Dr. Vess L. Johnson

Team of 4

Poorna Chandra Konduru (11443242)

Charitha Seelam (11449556)

Nikith Kumar Aragani (11479338)

Manisha Darshanam (11479991)

**U.S. Household Income Distribution by State**

**ABSTRACT**

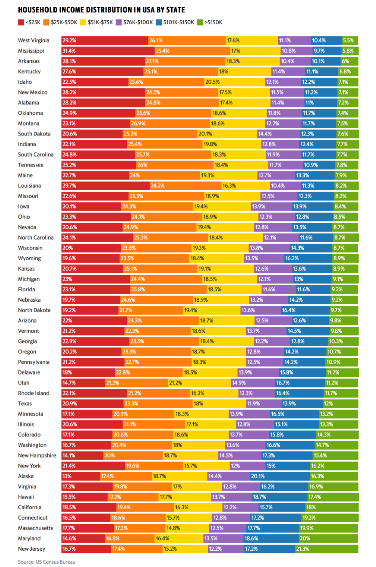
We are aiming to analyze US household income with the help of data set from the US Census Bureau/data.word. Data is about aggregated household income earned from different sources all members/family living in the same house by state wise ranging from 2009 to 2016 with respect to various income levels which is logically distributed between ($10000 to $14999) and ($200000 to more).

Dataset also gives information about what percentage of people of the specific state falls into different income level categories as well as the number of people categorized to various income levels. This will help us to understand income diversity among states and income distribution among the people in the state to calculate income inequality.

**INTRODUCTION**

Each state in the U.S. is remarkable, with various financial possibilities and openings accessible to its occupants. For instance, in a state like New York, there is an excess of lucrative positions accessible in tech and money areas. In the interim, in places like North Dakota and Alaska, there is an amazing enrichment of normal assets that assist with setting out freedom for individuals living there. The present representation from Reddit shows how extraordinary each state depends on yearly family pay dispersion information.

It's important that the underneath information doesn't consider the average cost for basic items, which can immensely affect how far that family pay goes.



The above realistic, utilizing information from the U.S. Registration Bureau, shows the family pay circulation for each state. The pay for each state is separated into six sections (<$25k, $25k-$50k, and so forth), and information is arranged by the particular rates of each state in the >$150k section.

Curiously, the state with the most noteworthy rate in the top section (>$150k) is New Jersey with 21.3% of families. The wellbeing and life sciences areas are blasting in the state – and 14 of 20 of the biggest biopharmaceutical organizations have tasks in New Jersey. In the mean time, significant provinces in the state likewise have vicinity to huge urban communities like New York City and Philadelphia, permitting individuals in the Garden State to effectively drive to lucrative positions in these metro habitats. It's the territory of Alaska, which has the most noteworthy extent of families (69.5%) acquiring >$50k each year in the whole country. The state is known for being wealthy in oil and gas, and part of the pay for every family comes from the yearly profits of the Alaska Permanent Fund, which is valued at $55 billion today.

**METHODS**

**TABLEAU**

In the Business Knowledge Industry, the incredible information perception device is the Scene. Crude information can be made to effectively reasonable by Scene. It is intended to assist clients with creating representations with no earlier information on programming. Worksheets and Dashboards are the representations that can be made by utilizing Scene. The information put away in different spots can be associated and separated utilizing Scene. By utilizing the Scene apparatus, we can plan our own perception. The best highlights of the Scene are Information Mixing, Ongoing examination, and Cooperation of information.

There are different strategies that can be utilized in Scene for representation. Dissipate Plot is a sort of plot with two factors of a bunch of information by utilizing cartesian directions to show the qualities. Box Plot is a sort of diagram with lines stretching out from the containers showing inconstancy outside the upper and lower quartiles. Tree map gives us the data of most noteworthy worth i.e., the most noteworthy square gives us the higher worth. The reference chart shows us rectangular bars which address classifications with their tallness or length as indicated by their qualities. Thus, we have line outlines, bar diagrams, image maps, disperse plot, box plot and different charts can be utilized for perception methods.

**EXPLORATORY DATA ANALYSIS**

A way to deal with investigating datasets with visual techniques is called an Exploratory Information Examination. I have utilized the Scene to picture this dataset from the arrangement of representation apparatuses. In this way, the informational collection subsequently taken is dissected with the accompanying perception diagrams and the interaction of exploratory information investigation can be portrayed by the accompanying representation charts. At the point when we take an informational collection various inquiries will be raised, to find a decent solution for the inquiries we need to picture the informational collection and take the fields in like manner, so we can find the right solutions for the inquiries.

**DATA SET DESCRIPTION**

Dataset also gives information about what percentage of people of the specific state falls into different income level categories as well as several people categorized to various income levels. This will help us to understand income diversity among states and income distribution among the people in the state to calculate income inequality.

