

Olympics Presentation

Devin Reeh, Godwin Rayen, Ralph Jiang, and Raymond Na

Agenda

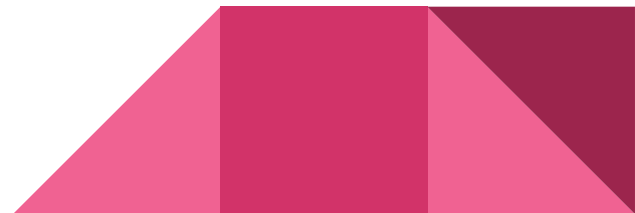
Research Question

Background Information

Exploratory Data Analysis

Model Results

Limitations/Further Research



Research Question and Model Chosen

Research Question : Which groupings of countries do well in certain types of sports.

Model Chosen: Dendrogram Clustering



Background Information

Olympics Dataset over the last 120 years

Variables in the Data :

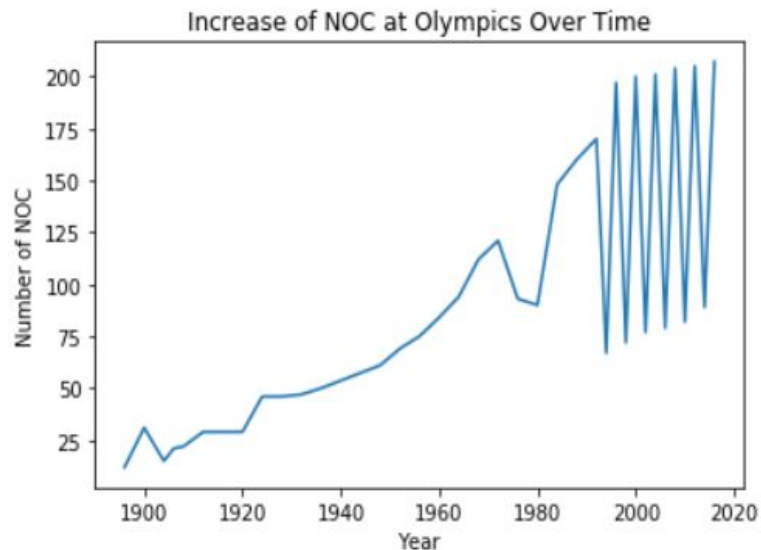
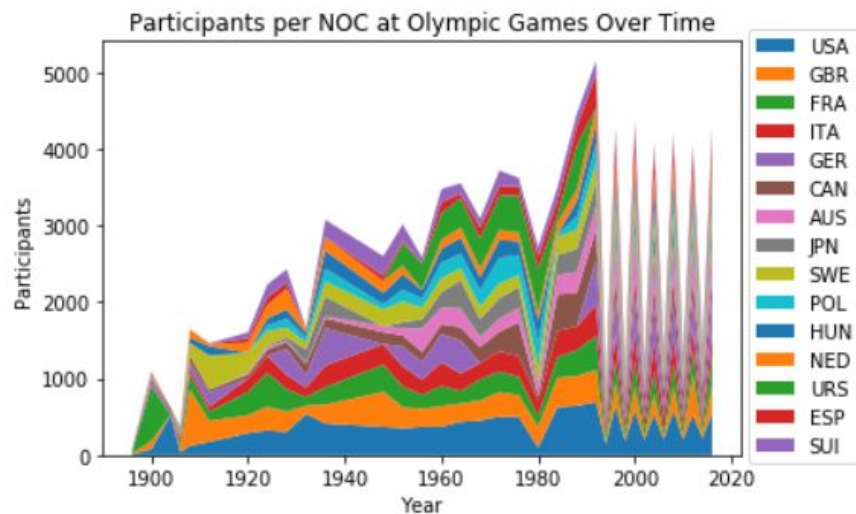
ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
----	------	-----	-----	--------	--------	------	-----	-------	------	--------	------	-------	-------	-------

Covers games between 1896-2016 - But starting at 1994

Every Olympic sport played by all participating countries(Winter and Summer)

