Assessing Environmental Inequality in Air Pollution Exposure

Preliminary Analysis for Final Research Paper

Ayaka Naganuma
University of Ottawa

November 26, 2025

Introduction

Background: Environmental Inequality

Lower socioeconomic status and racialized communities often face higher pollution exposure. In Canada, evidence exists but is fragmented and inconsistent across cities.

Research Gap in Canadian Context

- Canadian evidence on socioeconomic inequality in pollution exposure existed but was fragmented.
- Existing studies rarely control for key physical confounders (population density, industrial activity, road networks).
- No prior Canadian study has tested whether income and racial composition jointly shape pollution exposure.

Hypotheses

- **H1: Socioeconomic Inequality**

$$\text{Income} \downarrow, \ \text{Education} \downarrow, \quad \Longrightarrow \quad \text{PM}_{2.5} \uparrow$$

- **H2: Robust SES Effects (after controls)**

$$\text{SES} \Longrightarrow \text{PM}_{2.5} \quad \text{even after controlling for:} \begin{cases} \text{Industrial emissions} \\ \text{Major roads} \\ \text{Population density} \end{cases}$$

- **H3: Compounded Vulnerability (Interaction)**

$$\text{Income} \downarrow \ \times \ \text{Minority Share} \uparrow \Longrightarrow \quad \text{PM}_{2.5} \uparrow\uparrow$$

Data and Variables

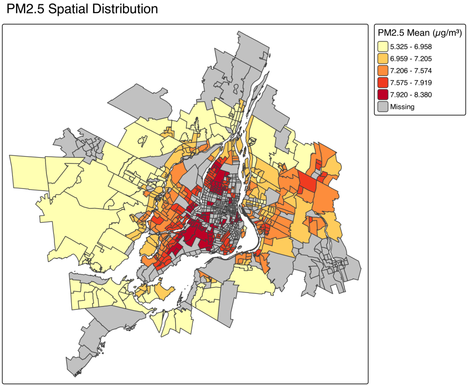
Data sources

- Census 2016: median income, education rate (share with university degree), visible minority share, population density
- NASA $PM_{2.5}$ (2016–2018): satellite-based annual average concentrations
- NPRI (2016–2018): industrial emissions (facility-level)
- Road Network File 2016: major road density

Study area: Montréal

- 970 Census Tracts in the original dataset
- After removing records with missing values: **478 complete CTs used for analysis**

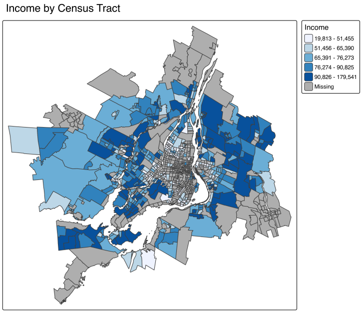
Explanatory Analysis: Spatial Distribution of $PM_{2.5}$:



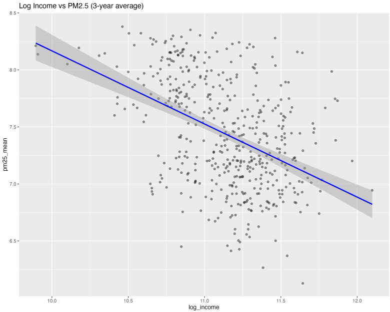
Key Observation

$PM_{2.5}$ concentrations are highest in the central and eastern parts of Montréal, showing a clear spatial gradient across the city.

Explanatory Analysis: Income and $PM_{2.5}$



Median Income (Map)

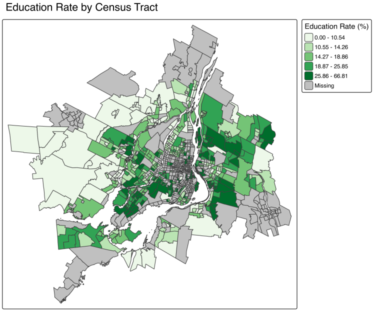


Income vs. $PM_{2.5}$ (Scatter)

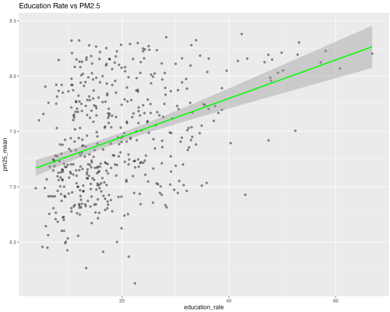
Key Observation

Lower-income Census Tracts tend to have higher $PM_{2.5}$ concentrations.