Dataset: <https://data.world/makeovermonday/2018-w-3-u-s-household-income-distribution-by-state>

**DATA CLEANING**

There were some missing qualities in the dataset, which makes an aggravation in examining the information. In this way, every one of the clear and invalid fields of the dataset is set to 0 for numerical tasks to be performed. Information with copy esteems will cause irregularity which brings about wrong expectations. Thus, the copy information will be erased to lessen irregularity in information. At first, my crude informational index has lines and sections in a chaotic way which doesn't assist with working with the scene. Hence, the information is orchestrated in a coordinated way to fit in the scene.

**US HOUSEHOLD DATASET ATTRIBUTES**

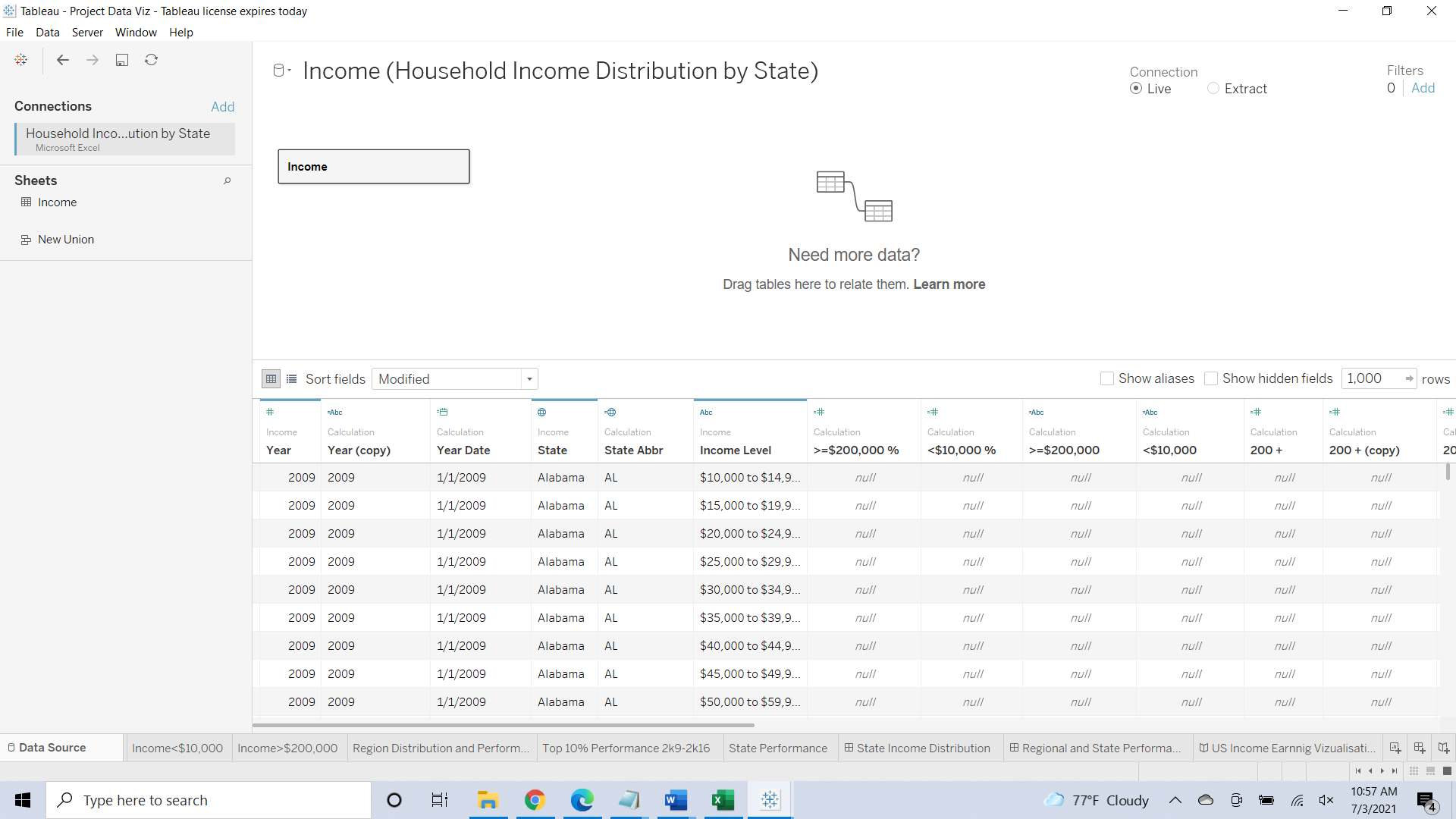
We are planning to break down US family pay with the assistance of an informational index from the US Statistics Bureau/data. Word. Information is about collected family pay procured from various sources all individuals/family living in the same house by state shrewd going from 2009 to 2016 regarding different pay levels which is consistently dispersed between ($10000 to $14999) and ($200000 to additional).

Dataset likewise gives data about which level of individuals of the explicit state falls into various pay level classes just as a few individuals ordered to different pay levels. This will assist us with understanding pay variety among states and pay appropriation among individuals in the state to ascertain pay disparity.

