TABLEAU DASHBOARD
TOPIC:120 Years of Olympic History Visualization
GROUP 9: Morteza K. Haftlang, Rituraj Singh, Samuel Devdas

1. Motivation

This visualization dashboard will give the user a historical overview of the Olympic Games over the past 100 years (1896 to 2016). It gives a big picture of the Games over time and seasons, with respect to participation from different countries, Sports, top Athletes, their performance and physical characteristics like height and weight. Various combinations of queries can be formulated like:

- How many athletes, sports, and nations participate?
- Sport achievements of nations and their global standing point ?
- Which are the top sports in medals and participation?
- Who won the most medals, when and where?
- What are the physical characteristics of the athletes (e.g., sex, height, weight)?

2. Data

Description of the Dataset: The downloaded file *athlete_events.csv* contains 271116 rows and 15 columns. Each row corresponds to an individual athlete competing in a particular sport event. The columns are:

Column Name	Description
ID	Unique identity number for each Athlete
Name	Athlete's name
Sex	Male or Female (M or F)
Age	Number of Years
Height	In Centimetres
Weight	In Kilograms
Team	Team name (Generally Country name)
NOC	National Olympic Committee 3-letter country code

Column Name	Description
Games	Year and season
Year	YYYY format (e.g. 2016)
Season	Summer Games or Winter Games
City	Host City
Sport	Name of the sport (e.g. Athletics)
Event	Name of event (e.g. Athletic men's shot put)
Medal	Type of Medal: Gold, Silver, Bronze or NA (No medal won)

3. Process

3.1 Topic selection and Data collection

The main idea was to choose a globally familiar topic with the purpose of ease of understanding for any audience. The dataset was downloaded from Kaggle. (*https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results*). The original author scraped it from '*www.sports-reference.com'* in May 2018.

3.2 Concepts and iterations

For the first concept the ideas included worksheets like World map displaying Total Athlete count by Country with proportionate size circles, Pie chart of Medal type distribution, Historical Stacked Bar chart showing Athlete count by sex.
In the next concept, a Tree map for Countries medal count and Funnel Chart for top sports was added. After Consultation and feedback, the world map was excluded, due to redundancy of the function 'Top countries by medals' with Tree-map.

3.3 Data Exploration

Data was explored using Shell tools, such as grep, cat, wc, head, tail etc.
The basic objects of interest were dataset size, shape, and column header names.
Desktop csv viewer and Tableau data views came in handy for reference views.

3.4 Data Preprocessing using Linux shell commands

1. **Finding Rows containing Outdated "Art Competitions":** Using Shell Command **grep** (search) and **wc -l** (Count lines)**,** 3578 rows were found which needed to be removed.
    - ***Command****:* grep "Art Competitions" Olympic_history_120yrs.csv | wc -l

2. **Removal of Rows containing Outdated "Art Competitions":** Using Shell Command **grep,** created a new file called "OlympicWoArts.csv (267539 rows) without the rows belonging to "Arts Competitions". (3578 rows were removed from "Olympic_history_120yrs.csv")
    - ***Command****:* grep -v "Art Competitions" Olympic_history_120yrs.csv > OlympicWoArt.csv

3. **Column name substitution:**Using Shell Command **sed** (find and replace)**,** created a new file called "Olympic_Update.csv", in which the Column headers 'ID', 'Name', 'NOC', 'City' were substituted with 'ID_Athlete', 'Athlete_Name', 'Country_Name', and 'Host_City' respectively.
    - ***Sh. Command****:* sed '1 s/ID/ID_Athlete/g; s/Name/Athlete_Name/g; s/NOC/Country_Name/g; s/City/Host_City/g' OlympicWoArt.csv > Olympic_Update.csv

3.5 Tableau Dashboard Design, Feedback, and iterations

For the initial dashboard design concept, some rough layouts were sketched to picture the sequence and placements of the 5 worksheets decided in section 3.2. The layout design starting from top right position and moving anticlockwise is as follows (Figure 1):

