

Prerequisites

- You have installed Tableau Desktop on your computer.
Available here: <http://www.tableau.com/academic/students>
- You have downloaded the data (athlete_events.csv)
available here: <https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results> or here
https://www.dropbox.com/s/6h5n8pcfwwc6wdl/athlete_events.csv?dl=0

Motivation

With Tableau you can quickly create visualizations of your data. Its ease of use makes it a valuable tool for initial data exploration and it also allows you to create complex interactive visualizations (e.g. for prototyping).

You can create dashboards consisting of multiple views on the data and options for interaction. Here are some examples:

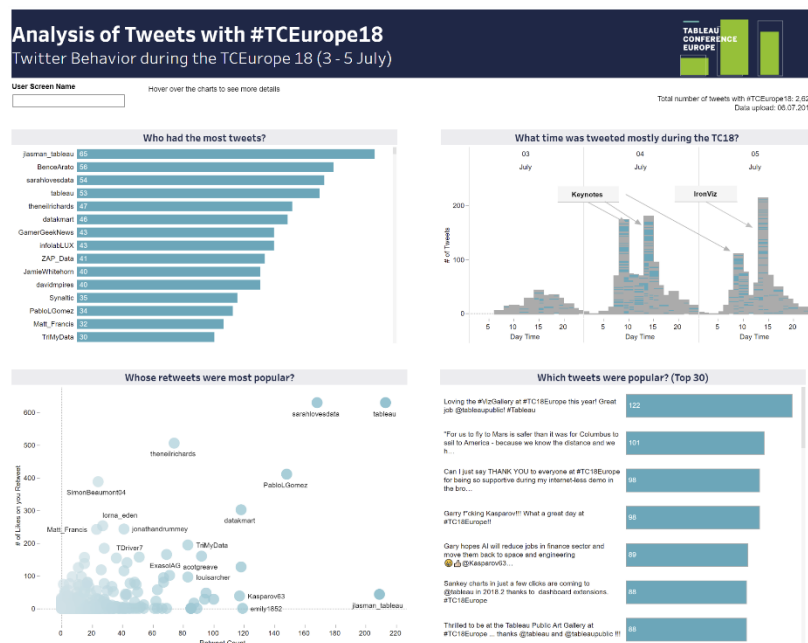


Figure 1: <https://public.tableau.com/en-us/s/gallery/tceurope18-tweets-analysis?gallery=votd>

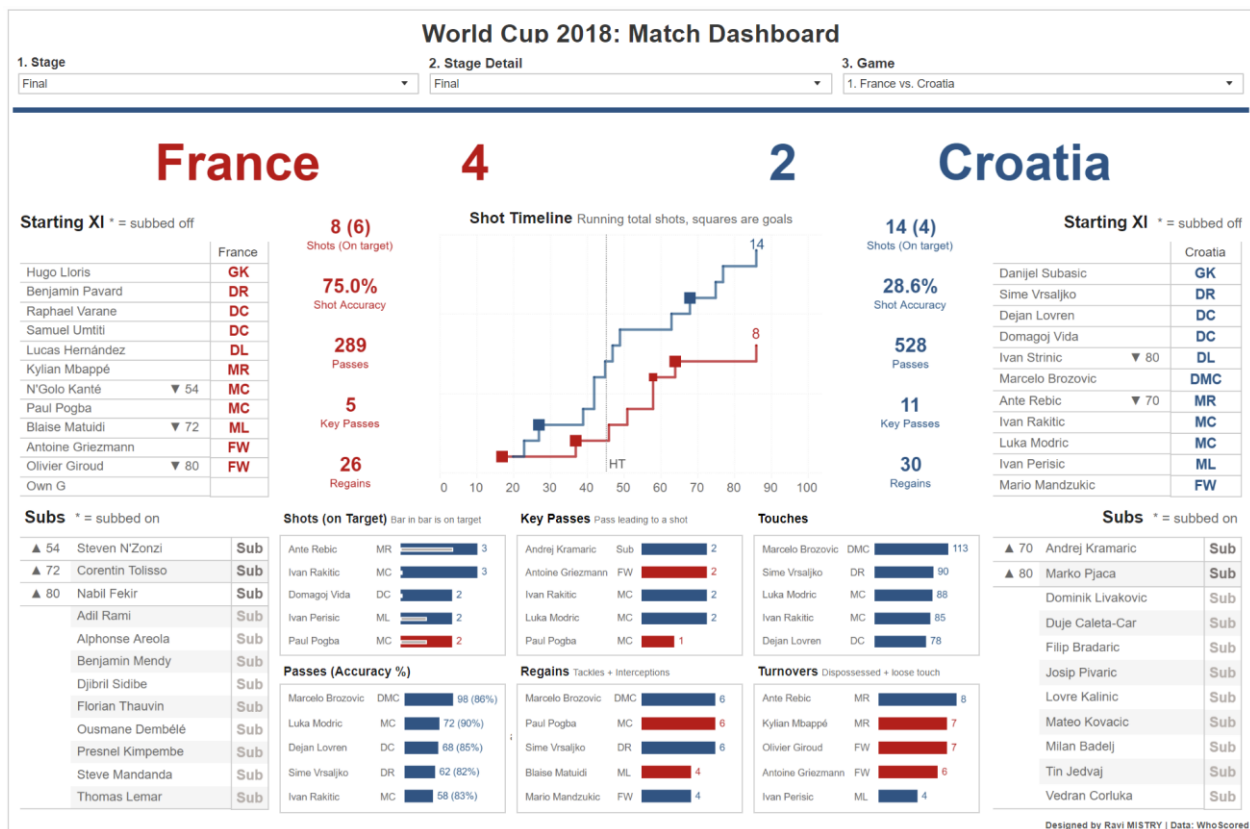


Figure 2: <https://public.tableau.com/en-us/s/gallery/world-cup-match-dashboard?gallery=votd>

the mediGAP

SOURCE | business insider • DESIGN | lindsay betzendahl • PROJECT | project health viz

Once you turn 65 you qualify for Medicare health insurance. However, Medicare doesn't cover everything. Medigap plans are sold by private companies to help pay for some of those extra costs. Depending on where you live, it will cost you.

