Insight 1 - Data Dashboard, breakdown by county

Link:
https://public.tableau.com/profile/christine7077#!/vizhome/BuildDataDashboardsProject5/Breakdownbycounty

Map:
The county map is most interesting to look at not when looking at the country as a whole necessarily (although when looking at it as a whole, there are a few counties in brown that really jump out with substantially higher unemployment rates), but when looking at it at a state by state level, or by comparing a few states at a time, such as those in the same region or sharing borders.

I decided to include average income per capita, total population, and average unemployment here, as these are 3 major categories that I believe are quite important to look at when making comparisons using a map.

Scatterplot:
The county scatterplot has a linear correlation coefficient of about 0.4442, meaning that the plot has a moderate, positive relationship between the number of counties in each state and the total population of each state.

There are 2 clear outliers in this plot, which are also the 2 largest states in terms of total population. Texas has by far the highest number of counties with 254, while California's number of counties is not an outlier in itself (58), but is an outlier because its total population per county is quite low due to its high population (about 662,439 residents per county).

Bar chart:
I thought it would be interesting to compare the average commute time across counties in each state to see if there were any differences within each state. Between the maximum and minimum average commutes in each state there are generally noticeable differences, and it would be interesting to learn more about what lies behind those differences. Maybe they are large versus small counties in terms of geographic size, or perhaps more people work from home in the counties with lower commute times? As geographic size of each county is not included in the dataset, I decided to have a quick look at the average percentage of people working from home versus the mean commute time to see if there was a correlation.

Insight 1 Sub-visualization:
https://public.tableau.com/profile/christine7077#!/vizhome/BuildDataDashboardsProject5/County-ScatterplotAvg_commute

If you have a look at the correlation for all counties, the linear correlation coefficient is negative and weak at about -0.2135. Perhaps in some counties there is a stronger correlation, but overall, this weak correlation does not provide much evidence for this being the reason for shorter commutes. Perhaps the city infrastructure is simply better, meaning less traffic and therefore shorter commutes. We would require more data in order to find more meaningful connections here.

Design:
I have chosen to use color marks that are generally easiest for colorblind individuals to see. These colors also help to differentiate the unemployment rate on the map and the different states on the scatterplot. I erased the singular titles from all graphs to reduce clutter, and only included 1 dropdown that functions for both the bar chart and the map.

References:
https://kb.tableau.com/articles/howto/finding-the-pearson-correlation
https://www.thedataschool.co.uk/nick-jastrzebski/rounding-numbers-in-tableau/
https://help.tableau.com/current/pro/desktop/en-us/trendlines_add.htm

Insight 2 - Data Dashboard, Breakdown by state

Link:
https://public.tableau.com/profile/christine7077#!/vizhome/BuildDataDashboardsProject5/Breakdownbystate

Map:
I included a state map to have a look at average income per capita, and the percentage of the population working in various industries (professional, construction, etc.), as these are the measures I am comparing in my scatterplots. The color mark for the map is set as the average income per capita, and it is clear to see that a few states in the northeast region have the highest income per capita (with the highest in Connecticut with $37,025). Puerto Rico clearly has the lowest income per capita at only $9,618. When we filter out Puerto Rico, we see that there are a few states with similar lows, but Mississippi seems to have the lowest with $18,805 (nearly double that of Puerto Rico).

Scatterplot:
I wanted to have a look at the correlation between the income per capita (or average income per person) and the percentage of residents working in the professional, service, office, construction and production census categories, in order to see what relationship there may be between these 2 measures.
Here is a breakdown of these categories as listed on the Kaggle page for this dataset:

- Professional: % employed in management, business, science, and arts
- Service: % employed in service jobs
- Office: % employed in sales and office jobs
- Construction: % employed in natural resources, construction, and maintenance
- Production: % employed in production, transportation, and material movement

The scatterplots are listed in terms of strongest to weakest correlation. 3 of the plots have a moderate correlation: 1. Professional with the strongest, and only positive relationship at about 0.6555, 2. Production, with a negative relationship of -0.3568, and 3. Service, with a negative relationship of -0.3513. The other 2 plots have weak relationships: 1. Construction, with a negative correlation of -0.2309 and 2. Office, with a very weak negative correlation of -0.0379.

This information shows us that some relationships exist between at least the moderate correlations. The strongest relationship shows that the higher the average percentage of individuals working in the professional category is in a given state, the higher the average income per capita seems to be.

Bar chart:
To follow up on the insights found in the scatterplots, I decided to also add a bar chart to the dashboard, to look at where each state falls in relation to the percentage of residents working in the professional category. The District of Columbia has by far the highest average percentage at 60.9% while Alabama has the lowest at 27.69%.

Design:
As my dataset is geographically focused in nature, I wanted to include a map in each dashboard so that the audience can click through the maps at a county, state, and regional level. I have added trend lines in all of my scatterplots in order to make it clear as to the negative or positive relationship of each small multiple plot, and set the color mark to state in order for the various states to be more clearly picked out. I also included a filter so the audience is able to only see certain states on the map and in the bar chart if they choose, or can see all states.

References:
https://kb.tableau.com/articles/howto/finding-the-pearson-correlation
https://www.kaggle.com/muonneutrino/us-census-demographic-data/data#acs2015_county_data.csv

Insight 3 - Data Dashboard, breakdown by region

Link:
https://public.tableau.com/profile/christine7077#!/vizhome/BuildDataDashboardsProject5/Breakdownbyregion

Map:
I created state groups based on the Census designation of the region divisions in the US (link below). The US regions are highly disputed, and are different depending on what source you refer to, but as my dataset is from the 2015 US Census data, I thought it best to go with their designations for this dashboard. I included the average household income and unemployment for each region as well of the total population. I sorted by the average income as this is something I then broke down in the histogram.

Histogram:
In this graph, I wanted to compare the frequency of the median household income (in terms of counties with each income) on the region level, in order to see which income has the highest distribution, and how incomes are generally distributed at the region level. When looking at the graph with Puerto Rico filtered out, the highest point of the graph is at $42,000, and it is clear to see that there is a wide range in the values, with both the minimum income ($17,000) and maximum ($122,000) in the south. It is interesting to see this large range

of $105,000 within the same region, but as the south region is by far the most populous, it is perhaps not so surprising.

Stacked bar graph:
With this graph, I wanted to compare the breakdown of race across regions, and to see which percentage of each race resides in each region. The South clearly has the largest share overall with nearly 260% of the total population while the Northeast has only about 30%. The South does not have the highest percentage for all races however, with the West having a significantly higher percentage of individuals claiming their race as Pacific (62.11% versus the South's 21.86%), and the West also having a larger proportion of Native (40.72% versus the South's 21.82%).

Design:
Across all of my visualizations I kept the color scheme consistent, so that the 3 could be shown together as a set and not clash, and made sure that the color scheme would likely be suitable for anyone who may be color blind. I maintained a large data-ink ratio by removing unnecessary items such as duplicate drop down menus and titles, and only used color to help differentiate between different dimensions and measures.

Links:
https://kb.tableau.com/articles/howto/stacked-bar-chart-multiple-measures
https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf