Eman Tahir

1718

**DAV Assignment 01**

**Data Download:**

Berkley Earth:  
<https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data>

**Data Cleaning:**

All the data has been cleaned during using tableau prep and then is used for visualizations. The workflow is submitted along with the folder.

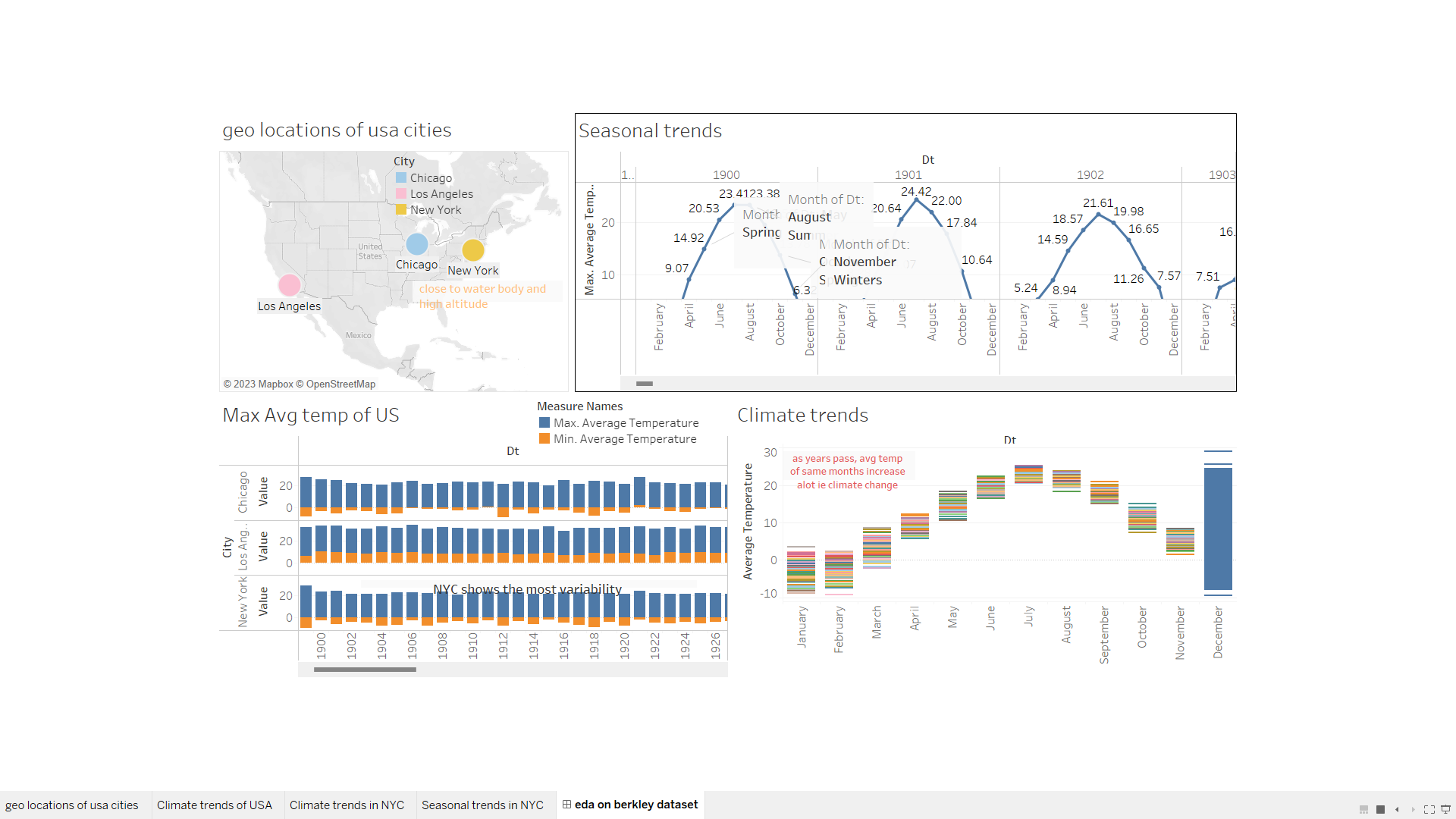
**Region Selection:**

The region I selected was NYC i.e. New York. The reason for this selection was

* It showed the most variability in terms of temperature
* The 4 seasons were observed
* There were drastic climatic changes over the years
* Its datasets were readily available

