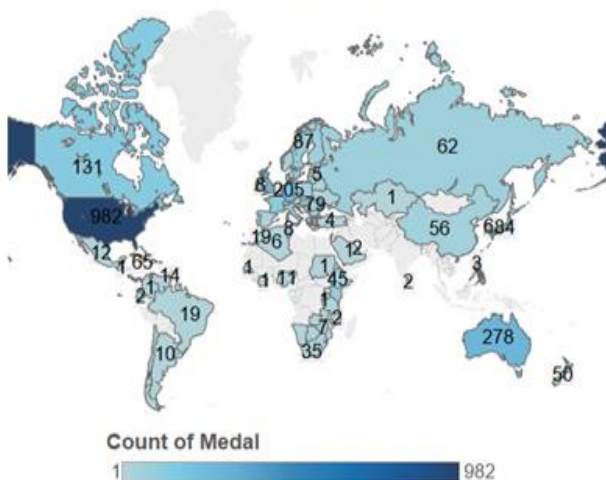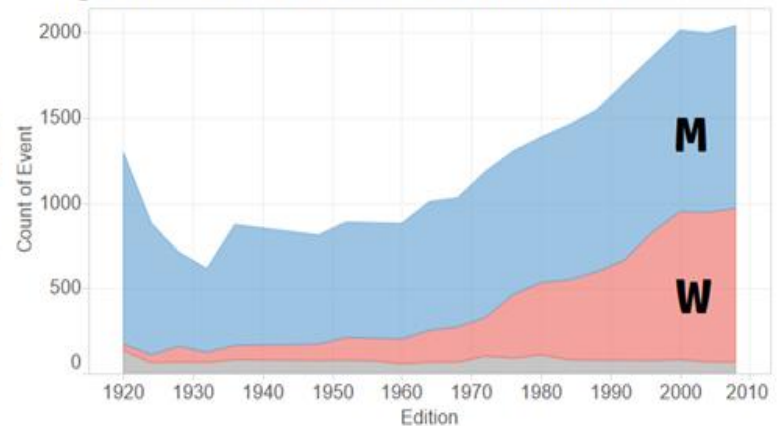# Assignment 2a Tableau Dashboard

Introduction

From the editorial perspectives developed in Assignment 1b, I was particularly intrigued by ideas that explored the Olympics data in relation to gender of the medallists. Given that MedalsData1 (the first sheet of the data) ranges from 1896 to 2012, and women gained an increasing level of rights and power over this time period, I expected to observe an overall increase in the number of events for women and the participation rate of female athletes in the Olympic Games. Furthermore, I wanted to check whether the strongest performing countries (e.g. United States of America) are different for male and female athletes, and whether individual top medallists from 1896 to 2012 actually represent those countries.

Some of the initial idea sketches are

Data Transformation

Although data transformation is not a required task in Assignment 2a, the dashboard is formulated with the transformed data of MedalsData1 (the first sheet of the Olympics data). MedalsData1 is chosen, as it contains information for the entire time period (1896-2012) of past Olympic Games. I will briefly discuss some of the main changes made to the original datasheet.

- The name of the top 10 male medallist, Emil Zatopek, was listed as, "Emil Z?topek". A change has been made to fix this type error.
- Some of the country names (e.g. West Germany, German Democratic Republic and Australasia) have been updated to their current names (e.g. Germany and Australia).
- The embedded gender element in the event column is filtered and appended as a separate column of gender of events/athletes.

These transformations are necessary in order to depict correct and updated information of the Olympic data in the dashboard. Also, given that the main focus of the dashboard is gender, the extraction of gender information from the event column is an integral part of data cleansing.

Dashboard Structure

The tableau dashboard consists of 6 visualisations. While this number is more than what has been required in the assignment, the dashboard focuses on the gender aspect of the Olympics data. Therefore, same visualisations have been reproduced with a different gender for the comparison of male and female athletes. In terms of the types of visualisations, there are 4 visualisation types in the dashboard. In the following section, I will explain the design decisions behind each of 6 visualisations, justifying the use of visual channels, visualisation techniques, and interactions.

