Final Report

due November 16, 2021 by 11:59 PM

Kaitlyn Lewars and Katie Meehl: The Exposure Experience

November 16, 2021

```
library(tidyverse)
```

```
## Warning in system("timedatectl", intern = TRUE): running command 'timedatectl'
## had status 1
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readr)

death_rates_total_air_pollution <- read_csv("~/R/Project Proposal/data/death-rates-from-air-pollution.cs
```

```
## Rows: 6468 Columns: 7
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Entity, Code
## dbl (5): Year, Deaths - Air pollution - Sex: Both - Age: Age-standardized (R...
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
death_rates_from_air_pollution <- read_csv("~/R/Project Proposal/data/death-rates-total-air-pollution.cs
```

```
## Rows: 6468 Columns: 4
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Entity, Code
## dbl (2): Year, Deaths - Air pollution - Sex: Both - Age: Age-standardized (R...
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
number_of_deaths_by_risk_factor <- read_csv("~/R/Project Proposal/data/number-of-deaths-by-risk-factor.
```

```
## Rows: 6468 Columns: 32
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (2): Entity, Code
## dbl (30): Year, Unsafe water source, Unsafe sanitation, No access to handwas...

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

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).

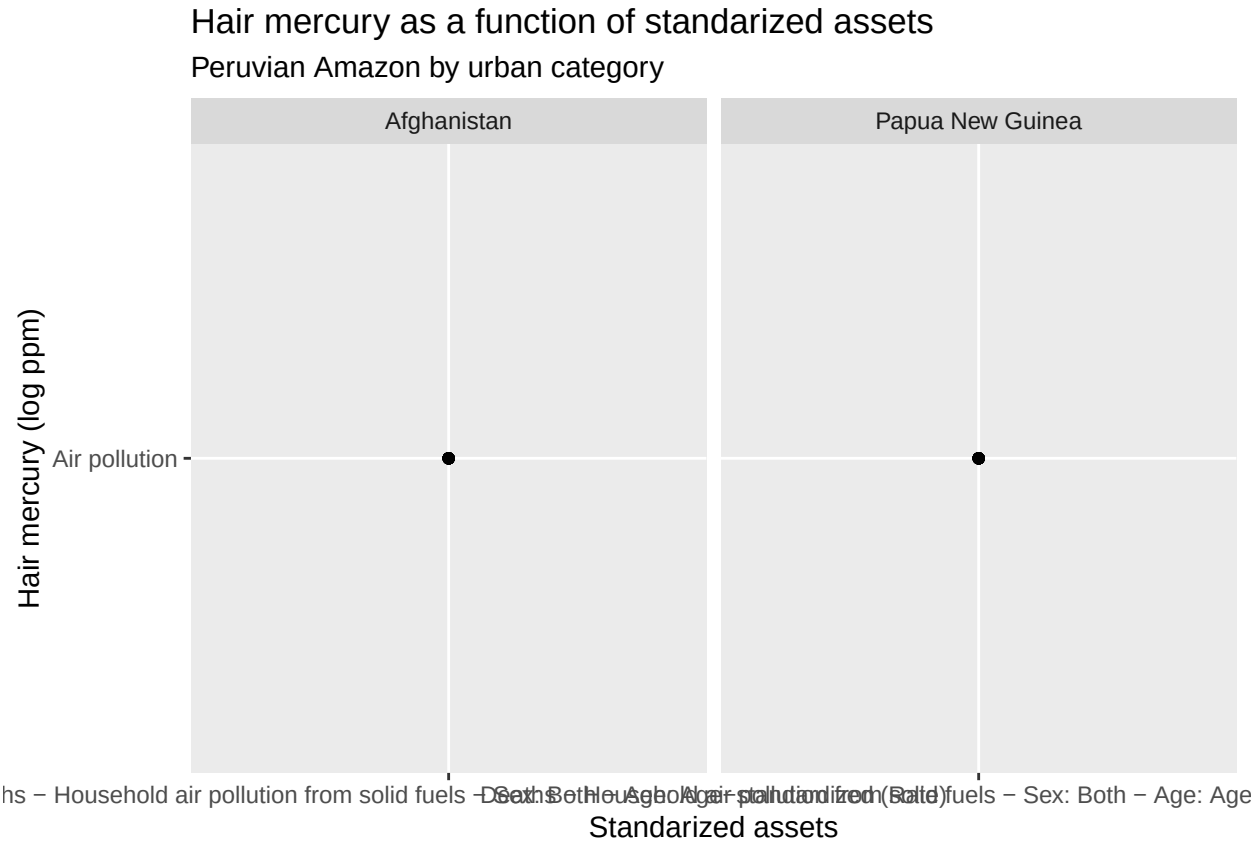
```
inbetween <- death_rates_total_air_pollution %>%
  left_join(death_rates_from_air_pollution, id = Entity)
```

```
## Joining, by = c("Entity", "Code", "Year", "Deaths - Air pollution - Sex: Both - Age: Age-standardize
```

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

```
## Joining, by = c("Entity", "Code", "Year")
```

```
joineddata %>%
  filter(Entity %in% c("Papua New Guinea", "Afghanistan", "Laus" )) %>%
  ggplot(aes(x = 'Deaths - Household air pollution from solid fuels - Sex: Both - Age: Age-standardized
    geom_point(alpha = 0.5) +
    labs(
      title = "Hair mercury as a function of standarized assets",
      subtitle = "Peruvian Amazon by urban category",
      x = "Standarized assets",
      y = "Hair mercury (log ppm)") +
    facet_wrap(~Entity)
```


Hair mercury as a function of standarized assets
Peruvian Amazon by urban category
Afghanistan
Papua New Guinea
Hair mercury (log ppm)
Air pollution
hs − Household air pollution from solid fuels − Sex: Both − Age: Age
Household air pollution from solid fuels − Sex: Both − Age: Age
Standarized assets