Proposal: Visualization and Analysis of UFO Sightings in United States

Basic info

Project Title

UFO Sighting Data Visualization and Analysis in United States

Team Members

Xue Zou (github ID: xuezzou)

Yunchao Liu (github ID: LanceKnight)

Link to Github

https://github.com/xuezzou/Vis-Project-fall-2018/

Background and Motivation

Whether aliens exist or not has been a lasting debate topic. People from all over the world claimed to spot UFO under various conditions. It is fun to explore UFO sightings since not many conclusions have been drawn regarding UFO sightseeing reports. We wish to discover some interesting facts from the data. Moreover, since this contains geospatial data (longitude and latitude), what we have learnt in class can also play a role in this project.

Inspiration

What areas of the state or country are most likely to have UFO sightings? Are there any trends in UFO sightings over time? Do they tend to be clustered or seasonal? Do clusters of UFO sightings correlate with landmarks, such as airports or government research centers? What are the most common UFO descriptions?

Objectives

We hope to find some interesting facts about UFO sighting. By visualizing where and when reports are filed, we hope to make a model for a trend or pattern for UFO sightseeing.

We intend to learn how to effectively and expressively present geospatial data and draw conclusions from the visual presentation of data.

Data

The data contains over 80, 000 reports of UFO sightings over the last century from 1910 to 2014. Since the reports date back to the early 20th century, some older data might be obscured. Data contains city, state, time, description, and duration of each sighting.

The dataset is originated from The National UFO Reporting Center (NUFORC), a non-profit corporation located in Seattle, Washington, which corroborates and documents from individuals who have been witness to unusual, possibly UFO-related events. Then the data is further scraped, geo-located, and time standardized by Sigmond Axel on his Github.

Data Processing

We decide to work only on U.S data so we need to filter out non-US data. Moreover, Sigmon Axel has already filtered out data that has erroneous or blank time (8.0237%). He also standardized duration time in the unit second. We may further filter out description and geo-position based on our progress.

Must-Have Features

The main visualization would be a U.S. map that has states sequentially colored based on the location and aggregated count of the report. When user click on a state, three additional plots would appear on the right side of the map. The first one is a line plot based on the time and the counts in that states. The second one displays shape and counts in a bar chart, whereas the third one explores the relationship between duration.

Optional Features

First, instead of using color to display the counts of the reports within each state, we could display individual points of each report on the map based on its latitude and longitude with map information of s. Secondly, we may consider more interaction such as adding zooming and brushing for the map. Thirdly, we could allow multiple state selection and then update the data accordingly. For example, when the user shift + click California and Texas on the map. Both of their data would appear in the time plot with two lines, shape plot and duration plot with stacked bar charts. Furthermore, besides U.S., we may also include data from Canada, Germany and Great Britain.

Project Schedule

Nov 5 – 7	Proposal
Nov 7	Proposal Submission
Nov 8 – 11	Data Cleanup
Nov 12 – 20	Prototype Coding
Nov 14	In-class Update 1
Nov 21	Prototype Submission
Nov 22 – 30	Complete Coding
Nov 28	In-class Update 2
Dec 1 – 6	Fix Bugs & Improve the Prototype based on Update 2 Feedbacks
Dec 7	Presentation & Final Presentation

Division of Work

Zou Final Proposal, Five Sheet Design, Updates and Notes, Coding

Liu Proposal Draft, Data Cleanup, Map Coding

Visualization Design

(See below)

Resources & References

Scraped data by Sigmond Axel https://github.com/planetsig/ufo-reports

Kaggle Dataset https://www.kaggle.com/NUFORC/ufo-sightings

NUFORC http://www.nuforc.org/

Some discussion of the dataset

https://www.kaggle.com/tanyavas/ufo-analysis-x-files/notebook

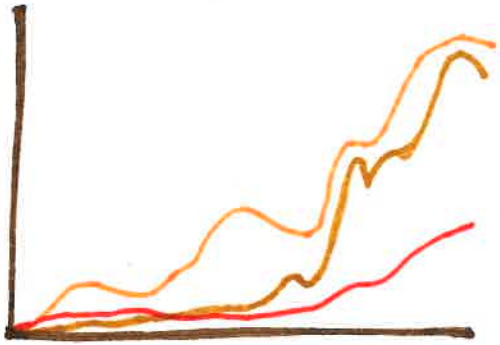
https://www.kaggle.com/abigaillarion/ufo-reports-in-united-states/notebook

https://www.kaggle.com/NUFORC/ufo-sightings/discussion

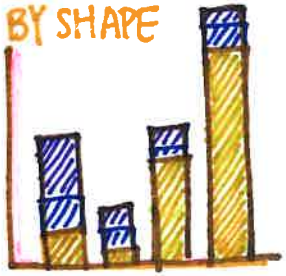
Five Sheet Methodology http://fds.design/index.php/2015/06/25/sheet-2-3-4-initial-designs/

SHEET 1

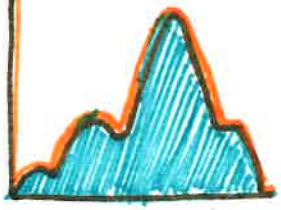
UFO REPORTS BY YEAR



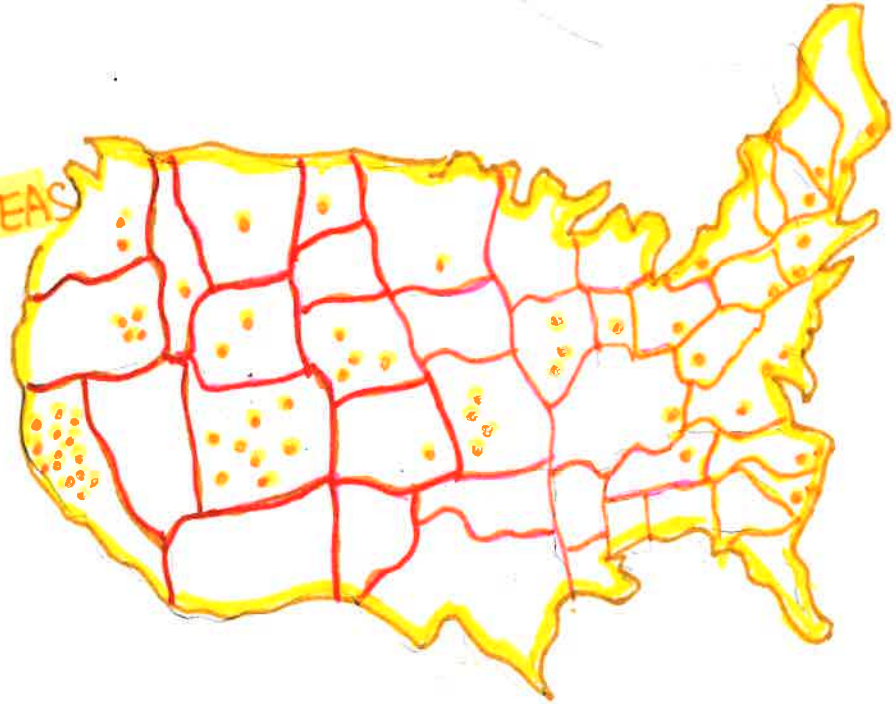
BY SHAPE



DURATION



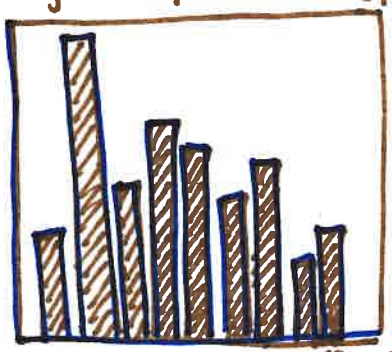
IDEAS

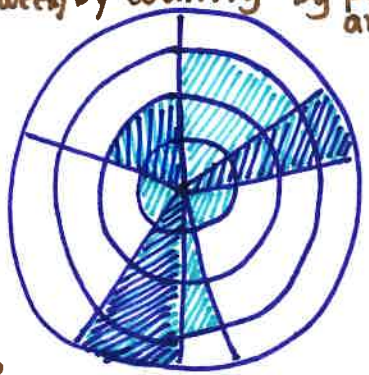


- What areas of the country are most likely to have UFO sightseeings?
- Are there any trends in UFO sightseeings over time? Do they tend to be clustered or seasonal?
- Do clusters of UFO sightseeings correlate with landmarks, such as airports or government research centers?
- What are most common UFO descriptions?