Design Decisions

Overall Dashboard

The use of colour, font and composition of visualisations in the dashboard is inspired by the official Olympic website (http://www.rio2016.com/en/olympic-games). The website presents visualisations, using the 5 main colours of the Olympic logo, on a simple white background. In this way, the visualisations are presented in the most minimal, yet effective style.

The dashboard is similarly designed to the website. Since there are a lot of messages trying to be delivered to the audience, I have purposely grounded the background to white. This will allow visualisations to receive more attention and focus. The main colours used in the dashboard title and visualisations resemble the 5 main colours of the Olympics logo and the 3 colours of medals (gold, silver, and bronze). This is to generate a perception that the dashboard is a representation of the Olympics Data.

The dashboard consists of 1 table, 1 stacked bar chart (histogram), 2 simple bar charts and 2 horizontal customized bar charts. The decision behind each visualisation and technique will be explained in details in the following parts. The composition and alignment of the visualisations (charts) are carefully arranged, so that the audience can easily follow from left to right, and top to bottom. The size of the visualisations are also adjusted, so that each chart can sufficiently deliver its message, and the audience can comprehend and interact with the story of the dashboard.

1) Events by Gender Stacked Bar Chart/Histogram (Continuous)

The visualisation shows the total number of events per Olympic Game (Years) by gender across the entire time period from 1896 to 2012. MedalsData1 does not contain the data for group events, and therefore, the category of mixed gender events does not exist in this chart. The chart will allow the audience to examine

      1) The change in the total number of events over time
      2) The change in the number of events for male over time
      3) The change in the number of events for female over time
      4) The overall change in the participation of female athletes in Olympic Games.

The visualisation deals with quantitative (interval) variable of Years and quantitative (ratio) variable of Events (the number of events per Olympic Game). It also compares the composition of the categorical (nominal) variable of gender over time. In order to depict the composition of male and female events,

we can either use **stacked area chart** or **stacked bar chart**. The use of stacked area chart was initially considered, as it would capture the overall trends in the number of events for each gender and the total events. However, although the quantitative variable of Years is continuous, we do not have data for every year; the Olympic Game runs every 4 years. Also, we know that the Olympic Game did not run during the periods of World War 1 and World War 2. Using stacked area chart is not appropriate to show these ideas; the chart will be depicted as if we have data entries for every year, even during the World War periods. Therefore, the relative difference of male and female events is depicted using **vertical stacked bar chart** over the time period from 1896 to 2012. The stacked bar chart is most appropriate in this case to show the relative difference of male and female events, and the overall change in the number of events.

The visualisation incorporates the visual mark of **bars (lines)** to show the number of events per Olympic Game. The magnitude channel of **length/size** is applied on the bars to reflect the magnitude of the number of events. The use of length/size allows the audience to easily distinguish the relative difference of the number of events and indicate when the number was at its largest or smallest.

Within the stacked bar chart, each bar is presented as either a blue line or blue and red segments. The identity channel of **colour hue** is applied to differentiate the number of events by gender (male and female). Events for male are depicted with blue, and events for female are depicted with red. Note that these colours resemble the colours of the official Olympics logo. This is to further strengthen the perception that the visualisation represent the Olympics Data. Also, the colours of blue and red each has the connotation for male and female. These colours will successfully differentiate the gender aspect of the visualisation, without confusing the audience. A separate label for the use of colours is included at the bottom of the dashboard. The use of colour channel for the categorical data of gender is effective, as we can selectively colour the marks (bars), and group the bars representing each year/game into gender (male and female).

Hovering over individual bars shows **labels** including information: Year of the event, Gender that the segmented bar represent, Total number of medallists during the Olympic Game (Year), and Distinct event types. The decision on the label was to show information that are most relevant to the motivation of the visualisation listed above. Also, hovering over the blue/red segmented bars **highlight** the Male/Female row of the simple text table, **"From 1896 to 2012 (Total)"**. This **interaction** allows the audience to quickly grasp the current (total) levels of event types and medallists, while differentiating the aspect of gender in the data.

