Influenza in the US

1. Objective

The objectives of the project are to visualize how influenza interact with different features and data we extract in the US.

We want to show the correlation between ILI (Influenza-like illness) and some features listed below:

- Temperature
- Vaccine receiving percentage
- Under age 65 without health insurance percentage
- Hospital number

Thus, in total we created 5 figures, which worked on the correlation between ILI and different features to show the result.

2. Members and Task Contributions

LIANG Yangxiao (3035418775) :

- Figure of monthly ILI percentage and monthly mean temperature of Alabama & Alaska in 2011
- The trend of US states (except Arizona & Florida) ILI percentage in 2011

LIN Shaozhen (3035420194) :

- Influenza vaccine receiving percentage of different age groups changes over years
- Figure of ILI report percentage and percentage of persons under age 65 without health insurance in the US (except Arizona & Florida) by state over 2011 - 2014

LI Qin (3035419511) :

- Figure of ILI & hospital numbers

3. Report Organization

This report is organized by showing these reports in succession and aims to offer help to health officials and application developers who seek visualized big data about influenza and to the public to seek information in order to get effective control of influenza and knowledge of prevention.

The trend of US states (except Arizona & Florida) ILI percentage in 2011

Visualizations that we have created

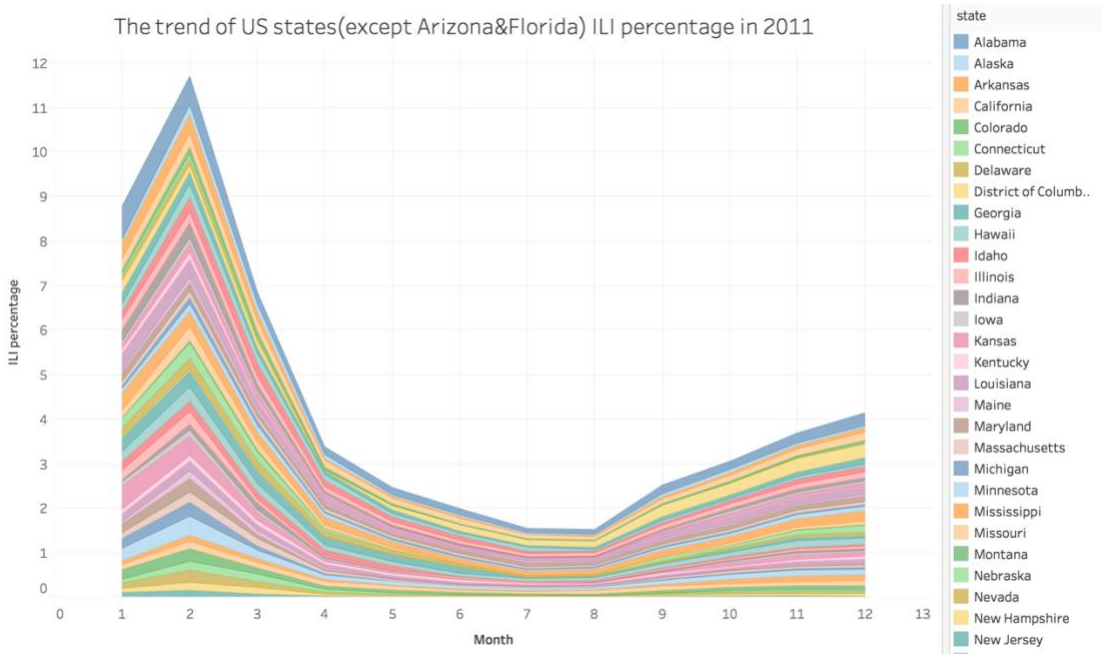


Figure 1

Figure 1 illustrates the trend of US states (except Arizona & Florida) monthly ILI percentage in 2011 in the form of zero-basedline stacked area graph.

There is only one main attribute-(Monthly) ILI percentage. For ILI percentage, it is the ratio of the number of cases of ILI to the total number of patients per month in all states (except Arizona & Florida).

the purpose and effective of your visualizations

Figure 1 has the main purpose of descriptive with showing the trend of monthly ILI percentage by states in 2011. It takes advantage of the feature of stacked area graph to show not only the monthly ILI percentage of the states which are color-coded, but also the overall annual ILI trend in the United States. Across the country, we can see outbreaks in January and march. It was the worst in February. But since March, the number of ILI percentage has

fallen sharply, until July and August, the period of ILI was low. By the end of the year there was a gradual rebound.

In the Tableau dashboard link we provide, you can use the mouse to select the state you want to focus on or use state filter to see some specific states with doing comparisons.

Through this graph, we can get some useful guidance. For example, the government and related health authorities can use the data from previous years to predict when the influenza will break out in the next year to take precautions.

Difficulties that you have encountered

The main difficulty is to use the weekly ILI Numbers and the weekly total number of patients for all the U.S. states to recalculate the monthly ILI percentage. This is the most time-consuming part.

