CS171 Final Project Process Book

Group Members:
Marie Konopacki
mariekonopacki@college.harvard.edu
Miu Kumakura
mkumakura@college.harvard.edu
Nicole Araya
naraya@college.harvard.edu

Team Name: *United by Emotion* #Tokyo2020-21

Project Title: Visualizing the Olympics

Project Abstract: The abstract should include your **motivation** for doing this project, what you are trying to do, and which **goals** you want to achieve. Also, include a description of your planned **datasets** and where you will acquire them. *

Our project aims to visualize multiple aspects of the Olympics over the years, ranging from records, gender, economics, and the sites of the competition, while also incorporating interactive elements to allow the user to understand certain athletic feats. We are interested in exploring the Olympics because of its relevance to this year and the multidimensionality of available data. While the Olympic Games are well-renowned throughout the globe, most people have a surface-level understanding of it that revolves around the types of sports included, specific athletes, and world records. We would like to create an interactive visualization that allows for a more insightful look at these events which are familiar but not too familiar. We hope to invite users to learn about how there is more at play at each Olympic Game than the sports themselves; issues such as gender, country resources, and politics play a role as well.

Our planned datasets include 120 years of Olympic history athletes and results, additional Olympic history data, and Olympic costs. These datasets detail data on Olympic locations, participants, winners, costs, and records over time.

Week 9: Team Agreement and Detailed Project Plan:

Team Agreement

- Although code will be written by individuals, all team members should be involved with the technical aspects of the project. All code should be documented well.
- Final design decisions will be discussed among all members; fair compromises should be made when necessary.
- Work hours should be split as evenly as possible (actual task output may differ based on an individual's ability / previous experience). This ensures not only fairness but also learning opportunities for everyone. We will keep each other accountable so that one person does not work too much / too little.
- We will use a Git workflow to aid our progress as a team and help us split up the work of coding.
- We may be responsible for individual workflows, but we will communicate via text to update each other on progress and concerns.
- We will set a time for a weekly work session to discuss weekly assignments and review progress.

Signatures: Marie Konopacki, Nicole Araya, Miu Kumakura
Date: 10/30/2020

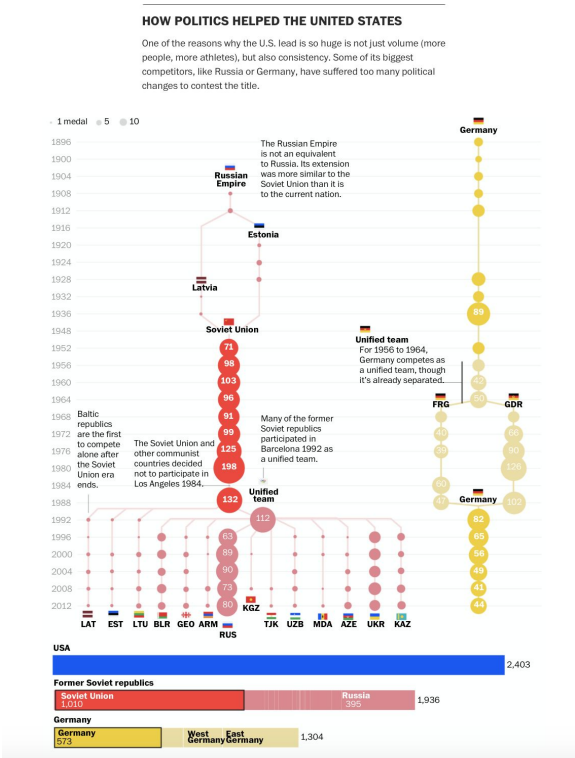
Background and Motivation

Our project aims to visualize multiple aspects of the Olympics over the years, ranging from records, gender, economics, and the sites of the competition, while also incorporating interactive elements to allow the user to understand certain athletic feats. We are interested in exploring the Olympics because of its relevance to this year and the multidimensionality of available data. While visualizations of the Olympics already exist, we found that most of them provide visualizations on specific records, such as number of medals won or the speeds of runners. For our project, we think it would be interesting to draw in data on social, economic, and political factors to develop insights into why traditionally powerful countries tend to be stronger and why countries tend to win in certain sports over others. We are also curious about personal-level data on athletes training schedules, backgrounds, and experiences to gain a deeper understanding about what has led to their success.

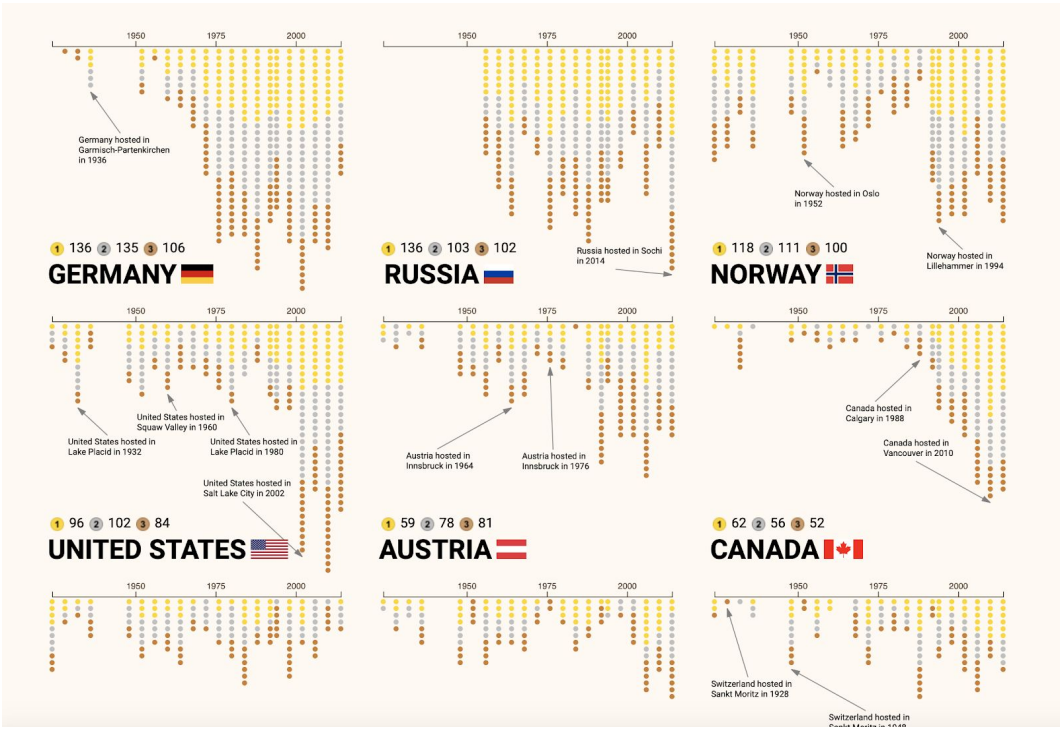
Related Work:

I. How Team USA Dominates the Olympics, The Washington Post

An example of incorporating contextual information to the data:

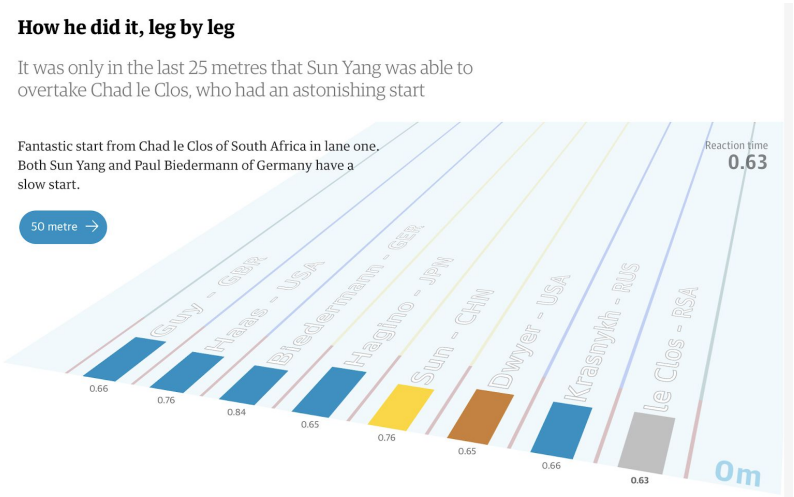


II. [The Winter Games](), Elbert Wang



III. [Olympic feathers](), Nadieh Bremer

IV. [How he did it, leg by leg:](#) Visualization of Sun Yang's gold medal swim in 2016, The Guardian

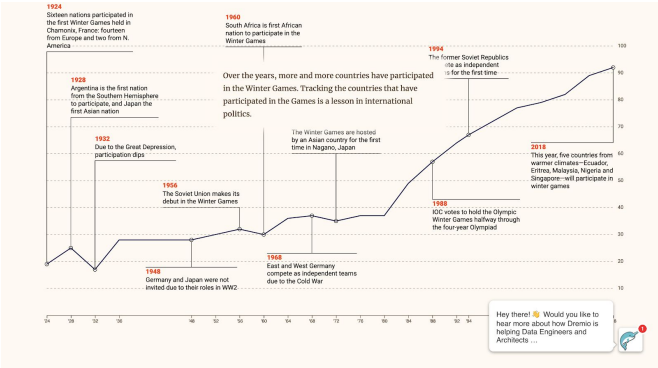


V. [5 Data Stories of the Olympic Games](#)

VI. [#Rio2016: 7 data stories and interactives from the first week of the Olympic Games](#)

Thoughts about the related work:

I. Not enough interactive components but good at drawing out and explaining patterns in the data.

II. 

Can occasionally get a bit wordy and can feel pretty monotone at times. But it does have a very clear story and progression.

III. Not a very guided visualization as it gives a lot of raw data at once. Not a lot of story. Great use of the Olympic color palette, though.

IV. We were inspired by the similarities between bar charts and race lanes, and are looking to create similar time-lapse charts to visualize changing records/increasing race speeds over time.

Audience and Questions

Our target audience is the **general public**. We invite users to compare themselves to the athletes through interactive elements of our story and encourage them to develop insights from: 1) cross-country comparisons of records/wins, 2) an analysis of how social, economic, and political factors influence the games, and 3) individual-level data that portray what it takes to become an olympic athlete.

Topics we would like to address

- Gender at the Olympics
    - What sports have equal representation?
- How records progress over time
    - How much better people have gotten over the past decades?
    - How frequently records are overturned (or infrequently)?
    - Do the same countries win each year? Each sport?
- How a regular person compares to a racer at the Olympics (Interactive, user inputs their time)
    - What sport would I/user play?
- Cheating at the Olympics (need to find dataset on doping)
    - Is there a pattern in cheating based on country, gender, sport?
- How money/geography of country correlates with wins/sports
    - Is success in Winter Olympics based on country's latitude?
    - Is GDP correlated with sports that are more expensive?
- How athletes were raised (financial background, when they started training, etc.)
    - Can also see what sports took the least amount of time to train for -- i.e. the "easiest" sport to compete in, in the Olympics
    - Where are athletes from?
    - What resources does the average Olympic athlete have access to?
- How do athletes train
    - How long do they train for (hours per day/years)?
- How much do athletes get paid
    - Is there a difference in payment between sports or gender?

Data:

1. 120 years of Olympic history athletes and results: identifies athlete data from Athens 1896 - Rio 2016 including: name, sex, age, weight, height, team, season, city, sport, event, medal
2. Olympic history data: calculated dataset focused on trends regarding the art competitions, women in the Olympics, geographic representation, and height and weight of athletes.
3. Olympic costs
4. List of articles detailing starting training age of athletes
    a. https://sportsretriever.com/uncategorized/starting-young-ages-worlds-top-athletes-start-training/

    b. https://venngage.com/blog/olympics/#:~:text=Athletic%20training%20does%20not%20start,for%20the%20past%20few%20Games.
    c. https://www.bustle.com/p/what-age-do-most-olympians-start-training-many-are-children-when-they-begin-learning-their-sport-7935404
5. Sports Accessibility:
    a. https://www.golfdigest.com/story/the-winter-olympics-are-amazing-but-also-a-shining-beacon-of-sports-privilege
    b. Tiny town in VT has a ton of olympic athletes:
        i. https://www.nytimes.com/2017/12/30/sports/olympics/a-tiny-vermont-town-is-a-big-cradle-of-olympians.html
6. Hosting the Olympic Games, an overstated advantage:
    a. https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/P0AEHQ
7. Country data:
    a. http://hdr.undp.org/en/data: Human Development Indicators that include relevant socioeconomic factors
    b. http://data.uis.unesco.org/Index.aspx?DataSetCode=DEMO_DS: dataset on GDP
8. Assessment of the "Olympic Effect on Tourism"
    a. https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/24231
9. Analyzing Elite Running Performance with Historical Data
    a. https://towardsdatascience.com/analysing-elite-running-performance-with-historical-data-b9c6bdd9c5d8
10. What I learned: Over 50 Olympic Athletes' Daily Routines
    a. https://owaves.com/what-i-learned-from-researching-50-olympians/
    b. https://owaves.com/day-plans-category/olympian/
11. Hometowns of USA Olympians (Can then get data on family income)
    a. https://esrimedia.maps.arcgis.com/apps/MapSeries/index.html?appid=c31dfaab1a8e4ab8badbc1a8bb9a8276
    b. https://www.bloomberg.com/news/articles/2016-08-18/the-geography-of-team-usa-at-the-2016-rio-olympics
    c. https://www.teamusa.org/road-to-rio-2016/team-usa/athletes
        i. ^^^ Data for hometowns/colleges in a table
    d. https://www.teamusa.org/athletes (Can webscrape potentially?)
12. Doping/cheating at the Olympics:
    a. https://en.wikipedia.org/wiki/Doping_at_the_Olympic_Games
13. Reimagining the olympics
14. Olympic sports data by country
15. Viewership
    a. NBC london data (need to web scrape)
    b. NBC rio data (need to web scrape, does not have data by sport)
    c. https://www.nielsen.com/us/en/insights/article/2010/how-different-genders-ages-races-and-regions-watch-the-olympics/
    d. Yearly change in viewership
    e. IOC broadcasting data

f. "What is your favorite Olympic event?" survey in US

Data Cleanup

We do not expect substantial data cleanup since most of our datasets are already cleaned. Some of our data come from articles, so we will have a web scraping step at the beginning. We do not anticipate having to do major calculations on our data after it has been scraped. Quantities that we plan to derive from our data include information on the family incomes of Olympic athletes, starting training ages, number of training hours per day, socioeconomic data of participating countries, athlete doping data, and economic effects of tourism on host countries, as well as the usual Olympics data of athlete country, event, and biological data.

