Title ✓

Should remind reader of hypothesis

A-

Daniel Malone
Elliane Hall
Md Sayemul Islam

Data Description

I wonder if the American Community survey or a different dataset will get you more years of data ✓ keep digging

Data from the United States Census Bureau is used to obtain information about population estimates and estimated migration rates by county. Estimates are based on collected Census Data from 2000, 2010, and 2020. The dataset encompasses a comprehensive compilation of data pertaining to 3,143 counties, covering years 2000 to 2022. Variables include data on population, migration, births, and deaths. Within this extensive dataset, individual observations were identified and documented for each county in each year, totalling 72,289 observations. ~~The dataset is cleaned and presented at a county-year level.~~

county-year

Employment data is sourced from the United States Department of Agriculture, and covers the time period 2000 to 2022. The dataset includes information about the number of people employed and unemployed, the unemployment rate, and the number of people in the labor force, all at a county level. The data has information on 3,143 counties and 23 years, thus making the total number of observations 72,289. ~~The dataset is cleaned and presented at a county-year level.~~

county-year

Precipitation data is sourced from the National Oceanic and Atmospheric Association, specifically the National Centers Environmental Information. The NOAA site hosts an archive of climate data gathered from 130 observing platforms. The dataset used contains precipitation data by each United States county for each year between 1895 and 2022. The data is restricted to the continental United States, so entries for Hawaii, Alaska, and the various overseas American territories were dropped. The difference between precipitation in the year of observation and the year before are calculated. Lastly, the years are restricted to the time period between 2000 and 2022, thus making 72,289 total number of observations.

All datasets were merged together by FIPS code and year, to create a unified dataset.

I wonder whether diff. from, say 3 yrs prior might be better

Our hypothesis tests the effect of droughts on population migration on a county level. Thus, the difference in precipitation between the current year and the prior year will tell us if there is consistently low precipitation in a county. After conducting some more analysis on the difference in precipitation, we will determine how many years of decreasing population will be defined as a drought. Population migration is defined as the total number of people that migrated in/ out of a county as a percent of the population in the county for each year. The data we chose is valid for testing our hypothesis because it tells us how many people are migrating in and out of a county compared to the total county population, as well as how precipitation is changing within a county over the set of years. This will allow us to derive estimates on the effect of low precipitation (droughts) on changes in population from migration per county.

The first shortcoming of our data is that it may not account for all potential confounding variables that could influence demographic changes. Omitted variables can lead to spurious correlations or incomplete explanations. Moving forward, we may want to include data on agricultural output and overall county GDP, since agricultural jobs would likely influence migration to counties with high agricultural output, and high agricultural output is likely correlated with high precipitation. Further, we assume that there will be higher migration rates to counties with higher overall county GDP. We plan to use state or county fixed effects, as well as year fixed effects, in our regressions. However, we do not think that county GDP and agricultural output would be controlled with these fixed effects.

The precipitation data we collected spans years 1895 through 2022 and the unemployment and migration data we collected spans 2000 to 2022. Due to the confinement in our unemployment and migration data, we may not capture long-term trends or the impact of events that occurred outside of the time frame. It would be ideal if we could have data for migration and precipitation spanning 1895 to 2022. Furthermore, we chose precipitation data instead of data on droughts to define whether a county is in drought. There are many different ways to define a drought, so we felt it was best to define it ourselves to create consistency. While we have not yet found the best way to define it, we will explore the precipitation data and read more literature on similar topics to find the best definition.

Lastly, precipitation patterns may not have an immediate effect on migration. There could be a time lag between changes in precipitation and changes in migration that needs to be considered. Furthermore, since our data includes information up to 2022, it's possible that our analysis may not fully capture the long-term effects of recent events like the COVID-19 pandemic, which might have influenced migration patterns. Thus, we could potentially face structural break problems in our data due to COVID-19 or other significant events.

Usually describe sample size for the merged data, so 1895 - 2000 data for precip. is irrelevant.