Different methods that you have tried and Justification of your choices

Prior to this, we had tried to use line graph and map to represent the states' ILI percentage. But we found that the line graph is not as beautiful as the stacked area graph, although we think their ability to display data is comparable. For a map, it is difficult to show the change of data on time scale, which cannot satisfy the requirements we want to meet.

Anything you wanted to do but haven't? Why?

We want to do population density distribution on the influence of the ILI outbreak, but after the figure analysis, we found that the population density for each state is not very uniform, such as some states of the population concentrate in a small area e.g. a city only. However, the regional granularity of the ILI data we have is based on state. In a word, because the ILI data of cities, counties and other geographical units cannot be obtained, the relationship between the population density distribution and the ILI percentage is distorted in the graph to a certain extent.

Figure of monthly ILI percentage and monthly mean temperature of Alabama & Alaska in 2011

Visualizations that we have created

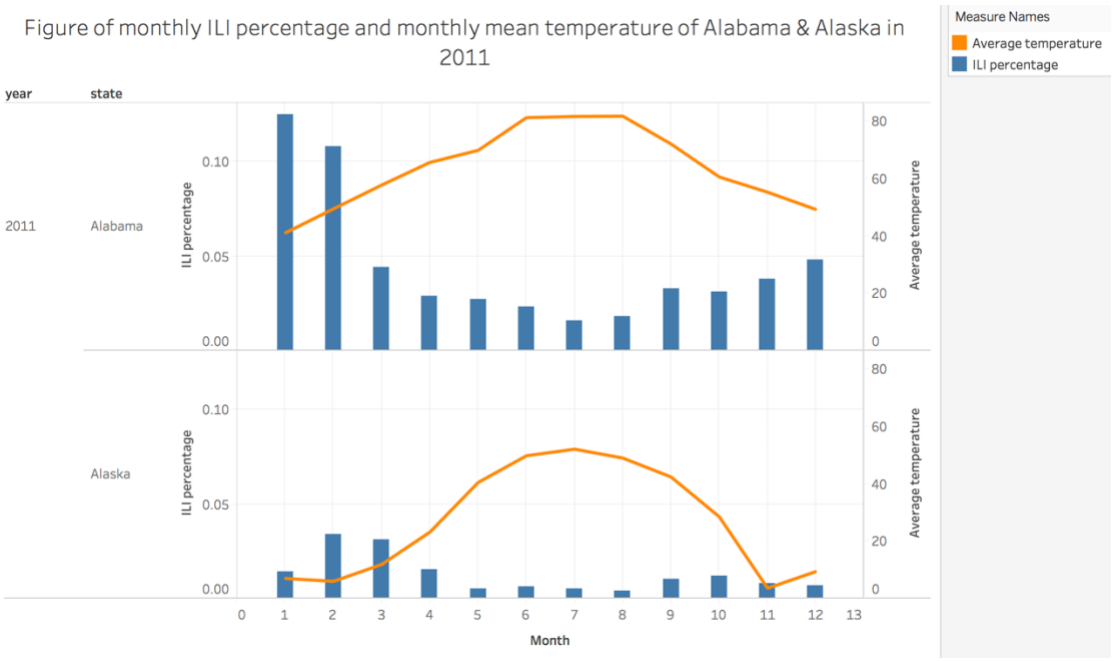


Figure 2

Figure 2 illustrates the correlation of monthly ILI percentage and monthly average temperature of Alabama & Alaska in 2011.

There are two main important attributes-(Monthly) ILI percentage and (Monthly) Average temperature. For ILI percentage, it is the ratio of the number of cases of ILI to the total number of patients per month in Alabama and Alaska. Based on these two attributes in this graphic, we find a new insight explained below.

the effective of your visualizations

In order to show temperature has a universal effect on ILI percentage, we chose to show the data of Alabama and Alaska whose geographical positions and population density are in different a lot. Under this context, we still find

the fact that the ILI monthly percentage is inversely proportional to the monthly average temperature. Take Alabama for analysis, the lowest temperature in a year occurs in January in which there was the highest ILI percentage. Behind significantly with the temperature rising, ILI percentage has plunged trend between February and march. Until June, July and August, because temperature had maintained a high level, ILI percentage was in a state of low and stable. By the end of the year, with the gradual decline of temperature, ILI percentage also gradually declined.

In the Tableau dashboard link we provide, you can use the mouse to select the month you want to focus on, and see the specific data, whether it's for the bar chart or the line graph.

New insight and purpose

In conclusion, Figure 2 has the purpose of explorative with showing the inverse relationship of monthly ILI percentage and monthly mean temperature of Alabama & Alaska in 2011.

Difficulties that you have encountered

Because ILI data and temperature data have different sources with different time units. In order to be able to show on the chart, we need to recompute the weekly ILI data to be monthly. Therefore, I divided 52 weeks into 12 months to calculate the number of ILI patients and the total number of patients. It's a lot of work for us. In order to save time and to show new insight clearly, we only show data from some states in 2011 like Figure 2 showing.

