

Regional Economies Almanac Data Series (READS)

Description

The Regional Economies Almanac Data Series provides estimates of many economic and demographic measures for regions throughout the United States. This document explains how to get started using the READS files.

Authors

Stephan Whitaker, Mitchell Isler, and Brett Huettner

Location

READS data are available on the Federal Reserve Bank of Cleveland website (clevelandfed.org) as a data appendix to the Regional Policy Report entitled “Fourth District Almanac 2026.”

Data Overview

File Inventory

We separate the READS into three files based on each measure’s data frequency:

- 01_census_demographics: Data only in Census years and, later, American Community Survey (ACS) years (annually)
- 02_annual_data: Data that are annual but not limited to Census and ACS years
- 03_monthly_data: Monthly data

File Structure

In the Fourth District Almanac, we use READS data to create line charts showing the value of a measure over time for a given set of places (for example, labor force participation rate in the Cleveland and Cincinnati metro areas over time).

Each row in each READS file is a date-measure observation, where the **var** column indicates the name of the measure. Each region is its own column. The numeric value in a given cell is the estimate of its row measure (**var**) on its row **date** and in its column region.

Percentile Distributions

The percentile columns (**p_10** through **p_90**) provide weighted percentile values of the national distribution for each **var** and on each **date**. They indicate the 10th, 25th, 75th, and 90th percentiles of the national data. For example, where **date** == "2000-01-01" and **var** == "youth_pop_pct", the value in column **p_10** is the 10th percentile value of the youth population in the year 2000, based on a population-weighted distribution of all places in the READS (all combined statistical areas and rural commuting zones (CZs)).

On Fourth District Almanac graphs for which percentile distribution values are available, we shade the region from **p_10** to **p_90** with low opacity and the region from **p_25** to **p_75** with slightly higher opacity. The percentile values displayed in the Fourth District Almanac differ slightly from the percentiles in the READS because the Fourth District rural CZs are combined into two rural regions (Plains and Appalachia) before the Almanac percentiles are calculated.

We have omitted the percentile distribution values for measures that are themselves percentiles.

Geographies

The READS places each county and county equivalent in the United States into either a combined statistical area (CSA) or a rural commuting zone, and we calculate percentile distribution values based on those regions. We also provide estimates for four geographic areas that are not CSAs or CZs: the United States as a whole, the Fourth Federal Reserve District, and two rural regions within the Fourth District. “Rural, Plains” and “Rural, Appalachia” are super-regions created for ease of interpretation and visualization in the Fourth District Almanac. We include these regions in the READS files for reproducibility.

We use Census delineation files to assign counties to CSAs. These files can be found here: <https://www.census.gov/geographies/reference-files/time-series/demo/metro-micro/delineation-files.html>

For the 2025 READS files, we used the 2000 Commuting Zone definitions, available here: <https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas>

We include the file `region_crosswalk.csv` with the READS files to allow users to look up which region contains a given county. Note that rural counties in the Fourth District are placed in their conglomerate regions and in their CSA or CZ.

Codebook

The file `00_codebook.csv` contains details about each of the measures in the READS data sets. For each measure, we provide the following information:

- `web_file`: which READS file contains the measure
- `title`: detailed measure description
- `units`: unit of measurement including any scaling of values
- `source`: original data source
- `note`: any methodological notes that change how the data are interpreted
- `label_format`: our suggestion of how to format a measure’s values (for convenience)

Example Usage

To further illustrate how to use the data set, here are some basic steps for plotting one of the measures as a time series line graph, as we do in the Fourth District Almanac.

1. Open or import the data set with the measure of interest.
2. Filter for the measure of interest and select the desired region(s).
3. Plot the filtered data set with `date` on the x-axis and the region’s value on the y-axis.
4. Add titles, labels, and captions. Format as desired.

