

# Project Proposal

Due October 11, 2021 by 11:59 PM

Kaitlyn Lewars and Katie Meehl: The Exposure Experience

October 11, 2021

## Load Packages

## Load Data

## Introduction and Data, including Research Questions

Climate change has been a recurring topic in the news in recent years as it becomes a more pressing problem. One of the important factors of climate change is air pollution. In 2017, air pollution was the 4th leading cause of mortality and the 5th leading cause of morbidity worldwide. Air pollution is a leading cause of morbidity and mortality, so we thought it would be important to explore a data set investigating this problem.

In general we would like to investigate air pollution as a cause of mortality. There are several different types of air pollution, but we will look at household pollution, ambient matter pollution, and ambient ozone pollution. With these variables we will compare them to see which air pollution is the most dangerous. We would also like to look into the trend of air pollution over the last 27 years. Lastly we would like to compare air pollution as a risk factor to other common risk factors. We downloaded this data from kaggle. There are several variables in this data including year, country, deaths by each type of air pollution, and deaths by other risk factors.

The data collection is a bit complicated. In order to estimate deaths caused by pollution they use “mathematical functions, derived from epidemiological studies from countries around the world, that relate different levels of exposure to the increased risk of death or disability from each cause, by age and sex, where applicable, estimates of population exposure to PM2.5, ozone, and household air pollution, country-specific data on underlying rates of disease and death for each pollution-linked disease, and a comprehensive set of population data, adjusted to match the UN2015 Population Prospectus and obtained from the Gridded Population of the World (GPW) database for each country” (<https://www.stateofglobalair.org/data/estimate-burden>).

## Glimpse

```
inbetween <- death_rates_total_air_pollution %>%
  left_join(death_rates_from_air_pollution, id = Entity)

joineddata <- inbetween %>%
  left_join(number_of_deaths_by_risk_factor, id = Entity)

glimpse(joineddata)
```

```
## Rows: 6,468
## Columns: 36
## $ Entity
```

<ch

```

## $ Code
## $ Year
## $ `Deaths - Air pollution - Sex: Both - Age: Age-standardized (Rate)`
## $ `Deaths - Household air pollution from solid fuels - Sex: Both - Age: Age-standardized (Rate)`
## $ `Deaths - Ambient particulate matter pollution - Sex: Both - Age: Age-standardized (Rate)`
## $ `Deaths - Ambient ozone pollution - Sex: Both - Age: Age-standardized (Rate)`
## $ `Unsafe water source`
## $ `Unsafe sanitation`
## $ `No access to handwashing facility`
## $ `Household air pollution from solid fuels`
## $ `Non-exclusive breastfeeding`
## $ `Discontinued breastfeeding`
## $ `Child wasting`
## $ `Child stunting`
## $ `Low birth weight for gestation`
## $ `Secondhand smoke`
## $ `Alcohol use`
## $ `Drug use`
## $ `Diet low in fruits`
## $ `Diet low in vegetables`
## $ `Unsafe sex`
## $ `Low physical activity`
## $ `High fasting plasma glucose`
## $ `High total cholesterol`
## $ `High body-mass index`
## $ `High systolic blood pressure`
## $ Smoking
## $ `Iron deficiency`
## $ `Vitamin A deficiency`
## $ `Low bone mineral density`
## $ `Air pollution`
## $ `Outdoor air pollution`
## $ `Diet high in sodium`
## $ `Diet low in whole grains`
## $ `Diet low in nuts and seeds`

```

## Data Analysis Plan

(Specify the outcome (response, Y) and predictor (explanatory, X) variables you will use to answer your question, as well as the comparison groups you will use, if applicable. You may include very preliminary exploratory data analysis, including some summary statistics and visualizations, along with some explanation on how they help you learn more about your data. Note the statistical method(s) that you believe will be useful in answering your question(s). What results from these specific statistical methods are needed to support your hypothesized answer?)

Monthly Birth Averages (2016–2021)

