policing. For instance, if data shows certain crime peak at specific times or cities, the Law Enforcement Agencies can optimize patrol schedules or community outreach programs.
- **Post-lecture exercise**: Load a dataset provided (e.g., local government spending data). Perform basic data inspection and manipulation tasks:
    1. Load the data using readr.
    2. Use dplyr to filter for entries where spending is above a certain threshold.
    3. Create a basic plot using ggplot2 to visualize the spending trends over time.

Unit III	8 hours	R for Data Analytics

● Data import basics, R data frames, cleaning, organizing, and transforming data, bias data functions, more structured data exploration with tibbles and Tidyr

● Data Manipulation with dplyr: Techniques for filtering, selecting, mutating, and summarizing data. Example: Cleaning and preparing a dataset on healthcare spending.

● Data Analysis Techniques: Applying basic statistical functions (mean, medium, mode), exploring correlation, and simple regression models.

● **Case Study**: Analyzing public transportation usage patterns in cities
  1. Objective: Using data manipulation techniques in R to analyze public transportation usage patterns, identify bottlenecks recommend improvements to enhance efficiency and accessibility.
  2. Task: Importing data that include details like travel time slots, route details, passenger counts and station information, etc. Cleaning data for any inconsistencies or missing values. Categorizing routes based on average passenger counts and travel time slots. Using 'dplyr' to find average median and peak passenger loads on different routes at different times. Analyzing routes with consistently high passenger loads and busiest peak time analysis.
  3. Discussion: how analysis can inform transit authorities about necessary adjustments in service frequency and capacity enhancement to meet passenger demands. Deliberation on how analysis can guide into decisions on where to allocate resources for improving public transport infrastructure such as buses, trains or stations. Evaluate how increasing efficiency of public transport can add value to sustainability goals like reducing vehicle emissions. Discussion on the role of public feedback in refining data analysis and transport planning. Also, deliberation on integrating real-time passenger counting and route mapping can enhance data accuracy and transport operations.

● **Post-lecture exercise**: Given a dataset on public school performance:

  1. Clean the data by handling missing values and removing duplicates.

  2. Create a new column that categorizes schools based on performance metrics.

  3. Summarize the data to find average performance metrics by category.

Unit IV	6 hours	**Advanced Data Visualization in R**

● Visualization basics in R and Tidyverse and common problems in visualization
● Using ggplot2 for Data Visualization: Constructing basic plots like histograms, scatter plots, and line graphs; customizing plots with themes, aesthetics, facets, and scales. Example: Visualizing economic growth over time
● Interactive Visualizations: Introduction to packages like plotly for interactive plots and shiny for developing web apps. Example: Creating a shiny app to explore the relationship between education levels and employment rates.
● **Case Study**: Impact of Education on Employment rates
  1. Objective: Analyzing the relationships between education levels and employment rates across different regions to inform targeted educational funding and job training policies
  2. Task: Import and clean employment and education data from government databases or reputable sources like CMIE. Handle missing values, outliers, format inconsistencies and merge datasets. Create Scatter plots using 'ggplot2' to compare the average highest education achieved with employment rates in each region. Develop a line graph showing trends over time in education attainment and corresponding employment rates to identify any shifts. Incorporate or add any more variable in the analysis like job sector growth using color or facets to represent different dimensions such as time, region or job sector.
  3. Discussion: patterns seen such as regions with higher education levels correlating with higher employment rates that show growth in regions with certain educational strengths. Deliberation on how visualization pinpoint the regions where increased educational funding could boost employment rates. Evaluate the need for targeted job training programs in regions where employment rates do not correlate strongly with higher education levels, suggesting possible skill mismatch in the job market.
● **Post- lecture exercise**: Using the same public school performance dataset:
  1. Use ggplot2 to create a scatter plot comparing student-teacher ratios to performance scores, adding a trend line.
  2. Customize the plot by changing color themes and adding labels for clarity.
  3. Create a series of visualizations in R to represent the data insights.
  4. Prepare a short presentation or report summarizing the findings and visualizations

Unit V	3 hours	**Communicating Dashboards to Stakeholders**

● Dashboard Design Principles: Focusing on layout design, choosing the right visuals, setting up user interactivity features, accessibility, etc.

● Storytelling with Data: Narrative techniques or emphasizing on key points.

● Presentation Skills: Understanding the stakeholder needs and aligning it with dashboard objectives, applying depth of data based on targeted stakeholder, effective and clear communication of insights, how to use reports or slides to complement the dashboard presentation.

● Iterative Designing: receiving feedback through user testing, version controls, etc.

● **Post-lecture Exercise**:

1. Students create a visualization from a dataset on public health, focusing on one disease
2. Prepare a short presentation or report summarizing the findings and visualization
3. Identify the relevant stakeholders who are communicated of these findings and share the insights
4. Practice iteration on the same incorporating the feedback from peer groups formed for this exercise

Unit VI	3 hours	Group Project

- **Capstone Project**: Students use the skills learned to analyze a dataset, create visualizations in R, and develop a dashboard.
- **Peer Review**: An interactive session where students provide feedback on each other's projects based on clarity, usability, and analytical depth.
- **Final Presentations**: Students present their projects to a panel comprising course instructors and invited stakeholders, simulating a real-world proposal.