

# ESOC 214 - Data Challenge 06

Jake Moore

## Background and approach

We are working with regional and state level milk production over time data. Population data is more recent than the milk production data, so we treat it as an indicator rather than an exact same-year match. The following process is straightforward for this brief case study:

First, we will import our data and analyze the contents, structure, and integrity. Next, we will identify the timeline of our data, finding the minimum and maximum year to correctly identify a range of time. Following that we will identify the top state by milk production for the year of 2017. Then we will visualize milk production in pounds over time using a scatterplot, coloring in different regions. Next, we will visualize milk produced in pounds on a choropleth map for the year of 2017. Finally, we will compare population vs. milk production in pounds using usmap::statepop using Pearson correlation. We will then dynamically link our data to a scatterplot.

## Limitations:

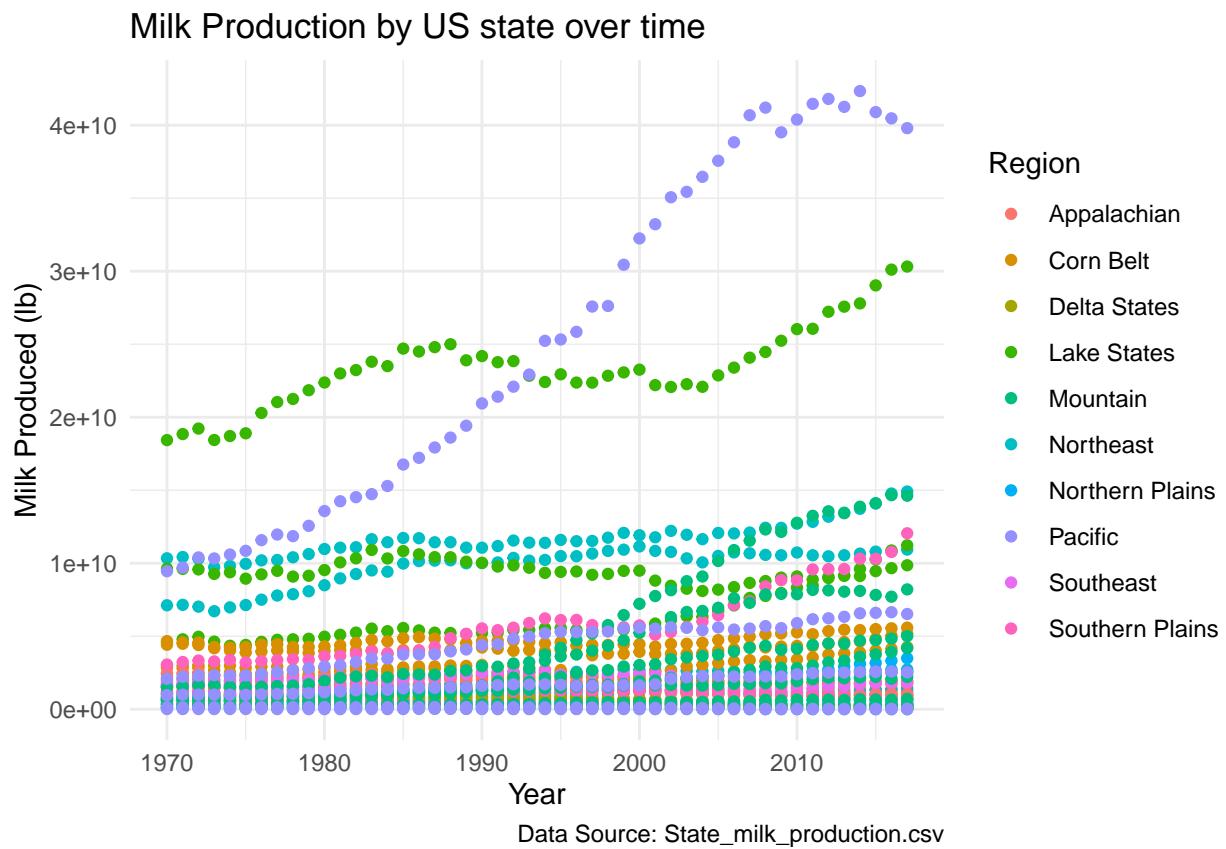
Our data has a few key limitations: - Our correlation value shows a moderate positive correlation (0.63). It is common practice for a moderate correlation to span from 0.4-0.6 and a strong correlation to span from 0.6-0.8. This results in our correlation value being on the very low end of that scale. - Milk production data is only recent to the year of 2017, while state population data returned values as recent as 2022. This could make it more difficult to confirm without a doubt that there is a positive correlation.