Annual Average Medigap Plan Cost by region

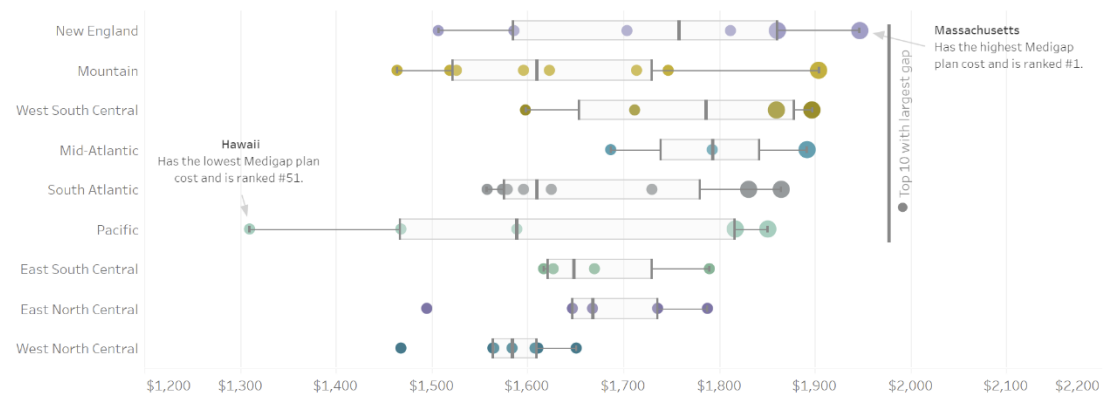
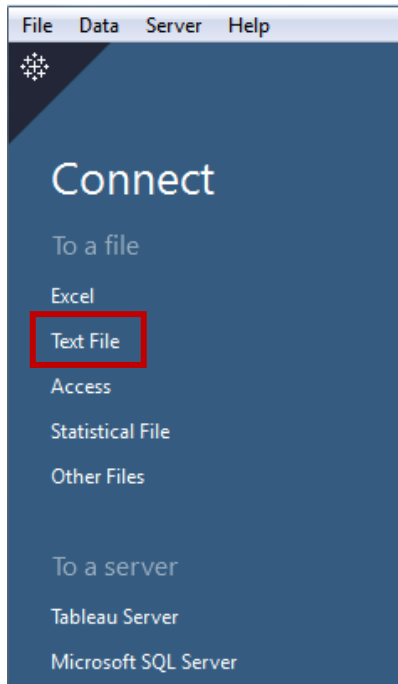


Figure 3: <http://public.tableau.com/profile/datalicious.ptv.ltd#!/vizhome/MalteSpitzCallData/MalteSpitzcalldatadashboard>

Loading Data



The first step to the visualization is to load the data. On the left side of the screen in the section **To a File** choose **Text File**. This way you can open the .csv file.

After you have selected the data file, you will see an initial table representation of your data. For the data entries to appear, you might have to click **Update Now**.

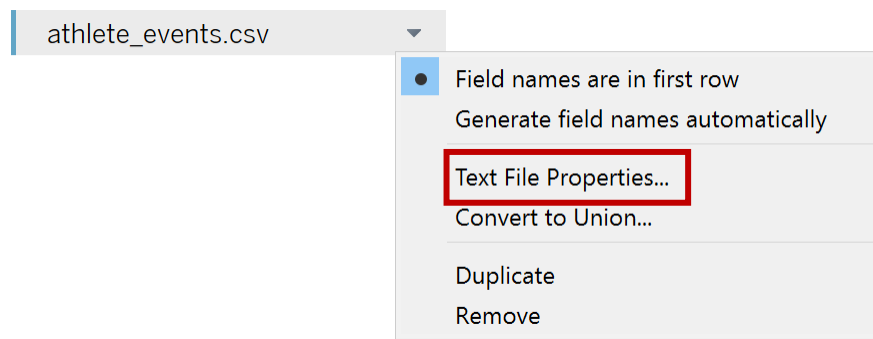
athlete_events

Connection: ☒ Live ☐ Extract Filters: 0 | Add

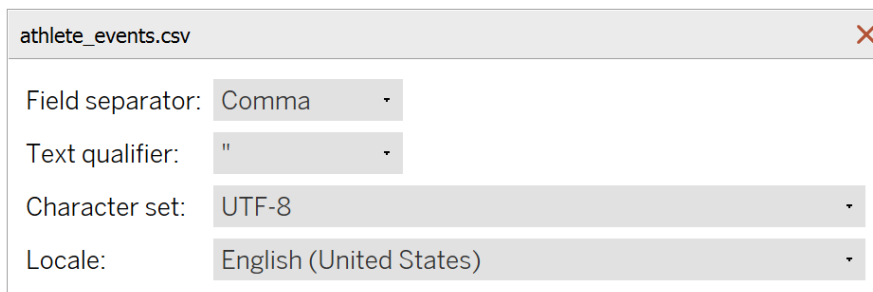
athlete_events.csv

#	athlete_events.csv	athlete_events.csv	athlete_events.csv	athlete_events.csv	athlete_events.csv	athlete_events.csv	athlete_events.csv	athlete_events.csv	athlete_events.csv	athlete_events.csv	athlete_events.csv	athlete_events.csv	athlete_events.csv	athlete_events.csv
ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
1	A Dijiang	M	24	180	80	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Bas...	NA
2	A Lamusi	M	23	170	60	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lig...	NA
3	Gunnar Nielsen Aabye	M	24	NA	NA	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NA
4	Edgar Lindenaau Aabye	M	34	NA	NA	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug...	Gold
5	Christine Jacobsa Aaft...	F	21	185	82	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Wome...	NA
5	Christine Jacobsa Aaft...	F	21	185	82	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Wome...	NA
5	Christine Jacobsa Aaft...	F	25	185	82	Netherlands	NED	1992 Winter	1992	Winter	Albertville	Speed Skating	Speed Skating Wome...	NA

To make sure our data is correct we need to check what encoding is used to load the data.



