Exploring Vital Signs and Medications

Yoon Choi, Kelvin Encarnacao, Zhao Zhang

# Introduction

Patient clinical notes are an important piece of information that can be used to analyze health data to solve an important problem. Our objective in this study was to determine whether we can gain insights from the dataset in regard to vital signs/physical exam readings and medication across the dataset as a whole and across the dataset at a patient level. We got the data from Kumar et al. paper, extracted relevant information for the health data analysis, and finally created charts to analyze it.

# Methods

We were given dataset in XML format from Kumar et al. paper titled “Creation of a new longitudinal mix of clinical narratives.” The dataset is a mix of texts and tags. With the data in our hands, we analyzed the free texts of each individual document with the annotation software MAE (Multi-purpose Annotation Environment), and Tag section of our XML contents with Python.

Our annotation process using MAE involved ----.

Our analysis process using Python involved iterative functions that took the XML documents as inputs and returned dictionaries as outputs which were later stored in the csv format.

# 2.1 Question 1:

Explained in detailed by Zhao.

# 2.2 Question 2 - 5:

With the data transformed from XML to CSV format, we began extracting information for question 2-5. In total, we created three functions. This included frequency distribution of the medications taken and a frequency distribution of categories of medications taken, 10 individuals taking the greatest number of medication types, 10 individuals taking the least number of medication types, and 10 individuals taking the least number of medications. Our main function loaded the CSV file to be iterated and called the three functions: question2, question3and4, and question5. These functions were given the first parameters as the data to iterate through, and also given specific dictionaries to store each of their results. The overall steps in each procedure are as below:

1. Create a dictionary for storing the results.
2. Iterate over the dataset.
3. Extract the desired data.
4. If the data is not yet present in our result dictionary, add the key value pair where item is the key and the starting frequency as one.
5. If the data is already in our result dictionary, increment the frequency by one.
6. When we finish traversing through the dataset, we return the dictionary to the main function.
7. Sort the dictionary so that we can easily see that data we want.
8. Save the sorted dictionary in a CSV file (in a folder where only CSV files are located for separation of concerns) for easier analysis and graphs.

# 2.6 Question 6:

Stuff.

Each of the functions that returned dictionaries as outputs were saved to CSV files using the custom python module named csvModule. This allowed for an easier graph creation and more intuitive result of our analysis. The final data analysis was done under Excel.

# Results

# 3.1 Question 1 Result:

Stuff.

# 3.2 Question 2 Result:

## 3.2.1 Frequency Distribution of the Medications Taken

# 3.3 Question 3 Result:

Stuff.

# 3.4 Question 4 Result:

Stuff.

# 3.5 Question 5 Result:

Stuff.

# 3.1 Question 6 Result:

Stuff.

# Limitations

Talk about limitations.

# Conclusions

Talk about conclusions.

# Bibliography

1. Bibliography here.