Bin Lama
Jihane Eter
Sammi Rowen

<div align="center">Final Project</div>

Our dataset showcases information about Ted Talks, popular conferences that explore ideas worth sharing, available to the ordinary person through either attending the talk or watching an online recording after the fact. These talks happen throughout the year on various continents and cover a diverse array of topics from technology to theater, medicine to personal development. They vary in length, with some talks lasting three minutes and others twenty minutes or more.

With all of this variability, we began to wonder how this affected viewership and each talk's popularity. Were viewers more interested in certain topics over others? Did videos that were longer lead to more comments, as viewers had more knowledge after watching or more interest to sit through a longer video? Did talks with more views necessarily correlate to more comments? Given our initial thoughts, we decided to answer some of these questions through visualizations. We filtered the original data to keep only the columns related to views, comments, the duration of the video, the number of languages available, the Unix timestamp of the filming, tags, the speaker's name, and the URL of the talk on Ted.com. After beginning our visualizations, we also added a column for the film date, in this date formatting (yyyy-mm-dd), as well as a string for month and year. These helped in building the line chart and calendar view in our visualization.

Our visualizations begin at a large granularity, showcasing the frequency Ted Talks are filmed over the years 2012 through 2017. We chose to use these years, as were closest to today's time and we felt five years would not overcrowd the visualizations.This data would be relevant to someone planning to attend a Ted Conference, as they could plan to take time off around these common times and travel wherever the conference was held. Upon clicking on a particular year in the line graph, a calendar view updates so viewers can see what days are most popular for filming. We originally planned to use the publish date of the video, but we found that the film date created a more diverse, interesting visualization as the dates spread more widely across the year. We also discussed the relevance of using the film date instead of the published date, because users would likely be more interested in the societal pressures and hot topics on the news, in education, and in the media during the time of the talk rather than when it was released. Additionally, in our initial design, we wanted to create a brush tool for the calendar view; however, this posed two main issues. The first issue involved the creation of the year as one big svg element with lines drawn on top to organize the months. Thus, we could not separate out the months, making it impossible to add a brush tool. So instead we chose to select one day by click, and that month would then be used to filter the next charts. At first, this seemed like a big issue; however, we believe the functionality is still apparent because the channels for the days are much smaller than the channels for the month, making it easy for users to still feel like they are clicking by month. Additionally, after seeing our visualization begin to work, we believe having the ability to choose multiple months would have cluttered the parallel coordinates graph, making it illegible and decreasing the usability of our system.

Once the month selection is made, there is an update to the parallel coordinates chart that shows every talk in that month as well as an updated bubble chart that depicts the tags correlating to those talks. Each line in the parallel coordinates chart represents one video and depicts the total number of comments, the duration of the film, the number of languages the talk is available in, and the total number of views the video has on the axes. The colors encode how many TED talks the speaker has given, which has numerous reasonings--to see if the speaker talks about the same topics, and thus be in the same tags, as well as to see if speakers with more talks were more popular overall. The channels for each line, the color, doesn't contain a direct meaning other than separating it from lines with different values.

Additionally, clicking on the calendar view also updates the bubble chart to the right of the parallel coordinates chart, showing the top thirty tags of videos filmed in that month. Tags describe the categories of the talks, using words such as culture, biology, internet, art, community, or creativity. We changed our design for this visualization, as we were originally going to create a treemap but did not want to group tags into a hierarchy, as this would create further confinement for the talks. We also felt users should be able to choose videos based on the tags the Ted Organization deemed, since these would be how a user would find the video on Ted.com. We initially planned to display only thirty tags so users could more distinctly see the tag names as well as a greater distinction in circle size. However, while we attempted to limit the frequency of bubbles, we could not figure it out and thus, all tags are present as bubbles.

Another change we made after implementing our visualizations was our updates between charts. Originally the calendar view was going to update the bubble chart, and upon clicking the bubbles the parallel coordinates chart would be updated. However, we wanted to have more back and forth interaction, so the parallel coordinates chart and the bubble chart now both update one another.

