**DSC 530 - Data Visualization**

**Phase 1-A**

*Group Members:*

*Name (Student ID)*

**Rohan Gonjari (01987448)**

**Abhijot Bedi (01908371)**

**Amogh Darekar (01937045)**

**Problem Statement:** Visualizing Olympic performance, by country, from 1960 to 2016 and its correlation with various indirect affecting parameters over the years.

**Description:** We aim to investigate factors involved in the performances of athletes and countries as a whole to pinpoint how individuals and groups can improve. We also plan on looking at any other potential factors (GDP, Life expectancy, etc.) that do not essentially play a direct role in affecting performances. There are many other factors outside of the anatomy of an athlete that might affect how well a country performs at the Olympics. These might include the total population, what income group the country is a part of, how big the economy is, and so on. There is also the potential to figure out whether some popular sports are dominated by certain age groups and genders.

**Data Sources:**

120 years of Olympic history from 1896 to 2016:

<https://maven-datasets.s3.amazonaws.com/Olympic+Athletes/Olympic_Athletes.zip>

GDP of countries from 1960 to 2020:

<https://www.kaggle.com/holoong9291/gdp-of-all-countries19602020>

Life Expectancy from 1950 to 2019:

<https://www.kaggle.com/sansuthi/life-expectancy>