

# Goals

Since 2016 Rio Olympics is approaching...



We want to:

1. Identify trends in past Olympic performance using data from 1896-2008
2. Analyze potential correlation between performance and economy & population
3. Make predictions on Olympic performance

# Metrics

- Historical medal data for individuals and countries
- Population by country
- GDP by country

# Data Extraction

1. Source from [Google Spreadsheet](#), manually publish each sheet to get html version (google spreadsheet came from this [article](#))
2. GDP & Population data from Quandl, call [API](#)
3. Put data in Pandas Dataframes, clean data, create graphs with matplotlib, build model with statsmodels API

Quandl [Search finance & economic data](#) [TOOLS](#)

[BROWSE DATABASES](#)

## World Bank World Development Indicators

[DATA](#) [DOCUMENTATION](#)

Datasets [US GDP](#)

Preview Unavailable	Coal rents (% of GDP) - American Samoa Coal rents (% of GDP) - American Samoa. CODE: <a href="#">WWDI/ASM_NY_GDP_COAL_RT_ZS</a>	Latest: 2013-12-31 Annual, since 1970
Preview Unavailable	Oil rents (% of GDP) - American Samoa Oil rents (% of GDP) - American Samoa. CODE: <a href="#">WWDI/ASM_NY_GDP_PETR_RT_ZS</a>	Latest: 2012-12-31 Annual, since 1970
	Merchandise trade (% of GDP) - United States Merchandise trade (% of GDP) - United States. CODE: <a href="#">WWDI/USA_TG_VAL_TOTL_GD_ZS</a>	Latest: 2014-12-31 Annual, since 1960

**Olympic Games 2012**  
London 2012 Olympics data

Olympic medal winners: every one since 1896 as open data

Which countries have won the most Olympic medals? Who has won the most too? Get the full data here

- [Get the data](#)
- [Interactive guide to this data](#)
- [Data journalism and data visualisations from the Guardian](#)

