written-report

Introduction and Data

Urbanization shapes the demographic and economic conditions of states across the U.S., influencing population structure and workforce characteristics. Our project examines how levels of urbanization may relate to the average age of working-age residents. To explore this relationship, we compare California, the most urbanized state at 94.2%, with Vermont, the least urbanized state at 35.1%. Understanding whether less urbanized areas have older working populations can offer insight into potential economic challenges, such as shrinking labor forces or reduced productivity.

Our analysis uses data from the Current Population Survey (CPS), a monthly survey conducted by the U.S. Census Bureau. The CPS includes demographic and labor-force information such as age, state, and employment status, but for this project we focus only on respondents' age and state of residence. We restricted the dataset to individuals living in California or Vermont to directly compare their working-age populations. This leads us to our research question: **Is the average age of Vermont's working-age population higher than that of California?**

Methodology

The data for this project was obtained from the CPS library in R, stored in our project's "data" folder as data.qmd. We filtered this dataset to include only residents of California and Vermont, resulting in 377 observations across 17 variables. For our analysis, we focused exclusively on the age and state variables to examine the average age of the working-age population in each state.

To test our research question, we conducted a one-sided t-test. In a one-sided test, the null hypothesis assumes that the parameter of interest is greater than or equal to a specific value, while the alternative hypothesis assumes it is less. This approach was appropriate because our study aims to determine whether the average age in Vermont is higher than in California, rather than simply testing for any difference. The t-test was conducted by creating a combined variable for California and Vermont (vt_ca) and testing the difference in means with the null hypothesis that the mean age difference is less than or equal to zero. Additionally, we visualized

the age distributions of the two states using box and whisker plots. Separate plots were created for California and Vermont, with age represented on the y-axis. These plots allow for a visual comparison of the central tendency and spread of ages within each state, highlighting differences in distributions and variability.

Results

The one-sided t-test comparing the average age of the working-age population in Vermont and California showed that Vermont has a higher mean age than California. The test provided statistical evidence to reject the null hypothesis, supporting the conclusion that Vermont's workforce is older on average.

The box and whisker plots further illustrate the differences between the two states. Vermont's age distribution is skewed toward older ages, with a wider interquartile range, indicating a larger proportion of residents above the mean age. In contrast, California's distribution is more compressed and centered around a slightly lower mean, reflecting a younger and more evenly distributed workforce.

Together, the t-test and visual analysis confirm that urbanization appears associated with differences in workforce age, with the less urbanized Vermont showing an older working population compared to the highly urbanized California.

Discussion & Conclusion

Our analysis indicates that Vermont, the less urbanized state, has a higher average age among its working-age population compared to California, the most urbanized state. This finding aligns with the expectation that lower urbanization may be associated with older populations, potentially due to lower residents migrating toward urban areas for education and employment opportunities. The difference in age distributions, as shown in the box and whisker plots, highlights how demographic trends vary with urbanization levels and may have implications for workforce availability and economic growth.

These results suggest that states with lower urbanization may face challenges related to an aging workforce, such as labor shortages, reduced productivity, or increased pressure on social services. Conversely, highly urbanized states like California may benefit from a younger and more evenly distributed workforce, which could support sustained economic activity and innovation.

In conclusion, our project demonstrates a clear relationship between urbanization and workforce age, showing that less urbanized areas tend to have older working populations. Understanding these demographic patterns is important for policymakers and planners, as it can inform strategies to attract younger workers, support aging populations, and maintain economic stability across states with varying levels of urbanization.

Outline & Breakdown of Our Project

Our primary project work and data analysis can be found in **Project.Rproj.qmd**

This file contains our **hypothesis testing**, and was where we were able to determine whether to reject or accept the null hypothesis**.** Here, you can finding the following information:

1. CPS Library Pull 2. One-Sided T-Test 3. Box and Whisker Plots 4. Confidence Intervals 5. T-Distribution Plot w/ Critical Value

Break down of each of the following is below:

CPS Library Pull:

What is this?

The data for our project is stored within an R Library. This data set can be found isolated in our "data" folder in data.qmd. To collect the data we needed for this project, we filtered the data set by two states: **CA and VT**. This data set provided us with 377 rows and 17 columns of information.

One-Sided Test:

What is a one-sided Test?

For a one-sided test of an unknown parameter, the null and alternative hypothesis are H0: Theta >= c, & H1: Theta < c, respectively. The hypothesis test of H0 determines whether there is statistical evidence to reject H0.

Why was this appropriate for our research?

Provided that we are seeking to compare two different values—the average age of a workforce eligible population–this is more appropriate than a two-sided test. For a two-sided test, the null hypothesis is instead H0: theta = c, and the alternative hypothesis is H1 != c. This form of testing is not appropriate given our research topic.

How did we conduct the T-Test?

