Ben Greenawald (bgh5yd)

Final Project Writeup/Reflection

Whenever the national conversation turns to the issue of gun violence, the issue of mental health inevitably follows. The idea is, any person who would commit an act of gun violence must be mentally ill, so addressing the epidemic of poor mental health should then mitigate gun violence. But is gun violence truly the consequence of poor mental health, or has mental health simply become a scapegoat to avoid finding the real, harder solutions? There have been a number of good analyses/visuals on these variables by themselves (for gun violence: https://www.vox.com/policy-and-politics/2017/10/2/16399418/us-gun-violence-statistics-maps-charts and for mental health http://www.mentalhealthamerica.net/issues/2017-state-mental-health-america-ranking-states) but I was unable to find any good visuals that explored these two variables together.
That is the primary question that I addressed in my visualization: the relationship, or lack thereof, between firearm violence and mental health rates.

Now, after some preliminary analysis, something became immediately clear: there really does not appear to be a meaningful relationship between these variables. Doing a simple linear regression analysis fails to produce meaningful results. This presents me with an interesting challenge in visualizing this data. Contrary to what we have done throughout this entire semester (revealing the patterns amongst data), I am trying to show that there is, in fact, no pattern amongst variables that many people believe to be correlated. Thus, instead of trying to show order, I am trying to show chaos. A user should walk away from my visual not with some newfound insight about the relationship between gun violence and mental health, but a newfound skepticism over whether such a relationship even exists. We will discuss this idea more throughout the write-up, but just wrapping my head around this challenge was one of my biggest struggles throughout this process.

The data

The backbone for any visual is the data. The data for this visual came primarily from 3 sources. For the firearm data came from the CDC Wonder database (https://wonder.cdc.gov/). The data is from the years 2012-2014 and tracks the number of deaths from homicides and suicides by state. An obvious issue here is that states have wildly different populations, which will have an effect on how many people died by gun violence. So, on top of raw counts, the normalized "rate" is found by taking the number of deaths per 100,000 people. The mental health data is a consolidation of a number of different datasets, all coming from the Substance Abuse and Mental Health Services Administration (https://data.world/samhsa/). These datasets estimate the prevalence of different mental illnesses amongst the population of each state over the years 2012-2014, reported as a percentage of the population. The final bit of data deals with mass shootings. Mass shootings are defined to be shootings where 4 or more people are injured/killed in a single shooting and the data here is the number of mass shootings by state over the years 2013-2014 (https://www.massshootingtracker.org/data). This data has also been engineered to contain a normalized "rate" which report number of mass shootings per 1 million people. Before being displayed, the "rate" variables are scaled between 0 and 1 to make filtering operations easier to work with.

The visualization

One thing that was important to me when making this visual was to not just have one visual, but a number of cohesive visuals (I ended up doing 2). The reason for this is that, as we have seen a number of times throughout the semester (especially during the "Visual Data Analysis" homework), different visuals reveal and hide different things. Thus, having more than one visual, if done right, can help tell a more complete version of the story. The two visuals I went with were an abstract map and a parallel coordinates graph.

The first visual component was an abstract map. On this map, each state is simply represented by a square where the size of that square represents the mental health rate and the color represents the firearm death rate. For the color scheme here, I chose I diverging blue-white-red color scheme. I found that a single gradient color scheme made it much harder to distinguish states. This particular color scheme was chosen because it felt natural. Blue indicates less violence, going all the way to red which indicates more gun violence. Though unintentional, it also worked out that my US map was red, white, and blue. The reason I went with an abstract map is that this data is not strictly geographic. "States" are just concepts that help us create discrete observations within our data, but having the states look like states serves no real purpose. Further, abstracting states as squares allows us to have an additional dimension (size) on top of just color. This allows us to view both variables in a single map. Now, a fair question to ask is, "if states are just concepts, why have a map at all?" This is a fair question, and one that I struggled with quite a bit. However, after prototyping a bit, something became clear: there does appear to be some sort of geographic component to this data, especially the gun violence data. In other words, states with similar gun violence rates tended to cluster geographically. Now, while this does not add anything to the question I am asking, it does help the user begin to ask a different interactive. In other words, a user can look at the map and see that there does not appear to be a strong relationship between the color of squares and the shape of squares (which there would be if gun violence and mental health were related), but they will see that the color of squares tends to cluster by geography. This should plant a seed in their mind that gun violence rates seems to have something to do with geographic location. The true underlying cause is beyond the scope of this visual, but using the map in this way should hopefully plant that seed.

The second visual strategy is the parallel coordinates chart. This is showing a lot of the same information as the map, but in a different, more precise way. So this chart has only two axes, the mental health rate and the gun violence rate, with a line for each state. What this particular graphical form does really well is to show the chaos amongst the data. This graph is sorted by the left side which is firearm violence, meaning there is a nice sense of order when the graph starts. If there were a strong relationship here, we would expect that order to be maintained as we draw over to the mental health side. There would, of course, be some overlap and noise, but overall the pattern should maintain. However, what we actually see is a pretty strong sense of chaos. In fact, it can be a bit overwhelming at first to try and parse through this graphic because of how much the lines just go everywhere. That is the point.

By default, the lines are just colored reddish-pink. The legend makes it clear that this color has nothing to do with the color on the map. However, a struggle I had here was how to utilize the color dimension for this graph. Since both my variables had their own axis, the color of the lines was still available to encode more data. This was tricky because any sort of continuous data would be too confusing to try and color, so I needed something discrete. Another issue I was having at this point was that mass shootings appeared nowhere in my graphic. The reason for this is that mass shootings don't happen enough that it is useful to view them on a continuous scale. However, looking at the data, the mass shootings did split up nicely into three categories (0 mass shootings, [0,1) mass shootings per 1 million, and >= 1 mass shooting per 1 million). So, I added the option to color the parallel coordinates graphic by the mass shooting categories and also filter on these. I had to be careful not to use the same colors as the map so that users did not try and relate these colors, so I used a black, green, and orange for these colors. Again, the legend makes it clear that these colors do not relate to the map colors.

As for the interaction, I gave the users the option to switch the firearm and mental health datasets. I also gave them filtering options so that they could break down the data in a number of different ways. One particularly interesting way is to look at the difference between the data. So the user sets a threshold, and can then look at states where the difference between the mental health and firearm data is greater than the given threshold. Since the data is all scaled to the interval [0,1], it is easy to look at this difference. States that pass the filtering operation will remain while those that do not will become transparent. Also, upon hovering any element in either the map or the parallel coordinates chart, the corresponding element in the other will highlight and more information about the given state will be displayed.

One big struggle I had was with this filtering operation, specifically in its presentation. Once the data is scaled to the interval [0,1], it becomes difficult to interpret individual data values. It becomes even harder to interpret the difference. So if I say, the difference between the mental health data and the firearm data is 0.4, what does that mean? Thinking about specific values has no real meaning, it's when you start looking at the trends (i.e, these states have the biggest differences) that patterns begin to emerge. What I wanted to do was remove these numbers altogether (as was done in the static visual), but I found that working with the threshold was almost impossible without including the numbers. So the inclusion of the scaled numbers, even thought they are difficult to interpret, was a necessary evil.

A final feature that I added were some presets for the interaction parameters to show some of the interesting things that I found in my data. Selecting a preset would automatically update the graphic and give you some notes from me about what you are seeing. This provided a simple way for the user to quickly get at some of the key insights that I identified.

The other big thing I struggled with was the complexity of the graphic, and trying to find a way to effectively explain exactly what is going on. Ideally, the graphic is intuitive enough that the user can just explore around and see what is going on. But after adding a good amount of interactive features, it is difficult to figure out what is going on solely by explanation, thus, some additional explanation was needed. I think having the preset helps a bit, but ultimately I just decided to have a very long tutorial and additional a "legend" for the interactive features that describes in great detail the visual and the ways to interact with it. These do not appear by default and only show up if the user requests them, so I don't think they were too big of a deal. However, I think if I had foreseen this issue at the start, I could have mitigated it a bit. This would have required me to rethink my filtering operation a bit, as this is the most confusing part of the graphic. But if I was able to rethink that operation (maybe look at something like the slope of the unscaled data), then that would simplify the differencing filter and would also remove the need to scale the data to the interval [0,1], which is also quite confusing.

Static Visualization

Don't want to say too much here, just that the static consolidated a number of the interactive presets described above to provide a compelling argument against the relationship between mental health and gun violence in a static form. I did have to change the background from light blue to white for printing reasons.

Insights

Like I mentioned at the start of this paper, the big key insight I identified was that there does not seem to be that strong of a relationship among mental health and firearm violence. This can be seen from the default graphics, which shows a good amount of chaos in the parallel coordinates graph and also in the map, where the color of the squares doesn't really seem to be related to the size. A further note here is that the color actually does seem to be related somehow to geography, hinting at what the true underlying cause of gun violence should be.

Looking at mass shootings and serious mental illness, an interesting insight occurs. Namely, that amongst the 8 states with a lot of mass shootings (1 or more per 1 million people), there did not seem to be any pattern in the rates of serious mental illness. In fact, one of the states with lots of mass shootings (Illinois), had one of the lowest overall rates of serious mental illness.

A final interesting insight was the relationship between deaths by suicide and thoughts of suicide. Despite the intuition that at least these variables should be related, the data shows otherwise. Half of the states show relatively large differences (relatively large here is subjective. I chose a decently high differencing threshold and the pattern was still clear). Utah, which had the highest rate of thoughts of suicide, had one of the lowest rates of gun death by suicide. On the other hand, Lousiana, Mississippi, and Alabama, which were all middle of the pack when it came to thoughts of suicide, were the three states with the most guns death by suicide. This again points to the idea that geography seems to be playing a larger component than mental health.

Conclusions

In this project, I attempted to visually show the lack of a relationship between mental health and gun violence. This was definitely my most ambitious visual of the semester, and because of that has some issues with explaining exactly what the visual can do, and displaying certain interactive elements. However, I think I succeeded overall. The parallel coordinates chart is effective at showing the chaos in the data, while the abstract map gives the users hints as to what the actual cause of increased gun violence might be. I provided a number of interactive tools to allow users to dive deeper into the data, as well as some presets that allowed the user to quickly glean insights from the chart.