Paper PP22

Automation of Exploratory Report Generation using R shiny and R Markdown

Denny Sebastian, Catalyst Clinical Research, Kochi, India
Vikash Jain, Catalyst Clinical Research, Hillsborough, New Jersey, USA

ABSTRACT

In the field of clinical data analysis and reporting, the demand for efficient and automated solutions has grown exponentially. This paper presents a novel approach to automate the generation of exploratory reports using the capabilities of R Shiny and the capability of R Markdown for document production. Traditional report generation often involves manual steps, which can be time-consuming and error prone. By leveraging the integration of R Shiny and R Markdown, we have developed a streamlined workflow that significantly reduces manual intervention and accelerates the production of insightful exploratory reports.

INTRODUCTION

Clinical data analysis is examining medical data for trends, patterns, and insights to improve treatment effectiveness, medical decision-making, and patient outcomes. Because of the requirement to analyze large datasets quickly, there is a growing need in the clinical data analysis and reporting space for simplified, automated solutions. Rapid data interpretation, pattern identification, and anomaly detection are made possible by efficient technologies that make use of machine learning and artificial intelligence algorithms. Error is decreased in crucial healthcare decision-making due to automation, which not only improves accuracy but also speeds up the extraction of insights. In the ever-changing field of medical research and healthcare delivery, these systems' scalability guarantees flexibility in response to changing demands for research, resulting in faster progress and better patient care.

Exploratory Data Analysis (EDA) in clinical research involves analyzing and summarizing data to extract insights, identify patterns, and produce hypotheses—all of which are essential for guiding subsequent research directions. EDA here seeks to understand treatment responses, potential correlations, and patient demographics. Researchers investigate data variability, distributions, and correlations between variables via statistical measurements, visualizations, and summary approaches. Data distributions, outliers, and trends are shown by graphical representations like histograms, boxplots, and scatter plots, while summary statistics like mean, median, and standard deviation provide light on central tendencies and dispersion.

EDA in clinical research often involves analysing baseline characteristics, assessing treatment effects, and investigating potential associations between clinical parameters and outcomes. Visualization aids like survival curves, Kaplan-Meier plots, and forest plots help assess treatment efficacy, while correlation analyses and regression models explore relationships between variables. Effective EDA not only helps in understanding the dataset's characteristics but also informs study design, hypothesis formulation, and aids in identifying potential confounding factors critical for robust clinical research outcomes.

VISUALIZATIONS IN EXPLORATORY DATA ANALYSIS

Exploratory data analysis (EDA) relies heavily on visualizations because they offer thorough and understandable insights into datasets. Analysts can easily understand trends, distributions, correlations, and outliers with the use of graphs, charts, and plots. Box plots show core central tendencies and variations, scatter plots show relationships between variables, and histograms clarify data distributions. Pie charts provide categorical comparisons, and heatmaps display correlations. These visual tools support a deeper knowledge of data structure and provide guidance for following analytical judgments and hypotheses in the EDA process by helping to discover patterns, anomalies, and possible topics for more inquiry.

In R, a powerful statistical programming language, numerous visualization libraries help to exploratory data analysis (EDA). Here are some major visualization tools commonly used in EDA with R:

- **ggplot2**: A versatile and widely used package for creating a variety of visualizations, including scatter plots, histograms, box plots, and more, using a layered grammar approach.
- **Plotly**: Allows for interactive visualizations, including scatter plots, line plots, bar charts, and 3D plots, enhancing EDA by enabling zooming, panning, and hover-over details.

- **ggvis**: Builds on ggplot2 but emphasizes interactive web-based visualizations, suitable for dynamic exploration and manipulation of data.
- **lattice**: Provides a powerful framework for creating conditioned plots like trellis plots, which are useful for visualizing relationships across multiple variables.
- **R base graphics**: The default plotting system in R offers various functions like plot(), hist(), and boxplot() for creating basic visualizations directly from R's base package.
- **Tidyverse** suite: Integrates several packages (including ggplot2) under a coherent framework, facilitating data manipulation and visualization in a consistent and intuitive manner.

These tools empower analysts to generate a wide range of visual representations to explore datasets, identify patterns, uncover relationships, detect anomalies, and gain a deeper understanding of the data's structure and characteristics during exploratory data analysis in R.

RMARKDOWN FOR REPORT GENERATION

Rmarkdown is a strong tool in R for creating dynamic reports that incorporate text, code, and visuals. It facilitates the smooth amalgamation of text, code, and output into a unified document format, hence enabling the automatic creation of reports. The integration of narrative, code chunks, and visuals improves document readability and reproducibility by streamlining the production of dynamic, repeatable reports in several forms such as Word, PDF, HTML, or presentations. The features it offers for creating reports are broken out as follows:

STRUCTURE:

Markdown Formatting: Allows easy creation of text with simple formatting using Markdown syntax.
Code Chunks: Integrates R code within the document using code chunks delineated by {r} ... .

FEATURES:

Reproducibility: Ensures reproducibility by embedding executable code within the document.
Flexibility: Supports various output formats like HTML, PDF, Word, presentations (e.g., PowerPoint), and more.
Templates: Offers customizable templates for specific output formats, providing a starting point for report creation.
Parameterization: Enables parameterization to create dynamic reports by changing input values interactively.

WORKFLOW:

Setup: Install the Rmarkdown package (install.packages("rmarkdown")) and any required formats or additional packages.
Document Creation: Compose your report using Markdown for text and incorporating R code chunks where needed.
Knitting: Use the "Knit" button or the render() function in R to execute the code and generate the final report in the desired format.

EXAMPLE USE CASES:

Data Analysis Reports: Combine analysis steps, results, and interpretations in a single document.
Documentation: Create documentation with code explanations and outputs.
Presentations: Develop slideshows with interactive or static content.
Dashboards: Build simple dashboards for quick data exploration.

TIPS:

Chunk Options: Utilize chunk options (echo, eval, include, etc.) to control what gets displayed or evaluated in the output.
Rmarkdown Extensions: Explore additional functionalities with extensions like bookdown, flexdashboard, or shiny for more complex and interactive reports.

Rmarkdown streamlines the process of generating comprehensive and reproducible reports, making it a valuable tool for communicating findings and analyses in a clear and organized manner within the R ecosystem.

EXPLORATORY REPORT FOR SDTM DOMAINS
The following figures shows some examples of the exploratory report for the domains DM and LB.

