

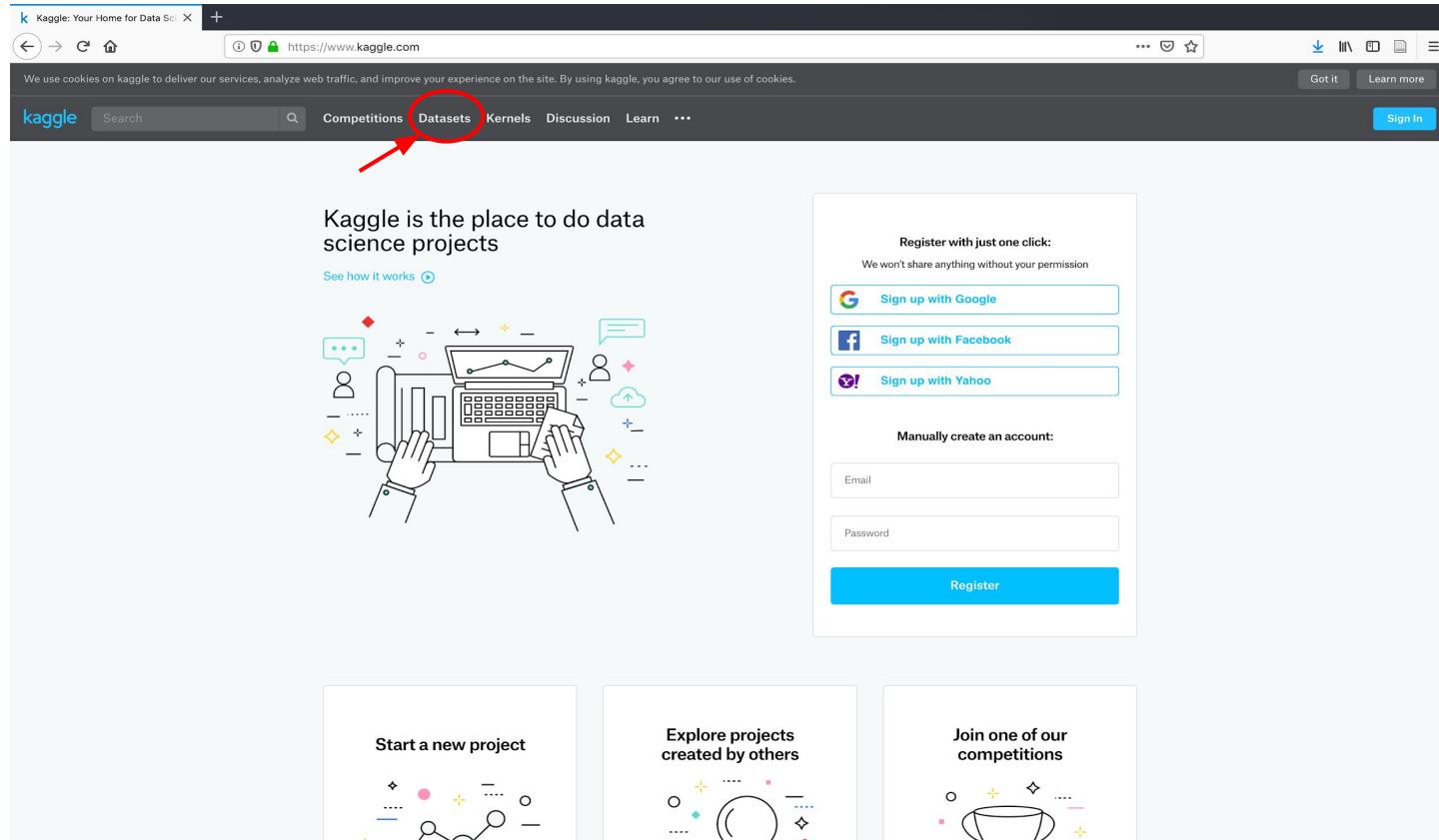
# Lab 1

Spring 2019 DS-1001

# Installing Tableau

- Go to <https://www.tableau.com/tft/activation> and select ‘Download Tableau Desktop’.
- After installing Tableau, open the application and choose ‘Activate’.
- Enter the Product key: **TCQI-0376-28C0-9E43-E64E** and click ‘Activate’.
  - Please do not share this code to people outside the class.
- Register with your NYU email.

# Find Dataset



The screenshot shows the Kaggle homepage. A red arrow points to the 'Datasets' tab in the top navigation bar, which is highlighted with a red circle. The page features a central illustration of hands interacting with a laptop displaying a line graph, surrounded by various icons representing data science concepts like databases, charts, and user profiles. Below the illustration, there are three main calls-to-action: 'Start a new project', 'Explore projects created by others', and 'Join one of our competitions'. To the right, a registration sidebar offers 'Register with just one click' via Google, Facebook, or Yahoo, or 'Manually create an account' with fields for Email and Password, and a 'Register' button.

Kaggle is the place to do data science projects

See how it works 

Start a new project

Explore projects created by others

Join one of our competitions

Register with just one click:

We won't share anything without your permission

 Sign up with Google

 Sign up with Facebook

 Sign up with Yahoo

Manually create an account:

Email

Password

Register 

# Find Dataset

The screenshot shows the Kaggle Datasets interface. At the top, there's a search bar with the query "olympics". Below the search bar, a table lists 14 datasets. A red arrow points to the first dataset in the list, which is highlighted with a red border. This dataset is titled "120 years of Olympic history: athletes and results". It has a thumbnail of two athletes, a rating of 555, and a brief description: "basic bio data on athletes and medal results from Athens 1896 to Rio 2016". The dataset was updated 7 months ago by Randi H Griffin. The table includes columns for Public datasets, Sizes, File types, Licenses, Tags, and a Sort by dropdown set to Relevance.

Rank	Dataset Title	Description	Tags	Size	File Type	Licenses	Comments	Views
65	Olympic Sports and Medals, 1896-2014	Which countries and athletes have won the most medals at the Olympic games?	olympic games, sports, CC0	493 KB	CSV	CC0	2	32
0	World Wushu Championships 2017	Chinese Martial Arts competition data	olympic games, sports, data visualization, Other	51 KB	CSV	Other	0	3
555	120 years of Olympic history: athletes and results	basic bio data on athletes and medal results from Athens 1896 to Rio 2016	olympic games, sports, history, CC0	5.4 MB	CSV	CC0	5	112
0	Olympic Results Biathlon	Biathlon Results from all Winter Olympic Games, 1960-2018	olympic games, sports	5.4 KB	CSV	Other	0	207
11	Top Running Times	Who are the fastest people in the world?	running, CSV, Other	279.6 KB	CSV	Other	0	9
0	olympics_2012	Seek4everd updated a year ago (Version 1)	Other, CC0	6.7 KB	CSV	CC0	0	451
0	olympics	Danan Dio updated a year ago (Version 1)	CC0	395.6 KB	CSV	CC0	0	1k

# Download Dataset

The screenshot shows a Kaggle dataset page for "120 years of Olympic history: athletes and results". The page features a banner with three athletes in action. Below the banner, the title "120 years of Olympic history: athletes and results" is displayed, along with a subtitle "basic bio data on athletes and medal results from Athens 1896 to Rio 2016". The dataset was created by Randi H Griffin and updated 7 months ago (Version 2). The main navigation tabs are Data, Overview, Kernels (112), Discussion (5), and Activity. A prominent blue button labeled "Download (5 MB)" is highlighted with a red circle and an arrow pointing to it. Below the download button, a message says "Your Dataset download has started. Show your appreciation with an upvote". The "Data" tab is selected, showing details about the dataset: Data Sources (athlete\_events.csv, 271k x 15; noc\_regions.csv, 230 x 3), About this file (No description yet), and Columns (ID, Name, Sex, Age, Height, Weight, Team, NOC, Games). At the bottom, a preview of the "athlete\_events.csv" file is shown, displaying columns like Season, City, Sport, Event, Medal, and percentages for Summer (82%) and Winter (18%) seasons.

Dataset Released Under CC0: Public Domain

120 years of Olympic history: athletes and results

basic bio data on athletes and medal results from Athens 1896 to Rio 2016

Randi H Griffin • updated 7 months ago (Version 2)

Data Overview Kernels (112) Discussion (5) Activity

Download (5 MB)

Your Dataset download has started. Show your appreciation with an upvote

555

Data (5 MB)

Data Sources

- athlete\_events.csv 271k x 15
- noc\_regions.csv 230 x 3

About this file

No description yet

Columns

- ID
- Name
- Sex
- Age
- Height
- Weight
- Team
- NOC
- Games

athlete\_events.csv (39.58 MB)

Season	City	Sport	Event	Medal
Summer	London	Athletics	Football Men's Fo...	NA
Winter	Athina	Gymnastics	Ice Hockey Men's ...	Gold

# Open file in Tableau

The screenshot shows the Tableau desktop application interface. On the left, the 'Connect' sidebar is open, listing various data source types. A red arrow points to the 'Text file' option, which is highlighted with a red box. The main workspace is titled 'Open' and displays sample workbooks like 'Superstore' and 'Regional'. To the right, the 'Discover' sidebar provides links to training, sharing, and resources.

