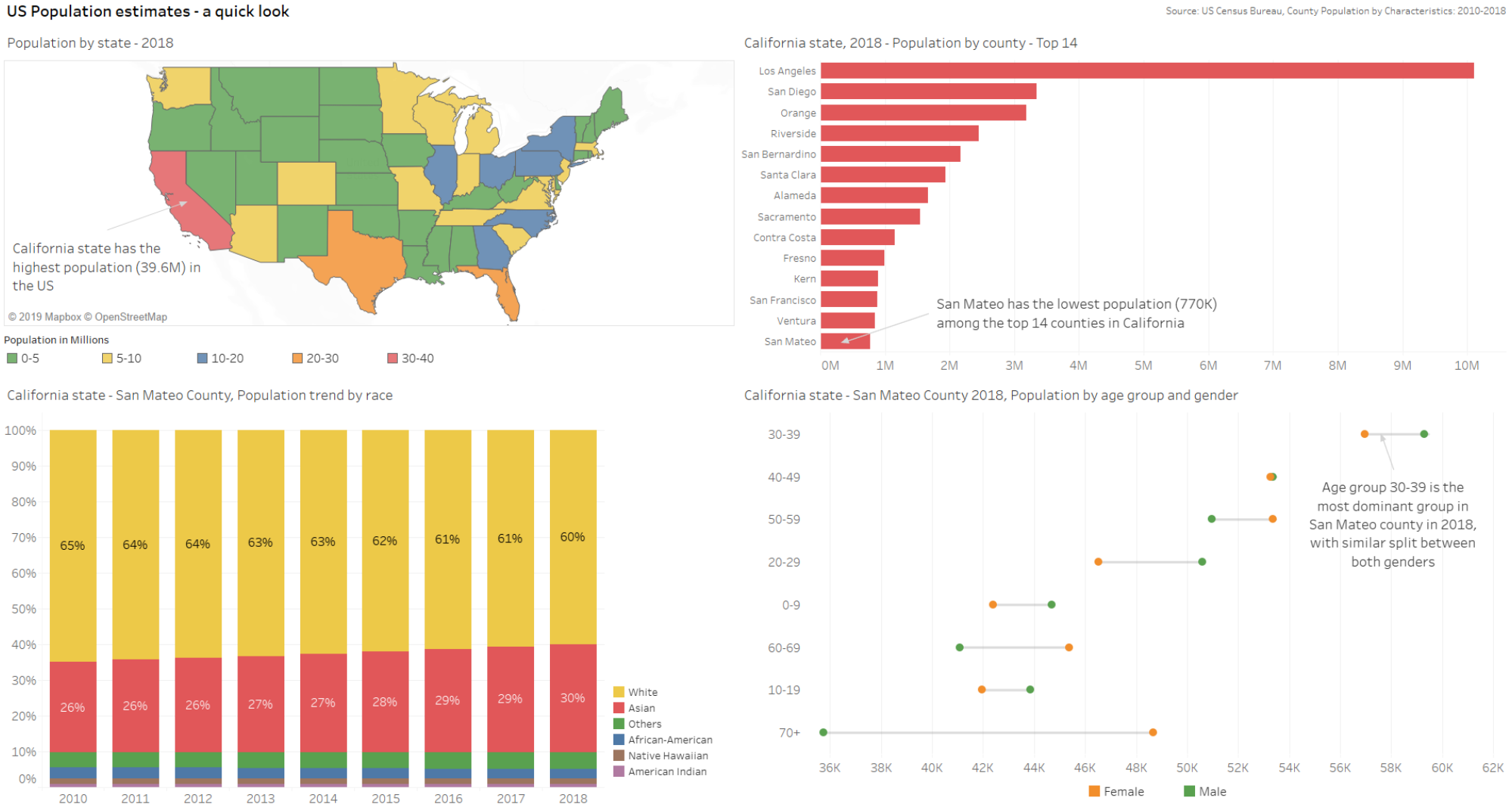

Data source: https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html#par_textimage_1383669527

For this visualization assignment I used the USA county population estimates that was created by the US Census bureau. The dataset has race, gender and age group wise population estimates for the years 2010-2018 for counties across states in the US. As a first step I read through the dataset documentation to understand the dataset. The dataset has a column for each combination of gender and race, ex: White male, white female etc. It also had column totals for the same. There were also multiple combinations of races in the data as people can report more than 1 race. For the sake of simplicity, I deleted all other columns and kept only single race wise columns in the data. I also deleted the column totals since I am going to pivot the data in tableau as it is easier for chart creation while visualizing dimension splits. When I imported the data in tableau, I used the pivot option to pivot all race-gender columns into a longer format. In the year column, I noted that the years were coded as numbers 1-11. I also noted that no.s 1,2,3 represented Census population, and estimates for the year 2010 while

the other values 4-11 represented only population estimates for years 2011-18. I used a data source filter in tableau to use only data for years 3-11 as these represented population estimates for 2010-18. The age group column also used no.s 0-18 to code the age groups. I also noted that 0 represented all groups, so I made sure to filter all my charts by this so that there is no double counting in the data. I created calculated fields using case statements to code the numbers for year and age group columns into appropriate values, as it would not make sense to view charts with the numeric codes. I converted the year field from numeric to date to aid in plotting time series graphs.

In the first chart I used a map to view population by state for the year 2018. Based on the population values, I created a calculated field with 5 population buckets for the population so that we have the right no of buckets to visualize in the map. Since the data has population by geographies, I felt the map would be the best way to visualize the data and get an overview of the population by state. From the map we can clearly see that majority of the states have a population of 0-5 million people. We can see clearly that California has the highest population in the US and is the only state in that particular population bucket (30-40Million). I decided to drill down and see the country wise population split in 2018 for California as a next step. I first thought of using the bar chart with above and below average view to get an idea of county wise population. However, there were too many counties and the average was skewed by LA county population. So I decided to only view the top 14 counties and used a bar chart to visualize the data. We can clearly see LA has the highest population by far in the state. In this case, I could have also used a map to see county wise population, but I felt the bar chart shows the population differences between counties better. The population of the other counties are slightly more similar, making LA look like an outlier.

Next I wanted to see the population split by race and see if there was any particular trend over time. I decided to use a stacked bar chart plotted by year and stacked by race for a particular county. Unsurprisingly, in many counties there was not much of a change in race distribution over time. However, I saw that in San Mateo county, we could see a clear trend over time. The % of white people was steadily decreasing over time and we can also clearly see that there was a simultaneous increase in the % of Asians in the population. Finally I wanted to explore the gender and age group characteristics in San Mateo county for 2018. I felt the best way to visualize both characteristics in one graph would be to use a dumbbell chart. The dots are colored by gender, so they show the population by gender. The y axis has the age groups, so we can see which age group was most prevalent in the population. The chart throws out a clear insight on the highest and lowest population age groups. The most dominant age group in the population was 30-39, with similar numbers for male and female. The age group with least population was 70+, and it is interesting to see that there were more females than males in this group. Overall, I tried to visualize all the characteristics of the population data and tried to drill down at each stage to explore the data at different granularities.