Outline:

1. Goals this week:
   1. To explore WHO and WB data and gain a better understanding of the type of research questions that can be asked during this research project.
   2. Brainstorm possible questions, with the associated initial data analysis to show their practicality.
2. Possible research questions
   1. Predicting women’s maternal mortality ratio using indicators about population health, healthcare system availability, and healthcare agency
   2. Asthma/chronic respiratory disease and environmental factors
3. Questions and to-dos

Possible research questions:

**Women’s Maternal Mortality Ratio (MMR)**

* The MMR quantifies ‘the number of maternal deaths during a given time period per 100,000 live births during the same time period’ (WHO).
* It ‘captures the risk of death in a single pregnancy or a single live birth’ (WHO)
* The MMR considers female deaths during pregnancy, childbirth, and within 42 days of termination of pregnancy.

The MMR can be affected by a wide variety of factors, from pre-existing conditions and availability of healthcare services to attitudes and health literacy.

To gain a better, preliminary understanding of the factors that can affect MMR, I plotted the covariance of MMR with a variety of other health indicators. Each point on the following scatterplot represents a single country in 2020 (e.g. Maldives).

This analysis suggested that MMR decreases when the following increase:

* Women participating in their own healthcare decisions
* Women participating in their own reproductive decisions
* Infants’ survival to 5 years old
* Knowledge about contraceptive methods
* Prenatal care
* Attendance of skilled professionals at births
* Percentage of doctors who are female
* Percentage of nursing personnel who are female
* Intermittent preventive treatment of malaria during pregnancy

The analysis indicated that MMR increased as the following increased:

* Rate of stillbirths
* HIV incidence
* Lack of participation in decisions about healthcare, visiting family, major household purchases

A graph with blue dots and red lines

AI-generated content may be incorrect.

These results indicate that we can use factors regarding the current healthcare system (e.g. availability of malaria treatments), coincident conditions (e.g. HIV prevalence), and decision-making agency to predict women’s health outcomes like MMR.

Many of these factors are related to each other, and there are numerous other, undescribed factors that can affect these outcomes. For example, there was a significant relationship (ANOVA testing) between the income level of a country (e.g. low income, lower middle income, upper middle income, high income) and MMR. Other examples of possible interactions include the effect of access to treatment centres, health literacy, sophistication of medical equipment and diagnostic practices on MMR.

To be able to decide how to use any of these factors to predict MMR, I will need a better understanding of how to isolate one variable from others.

Possible research question;

Can we predict MMR by using a mixture of indicators that describe population health and the state of the healthcare system as well as women’s knowledge about, and ability to make decisions regarding, their healthcare options?

* By separating these factors into three separate groups, we could additionally look at how interventions that improve indicators one of these groups would affect MMR rates for different countries.

**Environmental Indicators and Asthma/Chronic Respiratory Disease/COPD**

Possible research question:

Can we use information about population health and environmental factors (e.g. pollution levels) to predict prevalence of respiratory illness in a population?

* We can then 2estimate how interventions targeting these different environmental factors would affect respiratory health outcomes.

