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 (available at https://aqs.epa.gov/aqsweb/airdata/download_files.html) and the Population Migration data from the IRS (available at https://www.irs.gov/statistics/soi-tax-stats-migration-data) for the period 2012-2020.

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.

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 Air Quality Index dataset provides annual information at the State level about the number of people whose reported home address changed in their individual income tax returns.

Dataset Information

1. Air Quality Datasets

- The raw AQI datasets are available by year. Therefore, we create a dataset with the information from 2010 to 2020.
- This dataset is at the county level. We aggregate the information at the state level. Likewise, we calculate the state averages for the variables Unhealthy.Days, Max.AQI, and Days.PM2.5.

- We save the new dataset in the data/processed folder.

2. Inter-state migration datasets

- The Inter-state migration datasets do not include the variable "Year", so we create it. Then, we save each dataset in the folder Data/Raw.
- We create a dataset with information from 2012 to 2020.
- According to the dictionary for this dataset, the variable "y2_statefips" and the code "96" refers to the total outflows of migrants for each state in a given year. Therefore, we filtered the previous dataset by that value.
- We change this variable's name to "FIPS_Code", so later we can merge this dataset with the AQI dataset.
- We save the dataset with migration outflows per state for years 2010-2020 in the folder Data/Processed.

3. Scraping the FIPS codes

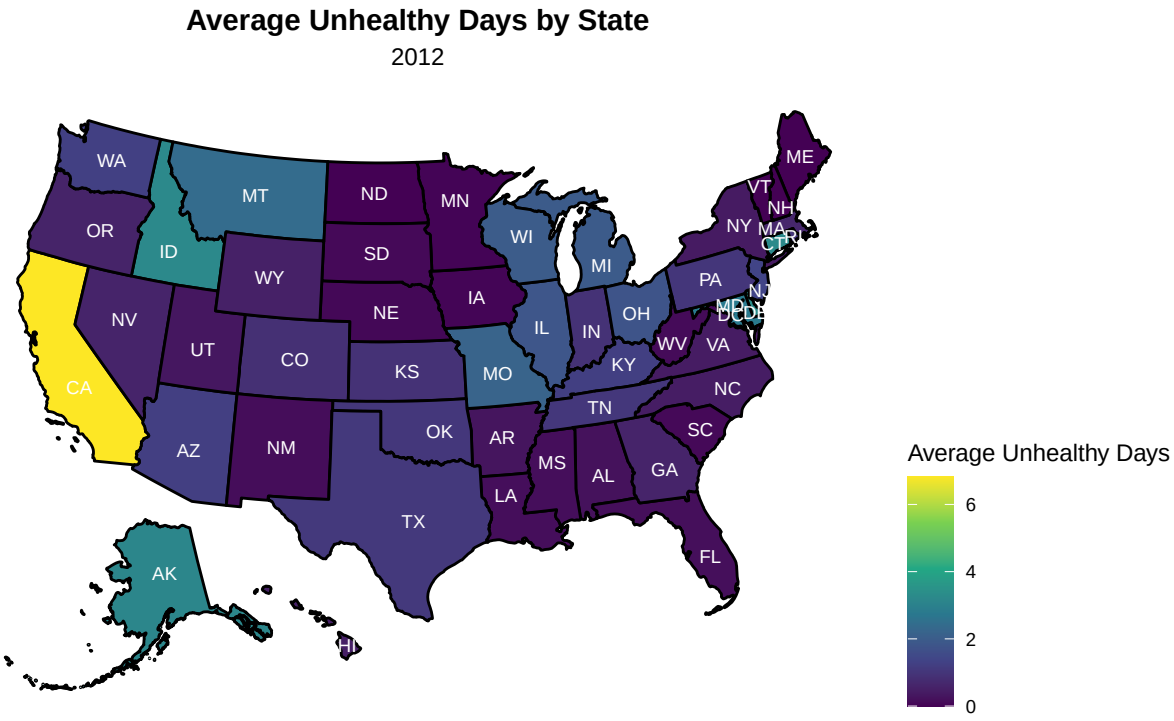
- The AQI Dataset has the names of each State in the US, but it does not have the code, which is the variable we need to merge this data with the migration one.
- In the object "the_website" we store the website direction where the FIPS codes are available (https://www.bls.gov/respondents/mwr/electronic-data-interchange/appendix-d-usps-state-abbreviations-and-fips-codes.htm).
- We scrape from the website the information for the states and codes, and we create a data frame.
- We merge this dataset with the AQI dataset by the variable "State".
- We save the dataset in the folder Data/Processed.

