

608_Final_Project

Karim Hammoud

The main purpose of this data is to explore the pollution data and understand the insights, the risk factors and deaths related to that.

Per Wikipedia Air pollution is the presence of substances in the atmosphere that are harmful to the health of humans and other living beings, or cause damage to the climate or to materials. There are many different types of air pollutants, such as gases (including ammonia, carbon monoxide, sulfur dioxide, nitrous oxides, methane, carbon dioxide and chlorofluorocarbons), particulates (both organic and inorganic), and biological molecules. Air pollution may cause diseases, allergies and even death to humans; it may also cause harm to other living organisms such as animals and food crops, and may damage the natural environment (for example, climate change, ozone depletion or habitat degradation) or built environment (for example, acid rain). Both human activity and natural processes can generate air pollution.

In this project I will build a shiny app for the data to show the pollution per country and year.

I acquired the data from Kaggle you can check it here the data consists of the Country and the number of deaths due to pollution and other factors.

Another thing I would like to analyze the risk factors and deaths and show visualization of them

read the data

```
total <- read.csv("https://raw.githubusercontent.com/akarimhammoud/CUNY-SPS/master/608_Data_Visualization/factor")
factor <- read.csv("https://raw.githubusercontent.com/akarimhammoud/CUNY-SPS/master/608_Data_Visualization/death")
death <- read.csv("https://raw.githubusercontent.com/akarimhammoud/CUNY-SPS/master/608_Data_Visualization/share_deaths")
share_deaths <- read.csv("https://raw.githubusercontent.com/akarimhammoud/CUNY-SPS/master/608_Data_Visualization")
```

Lets see the summary of the factors table

```
summary(factor)
```

##	Entity	Code	Year	Unsafe.water.source
##	Length:6468	Length:6468	Min. :1990	Min. : 0.0
##	Class :character	Class :character	1st Qu.:1997	1st Qu.: 10.2
##	Mode :character	Mode :character	Median :2004	Median : 279.0
##			Mean :2004	Mean : 31566.3
##			3rd Qu.:2010	3rd Qu.: 5301.7
##			Max. :2017	Max. :2111659.1
##				
##	Unsafe.sanitation	No.access.to.handwashing.facility		
##	Min. : 0.0	Min. : 0.1		
##	1st Qu.: 4.6	1st Qu.: 16.9		
##	Median : 160.2	Median : 252.5		
##	Mean : 23374.4	Mean : 18933.1		
##	3rd Qu.: 3832.3	3rd Qu.: 3811.4		
##	Max. :1638021.2	Max. :1239519.4		
##				
##	Household.air.pollution.from.solid.fuels	Non.exclusive.breastfeeding		
##	Min. : 0.0	Min. : 0.0		
##	1st Qu.: 87.6	1st Qu.: 4.6		
##	Median : 1091.7	Median : 102.4		
##	Mean : 43084.2	Mean : 6231.4		
##	3rd Qu.: 9162.0	3rd Qu.: 1367.8		
##	Max. :2708904.8	Max. :514102.4		
##				
##	Discontinued.breastfeeding	Child.wasting	Child.stunting	
##	Min. : 0.00	Min. : 0	Min. : 0.0	
##	1st Qu.: 0.26	1st Qu.: 41	1st Qu.: 1.9	
##	Median : 6.62	Median : 730	Median : 77.9	
##	Mean : 409.11	Mean : 43446	Mean : 11767.7	
##	3rd Qu.: 78.28	3rd Qu.: 10235	3rd Qu.: 1971.6	
##	Max. :34850.40	Max. :3365309	Max. :1001277.4	
##				
##	Low.birth.weight.for.gestation	Secondhand.smoke	Alcohol.use	
##	Min. : 0.3	Min. : 2.9	Min. : -2315	
##	1st Qu.: 144.6	1st Qu.: 278.1	1st Qu.: 364	
##	Median : 1220.7	Median : 1196.2	Median : 2803	
##	Mean : 30948.0	Mean : 24282.3	Mean : 50203	
##	3rd Qu.: 8708.1	3rd Qu.: 5963.7	3rd Qu.: 12891	
##	Max. :1976612.5	Max. :1260994.2	Max. :2842854	
##				
##	Drug.use	Diet.low.in.fruits	Diet.low.in.vegetables	
##	Min. : 1.2	Min. : 1.6	Min. : 0.8	
##	1st Qu.: 92.9	1st Qu.: 536.0	1st Qu.: 413.0	
##	Median : 408.6	Median : 2452.9	Median : 1837.8	
##	Mean : 8890.2	Mean : 45452.6	Mean : 28742.0	
##	3rd Qu.: 2170.8	3rd Qu.: 10521.8	3rd Qu.: 7612.3	
##	Max. :585348.2	Max. :2423447.4	Max. :1462367.4	
##				
##	Unsafe.sex	Low.physical.activity	High.fasting.plasma.glucose	
##	Min. : 1.0	Min. : 2.4	Min. : 21	
##	1st Qu.: 136.1	1st Qu.: 261.6	1st Qu.: 2035	
##	Median : 831.8	Median : 1189.4	Median : 7820	
##	Mean : 26764.5	Mean : 21141.5	Mean : 99556	
##	3rd Qu.: 5949.0	3rd Qu.: 5694.7	3rd Qu.: 34705	

```
## Max. :1771140.7 Max. :1263051.3 Max. :6526028
##
## High.total.cholesterol High.body.mass.index High.systolic.blood.pressure
## Min. : 10 Min. : 20 Min. : 21
## 1st Qu.: 839 1st Qu.: 1141 1st Qu.: 2665
## Median : 4005 Median : 4740 Median : 10993
## Mean : 51628 Mean : 68685 Mean : 174383
## 3rd Qu.: 17423 3rd Qu.: 21601 3rd Qu.: 47323
## Max. :4392505 Max. :4724346 Max. :10440818
## NA's :4907
## Smoking Iron.deficiency Vitamin.A.deficiency
## Min. : 12 Min. : 0.01 Min. : 0.0
## 1st Qu.: 1293 1st Qu.: 2.26 1st Qu.: 1.9
## Median : 5936 Median : 31.99 Median : 70.5
## Mean : 133548 Mean : 1878.75 Mean : 11908.6
## 3rd Qu.: 31638 3rd Qu.: 421.38 3rd Qu.: 2081.9
## Max. :7099111 Max. :125242.95 Max. :986995.0
##
## Low.bone.mineral.density Air.pollution Outdoor.air.pollution
## Min. : 0.4 Min. : 9 Min. : 5
## 1st Qu.: 40.6 1st Qu.: 1077 1st Qu.: 554
## Median : 246.8 Median : 6125 Median : 2242
## Mean : 4579.1 Mean : 95736 Mean : 55573
## 3rd Qu.: 1096.1 3rd Qu.: 22727 3rd Qu.: 12822
## Max. :327314.3 Max. :4895476 Max. :3408878
## NA's :1
## Diet.high.in.sodium Diet.low.in.whole.grains Diet.low.in.nuts.and.seeds
## Min. : 3 Min. : 9.3 Min. : 5.2
## 1st Qu.: 356 1st Qu.: 798.7 1st Qu.: 553.3
## Median : 1946 Median : 3504.3 Median : 2279.2
## Mean : 54241 Mean : 53348.8 Mean : 34967.0
## 3rd Qu.: 9691 3rd Qu.: 14463.7 3rd Qu.: 10038.8
## Max. :3196514 Max. :3065588.5 Max. :2062521.7
##
```

Lets see the summary of the share deaths

```
names(share_deaths)[4] <- "total"
```

```
summary(share_deaths)
```

```
## Entity Code Year total
## Length:6412 Length:6412 Min. :1990 Min. : 1.550
## Class :character Class :character 1st Qu.:1997 1st Qu.: 5.890
## Mode :character Mode :character Median :2004 Median : 7.365
## Mean :2004 Mean : 7.637
## 3rd Qu.:2010 3rd Qu.: 9.113
## Max. :2017 Max. :17.190
```

Now lets plot the top 20 areas with share of death caused of Air Pollution in 20

```
library(tidyverse)
```

```
## -- 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()

# filter just 2017
share_deaths1 <- share_deaths[share_deaths$Year == "2017",]

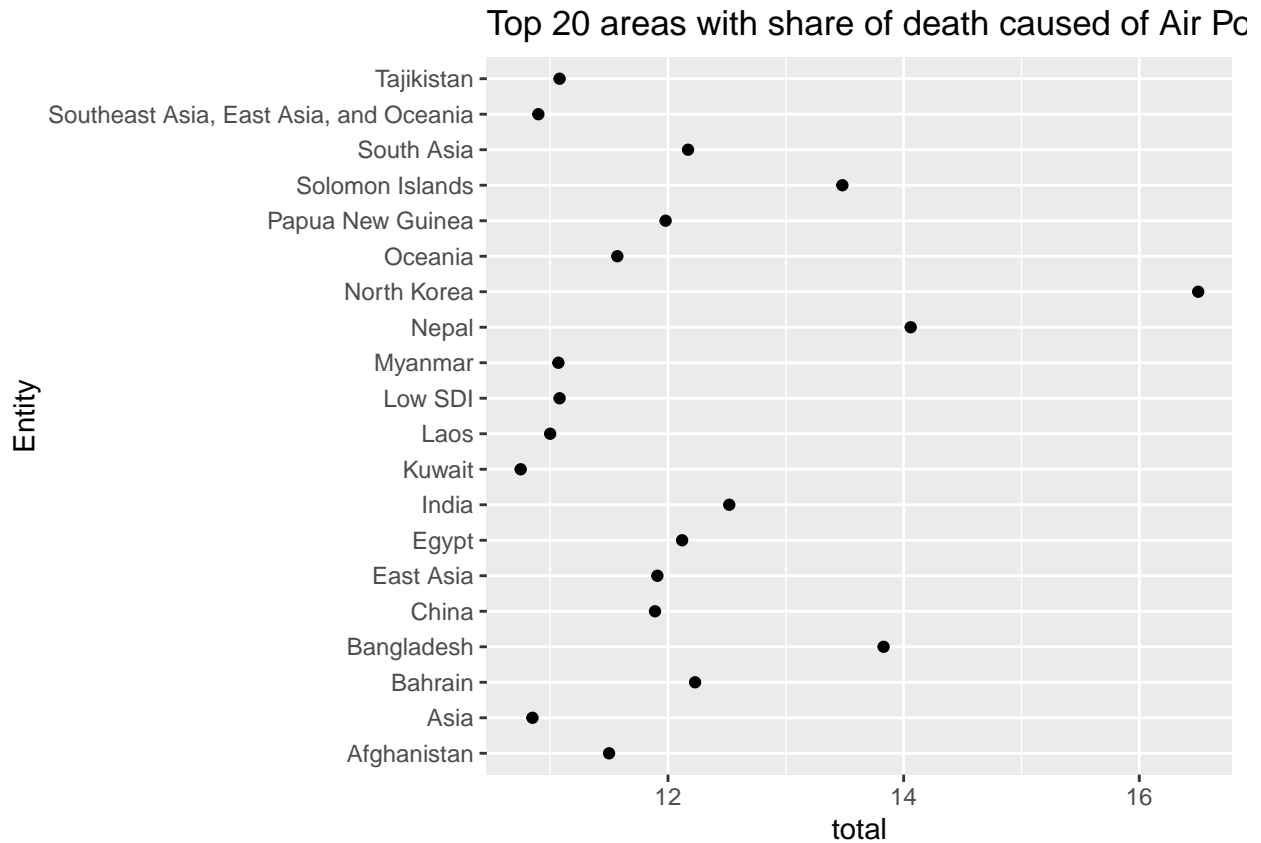
# get the top 20
share_deaths2 <- share_deaths1 %>% top_n(20)

## Selecting by total
share_deaths2

##
##      Entity Code Year total
## 1      Afghanistan AFG 2017 11.50
## 2              Asia      2017 10.85
## 3      Bahrain BHR 2017 12.23
## 4      Bangladesh BGD 2017 13.83
## 5              China CHN 2017 11.89
## 6      East Asia      2017 11.91
## 7      Egypt EGY 2017 12.12
## 8      India IND 2017 12.52
## 9      Kuwait KWT 2017 10.75
## 10     Laos LAO 2017 11.00
## 11     Low SDI      2017 11.08
## 12     Myanmar MMR 2017 11.07
## 13     Nepal NPL 2017 14.06
## 14     North Korea PRK 2017 16.50
## 15     Oceania      2017 11.57
## 16     Papua New Guinea PNG 2017 11.98
## 17     Solomon Islands SLB 2017 13.48
## 18     South Asia      2017 12.17
## 19 Southeast Asia, East Asia, and Oceania      2017 10.90
## 20     Tajikistan TJK 2017 11.08

# plot them
ggplot(data = share_deaths2, aes(x = total, y = Entity)) +
  geom_point(axes=FALSE) +
  labs(title = "Top 20 areas with share of death caused of Air Pollution")

## Warning: Ignoring unknown parameters: axes
```



Here we can see the top 20 areas and countries with Air Pollution.

This project seems really interesting especially in what we see in climate change and other factors are affecting our planet.