Figure 1: Report Introduction

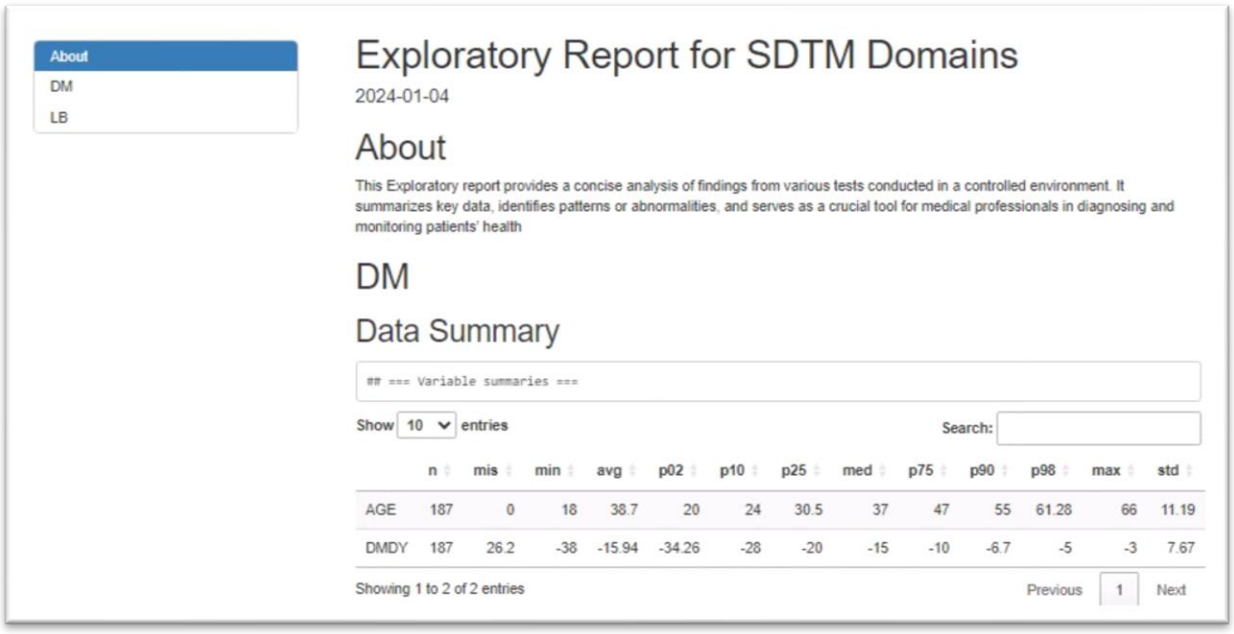


Figure 2: Cross Tabulation