Click the little triangle next to your data file in the large white area and select **Text file properties**.



Then select **Comma** as separator, " as text qualifier and **English (United States)** as locale. These settings depend on the dataset you have!

Sort fields: Data source order

#	athlete_ev... ID	athlete_events.csv Name	athlete_events... Sex	#	athlete_events.c... Weight
1	A Dijiang	M			80
2	A Lamusi	M			60
3	Gunnar Nielsen Aaby	M			null
4	Edgar Lindenau Aabye	M			null
5	Christine Jacoba Aafti...	F			82
5	Christine Jacoba Aafti...	F	21	185	82
5	Christine Jacoba Aafti...	F	25	185	82
5	Christine Jacoba Aafti...	F	25	185	82
5	Christine Jacoba Aafti...	F	27	185	82
5	Christine Jacoba Aafti...	F	27	185	82
6	Per Knut Aaland	M	31	188	75
6	Per Knut Aaland	M	31	188	75
6	Per Knut Aaland	M	31	188	75
6	Per Knut Aaland	M	31	188	75
6	Per Knut Aaland	M	33	188	75

Number (decimal)

Number (whole)

Date & Time

Date

String

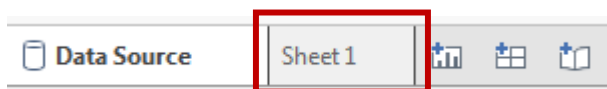
Boolean

Default

Geographic Role

Check the datatype of each column. Sometimes a NA in column results in a text column even though column describes something like the weight. Change the columns to the appropriate data types.

Spending some time here saves time afterwards. You may also go back to this screen any time to fix something.



After you have set the correct format, the table is useable. Click **Sheet 1** at the bottom of the window to proceed to your worksheet.

Basics

Dimensions & Measures

The screenshot shows the Tableau interface with the 'Data' pane on the left. The 'Data' pane is divided into two sections: 'Dimensions' and 'Measures'. The 'Dimensions' section lists 15 fields: Age, City, Event, Games, Height, ID, Medal, Name, NOC, Season, Sex, Sport, Team, Weight, and Year. The 'Measures' section lists 4 fields: Latitude (generated), Longitude (generated), Number of Records, and Measure Values. The 'Data' pane is titled 'Data' and 'Analytics' is selected. The dataset 'athlete_events' is selected. The 'Dimensions' section is titled 'Dimensions' and has a search icon. The 'Measures' section is titled 'Measures'.

Dimensions	Measures
# Age	🌐 Latitude (generated)
🌐 City	🌐 Longitude (generated)
Abc Event	=# Number of Records
Abc Games	# Measure Values
# Height	
# ID	
Abc Medal	
Abc Name	
Abc NOC	
Abc Season	
Abc Sex	
Abc Sport	
Abc Team	
# Weight	
# Year	
Abc Measure Names	

In the leftmost panel you will see the columns of your table as either dimensions or measures.

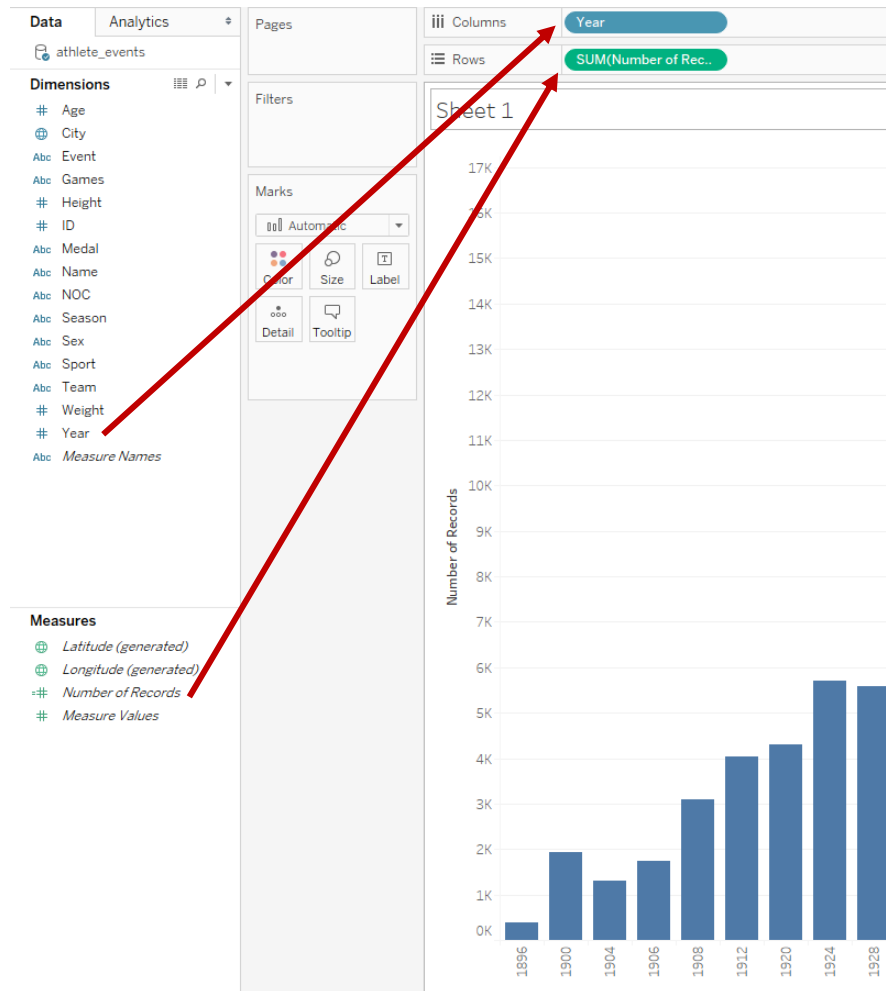
Dimensions are usually categorical datatypes. They can be used to separate your data by discrete tags.

Measures are the quantitative data that you will encode in your marks.

You can change the type by dragging a measure to dimension or the other way. This is useful if the auto detection of Tableau isn't working properly.

