Data Visualization Project 1 Proposal
Group: Max Demers, Ryan Li, Ethan Reeder

Document the exact data source you will use, what pre-processing will be needed, and how you will visualize the data. Use diagrams or even hand-made drawings to illustrate your intentions in a concrete way, so we can provide useful feedback.

Source: https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016

We want to investigate the relationship between GDP per capita and suicide rate. We plan to visualize suicide rates by country over time and as a function of GDP per capita. We will plot year on the x axis, GDP per capita on the Y axis, and each point will be a circle representing a country, with the radius of the circle proportional to suicide rate in that given year. We have provided a drawing as a reference for a basic representation of our concept.

Preprocessing will involve grouping data and removing incomplete relations. We will group data by country in each year, yielding a total suicide rate and GDP per capita per year per country. We may choose to bucket years or choose only select years to most effectively show change. This decision will be made once preprocessing is completed and we have a better sense for the data. Because the dataset includes data on the vast majority of countries each country will have a circle for each year, we will also choose a low integer number of countries with which an average American viewer is most familiar, e.g. US, China, Russia.

We believe that size of each circle will be an effective way to attract the viewer's attention to the most salient element of suicide rate. However, because it is difficult to directly compare surface area between circles of similar sizes, we will include a legend including rough sizes to assist the viewer.

If we find that the visualization is easy to cognize and uncrowded, we may turn each circle into a pie chart breaking down the suicide rate by generation, age, or sex.

Example: