

# Olympic Games

Can Data Predict How Countries will Perform?

# Introduction

- Is it possible to predict how well a country will do in the Olympics?
  - Outside of specific strategies and training for sports, is it possible to tell whether a country will do well or not based on population and economic data?
    - Are more populous countries more likely to win Olympic medals?
    - Are countries with higher GDP (Gross Domestic Product) winning more medals?
    - Are countries with growing population and growing GDP performing better over time?



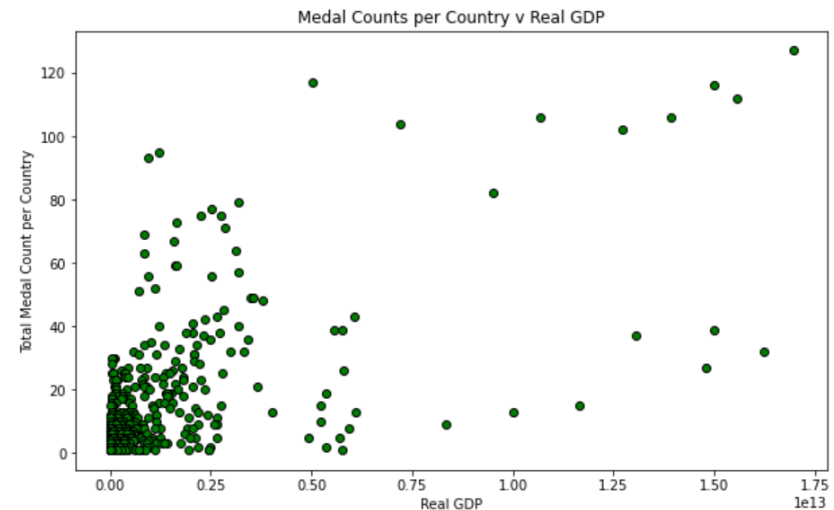
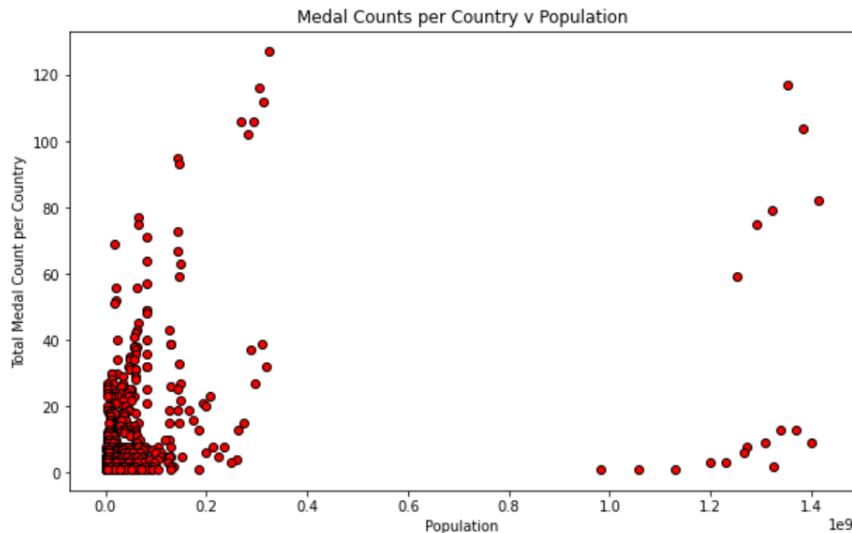
# Data

- Why these data sets?
  - We were looking for data sets that contained: number of medals, country, Olympic year, type of Olympic event, so that we could cross-reference with them with GDP and Population data trying to find trends or correlations.
    - Data Sets Used: Olympic Athlete Event Analysis, Olympic Games Medals 1896-2018, Country Wise GDP from 1994 to 2017 (Kaggle.com)
- Were there any problems with the data?
  - Overlapping information
    - Olympic Athlete Event Analysis was removed
    - Sorting data by athlete was confusing and produced duplicates
  - Discrepancies in country names, countries have changed names or become new independent countries.
    - Russian Olympic Committee, IOC Refugee Olympic team, Great Britain x United Kingdom, etc..
  - Missing total medal counts for event, country, or year.

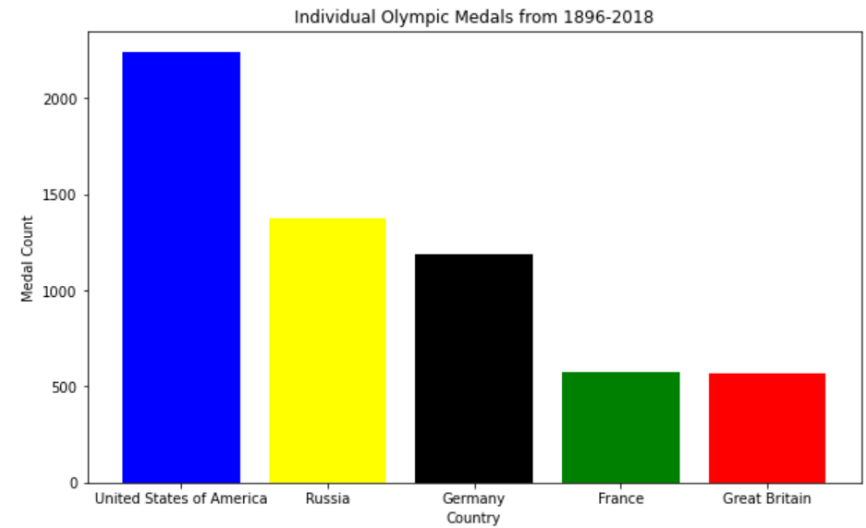
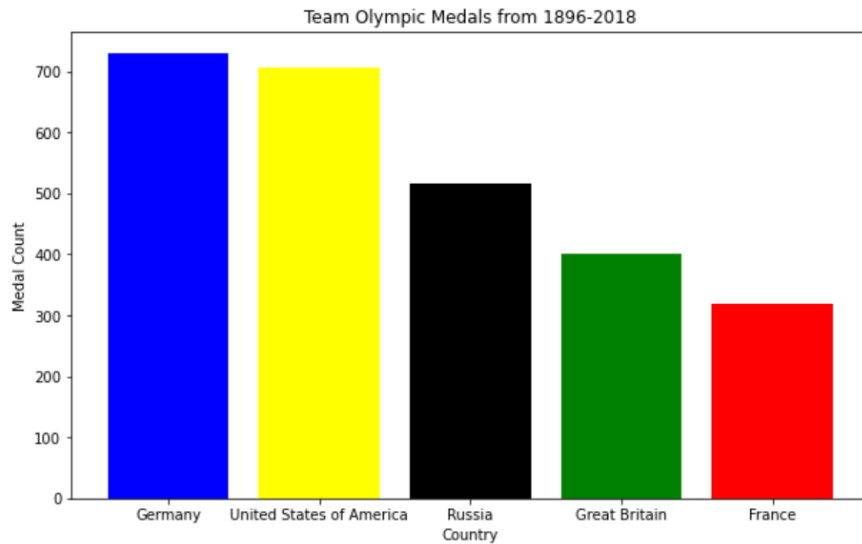


# What story is the data telling?

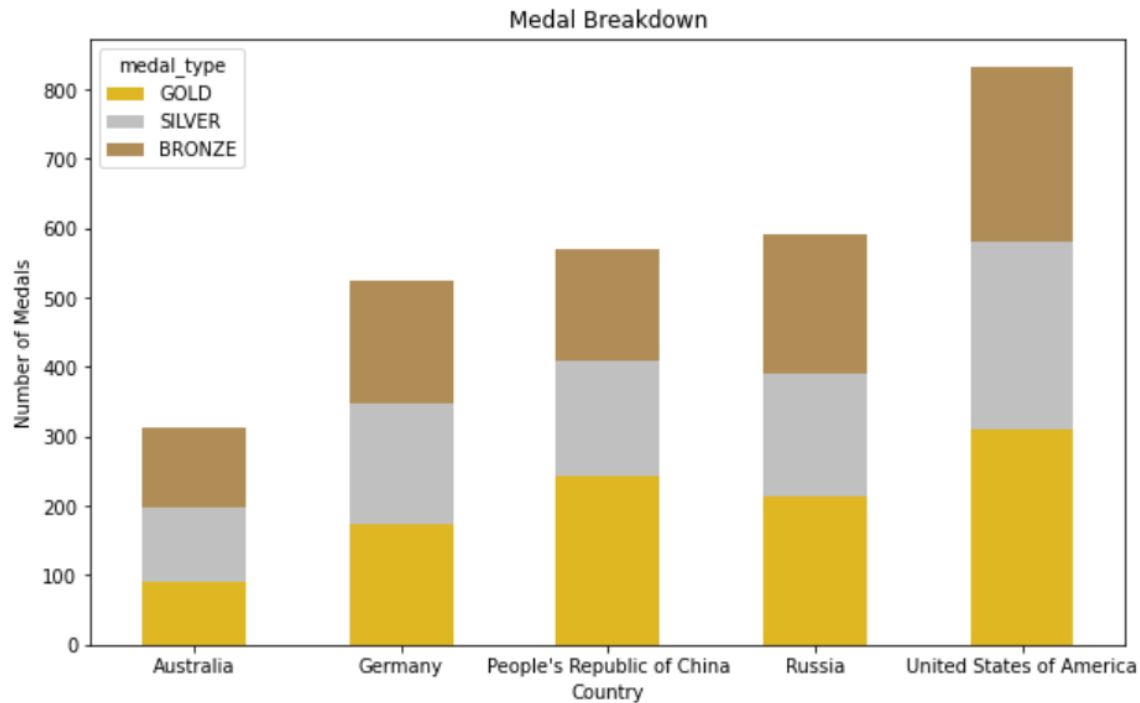
- What initial impressions do we have of the data?
  - The data seems to show that the model works well for countries with small population, but not necessarily for larger populations
- What does the data show?
- What conclusions can we take from the data and its visualizations?



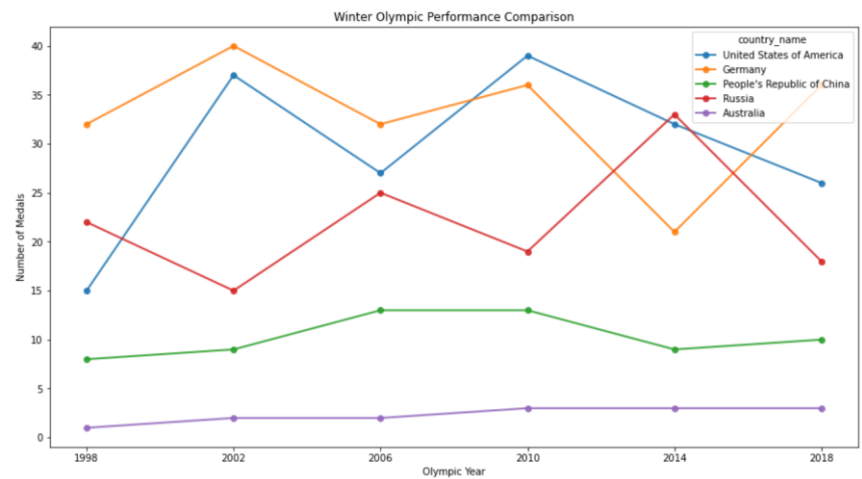
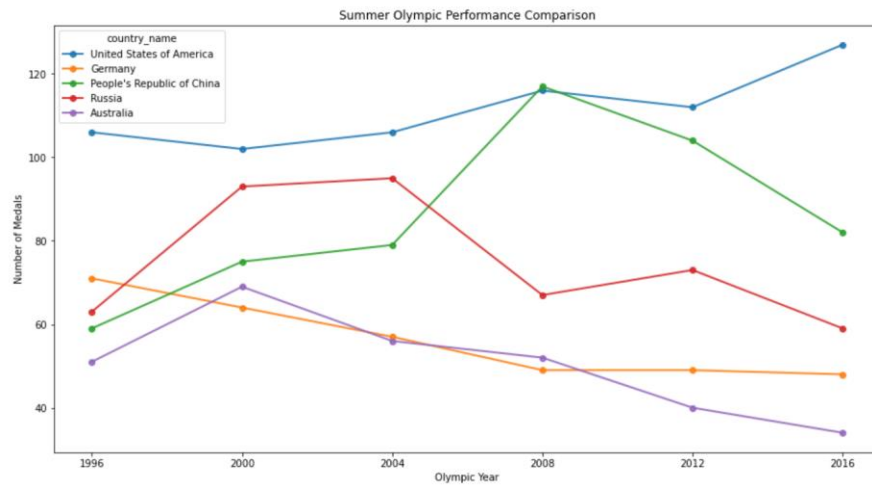
# Top 5 Countries with Total Medal Count from 1896-2018



# Top 5 Countries with Medal Breakdown from 1996-2018 (combined with GDP and Population data)



# Top 5 Countries with Total Medal Count from 1996-2018



# T-test with Population and GDP

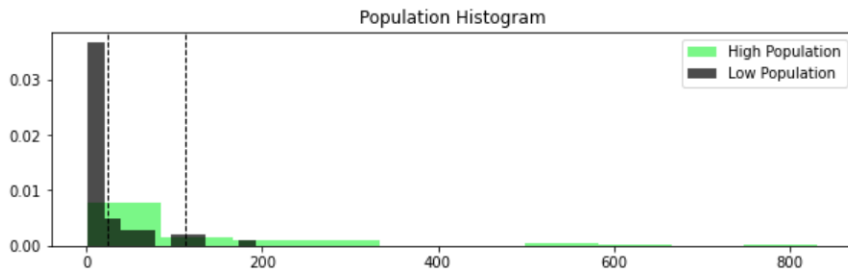
## High Population versus Low Population

### ❖ HYPOTHESIS

These samples came from different distributions - THERE IS a SIGNIFICANT difference in means.

### ❖ NULL HYPOTHESIS

These samples came from the SAME distribution - THERE IS NOT a difference in means.



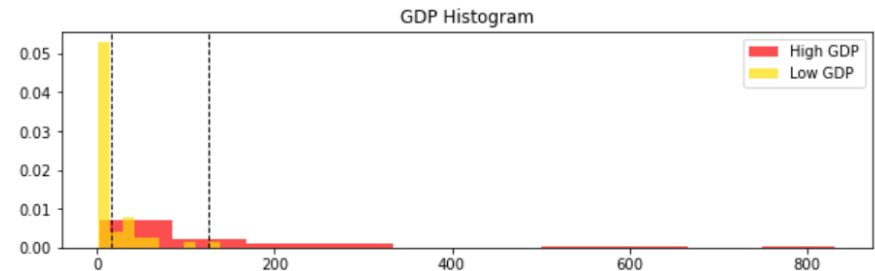
## High GDP versus Low GDP

### ❖ HYPOTHESIS

These samples came from different distributions - THERE IS a SIGNIFICANT difference in means.

### ❖ NULL HYPOTHESIS

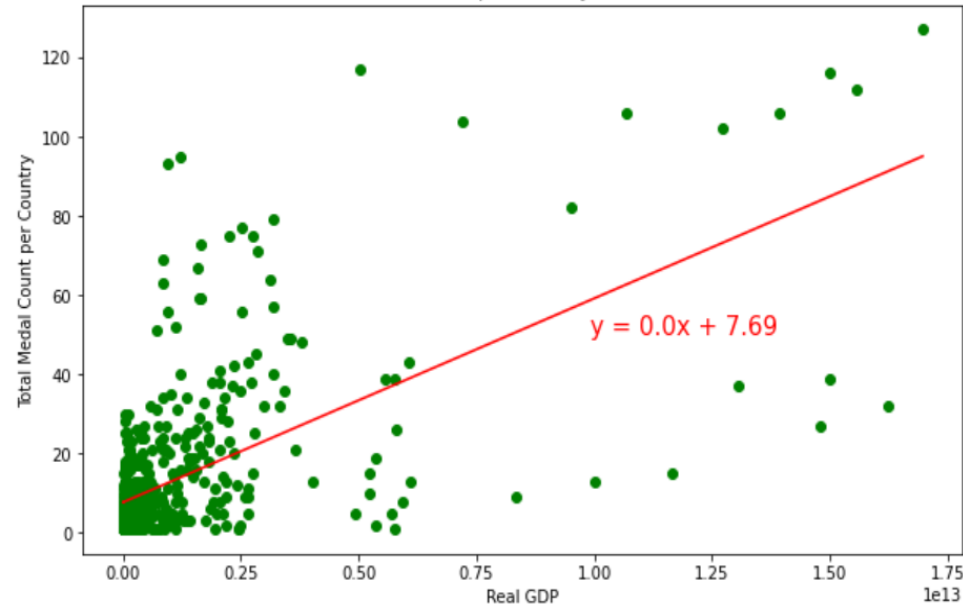
These samples came from the SAME distribution - THERE IS NOT a difference in means.



