

PV251 / Seminar #6

Katarína Furmanová

Tableau

HCI^{LAB}

visitlab



Pi Transition Paths by Martin Krzywinski

Today

- Project progress check
- Tableau tutorial
 - If working on your own laptop, download Tableau Desktop Public (Main menu ->Create->Download Tableau Desktop Public Edition)
 - Download data **athlete_events.csv** from IS
- Individual work





Semestral Project Reminder

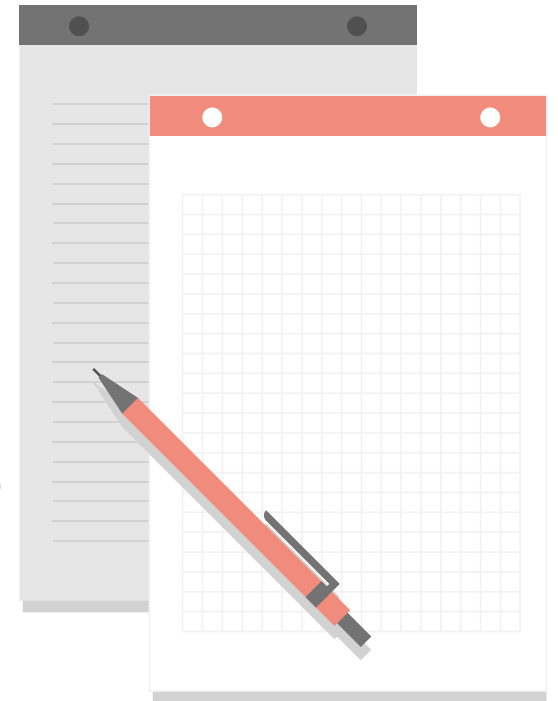
Semestral Project Requirements

- Interactive data visualization project
- 2-3 linked views, interactive features (e.g., filtering) + short report
- **Technology:** up to you
 - Preferably web-based (so you can share it), e.g., JavaScript + D3, Plotly/Dash, Tableau, ...
- **What to submit:**
 - sources
 - instructions for running (e.g., dependencies) or link to a deployed version
 - project report (see next slide)
- **Deadline:** ideally before attending exam, at the latest by **January 31st 2025**
- **Bonus points:**
 - Up to 3 points for project realisation
 - Up to 2 points for voluntary presentation

Semestral Project – Report

- Motivation – why did you choose the topic, what did you want to see/show
- Data sources + data preprocessing
- Explanation of the design choices
- Interesting observations in your visualization
- Screenshots
- Used technologies, requirements
 - Specify the environment in which you tested it (browser, resolution, ...)
- Lessons learned – what did you take away from the project

- Length: 1-2 pages (A4)



Pre-submission Checklist



- Before submitting your project, check:
 - All axes are labeled, and contain necessary info (e.g., units)
 - All charts have appropriate legends (e.g., explaining colors)
 - Your color choices make sense
 - Do not use same color for encoding different things
 - Do not over-use colors (e.g., 12+ colors)
 - All necessary files are included (e.g., there are no permanent links to local data files that will not work somewhere else)
- You will need to fix that anyway, so save yourselves (and me) the trouble of re-submission 😊

Bonus Points Presentation

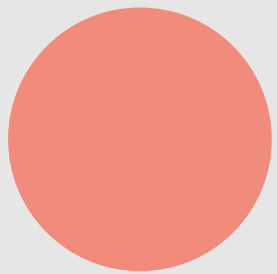
- 7-10 minutes presentation
- Content similar to project report
 - Motivation
 - Challenges
 - Design choices
 - Presentation of the final design
 - Possible extensions
 - Live demo at your own risk
 - Video with screencast might be safer, we all know how live demos (don't) work 😊
- Dates will be posted based on demand

Resources and Links

- Color selection: <https://colorbrewer2.org/>
- Maps:
 - [leaflet.js](#) – complete world maps (OpenStreetMaps), easy annotations, heatmaps, layers
 - Mapbox maps in Plotly: <https://plotly.com/python/maps/>
 - SVG: <https://simplemaps.com/resources/svg-maps>, <https://www.amcharts.com/svg-maps/>
 - GeoJSON: <https://observablehq.com/@deaxmachina/collection-of-maps>, Google 😊
- Tabular data: [Line-up.js](#) library for tabular visualization
- Deploying your application:
 - GitHub/GitLab pages
 - <https://glitch.com/> (<https://www.youtube.com/watch?v=3hVf4Giy5nc>)
 - <https://render.com/> (<https://github.com/thusharabandara/dash-app-render-deployment>)

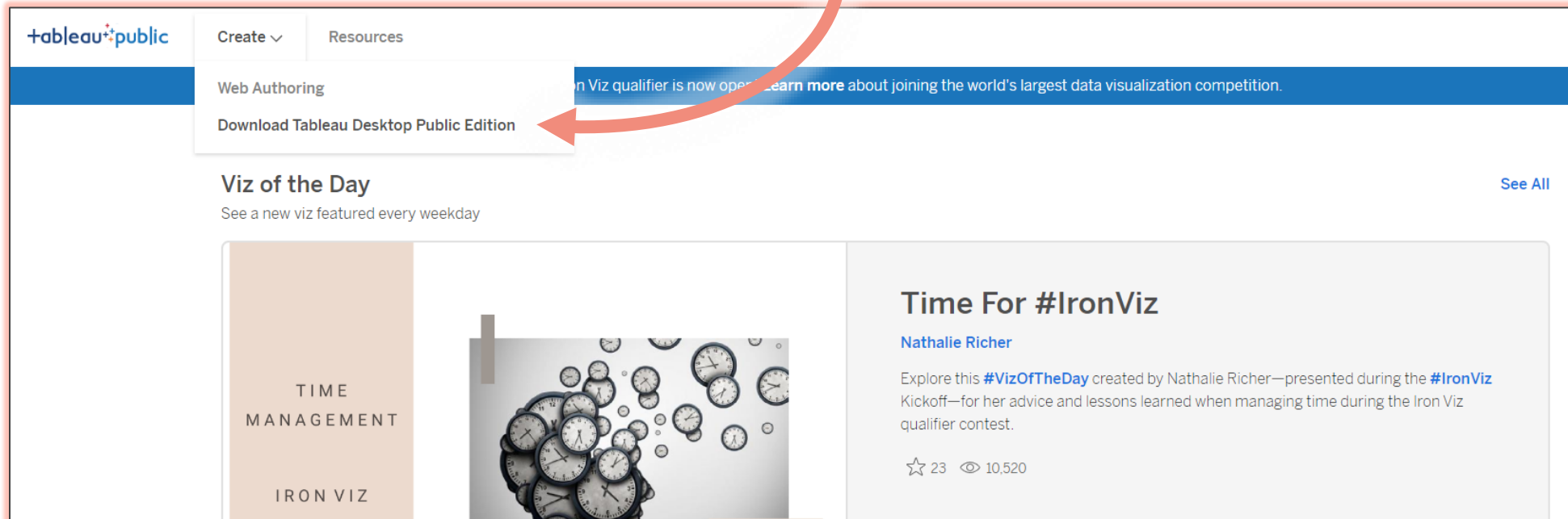


+ a b l e a u



Tableau

- **Tableau** is a powerful tool for building data visualizations
 - Supports wide spectrum of visualizations and interactions
 - Inspiration: [Tableau VIZ of the Day](#)
- If working on your own laptop, download [Tableau Desktop Public](#)
 - Main menu ->Create-> Download Tableau Desktop Public Edition

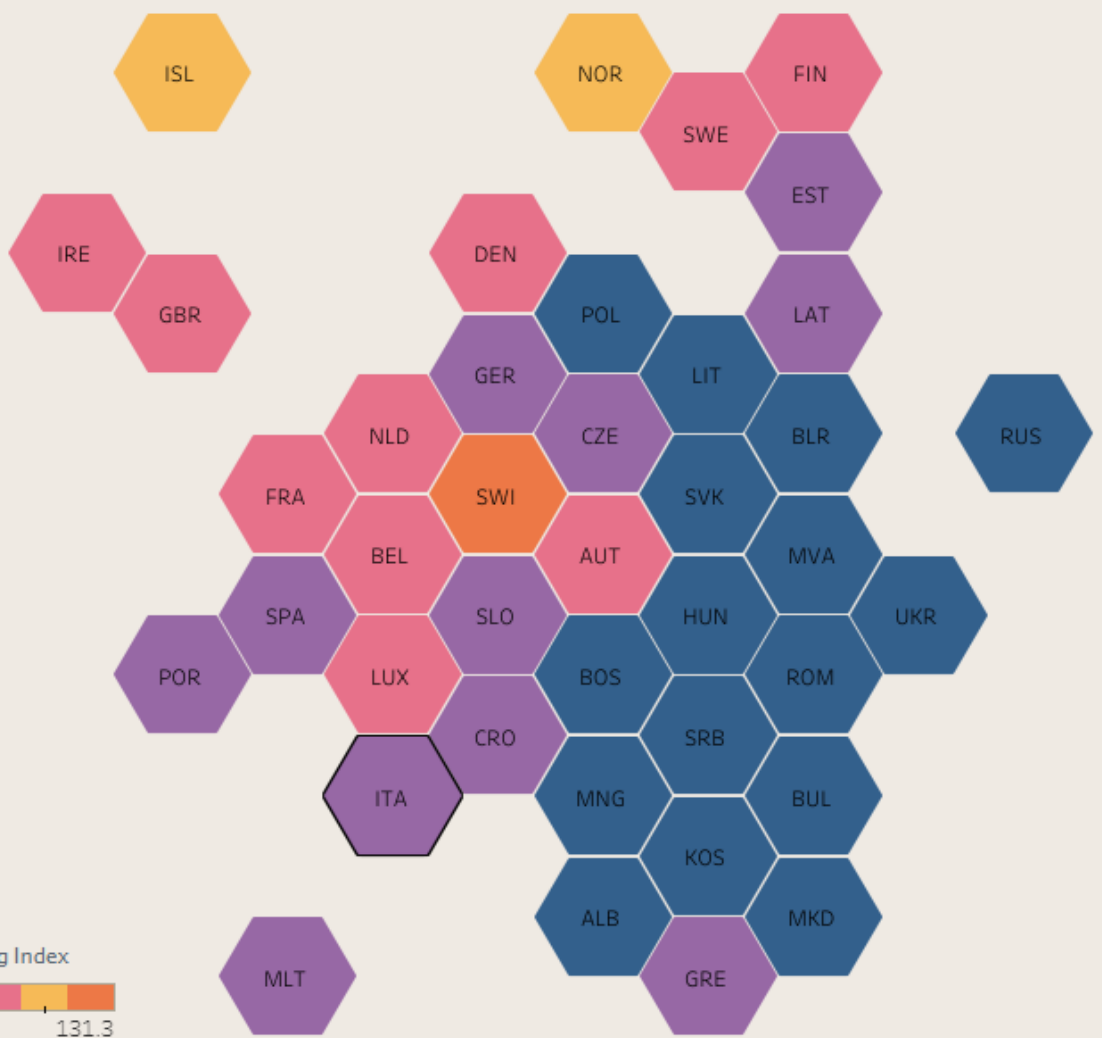


The screenshot shows the Tableau Public website interface. At the top left is the Tableau Public logo. To its right are two tabs: 'Create' and 'Resources'. The 'Create' tab is active, and its dropdown menu is open, showing three options: 'Web Authoring', 'Download Tableau Desktop Public Edition', and 'Download Tableau Desktop'. A red arrow points from the 'Create' tab to the 'Download Tableau Desktop Public Edition' option. Below the navigation bar is a blue banner with a notification about the Iron Viz qualifier. The main content area features a 'Viz of the Day' section with the text 'See a new viz featured every weekday' and a 'See All' link. Below this is a featured visualization titled 'Time For #IronViz' by Nathalie Richer. The visualization is a conceptual graphic of a human head filled with various clock faces. The text next to the visualization reads: 'Explore this #VizOfTheDay created by Nathalie Richer—presented during the #IronViz Kickoff—for her advice and lessons learned when managing time during the Iron Viz qualifier contest.' At the bottom of the visualization card, there are icons for a star (23) and an eye (10,520).

