Group #28

Group: Mortality Explorers

Group Members: Ndey Isatou Jobe (SM Biostatistics), Khondoker Nazmoon Nabi (PhD Biostatistics), Bennet Cho (MPH Clinical Effectiveness Program)

HW #4: Project Check-In

10/8/23

We have read and attest to the statement in the BST 210 HW assignment.

Q1.

"Under 5 mortality". Using the World Bank database, we will look at associations between CO2 emissions and mortality in children under 5 years of age controlling for socioeconomic factors.

Q2.

Data is sourced from the World Bank DataBank. Data is collected on all countries of the world, on yearly CO2 emissions, yearly under 5 mortality rate, and other variables to account for confounding and effect measurement.

Q3.

- What is the relationship between a country's CO2 emissions and under-5 mortality rates?
- Is there an association between higher income and lower under-5 mortality rates?

Q4.

Does the relationship between CO2 emissions and under-5 mortality vary across different countries?

Q5.

Under-5 mortality (Continuous)

Q6.

- Data Preprocessing:

Clean and preprocess the data to investigate missing values, outliers, and inconsistencies.

- Data Exploration:

Conduct exploratory data analysis (EDA) to understand the distribution of our variables and identify potential patterns or outliers.

- Hypothesis Formulation:

Formulate our hypotheses that higher CO2 emissions are associated with higher under-5 mortality rates, and that higher income is associated with lower mortality rates.

- Correlation Analysis:

Begin plotting scatter plots to assess individual relationships and by calculating correlation coefficients between CO2 emissions, income, and under-5 mortality rates. Use statistical tests to assess the significance of these correlations.

- Regression Analysis:

Perform regression analysis to examine the relationships more thoroughly. Use multiple regression to account for the influence of income while assessing the relationship between CO2 emissions and mortality rates.

- Interpret the coefficients (1 and 2) and assess their statistical significance:

Interpret the results of our regression analysis, focusing on the relationship between CO2 emissions, income, and under-5 mortality rates. Discuss whether the relationships are statistically significant and the magnitude of their effects.

- Stratified Analysis:

investigate the relationship across regions, consider stratified analysis. Analyse the relationship between CO2 emissions, income, and mortality rates separately for different regions or groups of countries.

- Interaction Effects:

Assess whether there are interaction effects between CO2 emissions and income that influence mortality rates differently in various regions. Include interaction terms in our regression model if appropriate.

- Model Evaluation:

Evaluate the goodness of fit of your regression models, such as adjusted R-squared, AIC, MSE, and check for model assumptions, including linearity, normality of residuals, and homoscedasticity.

- Interpretation of Findings:

Interpret the results of your regression analysis, focusing on the relationship between CO2 emissions, income, and under-5 mortality rates. Discuss whether the relationships are statistically significant and the magnitude of their effects.

- Regional Comparisons:

Compare the findings across regions to determine if the relationships hold consistently or if there are regional variations.Q7.

Q7.

Our biggest challenge foreseen is how to collectively agree on the best model, as modelling is art and subjective

Q8.

Boston Children's Hospital Global Health research team: https://www.childrenshospital.org/programs/global-health/meet-our-team

Q9.

We will be using the R software package to complete this task.

Q10.

