Template:

See design doc instructions for more information. (Design Doc - INFO 4602 - Fall 2024)

Examples from past classes: Example Design Docs

MEDIUM POST LINK

Beyond the Stars: Exploring UFO Sightings Data with Informative and Manipulative Visualizations

For this project, we utilized a *UFO Sightings* data set found on Kaggle for the purposes of constructing data visualizations that explore the significance of communicating data through differing perspectives. In order to do so, we constructed two categories of graphs: informative and manipulative. The goal of our informative visualizations is to accurately, clearly, and cogently communicate the data, as is to be expected of any other visualization, and the goal of our manipulative visualizations is to subvert expectations and effectively deceive the reader by deliberately altering conventional design choices.

The majority of your post should clearly walk through and reflect on the design stages that you went through to arrive at the final prototype.
What I'm looking for in design-process:

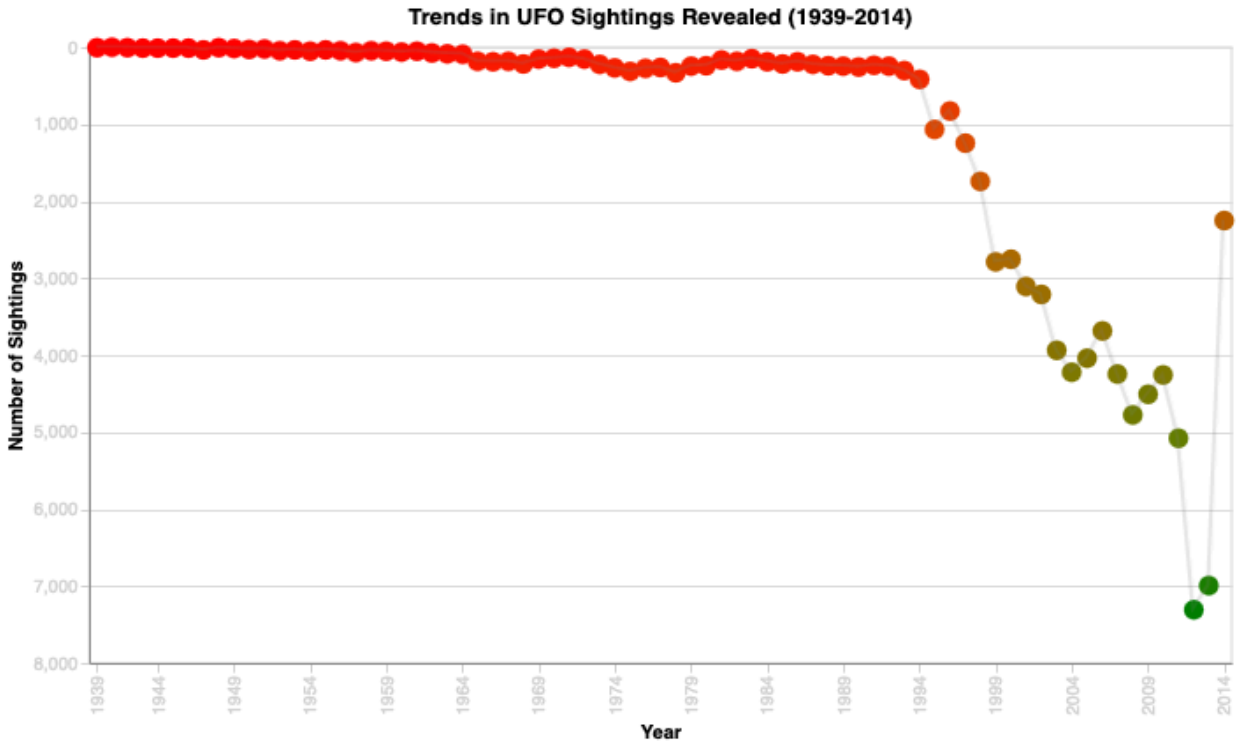
- You show written & visual evidence of sketching to explore different designs (ideation)
- You show written & visual evidence of sketching to prototype (prototyping)
- You show written & visual evidence of testing your prototypes with real people (user testing) and iteration of your design based on that feedback.
- You *concisely* describe why you chose among your possible designs (this does *not* have to be in-depth)

This could be helpful for design document structure

Ask yourself this: If one were to plot the trend of UFO sightings in the US over the past century, what might it look like? Further, what cultural phenomena or technological advances have contributed to these sightings and trends?

Might you guess that sightings would have peaked in the mid-50's or mid-60's as a result of Byron Haskin's *War of the Worlds*, America's infatuation with science fiction, or cultural hysteria over the Cold War?

If you had to choose between a trend upward in recent years or a massive dip, which would you choose?



Whatever you might have guessed, it's probably safe to assume that you didn't take the bait and lean towards a massive dip over the past 20 years. But, we do hope that the above data visualization gave you some trouble and challenged your intuition, at least initially.

The question you're asking…

Now you're probably asking yourself, "Why are you hoping that I was confused by the visualization *you* created?" and you would be asking the right question.

As mentioned in our initial summary, the purpose of this project was to "explore the significance of communicating data through differing perspectives." With how readily available data has become in the Age of Information, Michael Correll and Jeffrey Heer note that data is increasingly misrepresented through the breaking of data visualization conventions, data manipulation, obfustication, and nudging by individuals or groups hoping to tell stories other than the data actually tells.

And while this can be used for the purposes of true harm and the spreading of misinformation (we're looking at you, COVID graphs), we thought it might be fun to try our hand at differentiating between informative and manipulative visualizations using data that's a bit more trivial.

The Data

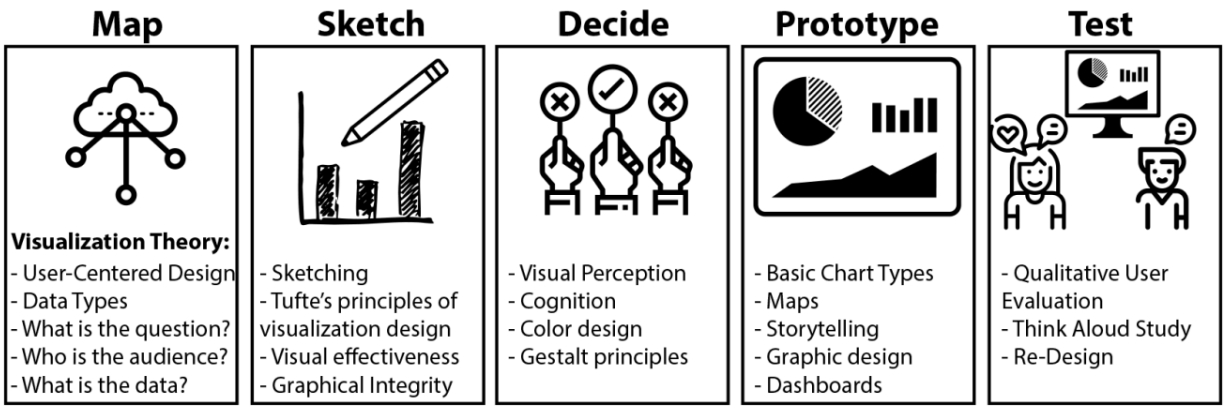
	datetime	city	state	country	shape	duration (seconds)	duration (hours/ min)	comments	date posted	latitude	longitude
0	10/10/1949 20:30	san marcos	tx	us	cylinder	2700	45 minutes	This event took place in early fall around 194...	4/27/2004	29.8830556	-97.941111
1	10/10/1949 21:00	lackland afb	tx	NaN	light	7200	1-2 hrs	1949 Lackland AFB&#44 TX. Lights racing acros...	12/16/2005	29.38421	-98.581082
2	10/10/1955 17:00	chester (uk/ england)	NaN	gb	circle	20	20 seconds	Green/Orange circular disc over Chester&#44 En...	1/21/2008	53.2	-2.916667
3	10/10/1956 21:00	edna	tx	us	circle	20	1/2 hour	My older brother and twin sister were leaving ...	1/17/2004	28.9783333	-96.645833
4	10/10/1960 20:00	kaneohe	hi	us	light	900	15 minutes	AS a Marine 1st Lt. flying an FJ4B fighter/att...	1/22/2004	21.4180556	-157.803611
...	...	...	...	...	...	...	...	...	...	...	...
80327	9/9/2013 21:15	nashville	tn	us	light	600.0	10 minutes	Round from the distance/slowly changing colors...	9/30/2013	36.165833	-86.784444
80328	9/9/2013 22:00	boise	id	us	circle	1200.0	20 minutes	Boise&#44 ID&#44 spherical&#44 20 min&#44 10 r...	9/30/2013	43.613611	-116.202500
80329	9/9/2013 22:00	napa	ca	us	other	1200.0	hour	Napa UFO&#44	9/30/2013	38.297222	-122.284444
80330	9/9/2013 22:20	vienna	va	us	circle	5.0	5 seconds	Saw a five gold lit cicular craft moving fastl...	9/30/2013	38.901111	-77.265556
80331	9/9/2013 23:00	edmond	ok	us	cigar	1020.0	17 minutes	2 witnesses 2 miles apart&#44 Red & White...	9/30/2013	35.652778	-97.477778

80332 rows × 11 columns

Our data comes from the National UFO Reporting Center's (NUFORC) Kaggle Dataset and contains just over 80k documented UFO sightings since 1910. As can be seen above, the data set is fairly simple as it contains exactly what one might expect in a reported UFO sighting. As can also be seen, this data set includes observations from multiple countries. And based on the scope of this project we decided that it would be best to focus on observations documented only in the US.

But we would not be fooled by this simplicity. Prior to crafting our visualizations, or even ideating, it was important that our team mapped, evaluated, and became acquainted with the data at hand, asking questions about what the data entailed, who our audience was, and how we might use it to visually articulate informative and manipulative stories.

The Design Process



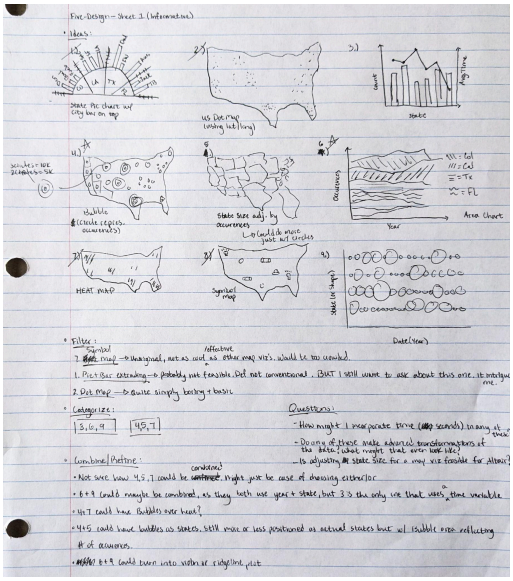
Mapping the data was the first step in our own design process, which heavily mirrored the Five Design-Sheet Method developed by Think Up Themes Ltd. and modeled by Kieran Tan Kah Wang. In essence, this method provides space for true creativity mainly through processes of of ideation/sketching, prototyping, and testing.

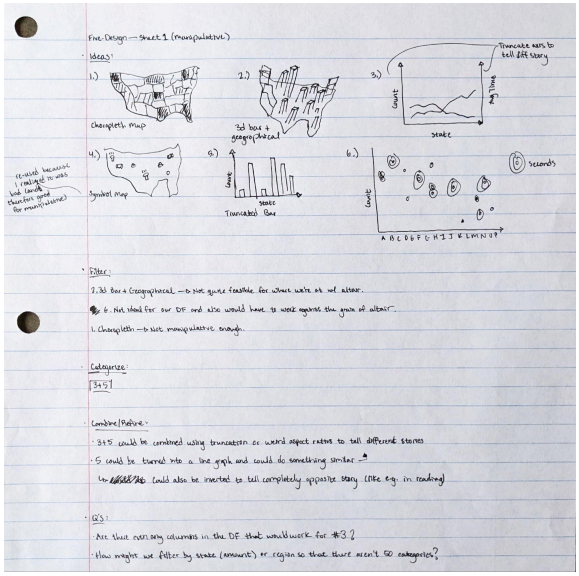
Ideation/Sketching

And so, we engaged in an extensive and crucial ideation/sketching process, challenging our initial thoughts and letting creativity flow freely. Our aim was to delve into our *UFO Sightings* dataset, uncovering interesting relations and patterns. This exploration fueled the creation of informative and manipulative visualizations. Our goal was to set aside time for each of us as individuals to generate and sketch ideas that we would later bring to the rest of the group.

Each of us took time on our own to generate up to ten ideas for both informative and manipulative charts. Then following the first sheet of the process, we filtered through, categorized, and combined/refined some of these ideas and then asked questions we might have had before then consolidating what we generated.

Max's Ideation





Max: Because I love structured processes, I followed the Five Design-Sheet Method, specifically Kieran Tan Kah Wang's rendition, almost "to a T," and engaged in practices of ideation, filtration, categorization, combination/refinement, and questioning.

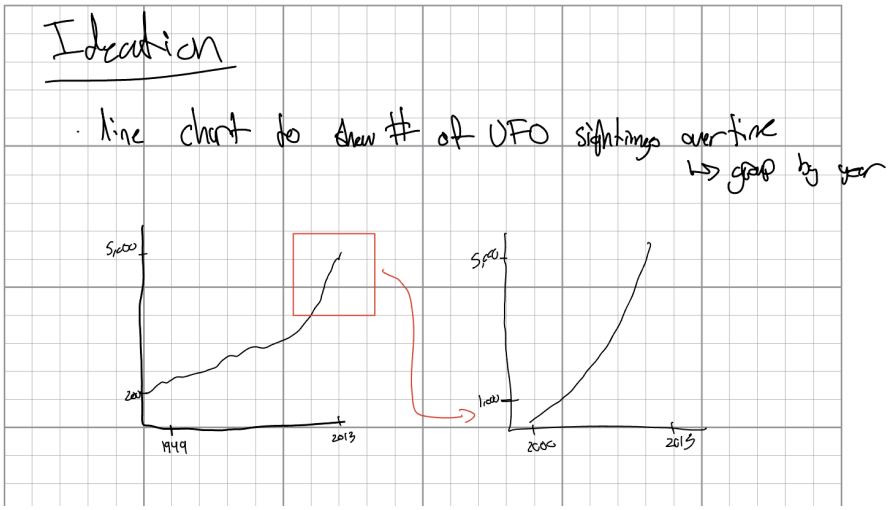
As can be seen, I set aside time to run through this process for both informative and manipulative designs, sketching ten and six ideas respectively. And while I would have liked to give more time to the manipulative designs, I knew that I could trust in the creativity and ideation of my team.

For my informative sketches, I included quite a few geographical data visualizations due to the data set's inclusion of coordinates. In addition, I also included some other graphs, such as a line/bar chart, a bubble chart, an area chart, and a concoction of my own in the top left corner (which was neither feasible nor effective).

For my manipulative sketches, I continued with geographical visualizations, even including the symbol map from my previous ideation after realizing that it would serve better to manipulate. In addition to these, I included a bar chart with a truncated y-axis, an odd bubble chart inspired by a chart evaluated in class, and a line graph with different y-axes.

Once I combined/refined and categorized my ideas, I settled on the area graph, the geographical bubble graph, the choropleth, and the symbol map to my team for discussion and evaluation. Further, I was prompted to ask questions of how we might make advanced transformations of the data in order to generate any of these graphs and of how we might reduce the state column in our data set so as not to chart an overwhelming number of categories.

Gustavo's Ideation



Ideation Phase: Shifting perspectives with a subset - the transformative power of selective data

Gustavo: One deceptive technique used during this ideation phase included choosing a specific range of data. By selecting a specific subset, we can change the perspective, compelling the visualization to illustrate a different story.

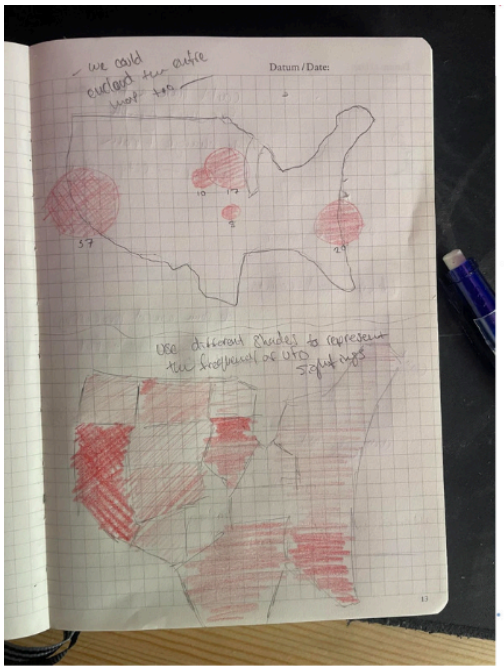
Upon lightly examining our dataset, it was clear that the number of UFO sightings increased as time progressed. Just before the year 2000, there were great increases in sightings. A potential

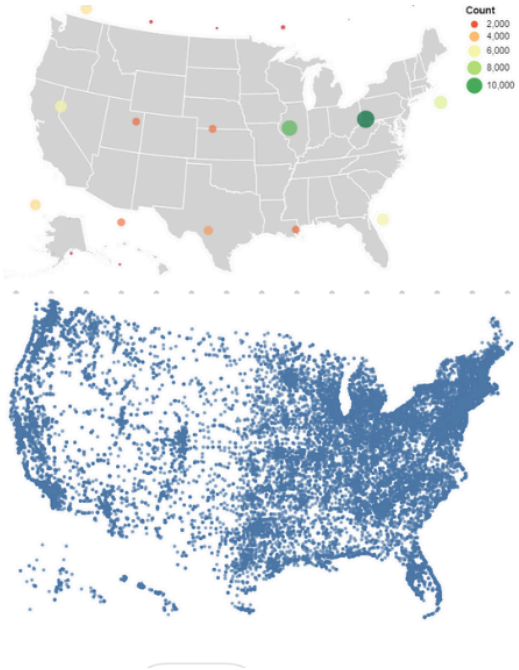
way to exaggerate this increase is to only examine data from 1990 - 2014. Plotting a graph with these ranges will make the increase in sightings appear more exponential.

The impact of this approach can influence certain perceptions, steering the reader to draw specific conclusions. For our dataset, this can emphasize a huge rise in UFO activity, which could lead to many narratives surrounding extraordinary events

After meeting as a team, we decided it would be interesting to try another deceptive technique rather than this one. We agreed to use an inverted y-axis to visualize the trend in UFO sightings over time. This technique will be employed during the prototyping/testing phase to craft a deceptive visualization that doesn't immediately reveal its manipulative nature to the audience.

Jackson's Ideation





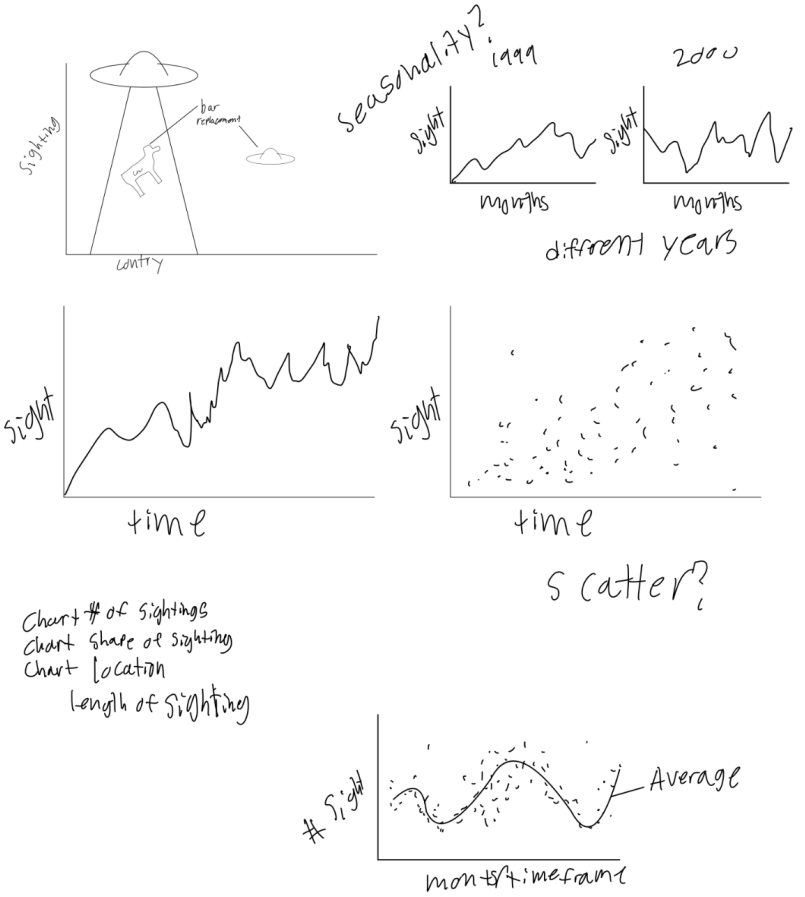
Jackson: Our dataset offered a lot of geographical information including: country, state, and even longitude and latitude for each UFO sighting. I decided to leverage this data and focus my visualizations on using the Geoshape mark in Altair to show the frequency of UFO sightings throughout the US.

I started my ideation process by sketching out various maps and exploring the different visuals I could make with the data. I experimented with various concepts, such as using circles of varying

sizes to indicate the sighting frequencies by state, and using color to represent the sightings as well.

Once I sketched a few ideas I decided to take my ideation into my Jupyter notebook and play with the capabilities of Altair. My first approach was to use the longitude and latitude data to place circles on the map. This task proved to be more challenging than anticipated, prompting me to investigate alternatives. This led me to explore plotting the longitude and latitude as points on the map, giving way to a great manipulative visualization.

Eric's Ideation



Initial ideation for different charts

With the volume of data available from our dataset I wanted to ensure that we could get a good overall picture on the number of sightings. Specifically I wanted to look for trends in the timeframe of spottings and outliers. Were there certain times of year, or day, that had spikes in

sightings? Did events like the war of worlds broadcast have a visible impact in reported sightings?

To this end, throughout my ideation process I focused on two main ideas. First I wanted charts that could show distinct differences in categories or years which led me to bar charts. Second I also wanted charts that could express seasonality, a line chart.

From here I created a few rough sketches and some notes on possible ideas of further development. The designs I came up with were a tad scattered, I mainly just wanted to get some ideas on paper, however eventually I decided to combine my line and scatter plots into a chart of sightings in scatter form and a line with a rolling average.

Choosing Charts

Charts:

- Informative
  - Chart 1: Area Chart (Max)
  - Chart 2: Rolling Mean (Eric
- Manipulative
  - Chart 1: Line chart with inverted y-axis (Gustavo)
  - Chart 2: Geographic Geo Scatter plotting UFO shapes (Jackson)

Once we each took time to sketch and ideate, we came together in order to evaluate our sketches based on design criteria learned in class. As such, we evaluated our sketches with the principles of marks, channels, and color theory in mind, while also taking into account feasibility and the constraints of Altair. Furthermore, we hoped to choose graphs that would effectively tell stories with an appropriate level of expressiveness.

With this in mind, we ultimately chose for Max's area chart, Jackson's informative and manipulative geographic maps, Gustavo's manipulative line chart, and Eric's informative rolling mean/scatter plot. And once we had agreed upon these visualizations, we continued to the next step of the Five Sheet-Design Method and sketched a low-fidelity prototype of each chart, which allowed us to explore each visualization's possibilities and consult friends and peers for feedback.