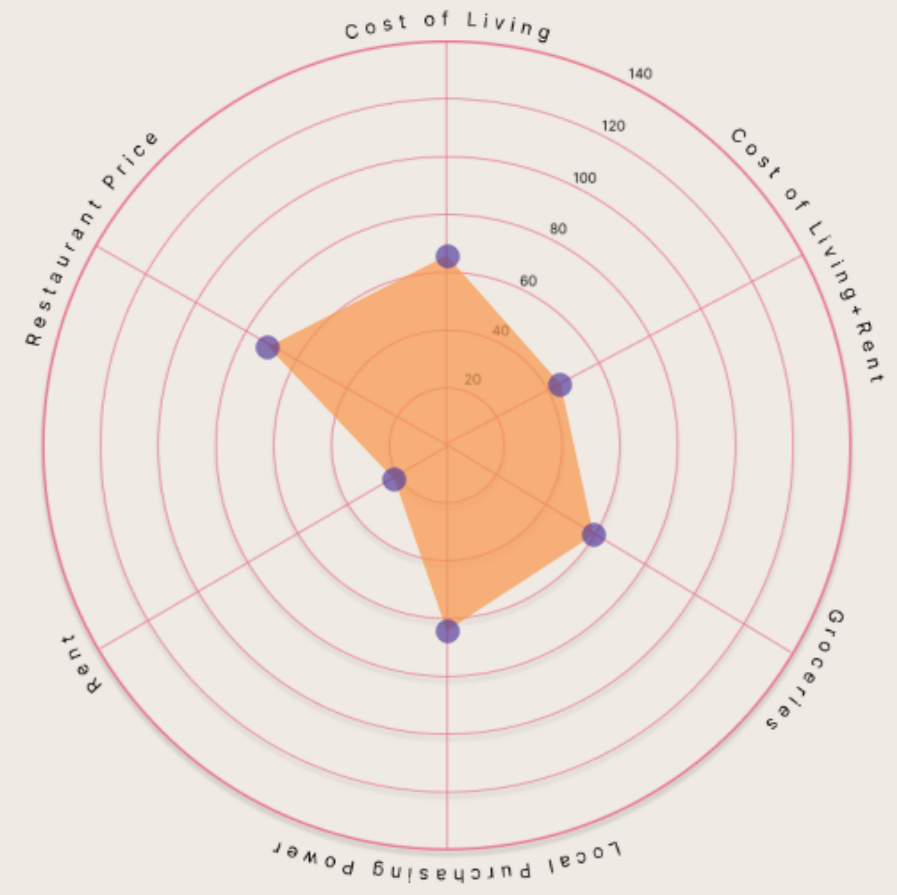
COST OF LIVING IN EUROPE

Cost of living index by country, in 2022

Indices are relative to New York City, value of 100. Click on a country to filter all the indices from the right.



All the indices for Italy

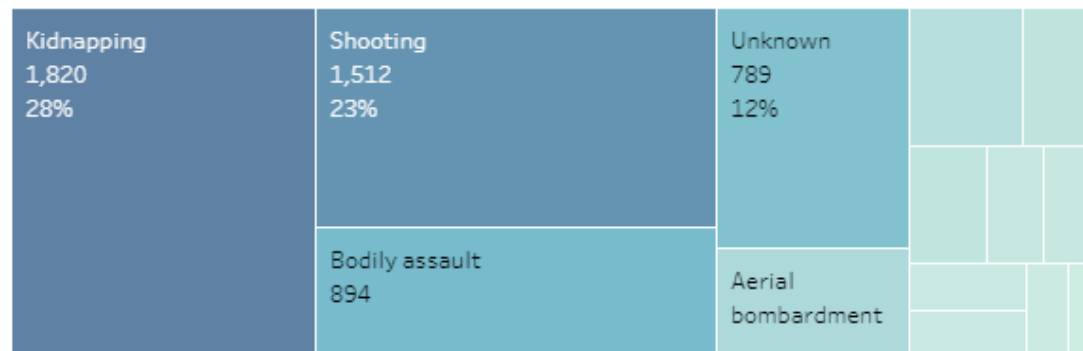


Aid Worker Security

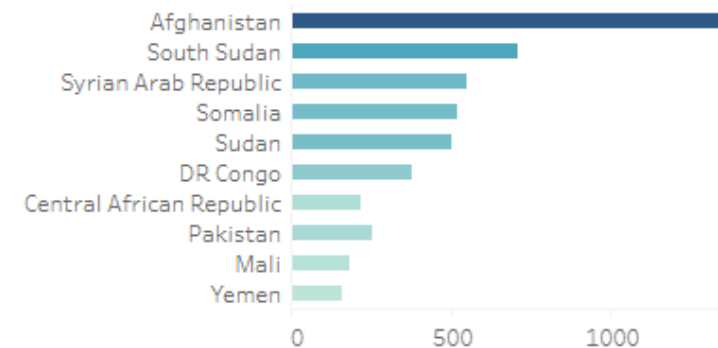
The Aid Worker Security Database (AWSD) records major incidents of violence against aid workers, with incident reports from 1997 through the present.

Initiated in 2005, to date the AWSD remains the sole comprehensive global source of this data, providing the evidence base for analysis of the changing security environment for civilian aid operations. Click on the diagram elements to filter the view, or focus your view by using the filter elements on the bottom left.

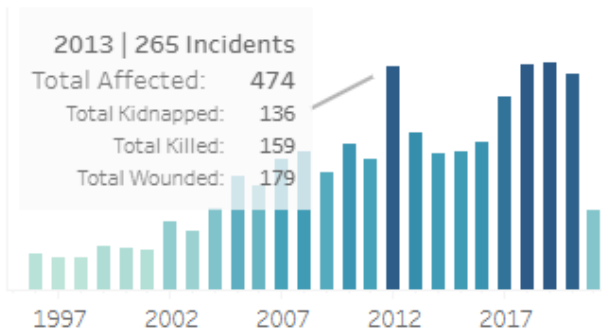
Means of Attack Distribution



10 most affected Countries



Incidents per Year

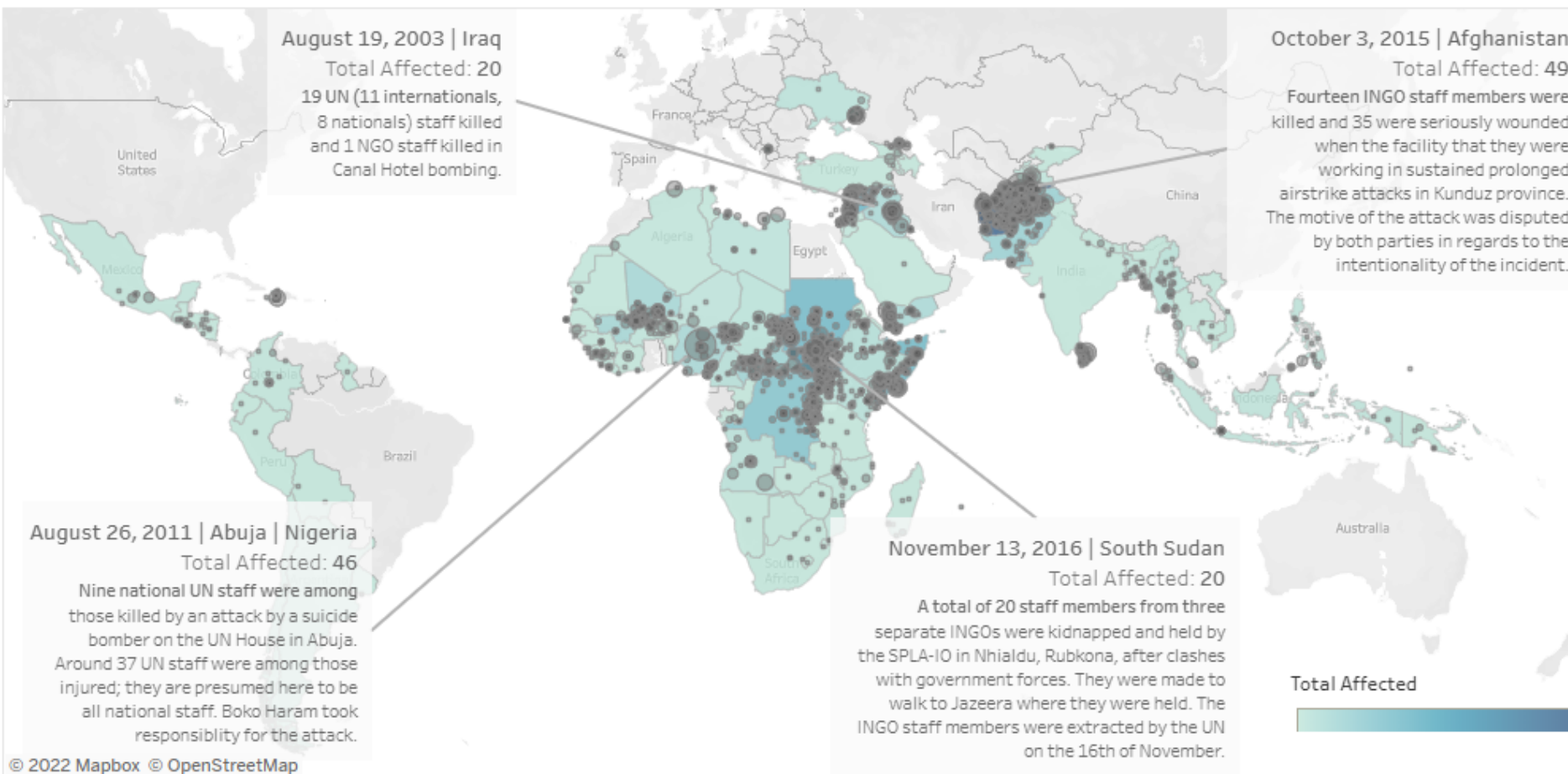


Filter by Country: (All)

Filter by Means Of Attack: (All)

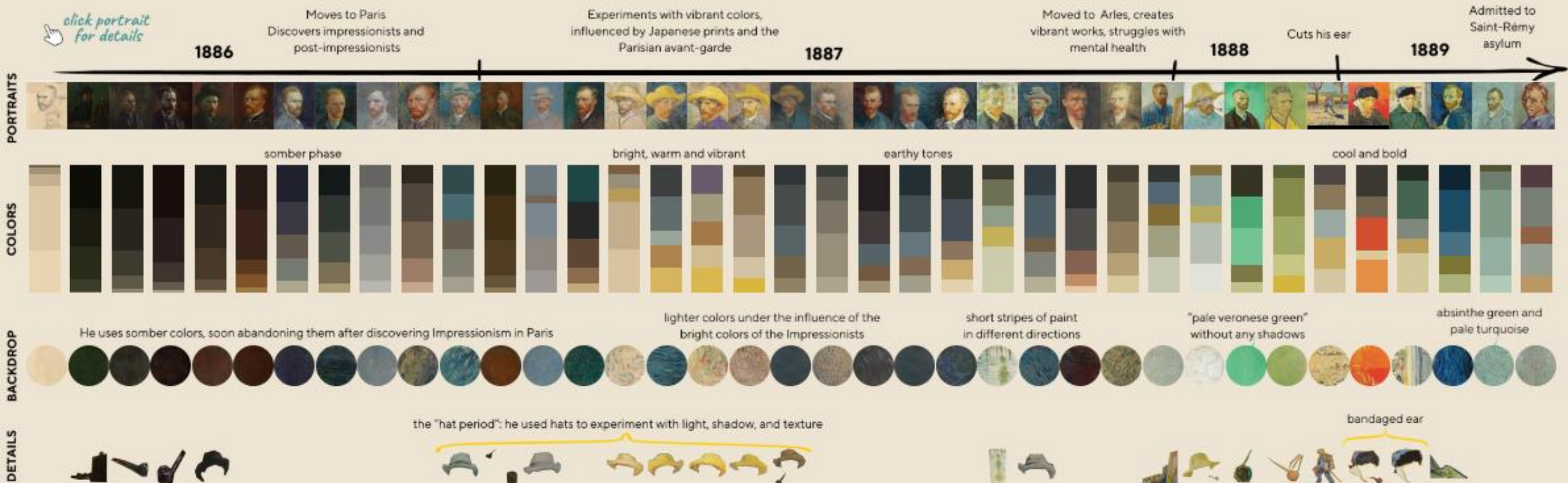
Filter by Attack Context: (All)

Filter by Y... 1997 ————— 2022

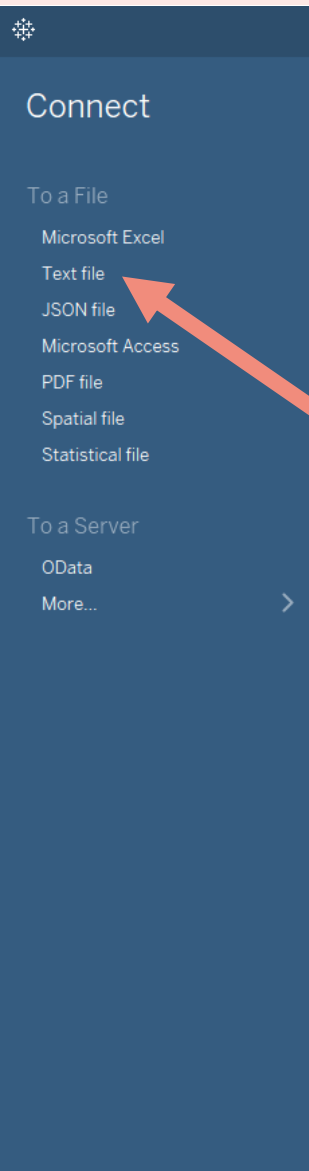


VAN GOGH IN 37 SELF-PORTRAITS

Vincent van Gogh, the legendary Dutch Post-Impressionist, used his self-portraits as windows into his deepest emotions, inviting us to see the world through his eyes. In this dashboard, explore Van Gogh's journey through 37 self-portraits, each piece unfolding a chapter of his life, emotions, and evolving artistry.



Getting started...



The screenshot shows the Tableau 'Connect' sidebar on the left. It is divided into two sections: 'To a File' and 'To a Server'. Under 'To a File', the options are Microsoft Excel, Text file, JSON file, Microsoft Access, PDF file, Spatial file, and Statistical file. An orange arrow points to the 'Text file' option. Under 'To a Server', the options are OData and More... with a right-pointing arrow.

Open

[Open from Tableau Public](#)

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**.

Discover

[How-to Videos](#)

Overview

Intro to the Interface

Chart Types

Dashboard

[More how-to videos...](#)

VIZ
OF THE
DAY

[See trending vizzes](#)
→

TIME
MANAGEMENT
IRON VIZ

Viz of the Day

[Explore viz of the day](#)

[Resources](#)

[Blog - Read latest post](#)

[Sample Data Sets](#)

[Current Status](#) 14

Connections Add

athlete_events
Text file

Files ⌵

Use Data Interpreter
Data Interpreter might be able to clean your Text file workbook.

athlete_events.csv

New Union

athlete_events

athlete_events.csv

- Open...
- Rename
- Remove
- Field names are in first row
- Generate field names automatically
- Text File Properties...**
- Convert to Union...

1. Click the little triangle next to your data file and select **Text File Properties...**

2. Select:
Comma as Field separator
" as Text qualifier
English (United States) as locale

athlete_events.csv

Field separator

Text qualifier

Character set

Locale

athlete_events.csv

Name
athlete_events.csv

Fields

Type	Field Name	Phys...	Rem...
	"ID","Name","Sex","Age","Hei...	athlet...	"ID", "...

athlete_events.csv
"ID","Name","Sex","Age","Height","Weight","Team","NOC","Games","Year","Season","City","Sport","Event","Medal"

3. For the data entries to appear, you might have to click **Update Now.**

1. Check the datatype of each column. Sometimes NA in column results in a String column even though column contains numerical data (e.g., age, weight, height). Click on the column type (Abc/#) and **change the columns to the appropriate data types** (in our case whole numbers).

athlete_events.csv 15 fields 271116 rows 100 rows

Name athlete_events.csv

Fields

Type	Field Name	Phys...	Rem...
#	ID	athlet...	ID
Abc	Name	athlet...	Name
Abc	Sex	athlet...	Sex
#	Age	athlet...	Age
Abc	Height	athlet...	Height
Abc	Weight	athlet...	Weight
Abc	Team	athlet...	Team

#	Abc	Abc	#	Abc	Abc	Abc	Abc	#	Abc
ID	Name	Sex	Age	Weight	Team	NOC	Games	Year	Season
33 582	George Louis ...	M					1904 Summer	1904	Summer
33 582	George Louis ...	M					1904 Summer	1904	Summer
33 582	George Louis ...	M					1904 Summer	1904	Summer
33 582	George Louis ...	M					1904 Summer	1904	Summer
33 582	George Louis ...	M					1904 Summer	1904	Summer
33 582	George Louis ...	M					1904 Summer	1904	Summer
33 582	George Louis ...	M					1904 Summer	1904	Summer
33 582	George Louis ...	M					1904 Summer	1904	Summer
33 582	George Louis ...	M					1904 Summer	1904	Summer
33 583	Natig Eyvazov	M					2000 Summer	2 000	Summer
33 584	Zohra Ez Zahr...	F					2016 Summer	2 016	Summer

Data Source Sheet 1

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

Dimensions and Measures

The screenshot shows the Tableau Data pane for the 'athlete_events' dataset. It is divided into two sections: 'Tables' and 'Measures'. The 'Tables' section lists various categorical fields: City, Event, Games, Medal, Name, NOC, Season, Sex, Sport, Team, and Measure Names. The 'Measures' section lists numerical and generated fields: Age, Height, ID, Weight, Year, athlete_events.csv (Count), Latitude (generated), Longitude (generated), and Measure Values. A red bracket on the right side of the pane groups the 'Tables' section as dimensions and the 'Measures' section as measures.

Table	Field
Tables	City
	Event
	Games
	Medal
	Name
	NOC
	Season
	Sex
	Sport
	Team
	Measure Names
Measures	Age
	Height
	ID
	Weight
	Year
	athlete_events.csv (Count)
	Latitude (generated)
	Longitude (generated)
Measure Values	

Blue = discrete

→ Add headers to the view

Green = continuous

→ Add axes to the view

Dimensions contain qualitative values (e.g., names, dates, locations). You can use dimensions to categorize or segment your data. They affect the level of detail in the view.

Measures contain numeric, quantitative values that you can measure. Measures can be aggregated. When you drag a measure into the view, Tableau applies an aggregation to that measure (by default).

Dimensions and Measures

The screenshot shows a data tool interface. On the left, there is a sidebar with a 'Data' tab and an 'Analytics' tab. Below the tabs, there is a search bar and a list of tables. The tables listed are: City, Event, Games, Medal, Name, NOC, Season, Sex, Sport, Team, and Measure Names. Below this list, there is a list of fields: Age, Height, ID, Weight, Year, athlete_events.csv (Count), Latitude (generated), Longitude (generated), and Measure Values. The 'Age' field is highlighted with a green bar. A context menu is open over the 'Age' field, showing options: Add to Sheet, Duplicate, Rename, Hide, Create, Transform, Convert to Discrete, and Convert to Dimension. The 'Convert to Discrete' and 'Convert to Dimension' options are highlighted with a red box.