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

Data Exploration

```
tibble [184 x 30] (S3: tbl_df/tbl/data.frame)
 $ Time
 $ Time Code
 $ Country Name
 $ Country Code
 $ CO2 emissions (kg per PPP $ of GDP) [EN.ATM.CO2E.PP.GD]
 $ CO2 emissions (metric tons per capita) [EN.ATM.CO2E.PC]
 $ GDP per capita (current US$) [NY.GDP.PCAP.CD]
 $ Compulsory education, duration (years) [SE.COM.DURS]
 $ female_employment
 $ male_employment
 $ GDP per capita (constant LCU) [NY.GDP.PCAP.KN]
 $ Mortality rate, neonatal (per 1,000 live births) [SH.DYN.NMRT]
 $ Mortality rate, under-5 (per 1,000 live births) [SH.DYN.MORT]
 $ Mortality rate, under-5, female (per 1,000 live births) [SH.DYN.MORT.FE]
 $ Mortality rate, under-5, male (per 1,000 live births) [SH.DYN.MORT.MA]
 $ People using at least basic drinking water services (% of population) [SH.H2O.BASW.ZS]
 $ People practicing open defecation (% of population) [SH.STA.ODFC.ZS]
 $ People using at least basic sanitation services (% of population) [SH.STA.BASS.ZS]
 $ Prevalence of current tobacco use, females (% of female adults) [SH.PRV.SMOK.FE]
 $ Prevalence of current tobacco use, males (% of male adults) [SH.PRV.SMOK.MA]
 $ Prevalence of moderate or severe food insecurity in the population (%) [SN.ITK.MSFI.ZS]
 $ Immunization, HepB3 (% of one-year-old children) [SH.IMM.HEPB]
 $ Immunization, DPT (% of children ages 12-23 months) [SH.IMM.IDPT]
 $ Immunization, measles (% of children ages 12-23 months) [SH.IMM.MEAS]
 $ Unemployment, female (% of female labor force) (modeled ILO estimate) [SL.UEM.TOTL.FE.ZS]
 $ Unemployment, female (% of female labor force) (national estimate) [SL.UEM.TOTL.FE.NE.ZS]
 $ Unemployment, male (% of male labor force) (modeled ILO estimate) [SL.UEM.TOTL.MA.ZS]
 $ Unemployment, male (% of male labor force) (national estimate) [SL.UEM.TOTL.MA.NE.ZS]
 $ Unemployment, total (% of total labor force) (modeled ILO estimate) [SL.UEM.TOTL.ZS]
 $ Unemployment, total (% of total labor force) (national estimate) [SL.UEM.TOTL.NE.ZS]
```

```
     Time             Time Code          Country Name       Country Code
 Length:184         Length:184         Length:184         Length:184
 Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character
```

```
CO2 emissions (kg per PPP $ of GDP) [EN.ATM.CO2E.PP.GD]
Min.   :0.03232
1st Qu.:0.11226
Median :0.16760
Mean   :0.20042
3rd Qu.:0.24507
Max.   :0.77353
NA's   :5
CO2 emissions (metric tons per capita) [EN.ATM.CO2E.PC]
Min.   : 0.03699
1st Qu.: 0.80176
Median : 2.94638
Mean   : 4.28644
3rd Qu.: 5.84861
Max.   :32.47447
NA's   :5
GDP per capita (current US$) [NY.GDP.PCAP.CD]
Min.   :   223.9
1st Qu.:  2113.2
Median :  5943.5
Mean   : 14422.0
3rd Qu.: 17470.0
Max.   :113218.7
NA's   :5
Compulsory education, duration (years) [SE.COM.DURS] female_employment
Length:184                                           Length:184
Class :character                                     Class :character
Mode  :character                                     Mode  :character



male_employment    GDP per capita (constant LCU) [NY.GDP.PCAP.KN]
Length:184         Length:184
Class :character   Class :character
Mode  :character   Mode  :character



Mortality rate, neonatal (per 1,000 live births) [SH.DYN.NMRT]
Min.   : 0.80
1st Qu.: 3.65
```

```
Median : 9.10
Mean   :12.53
3rd Qu.:20.65
Max.   :44.80
NA's   :5
Mortality rate, under-5 (per 1,000 live births) [SH.DYN.MORT]
Min.   :  2.00
1st Qu.:  6.65
Median : 16.50
Mean   : 27.52
3rd Qu.: 43.00
Max.   :118.30
NA's   :5
Mortality rate, under-5, female (per 1,000 live births) [SH.DYN.MORT.FE]
Min.   :  1.80
1st Qu.:  6.00
Median : 14.70
Mean   : 25.16
3rd Qu.: 38.85
Max.   :112.00
NA's   :5
Mortality rate, under-5, male (per 1,000 live births) [SH.DYN.MORT.MA]
Min.   :  2.20
1st Qu.:  7.20
Median : 17.80
Mean   : 29.76
3rd Qu.: 47.30
Max.   :124.00
NA's   :5
People using at least basic drinking water services (% of population) [SH.H2O.BASW.ZS]
Length:184
Class :character
Mode  :character




People practicing open defecation (% of population) [SH.STA.ODFC.ZS]
Length:184
Class :character
Mode  :character
```

```
People using at least basic sanitation services (% of population) [SH.STA.BASS.ZS]
Length:184
Class :character
Mode  :character
```

```
Prevalence of current tobacco use, females (% of female adults) [SH.PRV.SMOK.FE]
Length:184
Class :character
Mode  :character
```

```
Prevalence of current tobacco use, males (% of male adults) [SH.PRV.SMOK.MA]
Length:184
Class :character
Mode  :character
```

```
Prevalence of moderate or severe food insecurity in the population (%) [SN.ITK.MSFI.ZS]
Length:184
Class :character
Mode  :character
```

```
Immunization, HepB3 (% of one-year-old children) [SH.IMM.HEPB]
Length:184
Class :character
Mode  :character
```

```
Immunization, DPT (% of children ages 12-23 months) [SH.IMM.IDPT]
```

```
Length:184
Class :character
Mode  :character
```

```
Immunization, measles (% of children ages 12-23 months) [SH.IMM.MEAS]
Length:184
Class :character
Mode  :character
```

```
Unemployment, female (% of female labor force) (modeled ILO estimate) [SL.UEM.TOTL.FE.ZS]
Length:184
Class :character
Mode  :character
```

```
Unemployment, female (% of female labor force) (national estimate) [SL.UEM.TOTL.FE.NE.ZS]
Length:184
Class :character
Mode  :character
```

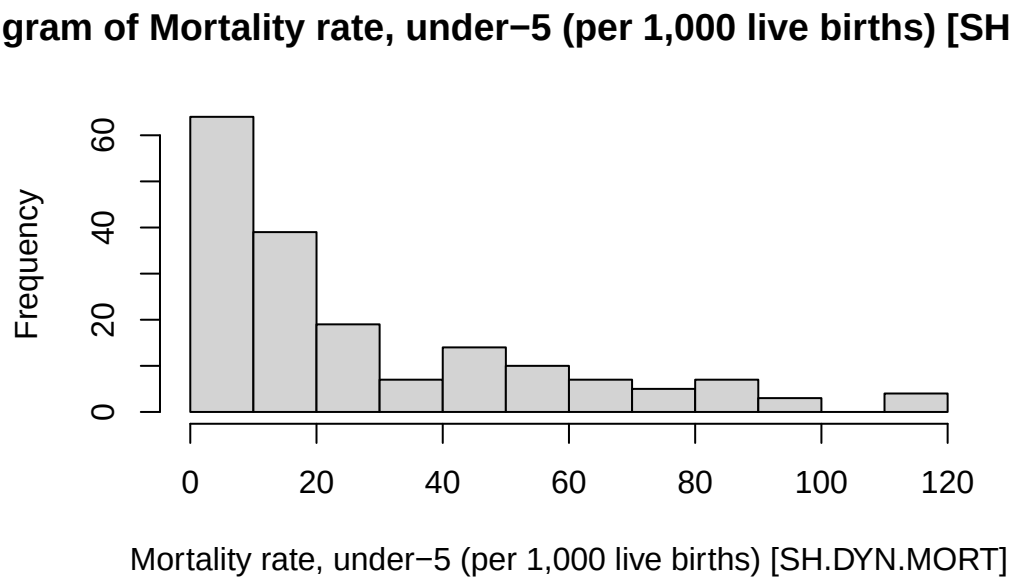
```
Unemployment, male (% of male labor force) (modeled ILO estimate) [SL.UEM.TOTL.MA.ZS]
Length:184
Class :character
Mode  :character
```

```
Unemployment, male (% of male labor force) (national estimate) [SL.UEM.TOTL.MA.NE.ZS]
Length:184
Class :character
Mode  :character
```