Prototyping

Max's Informative Prototype

---

1. ***Information***: meta-information such as title, author and dataset

2. ***Layout***: a sketch of the design visualisation

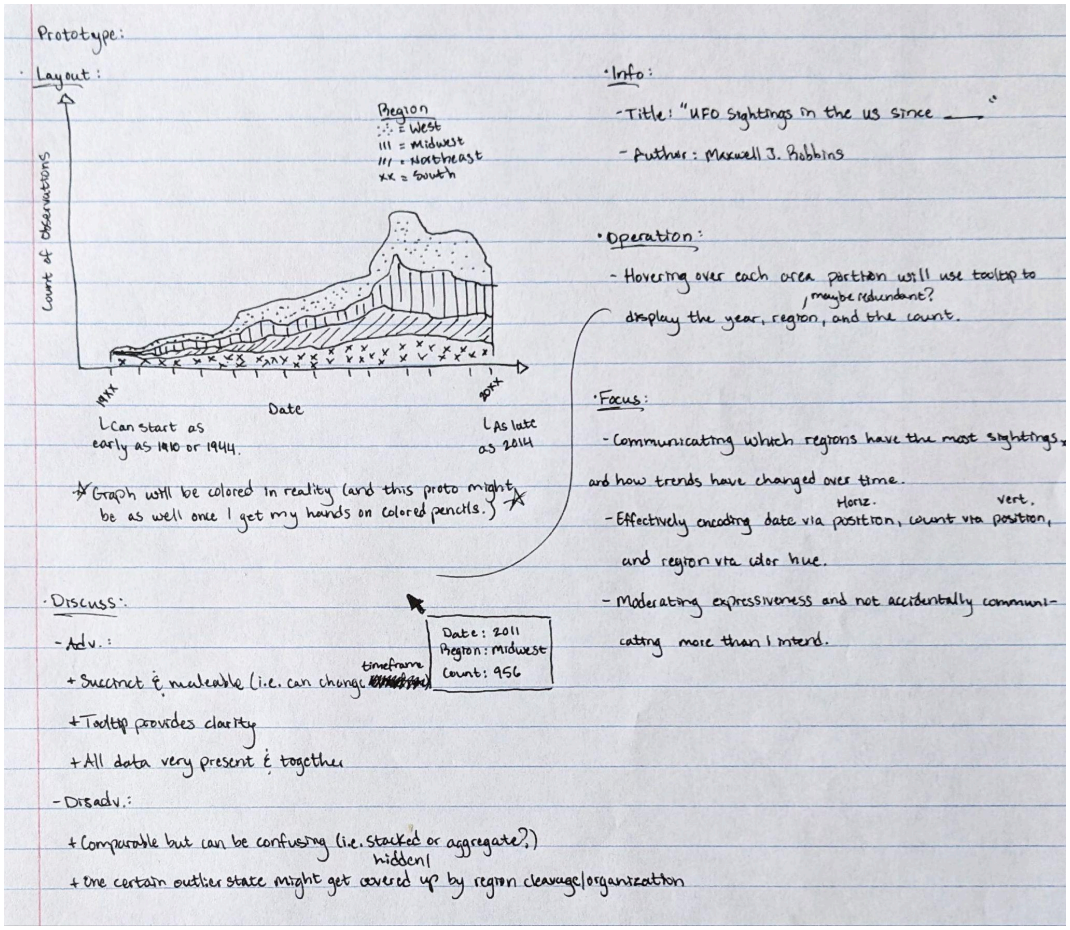
3. ***Operations***: list down the functions available in the visualisation and what the user is able to do with it

4. ***Focus***: identify the core portion of the visualisation design

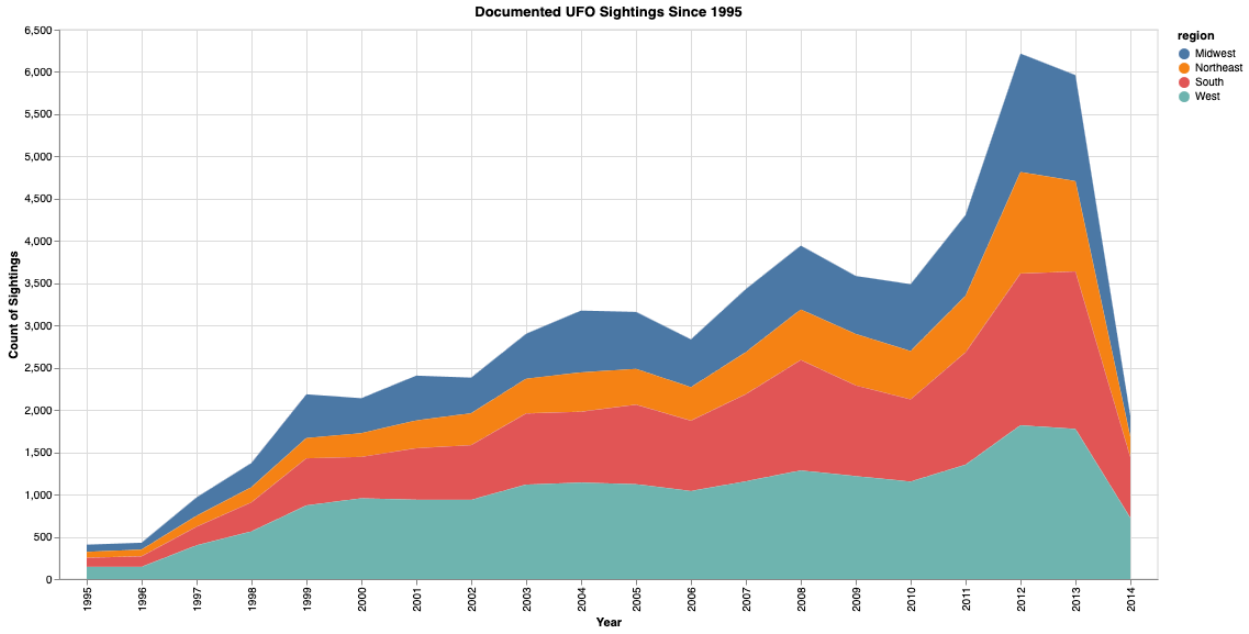
5. ***Discussion***: discuss the advantages, disadvantages, feasibility of the design

As previously mentioned, I was curious about how I might graph data related to each state without completely overwhelming the viewer with 50 different categories. And after deliberating with the team, we thought that it might be interesting to group each state by geographical regions (i.e. northeast, south, midwest, and west).

Once we decided to chart an area graph by region, I once again followed Kieran Tan Kah Wang's example, drawing a low-fidelity prototype that allowed me to get a feel for what it might truly look like, considering its operation, focus, advantages, and disadvantages.



As can be seen above, a big part of my focus was telling the story of the data effectively while moderating expressiveness. To do so, I considered that this visualization would utilize area as a mark while encoding "Date," an ordered attribute, via horizontal position on a common scale, "Count of Observations," a quantitative attribute, via vertical position on a common scale, and "Region," a categorical attribute, via color. And based on Tamara Munzner's effectiveness ranks of channels, these decisions are quite effective and do not communicate too many facts.



But while it is important for me to evaluate my own design choices based on design principles, none of it matters if viewers struggle to correctly perceive my data visualization. Accordingly, I took the next step in the design process by testing my prototype with people outside of class.

Justin:
- Not redundant to display year region, and count in tooltip
- He likes the clarity of the graph.
- Struggles to read it just a little bit.
- Not sure if stacked or not—not clear, necessarily.

Luke:
- Pretty neat
- He thinks that the increase is due to more people having access to the internet and smart phone cameras.
- Separating into regions is nice, but is there valuable info from that? They all increase and decrease similarly.
- He likes the idea of the interactive element (tooltip)
- Region might be redundant with the tooltip

Josh:
- People out here seeing a lot of UFOs
- Is it stacked or not?
- Understands that it is the amount of UFOs over time by region
- He likes the patterns (of the prototype)
- Pretty good
- Gridded graph might be good so that its possible to discern the value for each year

