

Project Progress Report 2

(due May 24th 11:59p.m)

You can start working on the project once your report is accepted and graded by your TA. The entire final project is worth **35%** of your final grade and this report accounts for **10%**. This project is done individually.

Submission Guideline

Download this google doc, fill the table. **Type** your answers, no handwritten answers will be accepted (except for the very last question). Submit it in **PDF** format on Gradescope.

If you need some inspirations please feel free to take a look at:

[Showcase of Information is Beautiful Awards](#)

[Bloomberg Year In Graphics Review](#)

[The Pudding](#)

[The New York Times](#)

Project Guidelines

Note: The guideline has been further clarified from Progress Report 1, so double-check whether your dataset choice still satisfies the updated guideline below.

1. You may use more than one dataset, however, regardless if you use one or multiple datasets, your visualizations must make use of at least three following data types - **link, position, and attribute**.
2. You cannot use any dataset from the class (Labs, Assignments, Lecture Exercises)
3. You can make your own dataset (Web scrape etc.) provided point 1. is satisfied.

Part 1 - Story and Narrative

| | |
|-------------------------------|---|
| Link to the dataset | <p>The old links don't work anymore for some reason so I updated them below in red.The datasets are the same.</p> <p>Dataset1(Ski Resorts in the USA, edited using Google Sheet): https://docs.google.com/spreadsheets/d/1k-1m6TX6ezqF5YTLGxr2pxxniN2hWZys4YaJbzsKtCk/edit?usp=sharing https://docs.google.com/spreadsheets/d/1usZbFY9cti2f0FWXv7bCi34CWKahhVo3vEzenpsSISs/edit?usp=sharing</p> <p>Dataset 2(Ski Resort Worldwide, edited using Google Sheet): https://docs.google.com/spreadsheets/d/171xv7XB24brR07RqCznmzTk1jFUWslC_E2cQqmmsVKI/edit?usp=sharing https://docs.google.com/spreadsheets/d/171xv7XB24brR07RqCznmzTk1jFUWslC_E2cQqmmsVKI/edit?usp=sharing</p> <p>Dataset 3: Ski podiums(OlympicWinterGame Super G): https://docs.google.com/spreadsheets/d/1luAcgYiaMp1oP9LJlt4ZhqwY2KavpsvTbvCTgwt5fM/edit?usp=sharing</p> <p>-this dataset is web scraped and transformed</p> <p>Geojson file source:https://eric.clst.org/tech/usgeojson/ named usa.json in submitted folder</p> |
| Example item from the dataset | <p>The bolded attributes are the attributes that are mainly used.</p> <p>Dataset1: Ski Resorts in the USA</p> <p>1) The list of attributes (each item is a distinct ski resort)</p> <ul style="list-style-type: none">- State: state name of the ski resort (categorical)- Summit: summit of the ski resort's location (numerical)- Lifts:number of lifts in the ski resort (numerical)- Runs: number of runs in the ski resort (numerical)- Acres: number of acres of the ski resort (numerical)- Green_percent: percent of green-level run/slope (numerical)- Green_acres: number of acres of green-level run/slope (numerical)- Blue_percent: percent of blue-level run/slope (numerical)- Blue_acres: number of acres of blue-level run/slope (numerical)- Black_percent: percent of black-level run/slope (numerical)- Black_acres: number of acres of black-level run/slope (numerical)- Lat: Latitude of the ski resort(numerical)- Lon: longitude o the ski resort(numerical) |