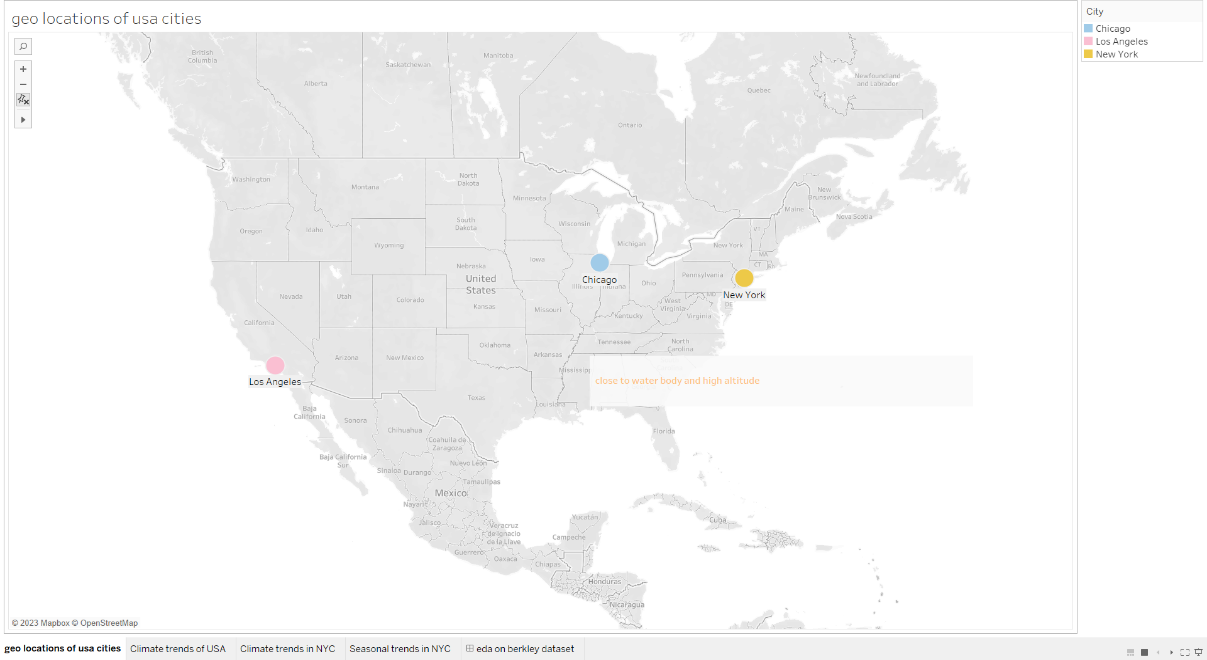
These findings are depicted in the dashboard created for Berkley Dataset.

