ECON 0150 Final Project

Exposure to PM2.5 air pollution vs. GDP per Capita

Name: Kurt Habazin, Jay Zheng, Darren Cen

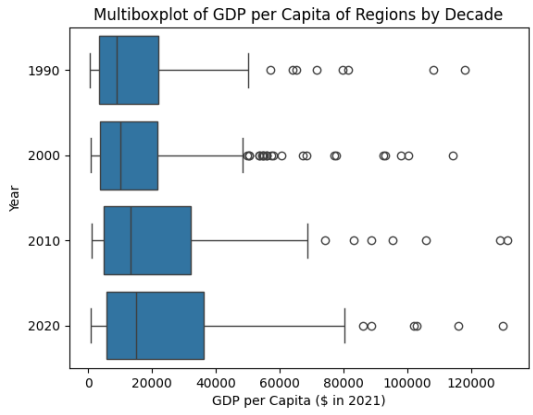
Introduction:

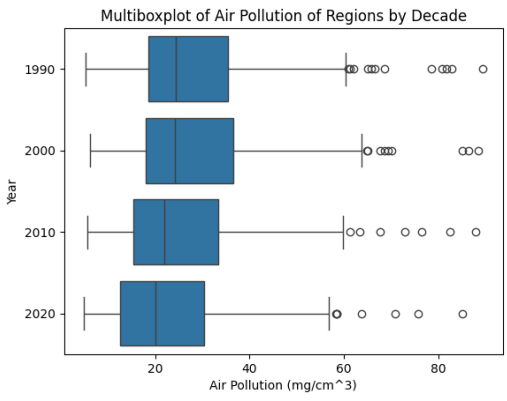
GDP per capita measures the average economic output per person in a country and can also be used to calculate the average standard of living. Lower standards of living are often associated with lower GDP per capita, worse health outcomes, and a lack of basic needs like clean water and food, as well as poor infrastructure and safety measures. Given these associations with lower GDP per capita, we thought it would be interesting to examine whether there is a specific relationship between poor air quality and GDP per capita. As a result, the research question that was brought up in our group was whether GDP correlates with air pollution across regions.

Data Description:

(μg/m³), Micrograms per cubic meter are the units used to measure the concentration of pollutants in the air. PM2.5 stands for particulate matter that is 2.5 micrometers or smaller in diameter. Thus, the "exposure to PM2.5 pollution" predictor variable measures the concentration of pollutants that are 2.5 micrometers or smaller in the air. The particle size, 2.5 micrometers or less, is important because it can penetrate the lungs and enter the bloodstream, causing serious health problems. The PM2.5 pollution data comes from the Global Burden of Disease Study and the Institute for Health Metrics and Evaluation and is processed by Our World in Data. The data was last updated on September 8th, 2025, and ranges from the years 1990-2020. Our predictor variable, GDP per capita, measures the average economic output per person in a country or region per year, adjusted for inflation and the differences in living costs between countries or regions. The data is expressed in constant international dollars, which is a hypothetical currency that is more advanced and accurate for comparing purchasing power than standard market exchange rates.

<u>Data Descriptives on the Statistical Distribution:</u>
Since we are running panel data with cross-sectional variables (GDP per capita and air pollution in different regions) across various periods, we develop a multi-box plot of figures separately between GDP per capita and air pollution to look at the distribution visually over four decades from 1990-2020;
*Based on the multi-boxplot, there seems to be a trend where GDP per capita tends to be increasing compared to the Air population per decade relative to the median and range.



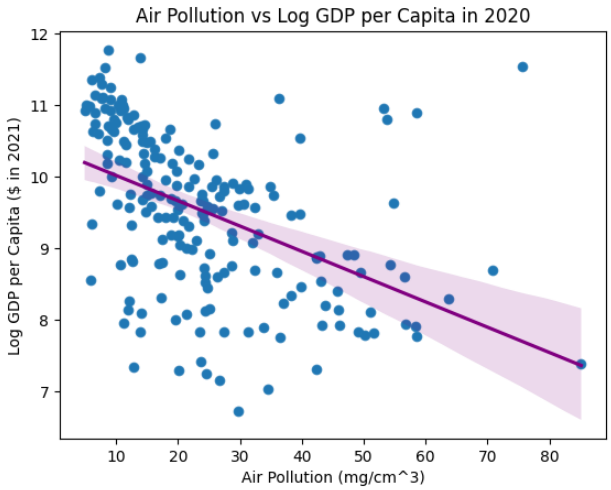


Methodology:

Regression Equation: Log of GDP per Capita = Beta_0+Beta_1*(air pollution)+epsilon (ε)
We estimated the linear relationship between log GDP per capita and exposure to air pollution (μg/m³) using a general linear model with a sample t-test. We used log GDP per capita as the dependent variable and air pollution (μg/m³) as the predictor variable. The general linear model compares how much GDP per capita changes for each additional unit of PM2.5 air pollution. To visualize the data, we used a scatterplot with a linear regression line to represent the relationship. Each dot on the scatterplot represents a country, including its GDP per capita and its average annual exposure to PM2.5 air pollutants. The downward slope of the regression line tells us that as exposure to PM2.5 air pollution increases, GDP per capita decreases. This is consistent with our general linear model, for every unit increase of PM2.5 air pollution, GDP per capita decreases by 6.3265 units.

Results and Analysis

Based on the individual regression test, we observed a p-value of nearly zero from both the intercept and slope from all four decades, raising statistical significance (reject our null that beta 1=0). For the intercept, this indicates that if the mean parameter of intercept is zero is true, the percentage of obtaining an extreme value of an intercept to this value is around 0%; highly confident that the air pollution would never be at zero when the log of GDP is zero. Additionally, if the parameter mean for the slope being zero is true, the probability we get an extreme value of a slope to this value is approximately 0% (0.000….); we are confident that the log GDP and air pollution slope is not around zero, indicating there is a correlation between the two variables. For example, filtering the year 2020, the coefficient value of 84.1976 for the intercept tells us that when log GDP per capita equals zero, we can expect to see 84.1976 units of PM2.5 air pollutants. The coefficient value of -6.3265 for the log_gdp predictor variable indicates that for every one-unit increase in exposure to PM2.5 air pollutants, log GDP per capita decreases by 6.3265 units with a p-value of 1.021743e-12 (approximately zero). Our scatterplots show slight curvature in the datapoints with some downward trend. However, near the 20(μg/m³) air pollution point, the data points don't show uniformity.




```
#regression test of 2020
model = smf.ols(formula='atm ~ log_gdp', data=data_2020)
print(model.fit().pvalues)
model.fit().summary()
```

```
Intercept    4.895539e-21
log_gdp      1.021743e-12
dtype: float64
```

OLS Regression Results

Dep. Variable:	atm	**R-squared:**	0.224
Model:	OLS	**Adj. R-squared:**	0.220
Method:	Least Squares	**F-statistic:**	57.96
Date:	Thu, 11 Dec 2025	**Prob (F-statistic):**	1.02e-12
Time:	20:38:58	**Log-Likelihood:**	-812.34
No. Observations:	203	**AIC:**	1629.
Df Residuals:	201	**BIC:**	1635.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	84.1976	7.969	10.566	0.000	68.484	99.911
log_gdp	-6.3265	0.831	-7.613	0.000	-7.965	-4.688

Conclusion

Our model does not guarantee that we can be 100% certain that there is a negative relationship between exposure to PM2.5 air pollution and GDP per capita, as there could be other confounding factors coming into play, depending on the unknown mechanism by which the data collected was executed, that were not tested, that contribute to the relationship shown in our model. In addition, there was still some slight curvature when transforming the logarithmic form of GDP per capita in the scatterplot; the linear regression model may not be accurate in addressing those curvatures. However, based on the experiment, it can be concluded that there is a strong correlation between GDP per capita and exposure to PM2.5 air pollutants, given a statistical significance of rejecting the null that log GDP per capita is not correlated with air pollution.

References

Global Burden of Disease Study (GBD 2021) - Institute for Health Metrics and Evaluation (IHME), via World Bank (2025); Eurostat, OECD, IMF, and World Bank (2025)
Link: https://ourworldindata.org/grapher/pm25-exposure-gdp