HOW DOES AIR QUALITY AFFECT DISEASE RATES?

GROUP 9

Daniel Tiessalo

Antoni Kajrys

Anssi Lampinen

Ignacio Wąsowicz-Peinado

Aleksei Filonov

CS-A1160 Beginner’s Python for Engineers

WHAT DO WE WANT TO FIND?

Hypothesis: High air pollution increases the prevalence of diseases

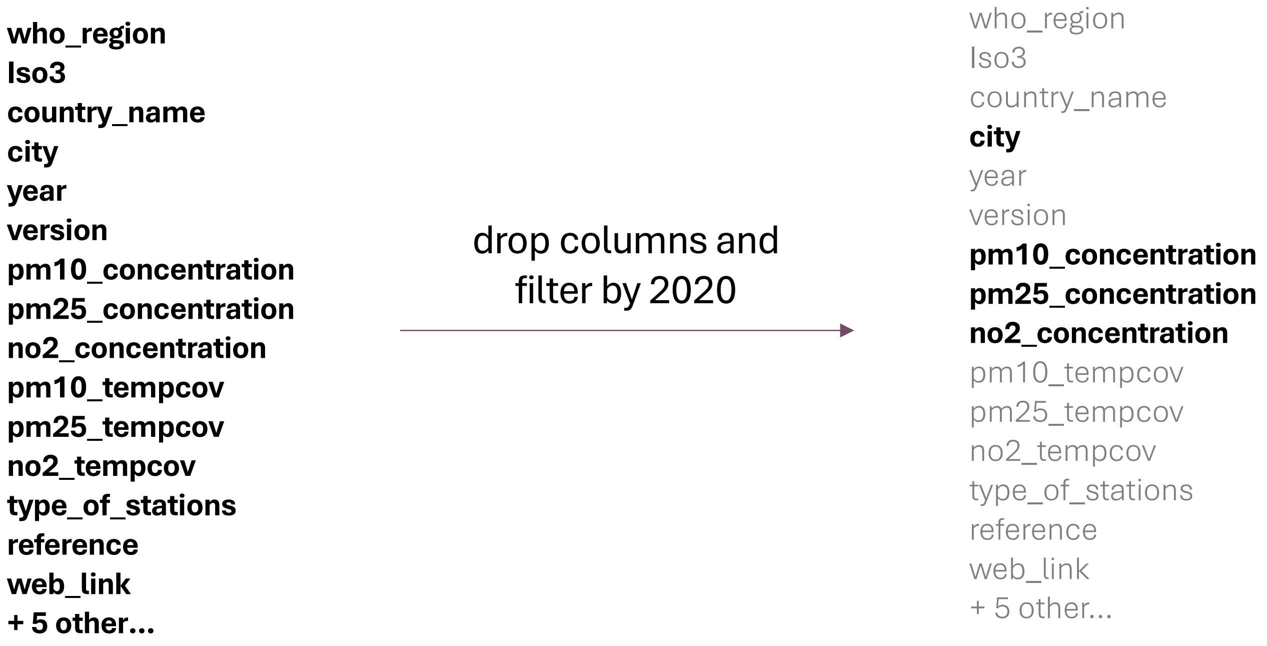
First, we plot the data against each other and use the visual method to see whether there exists any correlation

Then, we compute the correlation coefficients and their p-values

Finally, we interpret the results and their significance

MAIN DATASET

→ US Air quality data from around 400 cities (2020)