SHEET 2

LAYOUT ☆





FOCUS/ZOOM

shape: shape · size - duration

other variables:

- difference between the sightseeing & the date posted

TITLE

UFO sightseeings in U.S. viewing through geo map

TASK

- show the relationship between geo data and quantity of shape/duration, clustering etc. in U.S. map
- allow interaction to further zoom in the details in particular geo areas

ideas cont.

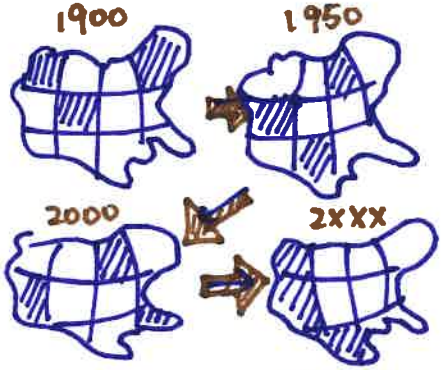
OPERATION

/* ANIMATION! */

- zoom in/out would display details in that particular region based on shape/position/duration in multiple views
- shape of the data point, sequential color-time...

DISCUSSION

- how about oth data attributes like date posted? (date posted - date discovered?)
- clustering makes the individual point hard to observe
- is time good for color mapping?
- the implementation of the animation? (time lapse, color etc.)



ANIMATION ★★★

A changing map based on time line, individual data points distributed based on time

(can also do shape/duration)

/* Animated from the first year to the last year. */

SHEET 3~4

LAYOUT (TEXT)

- wordcloud?

size-frequency
color-?
on a map?
corresponds to individual data point?
mouseover-?
POSITIONING?

OPERATION

- click on a state may zoom out/in more specific details about that state.
- moving the mouse /cursur may provide additional coloring/ highlight of that text

TASK

- to display descriptions of the reports by its corresponding positions
- to identify possible correlation between geo & text.

DISCUSSION

- How to choose these word is difficult.? frequency? figure of speech etc.
- clustering may make the text hard to visualize
- don't expect any correlation between description and other data attributes.
- how to clean up text data?
- how to position it to reduce clustering?

LAYOUT ( ) (lineplot)

OPERATION

- click, expand to its children's lineplot
- shift + click, back to its time plot
- (val would be view in density (/1000km²) eg. or (/100,000 population)

TASK

to identify pattern / correlation. between time and another variable. compare & contrast between different regeion.

DISCUSSION

- how should we assign the color?
- too many children for states in u.s.
  - limit the children's leaf level to county?
- frequency, time duration val as y-axis

