CS171 Final Project Process Book

Team Name: Piazza Posse
Team Leader: Katherine Deng (katherine_deng@college.harvard.edu)
Team Member: Christina Wang (chw6846@g.harvard.edu)
Team Member: Anna Li (annali2@college.harvard.edu)

Quick Links:

- Final Project Guidelines

Week 8: Process Book and Project Proposal

Project Title:
Asking the Burning Questions: Visualizing The West Coast's Devastating "New Normal"

Project Abstract:
In this project, we will analyze and visualize the wildfires that have been plaguing the West Coast in recent years. For each of the past few years, California, Oregon, Colorado, and other West Coast states have been hit by their worst wildfire seasons in modern history. Millions of acres have been destroyed, thousands of residents have been forced to relocate, and numerous lives have been lost. These tragedies have also sparked a fierce national conversation on climate change, and intense fears about the inhabitability of a warmer world. Our goal with this project is to better understand and visualize West Coast wildfires - their prevalence, spread, origins, impacts, and public perception, and how these have all changed over time. As a starting point for our project, we will work with official government wildfire datasets, such as the California Fire Perimeters dataset published by Cal Fire, and the Spatial Wildfire Occurrence dataset published by the US Forest Service. Moving forward, we may also consider linking county-level demographic data, to better understand the people who are most affected under this devastating "New Normal".

Week 9: Team Agreement & Detailed Project Plan

Team Agreement

- We've created a group chat, which will be our primary platform for communication on this project. We'll use it to discuss project work, ask and answer questions, share useful resources, set up group work sessions, and coordinate task delegation, outside of our weekly meeting time. We commit to being open and responsive in all communications.
- Weekly meetings: Wednesdays at 5:45pm (after lecture) over Zoom.
- Although code will typically be written individually, all team members will be involved with the technical aspects of the project. All code will be documented well and pushed frequently to our central repository on GitHub to ensure streamlined collaboration.
- Final design decisions will be discussed among all members; fair compromises should be made when necessary.
- Work hours will 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 hold each other accountable so that one person does not work too much or too little; if there are any issues about non-performing members, we will communicate our concerns and reiterate expectations sooner rather than later.
- We will use a Git workflow to aid our progress as a team and help us split up the work.

Signatures:
Christina Wang, Anna Li, Katherine Deng
Date: 10/28/20

Background and Motivation:
As discussed in our initial project abstract, many West Coast states have recently experienced their worst wildfire seasons in modern history, and most are still recovering from the tremendous loss of life, property, and natural resources. These increasingly severe wildfires are front of mind for many individuals right now, on the West Coast and beyond, and have also reignited conversations about the inhabitability of our changing world and the need to take concrete, decisive action on climate change. All three of us are very interested in better understanding these wildfires - their causes, prevalence, spread, impact, and perception. Our ultimate goal is to create greater awareness surrounding the devastation caused by these wildfires and to make their impact concrete and visually stunning to our audience. In doing so, we hope to spark an informed conversation about climate change.

Related Work:
Our initial inspiration for this project was a July 2018 piece by Buzzfeed, titled "How A Booming Population And Climate Change Made California's Wildfires Worse Than Ever." Their code and findings are available for reference in this Github repository. We plan to expand upon this work, by using more recent data, incorporating more advanced and interactive visualizations, and integrating additional sources of data (such as demographic information) to reveal further insights. As an example of one of

these more advanced visualizations, we're inspired by these animated wind movement maps, which, given data availability, we may use to visualize air quality changes in a more dynamic way.

Audience and Questions:

Since wildfires are a topic of broad interest, our primary audience will be the general public. We hope to address the following primary questions:

1. Where are wildfires primarily happening? How do they spread?
2. Has the prevalence and/or severity of wildfires increased over time?
3. What is the seasonality of wildfires? Do they always occur during certain months of the year?
4. What are the origins of these wildfires - are they naturally-occurring or sparked by human interference?
5. Have the primary sources of wildfires changed over time?
6. What are the demographics of those counties most affected by wildfires?
7. How are people on the West Coast being affected (loss of life, respiratory issues, property damage, destruction of natural resources, etc.), and have there been any efforts to provide relief?
8. Have these fires potentially been worsened by climate change?
    a. How do changes in the prevalence of wildfires coincide with droughts, rising temperatures, etc?
9. Has air quality in the burned and surrounding areas been affected?
10. How is the public perceiving the wildfires and their effects? Are peoples' behaviors changing due to the fires and smoke? Are people becoming increasingly concerned about climate change?
11. What is the damage being done by these fires? (evacuations, loss of life, acres burned, etc.)

Expected Data Sources:

Primary

- https://www.kaggle.com/captcalculator/wildfire-exploratory-analysis [compilation of US wildfires, 1992-2015]
- https://github.com/neuroslice/California-Wildfires/blob/master/caFire.json [d3 polygon mapping of CA fires, using geospatial data to retrieve exact fire perimeters]
- https://github.com/BuzzFeedNews/2018-07-wildfire-trends [compilation of CA wildfires through 2018 - data on structural damage and wildfires]
- https://gis.data.ca.gov/datasets/CALFIRE-Forestry::california-fire-perimeters-all/data?geometry=-156.025%2C24.627%2C-82.196%2C48.879 [current and historic CA wildfire perimeters, geographic data]
- https://www.usfa.fema.gov/data/statistics/order_download_data.html#download [building damage]
- https://ca.water.usgs.gov/wildfires/california-wildfire-data.html [current CA wildfires]
- https://github.com/codeforamerica/click_that_hood/blob/master/public/data/california-counties.geojson [CA county perimeters, geojson data for mapping]
- https://www.nifc.gov/fireInfo/fireInfo_stats_totalFires.html [total number of fires and acres burned]
- https://www.epa.gov/outdoor-air-quality-data/download-daily-data [daily outdoor air quality]

- https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-detail.html [CA census data through 2019, includes age/sex/race demographics]

Secondary [time permitting, kept here as a reference]

- https://gis.data.ca.gov/datasets/CAEnergy::california-electric-transmission-lines [electric lines map and data]
- https://www.epa.gov/air-trends/air-quality-cities-and-counties [aggregate outdoor air quality]
- https://www.fire.ca.gov/media/11416/top20_acres.pdf [largest CA fires]
- https://www.fire.ca.gov/media/11417/top20_destruction.pdf [most destructive CA fires]
- https://www.fire.ca.gov/media/5512/top20_deadliest.pdf [deadliest CA fires]
- https://west.stanford.edu/sites/g/files/sbiybj12076/f/blcsurveyoct2019.pdf [public opinion on CA wildfire policies]
- https://www.sciencebase.gov/catalog/item/5ee13de982ce3bd58d7be7e7 [1878-2019 spread, shape, and cause of fire]
- https://data-nifc.opendata.arcgis.com/ [current and historic US wildfire perimeters, ArcGIS data]
- https://data.world/abcotvdata/active-wildfires [active fires]
- https://www.fire.ca.gov/incidents/2020/ [2020 CA fires, data source for scraped Kaggle dataset]
- https://maps.nwcg.gov/sa/#/%3F/%3F/38.7955/-116.1293/5 [Fire Open Source Project Current Fires Map]
- https://www.geomac.gov/GeoMACTransition.shtml [Fire Open Source Project API references and drought data]
- https://www.usgs.gov/products/data-and-tools/real-time-data/wildfire [additional USGS Open Source APIs and Datasets]

