

What states do most remote workers live in?

- Analysis:

In order to find where most remote workers live, we pulled the data from the American Community Survey 5-Year Data (2009-2022) in the Census library through Python. First, we located what variables to use by finding the information in the American Community Survey 5-Year Data (2009-2022) website (<https://www.census.gov/data/developers/data-sets/acs-5year.2020.html#list-tab-1806015614>). After we do that, we confirm the years we wanted to take a look at. In this case, we wanted to look at the years 2020, 2021, and 2022 in order to see what changes there are for remote workers living in the United States. After that, we created a dictionary that maps state codes to state names. We retrieve the data by looping through each year in the years' list. We got our data for all states and stored them in a DataFrame.

Next we cleaned our data by converting the 'Percentage of Workers Who Work From Home' column to a numeric format and filling in the missing values, like NaN, in the column with 0. After doing that, we group our data by state and year, then calculate the average percentage of workers who work from home for each group. We reformatted our data by pivoting the data to make year columns (2020, 2021, 2022), with states as rows. We renamed the columns to include the year in the column name. For example, "Percentage of Workers Who Work From Home (2020)". After doing that, we converted the numeric values to percentage format with two decimal places.

Then we combined the data for all years by appending each year's data with the state names and year added to the list. We then combined it into a single DataFrame and cleaned it by renaming the columns, converting the total workers and workers who work from home to integers. Then we calculated the percentage of workers who work from home and created a DataFrame for percentages of workers who work from home.

- Conclusion/Findings:

What we found was that over the years that there has been a slow increase in remote workers living in the United States. There are locations/States where there are more remote workers living in than others. For 2020, we see that the top five locations where most remote workers live are the District of Columbia, Colorado, Oregon, Utah, and Vermont. For 2021, we see that the top five locations where most remote workers live are the District of Columbia, Colorado, Oregon, Massachusetts, and Maryland. Lastly, for 2022, we see that the top five locations where most remote workers live are the District of Columbia, Colorado, Washington, Oregon, and Maryland. On the other hand, we also see that there are locations where there are not many remote workers living there. However, there has been an increase over time with location having more remote workers living there. An example would be Michigan. In 2020, out of the total workers, the work from home workers was only 5.97%. In 2021, it increased to 8.33% and in 2022 it increased to 10.22%.

Have the numbers of female and male workers working from home changed over time?

- Analysis:

We define the variables for male and female workers for how many total workers there are and how many there are working from home through the Census Community Survey 5-Year Data (2009-2022) in the Census library in Python. We then collected the data by looping through the years to collect the total workers and the gender (male and female) that work from home. We then calculated the percentages of workers working from home for each state. After that, we stored the data into a list and created a DataFrame for it. To reorganize the data, we pivoted the DataFrame so the data is organized by year and gender instead. We cleaned the data by formatting them to be numeric values and percentages. We also created pie charts to show what the percentages of female and male workers working from home was like for the years 2020, 2021, and 2022.

To find the linear regression for female workers working from home, we filtered out the relevant columns from all three years. We also removed missing values so that the regression is performed on only states with data in there. We had “x” equal to the years 2020, 2021, and 2023 and “y” equal to the percentages values for female workers. The same process was done for finding the linear regression for male workers working from home. In order to perform the linear regression, we used the `linregress` function from `scipy.stats`. We then plotted the results and printed the r-squared value.

- **Conclusion/Findings:**

Why look at the relationship between the years and genders? There are many reasons but one of them is that it can help us identify whether remote work policies are equally benefiting all genders or not in remote work. This could be a step toward understanding issues like gender inequality in the workplace and whether remote work policies contribute to or reduce these inequalities.

The r-squared value for the relationship between female workers working from home and the years 2020, 2022, and 2023 is 0.3251 (32.5%). This tells us that while there is some relationship between the years and the percentage of female workers working from home, it's not a strong one and there are most likely other variables that are influencing this trend that are not explained in our regression model.

The r-squared value for the relationship between male workers working from home and the years 2020, 2022, and 2023 is 0.2764 (27.6%). This tells us that the relationship between the years and the percentage of male workers working from home is weaker than the relationship between the years and female workers working from home. We have the same conclusion as what we mentioned above for the relationship between the years and female workers working from home.

In conclusion, the numbers of males and females working from home did not change over time. Per looking at the linear regression, we can tell that there are likely more variables that are contributing to this trend and year cannot explain it alone for both genders. The other variables that could be influencing this would be the state policies, the employers' policies, government policies, the economy, and etc.

Q: How does the percentage of remote workers relate to the median household income across different states in 2022?

Analysis:

We pulled data from the census library (ACS 5-Year Estimates) and specifically looked at the variables `B08301_021E` containing work at home data and `B19013_001E` median household income.

*** Websites to data ***

B08301: Means of Transportation ... - Census Bureau Table (Work at home data)

B19013: Median Household Income in ... - Census Bureau Table (Median household income)

*** The US median estimate was 74,755.00 - note to self ***

*** Work from home estimate for 2022 was 24,381,732.00 - note to self ***

Cleaning the data we converted median income and work-at-home data to numeric as some columns contain non-numeric values, those values are converted to NaN which is easier to use. Assign state names to a separate column, Drop the 'NAME' column, then Append the data for the year to the all_data list.

We then created a scatter plot for a visual analysis in order to cross reference these two data sets. The result shows the percentage of the population in each state who completed the census working from home and the median income within those states.

Note: We know that the median income by state is not solely linked to the income of home-based workers. But that doesn't mean that home-based workers do not contribute to the median household income.

Conclusion: Looking at the chart, there is a possibility that the annual gross income is less for remote workers residing in the states with less work-at-home percentage vs a state with a higher work-at-home percentage. And with that we can possibly conclude that states with higher median income may offer more remote jobs vs states with lower median income as they may have more resources and wealth to offer remote positions.