SHEET 5

Final (Basic) Visualization

ORERATIONS

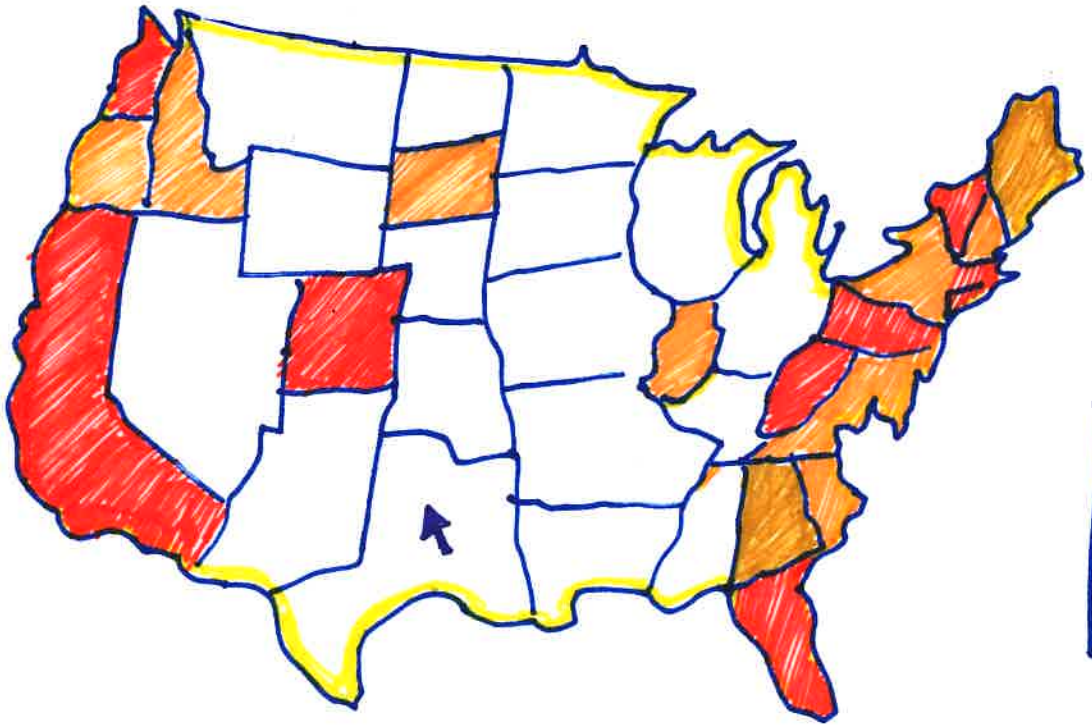
- click a point (a state) then display the three charts on the right side of the U.S. map
- hover through map would highlight the selection.

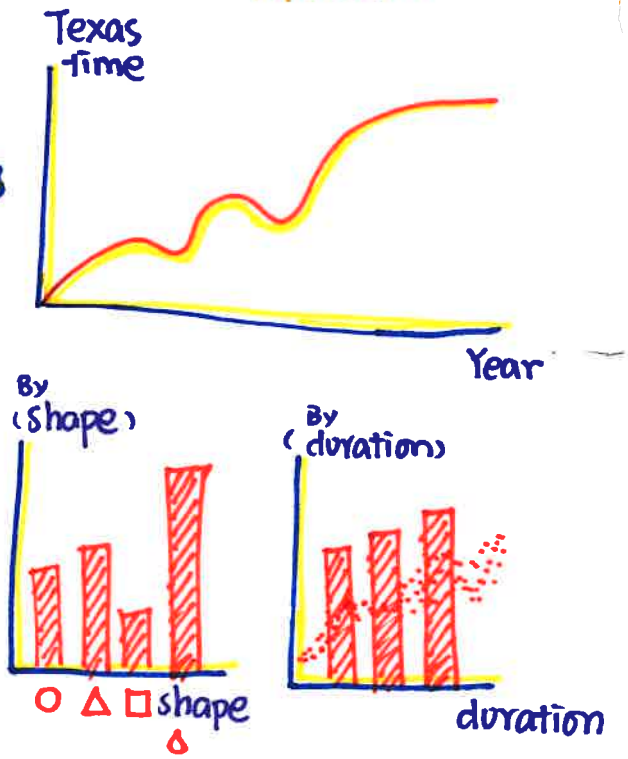
DETAILS

- the scale of the x-axis of duration subgraph would be log scale since most data's duration is short wheras some are really long
- the map would color based on LAB sequentially
- time scale in the time graph
- aggregate data based on state's population/area

LAYOUT

the map has sequential color by the quantity of data in a state (/population or area)





DISCUSSION

- Additional (optional features) /interaction
- zoom (in/out) • animation • display individual data on map
- world map /other countries (Germany, Canada, Great Britain)
- No text analysis • simultaneously click on multiple states
- how to set up the scale of the axes... (duration, log, scale?)
- too many shape categories