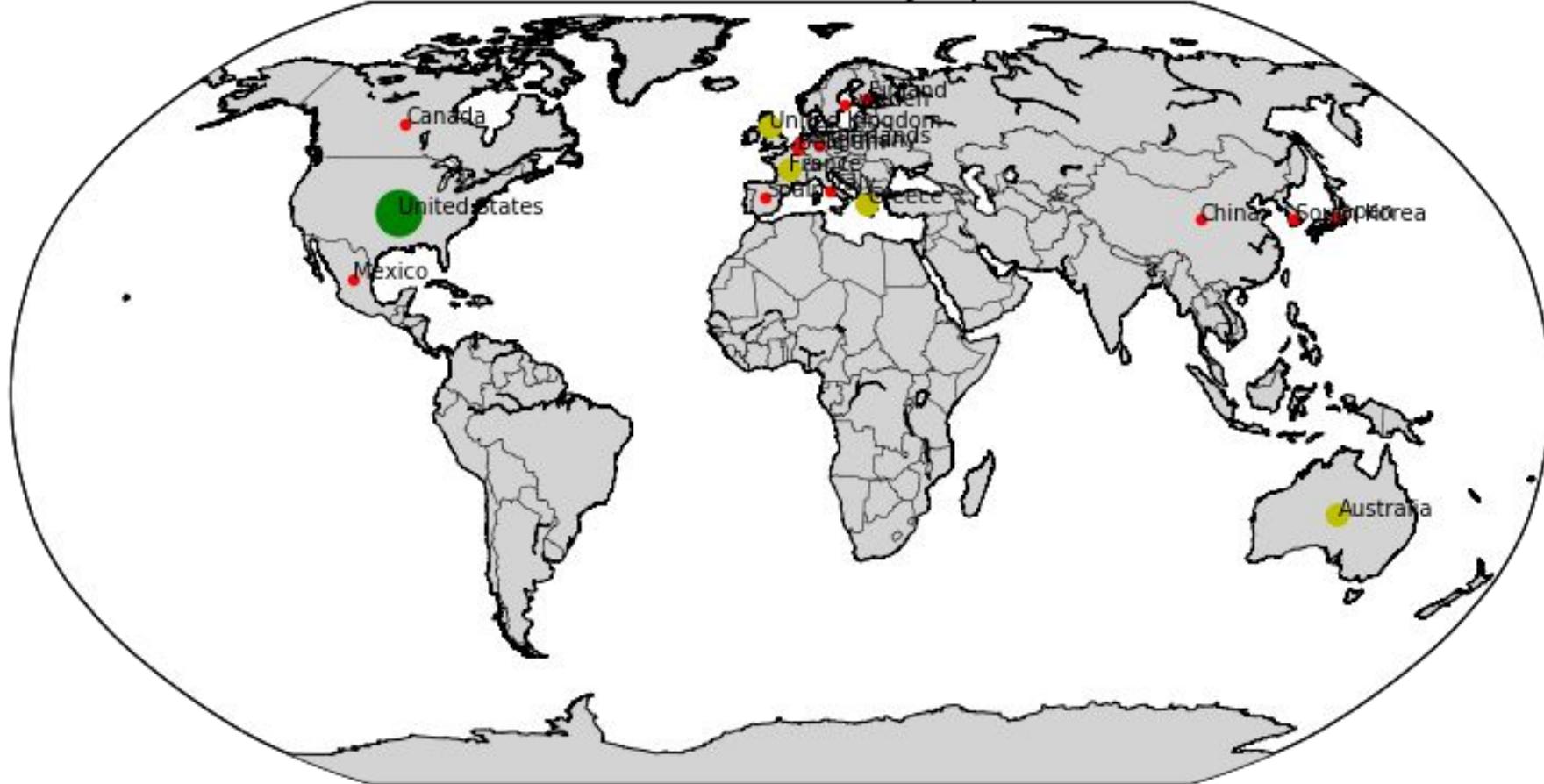


# Visualizations



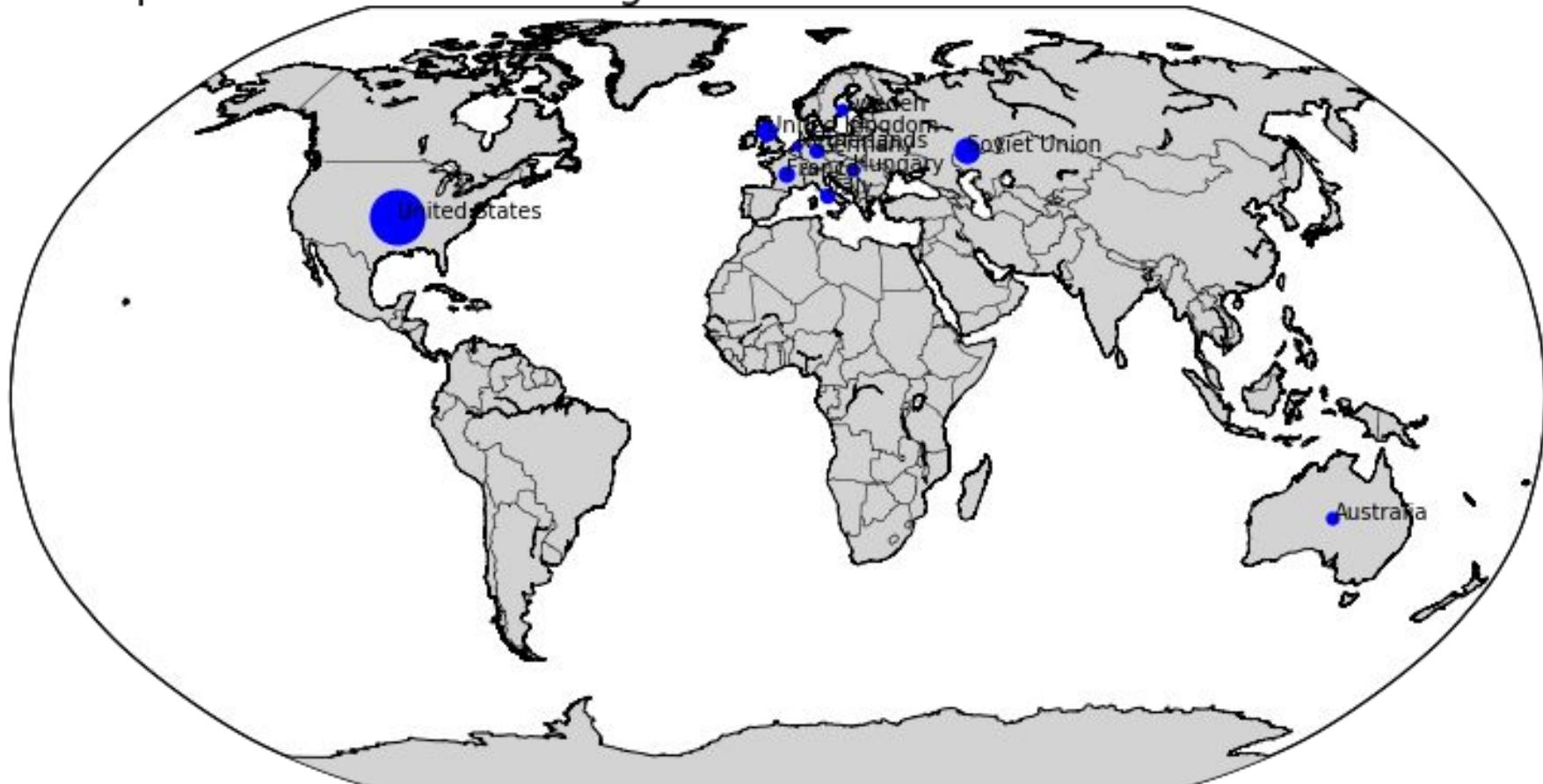
# U.S. has hosted the most times

All Countries that have Hosted Summer Olympic Games from 1896-2008

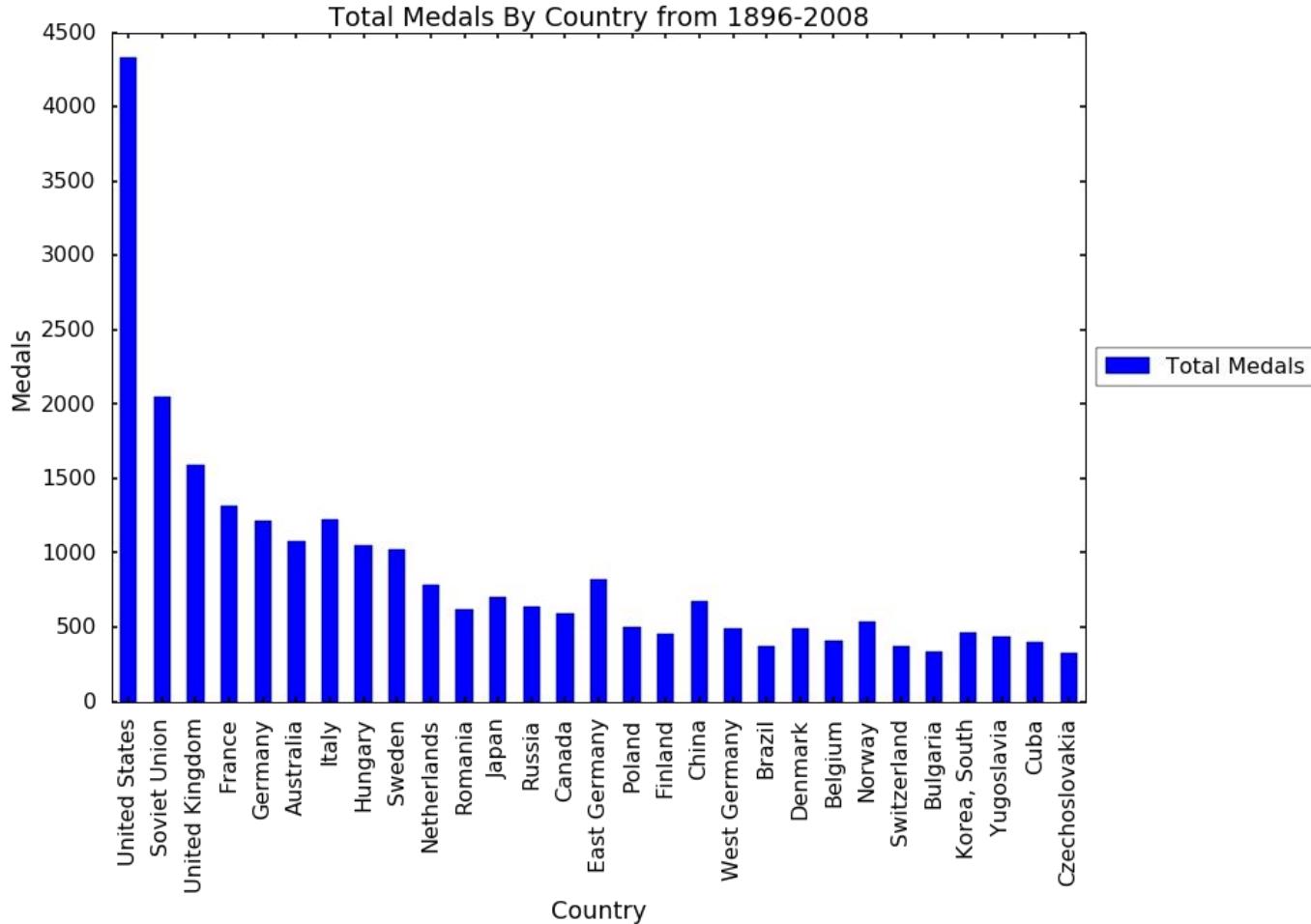