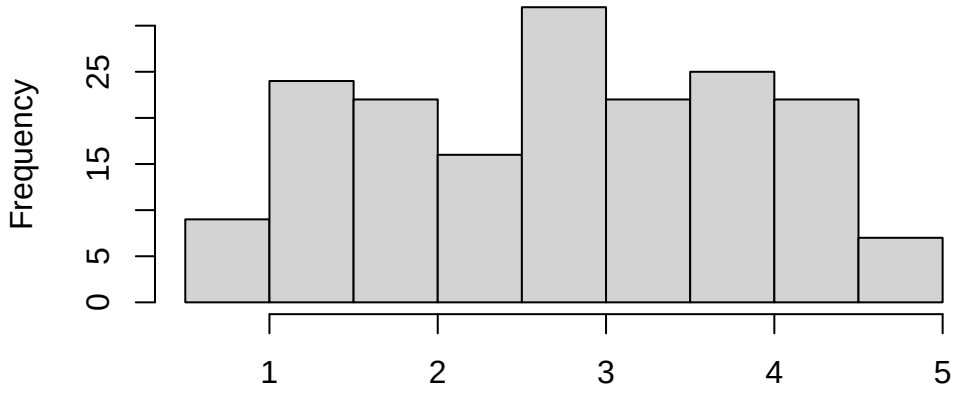
```
Unemployment, total (% of total labor force) (modeled ILO estimate) [SL.UEM.TOTL.ZS]
Length:184
Class :character
Mode  :character
```

```
Unemployment, total (% of total labor force) (national estimate) [SL.UEM.TOTL.NE.ZS]
Length:184
Class :character
Mode  :character
```

Data Visualization

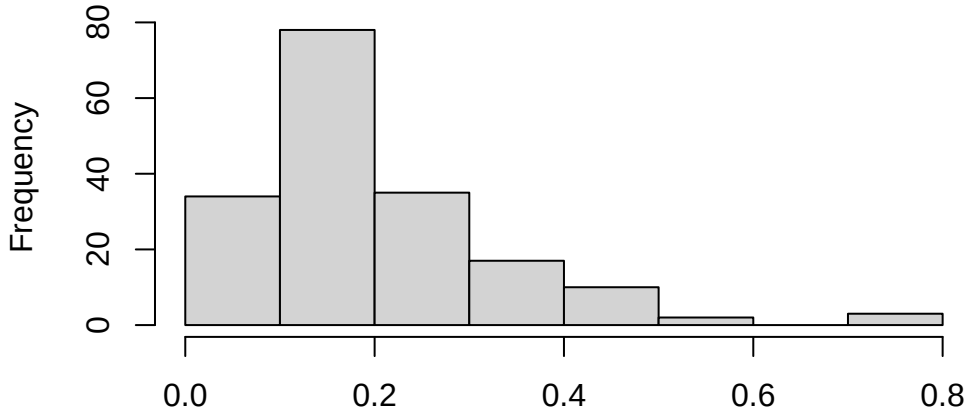


ım of log(‘Mortality rate, under−5 (per 1,000 live births) [SH.I



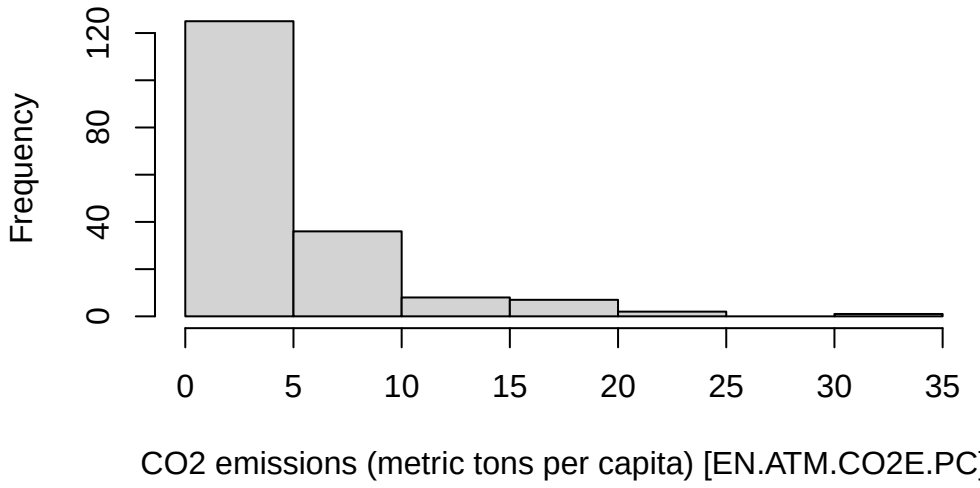
log(‘Mortality rate, under−5 (per 1,000 live births) [SH.DYN.MORT]‘)