Explanatory Analysis: Education and $PM_{2.5}$



Education Rate (Map)

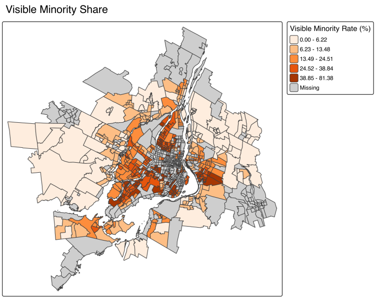


Education Rate vs. $PM_{2.5}$ (Scatter)

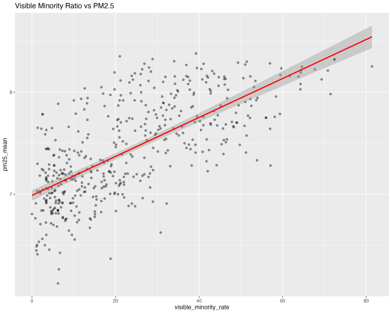
Key Observation

Higher education areas show higher $PM_{2.5}$, but this reflects urban core concentration of highly educated residents rather than SES disadvantage..

Explanatory Analysis: Visible Minority and $PM_{2.5}$



Visible Minority Share (Map)

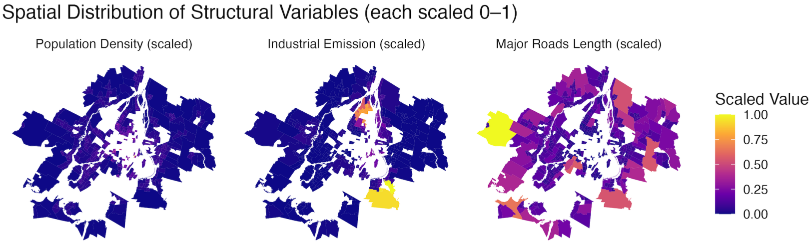


Visible Minority vs. $PM_{2.5}$ (Scatter)

Key Observation

Areas with a higher share of visible minority residents tend to have systematically higher $PM_{2.5}$ exposure.

Explanatory Analysis: Spatial Distribution of Structural Variables



Interpretation

Population density, road density and industrial activity serve as essential physical confounders.

OLS Model Specification

Baseline model (H1):

$$PM_{2.5,i} = \beta_0 + \beta_1 Income_i + \beta_2 EducationRate_i + \beta_3 MinorityShare_i + \epsilon_i$$

(Income and education capture socioeconomic status; MinorityShare represents racial composition. Education rate may function as a proxy for urban core structure rather than SES.)

Extended model with controls (H2):

$$+\beta_4 Industrial_i + \beta_5 RoadDensity_i + \beta_6 PopDensity_i$$

Interaction model (H3):

$$+\beta_7 (Income_i \times MinorityShare_i)$$

A log-income version of this model is also estimated.

OLS Results

	SES only	Full OLS	Interaction	Log OLS
(Intercept)	7.153 ***	7.200 ***	7.234 ***	10.480 ***
	(0.068)	(0.077)	(0.090)	(1.092)
income	-0.000 ***	-0.000 ***	-0.000 ***	
	(0.000)	(0.000)	(0.000)	
education_rate	0.015 ***	0.014 ***	0.014 ***	0.013 ***
	(0.002)	(0.002)	(0.002)	(0.002)
visible_minority_rate	0.014 ***	0.013 ***	0.011 ***	-0.051
	(0.001)	(0.001)	(0.003)	(0.027)
total_industrial_emission		0.000	0.000	
		(0.000)	(0.000)	
major_road_km		-0.003 ***	-0.003 ***	
		(0.001)	(0.001)	
pop_density		-0.000	0.000	
		(0.000)	(0.000)	
income:visible_minority_rate			0.000	
			(0.000)	
log_income				-0.334 ***
				(0.094)
log_major_road_km				-0.082 *
				(0.037)
log_pop_density				0.054 *
				(0.022)
log_total_industrial_emission				-0.007
				(0.006)
log_income:visible_minority_rate				0.006 *
				(0.002)
N	478	478	478	478
R2	0.562	0.583	0.583	0.592
Adj_R2	0.560	0.577	0.577	0.586
AIC	266.480	249.819	251.285	241.256
logLik	-128.240	-116.909	-116.642	-111.628

*** p < 0.001; ** p < 0.01; * p < 0.05.

Model Diagnostics: Why OLS is Not Enough

VIF (Model 2)

- income = 2.15; education = 1.47; minority = 1.79
- industrial = 1.03; roads = 1.41; density = 2.26
- $\rightarrow$ No multicollinearity concerns (all < 3)

RESET Test (Model 2)

- RESET statistic = 17.267, df = (2, 468), p-value = 5.817e08
- $\rightarrow$ Functional form misspecification detected

Moran's I for $PM_{2.5}$

- Moran's I = 0.89 ($p < 0.001$)
- $\rightarrow$ Very strong spatial autocorrelation

Implication

OLS residuals violate key assumptions (functional form + spatial independence). $\rightarrow$ OLS estimates may be biased, motivating the use of a Spatial Error Model (SEM).