Scraping options: https://www.octoparse.com/blog/scrape-data-from-a-table, Grepsr Chrome extension, Javascript row parsing

---------- TF feedback @ 11-03 ----------

Great job on outlining your topics, data source, audience and questions! Very nice job on collecting a variety of related work. I like that you select your target audience as the public users and outline different levels of visualization goals to draw attention to individual comparison with the olympians. This seems to be a less explored direction of olympia data, and can be very inspiring to make data more personal.

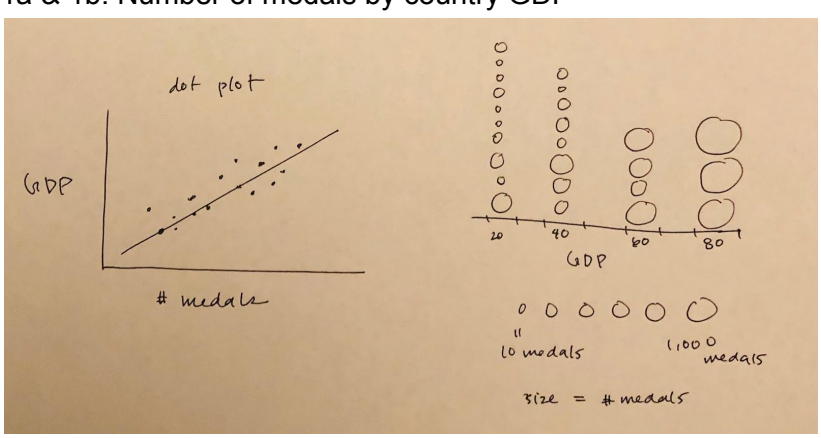
The data source you pointed towards seems sufficient to cover your questions, but since you have raised a variety of questions which require different dataset to answer, that could mean you need to spend more time preparing data for each visualization. Next step you can think about which part you want to focus on based on your storyline, which will guide where to put more design efforts.

Week 10: Data, Sketches, Decide and Storyboard

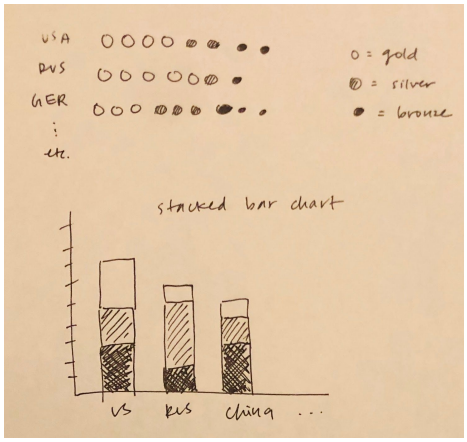
Sketches

Nicole:

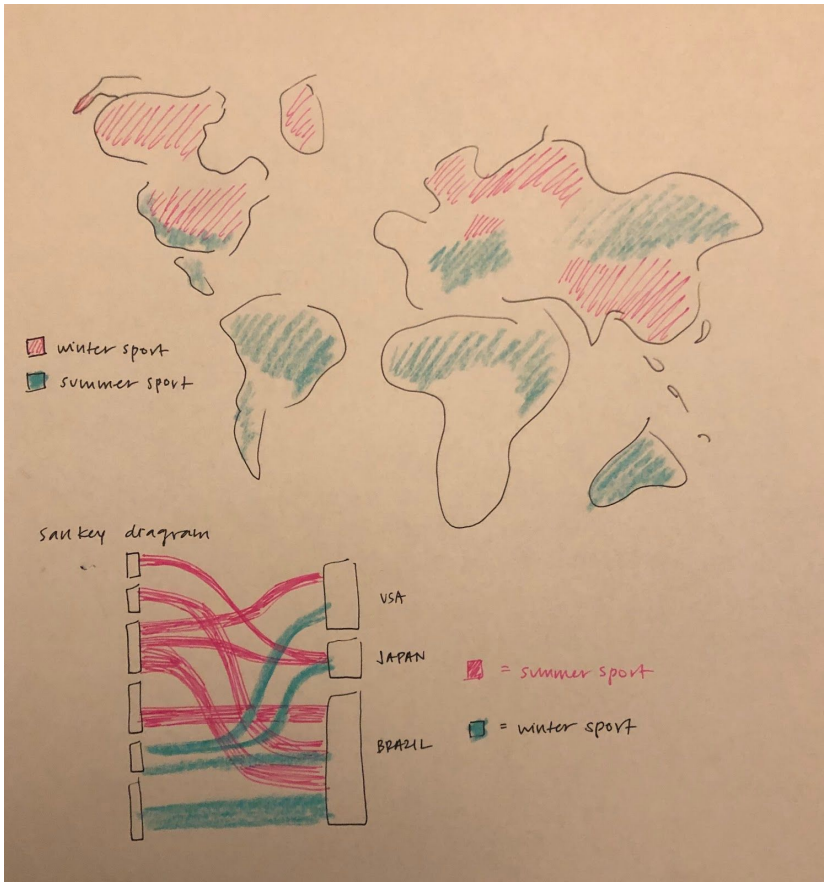
1a & 1b. Number of medals by country GDP



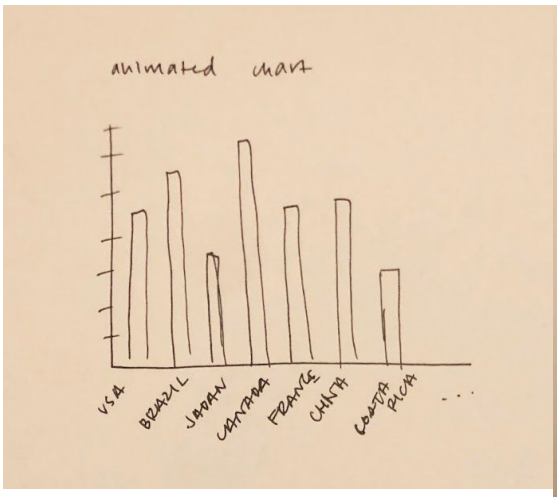
2a & 2b. Number of medals by country

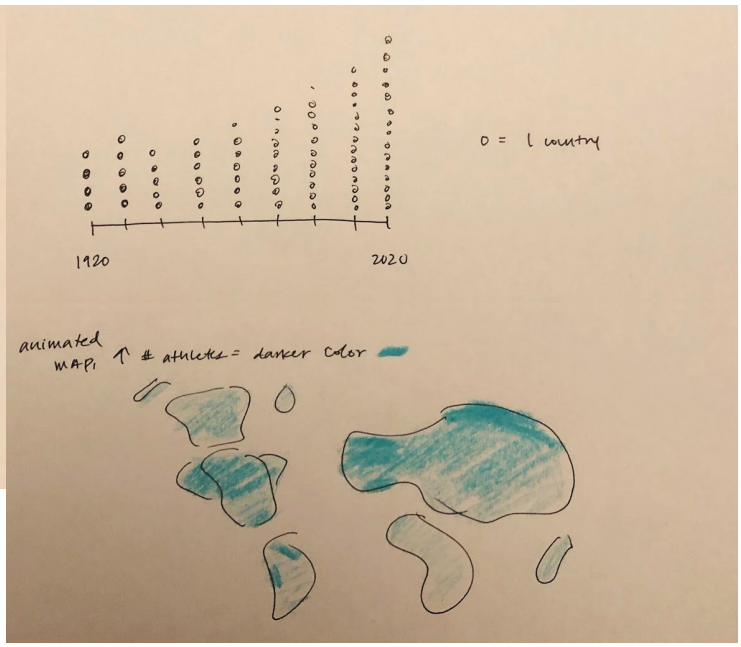


3a & 3b. Number of medals by country geography by sport

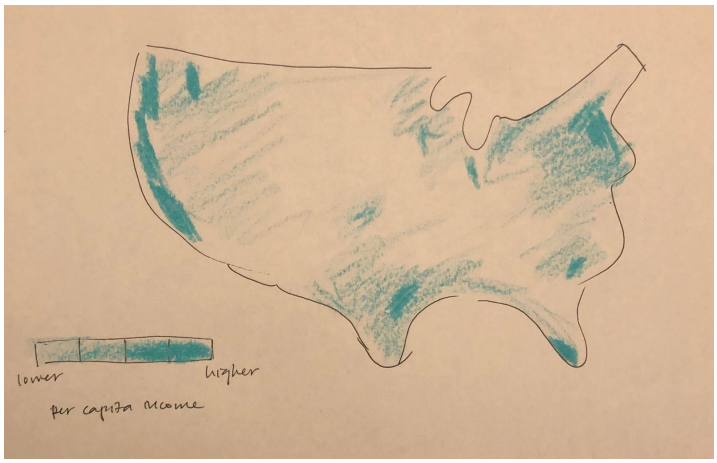


4a & 4b & 4c. Number of participants by geography by time



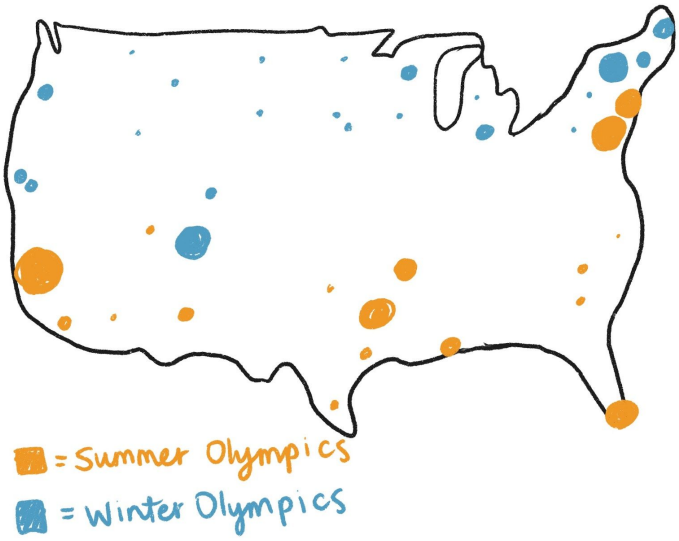


5. Athletes hometowns by average income

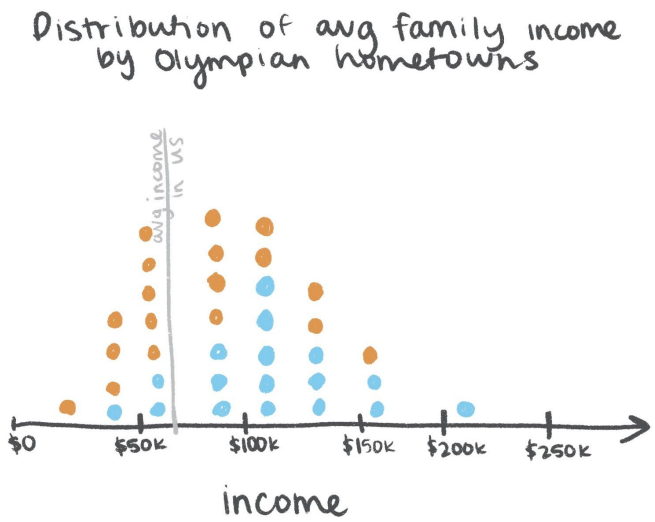


Marie:

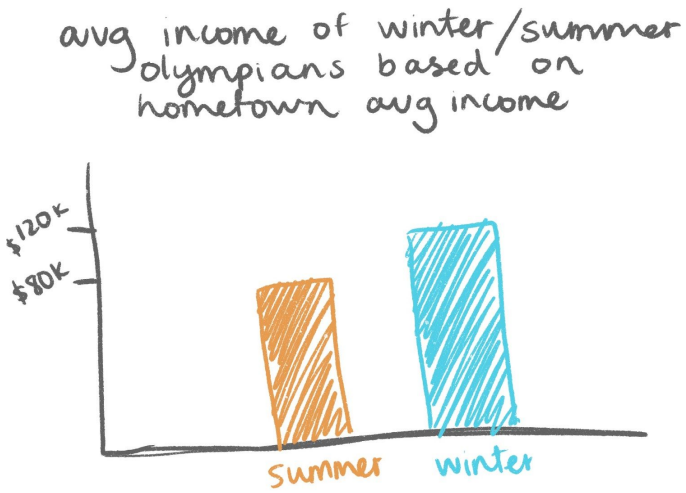
6. Map of athletes hometowns



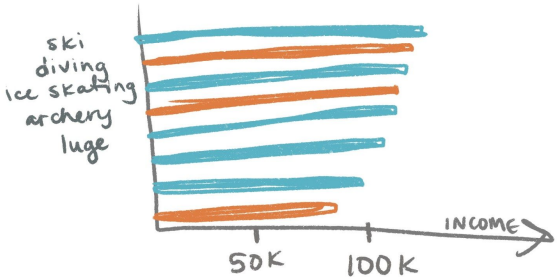
7. Athletes hometowns by average income



8. How Winter/Summer Olympics incomes compare



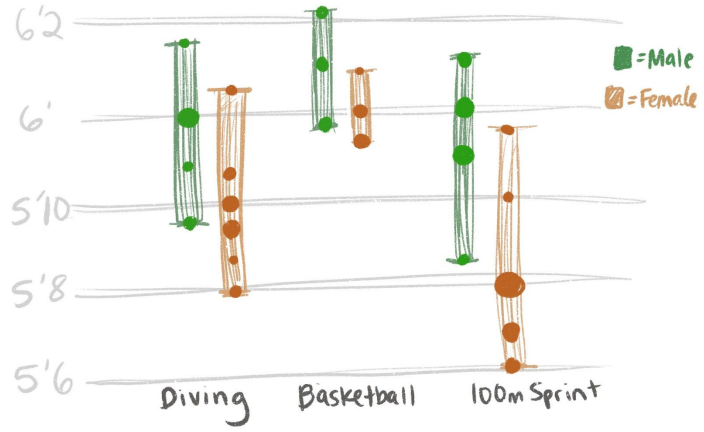
9. Predicted family income by sport



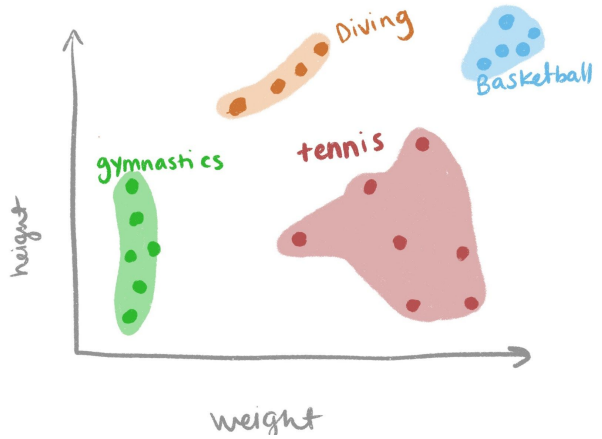
10. Map of athletes hometowns by wins in different sports

(What sports are most and least geographically diverse?)

