The Effects of Vaccines on the Prevalence of Infectious Diseases in America

Team Member: Lei Xu(lx557), Zhiqi Guo(zg475), Yanli Zhou(yz1349)

Contents

Problem Statements

Vaccinations has greatly reduced the prevalence of infectious diseases and the low price of vaccines continues to improve the health of millions around the world. To see how we are benefiting from vaccinations, it is necessary to compare the suffering before and after the introduction of the vaccine. In this project, we

acquired data on 50 infectious diseases in America and calculated the density of infected population in each state every year over a span of roughly 10 decades. Finally we use data visualization to show how vaccination changes people's life.

Methods

Using this data, we provide you with two ways of visualizing the before-after comparison: heatmap and geo-heatmap.

1. Heatmap:

Heat map is a chart-like figure that provides an overview of the prevalence of each infectious disease over the years. On the x-axis we have year and on the y-axis we have different states. Each grid is colored corresponding to the proportion of infected individuals in the state population in that given year. The intensity of the color represents the fraction of population infected, the more intense the color the higher number of people infected.

There is also a salient vertical black line plotted in each heatmap representing the year of vaccine introduction. The idea is that through comparing the two parts of the heat map separated by the black line, you should be able to clearly visualize a decline in that disease. In this case, six kinds of vaccines are available for visualization.

2. Geo Heatmap:

A geo heatmap is a visualization of the prevalence of a particular infectious disease in each state directly through the map of the United States. In this program, you can generate a geo heat map by choosing a disease, a year and even a color you like. The resulting graph is a map with each state colored according to the relative infected rate of that disease.

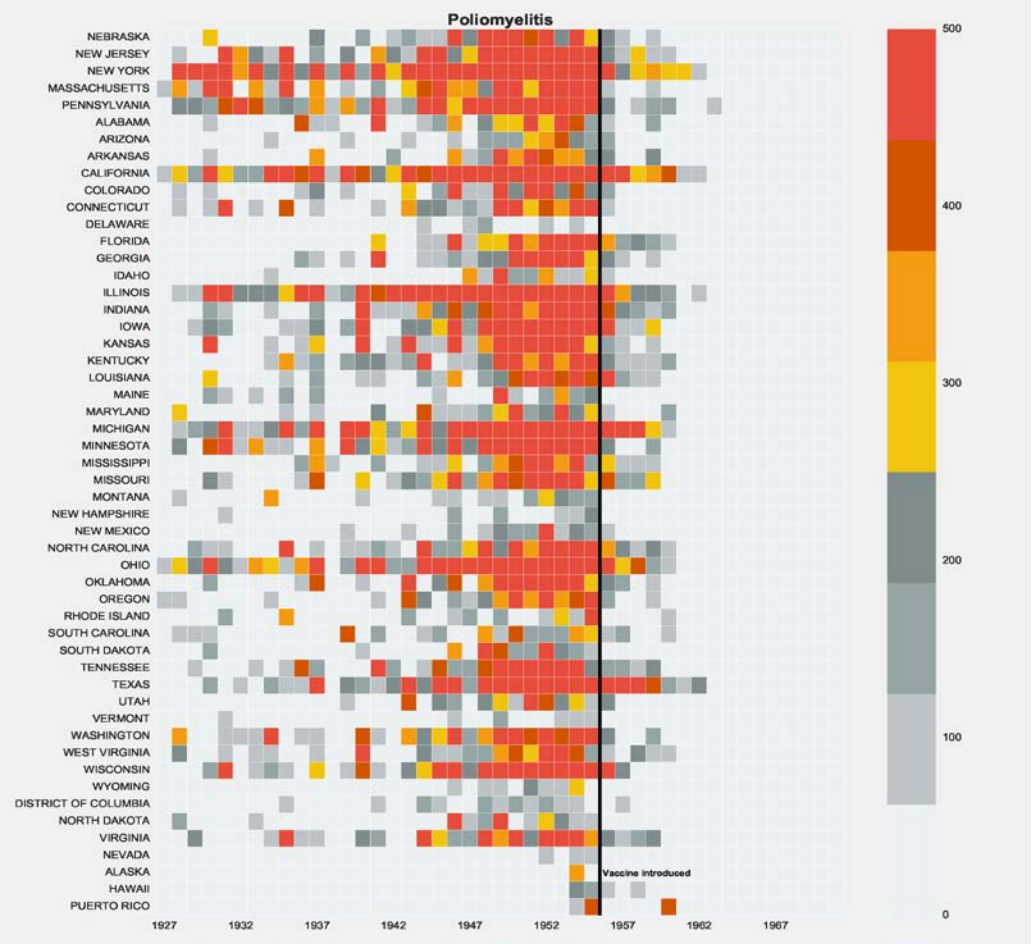
To have a better visualization effect, we modified the meaning of a strong color. Similar to heatmap but slightly different, a more intensed color in each graph represents a relatively higher infection rate in this year compared to other states. However, different graphs are not comparable in numbers of cases. In other words, two states with the same colors in two graphs don't mean they have the same number of cases in two years, but they both have relative high occurrence of cases. You will be able to see a difference in the geographical distribution generated from years before