To conduct the t-test we isolated the Vermont and California CPS data and creates a variable "vt_ca". The t-test was ran with the null hypothesis that the difference in means between California and Vermont was less than 0, which would indicate that the mean age of Vermont is greater than that of California. The alternative hypothesis was "greater".

Box and Whisker Plots:

What box and whisker plots did we have?

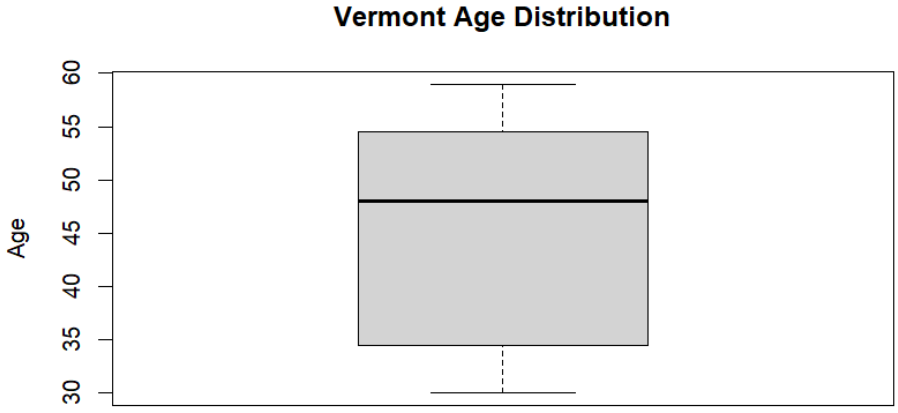
We used two box and whisker plots, one representing the age distribution in California and the other representing the age distribution in Vermont.

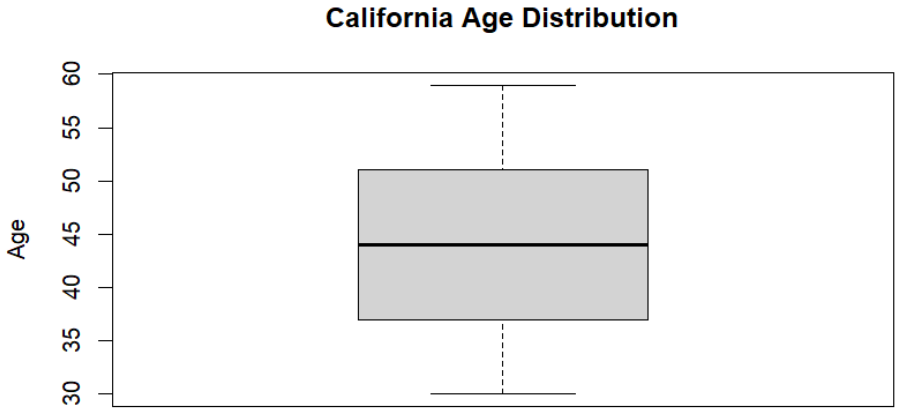
How did we create the box and whisker plots?

We created the box and whisker plots in R by isolating the Vermont and California age data. Each plot only has the data from one state. The y-axis shows the age of residents while the x-axis is unlabeled as each plot only measures one variable.

What did the box & whisker plots tells us?

The box and whisker plots show the overall age distribution of each state. The California plot shows that the mean age is slightly lower than that of Vermont, and the IQR is more compressed indicating a more even distribution. The Vermont plot is more skewed towards older ages, indicating there are more residents around and above the mean age given.





T-Distribution Plot w/ Critical Value:

What does a T-distribution Plot and CV tell us?

A t-distribution plot visually shows the range of possible t-values under the null hypothesis and how likely each value is. The critical value (CV) marks the threshold beyond which we would reject the null hypothesis at a chosen significance level. If the calculated t-statistic falls past the CV, it indicates the observed difference is unlikely due to chance. Together, the plot and CV help visually and quantitatively determine whether to reject the null hypothesis.

How did we show this as a plot?

We displayed our results using a t-distribution plot in R by plotting the probability density function of the t-distribution with 40.22 degrees of freedom. The critical value for our one-sided test at the 0.05 significance level was marked with a red dashed line. The calculated t-statistic (-0.448) was indicated with a solid blue line on the plot. This visual clearly shows that the t-statistic falls far below the critical value, illustrating that we do not have enough evidence to reject the null hypothesis.

What did this plot tell us?

The t-distribution plot shows the t-statistic of -0.448, which reflects that the mean age in California (44.51) is slightly less than that in Vermont (45.31). Because the t-value is negative, it aligns with the null hypothesis that the difference in means (CA-VT) is less than or equal to zero. The high p-value (0.6719) indicates that this difference is not statistically significant, meaning the null hypothesis is more likely to be true. Overall, the plot visually confirms that we do not have enough evidence to support the alternative hypothesis that California's working-age population is older than Vermont's.