

Olympic Champions

*How many total medals
will a country take home?*



Module 2: Statistical Inference

Adam Dick // FLATIRON SCHOOL

Exploratory Data Analysis

Dataset: Features and Challenges

120 Years of Olympic History: Athletes and Events

The Olympics

Individual Records

271,116 x 7
Competing Athletes
1896 - Present

Target Variable

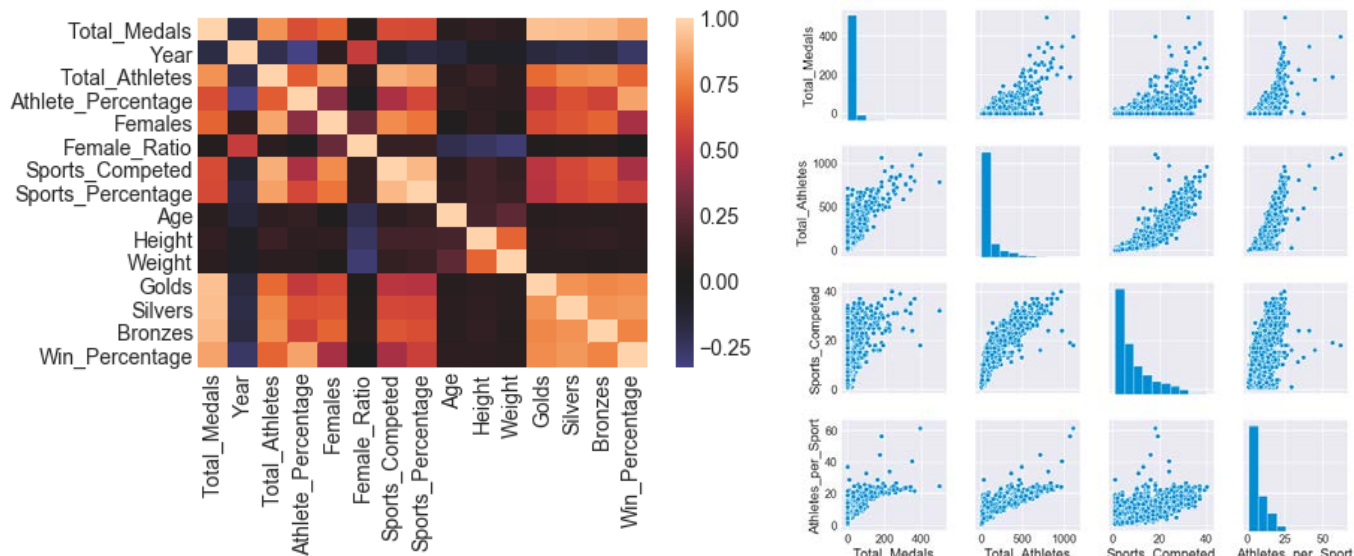
Total Medals Won

Data Grouping

Teams and Years

Key Features

Team Size & Sports
Superpowers



<https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results>

Multi Linear Regression

Predictive Modeling: Initial Model



Larger Olympic Teams Bring Home More Medals

Bring Numbers

Total Athletes

C_{TA} 0.1962

$P > |t|$ 0.0000

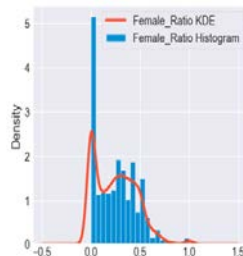
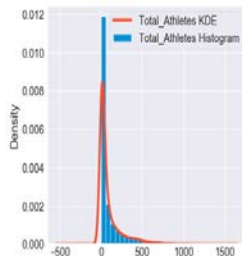
R^2 0.6530