**RESULTS**

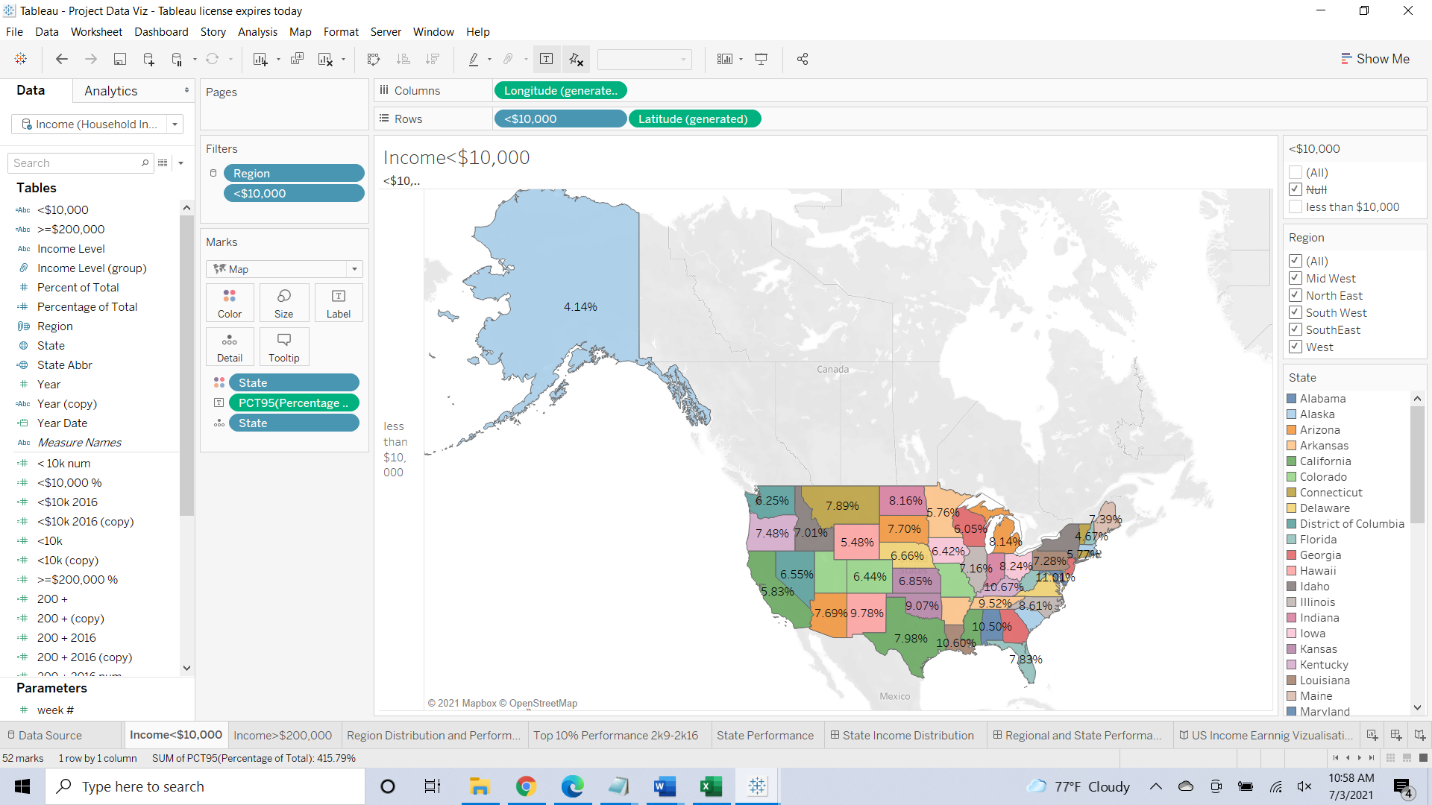
1)

From the below sheet we can be able to see the data source for the income (Household Income Distribution by state) for every year.