## Conclusion:

After running our brief analysis, we were able to successfully confirm that there is a positive correlation between state population and milk production in pounds. However, considering our limitations and our rather weak correlation value, it is likely there isn't a very strong positive relationship present. Further analysis would need to be conducted with more recent data.

Methods:

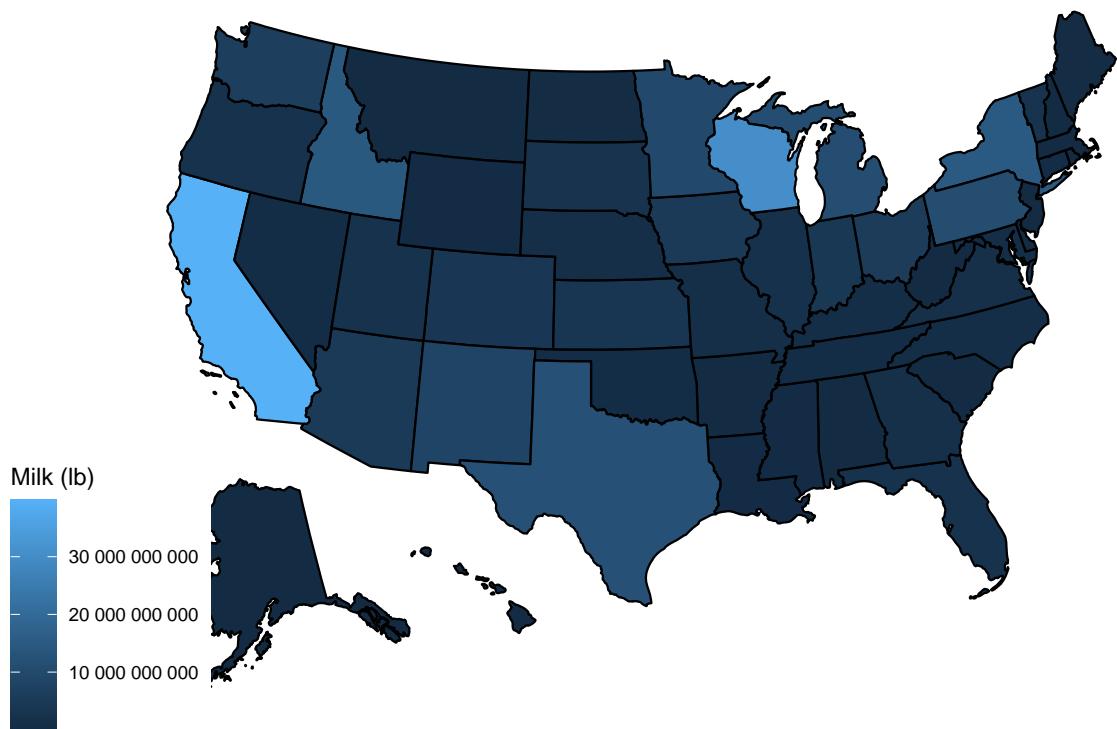
## Part A - Milk Production



```
## The scatter plot visualization above covers milk
## production in pounds from 1970 to 2017, colored by region. It is useful to
## visualize consistent regional differences and long-run trends.
```