U.S. has the highest medal count from 1896-2008

Top Ten Countries with Highest Total Medal Count from 1896-2008

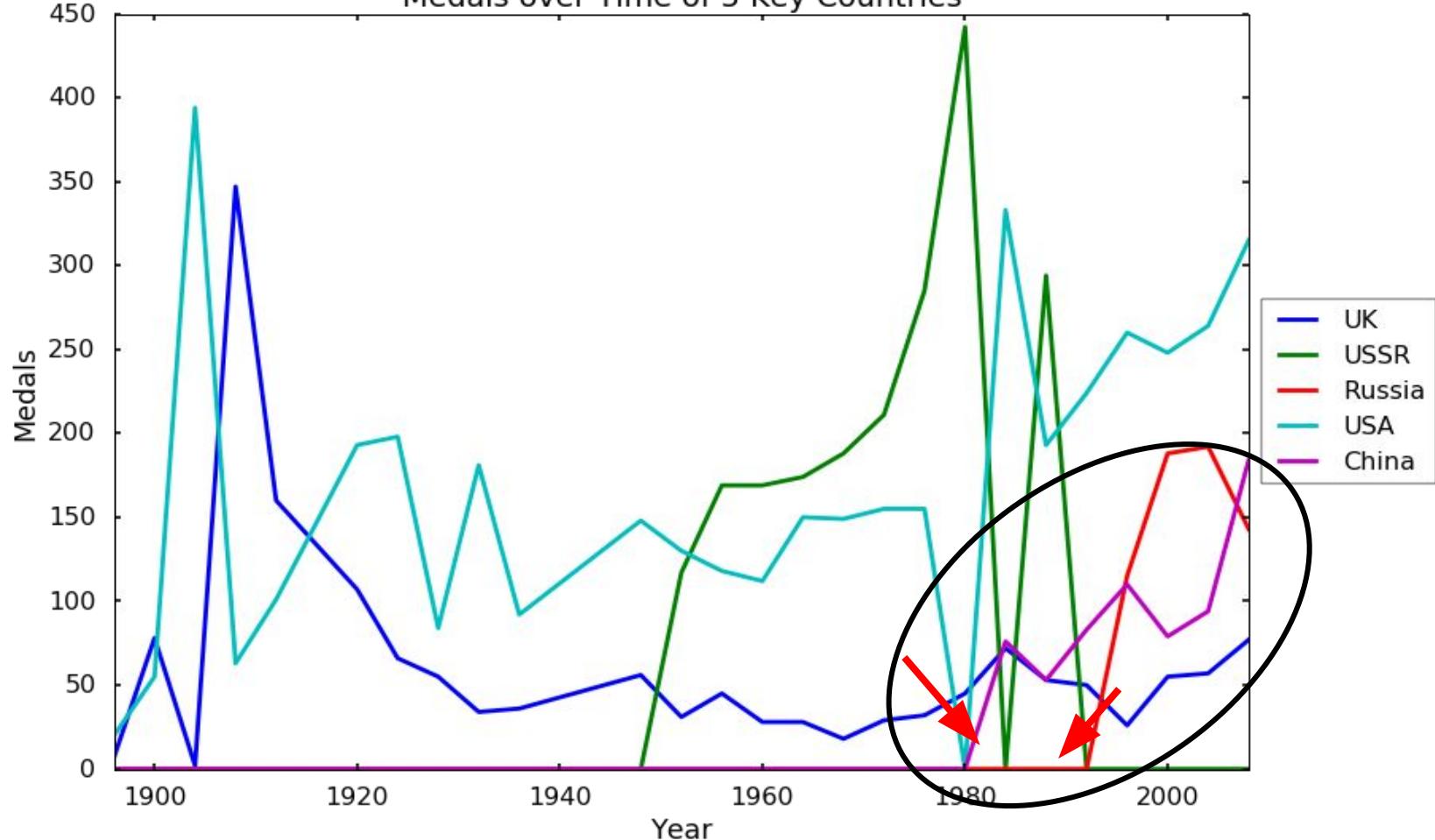


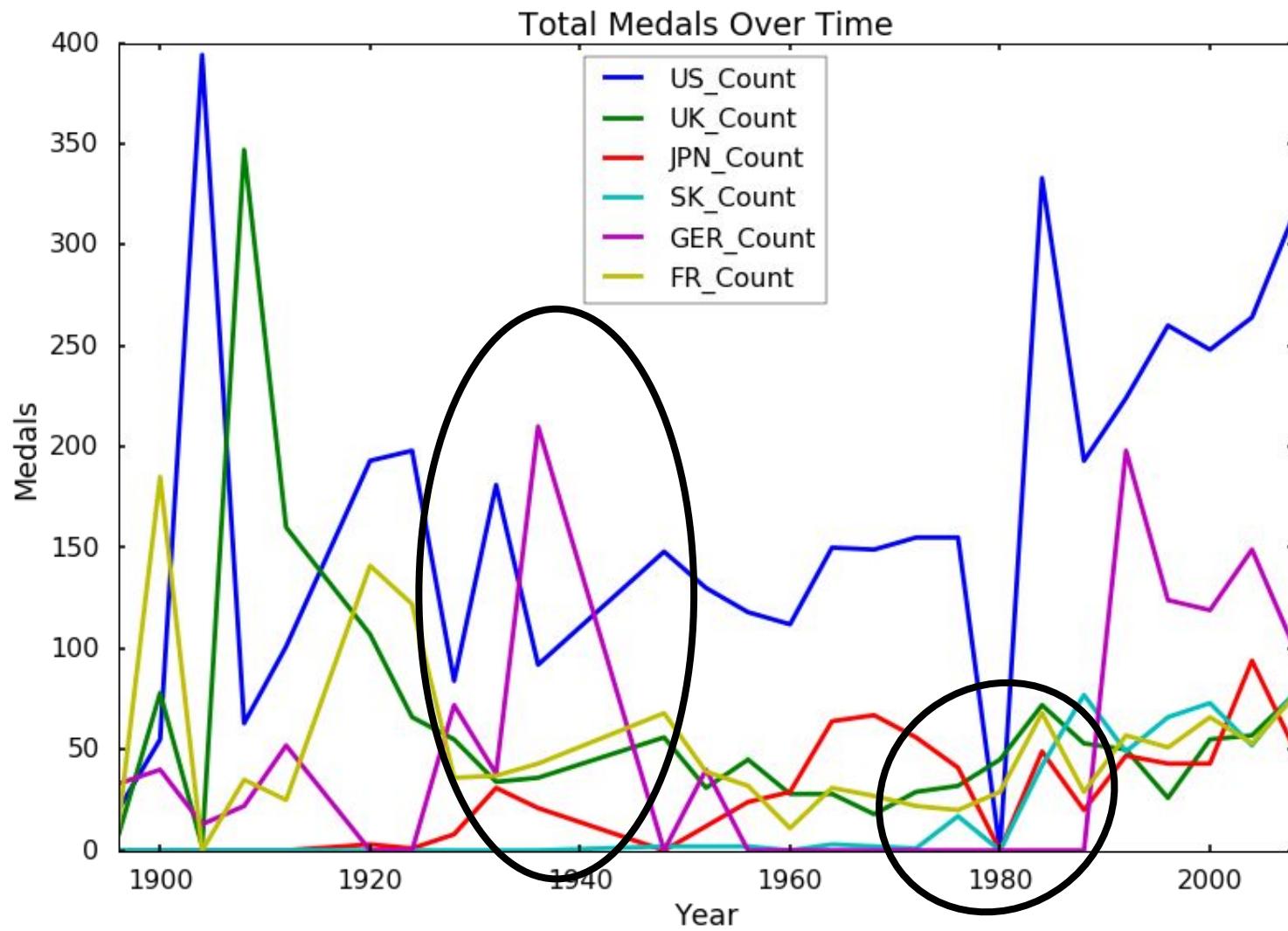
# U.S. has the highest medal count from 1896-2008 by a lot...



