Kleefstra Syndrome Disease Concept Model

Teaching Fellow - Minh Thu Bui   Supervisor - Professor Masanao Yajima

Wuge Li, Maysen Pagan, Amie Thomas

1   Project Background and Objectives

The objective of this project is to create an interpretable visualization of the results produced by Kristen Connors' Disease Concept Model on Kleefstra Syndrome. The visualization should show which symptoms and impacts were discussed the most during interviews with the healthcare providers and caregivers of children with the condition. The results should indicate which symptoms/impacts are the most significant for the specified age groups in the eyes of the caregiver. It should give insight in to how Kleefstra Syndrome impacts the lives of caregivers and those afflicted by the condition.

2   Visualizations

2.1   Total Frequencies Bar Plot

The most effective method for displaying the total number of references for every concept or impact is by using a bar chart. This visual representation includes four distinct categories: "KS Defining Concepts", "KS Individual Impact", "Caregiver Impact", and "Age", each depicted with varying colors to differentiate between them. Additionally, the data is organized in descending order to enable a clearer comparison of the references in each different categories.

2.1.1   Analysis

From Figure 1, we are able to make some simple observations. The most referenced KS Defining Concept across all age groups was Communication. The most referenced KS Individual Impact was health compared to emotional being the most referenced caregiver impact across all age groups. The Age bar plot allows to observe that in their interviews, caregivers most often referenced their children between the ages of 5 and 11.

2.2   Age Group Symptoms Frequencies Heatmap

Heatmaps are a great way to reveal relationships between variables. By representing numerical values with colors, heatmaps make it easier to identify areas of high and low values. For Figure 2 the frequency of each symptom was calculated by finding the sum of all symptoms mentioned per age group then dividing by the number that the specific symptom was mentioned. This normalizes the data so that the different response sizes are no longer an issue, making the age groups comparable.

2.2.1   Analysis

Interpreting the graph can be done like the following. For the "Impacts on Caregiver" heatmap (bottom right graph) we can see that the emotional impacts were mentioned at a higher frequency than all other impacts for the Under 2 category. Emotional impact continues to be the most talked about for ages 2-4 and 5-11. Then comes second place to social impacts for age category 12-17.

