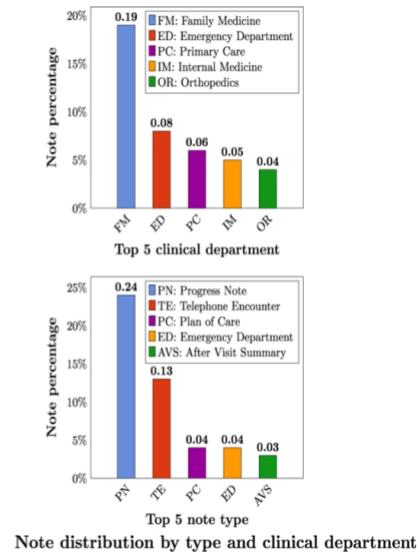
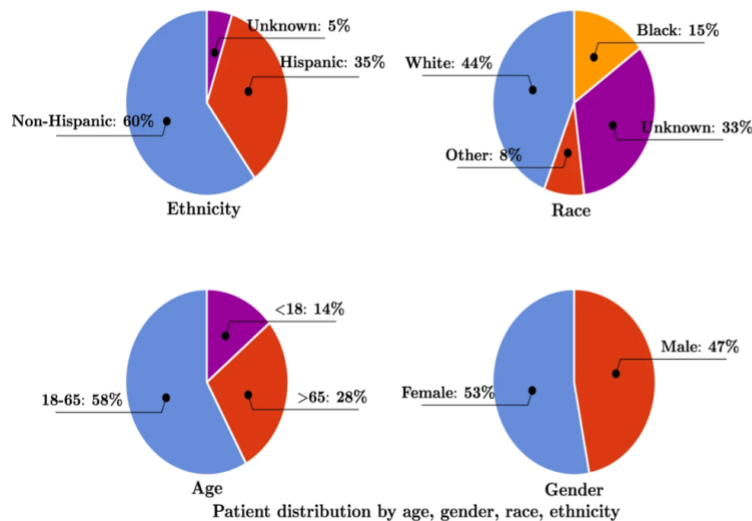


Homework One (1) - MATH 6490

Problem 5

- The article chosen for this work is found [here](#). A large language model for electronic health record.
- The printed article is attached to the submission of this assignment.



Apprehension

The graph shows a viewer the distribution of patients by age, gender, race and ethnicity on the left by pie chart while note distribution by type and clinical department on the right by bar plots. It can easily be noticed by the pie chart that the sizes of each group are compartmentalized and hence larger size of a compartment infer a larger group/distribution and they are also well differentiated by colors. On the right, longer bars tend to have higher distributions as against others with well-defined or differentiating colors coming along with their labels on the plots. A viewer can easily tell by looking at the plot which group or note type are dominant in the dataset and their various percentages boldly printed for clear comparisons. The graph indeed conveys a lot of information about patient dataset just by looking at the plots.

Clarity

The graph is straightforward. The labeling is legibly written and viably readable to the viewer. For instance, by looking at the pie charts, you can tell that non-hispanic, white, 18-65, female are the dominant groups for ethnicity, race, age and gender respectively. As well as, saying, family medicine and progress note dominate the note distribution by clinical department and type respectively. The distinctiveness of the colors also makes the graph easily interpretable.

Consistency

The graph, on the right is a pie chart with compartments as well as on the right, we have bar charts with vertical bars. These graphs are similar charts seen in the past. The way these graphs are interpreted is consistent with how other pie and bar charts are interpreted.

Efficiency

This statistical graph is efficient in summarizing the features of the dataset. It is simple and not bulky yet do a very good job at summarizing the statistics and comparisons of the distribution of the different features. A bar chart with horizontal bars and a dot plot could also be used as an alternative for clinical and note type distribution.

Necessity

The elements in the graph are important as they help in what message the data is sending out to the audience. The labeling communicates the proportions of each group or category in the dataset. The labeling on the x-axis of gives clear description to which bar it represents. But the presenter could have used horizontal bars with the names of the bar on the y-axis representing each group.

Truthfulness

The data is correctly plotted and well represented on the pie chart. The bar chart also represents the various groups and their distribution by length of the bars. However, the presenter could have again had the vertical axis on the horizontal axis and vice versa since the scale chosen do not really or clearly represent the true ratio with respect the entire dataset. An audience, cannot easily tell what those percentages on the vertical axis mean in terms of the entire dataset sine it is not summing up to 100%.

d. Could the graph standalone without the text?

Yes. The graph sends out information about a dataset and it is easily understandable. One will be able to tell or have an idea about the dataset without necessarily have to read any text. The values in the pie chart are labeled outside for legibility to a reader as well as the bar chart have legend to easily refer what each bar represents.

e. Explain how the graph visually depicts the information.

The graph shows a dataset of patients with largest non-Hispanic ethnicity group with whites being the majority. The graphs also explains that 58% of these patients are between the age group 18-65 as well as 53% of the patients being female. The graph also tells the audience that majority of these patients are within the family medicine department and most notes taken from these patients are progress notes amongst the five (5) types of notes taken.

f. The ratio between the height and width of the graph

By the Golden ratio for this graph is 2.0, a statistical graph should be 50% wider than taller. The graph looks good with the width being almost 10cm wider and the height about 5cm. Hence graph was made with the golden section in mind, since the width is 50% wider than the height.

g. Suggest how the graph could be improved.

I would not say there is much to do to improve the graph as the graph conveys a lot of information to the audience. However, my concern would be to change the scale on the y-axis if it is not necessarily representing the true ratio of notes type and department type with respect to the entire data size. Or a probably, the presenter could have indicated what those figures or numbers mean in terms of the data size.

Problem 10

a. Is the GPL an affix of grammar of graphics or more of a statistical

The GPL is an affix of graphical grammar. Wilkinson's GPL focuses on the concept of layers. These levels must be followed in a hierarchy in order to achieve the purpose of producing a statistical graph. Grammatical rules must be followed to build an acceptable or accurate sentence, just as they are in human language syntax. That is how the GPL behaves. Unlike ggplot2, GPL does not use pre-existing data to generate statistical graphics.

b. Is ggplot2 an affix of grammar of graphics or more of a statistical graphics language.

Ggplot2 can be seen either as an affix of grammar of graphics or as a statistical graphics language, depending on one's perspective. Wickham's ggplot2 lacks certain elements found in the Grammar of Graphics (GPL) because these features are already present in the R programming language, which ggplot2 is built on. This is fine if you view ggplot2 as a statistical graphics language. However, if you see it as an extension of the grammar of graphics, it could be considered incomplete since these R features, like the GPL algebra, trans, and data, are not part of ggplot2 itself. Nonetheless, R provides built-in data that can be used with ggplot2 to generate statistical graphics.

c. Which of GPL or ggplot2 are you likely to use. State your reasons

I will go for ggplot2.

Comparing to base R graphics, ggplot2 gives users more control over modifying plot components at a higher level. It offers flexibility by allowing data to be transformed across various scales and layers. Additionally, ggplot2 now has an official extension, making it easier for users to create custom geoms and positions and share them through other packages.

Most conveniently, for my basic knowledge of R, ggplot2 will make it even much easier of a choice to me compared to GPL giving that I am somewhat more familiar with the R programming language itself.