DATA VISUALIZATION

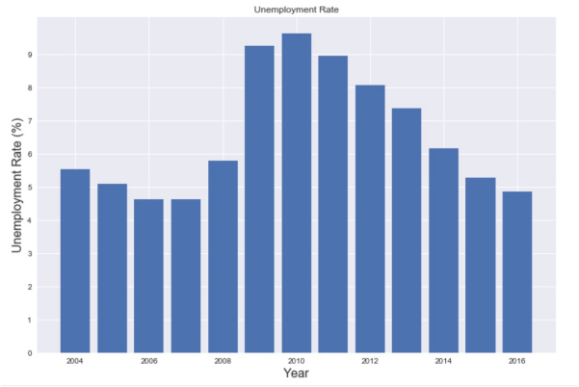
Causal effect of the Great Recession on Unemployment across the US

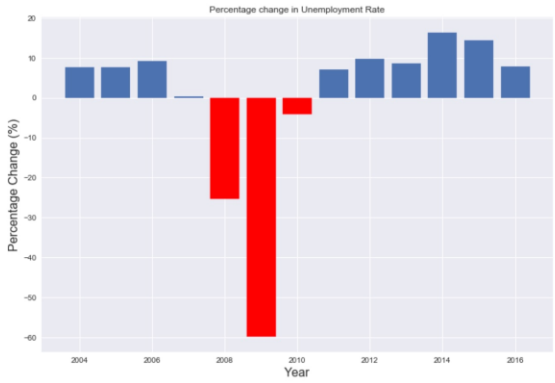
Karthik Anbazhagan (kartanba), Shailendra Patil (patilsh), Ankit Saxena (ansaxena)

To understand and reflect on the correlation of employment type, GDP growth, inflation, and demographics for all the states across the United States of America before, during, and after the Great Recession (2007–09). Leveraging these statistics we hope to debunk the misconception that the US economy is yet to recover from the great recession.

Introduction

The U.S. economy faced one of its unrelenting economic difficulties in 2007, the Great Recession, which lasted until late 2009. This was the worst recession[11] for the United States since the Great Depression of the 1930s. The recession mainly originated because of the real-estate market bubble and the lax lending of banks on low-interest rates. The Great Recession resulted in the scarcity of valuable assets in the market and the collapse of the financial sector. Americans suffered losses in wealth, especially in home investments, and the job market turned worse for many workers. The annual unemployment rate hit an all-time high of close to 9.5% during the great recession. Unemployment statistics have steadily gotten better in recent times, but the job market faced hardships in recovery, especially after the Great Recession. The job market took a significant tumble and the economic recovery of the US, a decade later, is still in the process of prolonged recovery.





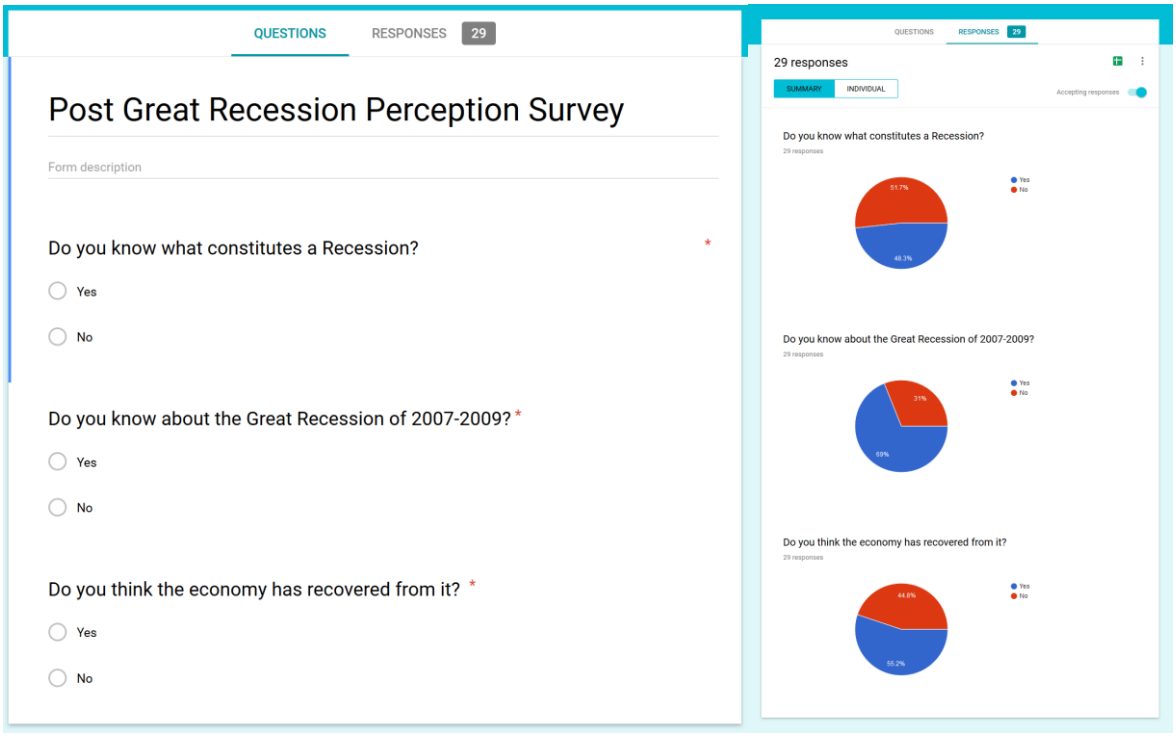
The Great Recession saw the US affected on two key economic variables: gross domestic product (GDP) and unemployment. The unemployment rate reached over the 6% mark for the first time. The US saw an increase of approximately 25% YoY change in 2008 and an added 60% YoY increase in unemployment during 2009. This crisis has been etched into the memories of people in the US and has created a long standing belief that the economy is yet to recover from the great recession.

"A decade after Great Recession, one in three Americans still feel as they haven't recovered from the recession" – ***CNBC***

"53 percent of Republicans think the unemployment rate has risen under Obama" – ***The Washington post***

"In 2015, six years from the end of the recession, total U.S. employment was 148 million compared to a pre-recession peak of 146 million. Meanwhile the population of working age people rose by 17.5 million, more than the 1.7 million employment gain" – ***Forbes***[10]

We conducted a survey, both in class and during the presentation at the Symposium at Indiana University, Bloomington. Our survey was a short questionnaire with three questions that asked students their understanding of the recession and if they believed the economy has recovered from it.



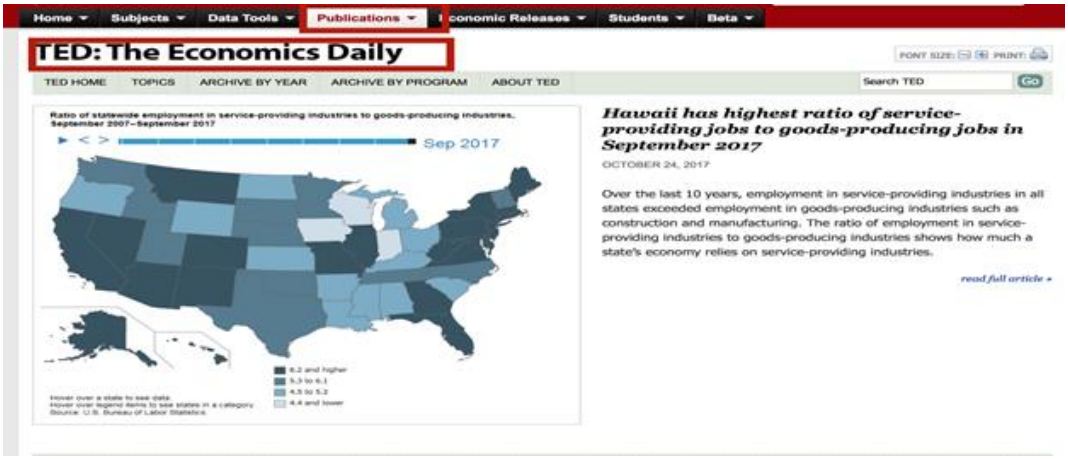
Approximately 44% of the people at our booth, the other day, believed that the people of the US still haven't recovered from the recession. These kind of stats, both from the web and from the people around us motivated us to take up this project, and help debunk the misconception that a few people have with regards to the recession of 2007-2009.

