**Process Book**

**Exploratory Data Analysis**

What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?

**Design Evolution**

What were the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?

**Analysis**

What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

**Overview and Motivation**

Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.

**Related Work**

We found some d3.js online examples really useful.

Here are the references we used for plotting the map data,

http://bl.ocks.org/michellechandra/0b2ce4923dc9b5809922  
https://bl.ocks.org/mbostock/4090848

https://gist.github.com/NPashaP/a74faf20b492ad377312

For the implementation of the tooltip, we find visualization in this link helpful,

http://bl.ocks.org/NPashaP/a74faf20b492ad377312

The assignment 2 that visualizes different time plots across different music categories and the design choices made using of stacked bar charts in assignment 4 inspires the subplots of our main view.

### Furthermore, here are some discussions of the dataset from Kaggle, which may also influence our design choice:

### 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

For the design and layout of the five Sheet Methodology in our proposal, we use http://fds.design/index for reference.

**Questions**

In our proposal, we outlined four main questions:

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?

Over time, we are more focused on the trend based on year, shape and duration by different states. We end up not process text from description although it’s a great source to visualize and analyze. We also didn’t dig into the seasonality of our data set, which may include interesting insights.

However, to investigate the potential correlation between government landmarks and our data, we introduced new data set consisted of geometric attributes (longitude, latitude) of some randomly chosen U.S. Air Force Bases.

**Data**

**Source:**

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.

**Scraping method**

The main scraping was done by Sigmond Axel on his github project. He took the data from NUFORC, which was further scraped, geo-located, and time standardized. He filtered out data that has erroneous or blank time (8.0237%). He also standardized duration time in the unit second.

**Clean up & Data Processing**

In the Python Script, we translated the states attributes of the data items to full state name instead of abbreviation in order to coordinate with the geo projector. Moreover, we only kept the attribute year (split from time attribute in the original data), state, country, shape, duration, latitude, and longitude. We further filter out items with no state. items without U.S., and items without shape specified.

**Links**

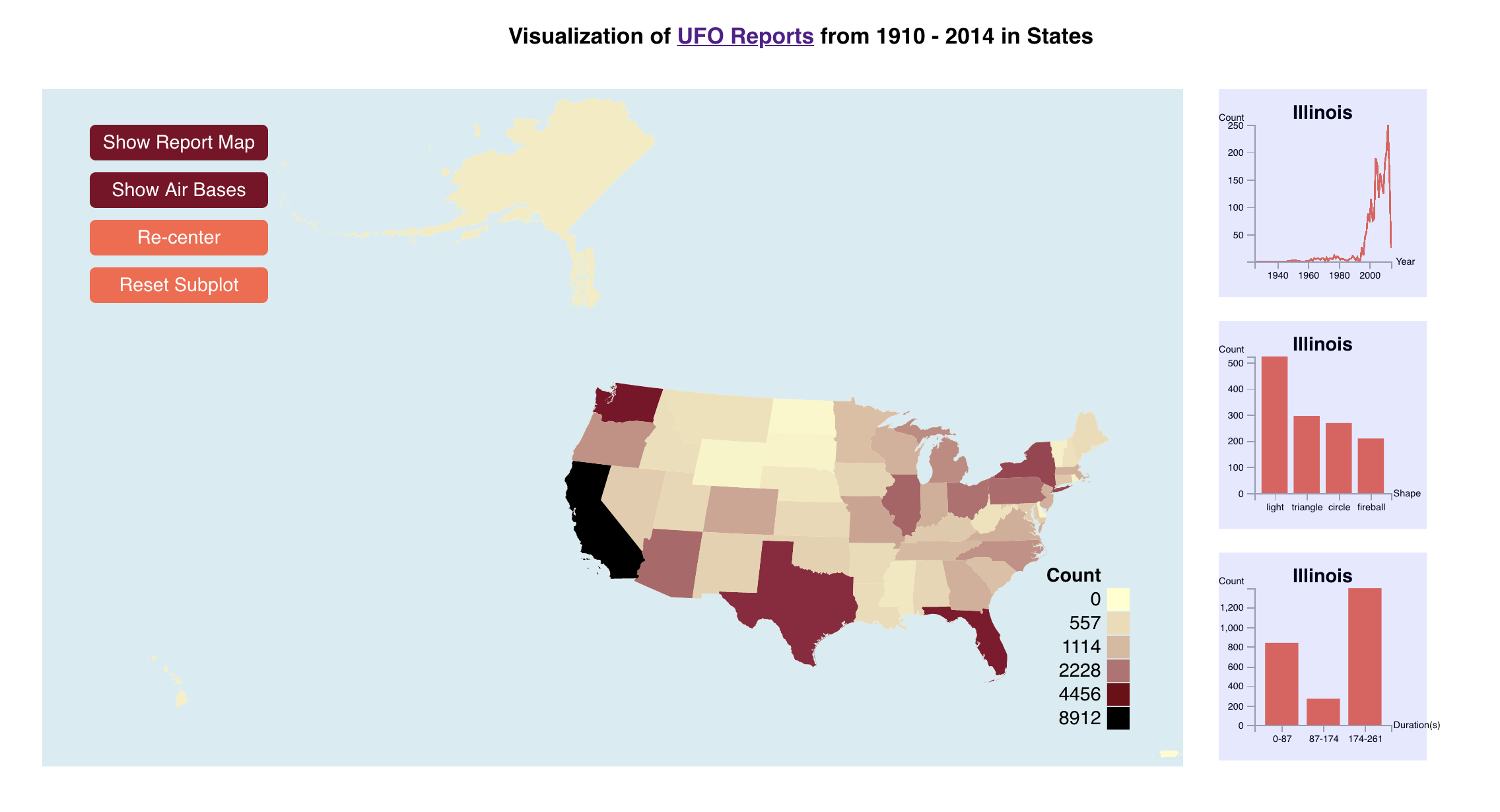
Scraped data by Sigmond Axel https://github.com/planetsig/ufo-reports