| | |
|---------------------------|--|
| | <p>2) A single item (row) in the dataset as an example: Washington,5774,6,68,2325,0.3,697.5,0.4,930,0.3,697.5, 49, -115.84</p> <p>Dataset2: Ski Resort Worldwide</p> <p>1) List of Attributes (each item is a distinct ski resort):</p> <ul style="list-style-type: none"> - Rate: SkiResort's Total rate (numerical) - Stars: star of ski resort's rating (ordinal) - Km Freeride:Freeride Kms (numerical) - Continent: continent of the ski resort (categorical) - Country: country of the ski resort (categorical) - Altitude: altitude of the ski resort (numerical) - Easy: Number of KMS of easy runs/slopes (numerical) - Intermediate : Number of KMS of intermediate runs/slopes (numerical) - Difficult: Number of KMS of difficult runs/slopes (numerical) - Slope offering, variety of runs: ratings of slope and runs (ordinal) - Snow reliability: ratings of snow reliability (ordinal) - Beginners: Evaluation criteria(beginners rating) (ordinal) - Total Kms: Total Number of KMS of runs/slopes (numerical) <p>2) A single item (row) in the dataset as an example. 4.9,5,45,Europe,Austria,2000,101,61,17,5,5,"5,0",179</p> <p>Dataset 3(Olympic Winter Game Super G Podiums):</p> <p>1) List of Attributes (each item is a podium record in one game):</p> <ul style="list-style-type: none"> - Date: date of Olympic Winter Games - Podium_athlete: the athlete who won a podium(any) - Athlete_nation: nationality of the athlete - Place: the place where the Olympic Winter game was - Nation: nation held the Olympic winter game <p>2) A single item (row) in the dataset as an example. 11-02-2022,1.GUT-BEHRAMI Lara, SUI, Beijing, CHN</p> |
| Story you want to deliver | <p>(a story should be in the form of a list of facts, insights, and messages - refer to the lecture slide)</p> <ul style="list-style-type: none"> - Facts: <p>Plot 1 (geometric distribution of ski resorts in USA): Features and characteristics (locations, summit, acres, runs...) of Ski Resorts</p> <p>Plot 2 (European ski resorts and difficulties): stacked bar chart Difficulties levels (easy, intermediate, difficult) of ski resorts by</p> |

| | |
|--------------------------------|--|
| | <p>country</p> <p>Plot 3 (network of ski athletes' countries): Connections between athletes' countries and the countries they won podiums in.</p> <ul style="list-style-type: none"> - Insights: <p>Plot 1: Some states have high density of ski resorts, such as Colorado, Michigan, and New York. Some ski resort that are really close on the map may have similar summit because their location have similar terrain. For example, resorts Vail and Aspen Mt in Colorado state have similar summit(top elevation) around 11000 feet.</p> <p>Plot 2: Different countries' ski resorts have different percentages of difficulty levels, which depends on the location and the terrain.</p> <p>Plot 3: the country that holds the Olympic Winter Games may have some connections with different countries' athletes' performances. All links of Australia are thick! This means Australia won podiums multiple times in each game</p> <ul style="list-style-type: none"> - Messages: Ski is a sport full of passion and courage. Exploring more places to exercise skiing skills is definitely a meaningful thing. Also, it is encouraging to see the places where our peers from our homeland won Olympic podiums in. <p>Processes - Data transformation, selection, aggregation</p> <p>I transformed the data from dataset 3. The original dataset has each item be one game. I split the podium records of each game so that each item is a podium record (one athlete in one game). In this way, I can build a network more directly because the changed dataset has links between each athlete's nationality and the country they have won the podiums in.</p> |
| Describe your target audience. | <p>(using the questions the lecture slide listed)</p> <p>My target audiences are skiers and ski lovers. They could get information about the ski resorts and the Olympic Winter Games of Super G ski event.</p> <ul style="list-style-type: none"> - For amateur skiers and ski lovers, this project is a handful of tools to find some new places to ski with |

