Comparing COVID-19 Data with an Interactive Mirror

Leader: Chris King (king.2229@osu.edu)
Team Members: Yifei Diao (diao.48@osu.edu), Da Gong (gong.475@osu.edu)
Professor: Dr. Jian Chen (chen.8028@osu.edu)
Department of Computer Science and Engineering at The Ohio State University

Abstract

The number of new coronavirus infections has increased exponentially since January affecting economic stability and social order throughout the entire world. Using a novel interactive mirror design introduced by our professor, Dr. Chen, we sought out to design an intuitive visualization that would effectively describe the COVID-19 propagation trends. After a great deal of hard work, we were able to develop a visual interface that compares the COVID-19 status of the most infected countries in the world.

1. Introduction

Novel Coronavirus, otherwise known as COVID-19, was declared a global pandemic by the World Health Organization on March 11, 2020. However, at the beginning of the Spring 2020 semester, it was only a frightening yet relatively isolated health crisis. The virus which broke out in Wuhan, China in December 2019, has only recently become the focus of world attention, but it has felt like an eternity ever since the stay at home orders and calls for social distancing were announced. Novel coronavirus has caused incredible stress on the global economy as people have lost their jobs and been forced to collect unemployment. Additionally, it has caused great panic as people desperately do what they can to prevent themselves from contracting the virus. Measures of all kinds are being taken including the wearing of face masks and even over the top hoarding of hand sanitizer and toilet paper.

Our goal with the interactive mirror was to present the global spread of the virus intuitively through methods of data visualization.

2. Project Proposal

Initially, we planned to create a visualization using a global geographical map combined with dots and color maps to show the propagation of the virus globally. We planned to acquire data with a variety of methods. First we would use Amazon's Mechanical Turk to survey mass amounts of people in order to get people's estimated idea of the number of infected, recovered, and deaths in different parts of the world. In addition, we would survey students and experts in the medical field at the Ohio State University. This data would then be used to display the difference between the actual numbers and the numbers people estimated.

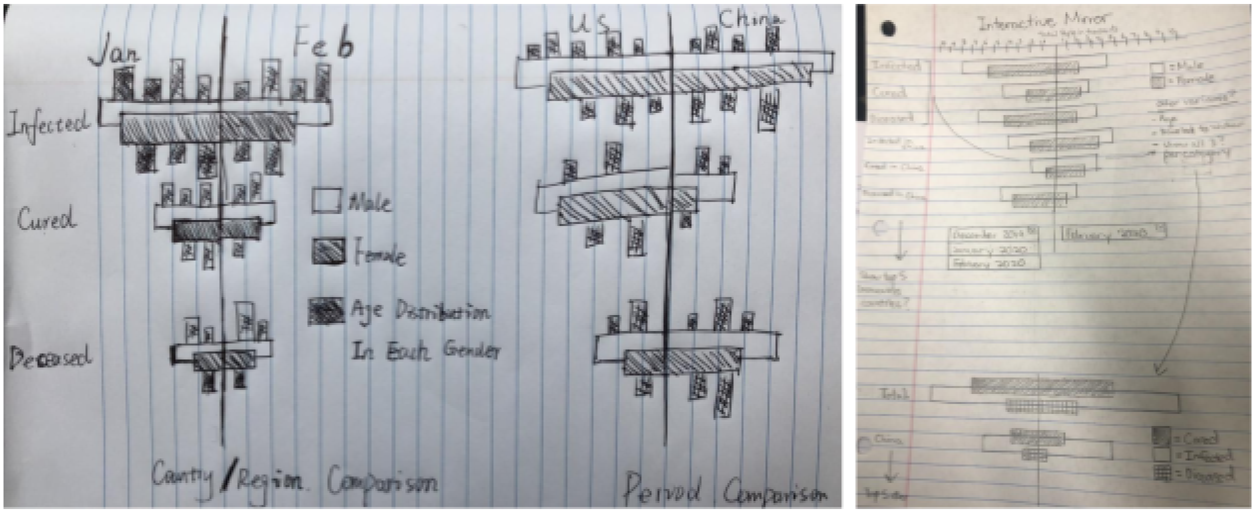

Figure 1: Early Interactive Mirror Design Sketches - Early designs included variables such as gender, age, infected, cured, and deceased. Several different comparisons were considered including by country and by period. Aditionally, a drop down menu was considered in order to select month. The final design ended up being quite a bit different, but many elements of these sketches were still used in some shape or form.