Each Additional
5 Athletes

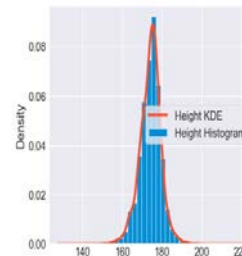
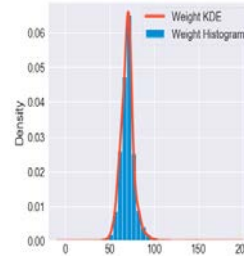
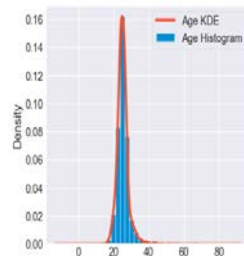
>>

Increases 1 Medal

Skewed Team Compositions



Normally-Distributed Individual Characteristics



Total_Medals	index	NOC	Year	Total_Athletes	Athlete_Percentage	Females	Female_Ratio	Sports_Competed	Sports_Percentage
7.0	4	GBR	1896	25	6.578947	0	0.0	7	77.777778
33.0	5	GER	1896	94	24.736842	0	0.0	6	66.666667
52.0	6	GRE	1896	148	38.947368	0	0.0	9	100.000000
20.0	11	USA	1896	27	7.105263	0	0.0	3	33.333333
5.0	13	AUS	1900	6	0.309917	0	0.0	2	10.000000



Multi Linear Regression

Predictive Modeling: Augmented Model

// FLATIRON
SCHOOL

Winning Teams Bring High-Quality Super-Athletes

Focus to Win

Sports Completed

C_{SC} -2.2385
 $P > |t|$ 0.0000

Athletes / Sport

$C_{A/S}$ -0.8649
 $P > |t|$ 0.0000

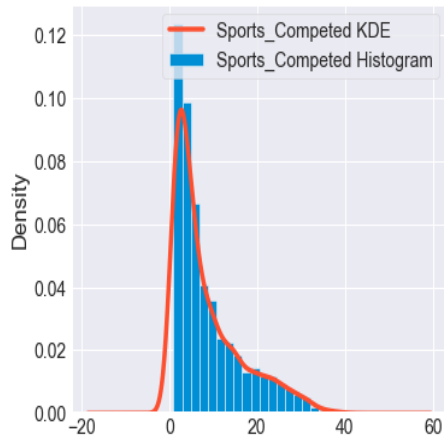
R^2 0.7070

Each Additional
2 Sports

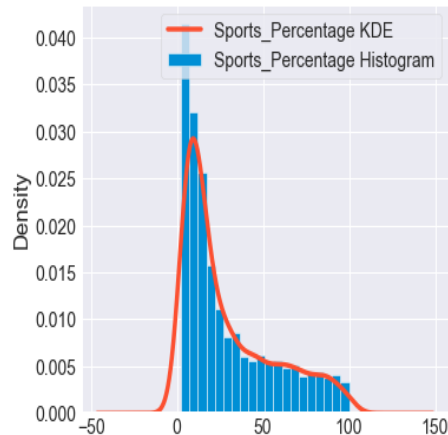
>>

Reduces 5 Medals

Most Countries Send Few Athletes



Few Countries Compete in 100% of Sports



Multi Linear Regression

The Olympics as a Proxy for the Cold War

Fight the Cold

Team USA

C_{USA} 50.9315
 $P > |t|$ 0.0000

Team USSR

C_{USSR} 147.890
 $P > |t|$ 0.0000

R^2 0.7900

Team USSR

97 Medal
Advantage Over
Team USA

Top Countries

	Total_Medals
Total_Medals	1.000000
Total_Athletes	0.808232
Females	0.681548
Sports_Competed	0.610092
NOC_USA	0.431405
NOC_URS	0.394856
NOC_GDR	0.196903
NOC_EUN	0.142606
NOC_GER	0.140874
NOC_GBR	0.124169
NOC_RUS	0.122100
NOC_FRG	0.107779
NOC_FRA	0.105887
NOC_ITA	0.103654
NOC_SWE	0.100369
NOC_CHN	0.088227
NOC_AUS	0.079652
NOC_CAN	0.079441
NOC_HUN	0.064314
NOC_NOR	0.058641

Final Regression Model

Dep. Variable:	Total_Medals	R-squared:	0.790			
Model:	OLS	Adj. R-squared:	0.789			
Method:	Least Squares	F-statistic:	1655.			
Date:	Thu, 18 Apr 2019	Prob (F-statistic):	0.00			
Time:	19:04:39	Log-Likelihood:	-12896.			
No. Observations:	3091	AIC:	2.581e+04			
Df Residuals:	3083	BIC:	2.586e+04			
Df Model:	7					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	1.9424	0.465	4.175	0.000	1.030	2.855
Total_Athletes	0.2398	0.005	49.613	0.000	0.230	0.249
Sports_Competed	-1.4266	0.080	-17.856	0.000	-1.583	-1.270
NOC_USA	50.9315	2.984	17.070	0.000	45.081	56.782
NOC_URS	147.8902	5.145	28.745	0.000	137.802	157.978
NOC_GDR	88.7256	6.488	13.675	0.000	76.004	101.447
NOC_EUN	129.5452	15.842	8.177	0.000	98.483	160.607
NOC_FRA	-16.7720	2.849	-5.887	0.000	-22.358	-11.186
Omnibus:	1065.380	Durbin-Watson:	1.879			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	128950.113			
Skew:	0.588	Prob(JB):	0.00			
Kurtosis:	34.620	Cond. No.	9.30e+03			

Western Bloc

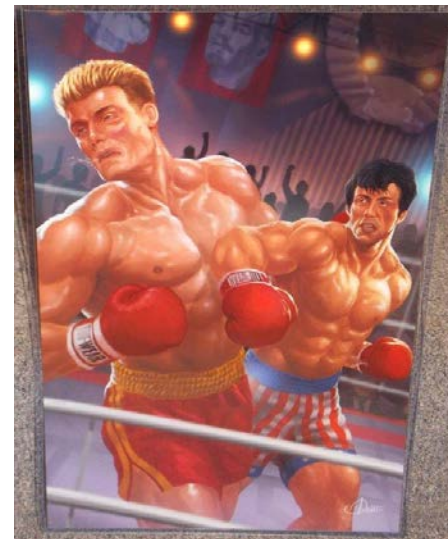
USA +51
France -17

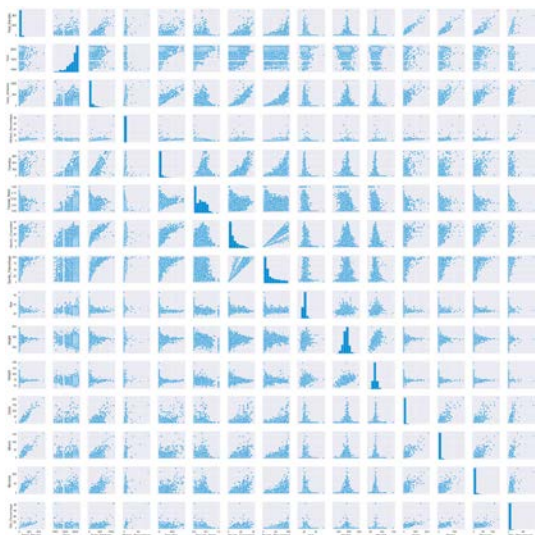
Eastern Bloc

USSR
+148

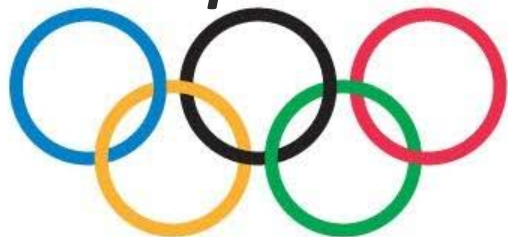
Former USSR
+130

E. Germany
+89





Olympic Champions



Module 2: Statistical Inference

Adam Dick // FLATIRON SCHOOL