gram of CO2 emissions (kg per PPP $ of GDP) [EN.ATM.CO2

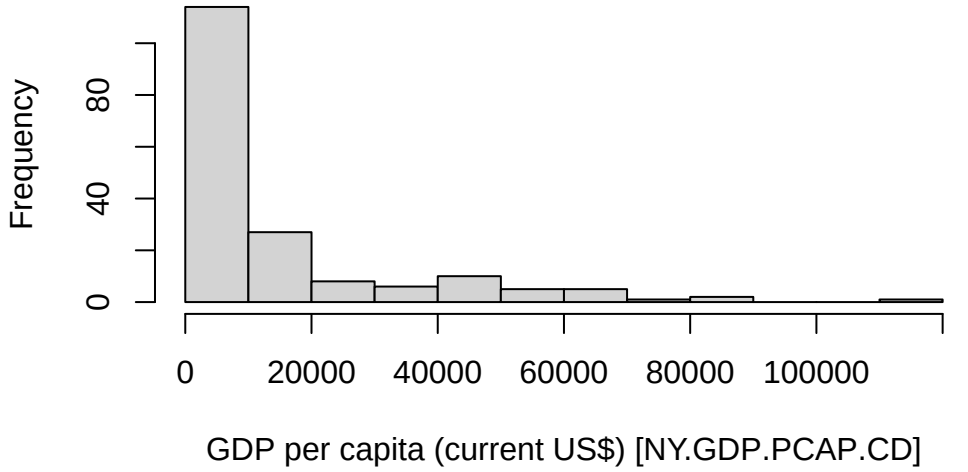


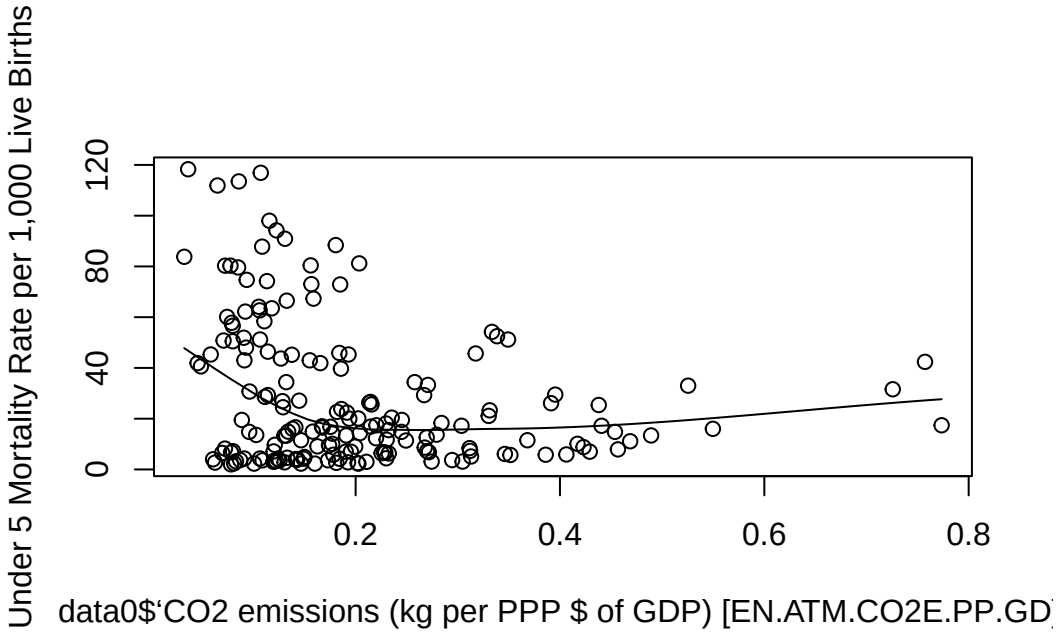
CO2 emissions (kg per PPP $ of GDP) [EN.ATM.CO2E.PP.GD]

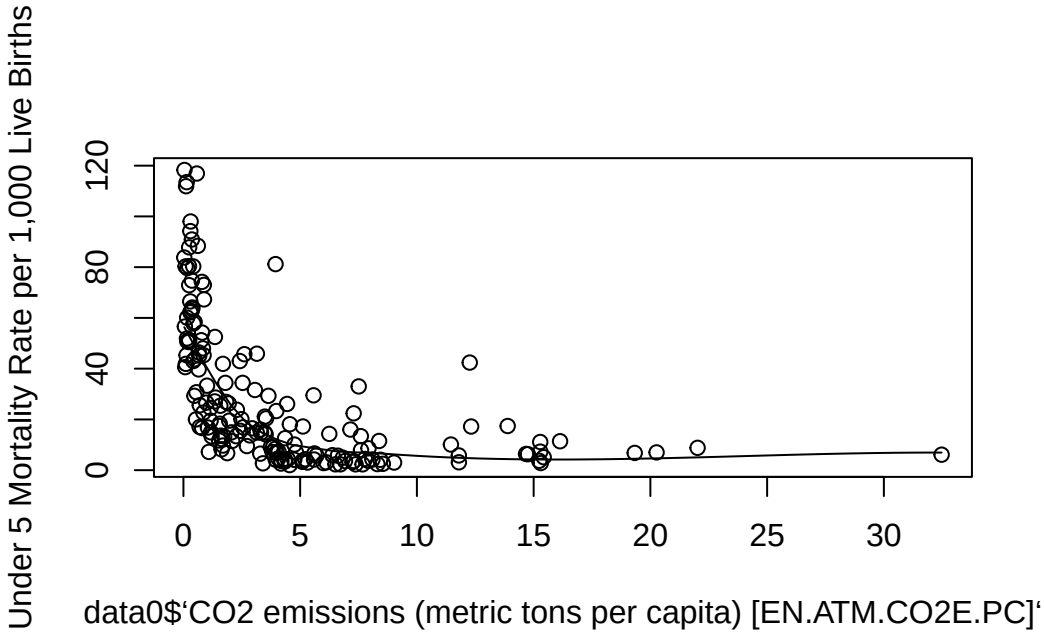
ogram of CO2 emissions (metric tons per capita) [EN.ATM.C

