**Impact of COVID-19 on Unemployment Rate (U.S.)**

ISM 6361 Data Visualization Project

By Satadipa Sarkar

Seattle Pacific University

**State a business reason for selecting your tools (problem you would like to solve).**

Whenever it comes to creating beautiful charts/maps, Tableau has always been my favorite tool because:

* Simple User Interface
* Interactive Dashboard building and story-telling through the vizzies created.
* It allows to connect from multiple data source an enables the user to write custom queries to join various files.
* Look and feel is very sophisticated and navigations are pretty easy.
* It helps me create a map in motion with animations and also forecast future data.
* The data used to create the vizzies for this project had to be collated and collected from many different files which wasn’t difficult.
* Tableau made it super easy to demonstrate the context of my project (unemployment statistics affected due to COVID-19 and the comparison over the year) fairly simple and beautiful.
* One of the major aspects of my project was to do a comparison of unemployment rate change in various states in U.S and plot the change in rates. Geographically. The map feature in Tableau made it look super intuitive and easy to relate to.
* Tableau shows the underlying data it used to create the vizz.

**Document how/where you got your data (if it is publicly available, or internal for a work project).**

* Dataset used for this project is publicly available from the URL <https://www.bls.gov/bls/unemployment.htm>
* Organization – U.S. Bureau of Labor statistics,
* The website had all the necessary data which I thought would be helpful for my visualization. I have used Unemployment rate data from the year (2015-2019), Industrywide unemployment (2020), Unemployment rate by state. The cost of living data was collected from <https://worldpopulationreview.com/states/cost-of-living-index-by-state/>.
* The data was fairly clean but I had to remove few columns (which wasn’t necessary for my project) without harming the integrity of the dataset. Few formatting was necessary. Also, I had to combine data from multiple years and split city name from raw data.
* Why this dataset?
* A week ago, when I was listening to NPR, the news that caught my attention was during this pandemic unemployment rate is going high. But we might think that healthcare workers are not affected due to this but in reality, this is not. Through my vizzies I could actually tell that story about the impact of COVID-19.

**Document how you used your tool. Many tools are super rich in features and you probably will not be exploring all the features, explain the parts you did use.**

Tableau is super rich in features. It had all the required features necessary for this project. The features which I ended up using are as follows:

* **Different charts (like histogram, horizontal bars, side by side bars, packed bubbles, tables, continuous lines, symbol maps, motion maps)**- ShowMe

1. side by side bars- used to display the unemployment rates of various states from 2015-2019.
2. Line chart – To display the change in unemployment rate from 2019-2020.
3. Horizontal bars – To display the map in motion to demonstrate the increase in unemployment rate of U.S. from January20-April 20.
4. Symbol Maps- High cost of living Index numbers, unemployment numbers. I used map layers feature to highlight unemployment rate for large metropolitan areas using stepped color (population) by state.
5. Bubble charts- Men-Women unemployment numbers, Unemployment by industry.

* **Forecasting and Trend lines**

The forecast of unemployment rate for the rest of the year 2020 from (May-Dec) was very important to predict. Even though Tableau did a forecast but I am not satisfied with the prediction because the user interface did not allow me to do more customized charts other than by seasonality, confidence intervals, aggregation, forecast model, length etc. I am bit skeptical about the outlier data of April and how did Tableau use this data to forecast for the rest of the months. It shows a forecast of 9.33 for the rest of the year.

* **Parameters**

Creating Parameters for Change over month, change over the year, deviation from national average in one vizz was super helpful and it allows the user to see the important information about change in rates from multiple perspectives.

* **Calculated field**

Calculated field helped me to do all the calculations used to create new attributes like change in rates, percentages, deviations from standard national average.

* **Filters**

Filters option was widely used in this project to show relevant information by state, rate, city etc.

* **Colors, tooltip, label, legend, size** – As studied in How charts Lie, the charts make a real difference when colors, size used to represent the attribute are used wisely and effectively.
* **The axis and scales used to represent the value are properly named and cited**.
* **Dashboard & stories**

Dashboard helped to picturize the summary/context of this project that is the current status of unemployment state-wise, rate of change, forecasts, industry-wise effect of COVID 19 on unemployment and showcase the majorly affected industries. The stories helped to document and elucidate the reason/outcome from the forecast.

* **Data source tables, attributes rename, changing the datatype.**
* **Quicktable calculations.**
* **Analytics -** for mapping the trend line (Polynomial) over the cell and forecasting.
* **Sort, swap, new worksheets tab.**

**Explain why you chose which visualization/charts**

The visualization charts used:

1. side by side bars- used to display the unemployment rates of various states from 2015-2019 to show the comparison between various years and get a quick idea about the trend in different states over the years.
2. Line chart – To display the change in unemployment rate from 2019-2020. The chart inclination from low to high provides an easy interpretation of the increase in unemployment rate between two years.
3. Horizontal bars – To display the map in motion to demonstrate the increase in unemployment rate of U.S. from January20-April 20. The horizontal bars change in height from March to April gives the idea of the sudden steep increase in unemployment due to COVID-19.
4. Symbol Maps- High cost of living Index numbers had state-wise data. The maps depicting HCL (High cost of living Index) with stepped colors and state name is perfect for understanding the impact. Likewise, for unemployment data.
5. Bubble charts- Men-Women unemployment numbers, Unemployment by industry- Bubble charts show the industries affected by COVID-19 using stepped colors and different charts varying from low to high unemployment rate.
6. Area Maps for forecasting was used to show the continuous monthly data from 2018-2020 and the spikes and lows in unemployment rate.

**Give an explanation/analysis of the output. What did you learn or uncover**.

Analysis of the output:

As already mentioned I wanted to understand the effect of COVID-19 on unemployment by industry and state in U.S. The major takeaways are as follows:

* The unemployment increased sharply for all industries particularly in leisure and hospitality.
* On broader classification by industry, the highest affected industry is Non-agricultural, private wages and salary workers.
* Amongst this pandemic the unemployment rate of health-care professionals is still higher than 2019. This data was surprising because I assumed that healthcare professionals are immune and won’t be impacted due to pandemic. Since they are the frontline workers now and in demand.
* The difference in unemployment rate amongst men and women is more in Wholesale and Retail, followed by Administrative and support.
* The unemployment rate rose to 14.7% in April 2020 due to the effects of COVID -19.
* Hawaii, New York, California have a very high cost of living index.
* But the unemployment rate of Hawaii is in the lower range as compared to California whose unemployment rate in 2020 has increased to 5.3.
* Amongst the large Metropolitan areas (cities) Pittsburgh has highest unemployment rate.
* Sharp increase in unemployment rate almost in all the state due to COVID-19.

**Conclude with the 3 W’s (What Went Well, What Did NOT go Well, What Would you do Differently Next Time).**

**What went well**:

* Plotting right charts for right context demonstration.
* Found all the necessary data from the publicly available websites with a bit of research.
* Dashboards and Stories depicts the story of the present scenario and aligns with the raw data perfectly.

**What did not go Well:**

* Forecasting unemployment rate for the rest of the year (May-Dec) 2020. Trend not satisfactory.
* Invested lot of time to determine the right factors which can be affected or correlated with the unemployment rate increase. Cleaning and formatting data took a considerable amount of time.

**What Would you do Differently Next Time**

* Spend less time in getting the right data and data cleaning.
* Try to plot the correlation between high cost of living index (with various factors involving it grocery, Transport, medical, housing) separately and see how it impacts the high unemployment rate.
* Deep dive on industry wise unemployment data.