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 2012 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.
- We save the dataset in the folder Data/Processed. This is the dataset that we will use in our analysis.

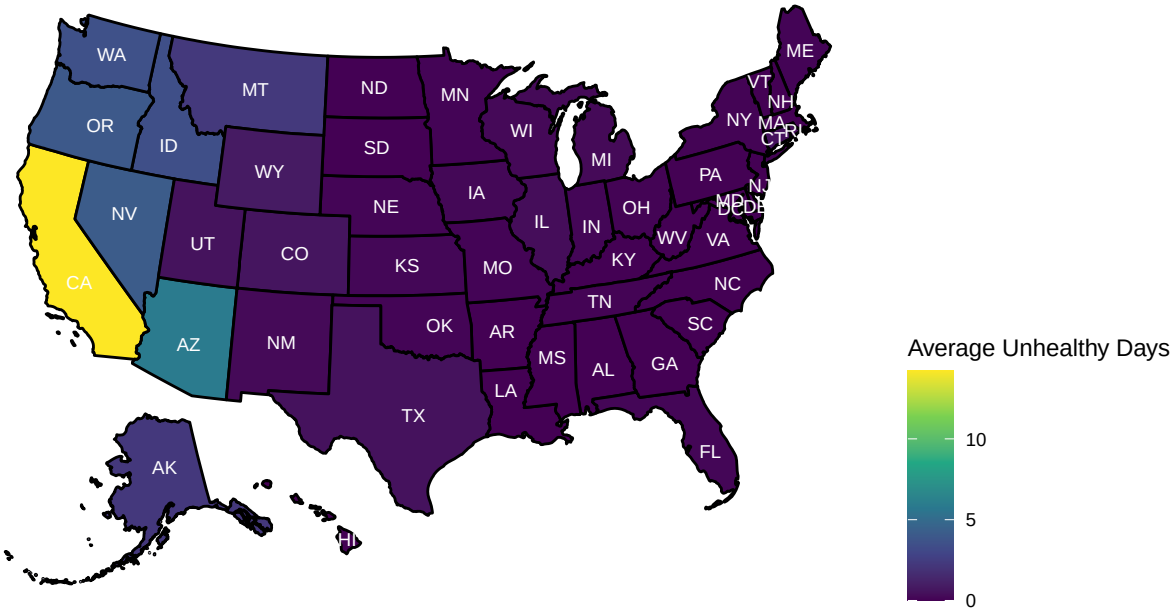
Exploratory Analysis

5.Maps AQI by State 2012 and 2020



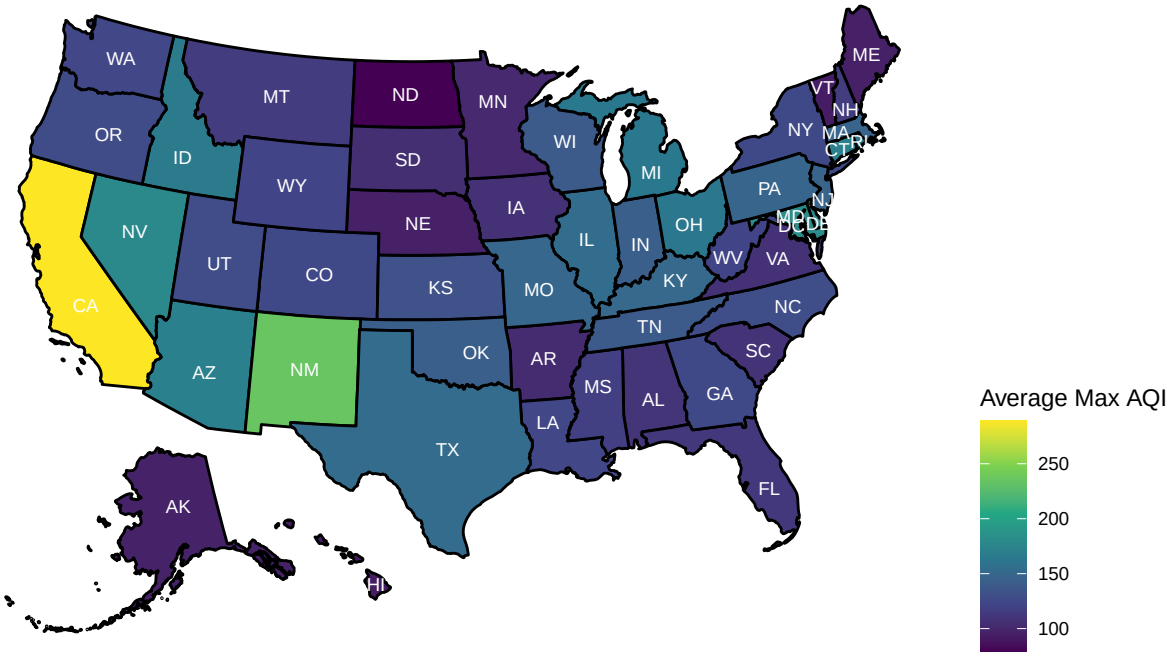
Average Unhealthy Days by State

2020