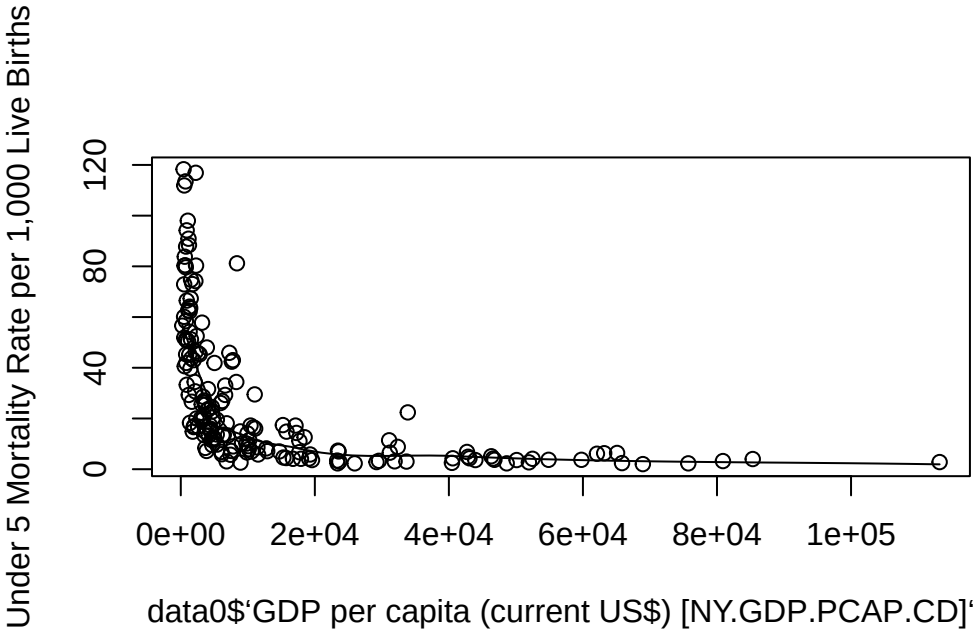


Histogram of GDP per capita (current US$) [NY.GDP.PCAP.

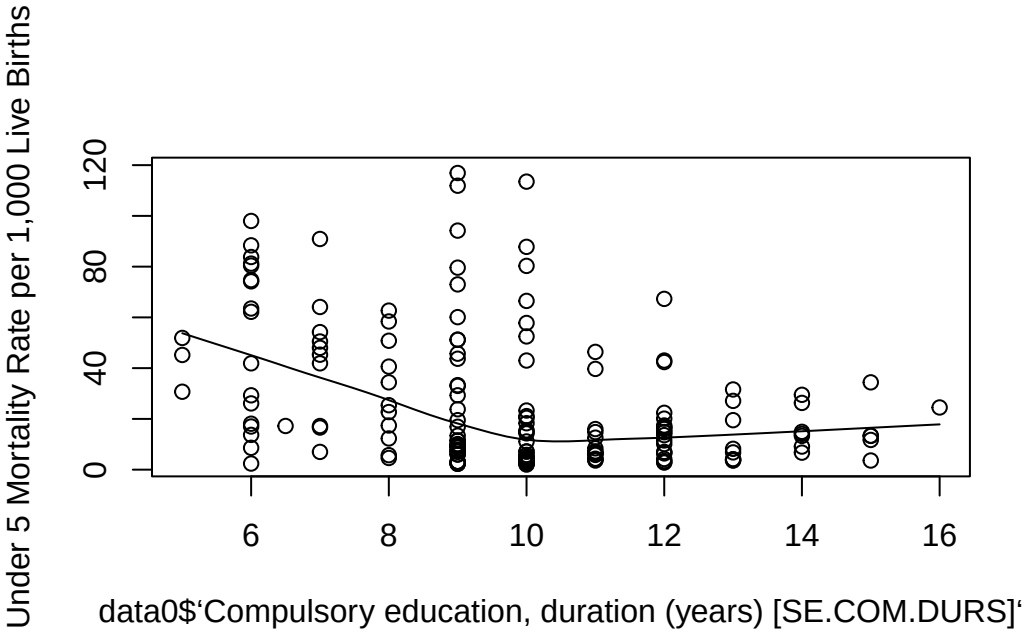



Under 5 Mortality Rate per 1,000 Live Births
0
40
80
120
0.2
0.4
0.6
0.8
data0$‘CO2 emissions (kg per PPP $ of GDP) [EN.ATM.CO2E.PP.GD

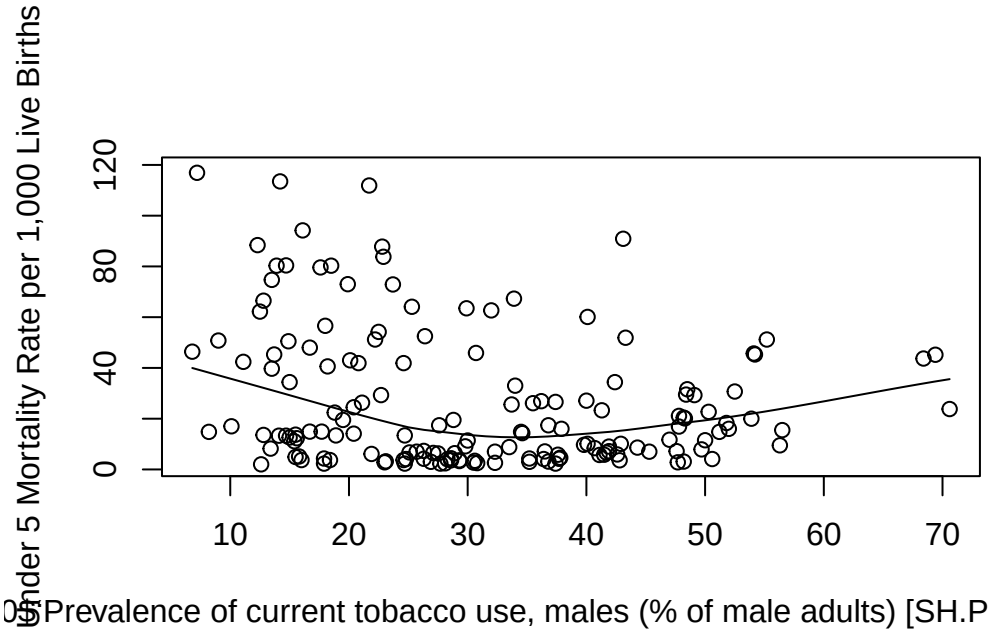

Under 5 Mortality Rate per 1,000 Live Births
0
40
80
120
0
5
10
15
20
25
30
data0$‘CO2 emissions (metric tons per capita) [EN.ATM.CO2E.PC]‘