| | |
|---|---|
| | <p>filtering features such as the difficulty levels of the snow runs and slopes.</p> <ul style="list-style-type: none"> - For Professionals and ski team members, can enjoy more professional data from the Olympic Winter game to be more proud of their peers! <p>If they are not familiar with skiing and the games, there will be annotations to help them through the visualizations.</p> <p>The key takeaway is to expand the horizons for skiers worldwide! Skiers can explore new places to enjoy skiing around the world. Also, to provide some pride for them, they can explore where their countries won podiums in Olympic games around the world.</p> <p>My audiences will encounter standard techniques of visualization through the types of networks, tables, and 2D plots. They can interact with them by exploring, searching, and summarizing the information that the visualization conveys.</p> <p>I assume basic mathematical background to look at the charts, tables, and basic percentages. My visualization will include some explanations to help them understand the data.</p> <p>My audiences can access the project on the webpage through electrical devices.</p> |
| The goal of your project outcome. And why? | <p>(exploratory vs. explanatory)</p> <p>My project is explanatory.</p> <p>It centers on human and targets general audiences like ski lovers.</p> <p>It will focus on communication through narrative explanations.</p> |
| Narrative structure you plan to use | Interactive Slideshow |
| Elaborate your choice of narrative structure. | <p>I plan to include several visualizations including</p> <ul style="list-style-type: none"> - a geographic visualization to show the distribution of ski resorts in the US - a stacked bar chart to show the difficulty levels of European ski resorts - A network to show Olympic athletes' countries <p>Interactive Slideshow is suitable for these different sections of projects. I will include narrative guides or explanations and allow users to interact with these visualizations about more aspects of the data.</p> |

| | |
|---|--|
| Narrative genre you plan to use | Partitioned poster |
| Elaborate your choice of narrative genre. | <p>I want to design a directed series of pieces to show my data and visualizations.</p> <p>A partitioned poster can incorporate my visualizations together with different stages. Using this genre can help me show the visualization while explaining them.</p> |

Part 2 - Outline

| | |
|---------------------------|---|
| Story you want to deliver | <p>(you can copy/paste from Part 1)</p> <ul style="list-style-type: none"> - Facts: <p>Plot 1 (geometric distribution of ski resorts in USA): Features and characteristics (locations, conditions, ratings, difficulties...) of Ski Resorts</p> <p>Plot 2 (European ski resorts and difficulties): stacked bar chart Difficulties levels(easy, intermediate, difficult) of ski resorts by country in Europe</p> <p>Plot 3 (network of ski athletes' countries): Connections between athletes' countries and the countries they won podiums in.</p> <ul style="list-style-type: none"> - Insights: <p>Plot 1:</p> <ul style="list-style-type: none"> - Some near ski resorts may have some similarities such as conditions and difficulties, which depends on the locations. Some states have high density of ski resorts, such as Colorado, Michigan, and New York. - Some ski resort that are really close on the map may have similar summit because their location have similar terrain. For example, resorts Vail and Aspen Mt in Colorado state have similar summit(top elevation) around 11000 feet. <p>Plot 2: Different countries' ski resorts have different percentages of difficulty levels, which depends on the location and the terrain.</p> <p>Plot 3:</p> <ul style="list-style-type: none"> - The country that holds the Olympic Winter Games may |
|---------------------------|---|

| | |
|--|--|
| | <p>have some connections with different countries' athletes' performances.</p> <ul style="list-style-type: none"> - All links of Australia are thick! This means Australia won podiums multiple times in each game - Messages: Ski is a sport full of passion and courage. Exploring more places to exercise skiing skills is definitely a meaningful thing. Also, it is encouraging to see the places where our peers from our homeland won Olympic podiums in. |
| Specifications on each plot in the order of how you lay out on your project | <p>(for each plot, include 1) clear task abstraction, 2) attributes used, 3) marks, 4) channels, and 5) how this plot adds to the story)</p> <ol style="list-style-type: none"> Plot 1: distribution of ski resorts of different difficulties (geographic visualization) (dataset1: Ski Resorts in the USA) <ol style="list-style-type: none"> task abstraction: <ol style="list-style-type: none"> compare ski resorts' attributes(acres, runs...) across the states locate a resort on the map explore other ski resorts near some areas attributes used: (dataset 1) <ul style="list-style-type: none"> - Lat: Latitude of the ski resort(numerical) - Lon: longitude o the ski resort(numerical) marks: point mark channels: horizontal position for latitude vertical positions for longitude how this plot adds to the story: <ul style="list-style-type: none"> - My visualization aims to deliver features of various ski resort across USA. - This plot will provide more detailed information including summit, runs, and acres data. Plot 2: Stacked Bar chart of Ski Resort Difficulties <ol style="list-style-type: none"> clear task abstraction: <ul style="list-style-type: none"> - compare one countrie's difficulty component to others (part-to-whole relationship) - lookup values such as difficulty components of resorts of a country |

- find trends of the difficulty levels around the world

2) attributes used:

2 keys: country(y-axis), difficulty(color: easy, intermediate, difficult)

1 value: count

3) marks: line mark

4) channels:

- separate bars by primary Key(country) with horizontal position
- separate secondary Key(difficulty levels) with unaligned vertical position
- express Value(count of resorts per country) with length (or unaligned position)

5) how this plot adds to the story: My visualization aims to deliver the information of ski resorts' difficulty levels. This plot will help to provide a direct way to compare the difficulty levels of ski resorts by country.