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

NUFORC http://www.nuforc.org/

**Implementation**

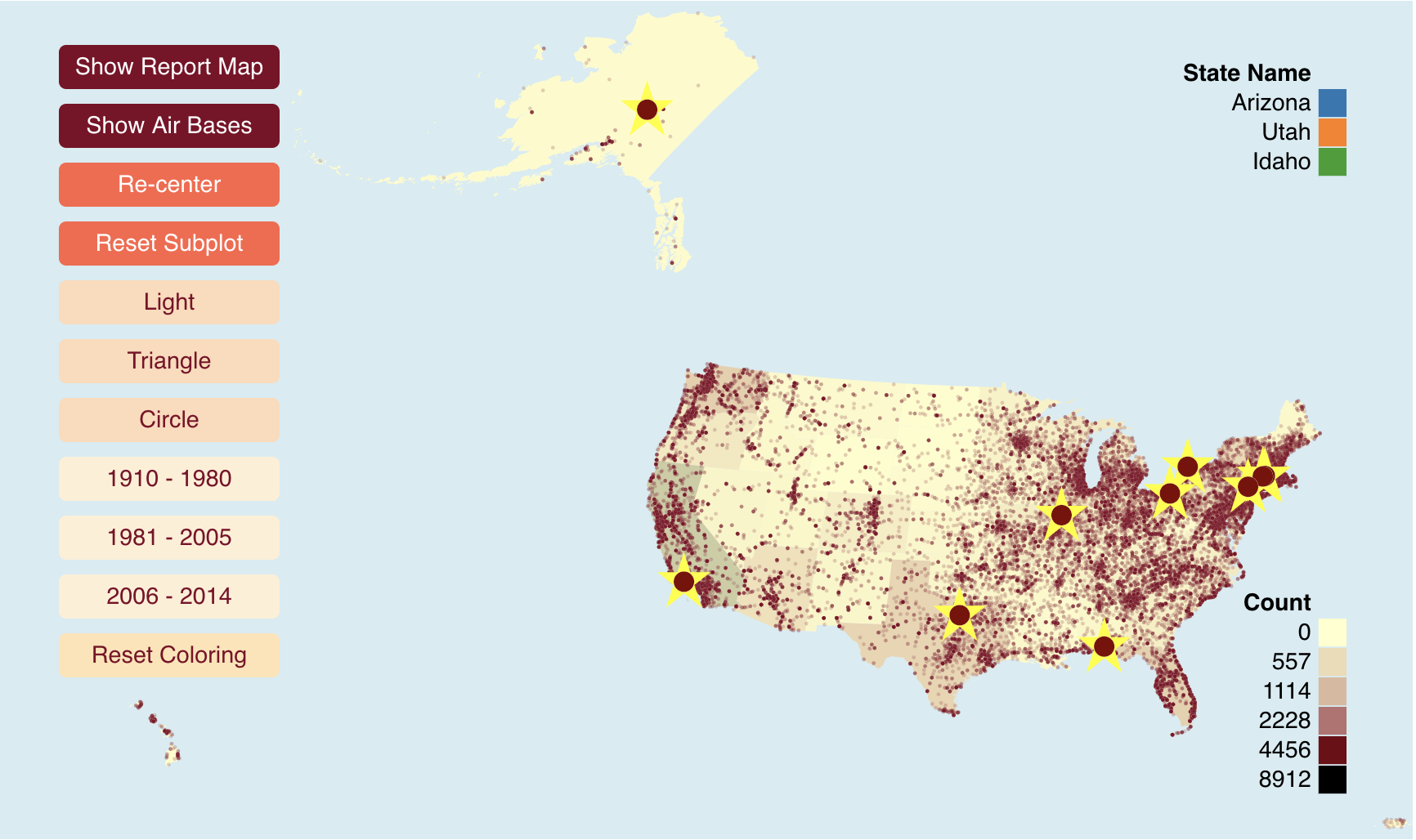
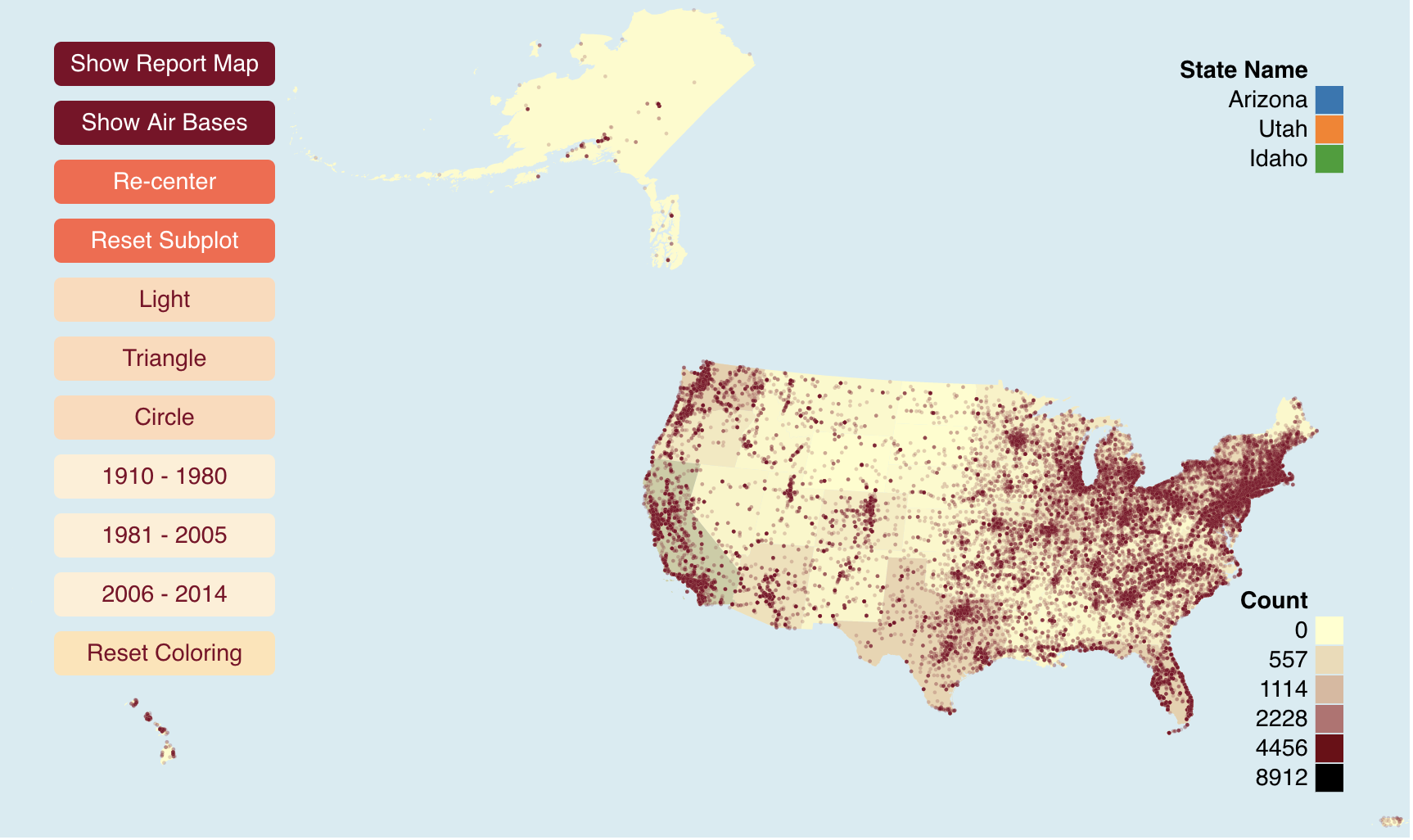
**Interaction: “Overview first”**