and years after the introduction of vaccine. Combined with heatmap, you can have a general idea of how many cases and how cases are distributed for each disease. In this situation, more than sixty kinds of vaccines are available for visualization.

Results & Conclusion

We picked up vaccinations 'Poliomyelitis' to show the results.

1. Heatmap



Conclusion: It is clear that after the vaccine was introduced in 1955[1], the number of poliomyelitis has been greatly reduced over different states.

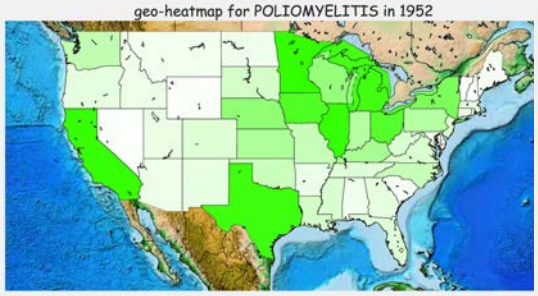
2. Geo heatmap

[1] From Wikipedia https://en.wikipedia.org/wiki/Polio_vaccine

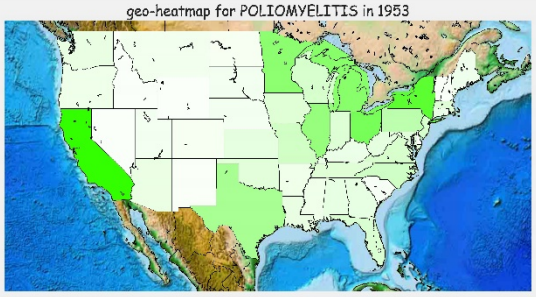
To see the effect of vaccines from geo-map perspective, we draw the geographical changes over the time period of 1952-1961.

We chose different degrees of color blue to show the relative density.