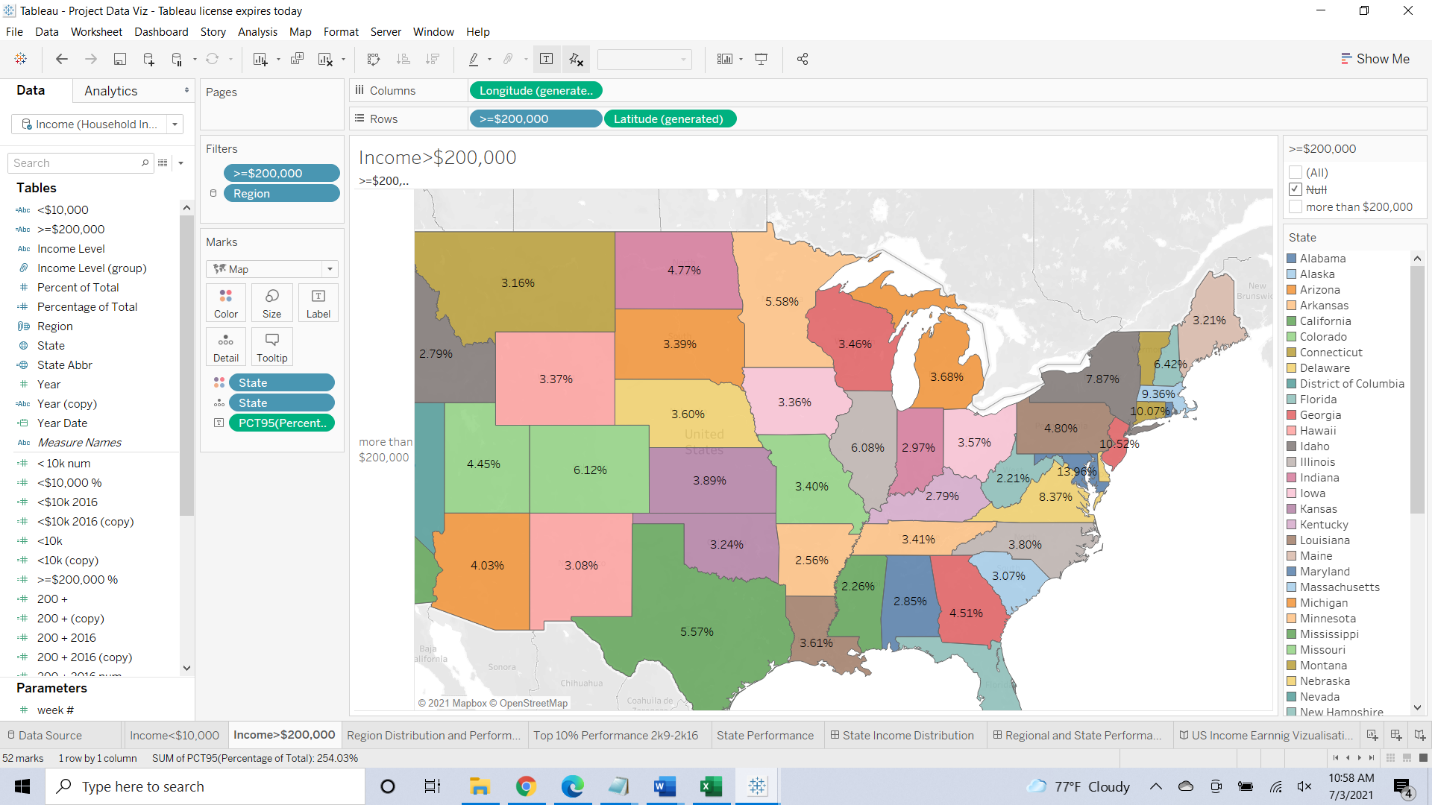
2)

From the below visualization, we illustrate that Map based on Longitude (generated) and Latitude (generated) broken down by <$10,000. The color shows details about State. The marks are labeled by percentile of Percentage of Total. Details are shown for State. The data is filtered on region, which keeps Mid-West, Northeast, Southwest, South-east, and West. The view is filtered on <$10,000, which keeps less than $10,000.



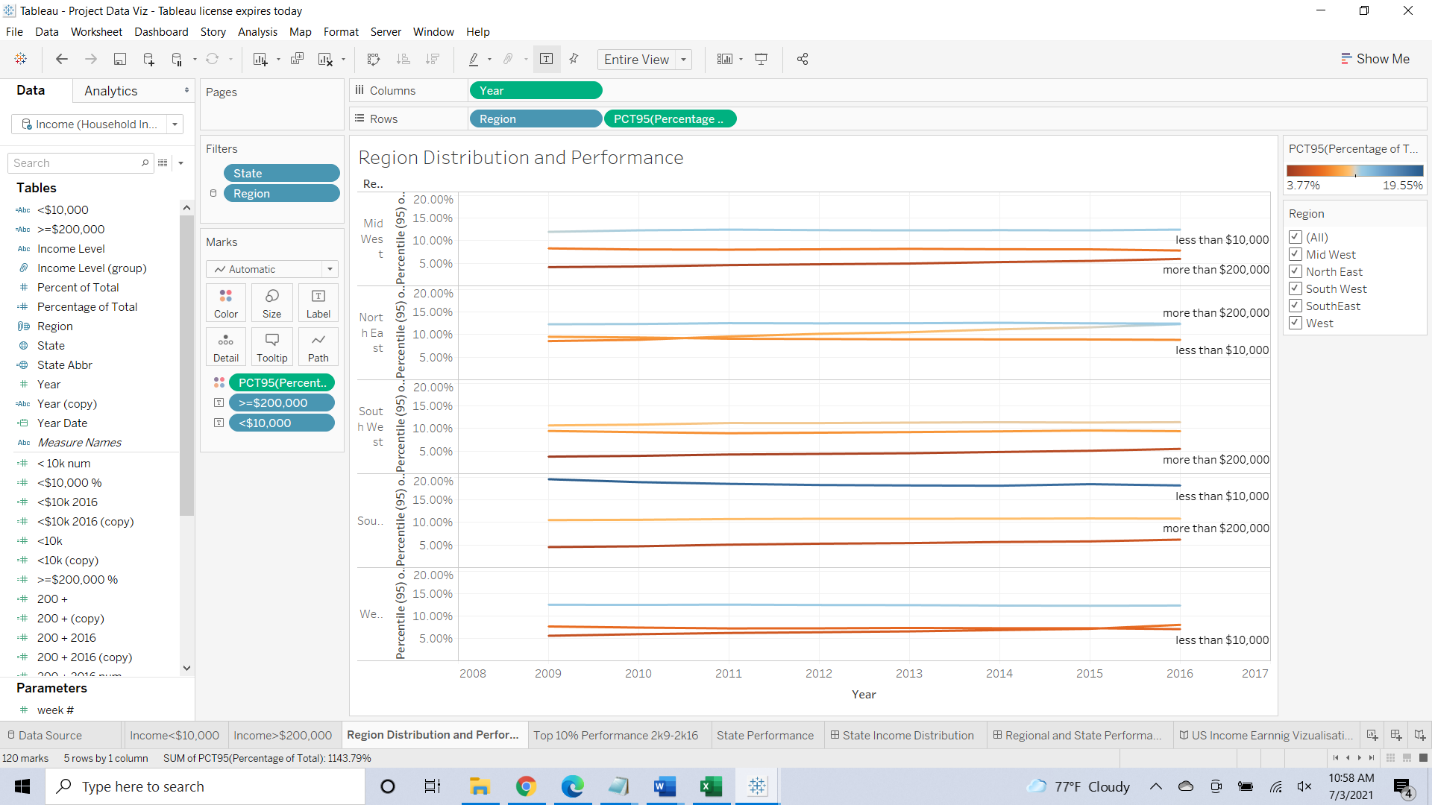
3)