Blue = discrete

→ Add headers to the view

Green = continuous

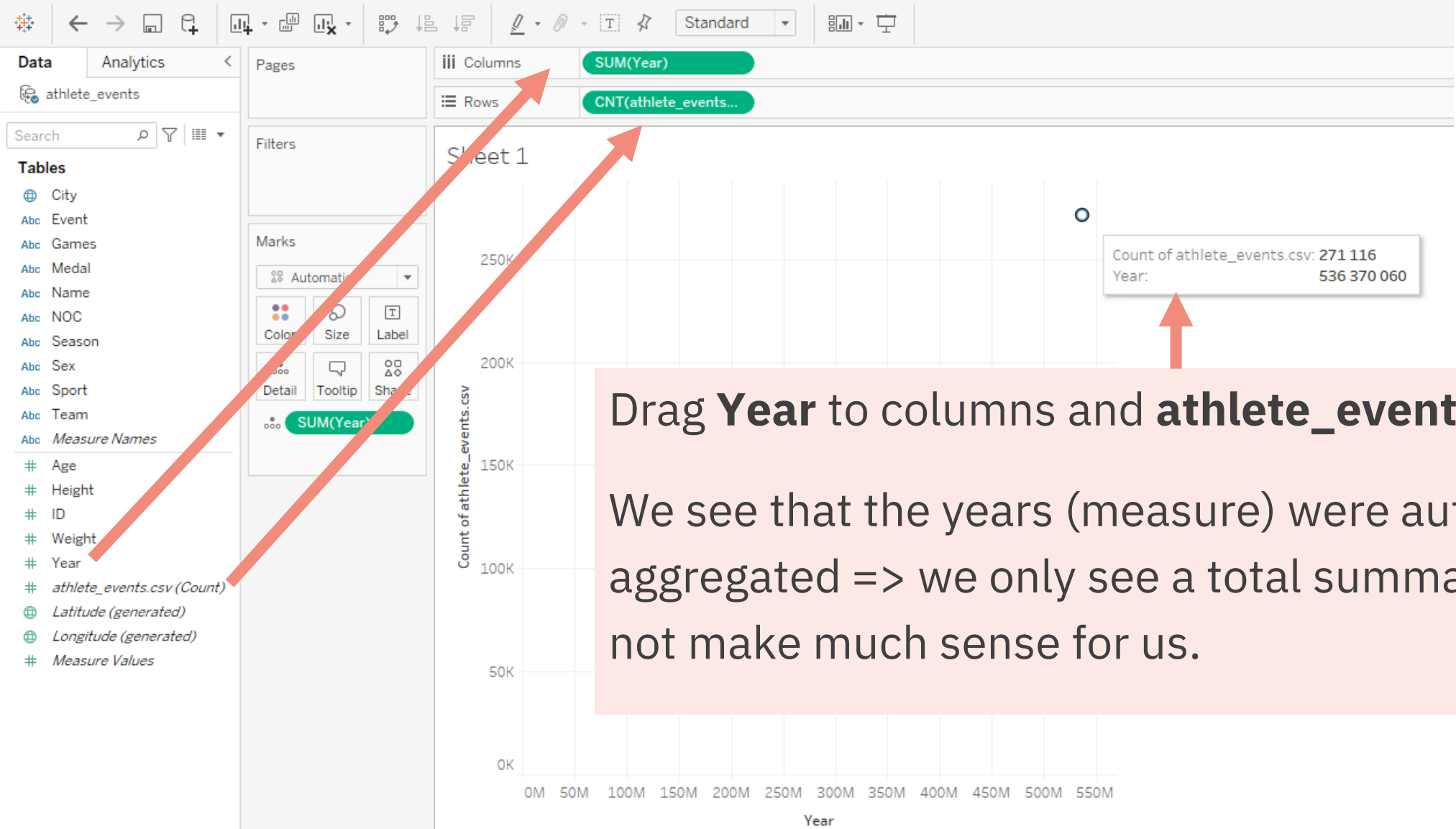
→ Add axes to the view

Dimensions contain qualitative values (e.g., names, dates, locations). You can use dimensions to categorize or segment your data. They affect the level of detail in the view.

Measures contain numeric, quantitative values that you can measure. Measures can be aggregated. When you drag a measure into the view, it is automatically aggregated (unless you change the default).

Can be changed by drag & drop or via context menu

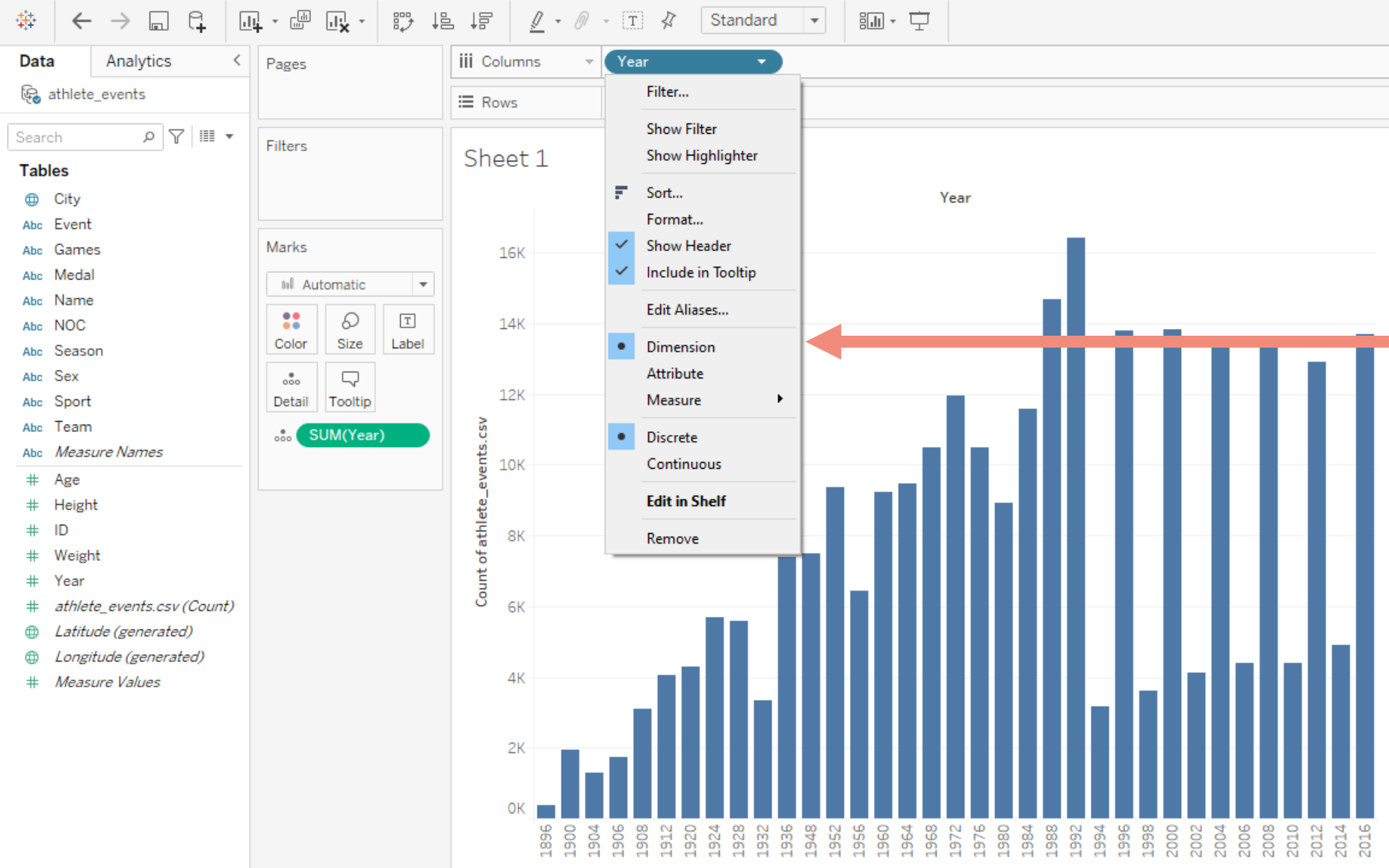
Let's create a **barchart** for number of **entries** for each **year**.



Drag **Year** to columns and **athlete_events.csv** to rows.

We see that the years (measure) were automatically aggregated => we only see a total summary, which does not make much sense for us.

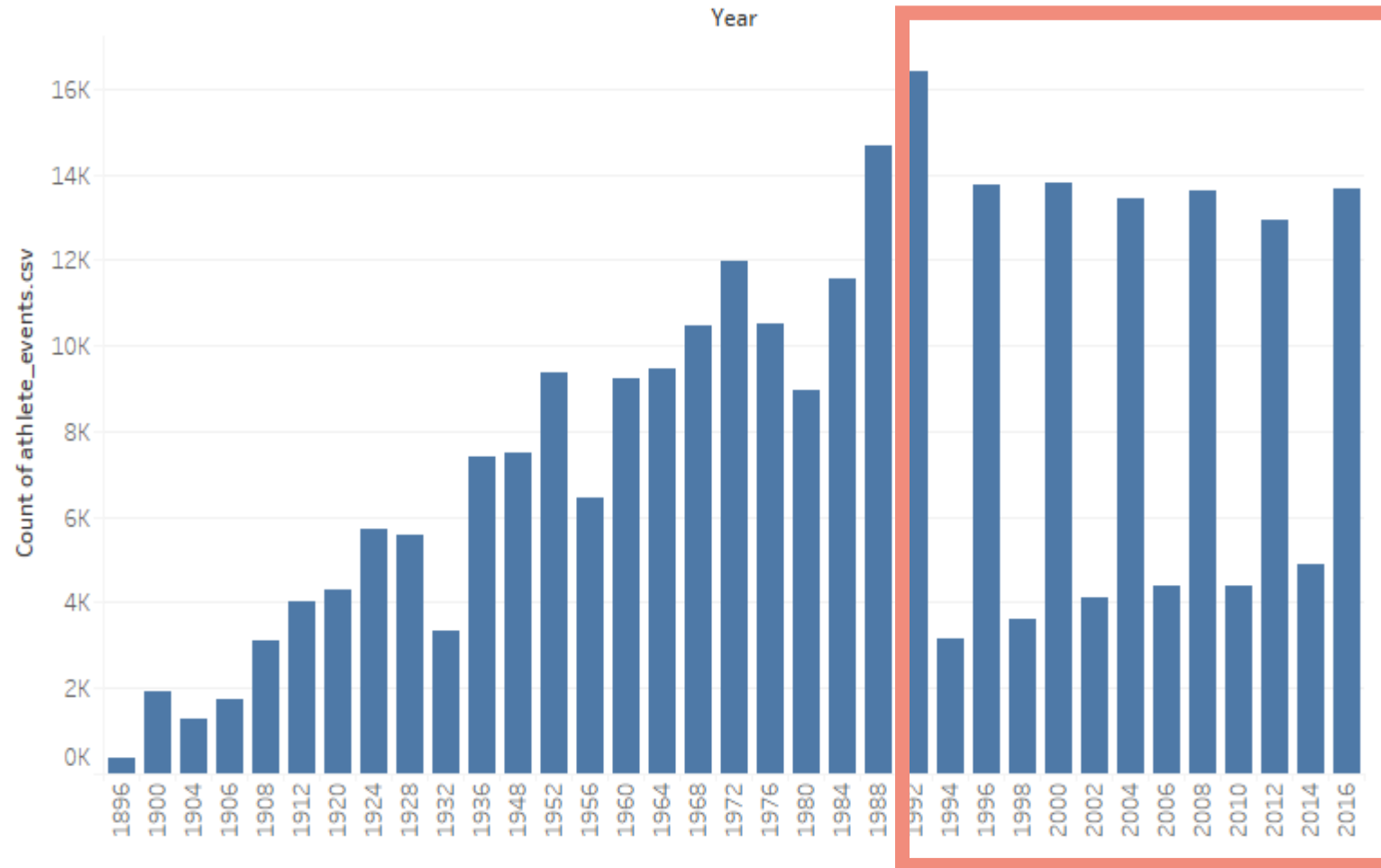
Let's create a **barchart** for number of **entries** for each **year**.



