Project\_1\_Group6

The goal of our project is to explore the diabetes data set from The National Institute of Diabetes and Digestive and Kidney disease and answer several questions related to diabetes factors. In our discovery we will focus on the following aspects:

1. Is there a correlation between the number of pregnancies and Diabetes? Using statistical analysis, we will determine whether there is a correlation between the number of pregnancies and diabetes. In other words, does the higher number of pregnancies mean there is an increased risk of diabetes. – Correlation plot
2. What are the most common factors associated with diabetes? We suspect that BMI, Age, and Genetic disposition are strongly associated and a future indicator that a person will have diabetes. In answering this question, we can hypothesize what preventive measures may help the patient. – 3 separate scatter plots, and one combined plot with all 3 factors
3. Which is the strongest single factor? We will look at each individual factor and determine which has the strongest correlation. – bar graph
4. What is the average age of Diabetic individuals? Understanding the average age will help us determine age distribution and determine realistic plans of care and appropriate interventions. For instance, if the patient is older, they may be in a different living situation and support environment compared to an individual in their 20s or 30s. – scatter plot

By the end of the project, we hope to have a better understanding of the risk factors associated with Diabetes and spread public health awareness.

1. Introduction slide
   1. Points for context
      1. All data points are women
   2. Present hypothesis
   3. Survey the class of how many people they know with diabetes
2. What are the most common factors associated with DM?
   1. 3 scatter plots, and one combined – will need regression line
3. What is the average age of Diabetic individuals?
   1. Scatter plot – no regression line
4. Is there a correlation between the number of pregnancies and Diabetes?
   1. Scatter plot- with regression
5. Which is the strongest single factor?
   1. Bar graph