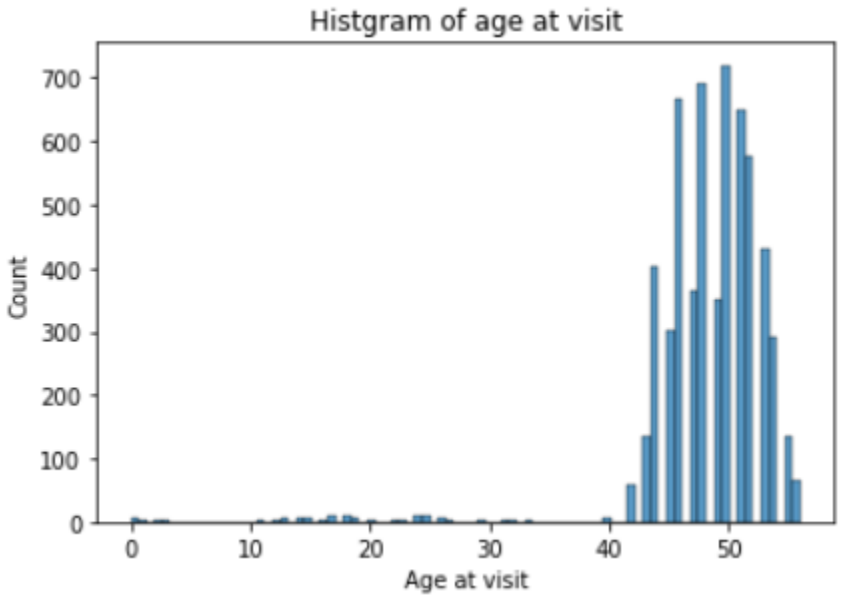
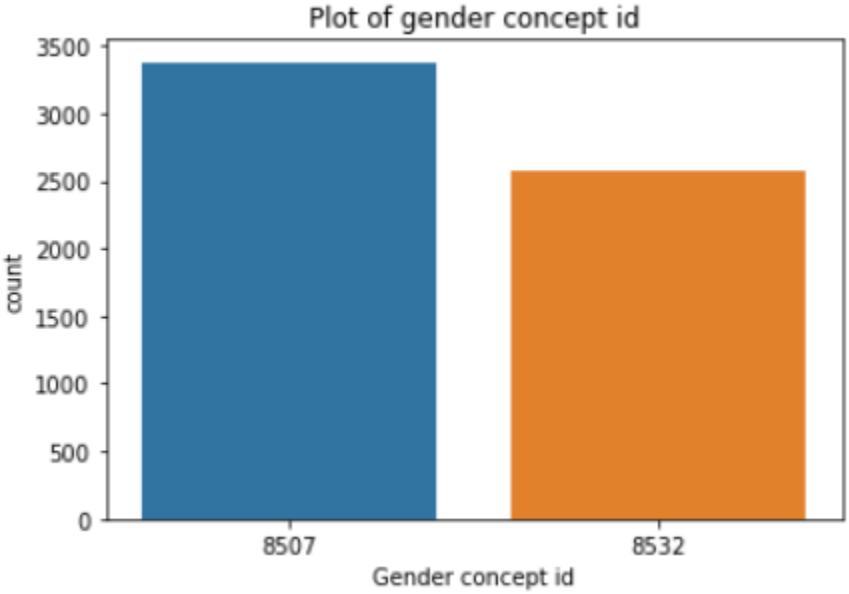
Peng Shen