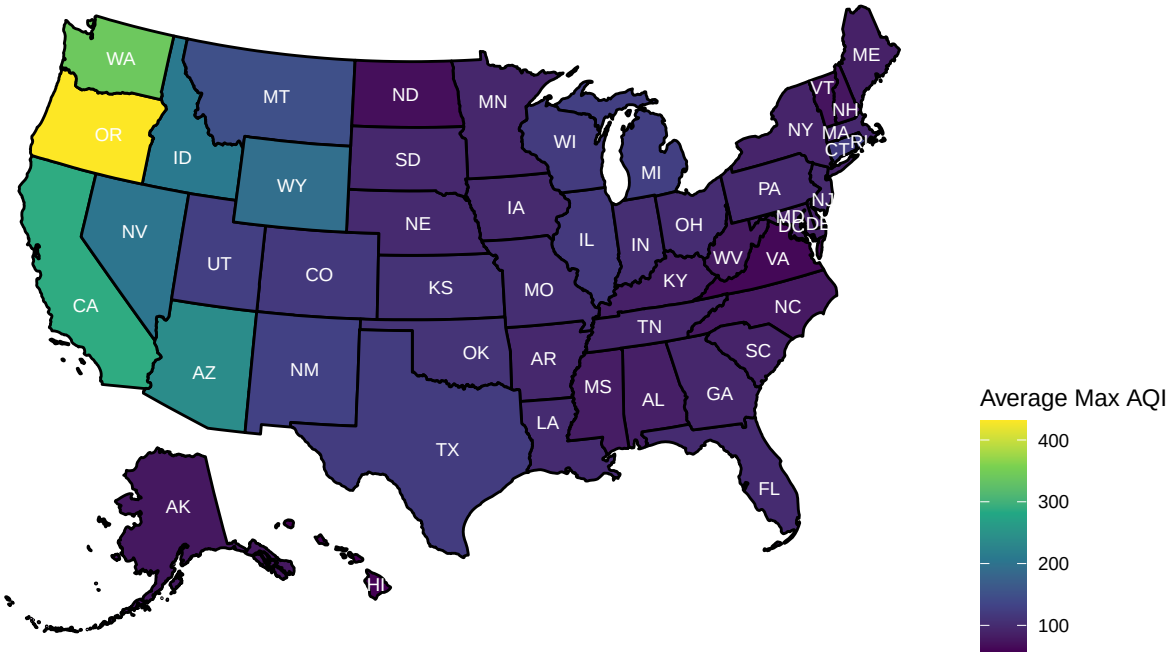
Average Max AQI by State

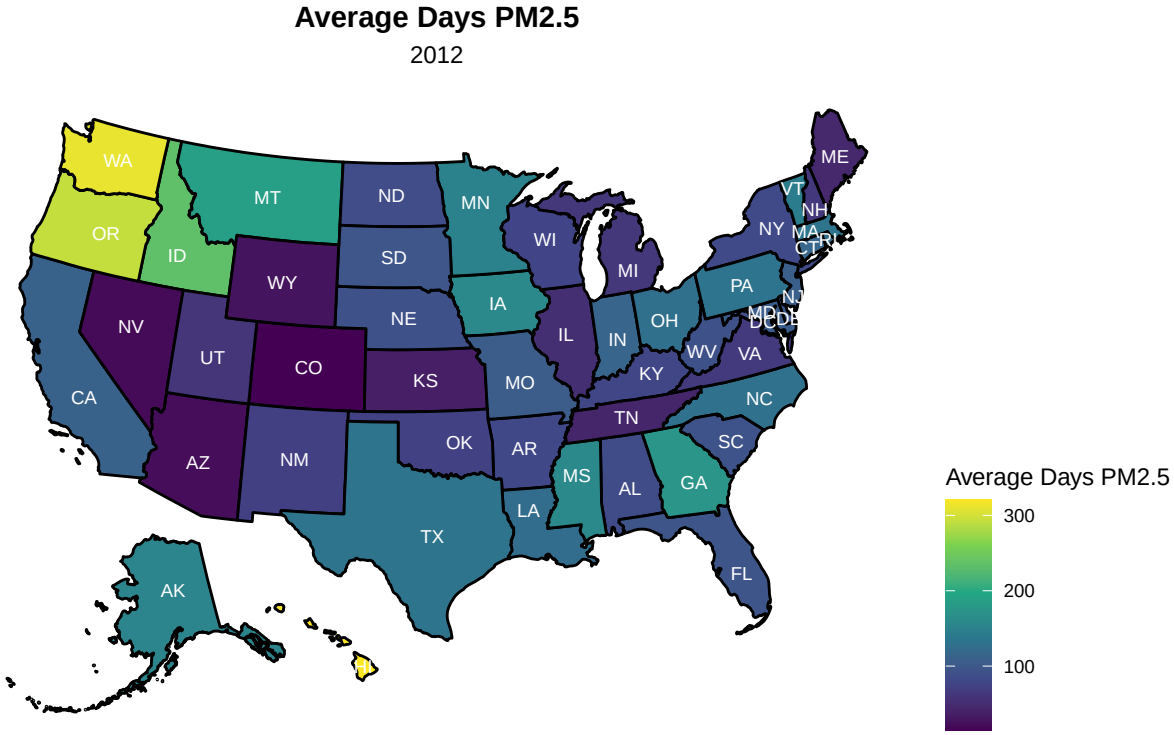
2012

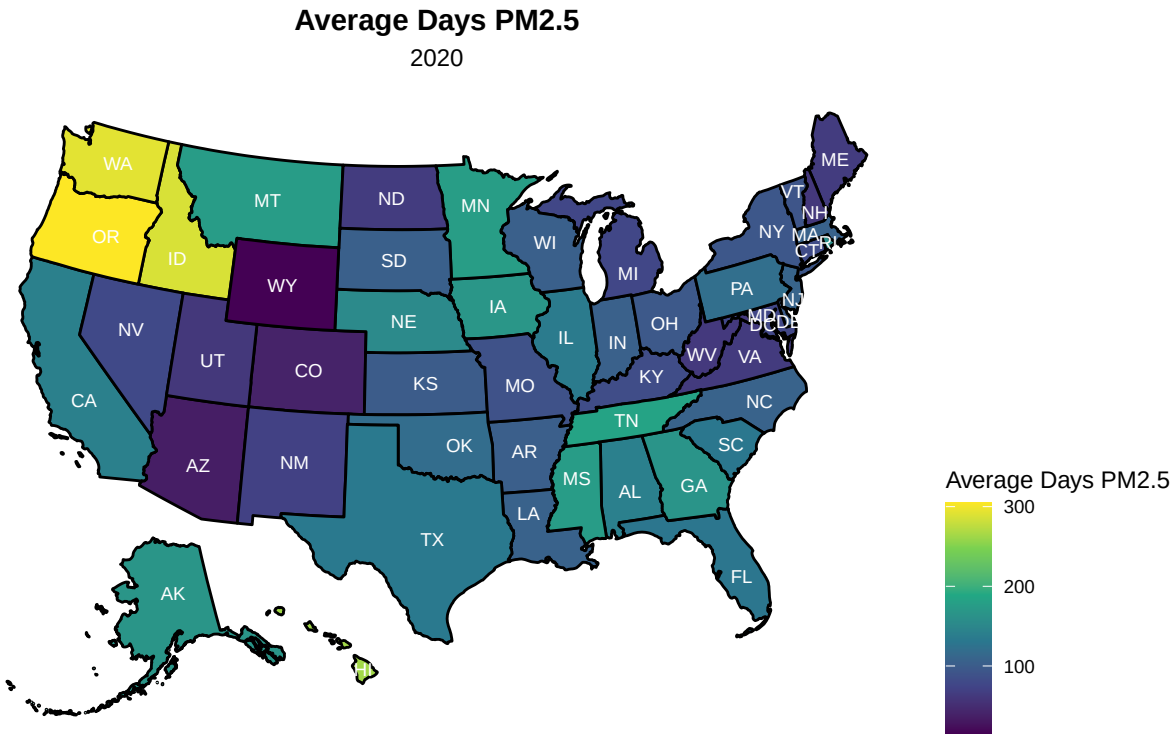


Average Max AQI by State

2020



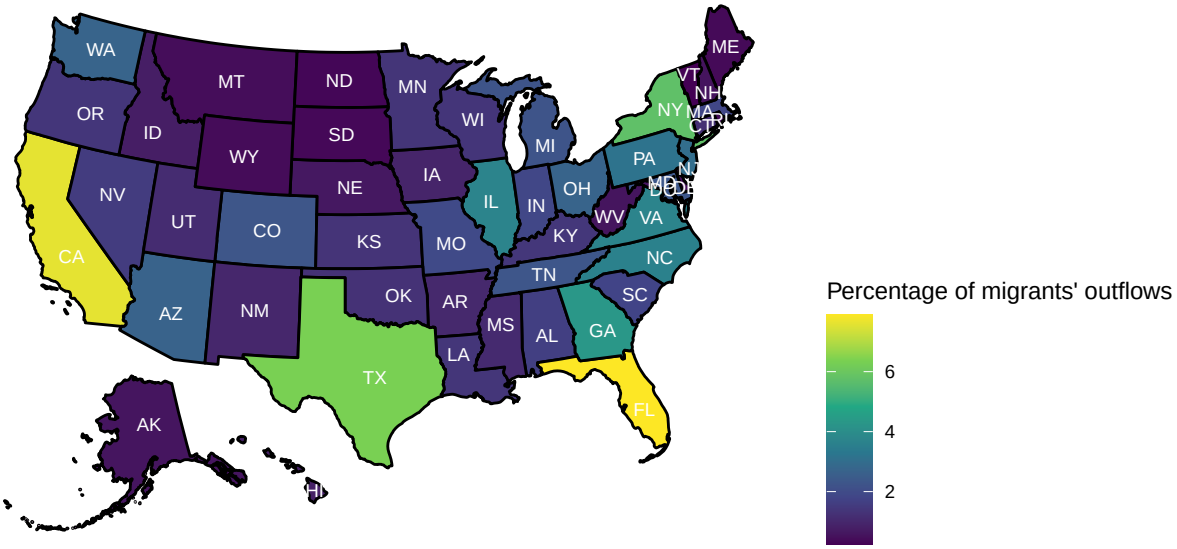

Average Days PM2.5
2012
WA
OR
CA
ID
NV
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MS
MI
IN
KY
TN
AL
OH
WV
GA
FL
SC
NC
VA
PA
NY
MD
DE
NJ
CT
RI
MA
VT
NH
ME
AK
HI
Average Days PM2.5
300
200
100