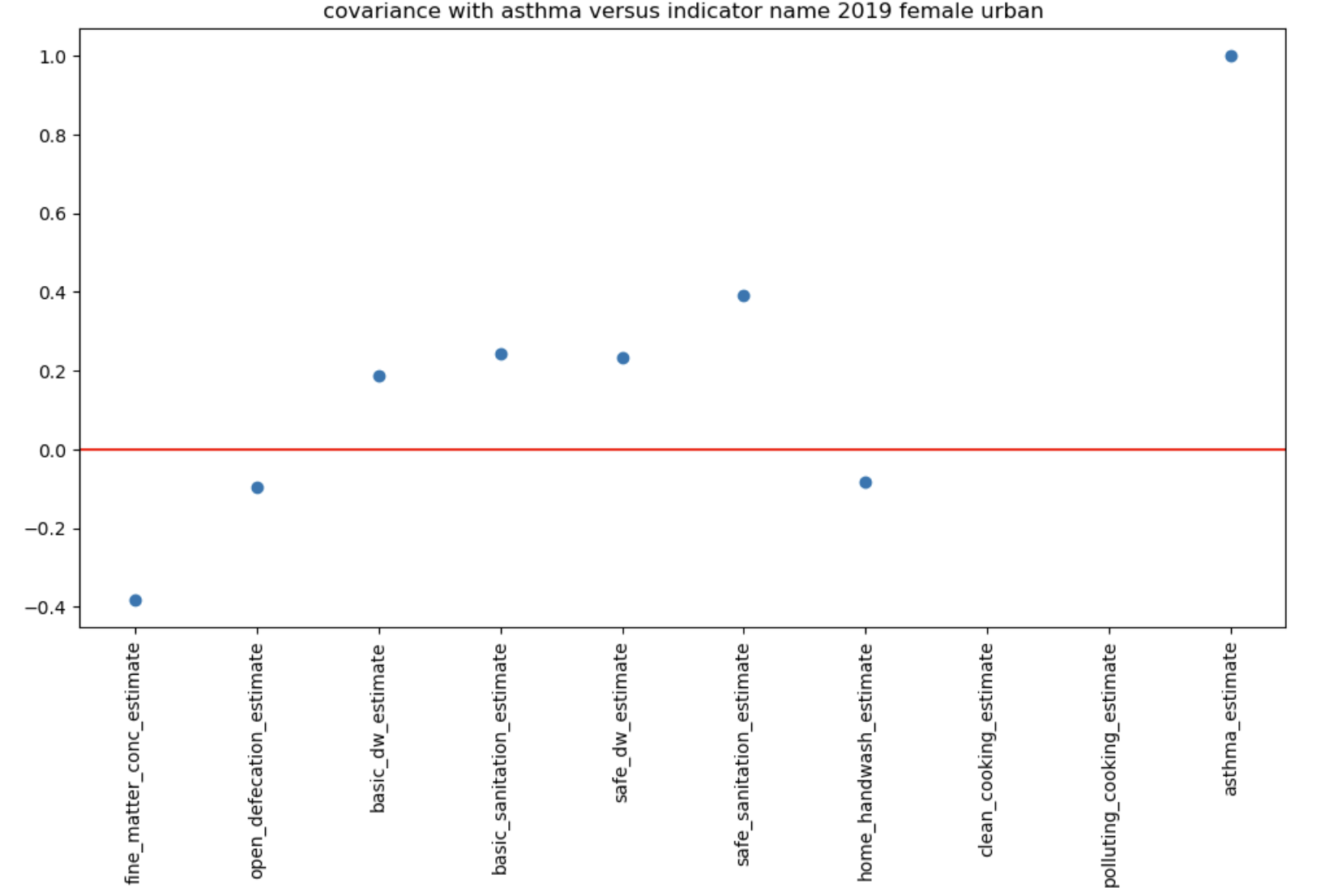
For example, we can use the following covariance plot to determine how DALYs associated with household pollution covary with other medical conditions (e.g. tuberculosis).

A graph with blue dots and red line

AI-generated content may be incorrect.

As another example of the scope of this question, we can see from the following scatterplot that:

* 2019 asthma prevalence for females in urban areas decreased as access to household handwashing facilities increased.
* However, there were some confusing results in this analysis, such as asthma prevalence tending to increase as the concentration of fine particulate matter (a metric for pollution levels) increased.
  + This could be a result of diagnosis rates as well as a myriad of confounding variables being included in this initial data visualisation.



Questions and To-Dos:

* How would you recommend best isolating the effect of one variable?
* Are either of these research questions suitable for this project?
  + If not, how would you recommend changing my approach?
  + If yes, what next steps would you recommend taking and/or how would you recommend refining them?
* Can we predict the future using this data?
  + For specific countries
* Research extrapolation methods
* Potential models:
  + Can we predict MMR using different subsets of data?
* Train model up to 2010, then predict for testable years
* Try different ML models for predictions
* Initially predict for 1 or 2 years
* Predict a single value and the variance
  + 95% confidence interval
    - Increases as we go into the future
* PCA
  + Initially focus on one year
  + Add columns to input for gender, income level
  + Colour plots for gender and countries, and see if there are clusters
* Supply names of datasets
* Upload to git