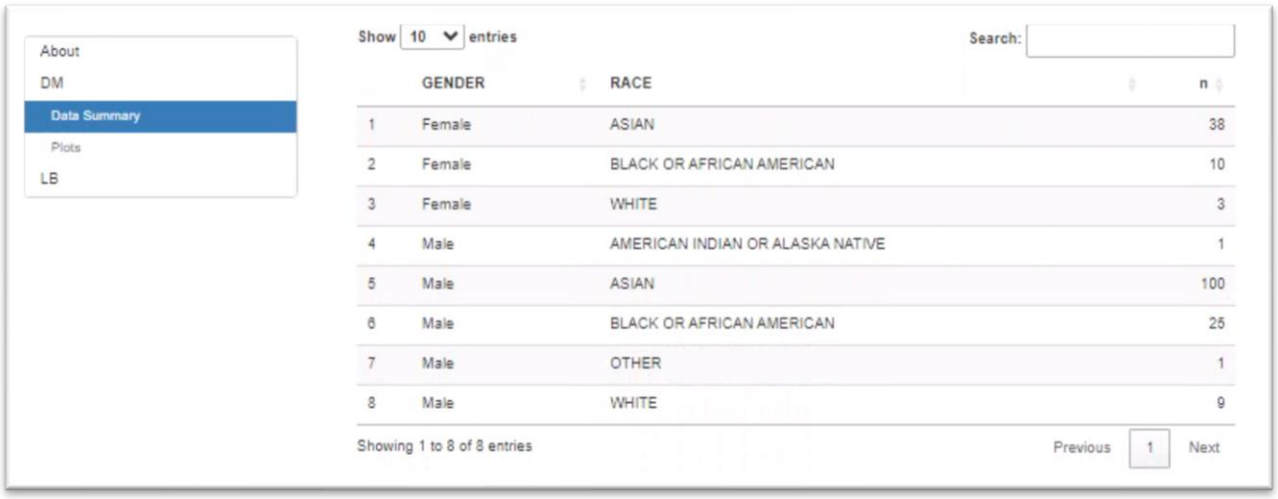


Figure 3: Violin Plot

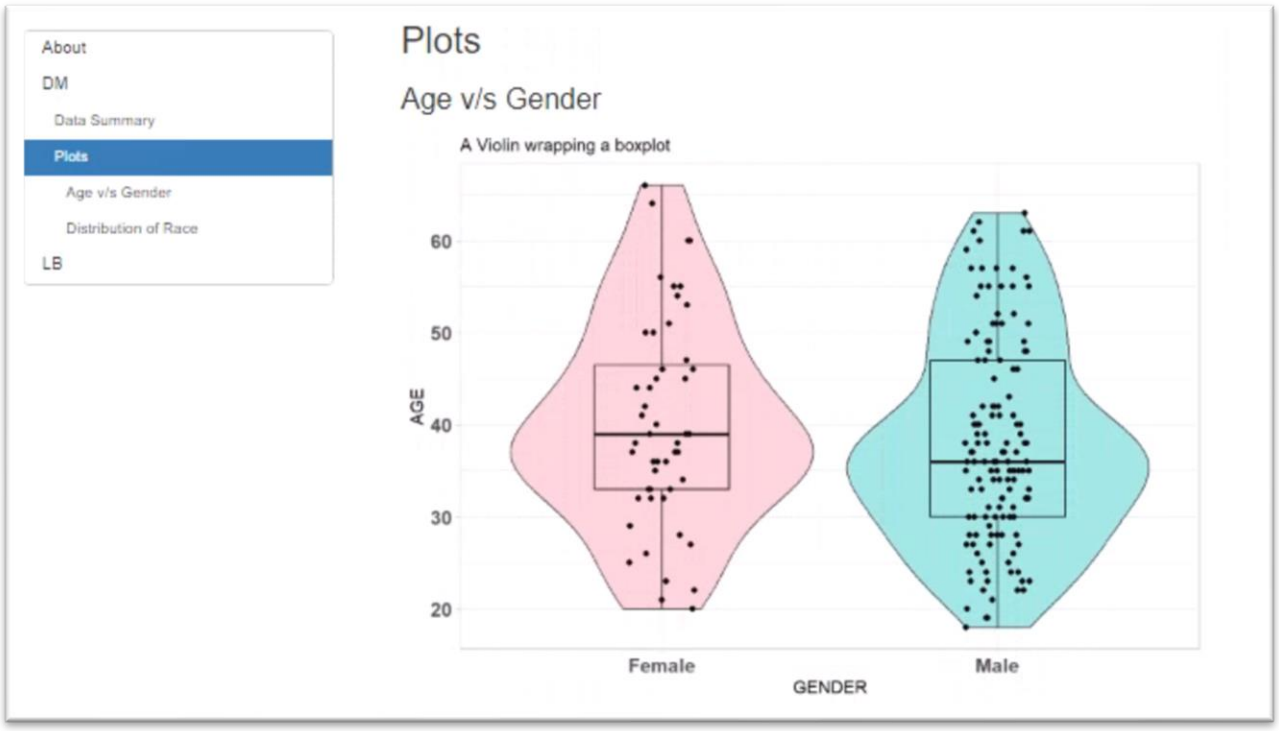