**Dashboard design:**

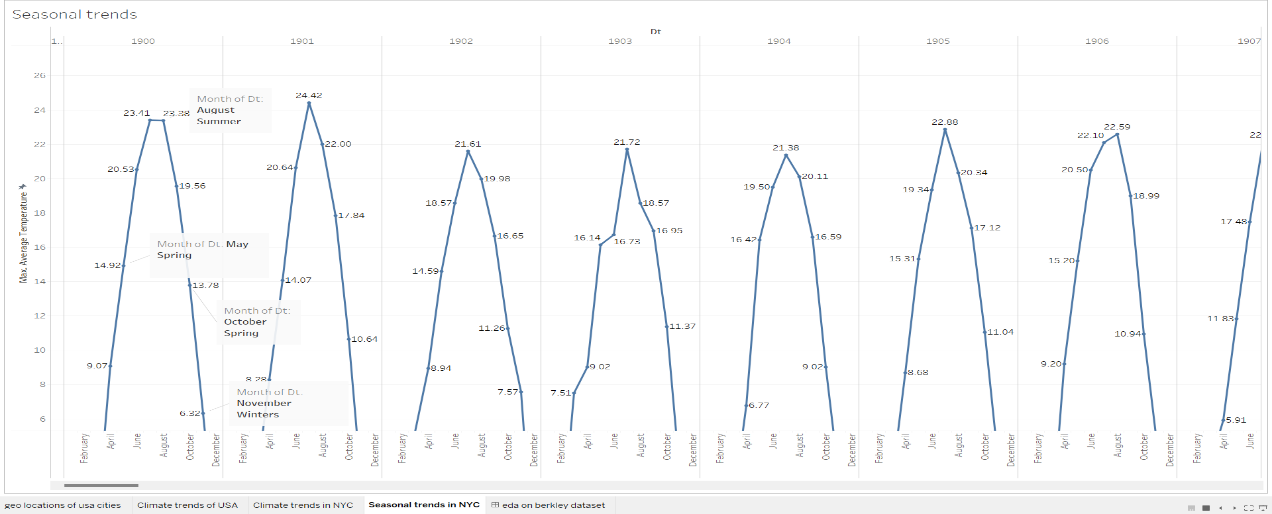


**Exploratory Data Analysis:**

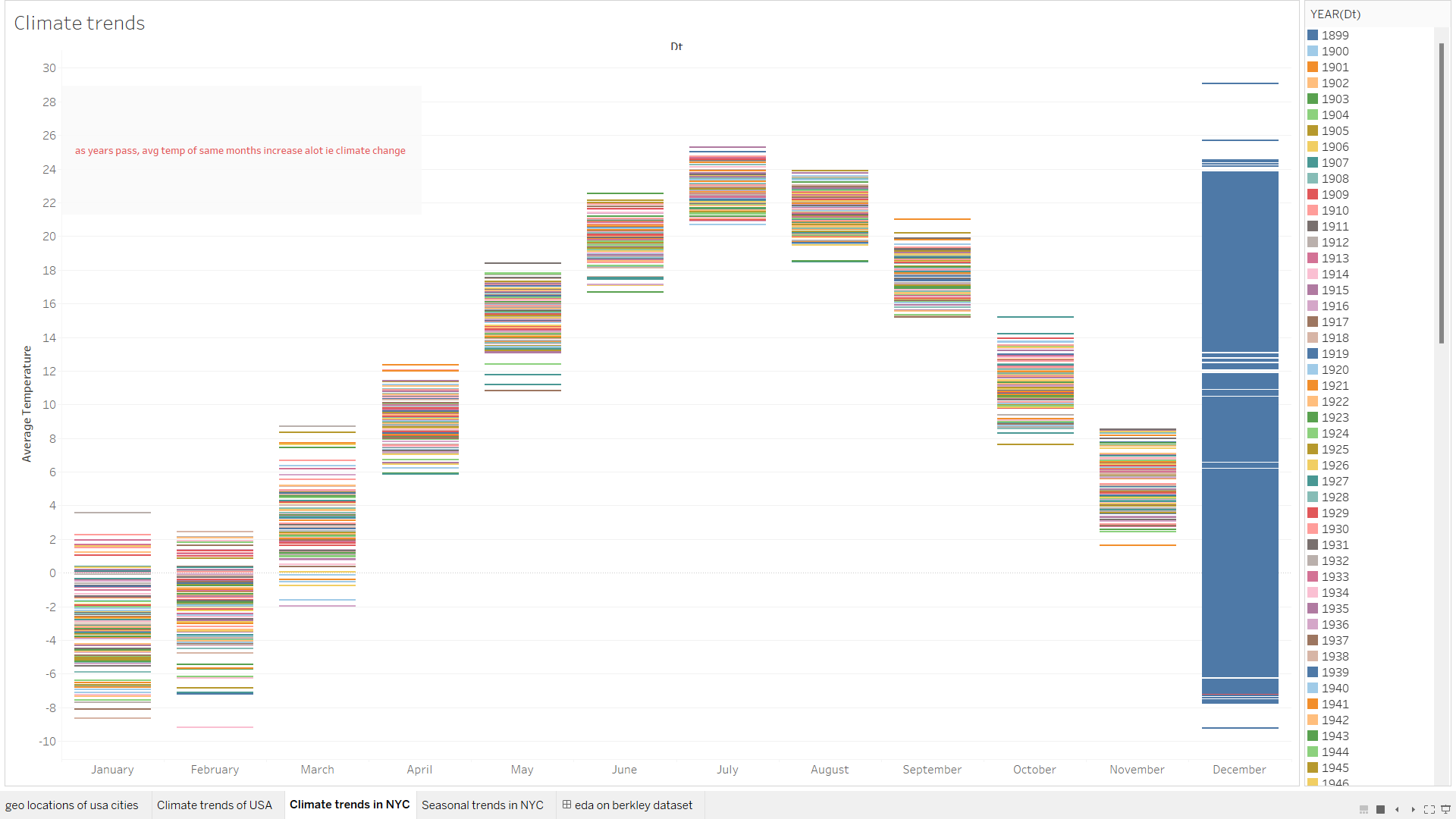
* + NYC is located at a latitude of approximately 40.7 degrees north of the equator. This places it in a temperate climate zone.
  + Its proximity to a water body has a moderating effect on its climate. The ocean acts as a heat reservoir, which means that NYC experiences milder temperatures than areas farther inland at similar latitudes.