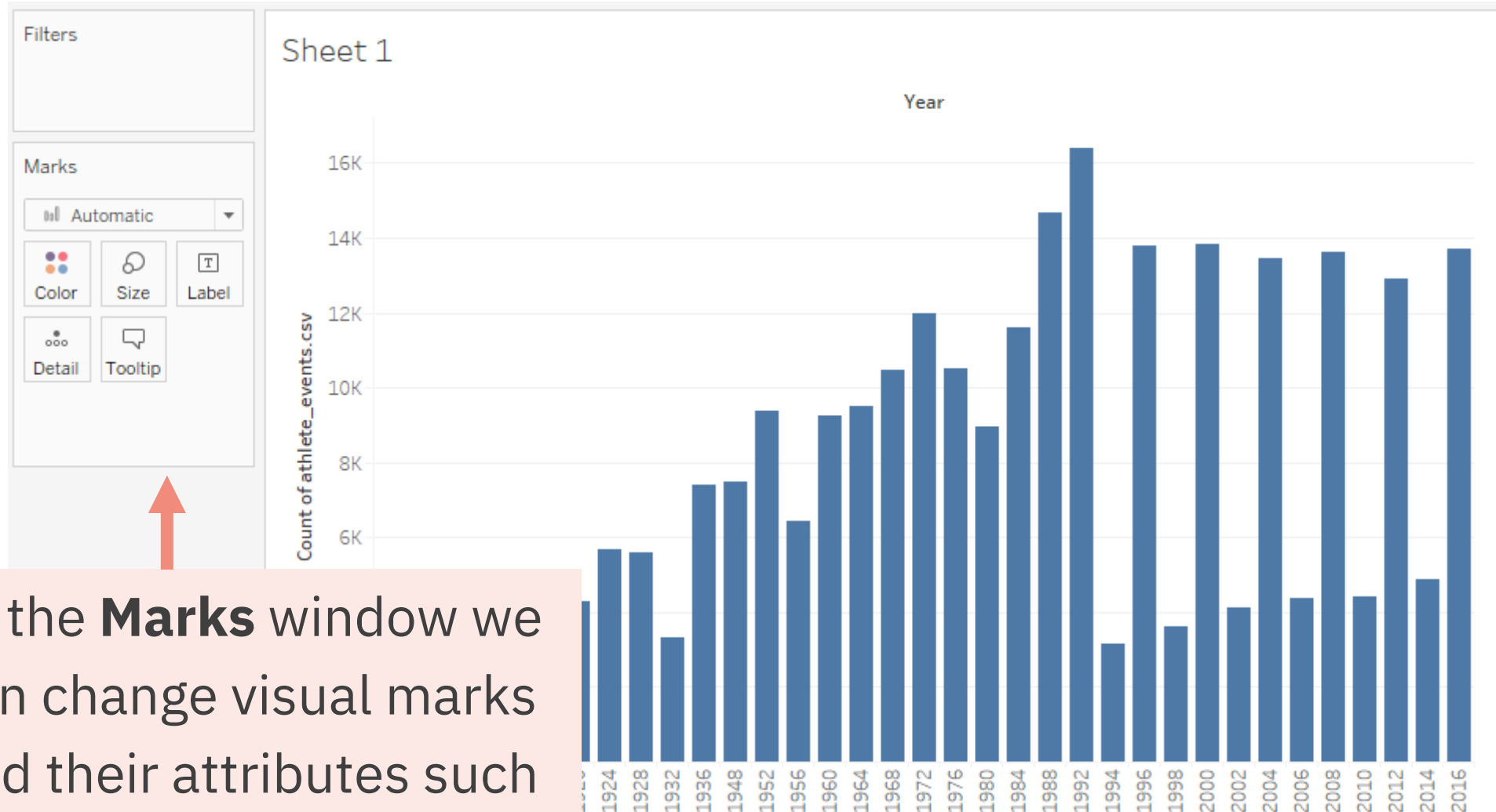
To segment the data by years we need **Year** to be a **discrete dimension**.

Can you guess why the numbers fluctuate so much from 1994?

Sheet 1



Let's color the entries by the season.



In the **Marks** window we can change visual marks and their attributes such as color or size.

We can see when winter games were introduced and when they started alternating with summer games.

Filters

Marks

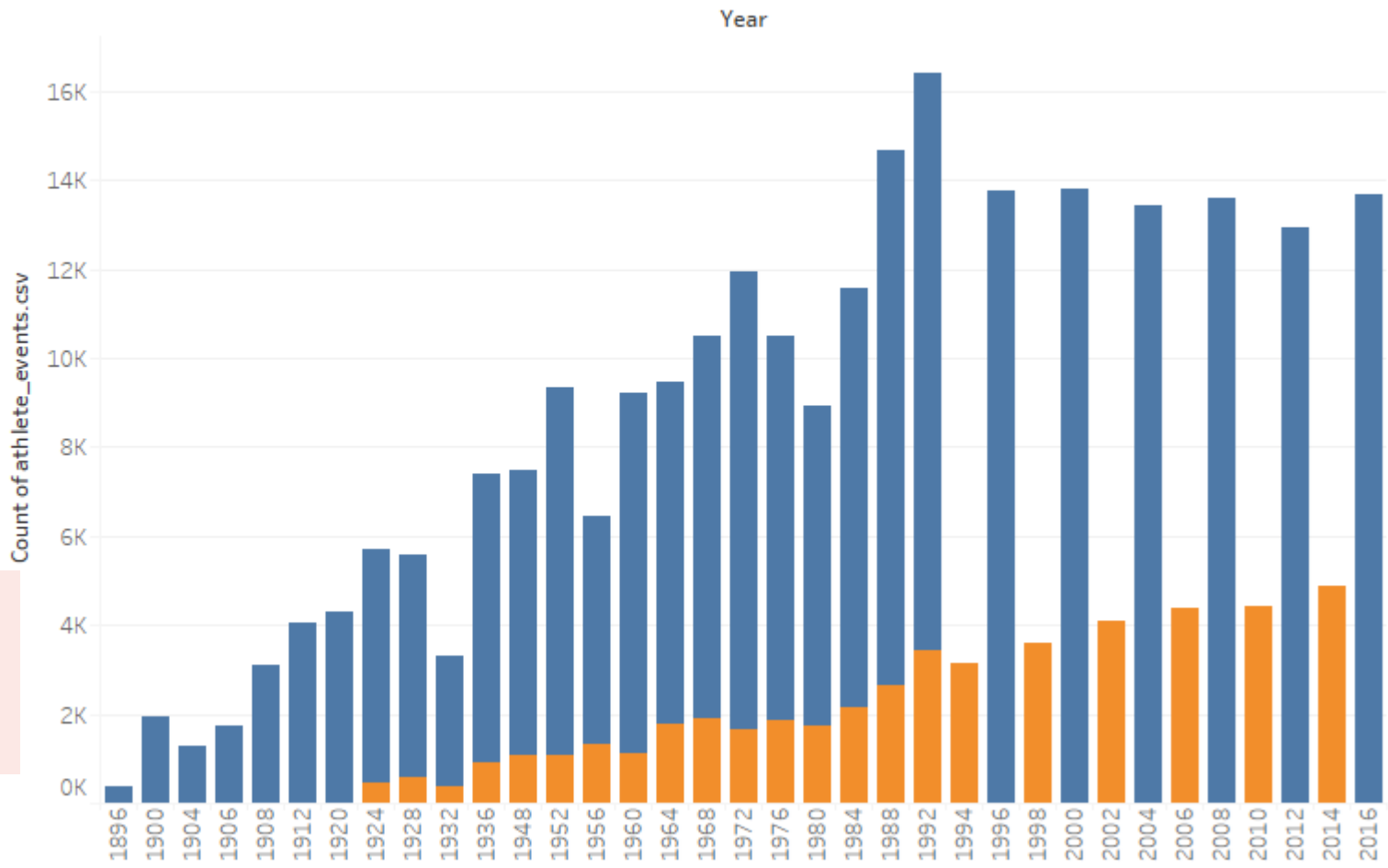
Automatic

Color Size Label

Detail Tooltip

Season

Sheet 1



Season

Summer

Winter

Drag **Season** field on **Color**.

Double click on legend to change colors.

Similarly, we can define filters by dragging fields into filter window.

The screenshot displays the Tableau interface. On the left, the 'Filters' shelf contains a pill for 'Season: Winter'. An orange arrow points from this pill to the 'Filter [Season]' dialog box on the right. The dialog box is open to the 'General' tab, showing 'Select from list' selected. The search text field is empty. Below the search field, 'Summer' is unchecked and 'Winter' is checked. At the bottom of the dialog, the 'Summary' section shows 'Field: [Season]', 'Selection: Selected 1 of 2 values', 'Wildcard: All', 'Condition: None', and 'Limit: None'. The 'OK' button is highlighted with a blue border.

The main view shows a bar chart titled 'Year' with the y-axis labeled 'Count of athlete_events.csv'. The x-axis lists years from 1924 to 2014. The bars show an overall increasing trend in the number of events over time.

Year	Count of athlete_events.csv
1924	400
1928	500
1932	300
1936	900
1948	1100
1952	1100
1956	1300
1960	1100
1964	1800
1968	1900
1972	1600
1976	1800
1980	1700
1984	2100
1988	2600
1992	3400
1994	3100
1998	3600
2002	4100
2006	4400
2010	4400
2014	4900

Changing Visualizations

The screenshot shows the Tableau interface. The main view is a box-and-whisker plot titled 'Season' showing the 'Count of athlete_events.csv' for 'Summer' and 'Winter'. The y-axis ranges from 0K to 14K. The 'Summer' box is significantly higher than the 'Winter' box. The 'Show Me' menu is open on the right, displaying various visualization options. A red arrow points from the 'Show Me' menu to the box-and-whisker plot in the main view.

Columns: Season
Rows: CNT(athlete_events...)

Sheet 1

Count of athlete_events.csv

Season

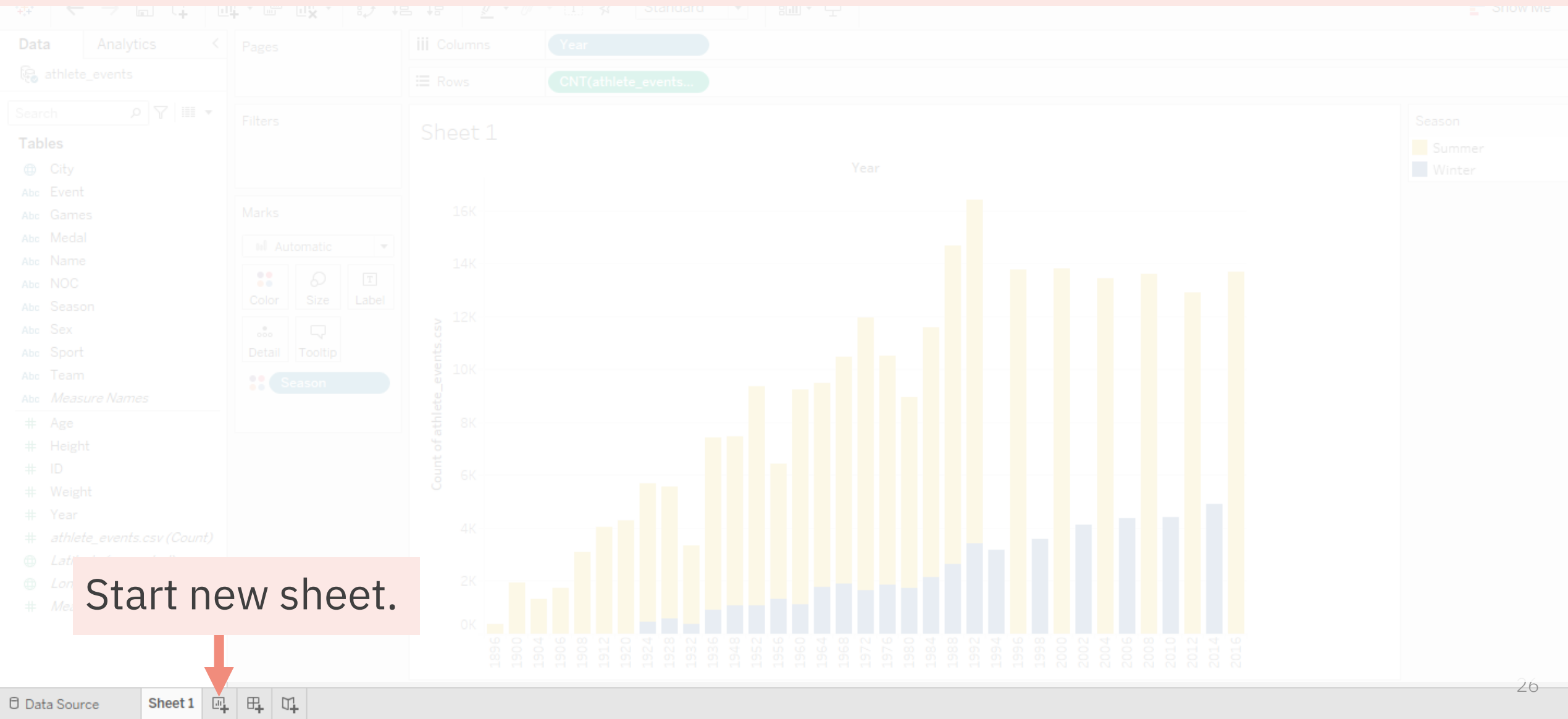
Summer Winter

Remove the filter and look into the **Show Me** menu. Here you can choose from all available visualizations for current fields.