Final Visualization System
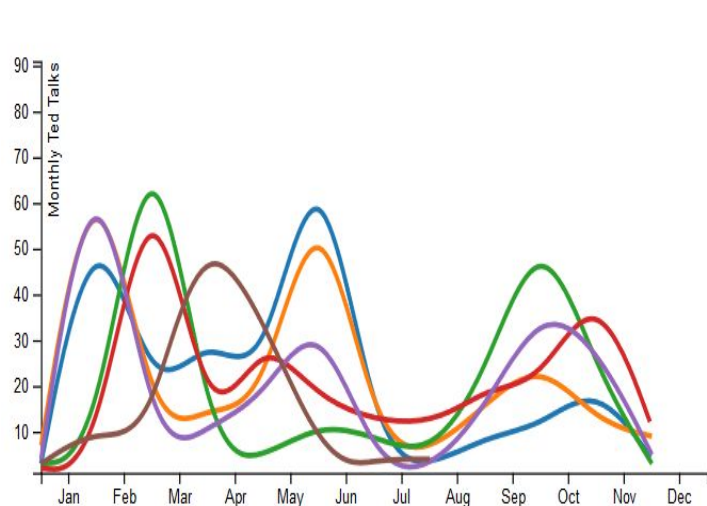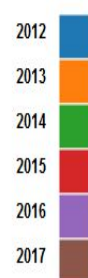
Overview of line chart: This is our first visualization, and it illustrates the total number of Ted Talks broken down by year and further on the X-Axis by month. Users can see definitive trends (low production in the summer months, higher in Feb/Mar for 2014 & 2015 and a change for 2016 & 2017).

Overview of calendar view and bubble chart:
These are two of our four final visualizations, the calendar view and the bubble chart. Both visualizations have the tooltip pictured, with the black background and the bolded text for category with the text in white. We chose to use the same color schemes for these two visualizations because our other two visualizations have five to six colors for their encodings and we wanted to mellow out this side of the viewport. We decided to outline the bubbles in black because of the light pink for low frequencies.

The interaction between these two graphs include clicking a month from the calendar view and updating the bubble chart to show all tags from those videos (and each video has numerous tags, making the bubble chart potentially have much more data than the parallel coordinates shows for frequency of talks.
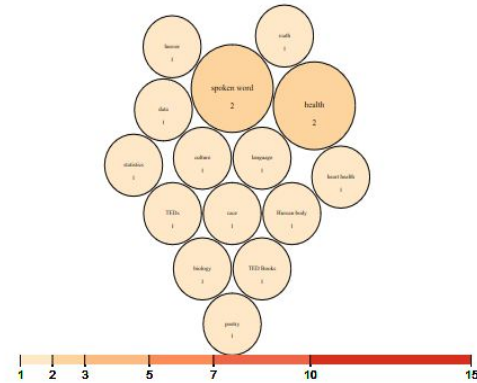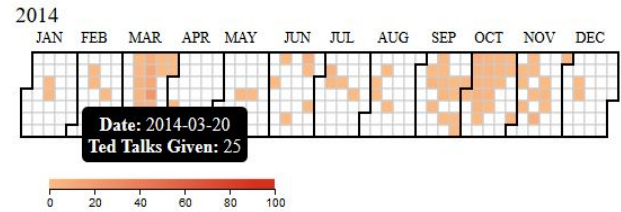
Examples of the interaction between parallel coordinates & bubble chart:
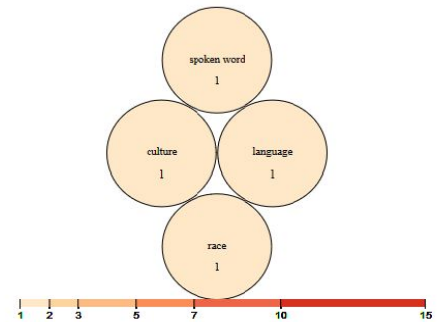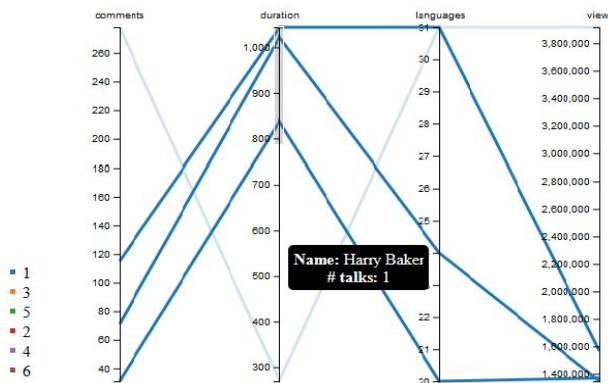
The parallel coordinates chart allows for multi-selection through brushing, so when the visualization gets too crowded users can filter the data they are interested in. To remove the brushed stroke, click on the axes the stroke was on. The brush stroke only works for the comments axis, but does update the bubble chart with the appropriate tags. If you want to filter using the tags, i.e. see the visualization from the opposite direction, select one or multiple bubbles and this will update the parallel coordinates chart with any lines that correlate to those tags. To return to all tags, click on a stroke. The colors correspond to tags, and the legend is updated to represent this as well. *Interaction shown below:*
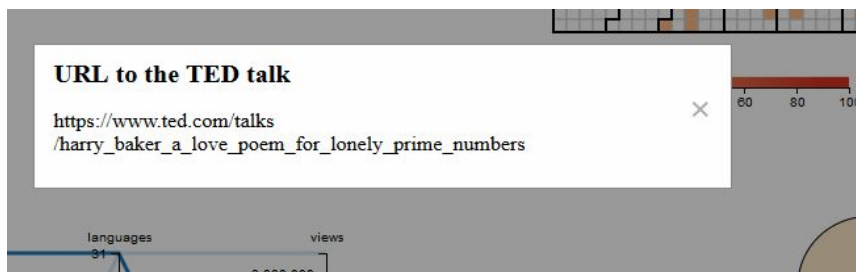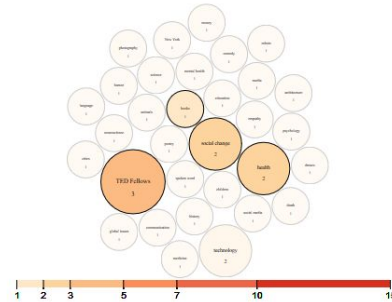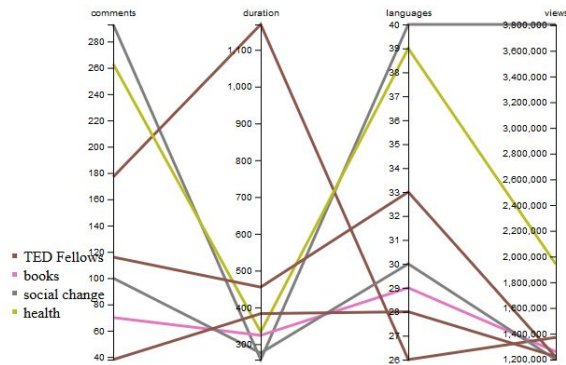
Clicking on a line in the parallel coordinates chart creates a pop up with the URL of the Ted Talk so users can watch the talk right then.
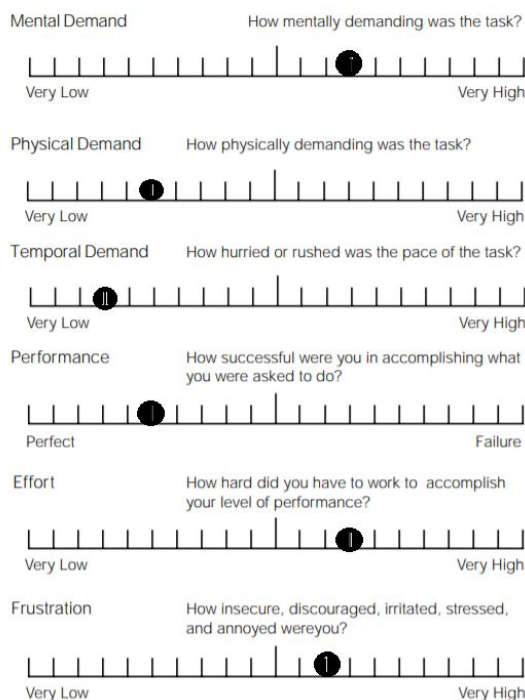
User Study:

Because we were limited in time for our user research, we were unable to interview classmates and other groups. However, we averaged responses from 2 family members and 3 friends (5 total) who collectively filled out a NASA-TLX for each task. The figure to the left shows the average response for the following tasks:

1. Can you please navigate through the system to find the pop-up URL.

2. Can you explain to me what each line means in the parallel coordinates graph (the bottom left)?

3. Can you figure out how to change the calendar view to a different year?

4. Can you tell me what the numbers under each bubble represents?

Based on their responses, we added more explanations subtly throughout our visualizations, like a tooltip for the bubble chart and parallel coordinates chart, to explain the questions that our users couldn't answer. When a large portion didn't know the answer to question four, we added a title in the tooltip, so users, if they hovered over the bubble, knew it was the frequency of talks. Additionally,

we found that users needed some help navigating through our system to end at the URL, but they did like that at the end, after choosing all of their filtering, they could watch the Ted Talk easily.

Tasks were divided among team members in the following way:
We all researched and created the design plan together.

Bin created the bubble graph, parallel coordinates chart, and coded the two way interaction between his two charts and added the popup too. He also contributed to the mile-stone write up, and collaborated on all interactions between charts.

Jihane created the line graph chart, created the presentation, contributed to the user study, collaborated on all interactions between charts and contributed to the mile-stone write up.

Sammi created the calendar view chart, wrote the final project write-up, created the presentation, contributed to the user study, collaborated on all interactions between charts, and contributed to the mile-stone write up.