

Bertelsmann Udacity Data Foundations Nanodegree Program Project # 4

Selma Yildirim
October 3, 2018

1st Insight:

Link to the dashboard:

https://public.tableau.com/profile/selma8149#!/vizhome/ChildPovertyintheUS_2/ChildPoverty?publish=yes

This dashboard shows the relation between the average percentage of child poverty and average income level across all of the United States. Here, by child poverty, we mean % of children living under the poverty level in 2015. According to the guidelines posted by the U.S. Department of HHS, the poverty threshold for a family of four in the year 2015 is given as \$24,250. The average child poverty rate across all of the US is 24.18% while the median child poverty is 22.70%. This indicates that the data for the whole population is slightly right-skewed. The histogram shows the distribution of the avg. child poverty across all of the United States and Puerto Rico.

As the bar graph shows, the highest child poverty rate is in Puerto Rico (60.21%) followed by Mississippi (36.26%) and Georgia (31.76%).

The third plot in this dashboard is a scatter plot showing the relation between child poverty and income levels. The general trend is that the higher the avg. income, the lower the avg. child poverty percentage in that state. However, in the District of Columbia, the child poverty rate percentage is high despite a higher avg. income. This may be correlated with many high income people living in that state which affects the average significantly.

Design choices:

- The best plot showing the distribution of the child poverty levels is a histogram.
- I chose a bar plot to look at the child poverty percentages to compare the values for the States and Puerto Rico. I sorted the bars from the tallest bar to shortest one as it makes it easier to compare and it is easy on the eyes.
- A scatterplot is one of the best options to show the relation between two quantities and to find values that are not following the general trend.
- I used US map, to look at the location of the counties in the US.
- Also, blue color palette works for the colorblindness.

Websites used:

<https://www.kaggle.com/muonneutrino/us-census-demographic-data/data>
<https://aspe.hhs.gov/2015-poverty-guidelines#thresholds>

2nd Insight:

Link to the dashboard:

<https://public.tableau.com/profile/selma8149#!/vizhome/Whowalksmore/walking-data?publish=yes>

In this dashboard, I explored the characteristics of people walking to work in the United States. I decided to look at the social classes defined by an Urban Institute Report (link: https://www.urban.org/research/publication/growing-size-and-incomes-upper-middle-class/view/full_report). According to that report, five social classes by 2014 income levels (for a family of three or equivalent) are given as follows: If income level is less than \$30,000, \$50,000, \$100,000 and \$350,000, the classes are defined as poor, lower-middle, middle, upper-middle and rich, respectively.

For this, I created a new calculated field using Income measure and I renamed this dimension as Social Class. Note that the largest class is lower-middle class, followed by middle class. There is no county in the rich class.

The bar plot shows the avg. walking percentage for these newly defined income levels. Observe that average % of people walking to work is the highest for the middle class followed by lower-middle class. Upper-middle class has the lowest average % among these social classes.

The dashboard also shows the relation between average % of people walking to their work and whether they are employed in public or private jobs are different for each social class. For the middle and middle-low class, the general pattern is that the higher the average % of people employed in public jobs, the higher the average % who walk to their work. This pattern is reversed for the people employed in private jobs. Note that for the middle class, the numbers in the States of Hawaii, District of Columbia and Alaska are significantly different than the rest of the other States and Puerto Rico and for the lower-middle class it is only Alaska. As for States with counties in poor and upper-middle classes, no pattern is observed either due to low number of counties in these classes.

Design choices:

- There are two plots in this dashboard using social class dimension: One shows the number of counties in each income level, the other one shows the relation between avg. percentages of walking for each social class. Since these two shows different calculations (i.e., one is count, the other avg. value) I decided to use different types of plots, packed bubbles and horizontal bars. I sorted bars from tallest to shortest for easy comparison.
- Since public work % and private work % are similar quantities and I was comparing both of them with avg. walking, I used the same plot choice. Since one of the best options to show a relation is a scatter plot, I used scatter plot to easily see relations and the differences in these cases.
- I used US map, to look at the location of the counties in the US.
- I used blue and orange colors because they work for colorblindness.

Websites used:

<https://www.kaggle.com/muonneutrino/us-census-demographic-data/data>

https://www.urban.org/research/publication/growing-size-and-incomes-upper-middle-class/view/full_report

3rd Insight:

Link to the dashboard:

<https://public.tableau.com/profile/selma8149#!/vizhome/TheMostandtheLeastpopulatedStates/Most-Least-populated?publish=yes>

In this dashboard, I wanted to explore the differences and similarities between the States with the most and least population. California is the State with the most population while Wyoming is the State with the least population. So, after I looked the total population for each State, I created a set including California and Wyoming only.

I looked at the median income levels and the relations between the % of production and poverty, % of production and unemployment for California and Wyoming separately.

Unlike Wyoming, there is a strong negative relation between the % of professional people (i.e., people employed in management, business, science and arts) and % of people living under poverty level in California. That is, the lower the % of people working in professional jobs, the higher the % of people under the poverty level.

On the other hand, this trend is different when we look at % in production, i.e., people working employed in productions, transportation and material movement. In both California and Wyoming, the general trend is that the highest the % of people employed in production jobs, the more the % of people under poverty level. This might be related to that production people earn less money resulting in a lower income level. Note that, Albany County is significantly different than other counties in Wyoming when we look at the production vs poverty plot.

When we consider unemployment vs production % and unemployment vs professional % in California, we see that these plots are very similar to their counterparts in the poverty plots. As for Wyoming, the county records are clustered more closely than poverty records and Albany County records are close to other counties in that state.

As for the income levels, we see that, on average, median income level of Wyoming residents are about \$1000 more than California even though there are no counties in Wyoming with median income level more than 90K. This may be caused by the % of people making less than 50K in California is higher than those in Wyoming.

Design choices:

- I used a Histogram for the distribution of median level incomes for California and Wyoming.
- I used Gantt plot for the values of counties in these States because it has a compact view and it is good for looking at individual values.
- As in the other insight, I used scatter plots to explore relation between two quantities and to compare general trends in each case.
- I used blue and orange colors because they work for colorblindness.

Websites used:

<https://www.kaggle.com/muonneutrino/us-census-demographic-data/data>