Tableau - Book1

File Data Server Help

Connect

To a File

- Microsoft Excel
- Text file**
- JSON file
- PDF file
- Spatial file
- Statistical file
- More...

To a Server

- Tableau Server
- Microsoft SQL Server
- MySQL
- Oracle
- Amazon Redshift
- More...

Saved Data Sources

- Sample - Superstore
- World Indicators

Open

Sample Workbooks

- Superstore
- Regional
- World Indicators

Discover

Open a Workbook

Training

- Getting Started
- Connecting to Data
- Visual Analytics
- Understanding Tableau
- More training videos...

Sharing

Learn more about ways to share

Resources

- Get Tableau Prep
- Blog - Tableau Conference Europe customer speaker applications are now open
- Forums

VIZ OF THE WEEK

Film Genre Popularity Since 1910 →

- ACTION
- WAR
- ROMANCE
- MUSICALS
- HORROR
- WESTERN

# Open file in Tableau

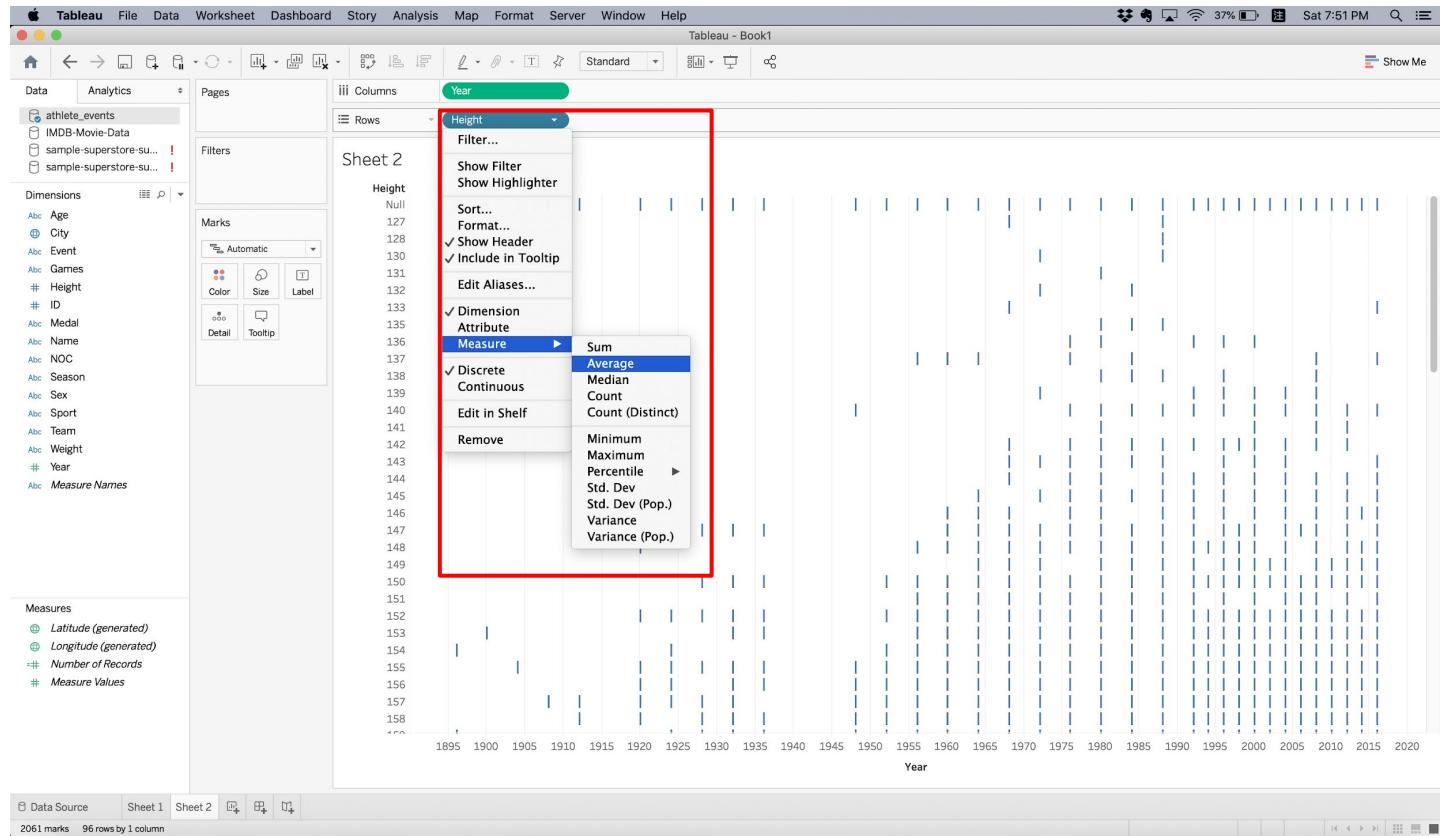
The screenshot shows the Tableau desktop interface. On the left, the 'Connections' pane displays a single connection named 'athlete\_events' (Text file). The main workspace shows a hierarchical tree under 'athletes\_events' with 'athlete\_events.csv' selected. The top right corner shows 'Live' selected for the connection. The bottom half of the screen shows a preview of the 'athlete\_events.csv' data. A red arrow points to the 'Sheet 1' tab at the bottom left of the preview area.

#	Abc athlete_events.csv ID	Abc athlete_events... Name	Abc athlete_events... Sex	Abc athlete_events... Age	Abc athlete_events.csv Height	Abc athlete_events.csv Weight	Abc athlete_events.csv Team	Abc athlete_events... NOC	Abc athlete_events.csv Games	#	Abc athlete_events.csv Year	Abc athlete_events.csv Season	Abc athlete_events.csv City	Abc athlete_events.csv Sport	Abc athlete_events.csv Event	Abc ath...
33539	Bndicte Evrard	F	17	176	38	Belgium	BEL	1992 Summer	1992	Summer	Barcelona	Gymnastics	Gymnastics Women'	N.		
33539	Bndicte Evrard	F	17	176	38	Belgium	BEL	1992 Summer	1992	Summer	Barcelona	Gymnastics	Gymnastics Women'	N.		
33539	Bndicte Evrard	F	17	176	38	Belgium	BEL	1992 Summer	1992	Summer	Barcelona	Gymnastics	Gymnastics Women'	N.		
33539	Bndicte Evrard	F	17	176	38	Belgium	BEL	1992 Summer	1992	Summer	Barcelona	Gymnastics	Gymnastics Women'	N.		
33540	J. Henri Evrot	M	46	NA	NA	France-1	FRA	1948 Winter	1948	Winter	Sankt Moritz	Bobsleigh	Bobsleigh Men's Two	N.		
33540	J. Henri Evrot	M	46	NA	NA	France-2	FRA	1948 Winter	1948	Winter	Sankt Moritz	Bobsleigh	Bobsleigh Men's Four	N.		
33541	Jens Ewald	M	21	185	77	Germany	GER	2004 Summer	2004	Summer	Athina	Canoeing	Canoeing Men's Kaya...	N.		
33542	Jacques Ewalds	M	24	180	76	Netherlands	NED	1964 Summer	1964	Summer	Tokyo	Volleyball	Volleyball Men's Voll...	N.		
33543	Essambo Ewane	M	27	NA	95	Cameroon	CMR	1980 Summer	1980	Summer	Moskva	Judo	Judo Men's Half-Hea...	N.		
33543	Essambo Ewane	M	31	NA	95	Cameroon	CMR	1984 Summer	1984	Summer	Los Angeles	Judo	Judo Men's Half-Hea...	N.		
33544	Maryse wanji-pe (-Ma...	F	19	176	62	France	FRA	1984 Summer	1984	Summer	Los Angeles	Athletics	Athletics Women's H...	N.		
33544	Maryse wanji-pe (-Ma...	F	24	176	62	France	FRA	1988 Summer	1988	Summer	Seoul	Athletics	Athletics Women's H...	N.		
33545	Monique wanji-pe-Tou...	F	21	173	65	France	FRA	1988 Summer	1988	Summer	Seoul	Athletics	Athletics Women's 1...	N.		
33545	Monique wanji-pe-Tou...	F	25	173	65	France	FRA	1992 Summer	1992	Summer	Barcelona	Athletics	Athletics Women's 1...	N.		

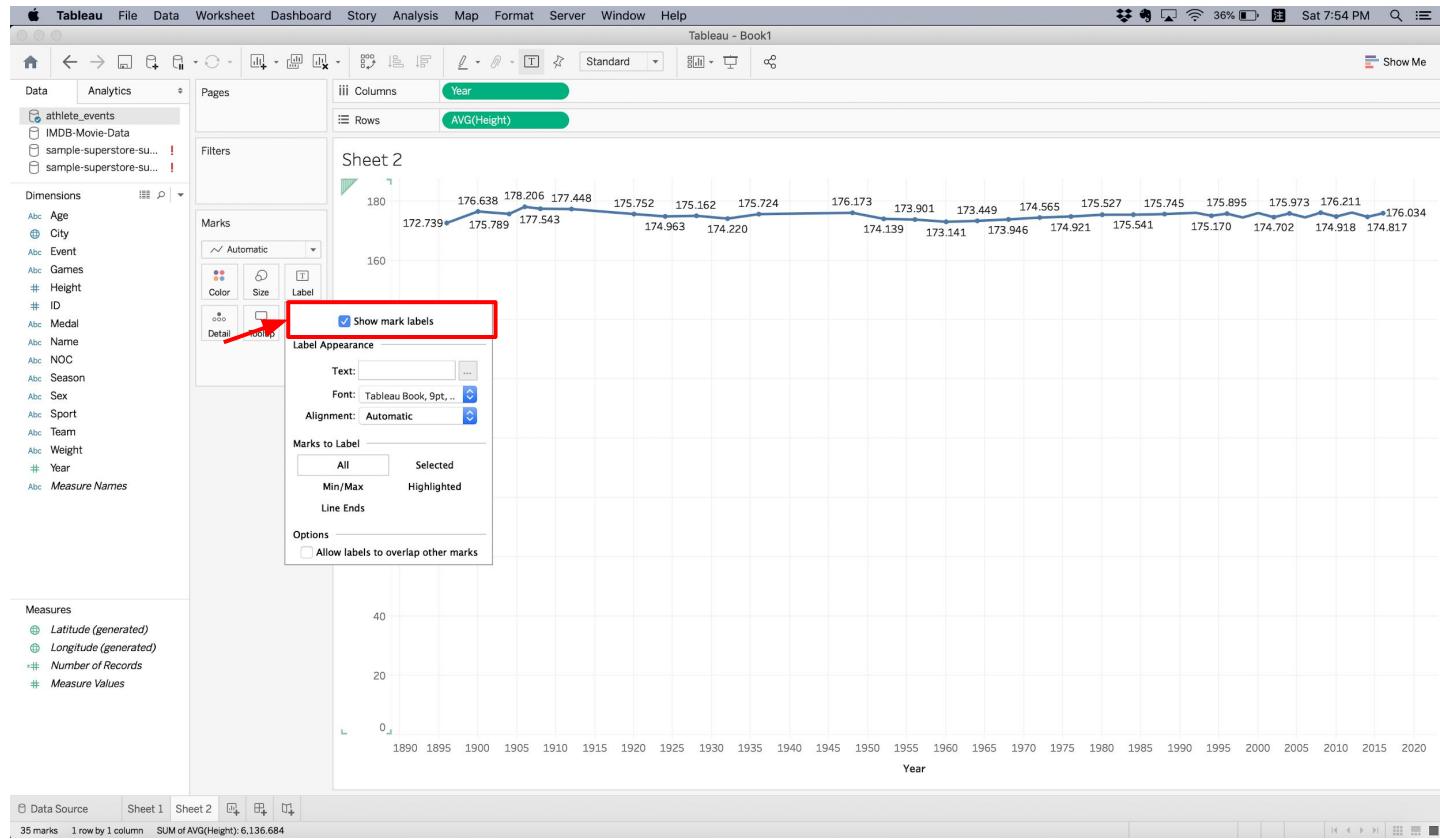
# Viz - Avg height of athletes by year

The screenshot shows the Tableau desktop application interface. The main area is titled "Sheet 2". On the left, the Data pane lists several data sources: "athlete\_events", "IMDB-Movie-Data", "sample-superstore-su...", and "sample-superstore-su...". The Dimensions pane lists various dimensions, including "Age", "City", "Event", "Games", "Height" (which is selected and highlighted in blue), "ID", "Medal", "Name", "NOC", "Season", "Sex", "Sport", "Team", "Weight", and "Year". The Measures pane lists generated measures: "Latitude (generated)", "Longitude (generated)", "Number of Records", and "Measure Values". A context menu is open over the "Height" dimension in the Dimensions pane. The menu options are: "Add to Sheet", "Duplicate", "Rename", "Hide", "Aliases...", "Create", "Transform", "Convert to Measure", "Change Data Type" (with "Number (decimal)" selected), "Geographic Role", "Default Properties", "Group by", "Folders", "Hierarchy", "Replace References...", and "Describe...". The "Number (decimal)" option is highlighted in blue. The status bar at the bottom shows "Data Source" and "Sheet 2". The top right corner displays the system tray with icons for battery, signal, and time ("Sat 7:49 PM").

