Data visualization in healthcare with analyzing medical cost using R programming

Sajib Ghosh Md. Nazmul Islam

Department of Computer Science Department of Computer Science

American International University-Bangladesh American International University-Bangladesh

Dhaka, Bangladesh Dhaka, Bangladesh

sajibg563@gmail.com nazmulislamdip@gmail.com

Akash Kumar Ghosh Punam Das

Department of Computer Science Department of Computer Science

American International University-Bangladesh American International University-Bangladesh

Dhaka, Bangladesh Dhaka, Bangladesh

akashghosh12300@gmail.com punamdas@gmail.com

**Abstract:** This paper delves into the application of data visualization techniques in analyzing datasets related to medical costs. By utilizing the R programming. A variety of visualization techniques are employed to uncover patterns, trends, and relationships within the data. The results provide valuable insight into the factors that impact medical costs and showcase the effects of visualization in healthcare research.

**Keywords**—**Data visualization, R programming, Healthcare, Medical cost.**

**I. INTRODUCTION**

Data visualization is an important part of our research. Our research paper in Data visualization starts with histogram, bar plot, boxplot, scatterplot, violin plot, line graphs: image for our supervisor and presentation for our department. We are working on our thesis paper and preparing. It is essential to summarize our data effectively. Creating figures is a crucial part of a scientist’s career, as good visualization can enhance the impact of your data and presentations. Data visualization is a valuable skill that extends beyond research, and having the knowledge and technical ability to create compelling visualizations is a highly beneficial transfer skill. Many professionals have secured new opportunities, both within and outside academia, by showcasing their proficiency in data analysis and visualization.

The medical cost dataset utilized in this study consists of a thorough compilation of anonymized patient records, encompassing a wide range of demographic,

clinical, and financial attributes. These attributes encompass patient age, gender, region, charge, medical, smoker, and sex. By utilizing this extensive dataset, researchers can access a wealth of information that provides valuable insights into the complex dynamics of healthcare expenditures.

**Data Visualization Using R**

Utilizing R programming for data visualization in this research paper is rooted in its adaptability, versatility, and reliability as statistical computing and graphics software. R provides a wide array of packages and libraries specifically crafted for data visualization, such as ggplot2, plot, among others. These tools offer a diverse range of visualization techniques, from basic plots to advanced interactive graphics, empowering researchers to create visually captivating and informative visualizations tailored to their analytical requirements.

Furthermore, R programming promotes reproducibility and transparency in data analysis, enabling researchers to document their code, share their methodology, and replicate their findings effortlessly. By leveraging R programming for data visualization in this research paper, we can effectively and efficiently analyze medical costs. Through the utilization of R programming, our goal is to extract actionable insights from the medical cost dataset that can guide evidence-based decision-making and contribute to enhancing healthcare delivery and outcomes.

# **Literature Review**

Mohammad Ahmad Alkatib, Amir Talaei-Khoei and Amir Hossen Ghapanchi et al. [1] have discussed about the paper The paper provides an in-depth analysis of the research field of healthcare data analytics, focusing on the utilization of data analytics tools and techniques to improve various aspects of healthcare performance. The authors advocate for the adoption of techniques that facilitate efficient data management, support decision-making processes, and improve patient engagement and outcomes. The paper provides a comprehensive overview of the current state of healthcare data analytics research, highlighting challenges, opportunities, and potential innovations to enhance healthcare performance and efficiency.

Lisa Meloncon and Emily Warner et al. [2] have analysis aptly summarize the paper is This paper provides an integrative literature review on data visualizations, particularly in health and medical contexts, analyzing 25 studies across disciplines. The paper highlights opportunities for TPC researchers to contribute to effective communication practices integrating data visualizations into health communication. By building on existing research and focusing on humanistic and user-centered aims, TPC can lead advancements in this area.

Paul Brennan et al. [3] have discussed about this paper discusses the creation of online courses, such as "R for Biochemists 101," aimed at teaching R to biochemists and molecular biologists. It emphasizes the importance of real-world data examples in learning. The paper serves as a comprehensive guide to leveraging programming languages like R for effective data visualization in scientific research, with practical advice, resources, and examples provided throughout.

Nishita Mehta, Anil Pandit et al. [4] this article's primary goal is to understand the current landscape of Big Data analytics in healthcare, highlighting both its potential and the challenges faced in its adoption. The authors emphasize the need for more quantitative studies and research in this field, especially in developing countries, to harness the full potential of Big Data analytics in healthcare. The review concludes by pointing out the scarcity of information on the real-world use of Big Data analytics in healthcare. It highlights the need for quantitative studies alongside qualitative approaches and emphasizes the necessity for research on Healthcare Big Data analytics in developing countries.