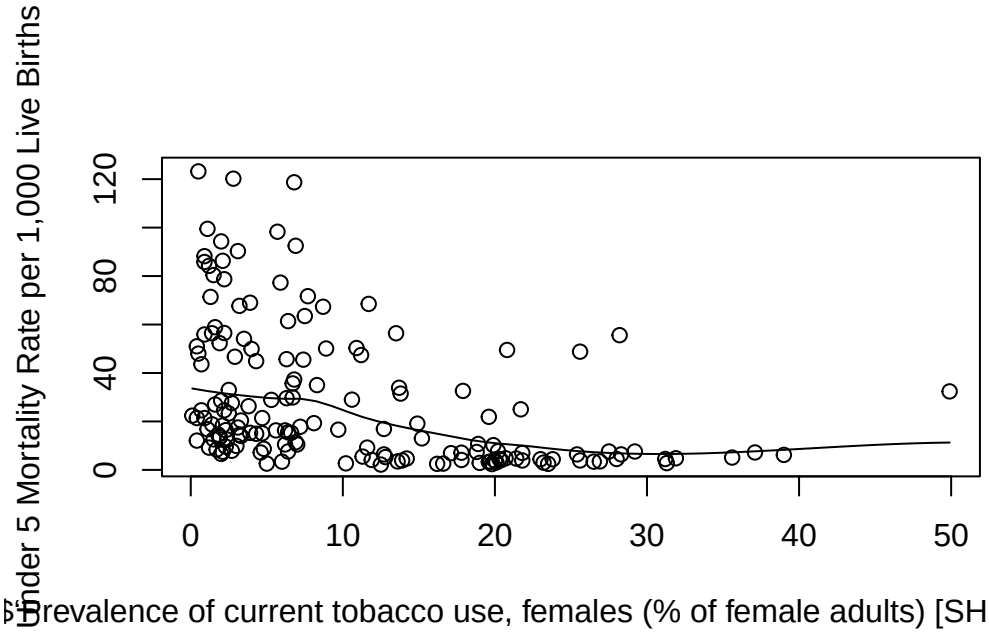

```
Warning in xy.coords(x, y, xlabel, ylabel): NAs introduced by coercion
```



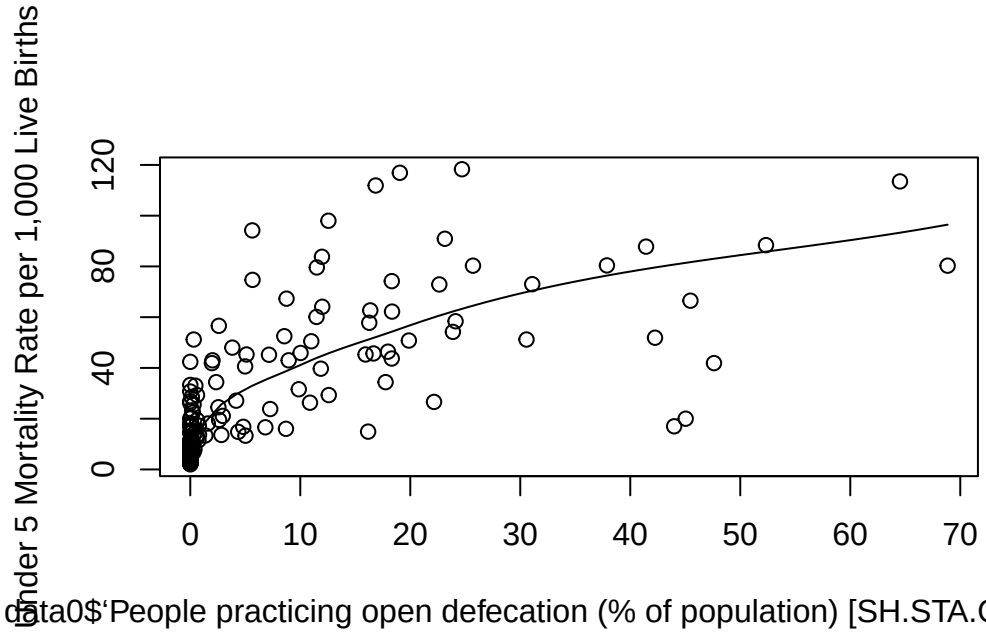

```
Warning in xy.coords(x, y, xlabel, ylabel): NAs introduced by coercion
```



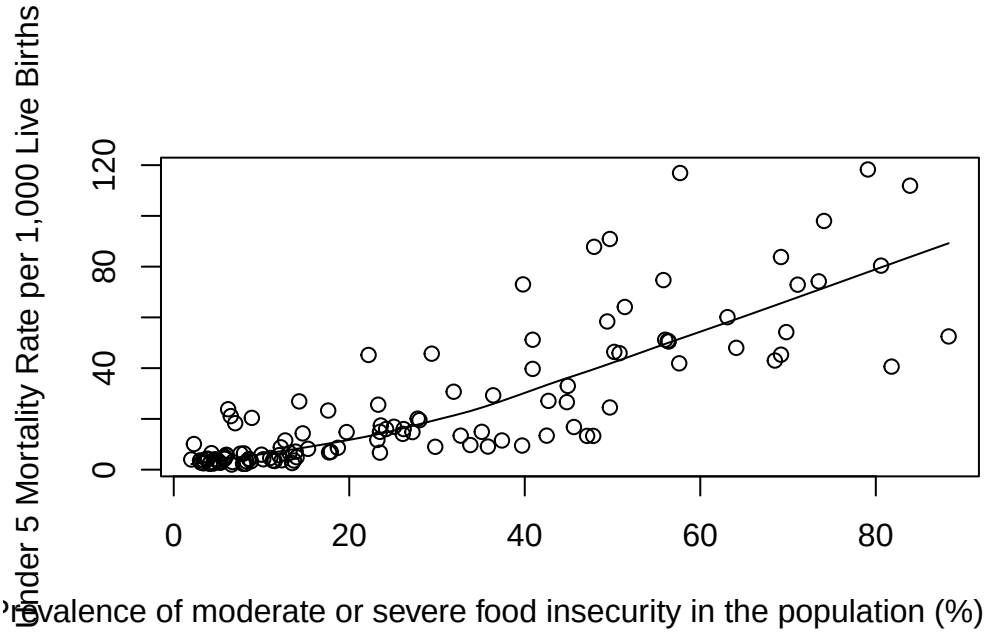

```
Warning in xy.coords(x, y, xlabel, ylabel): NAs introduced by coercion
```



