

# Population and poverty relationship in the U.S.\*

Nan An

06 May 2022

## Abstract

Population and poverty people percentage data in the U.S. is pulled from the United States Census Bureau and Census US Population department. It is used to analyze the relationship between population and poverty status by state, within the time from 2017 to 2020. In this report, we will use visualized data to explore the real situation of poverty and population in each state in the U.S. and apply the linear regression analysis to find the relationship equation in the U.S. In the end, we discuss the underlying economic and political problem related to population and poverty.

## 1 Introduction

There is always a question about how high fertility and related demographic variables affect and are affected by poverty. In the 1960s and 1970s, the popular view was that fertility decline would slow population growth in developing countries and thus reduce poverty(Merrick (2022)). Looking into the relationship between population and poverty is very important for the government to make population and economic policies. In the following research, we would choose the U.S. as our example, seeing the relationship between population and poverty status in terms of different states in the U.S.

In the U.S., People and families are considered poor when they lack the economic resources necessary to experience a minimal living standard. Official U.S. Census Bureau statistics estimate that 37 million persons, 11.4 percent of the total population, were poor in the United States in 2020(ASPE (2021)).

This report contains two main sections of data, one focusing on poverty and the other on population in 2017-2022, collected from the United States Census Bureau and Census US Population, after combing them together, we use data visualization to make two maps with population and poverty people percentage by state. Besides, we use ggplot to get the linear regression line and make a box plot to see how strong the relationship is between population and poverty.

In the end, we find there is a slightly positive relationship between population and poverty. It might mean that the adverse effects of rapid population growth on child health, and possibly on education, will likely increase poverty in the next generation(Ahlburg (2022)).

## 2 Data

The data set will be processed and analyzed in R (R Core Team 2020) primarily using the tidyverse (Wickham et al. 2019), dplyr (Wickham et al. 2018) packages and (Wickham et al. 2018) packages . Figures and tables will be created with ggplot2 (Wickham 2016). The packages readxl (Wickham and Bryan 2022), ggpubr (Kassambara 2020), usmap (Lorenzo 2022) and lubridate (Grolemund and Wickham 2011) are used to generate the R markdown report.

---

\*Code and data are available at: <https://github.com/gloria-an-1/Final-paper>.

## 2.1 Data Collection process

The majority of data is collected by the United States Census Bureau, through serveries called Income and Poverty in the United States: 2020(EMILY A. SHRIDER (2020)). The estimates are conducted in terms of income and poverty to evaluate national economic trends and understand their effect on the well-being of households, families, and individuals. Also, in order to see the relationship between poverty and population in the U.S., we collect the population data in the 2019 Census US Population Data By State from Kaggle(COHEN (2019)). This report contains two main sections, one focusing on poverty and the other on population.

### 2.1.1 Data Cleaning

In the data Percentage of People in Poverty by State Using 2- and 3-Year Averages: 2017-2018 and 2019-2020 from the United States Census Bureau, there are State names, Populations in thousands, Margins of error in percentage points, Population as of March of the following year, Information on confidentiality protection, sampling error and nonsampling error included. Here we select only a 3-year average Percentage of the population in poverty by state.

In the data 2019 Census US Population Data By State, there are State names, Population estimates, and Latitude and Longitude data for each state's capital city. We only select State names, Population estimates here. Then, we combine these two datasets together using the loop of states' names, making a new dataset including 4 columns: states' names, the abbreviation of states' names, the population in each state, and the poverty population percentage.