A decade later, the unemployment is back under the 5% that the US last saw during the pre-recession period. During this period, *"12 million jobs have been created since job losses stopped in Feb 2010 and the population of employable people has risen by 17.5 million compared to the 1.7 million back in 2006"*. This gives us a picture that employment recovery is moving in the right direction.

Employment is a right, not a privilege. The employment scenario of a country can explain the overall satisfaction of its citizens very well. In the United States, the employment data is maintained by the Department of Labor. The Bureau of Labor Statistics (BLS)[7] publishes information on employment and wages by demography, occupation, career information, employment types.

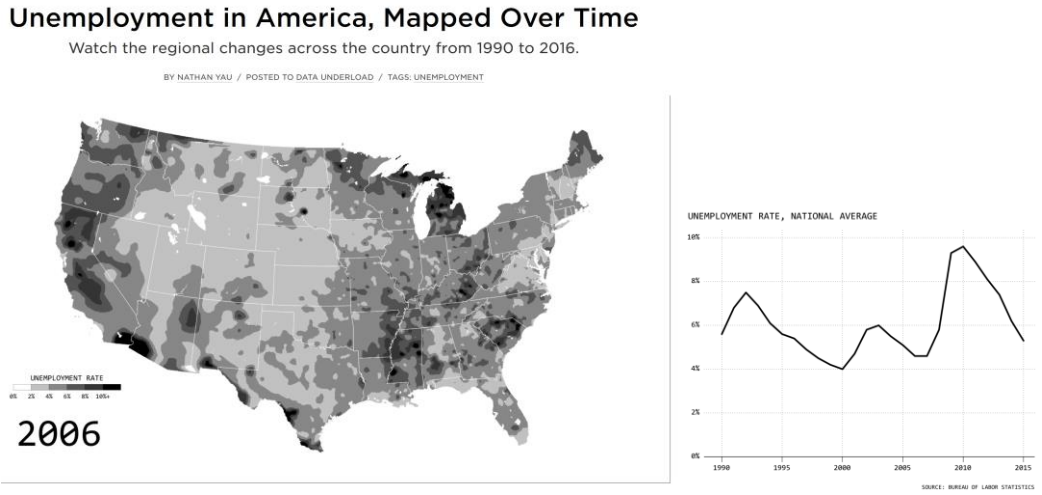
The state and local governments use the employment statistics for planning and budgetary purposes to organize local employment and training services. New policies are remodeled every year by the Federal Reserve, one such policy is to raise the interest rates by measuring the strength of the labor market.

The BLS[4] also publishes a new edition of "TED"[5][9] every business day. Each edition of TED typically includes a chart or map, sometimes both. One such edition below displays a visualization explaining which states provide what ratio of service-providing jobs to goods-producing jobs, on a monthly and weekly basis.

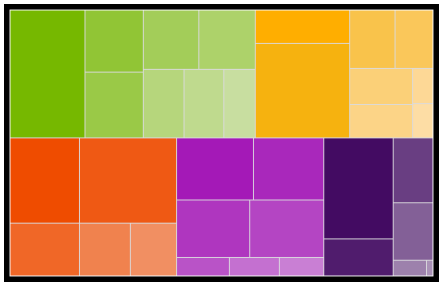


Effectiveness and techniques of visualizations: The US map specified in the daily report is print friendly and well described with the help of legends. We believe adding data labels or annotations on the map itself might help distinguish between the shades of the states.

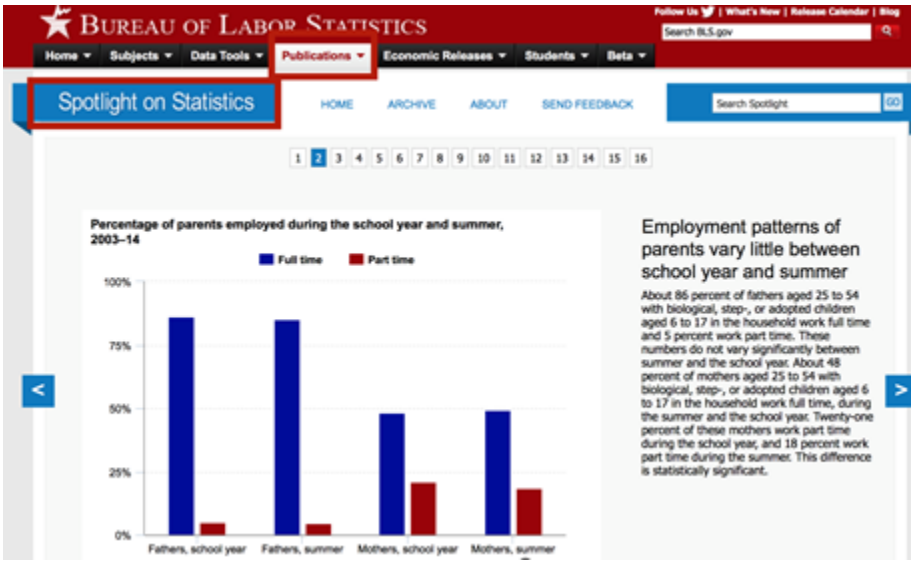
The Bureau of Labor Statistics estimates county-level unemployment on a monthly basis. It tracks annual averages that go back to 1990. The map above, shows the shifts over the past few decades. [https://www.youtube.com/watch?v=shqJR_0WdrI]



Effectiveness and techniques of visualizations: The US map specified in the daily report is print friendly just like the above map and well described with the help of legends. Since the map visualizes at a county level, it would be better to represent it as a tree map. As the map changes year on year, we tend to only focus on the really affected counties (darker shades) and the low values are not significantly visible.

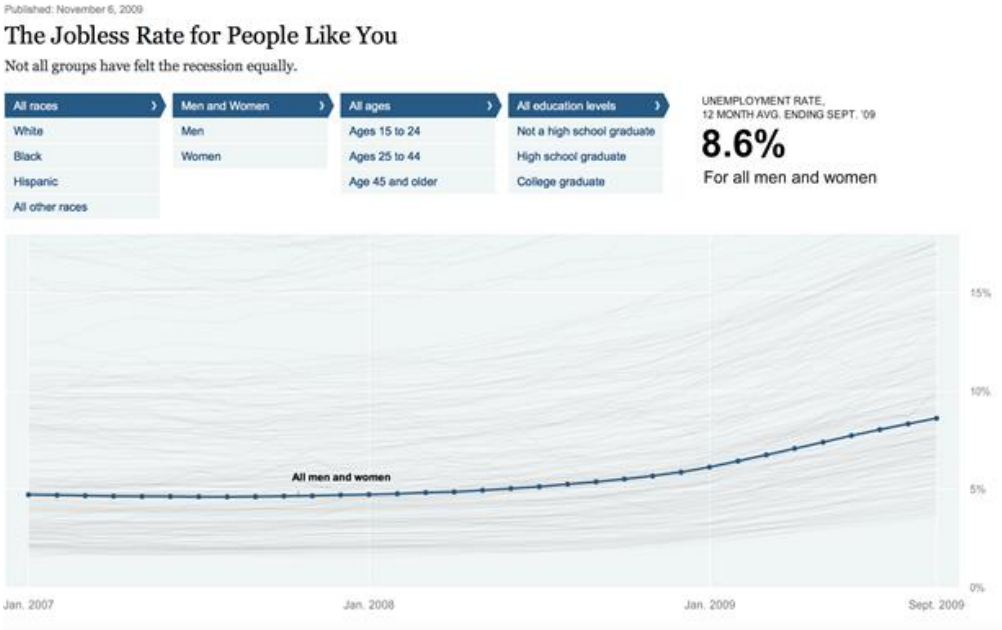


"Spotlight on Statistics"[6] is another publication that focuses on telling stories (about a topic) through visualizations.