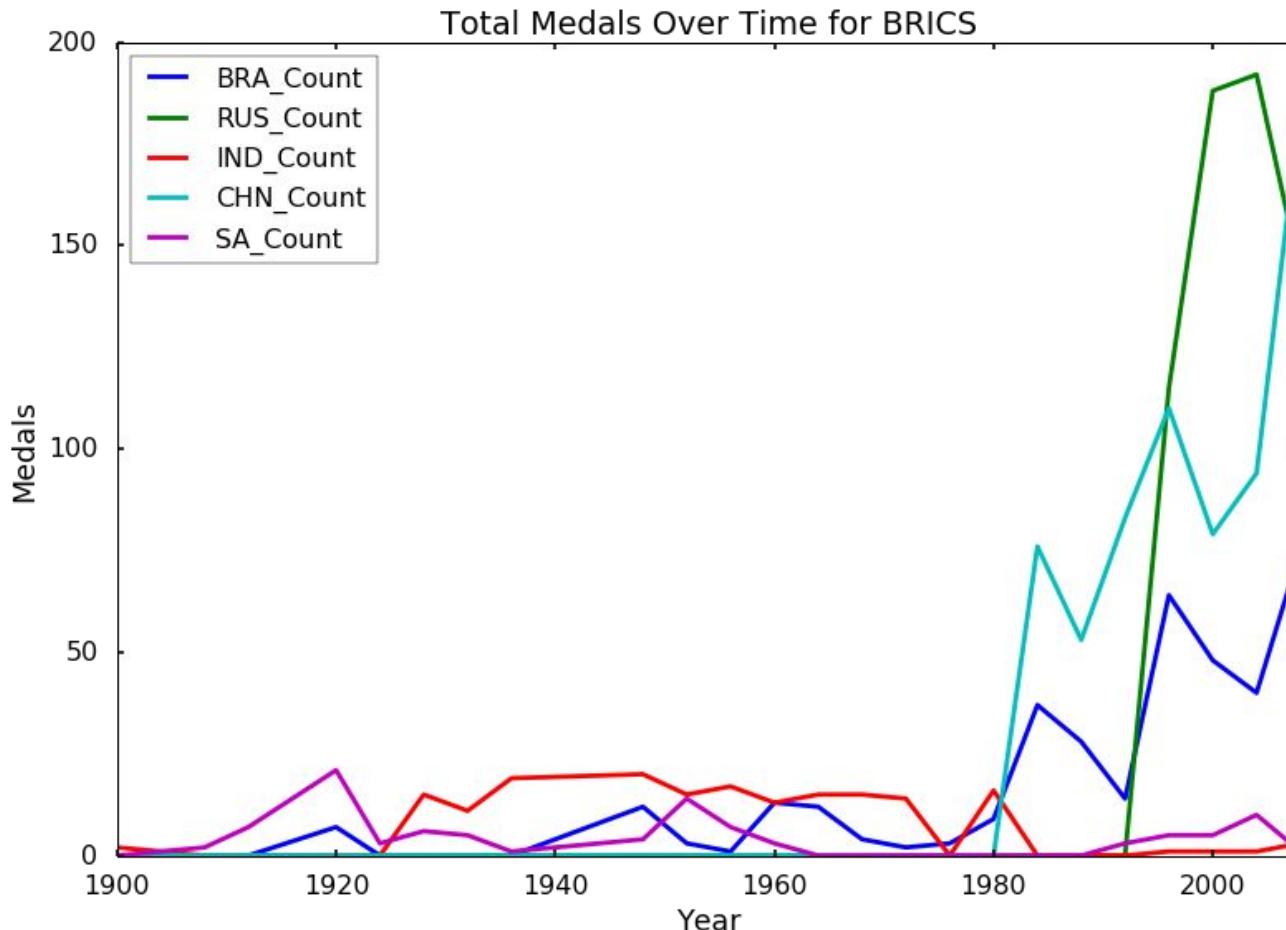
# History affects medal count

Medals over Time of 5 Key Countries

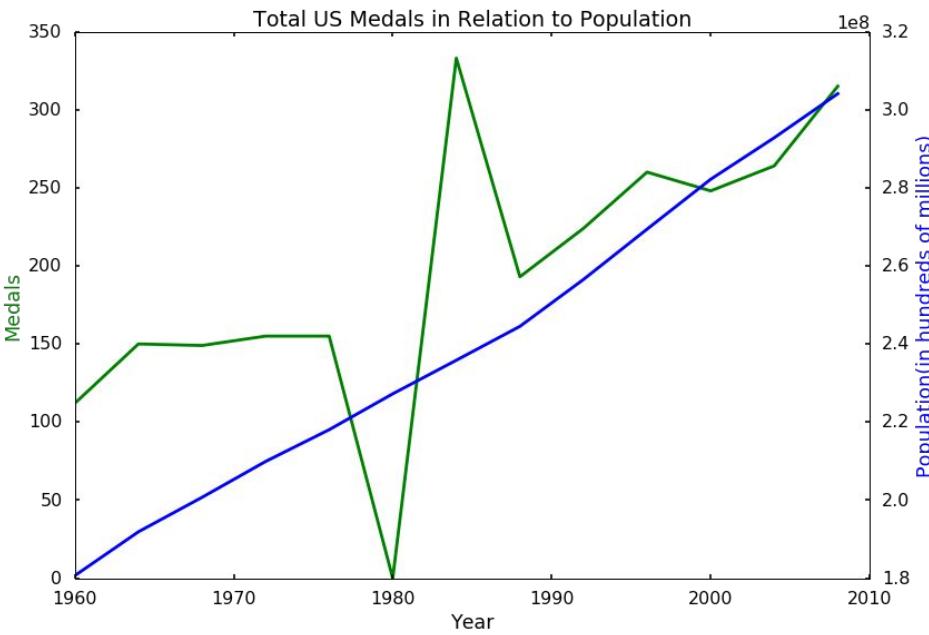
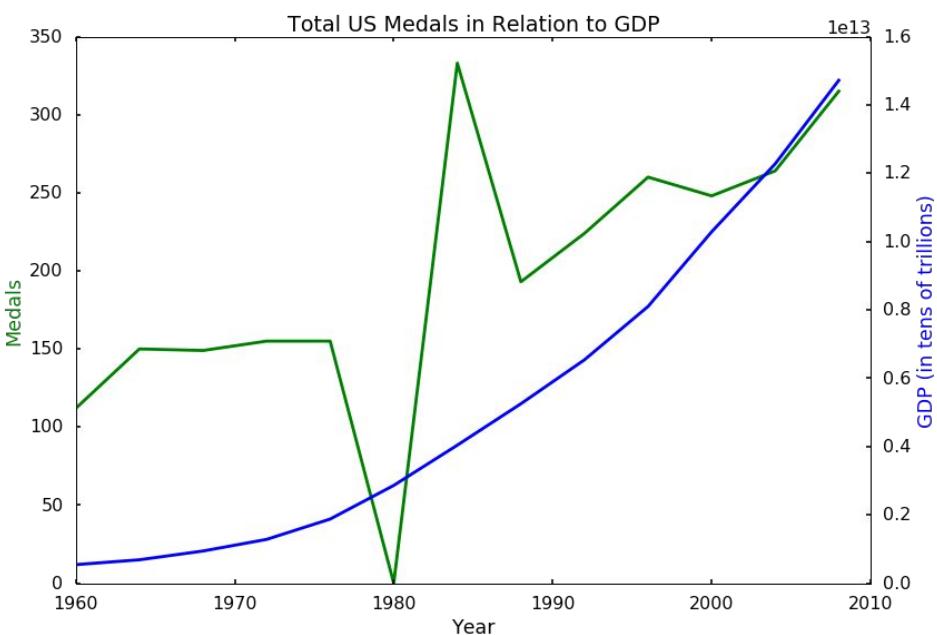