Main viewport includes a sequential color map that represents the total count of data items in each state. (Corresponding legend is on the lower right part of the main view)

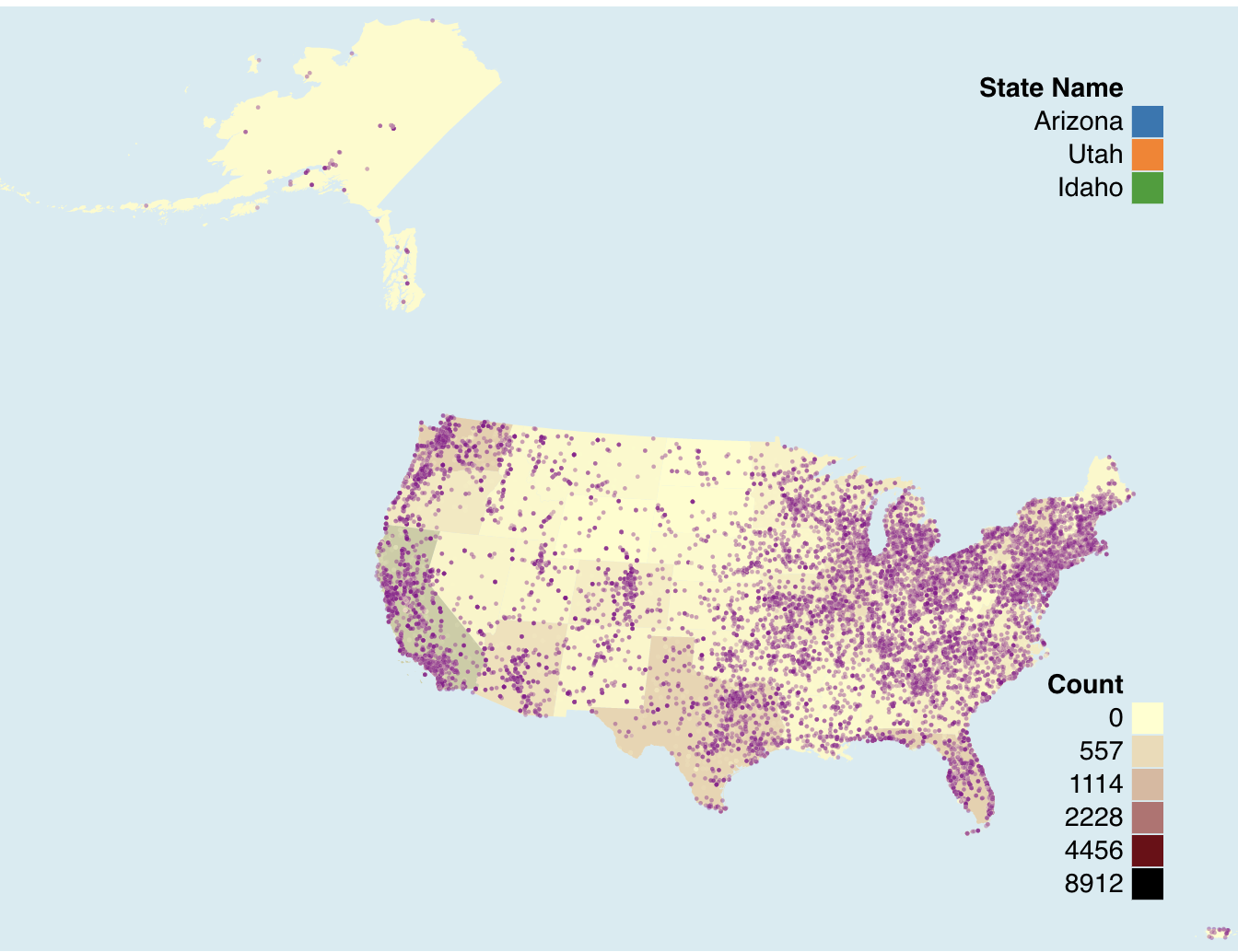
Interactively hovering across state would also display aggregated relevant year, shape, duration sub plots on the right.