Building Charts

You can drag and drop both measures and dimensions to the central panel to create plots.

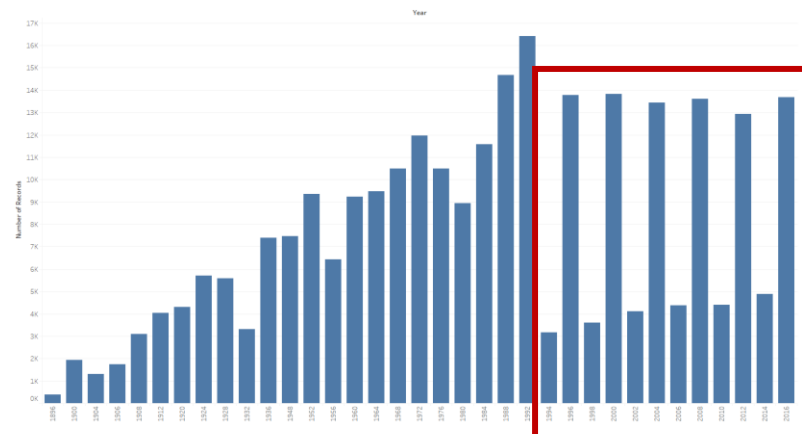


Let's suppose we are interested in the number of athletes per year .

Simply **drag** the dimension **Year** to the columns and the measure **Number of Records** to rows.

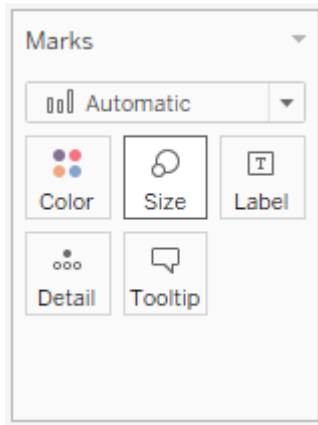
You will see that the records are automatically summed up and we see a bar chart of the athletes per year.

Looking at the chart, the number of participants seems to fluctuate in the last years. Any idea why?



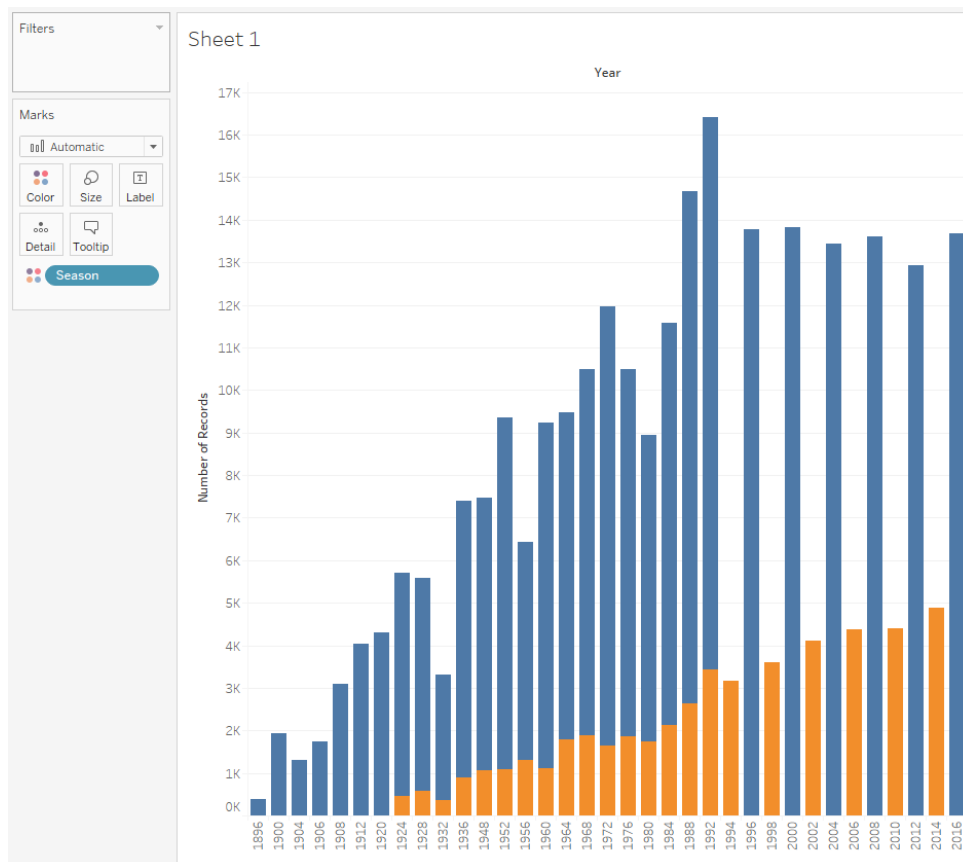
Encoding

We take now a closer look by using additional visual channels.



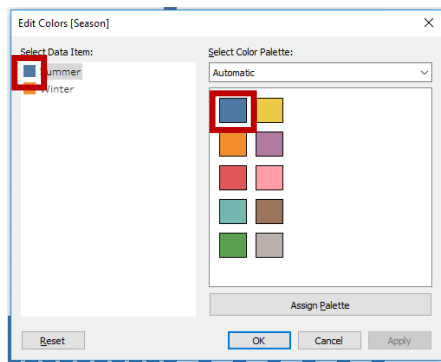
The **Marks** window shows all available channels in Tableau.

Let us drag and drop the **Season** field onto **Color**.