3. Plot 3: Network of Olympic athletes' countries (using dataset 3)

1) clear task abstraction:

- Explore athletes' nations and the nations that held the game on the podium board
- Look up an countries exact podium numbers by hovering on an athlete_nation node
- compare the connection between where the game took place and the nations' athletes' performance of different nations

2) attributes used:

- Nation: the country that held the Olympic Winter Game
- Athlete_nation: nationality of the athlete

3) marks: point marks(circles), lines

4) channels:

color(hue) represents the country

size(area): represents the number of podiums a country won in Olympic Winter Games

Thickness of links represent podiums a country won in Olympic Winter Games

Size and color together: Larger orange circles represents the host country, smaller blue circles represent the medalists'

| | |
|---|---|
| | <p>countries</p> <p>5) how this plot adds to the story:</p> <p>My visualization aims to illustrate the relationship between the athletes' nations and the nation of holding Olympic Winter Game. This plot will provide more facts and insights about this relationship.</p> |
| Elaborate the choice of their marks and channels for each vis | <p>Plot 1: distribution of ski resorts of different difficulties marks: point mark I use point mark(circles in D3) to encode/represent the individual item of the ski resort.</p> <p>4) channels: horizontal position for latitude and vertical positions for longitude These channels represent the exact location of the resorts on the map.</p> <p>4. Plot 2: Stacked Bar chart of Ski Resort Difficulties - marks: I will use line marks because I need the heights of bars to indicate the count of resorts. - Channels: Since my keys are countries and difficulty, I use them for horizontal and vertical positions. I used the value of count so I need a length of bar as a channel.</p> <p>5. Plot 3: Network of Olympic athletes' countries (using dataset 3) - marks: point marks(circles), lines - Point marks (Nodes) encode countries - Lines (Links) encode connections that indicate the podium took place. (i.e. Countries are connected to a larger node if the athlete's country won the podium in this larger node country)</p> <p>4) channels: I will use color(hue) to distinguish between host countries and participant countries in my data I will use size(thickness of lines): which represents the number of podiums a country won in the Olympic Winter Games, the more podiums a county won in a game, the thicker the line is</p> |

Following sample answer about a single plot shows how detailed your answers to part 2 should be.

1. Plot 1

- 1) Task: This chart a) analyzes trend between Height and Weight of patients with heart diseases and b) locates outliers within the patients
- 2) Attributes: Height, Weight
- 3) Marks: point mark
- 4) Channels:
 - aligned vertical position channel for Height
 - aligned horizontal channel for Weight
- 5) How this plot adds to the story:

My visualizations aim to deliver health characteristics of patients with heart disease. This plot will provide more specific insights on Height and Weight.

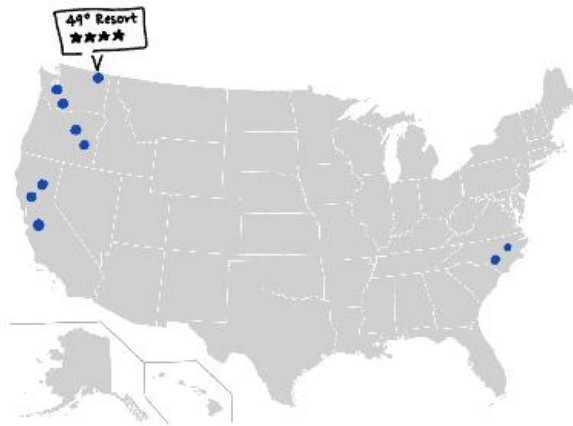
Part 3 - Prototype

Provide a photo or screenshot of your prototype. A prototype should depict how you place different components of your visualization. You may use pen-paper, or using tools like excalidraw, figma etc.