Different methods that you have tried

We have tried to use a line trend to represent the ILI percentage, and the average temperature of the month is represented by the line color of different scales. By contrast, this form of readability is not as good as Figure 2.

Justification of your choices

To compare the annual trends of two attributes, we use histograms and

graphs. We believe that the histogram and the graph are clearly showing the trend over time. At the same time, we believe that it is most appropriate to display one year's data per month instead of week. Choosing the right time granularity can help us control the details of the display.

Anything you wanted to do but haven't? Why?

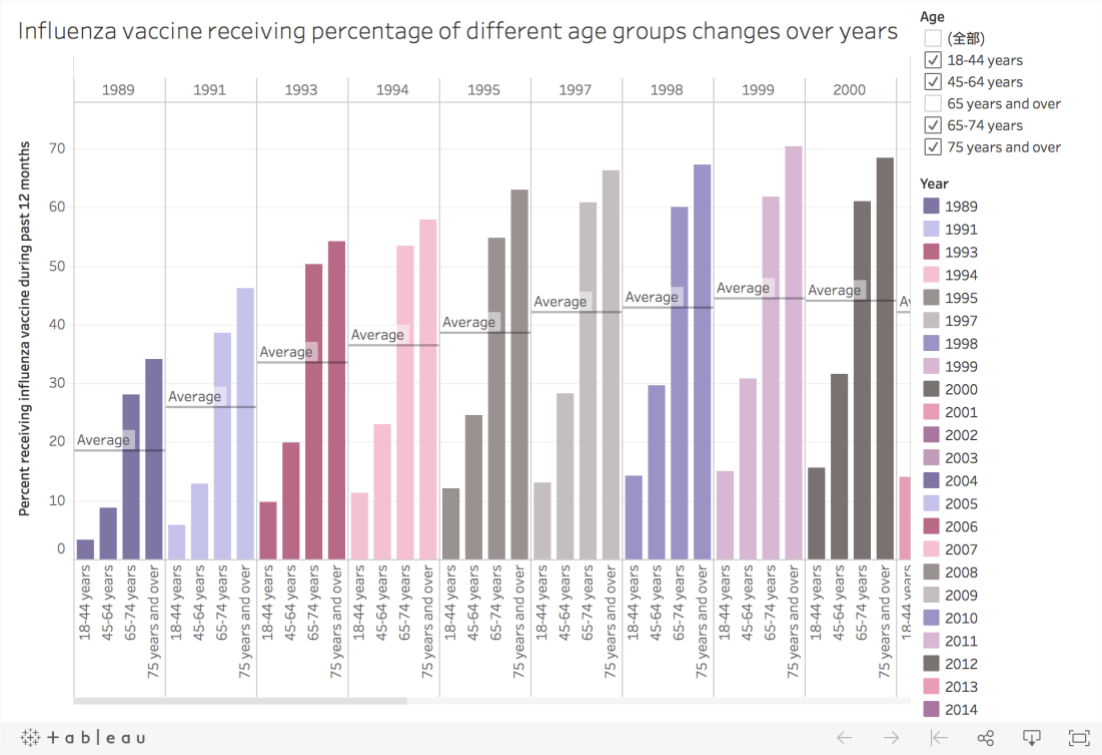
We would also like to show the effect of humidity on the monthly ILI percentage, one of the main factors affecting influenza. However, no suitable humidity data was found. Therefore, the PCP (Precipitation Index) which refers to rainfall was used instead. There was no special relationship between them after the graph was shown. This is the reason why we did not show this in the final visualization. In a word, we think PCP is not a suitable feature to replace the humidity feature.

Any limitations to the existing tool?

We think that Tableau has certain restrictions on data preprocessing, and we need to calculate it in excel in advance to a great extent.

Influenza vaccine receiving percentage of different age groups changes over years.