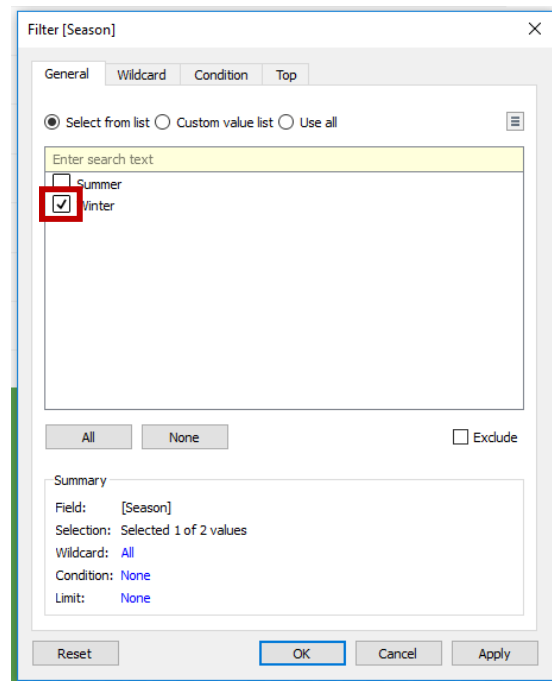
We can see now that in the beginning only summer games were held. After that both types happened during the same year and now they are during different years.

Double clicking the legend allows us to change the colors.

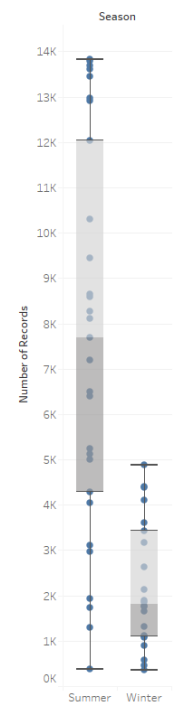


You may select a predefined color from a palette or you double click the color on the left side and select a custom color.

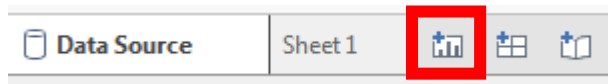
Dragging now **Season** onto the Filter area allows us to remove the summer season from this view.



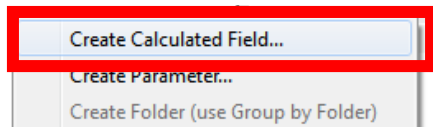
On the right upper side the **Show Me** window provides possible visualization given the current used fields. Let us **remove the Season filter** and select the **box-and-whisker** plot. Which leads to the plot on the right side and a different perspective on the dataset.



Next we create a new plot by selecting the **New Worksheet** button at the bottom of the window.

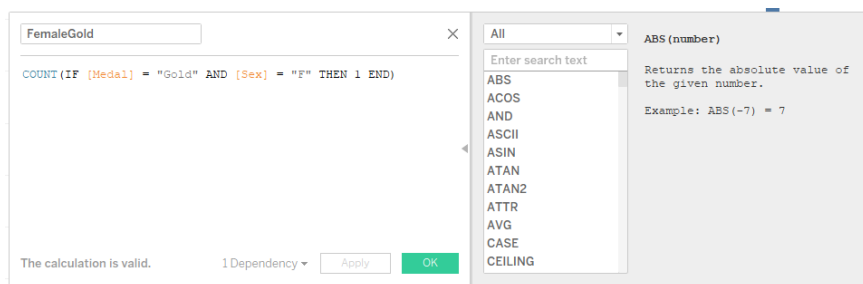


Now let's try to visualize how many women had gold medals during all those years. One way is to filter multiple times until we have what we want. Another option are **calculated fields**.



Create one by right click into the dimension or metrics and select **Create Calculated Field...**

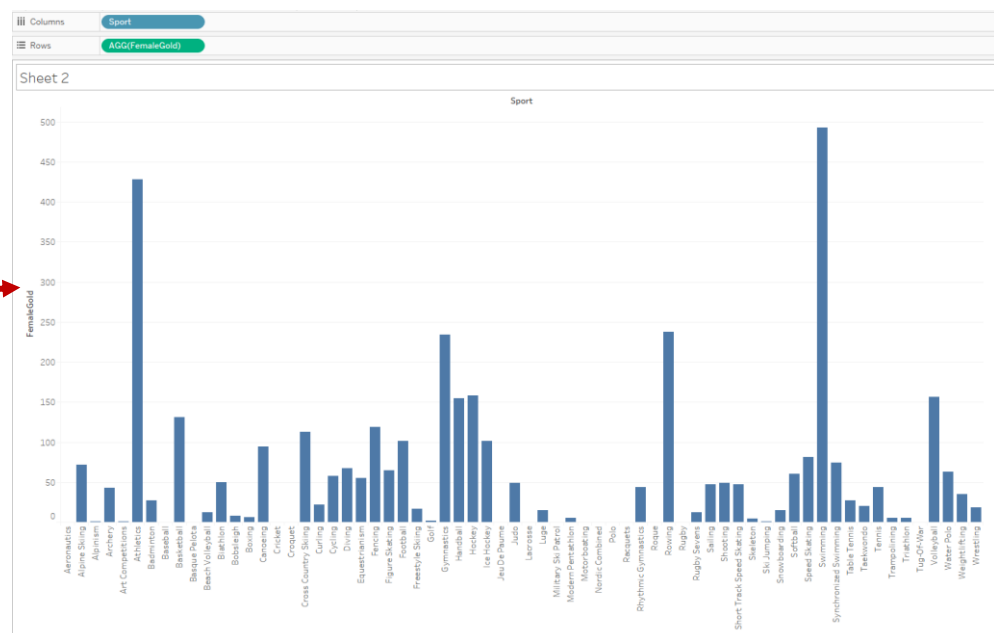
Let's name it FemaleGold and use the formula from the picture below. The square brackets allow you to target certain fields, like the Medal field.



The triangle at the right of this pop-up shows a list of all possible operations. Using this as row and Sport as column shows us the gold medals women won in different sports.

Let us now sort the x-axis by the number of gold medals. Click the small sort symbol directly on the y-axis

In addition, drag the created calculated field onto **Label** in Marks to show the numbers. Finally, we can drag **Season** onto color.