Ski Resorts & Ski Olympics

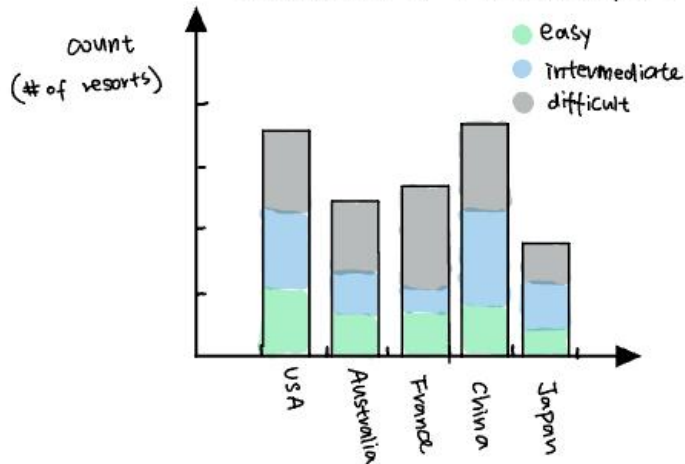
brief guidance : this project aims to provide information about ski resorts in USA , worldwide , and Olympics Winter Games . For both professionals and ski lovers or guests , explore some facts about ski .

Distribution of Ski Resorts in USA



From this plot, we can see many ski resorts on the map that you can mouse hover and explore them to see details (ratings, stars ...)

Distribution of Difficulties per Country

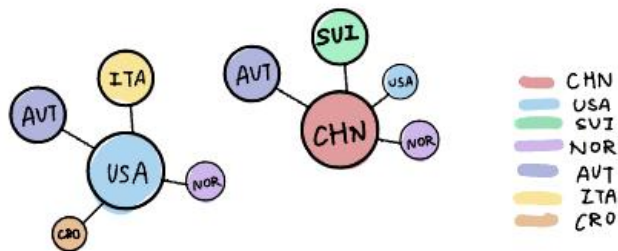


the difficulty components of ski resorts may vary from country to country. there is also some similarities among these components.

Olympic Winter Games Countries

some messages :

be proud of your nation winning podiums across the world !



Be Proud and Have Fun Skiing !

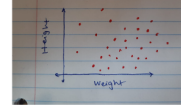
A basic, barebones sample prototype for this project

Heart Disease in the United States

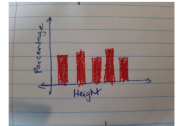
Leading cause of death for men, women, and people of most racial and ethnic groups in the United States.
One person dies every 33 seconds in the United States from cardiovascular disease
About 695,000 people in the United States died from heart disease in 2021 — that's 1 in every 5 deaths.
Heart disease cost the United States about \$239.9 billion each year from 2018 to 2019.

We will look into how height and weight plays a role in heart disease.

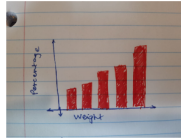
Here, we can see that weight overpowers height in terms of influence on heart disease.
Larger weight compared to height can lead to a person having a larger chance of having a heart disease.
(explanation continues)



From this plot, we can observe that height does not have a significant impact on heart disease.
Height is primarily determined by genetic factors and influenced by nutrition and overall health during childhood and adolescence.
It is important to note that height itself does not directly influence the functioning of the heart or the development of heart disease.
(explanation continues)



On the contrary, weight plays a significant role.
Weight puts a person at risk for type-2 diabetes, sleep apnea, metabolic dysfunction, high blood pressure, which in turn gives rise to heart disease.
(explanation continues)



Heart disease continues to be a leading cause of mortality globally, but the good news is that it is often preventable.
By understanding the role of weight in heart disease prevention, you can take proactive steps towards improving your cardiovascular health and enhancing your overall quality of life.

We can use the following strategies for the prevention of heart disease

- Adopt a healthy eating pattern
- Engage in regular physical activity
- Maintain portion control
- Limit sugary drinks and alcohol

Stay healthy, stay happy!