Project Metadata:

Title: UFO Sightings
Repo: https://github.com/sjakes201/cs571
Jacob Sweet (jdsweet@umass.edu) SPIRE ID: 32789437
David Gerard (dgerard@umass.edu) SPIRE ID: 32677808
Josh Daniel (jtdaniel@umass.edu) SPIRE ID: 33423068

Background and Motivation:

Our motivations stem both from curiosity and our past individual research. UFO sightings, while usually accepted as hoaxes, provide interesting insight into how different population variables correlate with metrics like superstition and belief sets. This means that even under the assumption that there has never been a real UFO sighting, we can still learn interesting things about different populations around the world. Multiple members in our group have also done past research in various geography & population based labs or studies, providing an interest in the nature of this data. We can apply our learned skills to make valuable observations about this dataset.

Project Objectives:

We would like to make observations about how different populations, defined by metrics like size and location, correlate with UFO sightings. We can also look at temporal relationships: how UFO sightings have changed over time. Some of the relationships we plan to learn about are:

1. UFO sightings quantity vs. population per land mass unit
2. UFO sightings quantity over time
3. UFO sightings in different geographical regions, conditioned on time
4. UFO shape distribution vs. geographical regions
5. UFO shape vs. duration of encounter
6. UFO shape vs. geographical region
7. UFO sightings quantity vs. GDP per capita in the region
8. UFO sightings quantity vs. education

There are a couple benefits to making these observations. The first one is that it is just fun data to look at. UFO sightings are usually interesting and good science-fiction content. There are also two more academic angles: from the angle that UFOs do and do not exist. Under the assumption that UFOs do exist, this data provides valuable real observations about UFO sightings. This can help predict future UFO sightings, to learn more about them or prepare for their arrival. On the contrary, under the assumption that UFOs do not exist, this provides valuable relationships that could be extended to other, real phenomena, such as superstition, air quality, and plane sightings.

Data:

From where and how are you collecting your data? Provide a link to your data sources, unless you cannot (e.g., for privacy).

The UFO sighting data originates from The National UFO Reporting Center (https://nuforc.org/), a non-profit Washington State corporation. The organization was founded in 1974 by Bob Gribble to collect and analyze unusual sightings of UFOs worldwide. There exist over 180,000 reported sightings over the past five decades, providing a plethora of data to visualize for this project. The data was cleaned by the NUFORC and published on Kaggle (https://www.kaggle.com/datasets/NUFORC/ufo-sightings), resulting in a total of 80,000 entries. Each data entry represents a single UFO sighting and contains the following information: datetime of sighting, city, state, country, shape, duration, comments, date of entry, and geographic location (longitude and latitude). The significant majority of the data originates from The United States, as well as Canada, Great Britain, and Australia. Since 70,293 entries of the 80,000 entries are from The United States, our data visualization and analysis will just consider The United States. To connect the UFO sightings with external data metrics, we plan to consider county and country-level poverty, population, unemployment, and education-level datasets. The four datasets are all found on the United States Department of Agriculture (USDA) Economic Research Service website (https://www.ers.usda.gov/data-products/county-level-data-sets). These are each csv files that we can directly access and utilize for this project.

Data Processing:

Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?

We do not expect to perform substantial data clean-up but rather intend to focus on integrating external data sets into the UFO sightings data set. The purpose of this is to compare UFO sightings with population, economic, and educational disparities across geographic regions. Thus the economic, education, and other related geographic-wise data will be preprocessed with the UFO data to properly visualize potential relationships between the variables. This will be implemented by reading in csv files for each of the data sets and performing SQL/pandas table join operations to match UFO sighting entries with external data of interest. The datasets will also be utilized to create the visualizations independently and provide context for creating maps and other charts. Some of the quantities we plan to derive from our datasets include sighting density, sighting duration statistics, yearly and seasonal trends in data, and sighting type/shape distributions.

Visualization Design:

How will you display your data? Use the Five Design-Sheet Methodology to generate ideas for your visualization design. Develop three alternative prototype designs for your visualizations (Sheets 2,3,4 from the FdS methodology). Create one final design (Sheet 5 from the FdS

methodology) that incorporates the best features of your three designs.Describe your designs and justify your choice of visual encodings.