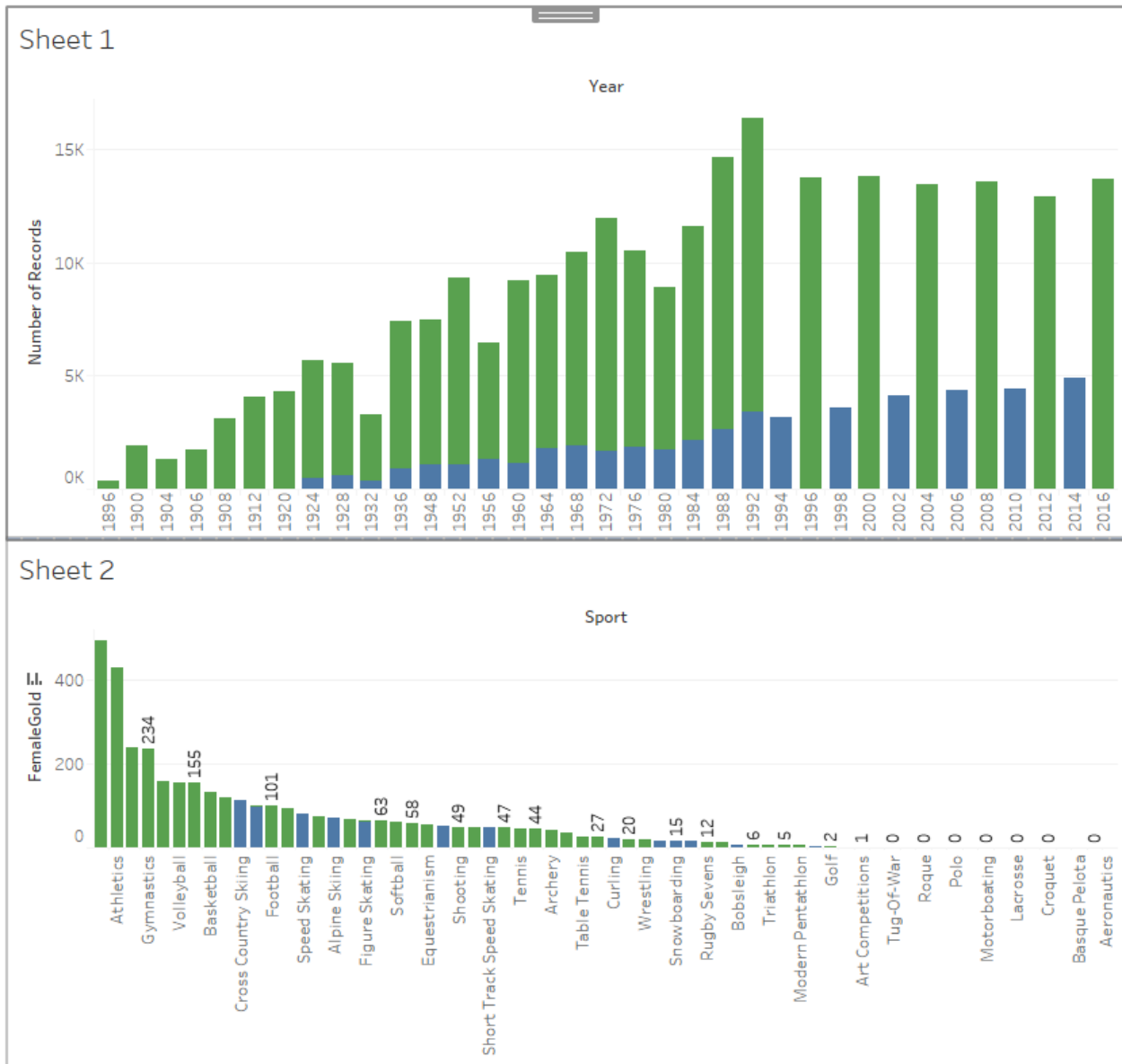
What are the top three sports for female athletes?

Dashboard

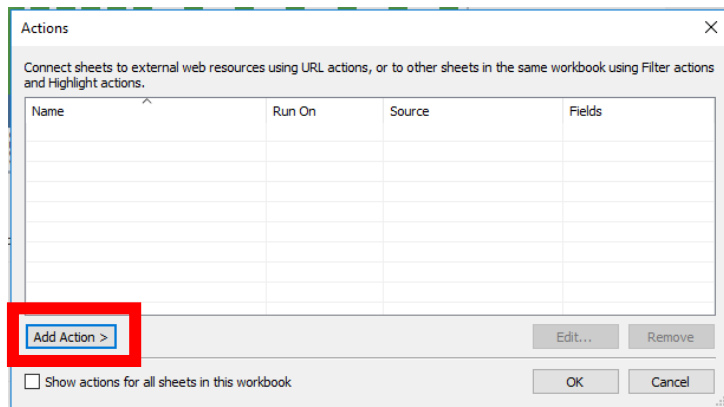
Create a dashboard by selecting the create Dashboard button at the bottom.



Drag and drop the two sheets into the empty space so it looks like this:

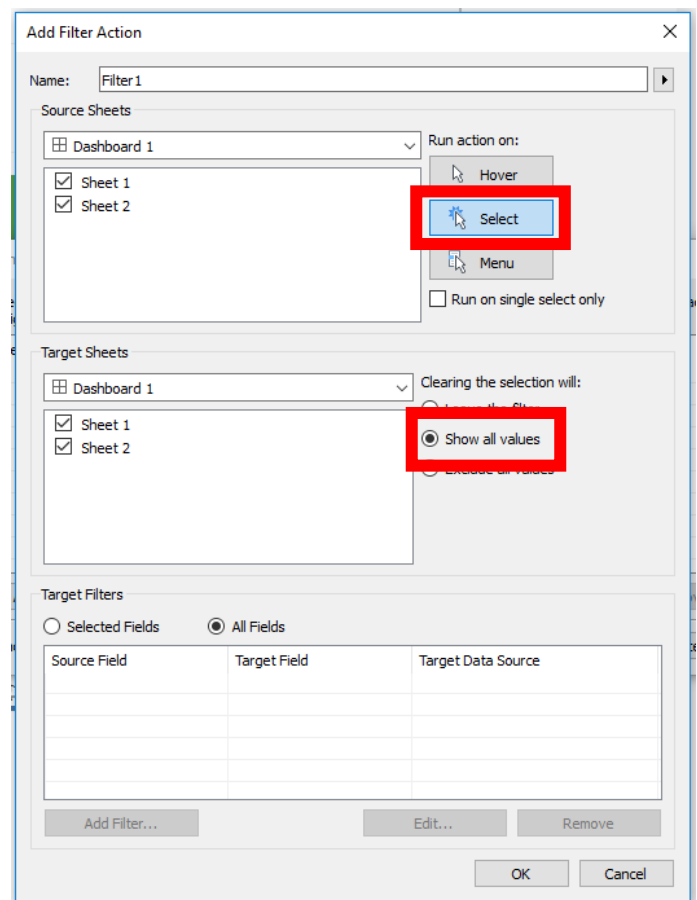


Let us now connect those two sheets though brushing and linking. Open **Dashboard** at the top and select **Actions**.

