Breathe Easy: Unveiling the Hidden Causes & Risk Factors of COPD (Chronic Obstructive Pulmonary Disease)

Flower Yang, Ting Tsai, Yu-Ting Weng, Vivian Hung

Background

What is COPD?

“Chronic Obstructive Pulmonary Disease(COPD) is a leading cause of death in the United States. Overall COPD prevalence declined during 1999–2011. But prevalence of COPD did not change significantly from 2011 (6.1%) to 2021 (6.0%). *”

How does GIS analysis help us understand COPD prevalence?

“The US counties with the highest COPD prevalence tended to be located in **non-metropolitan area.** Rural populations may have more COPD-related issues due to **more people smoking, increased exposure to secondhand smoke, and less access to smoking cessation programs compared with people living in more urban areas.** Rural residents are also more likely to be **uninsured and have higher poverty levels**, which may lead to less access to early diagnosis and treatment. **”

*Trends in the Prevalence of Chronic Obstructive Pulmonary Disease Among Adults Aged ≥18 Years — United States, 2011–2021, CDC

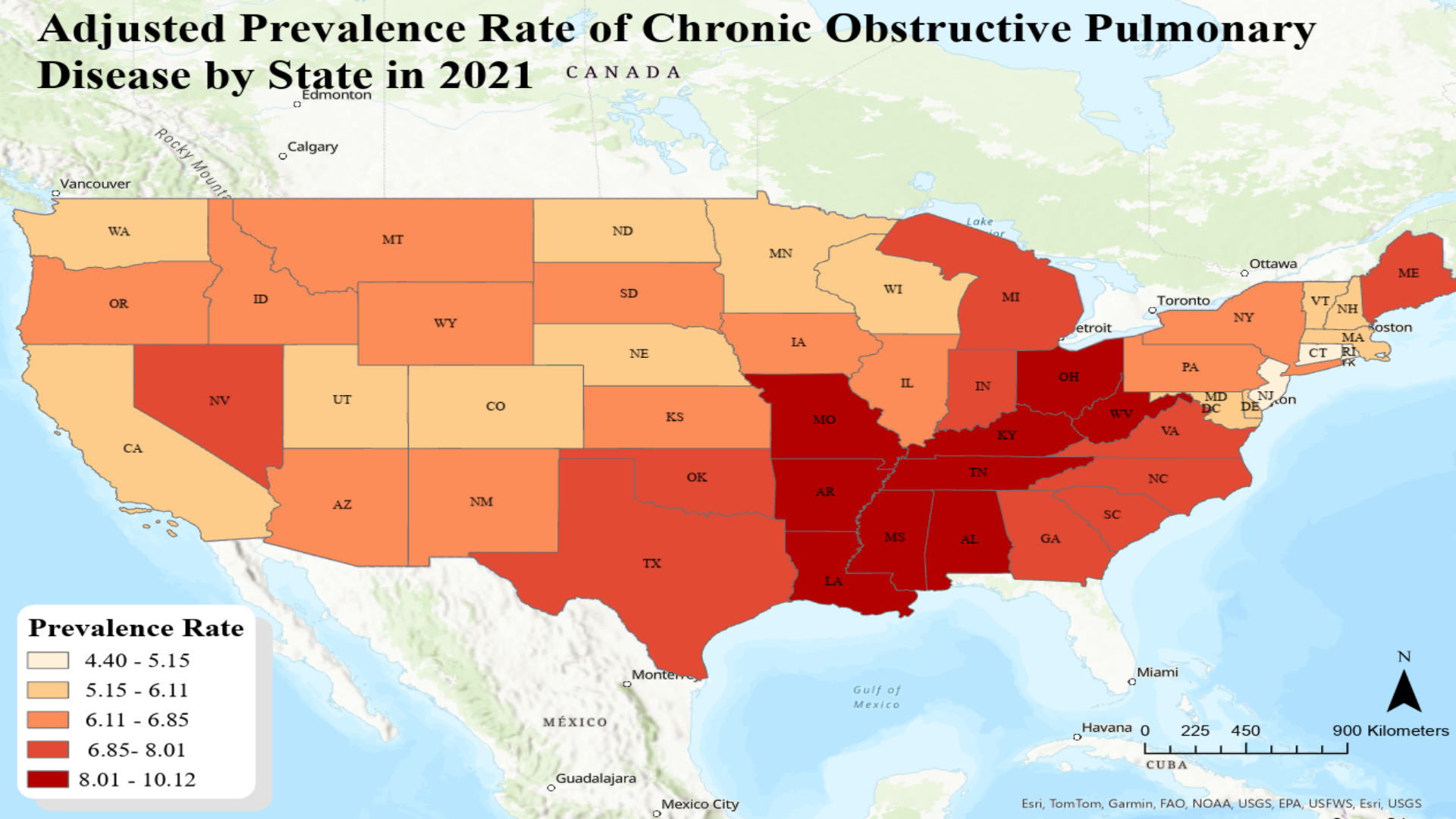
** Urban-Rural Differences in COPD, CDC

Research Question and Method Overview

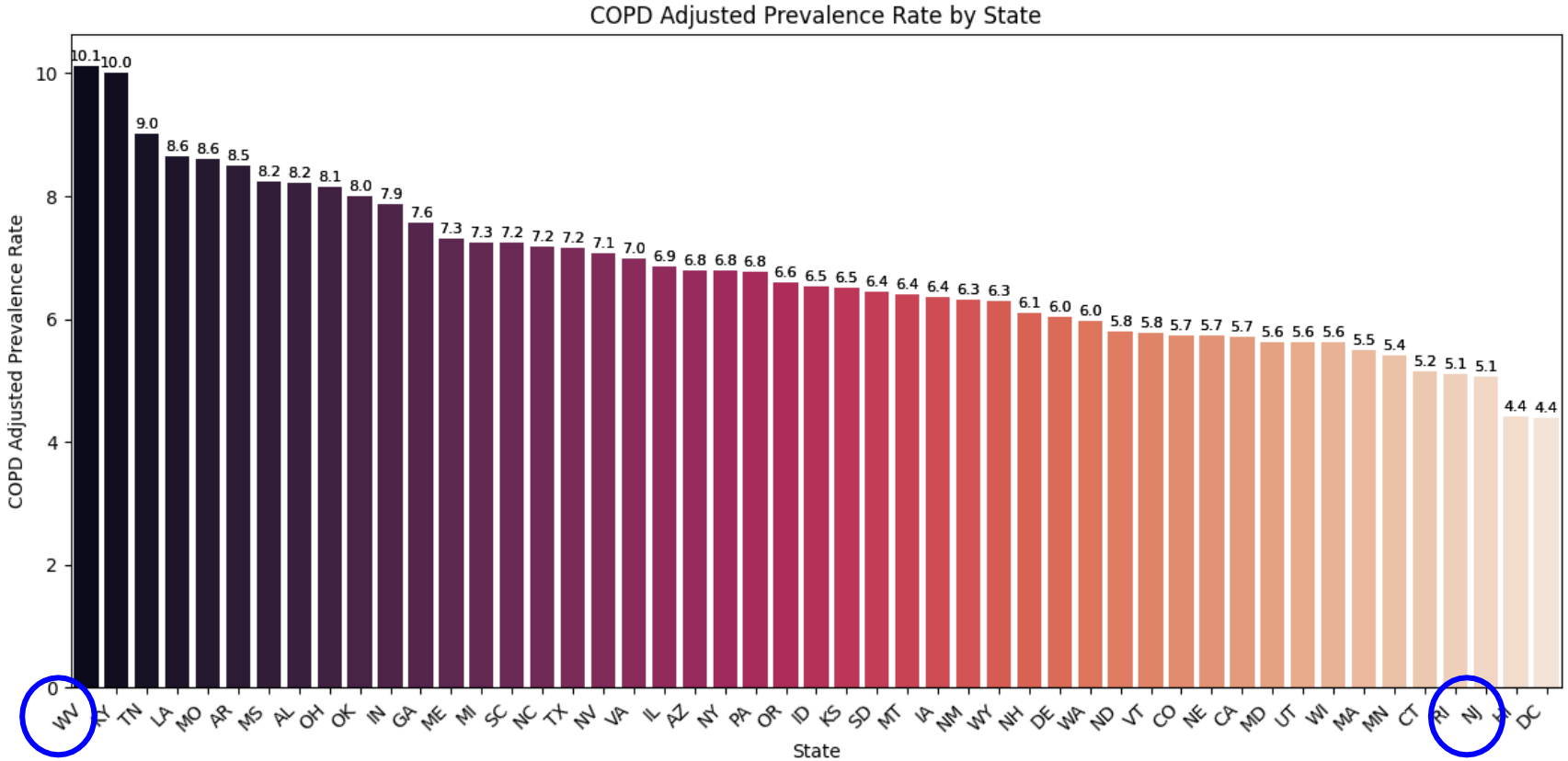
- Research Question
  - What **factors** contribute to chronic obstructive pulmonary disease (COPD) prevalence?
  - Are there **regional disparities**?
  - How can **policies** be formulated or adjusted to address these disparities?

- Method
  - Compare the state with the highest prevalence rate to the one with the lowest prevalence rate at the county level.
  - Identify key variables /socioeconomic factors by regression analysis
  - Compare patterns at the county level
  - Use residual to recheck the model fit

Adjusted Prevalence Rate of Chronic Obstructive Pulmonary Disease by State in 2021



COPD Prevalence Across States



Regression: Identifying Key Determinants of COPD

In order to further understand the key determinants of COPD, we conducted a regression analysis to explore the relationships between COPD and various factors across demographics, behavioral patterns, and healthcare resources.

Using Lasso feature selection and Ordinary Least Squares (OLS) methods, we identified statistically significant variables and mapped their distribution across New Jersey and West Virginia. This approach allows us to investigate potential disparities in the factors contributing to COPD in these regions.

Feature Selections - National Level

Target variable: COPD_AdjPrev

OLS Regression Results

Dep. Variable:	COPD_AdjPrev	R-squared:	0.945
Model:	OLS	Adj. R-squared:	0.925
Method:	Least Squares	F-statistic:	46.45
Date:	Mon, 13 May 2024	Prob (F-statistic):	3.77e-18
Time:	20:36:04	Log-Likelihood:	-11.010
No. Observations:	49	AIC:	50.02
Df Residuals:	35	BIC:	76.51
Df Model:	13		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025
const	1.3264	2.032	0.653	0.518	-2.799
ACCESS2_AdjPrev	-0.0202	0.032	-0.633	0.531	-0.085
BINGE_AdjPrev	-0.1537	0.032	-4.865	0.000	-0.218
CHECKUP_AdjPrev	0.0203	0.021	0.961	0.343	-0.023
CSMOKING_AdjPrev	0.3335	0.046	7.209	0.000	0.240
LPA_AdjPrev	-0.0454	0.033	-1.356	0.184	-0.113
SLEEP_AdjPrev	0.0919	0.034	2.721	0.010	0.023
Unemployment_rate_2021	-0.1175	0.061	-1.921	0.063	-0.242
Median_Household_Income_2021	-1.86e-05	1.26e-05	-1.478	0.148	-4.42e-05
Percent of adults with a bachelor's degree or higher, 2017-21	0.0229	0.017	1.369	0.180	-0.011
popn_wh_21	8.057e-08	3.94e-08	2.043	0.049	5e-10
popn_bl_21	-3.229e-07	1.39e-07	-2.320	0.026	-6.06e-07
popn_hsp_21	1.449e-07	7.17e-08	2.021	0.051	-6.5e-10
popn_asn_21	-2.783e-07	2e-07	-1.388	0.174	-6.85e-07

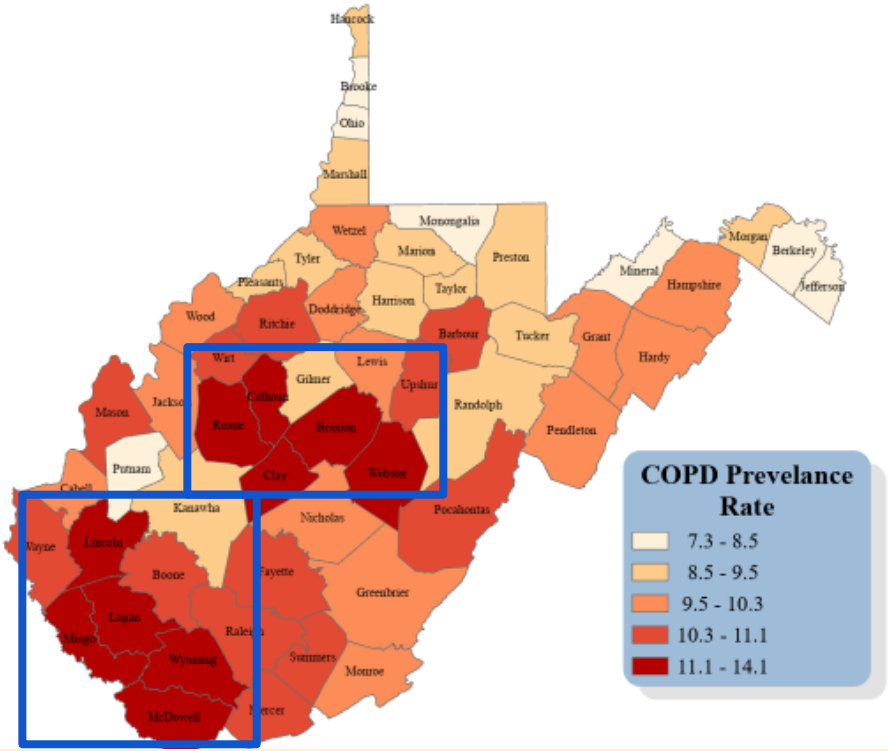
Statistically Significance:

