Data Visualisation using R & Tableau

Dr. Praveena Musunuru

December 15, 2024

विद्या परं दैवतम
IIM
VISAKHAPATNAM

About ggplot2

Description: ggplot2 is a powerful data visualization package in R.

- Allows customization and layering of plots.
- Supports multiple plot types like line, bar, scatter, bubble charts, and heatmaps.

$$\begin{gathered}\text{ggplot(data, aes(x = var1, y = var2, color = var3))} + \\ \text{geom\_point()} + \\ \text{facet\_wrap(category)} + \\ \text{scale\_fill\_gradient(low = "blue", high = "red")} + \\ \text{labs(title = "Title", x = "X-axis", y = "Y-axis")} + \\ \text{theme\_minimal()}\end{gathered}$$

ggplot2 Dictionary: Controllable Elements

What You Can Control in ggplot2:

- **Data:** Specify the dataset to be visualized.
- **Aesthetics (aes):** Map data variables to visual properties such as:
  - **x-axis and y-axis:** Define axes for your plot.
  - **color:** Map a variable to line or point color.
  - **size:** Define the size of points or lines.
  - **shape:** Specify the shape of points.
- **Geometries (geom):** Define the type of plot, such as:
  - `geom_line`: Line chart.
  - `geom_bar`: Bar chart.
  - `geom_point`: Scatter plot.
  - `geom_tile`: Heatmap.

ggplot2 Dictionary: Controllable Elements

- **Facets:** Create multiple panels for different subsets of data using:
  - `facet_wrap`
  - `facet_grid`
- **Scales:** Adjust the visual representation, including:
  - `scale_fill_gradient`: Gradient color mapping.
  - `scale_x_continuous`: Modify x-axis.
  - `scale_y_log10`: Apply logarithmic scaling to y-axis.
- **Themes:** Customize non-data elements such as:
  - Background, gridlines, font styles.
  - `theme_minimal`, `theme_dark`, `theme_classic`, etc.
- **Labels:** Add titles, axis labels, and legends using `labs`.
- **Annotations:** Highlight specific data points or areas.

विद्या परं दैवतम्
IIM
VISAKHAPATNAM

Line Chart: Nifty 50 Index

Description: A line chart visualizing the closing price of the Nifty 50 index over time.

- Represents trends over time.
- Useful for identifying market patterns.

Heatmap: Correlation of Financial Ratios

Description: A heatmap visualizing the correlation between various financial ratios.

- Correlation values range from -1 to 1.
- Helps identify relationships between variables.

Introduction to Wide and Long Formats

Wide Format:

- Data where each column represents a different variable.

Example - Wide Format:

symbol	Mcap	PB	PE	Beta
ABB	271221.5	8.05	81.36	1.31
ACC	271549.8	3.1	46.98	0.98

Long Format:

- Data where values are stacked into a single column with an identifying column for the variable.

Example - Long Format:

symbol	Metric	Value
ABB	Mcap	271221.5
ABB	PB	8.05
ABB	PE	81.36
ABB	Beta	1.31
ACC	Mcap	271549.8
ACC	PB	3.1

Bar Chart: Market Capitalization Comparison

Description: A bar chart comparing the market capitalization of companies.

- Horizontal bars make it easier to compare categories.
- Market cap values are plotted for visualization.

Bubble Chart: Multivariate Visualization

Description: A bubble chart visualizing Market Cap, PE Ratio, and Beta.

- Bubble size represents Market Cap.
- Color represents PB Ratio.
- Highlights multivariate relationships.

Scatter Plot: Correlation and Regression Analysis

Description: A scatter plot showing the relationship between PE Ratio and Market Cap.

- A regression line is fitted to identify trends.
- Highlights positive/negative correlations.

Word Cloud: Textual Data Visualization

Description: A word cloud visualizing the most frequent terms in textual data.

- Larger words indicate higher frequency.
- Highlights important keywords in the text.

Overview of Integrating R with Tableau

Why Integrate R with Tableau?

- Leverage Tableau's visualization capabilities with R's statistical and analytical power.
- Perform advanced calculations, machine learning, or custom statistical models directly within Tableau.

What is Rserve?

- A lightweight R server that allows Tableau to communicate with R.
- Acts as a bridge between Tableau and R for seamless integration.

Setting Up Rserve

Prerequisites:

1. Install R on your machine.
2. Install the `Rserve` package in R:
3. Start the Rserve server by running `Rserve()` in the R console.

Connecting Tableau to R:

1. Open Tableau Desktop.
2. Navigate to: **Help - Settings and Performance - Manage External Service Connection**.
3. Configure:
   - **Service:** Rserve.
   - **Host:** `localhost` (or the R server IP).
   - **Port:** Default is 6311.
4. Test the connection.

विद्या परं दैवतम्
IIM
VISAKHAPATNAM

Writing Calculations in Tableau using R

Syntax for R Calculations in Tableau:

- **SCRIPT_REAL:** Returns numerical values.
- **SCRIPT_STRING:** Returns text.
- **SCRIPT_INT:** Returns integer values.
- **SCRIPT_BOOL:** Returns Boolean values.
- R scripts in Tableau are recalculated every time the data changes.