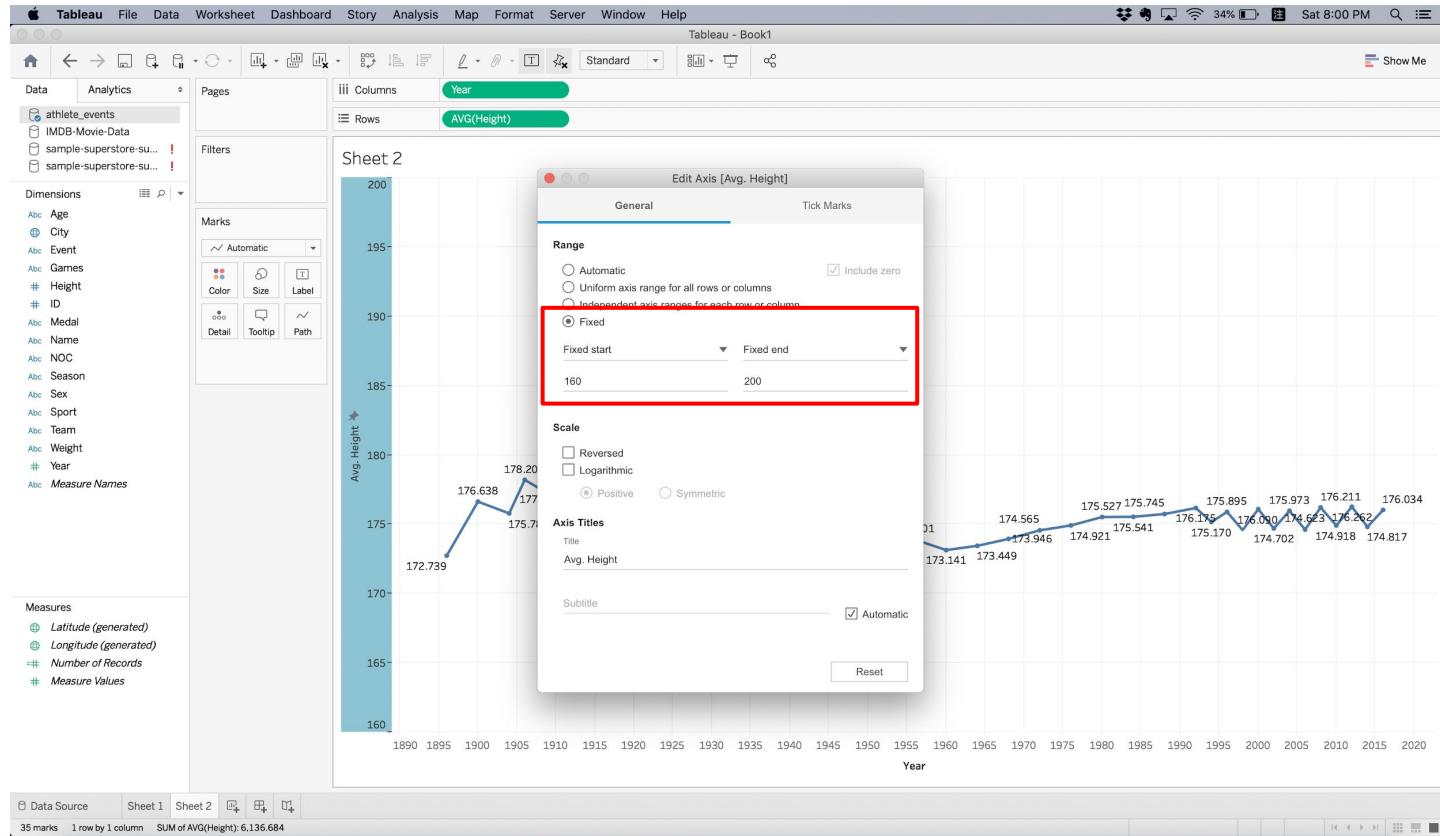
# Viz - Avg height of athletes by year



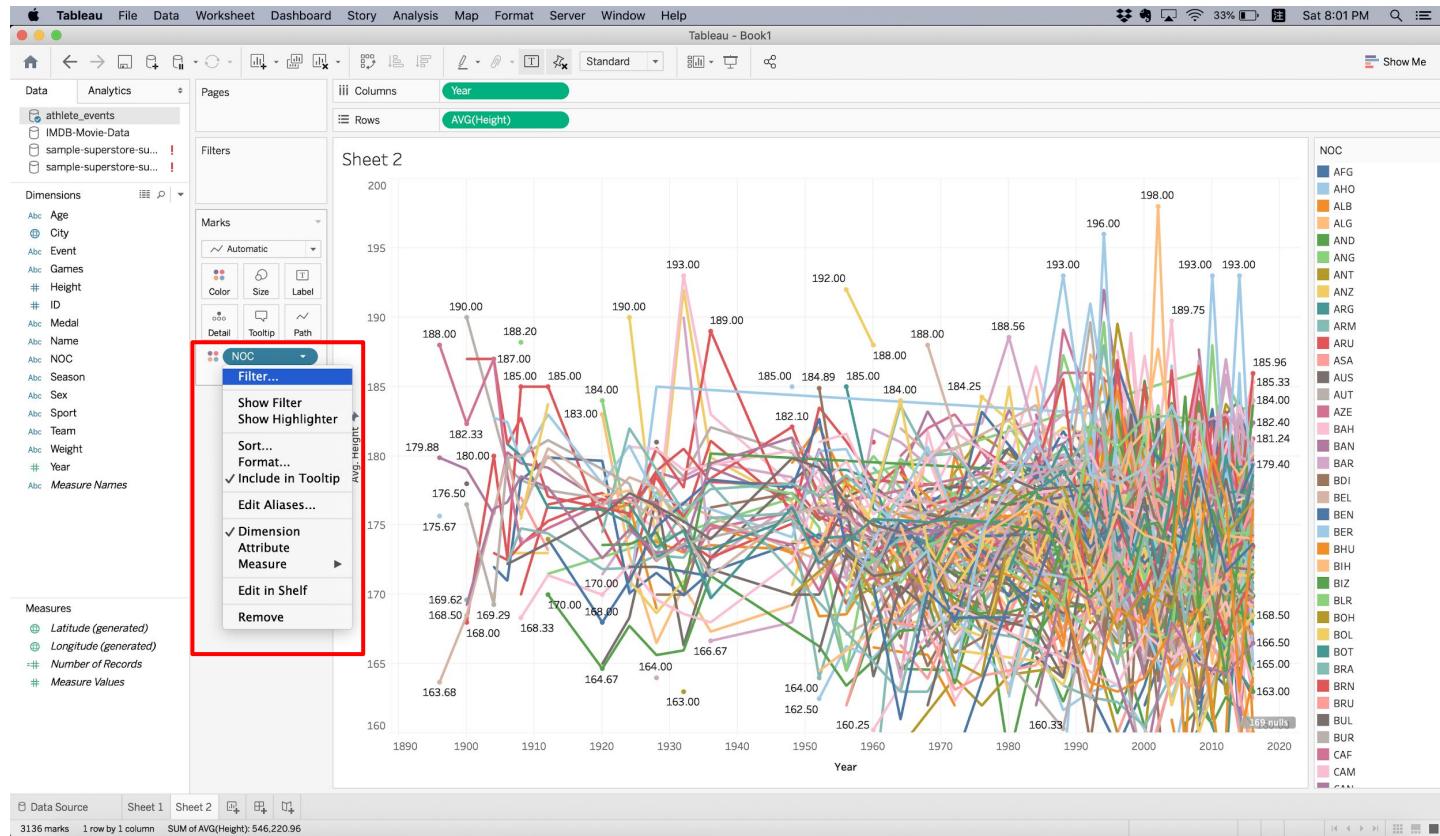
# Viz - Avg height of athletes by year



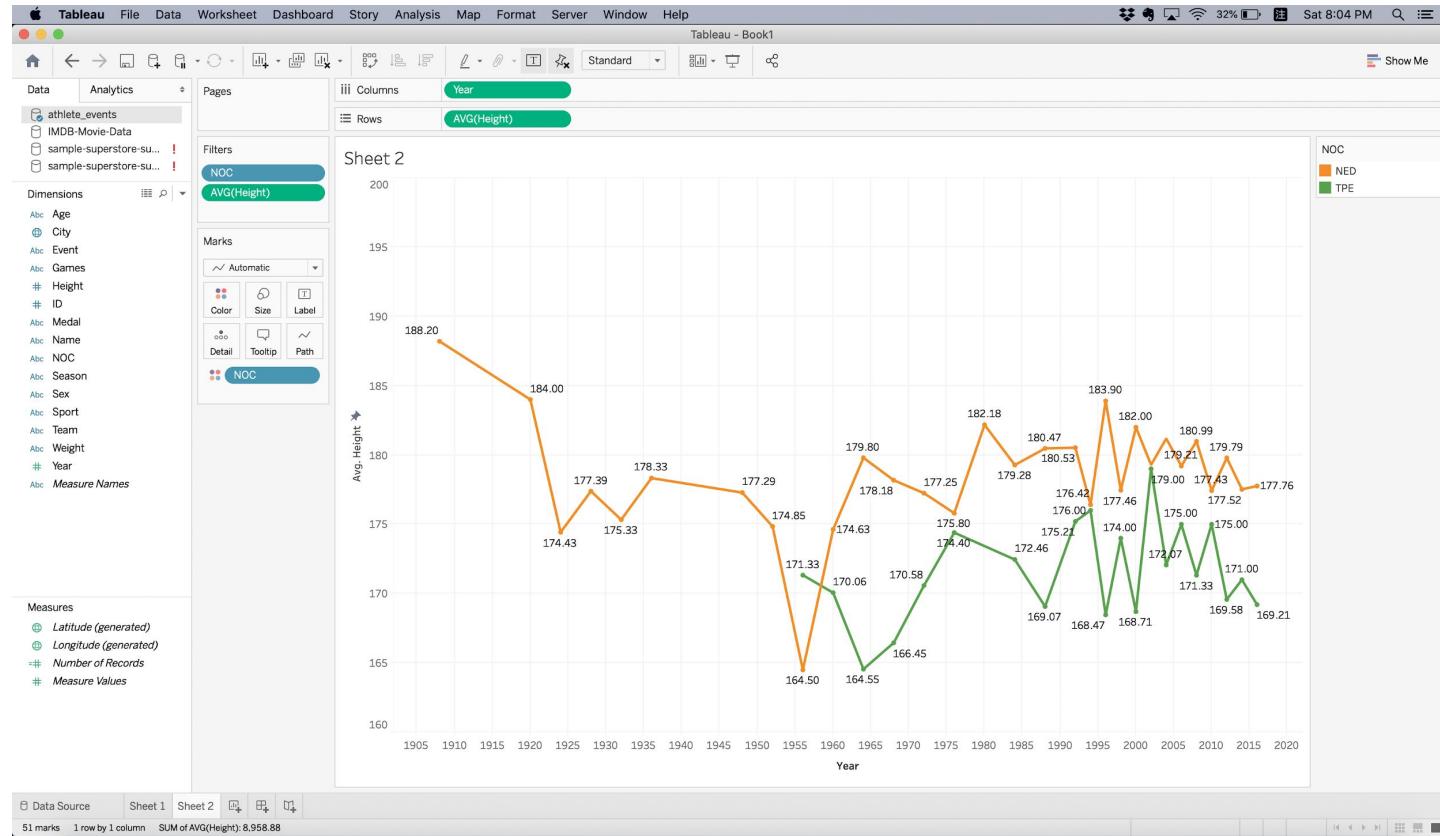
# Viz - change y-axis range



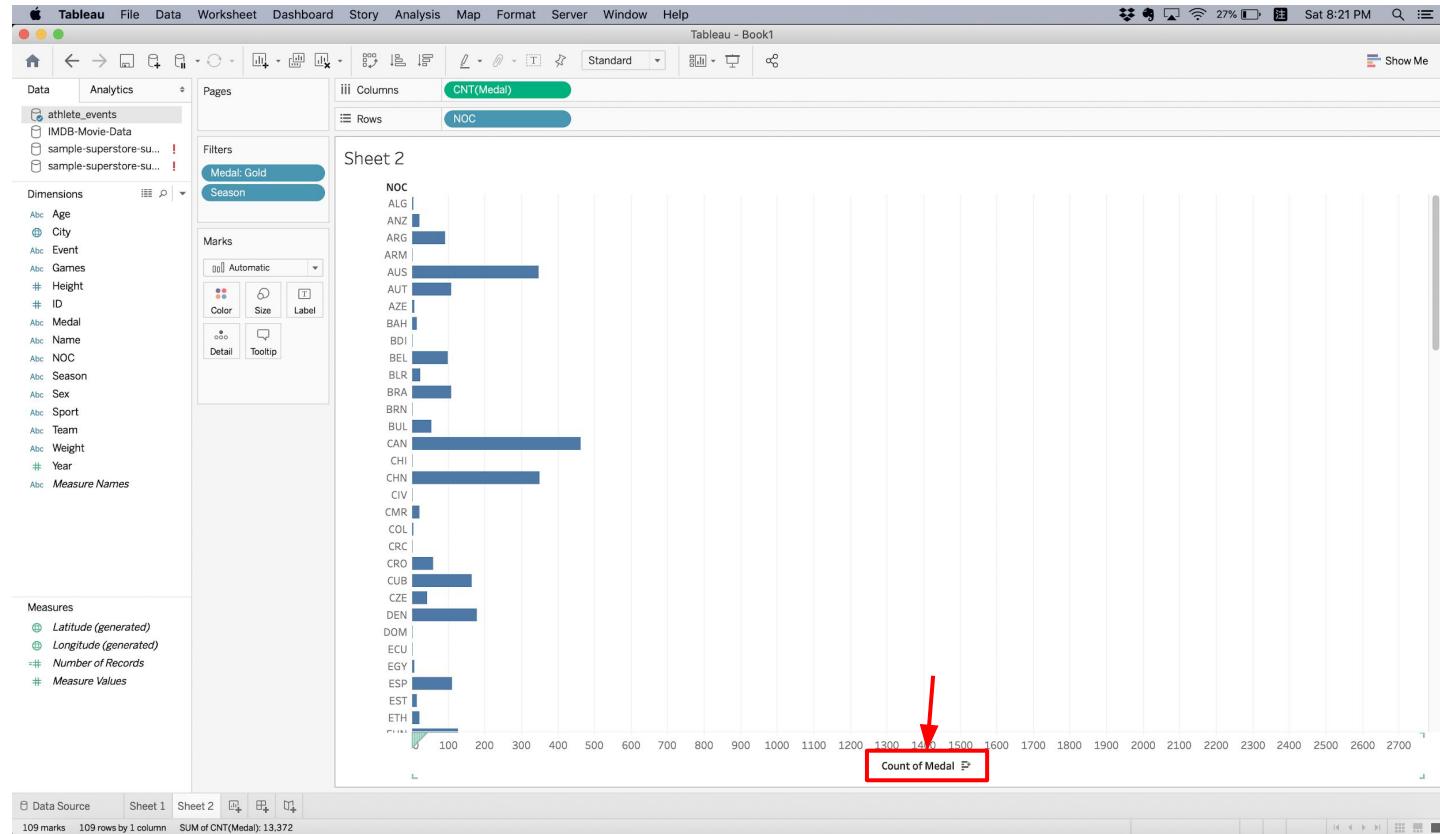