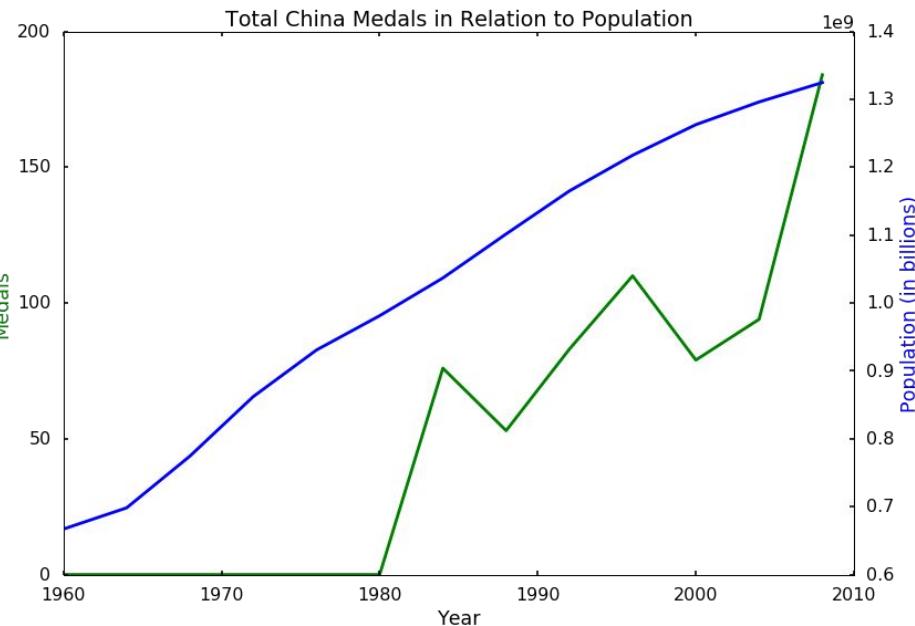
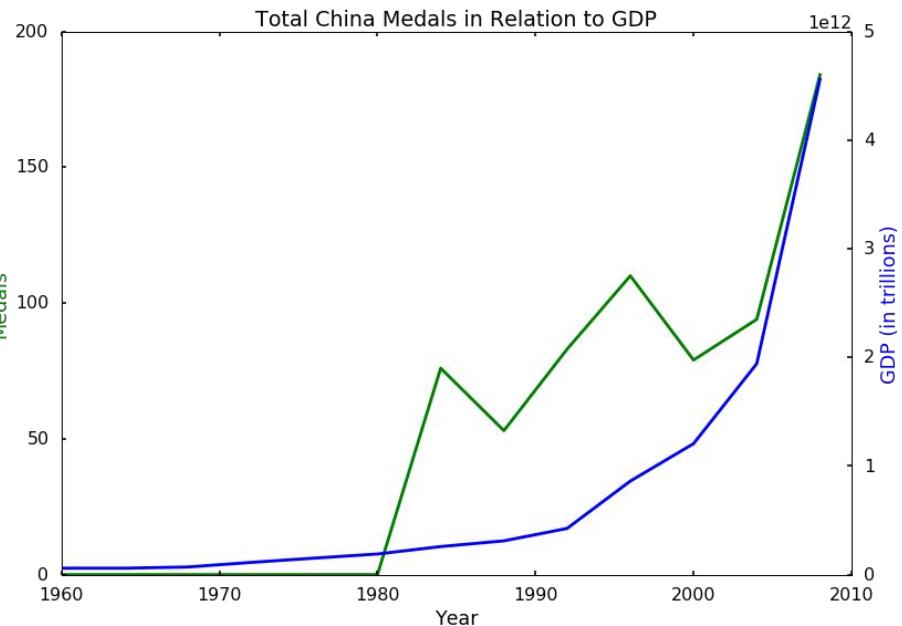
# Emerging countries have mixed results at Olympics



