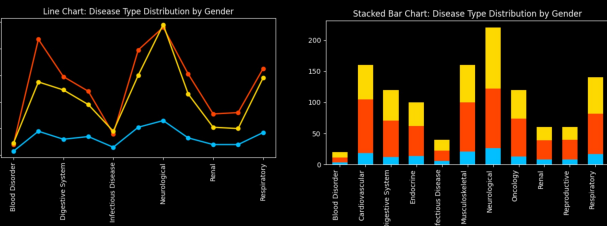


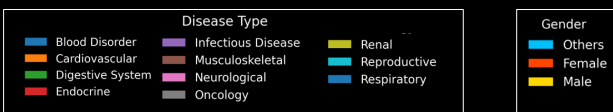
Introduction

In today's world, data plays a significant role in shaping decisions, innovations and powering technologies around us. In this process, data visualization is crucial in helping people make informed decisions. Therefore, misrepresenting data can lead to incorrect conclusions and poor decision-making, making the choice of data visualization crucial. In our study, we examined different types of graphs and relationships especially related to Categorical Data.

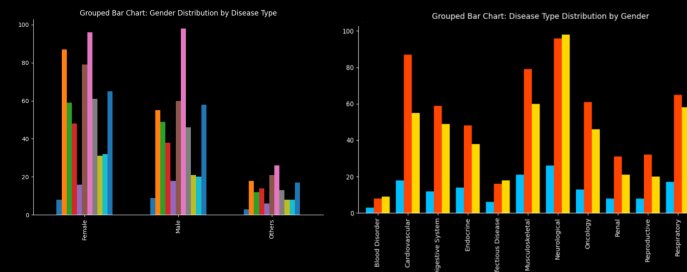
For instance, consider these graphs:



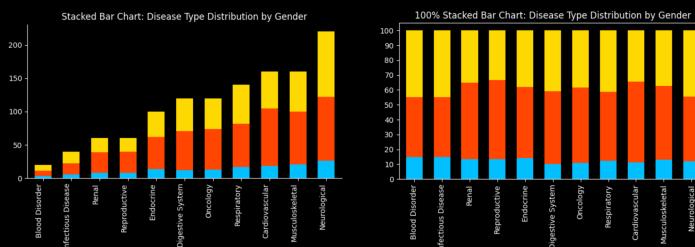
These two graphs illustrate the distribution of diseases across different genders. A line graph can be interpreted differently, potentially misleading viewers into perceiving a progression or pattern, even when none exists. In contrast, a bar graph focuses on individual values, making comparisons between genders more straightforward.



Here, we have an example of a grouped bar chart illustrating the distribution of patients by disease type. In the top graph, having three categories on the x-axis results in a large number of bars for each disease makes it challenging for the to user read. However, when we switch the axis by placing disease type categories on the x-axis, it enhances clarity and makes the data easier to interpret. This demonstrates that using larger number of categories on the x-axis can improve readability and overall presentation.



The two graphs below represent the same data, comparing disease type and gender. The left graph is a stacked bar chart that shows the total number of patients, but it makes it difficult to compare gender proportions across categories. As shown, the 100% stacked bar chart on the right simplifies this comparison, making it easier for the user to interpret the gender proportions within each category. However, 100% stacked bar graph cannot be considered as standalone graph.



Methodology

Our study involves a structured approach to analyze how non domain expert can interpret data visualizations, particularly for categorical data. We studied commonly used graphs such as bar charts, pie charts, stacked bar plots and grouped bar plots and identified different ways of drawing these graphs. We tried to find a rough guidelines for each type of graph specific to the data that the user wants to illustrate.

Conclusion and Future Plan

We have researched ways to develop systematic guidelines for the general audience to select appropriate visualizations for categorical vs. categorical data. Through this process, we recognized the critical role of choosing the right graph type and attributes in data analytics to ensure accurate and effective information presentation.

We focused on categorical data and are now in the process of expanding our research. In the future, we plan to explore visualizations beyond categorical data and investigate ways to enhance their effectiveness.