Figure 4: Distribution Plot 1

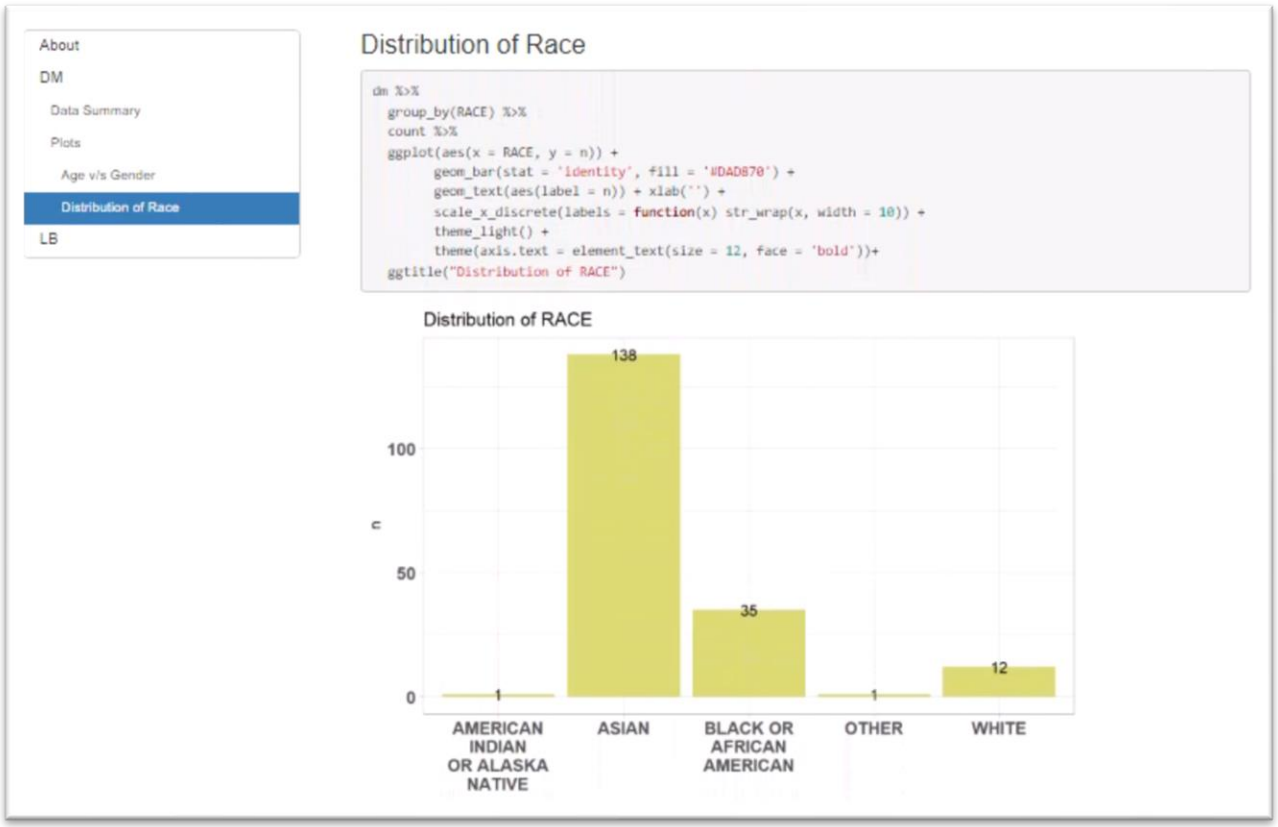
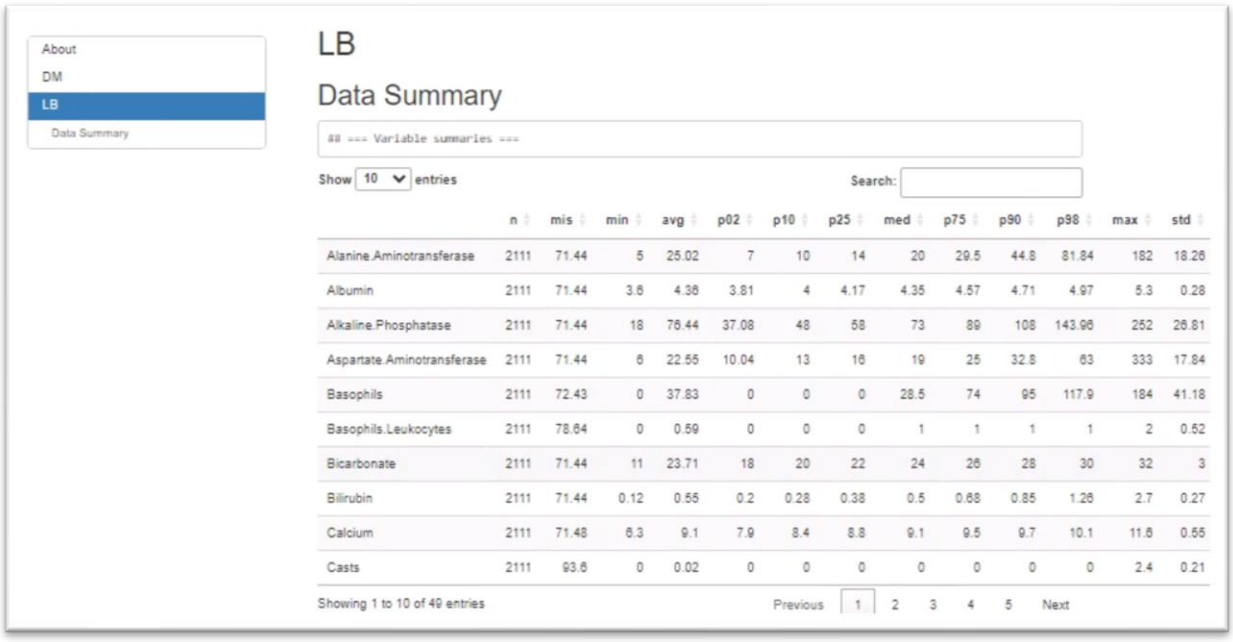