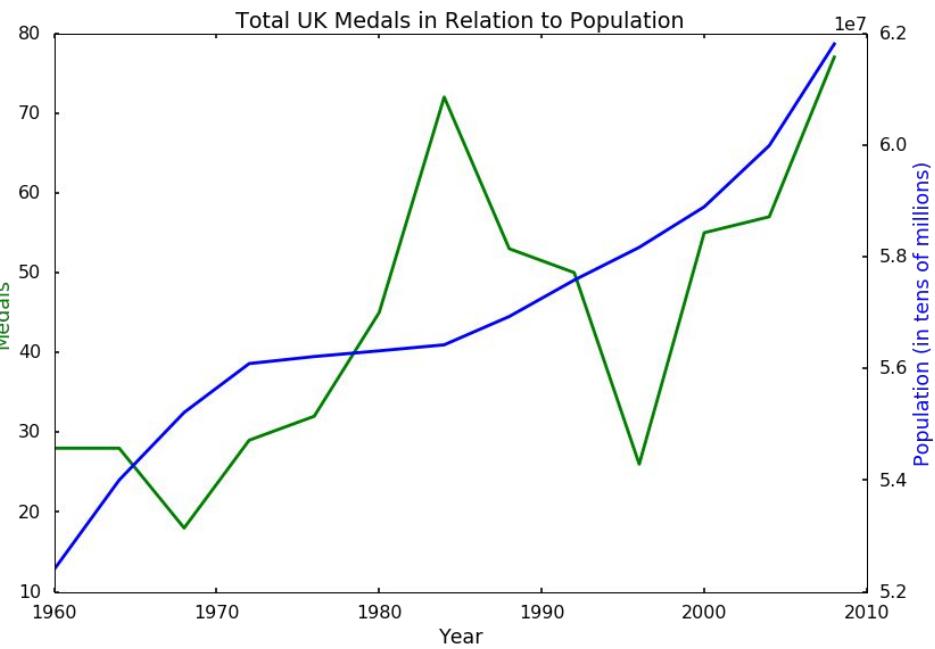
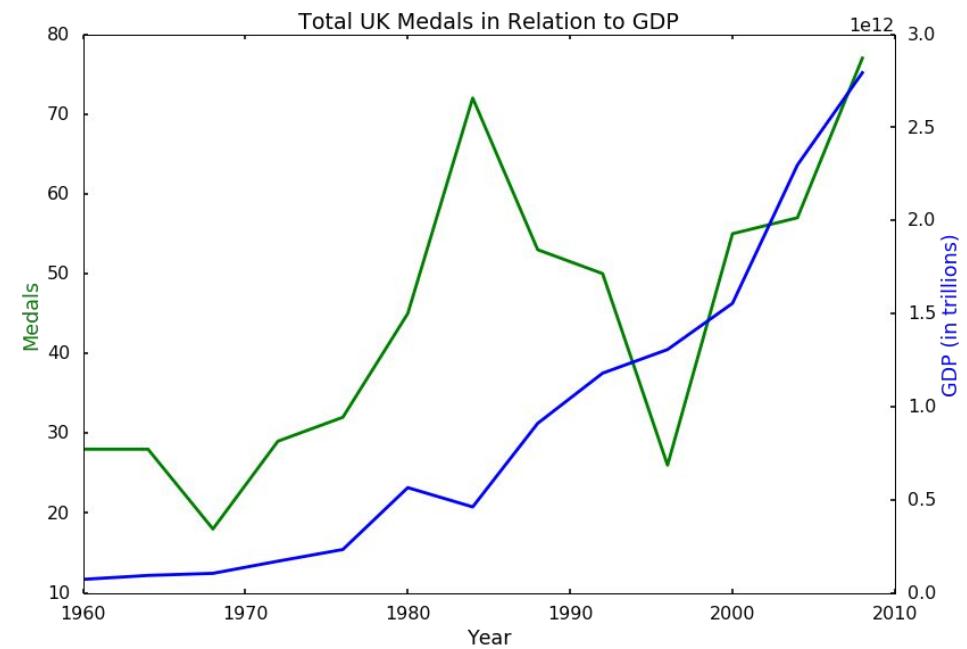
# Correlation to Economic/Social Factors: United States



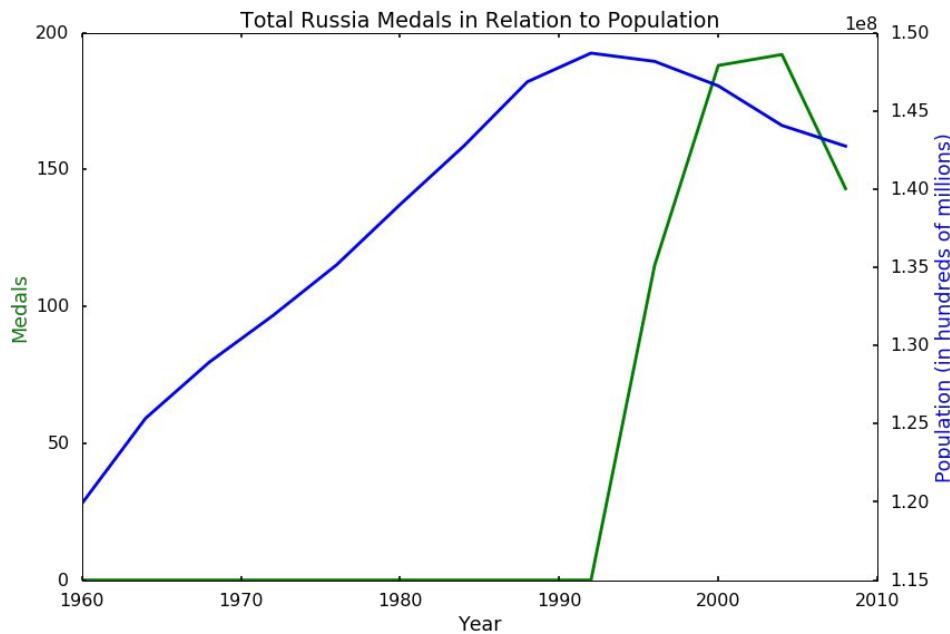
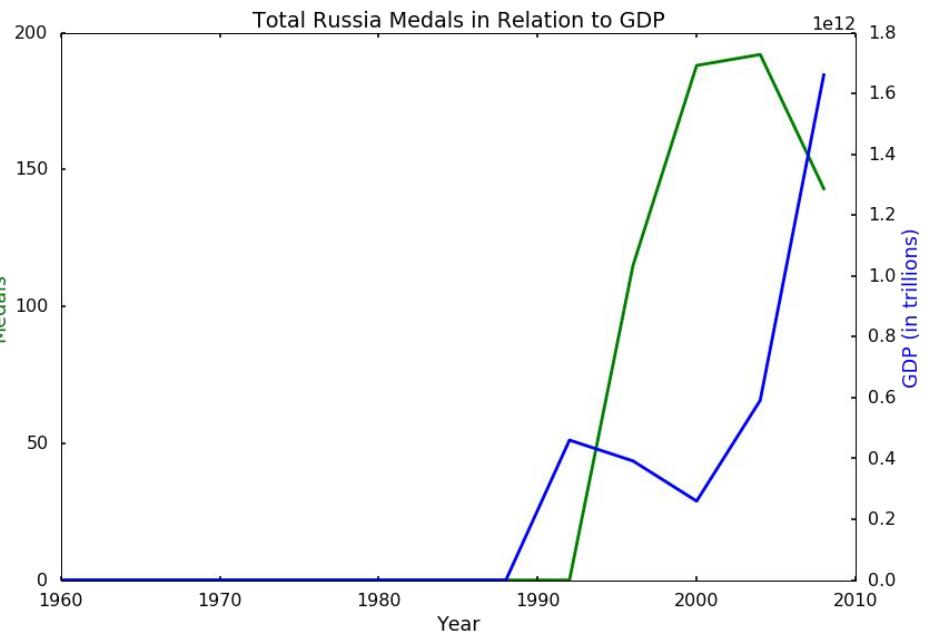
# Correlation to Economic/Social Factors: China



# Correlation to Economic/Social Factors: United Kingdom



# Correlation to Economic/Social Factors: Russia



# Predictions



# Prediction Method:

1. Create a model with OLS Regression
2. Fit a line to the model
3. Graph the model
4. Calculate prediction for medals using predict function

DISCLAIMER: Medal Count is ALL MEDALS GIVEN OUT (ex. Basketball team has 20 members, the data set counts all 20 medals)

```
In [62]: # Normalize/relabel the time variables
demo_table["Year_Norm"] = (pd.to_numeric(demo_table["Year"])-1896)/10
# Create the Constant variable
demo_table["Constant"] = [1]*len(demo_table)
```

```
In [63]: # Fit a function using the OLS, based on Year and Constant
model = sm.ols(formula = 'USA_Total_Medals ~ np.power(Year_Norm, 2) + Year_Norm + Constant', data = demo_table).fit()

model.summary()
```