16. Sport by height

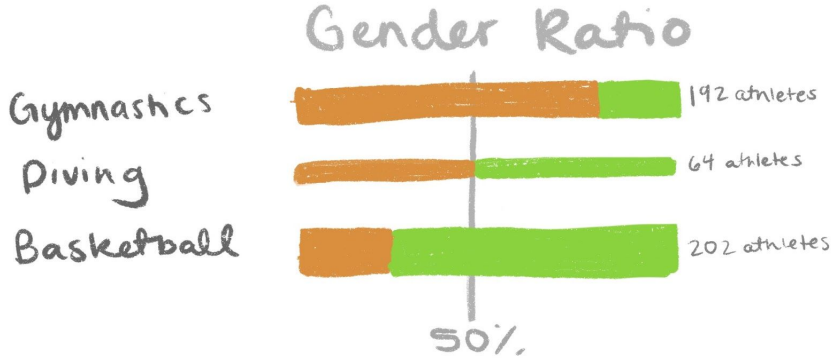


17. Sport by weight/height (toggle to divide by gender)



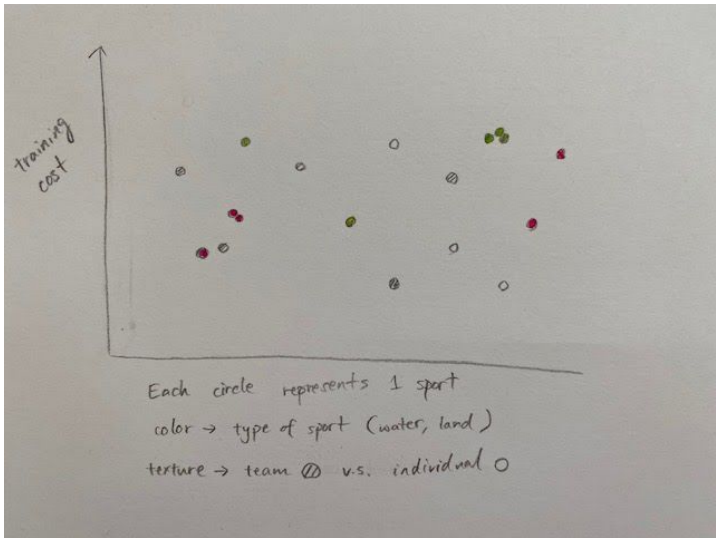
(Sports with biggest “area” have biggest diversity in body sizes)

19. Gender ratio of sports

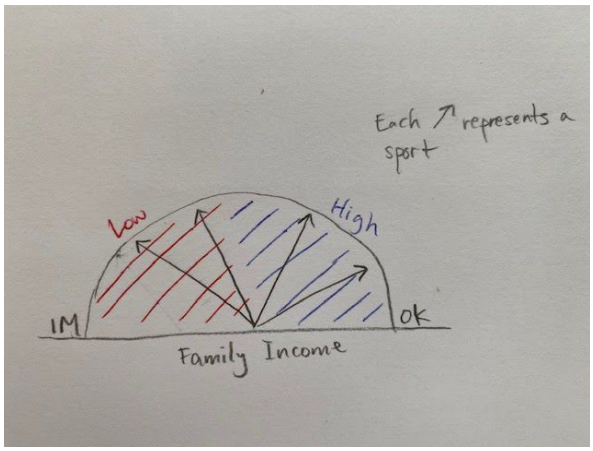


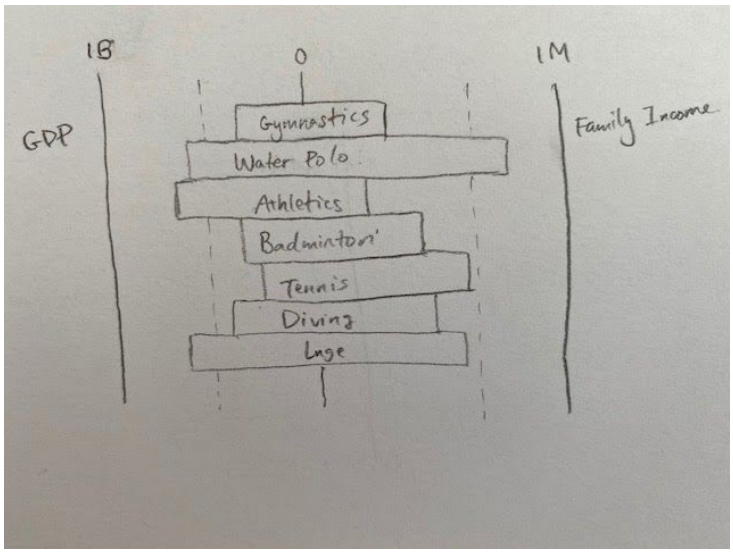
Miu:

11. Sports by training cost

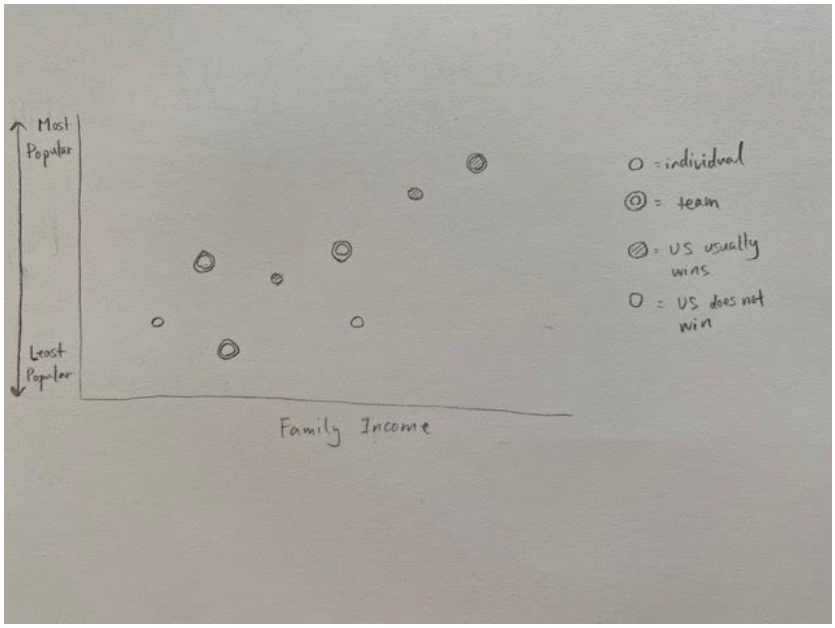


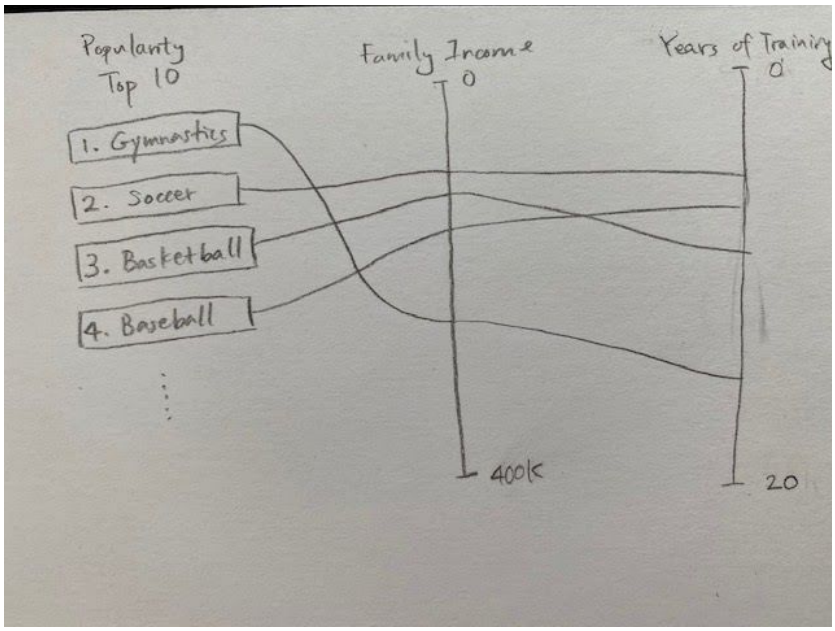
12a & 12b. Sports with high accessibility - average family income, country GDP



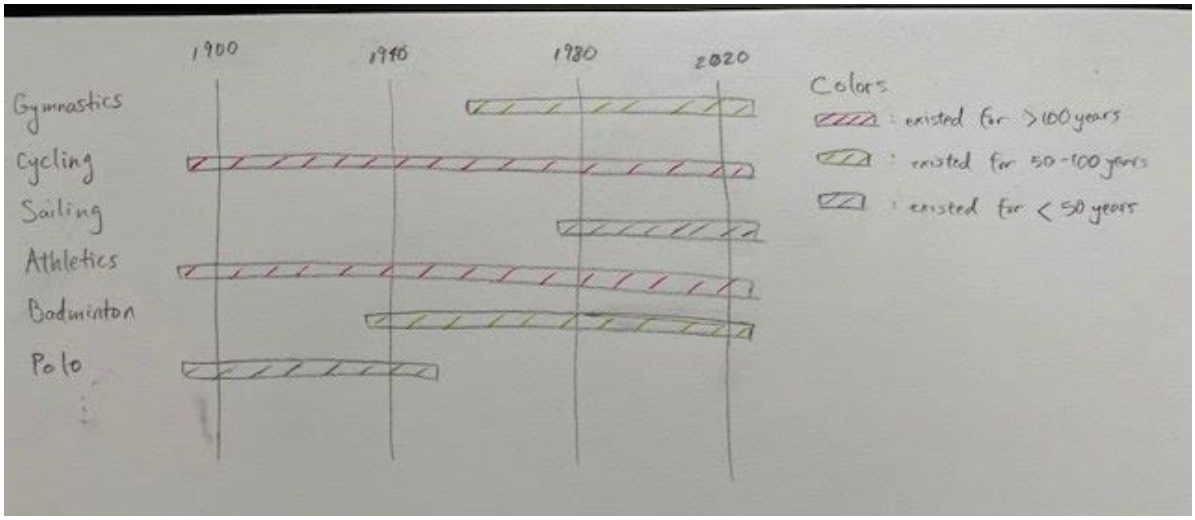


(GDP represents GDP of country that has the most medals in the relevant sport)
13a & 13b. Accessibility (based on factors we see above) of sport and its viewership (i.e. “popularity”)

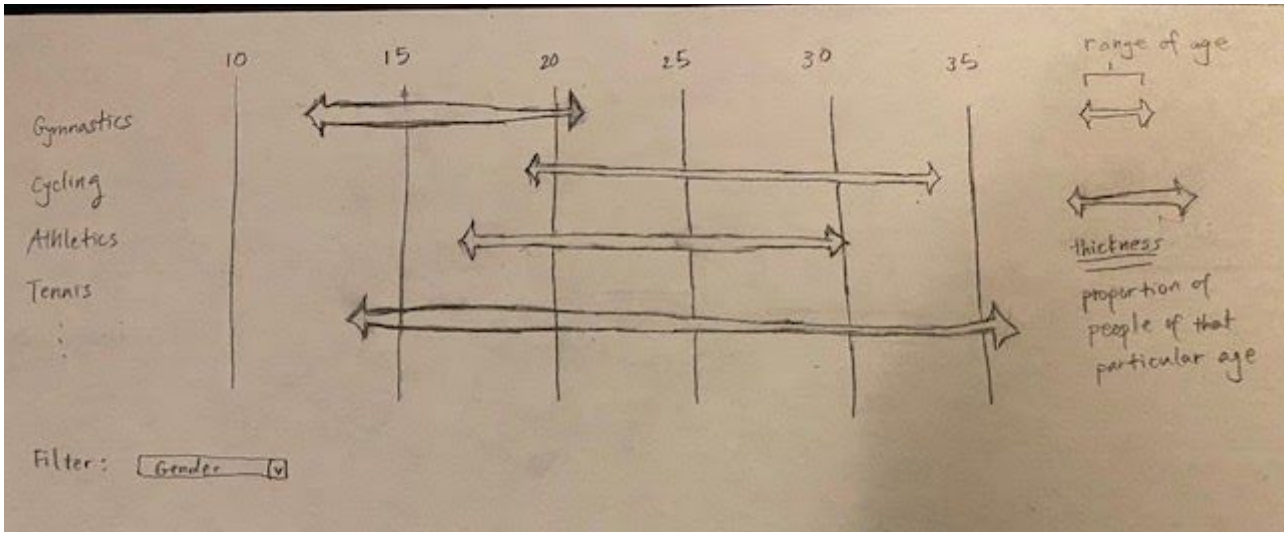




14. Number of events by time



18. Age by sport



<u>Datasets to use:</u>

Hometowns of athletes

- [2018 Winter Olympics Olympian Hometown + College]
- [2016 Summer Olympics Olympian Hometown + College]

Family income of towns in USA

- Link to formatted Google Sheet here
  - from US Census data 2009-2013
  - account for inflation
- **** More current income data to be used 2019 Median Household Income:
  - https://www2.census.gov/programs-surveys/acs/summary_file/2019/data/1_year_geographic_comparison_tables/GCT1901.csv

Viewership/Popularity

- Popularity based on survey about “best” olympic sport

- 120 years of Olympic history athletes and results: identifies athlete data from Athens 1896 - Rio 2016 including: name, sex, age, weight, height, team, season, city, sport, event, medal
- Olympic history data: calculated dataset focused on trends regarding the art competitions, women in the Olympics, geographic representation, and height and weight of athletes.

Sketches: Decision + Storyboard

Our visualizations will center around our main question: Which sports are the most accessible? When determining a sport's diversity/accessibility, we are taking multiple factors into consideration:

- Diversity of athletes
  - Height/weight (17)
    - represent sport distribution of athlete height/weight with points indicating a particular athlete and color indicating sport.
  - Gender (19)
    - Gender ratios of each sport with thickness corresponding to total number of athletes in that sport
  - Age (18) - age distribution of athletes with ribbon thickness corresponding to proportion of athletes
  - Hometown/predicted family income (5,7)
    - Combine 5 and 7 as linked graphs. Map of US color-coded by predicted income and bar chart showing income/medal distribution
- Diversity of countries winning
  - Geography (3a, 3b)
    - We will combine 3a(map) 3b(sankey diagram) as linked graphs. Each shows different granularity of information and are complementary to each other.
  - GDP (1a, 2a)
    - 1a: A line graph more aptly represents the relationship between our two quantitative variables: GDP and medal count
    - 2a: We will indicate winning countries by medal count and by type of medal won (color coded as gold, silver, bronze)
- Sport costs (11) - dot plot encoded with information about sport characteristics (land/water, team/individual)
- Popularity (12)
  - We'll compare sports' accessibility rankings to its popularity to identify any trends or interesting findings. We will use a scale similar to graph 12 to represent the accessibility ranking of a given sport.

**Instead of a group vote, we had a group discussion about the strengths of each sketch, ones that overlapped, and ones that we could link together. We also discussed what Gestalt and CRAP principles we could apply to our visuals by thinking through

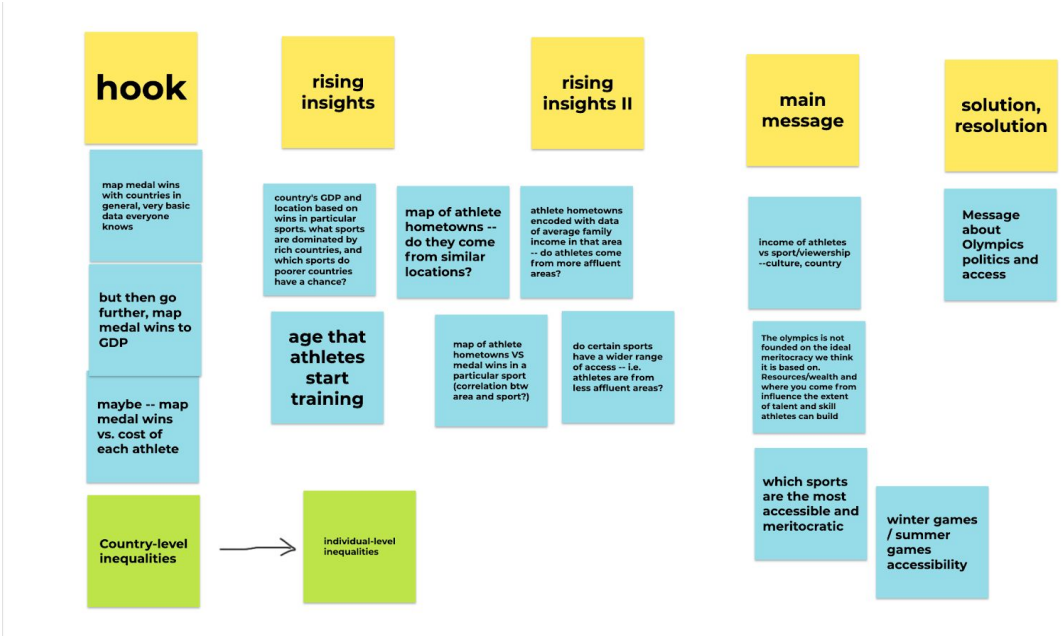
which elements would best represent the different layers of information we want to display in each one.

Big Picture

While the majority of our visualizations will include brushing and linked graphs, we also want to create an interactive element for the user at the beginning of our story. We want the user to input their own hometown (potentially family income?), and over the course of our visualizations, demonstrate how the user compares. This also serves as a grounding for the user, making the data more personal and reflective.

Initial Storyboard

Link to Jamboard



---------- TF feedback @ 11-12 ----------

Very nice job! I quite like your direction so far- pretty clear topic with your own perspective. The sketches are also nice with some creative / advanced charts.

I have a few questions after reading what you had for dataset and storyline-

1. The data you listed only contain one year of data for US athletes. Are you only focusing on US athletes for household income mapping?
2. How many years of data are you using to make analysis on and draw conclusions? Are they consistent with the GDP data you'll be using?
3. Average household income might not be representative of the resources these olympic athletes received, since the government fund and support, coaching & training resources, or sponsorship from industries are playing huge roles in the success of an athlete. However, it might still be interesting to visualize the data you have, only that you need the same level of granularity on income data from all countries.

Generally I think your questions and storyline are clear. Two things for attention are data consistency (since you have different data sources) and messages you can conclude from your analysis. You are covering a wide range of analysis so make sure they are structured and connected well in your story and visualization design.

Week 11: Prototype I

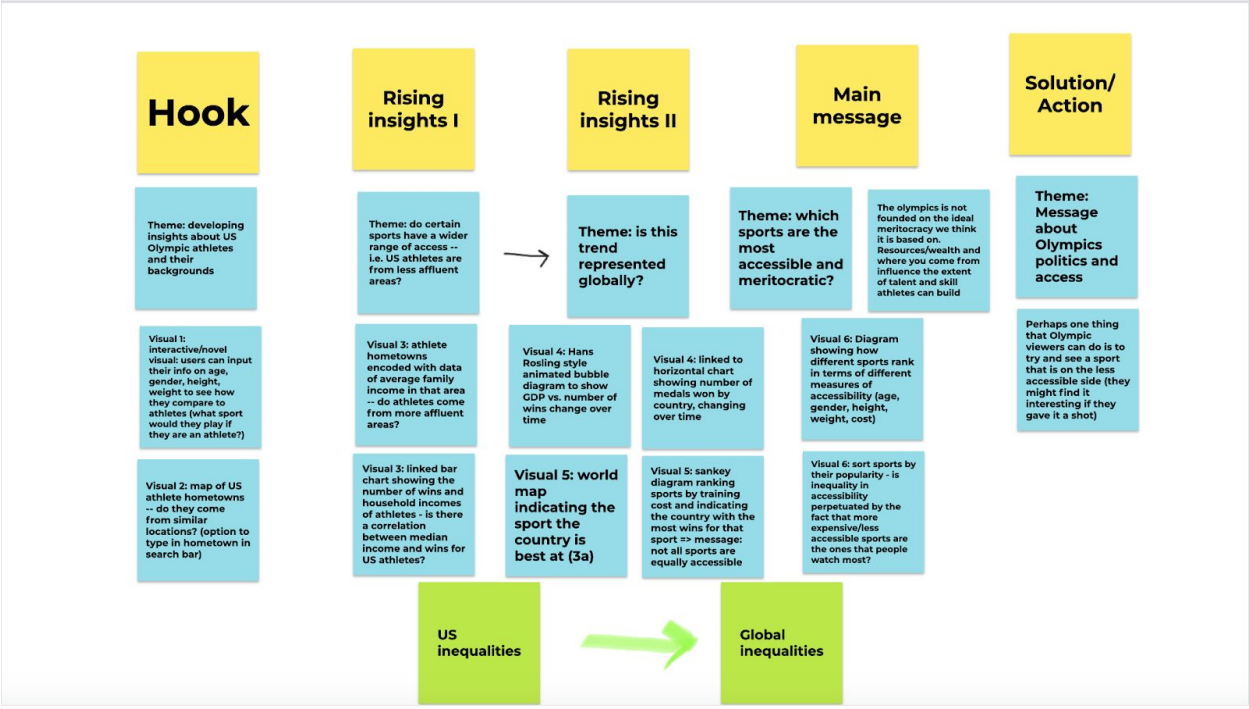
Requirements

- Data scraping and cleaning complete (using the real data sets)
  - Marie - popularity (538 survey, 2016 What is the best sport? - US data)
- At least one D3 visualization already partly implemented, and drafts for 2 more visualizations
  - Marie (weight/height by athlete/sport)
- Rough webpage design and structure has to be done and implemented (placeholders for visualizations, text, and images allowed)
  - Nicole create github repo and begin layout of slide scroll
  - Nicole create some temp filler tableau graphs
- Storytelling is clear
- The first design of an innovative view
  - Miu
- Interactions (e.g., filtering, brushing, etc.) have to be designed (at least in a textual description and some sketches)
  - Miu
- Up to date process book

Response to TF feedback:

- We will limit focus to years: 2018 Winter, 2016 Summer
- Story arc: focus on US -> correlation between GDP and wins -> focus globally

