



Agenda

- Theme ~ Olympic Case Study
- Coding Approach
- Data Munging Techniques
- Client Portal
- Dashboard
 - Plotly ~ Medal Count
 - Plotly ~ Female Stats
 - Leaflet Map
 - Costs
 - Sports
- Questions & Answers



Data Visualizing for important decision-making



- ~ Project Background
- ~ Topic Selection
- ~ Data Set Selection Process
- ~ Inspiration

IRA Story Tellers, INC.

Olympic Case Study ▾ About Us

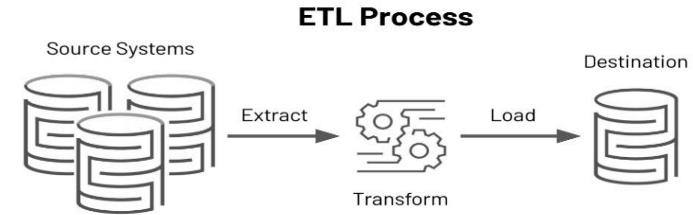
Welcome To IRA Story Tellers, INC.
Data Visualizing for your important decision making process.

IRA Story Tellers, INC. believes that story telling should be simple enough for a child to understand.

We are always turning complex data into valuable visualizations that help tell our shareholders the stories they need, in order to make important decisions.

Need more examples of what we can do for you? View our most recent case study in collaboration with the International Olympic Committee, and please reach out if you have any questions.





Finding Data

Three Data Sources: Kaggle & Data World

- This is a pre-ETL stage that involves data research and screening, as well as developing a project proposal.

Extract Data

CVS Files

- Extracted CSV files were inspected to identify possible flaws including different column titles, labelling format, and dates.
- On Jupyter, dependencies (Pandas, NumPy and SQL Alchemy) were imported dependencies.

Transform Data

Data Frames

- Created a Medals Data Frame and split Data Frame to three using groupby function by medal categories: Gold, Silver, and Bronze.
- Additional columns for medals count were inserted. To convert data to integers NaN values were replaced by zero(0) and columns were then changed to (int) data type.
- Medals and athletes' tables were merged. Columns were filtered, renamed and formatted. Columns with empty rows and duplicate insertions were dropped.
- The two cost Data Frames' columns were filtered and renamed. Duplicates were dropped, Standardized NaN values. Merged the two cost Data Frames based on year and city.

Load Data

ERD Depicting Tables

- Initialized engine to connect to postgres (pgAdmin) and checked for tables using engine.table_names().
- Through pgAdmin, we created tables, loaded data and joined them to get the relevant columns to our question. This database was chosen because the Olympics is a very structured event, and we believed that a relational database would be the best way to load the data. The data was structured in a way, that there were many common columns that we could use to merge the data together.