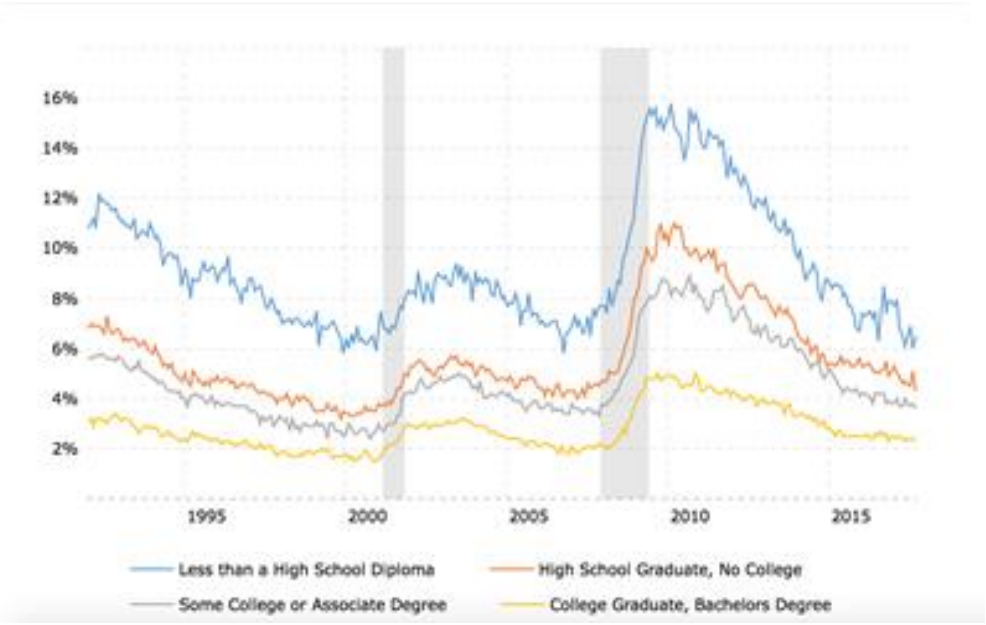
Effectiveness and techniques of visualizations: The plot shows simple red/blue shades. These are not print friendly. It is also hard to compare the 1st and the 2nd column. Instead we could have colors on the categorical values and have full-time/part-time as two segments.

Another interesting visualization is the NY Time's graph from 2007-2009, during the recession period, showing how unemployment rate affected people of the US across different demography.



The graph provides the smoothed curve of unemployment with every demographic detail as a filter[12]. One thing that can be improved in this graph is the scale of y axis. The maximum unemployment is up to 10% and also in case of unemployment even a small 1 or 2 % change in unemployment is significant, hence making the y axis label difference as 1% and y axis last value as 10% would be better here.

Another interesting visualization is by Macrotrends which published a time series, line plot of the unemployment rate changes across the US, between the period 1990 and 2017, based on education[13].



Effectiveness and techniques of visualizations: The visualization has two highlighted segments that show the period of recession. This directly get the users to focus on these grey areas. The choice of colors are ideal as they are not too vibrant or harsh on the eye and the labeling has been done well. The visualization has also been shaded for the period of recession which puts a spotlight on the area to focus.

In our project, we now hope to track the effect of this Great Recession on the unemployment rates across the US and the extent to which the economy has recovered post the recession. By mining the data from before and after the Great Recession, we throw some light on the section of the society that was most affected and explore in-depth the details of those who recouped from this quicker than the rest.

We hope to contribute by providing more than a multidimensional view of the data and compare different aspects affected by the recession. All data sources and visualizations we saw online with respect to the Great Recession was all one dimensional. Our hope is to create a dashboard that can help provide multiple visualizations which provide multiple comparable insights.

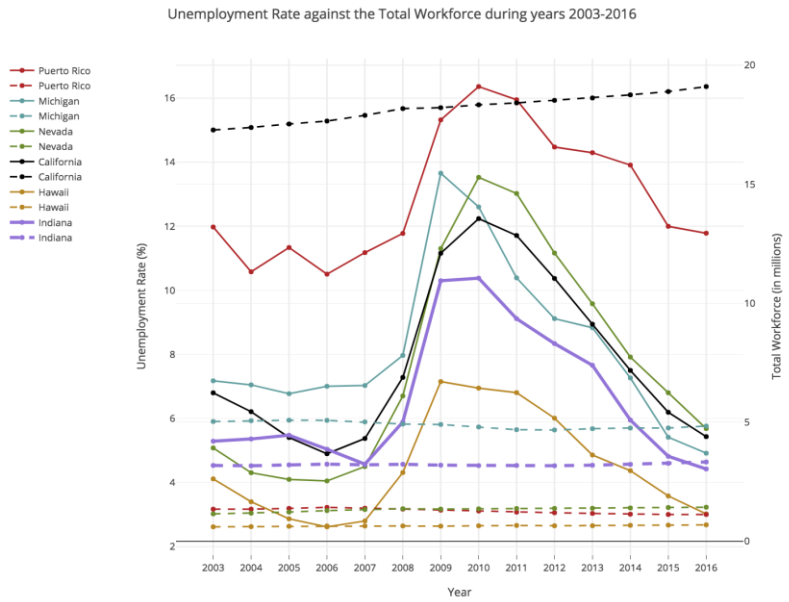


Fig: Labor force vs Unemployment percentage presented on a dual axis. The dotted lines represent labor force and the thick lines represent the Unemployment percentage change between 2003 and 2016.

The overall labor force of the US rose from approximately 147 million to 158 million between 2004 and 2016, a 11 million increase in this period. The unemployment percentage of 5.5% during 2004 is now recovered to about 4.87% in 2016, leaving no trace of the recession. This highlights a positive growth in the job market, generating close to 12 million jobs (based on BLS data) over the decade.

Some of the big states contributed majorly for the healthy recovery of the US, from recession. From the figure above, we can see California had a 1+ million increase of in workforce over the decade and also a decline in the unemployment rate from 7.2% to 5.6%. This signifies an increase in employment. These trends are similar across many states across the US.

Data and Methods

The BLS[2][3] is a governmental statistical agency that collects, processes, analyzes, and disseminates essential statistical data to the American public. The BLS has taken initiatives so that common people can understand the data easily. In recent years, they have done more to include visualizations in their publications. The dataset is available in the public database of 'Google Big Query'[1], which can be accessed using SQL.

Source of data

The data is fetched by the Current Population Survey (CPS), also referred to as the household survey, is a monthly sample survey of 60,000 eligible households conducted by the U.S. Census Bureau for the Bureau of Labor Statistics.

The basic monthly survey gathers demographic characteristics of people in the household and information to determine whether they are employed, unemployed, or not in the labor force.

Metadata

The data contains the following:

- Number of employed, unemployed, and labor force participation and unemployment rates by:
    - Age
    - Sex
    - Race and Hispanic or Latino ethnicity
    - Educational attainment
    - Veteran status

- Employed people by occupation, industry, self-employment, hours of work, full- or part-time status, including involuntary part time.

- Unemployed people by duration of unemployment and general reason for unemployment.

- Median weekly earnings by demographic characteristics, occupation, full- and part-time employment status, union and non-union status.

- Number of hourly-paid workers with earnings at and below the federal minimum wage.

- GDP and Inflation changes year over year.

- Unemployment rates across multiple industries.

Research Questions

How did the unemployment rates change over time?

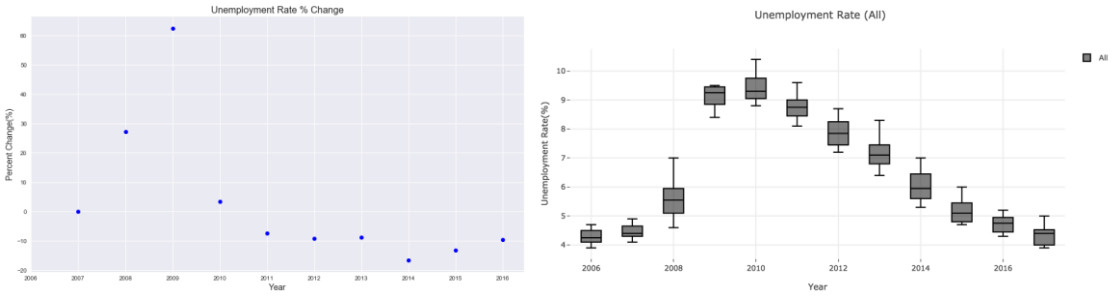


Fig: Percentage Unemployment change between 2006 and 2016.

We used the boxplot to show the monthly unemployment values (min, max, median unemployment percent in the year). Comparing the different unemployment values in a year, we will be able to compare the minimum, maximum, and median unemployment values across different years.

For example: When comparing box plot for different years, we can see that in 2008, there is a very large range of unemployment (between 4.5% and 7.0%). Also, the maximum unemployment rate in 2014 is lesser than the median rate in 2013.

- In this case, we had an option to use either scatter plot or bar chart. We used scatter plot to avoid the clutter that the bar chart produces.
- A scatter plot of the monthly data over 10 years would generate (12*10 =) 120 points and visualizing all the 120 points would make the scatter really hard to comprehend.