Below is a code snippet illustrating this process using base R:

```
# Read in the file.
df = read.csv(file = '01_census_demographics.csv', check.names = F)

# Filter for the measure and region of interest.
# Here, the share of the population in the Cleveland CSA born in the state in which they live.
chartdata = df[which(df$var == 'native_pop_pct'),c('date', 'Cleveland-Akron-Canton, OH')]

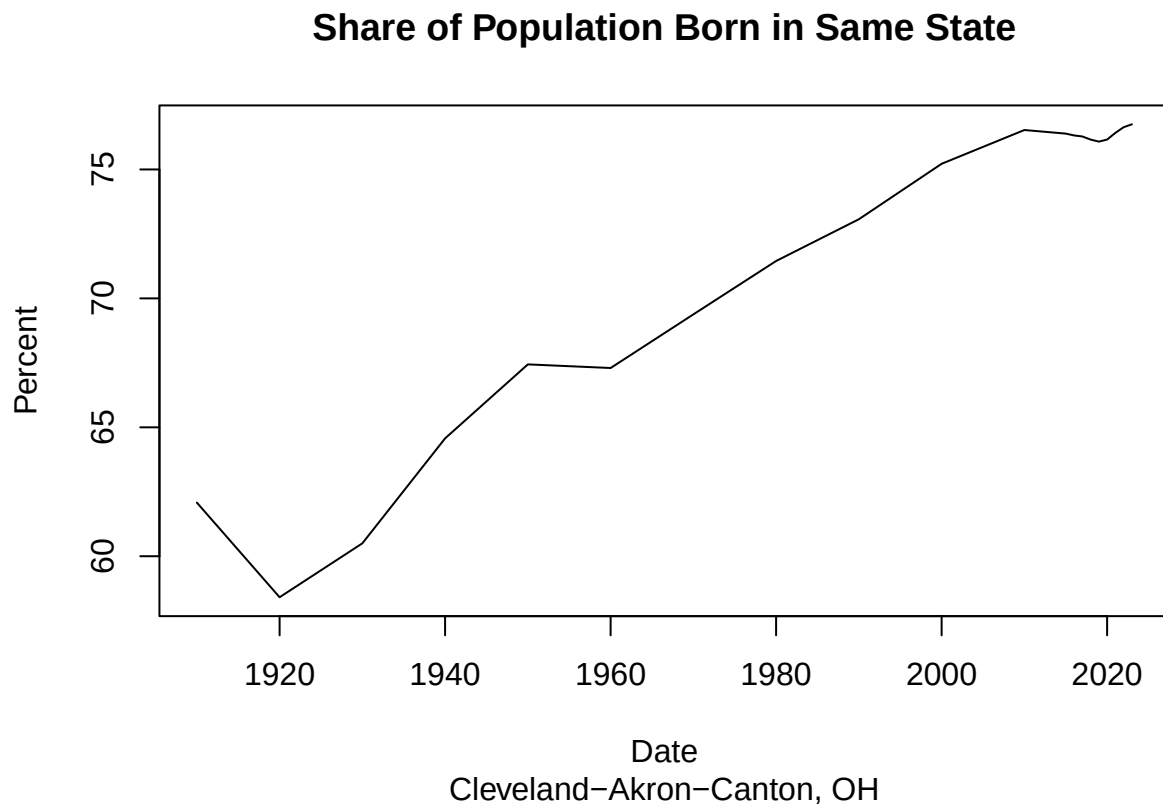
# Specify variable type for the date column.
chartdata$date = as.Date(chartdata$date)
```

```

# Define plot labels. Refer to the codebook for details.
plot_title = 'Share of Population Born in Same State'
plot_subtitle = names(chartdata)[2]
plot_yaxis = 'Percent'
plot_xaxis = 'Date'

# Create a basic plot.
plot(x = chartdata$date, y = chartdata$`Cleveland-Akron-Canton, OH`, type = 'l',
     main = plot_title, sub = plot_subtitle, ylab = plot_yaxis, xlab = plot_xaxis)

```



Data Notes

IPUMS

In the measures we estimated from IPUMS data, we used the following sample files:

- 1910 to 1950: 1% sample
- 1960: 5% sample
- 1970: 1% metro sample, Form 1
- 1980 to 2000: 5% state sample
- 2010 and later: ACS 1-year microdata

As noted in the codebook, we use linear interpolation to replace values from the 2020 ACS because the Census Bureau is uncertain of the data quality that year. We also apply a 5-year moving average to IPUMS data beginning in 2015.

CoreLogic

The house price indexes (HPIs) displayed in the Fourth District Almanac are calculated using proprietary CoreLogic data. The terms of the CoreLogic user agreement do not permit making the HPIs available for download.

Last Updated

This file was last updated in January 2026 by M. Isler.