Data Munging Techniques

Databases and Tools

- Python Flask
- SQL
- SQLite
- Pandas
- API

Visualization Tools

- HTML/CSS
- Bootstrap
- JavaScript
- Leaflet
- Plotly



Interactive User Experience



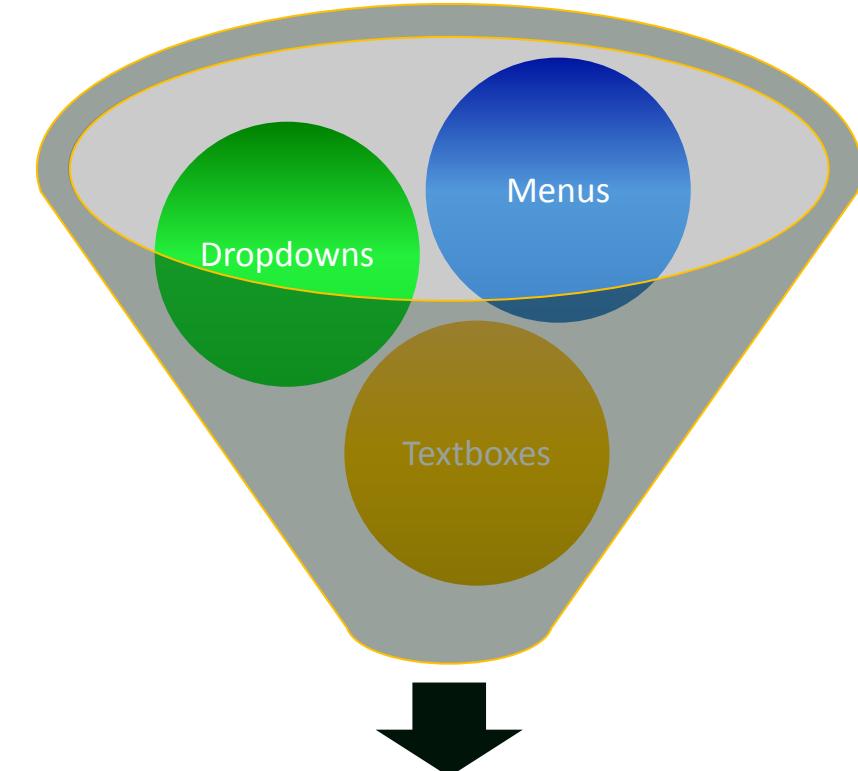
Olympic Case Study • About Us



About Olympic Case Study

Medal Count & Participation over time

Female presence over time



Dashboard

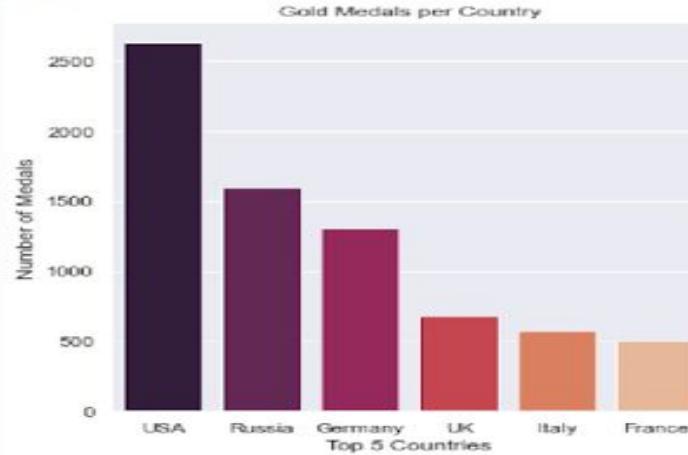


PLOTLY ~ BAR CHARTS

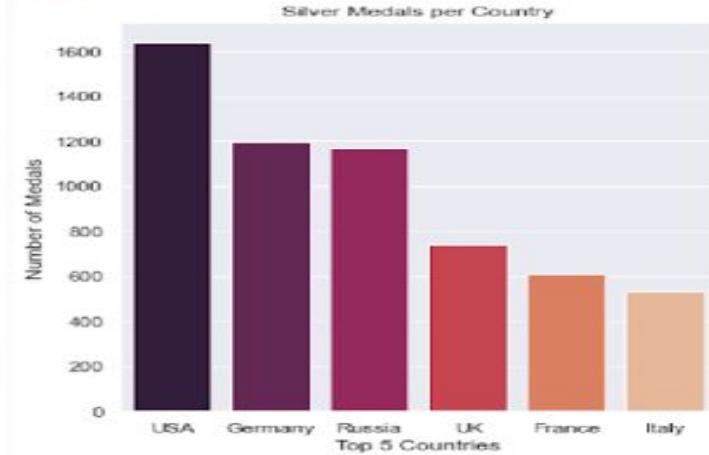
MEDAL COUNT

Comparison: Top Ten Medal Count Gold, Silver, and Bronze