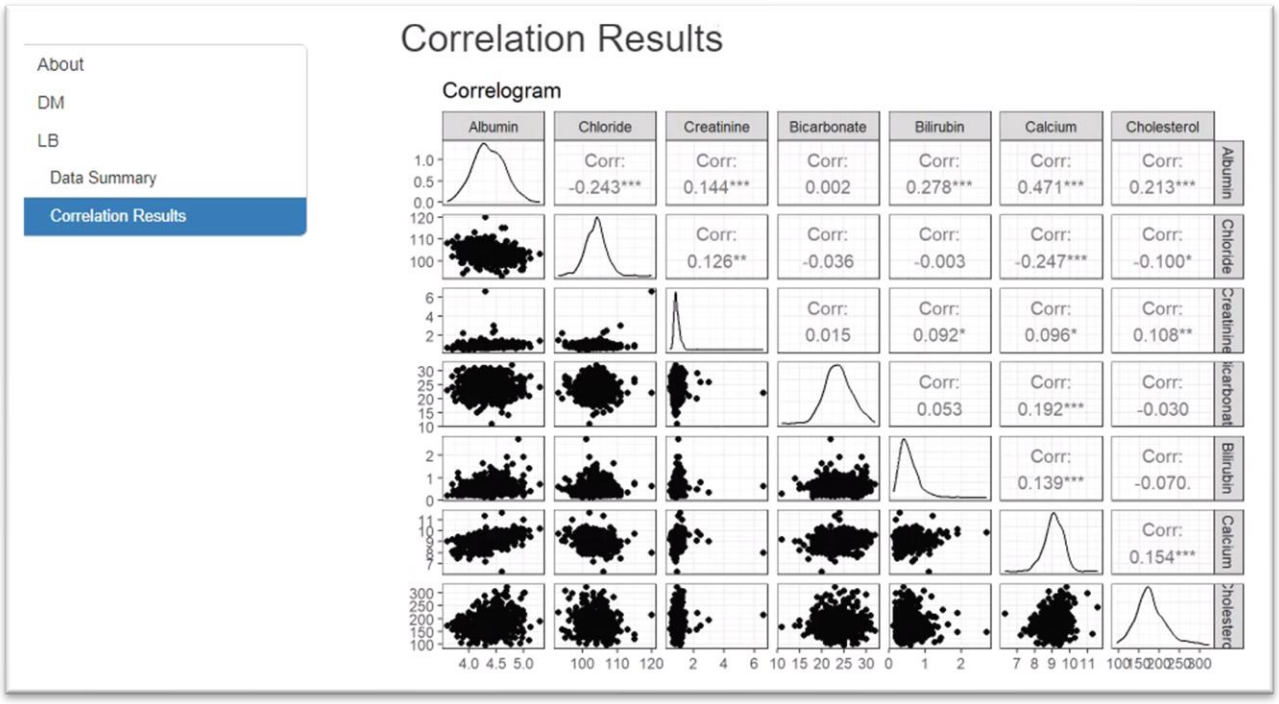
Figure 5: Distribution Plot 2

Figure 6: Data Summary Table



	n	mis	min	avg	p02	p10	p25	med	p75	p90	p98	max	std
Alanine.Aminotransferase	2111	71.44	5	25.02	7	10	14	20	29.5	44.8	81.84	182	18.26
Albumin	2111	71.44	3.6	4.36	3.81	4	4.17	4.35	4.57	4.71	4.97	5.3	0.28
Alkaline.Phosphatase	2111	71.44	18	76.44	37.08	48	58	73	89	108	143.96	252	26.81
Aspartate.Aminotransferase	2111	71.44	6	22.55	10.04	13	16	19	25	32.8	63	333	17.84
Basophils	2111	72.43	0	37.83	0	0	0	28.5	74	95	117.9	184	41.18
Basophils.Leukocytes	2111	78.64	0	0.59	0	0	0	1	1	1	1	2	0.52
Bicarbonate	2111	71.44	11	23.71	18	20	22	24	26	28	30	32	3
Bilirubin	2111	71.44	0.12	0.55	0.2	0.28	0.38	0.5	0.68	0.85	1.26	2.7	0.27
Calcium	2111	71.48	6.3	9.1	7.9	8.4	8.8	9.1	9.5	9.7	10.1	11.6	0.55
Casts	2111	93.6	0	0.02	0	0	0	0	0	0	0	2.4	0.21

Figure 7: Correlation Results



INTEGRATION OF RMARKDOWN REPORT INTO RSHINY DASHBOARD

Integrating Rmarkdown reports into an R Shiny dashboard enables dynamic and interactive reporting. Using the render function in Shiny, the report generated in Rmarkdown can be embedded within the Shiny app, allowing users to generate and view customized reports based on their inputs and selections in the dashboard interface. This integration enhances the dashboard's functionality by providing users with detailed, customized reports that update dynamically based on their interactions with the Shiny app. Combine R Markdown with the flexdashboard package to quickly assemble R components into administrative dashboards.

CONCLUSION

Automating the generation of exploratory reports through the integration of R Shiny and R Markdown marks a significant leap in efficiency and productivity. This integration makes it possible to create dynamic, data-driven reports with flexibility that not only accelerates the reporting process but also empower users to explore data insights in greater detail.

