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. 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                                                                          <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, ambient matter, and ozone) 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. 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. We didn't do the rate

because there wasn't any population data within this data set. We may want to use some population data in order to better compare death by country (calculating death rates) . Our null hypothesis is that air pollution has a bigger effect on mortality than other risk factors. As for statistical methods we would like to further explore the difference in proportion of deaths due to air pollution and the proportion of deaths due to other risk factors, using a simulation-based two-sample t-test to check if we can reject the null hypothesis at the 95% confidence level. We would also like to perform chi-squared test to check for independence between risk factors/predictors. These are needed to check if the predictors are interacting with each other.