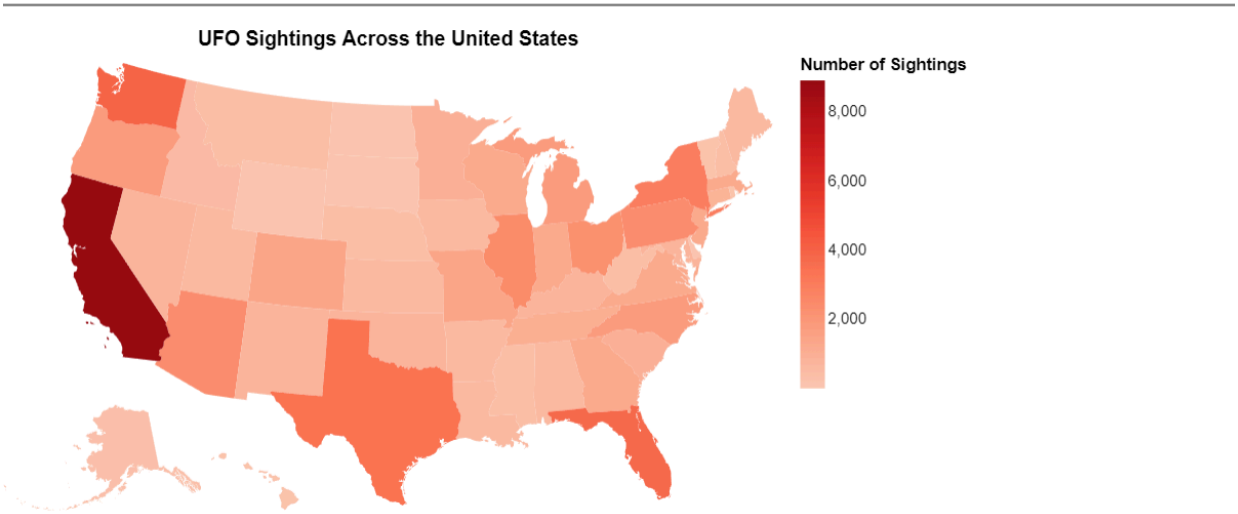
In doing so, I presented my prototype to three friends of mine, Justin, Luke, and Josh, each of whom have varying knowledge of data visualization. Their feedback is above.

My biggest takeaway from testing my prototype was that each of my friends were able to generally understand the visualization but weren't sure about whether the area portions were stacked or aggregated. Furthermore, each of them loved the idea of the tooltip.

Even though each of them liked the idea of the tooltip, I was worried that it would become something that viewers would depend on, especially after reading Lisa Charlotte Muth's article on when to use classed and when to use unclassed color scales. While classed and unclassed color scales aren't the issue in my case, Muth mentions in the article that it is important to consider that viewers might see graphs statically, like on a PDF, and not be able to depend on interactive tooltips to understand them.

With this feedback and knowledge in mind, I had to rethink how I might approach my graph while combining effectiveness and expressiveness and enabling viewers to perceive it correctly without depending on a tooltip.

Jackson's Informative and Manipulative Prototypes

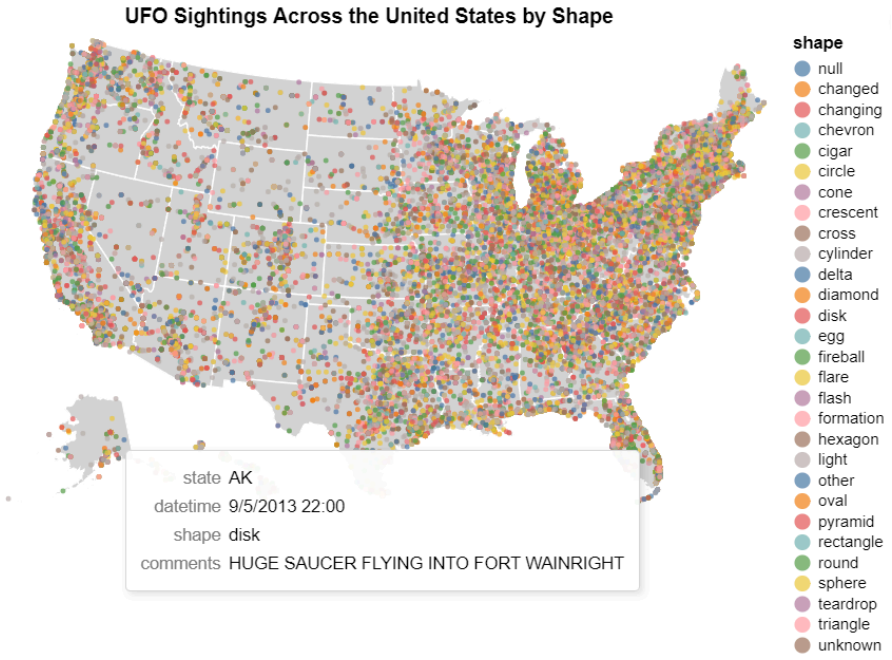


For this visualization I aimed to create a very simple easy to read choropleth map with the shade of color representing the amount of UFO sightings in each state.

I decided to use a monochromatic color scheme to bring attention to the states with the highest frequency of sightings. This technique, called color ramping, is also mentioned by Jon Schwabish in Better Data Visualizations where he says "This color palette is also easy to understand –smaller numbers correspond to lighter colors and larger numbers to darker colors."

The one big downside to using color ramps is the difficulty of comparing two states with similar values. To address this issue I added a “tooltip” so that when you hover over a state you will get information including the state and number of sightings.

Further information on creating this visualization can be found in the Altair documentation.

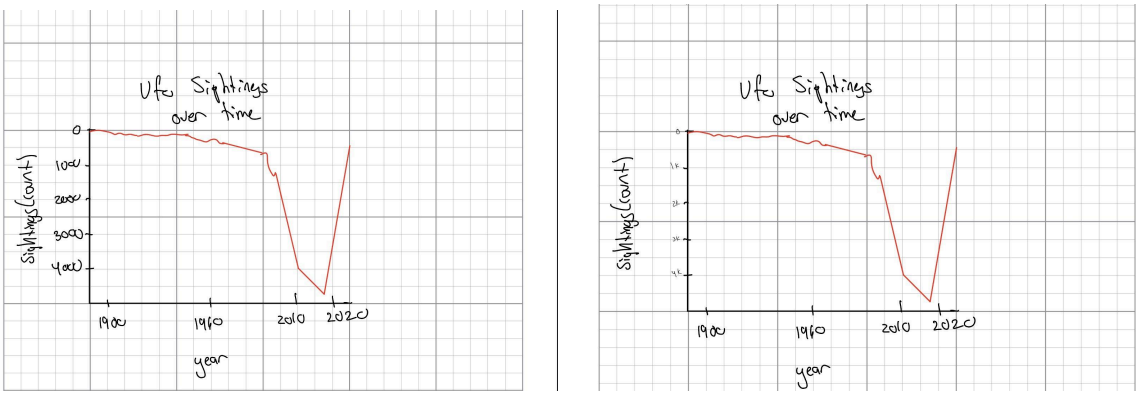


I went into my next visualization with the goal of creating a manipulative mapping of our UFO sightings data. I chose a geo scatter plot for its capacity to obscure the true distribution of data points. This technique exaggerated the apparent concentration of sightings in the Northeast, misleading viewers into overlooking the actual hotspot in California, where sightings are densely clustered along the coast.

To compound the manipulation, I color-coded the points according to the reported shape of each UFO sighting, adding a layer of complexity that could distract from the data.

Additionally, I incorporated “tooltips” providing information on each sighting (state, date, shape, and eyewitness accounts) to offer interactive accessibility to the data.

Gustavo’s Manipulative Prototype



Two paper prototypes depicting manipulative UFO visualizations with different y-axis.

One deceptive technique used included inverting the y-axis so that initially, people draw the wrong conclusions from any trends they notice. These two versions of this paper prototype are the same but their differences lie in the y-axis. One has more noticeable and ‘normal’ labels, while the other labels have smaller fonts and are a slightly lighter shade, slightly decreasing the chart’s effectiveness. The decision to test two prototypes stems from several considerations, including the need to address feedback from initial interviews.