For example, **box-and-whisker** plots.

For **box-and-whisker** plots to work:
0 or more **Dimensions**
1 or more **Measures**
Use at least 1 dimension or disaggregate

Calculated Fields: How many women received gold in different disciplines?



Calculated Fields: How many women received gold in different disciplines?

The image shows a Tableau interface with a data window for 'athlete_events'. A red callout box contains the text: "Right click into data window and select **Create Calculated Field...**". Below this, a context menu is open, with "Create Calculated Field..." highlighted in red. An arrow points from this menu item to a "FemaleGold" calculated field dialog box. The dialog box contains the following SQL expression: `COUNT(IF [Medal] = 'Gold' AND [Sex] = 'F' THEN 1 END)`. Below the expression, it says "The calculation is valid." and has "Apply" and "OK" buttons. To the right of the dialog box is a list of aggregation functions: "All", "count", "COUNT", "COUNTD", "RUNNING_COUNT", and "WINDOW_COUNT". The "COUNT" function is selected. To the right of the list is a partial description of the COUNT function: "COUNT (expression... Returns the number of rows in a group. NULL values are not counted. Example: COUNT([...])".

Right click into data window and select **Create Calculated Field...**

Create Calculated Field...

FemaleGold

```
COUNT(IF [Medal] = 'Gold' AND [Sex] = 'F' THEN 1 END)
```

The calculation is valid.

Apply OK

count

COUNT

COUNTD

RUNNING_COUNT

WINDOW_COUNT

COUNT (expression... Returns the number of rows in a group. NULL values are not counted. Example: COUNT([...])

Calculated Fields: How many women received gold in different disciplines?

' data-bbox="144 206 890 815"/>

FemaleGold

```
COUNT(IF [Medal] = 'Gold' AND [Sex] = 'F' THEN 1 END)
```

function field – in [] value – in “

The calculation is valid. Apply OK

All

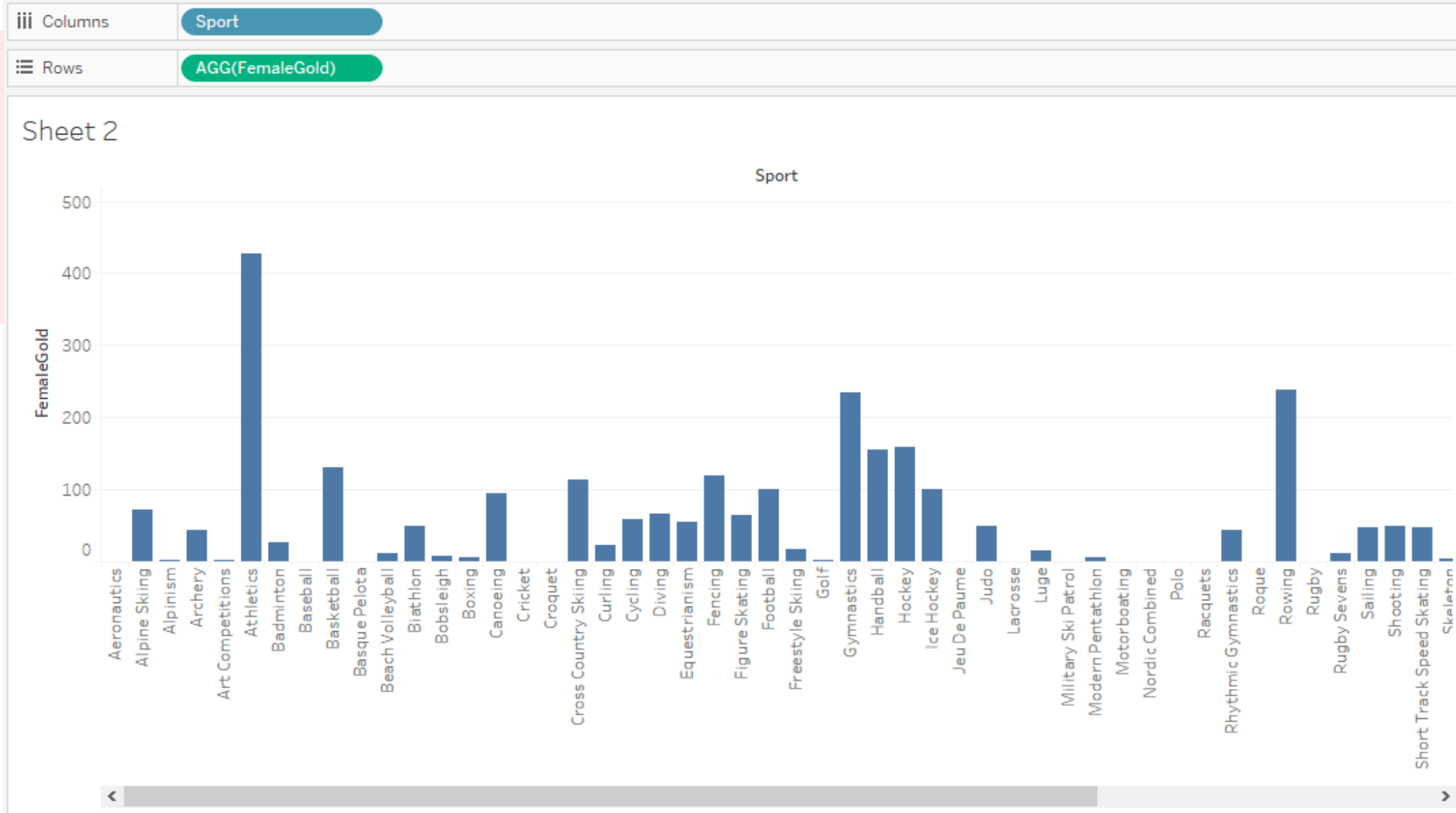
- count
- COUNT
- COUNTD
- RUNNING_COUNT
- WINDOW_COUNT

COUNT (expression)
Returns the number of items in a group. NULL values are not counted.
Example: COUNT([Customer ID])

COUNT(IF [Medal] = 'Gold' AND [Sex] = 'F' THEN 1 END)

Calculated Fields: How many women received gold in different disciplines?

Drag **Sport** to columns and **FemaleGold** to rows.



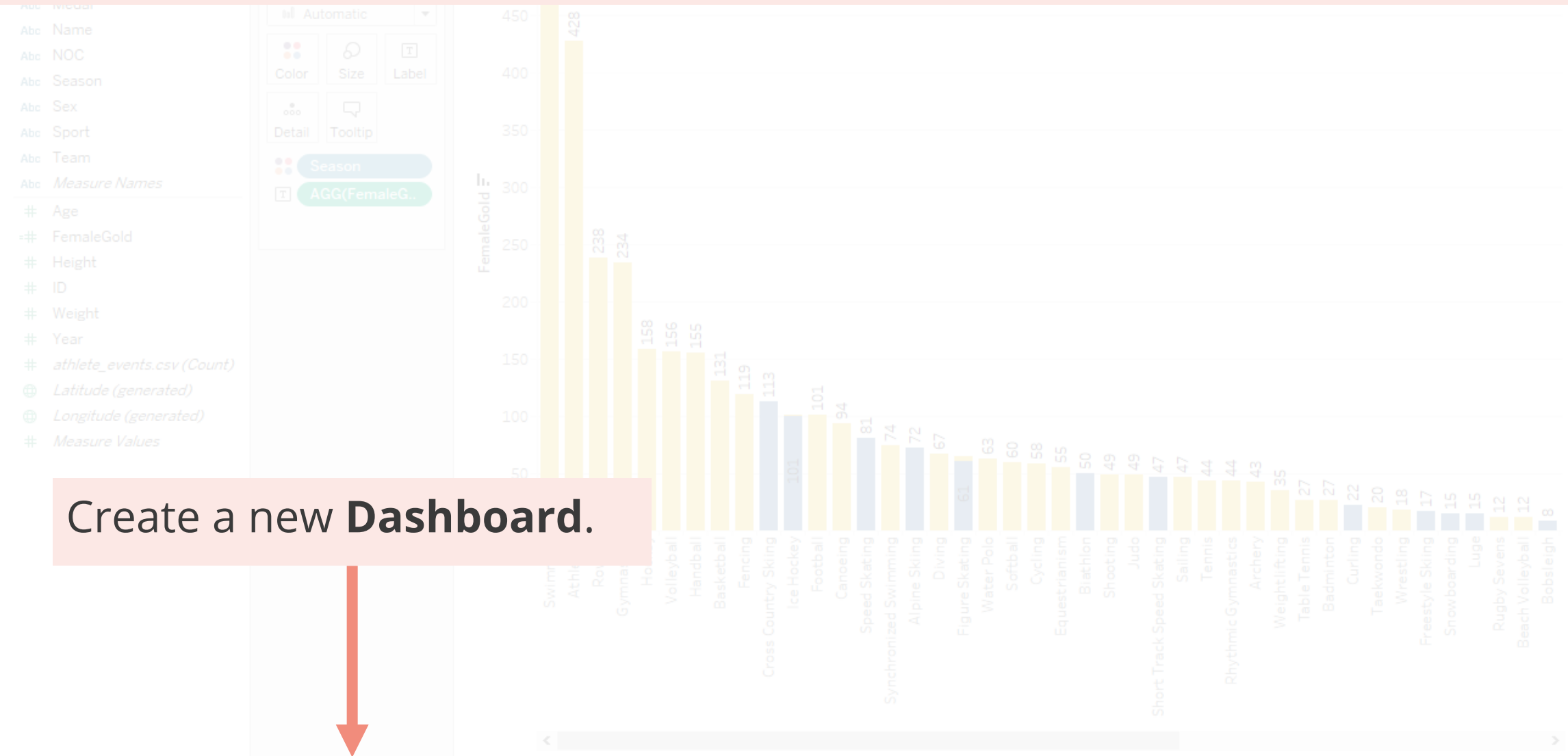
Calculated Fields: How many women received gold in different disciplines?

Sort disciplines by number of medals by clicking on the sort icon next to y-axis label.

Drag **FemaleGold** onto **Label** to see the exact numbers. Add coloring by **Season**.