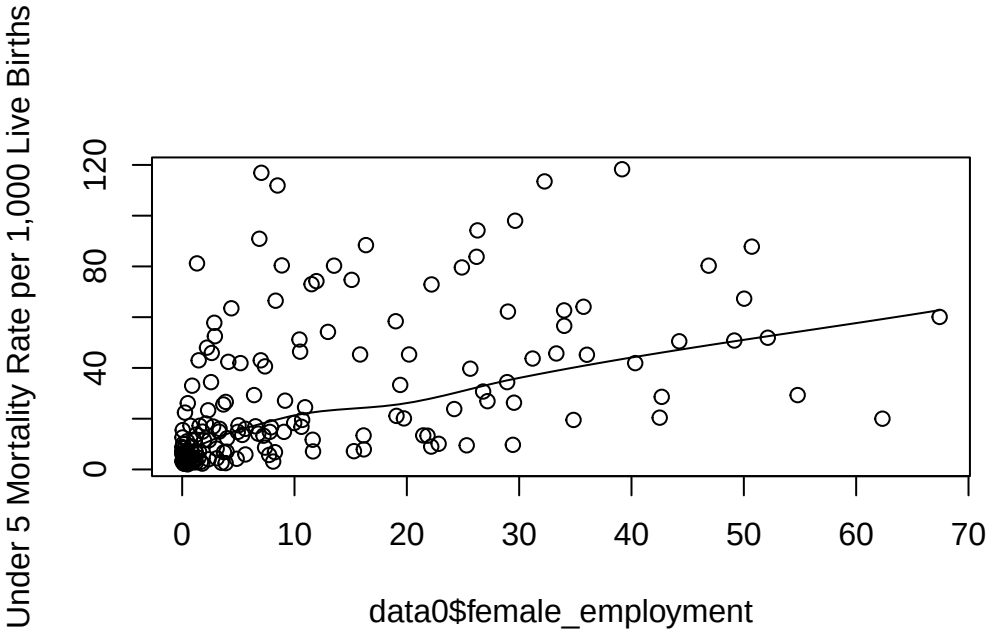

```
Warning in xy.coords(x, y, xlabel, ylabel): NAs introduced by coercion
```



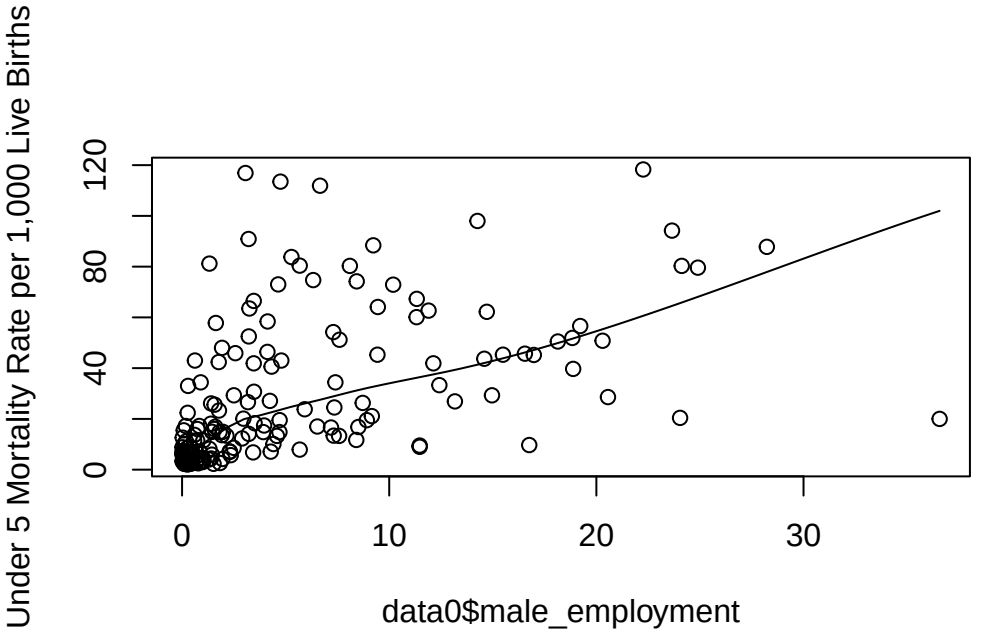

```
Warning in xy.coords(x, y, xlabel, ylabel): NAs introduced by coercion
```