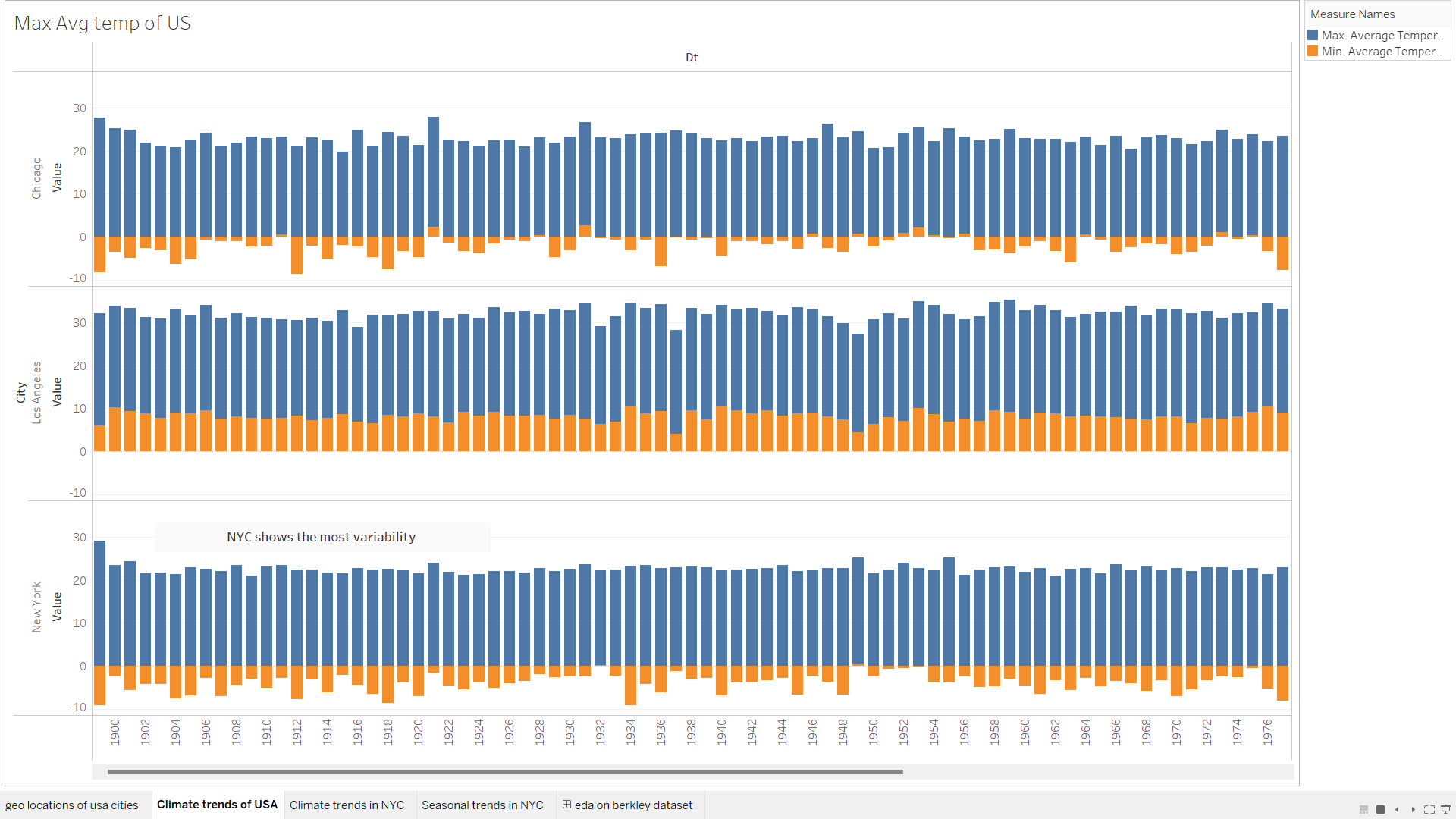
* + NYC's latitude places it in a region with distinct seasons, including cold winters and warm summers.



* + Moreover, over the years probably due to global warming, the overall climate has been drastically impacted.



* + Generally, areas closer to the equator receive more direct sunlight throughout the year and tend to have warmer temperatures. As you move away from the equator (north or south), temperatures tend to be cooler. As a result you experience a great variety of temperature differences.



**Data Download:**

Unemployment in NYC:

<https://catalog.data.gov/dataset/local-area-unemployment-statistics-beginning-1976>

**Data Cleaning:**

All the data has been cleaned during using tableau prep and then is used for visualizations. The workflow is submitted along with the folder.

Then it is joined with the Berkley earth dataset for further analysis. The joining basis were MONTH/YEAR and REGION.

It was then further cleaned and used for analysis.