Interviews

The next step in the design process was to have people, preferably outside of class, test out our prototypes and collect feedback to improve them. With this manipulative visualization, I was also curious to see if it was successful in deceiving/misleading people and how quickly they would notice any manipulative design features. Before testing, I asked people to give me any key insights they get from the visualization and that I am testing both informative and manipulative visualizations but will not be telling them which I am testing for

Initially, I tested the prototype with a bolder/more noticeable y-axis on two people. Both reported that the UFO sightings seemed to be increasing as time passed. They also commented on the y-axis:

- "At first I was a little bit confused because it was upside down, But I got it in a bit when I looked at it more"
- "The y-axis is a bit distracting, it was a bit confusing at first, but I noticed it right away"

After these two initial interviews, I thought it would be a good idea to alter the prototype a bit and make the y-axis a little less noticeable (decreasing its expressiveness) by making the font smaller and lighter, and making changes to how I display the numbers (1000 → 1k). I tested this version of the prototype on 3 different people all of whom drew a similar conclusion and did not notice the y-axis. All 3 individuals claimed that UFO sightings were decreasing and did not immediately notice the y-axis like in the first 2 interviews. Some notes included:

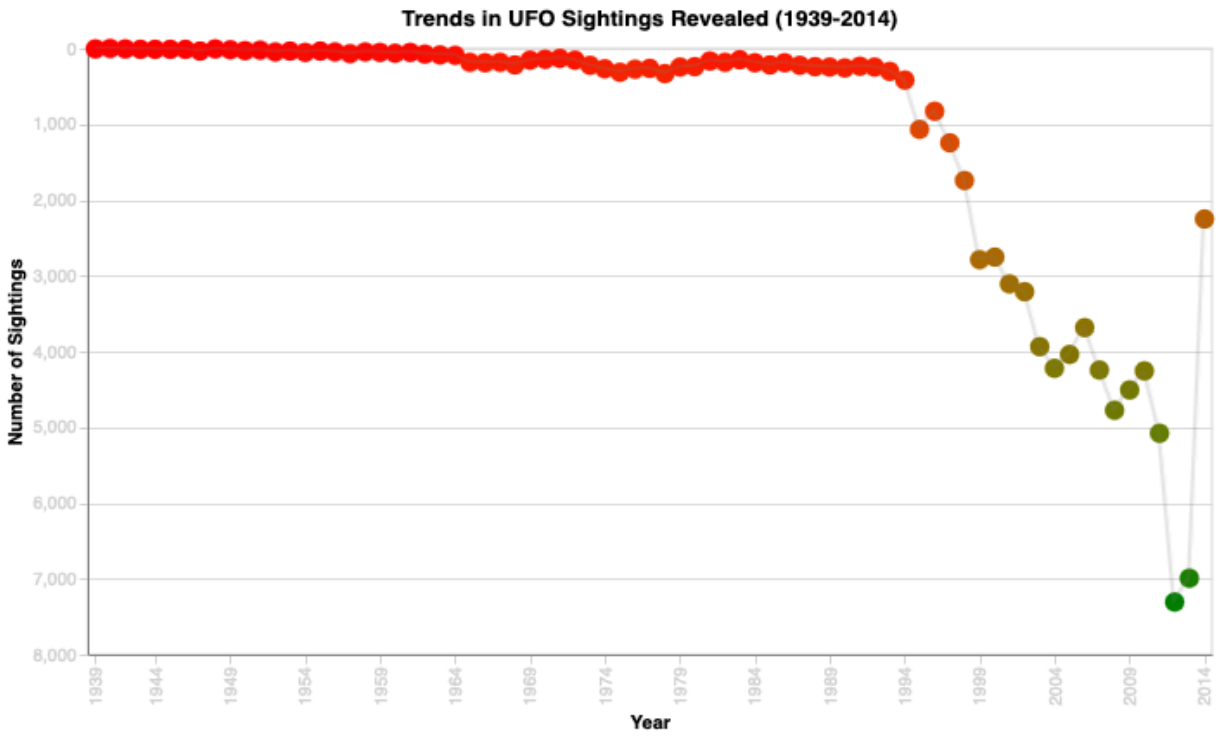
- "UFO sightings are decreasing as technology advances"
- "There are more sightings of UFOs during COVID"
- After telling them to look at the y-axis: "Oh wait what the? Yeah, that's deceptive, I wouldn't have noticed"

After I received this valuable feedback, I thought it was now appropriate to take my prototype and try programming it in Altair. One feature I wanted to add to my visualization that I had not added to my prototype was color scales. I believed that I could use color to aid in my manipulative design.

Using a diverging color scale would work best for my visualization. According to Lisa Charlotte Muth in her article *Which color scale to use when visualizing data*, Diverging color scales "have a bright middle value and then go darker to both ends of the scale in different hues. [They] are often used to visualize negative and positive values." In considering my color scale, I contemplate the utilization of a gradient from green to yellow to red. Red, commonly linked with danger or negative impacts, contrasts with green, symbolizing safety or positivity, while yellow assumes a neutral position in between. To add to the visualization's manipulative nature, I

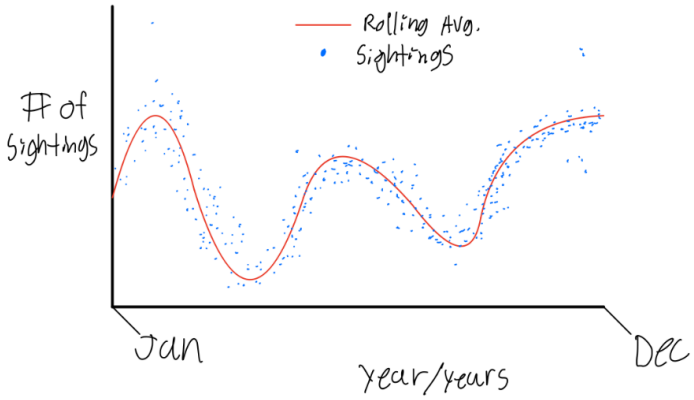
decided to use the opposite color associations to influence the audience to read the chart in a certain way.

In this design, my original intention was for red to signify periods of heightened UFO sightings, and green to represent low sightings. However, due to the inversion of the y-axis, this interpretation is reversed, creating an unexpected twist in the visual narrative. Furthermore, I made the labels a light gray color so they are harder to read and removed any keys that help the reader understand the visualization better, decreasing the chart's overall effectiveness.



Prototype implementation using Altair

Eric’s Informative Prototype



Titles

Sightings over time

Sightings during the Year 19XX/20XX

Seasonal Sinhgtings

Intent

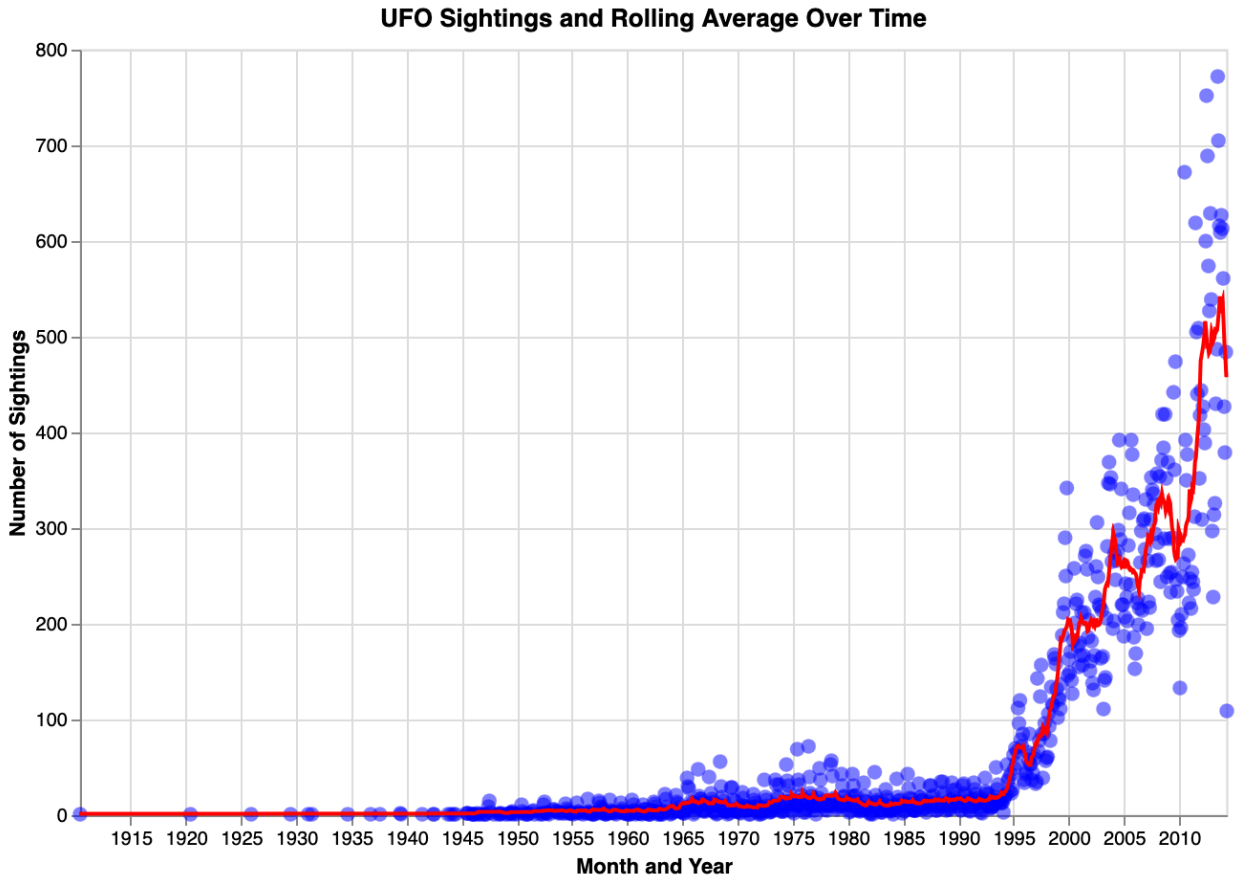
What times of the year see more reported sightings?

How do # of sightings change over time?

notes

- events in world affect # of sightings?
- year or years for x-axis
  - multiple charts for multiple years?
- junk data, & outliers?
  - identification, what they mean?

Prototype idea generation and notes



Initial Rolling mean chart prototype

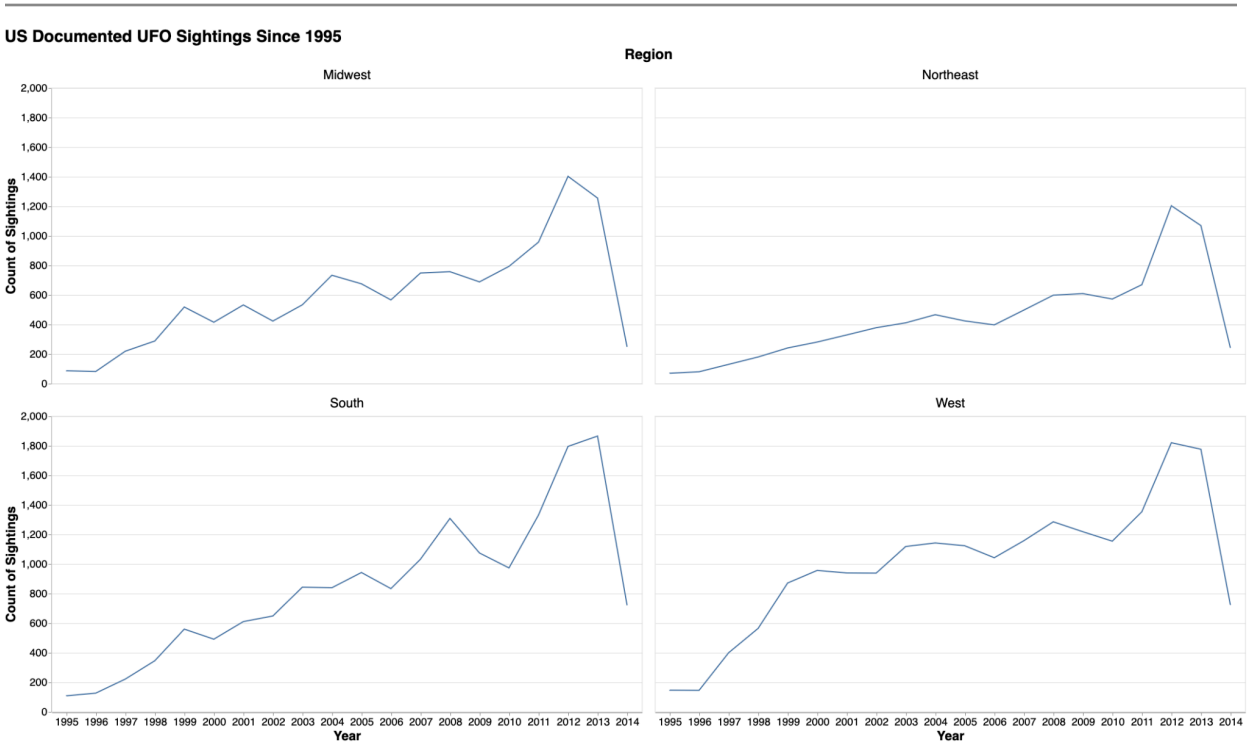
Using my decisions in the ideation phase my initial prototypes were a bit rough. I started the process by trying to figure out what exactly to plot on my charts. I hoped to plot all the available data onto a single year, the idea was to see if the data showed a common seasonality trend with recurrent spikes and dips depending on the season. I wasn't confident in this plan as the end result could have been uninformative junk, which it ended up being, so I generated a few other ways I could use the plot.

For the plot used in the demo day I decided to plot everything. While I felt there was merit in plotting a few years either in chunks or on individual charts I wanted to get a general idea of what others might think of the chart itself so I left these other ideas for refinement.

From some feedback I got, people generally interpreted the graph as showing a large increase in sightings starting in 1995 which falls in line with what I wanted. That being said people also seemed to just glide over the smaller rises present from around 1965 to 1980.

Final Designs & Takeaways

Max's Final Design



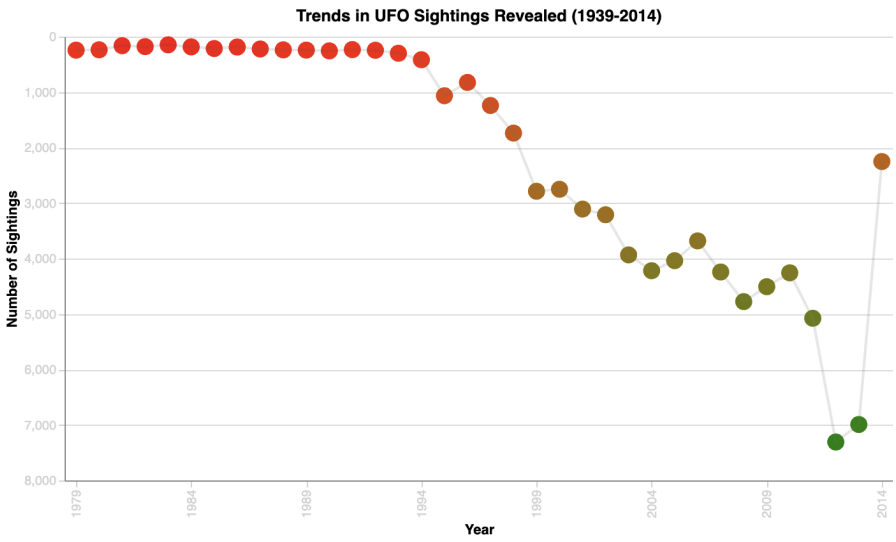
Much of the feedback I got from friends and peers surrounded the reasoning for choosing an area graph. In the case of my friends, who I tested my prototype with, they weren't sure if the areas were stacked or aggregated. In the case of peers, they found the stacked area graph to be slightly ineffective as it showed an obvious general trend but hindered the viewers' ability to compare between regions.

With this feedback in mind, I decided against an area graph in totality and ultimately landed on a faceted line chart spanning the years of 1995 and 2014. I specifically chose this timeframe because (1) there were minimal UFO sightings prior to 1979, (2) Gustavo's manipulative chart already covered a wider range, and (3) 1995 was about when UFO sightings substantially increased.

Additionally, this chart allows a clearer comparison between regions and is even more effective than my area graph prototype as "Region," a categorical attribute, is encoded via spatial region, the highest rank for identity channels and one rank above color according to Tamara Munzner, rather than color.

Although effective, I do regret the fact that three of our visualizations are line-oriented following my decision against an area chart.

