Is air pollution correlated to inter-state migration in the US?

EDA Final Project

Leonardo Rueda[1,2]   Theo Napitupulu[1,3]

[1] Sanford School of Public Policy [2] jr466@duke.edu [3] yosia.napitupulu@duke.edu

Abstract

This is my dream abstract. **Keywords: Air Pollution, Migration.**

Contents

List of Tables

List of Figures

Rationale and Research Questions

This project looks at the effects of air pollution on inter-state migration in the United States using the Air Quality Index datasets from EPA and the Population Migration data from the IRS for the period 2013-2020.

Evidence from middle-income countries shows that air pollution has negative impacts on several health and economic outcomes, such as mortality rates, health expenditures, mental health, hours worked, labor productivity and income. Additionally, other studies have shown how migration decisions are affected by air pollution. For instance, Chen, S., Oliva, P., & Zhang, P. (2022) found that a 10 percent increase in air pollution, holding everything else constant, reduces population through net outmigration by about 2.8 percent in a given county in China (see Chen, S., Oliva, P., & Zhang, P. (2022). The effect of air pollution on migration: Evidence from China. Journal of Development Economics, 156, 102833. https://doi.org/10.1016/j.jdeveco.2022.102833)

Our hypothesis is that there is a positive relationship between high air pollution and migration outflows, that is, the highest the pollution registered by the Air Quality Index (AQI) in a state in a given year, the higher the number of people leaving that state the same year.

Dataset Information

1. Air Quality Datasets

The Air Quality Index dataset provides annual information per county about the maximum values reached by the AQI, the number of days in which this index reached values considered unhealthy, and the number of days with PM2.5 particles recorded. The information is capture by the EPA local air quality stations. These datasets are available at https://aqs.epa.gov/aqsweb/airdata/download_files.html. We create a dataset with the information from 2013 to 2020 and aggregate the information at the state level. Likewise, we calculate the state averages for the variables Unhealthy.Days, Max.AQI, and Days.PM2.5.

2. Inter-state migration datasets

The State-to-State outflows dataset provides annual information at the State level about the number of people whose reported home address changed in their individual income tax returns from one year to the other. These datasets are available at https://www.irs.gov/statistics/soi-tax-stats-migration-data. The Inter-state migration datasets do not include the variable "Year", so we create it from 2013 to 2020.

According to the dictionary for this dataset, the variable "y2_statefips" and the code "96" refer to the total outflows of migrants for each state in a given year. Therefore, we filtered the previous dataset by that value. Additionally, we change this variable's name to "FIPS_Code", so later we can merge this dataset with the AQI dataset.

3. Scraping the FIPS codes

The AQI Dataset has the names of each State in the US, but it does not have the code, which is the variable we need to merge this data with the migration one. We scrape the FIPS codes from the webpage (https://www.bls.gov/respondents/mwr/electronic-data-interchange/appendix-d-usps-state-abbreviations-and-fips-codes.htm) and we create a data frame. Then, we merge this dataset with the AQI dataset by the variable "State".

4. Merging the AQI and Migration datasets

We merge and arrange the AQI and the migration datasets by the variables "FIPS_Code" and "Year". The resulting dataset has information from 2013 to 2020. According to the dictionary for the migration dataset, the variable "n2" refers to the number of individuals who migrated to other states. To facilitate the interpretation, we changed the name of the variable to "Migrants.outflows". We create a subset of the previous dataset with the variables of interest: FIPS_Code, Year, State, Avg.Unhealthy.DAys, Avg.Days.AQI, Avg.Days.PM2.5, and Migrants.outflows. This is the dataset that we will use in our analysis.

The final dataset has 7 variables: FIPS Code, Year, State, Avg.Unhealthy.Days, Avg.Max.AQI, Avg.Days.PM2.5, and number of migrants. In the table 1 we can see the characteristics of the dataset for years 2013 and 2020. Each year has information for the 50 states of the USA (we do not include in the analysis the Distric of Columbia, Puerto Rico and the Virgin islands). For 2013, on average each state showed 0.56 days with air quality classified as "unhealthy", the maximum value reach by the AQI was 114.89, the number of days with PM2.5 registered were 116.57, and 135,657 people migrated to a different state. For 2020, on average each state showed 0.90 days with air quality classified as "unhealthy", the maximum value reach by the AQI was 122.81, the number of days with PM2.5 registered were 121.52, and about 138,954 people migrated to a different state.

Table 1: Summary statistics for the final dataset 2013-2020

Variable	Type	Obs	Min	Median	Media	Max
Year=2013						
FIPS Code	int	50				
State	chr	50				
Avg.Unhealthy.Days	num	50	0.00000	0.03348	0.56471	6.13333
Avg.Max.AQI	num	50	73.00	107.22	114.89	240.85
Avg.Days.PM2.5	num	50	17.43	110.33	116.57	309.00
Migrants.outflows	num	50	16540	95343	135657	532619
Year=2020						
FIPS Code	int	50				
State	chr	50				
Avg.Unhealthy.Days	num	50	0.00000	0.09762	0.90793	14.20755
Avg.Max.AQI	num	50	58.75	98.25	122.81	430.35
Avg.Days.PM2.5	num	50	16.11	105.87	121.52	304.39
Migrants.outflows	num	50	16378	92877	138954	684935

Exploratory Analysis

Air Quality Indicators

In the figure 1 we can see some descriptive statistics for the air quality indicators between 2013 and 2020. At the top left, the average unhealthy days registered in all states and for all years show no days with unhealthy air quality. However, between 2013 and 2016 some states registered on average 2 days, and between 2018 and 2020 this registries dropped to 1 day. At the top right, the average maximum AQI registered for all states reached numbers close to 100, which means that the air quality is acceptable. At the same time, in some years this indicator reached numbers above between 150 and 200, meaning that the air quality was poor and people could experience negative health effects. Finally, at the bottom left the average number of days when PM2.5 were detected reached 100 during the period of time analyzed. This result is worrisome because a long exposure to these particles is harmful to human health. This indicator shows a high dispersion, with some states showing less than 50 days, while others show more than 150 days. It is important to mention that the analysis of averages by state per year allows to have an initial picture of the air quality in each year, but does not allow to identify the air quality by state. In the following analysis, some maps allows to see this.

It is important to mention that the state averages do not allow to understand what is the distribution per county. It could be the case that air quality registries in one county over influences the average for the entire state, or at the other extreme, it could be the case that the air quality is equally distributed in each state.