### 2.1.2 Drawbacks

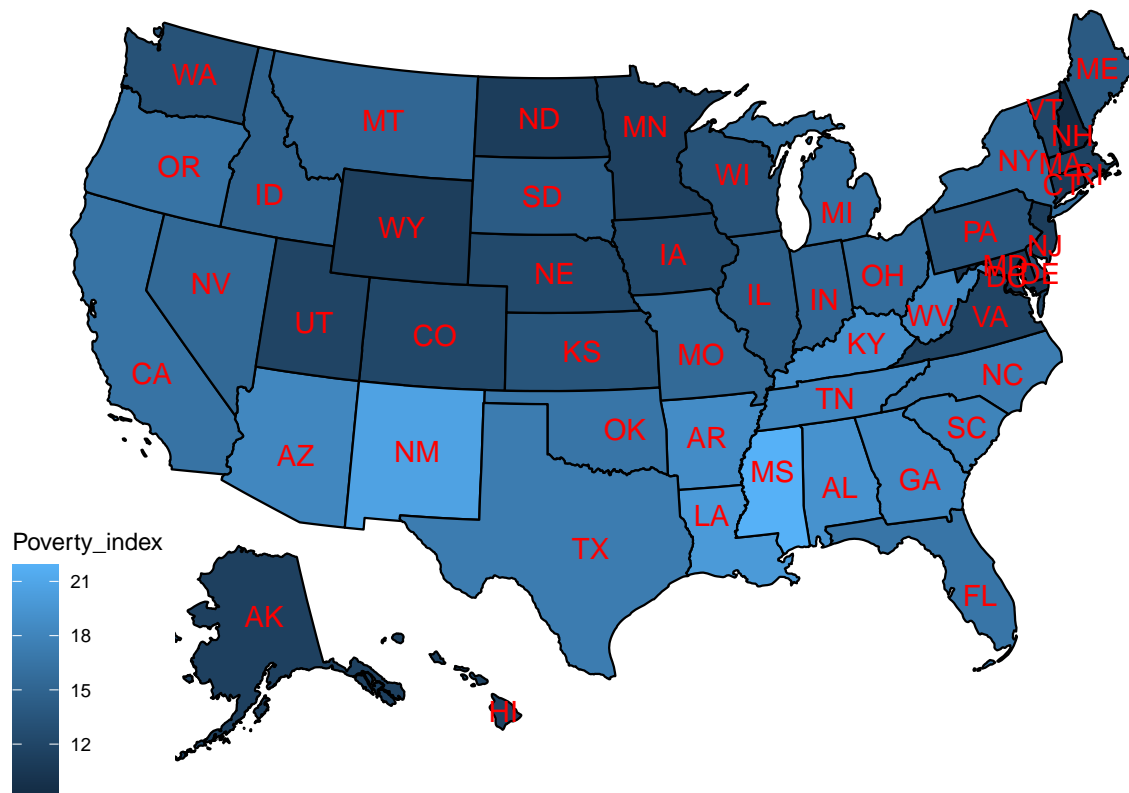
As there is no specific year that has exact data of both population by state and poverty percentage, we use the population data in 2019 and poverty information in 2017-2020, so there might be some discrepancy.

## 3 Data Visualization

### 3.1 Map of population by state in the U.S.

Then we plot a visualized map to see the difference in each province's population in the U.S. The following graph is the map with population values by states in 2014 in the U.S. The states with lighter colours means the ones with more population, for example, CA(California) has the largest population of 38.6 million in 2014. TX(Texas) has the second largest population of 26.96 million in 2014. NY(New York) province also has a large population of 19.65 million in the same year. The "e" means 10 million on the following map.

