Effect of Particulate Matter Measures on Personal Income

By Ryan Shehata
ECO 590
Dr. Kaltenberg

Background

- Fine particulate matter is one of the main pollutants that is monitored in the US, as it is an important component of air quality
- Particles come from a variety of sources of emissions, including various economic activities
- Exposure causes adverse health effects in both the short-term and long-term
- Short-term effects include throat irritation, coughing, sneezing, and shortness of breath
- Long-term effects can range from increased cardiovascular and respiratory hospital admissions to increased mortality from lung cancer or heart disease

Source: NY State Department of Health

Research Question

- Because exposure leads to adverse health effects, I aim to measure what the effect of particulate matter exposure is on personal income per capita levels
- While particulate matter is a byproduct of economic activity that can be seen as increasing income, the negative health effects caused by exposure will outweigh the positives of production
- Results of this study can hopefully help formulate a basis for policies that balance economic productivity and preserve human health

Data Sources

- Data from 3 different sources
- Quarterly particulate matter data extracted from the Air Quality System API provided by US Environmental Protection Agency
- Quarterly personal income per capita data sourced from the Bureau of Economic Analysis (BEA)
- Census data providing regional and divisional mappings for states sourced from Github repo

Variables

- Year
- Quarter
- State
- Period
- PM2.5 Measure
- Personal Income Per Capita Data
- Region
- Division

Air Quality System API

- Contains ambient air sample data collected by state, local, tribal, and federal air pollution control agencies from various monitors nationwide
- Provides access to daily summary, quarterly summary, annual summary, and quality assurance data
- Only 19 parameters to tweak in total, this number shrinks greatly depending on what kind of data you want
- Documentation provided step-by-step details on how to build out a URL for a specific data extract
- API is free to use, but a key is required

API Request

- Only 6 parameters needed to tweak for state quarterly data
- Documentation was easy to understand and provides a step-by-step on how to build out the URL

<table>
<tr><td rowspan="4">Quarterly Summary Data</td><td rowspan="2">By County</td><td>quarterlyData/byCounty</td><td>email, key, param, bdate, edate, state, county</td><td>cbdate, cedate</td></tr>
<tr><td colspan="3">Example; returns quarterly summary FRM/FEM and non-FRM PM2.5 data for Wake County for 2016:
https://aqs.epa.gov/data/api/quarterlyData/byCounty?
email=test@aqs.api&key=test&param=88101,88502&bdate=20160101&edate=20160228&state=37&county=183</td></tr>
<tr><td rowspan="2">By State</td><td>quarterlyData/byState</td><td>email, key, param, bdate, edate, state</td><td>cbdate, cedate</td></tr>
<tr><td colspan="3">Example; returns all benzene quarterly summaries from North Carolina for 1995:
https://aqs.epa.gov/data/api/quarterlyData/byState?email=test@aqs.api&key=test&param=45201&bdate=19950515&edate=19950515&state=37</td></tr>
</table>

API Request (cont.)

- Needed quarterly state data for 4 different years, meaning over 200 API requests needed to be made
- Challenge was to figure out how to efficiently extract data from over 200 calls with minimal code
- Documentation also delineated a limitation on how many calls can be made within a minute
- Had to figure out how to build out a loop so that can extract data for each state in each period without overloading the server

API Request (cont.)

```
#First created the list of states the API needs to retrieve information from using the document
state = [num1call, num2call, num4call, num5call, num6call, num8call, num9call, 10, 11, 12, 13,
         23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40,
         41, 42, 44, 45, 46, 47, 48, 49, 50, 51, 53, 54, 55, 56]

date = [2016, 2017, 2018, 2019] #Then created the list of the 4 years I needed as well

data = [] #Initialized an empty list to append the JSON output to

for year in date: #looping thru URL for different time periods,
    #delineating parameter (PM2.5) and that I want quarterly data
    for st in state: #nested for loop to also loop through states
        url = f"https://aqs.epa.gov/data/api/quarterlyData/byState?email=rshehata1000@gmail.com
        url += f"param=88101&bdate={year}0101&edate={year}0101&state={st}"
        response = requests.get(url) #Get request
        data.append(response.json()) #Appending each JSON output to the initialized list
        time.sleep(7) #Resting the loop every 7 seconds to avoid overloading the API server
```

API Output


```
{
  "Header": [
    {
      "status": "Success",
      "request_time": "2023-05-08T10:57:21-04:00",
      "url":
"https://aqs.epa.gov/data/api/quarterlyData/byCounty?email=test@aqs.api&key=test&param=88101,88502&bdate=20160101&edate=20160228&state=37&county=183",
      "rows": 58
    }
  ],
  "Data": [
    {
      "state_code": "37",
      "county_code": "183",
      "site_number": "0014",
      "parameter_code": "88101",
      "poc": 3,
      "latitude": 35.856111,
      "longitude": -78.574167,
      "datum": "WGS84",
      "parameter": "PM2.5 - Local Conditions",
```

API Output (cont.)

