

Week 14 - Final Touches

- Dec 10th Final Touches
 - Where do students come from
 - Legend for color
 - Better switch between public/private
 - Different text at the top that explains the visualization
 - Swap the order with the one before it
 - Add context between this and the one before it
 - Standardize all the section titles
 - Add names to the welcome container

Week 13 - Prototype v2 testing

- Add explanation of ELO
 - Some conceptualization of ELO
- Profile explanation is wrong - fix that
- Add more explanation of how debates are structured, what rounds look like, what LD is, who is debating against who, etc
- Add more explanation for topics
 - Aff/neg splits key disappeared, add the key back in
- Explain aff/neg and fairness
- Filter is broken
- Add key for the gender
 - Rewriting the first section to make it easier to digest, less sentences just less material but more content
 - Add a thesis somewhere ?
- More design elements in the last 10%
- Text is too much
 - Make it more interactive and engaging

	Notes (To be filled by project leads)
Tester Name	Jack Schwab
Describe any usability issues or confusion the tester encountered while using the prototype.	Confusion with sliders for filter
Was the tester able to understand the main message of the data story? (e.g., Yes/No + why/why not?)	Conceptual understandings —missing context on what ELO is, how debates are structured. Could interpret that our thesis was connecting gender/schooling to debate performance.
What parts of the interface or visualization did the tester find most engaging or effective?	The tester found the bar graph of the four groups most effective.
What parts did the tester find confusing or less	The tester found the "How fair are debate

effective?	tournaments" confusing because he couldn't discern the meaning of the green/red, and the meaning of the labels as tournaments (as opposed to schools).
Did the tester encounter any inconsistencies in design, data, or narrative?	The tester found that we refer to the average ELO as -200 and as 400.
Were there any unexpected interactions or insights that emerged during the session?	The slider tools did not work as expected.
What specific improvements or changes did the tester suggest for the prototype?	The tester suggested providing more context, such as a hypothesis about what we are investigating, and introductory message about what ELO is. Trying to avoid jargon-y language like "triple octa".
Did the tester suggest any additional insights or visualizations to include?	The tester suggested including legends, and including a filter for the last slide.
General observations or comments from the tester.	

Week 10 - Visualization Sketches

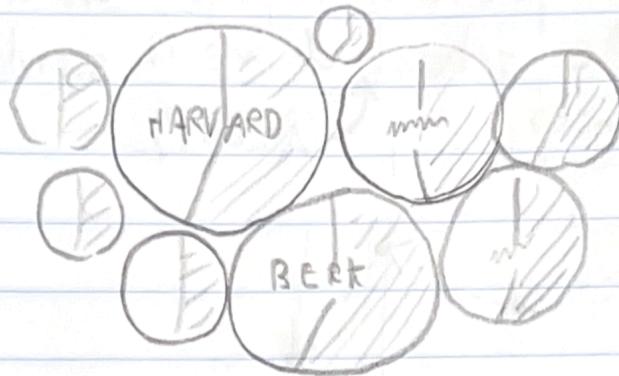
Vis. 1, Liam

10 Major Tournaments

Size: # of rounds

◻ Affirmative wins

■ Negative wins



Vis. 2, Liam. Interactive.

