Team 6 Rationale- Nutrition Study

1. Motivation

   The nutrition study lays out a variety of health factors concerning the concentration of plasma in the patients. We would like to look at these factors and see which ones have the largest effect on plasma concentration (Beta & Retinol). Does the effect of all the factors vary by age? By gender? What other connections are there between the variables? Do people who drink tend to smoke as well? What lifestyles lead to higher cholesterol and BMI? Can connections like that be used to identify lifestyles that promote a high/low plasma concentration?

2. What we expect to learn

   First, we would expect to answer the overarching question regarding plasma concentration. We would also like to see the relationships between the health factors. We also hope to learn if certain lifestyles put one at a greater risk.

3. Methods

   We would start with using ggplot to do exploratory data visualization, get a feel of what the data looks like, what trends and correlations are easy to see. After, we would have to use statistical methods to see the connections in the data. The data set we are working with has a low sample size of only 315 people, so we'll probably have to end up extrapolating information out of the data with linear regression or other fits, maybe comparing distributions together, not just raw points.

4. Potential issues

   The data set has a low sample size as previously stated, seeing the relationship between the variables might be difficult because of this. Also, there are a lot of variables in the data set, there might be complicated connections that take a lot of work to find.

5. Collaboration

   We plan to collaborate through Teams, in addition to Google Docs. We will distribute the workload within R, such as plots and statistical methods. When making the video, we will practice in Zoom before the presentation. We already meet weekly for the modules, so we will check on the project during the weekly meetings and classes.