# Viz - filter by countries



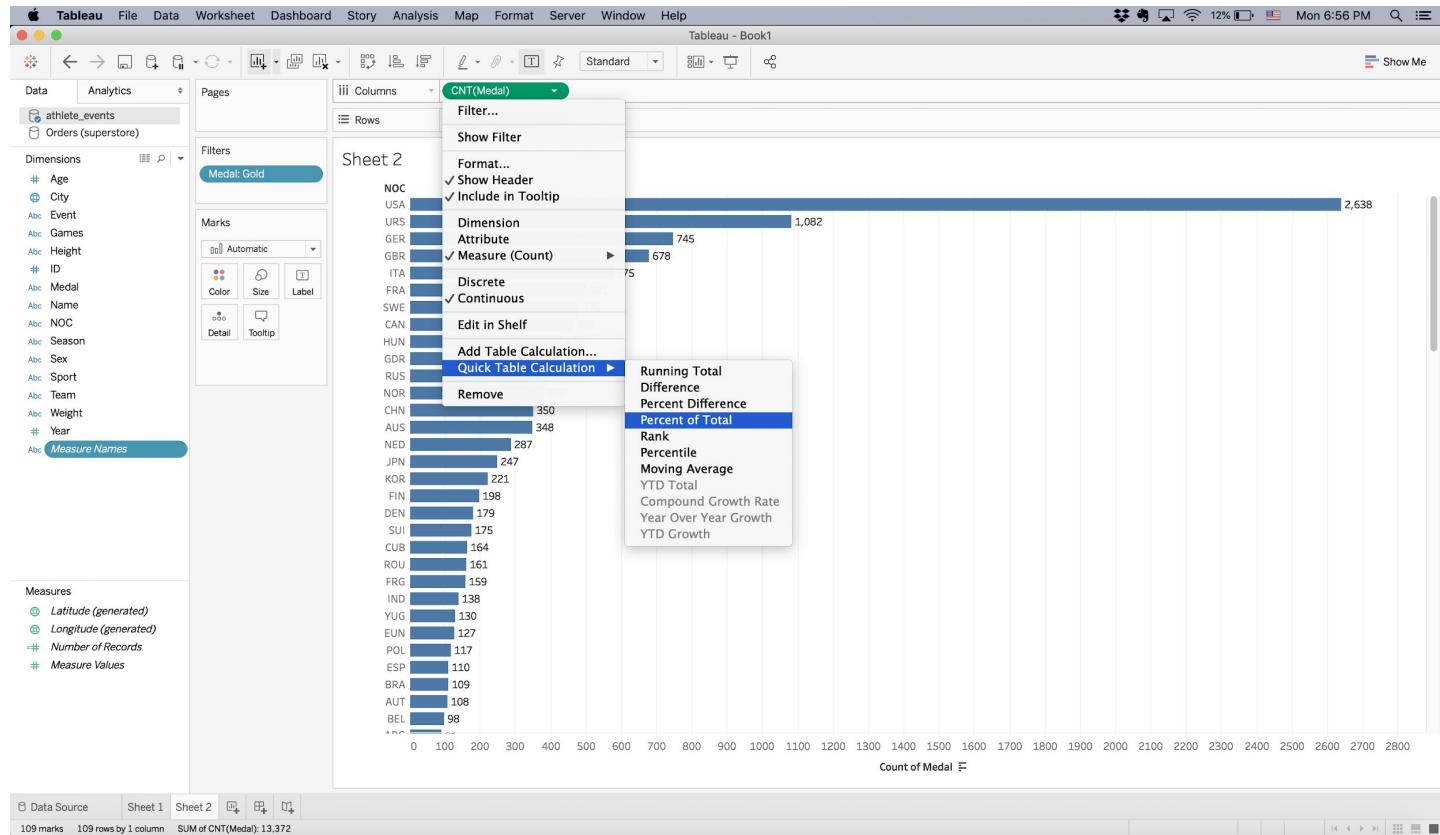
# Viz - Avg height of Denmark and Taiwan athletes



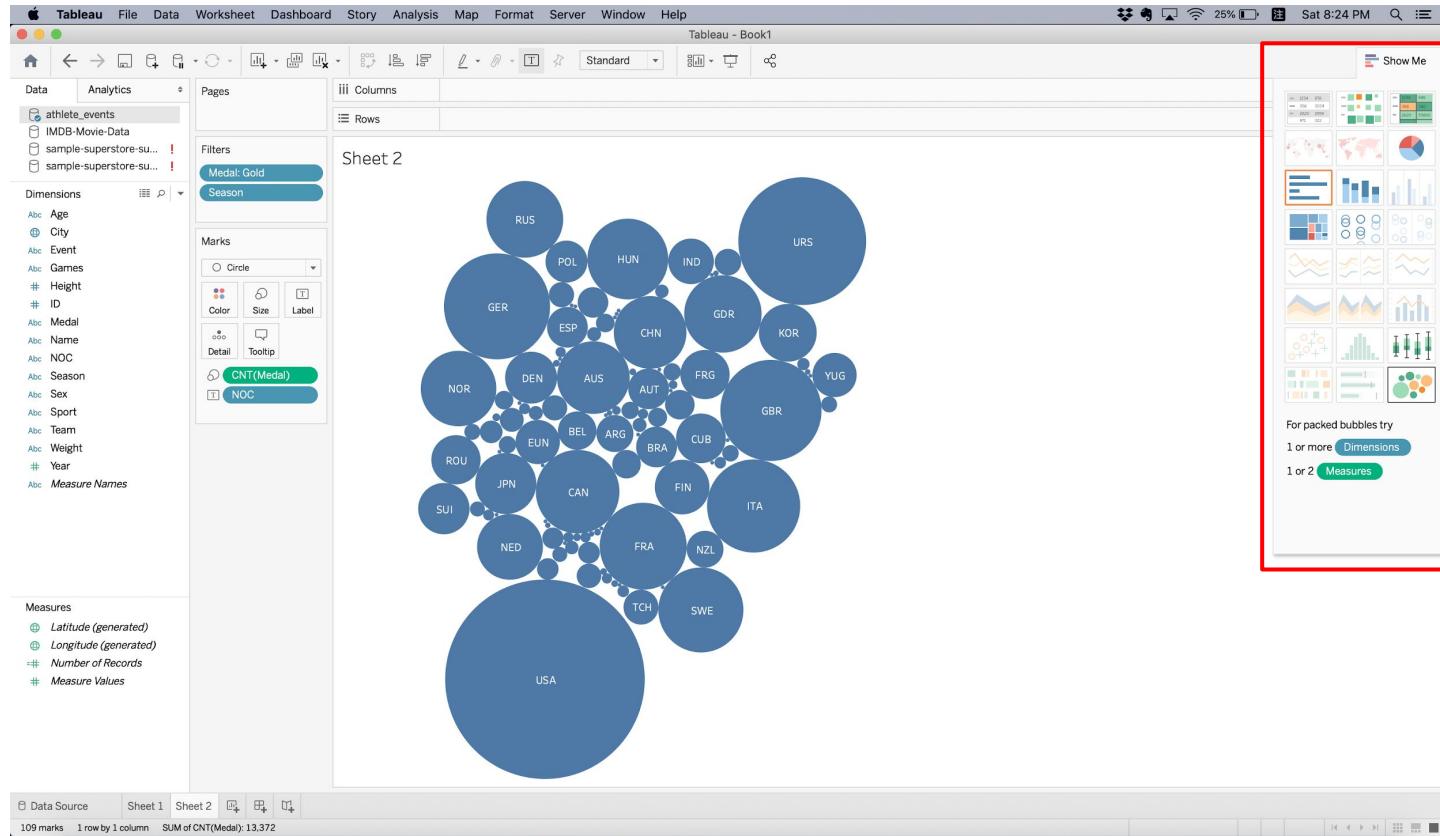
# Viz - # of gold medals by countries, sorted by counts



