# Project Proposal

Due October 11, 2021 by 11:59 PM

Kaitlyn Lewars and Katie Meehl: The Exposure Experience

October 11, 2021

## Load Packages

```
library(tidyverse)
library(readr)
```

## Load Data

```
death_rates_total_air_pollution <- read_csv("~/R/Project Proposal/data/death-rates-from-air-pollution.cs
death_rates_from_air_pollution <- read_csv("~/R/Project Proposal/data/death-rates-total-air-pollution.cs
number_of_deaths_by_risk_factor <- read_csv("~/R/Project Proposal/data/number-of-deaths-by-risk-factor.
```

## 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. As air pollution is a leading cause of morbidity and mortality, 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 globally to see 1) if the level of ambient air pollution concentration is correlated with number of deaths in a population or likelihood of an individual dying 2) what type of air pollution has the most effect on mortality. We hypothesize that ambient air pollution concentration and the likelihood of dying or number of deaths in a population are associated, with the higher the concentration of ambient air pollution the more likely an individual will die or more people from a population will die. We also hypothesize that ambient air pollution will have a greater effect on mortality because household air pollution is mainly a problem in low-income countries.

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)
```

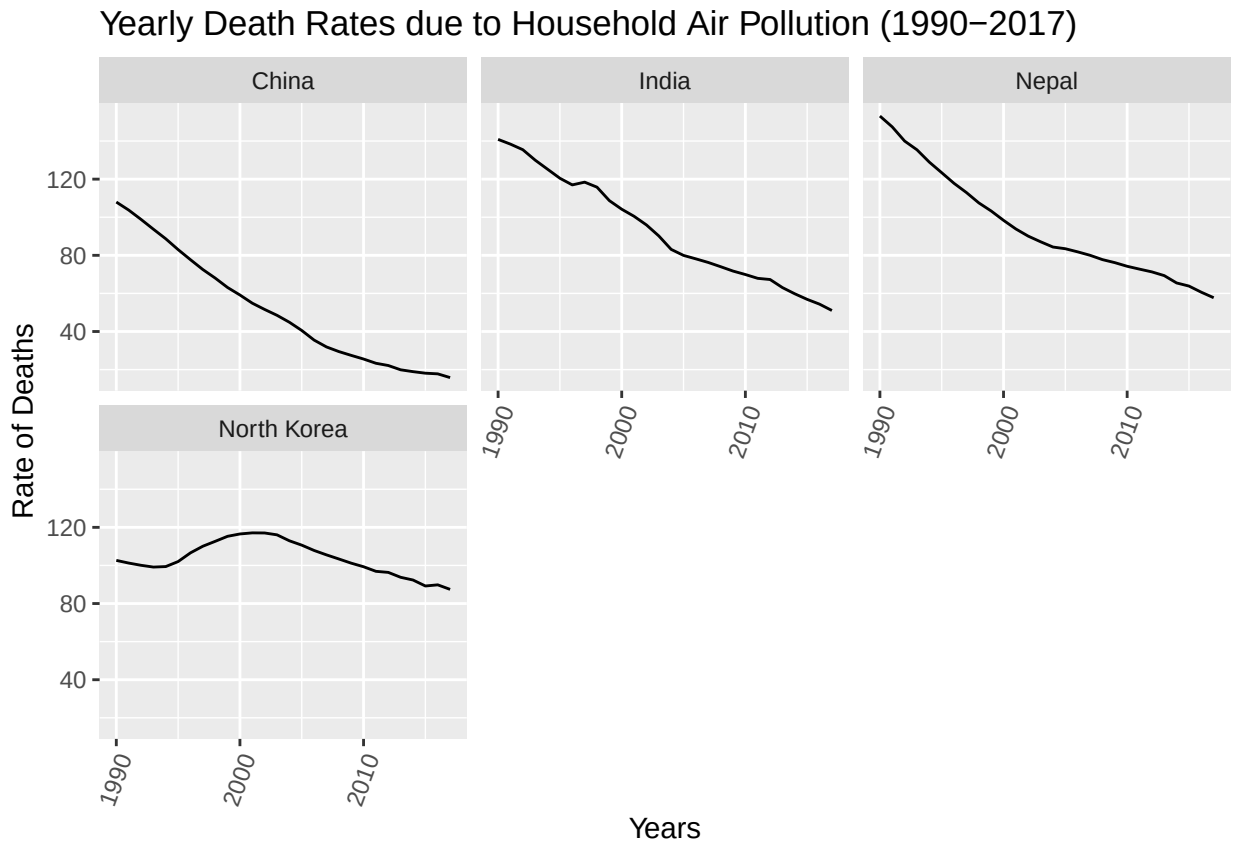
```
joineddata <- inbetween %>%
  left_join(number_of_deaths_by_risk_factor, id = Entity)

glimpse(joineddata)
```

```
## Rows: 6,468
## Columns: 36
## $ Entity                                                                        <ch
## $ Code                                                                          <ch
## $ Year                                                                          <db
## $ `Deaths - Air pollution - Sex: Both - Age: Age-standardized (Rate)`           <db
## $ `Deaths - Household air pollution from solid fuels - Sex: Both - Age: Age-standardized (Rate)` <db
## $ `Deaths - Ambient particulate matter pollution - Sex: Both - Age: Age-standardized (Rate)`    <db
## $ `Deaths - Ambient ozone pollution - Sex: Both - Age: Age-standardized (Rate)`                 <db
## $ `Unsafe water source`                                                         <db
## $ `Unsafe sanitation`                                                           <db
## $ `No access to handwashing facility`                                           <db
## $ `Household air pollution from solid fuels`                                    <db
## $ `Non-exclusive breastfeeding`                                                 <db
## $ `Discontinued breastfeeding`                                                  <db
## $ `Child wasting`                                                               <db
## $ `Child stunting`                                                              <db
## $ `Low birth weight for gestation`                                              <db
## $ `Secondhand smoke`                                                            <db
## $ `Alcohol use`                                                                 <db
## $ `Drug use`                                                                    <db
## $ `Diet low in fruits`                                                          <db
## $ `Diet low in vegetables`                                                      <db
## $ `Unsafe sex`                                                                  <db
## $ `Low physical activity`                                                       <db
## $ `High fasting plasma glucose`                                                 <db
## $ `High total cholesterol`                                                      <db
## $ `High body-mass index`                                                        <db
## $ `High systolic blood pressure`                                                <db
## $ Smoking                                                                       <db
## $ `Iron deficiency`                                                             <db
## $ `Vitamin A deficiency`                                                        <db
## $ `Low bone mineral density`                                                    <db
## $ `Air pollution`                                                               <db
## $ `Outdoor air pollution`                                                       <db
## $ `Diet high in sodium`                                                         <db
## $ `Diet low in whole grains`                                                    <db
## $ `Diet low in nuts and seeds`                                                  <db
```

## Data Analysis Plan

For our plan, we are looking at several different predictors, mostly focusing on different types of air pollution (household and ambient matter) and the outcomes we are looking at are deaths. We will compare the different types of air pollution to each other and to other risk factors/predictors. There will likely be a lot of correlation between the variables, so we would like to look into their interrelationships. Down below we created four graphs that examine the trends of ozone related deaths in four different countries. In general there seems to be a downward trend with deaths associated with household air pollution, which makes sense as household air pollution has been decreasing with the introduction of alternative clean fuel options. We would like to see

if that trend remains true in other countries and among other types of air pollution.




## What We Fixed

We got points off for having the wrong labels. So we fixed the labels from count to rates. We also added a bit more to our data analysis plan, explaining that we may need to look at the relationships between the data since they are highly correlated.