Spatial Error Model (SEM): Specification and Results

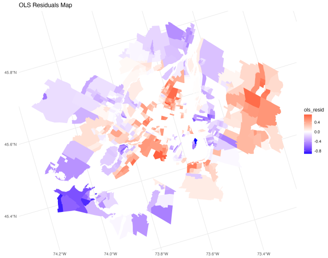
	SES
(Intercept)	7.715 ***
	(0.555)
log_income	-0.092
	(0.048)
visible_minority_rate	-0.025
	(0.014)
education_rate	0.005 ***
	(0.001)
log_major_road_km	0.019
	(0.016)
log_pop_density	0.062 ***
	(0.009)
log_total_industrial_emission	0.006
	(0.006)
log_income:visible_minority_rate	0.002
	(0.001)
lambda	0.921 ***
	(0.011)
N	478
R2	0.932
AIC	-409.406
logLik	214.703

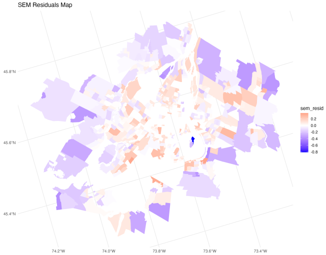
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Spatial Diagnostics: SEM Corrects Spatial Dependence

Spatial Autocorrelation (Residuals)

- OLS residual Moran's I $= 0.89$ ($p < 0.001$)
- SEM residual Moran's I $= 0.076$ ($p = 0.004$)





Interpretation

SEM greatly reduces spatial dependence in the residuals, indicating that the model successfully accounts for spatial error processes.

Conclusion

Conclusions

- **H1: Not supported.** Income is not significant in the SEM, and education's positive effect reflects urban-core density rather than SES inequality.
- **H2: Not supported.** After controlling for density, roads, and industry, SES variables lose significance. $PM_{2.5}$ variation is driven mainly by urban structure.
- **H3: Weak support.** The income $\times$ minority interaction appears only in the log-OLS model and disappears in SEM.
- **Population density is the strongest predictor of $PM_{2.5}$ $\rightarrow$ urban structure is central to explaining pollution hotspots.**

Additional Insights

- SEM drastically reduces spatial autocorrelation $\rightarrow$ confirming that spatial models are essential for valid inference.

Limitations and Next Steps

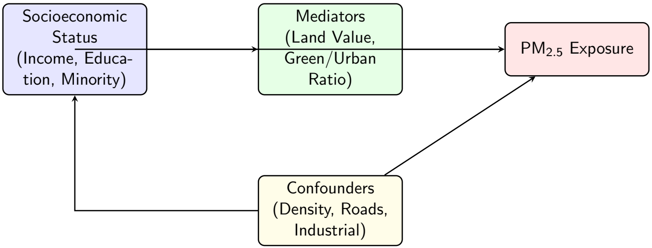
Limitations

- Findings are not yet generalizable across Canada.
- NO_2 and other pollutants are not included.
- Several Census Tracts were removed due to missing data, which may introduce spatial selection bias.
- Industrial emissions are aggregated at the tract level and may underrepresent point-source variation.

Next Steps

- Refine the SES specification in the final paper by reclassifying education as an indicator of urban spatial structure, rather than treating it as a core SES component.
- Explore whether mediators (e.g., greenness, land value, urban form) statistically amplify SES-based exposure disparities.
- Extend spatial analysis to other major cities in Canada.

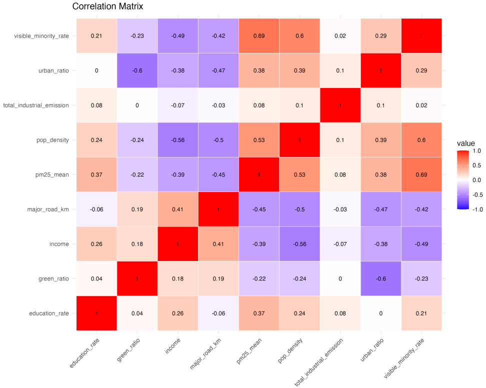
Appendix: SES, Confounders, and Mediators



Interpretation

- SES influences both neighborhood sorting and pollution exposure.
- Land value and green/urban ratio lie on the SES $\rightarrow$ PM pathway (mediators).
- Density, roads, and industrial activity affect both SES sorting and PM levels (confounders).

Appendix: Correlation Structure



- Socioeconomic variables correlate strongly with $PM_{2.5}$.
- Population density shows strong positive correlation (urban core effect).
- Green/urban ratios closely track SES $\rightarrow$ likely **mediators**.

Appendix: Spatial Error Model (SEM)

SEM structure

$$y = X\beta + u, \qquad u = \lambda W u + \varepsilon$$

- y: Outcome ($PM_{2.5}$).
- X: Observed predictors (log-income, education, minority share, roads, population density, industry).
- β: Effect of each predictor (similar to OLS).
- W: Spatial weights matrix (4-nearest-neighbors).
- u: Spatially correlated error term capturing unobserved spatial processes.
- λ: Spatial error dependence ($\lambda = 0.921$).
- ε: i.i.d. random noise.

Intuition

SEM corrects for unobserved spatial factors that violate OLS assumptions. It removes spatial clustering in residuals and produces unbiased estimates.