# Viz - % of gold medals

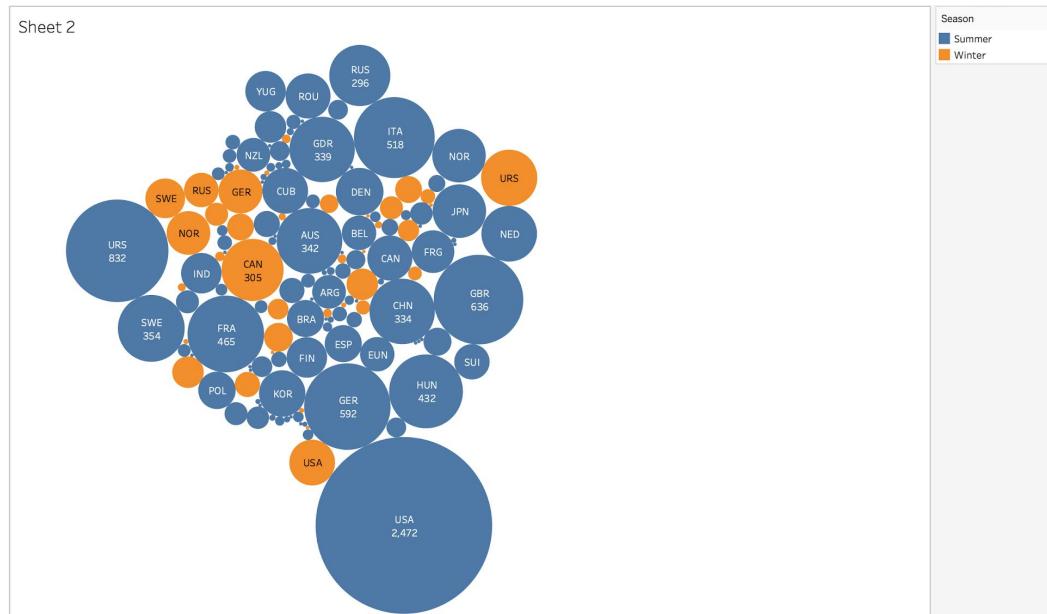


# Viz - # of gold medals by countries (bubbles)



# Viz - Practice

- Separate winter Olympics from summer Olympics
- Add number of gold medals as markers in addition to country names



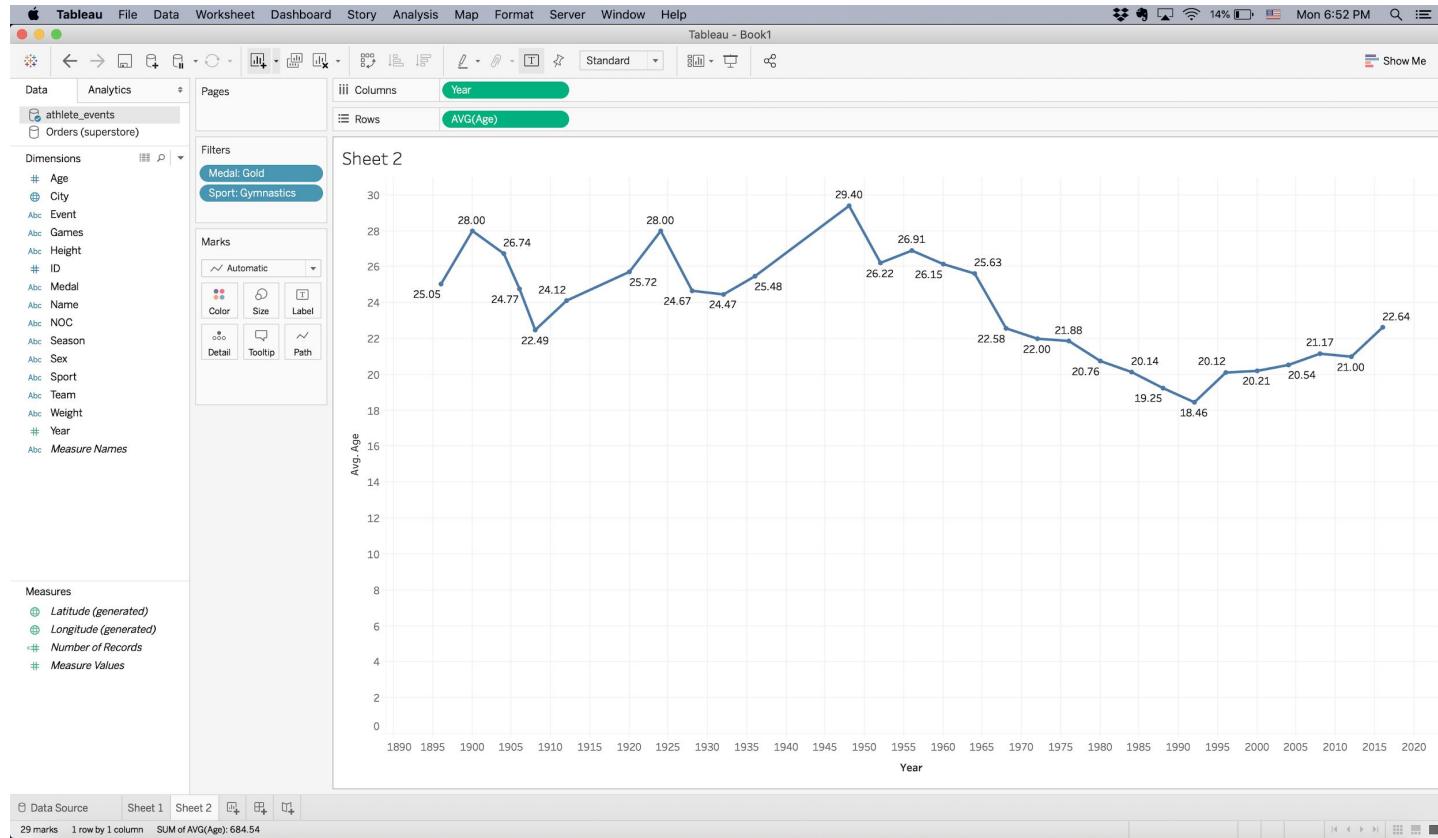
# Viz - highlight only countries with most gold medals

The screenshot shows a Tableau visualization titled "Sheet 2". The chart displays bubbles representing countries, where the size of each bubble corresponds to the number of gold medals won. The bubbles are colored according to the season: Summer (blue) and Winter (orange). A large blue bubble for the USA is labeled "USA 2,472".

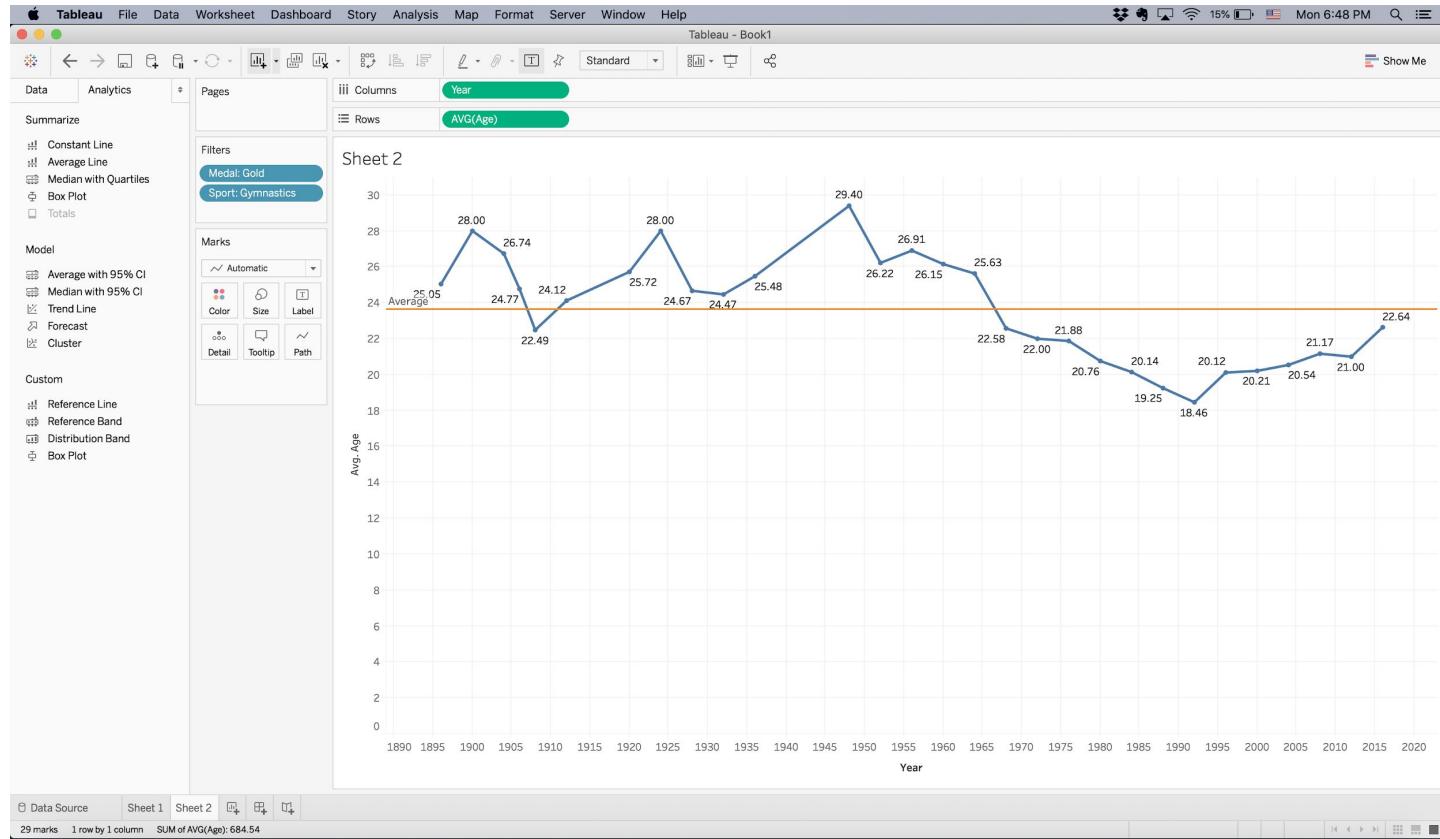
The interface includes:

- Dimensions:** Age, City, Event, Games, Height, ID, Medal, Name, NOC, Season, Sex, Sport, Team, Weight, Year.
- Measures:** Latitude (generated), Longitude (generated), Number of Records, Measure Values.
- Marks:** Circle, Color, Size, Label (highlighted with a red box and labeled 1).
- Filters:** Medal: Gold (highlighted with a red box).
- Sheet 2:** The main visualization area.
- Legend:** Season (Summer, Winter).
- Contextual menu:** A context menu is open over the USA bubble, labeled 2. It includes options like Select All, View Data..., Copy, Format..., and Mark Label. Under Mark Label, "Always Show" is checked.
- Bottom navigation:** Data Source, Sheet 1, Sheet 2, etc.

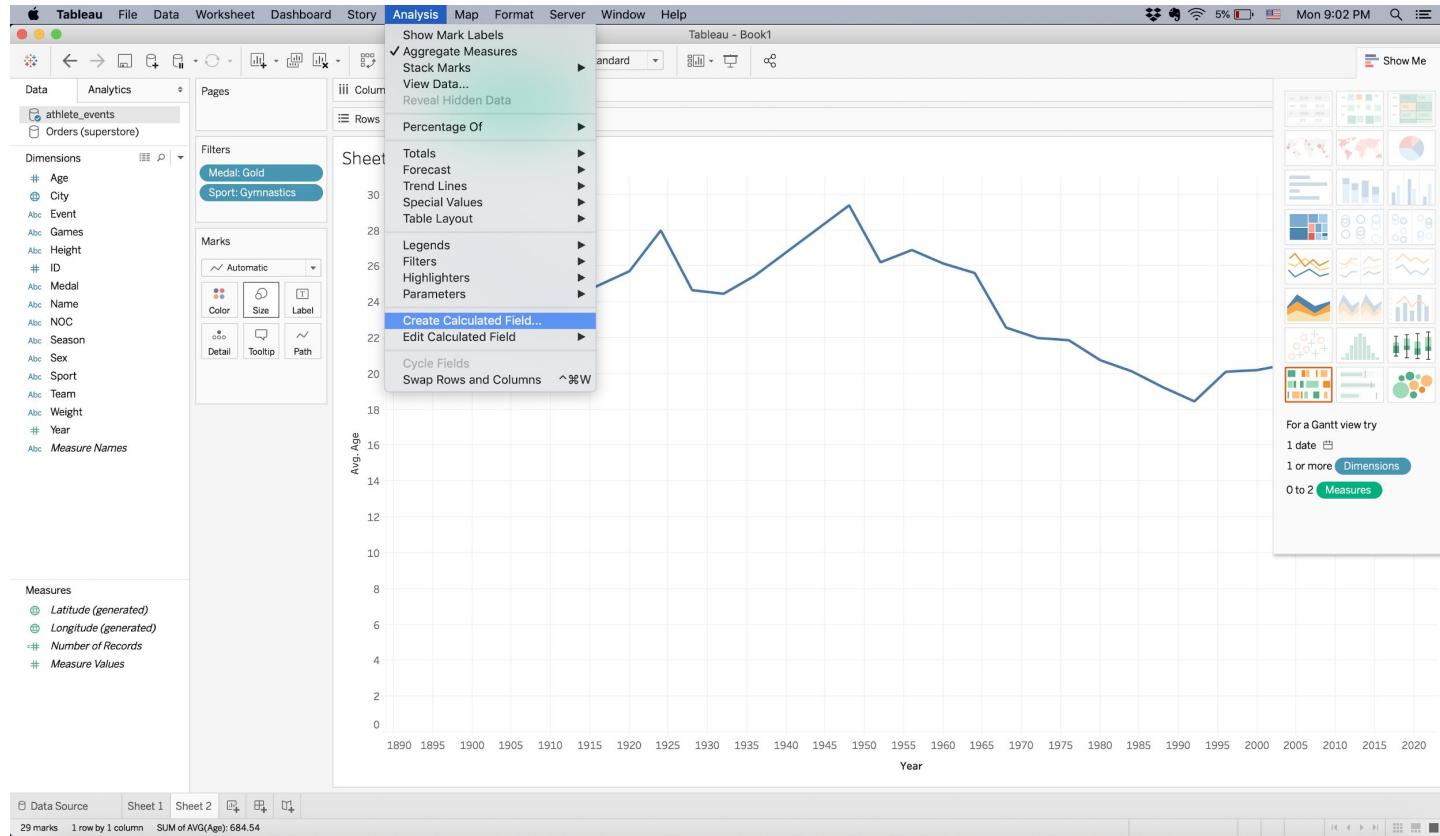
# Viz - Practice II: Avg age of gymnastics gold medalists