Visualizing Descriptive

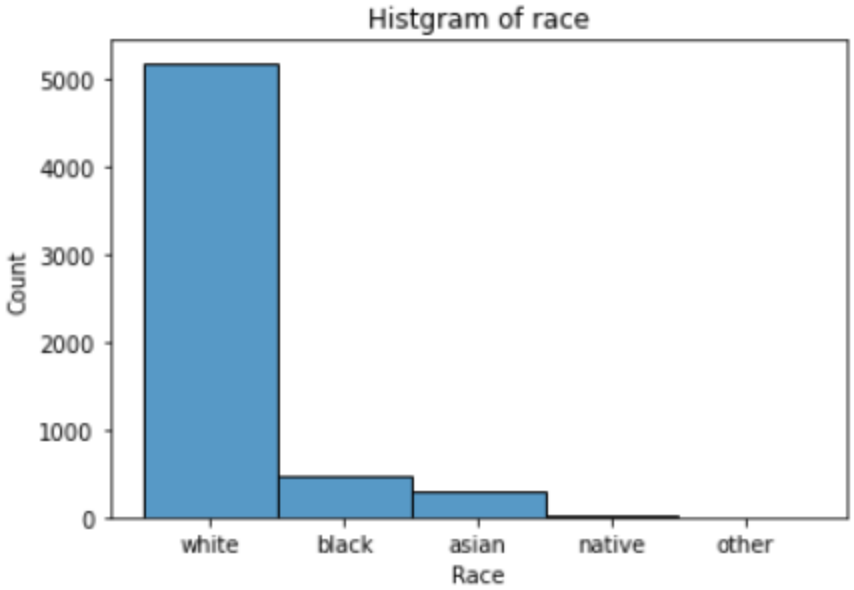
Part 1.



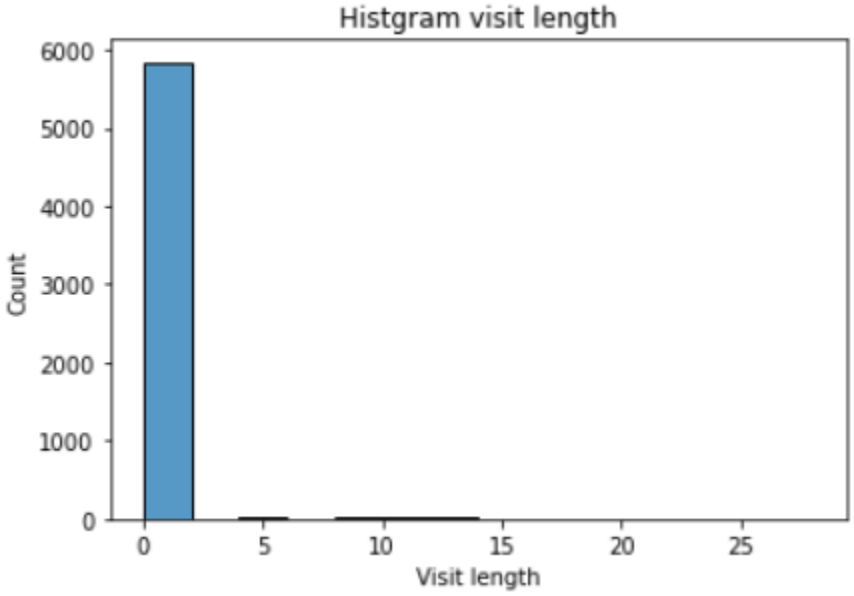
From the histogram of age at visit, most patients are visiting at the age between 40 to 56.



From the plot of gender concept id, there are more male patients than female patients.