As we can see from the plot, the percentage change in unemployment rates in 2006, 2007 is not as significant as the ones we see from 2007. From 2008 we can see a significant increase in the unemployment rate till 2010. This is the period where the US suffered through Great Recession. From 2011, we can see a decline in the percentage change in unemployment signifying a positive status over time.

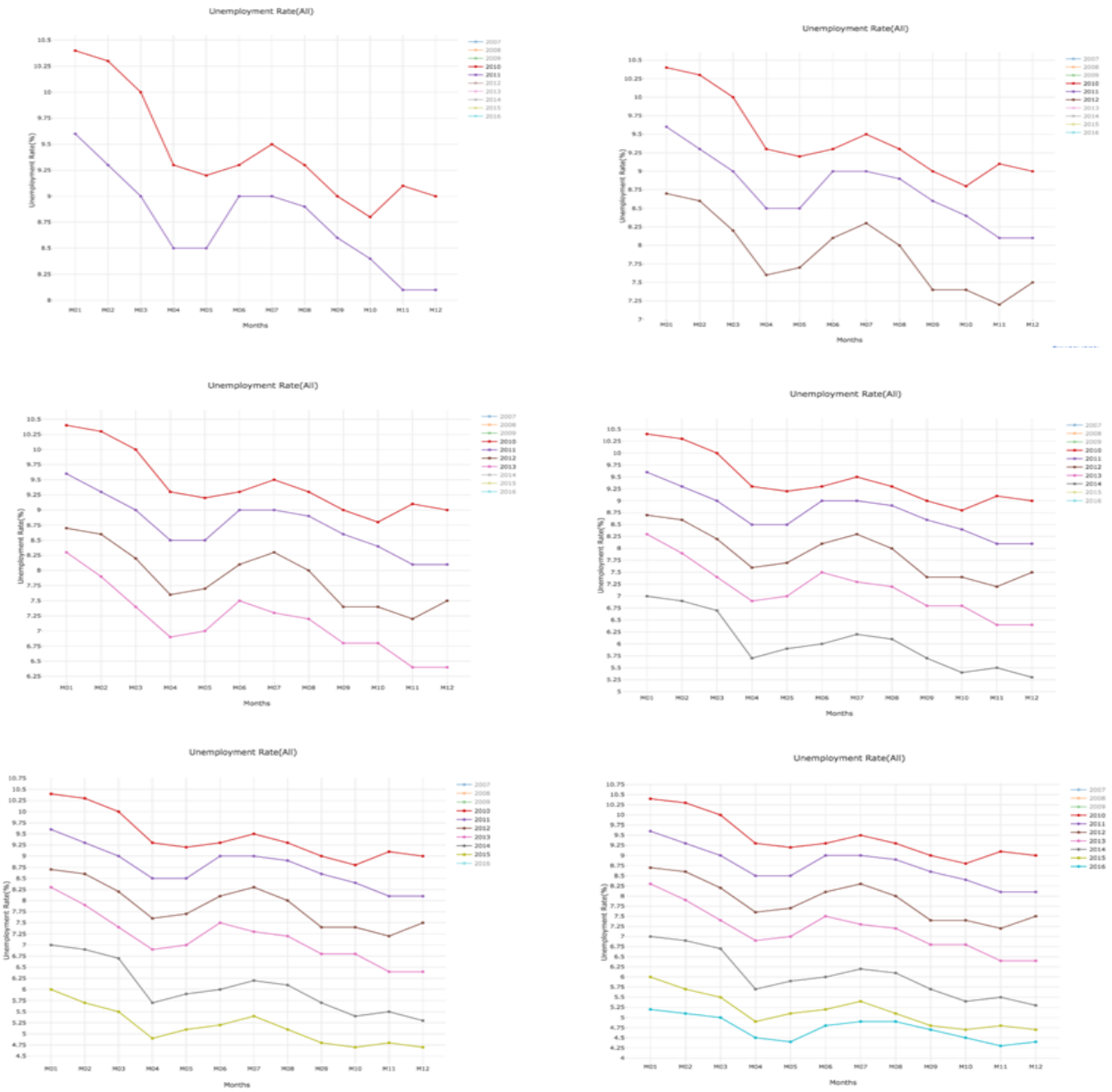
From the box plot, we can say that the unemployment rate in 2016 is as much as it was in 2006 and 2007, i.e. pre-recession period.

Did the unemployment change have a monthly pattern?

We wanted to capture the change in unemployment rate for all the months and see if there was a pattern in the data points of a particular month, where the unemployment rate was affected during this period.

Fig: Percentage Unemployment change across months in 2007, 2008, 2009, and 2010

The unemployment rates have increased from 2007-2010 i.e. during the recession period as shown from above graphs. In 2010, there is a fluctuation in unemployment i.e. it increases initially and starts to decrease from June showing that after the end of recession, the market took time to stabilize.



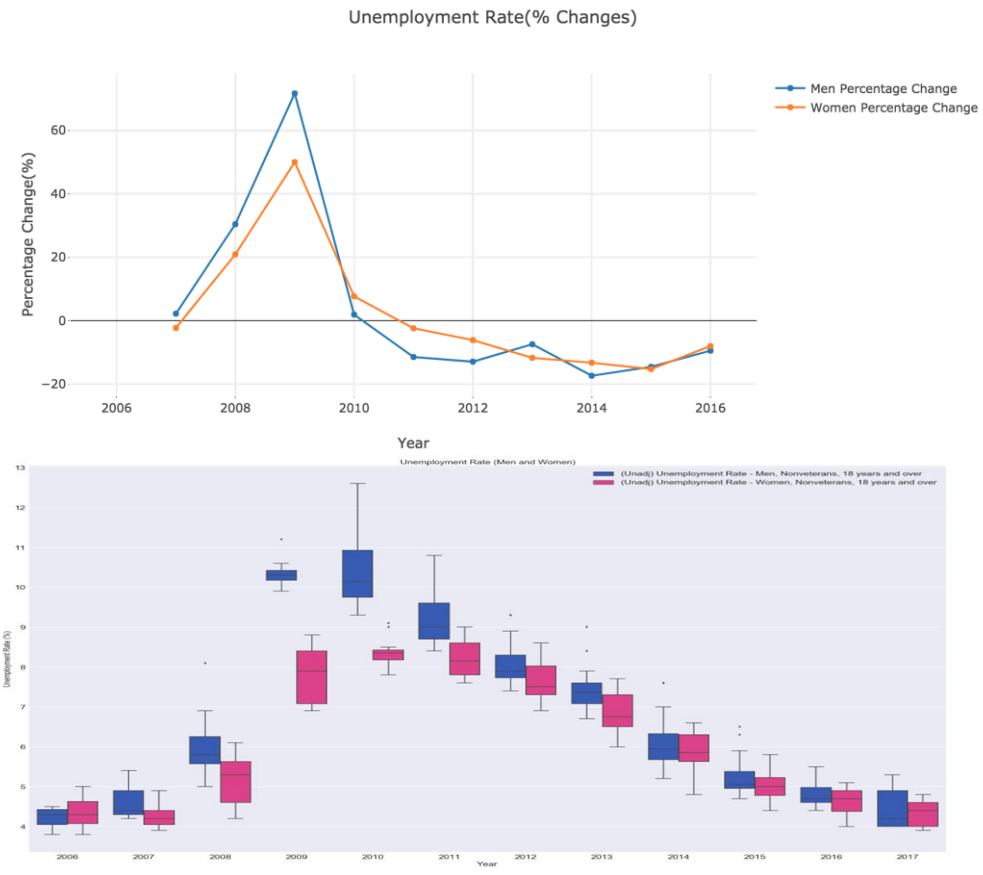
The unemployment rates start to decrease from 2011 as seen above. As seen, the unemployment rate has decreased every year and during 2016 the rate is nearly 4.5% which was last seen before the recession i.e. 2006 and 2007.

Since this is a time series data, we have decided to use a line chart. This helps us track the trends and also, the color helps us distinguish between the different years.

- The other options are scatter plots, but making 10 different classes i.e. for each year and plotting scatter plot would lead to just points on the graph with 10 different colors for each year. It makes it difficult to track the change between the different years, hence we opted for the line charts.

How were the unemployment rates affected by gender?

We wanted to see if there was a gender bias in the way the recession affected the unemployment rates for both men and women. We use a line chart to compare the trend and a box plot to see the min, max, and median distributions across every year.



From the line graph, we see both the men and women have similar unemployment rate change in 2006 and 2007. But in the years 2007-2010, the percentage increase in unemployment rate for men is significantly higher than women, and this highlights the fact that men were affected greater than the women during the recession. Also, during 2009, we see that the men have nearly same unemployment rates (9.8% - 10.5%) throughout the year whereas the women have varying unemployment rates (6.9% - 8.9%). Looking closer into the recovery from 2010, we see the men recovered quicker than the women. 2010 to 2012 saw a much greater decline in unemployment rates for men indicating that more men were hired.

We used a line chart to track the trend in the percentage change year-on-year and a boxplot to compare 10 years of monthly data and visualizing all the 120 points without clutter. Boxplot gives us a better picture of the minimum, maximum, and median unemployment values in a year and makes it more helpful in comparing between genders.

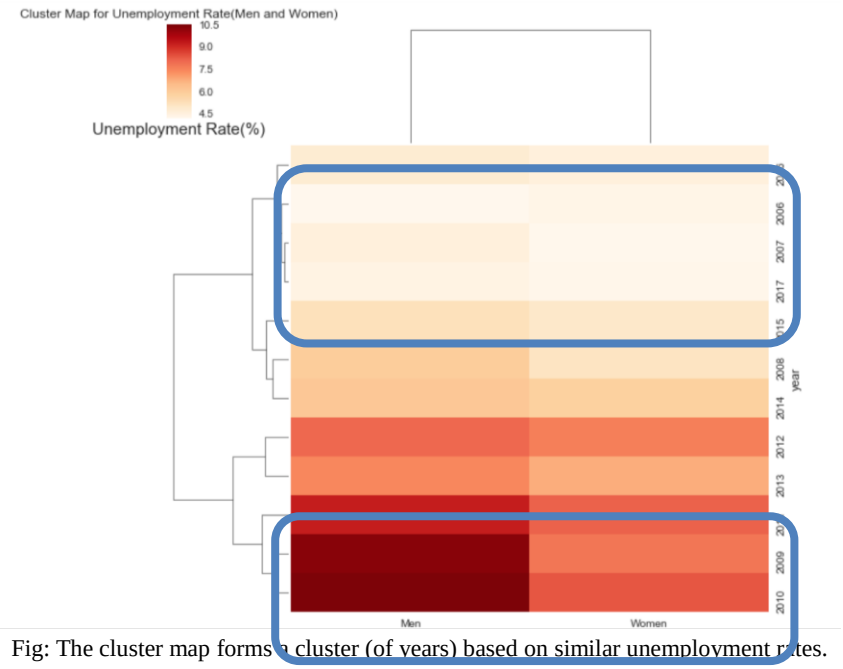
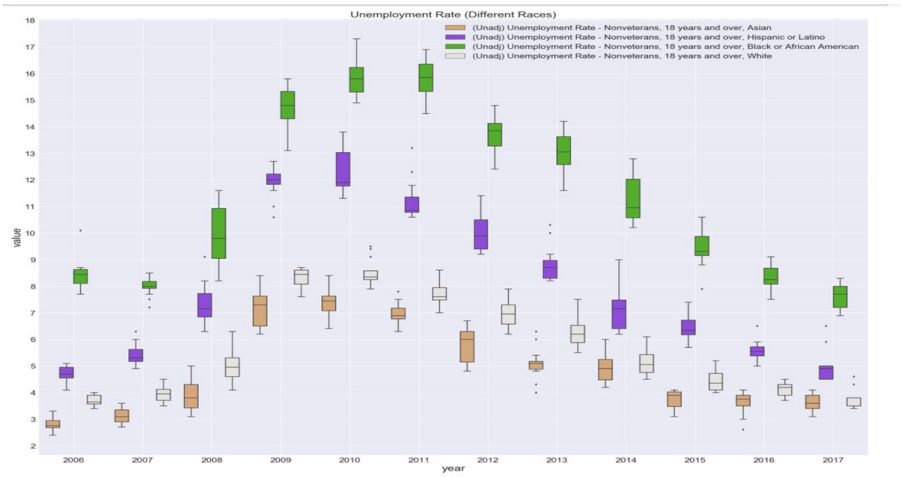


Fig: The cluster map forms a cluster (of years) based on similar unemployment rates.
The bottom highlighted clusters indicates a higher unemployment rate (around the time of recession). Likewise, the upper highlighted clusters indicate a lower unemployment rate (pre/post stability).

We can see from the clusters that the dark red band at the bottom (2009, 2010, 2011) are the periods affected by recession. The light band at the top (2017, 2007, 2016, 2006) are the pre/post periods of recession. These have formed one cluster which indicates that the rates are very similar if not equal, indicating the recession effect is no longer persistent. The cluster (years with less unemployment rate) shows that it took about 6 years post the Great Recession for the unemployment rates to reach such low rates that were last seen in 2006.

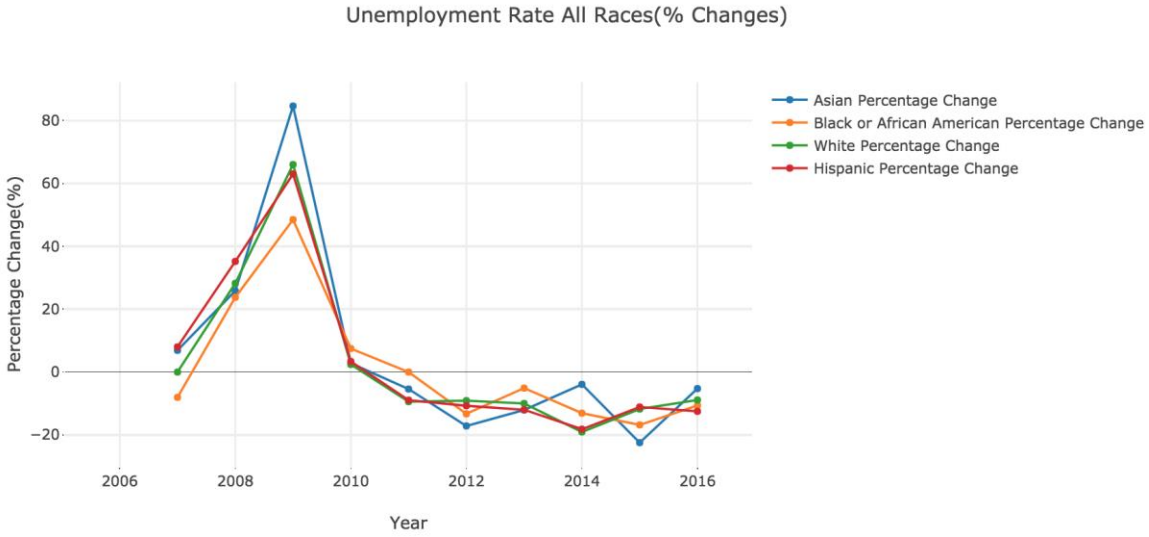
How were the unemployment rates affected by ethnicity?

We wanted to see if there was an ethnicity bias in the way the recession affected the unemployment rates. We wished to see if Asian, White, Hispanic or Latino, Black or African American are affected differently. We use a line chart to compare the trend and a box plot to see the minimum, maximum, and median distributions across every year.



We used the boxplot to show the monthly unemployment values (minimum, maximum, median unemployment percent in the year). Comparing the different unemployment values in a year, we will be able to compare the minimum, maximum, median unemployment values across different ethnic groups and different years.