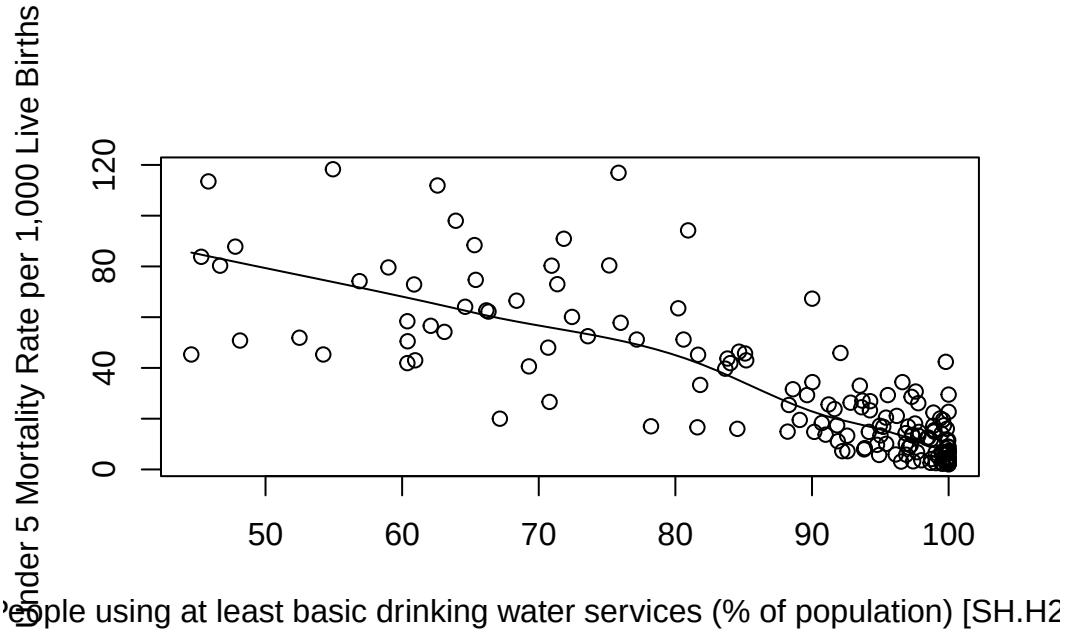

```
Warning in xy.coords(x, y, xlabel, ylabel): NAs introduced by coercion
```




```
Warning in xy.coords(x, y, xlabel, ylabel): NAs introduced by coercion
```




```
Warning in xy.coords(x, y, xlabel, ylabel): NAs introduced by coercion
```



Q11.

Project Attestation: No member of this group is using these data or same/similar questions in any other course or course project, at HSPH.

APPENDIX

```
#Load packages
library(readxl)
#Load packages
library(readxl)

# Load data
data0 <- read_xlsx("P_Data_Extract_From_Sustainable_Development_Goals_(SDGs)
                   (3) (1).xlsx")
attach(data0)

#Data exploration
# View the structure of the dataset
str(data0)
```

```r
# Summary statistics
summary(data0)

# Data Visualizations
hist(`Mortality rate, under-5 (per 1,000 live births) [SH.DYN.MORT]`)

hist(log(`Mortality rate, under-5 (per 1,000 live births) [SH.DYN.MORT]`))

hist(`CO2 emissions (kg per PPP $ of GDP) [EN.ATM.CO2E.PP.GD]`)

hist(`CO2 emissions (metric tons per capita) [EN.ATM.CO2E.PC]`)

hist(`GDP per capita (current US$) [NY.GDP.PCAP.CD]`)

scatter.smooth(data0$`CO2 emissions (kg per PPP $ of GDP)
               [EN.ATM.CO2E.PP.GD]`, data0$`Mortality rate,
               under-5 (per 1,000 live births) [SH.DYN.MORT]`,
               ylab = "Under 5 Mortality Rate per 1,000 Live Births")


scatter.smooth(data0$`CO2 emissions (metric tons per capita) [EN.ATM.CO2E.PC]`,
               data0$`Mortality rate, under-5 (per 1,000 live births) [SH.DYN.MORT]`,
               ylab = "Under 5 Mortality Rate per 1,000 Live Births")

scatter.smooth(data0$`GDP per capita (current US$) [NY.GDP.PCAP.CD]`,
               data0$`Mortality rate, under-5 (per 1,000 live births) [SH.DYN.MORT]`,
               ylab = "Under 5 Mortality Rate per 1,000 Live Births")


scatter.smooth(data0$`Compulsory education, duration (years) [SE.COM.DURS]`,
               data0$`Mortality rate,
               under-5 (per 1,000 live births) [SH.DYN.MORT]`,
               ylab = "Under 5 Mortality Rate per 1,000 Live Births")

scatter.smooth(data0$`Prevalence of current tobacco use,
               males (% of male adults) [SH.PRV.SMOK.MA]`,
               data0$`Mortality rate, under-5 (per 1,000 live births) [SH.DYN.MORT]`,
               ylab = "Under 5 Mortality Rate per 1,000 Live Births")

scatter.smooth(data0$`Prevalence of current tobacco use,
               females (% of female adults) [SH.PRV.SMOK.FE]`,
```

```
              data0$`Mortality rate, under-5, male (per 1,000 live births)
              [SH.DYN.MORT.MA]`,
              ylab = "Under 5 Mortality Rate per 1,000 Live Births")

scatter.smooth(data0$`People practicing open defecation
              (% of population) [SH.STA.ODFC.ZS]`,
              data0$`Mortality rate, under-5 (per 1,000 live births) [SH.DYN.MORT]`,
              ylab = "Under 5 Mortality Rate per 1,000 Live Births")

scatter.smooth(data0$`Prevalence of moderate or severe
              food insecurity in the population (%)
              [SN.ITK.MSFI.ZS]`, data0$`Mortality rate,
              under-5 (per 1,000 live births) [SH.DYN.MORT]`,
              ylab = "Under 5 Mortality Rate per 1,000 Live Births")

scatter.smooth(data0$`Contributing family workers, female
              (% of female employment) (modeled ILO estimate) [SL.FAM.WORK.FE.ZS]`,
              data0$`Mortality rate, under-5 (per 1,000 live births) [SH.DYN.MORT]`,
              ylab = "Under 5 Mortality Rate per 1,000 Live Births")

scatter.smooth(data0$`Contributing family workers,
              male (% of male employment) (modeled ILO estimate) [SL.FAM.WORK.MA.ZS]`,
              data0$`Mortality rate, under-5 (per 1,000 live births) [SH.DYN.MORT]`,

              ylab = "Under 5 Mortality Rate per 1,000 Live Births")

scatter.smooth(data0$`People using at least basic drinking water services
              (% of population) [SH.H2O.BASW.ZS]`,
              data0$`Mortality rate, under-5 (per 1,000 live births)
              [SH.DYN.MORT]`,
              ylab = "Under 5 Mortality Rate per 1,000 Live Births")
```