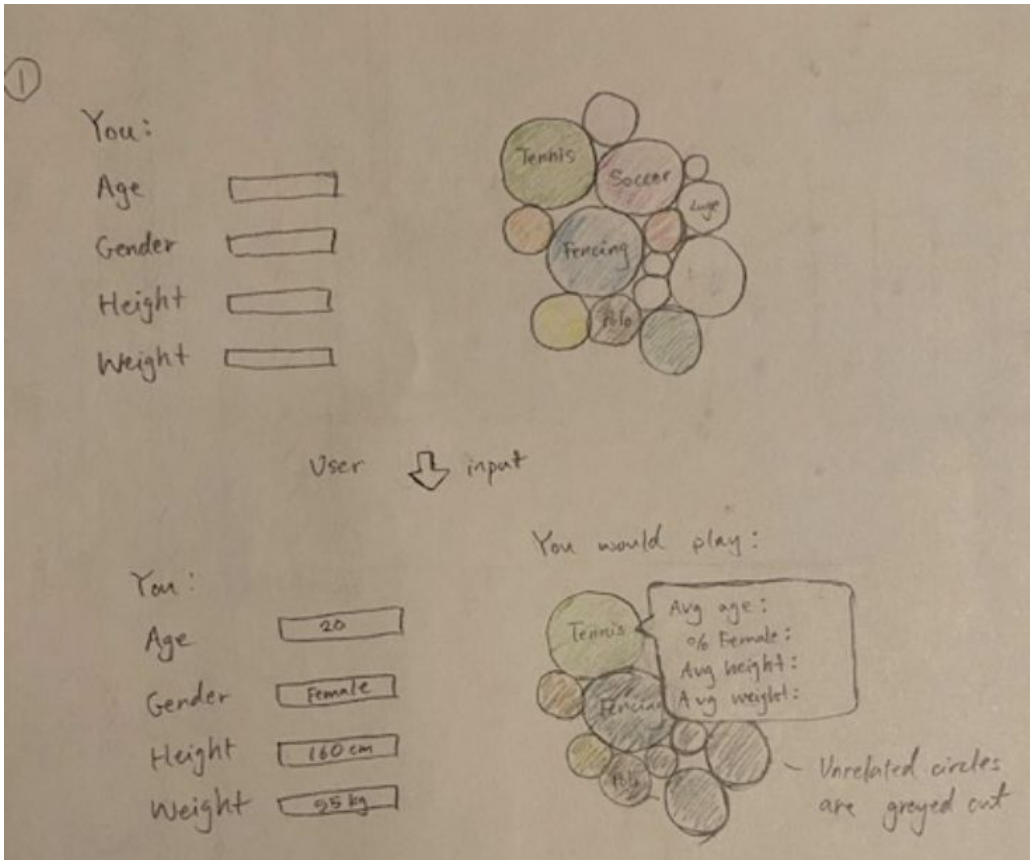
Revised Storyboard: story

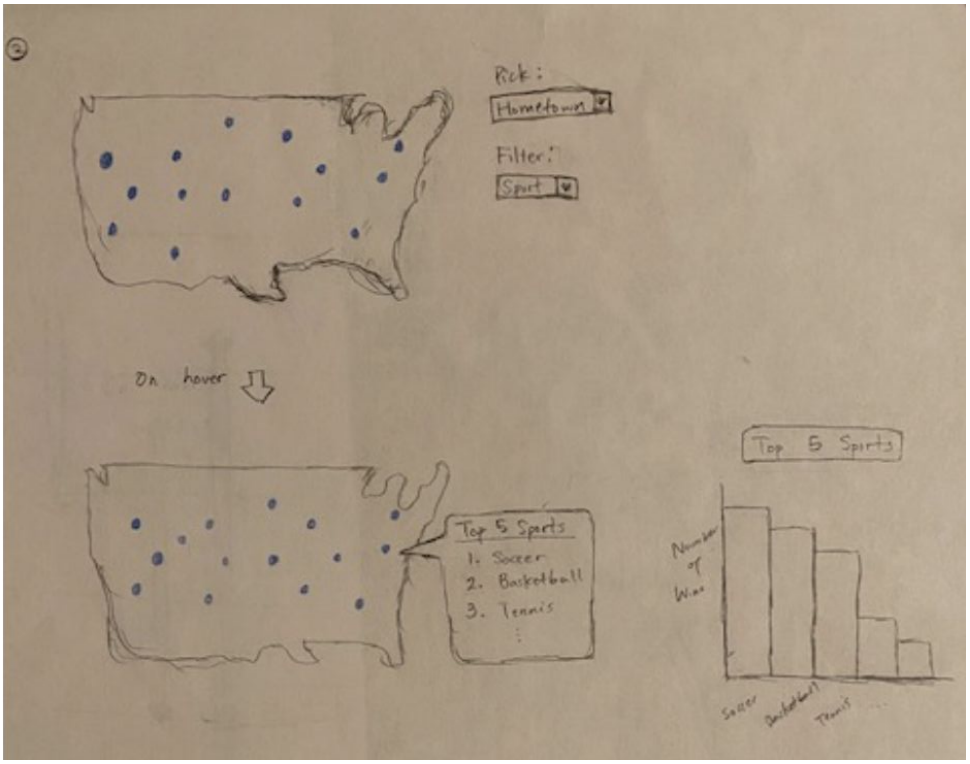


<u>Our Visualizations</u>

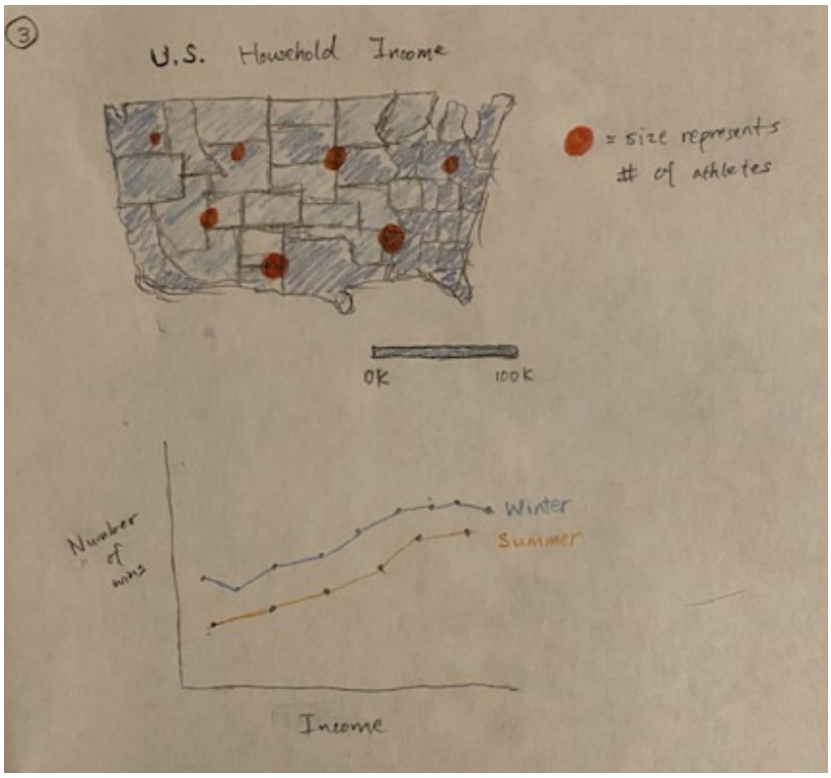
1. Bio Information: users enter age, sex, height, weight to find out what sport they would play if they were olympic athletes
2. Athlete US hometown: map of US athlete hometowns, map out where athletes come from. Search bar to filter by hometown.
    a. On hover display top 5 sports from that region
        i. Link to bar chart with top 5 sports and number of athletes
    b. Filter by sport
3. US Location Vs Median income Vs wins
    a. Choropleth income map overlaid with circles indicating number of athletes from a region
    b. Bar chart showing relationship between income and win
4. **Novel visualization** Country GDP and wins correlation (replace 1a and 2a)
    a. Dot plot: GDP versus wins, size of dots will show number of athletes sent
        i. This will be a video of change over time
    b. Link to horizontal stacked bar chart showing country medal wins
5. Sport vs world map
    a. global map of sports (wins) by country
    b. Sport costs -- which sports are the most expensive
6. Which sports are the most accessible and meritocratic
7. Sport popularity
    a. how does popularity correlate to accessibility

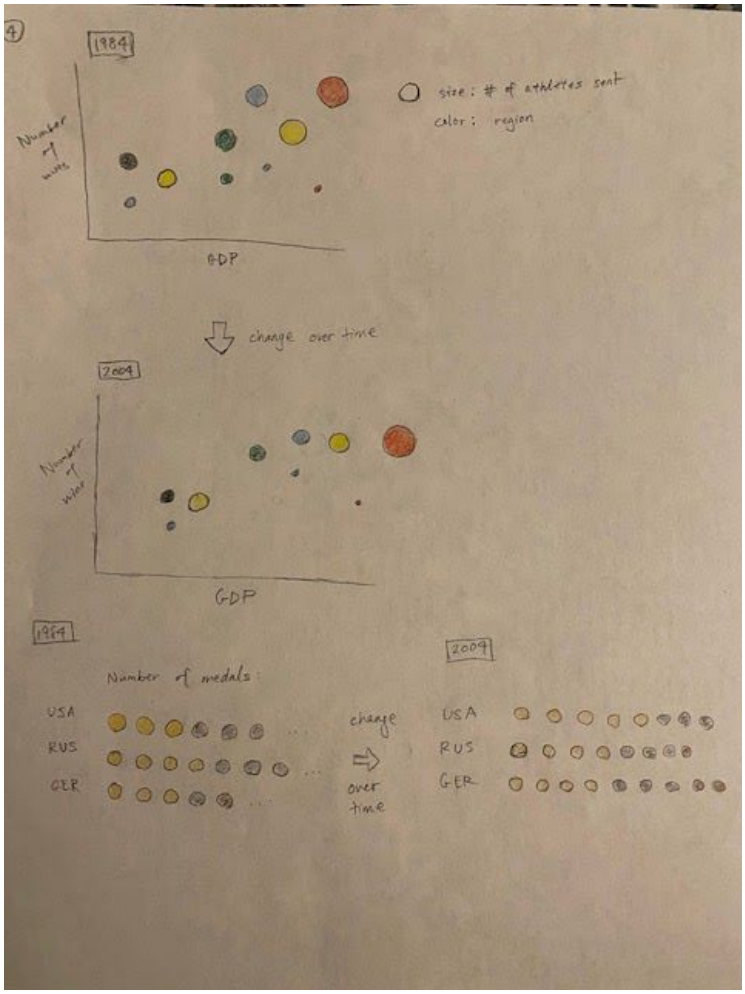
Interactions and Sketches



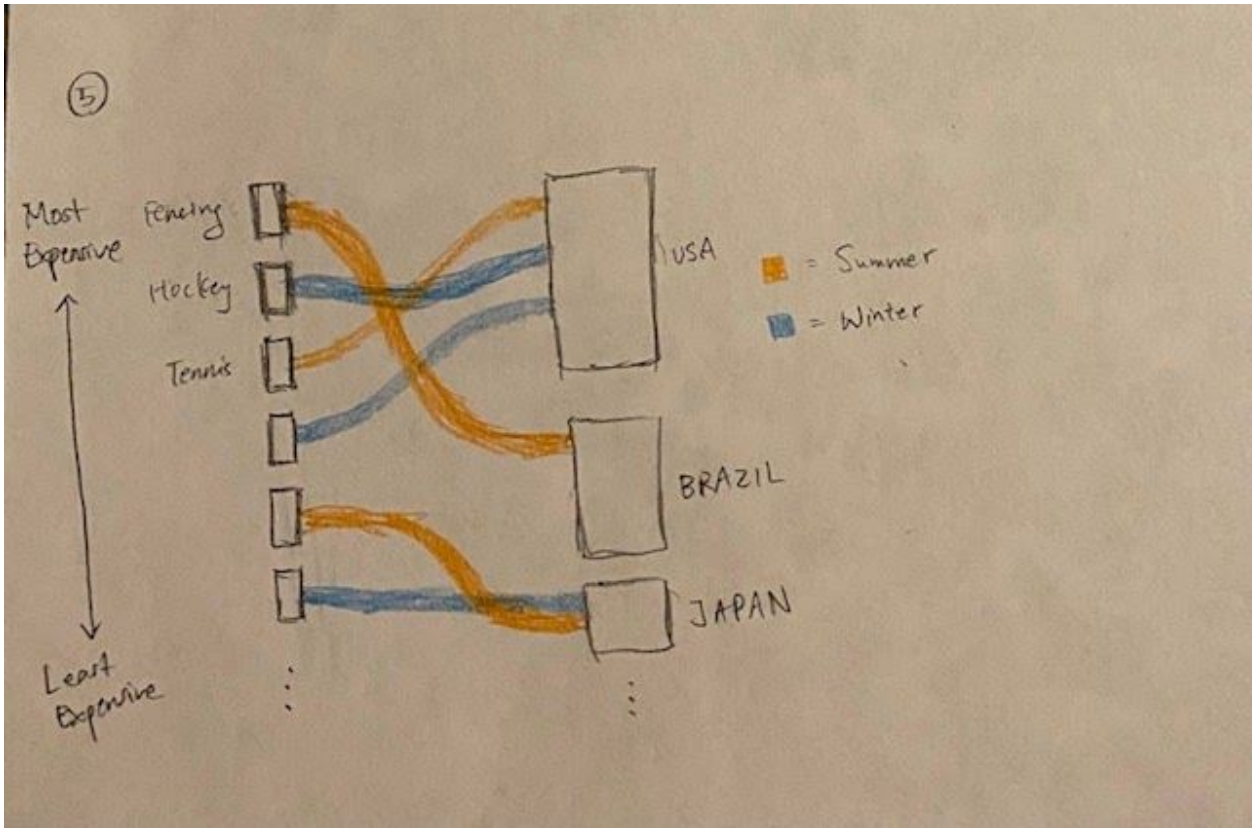


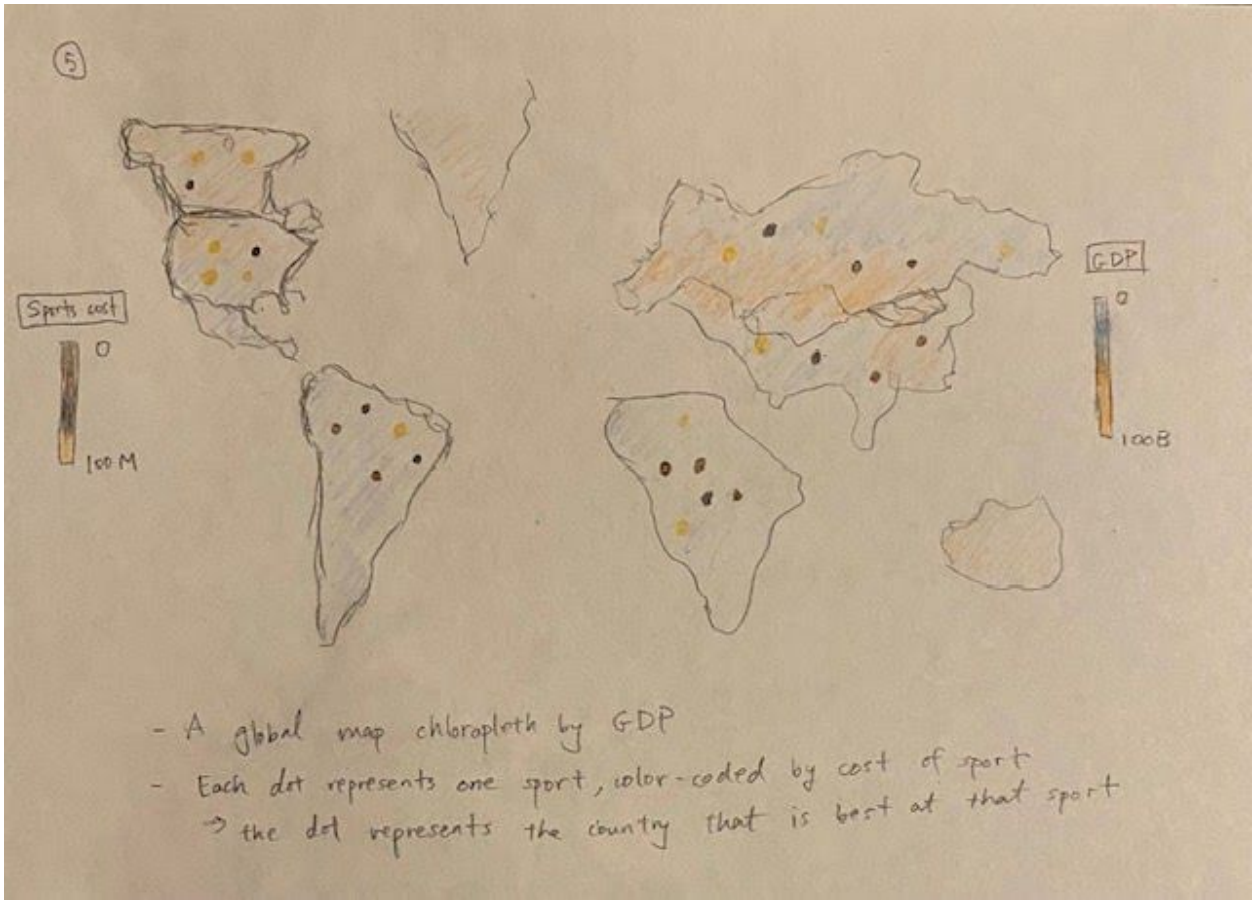
The bar chart is linked to the US map. ON hover, the tooltip and the bar chart reflect the info of the location is selected.