OLS Regression Results

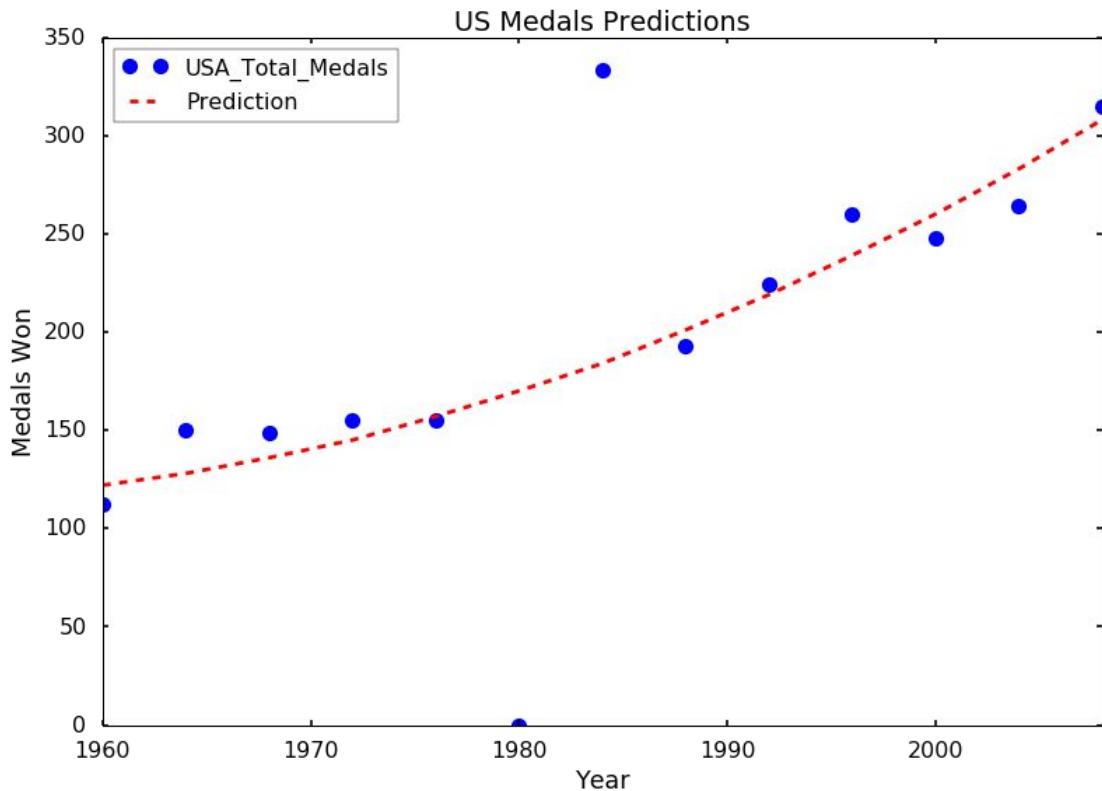
Dep. Variable:	USA_Total_Medals	R-squared:	0.462
Model:	OLS	Adj. R-squared:	0.355
Method:	Least Squares	F-statistic:	4.302
Date:	Tue, 03 May 2016	Prob (F-statistic):	0.0449
Time:	04:18:21	Log-Likelihood:	-72.462
No. Observations:	13	AIC:	150.9
Df Residuals:	10	BIC:	152.6
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[95.0% Conf. Int.]
Intercept	127.5954	386.487	0.330	0.748	-733.551 988.742
np.power(Year_Norm, 2)	5.3259	10.154	0.525	0.611	-17.298 27.950
Year_Norm	-54.8626	179.213	-0.306	0.766	-454.175 344.450
Constant	127.5954	386.487	0.330	0.748	-733.551 988.742

Omnibus:	8.434	Durbin-Watson:	2.981
Prob(Omnibus):	0.015	Jarque-Bera (JB):	5.897
Skew:	-0.493	Prob(JB):	0.0524
Kurtosis:	6.148	Cond. No.	3.87e+17

# Model and Prediction

## United States



### Predictions:

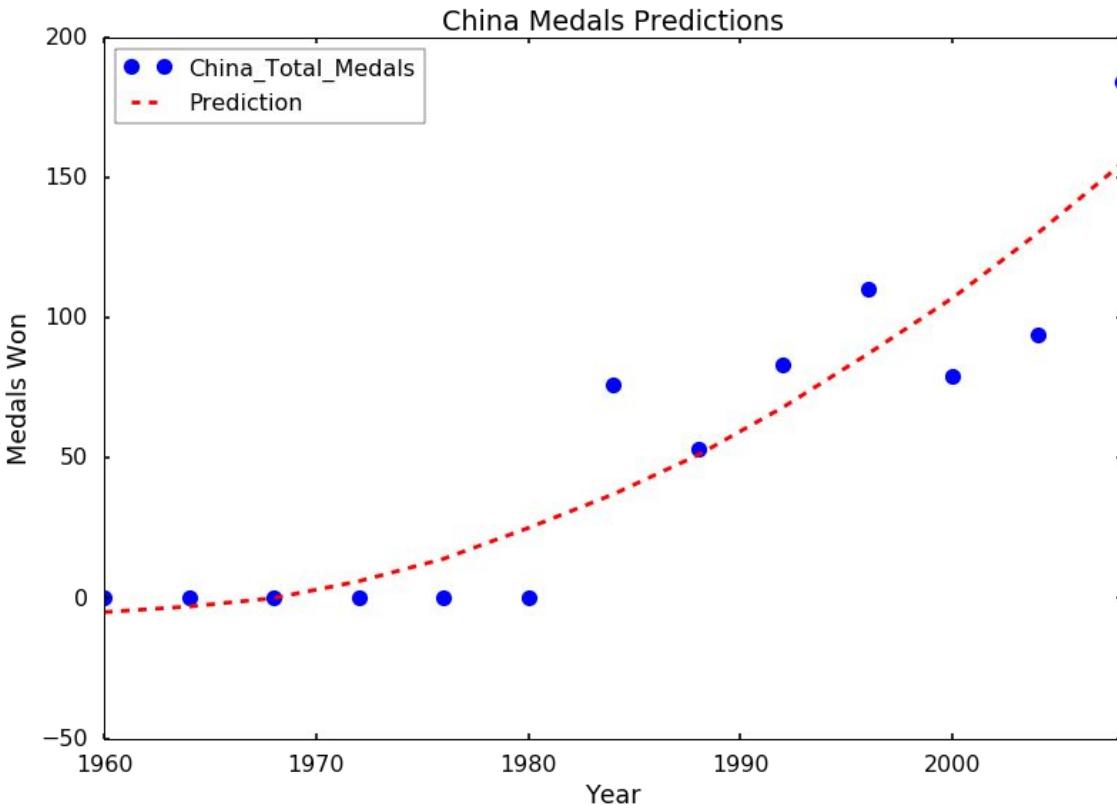
- 2004: 283 medals
- 2008: 308 medals
- 2012: 335 medals
- 2016: 363 medals