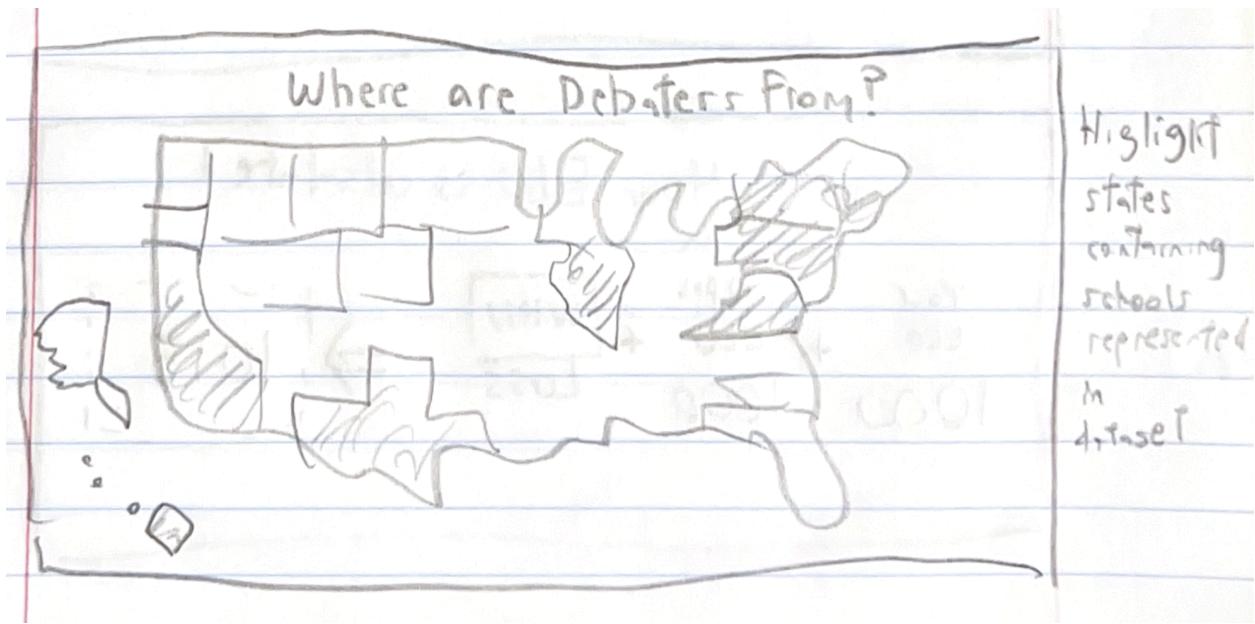
How ELO changes

Highlight
a specific
debater
color:
Gender

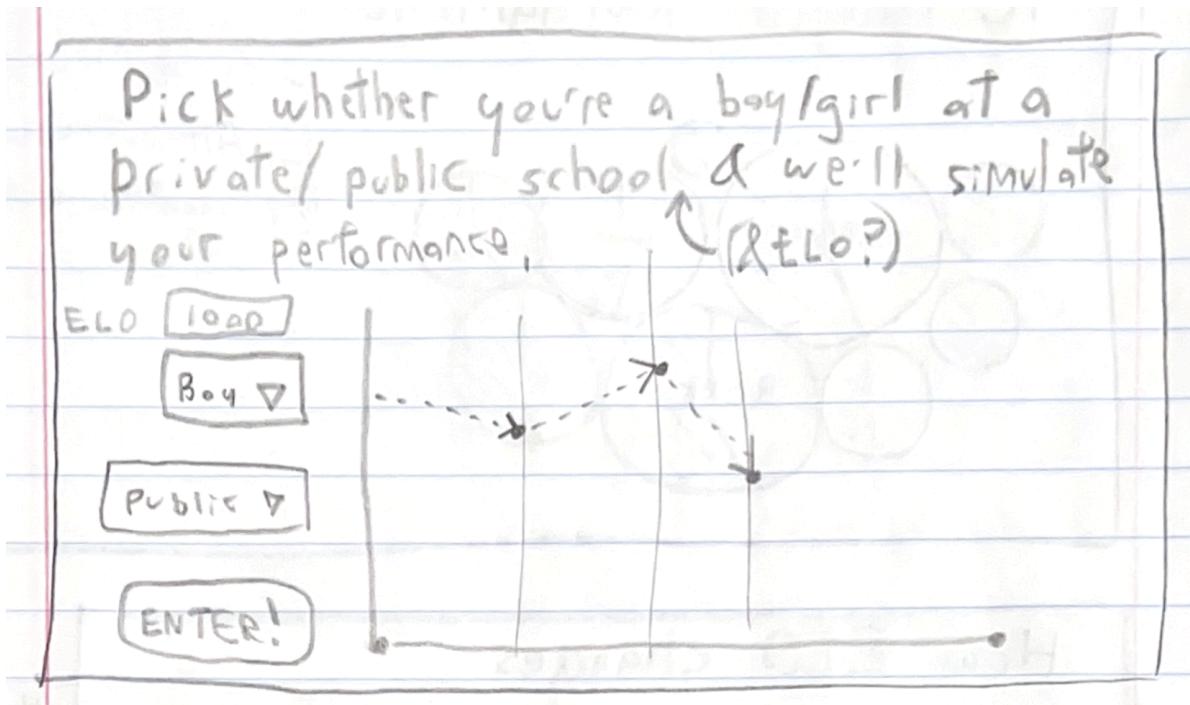
ELO

Tournament 1 2 3 4 5 6 7

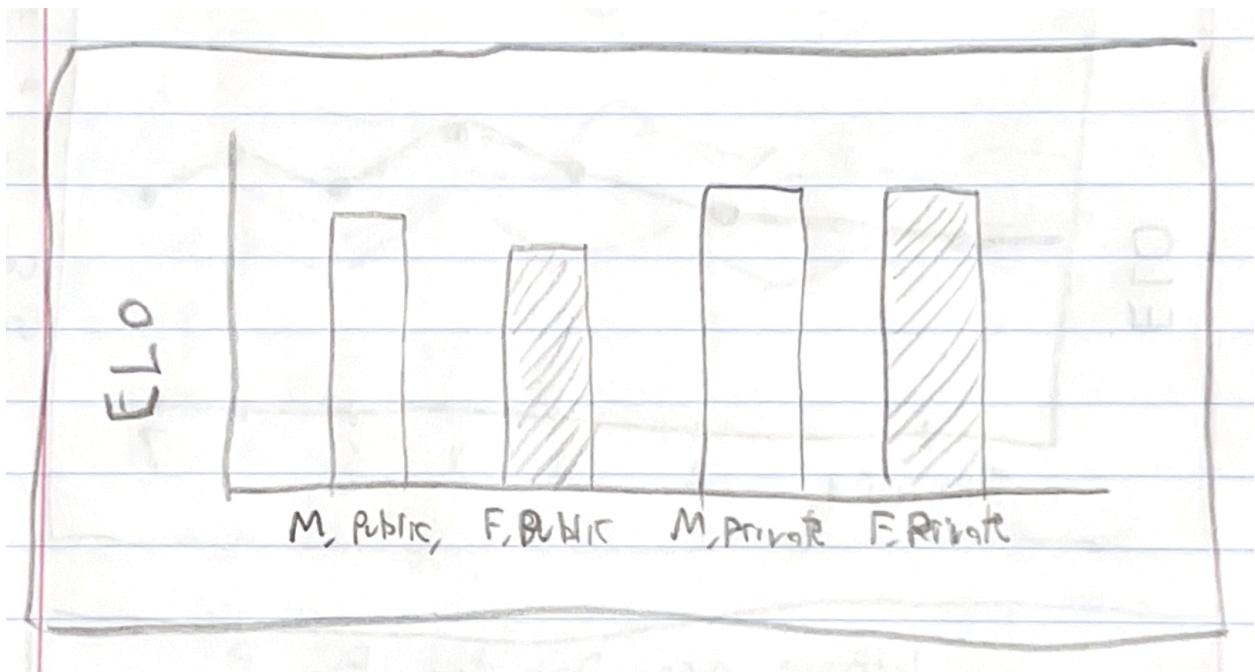
Vis. 3, Liam



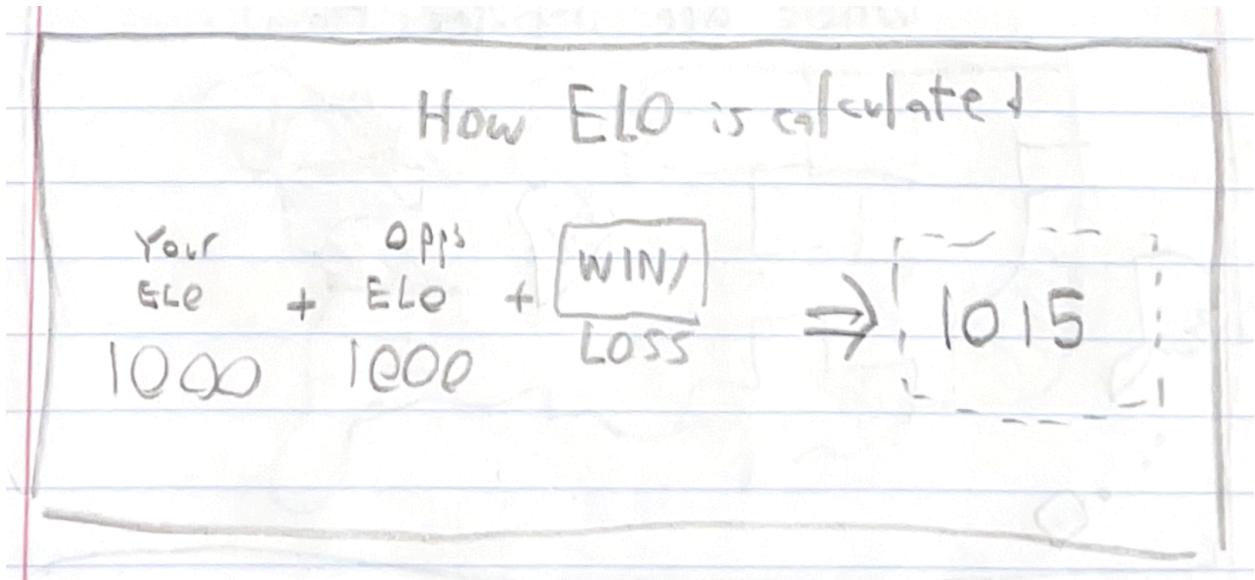
Vis. 4, Liam. Colors encode gender & public/private. Interactive.



Vis. 5, Liam. Comparison of average ELO conditioned on certain factors. Colors encode gender & public/private.

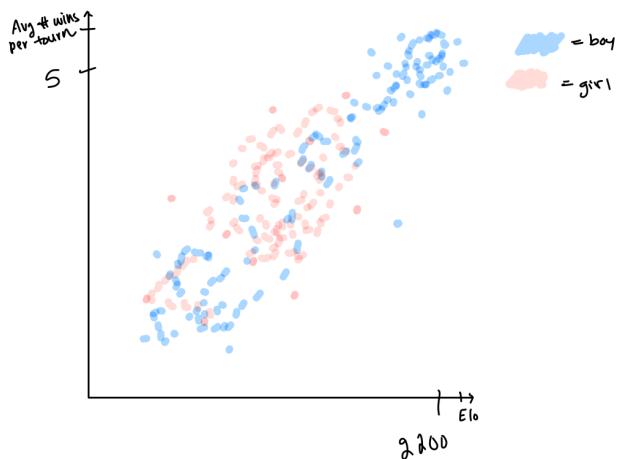


Viz. 6, Liam. Interactive.