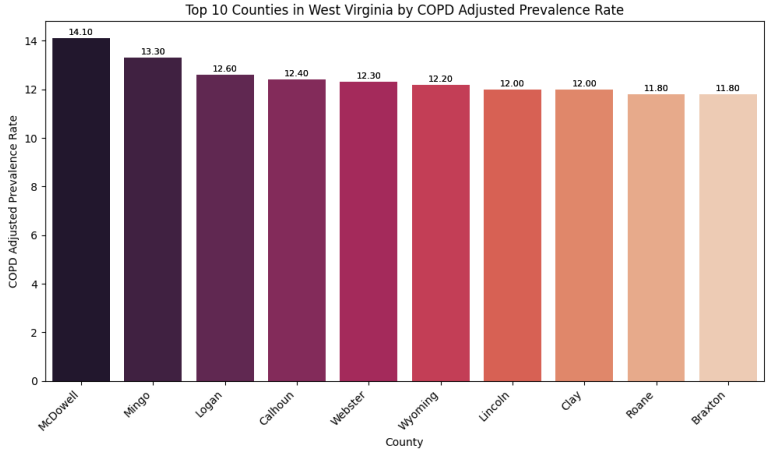
- Smoking
- Binge Drinking
- Sleep
- White Population
- African American Population

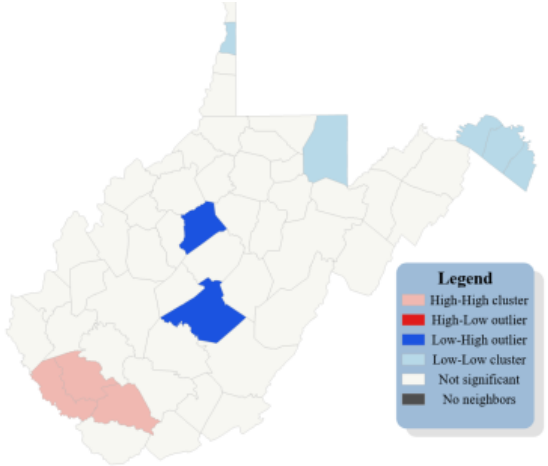
West Virginia - Diving into State

- Smoking
- Binge Drinking
- Sleep
- White Population
- African American Population

COPD Prevalence in West Virginia by County





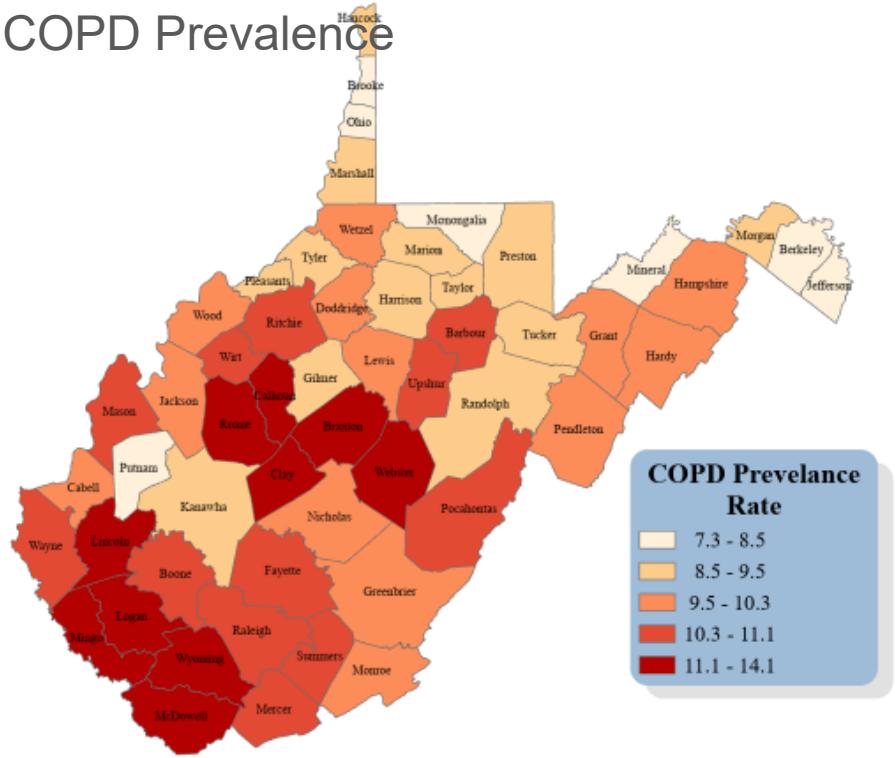


The COPD prevalence is concentrated in the central and southern regions. The farther south the county is, the higher the COPD prevalence rate.

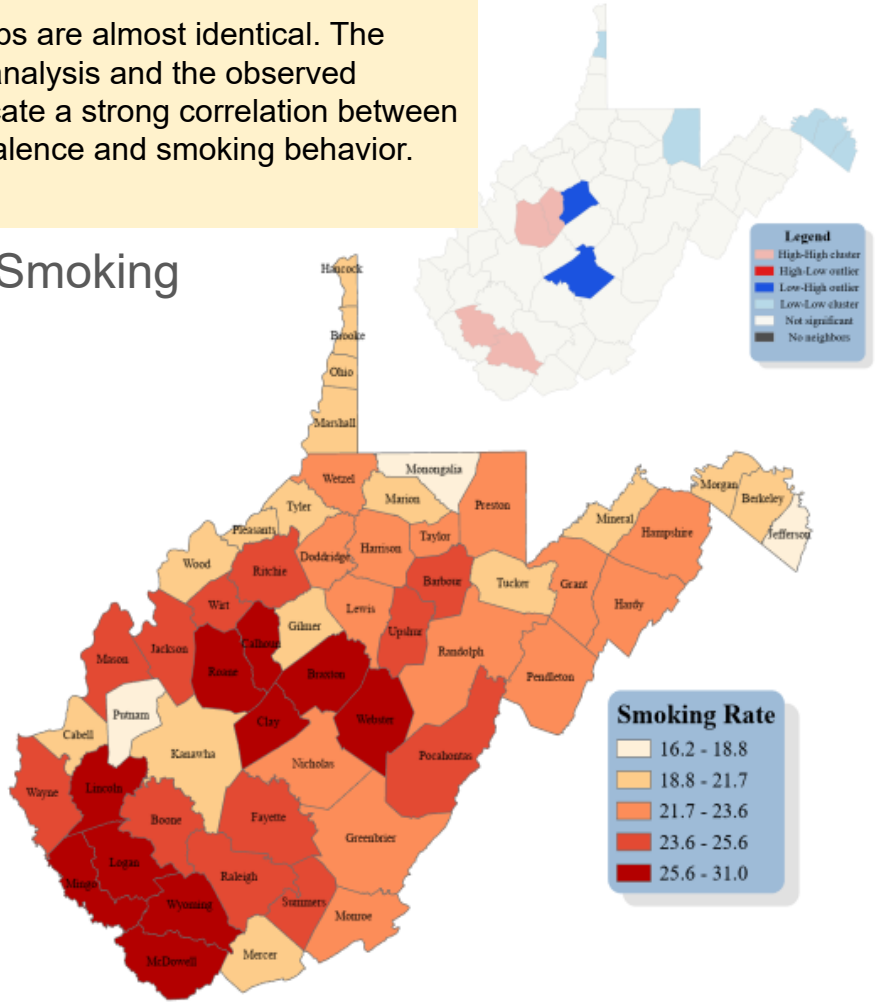
West Virginia - Smoking

The two maps are almost identical. The regression analysis and the observed pattern indicate a strong correlation between COPD prevalence and smoking behavior.

COPD Prevalence

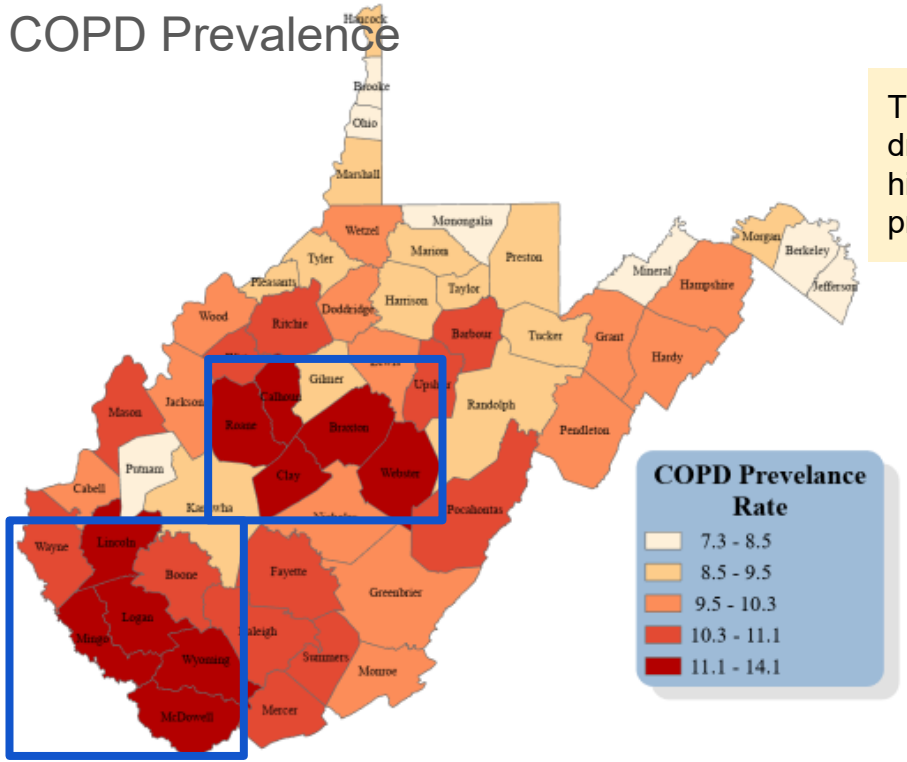


Smoking



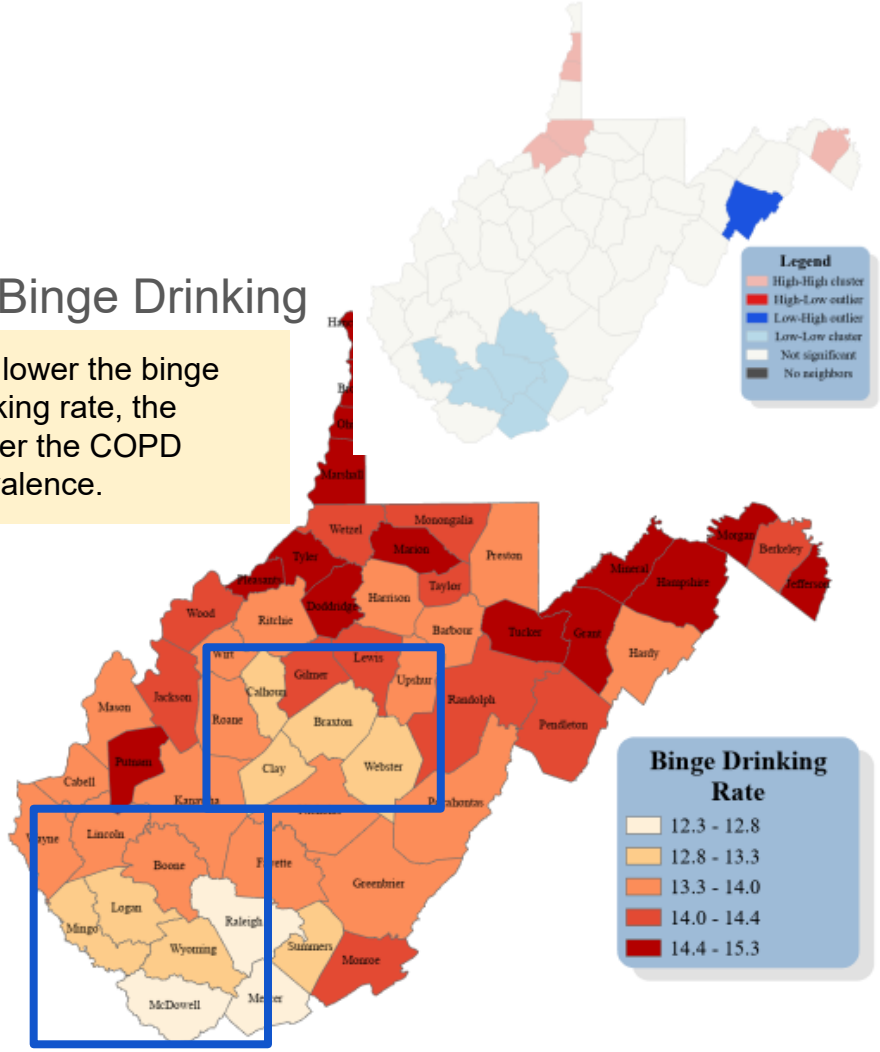
West Virginia - Binge Drinking

COPD Prevalence



Binge Drinking

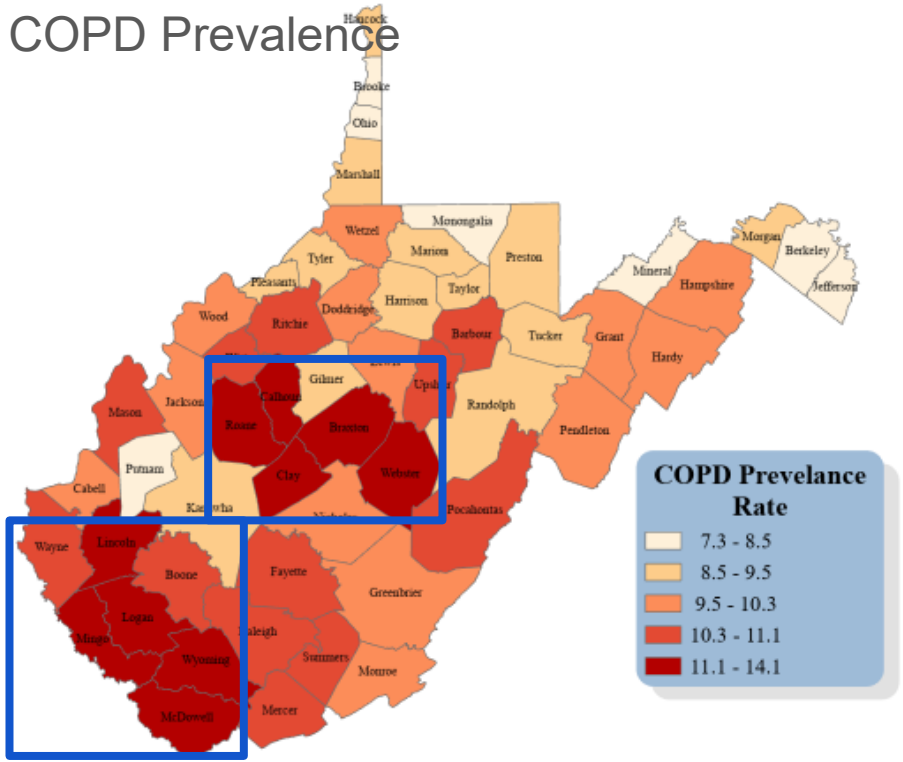
The lower the binge drinking rate, the higher the COPD prevalence.



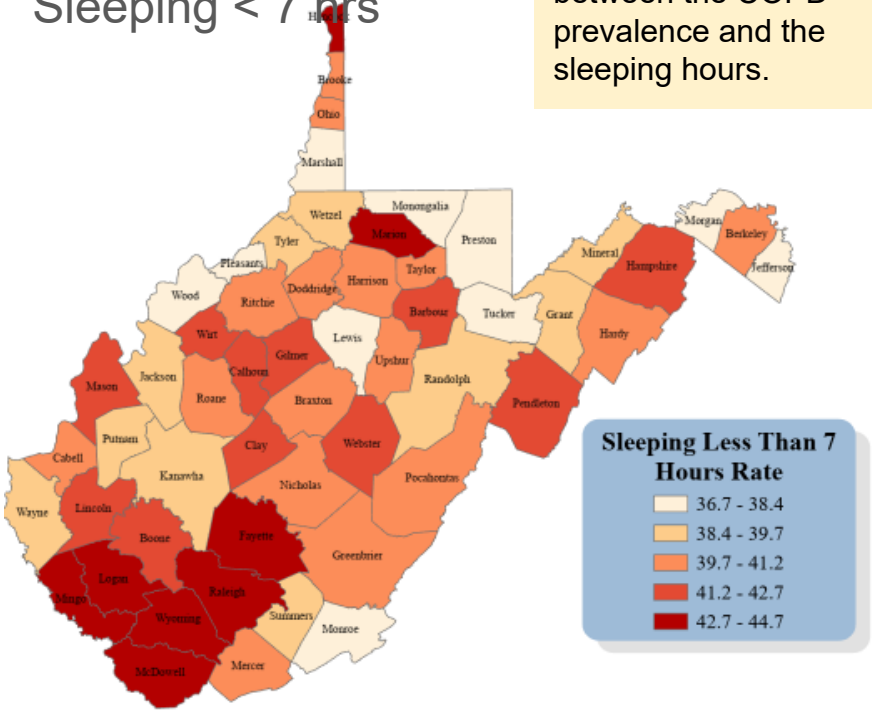
West Virginia - Sleeping Less Than 7 Hours

There is no clear pattern observed between the COPD prevalence and the sleeping hours.

COPD Prevalence

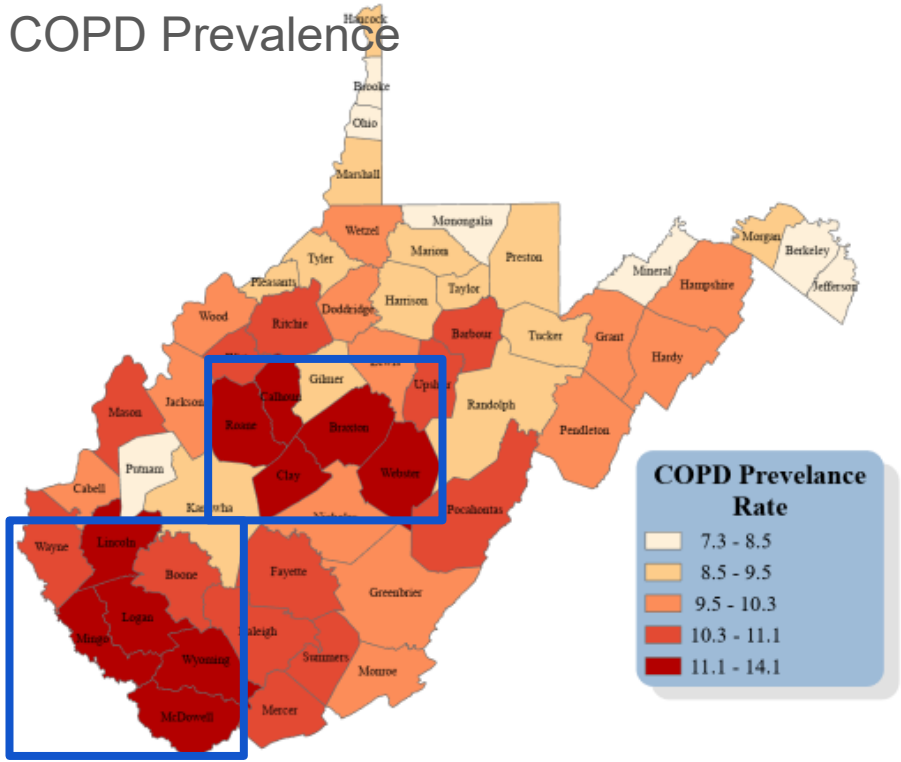


Sleeping < 7 hrs

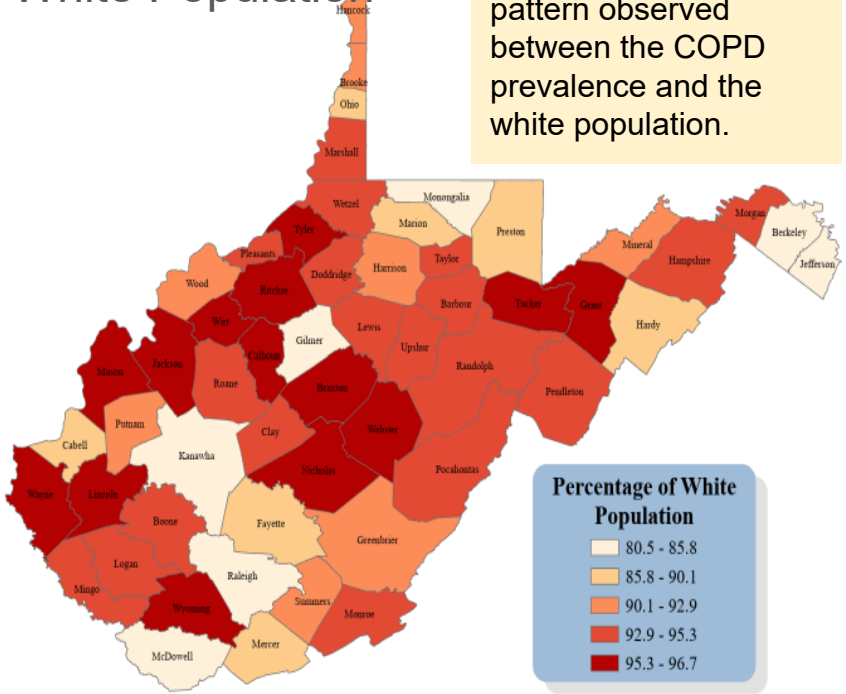


West Virginia - White Population

COPD Prevalence



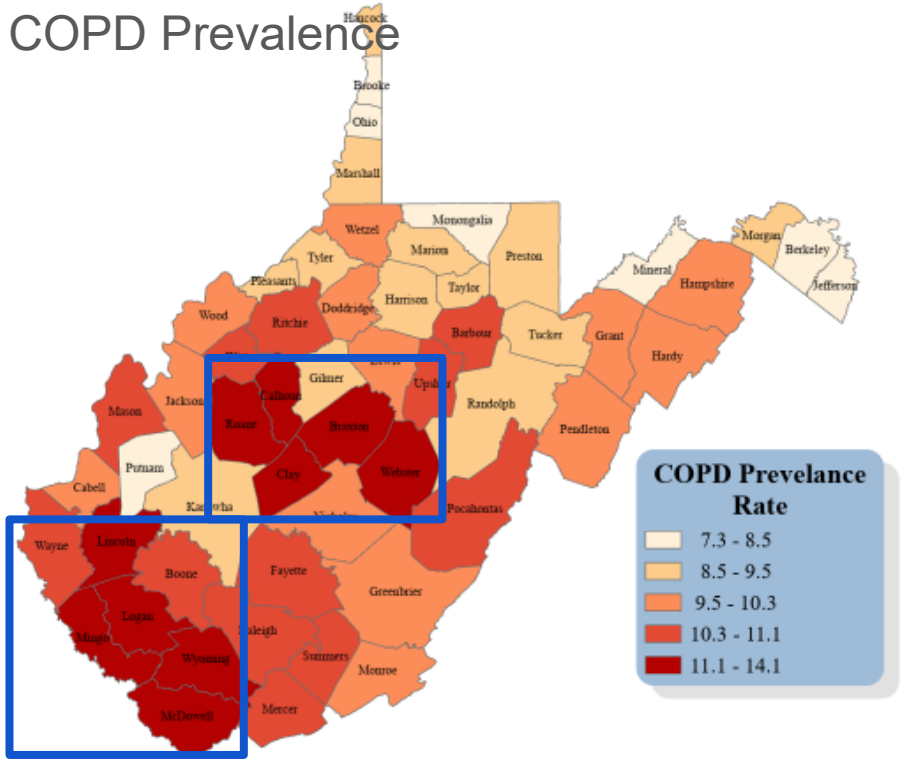
White Population



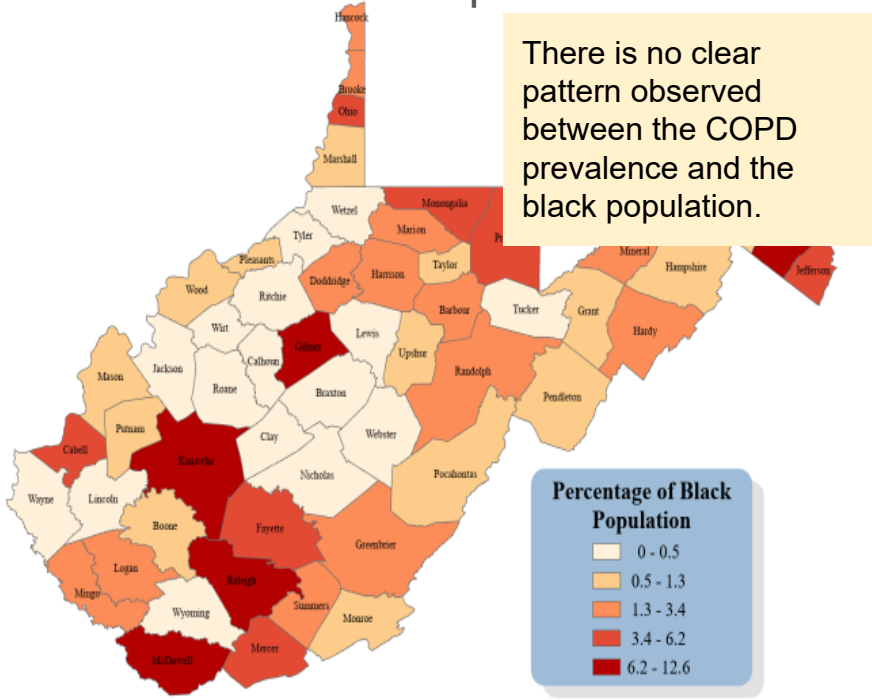
There is no clear pattern observed between the COPD prevalence and the white population.

