Using Visual Analytics to Understand Covid-19 Dataset

Zeyu Cai* Qingyun Wang† Qi Zeng‡

University of Illinois at Urbana-Champaign

ABSTRACT

To learn the status of pandemic more easily, people need to digest vast amount of relevant pandemic information every day. The project aims at providing a graphical representation of current covid cases for a region using color systems to present better visualizations for people. Furthermore, we explore various color systems and examine how different color systems affect information delivery efficiency. Our code are publicly avaliable at `https://github.com/ZeyuCai/cs519_project`.

Index Terms: K.6.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle; K.7.m [The Computing Profession]: Miscellaneous—Ethics

1 INTRODUCTION

Pandemic brings panic. The prevalence of corona virus has brought growing interests in acquiring and analyzing a huge amount of accessible relevant data to keep a close eye on this event. In this work, we aim at providing a graphical representation of current covid cases for a region using color systems to present better visualizations for people. Specifically we explore various color and shape systems and examine how they affect information delivery efficiency.

Our contributions are three-fold:

- We build a unified geographic choropleth maps and survey the most suitable color systems.
- We show that an animated line graph may helps the audience better perceive information progression than a static line graph.
- By temporal pattern visualization, we show that vaccine help reduce infection in a delayed timely manner.

2 GEOGRAPHIC CHOROPLETH MAPS

Geographic data usually visualize data with maps. During covid-19 pandemic, geographic visualization provide an effective way to show the pandemic information in United States. Most of the websites shows vaccination and infection data in two separate figures. Moreover, we find that data for some states (e.g. AL, HI) are not represented in those visualization. To address those problems, we want to integrate those two different data sources in a unified way and visualize data for all states in the United States. In this way, readers can have a better understanding of the effectiveness of vaccination.

We build a unified geographic choropleth maps using both vaccination and infection data collected at December 12, 2021. To simplify this situation, we only concentrate on state-level information. In the choropleth, we design two layers: one represents the infection rate, and another represents the vaccination rate. The legend on the upper right is automatically generated. We use United States state maps provided by `https://raw.githubusercontent.com/python-visualization/folium/master/examples/data/us-states.json`. Figure 1 shows a sample choropleth map or infection and vaccination rate in United States.

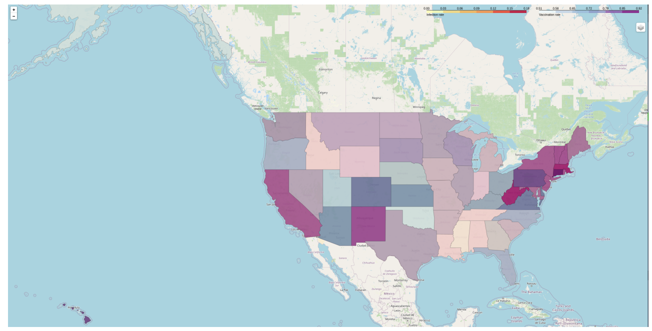

Figure 1: Choropleth map for Covid-19 for infection and vaccination rate in United States

One of the questions we are most interested in is learning how different colormap can change the effectiveness of the visualization. To address this problem, we test choropleth map with three different combination of colormap: *YlGn* & *BuPu*; *YlOrRd* & *BuPu*, and *YlOrRd* & *PuBuGn*. We evaluate the effectiveness of those three choropleth maps by asking them to answer the following questions:

- Which state has the highest infection rate?
- Which state has the highest vaccination rate?

We conduct a suvey with nigne participants. We divde them into three groups. Each group is asked to check one of those three choropleth maps for one minutes. They are then asked about the questions mentioned above. Unfortunately, none of those three groups answered the highest infection rate question correctly. For the highest infection rate question, the group with *YlOrRd* & *PuBuGn* achieves the best accuracy.

In the future, we will analyze more colormap combinations to investigate the best way to integrate those two different data sources.

3 ANIMATED LINE GRAPH

In class, we learned several static techniques to examine data change over time. Our project explores a new way - the animated line graph, which displays several charts states one after the other to illustrate the progression of information.

We build an animated graph based on over 300 days of covid data for five states: Washington, Illinois, California, Arizona, Alabama. The color red denotes Washington state; The color green denotes Illinois state; The color blue denotes California state; The color orange denotes Arizona state; The color yellow denotes Alabama state. The graph uses the auto locator and formatter from matplotlib to avoid creating an over-crowded x-axis. The results are shown in figure 2.

A simple survey of ten participants is conducted to understand better how well the audience perceives the information from the generated animated line graph compared to the static line graph.

