

Final Project Proposal

Garrett Birch, Zixin Ma, Yikai Mao
Breakout Group #3

Project Topic: Analysis of Beijing 2022 Olympic Winter Games

Dataset Description: This dataset contains athletes, coaches, medals, results for curling and hockey, and other records for Beijing 2022 Olympic Winter Games.

Dataset Link: <https://www.kaggle.com/piterfm/beijing-2022-olympics>

Reason to choose: We are all interested in sports topics, and the Beijing winter olympics just ended. It is interesting to see who won the games and which country got the most medals.

Screenshots of some table items we might use:

athletes.head()

	name	short_name	gender	birth_date	birth_place	birth_country	country	country_code	discipline	discipline_code	residence_place	residence_country	height_m/ft	url
0	AAGAARD Mikkel	AAGAARD M	Male	1995-10-18	FREDERIKSHAVN	Denmark	Denmark	DEN	Ice Hockey	IHO	ORNSKOLDSVIK	Sweden	1.84/6'0"	../en/results/ice-hockey/athlete-profile...
1	AALTO Antti	AALTO A	Male	1995-04-02	KITEE	Finland	Finland	FIN	Ski Jumping	SJP	KUOPIO	Finland	NaN	../en/results/ski-jumping/athlete-profile...
2	AALTONEN Miro	AALTONEN M	Male	1993-06-07	JOENSUU	Finland	Finland	FIN	Ice Hockey	IHO	PODOLSK	Russian Federation	1.80/5'10"	../en/results/ice-hockey/athlete-profile...
3	ABDELKADER Justin	ABDELKADER J	Male	1987-02-25	MUSKEGON, MI	United States of America	United States of America	USA	Ice Hockey	IHO	GRAND RAPIDS, MI	United States of America	1.87/6'1"	../en/results/ice-hockey/athlete-profile...
4	ABDI Fayik	ABDI F	Male	1997-10-07	SAN DIEGO, CA	United States of America	Saudi Arabia	KSA	Alpine Skiing	ALP	DHAHRAN	Saudi Arabia	NaN	../en/results/alpine-skiing/athlete-profile...

medals.head()

	medal_type	medal_code	medal_date	athlete_short_name	athlete_name	athlete_sex	athlete_link	event	country	country_code	discipline	discipline_code
0	Gold	1	2022-02-05 00:00:00.0	ECKHOFF T	ECKHOFF Tiril	X	../en/results/biathlon/athlete-profile-n...	Mixed Relay 4x6km (W+M)	Norway	NOR	Biathlon	BTH
1	Gold	1	2022-02-05 00:00:00.0	ROEISELAND MO	ROEISELAND Marte Olsbu	X	../en/results/biathlon/athlete-profile-n...	Mixed Relay 4x6km (W+M)	Norway	NOR	Biathlon	BTH
2	Gold	1	2022-02-05 00:00:00.0	BOE T	BOE Tarjei	X	../en/results/biathlon/athlete-profile-n...	Mixed Relay 4x6km (W+M)	Norway	NOR	Biathlon	BTH
3	Gold	1	2022-02-05 00:00:00.0	BOE JT	BOE Johannes Thingnes	X	../en/results/biathlon/athlete-profile-n...	Mixed Relay 4x6km (W+M)	Norway	NOR	Biathlon	BTH
4	Silver	2	2022-02-05 00:00:00.0	CHEVALIER-BOUCHET	CHEVALIER-BOUCHET Anaïs	X	../en/results/biathlon/athlete-profile-n...	Mixed Relay 4x6km (W+M)	France	FRA	Biathlon	BTH

TASKs with visualization plots:

Requirement: at least one scatterplot, one bar plot/line plot/dot plot, two heatmap/stacked bar graph/streamgraph, one boxplot/network/tree/adjacency matrix

Requirements: 1 viz w/ buttons, 1 viz w/ animation, 1 viz w/ tooltip, 1 viz w/ interactive event not discussed in class

1. Explore the relationship between the numbers of athletes by countries and the numbers of medals by countries-> scatter plot

2. Explore the numbers of athletes by countries ->bar plot;

Explore the numbers of each discipline -> bar plot

3. Explore the numbers of Gold, Silver, Bronze medals by countries ->stacked bar graph;

4. Explore the number of medals per country by day -> streamgraph

Possible alternative: Explore number of female, males per event -> stacked bar graph

5. Explore the ages of athletes ->box plot

6. (Extra Credit) Explore the distributions of gold medals across countries -> geomap-based plot, possibly a choropleth plot. (Could have a tooltip when hovering over a country to display the country + num of medals?)

Timeline:

Finding Dataset: Garrett Birch

Drafting project proposal: Garrett Birch, Zixin Ma, Yikai Mao

Data cleaning: Datasets are almost clean. Maybe do some joins/merges. No major null values present.

Coding: Garrett Birch, Zixin Ma, Yikai Mao

Documentation: Done by all members after/during coding

Interactivity Ideas:

Button: Will be used in plot #5, Explore the ages of athletes with a box plot. The button will serve as a filter for looking between female or male athletes.

Tooltip: Will be used in plot #6, choropleth geomap plot. The tooltip will allow users to hover over a certain country and see what medals they have won.

Animation: Used in conjunction with a button to sort the data in plot #2, explore the number of athletes by country using a bar plot. When the user hits the button, the plot will translate and animate to a sorted bar plot.

Not Learned in Class - D3 Brush mixed with zoom used in plot #1. Will allow the user to brush (select) a certain area of the scatter plot and the plot will zoom in on this area and update x/y axis/scale.