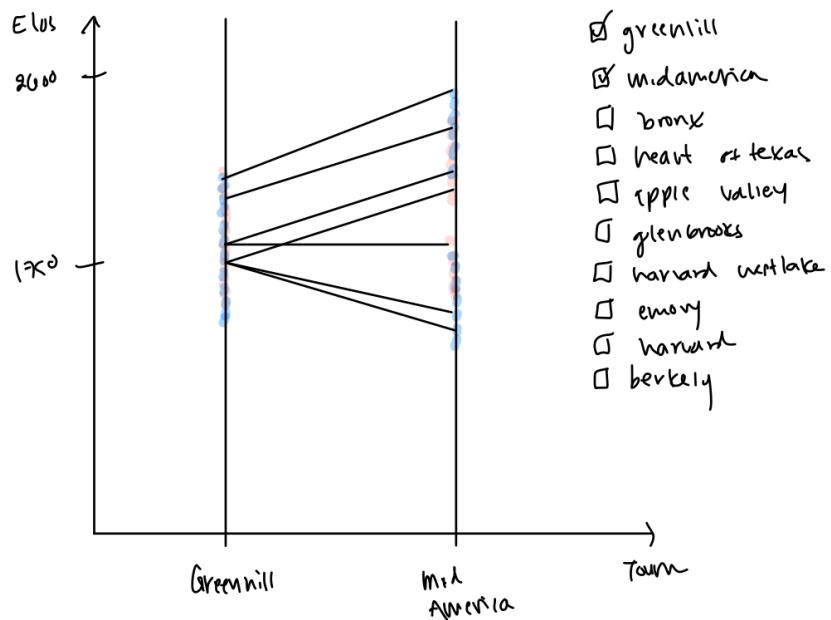
Viz. 7, Grace, Colors encode gender

Average Number of Wins vs. ELO score



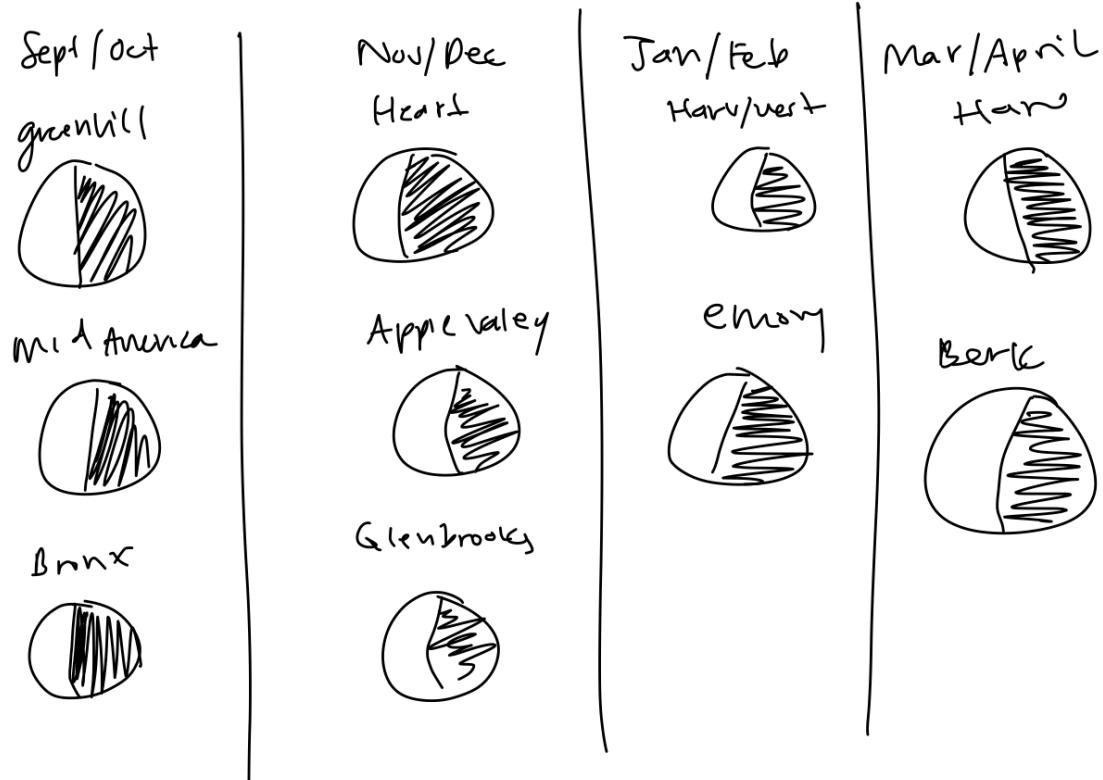
Viz. 8, Grace, Interactive choose which tournament, colors encode gender

How did a specific tournament affect ELOs?



Viz. 9, Grace

Were topics Aff/Neg skewed?



Viz. 10, Grace, What does the average debater look like?

Average Debater Profile.

XXXX

ELO

X.X

Avg WMC/Tourna



X

Tournaments
Attended

Viz. 11, Grace, Interactive, what percentile debater are you?

What percentile are you?