Data Cleanup:

- We expect to do minimal data cleaning on our primary datasets, which will provide us with basic information on the wildfires - date, location, acres burned, causes, etc. Since some datasets are outdated, we will have to do some work to extract information for more recent years up to 2020.
- We expect to perform more involved data cleaning when incorporating data on climate, air quality, damage, demographics, and public perception, and joining them with our existing fire datasets.
- Data Expectations: 1) fire locations, sizes, sources (human interaction vs. natural causes), and timeframes; 2) weather/climate data including drought, wind patterns, and air quality; 3) evacuations; 4) damages and impacts (human life, structures, environment); 5) demographics of those most affected; [Time Permitting] 6) public opinion on wildfire policies and climate change; 7) firefighter information; 8) the role of electricity and power companies in causing these fires.

Week 10: Data, Sketches, Decide & Storyboard

Final Datasets:

- [US fires data, 1992-2015]
    - https://www.kaggle.com/captcalculator/wildfire-exploratory-analysis
- [CA fires descriptions, 2013-2020, information on cause, agency, etc.]
    - https://www.kaggle.com/ananthu017/california-wildfire-incidents-20132020
- [CA fire polygons - geospatial data, allows us to retrieve exact fire perimeters for mapping]
    - https://github.com/neuroslice/California-Wildfires/blob/master/caFire.json
- [CA 2010s census demographic data, with information on age/race/sex/income]
    - https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-detail.html
- [US daily air quality data, all major pollutants]
    - https://www.epa.gov/outdoor-air-quality-data/download-daily-data

Note: All datasets linked above are clean and updated. Some straightforward data wrangling will be performed in d3 (for example, joining the CA demographics and fires data, likely using county name as the primary key).

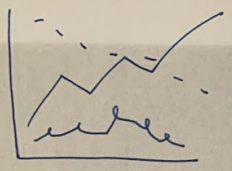
Sketches:

Anna:

Questions Addressed:

A1) What are the origins of these wildfires - are they naturally-occurring or sparked by human interference? Have the primary sources of wildfires changed over time?
A2) Has air quality in the burned and surrounding areas been affected?
A3) What is the damage being done by these fires? (evacuations, loss of life, acres burned, etc.)
A4) Where are wildfires primarily happening? How do they spread?
A5) Has the prevalence and/or severity of wildfires increased over time?
A6) What are the effects/damages of the fires?
A7) Have the primary sources of wildfires changed over time? Has the severity of wildfires increased over time?
A8) What is the seasonality of wildfires? Do they always occur during certain months of the year?

① Fatalities per year

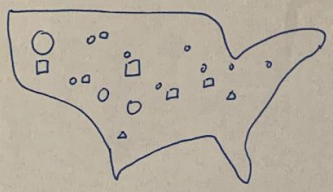


— = ~~fire~~ lightning

- - = arson

ℓ = campfire

② Level of Pollutants 2019

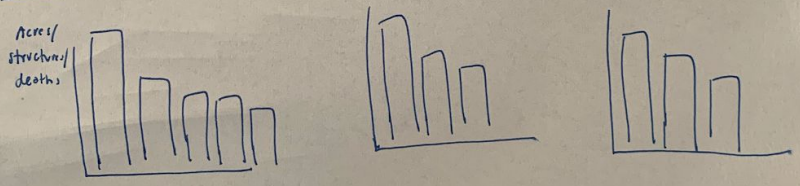


□ = CO

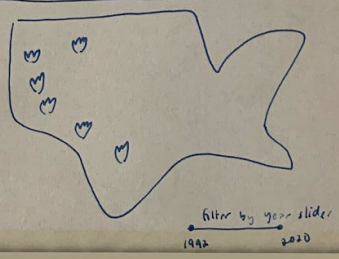
NO_2 = ○

△ = $PM_{2.5}$

③ Largest Cali wildfire → Most Destructive → Most Deadly



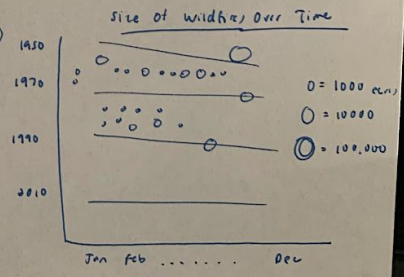
④ Wildfires 1992 - 2020



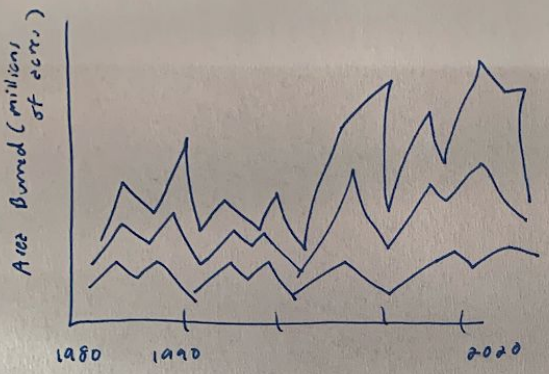
filter by year slider

1992 2020

⑤ Size of Wildfires Over Time

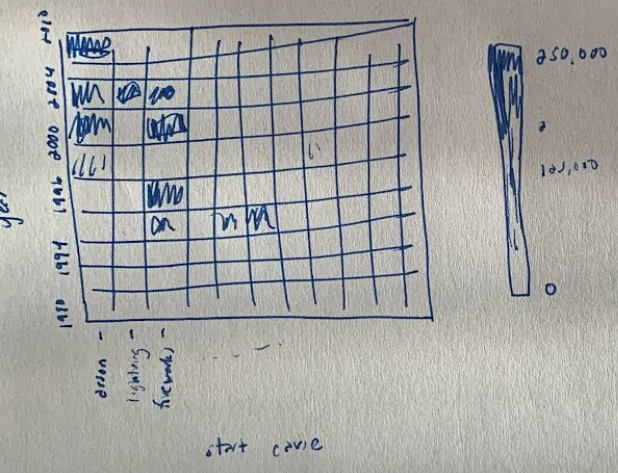


⑥ Damage caused by wildfires

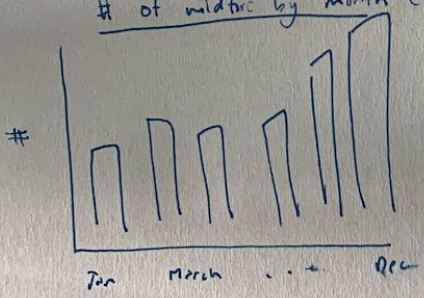


each layer can represent
- cause of fire
- severity of fire
- location of fire

⑦ heatmap of # of fires by ~~dance~~ start cause



⑧ # of wildfire by month (or year)



Katherine:

Questions Addressed:

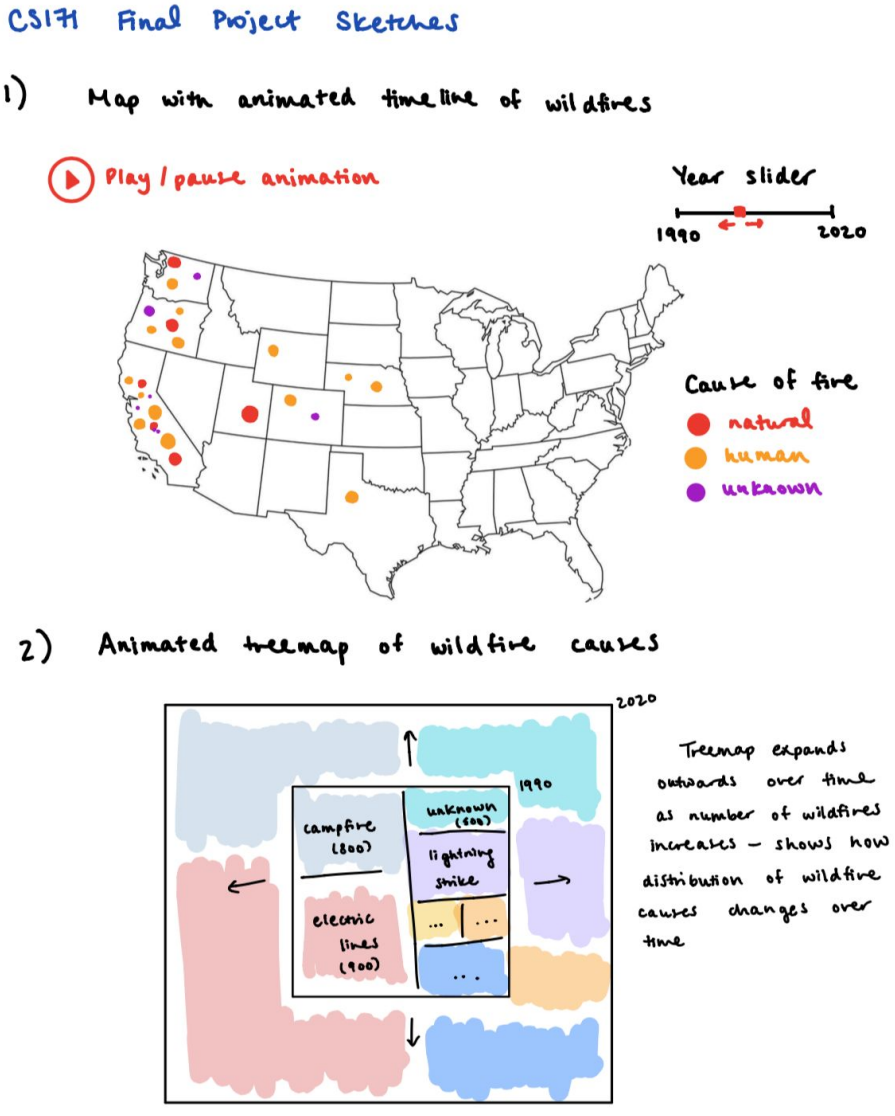
K1) Where are wildfires primarily happening? Has the prevalence and/or severity of wildfires increased over time? Have the primary sources of wildfires changed over time?

K2) Has the prevalence of wildfires increased over time? Have the primary sources of wildfires changed over time?

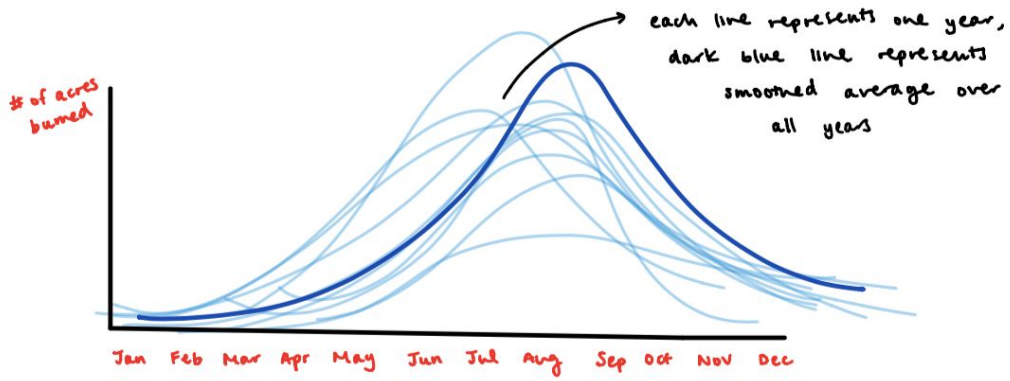
K3) What is the seasonality of wildfires? Do they always occur during certain months of the year?

K4) What are the demographics of those counties most affected by wildfires? Are they disproportionately affecting certain groups of people?

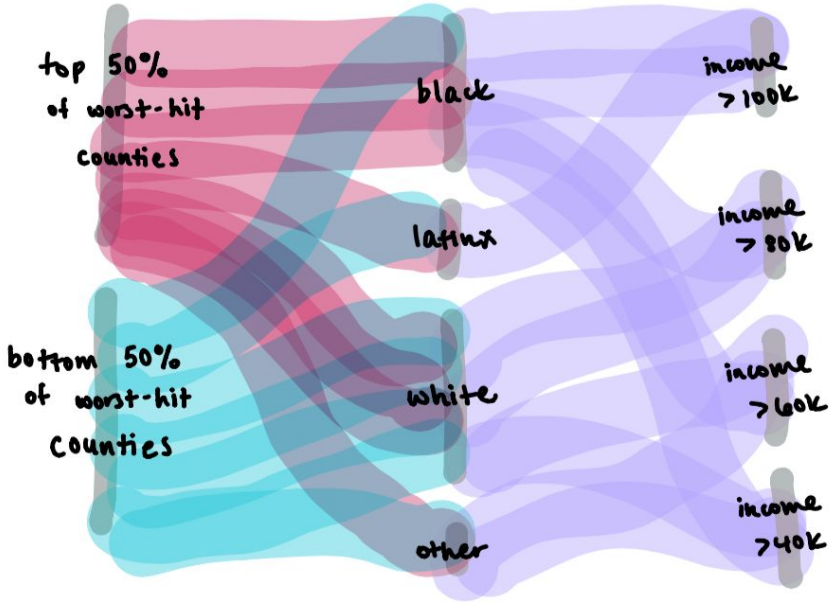
K5) Has the prevalence of wildfires increased over time? What are the origins of wildfires and have they changed over time?