**Dashboard design and visual story:**

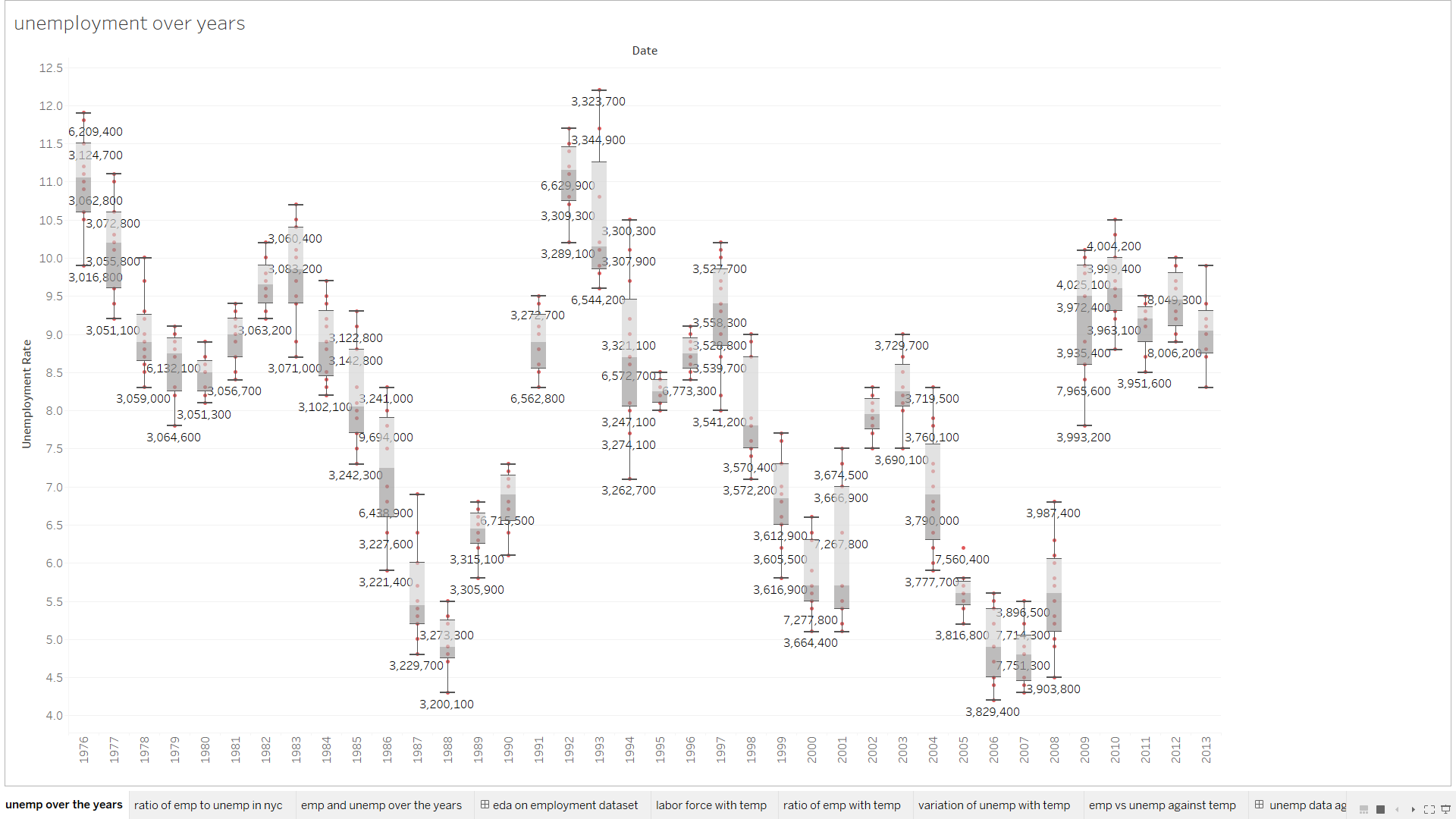
The necessary dashboards are created and joined to form an in intricate visual story. The dashboards contain eda on the acquired dataset first (trends of unemployment over the years) and the second dashboard contains eda on acquired dataset joined with Berkley earth. So it explains the relationships and correlation of unemployment with temperature.

The visual story depicts how employment rate has increased over time because of more opportunities and population. However if we look closely as NYC experiences temperatures from -14 to 28 on average. In extreme winters, the employment rate decreases conversely the unemployment rate increases. This may be due to blocked streets, discomfort in travelling, diseases and people may prefer to work remotely as seen in the labor force (employed + seeking employment) increase.

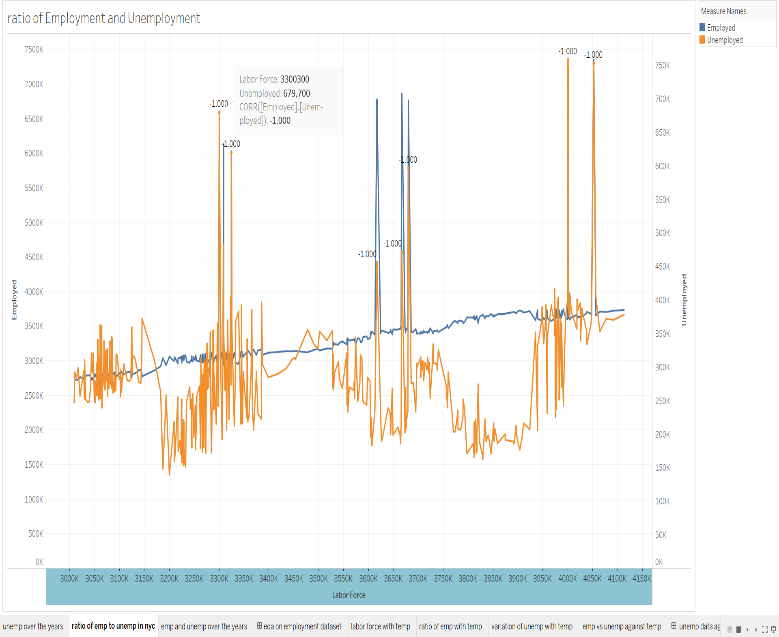
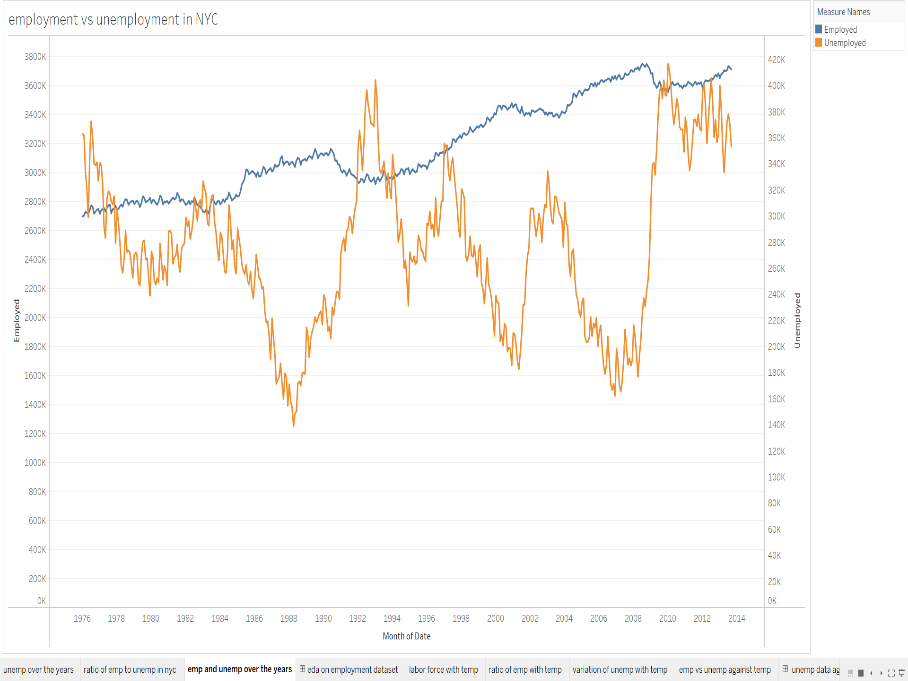
**Exploratory Data Analysis:**

I start by exploring the unemployment dataset. So the very first thing to do was to observe its trend. I did by making a box and whiskers plot. Box plots provide a clear representation of the central tendency and spread of data, allowing you to see how the unemployment rate has evolved over time.

On an average the unemployment rate decreased in NYC. There were some outliers suggesting otherwise but that is normal because unemployment is directly related to economic factors that can do up and down.

****

Then we went into further analysis by comparing the employment and unemployment ratio over time. We came to the conclusion that as time has progresses so has NYC. At first the unemployment rate was quite high but then the employment rate increased by leaps and bounds.

****

Then we began to compare it with temperature and it gave us a new insight that in extreme winters the employment rate begins to decrease while in summers and spring it increases. However, the overall employment rate does increase throughout the year but it very largely depends upon the temperature. Summers aren’t too harsh but in winters a significant dip is seen in the employment rate.

****

**Data Download:**

Influenza:

[**https://www.kaggle.com/datasets/titustitus/h1n1-new-york-2009**](https://www.kaggle.com/datasets/titustitus/h1n1-new-york-2009)

**Data Cleaning:**

All the data has been cleaned during using tableau prep and then is used for visualizations. The workflow is submitted along with the folder.

Then it is joined with the Berkley earth dataset for further analysis. The joining basis were MONTH/YEAR and REGION.

It was then further cleaned and used for analysis.