From the below visualization the Map based on Longitude (generated) and Latitude (generated) broken down by >=$200,000. The color shows details about State. The marks are labeled by percentile of Percentage of Total. Details are shown for State. The data is filtered on region, which keeps Mid-West, North East, South West, South East, and West. The view is filtered on >=$200,000, which keeps more than $200,000.



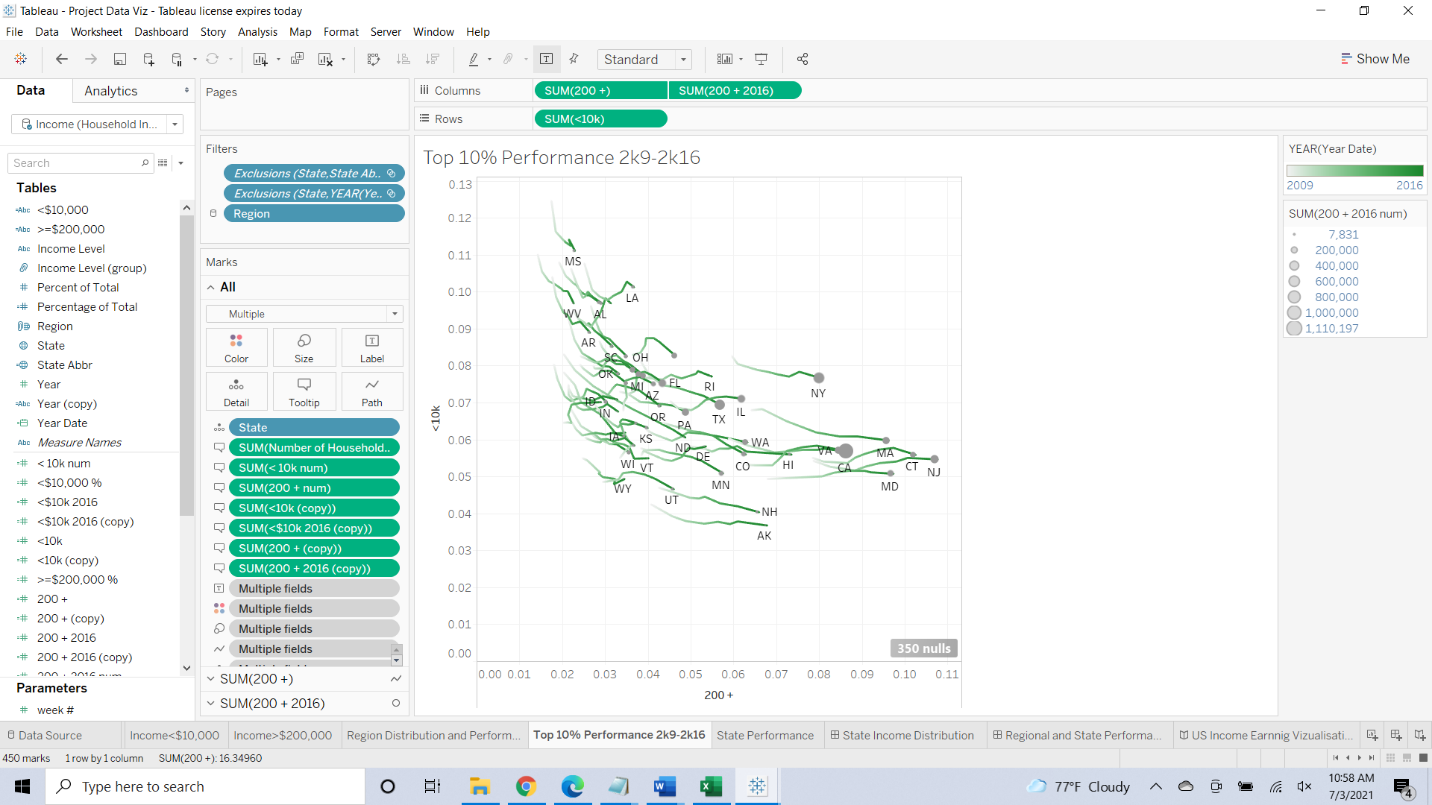
4)

From the below visualization, the sheet for the regional distribution and performance and the trend of percentile of Percentage of Total for Year broken down by region. The color shows the percentile of Percentage of Total. The marks are labeled by >=$200,000 and <$10,000. The data is filtered on State, which keeps 52 of 52 members. The view is filtered on region, which keeps Mid-West, Northeast, Southwest, Southeast, and West.



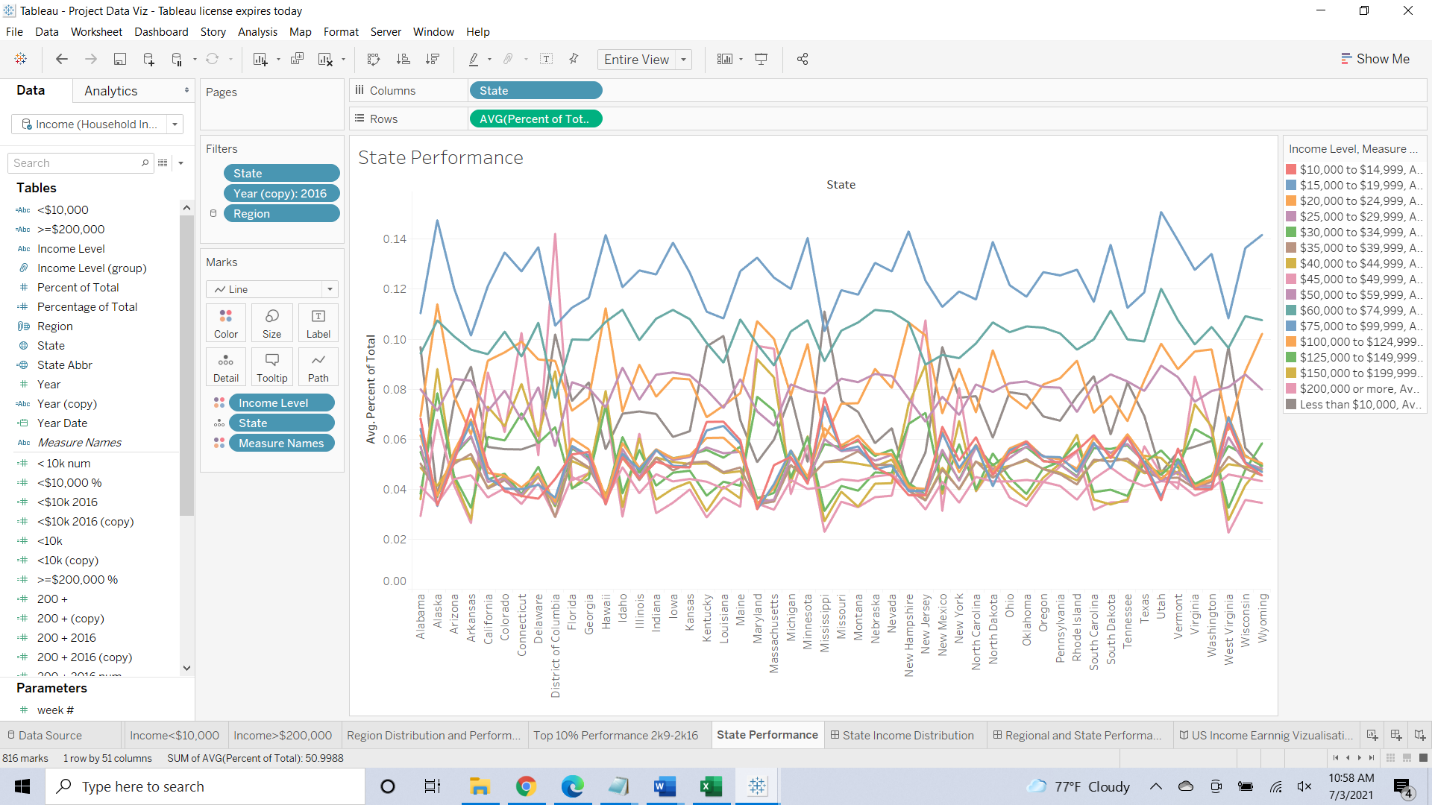
5)

This visualization demonstrates the top 10% performance 2k9-2k16 and the sheet gives us the sum of 200 + and the sum of 200 + 2016 vs. the sum of <10k. Details are shown for State. For pane Sum of 200 +: Color shows details about Year Date Year. For pane Sum of 200 + 2016: Size shows sum of 200 + 2016 num. The marks are labeled by State Abbr. Details are shown for State and Year Date Year. The data is filtered on region, which keeps Mid-West, Northeast, South West, south-east, and West. The view is filtered on Exclusions (State, State Abbr, YEAR (Year Date)) and Exclusions (State, YEAR (Year Date)). The Exclusions (State, State Abbr, YEAR (Year Date)) filter keeps 407 members. The Exclusions (State, YEAR (Year Date)) filter keeps 409 members.