Our goal was to show how panic was impacting people's perspective of the virus. The questions we planned to ask participants of Amazon's Mechanical Turk are as follows:

1. What city are you from?
2. What country are you from?
3. How many cases of COVID-19 do you think exist in your country?
4. How many people do you think have recovered from COVID-19 in your country?
5. How many people do you think have died from COVID-19 in your country?

Using a combination of D3, VTK, and Python we hoped to implement this design. Although we thought we had come up with a solid project idea, we had no clue how wrong we were!

After meeting with Dr. Chen, it came to our attention that this was maybe not the best plan of action. For starters, our design was deeply flawed. As we learned in our readings, the visualization pipeline has the opportunity for uncertainty to be introduced at each stage and we had significant uncertainty in just the first stage of the pipeline, the acquisition of data. We were counting on Amazon's Mechanical Turk to provide good data for us, but this was incredibly unlikely. Therefore, we switched paths and decided to focus on the actual data that was out there. In addition, it came to our attention that our geographic design was simply not a good way to display information. Dr. Chen recommended that we use her novel interactive mirror concept to display the information.

The premise of the interactive mirror was that you could move the extremities of a bidirectional bar chart towards each other so that you could better compare the value of similar bars. Now that we knew the premise of our design, we had to figure out how we would apply it. Some sketches of our initial design can be seen in Figure 1.

3. Literature Review

After we had come up with our design, the next part of the project was the literature review. Dr. Chen helped us find three papers related to the interactive mirror concept. Below are brief summaries of those papers.

3.1. Automatic Selection of Partitioning Variables for Small Multiple Displays

The goal of this paper was to understand multidimensional data using 2D techniques. The two types of visualizations analyzed include projective displays such as scatterplot matrices and small multiple displays which show 2D slices of data created by partitioning one or more variables.

The problem this paper focuses on is that as the number of variables grows, neither visualization type mentioned scales well since the number of plots increases quickly. The authors' solution involves showing only the subset of interesting variables, but this can be very time consuming. Therefore, they attempted to create an algorithm that will automate this process successfully based on a set of goodness criteria.

One study used to motivate this study was the linear relationship between ACT scores and admission rates for universities with high ACT scores. After analyzing several sets of small multiple displays given the data set, it was evident that there were no patterns related to the relationship of ACT scores and admission rates except for in universities with limited admission rates and high ACT scores. Based on this outcome, the authors' goal was to devise a method that examines variables in a data set and recommends effective small multiple displays such as the ones observed manually.

In order to devise this method, the authors' sought to describe a set of goodness criteria for evaluating small multiple displays, develop a method for measuring the quality of partitioning variables based on randomized tests while adjusting pattern detection sensitivities, and demonstrate the effectiveness in goodness criteria. In order to test the results, the authors decided to focus on small multiples based on scatter plots and scagnostics, a popular cognostics metric. The algorithm devised in this paper may be particularly helpful in our project.

Despite the focus of this paper being on small multiple displays, our project will also have a number of different partitioning variables based on our research questions. Therefore, we could use this algorithm to select interactive mirror comparisons that show valuable trends in data. The algorithm depends on four criteria in what is called Goodness-of-Split Criteria. First, the algorithm chooses data comparisons that convey rich visual comparisons. Second, it selects comparisons that are more informative than the input visualization. Third, it chooses well-supported comparisons that convey robust patterns as opposed to misleading patterns of chance. Lastly, it selects comparisons that are parsimonious in that it has fewer partitions.

One study that showcases the solution to a major challenge with this algorithm was based on the relationship between median value of home and proportion of houses built prior to 1940. The results of the study were several small multiple displays which showcased effective trends with varying levels of skew. Therefore, the challenge that the authors had was in using this algorithm to determine the likelihood of a small multiple displays cognostic value. The solution was to compute likelihood using randomized permutation tests. To evaluate how unlikely it was that these scores arose due to chance, the data was permuted into partitions a thousand times. By comparing the true skewed values to the randomized results, it was determined that certain results were more skewed than the random permutations which indicated that evidence was not due to chance.

The main takeaways I had from this paper in relationship to our project was that in order to select valuable comparisons in data, we need to use a large dataset to ensure that the trends we identify are well-informed and that we must permute the data into partitions at least a thousand times to ensure that the evidence of a trend is not due to chance.

3.2. Animated Transitions in Statistical Data Graphics

The goal of the paper is to investigate the effectiveness of animated transitions between common statistical data graphics. The author tries to find whether animated transitions can significantly improve graphical perception. Although animated transitions have received a lot of people's attention, little research has been conducted to characterize the design space of transitions between statistical data graphics and assess how animated transitions affect graphical perception. Therefore, this paper aims to do research on this.

Before the experiment, animation principles including Congruence Principle and Apprehension Principle were introduced first in order to guide the design of the experiment. As for other theoretical preparations, the author extended theoretical models of data graphics, aiming to mainly focus on syntax and semantics analysis in this experiment. Also, a taxonomy of animation transition types which can be applied to the experiment were introduced. These transitions included filtering, sorting, substrate transformation, timestamps, visualization changes, and data schema changes.

Two experiments were conducted to test the effects of animated transitions in the research. One is at the syntactic level of analysis, which is generally object tracking during animated transition. The other is at the semantic level of analysis, which is generally estimating changing values of graphs during animated transition. In both experiments, different kinds of transitions including static transition, animation

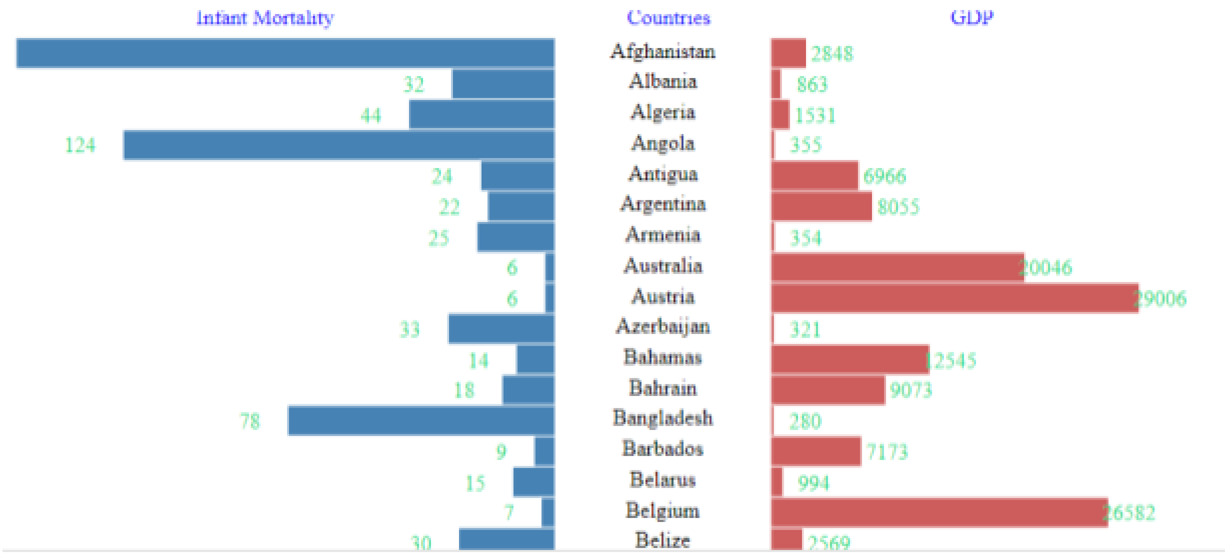

Figure 2: This simple bidirectional bar chart served as the foundation for our COVID-19 Interactive Mirror. The original can be found at the following link: https://bl.ocks.org/kaijiezhou/bac86244017c850034fe

transition and staged animation transition were implemented.

Statistical graphs including grouped bars, scatter plot, stacked bars and donut chart were used. The experiment result showed that the staged animation performed the best in both syntactic and semantic tasks, followed by animation, and static transition performed the worst. This proved that animated transitions with careful design can significantly improve graphical perception of changes between statistical data graphics.

3.3. Designing for Mobile and Immersive Visual Analytics in the Field

Data collection and analysis are important in disaster warning and emergency issues. The authors point out that, in field work, gaps between data collection and data analysis in spatial and temporal contexts. For example, data collection is usually in the field and data analysis and validation are usually in labs, which guarantees neither accurate results nor correct data. A new technique implemented on FieldView by the authors combines data collection and analysis together on a mobile device to solve these problems mentioned as above.