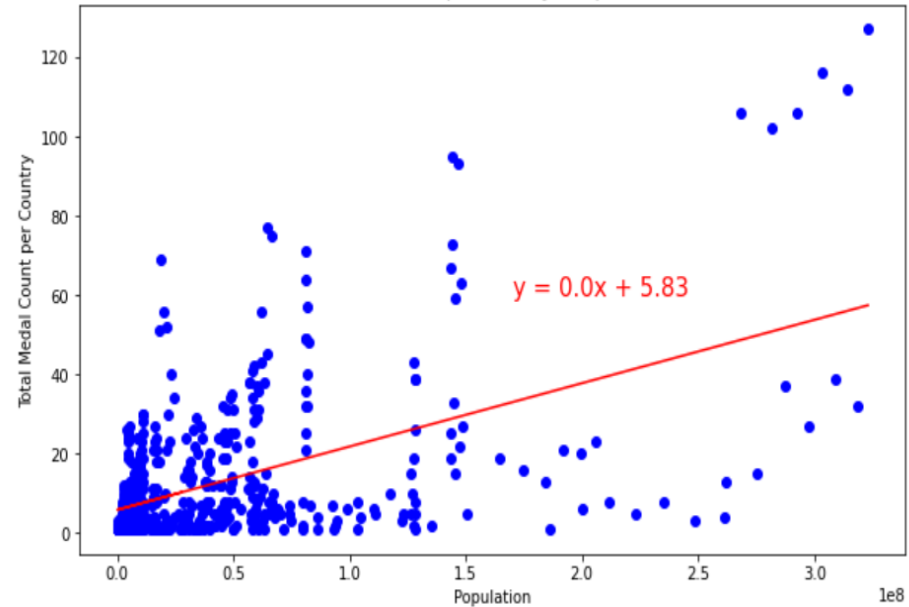


# Regression for GDP and Population

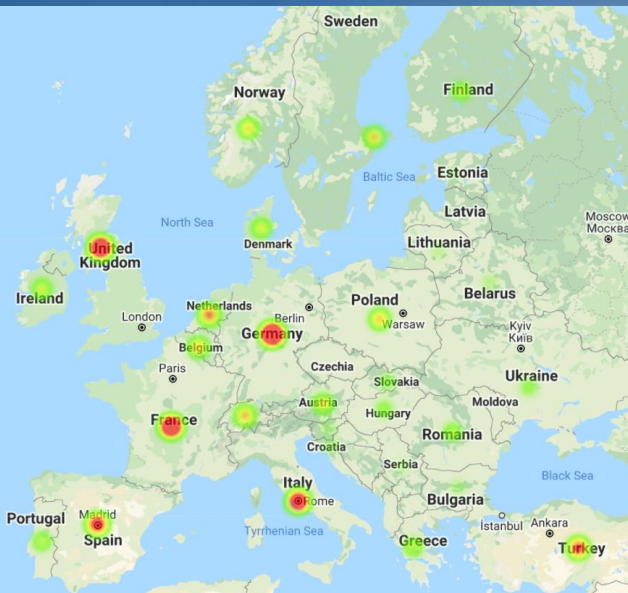
Medal Counts per Country v Real GDP



Medal Counts per Country v Population