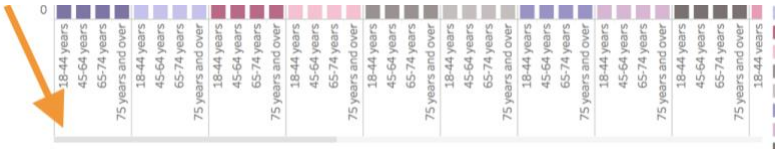
Visualizations that we have created



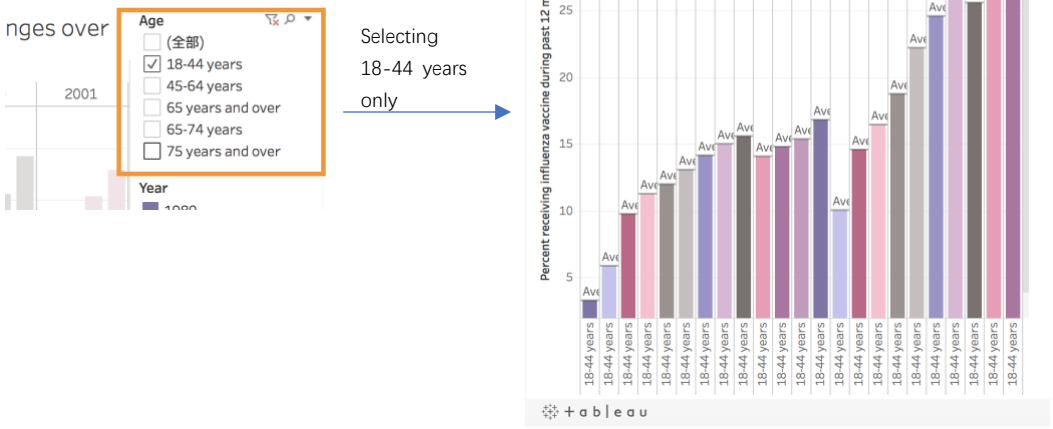
the effective of your visualizations

The visualization shows the figure of the influenza vaccine receiving percentage of different age groups changes over years. To be specific, the figure covers age groups in the following categories: 18–44 years, 45-64 years, 65 years and over, 65-74 years, 75 years and over. Time ranges from 1989 to 2014. And the measure value is the percentage receiving influenza vaccine during the past 12 months under different age groups and years. Additionally, the average values of different years are shown for reference.

By positioning mouse on the figure, the corresponding data will be shown. Users can use the horizontal scrollbars to see figure of other years.

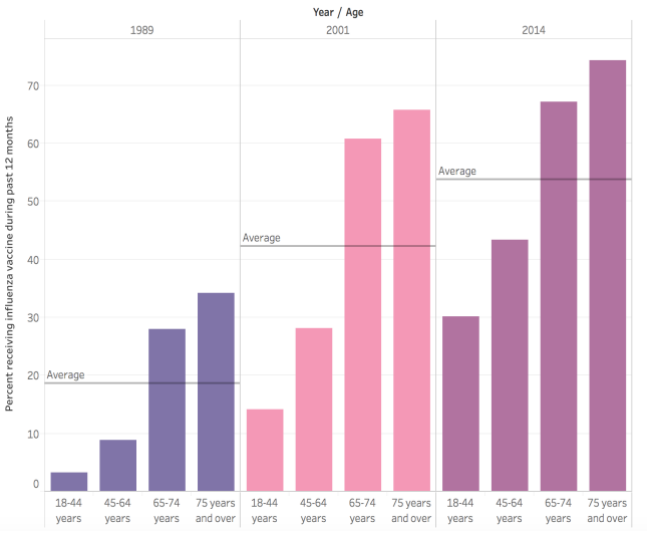


The tick boxes on the upper right enables users to select the age group they are interested for further exploration.



New insight and purpose

a. In general, the distribution patterns of the percentage receiving influenza vaccine during past 12 months (hereafter in this report will be abbreviated as 'flu vaccine percentage') under different age groups over years are similar, here the flu vaccine percentage increases as age becomes larger, and the greatest gap appears between 45-64 years group and 65-74 years group. In other words, it seems that 65 is a division, where people younger than 65 have lower flu vaccine than that of people aged 65 and above.

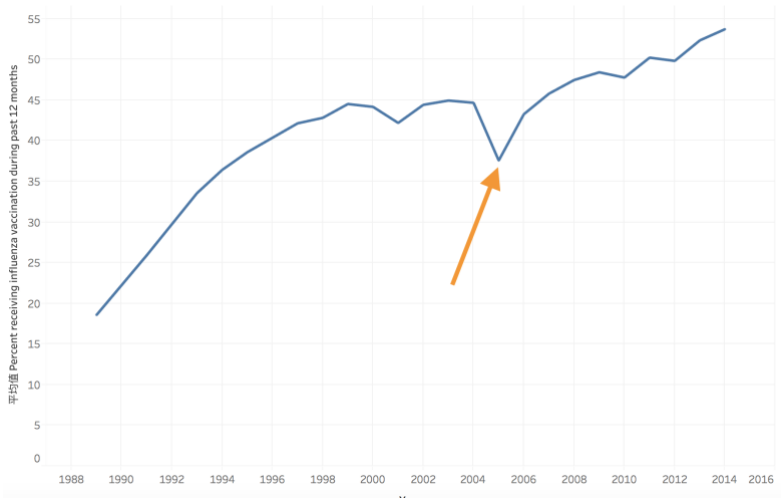


General distribution pattern of flu vaccine percentage

b. By analyzing the average flu vaccine percentage of different years, we can see that the overall trend of flu vaccine percentage in the US over years is growing,

which implies that more and more people are receiving flu vaccine and be aware of the importance of flu vaccine.

However, we noticed an interesting change in the figure. There is a sudden decline in 2005, which is very abnormal.