West Virginia - African American Population

COPD Prevalence



African American Population



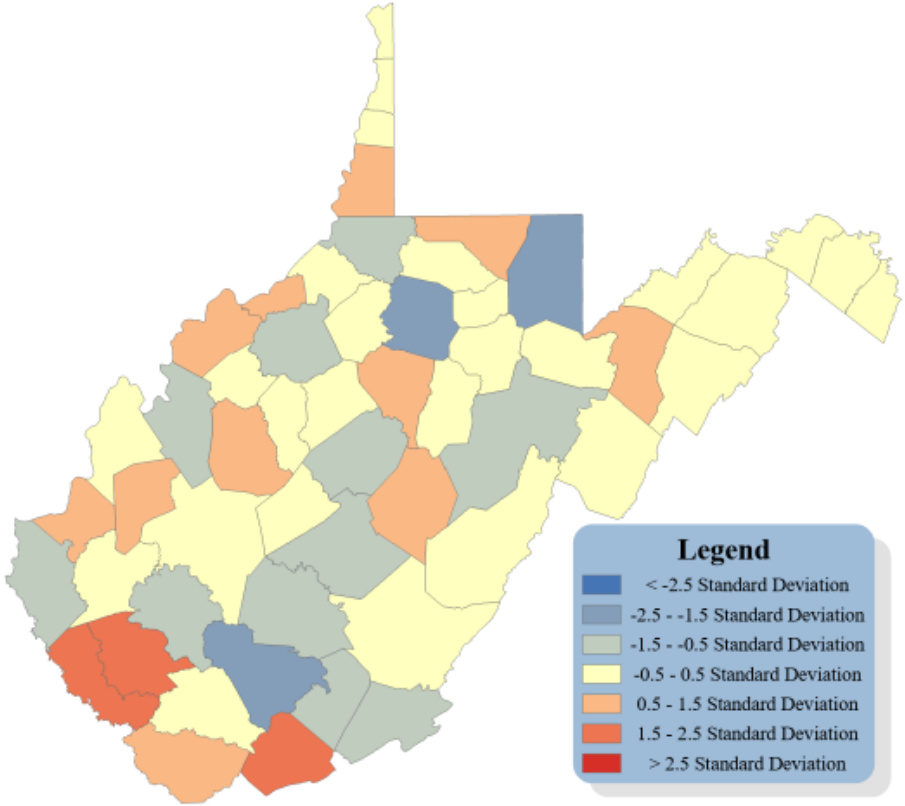
There is no clear pattern observed between the COPD prevalence and the black population.

West Virginia - Standardized Errors Map

Dependent variable: COPD prevalence

Independent variables:

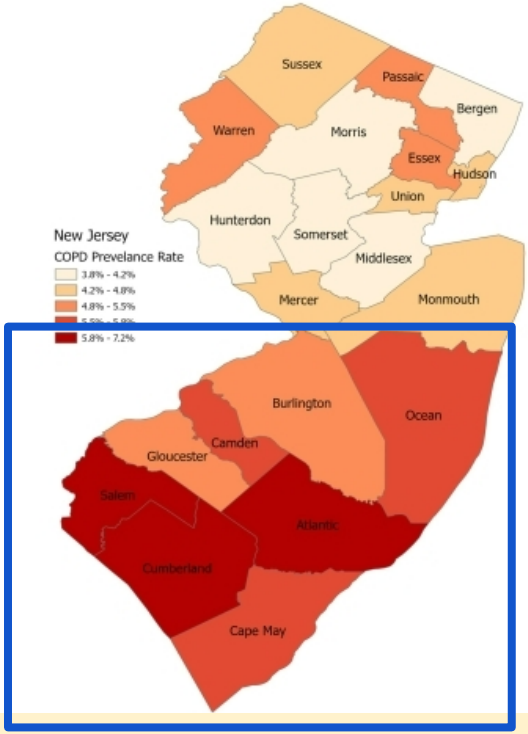
- Smoking
- Binge Drinking
- Sleep
- White Population
- African American Population

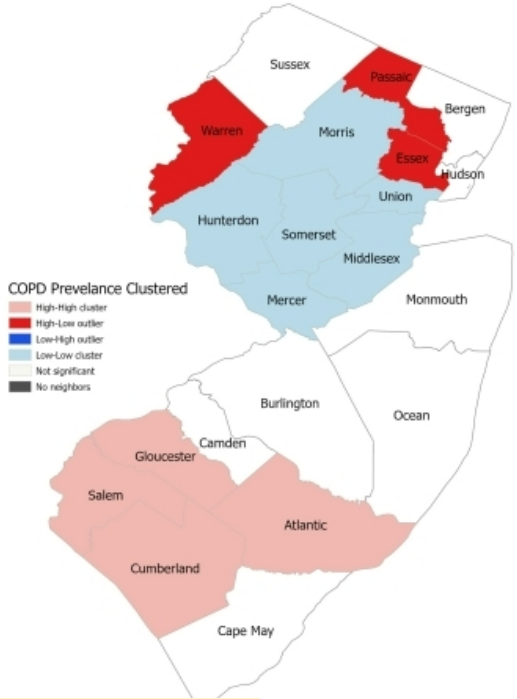


New Jersey - Diving into State

- Smoking
- Binge Drinking
- Sleep
- White Population
- African American Population

COPD Prevalence in New Jersey by County



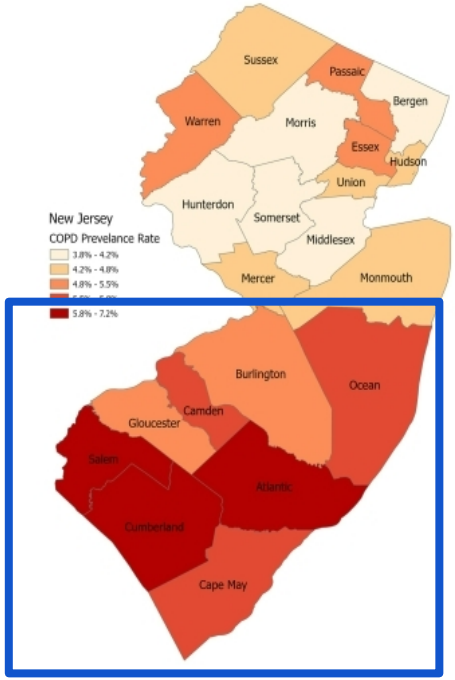


South part of NJ have higher rate of COPD.
Salem, Cumberland and Atlantic has the highest rate of COPD.

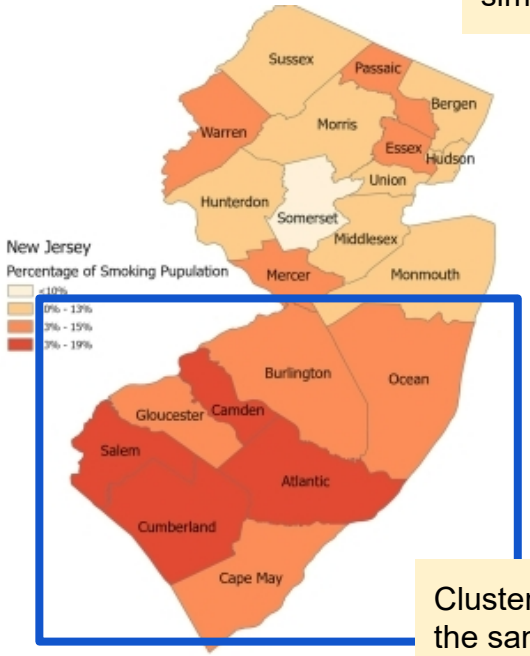
Looking at the clustered COPD, We found that except for south part of NJ, **Warren, Passaic, and Essex** with High-Low cluster are also worth investigating.

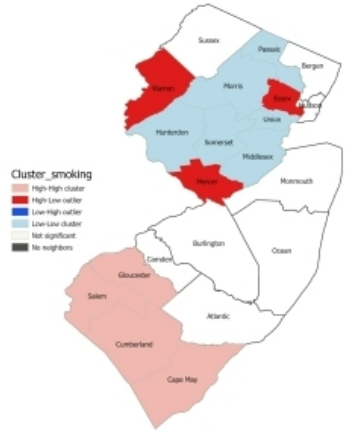
New Jersey - Smoking

COPD Prevalence



Smoking



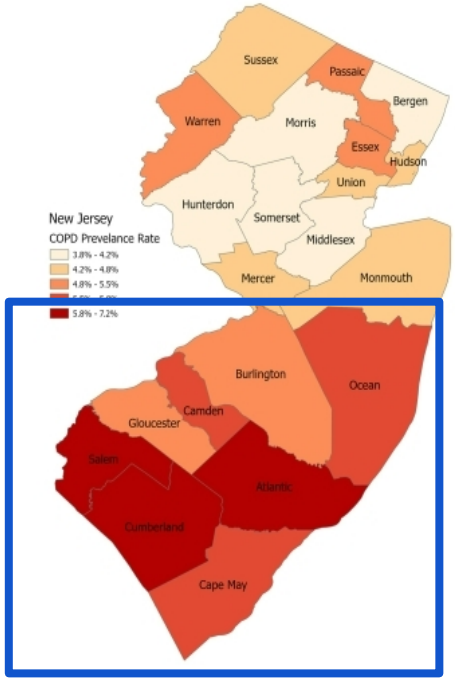


COPD and Smoking pattern are highly similar.

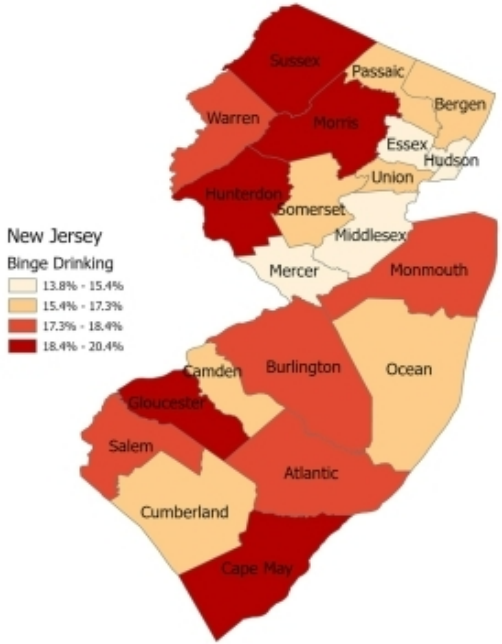
Clustered smoking also explain the same thing!

New Jersey - Binge Drinking

COPD Prevalence



Binge Drinking

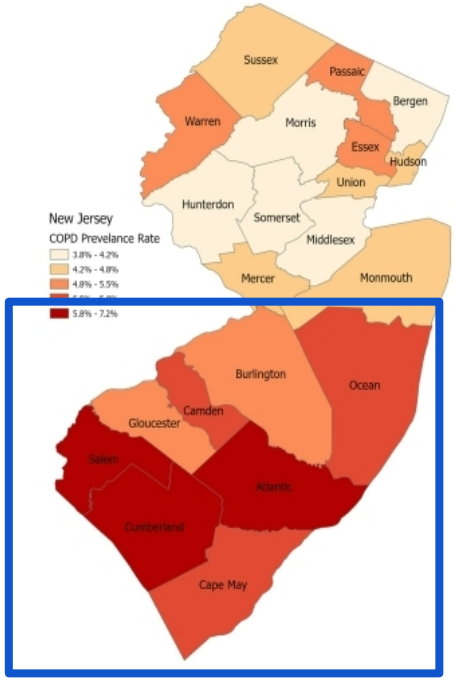


COPD and binge drinking should be negative correlated according to our OLS.

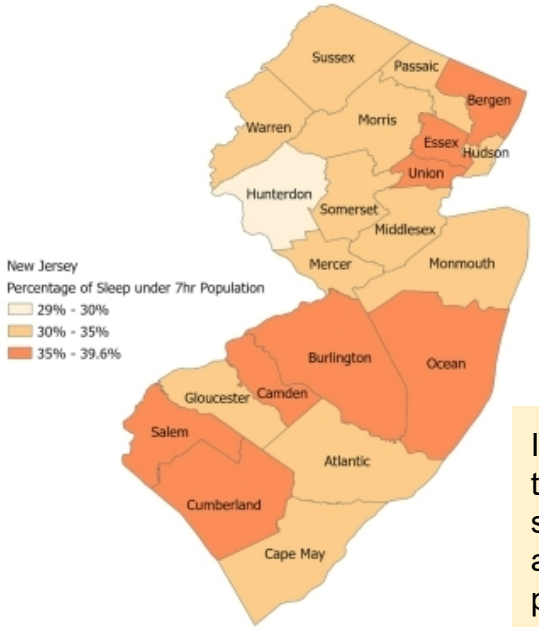
In the map, we can observe some high binge drinking areas have a lower COPD prevalence. But the pattern is not obvious.