FieldView is an open-source prototype platform for the mobile and situated visualization of field-collected data so that we can easily learn from their visualization design ideas. However, most of the visualizations designed by the authors are bar charts, scatter charts or line charts, which cannot appropriately describe the virus status which is also a kind of disaster or emergency issues.

For example, bar charts are not appropriate to describe a multidimensional data set, such as period, gender and age together, which are commonly used to describe people infected by the coronavirus. We will design several new forms of charts using interactive mirrors and try to optimize the design mentioned in this paper on FieldView to find a better way to present the status of virus.

4. Method

4.1. Tools

As we began our journey in developing the interactive mirror, we considered several different frameworks. The main two we came down to were WebGL and D3. After spending some time with each, we decided that our collective experience would be better utilized in a D3 environment.

After we selected D3 as our framework of choice, we began to search for existing D3 projects that could be used as the foundation of our project. We came across many incredible designs. One design we found was a dynamically animated bar chart that showed the population of the most populated cities in the world over the last 200 years. As a city's population would overcome another, it's bar would move above the other. This design has actually become very popular recently and we considered altering our design to accomodate this incredible visualization. However, ultimately we kept our original goal in mind. This design was too unnecessarily complicated to serve as our Interactive Mirror foundation. After continued searching, we found a simple but effective bidirectional bar chart built in D3. It was the perfect foundation for the Interactive Mirror. The original work can be seen in Figure 2.

In addition to D3, we used a variety of web design tools to develop the COVID-19 Interactive Mirror. We used the basic web design tools including HTML and CSS. In fact, we used tools within HTML5 in order to store variables in the local memory of a computer's browser. In addition, we used vanilla Javascript and several libraries in addition to D3 including JQuery and Ajax. In addition to languages, we used an external slider tool as the premise for our date slider.

The last tool to be discussed is the actual data. There are two sources of data that the COVID-19 Interactive Mirror draws from. The first dataset is Johns Hopkins' Github repository which keeps track of infections, recovered, and deaths on a daily basis with respect to COVID-19. The second dataset is Our World in Data's public csv that keeps track of COVID-19 testing data on a daily basis.

4.2. Development

The development of the COVID-19 Interactive Mirror at its base level is divided into three parts which include the design of the inital interactive mirror framework, the refinement of the existing framework, and the addition of features based on feedback from our professor, Dr. Chen, and our classmates.

The framework consisted of adjusting the original bidirectional bar chart

Figure 3: This was a pivotal point in the development of the COVID-19 Interactive Mirror. The design went from being a comparison of infected, recovered, or deaths between countries to the comparison infected, recovered, or deaths for one country over the span of all dates we had data on. The next stage would be to overlay infected, recovered, and deaths and compare countries on each side.

so that it read COVID-19 data rather than infant mortality data. Next, each horizontal bar chart had to be moved into its own respective html div. The width of which would be dynamically decided based on the size of the data. At first, the D3 scale being used was linear, but it quickly became apparent that the identity function would need to be used in order to compare data accurately. This meant that the lengths of the bars were way too long. Therefore, the idea for a scale factor was formed in order to reduce the size of the bar chart values. Next, the interactive mirror had to be coded. We used JQuery to select the scroll bars of each HTML div and move each accordingly when the other was moved. Once the mirror was working, the rest of the framework design consisted of manipulating the data in which we decided to first use text input fields and a button due to the ease at which it was implemented. This was however a terrible way to manipulate the data because the only way you could do so effectively is if you knew the fields present in the csv file.

The next stage of the design was refinement. In this stage, we switched the design. At first it was comparing the infections, recovered, or deaths between all countries. However, we changed it so that it was comparing infections, recovered, or deaths for one country over a period of time. This was just an early proof of concept though. The next step would be to overlay the infections, recovered, and deaths for each country and compare two countries. This pivotal moment of the development process can be seen in Figure 3. The refinement continued with the addition of drop down selection bars, the addition of a scale, and further CSS refinement.

The last part of the design process was the addition of more features. When we met with Dr. Chen, she gave us some very good feedback. She told us to introduce a fourth variable using a line as mentioned in the textbook she had directed to us earlier in the semester. Next, she told us to create a date slider so that we could select a time period. Lastly, she told us to make the current design available for mobile.