enter how many rounds you win

drop down male female

Tournaments Attended

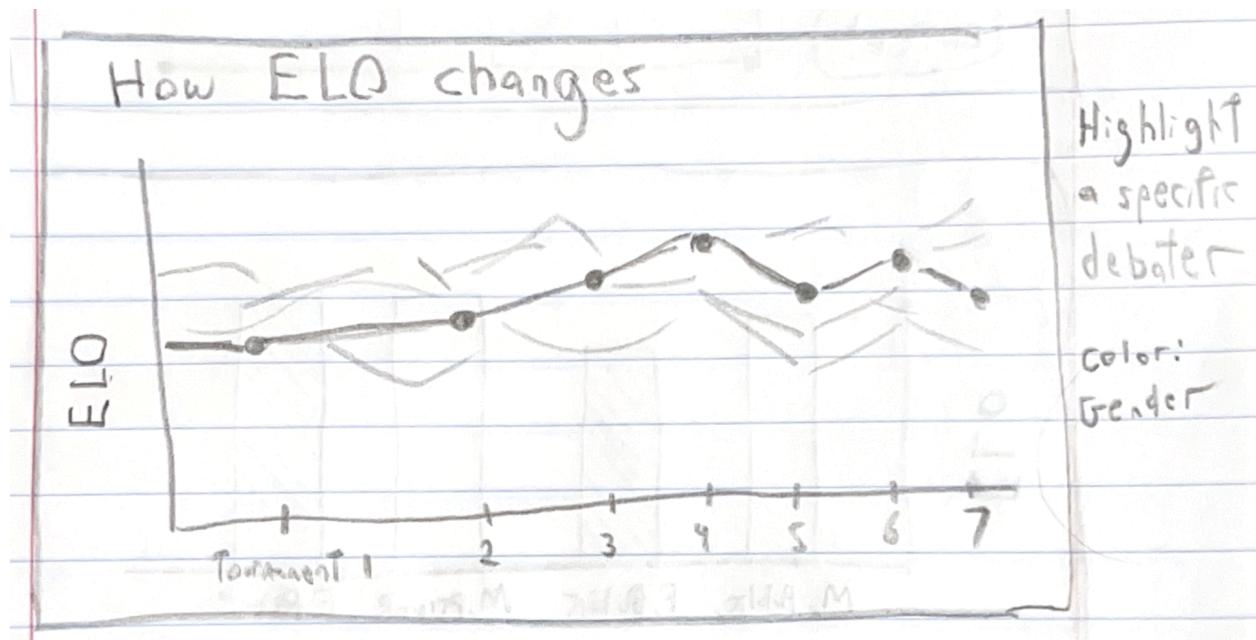
XX.X%

percentile of
all men

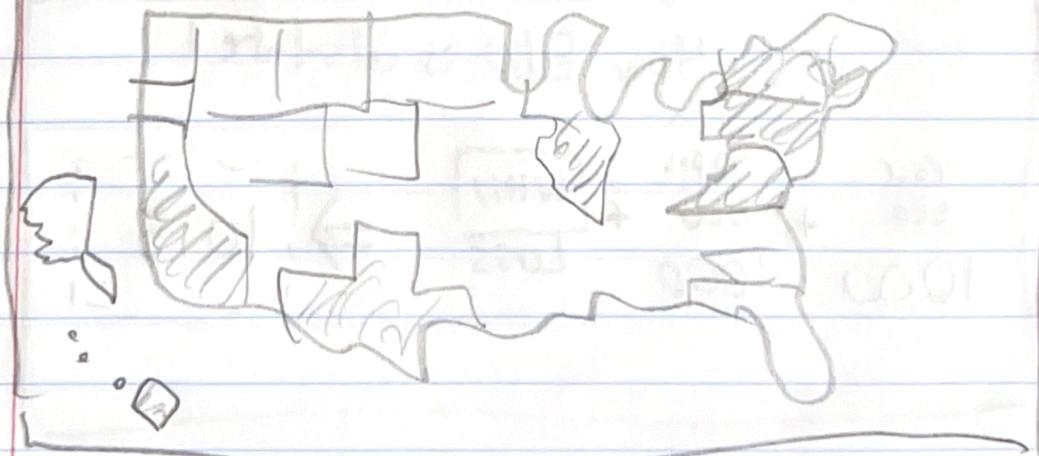
1.

Sketch ID	Question ID	Author	Votes
1	7,8	LN	
2	6	LN	1
3	10	LN	1
4	5	LN	2
5	1,2	LN	1
6	6	LN	
7	1	GK	
8	6	GK	
9	7,8	GK	1
10	11	GK	2
11	5	GK	2

Chosen:



Where are Debaters from?



Highlight
states
containing
schools
represented
in dataset

Pick whether you're a boy/girl at a
private/public school & we'll simulate
your performance.

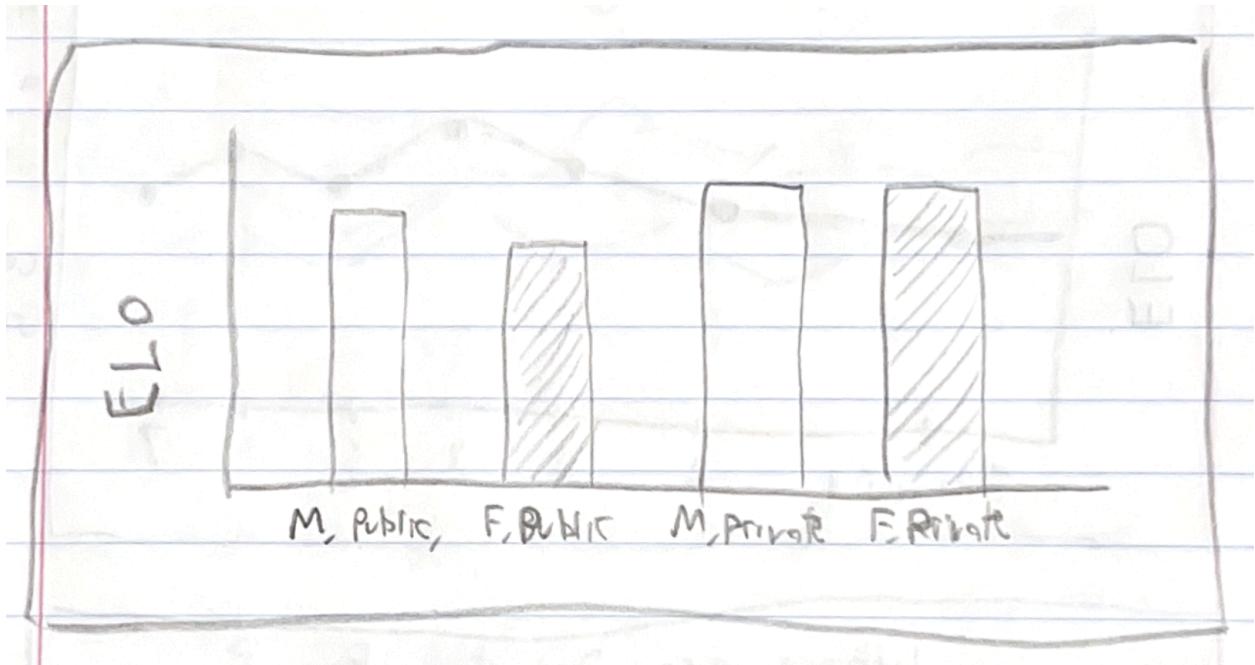
(& ELO?)

ELO 1000

Boy ▽

Public ▽

ENTER!



Were topics Aff/Neg skewed?

Sept/Oct

greenhill



mid America



Bronx



Nov/Dec

Heart



Apple Valley



Glenbrook



Jan/Feb

Harv/vert



emory



Mar/April

Harv



Berk



Average Debater Profile.

XXXX

ELO

X.X

Avg WMC/Tourna



Male

X

Tournaments

Attended

What percentile are you?