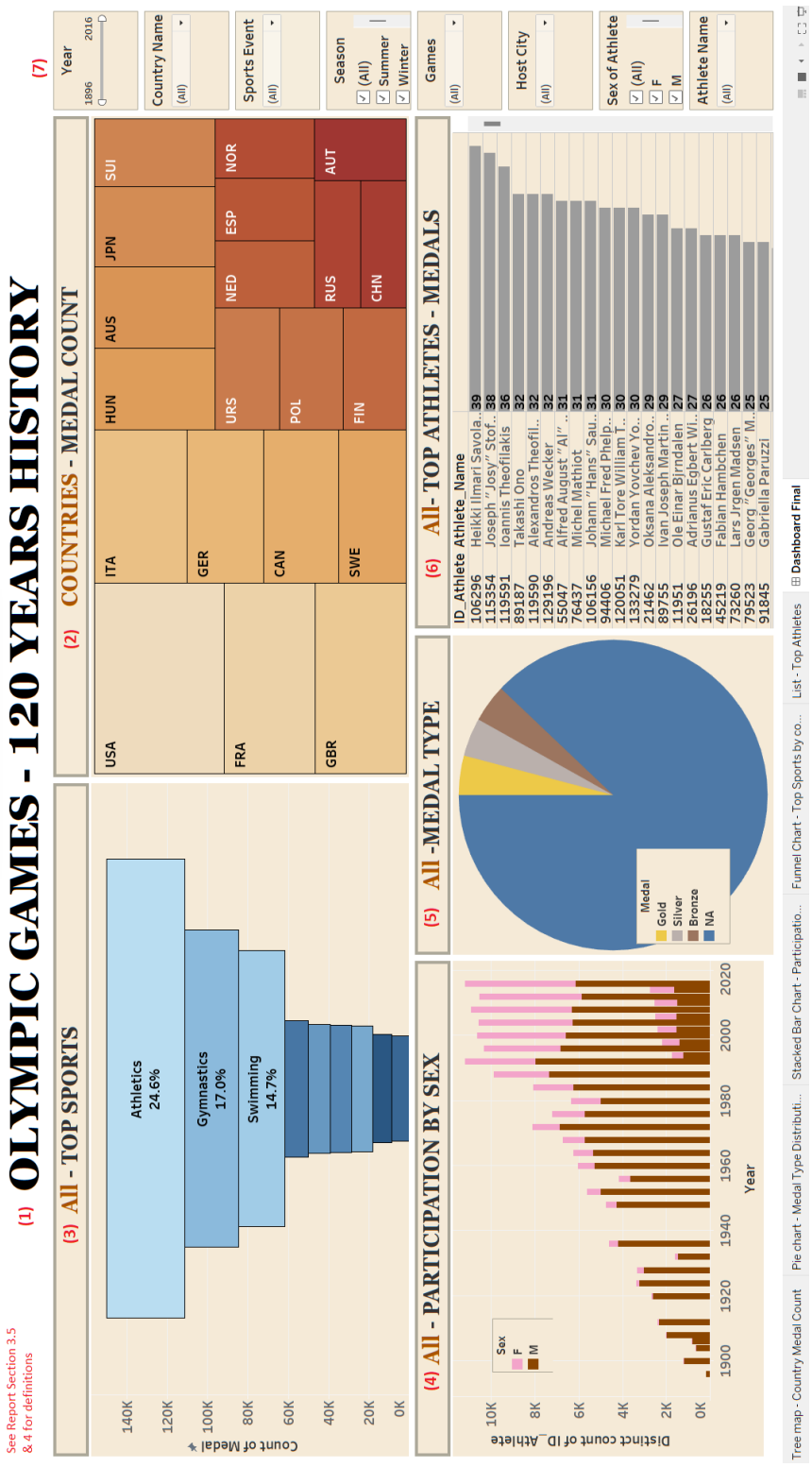
1. Top: Main Title – Olympic Games – 120 Years History
2. Top right: Tree Map – Top 20 Countries by Medal Count
3. Top left: Funnel Chart – Top Sports by Proportion of Medal Count
4. Bottom left: Historical Stacked Bar Chart – Participation by Sex
5. Bottom middle: Pie Chart – Medal Type Distribution
6. Bottom right: Top Athletes List with Horizontal Bar chart for Medal Count
7. Far Right: Vertical multiple filter panel for Customization

Note: Read more in '4. Usage'

3.6 Final Design and Optimizations

The Final Dashboard design took shape after multiple team meetings and coaching sessions with the course professor, where Visual design aspects were discussed in detail. The top discussion topics were: Chart type selection, Layout and placement, Color & Style details, etc.