# Heatmap Activity in Europe

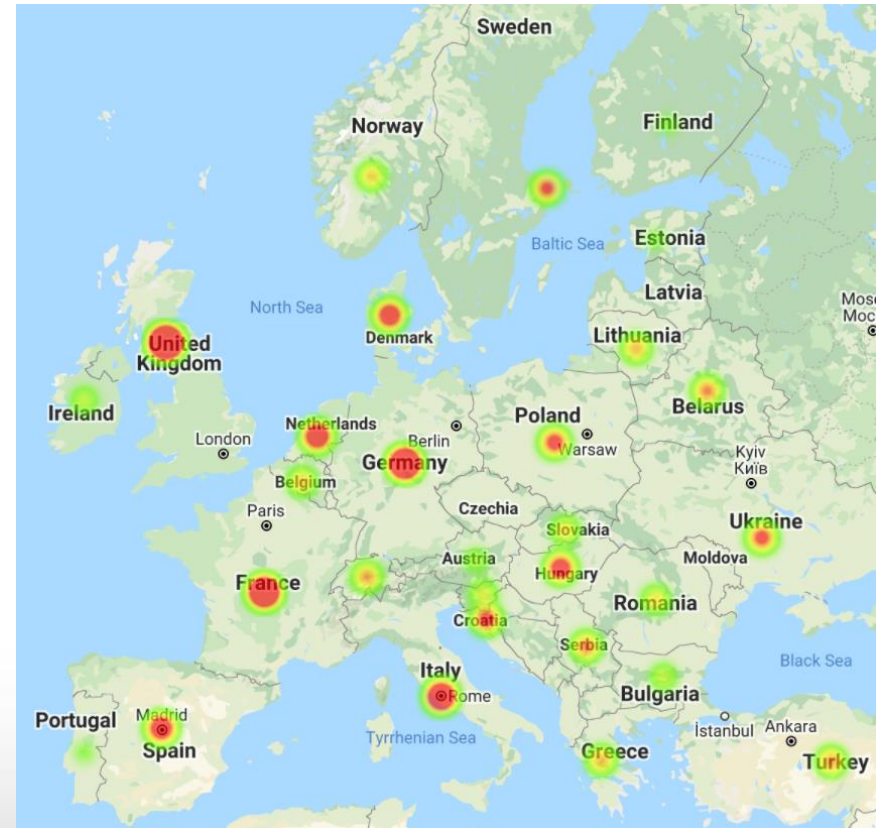


GDP



Population

Medals Won

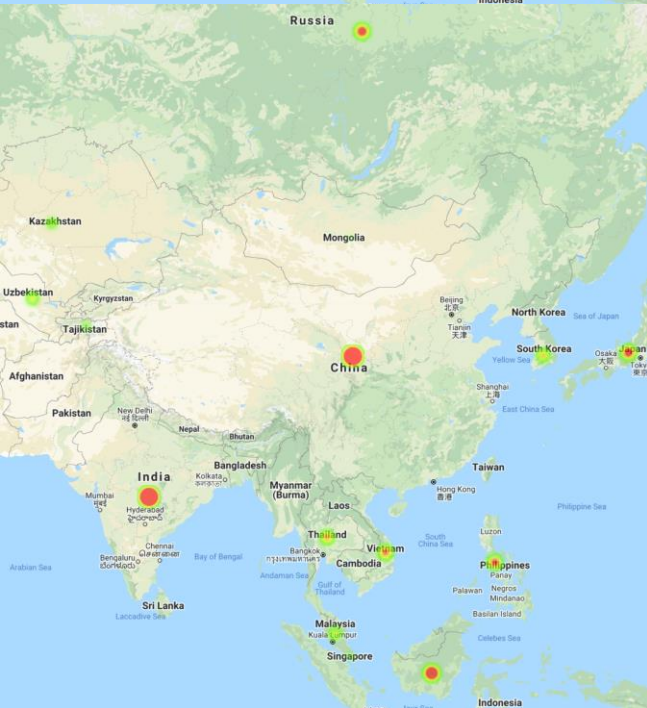
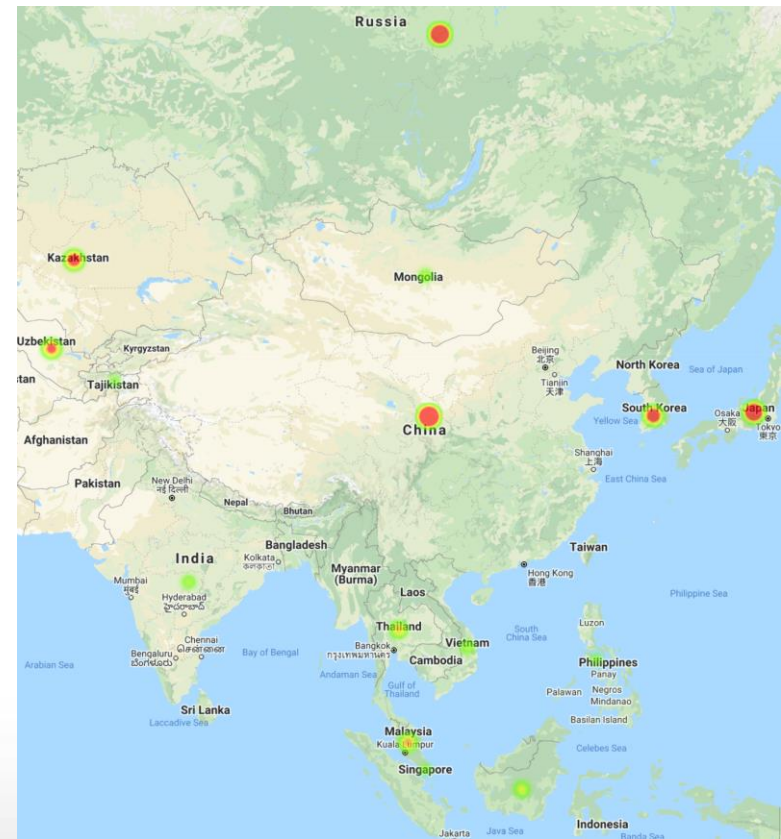
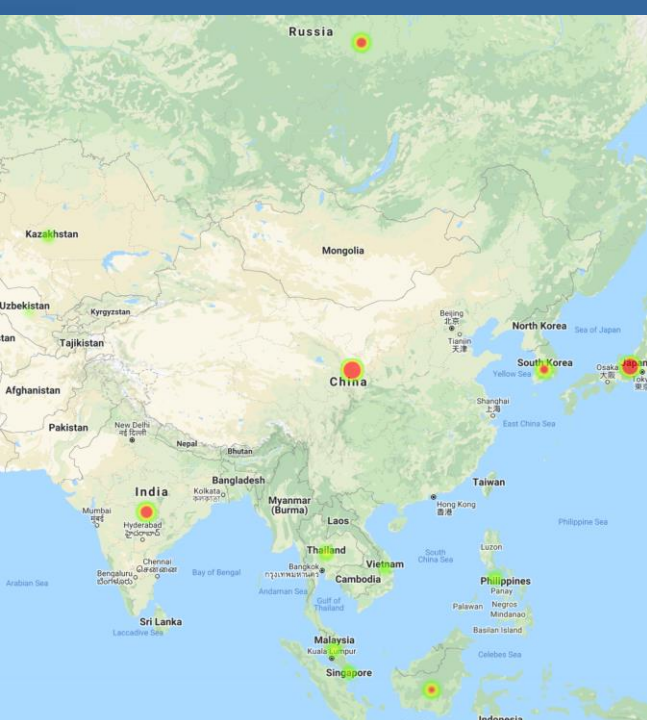


# Heatmap Activity in Russia/Asia

GDP

Medals Won

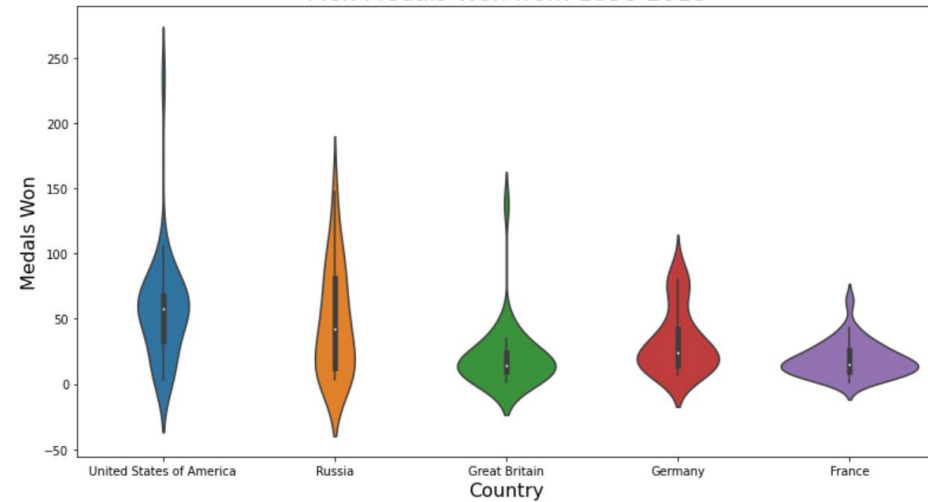
Population



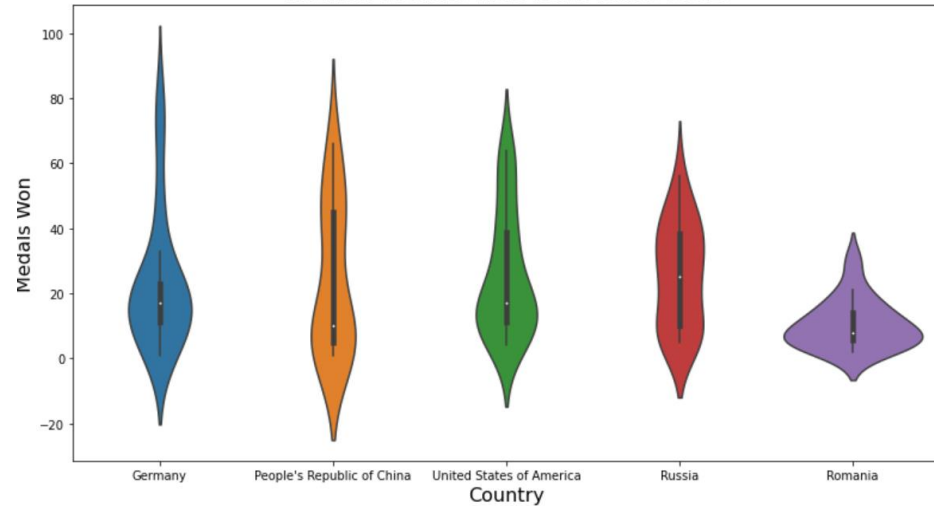


# Other Observations

Men Medals Won from 1896-2018



Women Medals Won from 1896-2018



# Calls to Action

- Can we predict how a country will do in the Olympics based on their GDP or Population
- If a country increases their population by  $X$  would they perform better in the Olympics?
- If a country increases their GDP by  $X$  will they perform better in the Olympics?



# Limitations

- If resources were unlimited, what would you do differently? What would you have explored? What would you have attempted?
  - Performance comparisons between N. America x Western Europe
  - GDP/Population influencing performance by gender
  - Can demographic data and GDP help predict if Olympic records would be broken?
  - Sub-groups: team sports x individual sports, track and field, indoor x outdoor sports...