4

enter how many rounds you win

male v

drop down male female

3

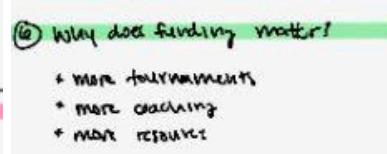
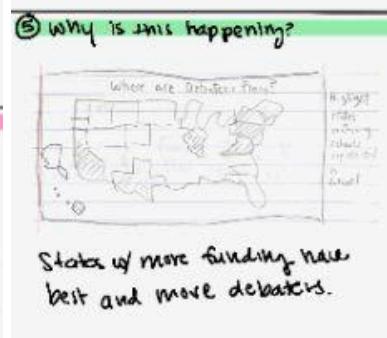
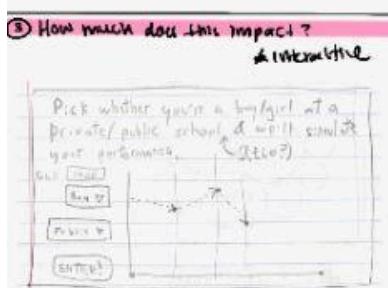
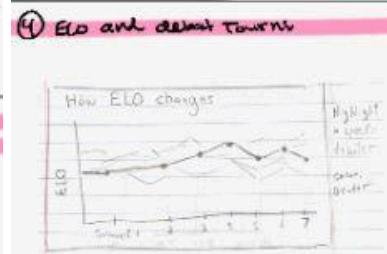
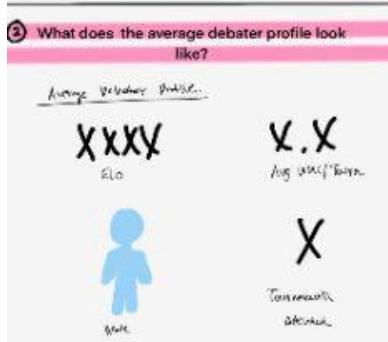
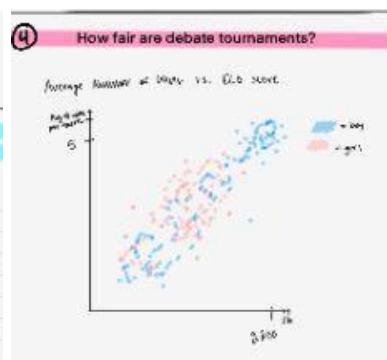
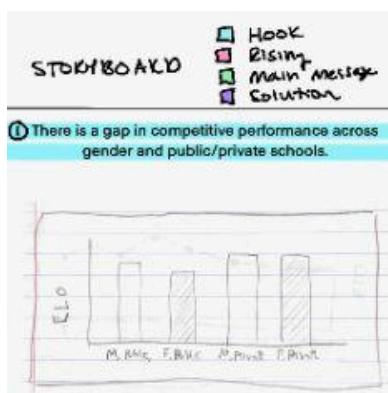
Tournaments Attended

XX.X%

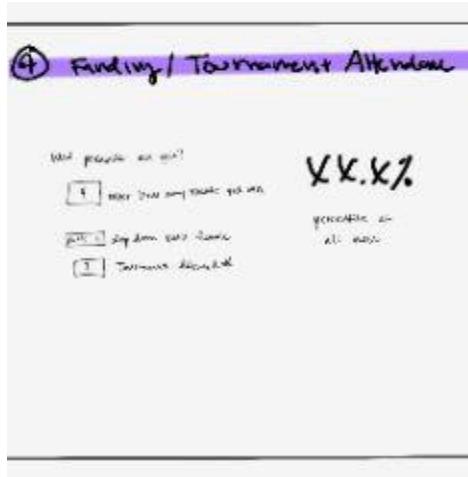
percentile of
all men

We chose these visualizations because we think they answer the most pertinent questions regarding debate data. Some visualizations would have been repetitive to keep (ie. viz 1 and viz 9) as they depict pretty much the same data and answer the same question, so we tried to limit that type of overlap. We also tried to keep a mix of interactive elements and just informative visualizations.

Storyboard



④ Funding / Tournament Attendance



Mapping

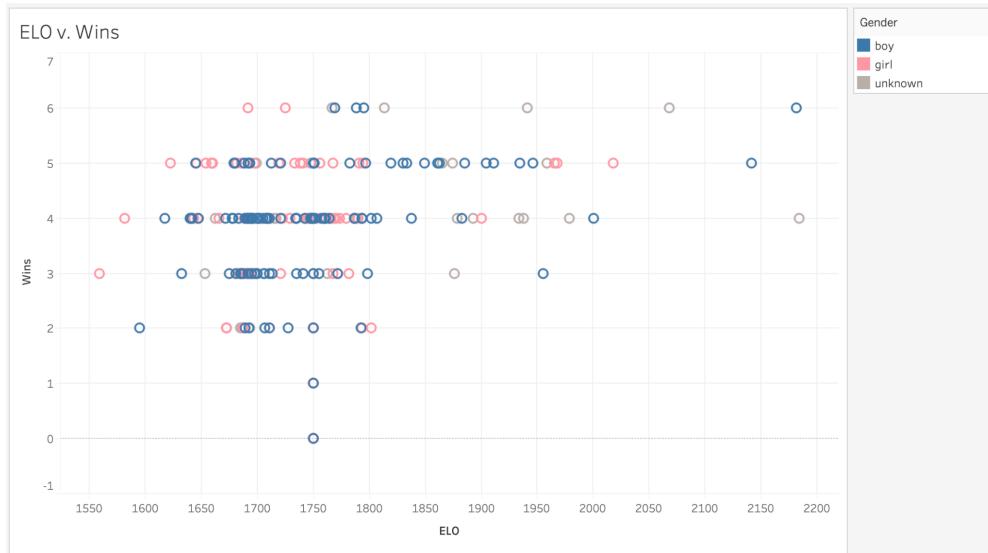
Team

1. Who is your audience? Come up with **at least three** options and pick one target audience.
 - a. Debaters
 - b. Debate instructors
 - c. Policymakers
 - d. **Casual audience interested in the culture of debate (our pick)**
2. Describe your target audience in more detail. What do they know? What are their interests? What visualization literacy do they have? At what level of detail will you present information to them?
 - a. Since the audience may not have experience with debate, they may lack context about the practice of debate. They may not be familiar with the format of debate competitions, the shorthand language used, the popularity of debate, an idea of who competes in debate, or common topics of debate. They may not know which debate competitions are the most important, and what ELO is. We can probably assume that a casual audience will understand that debate is a form of competition with winners and losers, and that it follows a similar format to other competitive practices like sports.