Variable List

Source: U.S. Census Bureau

VARIABLE	DESCRIPTION
fips	FIPS code for county
year	Year of observation
pop_estimates	Census estimates of the population in county
pop_change	Population change from prior year to current in county
births	Number of births in county during year of observation
deaths	Number of deaths in county during year of observation
natural_inc	Natural increase in population from prior year to current in county
int_migration	Net international migration in county during year of observation
dom_migration	Net domestic migration in county during year of observation
net_migration	Net migration in county during year of observation
residual	Residual population not explained by demographic component
gq_estimate	Population of people in county living in group quarters
birth_rate	Birth rate in county during year of observation
death_rate	Death rate in county during year of observation
natural_inc_rate	Natural increase rate in county during year of observation
int_migration_rate	International migration rate in county during year of observation
dom_migration_rate	Domestic migration rate in county during year of observation
net_migration_rate	Net migration rate in county during year of observation
fips_year	Interactive term with FIPS code and year variable
migration_pop	Migration as percent of population in county in year of observation

Source: U.S. Department of Agriculture

VARIABLE	DESCRIPTION
fips	FIPS code for county
year	Year of observation
employed	Number employed people in county during year of observation
unemployed	Number unemployed people in county during year of observation
unemploy_rate	Unemployment rate in county during year of observation
labor_force	Number people in labor force in county during year of observation
fips_year	Interactive term with FIPS code and year variable

Source: National Oceanic and Atmospheric Administration

VARIABLE	DESCRIPTION
fips	FIPS code for county
year	Year of observation
precip	Total precipitation for county during year of observation
fips_year	Interactive term with FIPS code and year variable
precip_diff	Difference in current year and previous year precipitation

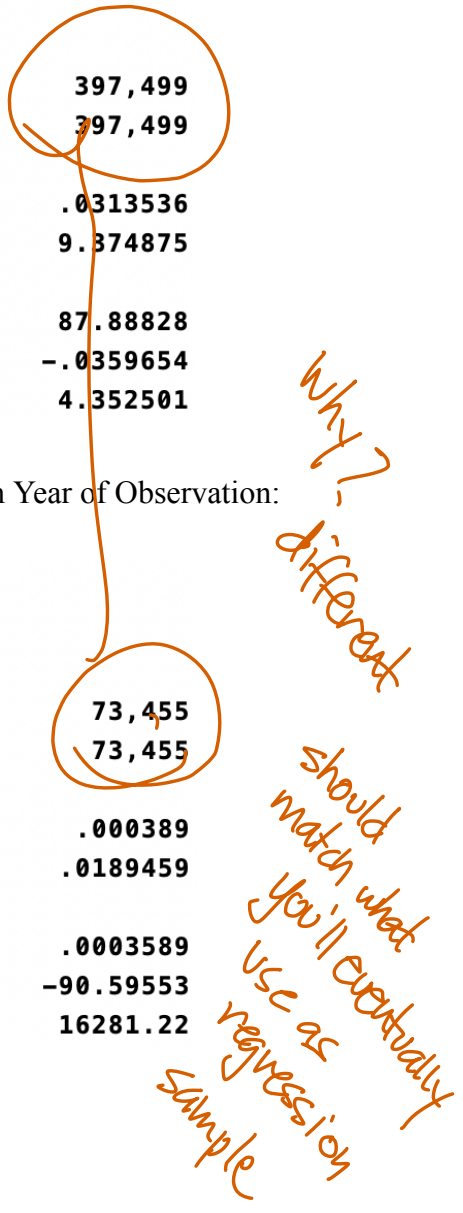
Descriptive Statistics

Descriptive Statistics for Difference in Current Year and Previous Year Precipitation**:**

	Percentiles	Smallest		
1%	-24.05	-70.55		
5%	-15.35	-68.56		
10%	-11.26	-62.01	Obs	397,499
25%	-5.46	-61.77	Sum of wgt.	397,499
50%	.03		Mean	.0313536
		Largest	Std. dev.	9.374875
75%	5.54	72.83		
90%	11.41	76.25	Variance	87.88828
95%	15.4	77.25	Skewness	-.0359654
99%	23.77	83.91	Kurtosis	4.352501

Descriptive Statistics for Migration as Percent of Population in County in Year of Observation:

	Percentiles	Smallest		
1%	-.0336257	-3.510246		
5%	-.0174546	-1.189678		
10%	-.011919	-.5345401	Obs	73,455
25%	-.0050484	-.425	Sum of wgt.	73,455
50%	0		Mean	.000389
		Largest	Std. dev.	.0189459
75%	.0055639	.1708185		
90%	.0136286	.1886489	Variance	.0003589
95%	.0205547	.1886792	Skewness	-90.59553
99%	.0378583	.2436975	Kurtosis	16281.22



Scatter Plot Between Precipitation and Migration