Spie Chart Analysis of the Cost of Healthcare for the 60 Years Old Age Group

Keaton Raymond
kraymond@uccs.edu

Abstract

This analysis aims to analyze the differences in healthcare costs between eleven different diseases. It demonstrates the cost to the patient, the cost to insurance, the total cost, and the number of patients affected all in one concise chart. Utilizing this visualization, it is possible to draw conclusions based on the proportional differences between each of the slices. The visualization (fig. 1) shows that Hypertension affects the most patients by a wide margin and that Emphysema has the largest cost per case but affects the least amount of patients, as well as many other conclusions that could potentially be drawn depending on the focus.

Introduction

For this assignment, I was tasked with recreating a spie chart that was given to us using R Base Graphics. A spie chart is a sort of variant on a pie chart that adds more dimensionality to it. A typical pie chart simply shows what percentage of a pie different values take up, while a spie chart allows you to visualize more than that. It allows you to change the radius of the slices to show a second dimension of the data. You are also able to add inner slices that represent a third dimension of the data and include information in the gap between that outer slice and the inner slice, allowing for a fourth dimension. Essentially, using a spie chart, you are able to show in one concise chart what would take four separate pie charts, as well as more analysis compared to what is visually apparent in the spie chart.

Data

The data that was given to us represents the healthcare cost data for eleven different diseases. This data is given for two separate groups, the 60-year-old age group and the 25-year-old age group. For each age group there were five different variables: the full cost, the cost for the patient, the cost for the insurance, the number of patients observed, and the total cost for that disease in that age group, which is calculated by multiplying the number of patients by the full cost of that disease.

The data was already cleaned; however, I decided to drop the columns for the 25-year-old age group because the spie chart is only analyzing the 60-year-old age group. While this wasn't necessary for the analysis, I found it to be more concise for the data frame to only have the data pertaining to the group that was being analyzed.

Chart

To create the chart, I first looked at what visualizations were included with R Base Graphics and found that there was no built-in function for creating a spie chart, meaning I would have to create it manually. Because of this, I visually analyzed the spie chart and looked into the documentation for the pie function [1], which is used to create pie charts, as that is most similar to the spie chart. By using this method, I was able to find that you could create a pie chart slice by slice and modify the "radius" option in order to change the radius of that specific slice, allowing for varying sizes.

The next issue that I ran into was that the chart was scaled up by a large margin, taking up the entire PNG file. This informed me that I would have to scale down the radius so that the graph would fit onto the page. At first, I simply tried dividing the radius by increasingly large numbers, eventually ending on dividing it by around 9,000-10,000. However, I found that that approach gave me strange proportions when I later added the inner slices. So, I considered the data and realized that the full cost, which is the variable that determines the radius of the outer slice, actually represents the area of that slice. With that realization and knowing that the area of a circle is equal to πr^2, it follows that the radius would be equal to $\sqrt{A/\pi}$. I then applied that equation to the radius calculation. Unfortunately, the chart was still too large, so I additionally used my first approach of scaling it down by dividing it by an additional integer value. With this, I found that as opposed to having to divide it by 9,000-10,000, I now only had to divide it by seventy, which maintained seemingly normal proportions when I added the inner slices.

This still left me with the problem of showing the inner slices. Upon thinking deeper about what a spie chart, I realized that it is really just two overlapping pie charts, with each one having slices with a different radius but the same angle. This means that the two pie charts either must have separate color palettes to differentiate the outer and inner slices, or the outer slices have to have a lower opacity, so the inner slices are visible. Looking at the given chart that we are replicating, it seemed that they were the same color, meaning that I would have to lower the opacity of the outer slices. To accomplish this, I found that the pie function also has an argument “col” which is used for changing the color of the slice. So, when looking into how to create a color vector, I stumbled across the “rgb” function [2] which allows you to create a vector for a specific color with given red, green, and blue values. It also has an argument “alpha” which allows you to change the opacity of your color.

The next problem with creating the chart that I ran into was the color palette. The default R color palette only has six colors, meaning that when creating the chart, six of the slices had a color, while the other five were blank because there was no color assigned to them. This meant that I either had to find a new palette that had enough colors, or set custom rgb values. So, I first looked for other color palette packages and the first one recommended was RColorBrewer [3]. I then decided to look up a list of color palettes in the package [4] and found that it had nine palettes with enough colors where I wouldn’t have to use any custom colors. Two of those palettes, “Set3” and “Paired” looked similar to the colors in the given chart. I first tried “Paired”; however, I found that it didn’t look quite the same and put two similar colors next to each other, which I imagine could be useful for something like a bar chart but caused some confusion for a spie chart. I then tried the palette “Set3”, which looked much better and seems like it might be the same color palette that was used for the given chart we are replicating.

Results and Analysis

The four dimensions that are represented in the chart are: the angle of the slice (representing the total number of patients), the entire slice (representing the total cost and is also what is reflected in the label), the darker inner slice (representing the cost to the patient), and the leftover lighter part of the slice (representing the cost to insurance, which can also be viewed as the difference between the total cost and the cost to the patient).

There are many different conclusions that could be drawn from this chart, depending on what you would like to focus on. Of the eleven diseases accounted for, Hypertension affects the largest number of patients and has the highest total cost by a large margin but has an overall average-seeming cost to the customer and insurance. Emphysema, on the other hand, affects the smallest number of patients but sticks out as having the highest cost overall; but, as you look closer, you can also see that, compared to the other ten diseases, the cost to the patient seems to be lower proportionally to the cost to insurance. Through all eleven diseases, you can see that the cost to the patient seems to stay relatively consistent around 20-25% of the total cost.