Figure 1.0


(1) OLYMPIC GAMES - 120 YEARS HISTORY
See Report Section 3.5 & 4 for definitions
(2) COUNTRIES - MEDAL COUNT
USA
FRA
GBR
ITA
GER
CAN
SWE
HUN
AUS
URS
POL
FIN
JPN
NED
RUS
CHN
ESP
SUI
NOR
AUT
(3) All - TOP SPORTS
Athletics 24.6%
Gymnastics 17.0%
Swimming 14.7%
Count of Medal
(4) All - PARTICIPATION BY SEX
Sex
F
M
Distinct count of ID_Athlete
Year
(5) All -MEDAL TYPE
Medal
Gold
Silver
Bronze
NA
(6) All- TOP ATHLETES - MEDALS
ID_Athlete
Athlete_Name
106296 Heikki Ilmari Savola 39
115354 Joseph "Josy" Stof 38
119591 Ioannis Theofilakis 36
89187 Takashi Ono 32
119590 Alexandros Theofil 32
129196 Andreas Wecker 32
55047 Alfred August "Al" 31
76437 Michel Mathiot 31
106156 Johann "Hans" Sau 31
94406 Michael Fred Phelp 30
120051 Karl Tore William T 30
133279 Yordan Yovchev Yo 30
21462 Oksana Aleksandro 29
89755 Ivan Joseph Martin 29
11951 Ole Einar Bjrndalen 27
26196 Adrianus Egbert Wi 27
18255 Gustaf Eric Carlberg 26
45219 Fabian Hambchen 26
73260 Lars Jrgen Madsen 26
79523 Georg "Georges" M 25
91845 Gabriella Paruzzi 25
(7)
Year
1896
2016
Country Name
(All)
Sports Event
(All)
Season
(All)
Summer
Winter
Games
(All)
Host City
(All)
Sex of Athlete
(All)
F
M
Athlete Name
(All)
Tree map - Country Medal Count
Pie chart - Medal Type Distributi...
Stacked Bar Chart - Participatio...
Funnel Chart - Top Sports by co...
List - Top Athletes
Dashboard Final

4. Usage

The dashboard design is intended to be self-explanatory, so that the user quickly gets familiar with the interfaces of the individual charts.

4.1 Tree Map – Top 20 Countries by Medal Count

The Tree Map displays the top 20 countries ranked in descending order through the relative size of the respective box. On hovering above any box, the tooltip displays the Integer Count of the total medals won by the respective country. On clicking the box of a particular country, it will act as a filter and all other Worksheet titles will display the currently selected country name leading its own Title.
E.g. Clicking on USA > All – Top Sports >> USA – Top Sports
Additionally, the sheet information displayed will also be updated for the currently selected country.

4.2 Funnel Chart – Top Sports by Proportion of Medal Count

Navigating counterclockwise, by default the Funnel Chart displays the percentage proportion of medals won per sport category in descending order.
E.g. No filter applied – Global percentage proportion of medals won by sport till 2016 > Athletics 24.6%, Gymnastics 17.0%, Swimming 14.7% and so on. On hovering above any Sport category, the tooltip displays the total count of medals as well.
When a filter is applied, like in case USA Treemap box is selected, then the Funnel chart shows information for USA only > USA > Athletics 29.1%, Gymnastics 18.3%, Swimming 14.9% and so on.

4.3 Historical Stacked Bar Chart – Participation by Sex

Now, navigating down, the Historical Stacked Bar Chart – Participation by Sex, shows the historical growth of participation in the Olympics Games 1896 to 2016, with respect to the sex categories male (Brown) and female (Pink). In the default state, the chart shows the Global Count of unique Athletes year by year. On hovering above any stacked bar, the tooltip displays the total count of male or female Athlete participation. If the filter is Applied by Tree map and/or Funnel chart, the Stacked bars accordingly update.

4.4 Pie Chart – Medal Type Distribution

Next moving right, the Pie Chart – Medal Type Distribution, presents the proportion of Gold, Silver, Bronze and NA i.e. No result, for Global Olympic performance history by default. Again, the Pie chart will update on filtering by other sheets or explicit filter combinations.

4.5 Top Athletes List – Horizontal Bar chart for Medal Count

Finally, the last visualization panel is Top Athletes List with Horizontal Bar chart for Medal Count. It displays a ranked list of Top performing athletes with their integer total medal count next to the names. On hovering above any bar, the tooltip additionally displays the Name, sex, height, weight, and Country of the respective athlete. Here, as well, the list will update on filtering by other sheets or explicit filter combinations.

4.6 Vertical multiple filter panel for Customization

Lastly, the most important feature is the multi filter panel located vertically on the far right. The filters in top to bottom order are Year, Country Name, Sport Event, Season, Games Host City, Sex of Athlete, and Athlete Name.

5. Problems & Solutions

Some Challenges and Solutions:

- Finding an approach to preprocess data using shell tools : Due to our dataset's fairly simple cleaning needs, just 3 shell tools namely 'grep', 'wc -l', and 'sed' sufficed.
- Utilizing Height and Weight attributes meaningfully within the chosen visualizations : A simple way in the end was to use those in the 'Top Athletes list' tooltip.
- Due to the non-unique named Athletes : The 'Top Athletes list' needed to include the unique 'Athlete ID' column, even though it seems redundant.

6. Outcome and Reflection

Ultimately, a multi panel, cross Interactive Dashboard visualization with linked and customizable filtering was realized, which yields insights in a few clicks. Some possible future improvements could be utilizing the age attribute of Athletes, making sports section more detailed (e.g. Athletics > Men's Discus Throw) using the 'Event' attribute, and bringing back the World map sheet to show countries and host cities.