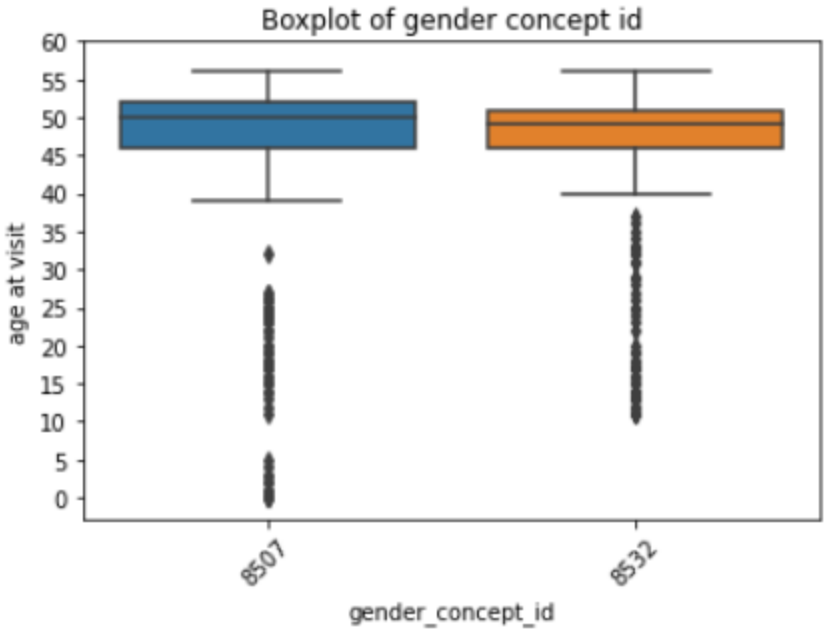


From the histogram of race, most of the patients are white and very few are black, asian or native and other.

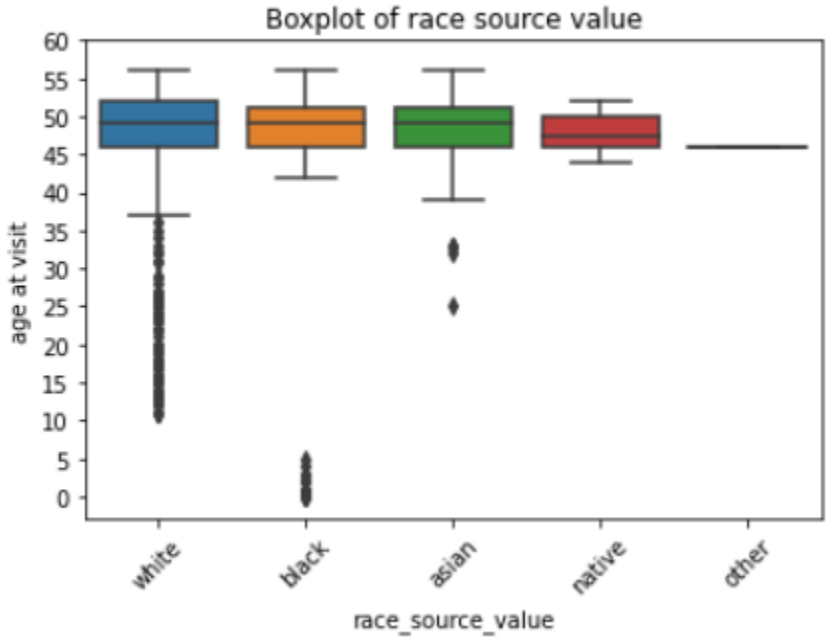


From the histogram of visit length, most patients are not staying at the facilities for more than 1 day.

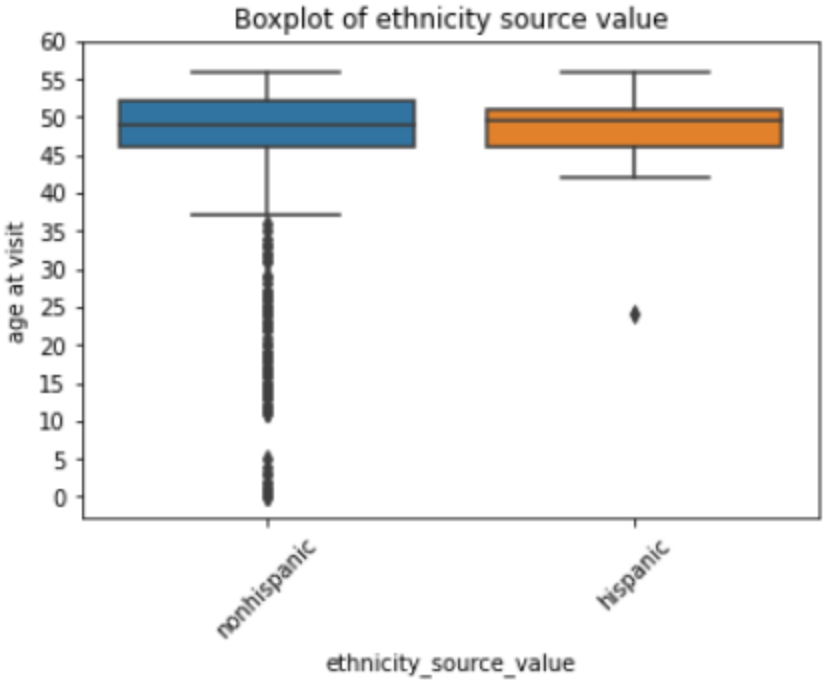
Part 2.