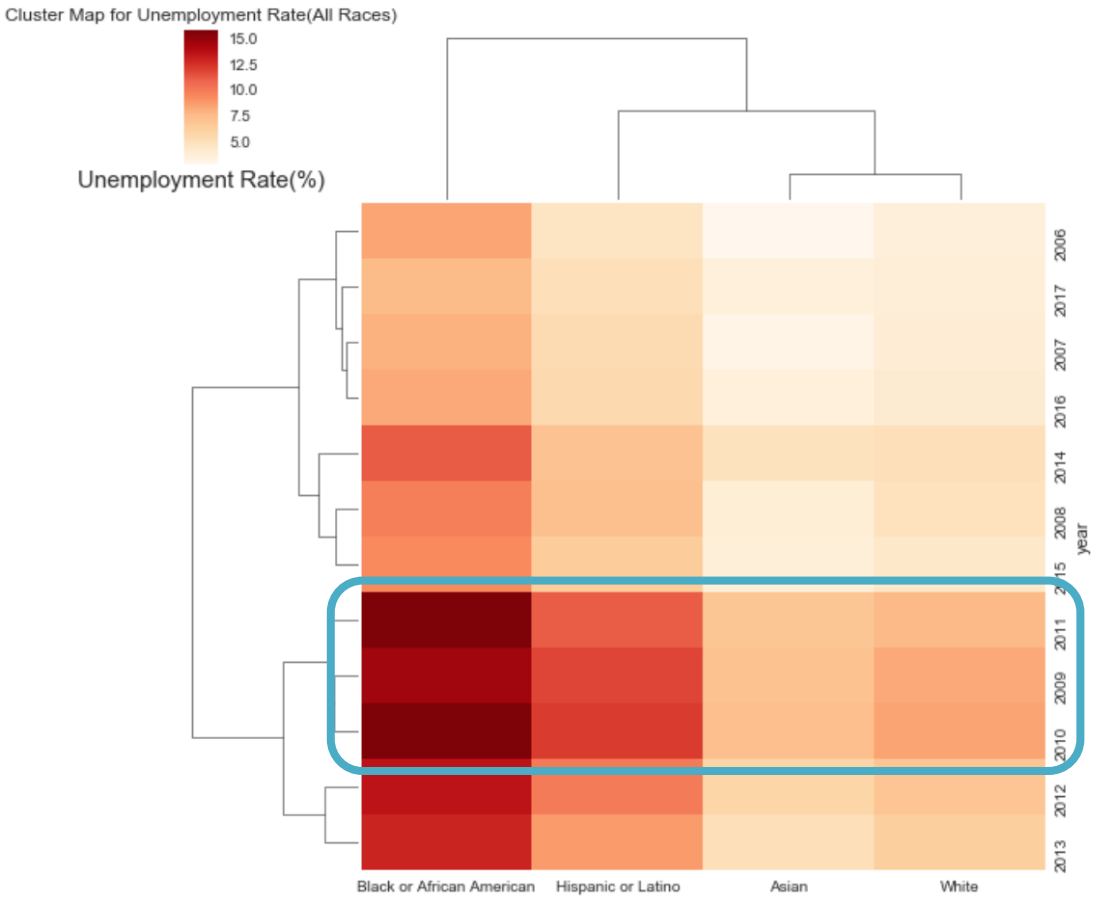
For example: When comparing box plot in the years 2006, we can see that the median unemployment rate is 8.5% whereas the median unemployment rates for the other groups are much lesser.



During the period 2008-2010, the percentage increase in unemployment rate for Asian was significantly higher compared to other races and for Blacks or African Americans it was least. Even though they have

the highest unemployment rates as seen in the boxplot above, the change in unemployment rates over the year is much lesser than the other groups. Based on the percentage change we can say that Asians were affected the most and Blacks or African Americans the least.

We used a line chart to track the year-on-year percentage change across the four ethnic groups and a boxplot to compare 10 years of monthly data and visualizing all the 480 points without clutter. Boxplot gives us a better picture of the minimum, maximum, and median unemployment values in a year and makes it more helpful in comparing between the different groups.



Even before the Great Recession, the employment opportunities for the African Americans were scarce. During the recession period the highest unemployment rates were noticed for African Americans and Hispanic, with Asians affected the most with unemployment rate increasing from 2.9% in 2006 to 7.4% in 2010. Likewise the recovery was quicker for Asians and Whites and much slower for the African Americans and Hispanics.

Comparing the job market during and post the Great Recession period:

The cluster map forms clusters of years and gender/ethnicity. As seen from first cluster map, the darkest region corresponds to the highest unemployment range and is for men in the year 2009, 2010, and 2011. This proves the above two sections and we can do this as an alternative to derive better insights from a single graph. Similarly, the least unemployment rate corresponds to the year 2006, 2007 (pre-recession) and 2016, 2017. This shows that unemployment rate in 2016 and 2017 has now reached percentages that was before recession.

The similar trends can be observed for the cluster map represented for the different races.

- Another way to represent this is a heat map. You can just see the change of values over the years. And suppose you have more than 20 years and you make a Heat Map in order to check how the values have been changed over the years but in a cluster map, we derive similar clusters. The period with similar values are drawn together as a cluster thereby making it an easy task to analyze the data.

Which regions were affected the most in terms of employment? And which regions helped most in the recovery of the US economy?

We wanted to check the unemployment throughout the United States for different years and thereby trying to show that unemployment was highest during recession. And hence we decided to go with Maps. We also wanted to see if the effect was regional and how the recovery took place. Was this also regional and did any states in particular help alleviate the condition.

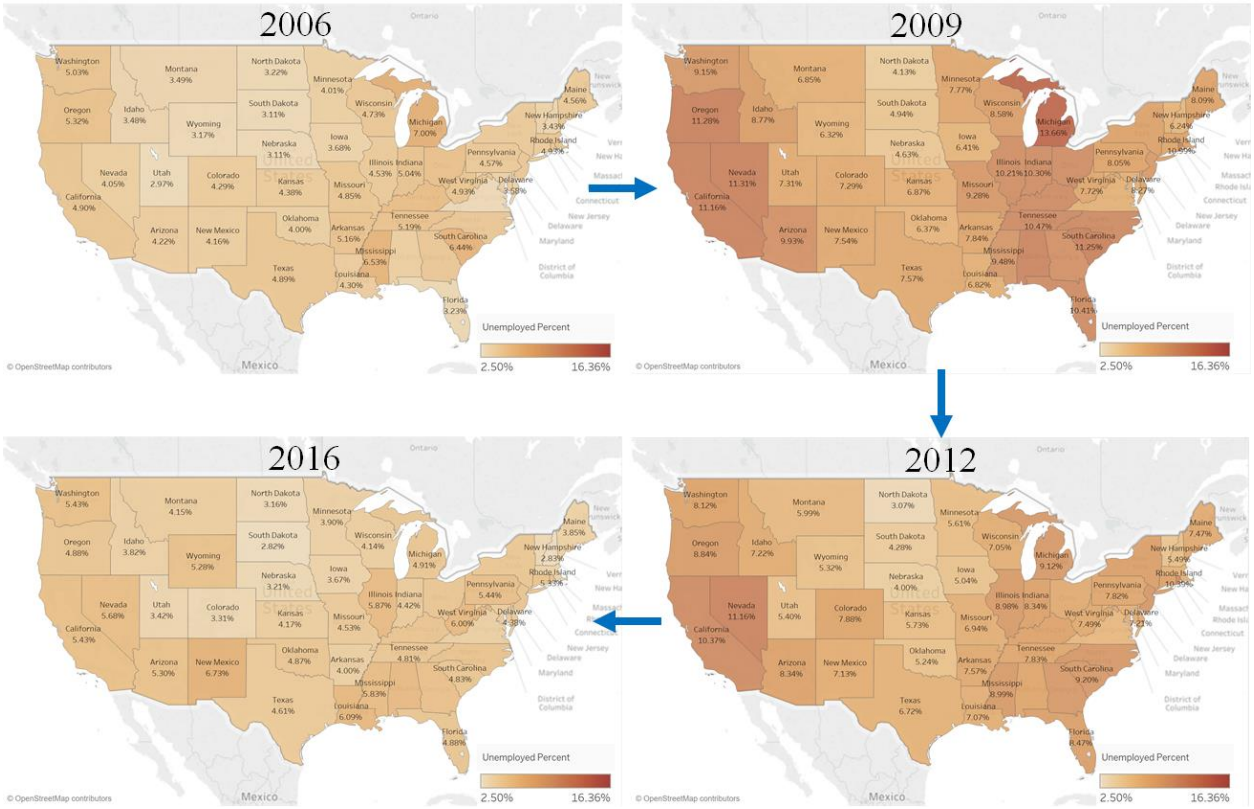
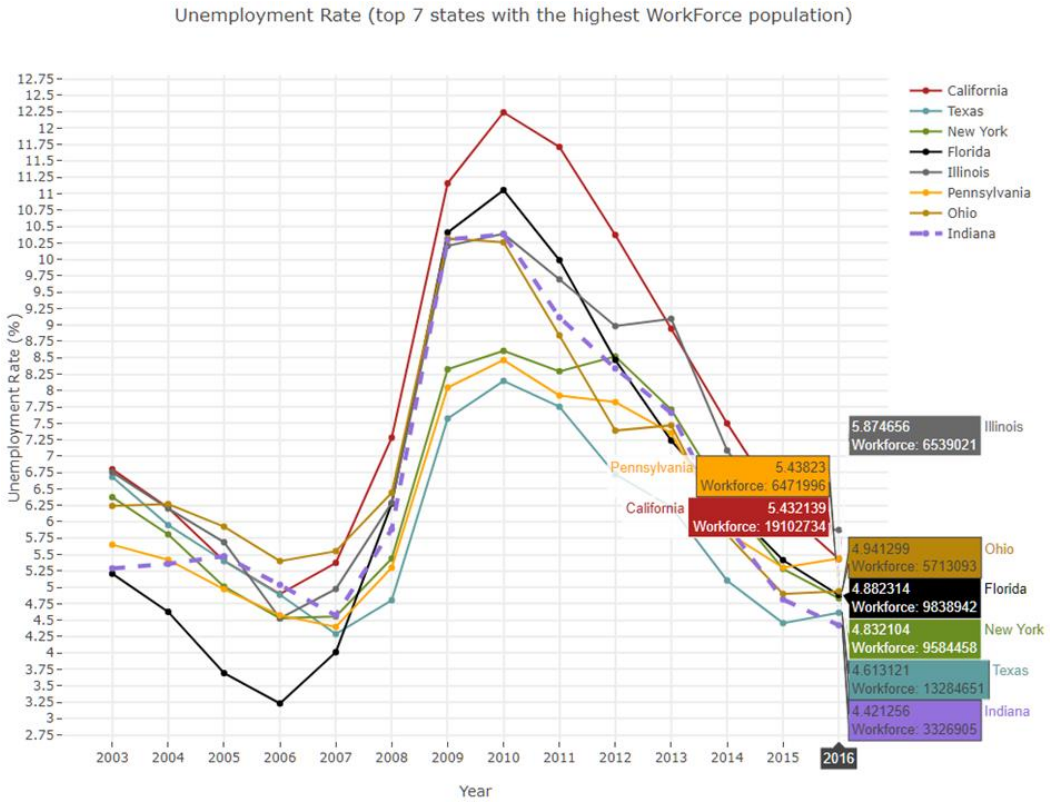