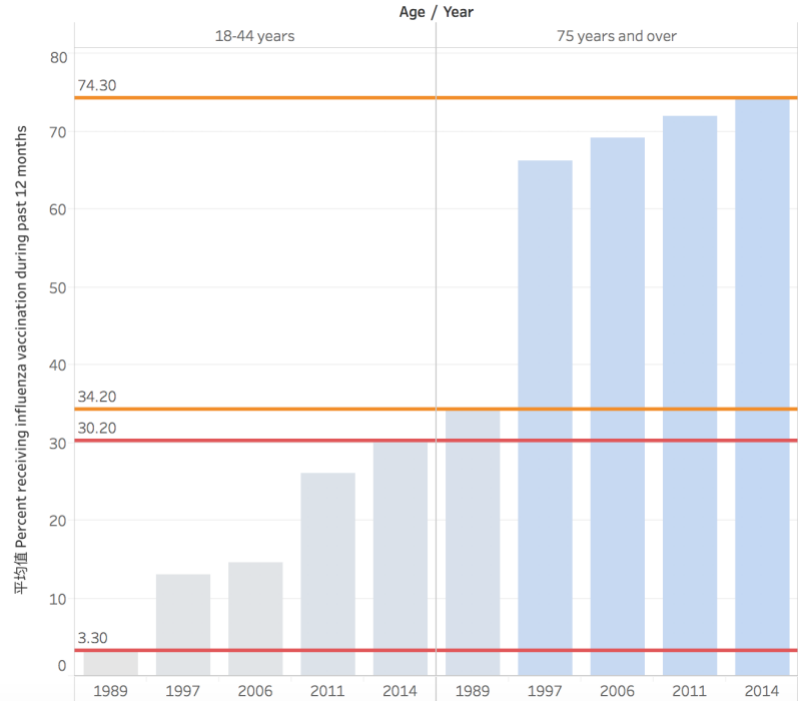
Average flu vaccine percentage in the US over 1989-2014

By doing research, we believe that the sudden decreasing in the figure is very possible as the result of the global spread of (highly pathogenic) H5N1 in 2005, which is considered a significant pandemic threat.

c. By analyzing different age groups' flu vaccine percentage over years, we found out that two age groups contribute most in the growth of the flu vaccine percentage in the US.

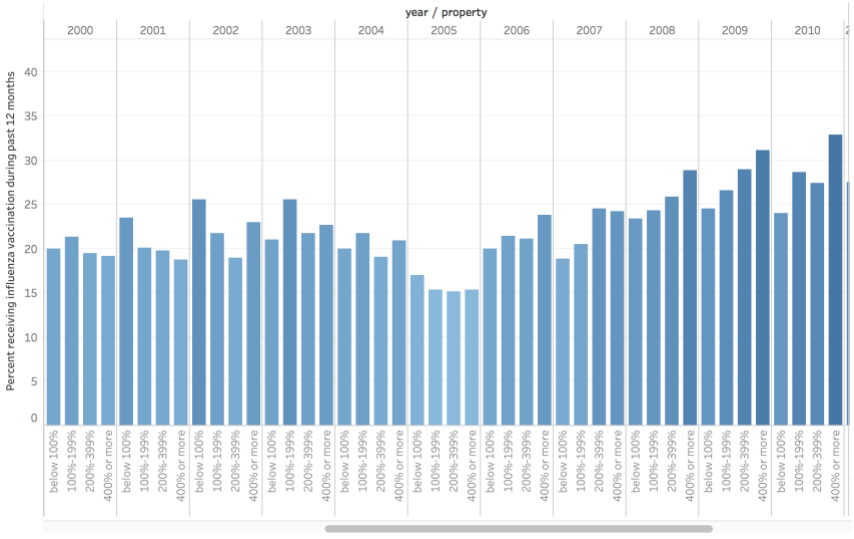
The first age group is the 18-44 years, of which the flu vaccine percentage is 9.2 times bigger in 2014 than in 1989. 18-44 years group owns the greatest growth compared with its initial value.

The second age group is the 75 years and over, which owns the greatest amplitude of 40.1%.



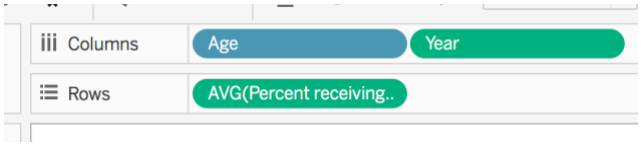
The visualize process (with difficulties, different methods and justification)