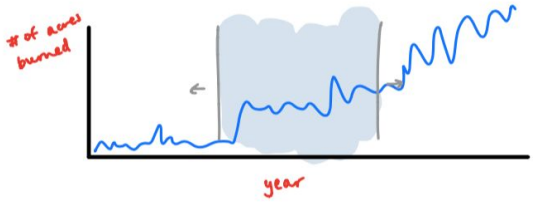
3) Wildfire seasonality "probability distribution"

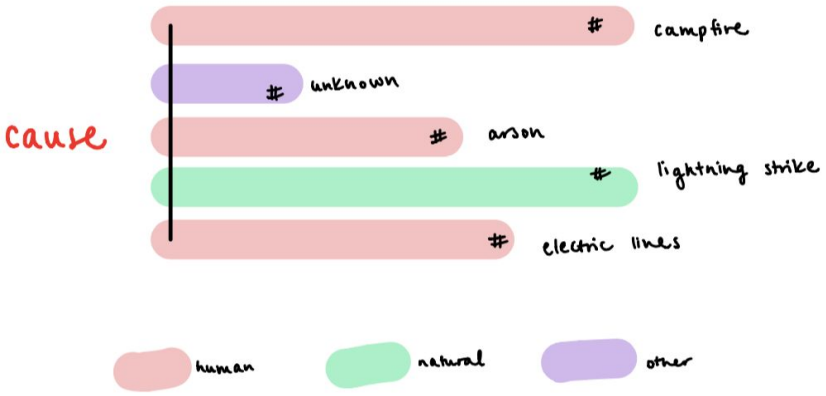


4) Demographics most affected by wildfires, using "parallel sets"

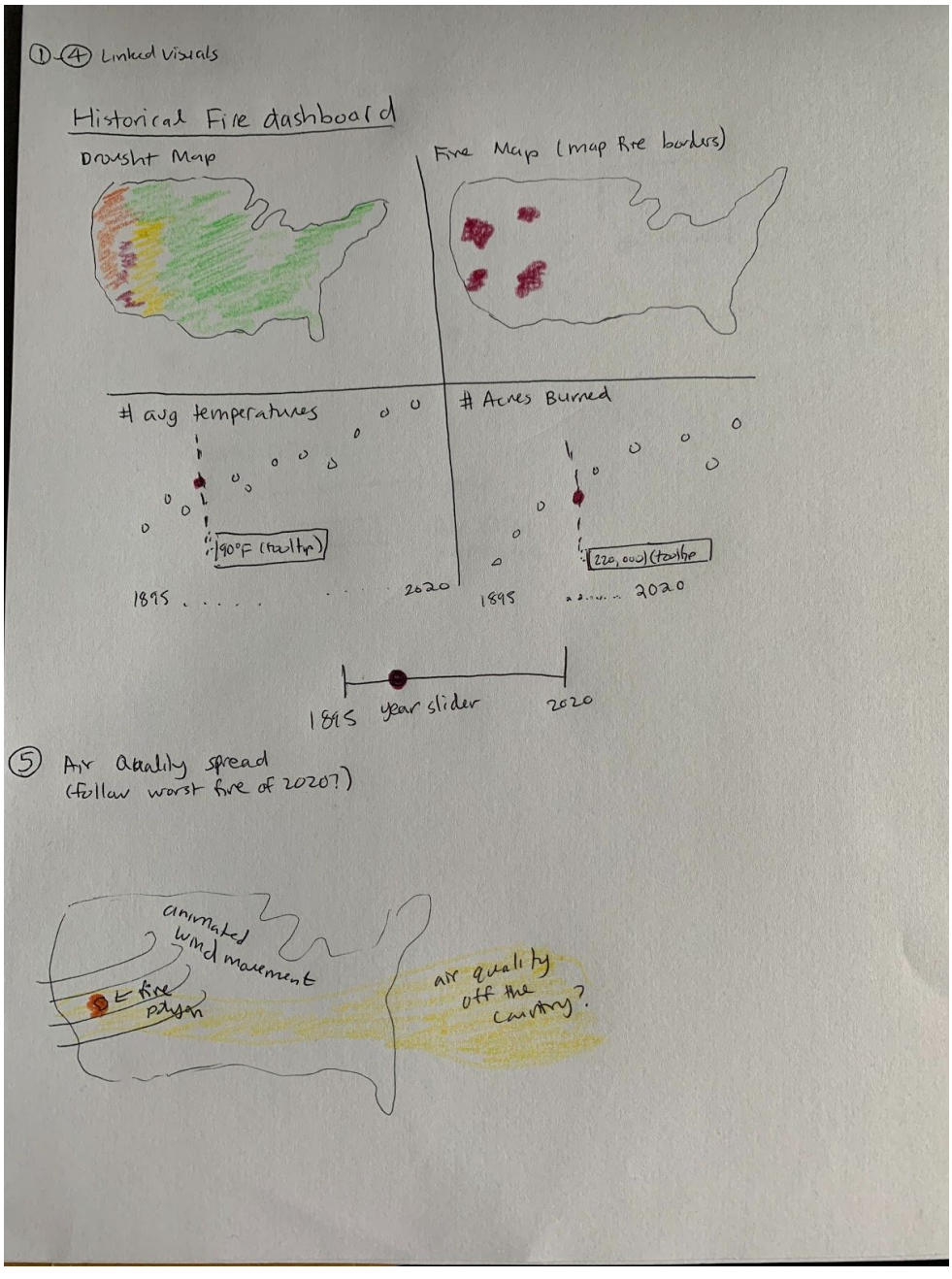


5) Brush and context





Christina:



-

Questions Addressed:

C1)