- Had some issues figuring out how to convert the JSON output to a dataframe, used the JSON library and normalize function
- After conversion to a dataframe, over 140K rows and 42 columns of data were populated
- Row-level data is at the site-level within a specific state, each state and time period has hundreds of monitoring sites collecting data in different sample durations and pollutant standards
- Had to trim down by specific pollutant standard and sample duration, and also calculate a state average for each time period based on all site values

Data Cleaning

Dropping columns

- df = df.loc[:, df.columns.intersection(['site_number','datum', 'sample_duration', 'pollutant_standard','year', 'quarter', 'arithmetic_mean', 'local_site_name', 'state', 'county', 'city', 'cbsa'])]

Dropping sample duration duplicate rows

- sample_duration_list = ['1 HOUR', '24-HR BLK AVG']
- df = df[df.sample_duration.isin(sample_duration_list) == False]

Creating period value

- df['period'] = df['year'].astype(str) + " Q" + df['quarter']

Calculate average PM2.5 for each state and period combination

- df['pm2.5_measure'] = df.groupby(['period', 'state'])['arithmetic_mean'].transform('mean')

Data Cleaning (cont.)

Dropping duplicates to get to 816 rows

- df = df.drop_duplicates(subset=['state', 'period'])
- df = df.reset_index(drop=True)

Dropping final columns

- df = df.loc[:, df.columns.intersection(['state', 'period', 'pm2.5_measure', 'year', 'quarter'])]

	year	quarter	state	period	pm2.5_measure
0	2016	1	Alabama	2016 Q1	7.596292
1	2016	2	Alabama	2016 Q2	9.191956
2	2016	3	Alabama	2016 Q3	8.421384
3	2016	4	Alabama	2016 Q4	8.758739
4	2016	1	Alaska	2016 Q1	13.757600
...	...	...	...	...	...
811	2019	4	Wisconsin	2019 Q4	7.828325
812	2019	1	Wyoming	2019 Q1	5.114842
813	2019	2	Wyoming	2019 Q2	2.906546
814	2019	3	Wyoming	2019 Q3	4.391925
815	2019	4	Wyoming	2019 Q4	4.044521

816 rows × 5 columns

Bringing in External Data Sources

Merging the Income Data

- merged_df = pd.merge(df, incomedf, on=['state', 'period'])

	year	quarter	state	period	pm2.5_measure	personal_income_per_capita
0	2016	1	Alabama	2016 Q1	7.596292	38849
1	2016	2	Alabama	2016 Q2	9.191956	38941
2	2016	3	Alabama	2016 Q3	8.421384	39004
3	2016	4	Alabama	2016 Q4	8.758739	39243
4	2016	1	Alaska	2016 Q1	13.757600	56378

Bringing in Census Data

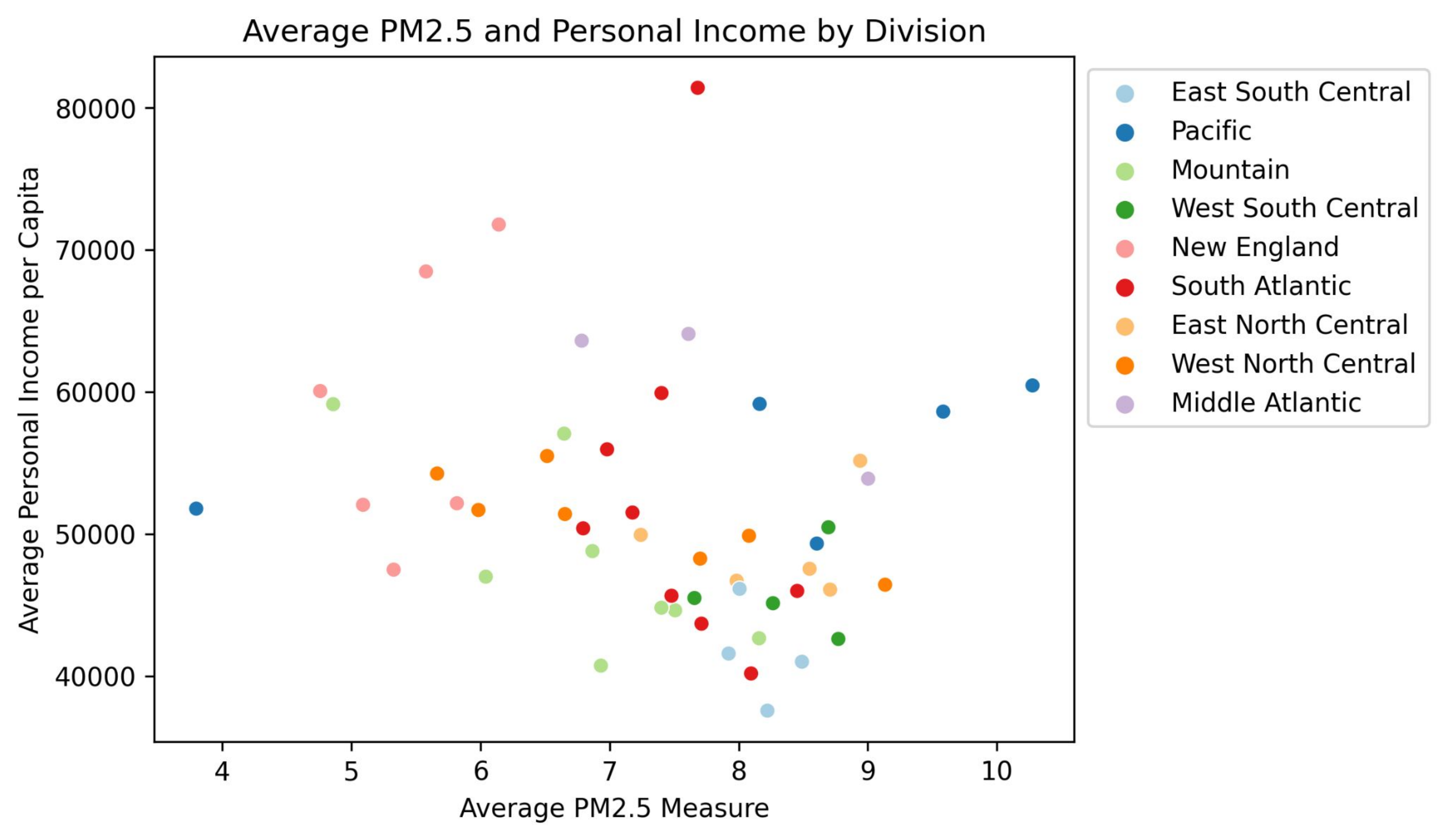
- region_map = dict(zip(census['State'], census['Region']))
- divisions_map = dict(zip(census['State'], census['Division']))
- merged_df['region'] = merged_df['state'].map(region_map)
- merged_df['division'] = merged_df['state'].map(divisions_map)