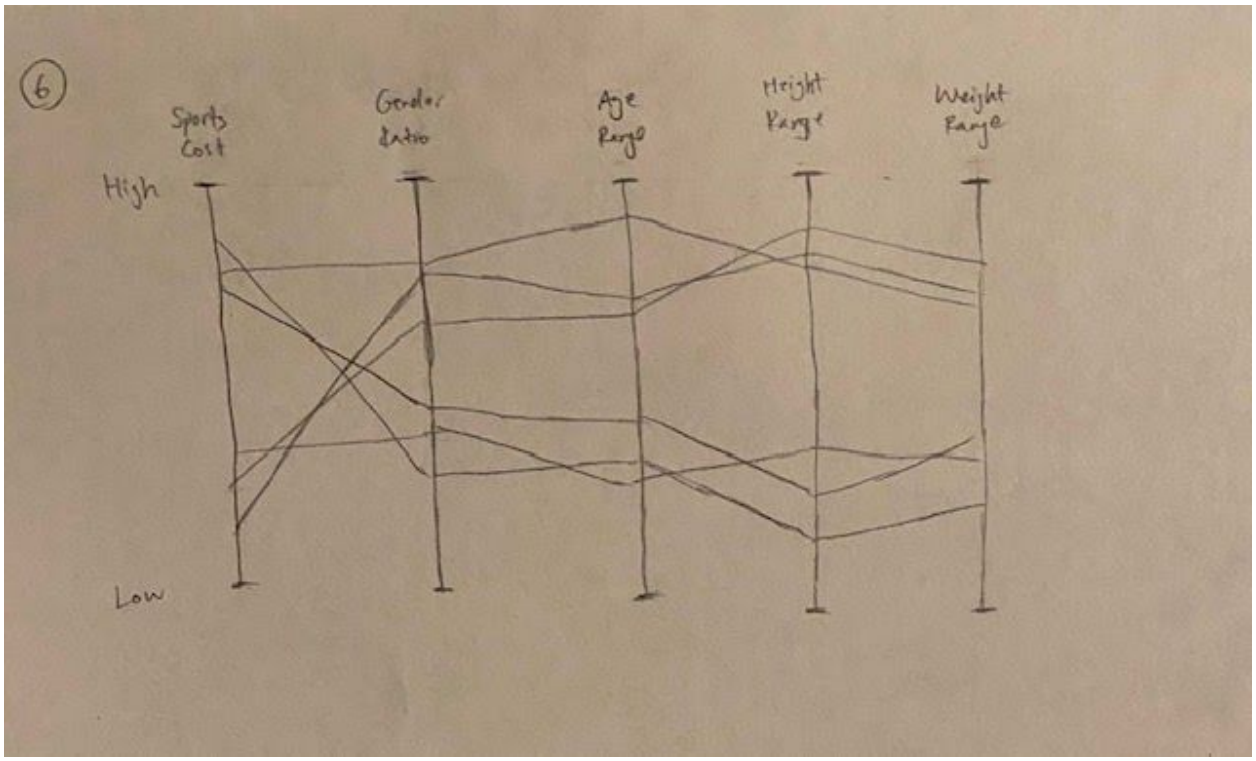


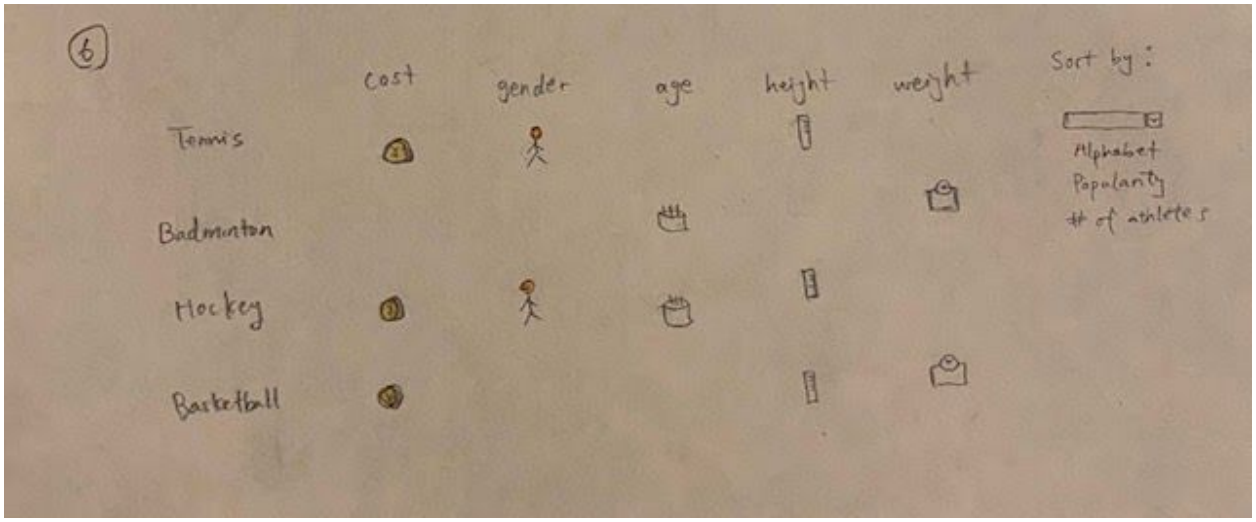


Animated visualization in which the bubble diagram and horizontal bar chart change with time. This is our innovative view.


5
Most Expensive
Fencing
Hockey
Tennis
Least Expensive
USA
BRAZIL
JAPAN
= Summer
= Winter


5
Sports cost
0
100 M
GDP
0
100B
- A global map chloropleth by GDP
- Each dot represents one sport, color-coded by cost of sport
→ the dot represents the country that is best at that sport


6
Sports Cost
Gender Ratio
Age Range
Height Range
Weight Range
High
Low



---------- TF feedback @ 11-19 ----------
Nice job for the prototype I overall! I think you have made the storyline very clear with the questions and sketches nicely laid out on the web page. Also your current storyboard is more focused and works pretty well from a reader's perspective- starting with personal data, stretching to cover the national and international topics. At this stage, I think if you can implement what you designed, it will be a really nice data story with appealing visualization design.

One suggestion on the current visualization #3, you present the number of wins vs. income for olympics in a line chart. I think the data encoding is unclear- are you comparing both the number of wins and number of olympians based on income level? I am not sure line chart is a good choice because the income falls in a range and you are basically grouping data into bins. Two things to consider, 1. How do you bin the income data range? 2. Number of wins might be highly skewed by certain athletes. In that case, maybe you need to highlight the wins by athletes.

The other suggestion is you might want to play around with the color palette of the physical distribution scatter plot you have implemented. The dark outline makes dots with darker fill difficult to see.

Week 12: Prototype II

Draft text for the website

slide1: Intro

The Olympic Games are one of the most viewed sporting events around the world. More than 200 states participate across 35 sports in more than 400 events. In the 2016 Rio Summer Olympics 3.6 Billion people tuned in to the 2016 Rio Summer Olympics while 1.9 billion watched the 2018 PyeongChang Winter Games. That means over half of the world watches the Summer Olympics, and this is without considering the increasing number of people who watch on streaming services instead of TV! But even if the Olympics are easily accessible to interested viewers, how accessible are the Games to athletes? More specifically, what does it take to be an Olympic athlete? Let's take a look.

https://www.statista.com/statistics/287966/olympic-games-tv-viewership-worldwide/
https://www.statista.com/statistics/531768/global-audience-of-the-winter-olympic-games/

slide2: What sport would you play?
First, let's see how you compare to the average Olympic athletes of 31 sports.

Include in note on bottom how this was calculated.
We are able to guess which sport you'd play based on the fact that each sport advantages a different physique, gender, or age. We can see how certain sports are more inclusive and diverse than others in terms of these physical attributes.

slide3: Physical distribution
No text? [comment on trends for gender, age]

For example, volleyball favors taller athletes and gymnastics shorter (unsurprising). Water polo players tend to be taller than hockey players (surprising?). Meanwhile, there are no visible trends for athletes who compete in athletics, which refers to track and field events such as high jumps and relays.

Slide4: US Olympic hometown
We can also think about inclusivity and access in terms of geographic representation. Let's first take a look at the U.S. Where do U.S. Olympic athletes come from?

=> The colors on the map show the median household incomes of US counties and each red dot indicates the hometown of a US Olympic athlete. We see that, on the whole, athletes are more likely to come from areas with higher median household incomes. This is interesting…might income be an indicator for one's potential to become an olympic athlete?

Slide5: Income and hometown
Athletes are more likely to come from households with higher income levels than the median household income. Perhaps participating in the Olympics is not just determined by natural talent, but by the ability to access resources that hone, train, and build that talent. Is this trend also represented globally?

Slide6: Country GDP
Yes it does! Countries with higher GDP have won more olympic medals over the years.

Slide7: Sports cost

Countries with higher GDP can also afford to host the games. Hosting the games gives countries the home field advantage and increases chances of winning medals.

Slide8: Conclusion
So what does this all say about the accessibility of the Olympic Games? Olympic sports are only biased in terms of the physical traits they favor, they are also biased in terms of the resources that athletes and countries have access to. Being of a particular physique is an advantage but so is being of a particular income background. The Olympic Games might not be the truly meritocratic competition we think it is.

As Olympics viewers we might be surprised by this - all we see on the screen is the test of athletes' talents and not much commentary on background. This might be because the sports we tend to watch are the ones that are more costly and inaccessible.

[introduce action point] Tackling inequality at this level requires a global effort, but one place we can all start is to watch a post we never have before…Judo? archery? Trampolining? Take your pick and start next year with the Tokyo 2021 Summer Olympics!

Prototype II Website
Slide1

Slide2

Introduction

The Olympic Games are one of the most viewed sporting events around the world. More than 200 states participate across 35 sports in more than 400 events. In the 2016 Rio Summer Olympics 3.6 Billion people tuned in to the 2016 Rio Summer Olympics while 1.9 billion watched the 2018 PyeongChang Winter Games. That means over half of the world watches the Summer Olympics, and this is without considering the increasing number of people who watch on streaming services instead of TV! But even if the Olympics are easily accessible to interested viewers, how accessible are the Games to athletes? More specifically, what does it take to be an Olympic athlete? Let's take a look.

Sources:[1] Viewership summer [2] Viewership winter.

Slide3

What Olympic sport would you play based on your physical traits?

First, let's see how you compare to the average Olympic athletes of 31 sports.

Your traits:

Age e.g. 20

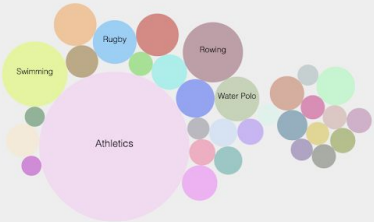
Gender Select

Height (in) e.g. 70

Weight (lb) e.g. 170

Find

Olympic athletes' traits:



We are able to guess which sport you'd play based on the fact that each sport advantages a different physique, gender, or age. We can see how certain sports are more inclusive and diverse than others in terms of these physical attributes.

Slide4

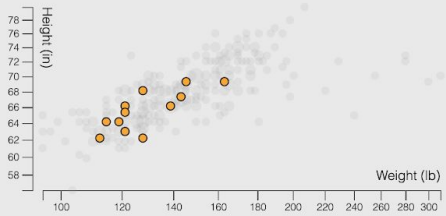
Physical Distribution of Rio 2016 Athletes

Toggle by Sex: Female

Male Female

Sport	Avg Weight	Avg Height	Avg Age
Cycling	129.7	65.6	28.2
Swimming	154.0	70.2	22.5
Volleyball	165.0	73.1	26.9
Trampolining	127.9	64.2	18.0
Athletics	145.3	67.3	26.0
Fencing	157.6	68.1	27.1
Synchronized Swimming	119.0	65.9	22.5
Weightlifting	193.3	65.2	26.7

Athlete Weight & Height by Sport



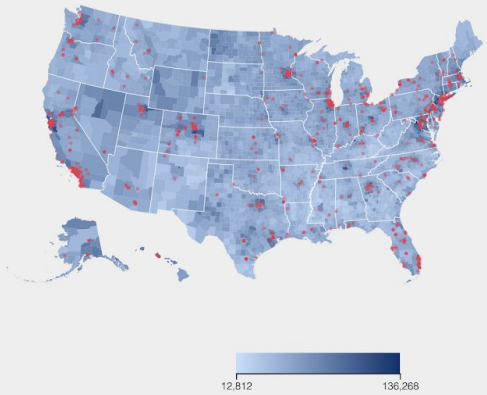
Age Distribution of Athletes

Gender Ratio for Sport

Slide5

US Olympic Athlete Hometowns

How many athletes are from your home state?



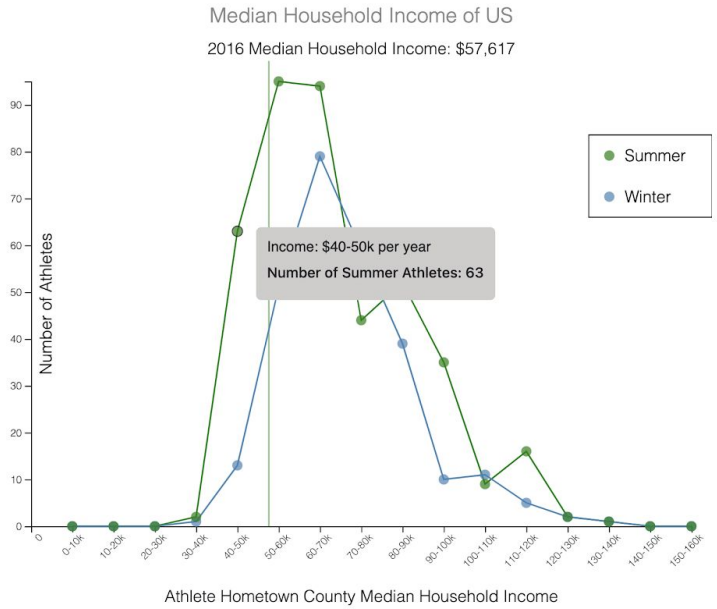
We can also think about inclusivity and access in terms of geographic representation. Let's first take a look at the U.S. Where do U.S. Olympic athletes come from?

The colors on the map show the median household incomes of US counties and each red dot indicates the hometown of a US Olympic athlete. We see that, on the whole, **athletes are more likely to come from areas with higher median household incomes.** This is interesting…might income be an indicator for one's potential to become an olympic athlete?

Slide6

How does household income of a hometown affect number of athletes sent to the Olympics?

Athletes are more likely to come from households with higher income levels than the median household income. Perhaps participating in the Olympics is not just determined by natural talent, but by the ability to access resources that hone, train, and build that talent. Is this trend also represented globally?



Slide7

Country Wins VS GDP

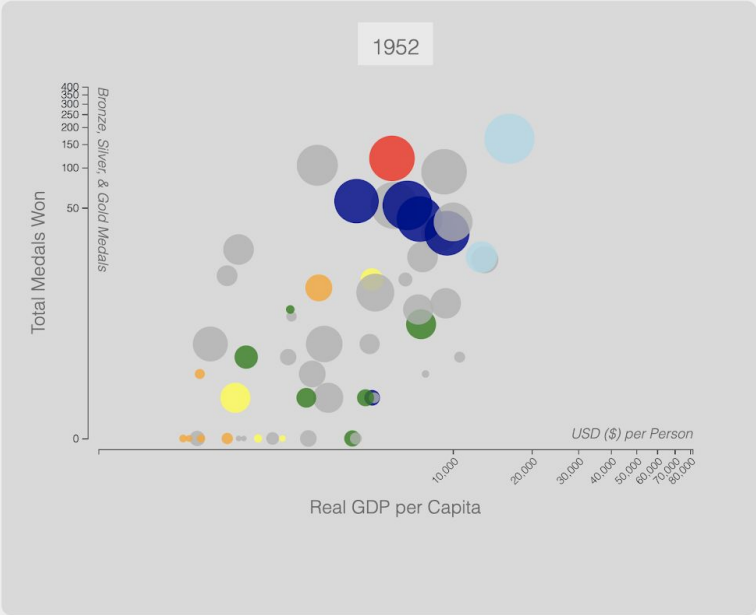
How does a country's wealth affect the probability of winning medals?