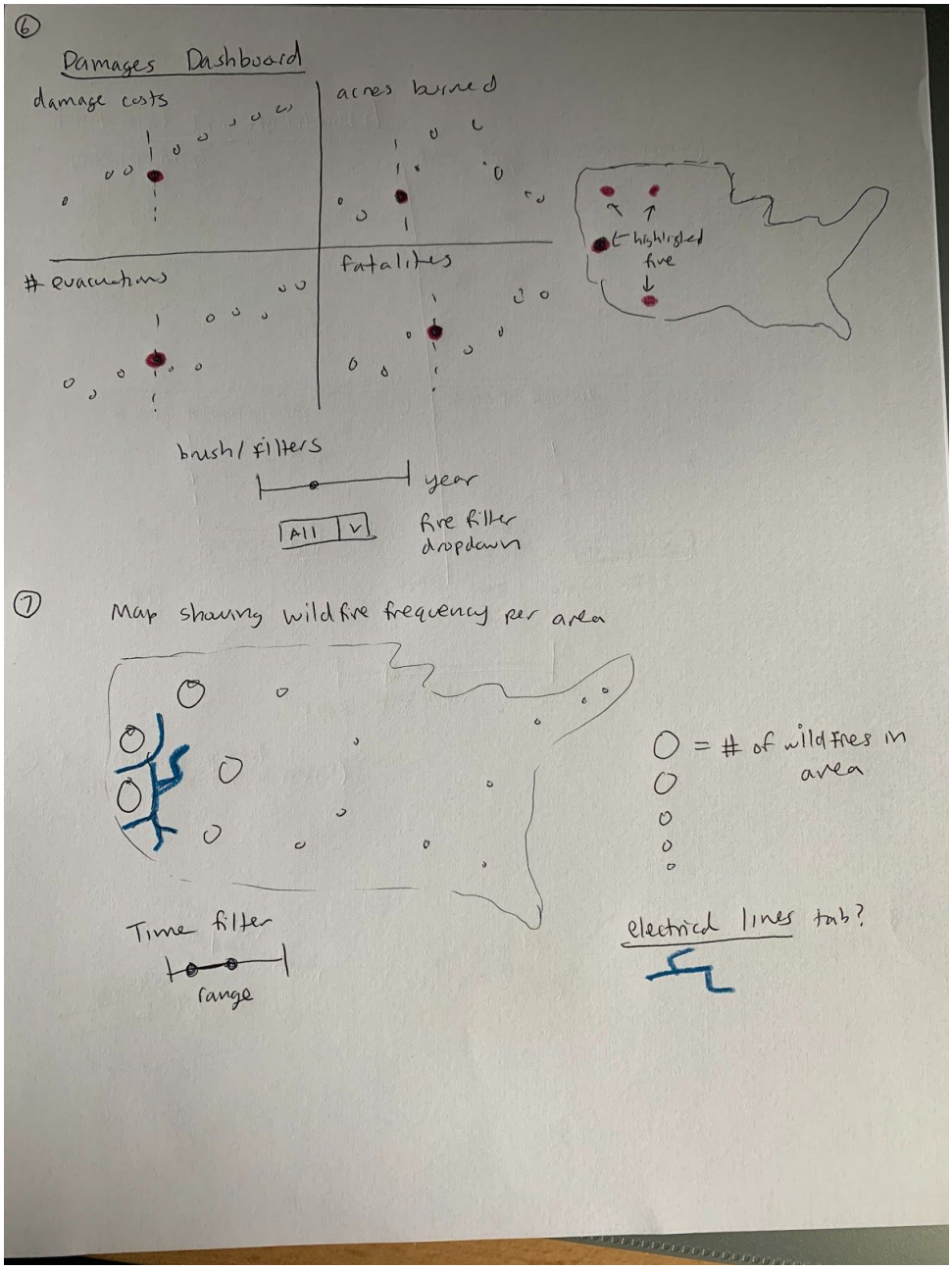
How has climate change affected fires over time?

- Question IDs Answered: 1, 2, 8, 11

C2)

How far of an effect do wildfires have? What are their human impacts?

- Wildfires spread when wind is above 30mph. Show how much a fire grows with wind.
- Show how air quality can be affected even beyond the country (off the map).
- Question IDs Answered: 1, 9



C3) How have wildfire damage trends changed over time?
- Show tooltip-linked damage costs, acres burned, # evacuations, fatalities, and wildfires mapping.
- Question IDs Answered: 7, 11

C4) How have electrical lines affected fires? Do the same counties get burned every year or different ones? Is the range growing?
- This can probably be separated into two different visuals: one with the electrical line map and an overlay of fires caused by electrical lines, and another with a symbol map of absolute total number of wildfires in area.
- Question IDs Answered: 1, 2, 4

Decide:

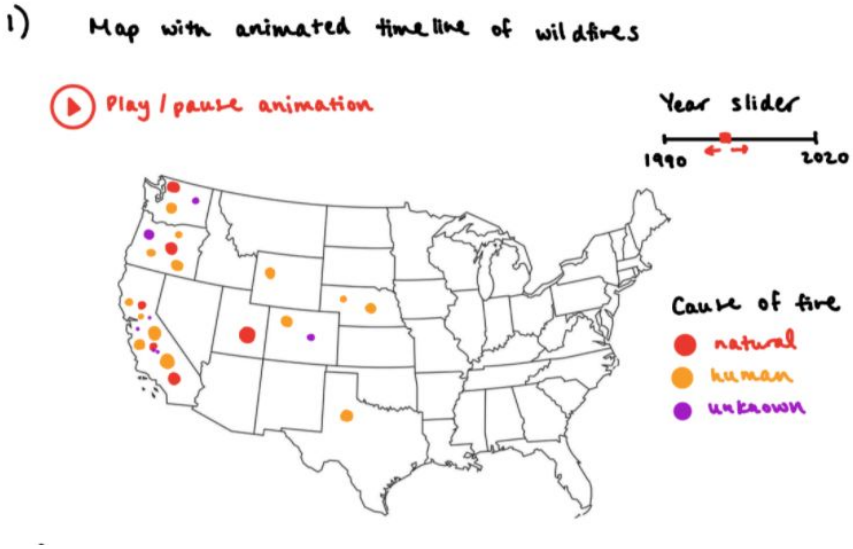
Sketch ID	Question ID	Author
K1, A4	1, 2, 4, 5	KD, AL
K2, K5	4, 5	KD
K3, A8	3	KD, AL
K4	6	KD
A1, A3, A7	4, 5, 7	AL
A2	9	AL
A5	2, 3	AL
A6	7	AL
C1	1, 2, 8, 11	CW
C2	1, 9	CW
C3	7, 11	CW
C4	1, 2, 4	CW

Voting Results:

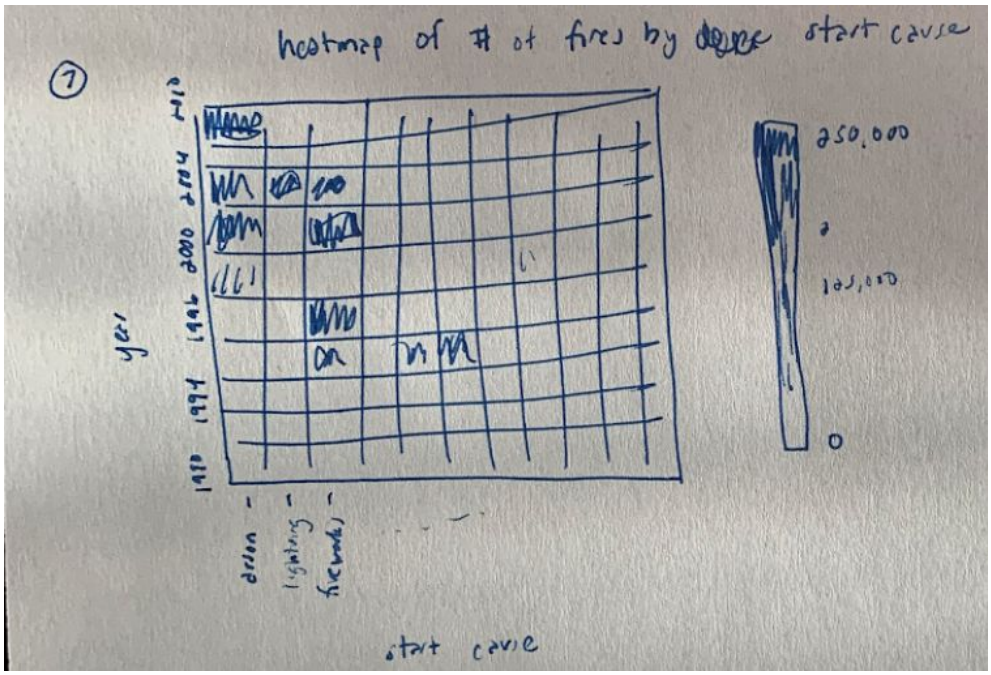
K1: 2
K2: 1
K4: 3
K5: 1
A7: 2
C1: 2
C2: 1
C3: 1
C4: 1