By clicking the button “Show Report Map”, the individuals would be displayed in the main view based on its geometric attributes. By clicking the button “Show Air Bases” for comparison, the corresponding air force bases in our data set would also appear on the plot. By clicking them again, individual points would disappear and the main view would back to the original view.



**Interaction: “zoom and filter”**

In the detailed circles mode, we have filter condition “Light”, “Triangle”, “Circle” (the three most reported shapes) to filter the data items by attribute shape, and year “1910 – 1980”, “1981 – 2005”, “2006 – 2014” to filter the data items by time.

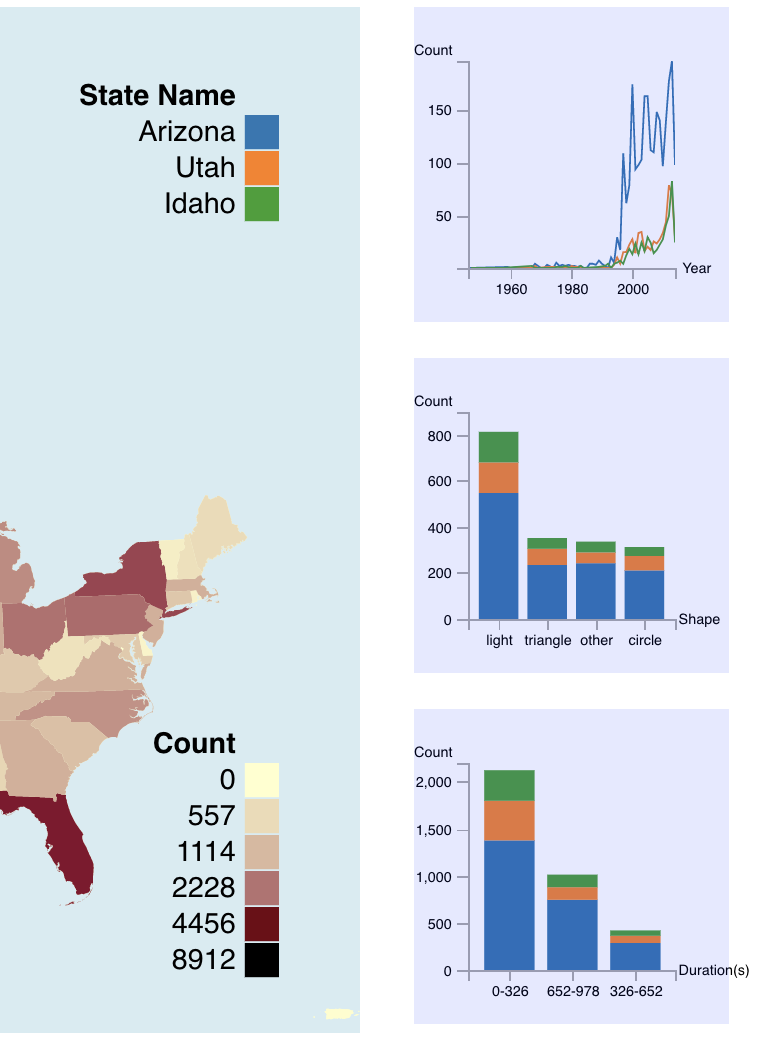


**Interaction: “details-on-demand”**

In the main viewport, hovering above a state would enable small tooltip that represent the numeric count value of that state.



Furthermore, by clicking on each state, comparison among states is presented. Max displayed state number would be 3. After the third click, further click would pop up the earliest selection of the state. “Reset Subplot” button would clean the current stack and change the view back to hover mode.



**Limitation**

First, we aggregate the data to state level, but we can narrow it further down to city information. Second, the viewport could be more interactive. The interaction between main view and sub plot could include brushing and filtering (linking) based on the attribute in the subplot. Third, air force bases could include more samples. Fourth, the plot should apply effective color channel as we only do a linear interpolation on two different color. Fifth, area or population of a state could be considered when we visualize the count. Sixth, the plot could contain more information in the tooltip. Last but not least, we could apply data kernel density estimation to mitigate cluster.