Exploratory Data Analysis: Hometown Data for Luzerne Country, Pennsylvania

# Objective

Explore three variables for Luzerne Country, Pennsylvania from the U.S. Census Bureau’s American Community Survey (ACS) 5-Year Estimates.

Trying to answer the questions:

* How has the total population changed from 2013 to 2022?
* Has the age distribution shifted (e.g., aging population)?
* Is the population becoming more or less educated?

# Data Access

Geography is Luzerne County, Pennsylvania for the period 2013-22.

FIPS codes: 42 (Pennsylvania) and 079 (Luzerne County)

Accessing data via the get\_acs() method from the tidycensus package with my Census API key.

# Variables

Selected three variables for initial analysis and can add more later.

| Variable | Description | ACS Variable Code |
| --- | --- | --- |
| median\_income | Median household income in the past 12 months (in inflation-adjusted dollars) | B19013\_001 |
| education\_pct | Percent of population age 25+ with a bachelor’s degree or higher | Derived from B15003 (education attainment counts) |
| median\_age | Median age of the total population | B01002\_001 |

Note: education\_pct is calculated manually by dividing the count of people with bachelor’s+ by the total population age 25+ from the detailed education table (B15003).

# Analysis

Load libraries and data.

library("ggplot2")  
library("knitr")  
  
survey\_data <- readRDS("../data/survey\_data.rds")

View data.

kable(survey\_data,  
 col.names = c("Year", "Median Age", "Total Population", "Education Percent"),  
 caption = "Survey Data",  
 format.args = list(big.mark = ","),  
 align = c("l", "r", "r", "r"))

Survey Data

| Year | Median Age | Total Population | Education Percent |
| --- | --- | --- | --- |
| 2,013 | 42.8 | 320,827 | 20.99704 |
| 2,014 | 42.9 | 320,392 | 21.41150 |
| 2,015 | 42.9 | 320,095 | 21.38310 |
| 2,016 | 43.0 | 318,917 | 22.05255 |
| 2,017 | 43.0 | 318,222 | 22.80427 |
| 2,018 | 43.0 | 317,884 | 23.16397 |
| 2,019 | 42.8 | 317,663 | 22.87561 |
| 2,020 | 42.7 | 317,547 | 23.55841 |
| 2,021 | 42.4 | 324,825 | 24.02065 |
| 2,022 | 42.2 | 325,396 | 24.53856 |

Looks like education percent (percent of people with a bachelor’s degree plus) has steadily increased over this ten-year period.

Total population decreased through 2020, then increased in 2021 and 2022.

Median age increased through 2018, then started declining. Median age is at it’s lowest point in 2022.

## COVID Pandemic Impact

Potentially important to note that this period overlaps with the COVID pandemic, which was a major disruption around the world.

While the pandemic started in Wuhan, China in December 2019, the first confirmed US case in Washington state was in January 2020, and the WHO declared it a global pandemic in March 2020. I’ll consider 2020 the first fully pandemic-impacted data year.

Widespread vaccine availability was in place my mid-2021 and by late 2022, emergency response policies and mandates were largely lifted.

For this analysis, I’ll consider 2022 the first mostly post-pandemic data year.

# Plots

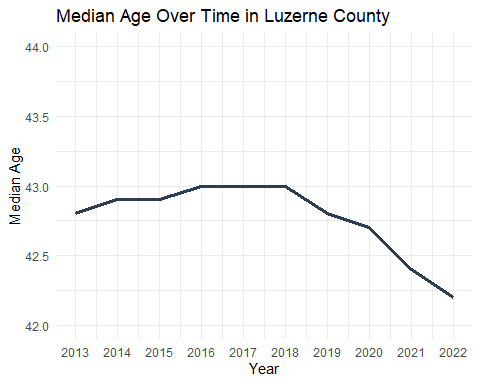
## Median Age

This plot confirms notes above. Overall, median age declined, but peaked in 2016-18.

Possible explanations? Impact of COVID deaths? Primarily impacted older people?

ggplot(survey\_data, aes(x = year, y = median\_age)) +  
 geom\_line(color = "#2C3E50", size = 1.2) +  
 scale\_x\_continuous(breaks = unique(survey\_data$year)) +  
 coord\_cartesian(ylim = c(42, 44)) +   
 labs(title = "Median Age Over Time in Luzerne County",  
 y = "Median Age", x = "Year") +  
 theme\_minimal()

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## ℹ Please use `linewidth` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

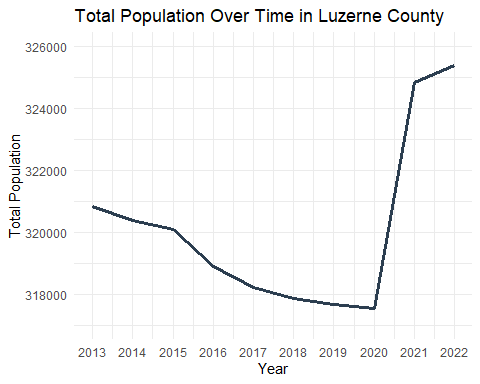


## Total Population

Overall increase, but decline prior to 2021.

Possible explanations?

ggplot(survey\_data, aes(x = year, y = total\_pop)) +  
 geom\_line(color = "#2C3E50", size = 1.2) +  
 scale\_x\_continuous(breaks = unique(survey\_data$year)) +  
 coord\_cartesian(ylim = c(317000, 326000)) +   
 labs(title = "Total Population Over Time in Luzerne County",  
 y = "Total Population", x = "Year") +  
 theme\_minimal()

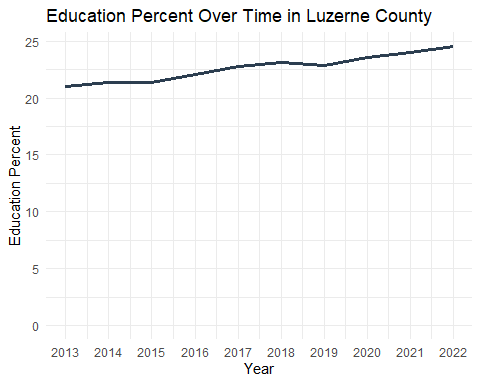


## Education Percent

Steady increase over the period. More people with a college education.

Reflect general pattern? How compare with the rest of the US?

ggplot(survey\_data, aes(x = year, y = education\_pct)) +  
 geom\_line(color = "#2C3E50", size = 1.2) +  
 scale\_x\_continuous(breaks = unique(survey\_data$year)) +  
 scale\_y\_continuous(limits = c(0, NA)) +  
 labs(title = "Education Percent Over Time in Luzerne County",  
 y = "Education Percent", x = "Year") +  
 theme\_minimal()



# Next Steps

How identify causes? Deeper into data?

Trends unique to Luzerne County? Compare to US? Surrouding counties? Pennsylvania?