Fig: Change in unemployment rate over the years across various states in the US.
We observe a very similar unemployment rate in 2016 (post-recession) as compared to 2006 (pre-recession)

Across the states, as seen from the graph the colors are darker in 2009, signifying in 2009 the unemployment rate was high through the nation compared to other years. And in 2016, the map again looks similar to the one which was in 2006 indicating by 2016 unemployment rate had improved and it was similar to what it was pre-recession.

We can also see that the west coast and the southeast regions got affected the most because of the Great Recession. Likewise, they recovered quicker than the other regions.



As seen in the above graph, the 4 of the many states that had the largest labor force are from the West and these big states had the quickest recovery compared to all the 50 states of the United States. These are some of the biggest IT capitals of the US and could hint that the IT sector helped in the recovery (an opportunity for further analysis).

Unemployment change across major Industries?

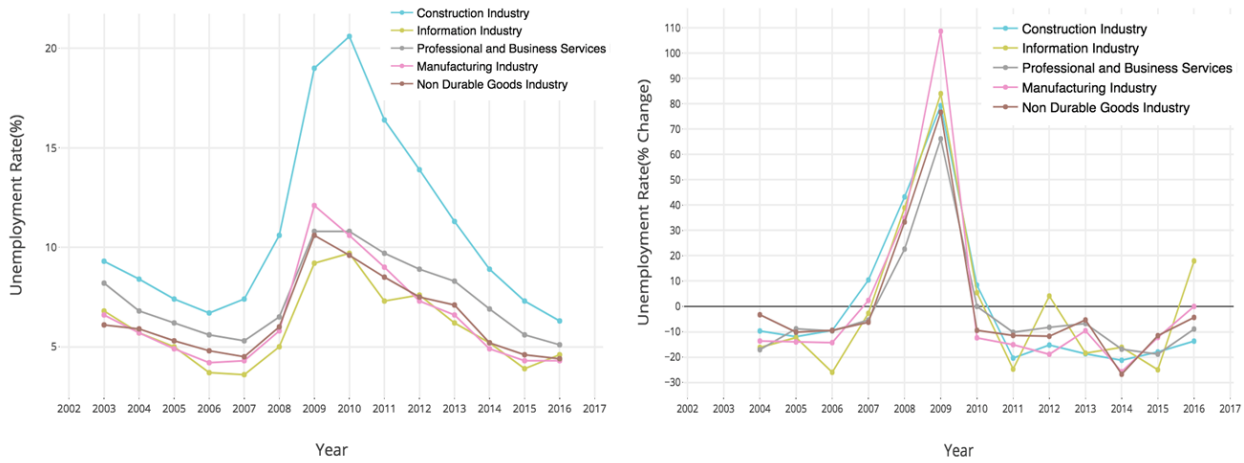


Fig: Unemployment rates across the five most affected industries in the US.

During an economic downturn, people hold off on making major purchases. Furniture usually falls into that category, and home furniture stores suffered as a result. Building Material and Supplies Dealers - construction projects dipped during the recession saw sales decrease during both 2009 and 2010. Construction Materials Wholesalers industry including establishments that sell plywood and bricks, and wholesalers of roofing, siding, and insulation materials faced heavy loses. Like the cement and concrete product manufacturing industry, this one is also dependent upon construction projects, and when those came to a grinding halt in 2009, the industry took a major hit.

“If there are fewer construction projects in progress, contractors would be buying fewer supplies.” This made the building material and supplies dealers industry the hardest hit of the recession.

Also, vacation or travel is one of the first things that are cut when people are tightening their budgets; travel industry and hospitality industries have been greatly affected by the recession.

What were the effect of GDP growth and Inflation on the unemployment rates?

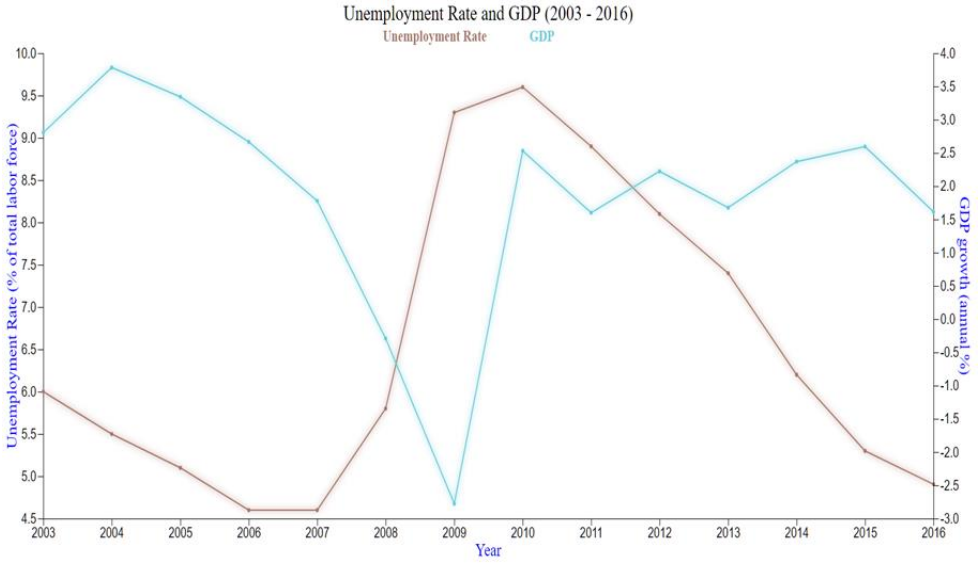


Fig: GDP growth rate vs Unemployment rate change.

GDP growth percentage started to decline in 2004 but 2007-2009 saw the greatest decline, starting the Great Recession. Since then, the US has greatly recovered compared to other countries around the world.