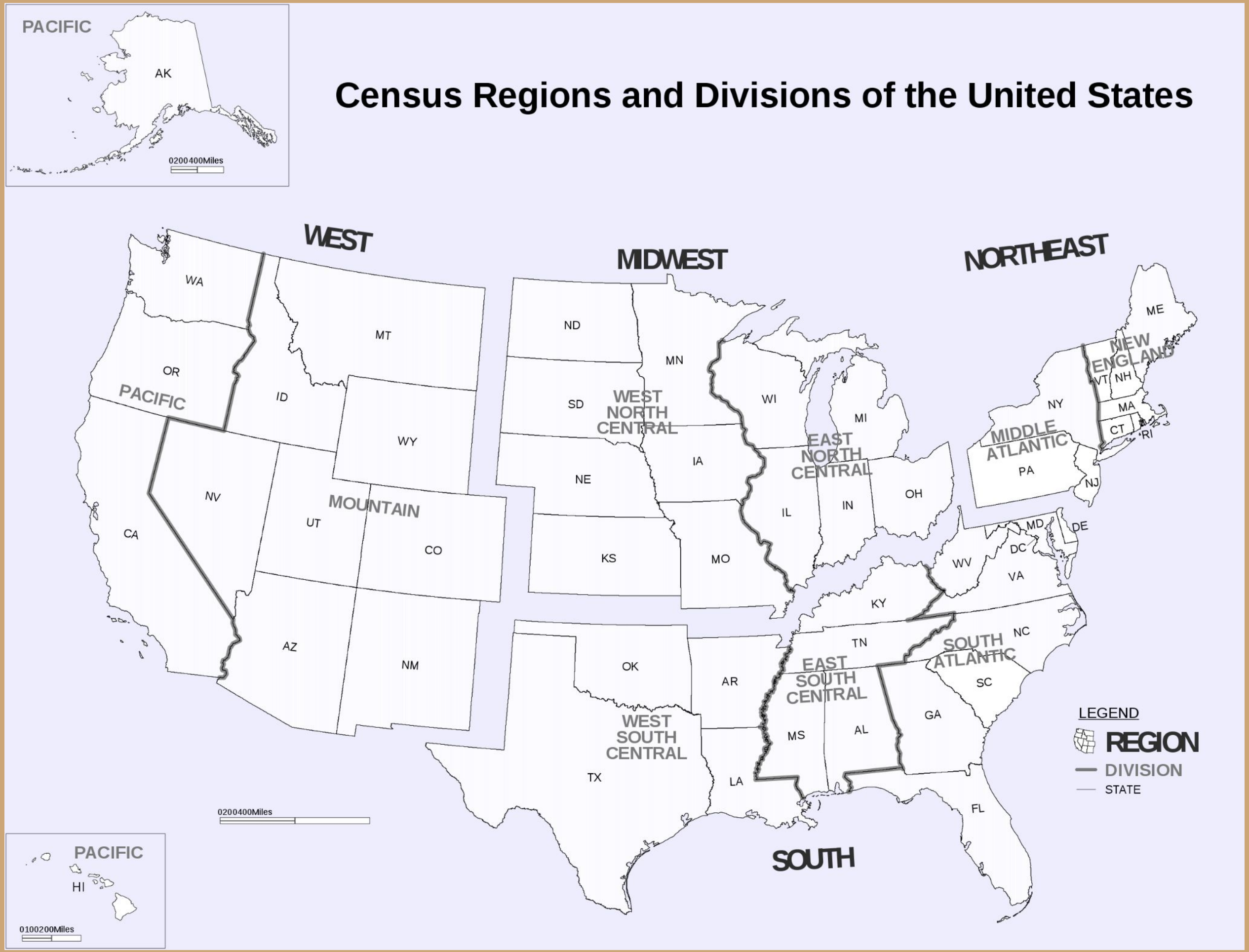
	year	quarter	state	period	pm2.5_measure	personal_income_per_capita	region	division
0	2016	1	Alabama	2016 Q1	7.596292	38849	South	East South Central
1	2016	2	Alabama	2016 Q2	9.191956	38941	South	East South Central
2	2016	3	Alabama	2016 Q3	8.421384	39004	South	East South Central
3	2016	4	Alabama	2016 Q4	8.758739	39243	South	East South Central
4	2016	1	Alaska	2016 Q1	13.757600	56378	West	Pacific