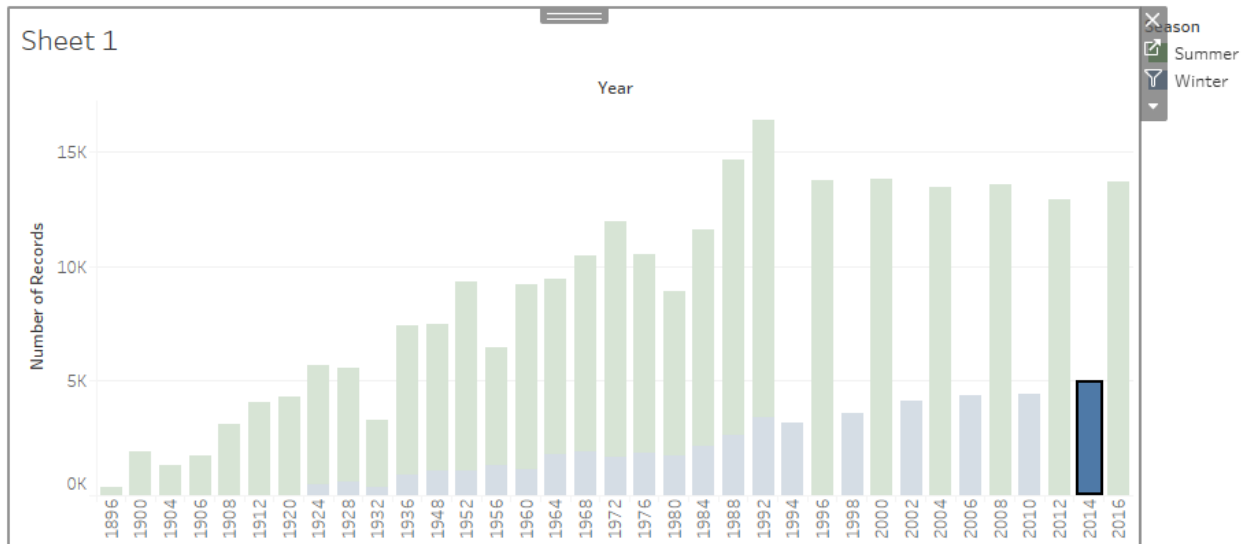


Select **Add Action** and create a **Filter**.

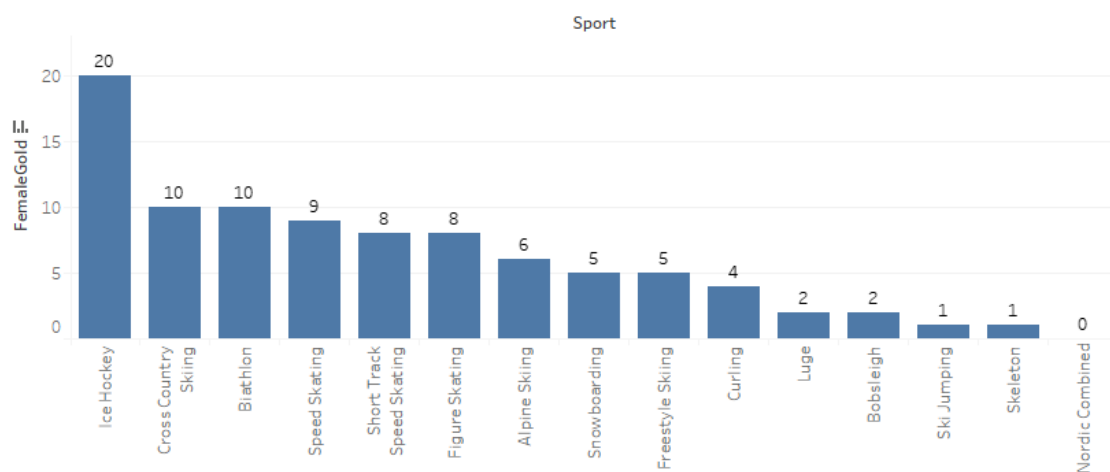
We use both sheets as source as well as target. You might say one sheet is used as a filter for all other sheets as an example. We also use **Select** as method and **show all values** after deselection.



This allows us to analyze the data further. Shown below I selected the winter games from 2014 in the upper chart and the lower chart shows only the female gold medals from this specific game.