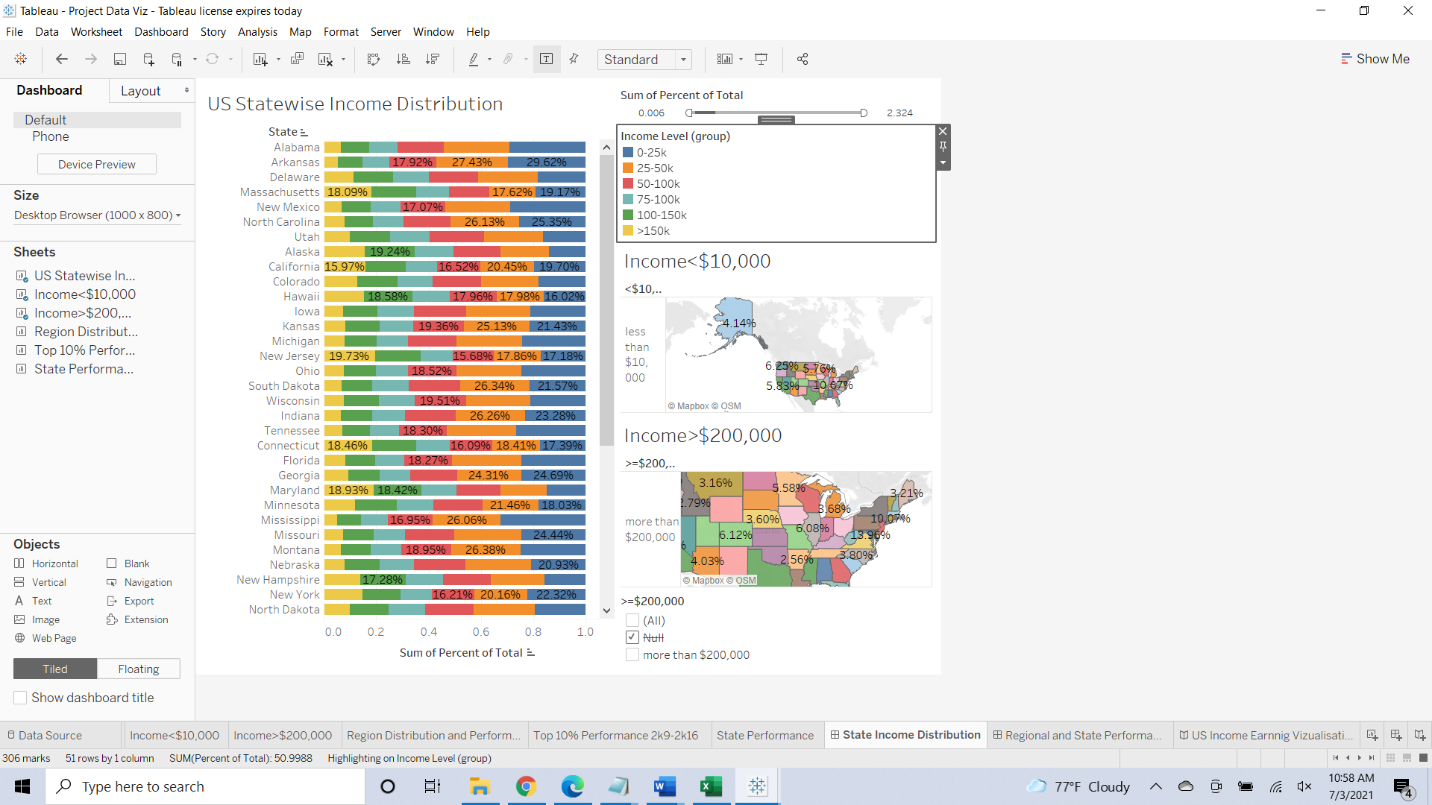
6)

From the below visualization of the state performance, the sheet demonstrates the trend of Avg. Percent of Total for State. The color shows details about Income Level and Avg. Percent of Total. Details are shown for State. The data is filtered on Year (copy) and Region. The Year (copy) filter keeps 2016. The Region filter keeps Mid-West, Northeast, Southwest, south-east, and West. The view is filtered on State, which excludes Puerto Rico.