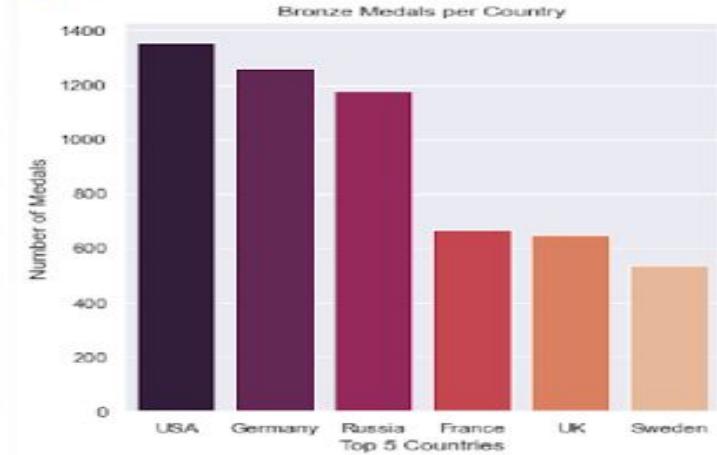
Gold



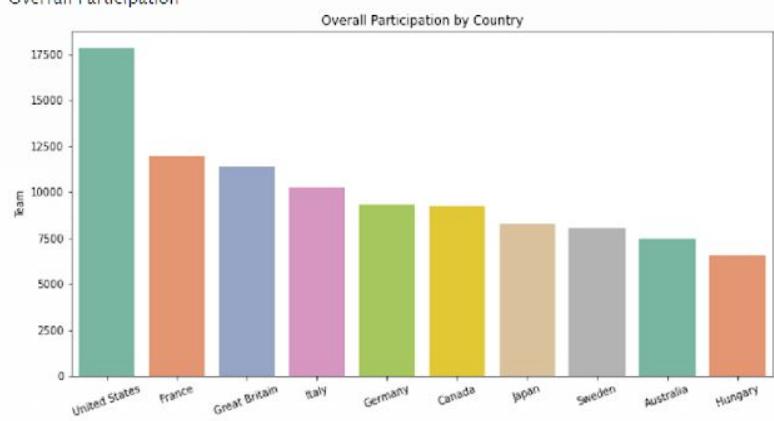
Silver



Bronze



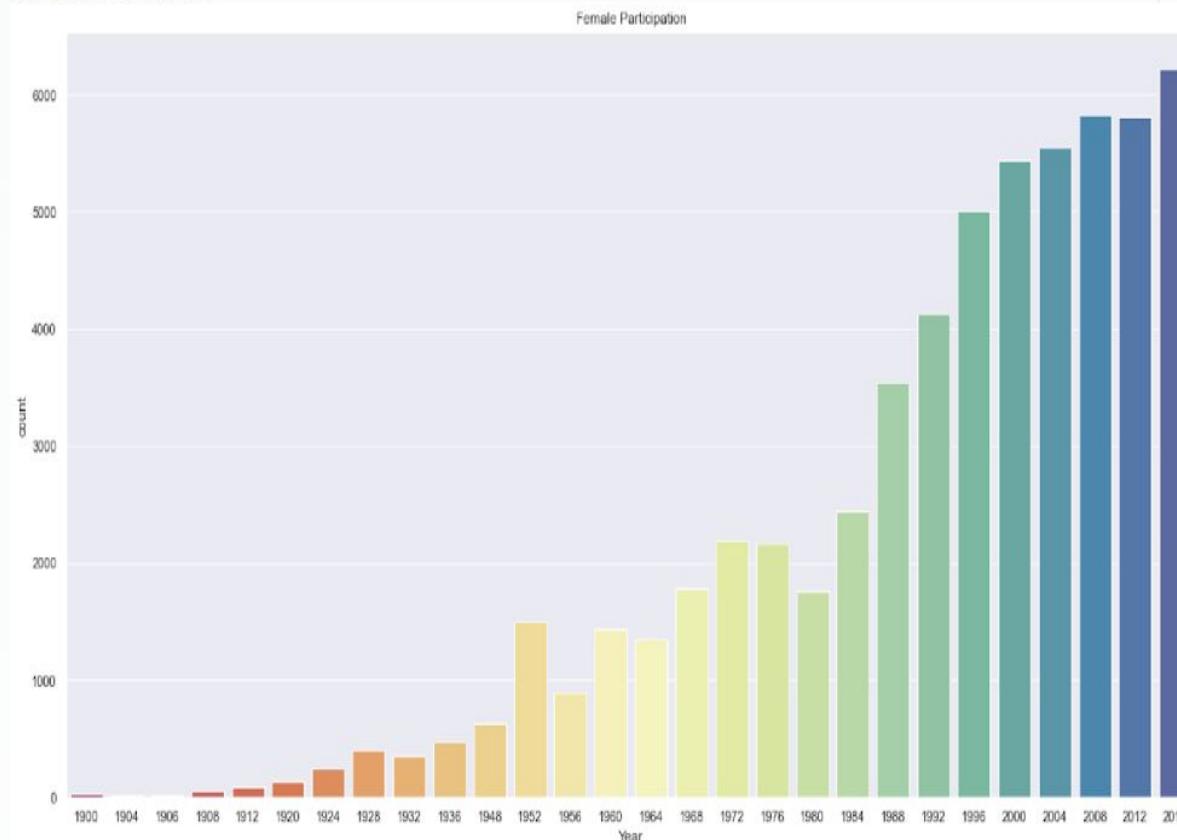
Overall Participation



Female Participation in the Olympics over 120 years

In this analysis we use a bar plot in order to visualize the growth of female participants, which has increased tremendously in the last 100 years. The number of female participants has tripled in the last 40 years. Based on the analysis we would recommend a country that is inclusive towards women and supports women's rights.