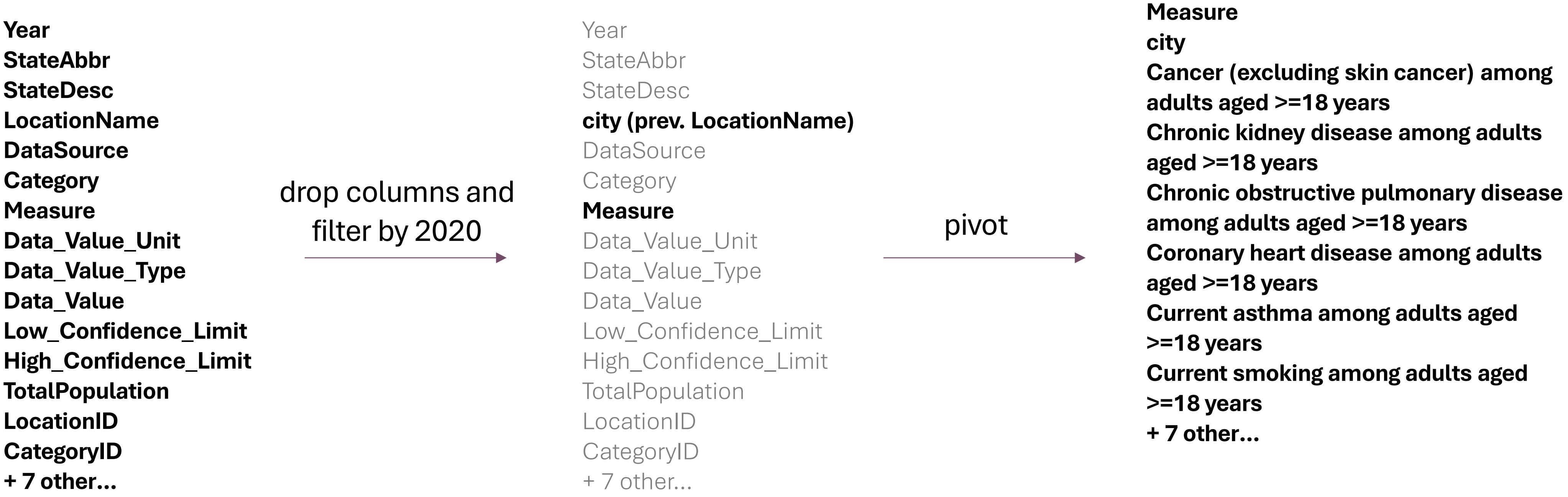
who_region
Iso3
country_name
city
year
version
pm10_concentration
pm25_concentration
no2_concentration
pm10_tempcov
pm25_tempcov
no2_tempcov
type_of_stations
reference
web_link
+ 5 other…

drop columns and filter by 2020

who_region
Iso3
country_name
city
year
version
pm10_concentration
pm25_concentration
no2_concentration
pm10_tempcov