The "Symptoms" heatmap is an a scale from 0.00-0.25, whereas the "Impacts on Caregiver" and "Impacts on Individual" are on a 0.00-1.00 scale. The reasoning is that the symptom most mentioned in interviews had a
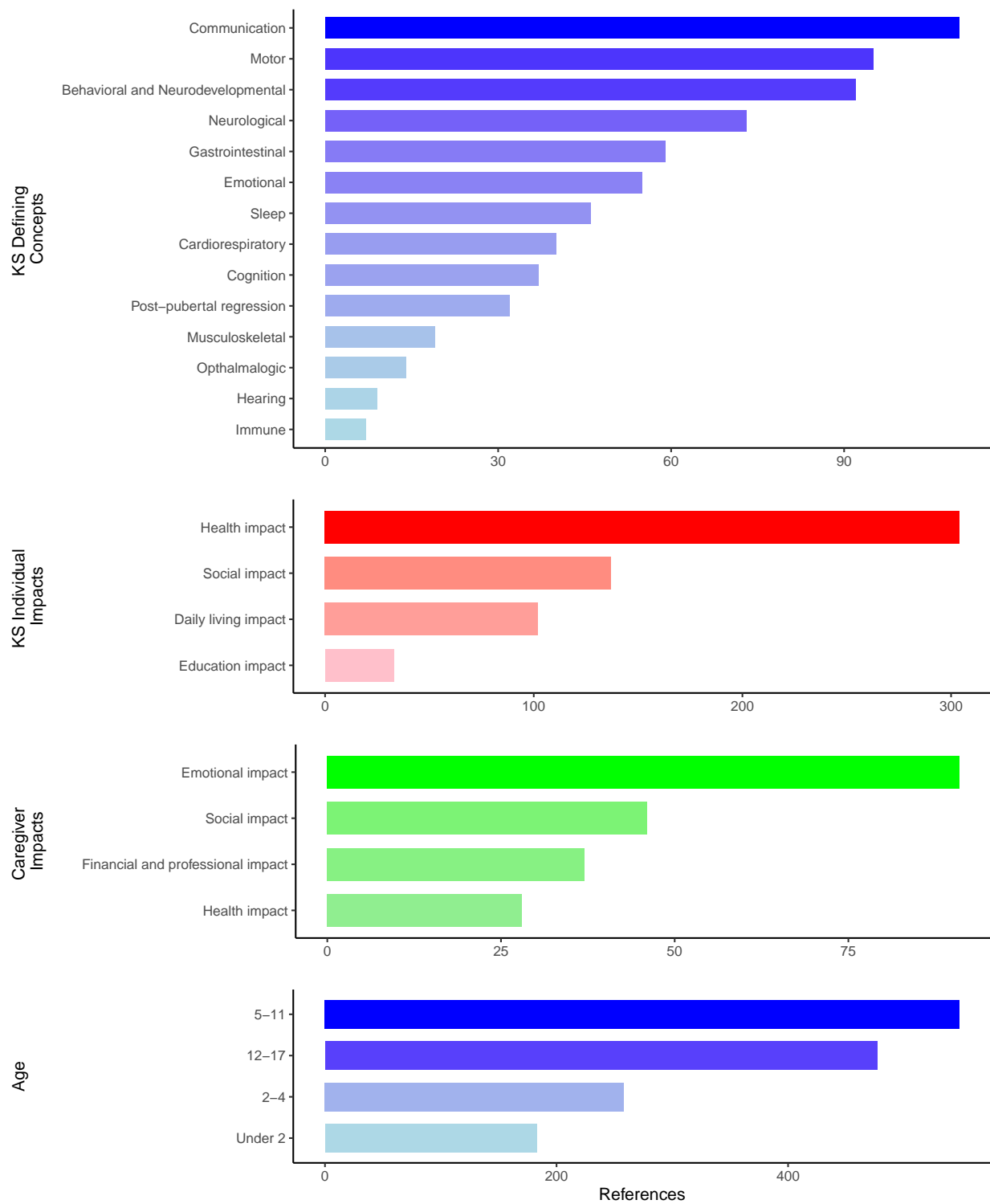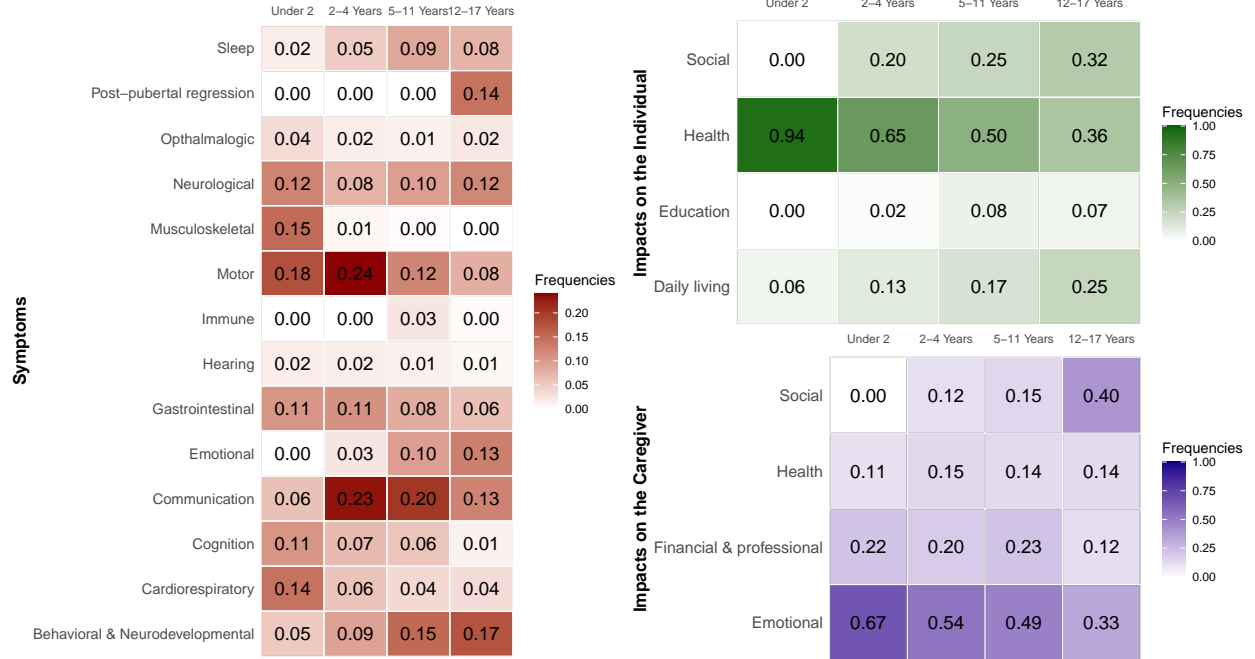
Figure 1: Total references bar chart.

Figure 2: Heatmap of references frequencies.

frequency of 0.24. Putting the heatmap on a 0.00-1.00 scale made it less visually effective for comparison purposes.

2.3 Age Group Frequencies Bar Plot

Another visualization that allows you to observe changes in the frequencies of references over different age groups is a stacked bar plot. Figure 3 shows stacked bar plots for the frequencies of referenced KS Defining Concepts, KS Individual Concepts, and KS Caregiver Impacts. The different colors in each plot represent different concepts or impacts and the labeled frequencies represent the most referenced concept or impact for that age group.