Must-Have Features:

Our visualizations with data and appropriate representations:

- D3 generated map of the US with states that are dynamically colored based on variable comparisons. For displaying certain variables such as raw UFO sightings, UFO sightings per capita, etc.
- A graph that compares education (all four levels in the database) to ufo sighting quantity using a visualization that compares ordinal and continuous data
- Trend line that uses position to show UFO sightings over time
- Various visualizations for the rest of our data comparisons that we will determine after exploring D3 more

An intuitive compartmentalized interface, possibly with different tabs / pages for different visualizations

Optional Features:

Dynamic variable comparison

- Compare any variable in the data set to any other variable in the data set
- Requires dynamically choosing visualization formats (histogram vs heatmap vs line graph etc) based on what makes sense for the data
- Reach feature

County-level data

- We found datasets that it exists in but it would be a lot of work to visualize something this granular
- Reach feature

Project Schedule:

Week 4:

- Whole group: finish and submit project proposal

Week 5:

- Jake: Setup ReactJS application
- David: Explore D3 and choose specific visualizations we will use for different diagrams
- Josh: Merge/join and filter through provided datasets

Week 6:

- Jake: Make general ReactJS app compartments
- David: Start to implement individual visualization (each in a React component for dynamic loading)
- Josh: Start to design script that queries for data to be used in frontend

Week 7:

- Jake: Continue implementing individual visualization components
- David: Continue implementing individual visualization components
- Josh: Integrate data querying system into finished visualization components

Week 8:

- Jake: Continue implementing individual visualization components
- David: Continue implementing individual visualization components
- Josh: Integrate data querying system into finished visualization components

Week 9:

- Jake: Integrate individual visualization components into containers
- David: Test data integration with different variables in different visualizations
- Josh: Test data integration with different variables in different visualizations

Week 10:

- Jake: Apply finishing touches to individual graphics (CSS styling)
- Josh: Apply finishing touches to general website division / compartments (CSS styling)
- David: Review project proposal & ensure we met all self-defined requirements and stuck to goal

Week 11:

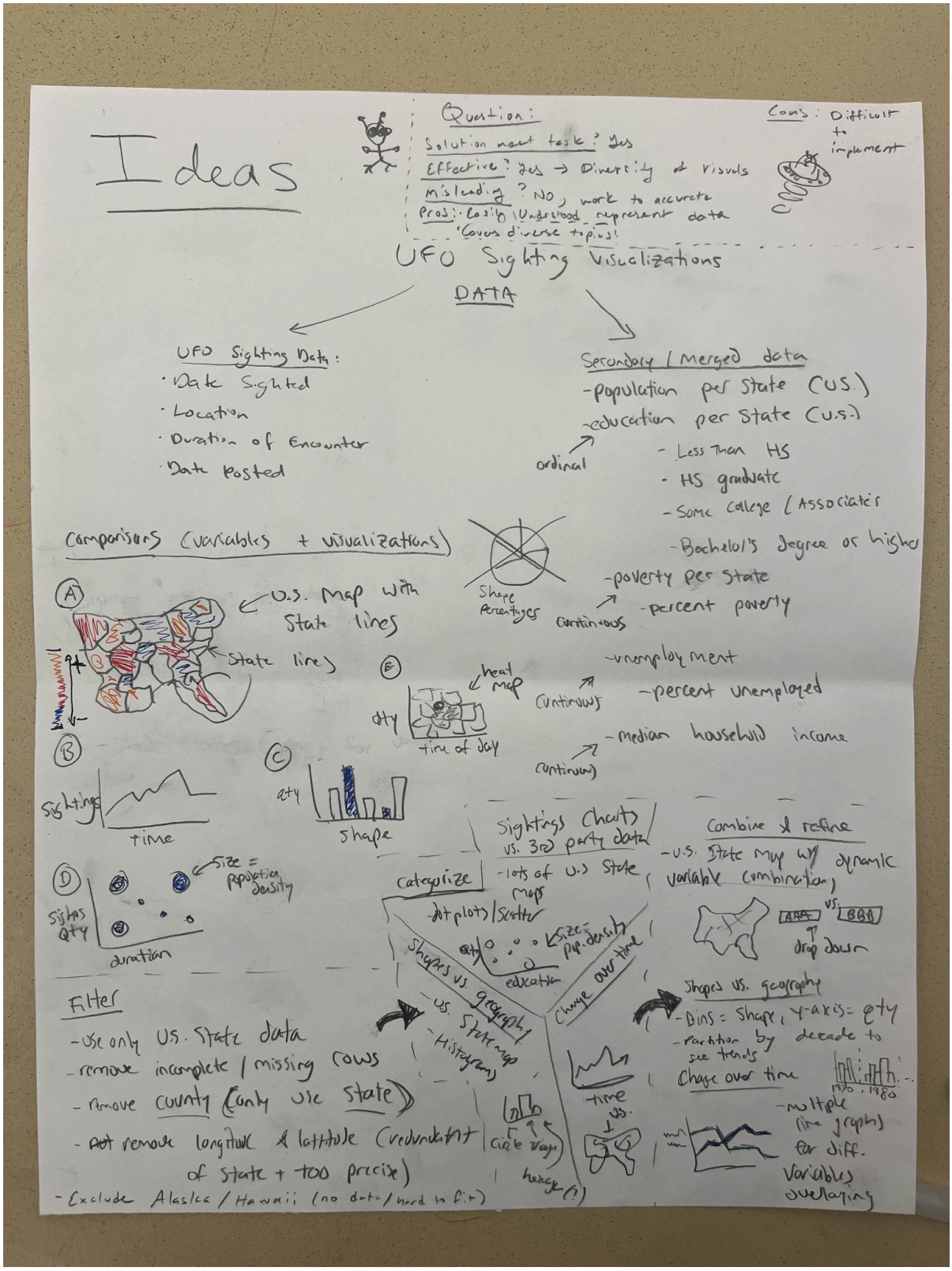
- Whole group: Upload to free hosting application (e.g. MS Azure)

