**Mason Walters and Laytoamb Wiggins**

**Data Visualization UT**

**ETL Project:**

Etract: The data sources we used was from Kaggle.com and Data.World.com

The dataset we extracted from Kaggle: 120 years of Olympic history: athletes and results. From the data we pulled athletes, regions and their events.

The dataset we extracted from DataWorld: Total population per county. This is census reports, Eurostat demographics, and UN Population divisions. This data was pulled from an API

All data was extracted from CSV.

Transform:

The athlete events csv had 15 columns which was narrowed down to 7 columns ( team, games, season, city, medal, sport and events).

Original data included 1896 to 2016. Years selected was 1960-2016.

The region data had 3 columns and was reduced to two columns.

The population data from an API was converted to csv

All null or n/a values were dropped from all columns

Load:

The final project was loaded into into MySQL and SQL lite and is uploaded in the repo.

120 years of Olympic History.ipynb (MySQL)

Olympic\_db.sqlite (SQLite)

# 2 120 years of Olympic history: athletes and results

## basic bio data on athletes and medal results from Athens 1896 to Rio 2016

0 years of Olympic history: athletes and results

**basic bio data on athletes and medal results from Athens 1896 to Rio 2016**