Female Athletes Bar Chart



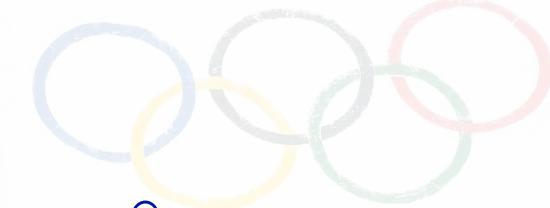
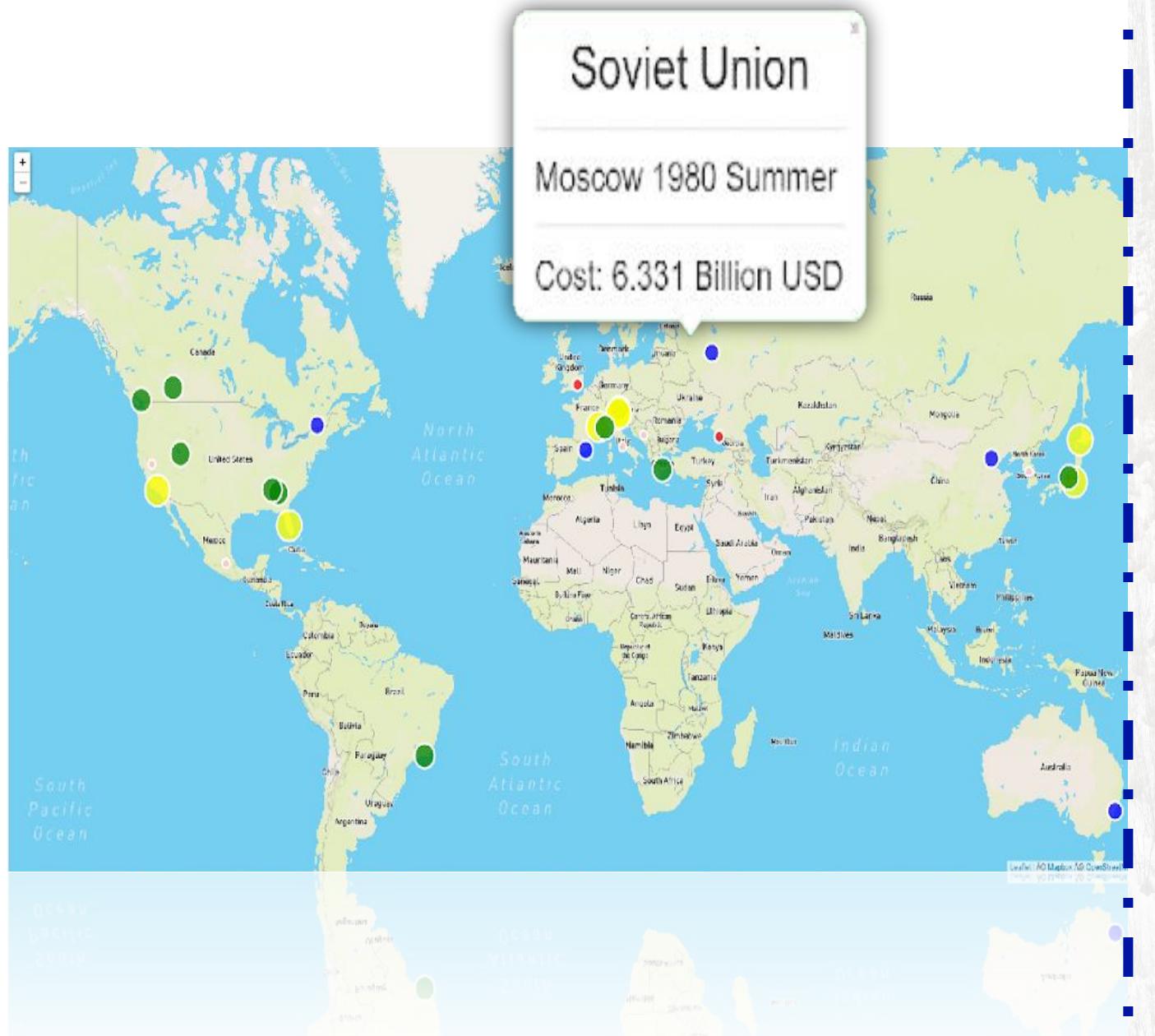
Male & Female participation comparison over time.

Male & Female Athletes over time



LEAFLET MAP VISUALIZATION

COST, YEAR, AND EVENT OF OLYMPICS



Map and Circle Vectors



Pop-Up Display Textboxes



Questions

DATA EMPOWERED