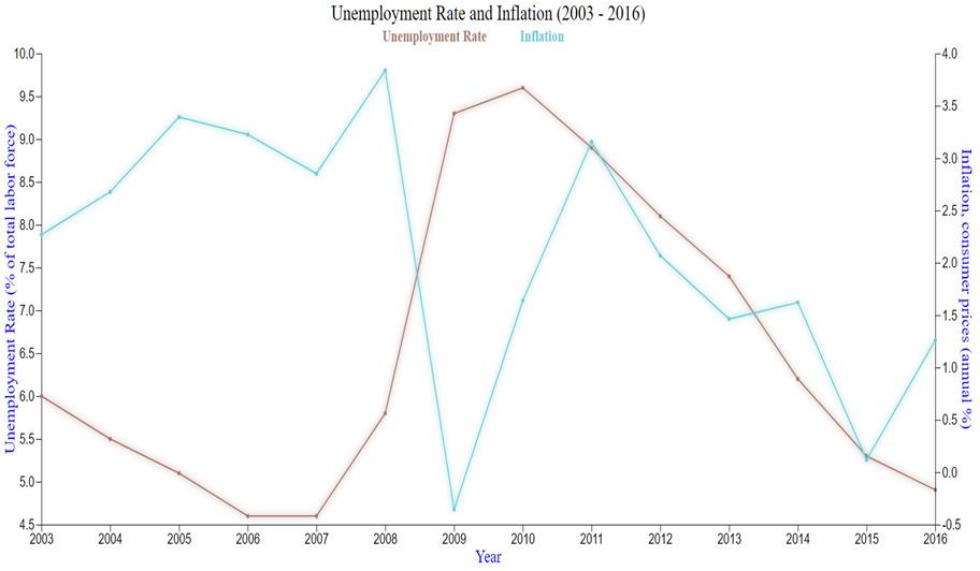


Fig: Inflation (consumer price index) with changes in unemployment rates.

During the year 2008, US saw the maximum inflation which lead to a decline in GDP growth percentage starting the Great Recession. Since then, the inflation has fluctuated and caused instability in the US markets.

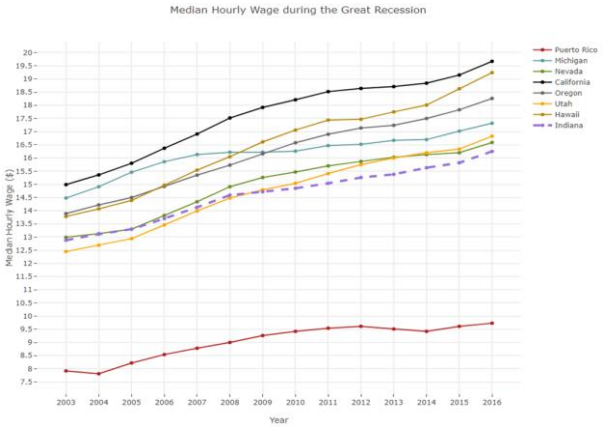
Inflation:

Inflation is either an increase in the production (printing) of money or an increase in price levels. Increase in money production will increase prices of products and services because of the availability of ample cash which will encourage people to spend it fast and increase demand of products, causing their prices to increase. To an extent inflation is good as it encourage companies to increase production because of the increase in demand and this increases GDP. But excessive or uncontrolled inflation is not good for any economy. Hyperinflation is caused when the increase in the amount of money printed exceeds the growth in the amount of goods and services, leading to debasement of the currency. The year 2008 saw the high value of inflation at 4.1% (Consumer Price Index). According to the economy theory, if there is a slack in the economy, the inflation should slow. This is not what we have found in the data. This problem arises from the fact that the Bureau of Labor Statistics measures inflation for owner-occupied housing by looking at the average rents paid at a sample of rental units that are tracked over time. Since rental prices for many leases were set a long time back, they only slowly reflect changes in market prices [15].

GDP:

Gross domestic product in the United States represents the total aggregate output of the U.S. economy [14]. A very high Gross domestic product values is also a problem, as it comes with inflation. An ideal GDP is between 2.5-3.5% GDP growth per year. The year 2009 saw a very low value of GDP growth rate at -2.8%.

Has there been an impact on the hourly wages over time and across the states?



As we can see the hourly wages are increasing for many states that has shown a positive recovery but states like Indiana and Michigan hasn't had much increase in hourly wages as the state hasn't recovered much, as seen from the maps. This gives us the opportunity to look into the state policies of these states and see where they went wrong.

Apart from the above visualization related to wages, we checked mean and median wages for different industries. These visualizations didn't have any convincing trend, hence we decided not to add them here. These visualizations are included in the jupyter notebook file.

Insights and Summary

After carefully examining the data, we have seen a trend of recovery in the unemployment rate across all the states in the US. During the great recession period, the average wages decreased by 1.6% compared to pre-recession period. However, by the year 2016, the salaries rebounded and experienced a growth close to 16% compared to pre-recession period (2007).

The year 2014 marks the actual recovery from the effects of recession across the nation. The recovery started from 2011 as seen from our visualizations, but the recovery rate was slow and it took around 6 years to recover from the effects of recession.

Geographically larger states like California, Texas, Florida, and New York were amongst the most affected states due to recession, as they are amongst the most populated states as well. The most affected region was the West coast, as states like California, Nevada, and Oregon, were amongst the states with the highest unemployment rates during the recession. Even though the Great Recession affected unevenly across different demographics, the effects trickled down to all the sections of the society.

The unemployment rates for both men and women before the recession were similar. But during the recession the unemployment rate of men significantly increased compared to the unemployment rate for women and thereby suggesting that men were more affected during the recession. Even though men were the most affected, looking closer into the recovery from 2010, we see the men recovered quicker than the women. From the year 2010 to the year 2012, there was a much greater decline in unemployment rates for men indicating that more men were hired during this period compared to women.

Even before the Great Recession, the unemployment rates for the African Americans were high. During the recession period the highest unemployment rates were noticed for the African Americans and Hispanic, with Asians undergoing the worst increase in unemployment rate from 2.9% in 2006 to 7.4% in 2010.

From the dual axis graph, the unemployment percentage of 5.5% during 2004 is now recovered to about 4.87% in 2016, leaving no trace of the recession. This highlights a positive growth in the job market, generating close to 12 million jobs (based on BLS data) over the decade.

Since the Manufacturing industry is one of the most affected industries, a lot of metro areas encountered a gap in employment opportunities during the recession. The employment was severely affected for workers with at least a high school diploma, but less than a Bachelor’s degree.

Year 2007-2009 saw the steepest decline in the GDP growth rate for US. This decline lasted for 9 continuous financial quarters starting from December 2007. The real GDP or inflation adjusted GDP started improving after the policymakers implemented the Financial Stabilization Bill (TARP) and the American Recovery and Reinvestment Act. Post 2009, the actual GDP has averaged around 2.2 percent annual growth, as desired. Also, the Recovery Act boosted the demand for many goods and services thereby preserving jobs in the recession.

Conclusion

Leveraging the detailed analysis and the statistics that we have plotted, we are able to conclude that the Great Recession is no longer in effect or has a very diminished effect on the overall unemployment.

We hope that this project is able to debunk the misconception that the people of US have about how the economy is yet to recover from the great recession.

Demographic statistics highlights majority of states to perform better than the pre-recession period. Even with the surge in workforce, strong decline in the unemployment rates indicates that more jobs are generated and the economy is having a stable growth. We would like to conclude that more jobs are being generated and the economy is booming.

Here are some interesting visualizations that we have created for this project using different data sources:
Tableau Visualizations
D3 Visualizations
Python Visualizations

Project Repository Links:
https://github.com/ankit2saxena/DV-Project-Report
https://github.com/shailendrapatil92/Data-Visualization

References

[1] Bureau of Labor Statistics Data, Google Cloud Platform

[2] Occupational Employment Statistics

[3] Local Area Unemployment Statistics

[4] https://blog.dol.gov/2015/05/22/visualizing-bls-data-to-improve-understanding

[5] https://www.bls.gov/opub/ted/

[6]https://www.bls.gov/spotlight/2017/expenditures-on-arts-movies-sporting-events-and-other-entertainments-and-recreational-activities/home.htm

[7] https://en.wikipedia.org/wiki/Bureau_of_Labor_Statistics

[8] https://www.bls.gov/opub/btn/home.htm

[9] https://www.bls.gov/opub/reports/

[10] https://www.forbes.com/sites/jeffreydorfman/2015/04/14/the-unemployment-rate-does-not-measure-labor-market-strength/#272dffee12cd

[11] https://en.wikipedia.org/wiki/List_of_recessions_in_the_United_States

[12] http://www.nytimes.com/interactive/2009/11/06/business/economy/unemployment-lines.html

[13] http://www.macrotrends.net/2509/unemployment-rate-by-education

[14] https://www.investopedia.com/articles/06/gdpinflation.asp?lgl=af-top-textlink-baseline

[15]https://www.economy.com/dismal/analysis/datapoints/284051/Why-Inflation-Didnt-Slow-Much-During-the-Great-Recession/