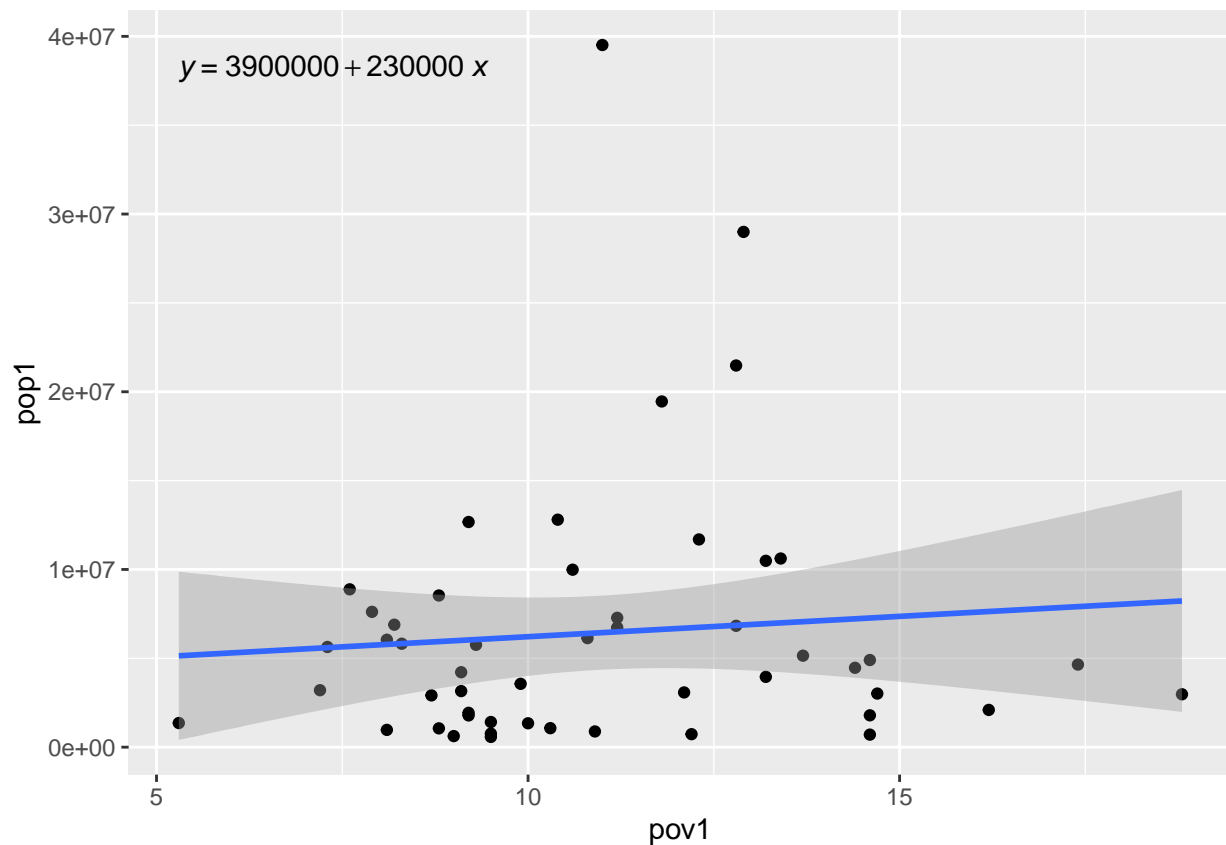




## 4 Model

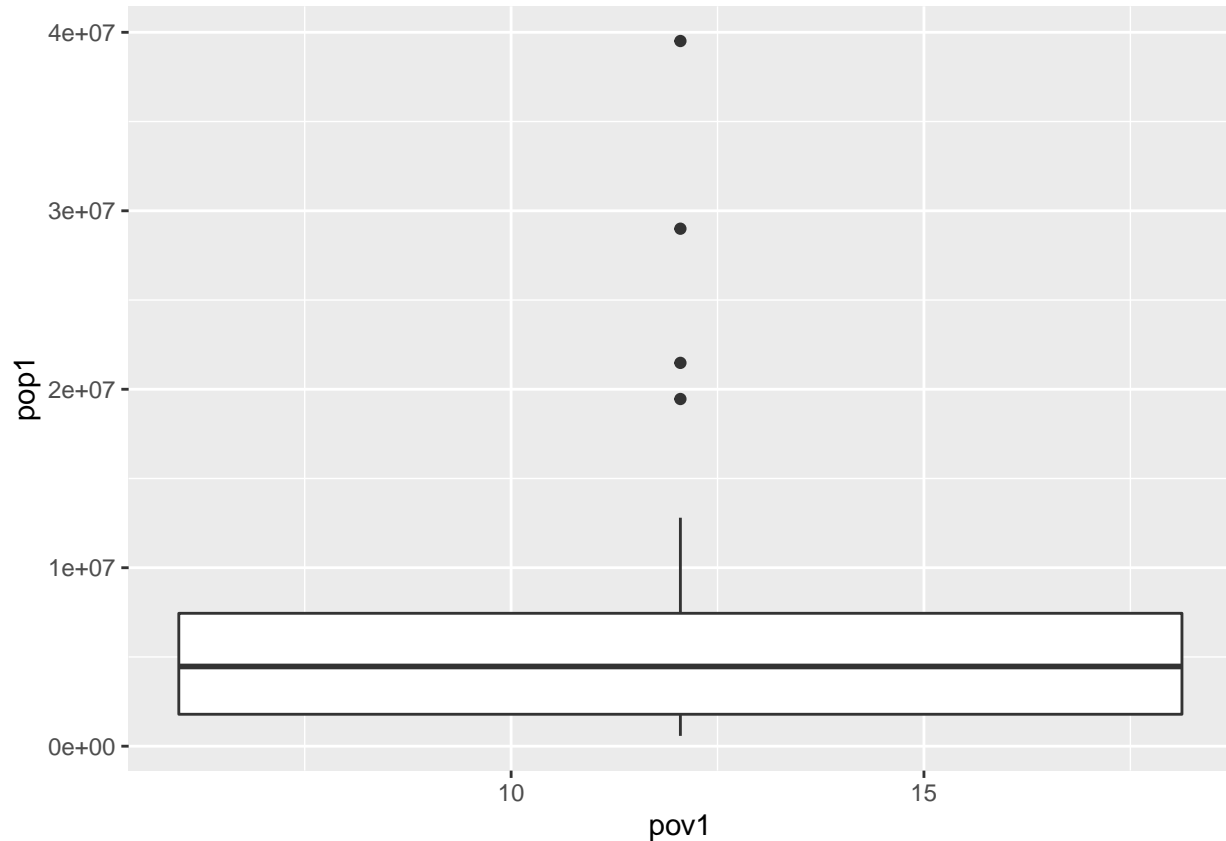
Then we use ggplot to make a model see the relationship between population and poverty situation in each state in the U.S., then use a regression line to make the relationship much clearer. The final equation is  $\text{pop} = 3900,000 + 230,000x$ , we can see there is still a positive relationship technically. Little direct evidence on the impact of population growth exists. Rapid population growth is likely to reduce per capita income growth and well-being, which tends to increase poverty. Also, in densely populated poor nations with pressure on land, rapid population growth increases landlessness and hence the incidence of poverty (Ahlburg (2022)).

```
## `geom_smooth()` using formula 'y ~ x'
```