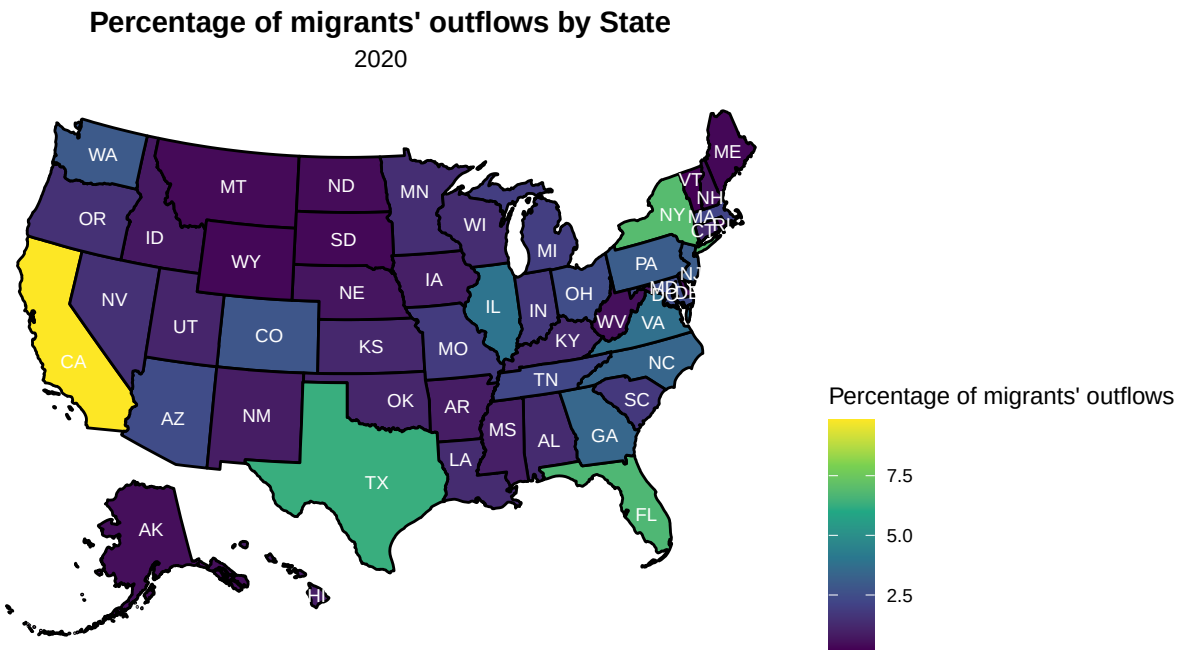


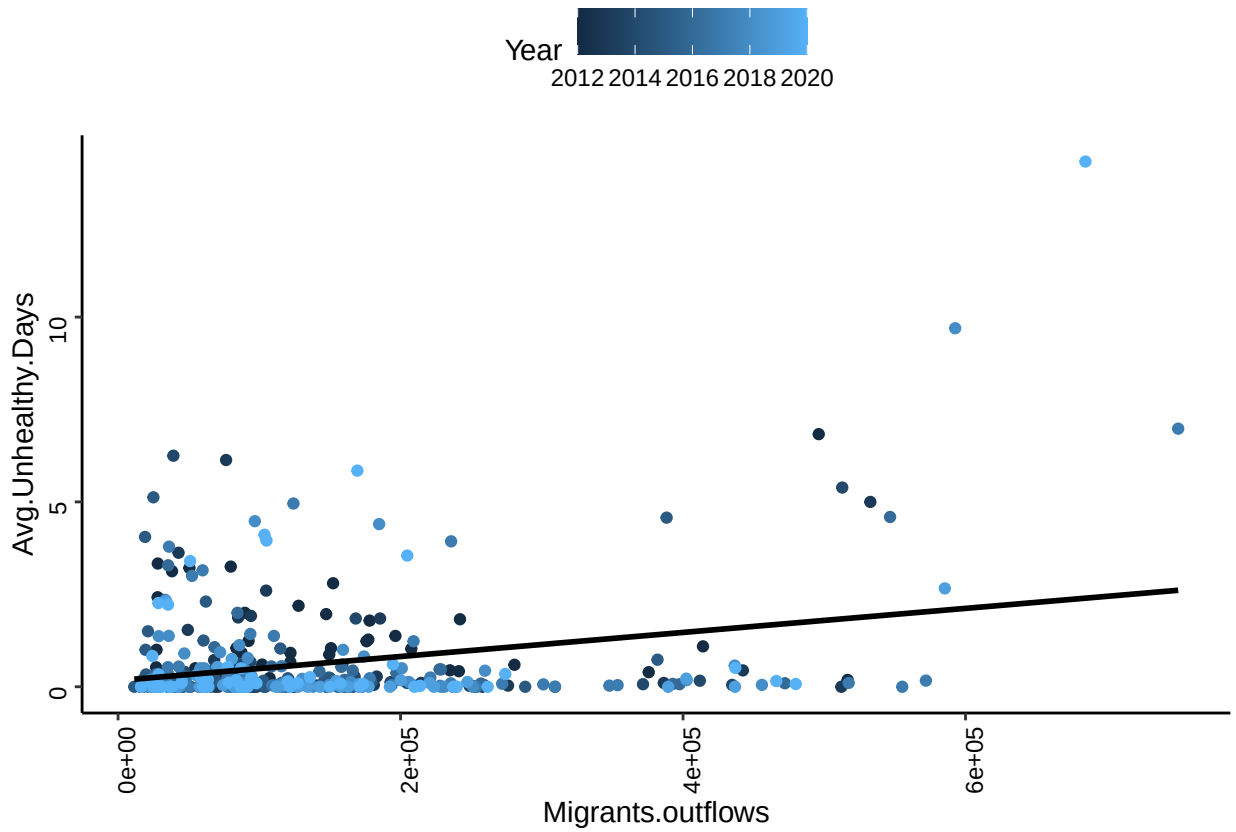
6. Maps migration by State 2012 and 2020

Percentage of migrants' outflows by State

2012




Percentage of migrants' outflows by State
2020
WA
MT
ND
MN
OR
ID
SD
WI
MI
WY
NY
VT
NH
ME
MA
RI
CT
PA
NJ
NE
IA
OH
NV
IL
IN
MD
DE
UT
CO
WV
VA
KS
MO
KY
CA
NC
TN
SC
OK
AR
MS
AL
GA
AZ
NM
LA
TX
FL
AK
HI
Percentage of migrants' outflows
7.5
5.0
2.5



Analysis

6. Correlations

```
##
##  Pearson's product-moment correlation
##
## data:  AQI_Mig.outflows_by.state_2012_2020$Avg.Unhealthy.Days and AQI_Mig.outflows_by.state_2012_202
## t = 6.3515, df = 448, p-value = 5.243e-10
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2003012 0.3700254
## sample estimates:
##      cor
## 0.287418

##
```

```
##  Pearson's product-moment correlation
##
## data:  AQI_Mig.outflows_by.state_2012_2020$Avg.Max.AQI and AQI_Mig.outflows_by.state_2012_2020$Migra
## t = 5.7931, df = 448, p-value = 1.303e-08
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1758409 0.3479362
## sample estimates:
##       cor
## 0.2639883

##
##  Pearson's product-moment correlation
##
## data:  AQI_Mig.outflows_by.state_2012_2020$Avg.Days.PM2.5 and AQI_Mig.outflows_by.state_2012_2020$Mi
## t = -0.48538, df = 448, p-value = 0.6276
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.11512036  0.06966023
## sample estimates:
##         cor
## -0.02292586
```