pm25_tempcov
no2_tempcov
type_of_stations
reference
web_link
+ 5 other…

ADDITIONAL DATASET

→ US health data from over 3000 cities (2020)



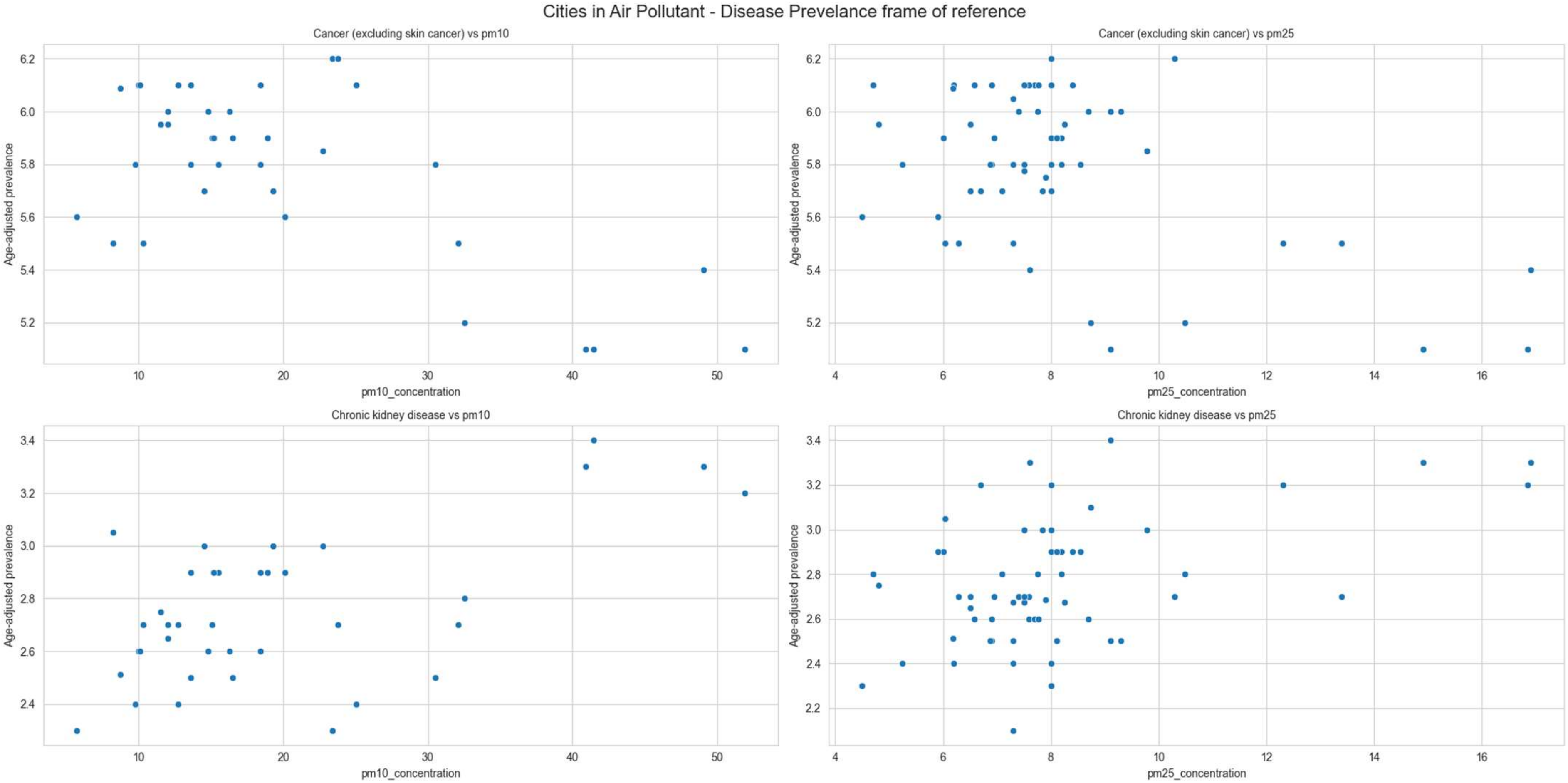
Year
StateAbbr
StateDesc
LocationName
DataSource
Category
Measure
Data_Value_Unit
Data_Value_Type
Data_Value
Low_Confidence_Limit
High_Confidence_Limit
TotalPopulation
LocationID
CategoryID
+ 7 other…

drop columns and filter by 2020

Year
StateAbbr
StateDesc
city (prev. LocationName)
DataSource
Category
Measure
Data_Value_Unit
Data_Value_Type
Data_Value
Low_Confidence_Limit
High_Confidence_Limit
TotalPopulation
LocationID
CategoryID
+ 7 other…

pivot

Measure
city
Cancer (excluding skin cancer) among adults aged >=18 years
Chronic kidney disease among adults aged >=18 years
Chronic obstructive pulmonary disease among adults aged >=18 years
Coronary heart disease among adults aged >=18 years
Current asthma among adults aged >=18 years
Current smoking among adults aged >=18 years
+ 7 other…



Cities in Air Pollutant - Disease Prevelance frame of reference

Cancer (excluding skin cancer) vs pm10

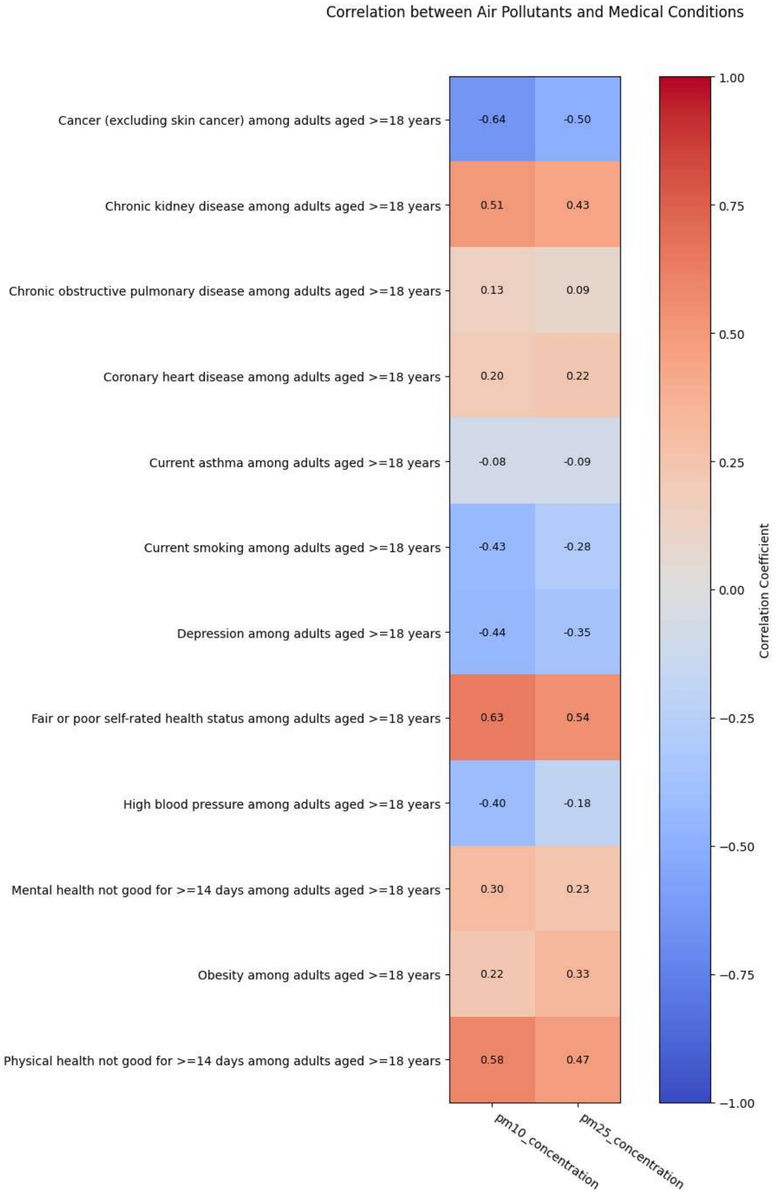
Cancer (excluding skin cancer) vs pm25

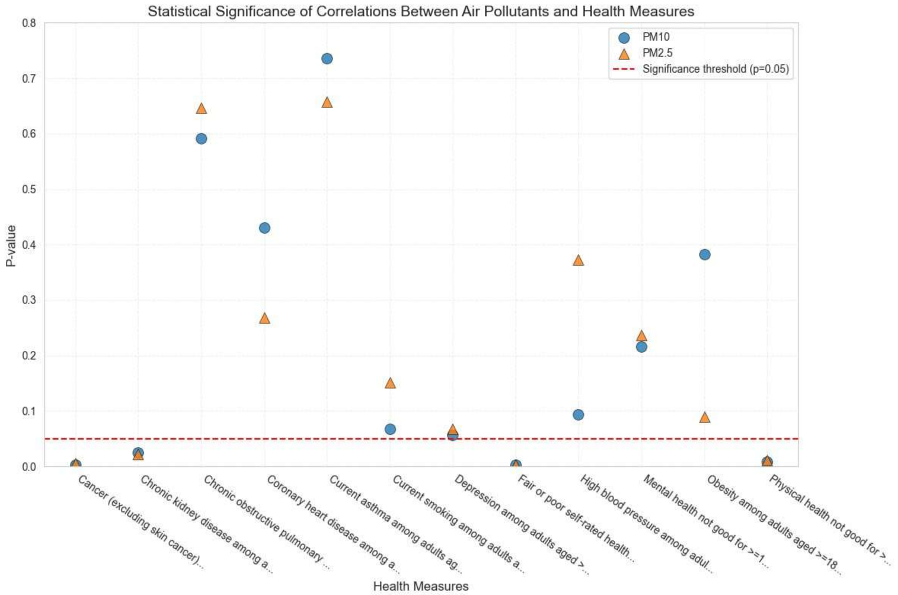
Chronic kidney disease vs pm10

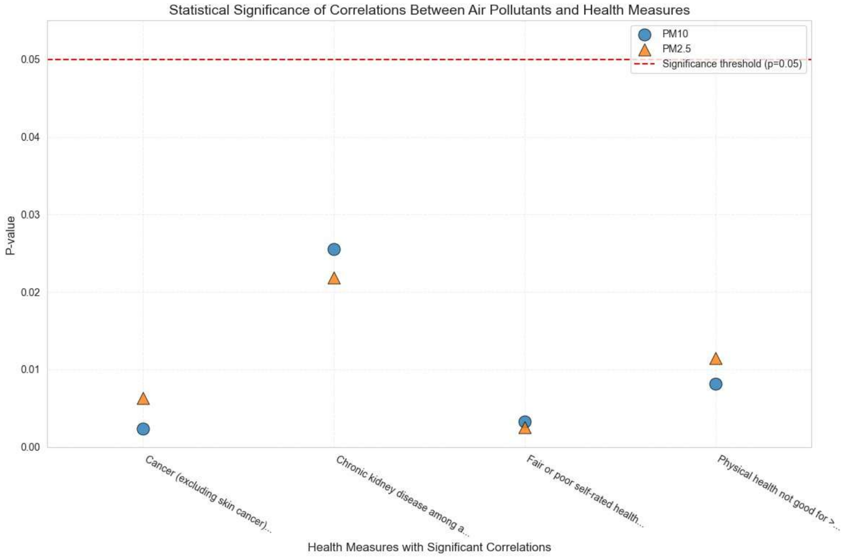
Chronic kidney disease vs pm25

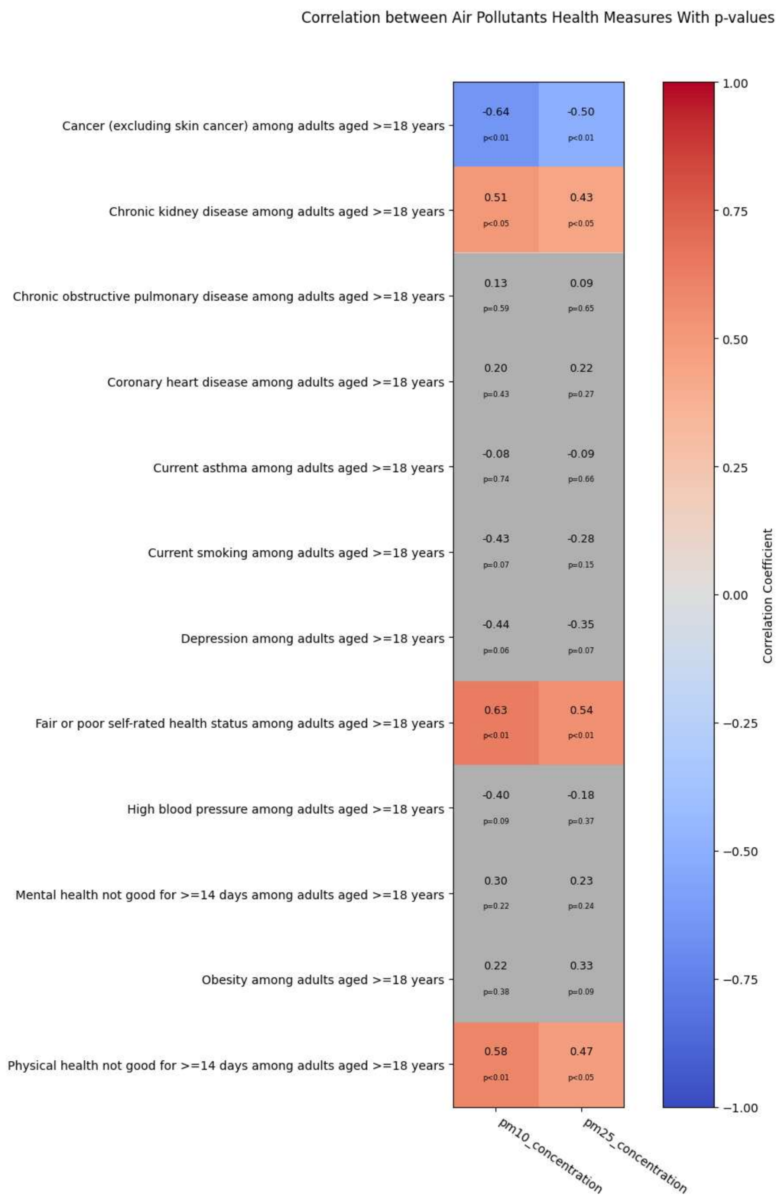
CALCULATE CORRELATIONS AND CHECK THEIR SIGNIFICANCE

- We first check if our data is normally distributed with a **Shapiro-Wilk test** → it is not
- We choose the **Spearman correlation test** because it does not require the normality of the data
- We use the **weighted test** to represent the difference in significance of different cities in the statistic based on their population
- We calculate the coefficients and their p-values
- We filter the coefficients so that only the statistically significant ones remain (p-value < 0.05)


Correlation between Air Pollutants and Medical Conditions
Cancer (excluding skin cancer) among adults aged >=18 years
-0.64
-0.50
Chronic kidney disease among adults aged >=18 years
0.51
0.43
Chronic obstructive pulmonary disease among adults aged >=18 years
0.13
0.09
Coronary heart disease among adults aged >=18 years
0.20
0.22
Current asthma among adults aged >=18 years
-0.08
-0.09
Current smoking among adults aged >=18 years
-0.43
-0.28
Depression among adults aged >=18 years
-0.44
-0.35
Fair or poor self-rated health status among adults aged >=18 years
0.63
0.54
High blood pressure among adults aged >=18 years
-0.40
-0.18
Mental health not good for >=14 days among adults aged >=18 years
0.30
0.23
Obesity among adults aged >=18 years
0.22
0.33
Physical health not good for >=14 days among adults aged >=18 years
0.58
0.47
pm10_concentration
pm25_concentration
Correlation Coefficient
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00


Statistical Significance of Correlations Between Air Pollutants and Health Measures
PM10
PM2.5
Significance threshold (p=0.05)
P-value
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Cancer (excluding skin cancer)...
Chronic kidney disease among a...
Chronic obstructive pulmonary ...
Coronary heart disease among a...
Current asthma among adults ag...
Current smoking among adults a...
Depression among adults aged >...
Fair or poor self-rated health...
High blood pressure among adul...
Mental health not good for >=1...
Obesity among adults aged >=18...
Physical health not good for >...
Health Measures


Statistical Significance of Correlations Between Air Pollutants and Health Measures
PM10
PM2.5
Significance threshold (p=0.05)
P-value
0.05
0.04
0.03
0.02
0.01
0.00
Cancer (excluding skin cancer)...
Chronic kidney disease among a...
Fair or poor self-rated health...
Physical health not good for >...
Health Measures with Significant Correlations

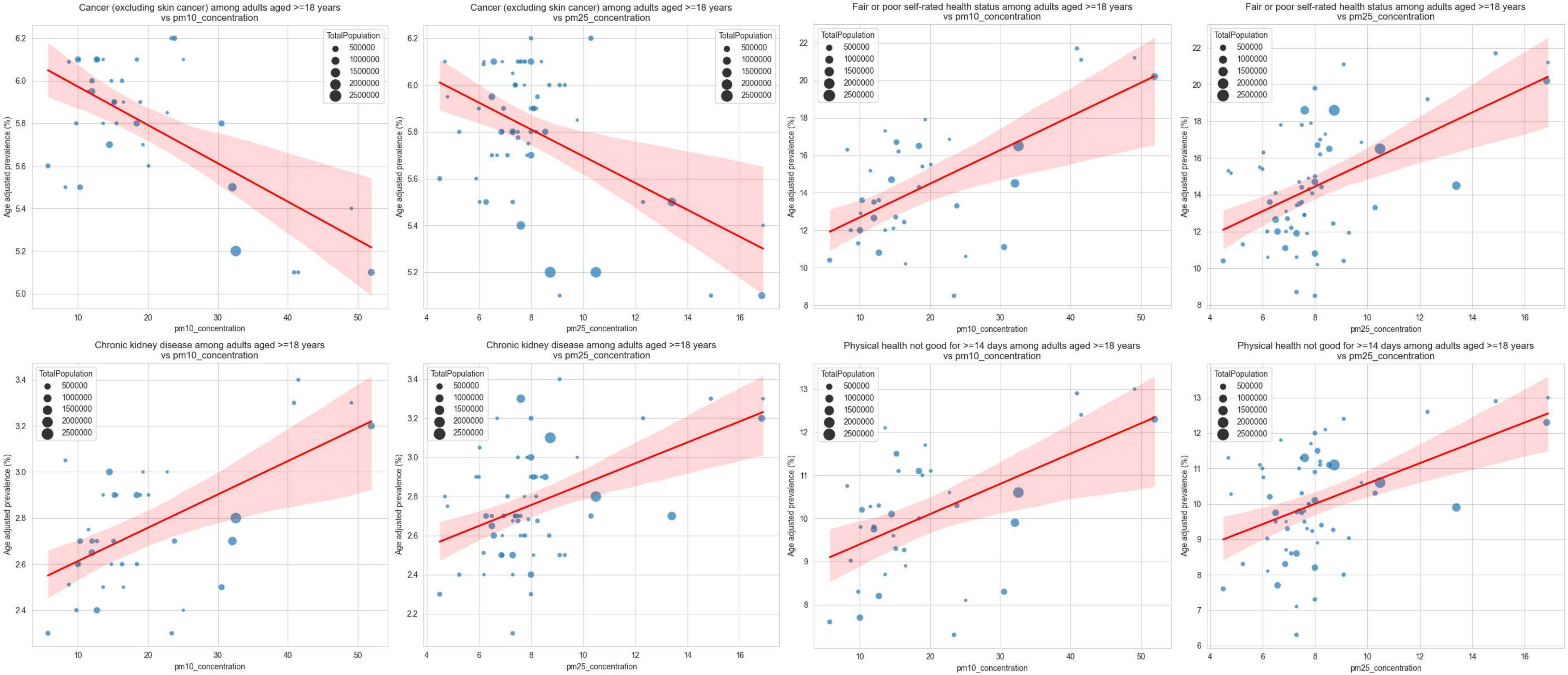


Correlation between Air Pollutants Health Measures With p-values

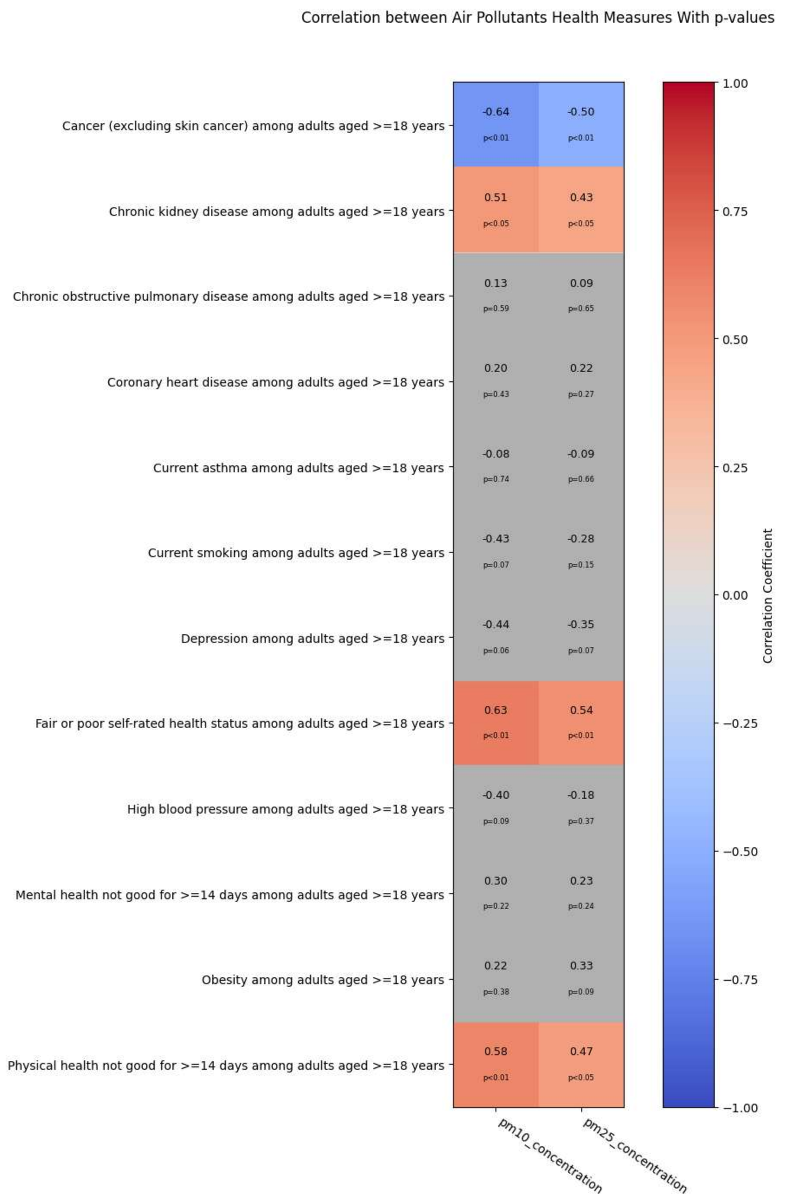
Conclusions

- Prevalence of cancer among adults is negatively correlated with air pollution
- Chronic kidney disease is positively correlated with air pollution
- Poor self-rated health status among adults is positively correlated with air pollution
- Bad physical health is positively correlated with air pollution
- Prevalence of neither asthma nor chronic obstructive pulmonary disease have significant correlation with the pollution levels, even though both are respiratory track conditions

BUBBLE PLOTS



CS-A1160 Beginner’s Python for Engineers



Conclusions

- Prevalence of cancer among adults is negatively correlated with air pollution
- Chronic kidney disease is positively correlated with air pollution
- Poor self-rated health status among adults is positively correlated with air pollution
- Bad physical health is positively correlated with air pollution
- Prevalence of neither asthma nor chronic obstructive pulmonary disease have significant correlation with the pollution levels, even though both are respiratory track conditions

IF YOU DON'T WANT CANCER, LIVE IN CRACOW!!!

CORRELATION IS NOT CAUSATION!

If we believed this study blindly, we would find that:

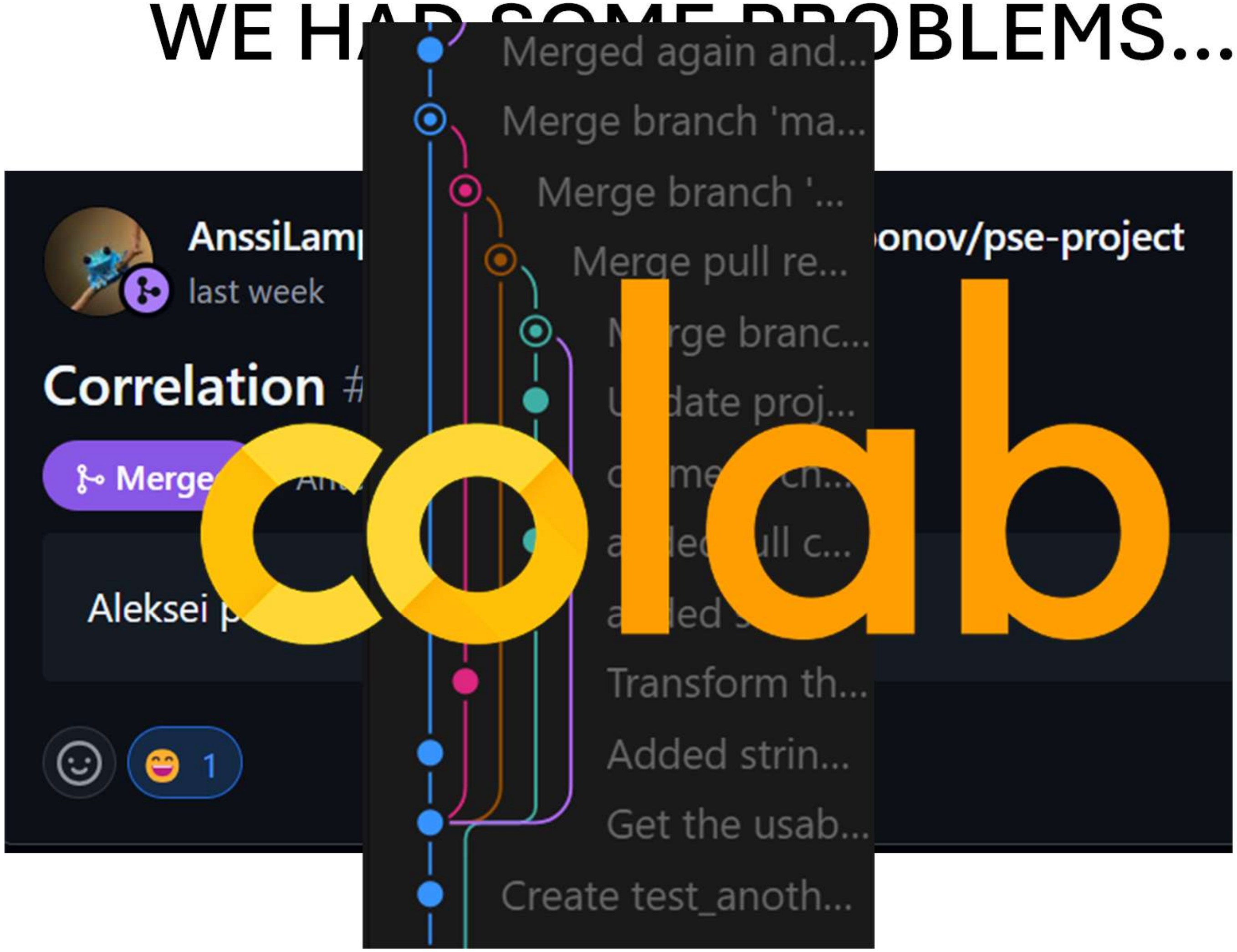
- Living in cities with high air pollution significantly reduces the chance of having cancer
- Depression is less prevalent in the cities with poor air quality
- More air pollution can decrease the prevalence of asthma and chronic obstructive pulmonary diseases

BUT

THIS DOES NOT MEAN THAT THE STUDY IS MEANINGLESS

- Smoking negatively correlates with air pollutants in this dataset.
- The conclusions of this analysis are inspirational.
- Correlation studies usually provide a starting point instead of concrete answers.

WE HAD SOME PROBLEMS...



AnssiLamp

last week

…onov/pse-project

Correlation #

Merge

Aleksei p

1

Merged again and...

Merge branch 'ma...

Merge branch '...

Merge pull re...

Merge branc...

Update proj...

c me ch...

added full c...

added s...

Transform th...

Added strin...

Get the usab...

Create test_anoth...

WE ALSO LEARNED SOMETHING

THANK YOU FOR YOUR ATTENTION

Aleksei Filonov
Presentation
Cleaning of datasets and merging
Correlation plot

Anssi Lampinen
Analysis
Weighing of statistics
Presentation

Daniel Tiessalo
Presentation
Cleaning of datasets and merging
Finding datasets

Antoni Kajrys
Finding datasets
Statistics
Checking of significance
Seeker of methods
Presentation

Ignacio Wąsowicz
Statistics
Presentation
Checking of significance