Week 12:

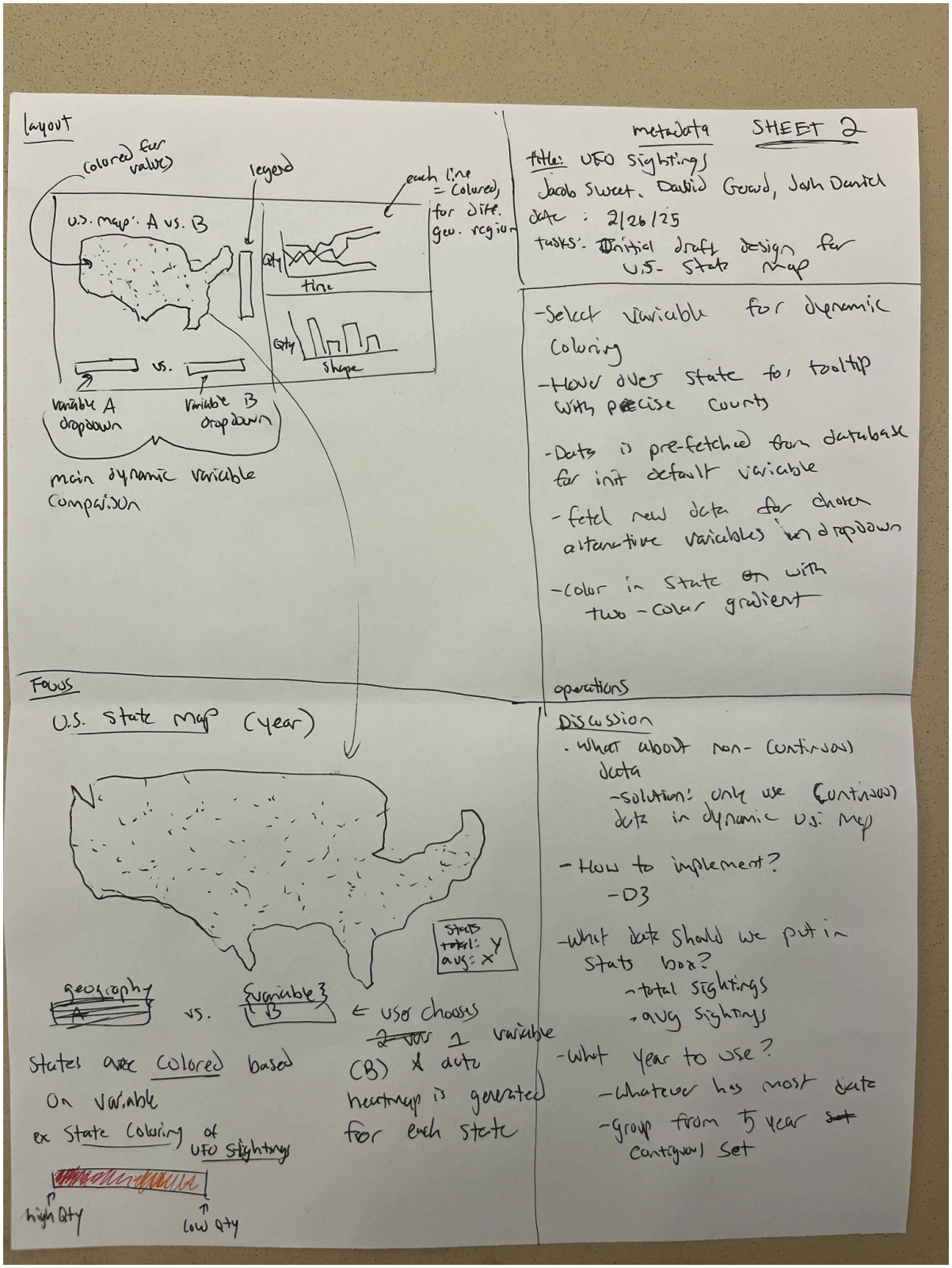
- Jake: test website visualizations and check for robustness and accuracy
- David: Plan and record video
- Josh: Start to plan for presentation