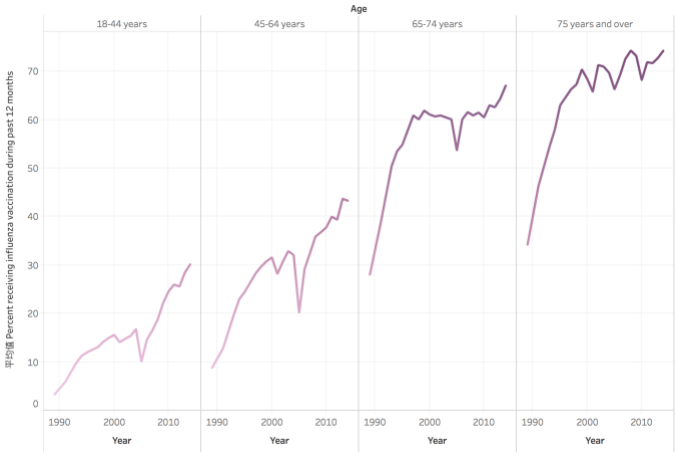
When designing the visualization of this part, I encountered some difficulties. The first is data filtering. The source data from Centers for Disease Control and Prevention has multiple characteristic including age, sex, race, percent of poverty level etc. I firstly selected age and percent of poverty level out of them because the data of others is not enough to discover insight or irrelevant to the topic. Then I tried to visualize both. The visualization in terms of the percent of poverty level doesn't give too much insight, thus I pick age as my choice.

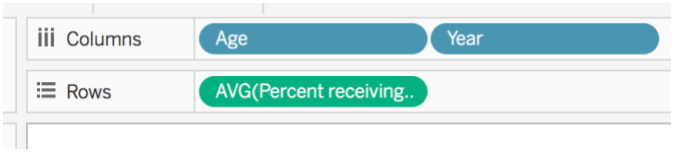


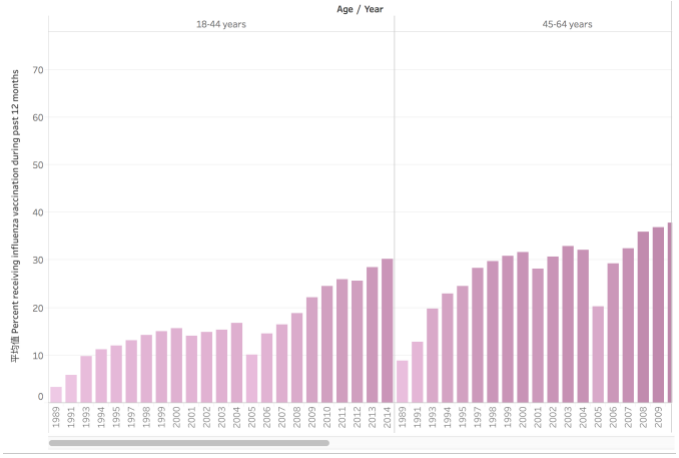
The second happened during the usage of Tableau. Being unfamiliar with Tableau at first, I failed to create the ideal visualization in the beginning. There are lots of key tips in Tableau, including filtering, dimensions order, discrete or continuous setting etc. that we should pay attention to.

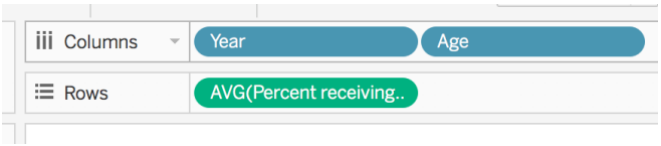
The following example shows the one of the processes that I encountered difficulty, solved the problem step by step and got the corresponding visualization results.