2.3.1 Analysis

These stacked bar plots are a great way to observe changes in the frequency of references across different age groups at a high level. For example, in stacked bar plot B, it is clear to see that health is the most referenced individual impact for children under the age of 2. However, as the age groups increase, the references to health decreases. This suggests that the significance of the child's health decreases as the child gets older when other impacts like social and daily living impacts begin to matter more. A similar pattern can be seen in bar plot C. As the age groups increase in age, the frequencies of references to emotional caregiver impacts decreases while social impacts increase.

There is a challenge with the interpretability of stacked bar plots. While the patterns mentioned above are clear to see looking at the plots, there is a difficulty in comparing concepts or impacts when they do not start at a common baseline. For example in bar plot A, it is hard to compare the references to gastrointestinal concepts between the age groups of less than 2 and 2 to 4 year olds. Additionally, due to the stacked nature of the bar plots, it is also difficult to determine what the exact frequency is of the referenced concept or impact unless it is the bottom bar.
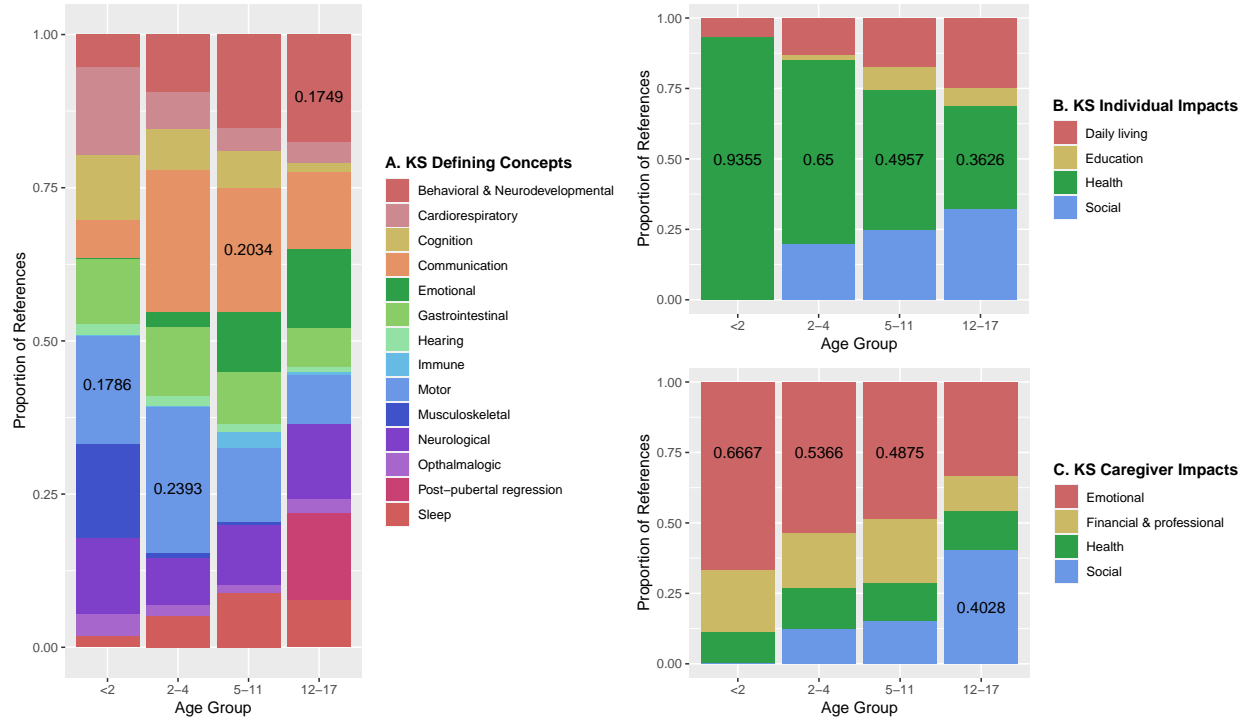
Figure 3: Stacked bar plot of references frequencies. Labeled frequencies represent the most frequent concept or impact for that age group.

3 Conclusion

The horizontal bar plot is a great visualization to observe patterns in the overall count of references to certain concepts, impacts, and age groups. The heatmap and stacked bar plot both are visualizations that allow one to compare the frequencies of references to concepts and impacts across different age groups. Line graphs or area graphs would not be representative of the data as the x-axis for line graphs and area graphs are continuous. The x-axis for the data provided is categorical as the ages of the children are divided into groups. As a result, a heatmap or stacked bar plot is the categorical alternative to area graphs. It is important to note that the cells of the heatmap provide the exact frequencies of each reference, a characteristic that the stacked barplot does not have.