5 sheet methodology sketches:

Sheet 1:



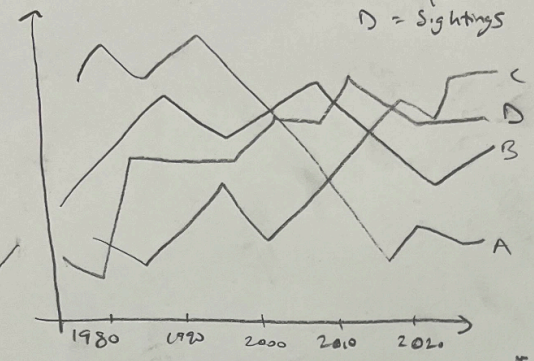
Sheet 2:

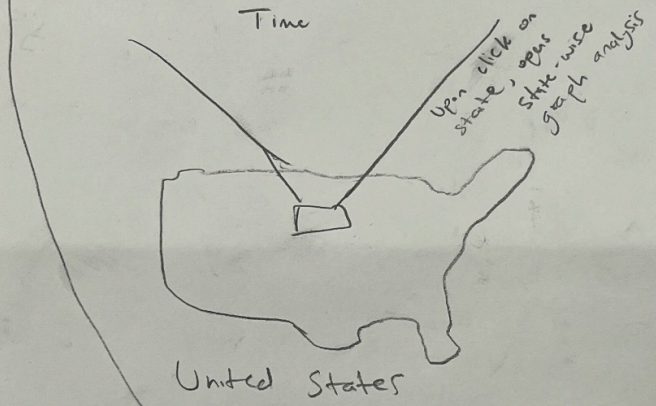

layout
colored for value
legend
each line = colored for diff. geo. region
U.S. map: A vs. B
Qty
time
Qty
shape
vs.
variable A dropdown
variable B dropdown
main dynamic variable comparison
metadata
SHEET 2
title: UFO sightings
Jacob Sweet, David Gerard, Josh Dantch
date: 2/26/25
tasks: Initial draft design for U.S. State map
-Select variable for dynamic coloring
-Hover over state to, tooltip with precise counts
-Data is pre-fetched from database for init default variable
-fetch new data for chosen alternative variables in dropdown
-color in State on with two-color gradient
Focus
U.S. State map (year)
State total: X
avg: Y
geography A
vs.
variable B
← user chooses 1 variable
States are colored based on variable
ex State Coloring of UFO Sightings
high Qty
low Qty
(B) X data heatmap is generated for each state
operations
Discussion
- what about non-continuous data
-solution: only use continuous data in dynamic U.S. Map
- How to implement?
-D3
-What data should we put in State box?
-total sightings
-avg sightings
-What year to use?
-whatever has most data
-group from 5 year contiguous set