Dashboards – combining multiple visualizations

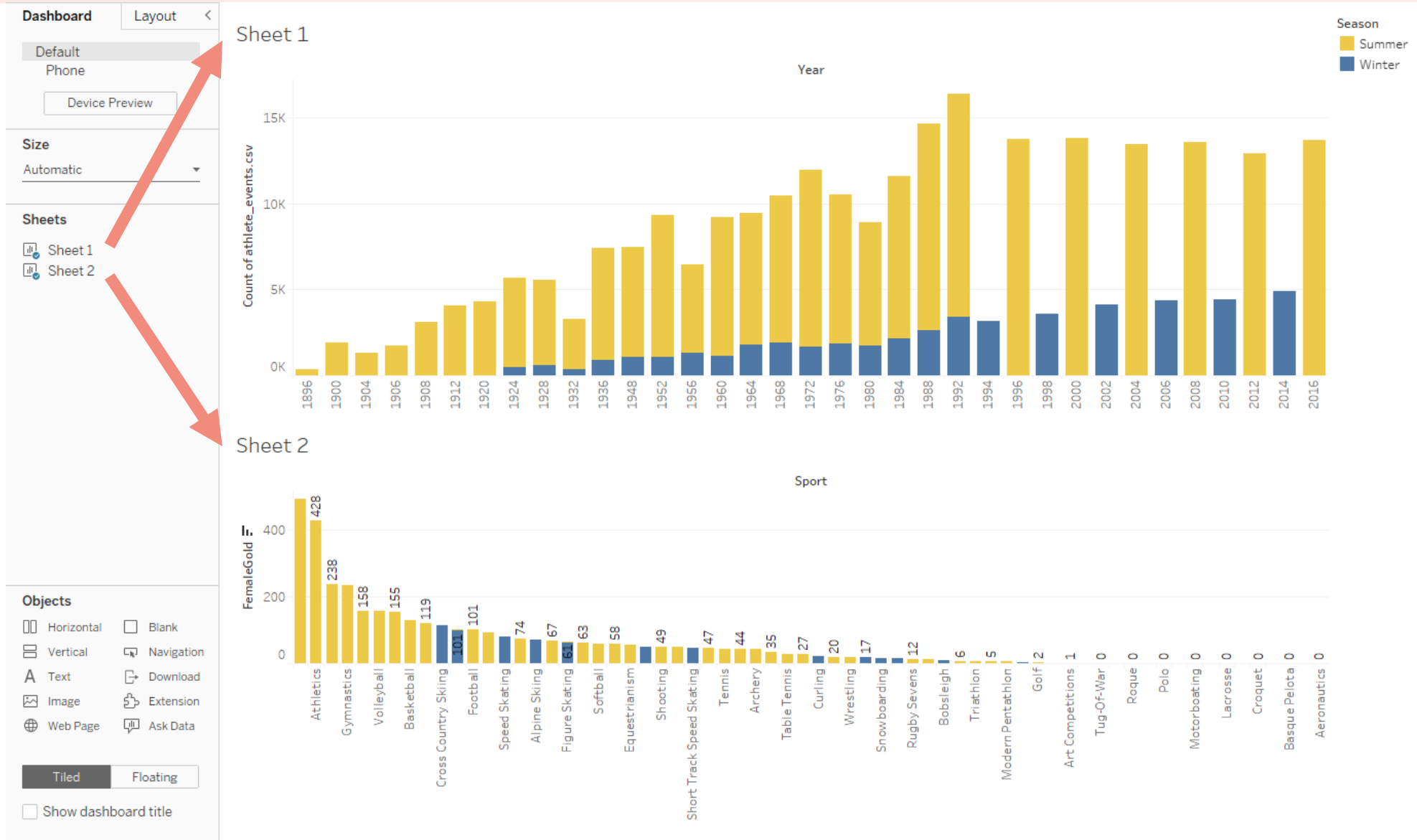


Create a new **Dashboard**.



Dashboards – combining multiple visualizations

Drag and drop the two sheets to the empty space.



Brushing and Linking Two Sheets

The screenshot shows the Tableau Desktop interface. The 'Dashboard' menu is open, and the 'Actions...' option is selected. The 'Actions' dialog box is also open, showing the 'Filter...' option selected in the list of actions. A red arrow points from the 'Actions...' menu item to the '1. Open Dashboard - Actions...' text box. Another red arrow points from the 'Filter...' option in the 'Actions' dialog to the '2. Select Add Action - Filter...' text box.

1. Open **Dashboard - Actions...**

2. Select **Add Action - Filter...**

Name	Run On	Source	Fields
Filter...			
Highlight...			
Go to URL...			
Go to Sheet...			
Change Parameter...			
Change Set Values...			

Brushing and Linking Two Sheets

Use both sheets as source as well as target so the interaction works both ways.

Use **Select** as a method and **Show all values** as a result of clearing the selection.

Add Filter Action

Name
Filter1 Insert

Source Sheets
Dashboard 1
Sheet 1
Sheet 2

Run action on
 Hover
 Select
 Menu
 Single-select only

Target Sheets
Dashboard 1
Sheet 1
Sheet 2

Clearing the selection will
 Keep filtered values
 Show all values
 Exclude all values

Filter
 All fields Selected fields

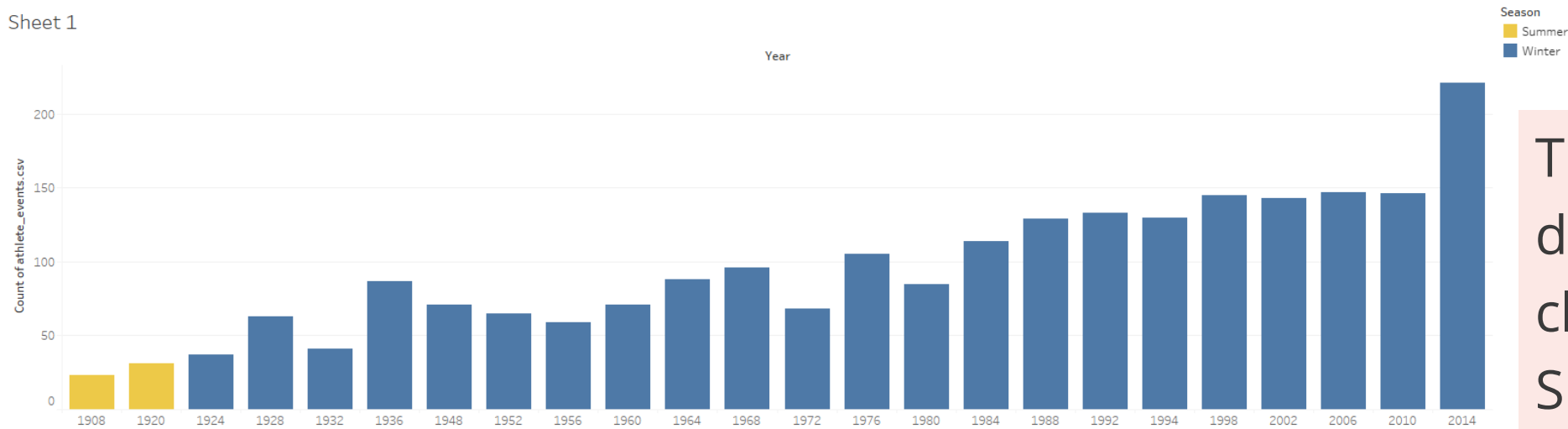
<input type="checkbox"/>	Source Field	Target Data Source	Target Field
<input type="checkbox"/>	Click to add		