2) From 1896 to 2012 (Total) Simple text table
The visualisation located at the right top corner of the dashboard is a simple text table, showing the total number of event types and medallists from 1896 (the first Olympic Game) to 2012 (the last Olympic Game) by gender. Note that the **table** has a very simple structure, presented in a **greyscale** with no distinct colour, and contains numerical information.

The visualisation deals with categorical (nominal) variable of gender and two quantitative (ratio) variables of event types and medallists. Although other statistical graph types (e.g. bar chart, area chart, line chart) could be used to show the incremental changes of the variables over time, the motivation

behind the **simple text table** is to fast-deliver the necessary numerical information (the total number of event types and medallists by gender) to the audience, which supplements to the visualisation, **"Events by Gender"**.

Also, the table should not receive much attention own its own. For this reason, the visualisation does not incorporate any particular visual marks or channels, and is presented in a greyscale with no colouring of cells.

3) Top 10 Countries of Male Medallists Simple Bar Chart (Discrete)

The visualisation shows the top 10 countries of male medallists across the entire time period from 1896 to 2012. Focuses are on the names of the countries, their order (United States of America… Netherlands), and the count of medals (Male Medallists) for each country represented by bars. The visualisation allows us to examine which are the 10 strongest performing countries in the Olympic Games, using the total count of medals for male athletes as an index.

The visualisation deals with categorical (nominal) variable of Country Name and quantitative (ratio) variable of Count of Medals for male athletes (all gold, silver, and bronze medals). In order to list the top 10 countries, we can use **choropleth map, packed bubble,** or **simple bar chart**. The use of choropleth map was initially considered as it would generate a geographical data with a continuous/saturated colour scale representing the number of medals. However, given that the aim of the visualisation is to list the top 10 countries, the map is inappropriate as it would consist of too many empty regions without any meanings. Then, the use of packed bubble was considered. The packed bubble can depict the count of medals with the visual mark of bubble/circle and channel of size. However, it is difficult to order the countries from first to tenth with the packed bubble visualisation, so this type of graph was also rejected. Therefore, the top 10 countries of male medallists are presented using a **simple bar chart**. This method is most appropriate to list the 10 countries in the counts of medals each country earned and compare the relative magnitude of each count.

The visualisation incorporates the visual mark of **bars (lines)** to show the total count of medals each top 10 country received from 1896 to 2012. The magnitude channel of **length/size** is applied on the bars to show the relative magnitude of medal counts. The use of bars and length/size allows the audience to quantify the difference between two or more bars (countries), and order the countries from highest to lowest (10th highest) medal recipient. The country names are written in a vertical side-way alignment to minimize unnecessary use of dashboard spaces. The count of medal shows tickers for every 100 medals.

Within the bar chart, each bar is presented as a blue line. The identity channel of **colour hue** is incorporated to show that the visualisation is representing male. Once again, the colour (blue) used in the visualisation resembles the colour of the official Olympics logo. This is to further strengthen the perception that the visualisation deals with the Olympics Data. The colour is coherent with the blue colour used in **"Events by Gender"** and the bottom label.

Hovering over individual bars shows **labels** including information: Country name and Total count of medals from 1896 to 2012. The decision on the label was to show information that are most relevant

data for the motivation of the visualisation: top 10 strongest performing countries in the Olympic Games, using the total count of medals for male athletes as an index.

Interactions among **Top 10 Countries of Male Medallists**, **Top 10 Countries of Female Medallists**, **Top 10 Male Medallists**, and **Top 10 Female Medallists**, will be explained in a separate part below.