Sheet 3:

Layout

A = Income
B = Education
C = Population
D = Sightings





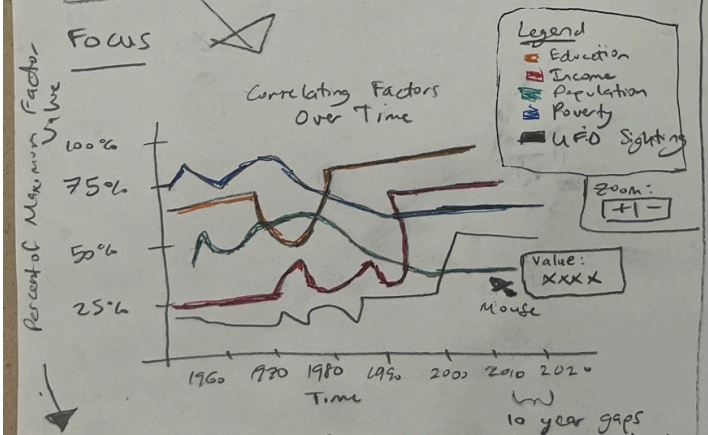
Metadata

Title: UFO Sighting Trends
Names: David Gerard, Jacob Sweet, Josh Daniel
Date: Feb. 27, 2025
Tasks: Analysis of UFO data over time of the last few decades
Sheet: 3

Operations:

- Click on a state on the map → opens data in main chart for just that state
- No state selected / deselect state → Full US trends (all states)
- Zoom into chart → finer grained details
- Hover over chart lines → display actual number at the hover point
- When state selected → changes website data selection

Focus



To properly compare the factor which all have different units, the data values are used to derive the following: the percentage value of the maximum feature