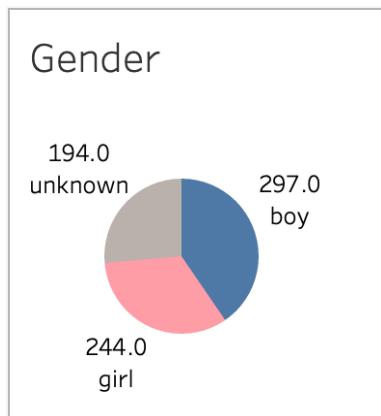
- b. Our audience most likely has a middling visualization literacy. We don't want to make visualizations too complicated. Since this is for casual readers, readers should be able to understand the big-picture message of each of our visualizations with a 5 second glance, and the visualizations should look appealing.
 - c. It would be nice if our visualizations provided expository information about the culture of debate so that casual readers can get a sense of it.
- 3. What questions about your data will be interesting for your audience? Come up with a list of interesting questions that your audience may have about your data. The more, the better, but your team should come up with **at least ten questions**.
 1. What is the difference in competitive success between boys and girls in debate?
 2. What is the difference in competitive success between private and public school students?
 3. Why is there a disparity in competitive success between private and public schools students?
 4. Which schools are the most competitive?
 5. How might I fare against the best debaters?
 6. How did student ELO scores change throughout the competitive season?
 7. Are some debate motions more likely to just let the affirmative side or negative side win (are they skewed)?
 8. How much of tournament results are based on luck?
 9. Can we predict the outcome of a tournament accurately?
 10. Where are the best debaters coming from?
 11. What does the average debater profile look like?
- 4. What data do you have? Download the data you picked from the website linked in the PDF that describes the data (available on Canvas, week 2). Look at it in Excel or Google spreadsheet and briefly describe each attribute and its data type (categorical, ordinal, or quantitative) in your process book. It's OK if you are unsure about the data type for some attributes - you can simply describe them (e.g., geographic location).
 - a. We are scraping data from tabroom.com from the 2023-2024 debate season. We are focusing on Lincoln Douglas Debate. The code for our scraping and cleaning can be found [here](#). We will likely scrape a few extra things to get more insight into other questions we have about debate, but the initial dataset is complete and can be downloaded [here](#).
 - i. Code - Categorical, Nominal
 - ii. Gender - Categorical, Nominal
 - iii. Name - Categorical, Nominal
 - iv. Private/Public - Categorical, Nominal
 - v. School - Categorical, Nominal
 - vi. ELO - Quantitative, Continuous
 - vii. Wins - Quantitative, Discrete
 - viii. Average Speaks (in progress) - Quantitative, Continuous

- ix. Aff win percentage (in progress) - Quantitative, Continuous
- x. Neg win percentage (in progress) - Quantitative, Continuous

Individual - Liam



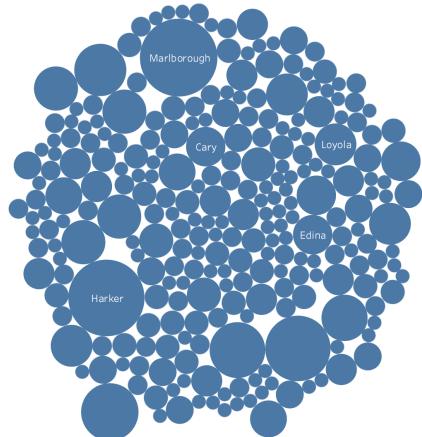
This chart compares ELO vs. wins. It has some unexpected results. As expected, players with high ELOs have won a higher number of games. However, some players with low ELOs have also won a fair number of games. This chart shows the general relation between wins and ELOs in general, and it highlights how winning is not only a factor in ELO, but also who you win against. Even with gender color coded, it is hard to spot a pattern related to gender. Compared to some of the other visualizations we have, this one says less about disparity in debate since the only variables are ELO, win number, and gender.



Gender breakdown. This provides some expository information about the demographics of debate. If the data had more people labeled, it would be easier to see if one gender was more prevalent than another. A simple chart like this at the beginning of a visualization series could

give casual readers a clue as to which people participate in debate, which would be helpful if they had not participated in the activity before.

Most Represented Schools

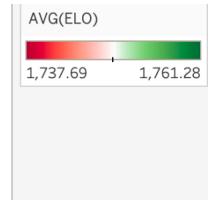


This is a bubble chart showing the schools who send the most students to debate competitions. The data for each school is currently not labeled substantially enough to color each school by private/public, but that could be a good addition to this visualization. This does a good job of answering which schools send the most people to debate. However, it raises the question of why these schools in particular are the biggest, where these schools are located, and what their connection to debate are. If we included a graph like this in our final project, it would be best to annotate it with short biographies of some of the bigger schools.