*e-mail: zeyucai2@illinois.edu
†e-mail:qingyun4@illinois.edu
‡e-mail:qizeng2@illinois.edu

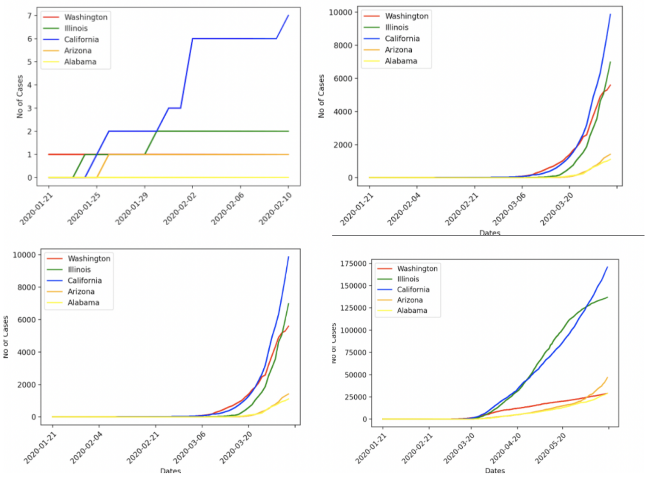

Figure 2: Animated Line Graph for Covid-19

We show five participants the animated graph and the other five participants for two minutes and then ask each of them identical five questions. The accuracy rate measures how well the participant perceives the graph. The group shown with the animated line graph has an 80 percent accuracy, and the group shown with the static line graph has a 40 percent accuracy rate.

Based on the result, We can roughly conclude that an animated line graph may helps the audience better perceive information progression than a static line graph.

4 Temporal Patterns

Vaccine is believed to be an effective way to stop the pandemic. When the number of distributed vaccinations grow, intuitively the affection numbers will gradually and eventually drop. Questions are posted that whether the vaccine has come into action in time as expected and how effective it is to stop the pandemic. To answer these questions, we propose to observe data in the perspective of temporal patterns. Specifically, we compare the vaccination and infection cases along timeline, as shown in Figure 3. Here we have a strong assumption that people get fully vaccined are unlikely to be overlapped with people get positive.

We collect covid-19 data from two sources: The daily state-wise vaccine data is from NYtimes `https://github.com/nytimes/covid-19-data` and the daily state-wise infection data is from the Covid Tracking Project `https://covidtracking.com/data/download`. Before data aggregation, we align the data by date and state and filter those invalid or incomplete ones.

Based on the above observation, we make hypothesis that the reduction of infectibility will reflect on later reduction of infectionability. To further analyze the delayed effect, we shift one of the distributions with a hyper-parameters to counteract the delayed effect, as shown in Figure 4.

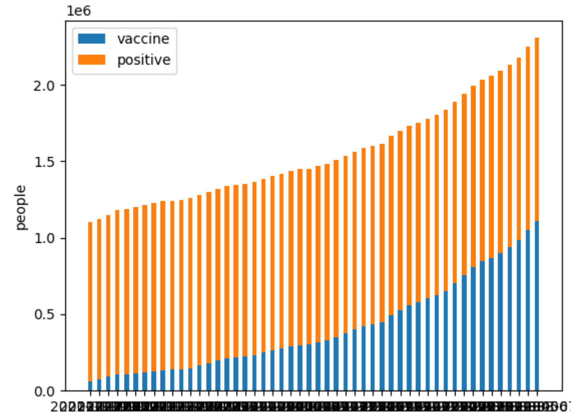

Figure 3: Comparison of the vaccination and infection cases along timeline in Illinois.

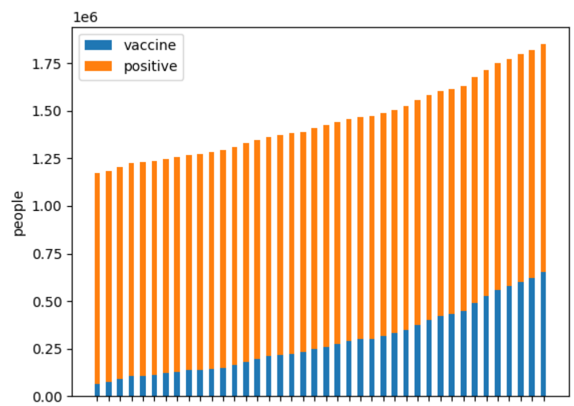

Figure 4: Comparison of the vaccination and infection cases with time shifting along timeline in Illinois.