Completed Dataset

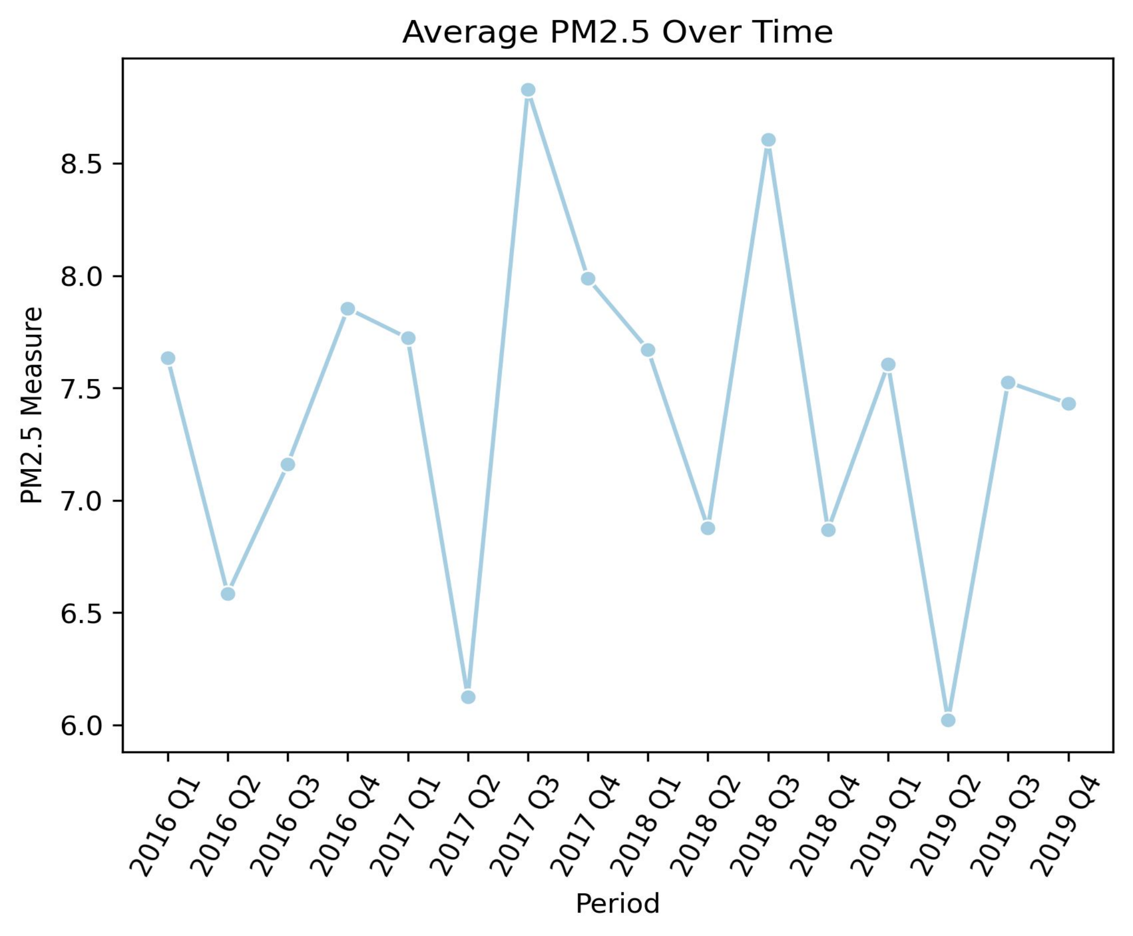
	year	quarter	state	period	pm2.5_measure	personal_income_per_capita	region	division
0	2016	1	Alabama	2016 Q1	7.596292	38849	South	East South Central
1	2016	2	Alabama	2016 Q2	9.191956	38941	South	East South Central
2	2016	3	Alabama	2016 Q3	8.421384	39004	South	East South Central
3	2016	4	Alabama	2016 Q4	8.758739	39243	South	East South Central
4	2016	1	Alaska	2016 Q1	13.757600	56378	West	Pacific
...	...	...	...	...	...	...	...	...
811	2019	4	Wisconsin	2019 Q4	7.828325	53502	Midwest	East North Central
812	2019	1	Wyoming	2019 Q1	5.114842	63294	West	Mountain
813	2019	2	Wyoming	2019 Q2	2.906546	63786	West	Mountain
814	2019	3	Wyoming	2019 Q3	4.391925	64536	West	Mountain
815	2019	4	Wyoming	2019 Q4	4.044521	64732	West	Mountain

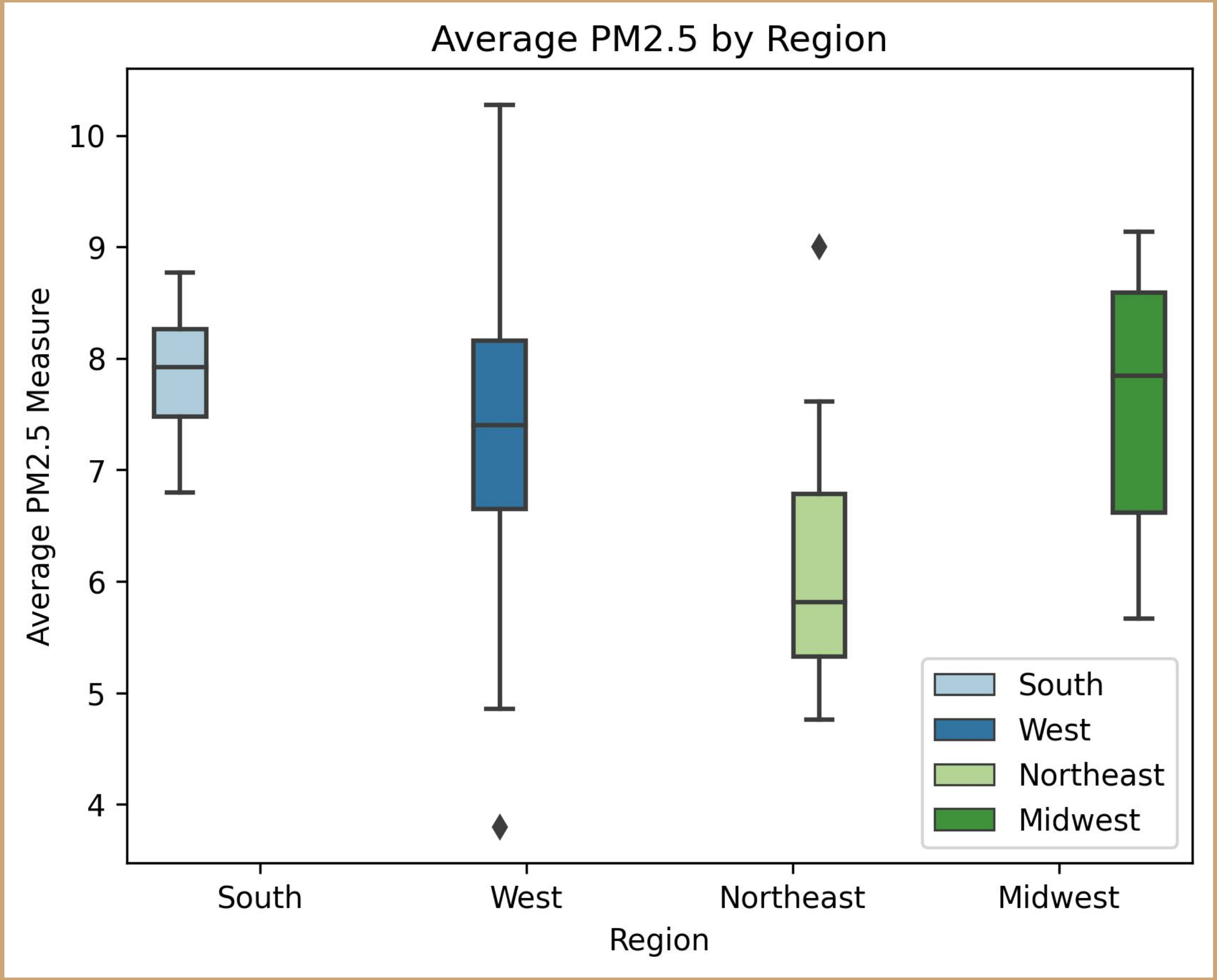
816 rows × 8 columns


Average PM2.5 and Personal Income by Division
Average Personal Income per Capita
80000
70000
60000
50000
40000
4
5
6
7
8
9
10
Average PM2.5 Measure
East South Central
Pacific
Mountain
West South Central
New England
South Atlantic
East North Central
West North Central
Middle Atlantic



Census Regions and Divisions of the United States
PACIFIC
AK
0200400Miles
WEST
MIDWEST
NORTHEAST
SOUTH
PACIFIC
MOUNTAIN
WEST NORTH CENTRAL
EAST NORTH CENTRAL
WEST SOUTH CENTRAL
EAST SOUTH CENTRAL
SOUTH ATLANTIC
MIDDLE ATLANTIC
NEW ENGLAND
WA
OR
CA
NV
ID
MT
WY
UT
CO
AZ
NM
ND
SD
NE
KS
MN
IA
MO
WI
IL
MI
IN
OH
OK
TX
AR
LA
MS
AL
TN
KY
WV
VA
MD
DC
DE
NC
SC
GA
FL
NY
PA
NJ
ME
VT
NH
MA
CT
RI
0200400Miles
PACIFIC
HI
0100200Miles
LEGEND
REGION
DIVISION
STATE


Average PM2.5 Over Time
PM2.5 Measure
8.5
8.0
7.5
7.0
6.5
6.0
2016 Q1
2016 Q2
2016 Q3
2016 Q4
2017 Q1
2017 Q2
2017 Q3
2017 Q4
2018 Q1
2018 Q2
2018 Q3
2018 Q4
2019 Q1
2019 Q2
2019 Q3
2019 Q4
Period



Average PM2.5 by Region
Average PM2.5 Measure
10
9
8
7
6
5
4
South
West
Northeast
Midwest
Region
South
West
Northeast
Midwest

Models

Model 1: $ln(personal\_income\_per\_capita_{it}) = \beta_0 + \beta_1 pm2.5\_measure_{it} + \lambda_t + \mu_{it}$

Model 2: $ln(personal\_income\_per\_capita_{it}) = \beta_0 + \beta_1 pm2.5\_measure_{it} + \lambda_t + \mu_{it}$ *(excludes West states)*

Model 3: $ln(personal\_income\_per\_capita_{it}) = \beta_0 + \beta_1 pm2.5\_measure_{it} + \lambda_t + \alpha_i + \mu_{it}$

Ryan Shehata Regression Results

	Dependent variable:		
	Log of Personal Income per Capita		
	Time Controls Only	Non-West States Only	State FE Added
	(1)	(2)	(3)
PM2.5 Measure	-0.015^{***}	-0.042^{***}	-0.0002
	(0.003)	(0.004)	(0.0003)
2016 Q2	-0.012	-0.017	0.003
	(0.031)	(0.035)	(0.002)
2016 Q3	0.003	0.008	0.010^{***}
	(0.031)	(0.035)	(0.002)
2016 Q4	0.019	0.024	0.016^{***}
	(0.031)	(0.035)	(0.002)
2017 Q1	0.029	0.022	0.028^{***}
	(0.031)	(0.035)	(0.002)
2017 Q2	0.014	-0.012	0.036^{***}
	(0.031)	(0.036)	(0.002)
2017 Q3	0.062^{**}	0.077^{**}	0.044^{***}
	(0.031)	(0.035)	(0.002)
2017 Q4	0.058^{*}	0.052	0.053^{***}
	(0.031)	(0.035)	(0.002)

2018 Q1	0.066**	0.071**	0.065***
	(0.031)	(0.035)	(0.002)
2018 Q2	0.063**	0.063*	0.074***
	(0.031)	(0.035)	(0.002)
2018 Q3	0.101***	0.123***	0.087***
	(0.031)	(0.035)	(0.002)
2018 Q4	0.088***	0.049	0.099***
	(0.031)	(0.036)	(0.002)
2019 Q1	0.116***	0.114***	0.116***
	(0.031)	(0.035)	(0.002)
2019 Q2	0.098***	0.073**	0.121***
	(0.031)	(0.036)	(0.002)
2019 Q3	0.129***	0.146***	0.130***
	(0.031)	(0.035)	(0.002)
2019 Q4	0.134***	0.112***	0.137***
	(0.031)	(0.035)	(0.002)
Constant	10.885***	11.093***	10.559***
	(0.031)	(0.042)	(0.004)
State FE?	No	No	Yes
Observations	808	608	808
R^2	0.108	0.191	0.996
Adjusted R^2	0.090	0.169	0.996
Residual Std. Error	0.156 (df = 791)	0.154 (df = 591)	0.010 (df = 741)
F Statistic	5.993*** (df = 16; 791)	8.695*** (df = 16; 591)	3,193.266*** (df = 66; 741)
Note:			*p<0.1; **p<0.05; ***p<0.01

Regression Results

- PM2.5 is statistically significant at the 1% level in Model 1 where only time fixed effects are included
- A one microgram per cubic meter increase in the particulate matter measure leads to a -1.5 percent change in personal income per capita holding all else constant
- PM2.5 is again statistically significant at the 1% level in Model 2 Census-designated Western states are excluded
- A one microgram per cubic meter increase in the particulate matter measure leads to a -4.2 percent change in personal income per capita holding all else constant

Regression Results (cont.)

- In Model 3 where time fixed effects and state fixed effects are included, particulate matter is no longer statistically significant
- A one microgram per cubic meter increase in the particulate matter measure leads to a -.02 percent change in personal income per capita holding all else constant

Discussion & Conclusion

- While the effect is small, the sheer presence of a pollutant in the air lowering income levels indicates that its effects on health in the long-term can't be ignored
- When health is negatively impacted, economic productivity at an individual level decreases in the long-run
- In the future, studies should try isolating other pollutants as well and see if the effect on income is also negative
- Some missing variables including state population and largest source of economic productivity within a state
- Interested in measuring the impact of particulate matter or other air quality indicators on personal consumption expenditures

Thank you!

Source

- https://www.health.ny.gov/environmental/indoors/air/pmq_a.htm