## Part B - US map for 2017

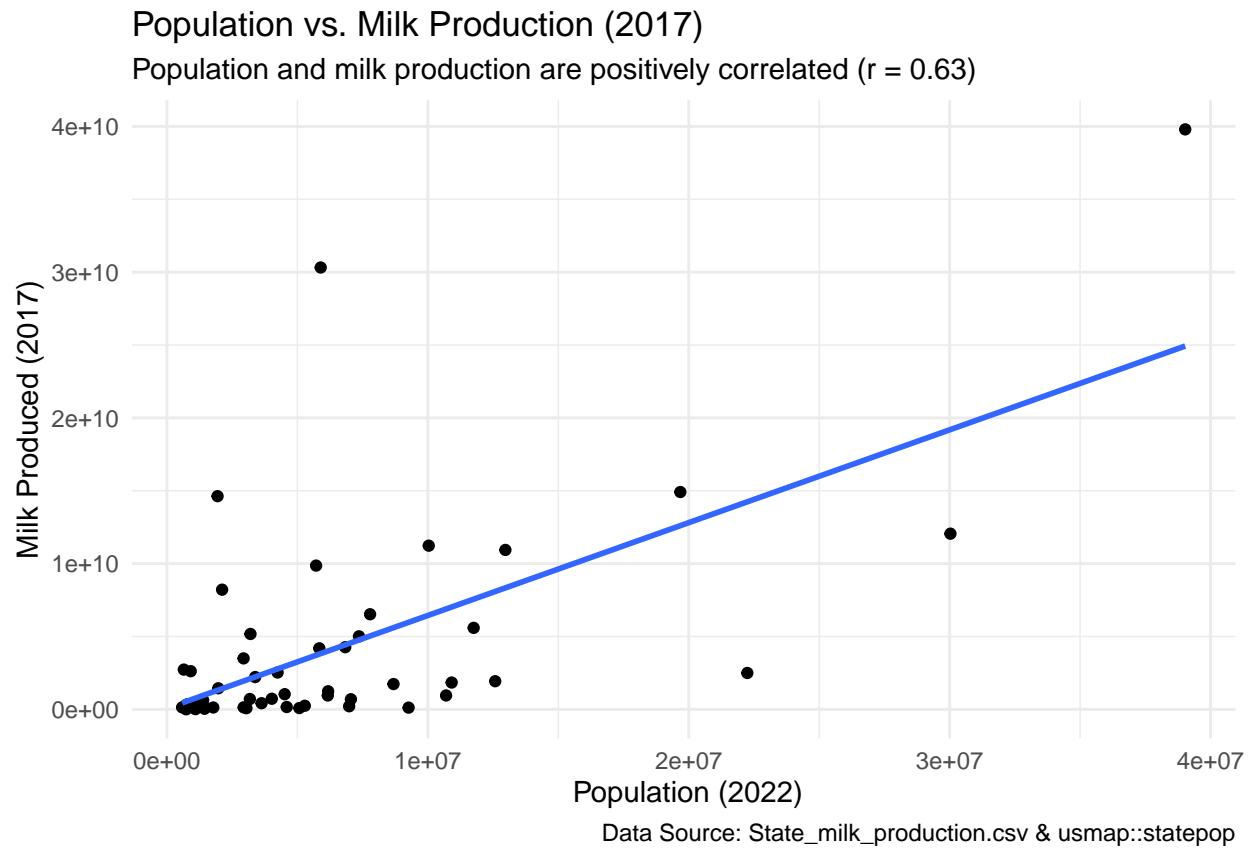
US Milk Production (2017)



Data Source: State\_milk\_production.csv

```
## The 2017 choropleth map above shows where total milk production
## in pounds concentrates across U.S. states. Due to the totals not being
## normalized by population, interpret this as volume, not per-capita production.
```

## Part C — Population vs milk



```
## Population uses the most recent field pop_2022
## (year 2022), while milk is from 2017; treat population as an indicator,
## not an exact match. Population and milk production are positively
## correlated ( $r = 0.63$ ), suggesting larger states tend to produce more milk.
```