How Life Expectancy is Impacted by Air Quality

Haisu Huang *University of Oregon, Sociology*

Introduction

Previous studies have examined the relationship between exposure to air pollution and its impact on health. A few studies on population scale influenced by air pollution have focused on consequences such as lower birth weight and higher death rate. But little research explores the direct association between air pollution and human suffering. I intended to use cancer data to indicate human suffering, however, due to lack of resources and limitation on time searching for such resources, I replaced the measurement of human suffering with life expectancy. This paper is to demonstrate how I conduct the statistical analysis of the relationship beween air quality and life expectancy.

Data and Methods

The data for this project is from the world development indicators. The dataset includes the following 12 variables with brief descriptions: 1. country name: all countries are selected 2. country code 3. time: year from 2008 to 2018 4. time code 5. pm2.5: air pollution, mean annual exposure (micrograms per cubic meter) 6. mortality rate: attributed to household and ambient air pollution, age-standardized (per 100,000 population) 7. life expectancy: at birth, total (years) 8. co2 emissions: metric tons per capita 9. gdp per capita: (current US$) 10. gdp per capita growth: (annual %) 11. urban population: (% of total) 12. mortality: from CVD, cancer, diabetes or CRD between exact ages 30 and 70 (%)

After studying the dataset, I processed data cleaning. First, from row 218 to 267, the geographical locations are regions, not individual countries. These rows do not contribute to my study so they were deleted. Second, the years of 2008, 2009 and 2018 do not contain valuable data on particular matters 2.5 (PM2.5). Since PM2.5 is a very important variable for air quality, these years were deleted for the convenience of analysis. Third, the years after 2014 also were deleted because they do not contain data on CO2 emission which is another important variable to indicate air quality. Due to missing values in several year variables, CO2 emission and PM2.5, this dataset can only interpret the relationship between life expectancy and air quality within a limited scope.

After the dataset is cleaned, I conducted 5 imputations and adjusted weight and cluster in order to address the design effect in the dataset. I used logistics regression models to analyze the relationships between each independant variable and life expectancy.

```
## -- Imputation 1 --
##
##   1  2  3  4  5  6  7  8
##
## -- Imputation 2 --
##
##   1  2  3  4  5  6  7  8  9 10
##
## -- Imputation 3 --
##
```

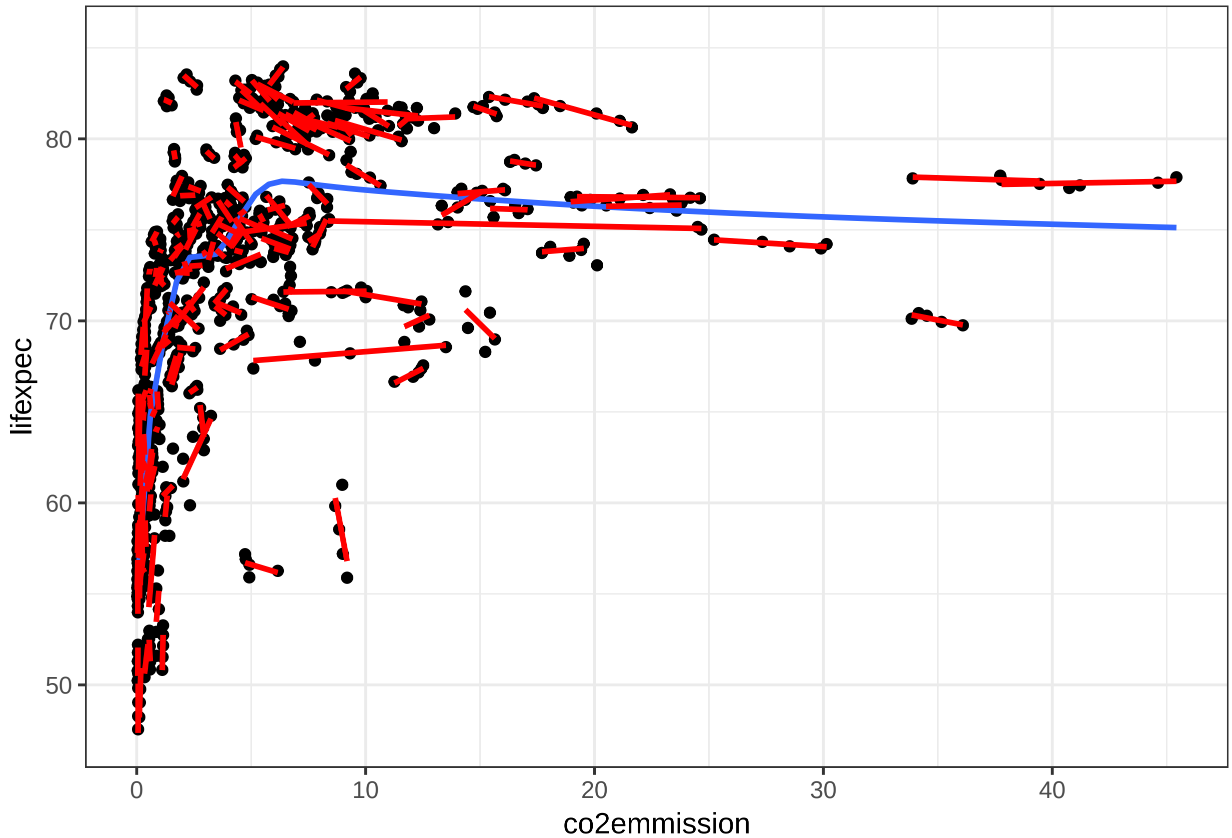

Figure 1. Relationship between CO2 Emission and Life Expectancy

```
##   1  2  3  4  5  6  7  8  9 10 11
##
## -- Imputation 4 --
##
##   1  2  3  4  5  6  7
##
## -- Imputation 5 --
##
##   1  2  3  4  5  6  7  8  9 10
```

Results

In the process of analysis, I examined the relationship between each variable and life expectancy. I also synthesized a few variable combinations to see how life expectancy is impacted. Both CO2 emission and GDP per capita seem to have similar positive effects on life expectancy, while PM2.5 shows some negative association with life expectancy.

Since both CO2 emission and PM2.5 indicate the quality of air, I sythesized these two variable and examined the relationship between this sythetic scale with life expectancy.

I also respectively synthetized (1) PM2.5 and GDP per capita (2) CO2 emission and GDP per capita (3) PM2.5, CO2 emission and GDP per capita. All three charts indicate similar trends of life expectancy with changes of the independant variables. One thing notable is the outlier with high

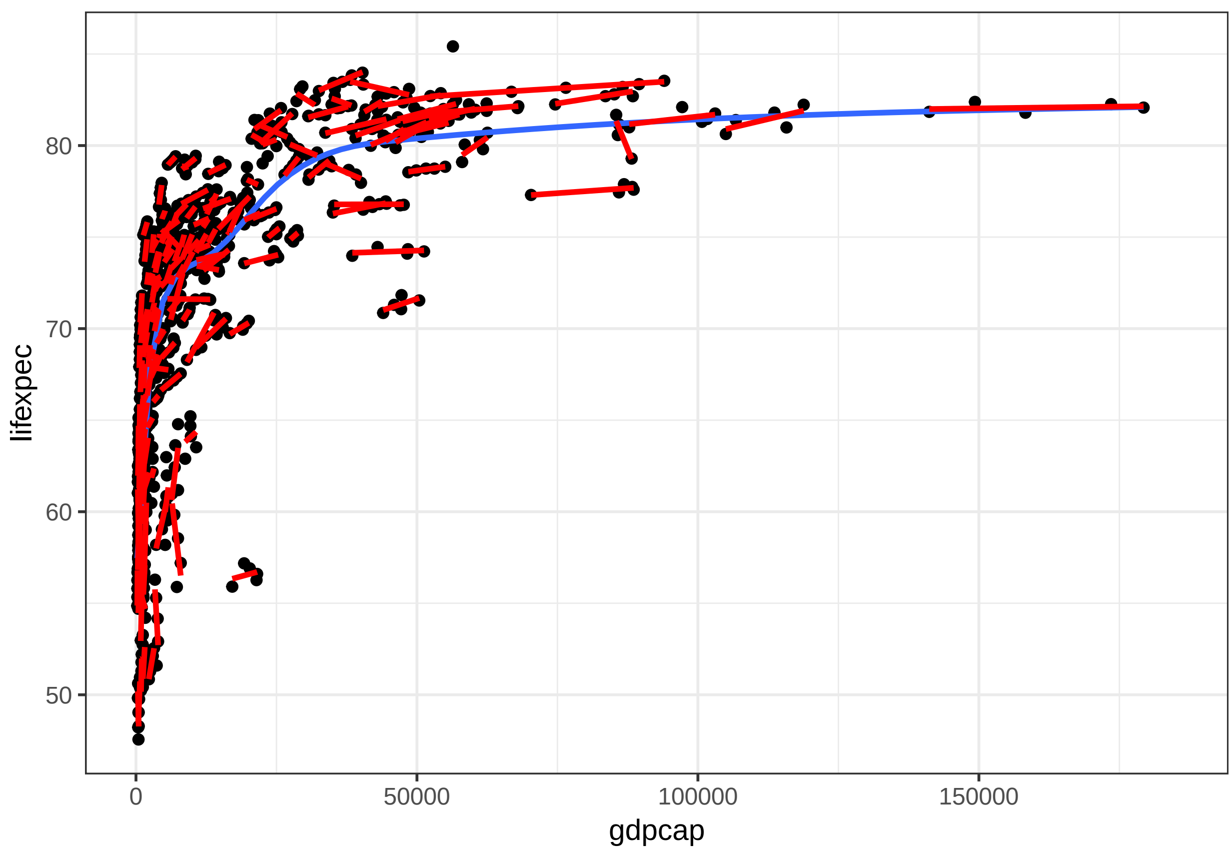

Figure 2. Relationship between GDP per capita and Life Expectancy

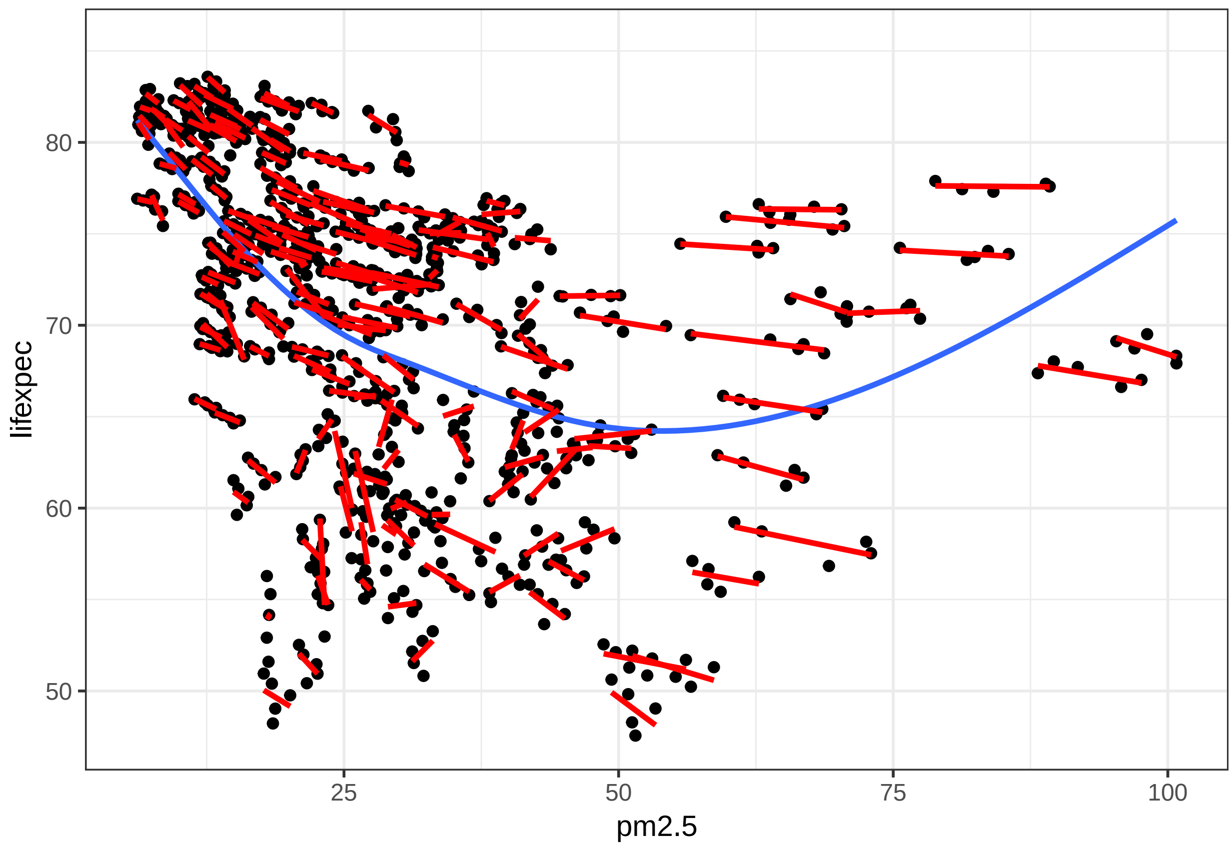

Figure 3. Relationship between PM2.5 and Life Expectancy

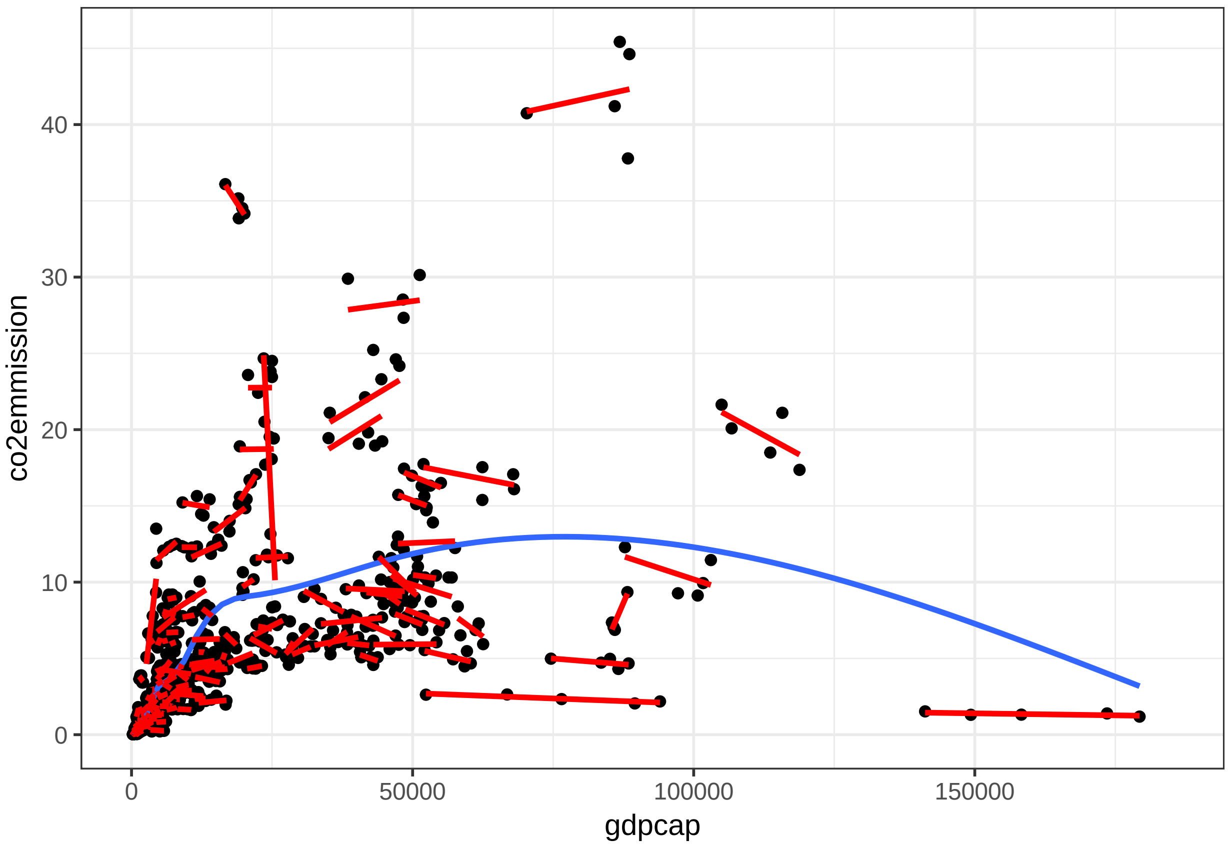

Figure 4. Relationship between GDP per Capita and CO2 Emission

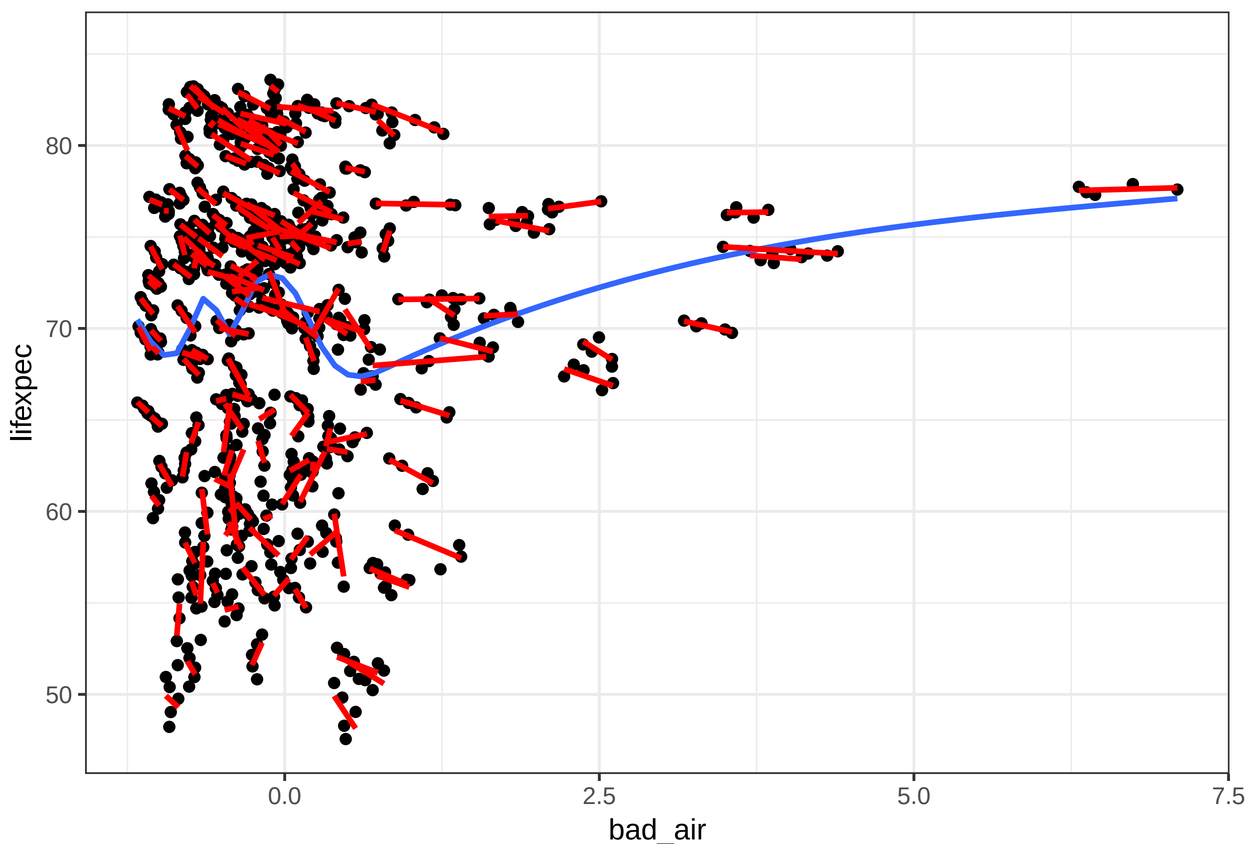

Figure 5. Synthetic PM2.5 & CO2 Emission and Life Expectancy

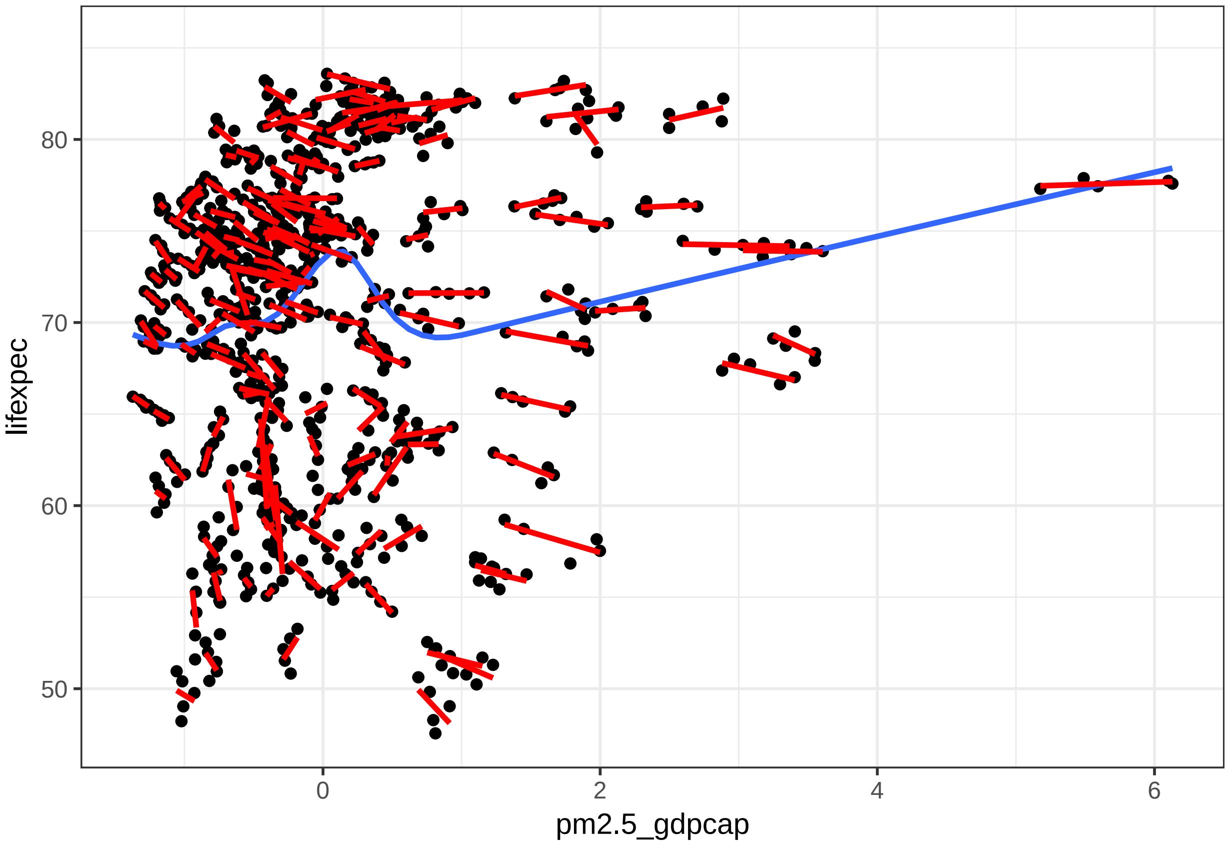

Figure 6. Relationship between Synthetic PM2.5 & GDP per capita and Life Expectancy

PM2.5 concentration and also high life expectancy. The country is Qatar, a wealthy middle eastern country. According to the data of Qatar, air quality does not cause life expenctany noticeably low. On the contrary, people in this country tend to live longer, despite of the poor air quality. This seems to contradicts with the common sense that poor air quality should lower life expectancy, however, it is suspected that GDP per capita may play a part in this contradictive phenomenon.

After the preliminary examination of how each indepedant variable and each synthetic scale influence life expetancy, I ran five models using log ratios and without log ratios. Bewteen models with log ratio and without log ratio, I kept the one without log ratio as it seems more intuitive for interpretation. For example, holding both CO2 emission and GDP per capita constant, on average, every microgram per cubic meter increase of PM2.5 per year is associated with .052 year decrease of life expectancy at birth.

Because PM2.5 seems to negatively associated with life expectancy, I ran five models with PM2.5 as the first variable for analysis. The influence of PM2.5 on life expectancy drops from .0672 to .0521 and remains little change after adding other variables. CO2 emission remains postively associated with life expectancy. According to model 3 in table 3, holding PM2.5 and GDP per capita constant, on average, every metric tons CO2 emission increase per capita is associated with .0131 year increase of life expectancy. The only time when CO2 emission is negatively associated with life expectancy is to hold GDP per capita constant. Holding GDP per capita constant, on average, every metric tons CO2 emission increase per capita is associated with .0248 years decrease of life expectancy. On the other hand, holding GDP per capita constant, on average, every microgram per cubic meter PM2.5 increase is associated with .521 year decrease of life expectancy.

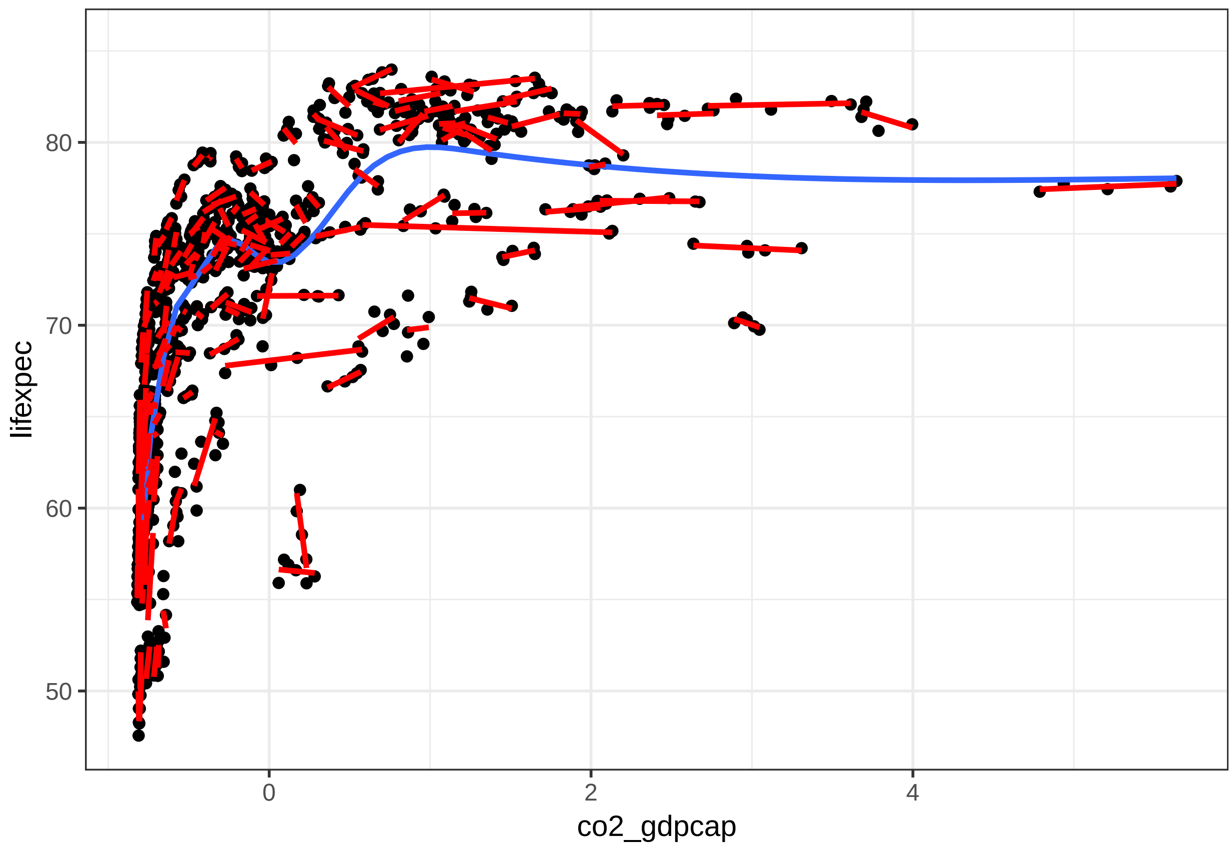

Figure 7. Relationship between Synthetic CO2 Emission & GDP per capita and Life Expectancy

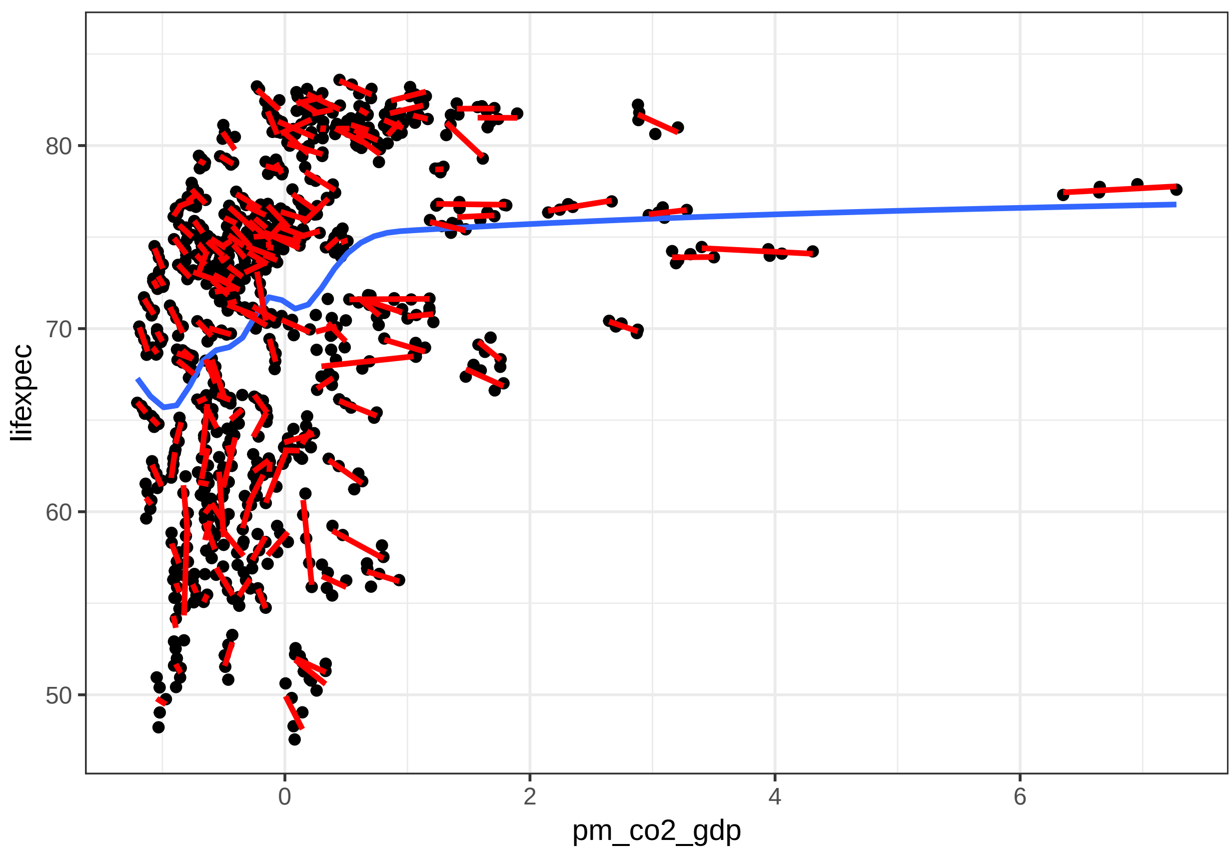

Figure 8. Relationship between Synthetic PM2.5, CO2 Emission & GDP per capita and Life Expectancy

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	4.1347^{***}	4.0551^{***}	4.0716^{***}	4.0730^{***}	3.9483^{***}
	(0.0122)	(0.0721)	(0.1038)	(0.1336)	(0.1826)
log(co2emmission)	0.0058	0.0005	0.0001	0.0010	0.0002
	(0.0049)	(0.0067)	(0.0058)	(0.0080)	(0.0078)
log(gdpcap)		0.0103	0.0098	0.0048	0.0051
		(0.0103)	(0.0111)	(0.0126)	(0.0122)
as.factor(time)2011		0.0035	0.0059	-0.0503	-0.0627
		(0.0037)	(0.0055)	(0.0448)	(0.0566)
as.factor(time)2012		0.0068^{*}	0.0097	-0.0576	-0.0622
		(0.0034)	(0.0054)	(0.0661)	(0.0710)
as.factor(time)2013		0.0105^{**}	0.8355	0.0075	0.0069
		(0.0035)	(2.0157)	(0.0043)	(0.0043)
as.factor(time)2014		0.0162^{***}	0.8280	0.0124^{*}	0.0116^{*}
		(0.0031)	(1.9957)	(0.0051)	(0.0052)
log(pm2.5)			-0.0034	0.0042	0.0040
			(0.0087)	(0.0190)	(0.0193)
log(gdpcapgro)				-0.0010	-0.0010
				(0.0010)	(0.0010)
log(urbanpop)					0.0386
					(0.0491)
R2	0.9697	0.9730	0.9748	0.9824	0.9825
BIC (null)	-2250	-2300	-2308	-1777	-1772
N	1075	1063	1046	802	802

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1. Statistical models

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	62.0769^{***}	61.5596^{***}	63.9004^{***}	63.9050^{***}	63.3445^{***}
	(0.8124)	(0.7993)	(2.3055)	(2.3047)	(2.4790)
co2emmission	-0.0104	-0.0439	-0.0216	-0.0217	-0.0243
	(0.0624)	(0.0628)	(0.0520)	(0.0519)	(0.0520)
I(gdpcap/1000)		0.0369	0.0323	0.0323	0.0327
		(0.0196)	(0.0190)	(0.0190)	(0.0191)
as.factor(time)2011		0.2590	0.2628	0.2620	0.2531
		(0.2750)	(0.2752)	(0.2746)	(0.2753)
as.factor(time)2012		0.4600	0.4468	0.4458	0.4327
		(0.2493)	(0.2578)	(0.2572)	(0.2597)
as.factor(time)2013		0.7057^{**}	0.6779^{**}	0.6762^{**}	0.6511^{**}
		(0.2287)	(0.2208)	(0.2211)	(0.2188)
as.factor(time)2014		1.1044^{***}	1.0369^{***}	1.0353^{***}	1.0075^{***}
		(0.2120)	(0.2397)	(0.2395)	(0.2400)
pm2.5			-0.0365	-0.0365	-0.0362
			(0.0327)	(0.0328)	(0.0329)
gdpcapgro				-0.0016	-0.0018
				(0.0103)	(0.0102)
urbanpop					0.0233
					(0.0410)
R2	0.9641	0.9670	0.9675	0.9675	0.9676
BIC (null)	-2085	-2141	-2151	-2145	-2139
N	1080	1080	1080	1080	1080

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2. Statistical models

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	65.4863^{***}	64.0059^{***}	63.9004^{***}	63.9050^{***}	63.3445^{***}
	(2.3118)	(2.3890)	(2.3055)	(2.3047)	(2.4790)
pm2.5	-0.0536	-0.0382	-0.0365	-0.0365	-0.0362
	(0.0340)	(0.0343)	(0.0327)	(0.0328)	(0.0329)
I(gdpcap/1000)		0.0303	0.0323	0.0323	0.0327
		(0.0191)	(0.0190)	(0.0190)	(0.0191)
as.factor(time)2011		0.2659	0.2628	0.2620	0.2531
		(0.2729)	(0.2752)	(0.2746)	(0.2753)
as.factor(time)2012		0.4441	0.4468	0.4458	0.4327
		(0.2563)	(0.2578)	(0.2572)	(0.2597)
as.factor(time)2013		0.6797^{**}	0.6779^{**}	0.6762^{**}	0.6511^{**}
		(0.2205)	(0.2208)	(0.2211)	(0.2188)
as.factor(time)2014		1.0362^{***}	1.0369^{***}	1.0353^{***}	1.0075^{***}
		(0.2405)	(0.2397)	(0.2395)	(0.2400)
co2emmission			-0.0216	-0.0217	-0.0243
			(0.0520)	(0.0519)	(0.0520)
gdpcapgro				-0.0016	-0.0018
				(0.0103)	(0.0102)
urbanpop					0.0233
					(0.0410)
R2	0.9651	0.9675	0.9675	0.9675	0.9676
BIC (null)	-2115	-2157	-2151	-2145	-2139
N	1080	1080	1080	1080	1080

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3. Statistical models

In both table 2 and 3, GDP per capita growth and urban population impact life expectancy differently. Holding PM2.5, GDP per capitca and CO2 emission constant, on average, every one percent increase of GDP per capita growth per year is associated with .0094 decrease of life expectancy. Holding all other variables constant, on average, every one percent increase of urban population is associated with .0359 years increase of life expectancy.

Conclusions

This paper aims to find out how air quality impacts life expectancy, while taking other confounding factors such as GDP per capita into considerations. According to above statistical results, both CO2 emission and PM2.5 impact life expectancy to different extents. When GDP per capita is held constant, both CO2 emission and PM2.5 have negative impact on life expectancy, the impact of PM2.5 being generally larger than CO2 emission. It is indicated that GDP per capita plays a part in expanding life expectancy despite of poor air quality. However, since this research does not address how much suffering there is among populations, it is unknown whether the people who live long actually enjoy their life or not. This project has its limitations. The data collected for analysis contain a substanitial amount of missing values. For example, in this dataset, I was only able to analyze the years from 2010 to 2015. The air quality from 2016 to now may have changed signifianctly due to rapid global warming. However, after running different tests and models, this project offers a glimpse of the relationship between air quality and life expectancy.

References

Matthew Gentzkow and Jesse Shapiro, “Code and Data for the Social Sciences: A Practitioner’s Guide”