Data visualization in Healthcare with Analyzing Medical cost using R programming.

Abstract: In this study, we analyze medical cost datasets applying data visualization methods. by the R programming. Several visualization techniques are used to identify links, patterns, and trends in the data. The results show the benefits of using images in healthcare research and offer insightful information about the variables influencing healthcare expenses.

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 figure 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.

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, plotly, gganimate, and shiny, 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 method:

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: The graph provided is a bar graph since it displays a ratio of individuals in various age groups. "Age" is the label on the x-axis, and it is separated into five categories (20-30, 30-40, 40-50, 50-60, and 60+).The number of individuals in each age group is displayed on the "Frequency" labeled y-axis.

A graph of age and age

Description automatically generated

Fig.1 Histogram of Age.

Bar plot: The number of persons classified by sex who have had sex appears to be displayed on the graph. One data point pertains to ladies, while the other is for males. The fact that the female bar is taller than the male bar suggests that more females than males have had sex, based on the data set.

A graph of a person and person

Description automatically generated

Fig.2 Bar plot of sex

Box plot: The median BMI, which is 30.58 in this instance, is shown by the line in the center of the box. The data set is divided in half by the median, with half of the values being greater and the other half being lower than the median. The data's middle quartiles are shown by the box. Third quartile (Q3), 38.33, is the upper edge of the box. The box's lower edge, 24.93, represents the first quartile (Q1). The interquartile range (IQR), or the variation between Q3 and Q1, is represented by the box. The IQR in this instance is 13.4. This indicates that the median is 13.4 points away from the middle 50% of the data points.

A diagram of a box plot

Description automatically generated

Fig.3 box plot of bmi

Scatter plot: To see a connection between two variables, apply scatter plots. Body Mass Index (BMI) is plotted on the x-axis, while charges are plotted on the y-axis in this scatter plot. Each data point shows the number of charges brought against an individual as well as their BMI.

A graph with black dots and blue line

Description automatically generated

Fig.4 scatterplot of BMI vs Charges

Violin plot: On the graph, there are four data points. Every data point is a circle, and while it looks like they are colored differently, it is hard to tell which colors they are.  
Nonetheless, it is evident that the northeast and southwest regions have the greatest number of charges, respectively. With maybe a few less than in the northeast region, the numbers of charges in the southeast and northwest regions seem to be comparable.

A chart of different colored shapes

Description automatically generated

Fig.5 violin plot region vs charge

Line graph: "Charges" and "Smoker" are labeled on the x- and y-axes respectively. There are more smokes than there are charges, according to the data points themselves.  
It's hard to discern for sure what this chart is attempting to depict without additional context. The labels might not be accurate, or the chart might be showing a situation where there are more smokers than there are charges made against them.  
The number of smokers who have been prosecuted for smoking-related offenses over time may be depicted in the chart.It might be keeping tabs on how many people in a given location receive penalties for smoking offenses relative to the overall number of smokers.

A graph with lines on it

Description automatically generated

Fig.6-line graph of smoker vs charges

Discission: This study's data visualization analysis offers grateful information about the trends and patterns in healthcare spending on medical treatment. Using R programming and other visualization methods, some important conclusions have been drawn. Significant differences in medical expenses between various treatments, patient demographics, and healthcare providers are shown by the visualizations. This variety highlights the need for focused measures to alleviate cost discrepancies and highlights the complexity of healthcare pricing. The study identifies various factors that may influence healthcare system costs, including patient age and health status, geographic location, and the kind and complexity of medical operations. Policymakers and healthcare administrators must comprehend these elements in order to create cost-effective interventions and resource allocation plans.

Conclusion: Understanding the intricacies of medical costs in the healthcare sector requires the use of R programming and data visualization. Partners can effectively collaborate to develop a healthcare system that is both more efficient and cost-effective by using a data-driven strategy. The importance of data security and quality in healthcare analytics, as well as the moral issues related to using patient data for cost analysis, must be emphasized. It's also critical to understand the possible drawbacks of R programming and the ongoing learning process that data scientists must adhere to. There are countless opportunities for this technology's future as the field of healthcare data visualization develops.

Reference:

1) Mohammad Ahmad Alkatib, Amir Talaei-Khoei and Amir Hossen Ghapanchi et al. <https://aisel.aisnet.org/acis2015/118/>

2)Lisa Meloncon and Emily Warner et al. <https://www.researchgate.net/publication/319285022_Data_visualizations_A_literature_review_and_opportunities_for_technical_and_professional_communication>

3)Paul Brennan et al. data visualization with the R programming language. [Data\_visualization\_with\_the\_programming\_language\_R.pdf](file:///C:\Users\User1\Downloads\Data_visualization_with_the_programming_language_R.pdf)

4) 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>

5) Anjali pant and R.s Rajput et al. introduction to research data it’s visualization using R. <file:///C:/Users/User1/Downloads/Ch-DV.pdf>

6) Nehad A. abudiyab and Abdullah T.Alanazi et al. visualization techniques in healthcare application. <https://www.cureus.com/articles/124334-visualization-techniques-in-healthcare-applications-a-narrative-review#!/authors>

7) 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>

8) 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)