Anjali pant and R.s Rajput et al. [5] The paper discusses the significance of data visualization in research and business activities. It highlights that data serves as raw material and is essential for making informed decisions. The volume of data being generated is increasing rapidly, emphasizing the need for effective visualization techniques. Data visualization simplifies complex data sets into accessible formats, aiding in understanding and decision-making processes.

Nehad A. abudiyab and Abdullah T.Alanazi et al. [6] This review explores the application and benefits of visualization techniques in healthcare settings, focusing on recent studies from 2018 to 2021. It highlights the role of data visualization in facilitating evidence-based medical practice and improving healthcare management systems. The review methodology involved a thorough search of databases such as the SLD portal, PubMed, and Google Scholar. The review underscores the significance of data visualization in healthcare, emphasizing its role in enhancing patient safety, quality of care, and decision-making processes.

Fatemeh Soleimani-Roozbahani, Ali Rajabzadeh Ghatari and Reza Radfar et al. [7] The article discusses the significance of conducting a scient metrics study to analyze the trends, methodologies, and focus areas of research in the field of Big Data in healthcare. It notes the rapid expansion of Big Data analytics in healthcare and its impact on research and healthcare practices.: It emphasizes the importance of information technologies and Big Data in revolutionizing healthcare systems. The increasing availability of healthcare data offers substantial potential to enhance healthcare efficiency, accuracy, and quality. The article discusses the various aspects of healthcare data, including its volume, velocity, variety, veracity, and value.

Sabyasachi Dash, Sushil Kumar Shakyawar, Mohit Sharma and Sandeep Kaushik et al. [8] Details the components and levels of the healthcare system and the transition from traditional paper-based medical records to electronic health records (EHRs). It outlines the benefits and importance of digitizing clinical records. The text seems to provide a foundation for understanding the importance and challenges of big data in healthcare, emphasizing the transition from traditional record-keeping methods to electronic systems like EHRs for better data management and improved healthcare services.

# **Data visualization techniques**

In this research paper we try to find out medical cost personal data set. We completed our project in R programming. Our teamwork mainly tries to find out, Data visualization in health care with analyzing medical cost using R programming. We visualize medical cost data set using graph. Here we display our all graph and showcase how we applied data visualization criteria in our project. We applied some criteria, now we discuss.

**Histogram:** A histogram is one of the most basic visualizations techniques to understand the frequency of the occurrence of value. A histogram is the graphical representation of quantitative data. It shows the distribution of the data by plotting the frequency of occurrences in a range. The value with longer plots signifies that more values are concentrated there.

A graph of age and age

Description automatically generated

**Fig.1 Histogram of Age.**

**Bar plot:** Bar charts are best suited for the visualization of categorical data because they allow you to easily see the difference between feature values by measuring the size(length) of the bars. A bar plot is used to make comparisons among discrete categories. Bar plot based on frequencies to create bar plot two steps required in R first to create a frequency vector and use frequency vector in a bar plot function.

A graph of a person and person

Description automatically generated

**Fig.2 Bar plot of sex**

**Box plot:** Box plots, also known as box and whisker plots, are an easy way to observe the distribution of the data you’re analyzing. The box demonstrates the central 50% of the data, with a line in the middle that shows the median value. In the plot below, we can see that the black dotsitting outside of the boxplot is an outlier.

A diagram of a box plot

Description automatically generated

**Fig.3 Box plot of Bmi**

**Scatter plot:** A scatter plot, also known as a scatter graph or a Scatter chart, is a two-dimensional data visualization that uses dots to represent the values obtained for two different variables - one plotted along the x-axis and the other plotted along the y-axis.