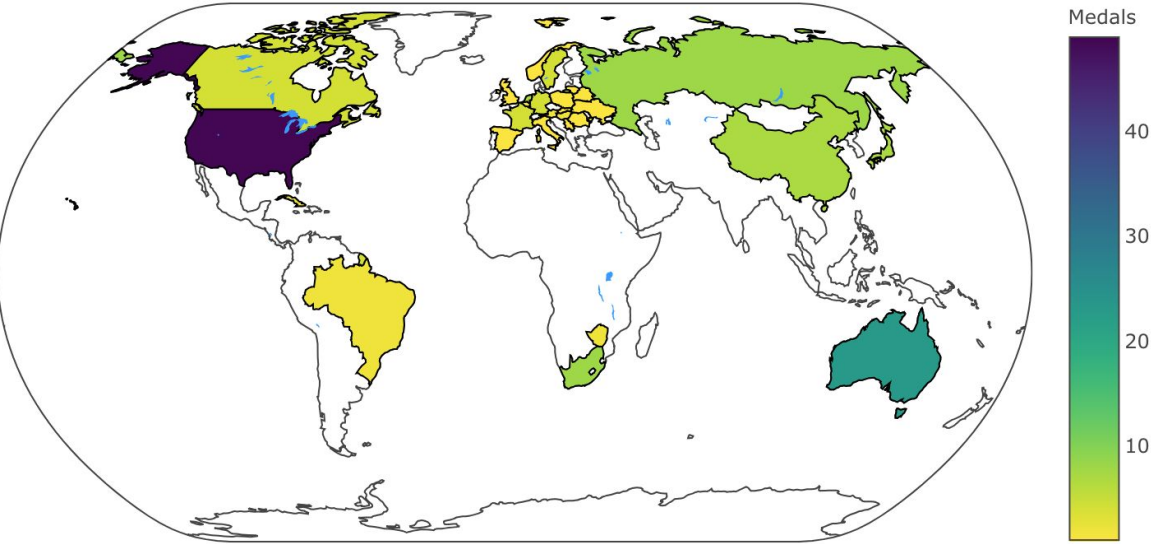


EDA - Pt 1



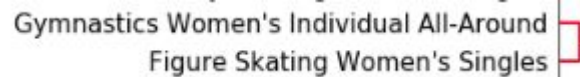
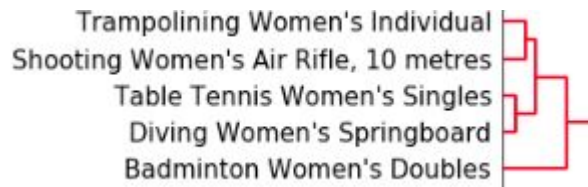
EDA - Pt. 2

Medal Density of Swimming 100m Events Men and Women

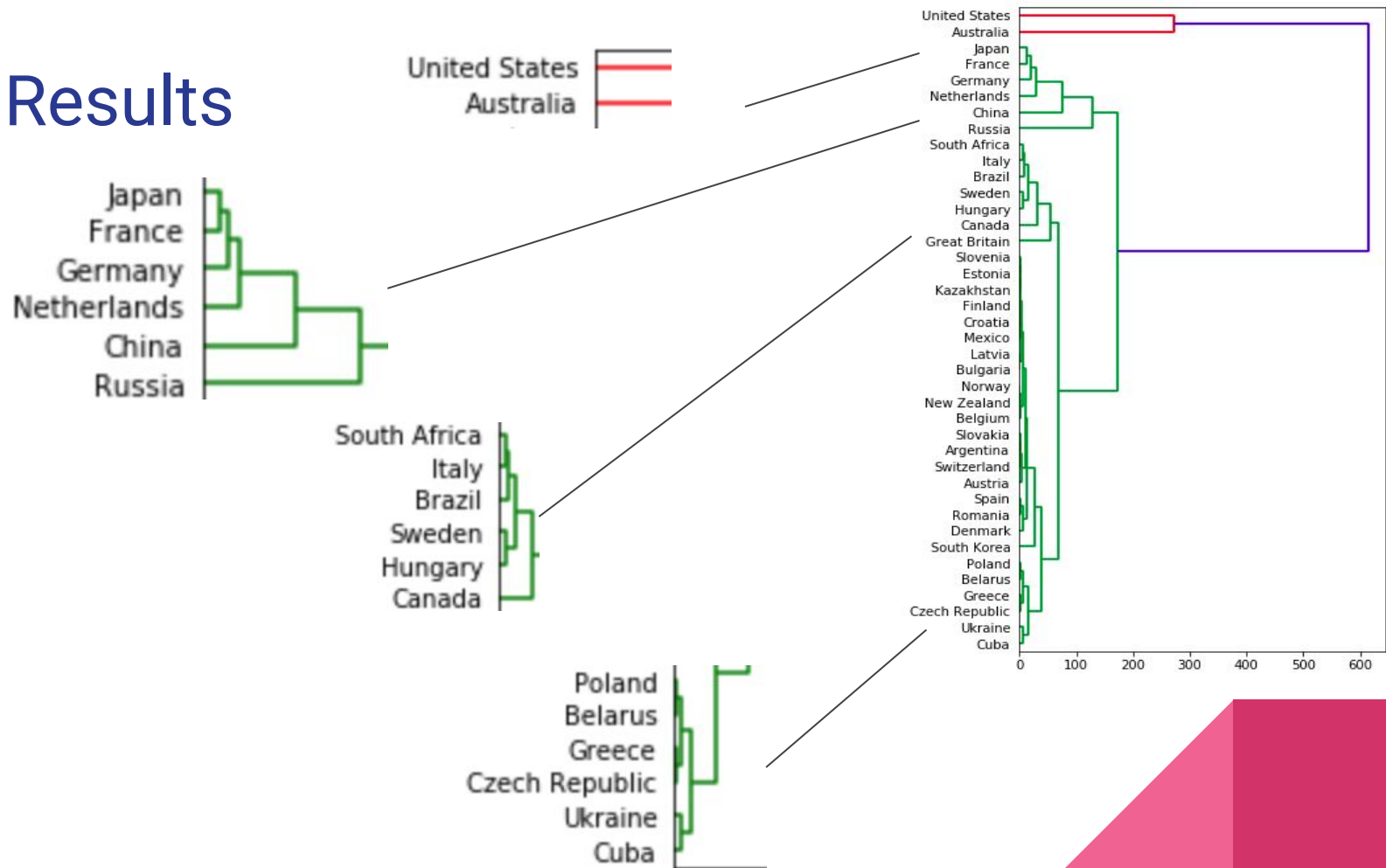


Lithuania, Ukraine, Belarus, Spain, Slovakia, Singapore, Serbia, Romania, Poland, Norway, Great Britain, Italy	1
Zimbabwe, Cuba, Brazil, Belgium, Austria, Hungary	2
Germany, France, Canada, Sweden	4
Japan, Netherlands, China	7
Russia, South Africa	8
Australia	28
United States	49

Results - Interesting Correlations



Results



Further Research

Are the winners recurring athletes?

Do certain countries have better training programs/facilities for some sports?

How would our results change when we add the additional years?





The End