Conclusion

Overall, I believe that the spie chart is a visualization that should be used more often, as it is a very easy to read and interpret and is an interesting visualization that is able to showcase more dimensions of a dataset compared to a typical single dimensional pie chart. The only downside that I see compared to a pie chart is that you only get specific values for one of the four dimensions, meaning that it is only particularly useful in the case of when you want to see the proportions of the variables between different observations. If you want to see specific values for all four dimensions as opposed to a specific value for one and a general proportion for the other three, but still wanted to use the pie chart format, you would instead need to show four pie charts, one for each of the dimensions. In this case, you would need a pie chart showing the patients for each disease, a pie chart for the total cost of each disease, a pie chart for the patient cost of each disease, and a pie chart for the insurance cost of each disease. For these reasons, a spie chart seems to be the best fit for this data. Potential conclusions and interpretations are clearly visible since the data is laid out in a concise manner. Upon viewing, information like the average cost to the patient, the disease(s) with the highest or lowest costs to the patient, and the disease(s) with the smallest or largest population sizes is clearly apparent. In a case like this where the goal is to analyze the proportions between the diseases, a spie chart is a simple to read yet informative visualization.

References

[1] *Pie: Pie charts*. RDocumentation. (n.d.). Retrieved April 5, 2023, from https://www.rdocumentation.org/packages/graphics/versions/3.6.2/topics/pie

[2] *RGB: RGB Color Specification*. RDocumentation. (n.d.). Retrieved April 5, 2023, from https://www.rdocumentation.org/packages/grDevices/versions/3.6.2/topics/rgb

[3] *RColorBrewer: Colorbrewer palettes*. RDocumentation. (n.d.). Retrieved April 5, 2023, from https://www.rdocumentation.org/packages/RColorBrewer/versions/1.1-3/topics/RColorBrewer

[4] Holtz, Y. (n.d.). *R Color Brewer's palettes*. – the R Graph Gallery. Retrieved April 5, 2023, from https://r-graph-gallery.com/38-rcolorbrewers-palettes.html

Appendix

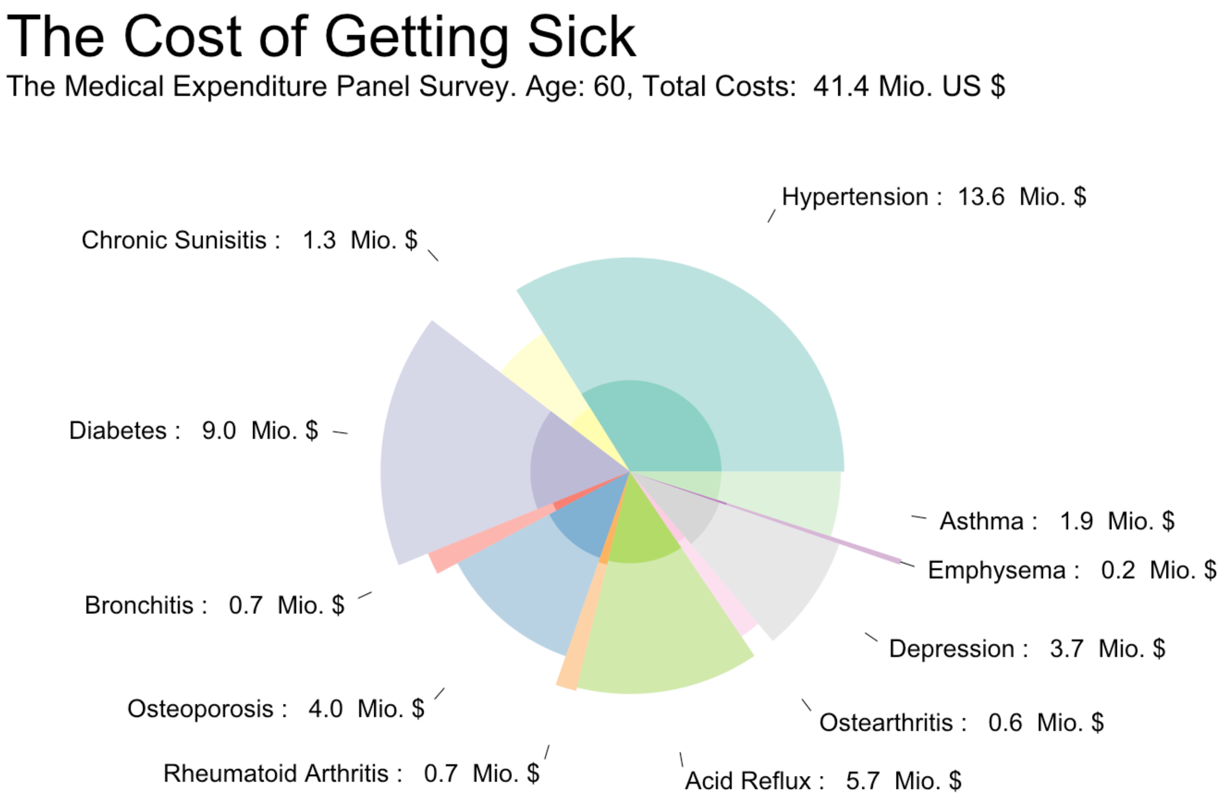

Figure 1: Spie Chart