Remove

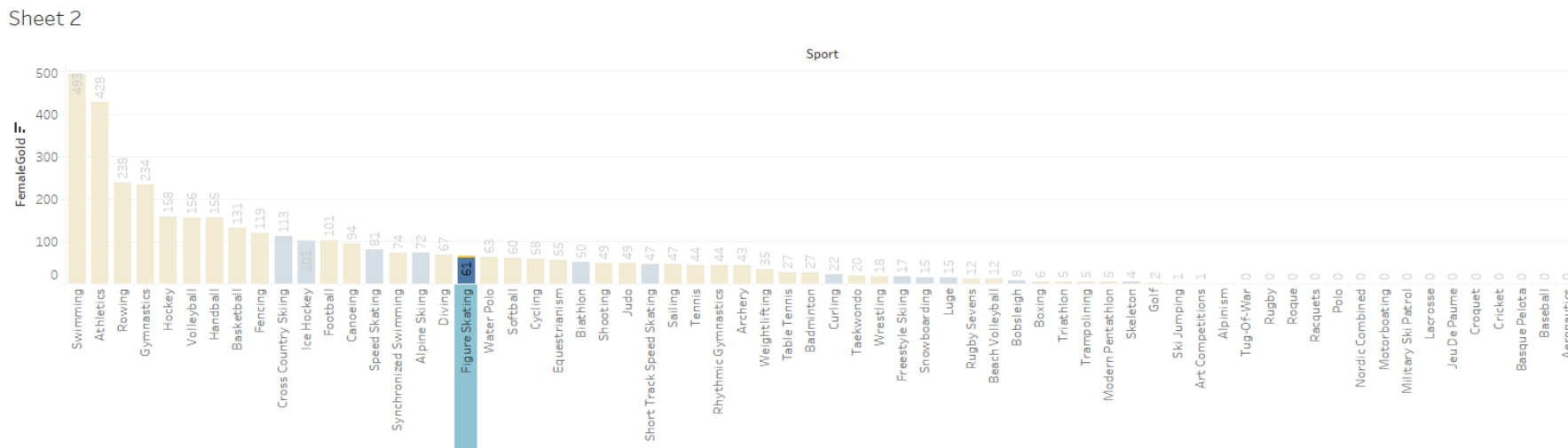
Cancel OK

Brushing and Linking Two Sheets

Sheet 1



Sheet 2

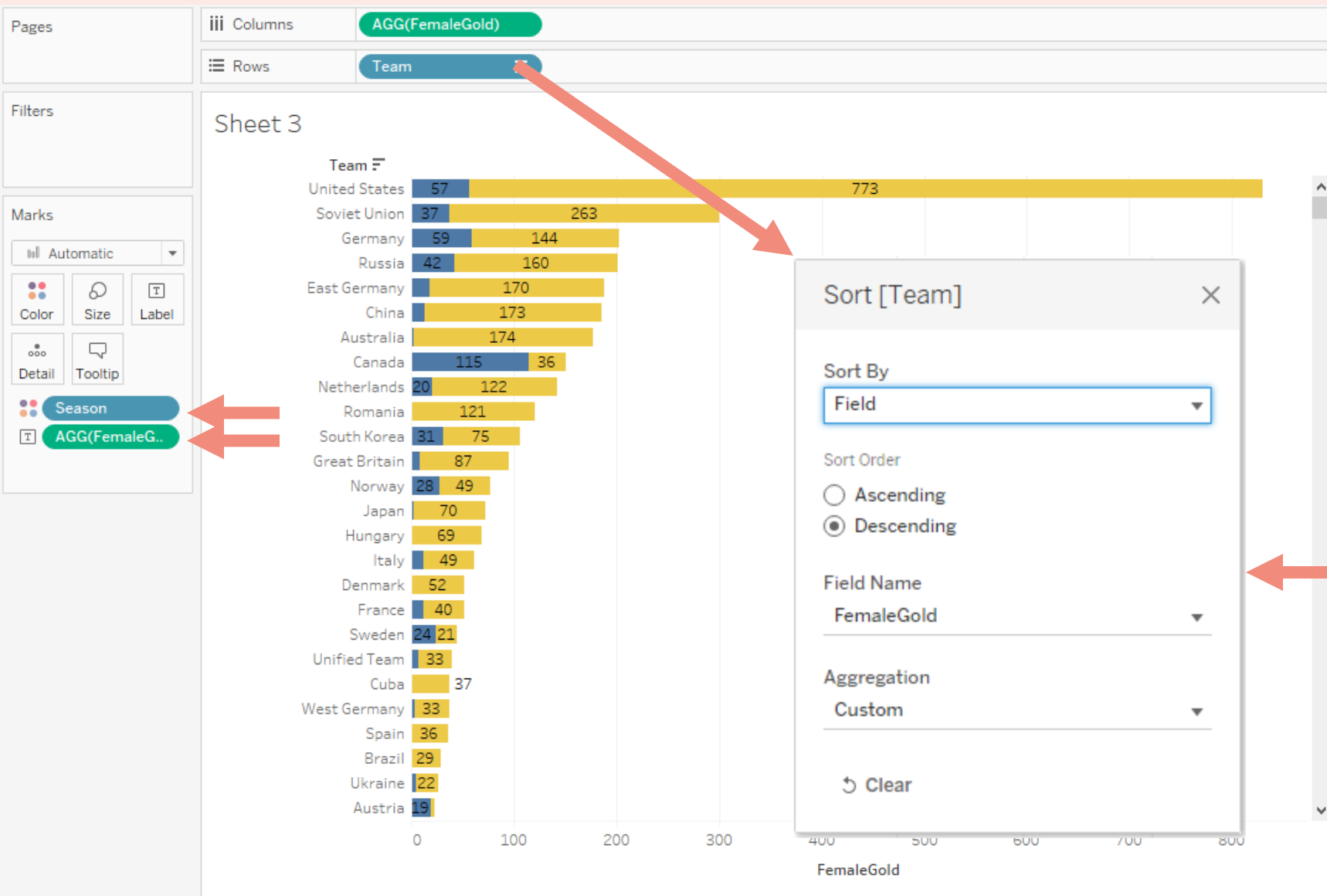


This allows further data analysis. E.g., clicking on Figure Skating updates the top chart to show data only for this discipline. We see that in years 1908 and 1920, it was part of the Summer games.

Questions

- How many gold medals were awarded to female competitors in 1972 in Athletics?
- Which female team received the most gold medals in Basketball?

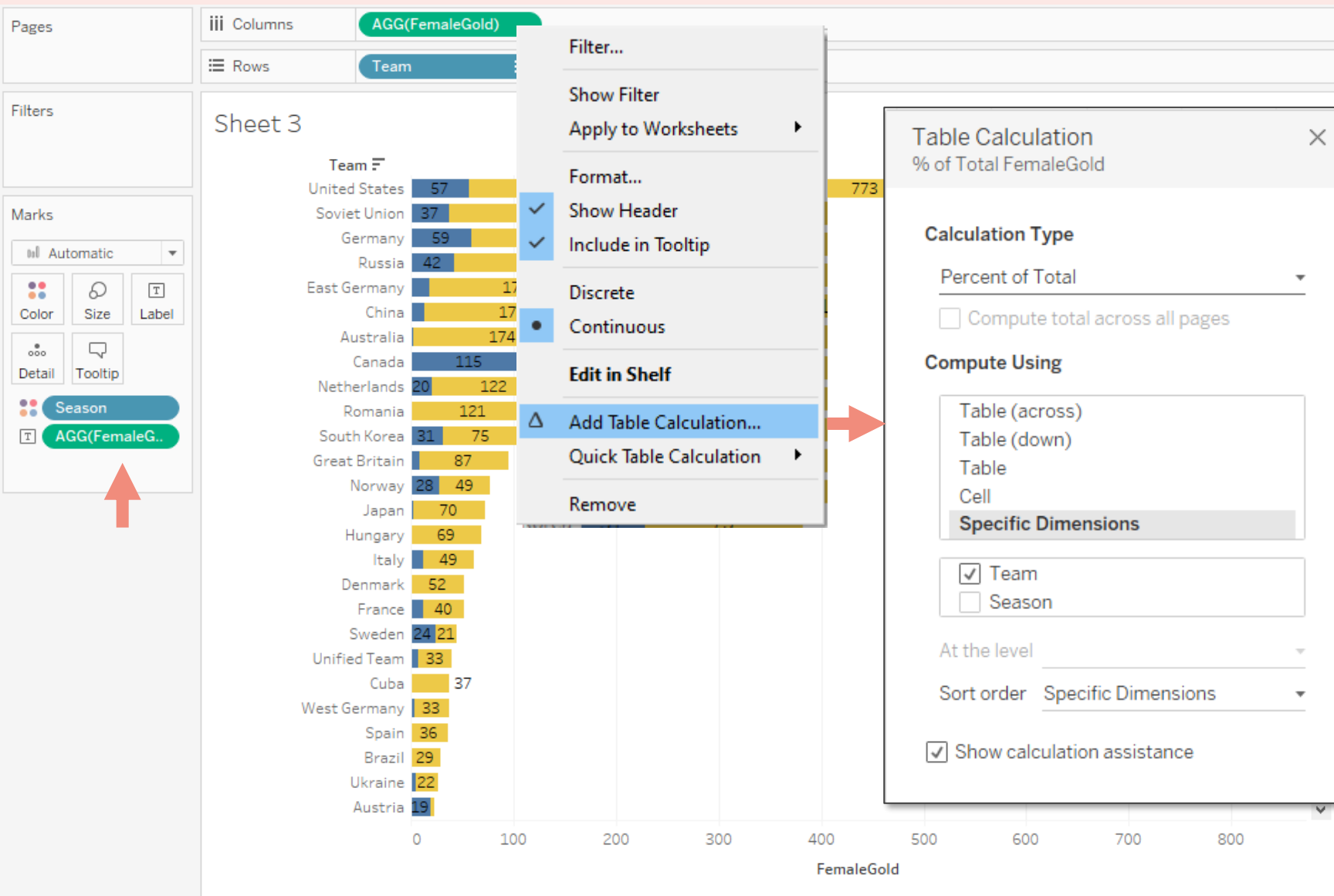
Team Statistics



1. Add new sheet with **FamaleGold** as columns and label, **Team** as rows and color by **Season**.

2. Right click **Team** (in rows header), select **Sort...** In the opened window set Sort By **Field** and **Descending** order.

Let's compare relative amounts of medals...

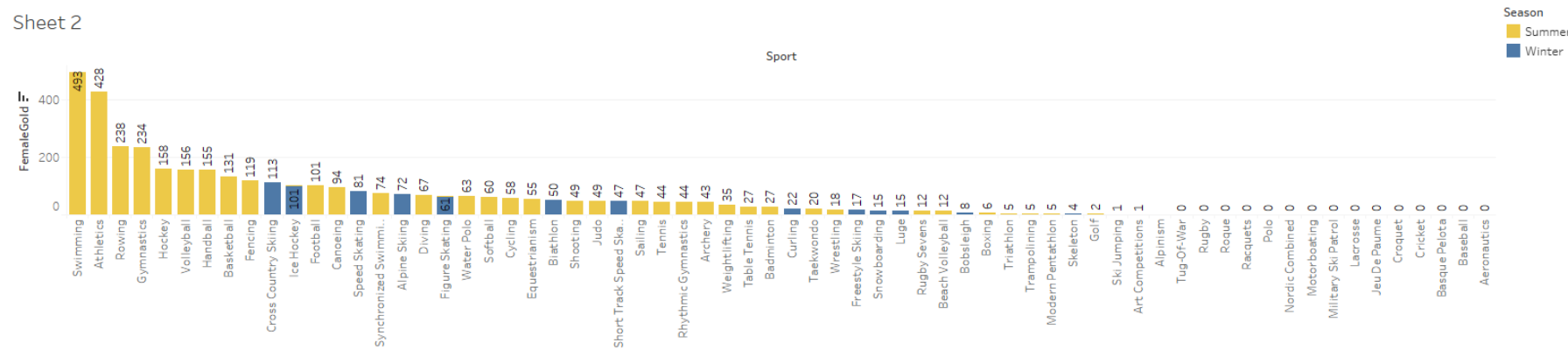


Right click **FamaleGold** (in rows header), select **AddTableCalculation...** Set calculation type as **Percent of Total**. Use **Specific Dimensions – Team**.

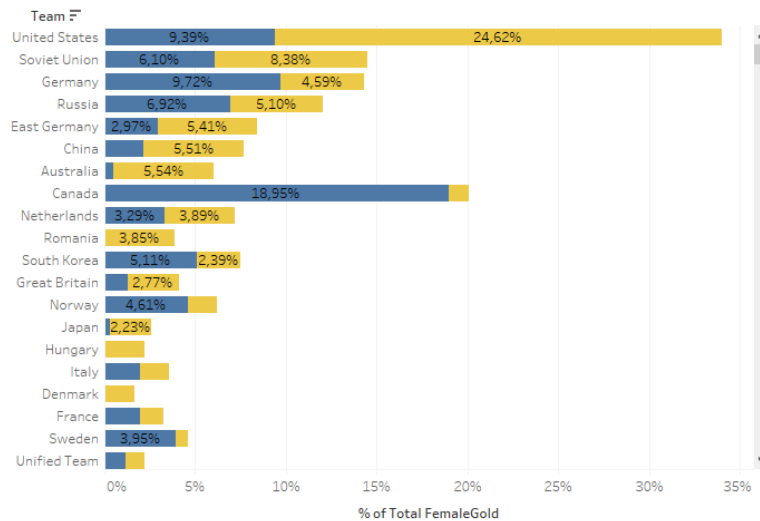
Repeat for FemaleGold label in Marks.

Let's compare relative amounts of medals...

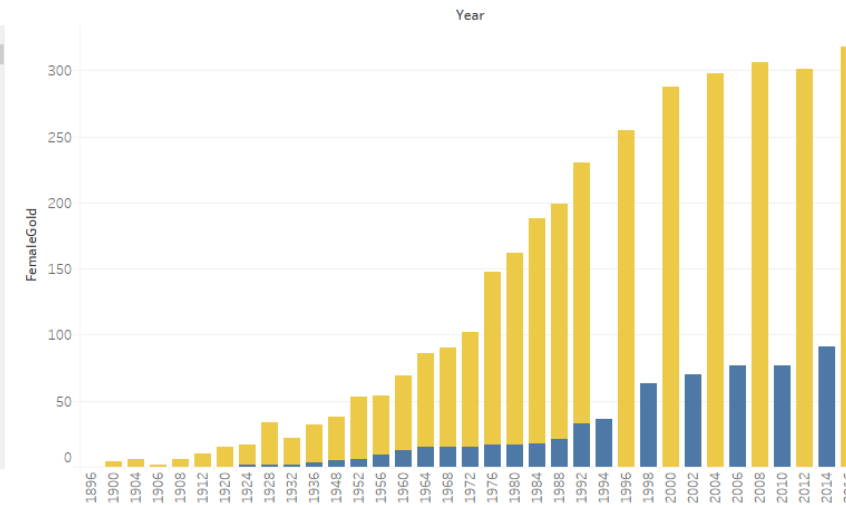
Sheet 2



Sheet 3



Sheet 1



Now we can compare performance of countries.

E.g., we see that Canada is doing much better in winter Olympics than in summer, while for USA it is the other way around.

Individual Tasks

1. Which age and weight combination results in the most medals in judo (irrespective of sex and medal type)?
2. Is age and medal somehow correlated (for specific sports or generally)?