Migration outflows by State

Air Quality and migration outflows

Analysis

6. Correlations

Table: Correlation between air quality indicators and migration outflows 2013-2020 | Confidence Interval | | 95% | |Correlations | Migrants Outflows | t | p-value | Min | Max | |—————- |——————|———|———————|————|————| |Avg. unhealthy days | 0.2921202 | 6.0936 | 2.609e-09 | 0.1997911 | 0.3793075 | |Avg. maximum AQI | 0.2586106 | 5.341 | 1.558e-07 | 0.1647361 | 0.3478420 | |Avg. days PM2.5 | -0.02109288 | -0.4209 | 0.6741 | -0.118899 | 0.07711834|
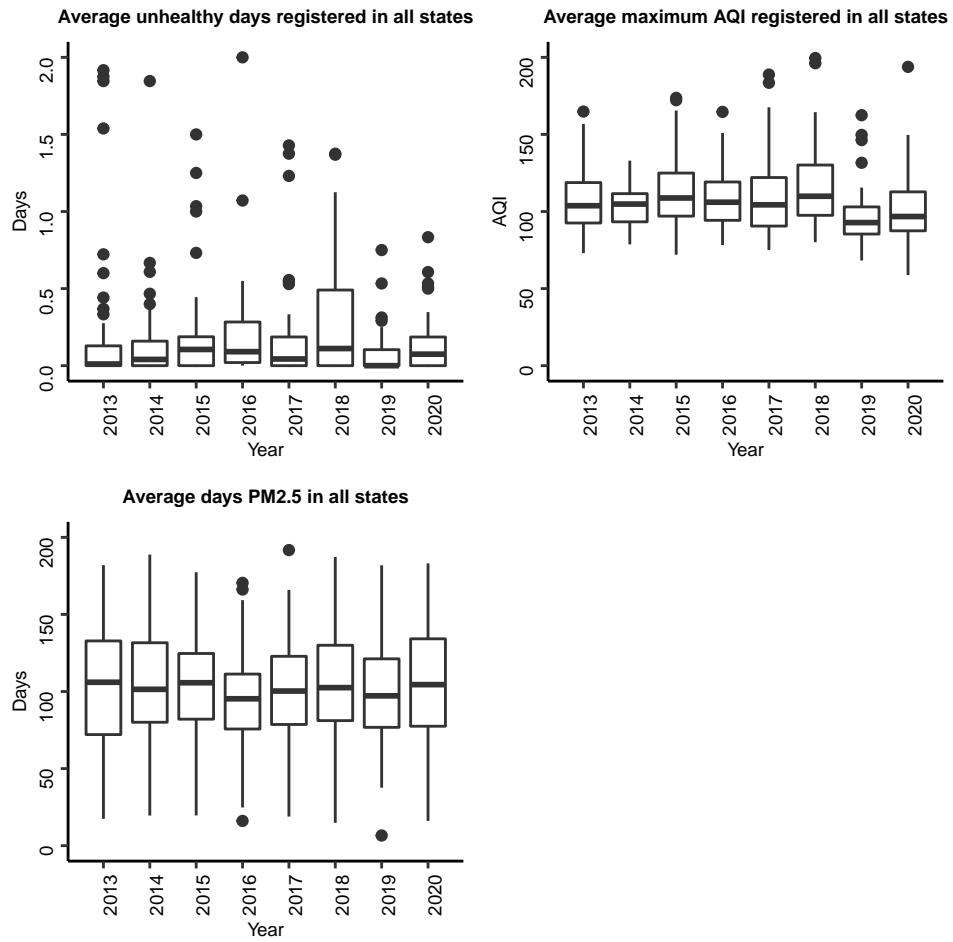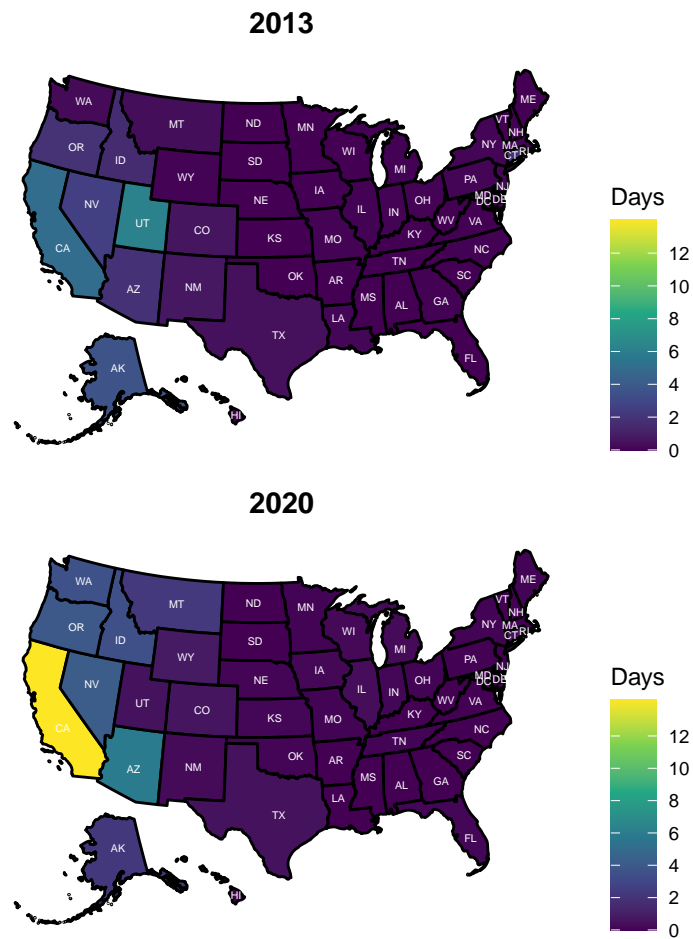
Figure 1: Air Quality Indicators 2013-2020