Then we plot a box plot to see how strong the relationship is, so we can see it is not a strong positive relationship and still have four outliers.

```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```



## 5 Results

In the end, after using linear regression, we find that the final equation of the relationship between population and poverty is  $\text{pop} = 3900,000 + 230,000\text{pov}$ . However, as the variance is big and outliers exist, the relationship is not that strong and clear, and from this research, we cannot tell whether the population is influenced by poverty or is influencing poverty.

Poverty would be both influenced by – and influences – population dynamics, including population growth, age structure, and rural-urban distribution. All of this has a critical impact on a country's development prospects and prospects for raising living standards for the poor. Investments in better health, including reproductive health, are essential for individual security and for reducing mortality and morbidity, which in turn improve a country's productivity and development prospects (UNFPA (2014)).

## 6 Discussion

This chapter presents estimates of the number of people in poverty and identifies factors that are related to poverty such as population.

Actually, a country's economic growth and poverty situation are often shaped by many other factors, such as different age groups, family planning etc.

For example, seeing in terms of overarching demographic trends, the young group does matter. Developing countries with large youth populations and declining fertility rates could see their economies soar, provided they invest heavily in young people's education and health and protect their rights. Potential economic gains could be realized through a 'demographic dividend,' which can occur when a country's working-age population grows larger relative to dependent populations, the report shows (UNFPA (2014)).

Family planning is important an important part of this process because many countries have large youth populations that will almost ensure continued rapid population growth unless fertility declines, which is what offers the possibility of a demographic dividend(UNFPA (2014)).

## References

- Ahlburg, Dennis A. 2022. *Population Growth and Poverty*. [https://link.springer.com/chapter/10.1007/978-3-662-03239-8\\_7#:~:text=First%2C%20rapid%20population%20growth%20is,hence%20the%20incidence%20of%20poverty](https://link.springer.com/chapter/10.1007/978-3-662-03239-8_7#:~:text=First%2C%20rapid%20population%20growth%20is,hence%20the%20incidence%20of%20poverty).
- ASPE. 2021. “Poverty Estimates, Trends, and Analysis.” <https://aspe.hhs.gov/topics/poverty-economic-mobility/poverty-estimates-trends-analysis#:~:text=People%20and%20families%20are%20considered,the%20United%20States%20in%202020>.
- COHEN, PERETZ. 2019. *2019 Census US Population Data by State*. <https://www.kaggle.com/datasets/0b157d370b5b46f60235fee6bfe7d8be3581a5c030fd0cd87c8ba4b32cee4684?resource=download>.
- EMILY A. SHRIDER, FRANCES CHEN, MELISSA KOLLAR. 2020. *Income and Poverty in the United States: 2020*. <https://www.census.gov/library/publications/2021/demo/p60-273.html>.
- Grolemund, Garrett, and Hadley Wickham. 2011. “Dates and Times Made Easy with lubridate.” *Journal of Statistical Software* 40 (3): 1–25. <https://www.jstatsoft.org/v40/i03/>.
- Kassambara, Alboukadel. 2020. *Ggpubr: 'Ggplot2' Based Publication Ready Plots*. <https://CRAN.R-project.org/package=ggpubr>.
- Lorenzo, Paolo Di. 2022. *Usmap: US Maps Including Alaska and Hawaii*. <https://CRAN.R-project.org/package=usmap>.
- Merrick, Thomas W. 2022. *Population and Poverty: New Views on an Old Controversy*. <https://www.gutmacher.org/journals/ipsrh/2002/03/population-and-poverty-new-views-old-controversy>.
- Poverty Rate by State 2022*. 2022. <https://worldpopulationreview.com/state-rankings/poverty-rate-by-state>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- UNFPA. 2014. *Population and Poverty*. <https://www.unfpa.org/resources/population-and-poverty>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, and Jennifer Bryan. 2022. *Readxl: Read Excel Files*.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2018. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.