value = $\frac{\text{Value}}{\text{max value of column}}$

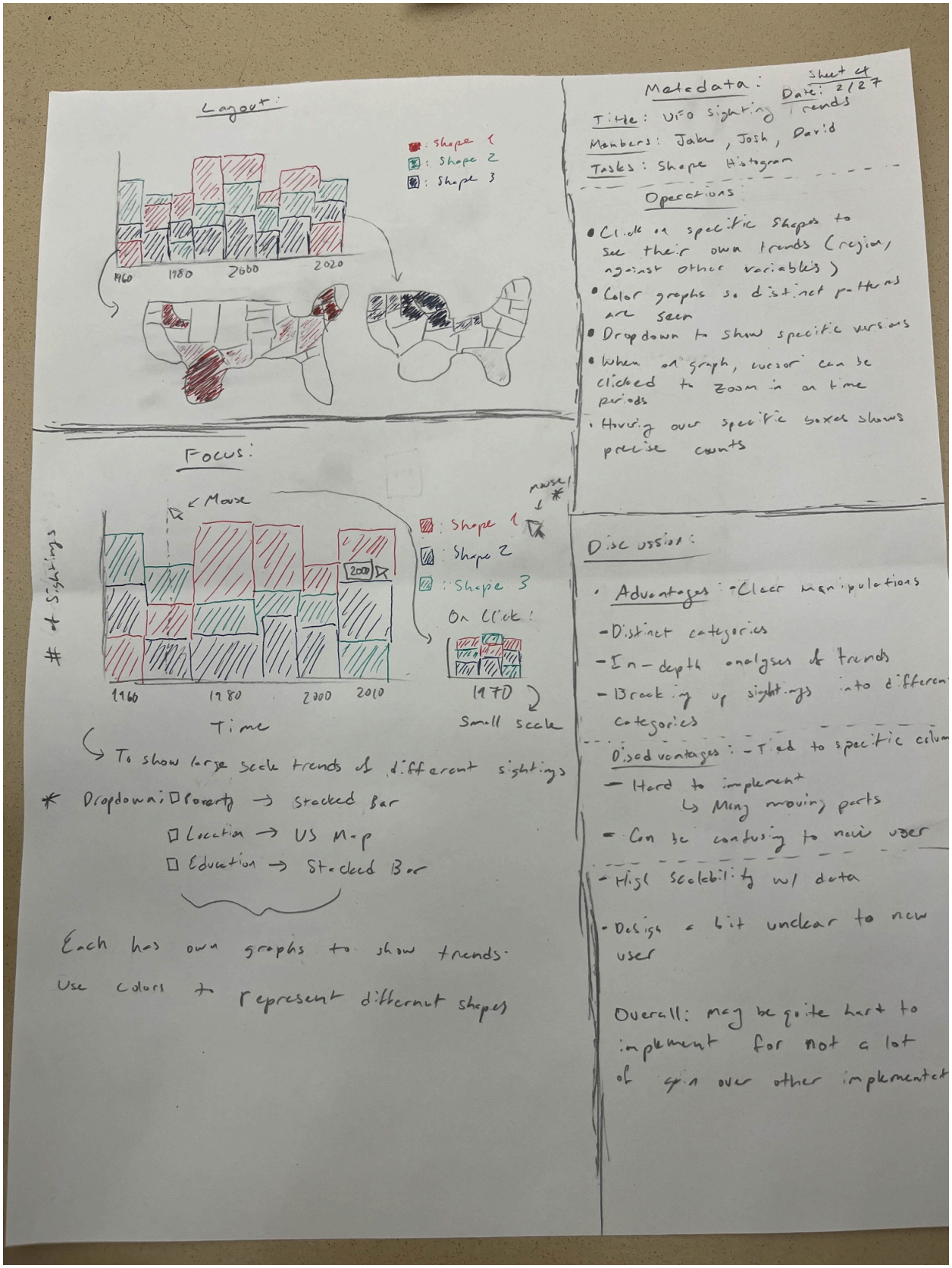
Discussion

Advantages:
- Fine grained detail: the pop-up selected chart allows users to truly look into small details of state-wise data over time.
- Dynamic Actions: User control over what they can/can't see

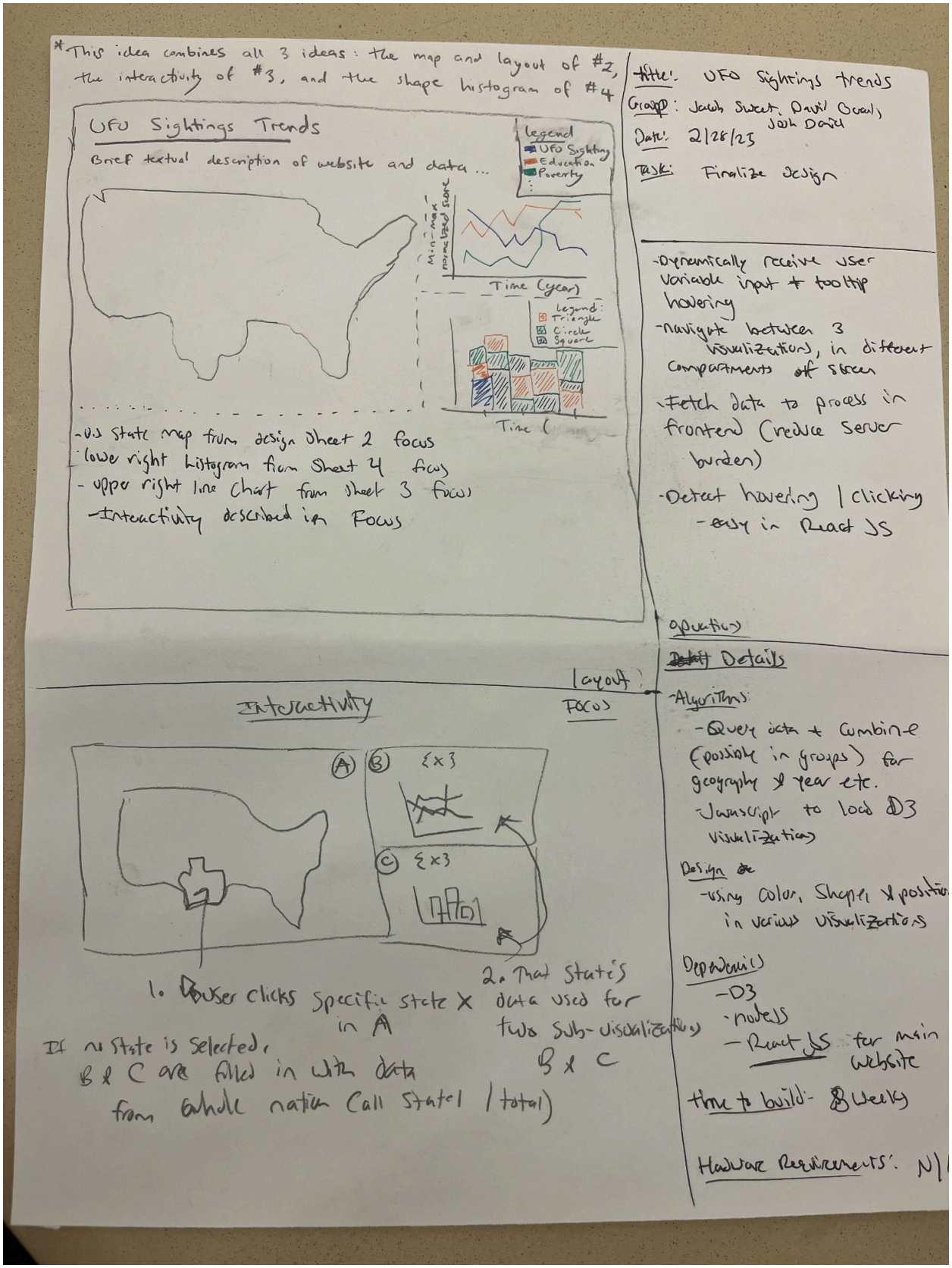
Disadvantages:
- The y-axis of the chart can be unclear and vague, as it only provides relative rather than absolute comparisons.