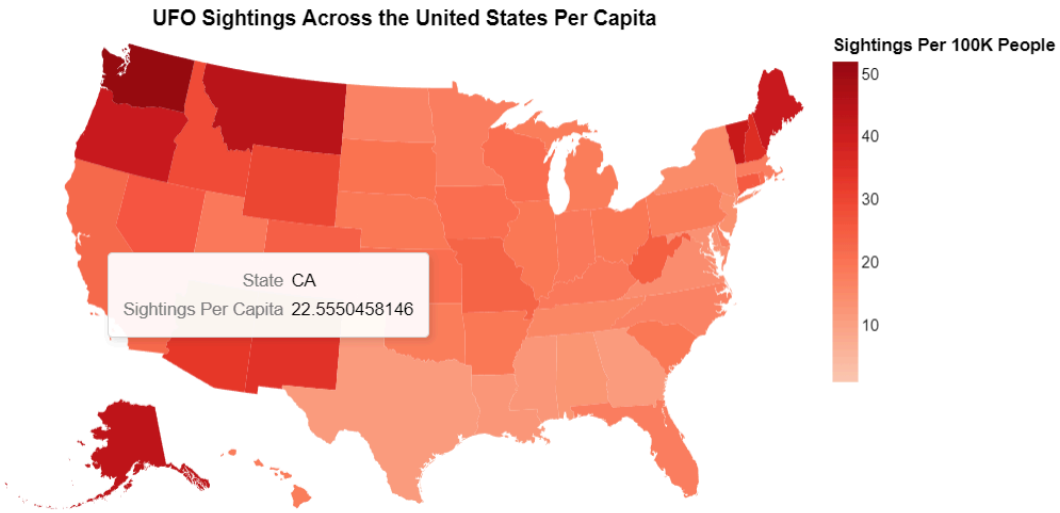
Gustavo's Final Design



Final Prototype Design

Following the feedback I received from peers and my team I decided to shorten the data range to paint a more effective picture. This shortened range shows a clearer trend and might influence the audience/reader to describe a decreasing rate in UFO sightings when in reality, UFO sightings are increasing.

Jackson's Final Designs



In the process of revising my initial choropleth visualization, I took into account feedback from my team and classmates, leading me to normalize the data on a per 100,000 pop scale. This adjustment allows for more meaningful comparisons between states, ensuring that the visualization reflects sighting frequencies in relation to state population and reducing misinterpretation as much as possible.

“The informative choropleth, possibly use a different gradient or color scheme because the lower values tend to blend together.” - Classmate

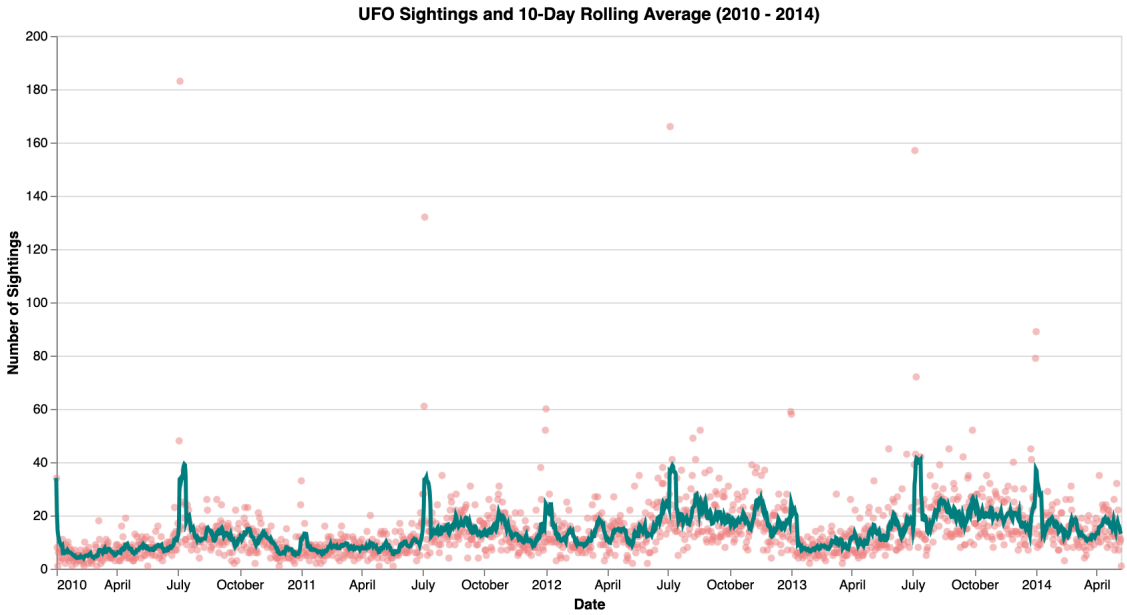
The choice of a monochromatic palette was deliberately aimed at emphasizing the states with higher frequencies of sightings. It also allows for a much more intuitive interpretation of the data whereas other color palettes may take some back and forth with the legend to understand the values, a monochromatic palette is widely understood. Additionally, it’s important to note the absence of the “tooltip” during the demo day, a feature that significantly enhances the interpretability of subtle differences between states.

When finalizing my geo scatter plot, feedback received during the demo day prompted me to streamline the visualization for a more subtle presentation. I opted to eliminate the shape-based color coding, adopting a uniform color palette instead. This adjustment aims to reduce the presence of manipulation without detracting from the manipulative nature the visualization already has.

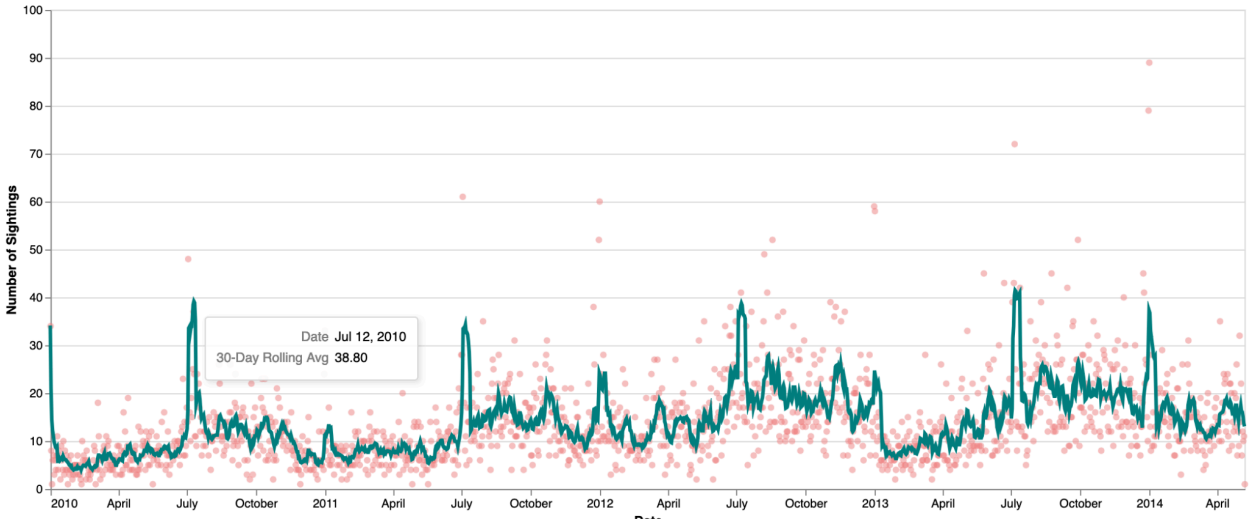
Eric's Final Design

Using feedback given during the in class demo and personal desires for making the chart more meaningful I first changed the colors. This was a fairly consistent point of feedback I got as people found the blue and red to be either hard to read or somewhat unpleasant to look at. To fix this I experimented with a few different colors before eventually deciding on a color pallet that was more subdued, a dark green and a peach color. With this new color pallet I was also able to emphasize the rolling mean line and make the scatter plot data less prominent.

Following this I decided to focus on a specific set of the data limiting myself to 2010-2014 as there is a more consistent date starting in 2000 with the largest spikes in sightings being from 2010 onward. I also found that if I limited the y-axis I could exaggerate the data making spikes more distinct and overall more visually interesting. Finally the last bit of feedback was suggesting the inclusion of a tooltip to which would improve the understanding of the chart while also making it more interactive.



Refined chart with new colors no vertical gridlines and a tooltip (showed in second image)



Manipulated Y-axis to exaggerate the data

As seen in the chart there is in fact a recurring pattern for UFO sightings in the United States. Around the time of the fourth of July, within a few day range, routinely every year sees a spike in sightings most likely due to the number of fireworks. Start of the year also sees a similar spike, again likely due to fireworks, although 2014 is a bit of an outlier, a bigger spike than that of other years. 2012 was seeing higher numbers than other years in general.

Takeaways