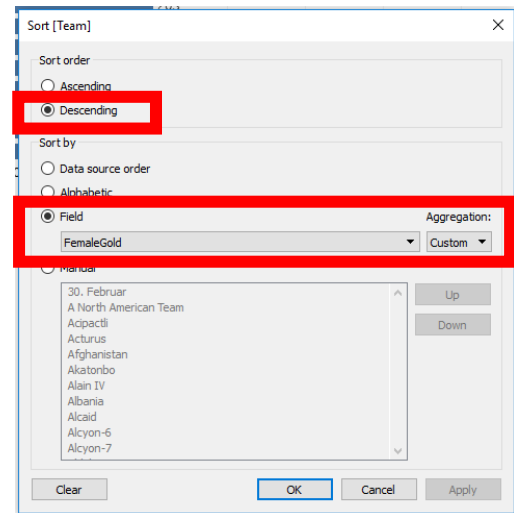
Sheet 2



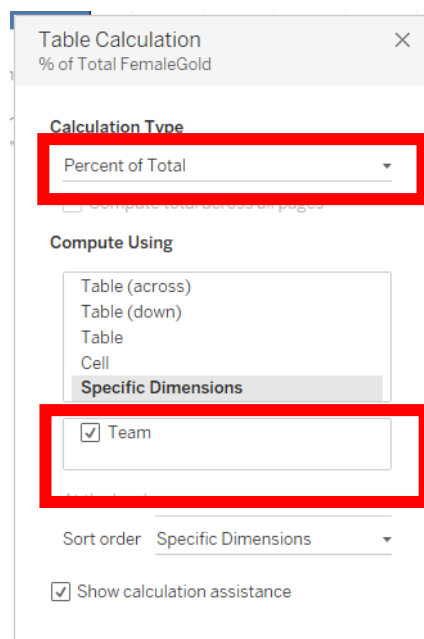
- Which year had most Basketball gold medals for female athletes?
- How many gold medals in athletics had the women in the games 1972?

Let us have a look at the percentages of medals for different countries. Create a new sheet and use **Team** as row, **FemaleGold** as column and label.

Right click **Team** and click at **Sort...** There, select **Descending** and **Sort by Field** as shown on the right side.



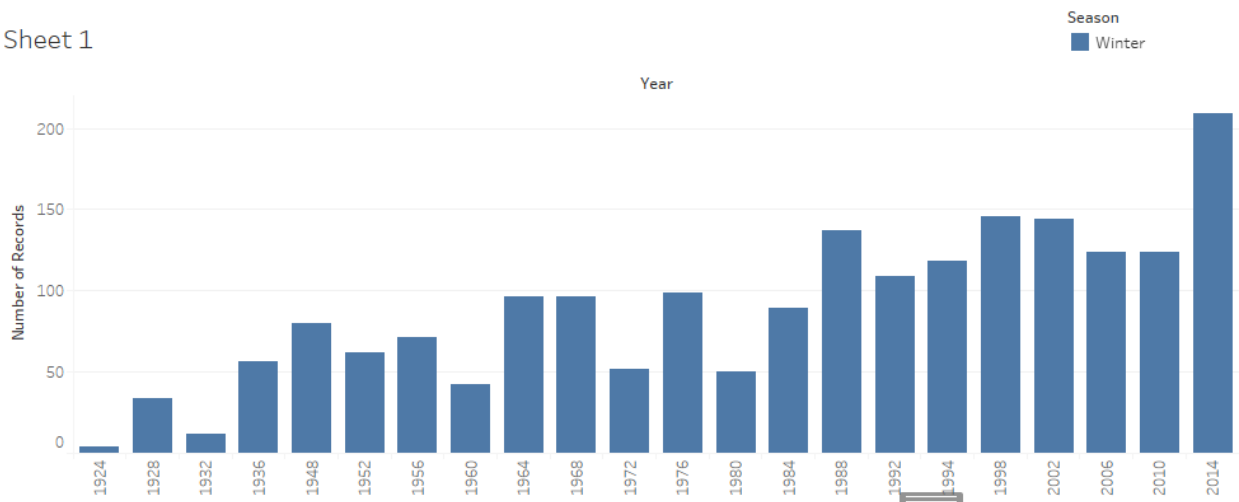
This shows us the total amount of female medals for each team or country.



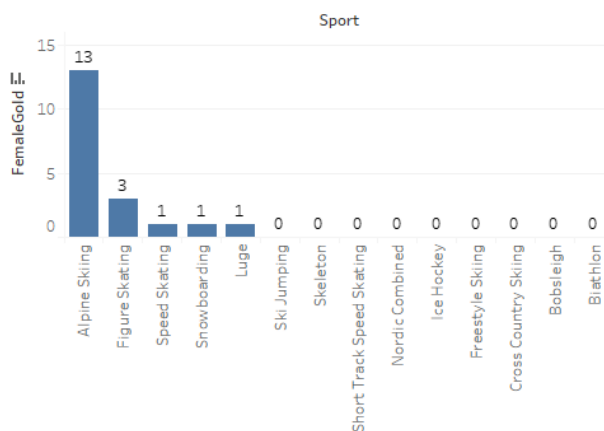
Let us now compare the relative amount of medals between countries. Right click **AGG(FemaleGold)** and select **Add Table Calculation**. Change the calculation type to **Percent of Total** use **Specific Dimensions** with **Team** as dimension. Do the same for the **AGG(FemaleGold)** for the label to see percent instead of the count.

Finally drag **Season** onto color and add the sheet to the dashboard.

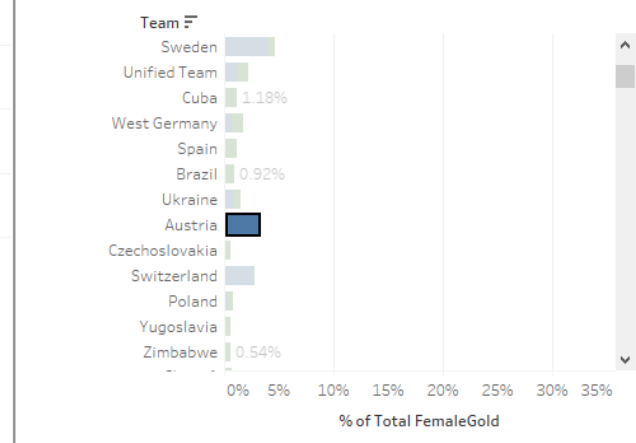
Sheet 1



Sheet 2



Sheet 3



This allows us now to look at the performance of certain countries. Here we see that Austria had only female gold medals in the winter seasons (shown by the color) and most of the medals are won in Alpine Skiing. Play around and look for your own country and compare it to your neighbors.

Create new sheets and try to answer questions like:

- Which age and weight combination results in the most medals in judo?
- Is age and medal somehow correlated (for specific sports or generally)?
- We saw the female gold medals, can you extend it to analyze gold, silver and bronze medals?
 - Show it for male and female athletes