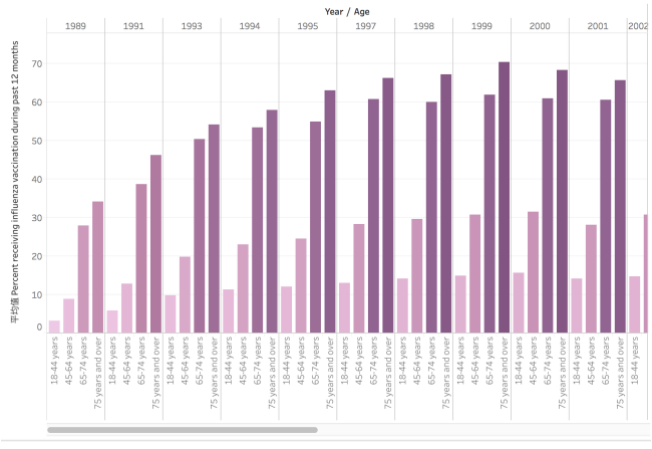
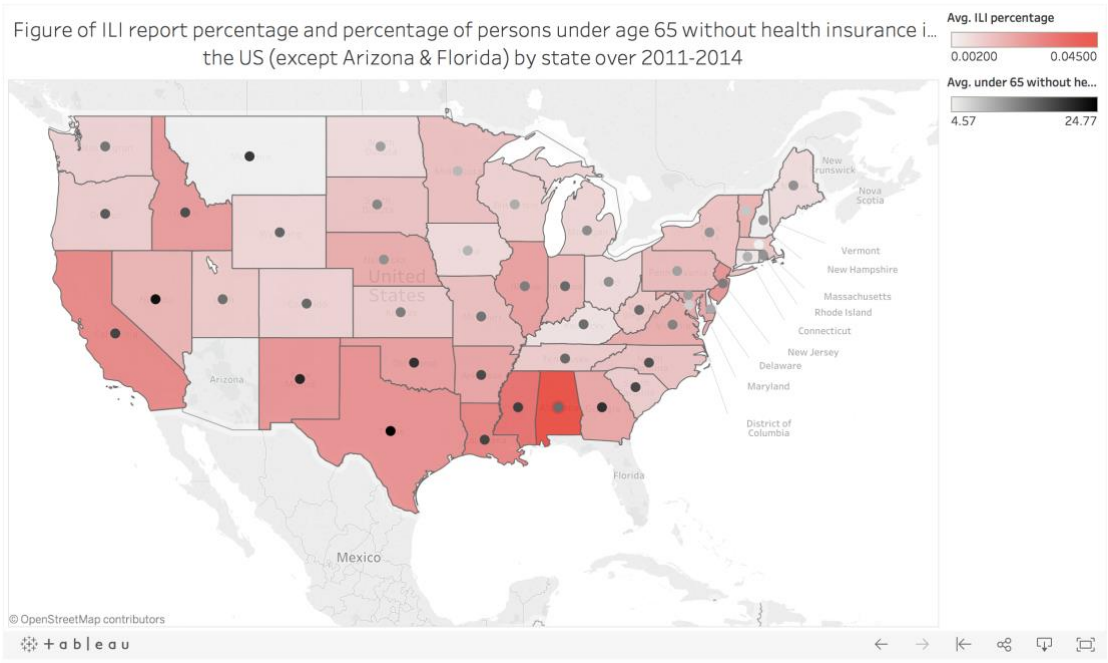


Figure of ILI report percentage and percentage of persons under age 65 without health insurance in the US (except Arizona & Florida) by state over 2011-2014

Visualizations that we have created

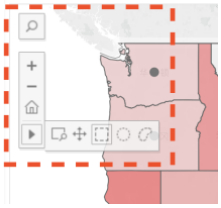


the effective of your visualizations

The visualization shows the figure of ILI (INFLUENZA-LIKE-ILLNESS) report percentage and percentage of persons under age 65 without health insurance in the US (except Arizona & Florida) by state over 2011-2014. To be specific, the color covering the states indicates the ILI percentage, where the darker the color is, the higher the ILI percentage is. The round dot on the state indicates the percentage of persons under age 65 without health insurance (hereafter in this report will be abbreviated as 'no insurance percentage'), where the darker the color is, the higher the no insurance percentage is.

By positioning mouse on the figure, the corresponding data will be shown.

Functions of search, zoom in, zoom out etc. can be realized on the upper left.

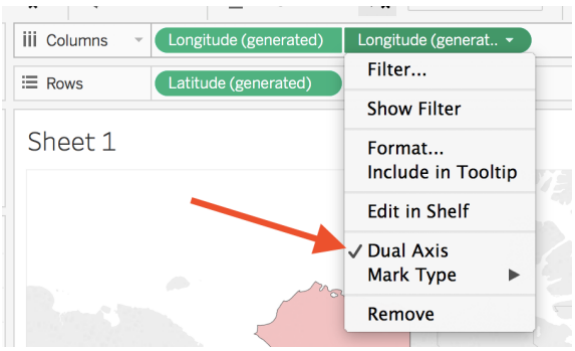


New insight and purpose

a. Overall, the state with dark dot has dark background color.
b. However, Montana is an abnormal case with high no insurance percentage but low ILI report percentage.
c. Since no insurance percentage can be used to describe the medical development level to some extent, we can draw the conclusion that in general, greater development of medical and the rising awareness of public health can be helpful to reduce the odds of influenza like illness, which will definitely be helpful to reduce the probability that the citizen has influenza.

The visualize process (with difficulties, different methods and justification)

The difficulty I encountered also happened during the usage of Tableau when I tried to show two measures in the same state in the map (show both the background color and the black dot in one state in this case). Dragging two longitudes to the columns didn't help. After searching on the Internet and exploring by myself, I figured out that the key to fix it is to select Dual Axis on the second longitude. By doing so, I am able to visualize the data the way I want.