2013

2020

Figure 2: Average Unhealthy Days by State 2013 and 2020
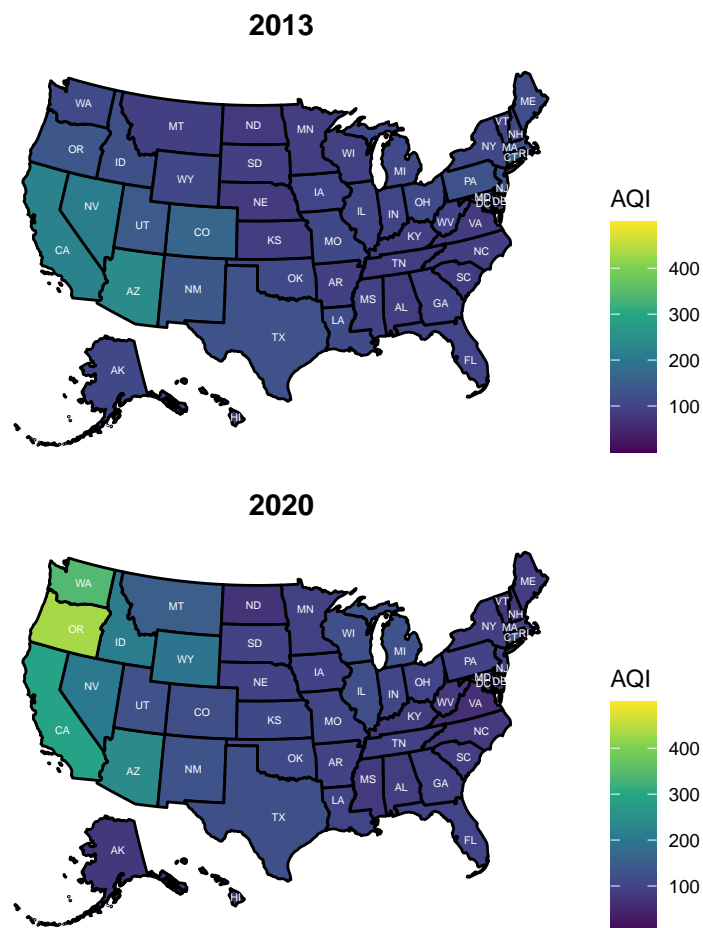
2013

2020

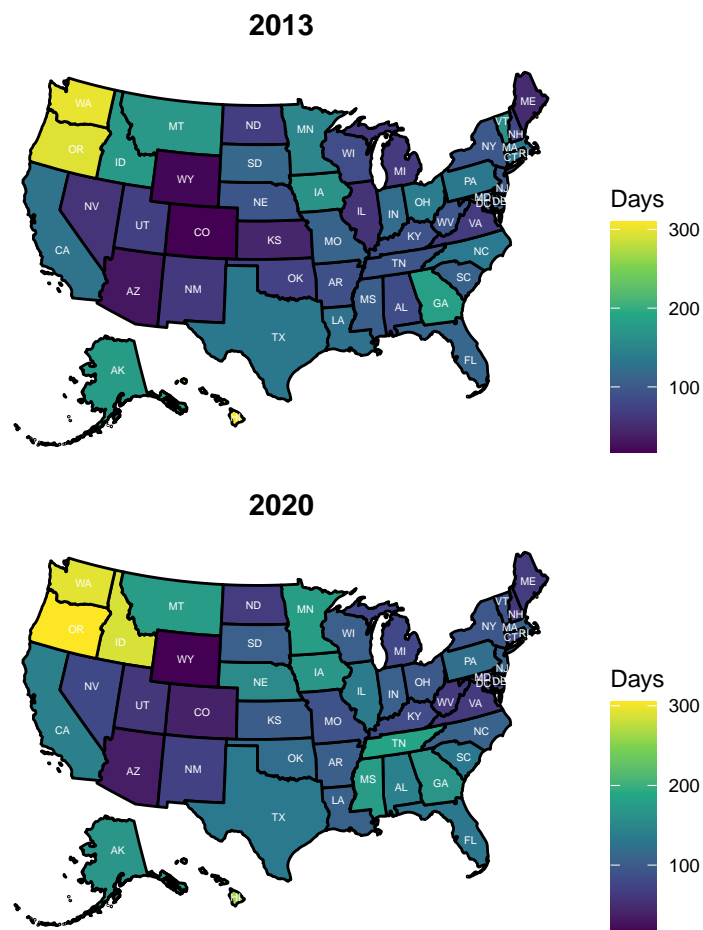Figure 3: Average Max AQI value by State 2013 and 2020