**Dashboard design and visual story:**

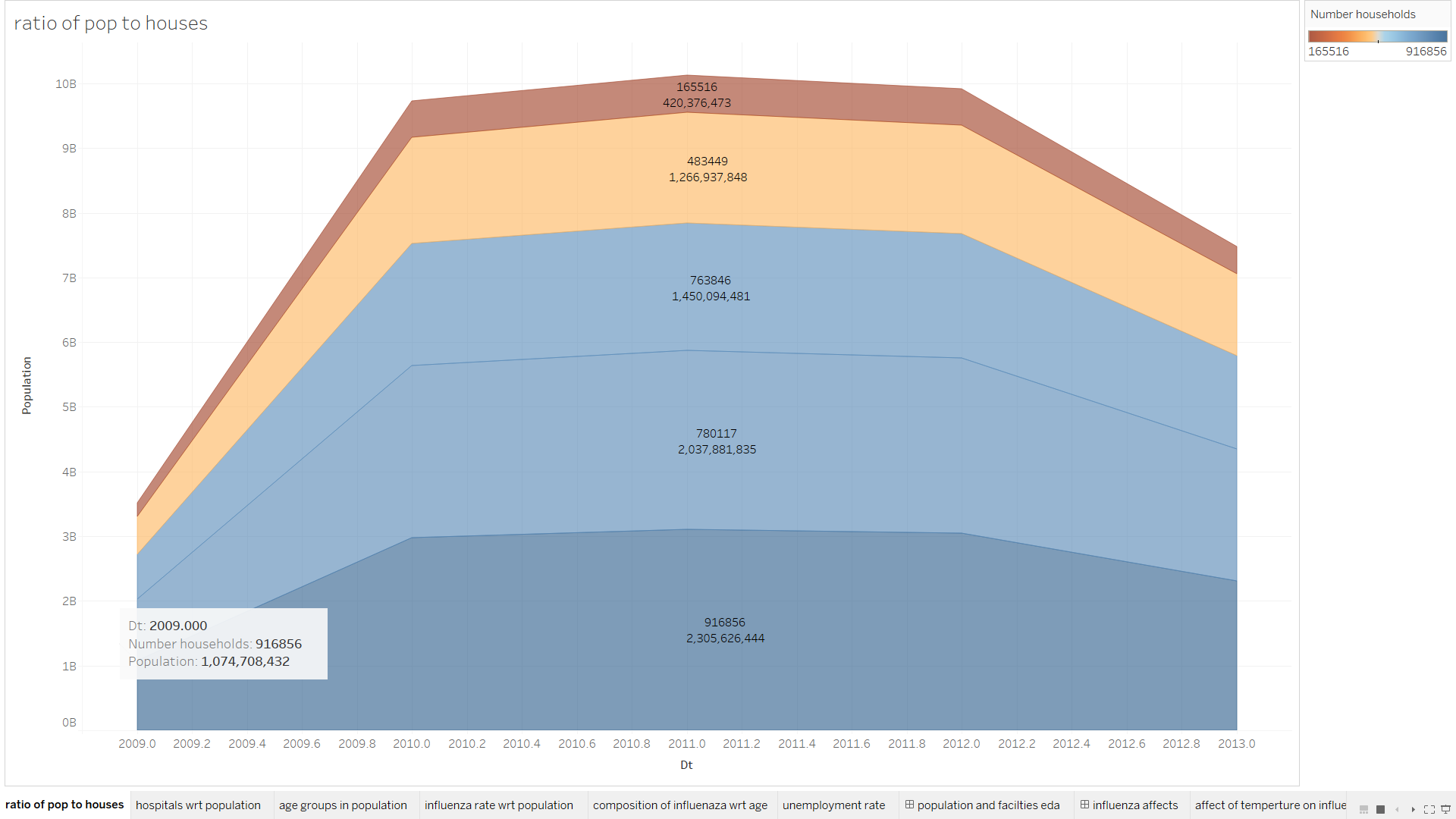
The necessary dashboards are created and joined to form an in intricate visual story. The dashboards contain eda on the acquired dataset first (effect of influenza on the population over the years) and the second dashboard contains eda on acquired dataset joined with Berkley earth. So it explains the relationships and correlation of influenza with temperature.

The visual story how influenza cases have increased over time extremely due to temperature. As the temperature increases (heat) the viruses begin to multiply and its number of cases increase.

**Exploratory Data Analysis:**

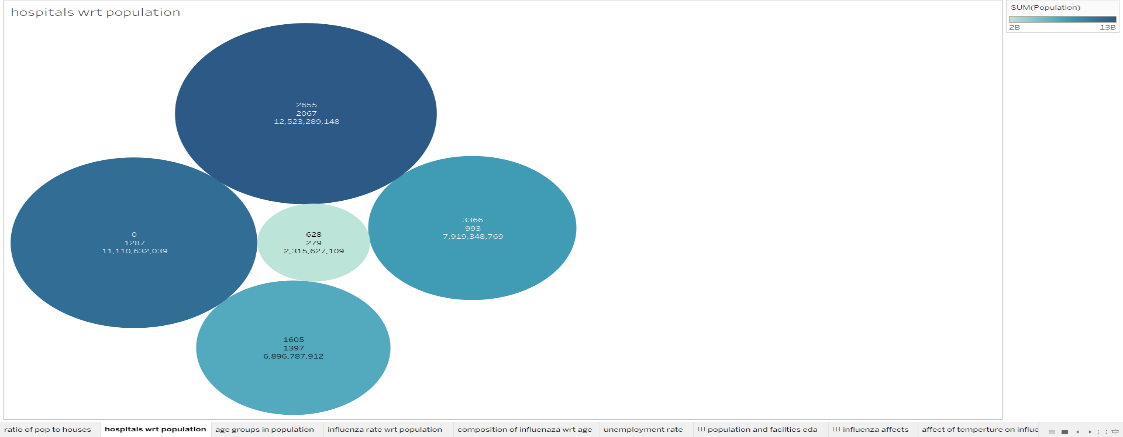
I start by analyzing the influenza dataset. The population, number of families, hospitals and different age groups at that particular time.

The population is quite huge however the population to houses ratio is quite good. It means family system is prevalent there and the economic situation isn’t too bad.