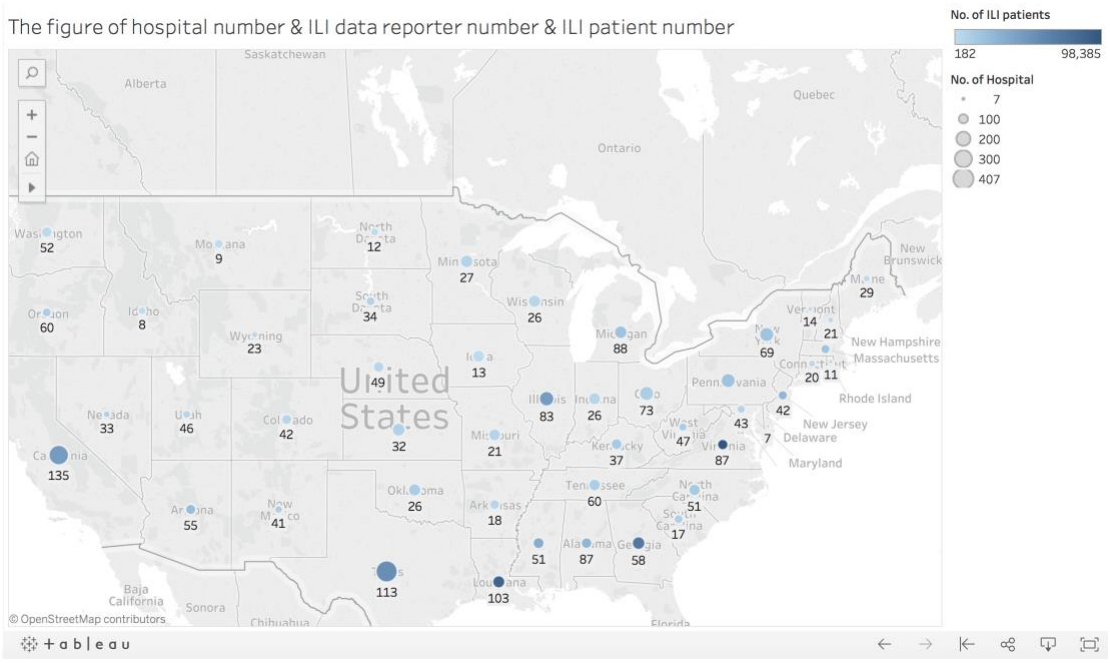
Small regrets

The source data of this visualization includes two files, which are both states. But data for ILI report percentage is weekly data so it took us some efforts to organize the data by year.

However, we failed to show yearly change in this visualization because of the limits of Tableau and our capability.

The figure of hospital number & ILI data reporter number & ILI patient number

Visualizations that we have created



the effective of your visualizations

In this graph named “figure of ILI & hospital numbers”, we have extracted the figure of how many hospitals in each state and the number of ILI patient of each state, as well as the number of data reporter statistics, that means information provider have been visited within hospitals in each state.

We can see from the visualized information graph, where circle dots used to illustrate the number of hospital, and we used color that changes from light to darker to indicate the change in data from small to large, this to say the darker the color the bigger the ILI patient number.

Form this ratio we can observe that influenza density in each state, this is meant to figure out the availability of medical equipment if it is sufficient. Where there are small dots with darker color, it is obviously conveyed the message that in these specific states the availability of medical equipment

might be in tense, because there are only a handful of hospitals in the city, but a large number of patients are received.

New insight and purpose

We also observed a particular phenomenon that dots with darker color are likely alongside the coastline. Based on this observation, we might conclude that humidity and wind speed have a direct impact on influenza infection.

We suggest that the government should strengthen the defense measures in the area of reinfection and improve the construction of medical equipment and equipment in cities with high morbidity density and less hospital construction. The most severe cities are Virginia, Georgia and Louisiana, as we can see on the information graph, these cities are represented by very small dots with really deep color, especially Virginia, where 87 hospitals received 98,385 ILI patients.

However, there is also a particular finding, in inland city Illinois, which do not along the coast, but the density of the disease is extremely high. This is an interesting finding that is different from the observed pattern we call it an outlier.

Difficulties that you have encountered

Accurate data search is difficult. For example, the hospital that gives data reports every week has some repeated reports, it must be done with excel skills, such as sum up and filtering. Some data are missing, such as Florida, where there has 0 reporter and ILI patient number, so we must delete the data of this area. Moreover, make full use of geographical classification includes relations show different geographical levels compared to the charts and the graphical interface, chart has only a limited word graphics guideline, cannot explain the context of the data very well

Different methods that you have tried

We tied use color to represent hospital number, but if use this method there are not obvious difference, because the range of hospital number is from 7-

407, whereas ILI number range is from 0-98385. So, the size of the dots representing them can be obvious.

Justification of your choices

We used symbol map to compare the number of hospital & ILI data reporter number in each state and the number of ILI patient. There are three attributes, reader of this graph can move the cursor to see which part's figure they interested. We believe that the symbol map can vividly illustrate the hospital's work overload by directly look at the size of the dots and the color on them. At the same time, we believe that it is the most appropriate way to display distributed figures and observe findings and patterns.

Anything you wanted to do but haven't? Why?

We wanted to find out how gender could differentiate affecting the spread of influenza, we wanted to see if female or male are more easily infect with flu, because lack of supporting data we haven't do the comparison, we can only obtain the total number of patients, but not the divided proportion on male or females.

Any limitations to the existing tool?

We think the Tableau has certain restriction on show me panel, there are limited layout to be choose, the function is simpler compared to D3.JS.