2013

2020

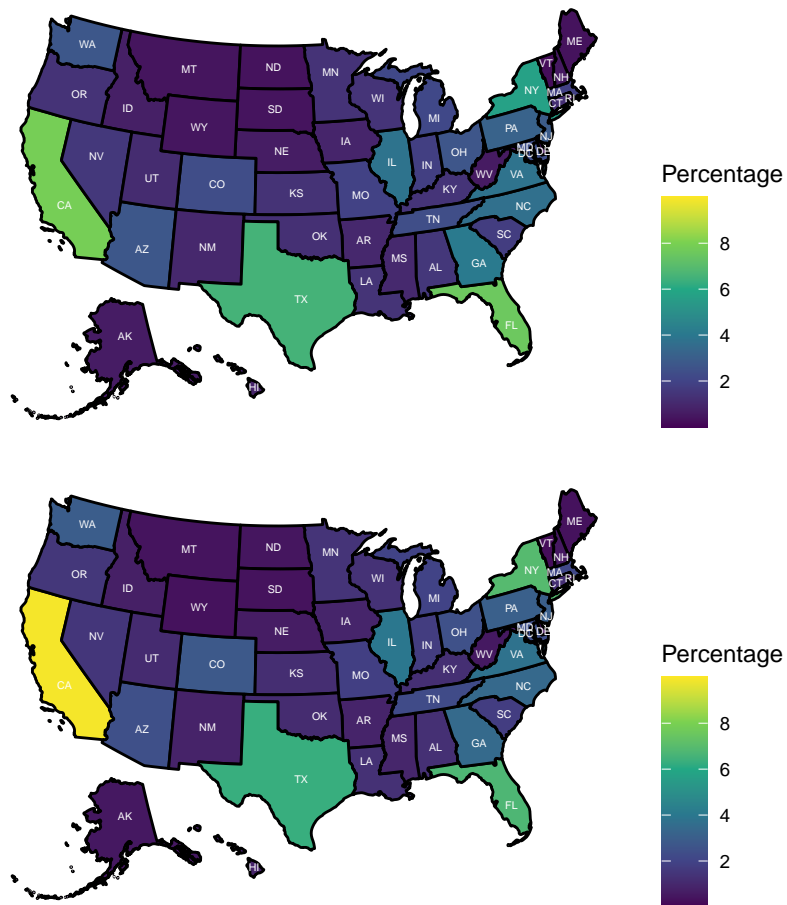Figure 4: Average PM2.5 by State 2013 and 2020

Figure 5: Percentage of migrants' outflows by State in 2013 and 2020
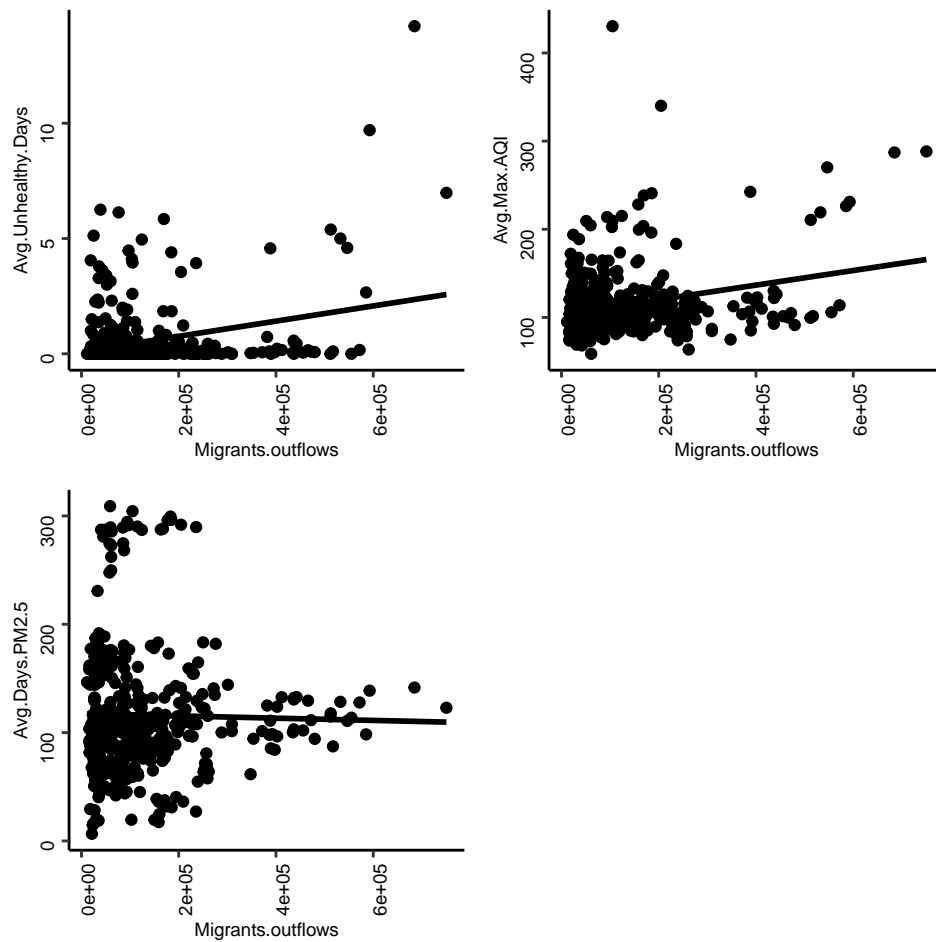
Figure 6: Relationship between air quality indicators and migrantion outflows between 2013 and 2020