Winning Sketches:

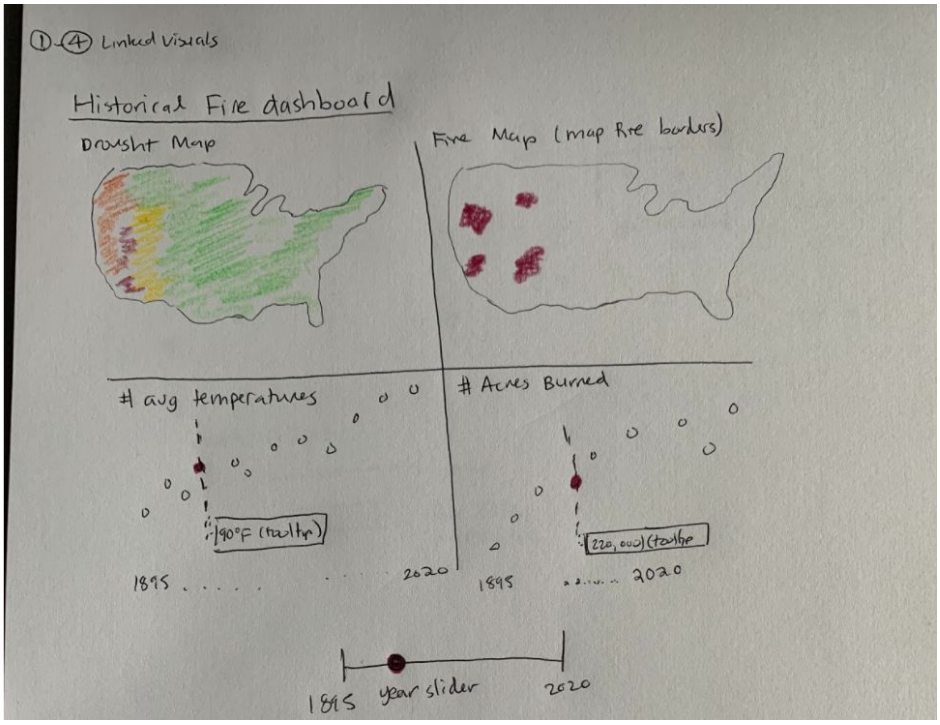
K1) Where are wildfires primarily happening? Has the prevalence and/or severity of wildfires increased over time? Have the primary sources of wildfires changed over time?



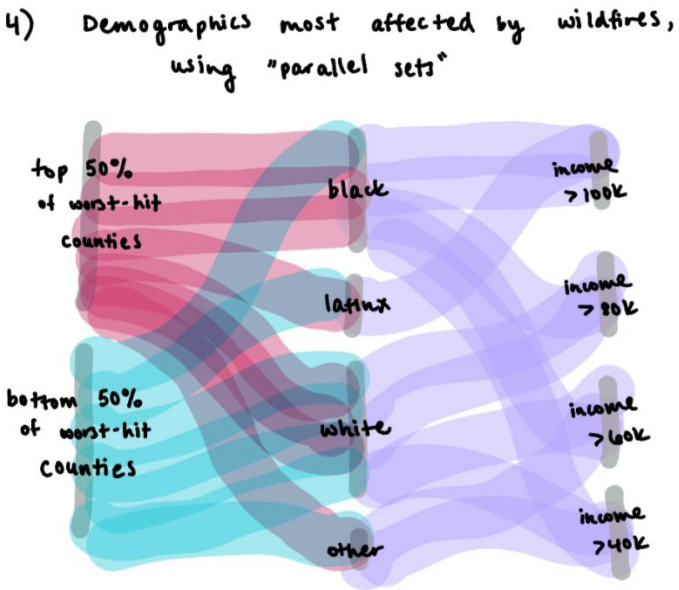
A7) Have the primary sources of wildfires changed over time? Has the severity of wildfires increased over time?



C1) How has climate change affected fires over time?



K4) What are the demographics of those counties most affected by wildfires? Are they disproportionately affecting certain groups of people?



Justification Paragraphs:
We chose Sketch **K1** because we thought it provided a comprehensive introduction to the data that would serve as a good starting point for our data story. It allows viewers to see an animated timeline map of where wildfires are occurring over time, and observe general trends in the prevalence, location, size, and cause of wildfires. In particular, it highlights the fact that wildfires pose the greatest threat to the West Coast of the United States, and that wildfires have become more common and more severe in recent years.

We chose Sketch **C1** because it addresses a key story point on climate change and how it may be linked to wildfire trends. It allows one to see correlations between drought, average temperatures, fires, and acres burned. In particular, the time slider allows one to see changes over time with each of the visuals in the panel having linked tooltips. It highlights how fires, damages, temperatures, and drought levels have all increased over time.

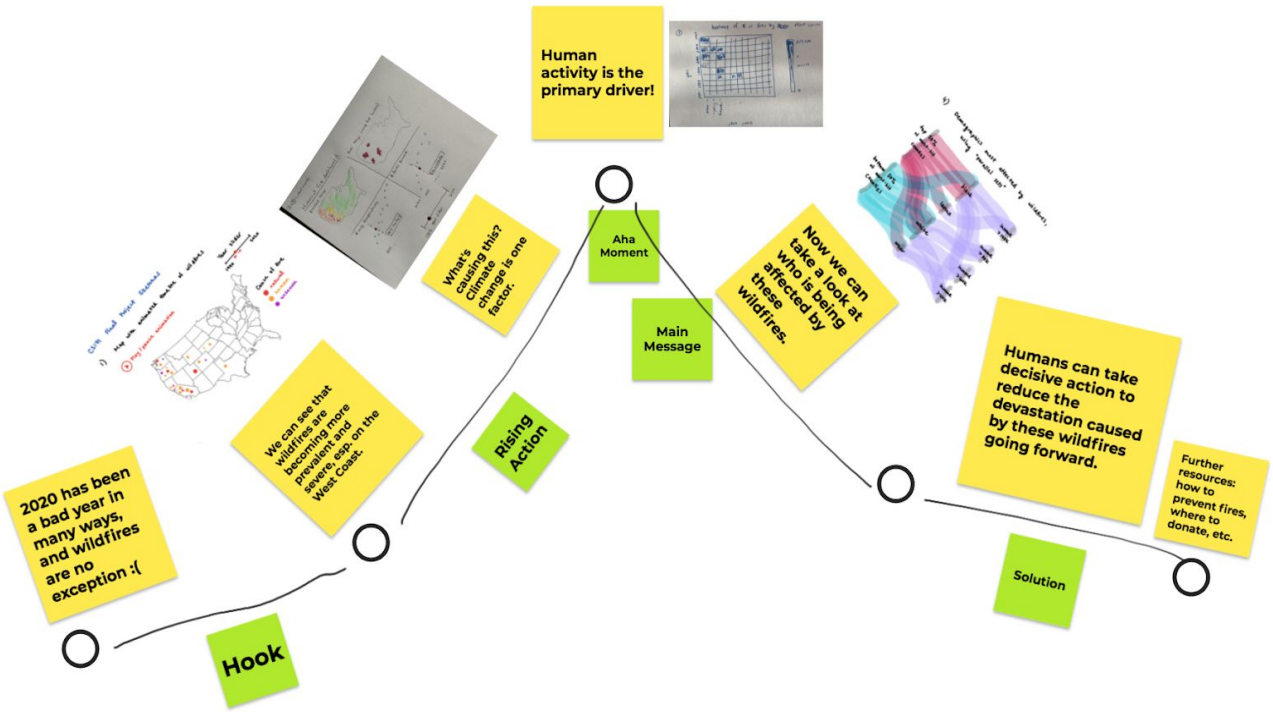
We chose Sketch **A7** because we thought that a heatmap would be an appropriate and effective way to analyze causes of wildfires and how they have changed over time. In addition, we could use different colors/shades to overlay the severity of wildfires. This would allow us to look at how much damage these wildfires have caused over time, and how this might relate to the causes of the wildfires.

We chose Sketch **K4** because a key question we are hoping to answer with this project is which demographic groups are most impacted by wildfires. We thought a parallel sets / Sankey diagram-like design would be a clean and intuitive way to visualize these flows and quickly see whether flows are going disproportionately to one demographic group over another.

Storyboard:

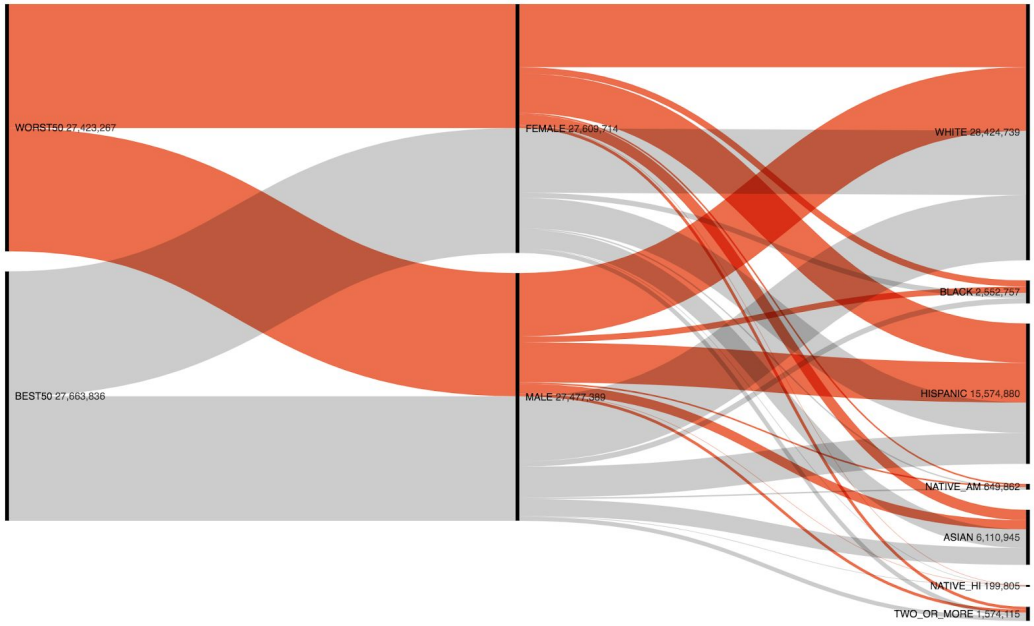
Main Message:
Human activity is the primary driver of the more prevalent and severe wildfires that we have witnessed in recent years. This insight suggests that these wildfires are not necessarily inevitable, and that decisive action by humans can help reduce the devastation caused by these wildfires going forward.

Jamboard:
https://jamboard.google.com/d/1fd56SDZSQoEPSjvWoOdASkWVU3mZUXXKdJnLIct50MQ/edit?usp=sharing


2020 has been a bad year in many ways, and wildfires are no exception :(
Hook
We can see that wildfires are becoming more prevalent and severe, esp on the West Coast.
Rising Action
What's causing this? Climate change is one factor.
Human activity is the primary driver!
Aha Moment
Main Message
Now we can take a look at who is being affected by these wildfires.
Humans can take decisive action to reduce the devastation caused by these wildfires going forward.
Solution
Further resources: how to prevent fires, where to donate, etc.

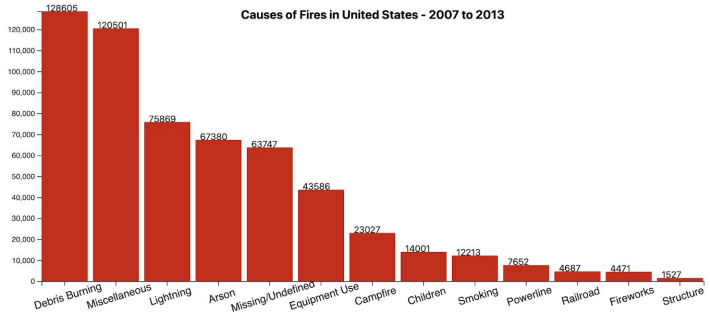
Week 11: Prototype V1

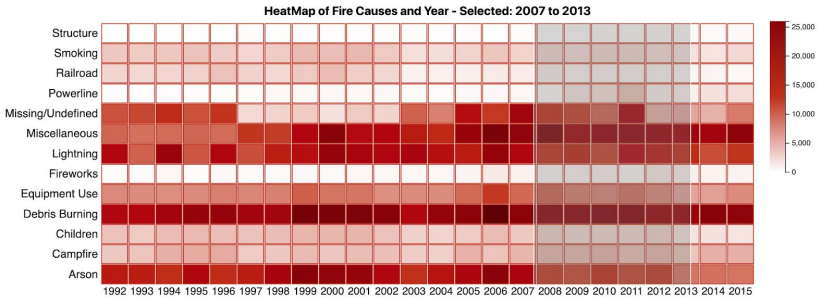
- Name of students who worked on Prototype V1 submission:
  - Katherine Deng, Anna Li, Christina Wang
- Data scraping and cleaning complete (using the real data sets):
  - Data scraping and cleaning completed - the final data sets are included in the /data folder in our submission.
- At least one D3 visualization already partly implemented, and drafts for 2 more visualizations:
  - We've partly implemented three of the four D3 visualizations in our storyboarding step, which you can see when running our code on Webstorm. The remaining visualization is included as an image in our index.html file.
- Rough webpage design and structure has to be done and implemented (placeholders for visualizations, text, and images allowed):
  - We have a rough design for our webpage, where we start with a dashboard showing a general animated overview of the wildfires. Then we dive more into the causes of the wildfires (human vs. nature, and connections to climate factors), and we conclude by analyzing the demographics of those most affected by the wildfires.
- Storytelling is clear:
  - Hopefully! Would love any feedback :D
- The first design of an innovative view:



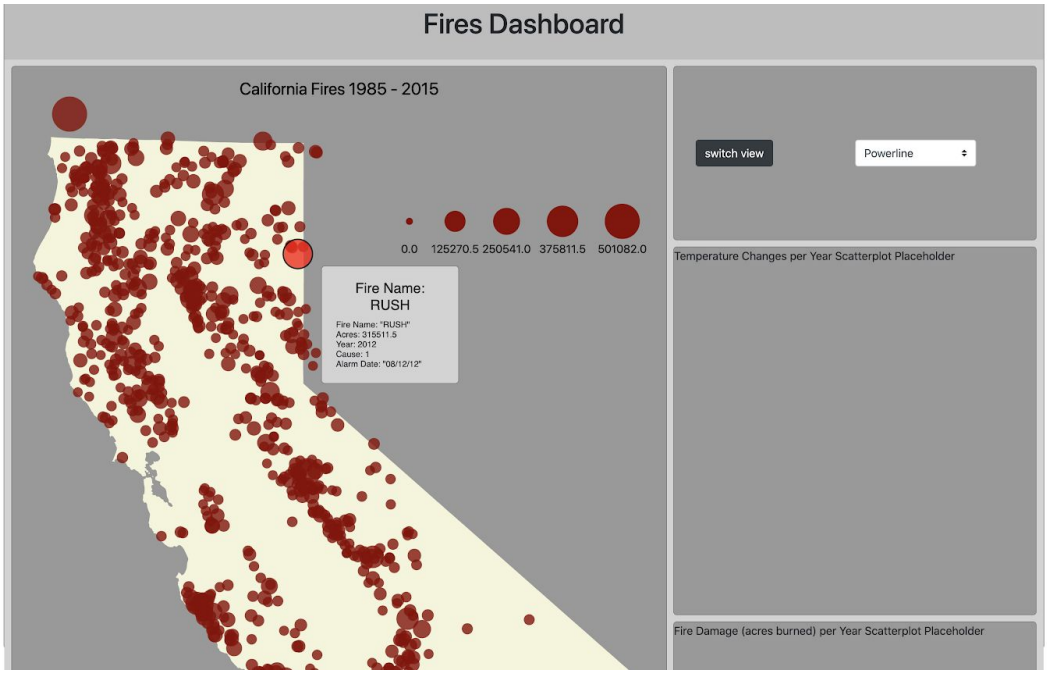
  - We've partly implemented our innovative view, the parallel sets diagram, which illustrates which demographics are most impacted by the fires. The first column

represents the total populations of the top 50% most- and least- affected counties (in terms of acres burned) in California. The second column shows how these counties' populations flow into male and female splits, and the third column shows how these flows further split into the dominant racial groups.
    - In the next round, we're hoping to swap the race and sex columns, since the male/female split is nearly 50-50 and relatively less informative. This will involve a few quick adjustments to the R script to change how the data is aggregated during wrangling.
    - To make this visualization more advanced and innovative, we're planning to add more interactivity in the next round of prototyping. One feature we'll add is allowing users to hover over a particular bar to highlight its flows across the entire graph. Another feature we're hoping to add is informative tooltips with county-level demographic information. For example, if a user hovers over the orange part of the "WHITE" bar, we would like for a tooltip to appear with text on the top three counties with the largest white populations, and how many acres were burned in each of those three counties.
- Interactions (e.g., filtering, brushing, etc.) have to be designed (at least in a textual description and some sketches):





    - Here is an example of brushing, where we can choose which years we want to include in looking at wildfire causes.



- Here is an example of filtering. On the top right, we can choose to filter out fires by causes such as powerlines, campfires, arson, etc. Additional filtering we plan to add is a time slider by year, and also the option to switch this map view into the fire polygons (the actual geographical perimeters of each fire).
- This graphic also features the tooltip interaction of hovering over each point to get more information on each fire. In the next round I hope to convert the fire cause descriptions from its numeric categories to actual descriptions such as 'Powerline'.
- Finally, there are several other placeholder boxes with descriptions inside. The full notes are in the GitHub repo and V1 demo page. Essentially, there will be 3 more linked plots: 1) a scatterplot of average annual temperature changes, 2) total acres burned each year, and 3) drought changes via a map or scatterplot. These views will all be linked based off of the filter selections and will together have a climate change impact theme.
- Additional interactions we hope to implement are detailed in the "Innovative View" bullet point above.

- Up to date process book:
  - Our process book is up to date :)

Data Modifications Log:

- Katherine:
  - To create the parallel sets visualization, I had to join the California_Fire_Incidents.csv and cc-est2019-alldata-06.csv datasets. Before doing so, I performed some light data cleaning on the census dataset - I combined the census data for all age groups (so we would only consider the

totals for each county), and also removed the word “County” from the county column so that it would join on that column with the CA fire incidents dataset.
    - I also removed some uninformative columns from the CA census dataset (e.g. state ID).
    - To aggregate the data into the canonical format expected for Sankey diagrams, I performed extensive data cleaning and wrangling in a separate R script, called data_cleaning.R, since many of these operations are more readily implemented in R than in Javascript.
        - This script writes to demographics_agg.csv in the /data folder.
- Anna:
    - The US Fires dataset is too large to be read into D3 directly, so I used Python to create smaller datasets that were helpful for the visualizations.
- Christina:
    - The original geoJSON dataset used for mapping all the wildfires within the US was too large and would ‘crash’ in d3. Thus, an API call for receiving parts of the geojson was set-up to help solve this issue, but for the initial draft only California fire data was used.
    - The combined dataset and API call for all agencies, including CalFire as commented below, was taken from https://data-nifc.opendata.arcgis.com/datasets/interagency-fire-perimeter-history-all-years.
    - California, specifically the CalFire agency, has collected more detailed information per wildfire (e.g. cause of fire); thus, it may be used as the final main map instead of all states. https://gis.data.ca.gov/datasets/CALFIRE-Forestry::california-fire-perimeters-all?geometry=162.799%2C24.627%2C-41.020%2C48.879