**Title**: Olympic Success and How it is Determined

**Group Members**: Jacob Volz, Zebediah Millslagle

**Purpose**:

Describe the relationship between the number of medals won and the GDP per capita of the country that won them. Also look at how this relationship changes with different sports and through time. It is an interesting subject for many Middlebury students considering how involved we are in sports and also provides a good study of how wealth influences opportunity.

**Scientific Question**:

Does wealth influence Olympic success? How does this change across sports and time periods?

**Data Sources**: Data for Olympic medals and who won them from <https://www.kaggle.com/the-guardian/olympic-games>. Contains three csv files:

* dictionary.csv, which contains countries and three letter country codes
* summer.csv, which contains who won what medal at summer Olympics
* winter.csv, which contains who won what medal at summer Olympics

Data for GDP per capita came from <http://data.worldbank.org/indicator/NY.GDP.PCAP.CD>. Contains a csv file with GDP for 263 locations from 1960 to 2016 (some data wasn’t collected for some locations during some years).

Because the Olympic data set and the GDP data set used different country codes, we needed a way to compare them. For this, we used the table from <https://en.wikipedia.org/wiki/Comparison_of_alphabetic_country_codes>, converted it to an html file, and opened it as a csv.

For looking at how different regions compared to others, we used the data from <https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv> to add to columns to our data based on ISO country code: Region and sub-region.

**Data Format**: After organizing the above data into tidy data format, we will have:

* + We have table, the observational unit being an athlete who participated in an event and won a medal (so people like Michael Phelps will be in multiple rows)
  + The table has 30,540 rows
  + The table has 12 columns:
    1. Year – years (int)
    2. City – city name (char)
    3. Sport – sport name (char)
    4. Discipline – discipline name (char)
    5. Athlete – athlete name (char)
    6. Gender – male/female
    7. Event – event name (char)
    8. Medal – type of medal (Gold/Silver/Bronze)
    9. Country – full country name (char)
    10. ISO – country ISO alpha-3 code (char)
    11. Region – name of region (char)
    12. SubRegion – name of sub-region (char)
    13. GDP – GDP per capita, USD (double)