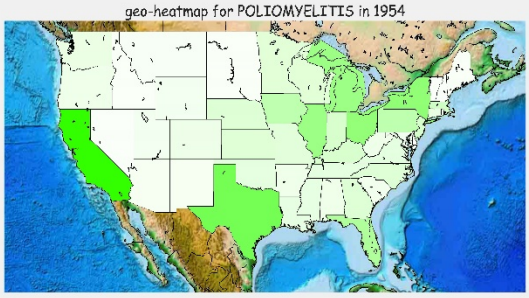
geo-heatmap for POLIOMYELITIS in 1952



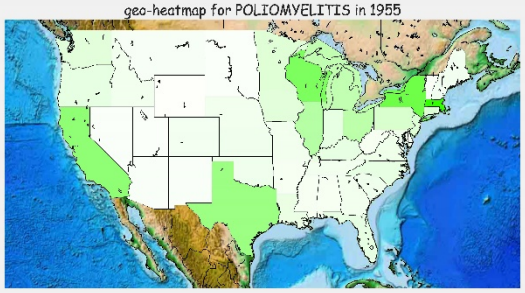
geo-heatmap for POLIOMYELITIS in 1953



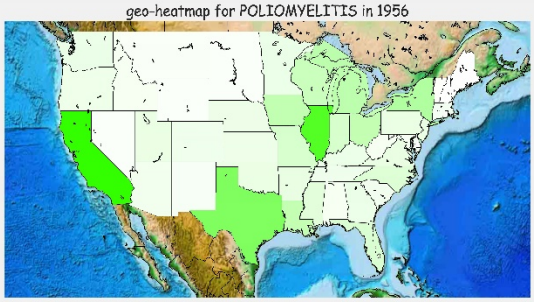
geo-heatmap for POLIOMYELITIS in 1954



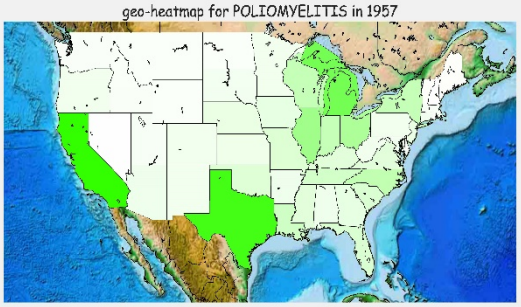
geo-heatmap for POLIOMYELITIS in 1955



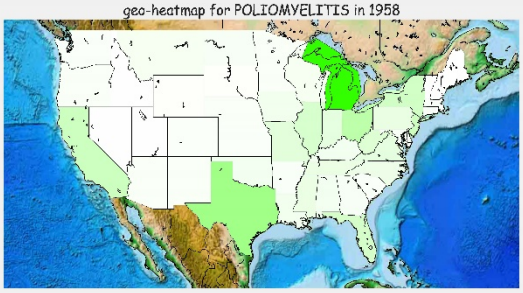
geo-heatmap for POLIOMYELITIS in 1956



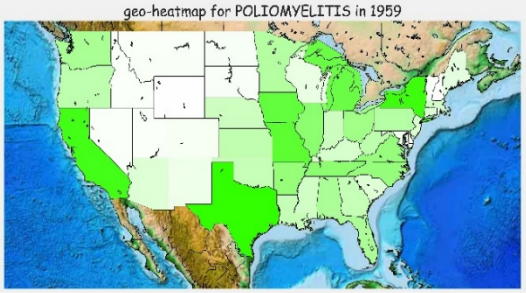
geo-heatmap for POLIOMYELITIS in 1957



geo-heatmap for POLIOMYELITIS in 1958



geo-heatmap for POLIOMYELITIS in 1959



Conclusion:

From the above, we can see there are three phases in the distribution of Poliomyelitis.

1. Before-introduction Phase: Most states have high cases of Poliomyelitis and it is quite evenly distributed;
2. Within-introduction Phase: There shows gap between different states. What's more, with heatmap we can see almost all states have fewer cases and different states have different reaction time. For those with high occurrence before, it took longer for the number of cases to decline.
3. After-introduction Phase: Most states have much lower cases of Poliomyelitis than before and it is quite evenly distributed again, with more in light color.

Appendix

1. **Website:**

To see the code, get the package and try it itself, please visit:

https://github.com/xlxulei1005/Infectious-Diseases-and-Effects-of-Vaccines

2. **Required packages**

Requires Python 3, and the following packages to function successfully:

- numpy
- pandas
- matplotlib
- Seaborn
- Basemap

3. **Sample walk-through of the main interface**

Execute in command line:

```
$ python vaccines_interface.py
You can type 'Quit' to exit the certain plot type. For example, if you are in
heatmap plotting mode, by input 'quit' you can exit the heatmap mode and back
to choosing plot type again
You can type 'Exit' to end the program
```

```
Please choose one of the following kind of plot:
 Heatmap or Geo-heatmap

> Heatmap

-Available diseases are Measles, Hepatitis_A, Rubella, Poliomyelitis, Smallpox
and Mumps.
Please enter the name of disease:

> Measles

Mean number of infected people before vaccine invented is:  313904.8
Mean number of infected people after vaccine invented is:  44216.0
```

Upon entering Measles as your chosen disease, the program will generate and store a pdf file that contains the heatmap visualization of the data of measles form 1928 to 1998. The program also prints out the mean number of infected people before and after vaccine introduction.

```
-Available diseases are Measles, Hepatitis_A, Rubella, Poliomyelitis, Smallpox
and Mumps.
Please enter the name of disease:

 > quit

Please choose one of the following kind of plot:
Heatmap or Geo-heatmap

> Geo-heatmap

-Enter the disease name to see the geo-heatmap of disease and vaccines.
-There are more than fifty diseases offered.

Please enter the color of heatmap, the color should be a float number from 0
to 1:

> 0.5

Please enter the disease you want to see, for example: MUMPS,MEASLES or input
'help' to see full list of available diseases

>MUMPS

The year range for this disease is from 1968 to 2014
Please enter the year

> 1970
```

In geo heatmap mode, the user is prompted to enter a disease, a color, and a year. For the program to run successfully, please make sure you enter: 1. a valid disease name from the provided list, which you can obtain by inputting 'help' in the command line to see a full list of available diseases; 2. a valid floating number between 0 to 1 (corresponding to 0-360° HSV cylindrical-coordinate

representations of points in an RGB color model.) that represents the color you wish the map to have; 3. a valid year from the range provided by the program. The program should then produce and save a pdf file that contains a geo-heatmap of the corresponding disease, color and year. Note that this process might take up to 15 seconds.

Unittest is also supported in this program. Simply call in command line:

```
python -m unittesttest.py
```

4. Acknowledgements

- All disease data was obtained from: Project Tycho
- A note on 'shapefile.py': This is not original code by members of this team. This file is added to support the geo-map plotting function.