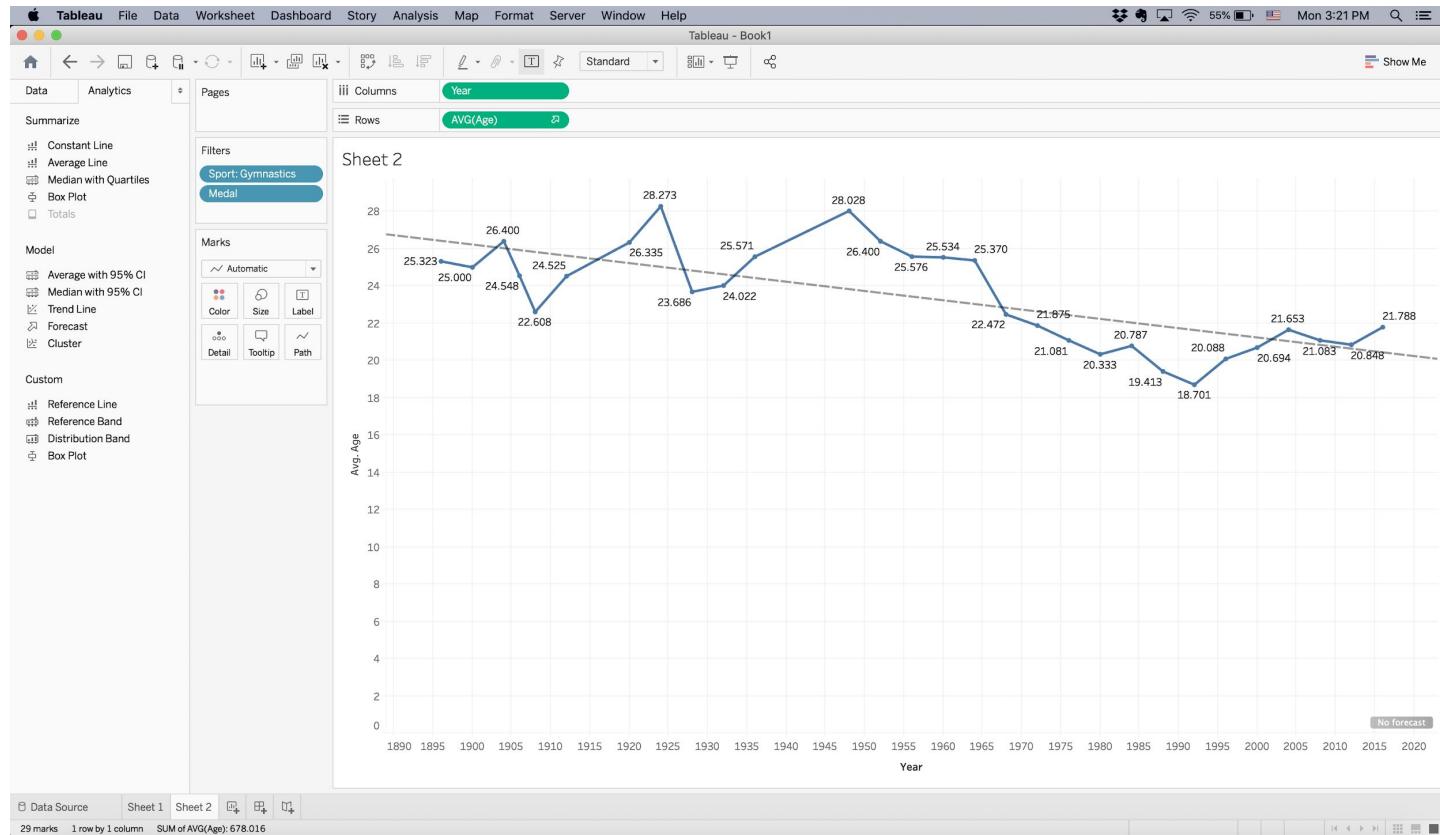
# Viz - Add reference line



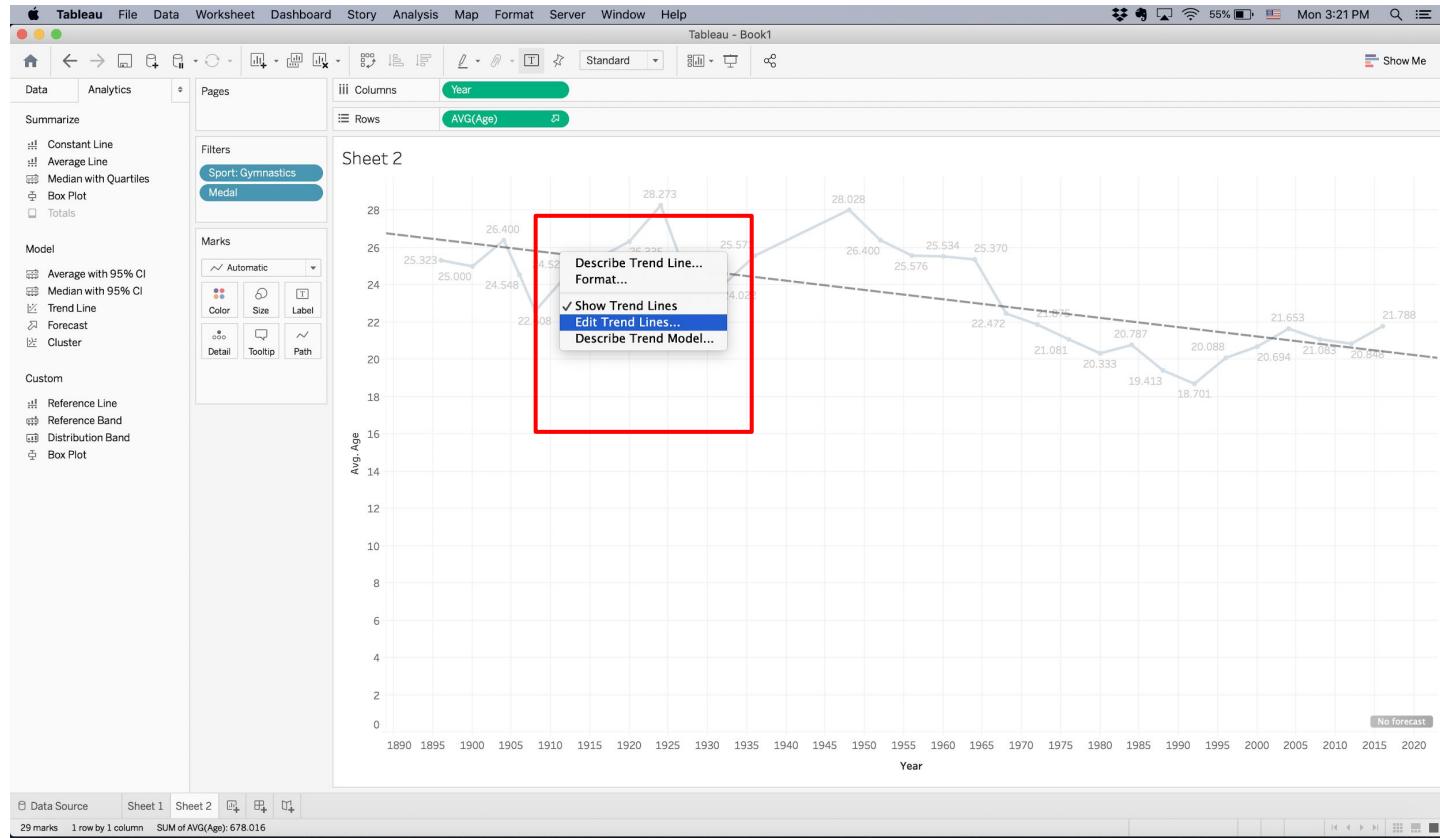
# Viz - Calculated fields



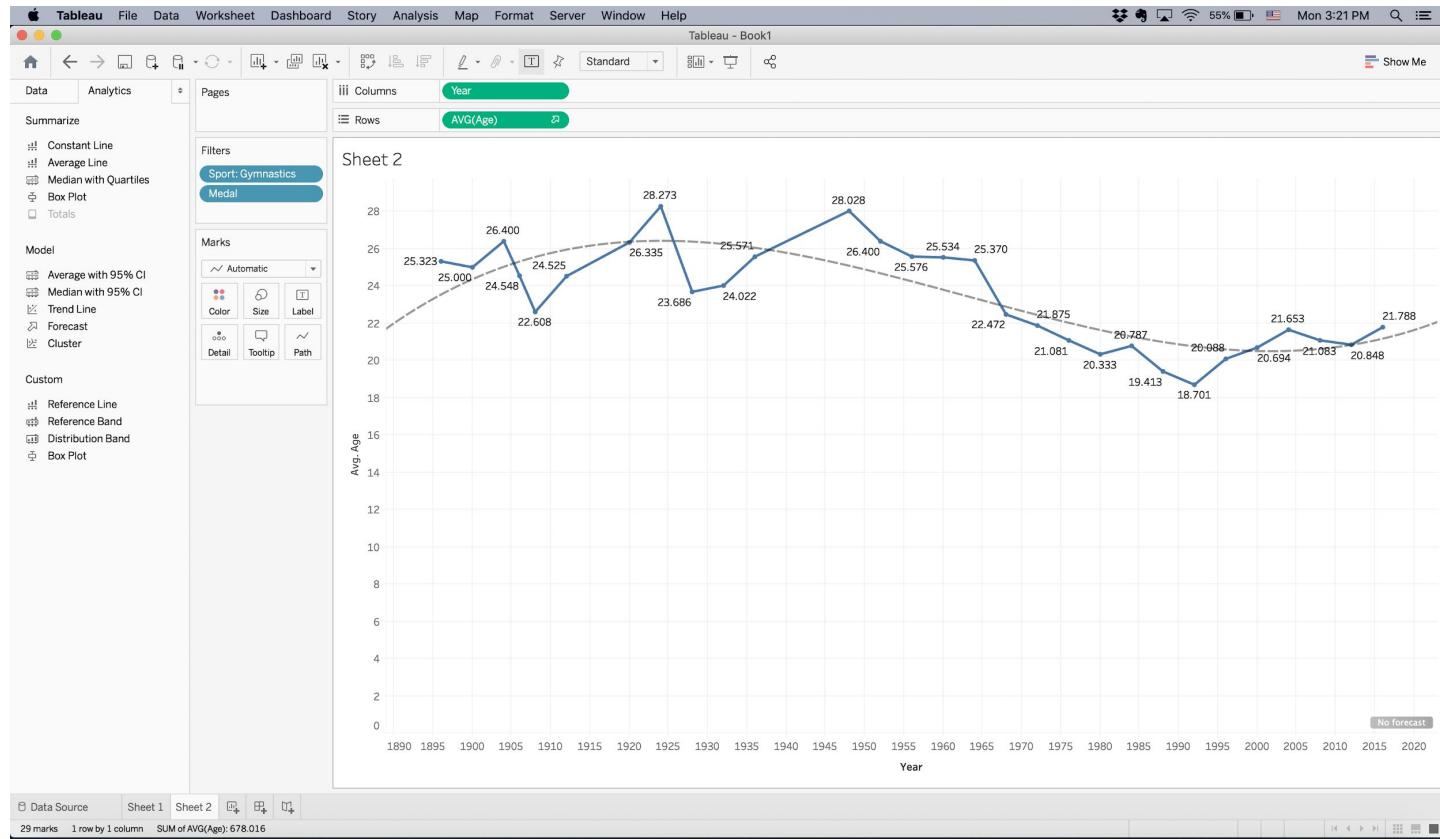
# Analytics - Regression lines



# Analytics - Editing Regression lines



# Analytics - Editing Regression lines



# Other resources for Tableau

- <https://www.tableau.com/learn>
- <https://www.tableau.com/solutions>