A graph with black dots and blue line

Description automatically generated

**Fig.4 Scatterplot of Bmi vs Charges**

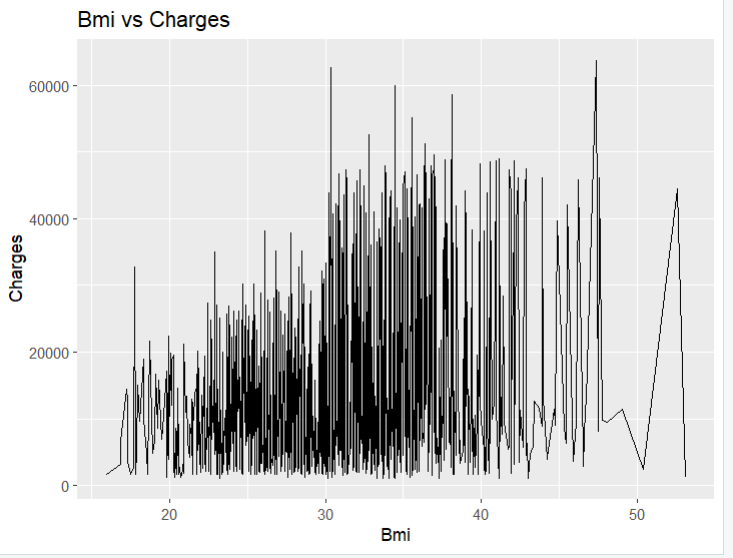
**Violin plot:** Violin Plot is a method to visualize the distribution of numerical data of different variables. A violin plot is a hybrid of a box plot and a kernel density plot, which shows peaks in the data. Violin plots depict summary statistics and the density of each variable. The circle dot represents the median.

A chart of different colored shapes

Description automatically generated

**Fig.5 Violin plot of Region vs Charge**

**Line graph:** Line graphs, also called line charts, are used to represent quantitative data collected over a specific subject and a specific time interval. All the data points are connected by a line. The line graph gives a graphical representation of the changes that had occurred over a given period.



**Fig.6 Line graph of Bmi vs Charges**

# **Discussion**

The data visualization analysis presented in this study provides valuable insights into the patterns and trends of medical costs in healthcare. By utilizing R programming and various visualization techniques, several key findings have emerged. The visualizations reveal significant variations in medical costs across different procedures, patient demographics, and healthcare providers. This variation emphasizes the complexity of healthcare pricing and underscores the necessity for targeted strategies to address cost disparities. The analysis highlights potential cost drivers within the healthcare system, such as the type and complexity of medical procedures, patient age and health status, and geographic location. Understanding these factors is crucial for policymakers and healthcare administrators in developing cost-effective interventions and resource allocation strategies.

# **Conclusion**

R programming and data visualization are essential tools in understanding the complexities of medical costs within the healthcare industry. By utilizing a data-driven approach, stakeholders can effectively collaborate to create a more efficient and cost-effective healthcare system. It is crucial to emphasize the significance of data quality and security in healthcare analytics, as well as the ethical considerations surrounding the use of patient data for cost analysis. Additionally, it is important to recognize the potential limitations of R programming and the necessity for continuous learning in the field of data science. As the landscape of healthcare data visualization continues to evolve, there are endless possibilities for the future of this technology. By staying informed and adapting to new advancements, we can further enhance our understanding of medical costs and improve the overall efficiency of the healthcare system.

# **REFERENCES**

1. Mohammad Ahmad Alkatib, Amir Talaei-Khoei and Amir Hossen Ghapanchi et al. “[Analysis of Research in Healthcare Data Analytics](https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1117&context=acis2015)”.

<https://aisel.aisnet.org/acis2015/118/>

1. Lisa Meloncon and Emily Warner et al. “Data Visualizations: A literature review and opportunities for technical and professional communication”.

<https://www.researchgate.net/publication/319285022_Data_visualizations_A_literature_review_and_opportunities_for_technical_and_professional_communication>

1. Paul Brennan et al. “Data visualization with the R programming language”.

[(PDF) Data visualization with the programming language R (researchgate.net)](https://www.researchgate.net/publication/354993237_Data_visualization_with_the_programming_language_R)

1. Nishita Mehta, Anil Pandit et al. “Concurrence of big data analytics and healthcare: A systematic review”. Volume 114, June 2018, Pages 57-65.

<https://www.sciencedirect.com/science/article/abs/pii/S1386505618302466>

1. Anjali pant and R.s Rajput et al. “Introduction to research data its visualization using R”.

<file:///C:/Users/User1/Downloads/Ch-DV.pdf>

1. Nehad A. abudiyab and Abdullah T.Alanazi et al. “Visualization Techniques in Healthcare Applications: A Narrative Review”.

<https://www.cureus.com/articles/124334-visualization-techniques-in-healthcare-applications-a-narrative-review#!/authors>

1. Fatemeh Soleimani-Roozbahani, Ali Rajabzadeh Ghatari and Reza Radfar et al. “Knowledge discovery from a more than a decade studies on healthcare Big data systems: a scient metrics study”. Article number: 8(2019).

<https://journalofbigdata.springeropen.com/articles/10.1186/s40537-018-0167-y>

1. Sabyasachi Dash, Sushil Kumar Shakyawar, Mohit Sharma and Sandeep Kaushik et al. “Big Data in Healthcare: management, analysis and future prospects”. published: 19 June 2019, Article number: 54(2019).

[https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019 0217-0](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019%200217-0)