

Thanksgiving Data Visualization

1. Data source information

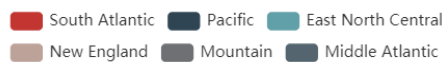
The dataset used in the homework is a survey about thanksgiving holiday in 2015. The researchers asked approximately 1000 respondents on Nov. 17, 2015 questions about their Thanksgiving in aspects of eating and traveling etc. In this visualization work, I used several variables including the **type of the main dish**; **types of side dishes** typically served at Thanksgiving dinner; **how far people travel** for Thanksgiving and their **US region**.

Original data link: <https://github.com/fivethirtyeight/data/blob/master/thanksgiving-2015/thanksgiving-2015-poll-data.csv>

This directory contains the data behind the story [*Here's What Your Part of America Eats On Thanksgiving*](<http://fivethirtyeight.com/features/heres-what-your-part-of-america-eats-on-thanksgiving>).

2. Visualization result

The visualization tool is python3, API: pyecharts(<https://pyecharts.org/#/>). Pie, Map and Heatmap are 3 types of graphs provided to show the data. These 3 graphs are saved as html format. **All of them is interactive**. Legends bottom and visual map module can be clicked and slide respectively to show the related data or not.

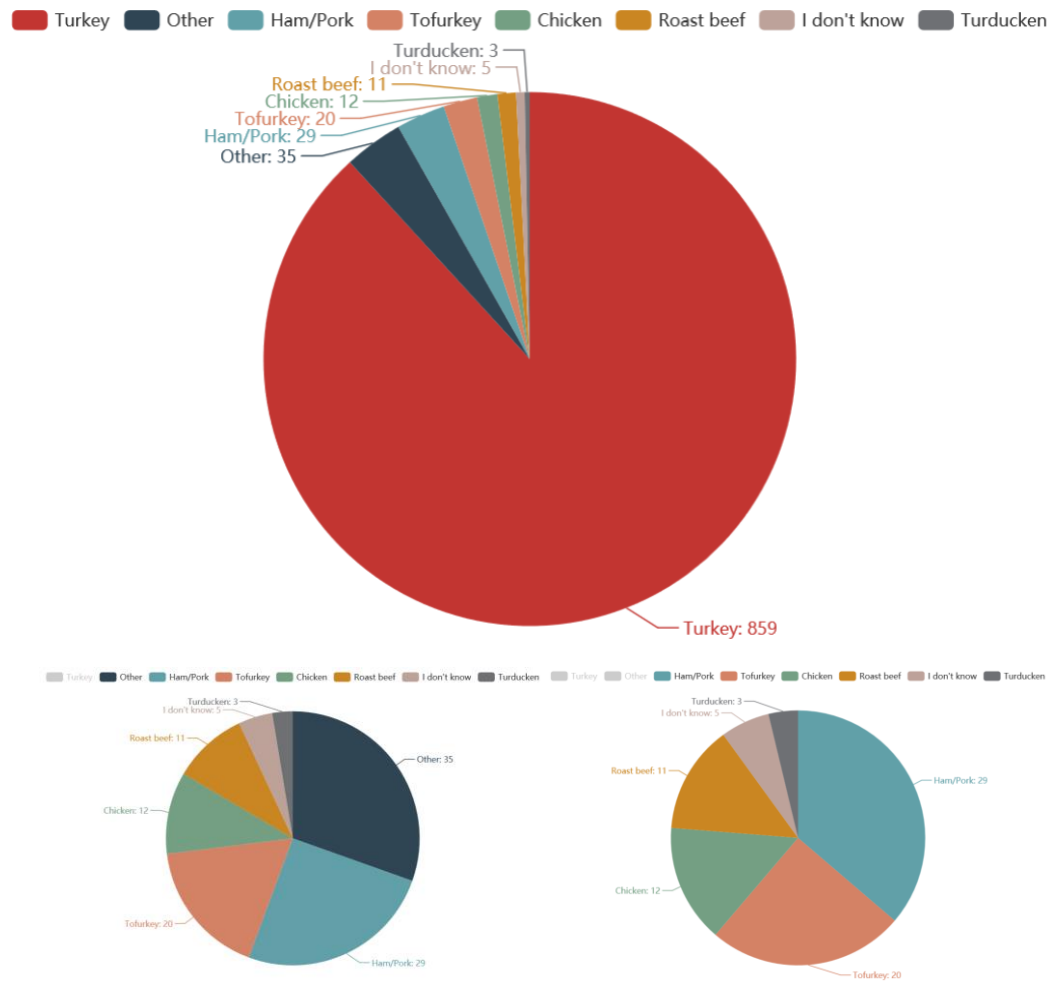


(Legend bottom)