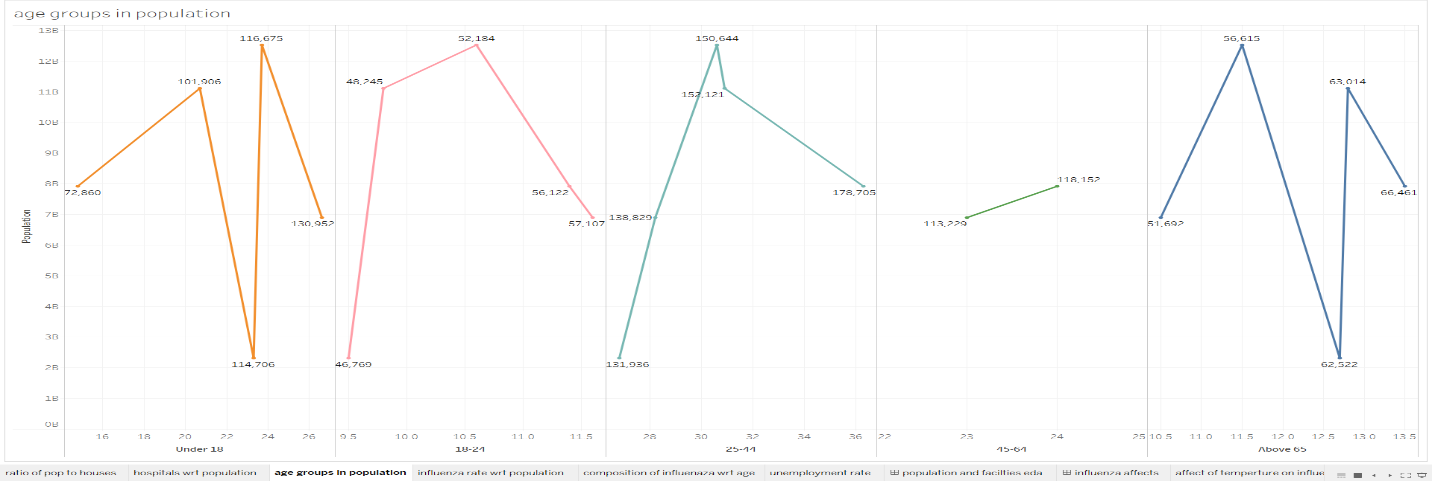
However the hospital to population issue is not good. Both the county and service hospitals combined are not enough especially to treat an epidemic such as influenza.

So, there are approximately 26,532 people for each hospital in the given population.

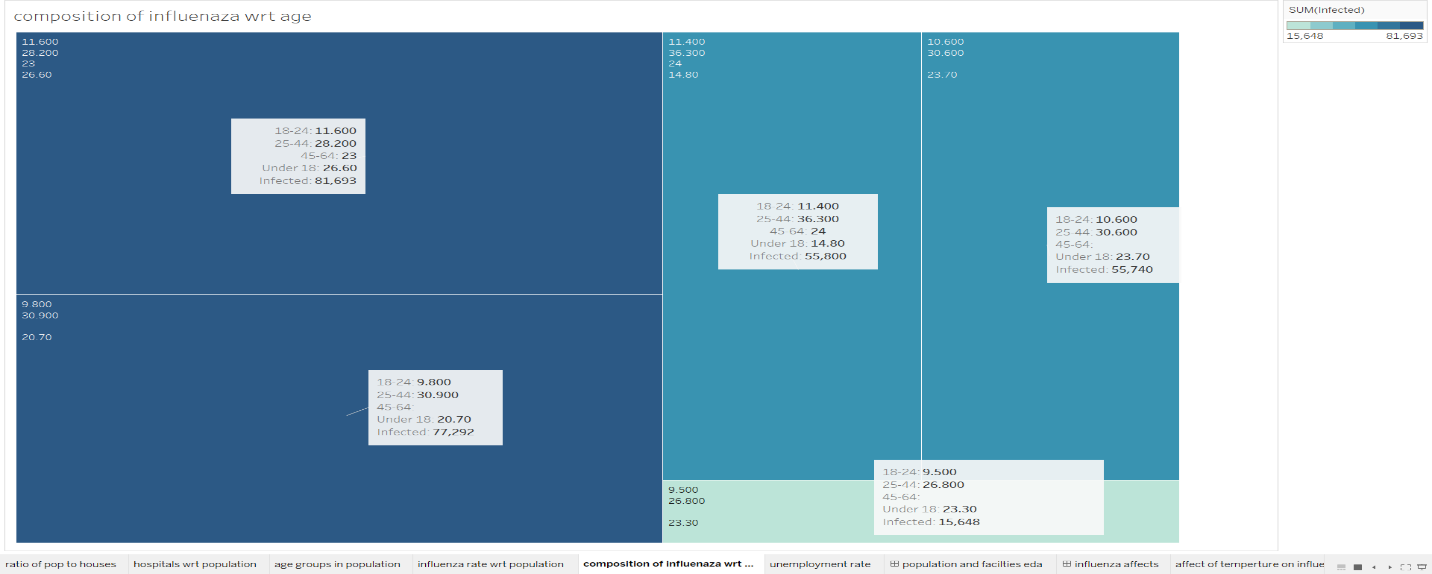


Then we looked at the age groups of the population. So the most residing age group for that particular time was 25-44.

So automatically the age group with the highest number lead to the most influenza victims.



It can be seen from the composition of influenza cases too that it affected the age group of 25-44. It can be due to number of reasons. First as this age group has more interactions and bigger social circles. Moreover they acquire the highest percentage of population.



Lastly we come to the conclusion that influenza is directly influenced by temperature. As the temperature increases (summers) the virus spreads and the influenza cases increases whereas in winters they decrease. And the overall increase of influenza cases can be vividly seen in the age group of 25-44.