Figure 4: Here is a screenshot of the final COVID-19 Interactive Mirror. The visualization displays infections, recovered, deaths, and tests for each country selected. On the left, the data for Italy is displayed and on the right, the data for France is displayed. A date slider is included to select the period of dates desired and the scroll bar at the bottom can be adjusted to compare the bar values for each country more accurately.

5. Results

We were able to successfully integrate Dr. Chen's feedback and in general were able to successfully integrate the COVID-19 Interactive Mirror. You can see a screenshot of the final design implemented in Figure 4. You can also use the tool if you visit the following link listed below: https://dacking15.github.io/InteractiveMirror/. Considering our design is developed with D3 and is on a live website, it should work on mobile, but for some reason, it is not quite there yet. This is something that could be worked on in the future because even if we can get it to work on mobile, the design will need to be adjusted so that it works intuitively on a small touch screen. So far, we can confirm that the link works on Google Chrome on Windows and Mac.

6. Discussion

The design purpose of the interactive mirror was to create a mirror-like visualization divided into two parts. When users move one side, the other side of the mirror would reflect that movement. Users can not only retrieve corresponding daily data on a specific date, but also can observe the trend of the number of infections, deaths, recovered, and tests over time. The mirror interaction allows users to drag the left or right side to compare and observe the difference between two bar lengths which is especially convenient for bar charts with similar data.

Five variables are reflected in the interface including date, country, number of infections, recovered and deaths. On each side of the webpage, the bar graph is presented horizontally and extends vertically representing our quantitative daata. Date is presented on the left-most side of the page representing our temporal data. Through this interface arrangement, users can observe the trends in each country easily.

The mirror fuction help users to locate data quickly and easily. When users try to locate data by date, the number of infected people displayed on both sides always stays in the same date range. In addition, users can drag one of the sides until the gap between the two sides can be clearly observed.

The bar charts are designed to be overlapped with each other. Users can observe significant differences between the four types of data. Additionally, we added a horizontal coordinate axis to help people understand data through scale. We chose the countries with the top 20 highest number of infections for comparison. Users can select the country on the left and the right side in the drop-down menu on the top.

The data is generated from a csv file thatmust be maintained regularly. However, we also have a python program that can be run to update the csv file automatically. We also implemented the scale factor function in the menu bar, allowing users to adapt the data to their screen. After changing the scale factor, the coordinate axis will be adjusted as well.

In terms of work, Chris King developed the interactive mirror framework, initial csv file, initial input fields, the scale coordinate axis, and scale factor. Yifei Diao developed the overlapping bar chart design and the automatic python csv update program. Da Gong developed the drop-down menu design and the date slider.

7. Conclusion

After much tireless work, we have successfully created a visualization of the global spread of COVID-19 using Dr. Chen's interactive mirror design. We implemented a program to automatically obtain COVID-19 data of total infections, recoveries, deaths, and tests from Johns Hopkins and Our World in Data. We built an interactive mirror framework in D3 using our collected data. Users can select countries within the top 20 of most affected to compare using drop down menus. They can easily compare the data within a certain time range by simply dragging the bidirectional chart towards each other. We also added some useful features in our visualization to make it more effective. A coordinate axis scale and scale factor were implemented to help users adjust the data accordingly. A date slider was created to allow users to choose a range of dates that they are interested in. What's more, we made our project adaptable so that it could be implemented for mobile devices in the future.

Overall, our interactive mirror design can help viewers compare COVID-19 data effectively between countries in an extremely intuitive fashion. We believe that our implementation visualizes the COVID-19 data in a way that upholds the fundamental research in the data visualization field.

8. References

A. Anand and J. Talbot, "Automatic Selection of Partitioning Variables for Small Multiple Displays," in IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 1, pp. 669-677, 31 Jan. 2016.

Whitlock, Matt, et al. “Designing for Mobile and Immersive Visual Analytics in the Field.” IEEE Transactions on Visualization and Computer Graphics, 2019, pp. 1–1., doi:10.1109/tvcg.2019.2934282.

Heer, Jeffrey, and George Robertson. “Animated Transitions in Statistical Data Graphics.” IEEE Transactions on Visualization and Computer Graphics, vol. 13, no. 6, 2007, pp. 1240–1247., doi:10.1109/tvcg.2007.70539.