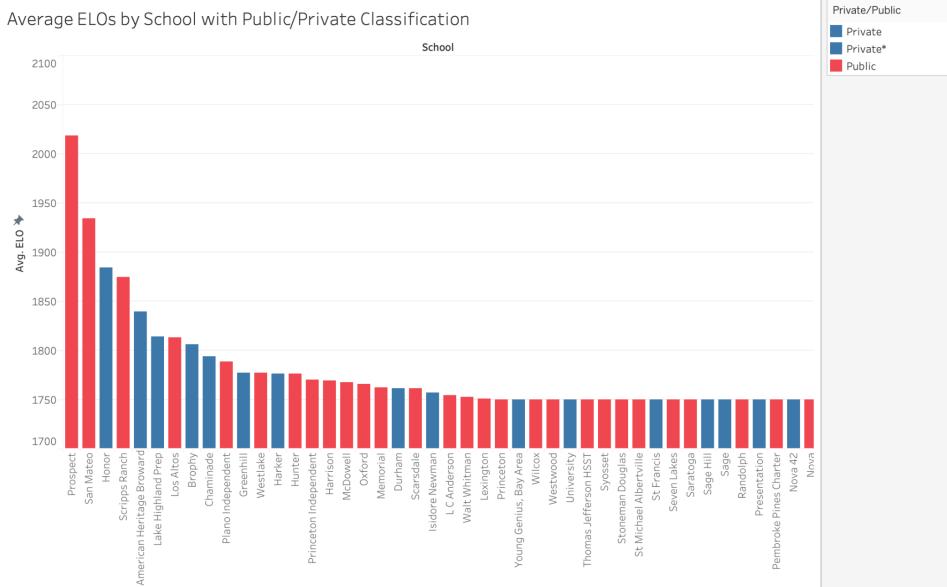
Individual - Grace

Private/Public and Gender ELO Averages

Gender	Private/Public	
	Private	Public
boy	1,761.28	1,743.92
girl	1,737.69	1,745.66



- This one is pretty simple, it shows a table of average ELO scores segmented by private/public schools and gender. While this is a simple table and it clearly answers the question of the gender disparity, it also begs the question of why private school girls are performing worse than public school girls. Does this discount the idea that private school students are generally more competitively successful? So, while it helps answer one question, it also poses more.



- This chart helps us answer two questions, (1) which schools are the most competitively successful and (2) whether public or private schools are. It is kind of hard to tell based on this chart though a solid answer to each of these questions. Considering that there is such a big spike for the first place school, it leads me to believe that there was just a single good student from that school which would potentially skew our results. So, while it does answer those questions, we should take a second to think about whether those answers are valid and look for other ways to legitimate them,



- This chart helps look at individual students and their ELO scores to see who is the most competitively successful. We see that there is more blue on the very upper ends, but there is also a little pink patch that is slightly above the little blue patch. This helps answer the question of how competitively successful boys and girls are in debate. This shows that the most competitively successful are men, but that there are some girls that are also fairly competitive. We might want to look at other ways to visualize this.

Team Agreement & Detailed Project Plan

Basic Information

- Project Title: Debate Dashboard
- Group Member Names and Emails:
 - Liam Norman lnorman@college.harvard.edu
 - Grace Kim gjkim@college.harvard.edu
- Team Name: Grace and Liam

Background and Motivation

- The idea for this project came from a few projects done throughout college regarding debate where we've collected tournament data, and analyzed competition success. These range from measuring debate performance disparities between public and private, measuring the impact of sexism on competitive debate, or even just creating ELO scores for individual debaters to predict tournament outcomes. For our final project in CS171, we want to make a website that accumulates all this data onto a website to visualize the analysis we have done throughout college. We hope to make a website that highlights different disparities in debate, as well as other interesting trends.

Related Work

- <https://www.debate.land/>
- <https://colab.research.google.com/drive/1sZCpos93lYksi9liCAQt3ZJMDdYQmRpX?usp=sharing>

Data

- Data is being collected from Tabroom.com using a tournament scraper that I created for a math modeling class. Using tournament result data, we create ELO scores and also attach classes of Public/Private School or Male/Female for each individual debater.
- The ELO scores are created using LD tournament result data from octafinal bid tournaments. This is because octafinal bid tournaments are the largest and most relevant tournaments. This classification also ensures that the tournaments are about the same level of difficulty.

Data Cleanup

- Aside from code that was used for past classes, we scraped extra data to make it easier to attach first and last names to each speaker code. This will allow us to use a gender function to assign gender to each debater based on their first name. We acknowledge this is an imperfect method since there are some asexual names like Alex, however we think this will still allow us to visualize trends for the whole of the dataset.
- We also have a dataset of school names and whether it is a Public/Private School so we will combine these datasets.
- In the end, the dataframe will have the following columns:
 - Name | Code | ELO | Gender | School Type
- We can then export this as a csv to process it and visualize it like we have done with past class problems.

Team Agreement

We will check in weekly about progress in our project. We plan to each do half of the visualizations. We made a [project github](#) so that we can both collaborate on it asynchronously and keep track of project progress.

If one of us is falling behind, they should tell the other ASAP. Then, they must pick up on their work.

Grace Kim

Liam Norman

Project Proposal - Debate Details

Throughout college, I've done a few projects regarding debate where I've collected tournament data, and analyzed competition success. These range from measuring debate disparities between public and private, measuring the impact of sexism on competitive debate, or even just creating ELO scores for individual debaters to predict tournament outcomes. For my final project in CS171 (and as a senior this year), I want to make a website that accumulates all this data onto a website to visualize the analysis I have done throughout college. Since I have most of the functions to scrape tournament data from tabroom.com already, I will likely start by adjusting those functions to scrape tournaments for this project and cleaning the data to suit my needs. I hope to make a website that is both enlightening when it comes to disparities in debate, and feels fulfilling as I close out my senior year.