Overall, this is suitable to visualize our data, but the chart y-axis is somewhat unclear. It is feasible to implement.

Sheet 4:


Layout:
Shape 1
Shape 2
Shape 3
1960
1980
2000
2020
Focus:
Mouse
# of Sightings
Shape 1
Shape 2
Shape 3
On Click:
2000
1960
1980
2000
2010
1970
Time
Small scale
To show large scale trends of different sightings
* Dropdown: Poverty → Stacked Bar
Location → US Map
Education → Stacked Bar
Each has own graphs to show trends.
Use colors to represent different shapes
Metadata:
Sheet 4
Date: 2/27
Title: UFO sightings / trends
Members: Jake, Josh, David
Tasks: Shape Histogram
Operations
• Click on specific shapes to see their own trends (region, against other variables)
• Color graphs so distinct patterns are seen
• Dropdown to show specific versions
• When on graph, cursor can be clicked to zoom in on time periods
• Having over specific boxes shows precise counts
Discussion:
• Advantages: -Clear manipulations
-Distinct categories
-In-depth analyses of trends
-Breaking up sightings into different categories
Disadvantages: -Tied to specific colum
– Hard to implement
↳ Many moving parts
– Can be confusing to new user
-High scalability w/ data
-Design a bit unclear to new user
Overall: may be quite hard to implement for not a lot of gain over other implemented

Sheet 5:


*This idea combines all 3 ideas: the map and layout of #2, the interactivity of #3, and the shape histogram of #4
UFO Sightings Trends
Brief textual description of website and data ...
Legend
UFO Sighting
Education
Poverty
Min-max normalized score
Time (year)
Legend:
Triangle
Circle
Square
Time (
-U.S state map from design sheet 2 focus
-lower right histogram from sheet 4 focus
- upper right line chart from sheet 3 focus
-Interactivity described in Focus
Title: UFO Sightings trends
Group: Jacob Sweet, David Ocon, Josh David
Date: 2/28/23
Task: Finalize design
-Dynamically receive user variable input + tooltip hovering
-navigate between 3 visualizations, in different compartments of screen
-Fetch data to process in frontend (reduce server burden)
-Detect hovering / clicking - easy in React JS
Interactivity
Focus
layout
A
B
{x3}
C
{x3}
1. User clicks specific state X in A
2. That state's data used for two sub-visualizations B & C
If no state is selected, B & C are filled in with data from whole nation (all states / total)
Operations
Details
-Algorithms:
-Query data + combine (possibly in groups) for geography & year etc.
-Javascript to load D3 visualizations
Design of
-using color, shape, & position in various visualizations
Dependencies
-D3
-node.js
-React JS for main website
time to build- 3 weeks
Hardware Requirements: N/