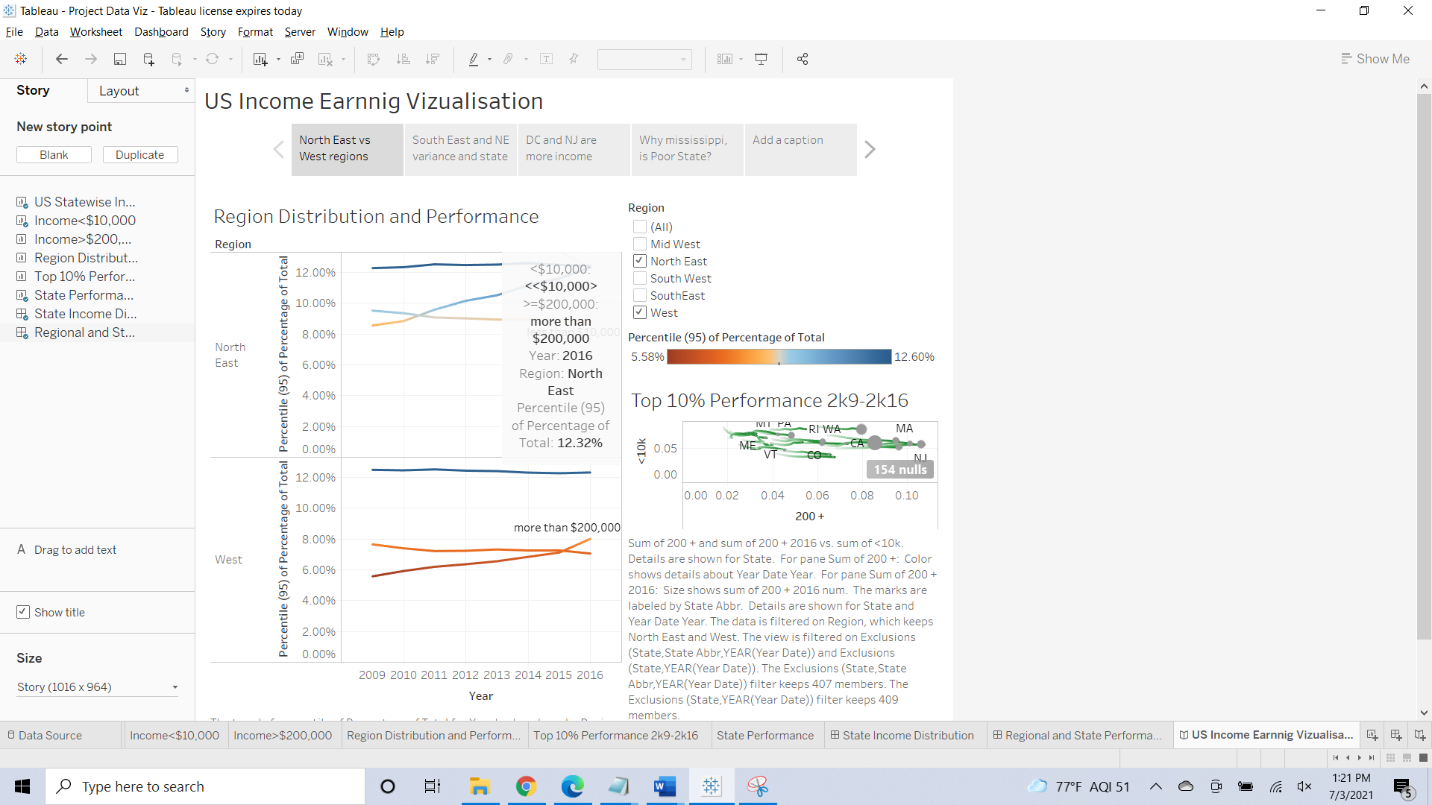
7)

This visualization illustrates the overall US state wise income distribution, and this dashboard gives us the Sum of Percent of Total for each State. The color shows details about Income Level (group). The marks are labeled by the sum of Percent of Total. The data is filtered on Year Date Year and Region. The Year Date Year filter keeps 2016. The Region filter keeps Mid-West, Northeast, Southwest, South East and West. The view is filtered on State and sum of Percent of Total. The State filter excludes Puerto Rico. The sum of Percent of Total filter ranges from 0.0062 to 2.3236.



8)

This visualization gives the US earning visualization This dashboard represents the region distribution and performance of top 10% from 2k9-2k16 for percentile 95% percentage of total.



**DISCUSSION AND CONCLUSION**

Visualization is one of the important stages in any data analysis project. Most of the results to our understanding can be achieved by the visualization techniques. As visualized above, we have proved d that How different states are performed from 2006 to 2016 to increase average household income. How different states performed reducing/stabilizing income inequality between Top 5% income earners and below 50%.How all the income levels earning people performed over the time of 2009 to 2016.States wise earnings over time which is proved with the help of visualization techniques. It gives us information on US household income distribution by state.

**ACKNOWLEDGEMENT**

I would like to thank Professor Dr. **Vess L. Johnson for giving me this opportunity in writing a report of our project with the proper guidelines. This report is written from my understanding using visualization techniques to give a clear explanation of the project.**

**REFERENCES**

[1] Guzman, G. G. (2017). *Household income: 2016*. US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

[2] Bee, A. (2012). *Household income inequality within US counties: 2006-2010*. US Department of Commerce, Economics and Statistics Administration, US Census Bureau.