This automated fusion meets a variety of user requirements by utilizing R Markdown's adaptable report layout and R Shiny's interactive features. It provides the freedom to quickly produce thorough reports, interactively examine data, and alter graphics.

The integration's ability to facilitate real-time adjustments and updates ensures that reports remain relevant in rapidly evolving scenarios Moreover, the user-friendly interface lowers the entrance barrier for non-technical users, enabling them to gain insightful knowledge without having a deep understanding of code. Overall, the automation of exploratory report generation using R Shiny and R Markdown is a revolutionary method that improves accessibility, effectiveness, and the breadth of insights obtained from data research.

REFERENCES

http://rmarkdown.rstudio.com
https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf

ACKNOWLEDGMENTS

We would like to express our gratitude to Catalyst Clinical Research team for their invaluable support and encouraging in the development of the research work. Our heartfelt thanks Limna Salim for providing us with guidance and feedback throughout the research work. And finally, Our families and friends for their unwavering support and encouragement throughout the research work.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Denny Sebastian
Company: Catalyst Clinical Research
Address: Phase II, Infopark, Kochi, Kerala, India
Work Phone: +91 9539693707
Email: denny.sebastian@catalystcr.com
Website: https://catalystcr.com/

Author Name: Vikash Jain
Company: Catalyst Clinical Research
Address: Hillsborough, New Jersey, USA
Work Phone: +1 (703) 577-0496
Email: vikash.jain@catalystcr.com
Website: https://catalystcr.com/