### Actual:

- 2004: 264 medals
- 2008: 315 medals
- 2012: 103 medals (w/o counting individuals)

# Model and Prediction

## China



### Predictions:

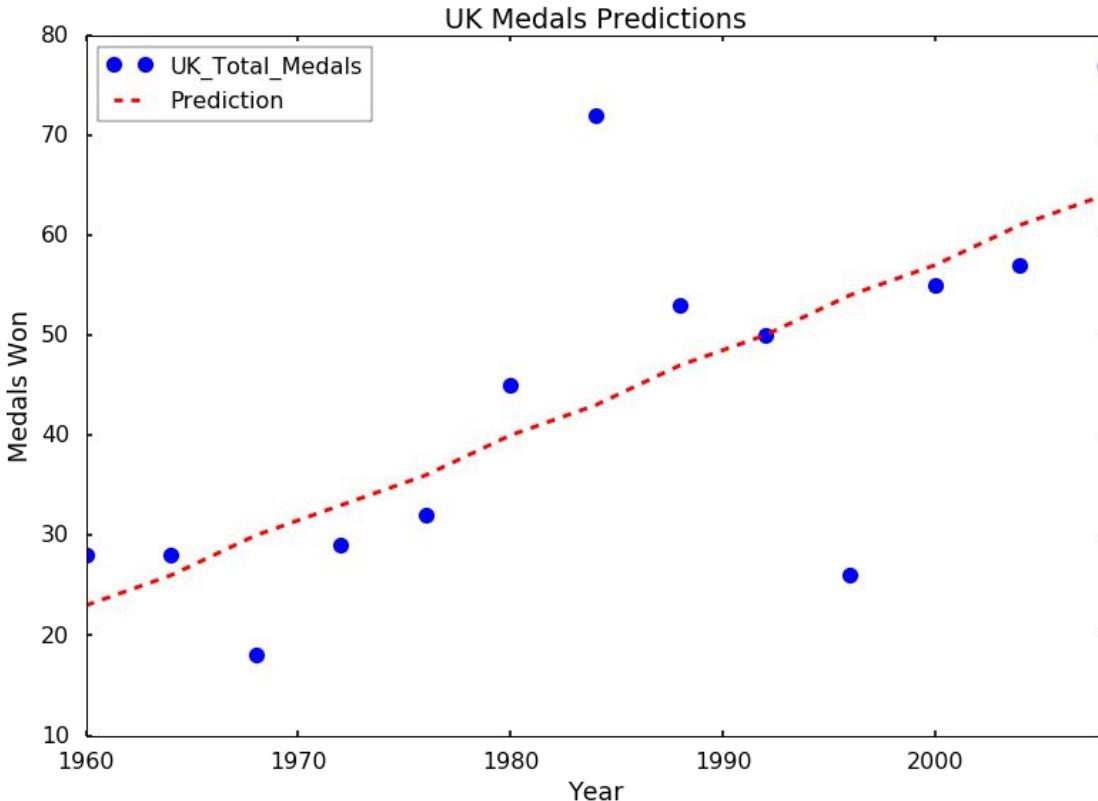
- 2004: 130 medals
- 2008: 155 medals
- 2012: 182 medals
- 2016: 212 medals

### Actual:

- 2004: 94 medals
- 2008: 184 medals
- 2012: 88 medals

# Model and Prediction

## United Kingdom



### Predictions:

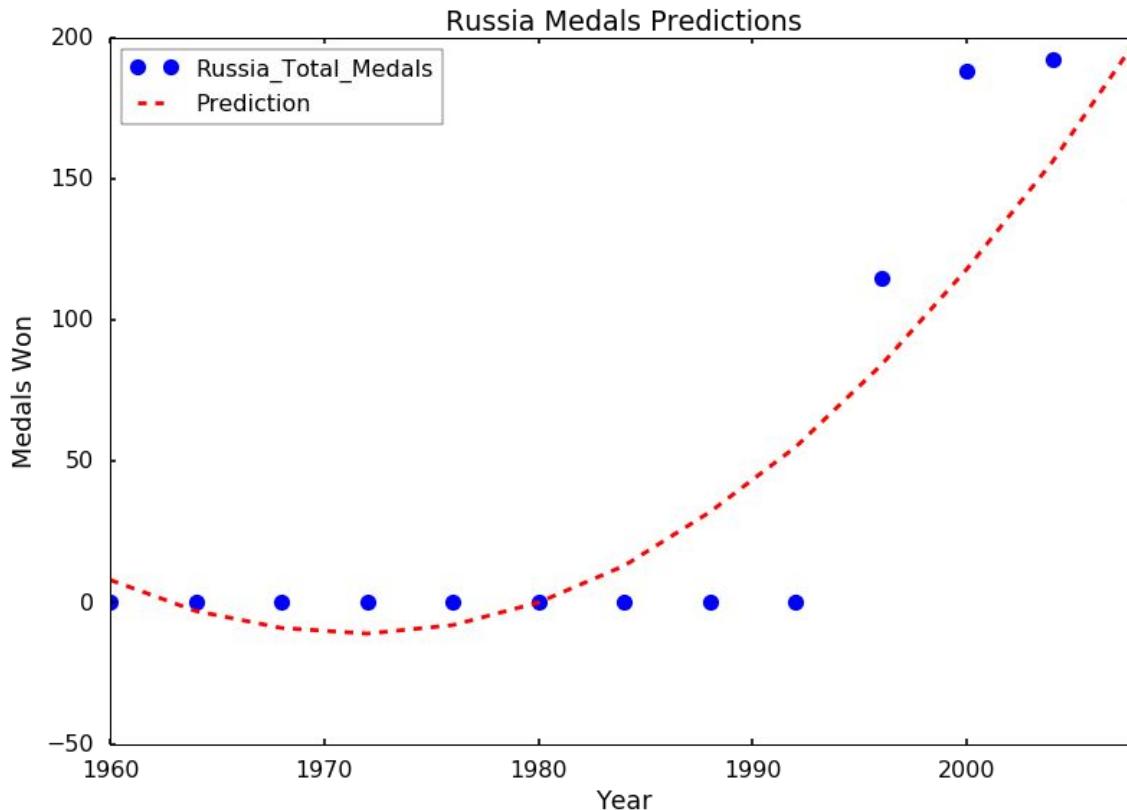
- 2004: 61 medals
- 2008: 64 medals
- 2012: 68 medals
- 2016: 71 medals

### Actual:

- 2004: 57 medals
- 2008: 77 medals
- 2012: 65 medals

# Model and Prediction

## Russia



### Predictions:

- 2004: 156 medals
- 2008: 200 medals
- 2012: 249 medals
- 2016: 303 medals

### Actual:

- 2004: 192 medals
- 2008: 143 medals
- 2012: 81 medals

# Improvements

- Normalize the count of medals over time
- Include more demographic factors - GDP per Capita, age groups, etc.
- Test more demographic factors into predictive model

