**Unit 4 Assignment: Data Visualization**

Joshua Schwarz

Department of Health Science and Information, Purdue Global

HI560: Health Care Data Analytics

Professor Zhang

March 25, 2025

After exploring hospital data on daily costs, I created several visualizations for everyone to better understand the data I was given. I created a text table with the number of patients and total expense to create a bar graph showing cost per day for May. I then created a bar graph demonstrating the number of patients seen. I also created a statistical graph that shows the May total of all patients' Maximum, Standard Deviation, and Minimum.

In the first graph I made, I represented the daily amount of patients in conjunction with the daily amount of expenditures. I discovered that the different patients and expenditures appeared to rise along with the number of days between the beginning and end of each month. Additionally, I noted that the total expenditures everyday greatly increased on days 3 and 6, and then continued to increase consistently.

The second image provides a bar graph that expresses the daily total cost. The visual revealed to me that there are a few days each month where total expenses rise during the month. Additionally, the expenses seem to rise at a consistent, slow pace in the areas we see the increases in.

The third visual aid is a bar graph of the total number of patients we saw in May. The image shows numerous spikes in patients followed by a sharp decline in patients.

For the fourth graphic, we collected some data on the daily patient influx. At the beginning of the month, we noted a maximum of 45 patients. On average, we saw around 24 patients daily with a standard deviation of 14.7 patients. The minimum number of patients seen in a day was 10.

For enhanced data analytics capability, the dataset should have an organized mechanism to accurately represent entries with dates, numbers, and categorical representation. Consistency and record deletion will speed up appropriate analysis, and the application of normalization processes can convert categorical entries into numeric entries to facilitate better analysis. Selection of a unique identifier will allow quick tracking of a record and referencing to a new dataset. Besides, automated data validation can reduce point of entry errors. A time series format will involve maintaining records in chronological order, and the date of a record can be stored in a standard format for ease and to enable trend analysis. Increased data granularity in collection, such as collecting when costs are separated by cost type, can provide improved cost insight by cost type and cost trends. These factors will greatly improve the data utility for predictive modeling and strategic decision-making. A unique identifier choice will allow quick tracking of a record and joining to a new dataset. Furthermore, utilization of automated data validation can reduce errors in point of entry. The time series format will have the maintenance of records in time order, and the timestamp of a record can be stored in a consistent format for convenience and trend analysis. More detailed data capture, for example, capturing where spending is disaggregated by type of cost, will provide greater detail in insights into costs by type and cost trends. Such characteristics will improve data value substantially compared to predictive models and strategic decision-making.