This does seem to be a wider trend. **Countries with higher GDP have won more olympic medals over the years.**

To Do:

Add key: Orange for Eastern Asian countries, red for Russia/North Korea, Dark blue for Europe, Light Blue for North America, Yellow for African countries, and Green for Latin American countries

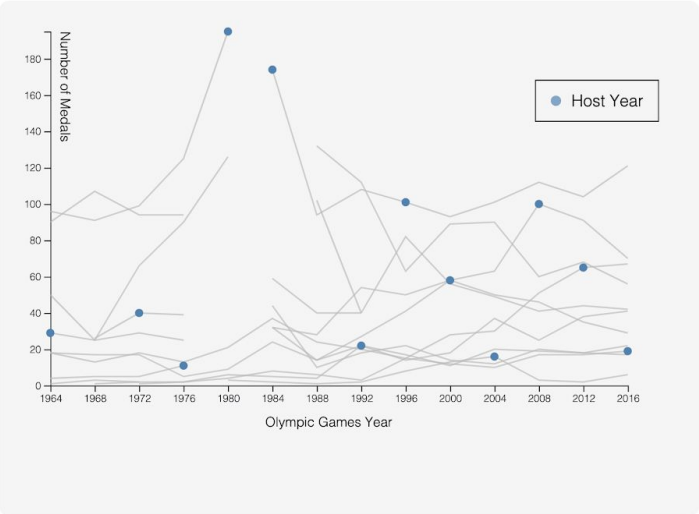
1952



Slide8

How does hosting the Olympics affect a country's chances of winning medals?

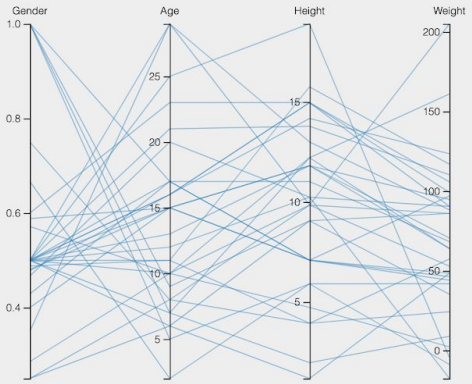
Medal Wins by Host Countries



Slide9

Which sports are the most truly accessible and meritocratic?

Trends of High and Low Accessibility Features



Sport Report Card

SILVER

Sport	Volleyball
Gender Ratio	50% Female
Age Range	22 - 38
Height Range	66.9 - 81.9 in
Weight Range	145.5 - 235.9 lb

Sports are assigned a GOLD, SILVER, BRONZE ranking based on their level of accessibility, which we calculated based on level of female representation and the range of athletes' ages, heights, and weights.

Slide10

A meritocracy?

So what does this all say about the accessibility of the Olympic Games? Olympic sports are only biased in terms of the physical traits they favor, they are also biased in terms of the resources that athletes and countries have access to. Being of a particular physique is an advantage but so is being of a particular income background. The Olympic Games might not be the truly meritocratic competition we think it is.

By Nicole Araya, Marie Konopacki, Miu Kumakura

---------- TF feedback @ 11-29 ----------
Overall, nice story and visualization implementation!
The storyline is pretty clear and nicely presented into different sections. The visual style is very neat and clean, with nice use of white space and font contrast. In general, the data story works well, but I want to point out some details that can make your visualization better.

They are suggestions for you to think about and take on what you can improve on, and feel free to reach out if you have any questions.

My feedback for the design & implementation is below:
(1) Introduction- The intro is nicely written with a good amount of information but there is too much text on the first page. Consider adding bold text, color, or break to highlight certain sentences/numbers to make it easier to read.

(2) The first bubble chart visualization is catchy, but it is unclear what the bubble size represents (athletes number of US?). The purpose of this chart is to compare the physical traits among different sports. I suggest making the size of circles representing one of the traits to allow easier comparisons and try to make the text labels visible when possible. E.g. circle size represents the average age, height, or weight, per user selects. This is minor, but I also suggest you can replace the text fields with sliders, and add a reset button, so the user can explore different options easily (plus the height in inch is not as common as ft+in expression). The footnote is better with left alignment based on CRAP.

(3) Physical Distribution of Rio 2016 Athletes chart is very nicely implemented! The only comment is to make the title clear that it's only among US athletes, same as (2).

(4) Hometown map-
The question title (How many athletes are from your home state?) doesn't represent the information and needs modification. The right paragraph is better with left alignment. The color legend needs explanation (household income of which year?). The interaction seems a little laggy- the tooltip is based on states but it only shows up when hovering at a certain county is a state or so, at least that's how it works for me. I think you use color contrast nicely on the map, but the red dots completely cover the income colors. Maybe add a toggle to show & hide the red dots.

(5) Household income line chart-
I think the chart means well and present the data clearly. However, the claim that more athletes come from high-income county is not quite obvious with the current chart. First, it is better if you show the vertical line of median income by default (you can adjust transparency so it's less intrusive). Second, maybe consider making y-axis as cumulative athletes' numbers so you can tell the proportion of athletes below and above the median income.

(6) GDP
This is a very nice chart! I think the animation adds extra insight towards the trend. However, make sure you highlight US or other specific countries you want to point out, so the user knows which dots to follow.

(7) Host countries
I think it is a very interesting chart and brings nice insights! You should explain the gap in the lines. Also the interaction on the thin lines is not easy and the country names are hidden- I would make the dots and lines interaction area thicker and add the country labels to the right of the line chart (besides the lines of 2016).

(8) parallel chart
Nice implementation. I would add a highlight to the hovered line, and color-coded the lines based on the ranking assignment.

(9) conclusion- you can add some reference links to support and extend your statement.

These are suggestions and I don't expect you to implement all of them. Overall, you need to improve on the first bubble chart, the general tooltip interaction on charts, and finalize the title and texts to support the storyline.
Feel free to incorporate them with user testing's feedback to update your final project submission.

Week 13: Feedback

Tester Name: Joshua Lee
Tester Email: joshua_lee@college.harvard.edu

General Observations from the think-aloud study:

He really enjoyed our clean design and thought we were in a good place.

What does the tester like about your data story?

He thought the topic was interesting and the progression made sense.

What improvements does the tester point out?

For the first slide, he didn't know he had to hover over the "You Play…" circle, since he thought it was a loading bubble. He also didn't know his height in inches, so it would be helpful to implement a feet/inches input instead of just inches. Can possibly do this via 2 drop downs (one for feet, one for inches)

For the US Olympic Athlete Hometowns slide, he wasn't sure if the choropleth map was income or population density. Make the key more clear

For the Hometown income line graph, make the text for the Median Income line text stand apart from the chart's title text-- kind of crowded.

For medals hosting, he was confused by the gaps in the lines but thought the data was super interesting. Was a little difficult to hover over the line since it's so thin

For the Accessibility scores vis, he really enjoyed the data but thought the scales were confusing. He didn't realize that Age, Height, and Weight was the range of those numbers (he thought it was the average or something). He also thought that it would make more sense for .50 to be at the top for the Gender score, since that would be the more accessible number, and 1 would be at the bottom since it's the least diverse/accessible value. Didn't know if he was hovering over a value at first though, since the report card didn't catch his eye immediately, and the lines had no indication they were being hovered on. Also include a tooltip on the hovered line just to state the sport. Also color the lines according to their score (gold, silver, bronze)

Was the intended key message clear to the tester? Why or why not?

Yes, he thought it was clear. Sometimes he thought we overstated it or was a little too wordy/heavy-handed at times

Did the tester get your next steps or call to action? Why or why not?

Yes, but he thought maybe we should attach resources to orgs that maybe address this type of sport inequality. But when I mentioned our planned ending which would talk about sport popularity, he thought that was a good idea too

Tester Name: Toby Satterthwaite
Tester Email: tsatterthwaite@college.harvard.edu

General Observations from the think-aloud study:

Make sure to clearly define to the user exactly what data is being represented (year, source, any data manipulations/calculations). Also, a lot of information is encoded into the tooltips which is not always obvious and the user might miss it if they are not looking.

What does the tester like about your data story?

He really liked that the story was understandable with very little text and for the most part the visualizations spoke for themselves with little need for context. The build up from US - wide economic factors to global economic factors in the olympics was clear and easy to follow.

What improvements does the tester point out?

Again, for some visualizations, he did not know that he needed to hover, so that functionality should be more clear. Add toggle for Income plot and host medals plot so that overlapping line charts are not overwhelming. Make trend lines thicker in host medals plot and final accessibility ranking so that important information immediately

jumps out to the viewer. For the host medals plot, add a vertical dashed line to indicate why there are line breaks (countries boycotting Olympics due to political reasons).

Was the intended key message clear to the tester? Why or why not?

Yes, the tester understood the overall message that the Olympics are not the pure meritocracy that we may think of it as. However, he was having a hard time differentiating specific sport winners and losers ranked by accessibility and diversity, and suggested a final slide visually showing the highest and lowest ranked sports in terms of diversity.

Did the tester get your next steps or call to action? Why or why not?

The call to action/next steps (to consider the Olympics more qualified light) were understandable but not super compelling from this prototype. For the next one we will include a visualization to support the message of linking viewership and accessibility, to clearly show the viewer their actionable next steps in viewing the next Olympics.

Tester Name: Emma Freeman
Tester Email: efreeman@college.harvard.edu

General Observations from the think-aloud study:

- Hook, message, and flow of the story is clear
- Would be useful to include a sentence encouraging users to interact with visualizations (i.e., for tooltips)
- Amount of text is good

What does the tester like about your data story?

- The intro and hook, especially that we contextualize the Olympics with data in the first slide and allow user to put in personal info
- The linkages between different graphs
- Text that reiterates the main point of each visual
- The fact that our story goes from the US to the rest of the world

What improvements does the tester point out?

- Finding user's sport:
    - include sentence that user should hover over circle

        - Enable user to type in different values and play around with vis
    - Physical attribute distribution:
        - Move legend about gender to the gender ratio map because it is confusing that you can toggle between genders but still have both mentioned in the legend next to it
        - Include units in the columns for weight and height
    - Household income map:
        - Hover is a bit sticky
        - Maybe add income info in the tooltip -> avg household income of each state?
        - Make map larger
    - Median household income:
        - The trend could be made clearer because the U-shape is a bit confusing
        - Include sentence about the trend we are seeing
        - Wasn't aware there was a hover function so pointing that out (or, allowing user to see the line at all times?)
    - Country wins vs. GDP:
        - Clarify what size of circle represents
    - Medal wins:
        - Make lines thicker
        - Include a sentence about what the main message/trend is
    - Accessibility:
        - Note how we arrived at thresholds for gold, silver, bronze
        - Add income or GDP info
    - Conclusion
        - Include data points in the paragraph
        - List 5 most accessible/least accessible or another summary of the data story

Was the intended key message clear to the tester? Why or why not?

- Key message was clear and reinforced by the text in intro/conclusion

Did the tester get your next steps or call to action? Why or why not?

- There was no clear call to action in the prototype, but we will include a sentence at the end about an actionable step that users can take based on the main message of our data story

<u>Test summary</u>

The specific edits we would add to our visualizations are listed in the notes above, but the main points of focus are:

- There are some interactions that are not intuitive so clarify them by text or by making functionality more explicit
- Need to clarify where data comes from (year, region represented), what elements such as size or color represent, the scales we use, and backend calculations we did in wrangling data (especially for how we calculated accessibility)

- Minimize and summarize text. Try an answer-first approach where main message is stated first or bolded to stand out.
- Think of a way to end our data story. We don't think a call to action is extremely necessary but perhaps tying the data back to how we started our story about how the general public engages with the Olympics. We can help the user glean insights about which sports are popular, which is something our targeted audience - general public - decides and thereby brings the story back to the user.

Team meeting notes

- Color palette
    - alternative (similar but slight differences - more blueish dark and brighter blue)
- Fonts
    - Cocomat
    - Sofia

---------- TF feedback @ 12-03 ----------
Nice record of user testing. It seems you got quite a few useful insights to improve on usability (interaction) and data clarity (annotation or explanation of data source). I like your complete record but it is required that you decided as a team what actions to take from user feedback in a separate list (-1). I am sure you have extracted user insights and improved your design but I just couldn't tell what your exact next steps are from the record itself.
Look forward to your final submission!

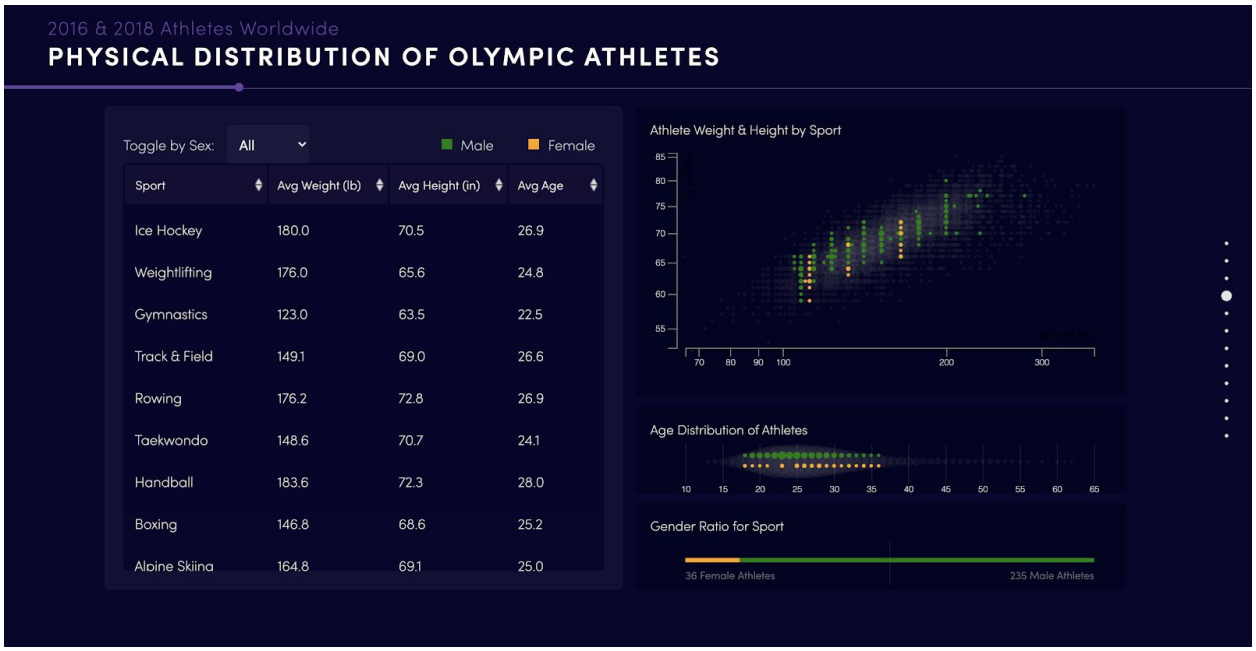
Week 14: Final Project Submission

Final Website
Slide1

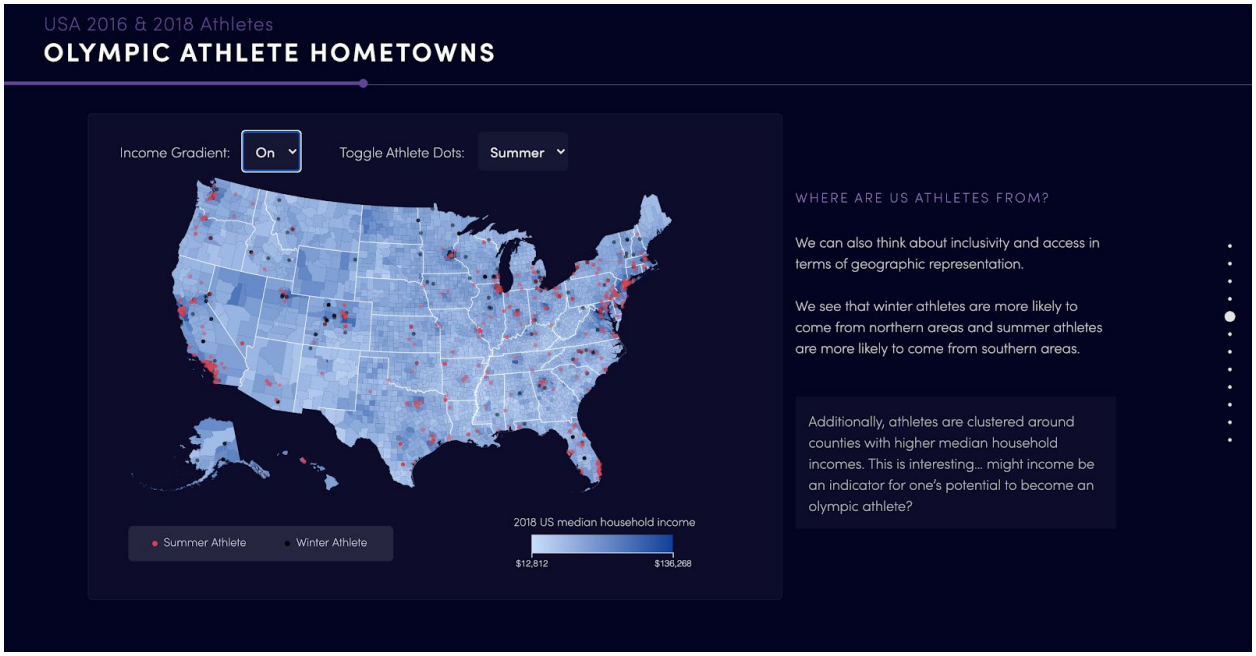
Slide2

Slide3

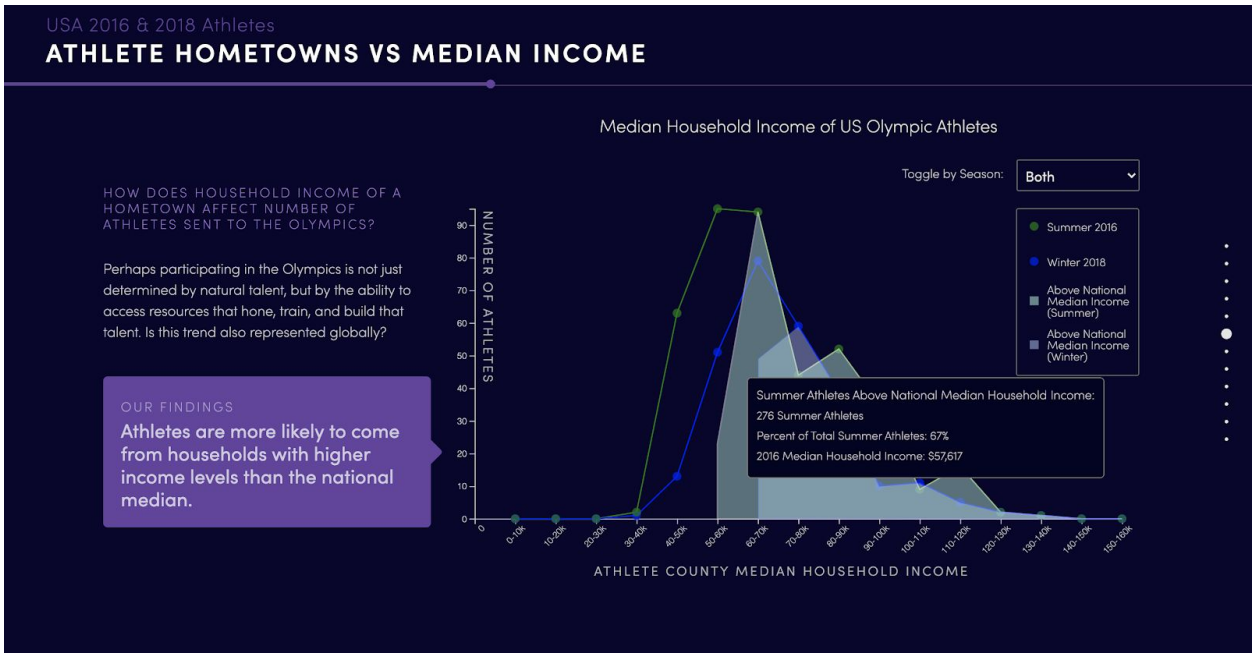
Slide4



Slide5



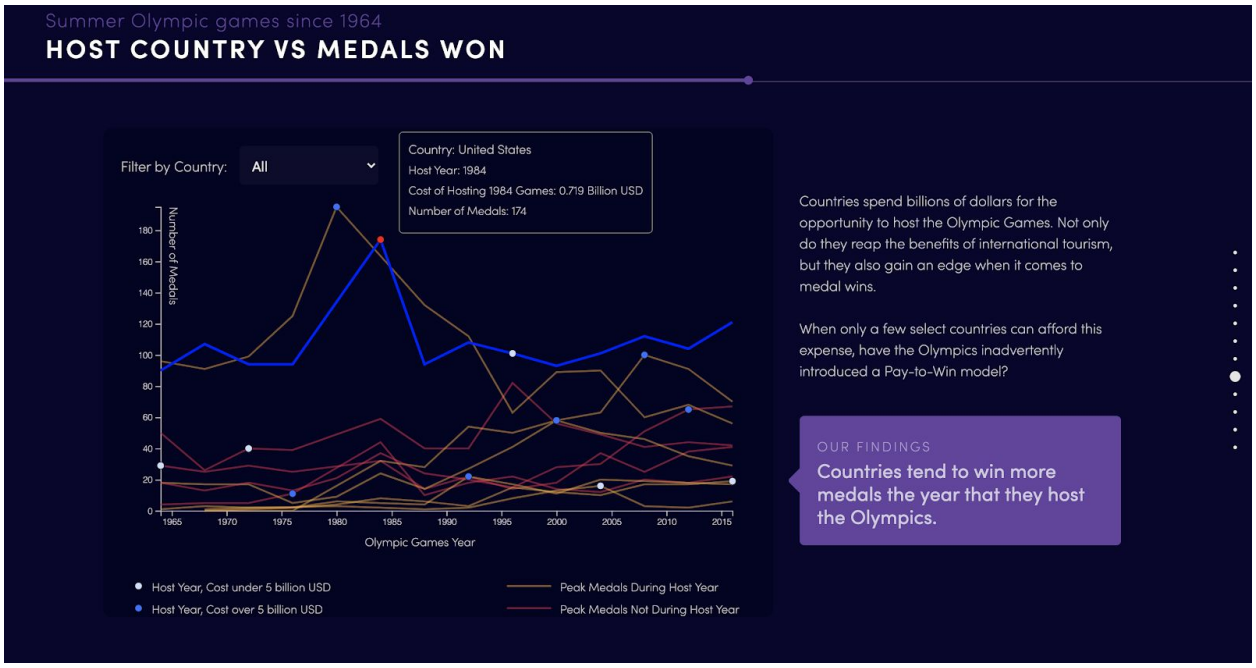
Slide6



Slide7

Olympic games since 1952
COUNTRY WINS VS GDP
Play
1952
EU & NATO
Former USSR
East Asia
Central & West Asia
Latin America
Africa
700
200
Athletes Sent to Compete
OUR FINDINGS
As non-Western countries grow in GDP, they become strikingly competitive.
1952
TOTAL MEDALS WON
Bronze, Silver, & Gold Medals
USD ($) per Person
REAL GDP PER CAPITA

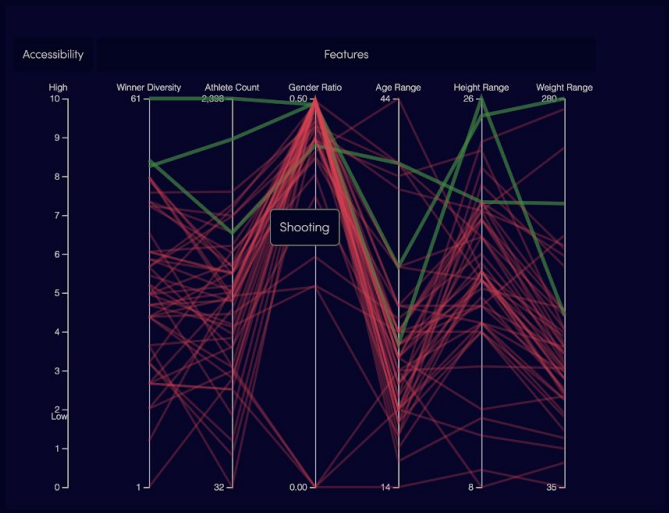
Slide8


Summer Olympic games since 1964
HOST COUNTRY VS MEDALS WON
Filter by Country:
All
Country: United States
Host Year: 1984
Cost of Hosting 1984 Games: 0.719 Billion USD
Number of Medals: 174
Number of Medals
Olympic Games Year
Host Year, Cost under 5 billion USD
Host Year, Cost over 5 billion USD
Peak Medals During Host Year
Peak Medals Not During Host Year
Countries spend billions of dollars for the opportunity to host the Olympic Games. Not only do they reap the benefits of international tourism, but they also gain an edge when it comes to medal wins.
When only a few select countries can afford this expense, have the Olympics inadvertently introduced a Pay-to-Win model?
OUR FINDINGS
Countries tend to win more medals the year that they host the Olympics.

Slide9

Olympic Games since 1952

ACCESSIBILITY RANKING BY SPORT



Sport Report Card

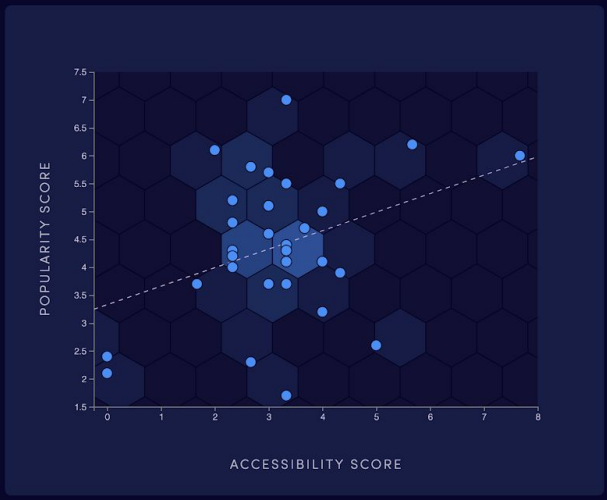
PASS 5.0/10

Sport	Shooting
Winner Diversity	35 Different Countries Won (Russia won 15% of the time)
Athlete Count	543
Gender Ratio	61% M, 39% F
Age Range	16 - 55
Height Range	4ft 11in - 6ft 8in
Weight Range	95 - 309 lbs

Slide10

Summer Olympics Games

POPULARITY VS ACCESSIBILITY



POPULARITY DATA

In 2016, the polling organization FiveThirtyEight surveyed thousands of people to find the most popular summer Olympic sports in order of relative favorability.

For example, Volleyball's score of 7 means that Volleyball was chosen 70% of the time when pitted against another sport.

ACCESSIBILITY DATA

The accessibility data, elaborated in the previous slide, aggregates athlete count, gender, size, age, and winner diversity.

OUR FINDINGS

Based on our metrics of accessibility, more accessible games are more popular.

Slide11

London Olympics in 1948. Photo by AFP/AFP/Getty Images.

A meritocracy?

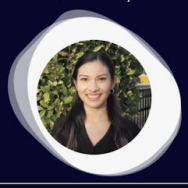
So what does this all say about the accessibility of the Olympic Games? Olympic sports are only biased in terms of the physical traits they favor, they are also biased in terms of the resources that athletes and countries have access to. Being of a particular physique is an advantage but so is being of a particular income background.

The Olympic Games might not be the truly meritocratic competition we think it is.

Slide12

Authors

Nicole Araya



Marie Konopacki

Miu Kumakura

Harvard CS171: Visualization

Fall 2020

Data Sources:

[1] 120 years of Olympic athletes and wins [2] US Athletes' Hometowns [3] Median US Household Income [4] UNESCO Institute for Statistics Data on Country GDP [5] Advantages of Hosting the Olympics [6] Olympic Sport Popularity