New Jersey - Sleeping Less Than 7 Hours

COPD Prevalence



Sleeping less than 7 hour

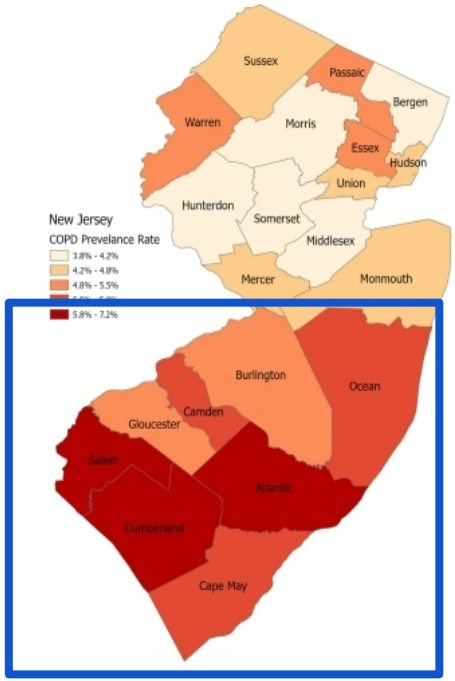


Sleeping less is associated with higher COPD.

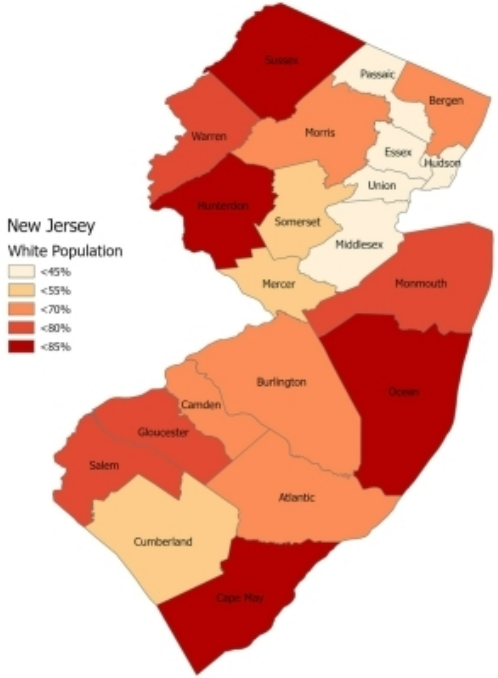
In the map, we can see that those have higher sleep rate area, does align with higher COPD prevalence.

New Jersey - White Population

COPD Prevalence



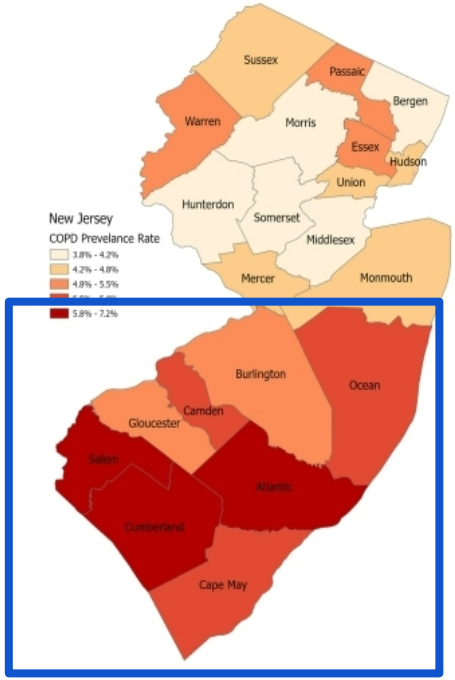
White Population



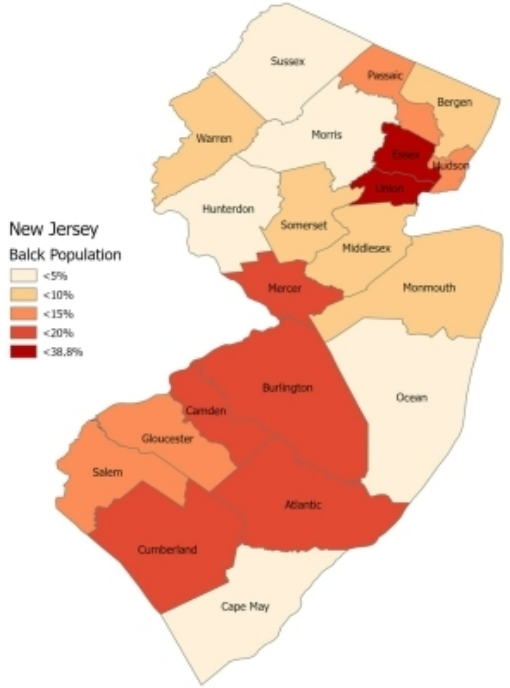
There is no clear pattern observed between the PD prevalence and the white population.

New Jersey - African American Population

COPD Prevalence



African American Population



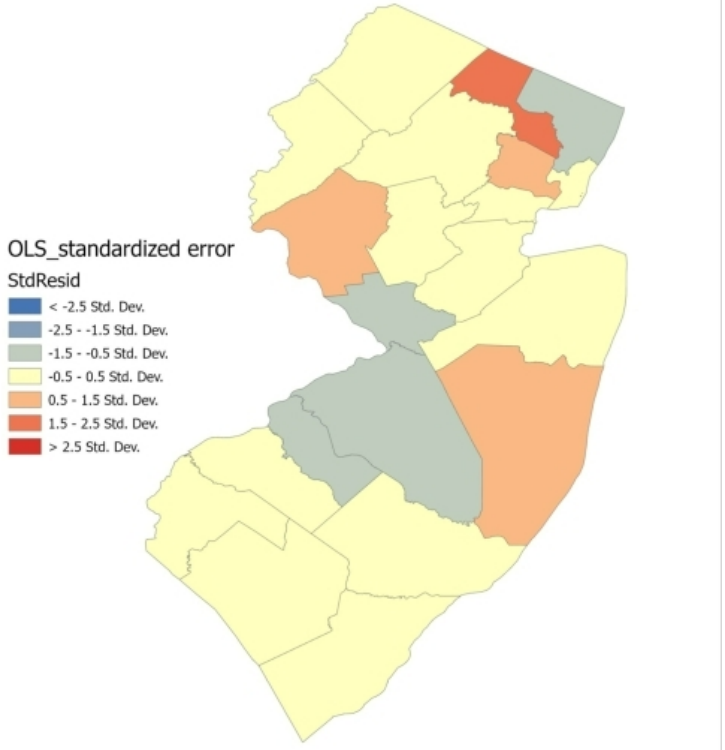
COPD and African American have relatively similar pattern, with outliner counties like Ocean and Cape May

New Jersey - Standardized Errors Map

Dependent variable: COPD prevalence

Independent variables:

- Smoking
- Binge Drinking
- Sleep
- White Population
- African American Population



Results: Feature Selections vs Spatial Analysis

West Virginia (Highest COPD state)

- Smoking
- Binge Drinking
- Sleeping hours < 7
- White population
- African American Population

New Jersey (Loewet COPD state)

- Smoking
- Binge Drinking
- Sleeping hours < 7
- White population
- African American Population

After using the state-level variables to run the spatial analysis, not every statistically significant variable showed a pattern aligned with COPD within the state.

Run OLS analyses for West Virginia and New Jersey.

Regression

- West Virginia
- New Jersey

Feature Selections - West Virginia

Target Variable: COPD_AdjPrev

OLS Regression Results

Dep. Variable:	COPD_AdjPrev	R-squared:	0.989
Model:	OLS	Adj. R-squared:	0.984
Method:	Least Squares	F-statistic:	226.0
Date:	Mon, 13 May 2024	Prob (F-statistic):	4.58e-33
Time:	22:04:57	Log-Likelihood:	25.961
No. Observations:	55	AIC:	-19.92
Df Residuals:	39	BIC:	12.20
Df Model:	15		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	9.1614	2.897	3.163	0.003	3.303	15.020
CSMOKING_AdjPrev	0.1716	0.043	4.016	0.000	0.085	0.258
ACCESS2_AdjPrev	0.3095	0.088	3.508	0.001	0.131	0.488
LPA_AdjPrev	0.0741	0.032	2.281	0.028	0.008	0.140
BINGE_AdjPrev	-0.3943	0.068	-5.785	0.000	-0.532	-0.256
CHECKUP_AdjPrev	-0.0428	0.027	-1.586	0.121	-0.097	0.012
SLEEP_AdjPrev	0.0070	0.018	0.384	0.703	-0.030	0.044
stgh_fte_phys_dent_incl_nh_21	-0.0034	0.001	-3.901	0.000	-0.005	-0.002
Unemployment_rate_2021	0.0250	0.024	1.038	0.306	-0.024	0.074
R_INTERNATIONAL_MIG_2021	-0.1321	0.120	-1.104	0.277	-0.374	0.110
R_NET_MIG_2021	-0.0031	0.005	-0.624	0.536	-0.013	0.007
stnglth_fte_rn_incl_nh_21	0.0023	0.001	2.244	0.031	0.000	0.004
popn_est_21	8.618e-06	2.3e-06	3.740	0.001	3.96e-06	1.33e-05
stgh_resp_ther_ft_incl_nh_21	-0.0021	0.003	-0.720	0.476	-0.008	0.004
popn_bl_pct_20	-0.0605	0.014	-4.309	0.000	-0.089	-0.032

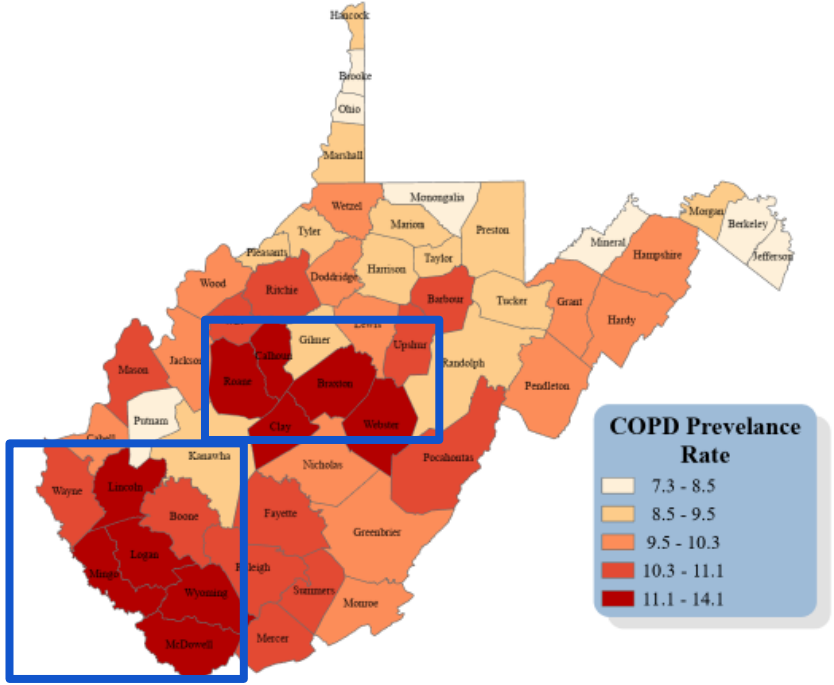
Statistically Significance:

- Smoking
- Lack of Health Insurance
- No Leisure-Time Physical Activity
- Binge Drinking
- Physicians
- Registered Nurses
- Total Population
- African American Population

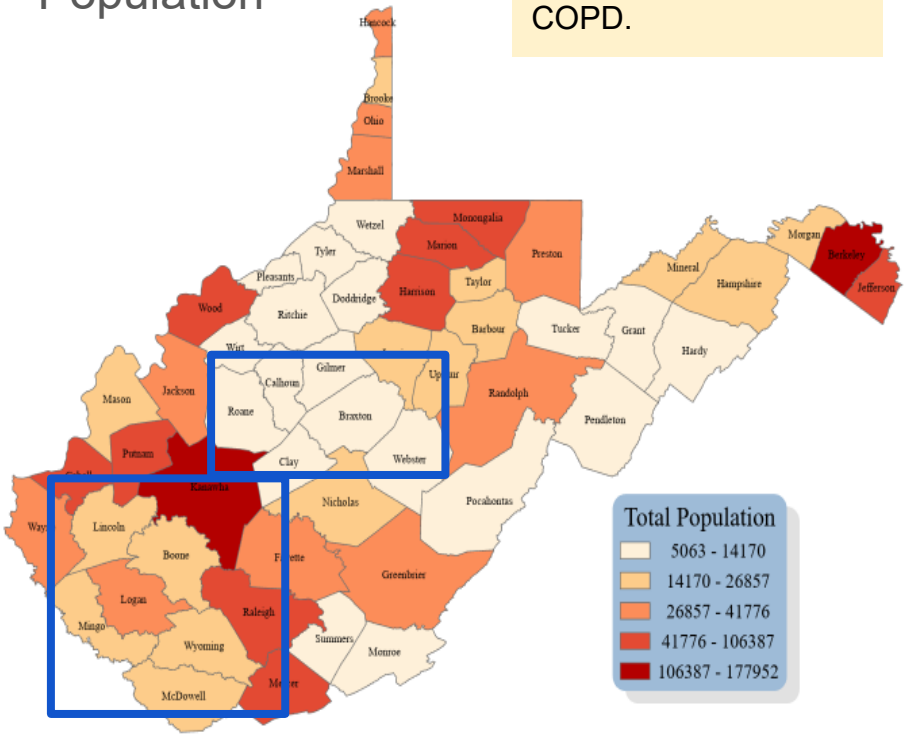
West Virginia Demographic Overview - Population

The lower the population, the higher the prevalence of COPD.

COPD Prevalence

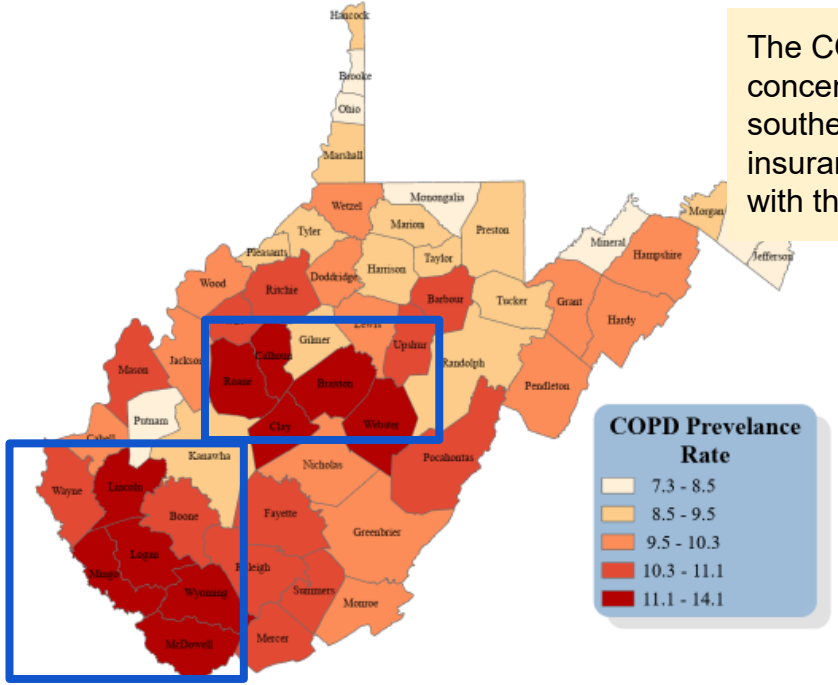


Population



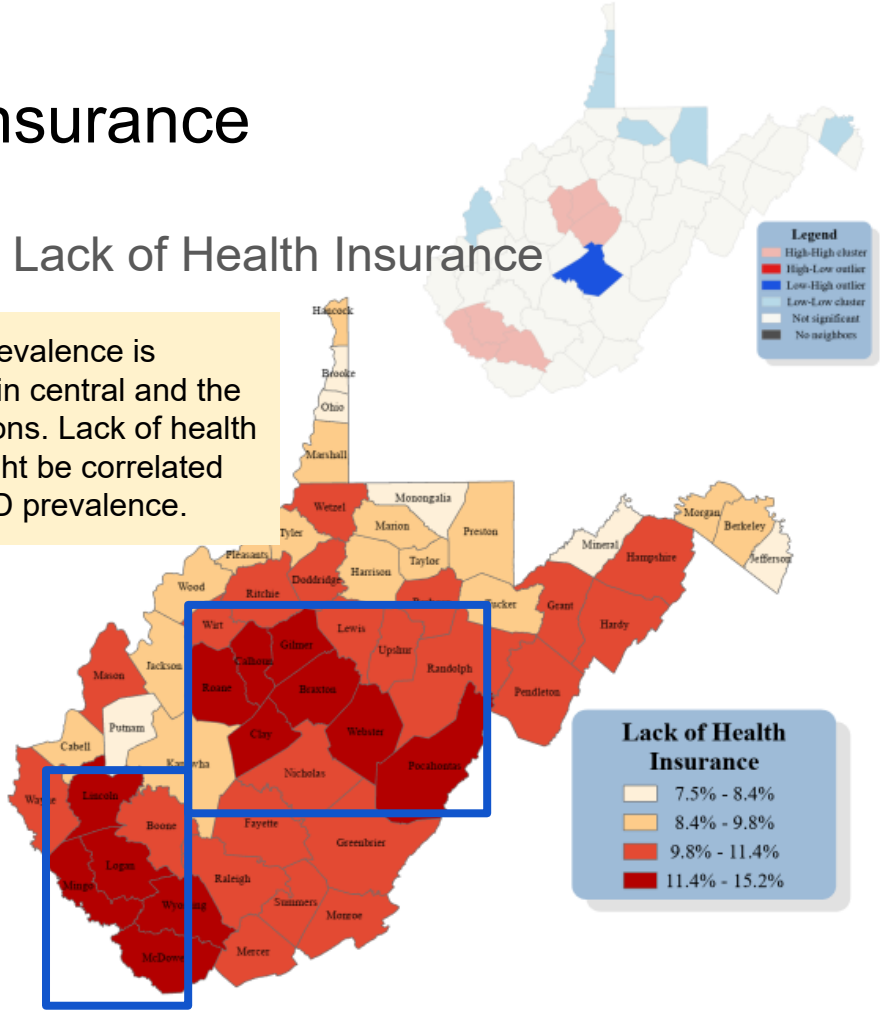
West Virginia - Lack of Health Insurance

COPD Prevalence Rate

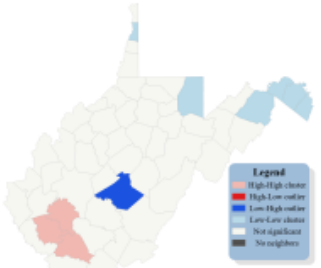


The COPD prevalence is concentrated in central and the southern regions. Lack of health insurance might be correlated with the COPD prevalence.

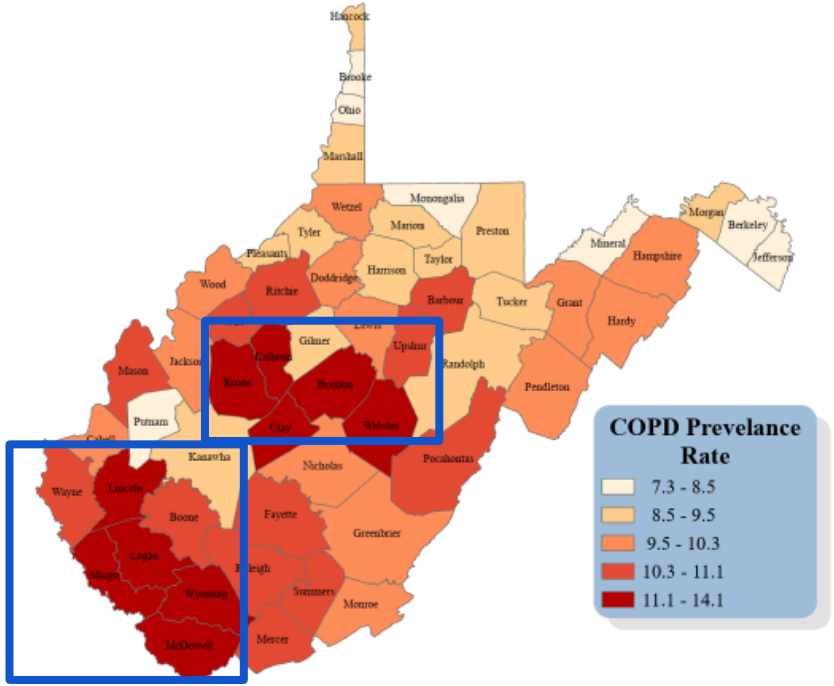
Lack of Health Insurance



West Virginia - No Leisure-Time & Physical Activity

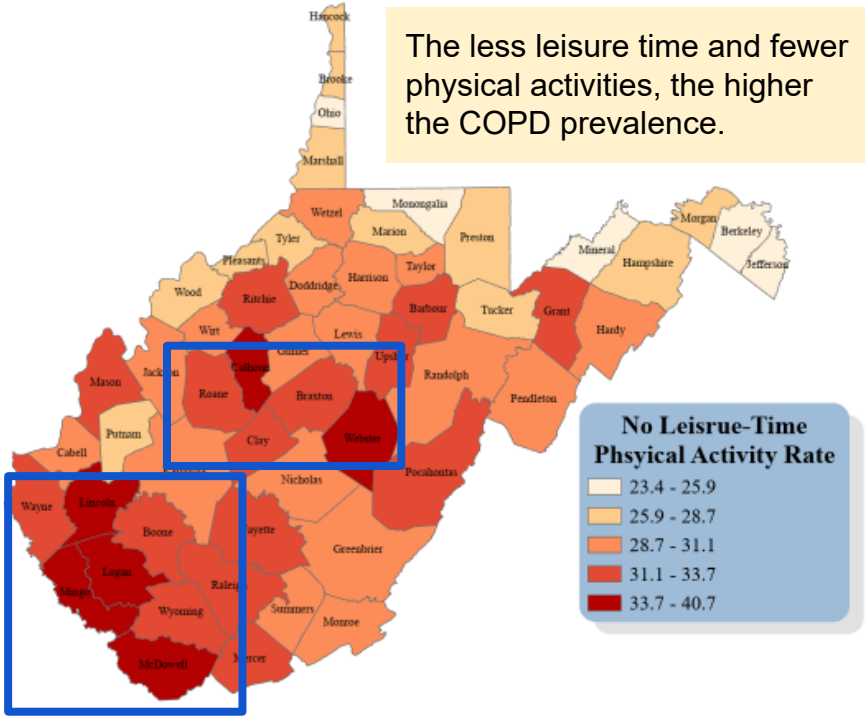


COPD Prevalence Rate



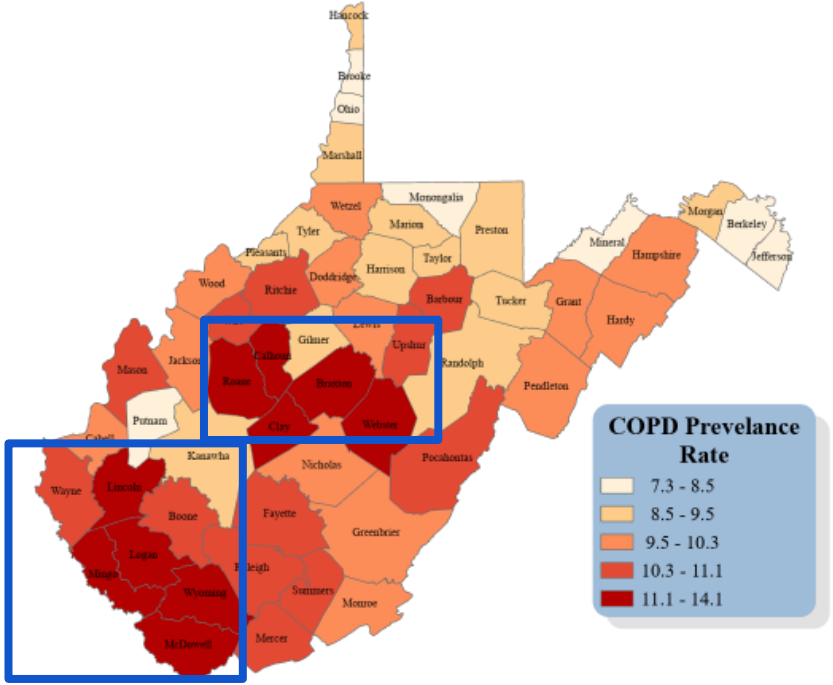
No Leisure-Time & Physical Activity

The less leisure time and fewer physical activities, the higher the COPD prevalence.



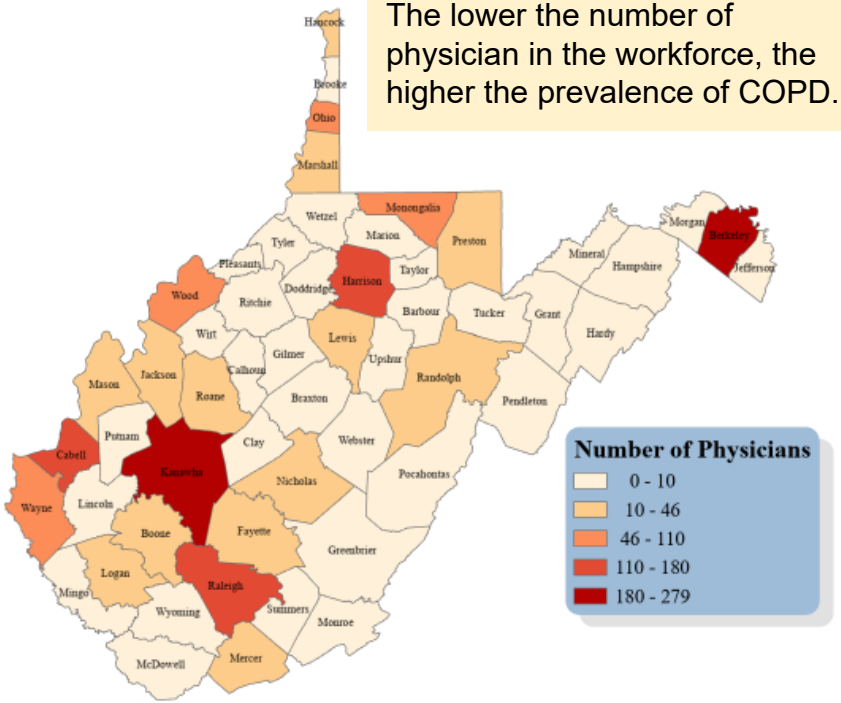
West Virginia - Physician Workforce

COPD Prevalence Rate



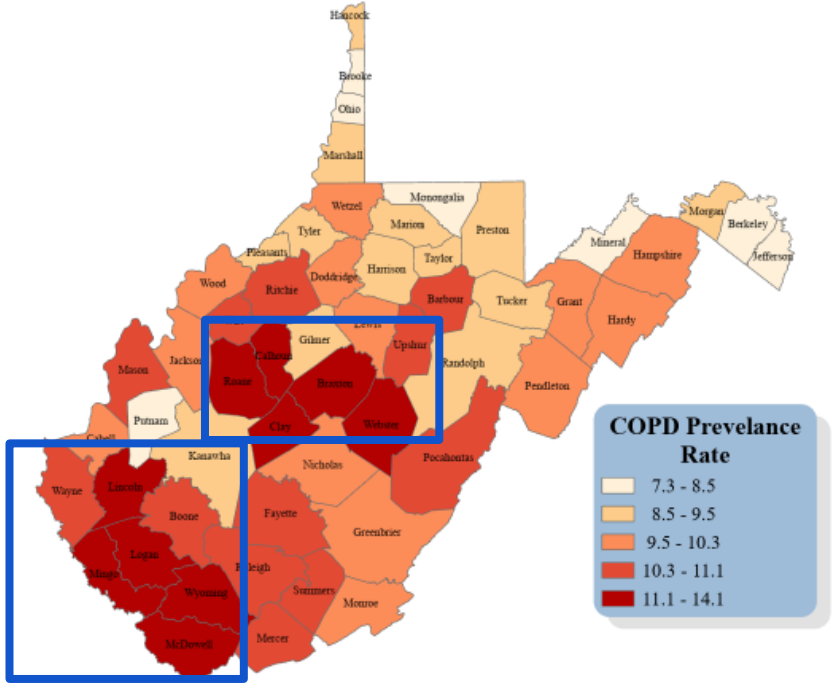
Physician Workforce

The lower the number of physician in the workforce, the higher the prevalence of COPD.

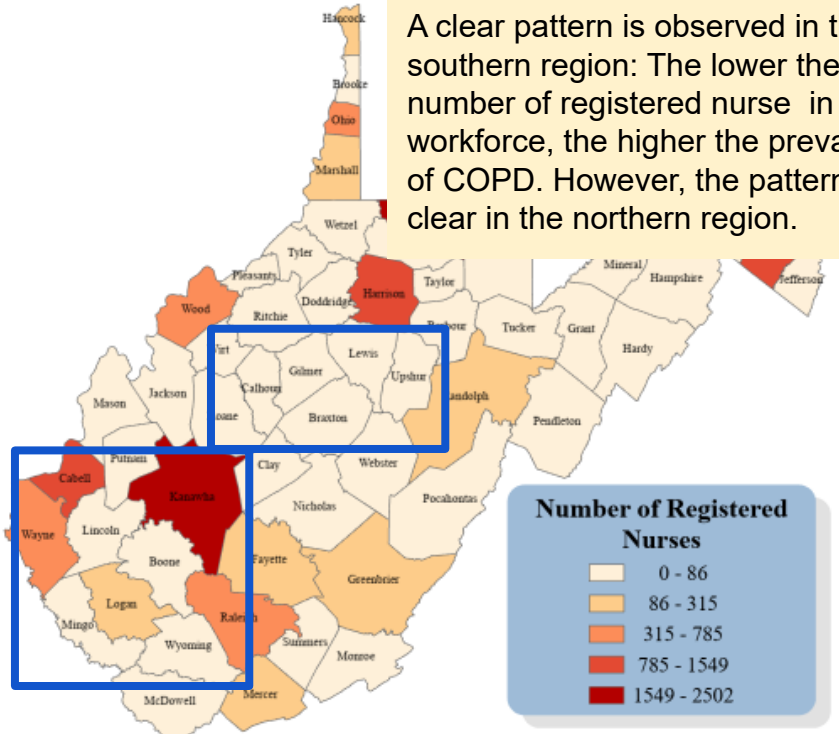


West Virginia - Registered Nurse Workforce

COPD Prevalence



Registered Nurse Workforce



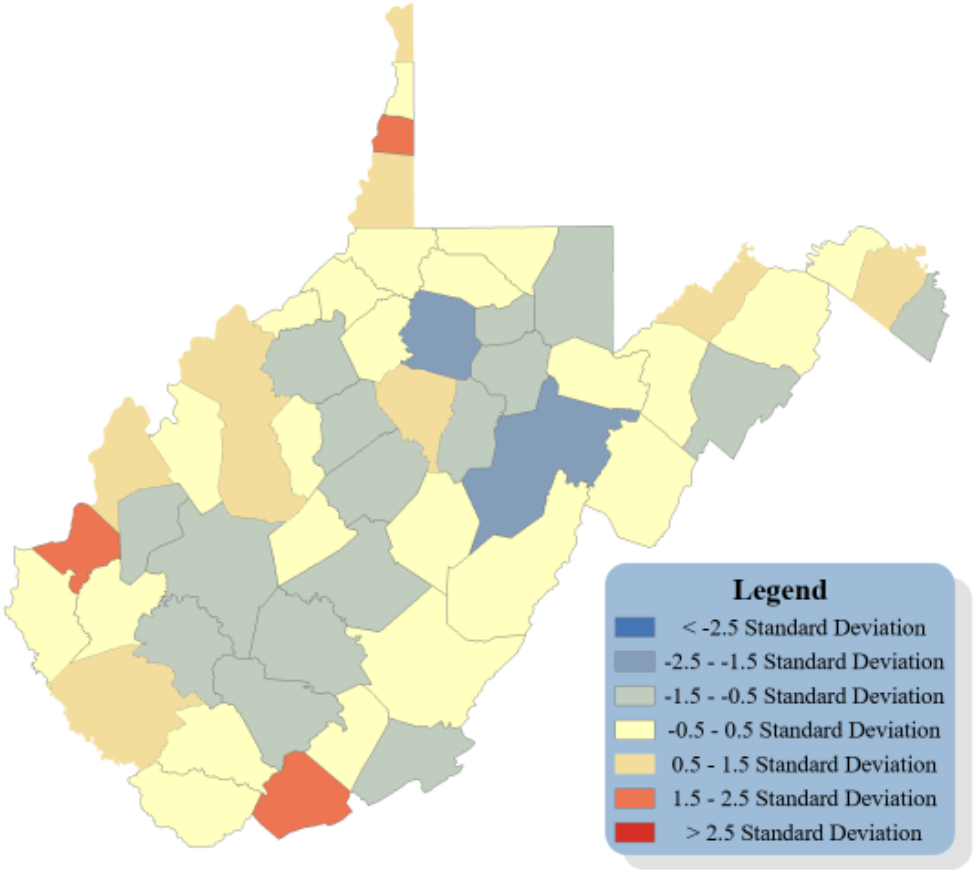
A clear pattern is observed in the southern region: The lower the number of registered nurse in the workforce, the higher the prevalence of COPD. However, the pattern is not clear in the northern region.

West Virginia - Standardized Errors Map

Dependent variable: COPD prevalence

Independent variables:

- Smoking
- Lack of Health Insurance
- No Leisure-Time & Physical Activity
- Binge Drinking
- Physicians
- Registered Nurses
- Total Population
- African American Population



Feature Selections - New Jersey

Target variable: COPD_AdjPrev

OLS Regression Results

Dep. Variable:	COPD_AdjPrev	R-squared:	0.974
Model:	OLS	Adj. R-squared:	0.966
Method:	Least Squares	F-statistic:	113.6
Date:	Mon, 13 May 2024	Prob (F-statistic):	2.26e-11
Time:	20:51:23	Log-Likelihood:	11.824
No. Observations:	21	AIC:	-11.65
Df Residuals:	15	BIC:	-5.381
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025
const	2.2089	1.117	1.977	0.067	-0.172
CSMOKING_AdjPrev	0.2122	0.044	4.878	0.000	0.119
LPA_AdjPrev	0.0346	0.017	2.089	0.054	-0.001
stnglth_fte_rn_incl_nh_21	0.0002	0.000	0.731	0.476	-0.000
Percent of adults with a bachelor's degree or higher, 2017-21	-0.0167	0.012	-1.404	0.181	-0.042
popn_asn_pct_20	-0.0225	0.008	-2.680	0.017	-0.040

Omnibus:	0.313	Durbin-Watson:	2.852
Prob(Omnibus):	0.855	Jarque-Bera (JB):	0.480
Skew:	-0.140	Prob(JB):	0.787

...

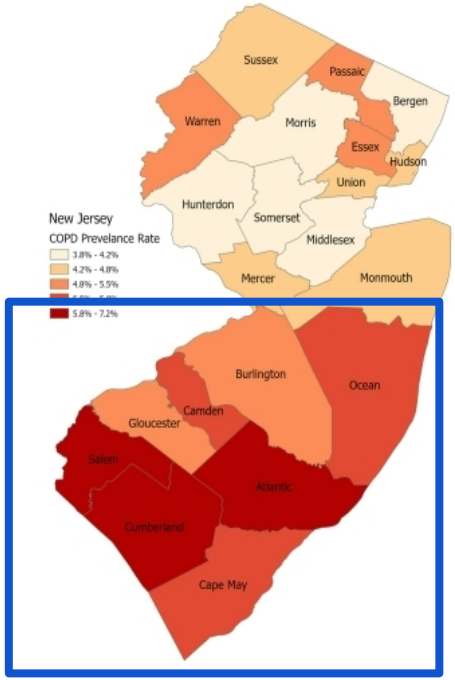
[2] The condition number is large, 6.43e+03. This might indicate that there are

Statistically Significance:

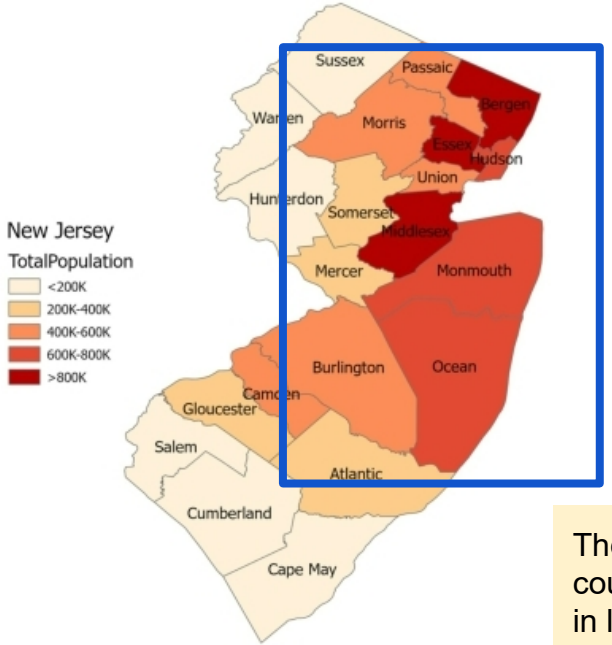
- Smoking
- Asian Population

New Jersey Demographic Overview: Population

COPD Prevalence



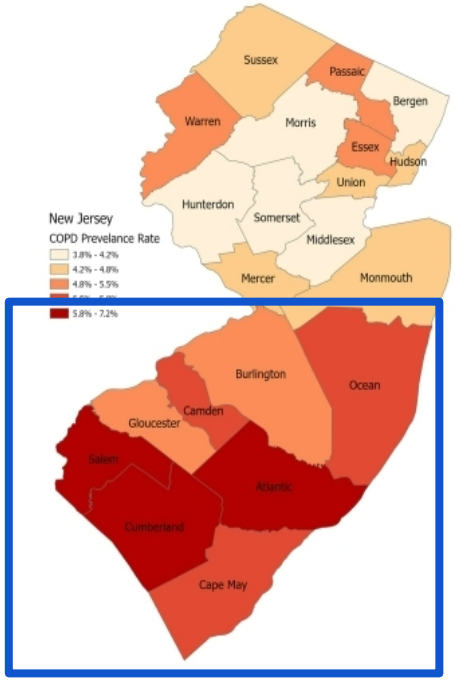
Population



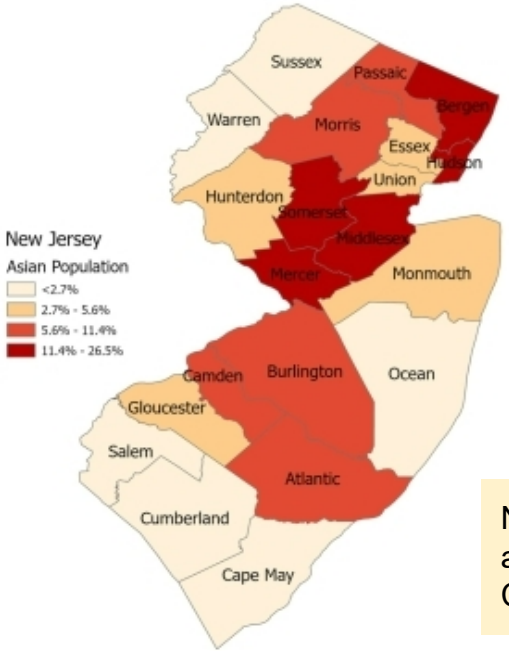
The higher rate of COPD county appears to locate in lower population area.

New Jersey - Asian Population

COPD Prevalence



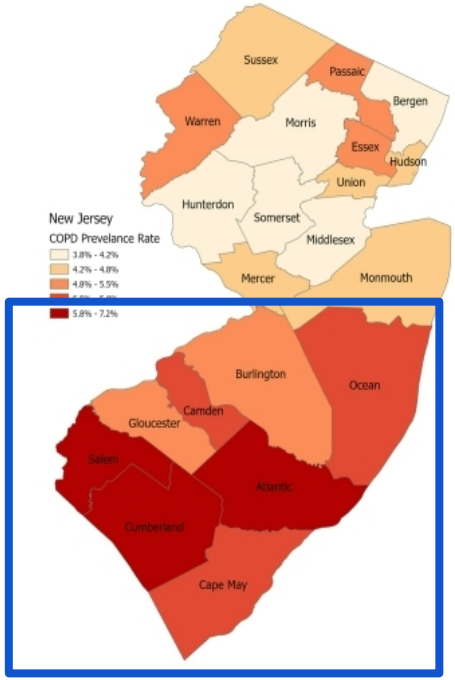
Asian Population



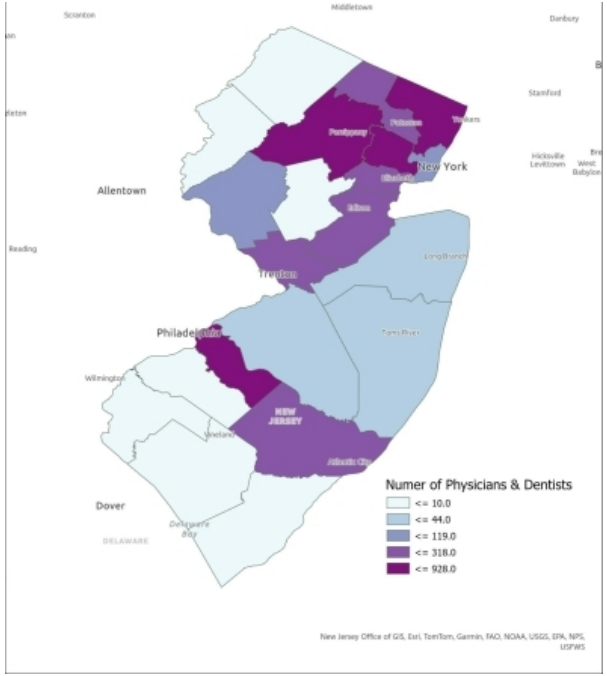
No clear pattern between asian population and COPD prevalence.

New Jersey - Physicians & Dentists, FT

COPD Prevalence

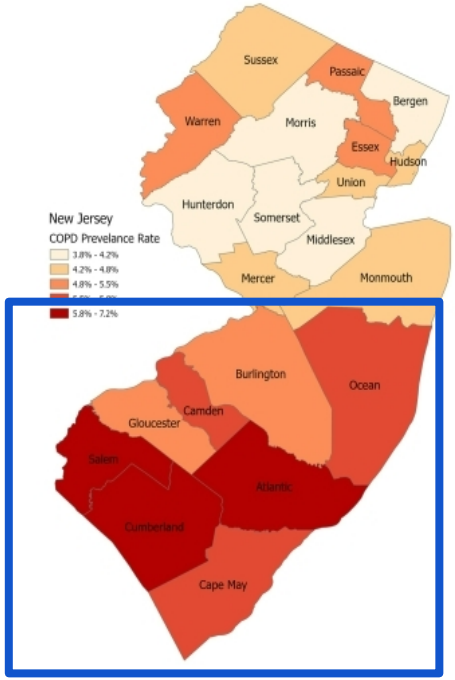


Physicians & Dentists

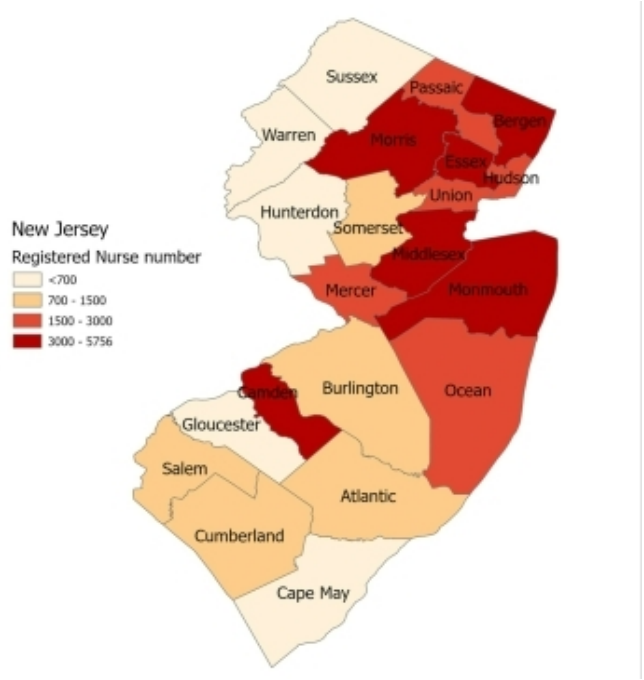


New Jersey - Registered Nurse Workforce

COPD Prevalence

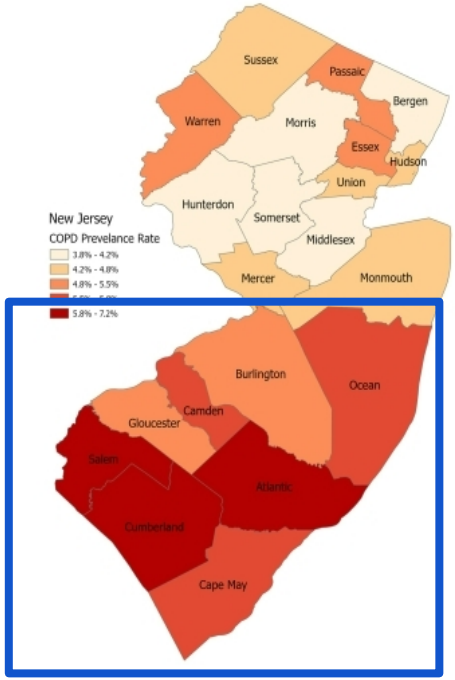


Registered Nurse number

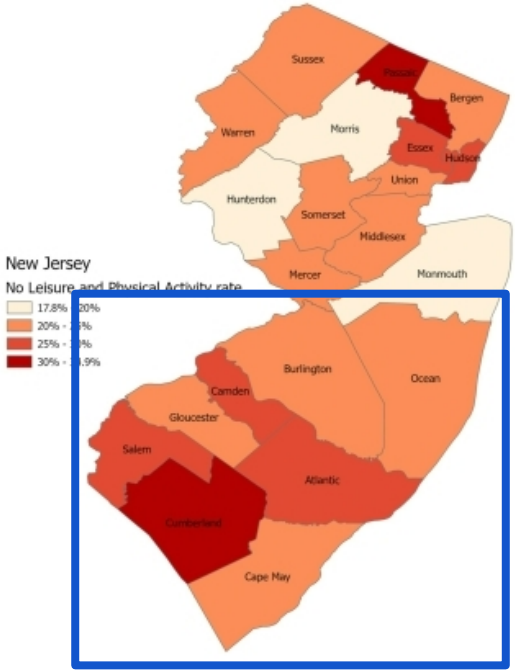


New Jersey: No Leisure Time and Physical Activity

COPD Prevalence



No Leisure Time and Physical Activity



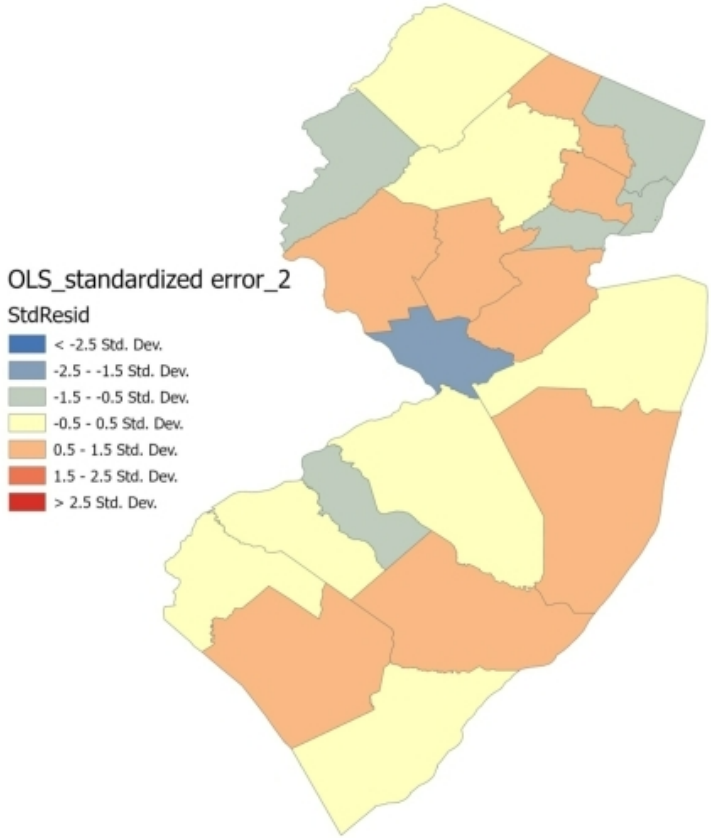
Two pattern are highly similar.

New Jersey- Standardized Errors Map

Dependent variable: COPD prevalence

Independent variables:

- Smoking
- Binge Drinking
- Sleep
- White Population
- African American Population
- Asian Population



Key Findings

The following findings are found in both West Virginia and New Jersey.

- There’s a strong correlation between COPD prevalence and smoking behavior.
- COPD is more prevalent in urban/highly populated area than rural area
- The less leisure time and fewer physical activities, the higher the COPD prevalence.

Policy Suggestion

Smoking Cessation Programs:

- **Increase Access to Smoking Cessation Resources**: Provide more resources such as counseling, support groups, and smoking cessation medications. These resources should be easily accessible and affordable to encourage more people to quit smoking.
- **Public Awareness Campaigns**: Launch public health campaigns to educate the population about the dangers of smoking and the benefits of quitting. Tailor these campaigns to specific demographics to maximize their effectiveness.

Region Targeted Health Strategies:

- **Targeted Healthcare Services**: Provide targeted healthcare services in rural areas where COPD prevalence is higher. This could include mobile health clinics, telehealth services, and community health worker programs.

Promotion of Physical Activity and Leisure Time:

- **Physical Activity Programs**: Implement programs that encourage physical activity, such as community fitness classes, walking groups, and sports leagues. Ensure these programs are accessible to all age groups and physical abilities.
- **Work-Life Balance Initiatives**: Encourage employers to adopt policies that promote work-life balance, allowing individuals more leisure time to engage in physical activities. This could include flexible working hours, remote work options, and mandatory break periods.

Data Source

Raw Data:

- data.HRSA.gov, Area Health Resources:

  https://data.hrsa.gov/data/download?hmpgtitle=hmpg-hrsa-data

- Economic Research Service, U.S. DEPARTMENT OF AGRICULTURE:

  https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/

- PLACES: County Data (GIS Friendly Format), 2023 release:

  https://data.cdc.gov/500-Cities-Places/PLACES-County-Data-GIS-Friendly-Format-2023-releas/i46a-9kgh/about_data

Paper Reference:

- https://www.cdc.gov/mmwr/volumes/72/wr/mm7246a1.htm
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10684355/