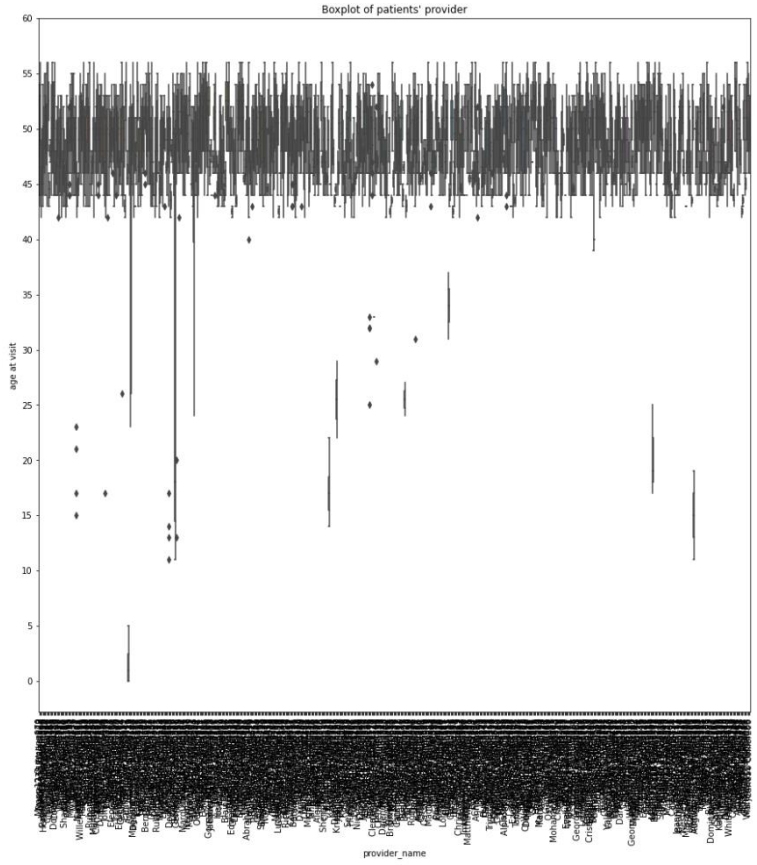
From the boxplot, the age of visit for male and female have similar distribution. But the minimum age at visit is smaller for male than for female.



From the boxplot, the ages of visit for all races have similar distribution, all with mean around 50.



From the boxplot, the ages of visit for both ethnicities are similar.



There are too many providers for patients. Not much information can be obtained from the boxplot above.

Short Answer Questions

Part 1.

The concept table contains definition for the concept ids, including information on what is measured, the name of the facilities, the type of drug that are using, which concept id is female, and which is male, etc. It helps to definition what actions are conducted during a patient's visit.

The person table contains information of the patients' demography, such as gender, birthday, race, etc. and it has a unique identifier for patient called 'person_id'.

The measurement table contains the number of the measurements for every patient's visit along with the measurement's unit. The information on what is being measured can be retrieved from joining with the concept table.

Part 2.

One question that can be answered from the data frame and visualizations is to predict the possibility of patient diagnosed with high systolic blood pressure.

The data contain demographic information and the number of visits, conditions of hypertension and diabetes. One machine learning method to predict such probability is gradient boosting tree such as XGBoost. As a boosting machine learning algorithm, it can learn from previously constructed tree and it can handle classification and predication of possibilities all at once.