4) Top 10 Countries of Female Medallists Simple Bar Chart (Discrete)

This visualisation is a reproduction of **3) Top 10 Countries of Male Medallists**, using the data for female medallists instead of male medallists. The only differences are

1) The use of female data instead of male data,
2) The use of red colour representing female medallists instead of blue colour,
3) The order of the countries, which is based on the total count of medals (female medallists) each country earned over the time period.

Same design decisions were applied for this visualisation.

5) Top 10 Male Medallists Customized Horizontal Bar Chart (Discrete)

The visualisation shows the top 10 male medallists across the entire time period from 1896 to 2012. Focuses are on the number of medals received, the types of medals received, the name of the athlete, and the country that the athlete represent. The visualisation presents the attributes of athlete(s) (names) and country name in the first two columns and charts medals in the third column. Each row represents items, which are the top 10 male medallists of MedalsData1.

The visualisation deals with categorical (nominal) variables of Athlete(s) (names) and country name, categorical (ordinal) variable of medals, and quantitative (ratio) variable of medal counts. Given that the ultimate aim of the visualisation is to compare the medal count of top 10 male medallists and sort them into the list in a descending order, we know that **simple bar chart** is most applicable as explained in 3) and 4). A horizontal layout will be applied instead of a vertical layout. This is to give more emphasis on the names and countries of the athletes, and compare the count and type of medals more efficiently.

The visualisation incorporates the visual mark of **circles (medals)** to show the count of medals for top 10 male medallists and the magnitude channel of **size/area** to adjust the circles to an adequate, same size that represent medals. This will allow the audience to quickly recognise that the visualisation represents Olympic medals, and easily quantify the difference between the counts of medals for two or more medallists by simply counting the number of circles. Furthermore, they can order the medallists from the highest to the lowest using the visual mark.

The identity channel of **colour hue** is incorporated to represent the categorical (ordinal) data of medals (gold, silver and bronze). This allows the audience to draw a stronger resemblance to the Olympic Games from the visualisation and compare the relative count of medals among the category of gold, silver, and bronze. By colouring the different types of medals, the audience can now compare two or more medallists and observe stronger performers in each ordinal category (gold>silver>bronze).

Hovering over individual bars shows **labels** including information: Event (sports type), Medal (type), and Games (Year of the Olympic Game). The decision on the label was to show information that are most

relevant data for the motivation of the visualisation: top 10 medallists using the total count of medals for male athletes as an index. The information on year and event was included to give an indication when and which event each medal is from.

6) Top 20 Female Medallists Customized Horizontal Bar Chart (Discrete)

This visualisation is a reproduction of **5) Top 10 Male Medallists**, using the data for female medallists instead of male medallists. Same design decisions were applied for this visualisation.

7) Interactions among Top 10 Countries of Male Medallists, Top 10 Countries of Female Medallists, Top 10 Male Medallists, and Top 10 Female Medallists

The interactions in the dashboard are designed so that they do not interrupt the process of understanding of the audience, but support it by either filtering or highlighting relevant charts when the audience hover around item entries of the MedalsData1 dataset.

The visualisations above listed are interacted to each other to enhance the understanding of the audience. Hovering around bars or country names of **Top 10 Countries of Male/Female Medallists** will filter the medallist entries in **Top 10 Male/Female medallists** to the country specific level. This will allow the audience to examine whether individual top medallists from 1896 to 2012 belong to top 10 countries of male/female medallists (the strongest performing countries). This is a good indication for whether countries with high total counts of medals also produce top performing athletes (e.g. Michael Phelps from United States of America). In the dashboard, this interaction allows the audience to see that this is not always the case; there are no top medallists filtered out for Germany and Great Britain.

Another interaction is between **Top 10 Countries of Male Medallists** and **Top 10 Countries of Female Medallists.** Hovering around bars or country names of one of these visualisations will highlight if the country is also present in the male/female counterpart of the visualisations. This allows the audience to examine whether the country of Top 10 Countries of Male/Female Medallists is also listed in the female/male counterpart of the visualisation. In the dashboard, we can observe the top 10 countries are same for both male and female medallists, although their orders are different.