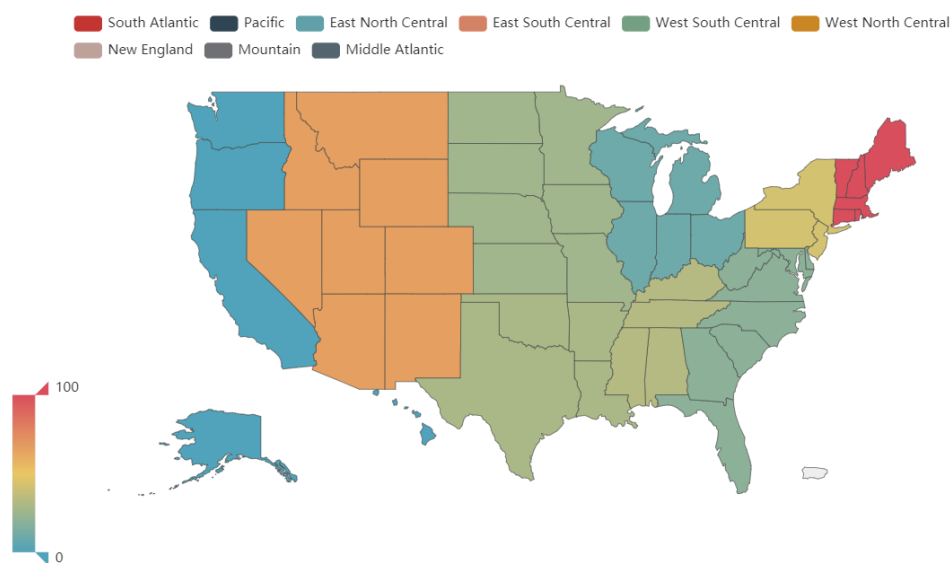
(Visual map module)

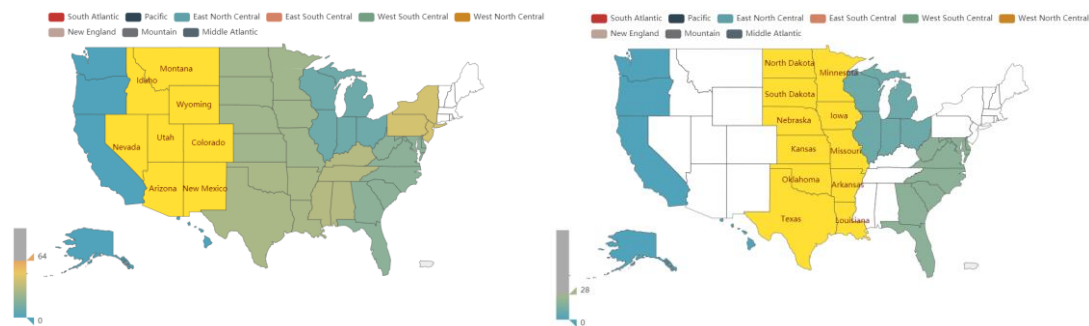
The screen shoot of 3 data visualization results are as follow:



(Source file: Pie.html)

Pie: In this graph, the variables “main dish” is used. The numbers behind each dishes are the amount of each type main dishes respondents choose.





(Source file: Map.html)

Map: In this graph, I put 4 types of answer into 4 different scores. The table below shows the switch:

Thanksgiving is happening at my home--I won't travel at all	0
Thanksgiving is local--it will take place in the town I live in	25
Thanksgiving is out of town but not too far--it's a drive of a few hours or less	75
Thanksgiving is out of town and far away--I have to drive several hours or fly	100

Then I counted the responses from different regions and calculated the average score. Finally, I normalized the scores and colored this map according to the size of the scores.



(Source file: HeatMap.html)

HeatMap: In this graph, I evaluate people's choices of 14 side dishes in 9 regions of the United States. The numbers shown in the figure are amount of chosen side dishes per 100 people. Due to the restriction of API version, names of side dishes on X axis are not completely shown. But these names can be seen when clicking the data bar.

3. Analysis

For the Pie figure, the proportion of each main dishes is shown clearly. Data is one dimensions and contain different names of main dishes. We can easily have a view of the relationship between the magnitude of the terms in the pie chart. Pie chart is used to study the proportion of components in a population. However, this type of graph is not suitable for large data set (classification) presentation. And the data cannot have negative values. In addition, when the proportions are close, it is difficult for the human eye to distinguish accurately.

For the Map figure, the data is two dimensions, one dimension is respondents' region and the other dimension is the evaluation of their willing to travel on thanksgiving day. Although this dimension the 4 types of answer rather than numbers, I transform the answers into 4 scores in ascending order. Map information is visual and intuitive. Maps are suitable for displaying data sets with geographic information, which is usually aggregated and continuous information in a region. Combined with the map, it can display the geographic data step by step intuitively. It is easy to judge the size of the measurement by color depth. The disadvantage of map includes: the data must contain geographic information and is aggregated; The values displayed are imprecise, and the bubble size and color depth are close to each other; The geographic area size is independent of the measurement and is easily misinterpreted.

For the Heatmap, the data is three dimensions: US region, types of side dishes, and the number of chosen dishes. In original data, the number of chosen dishes is not displayed. We counted the numbers from different areas. Compared with the other two charts, the heatmap can contain more information in dimensions and amount. It can show different meaning when we looking in horizontal or vertical direction. While the numbers of chosen dishes show the preferences of every single area, it can also be seen that which kinds of dishes is more popular than the rest on a national scale, which is judged by the color depth in vertical direction. For example, mashed photos and rolls/biscuits are almost red or yellow. And in horizontal direction we can tell that which dishes one area likes the most.