Final Assessment

Jesus Ruiz Flores

2022-11-12

I compared the CO2 and Methane Emissions of eighteen countries from 1990 to 2019.

Import libraries.

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 0.3.5   
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.4.1   
## ✔ readr 2.1.3 ✔ forcats 0.5.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(gtsummary)  
library(readxl)  
library(janitor)

##   
## Attaching package: 'janitor'  
##   
## The following objects are masked from 'package:stats':  
##   
## chisq.test, fisher.test

library(ggplot2)  
library(ggbeeswarm)   
library(forcats)

I downloaded the World Development Indicators data from the World Bank website (<https://www.worldbank.org/en/home>).

I arranged data in the sheet: “Emissions”.

Emissions <- read\_xls("data/API\_19\_DS2\_en\_excel\_v2\_4700532.xls", sheet = "Emissions")%>%  
 clean\_names()%>%  
 rename(Country = "country\_name",  
 Code = "country\_code",  
 Indicator = "indicator\_name",  
 `1990` = "x1990",  
 `1991` = "x1991",  
 `1992` = "x1992",  
 `1993` = "x1993",  
 `1994` = "x1994",  
 `1995` = "x1995",  
 `1996` = "x1996",  
 `1997` = "x1997",  
 `1998` = "x1998",  
 `1999` = "x1999",  
 `2000` = "x2000",  
 `2001` = "x2001",  
 `2002` = "x2002",  
 `2003` = "x2003",  
 `2004` = "x2004",  
 `2005` = "x2005",  
 `2006` = "x2006",  
 `2007` = "x2007",  
 `2008` = "x2008",  
 `2009` = "x2009",  
 `2010` = "x2010",  
 `2011` = "x2011",  
 `2012` = "x2012",  
 `2013` = "x2013",  
 `2014` = "x2014",  
 `2015` = "x2015",  
 `2016` = "x2016",  
 `2017` = "x2017",  
 `2018` = "x2018",  
 `2019` = "x2019")%>%  
 mutate(Indicator\_f = factor(Indicator, labels = c("CO2", "Methane")))  
   
  
str(Emissions)

## tibble [36 × 34] (S3: tbl\_df/tbl/data.frame)  
## $ Country : chr [1:36] "Angola" "Angola" "Albania" "Albania" ...  
## $ Code : chr [1:36] "AGO" "AGO" "ALB" "ALB" ...  
## $ Indicator : chr [1:36] "Methane emissions (kt of CO2 equivalent)" "CO2 emissions (kt)" "Methane emissions (kt of CO2 equivalent)" "CO2 emissions (kt)" ...  
## $ 1990 : num [1:36] 21220 6560 3540 5980 20 ...  
## $ 1991 : num [1:36] 21490 6670 3500 4060 20 ...  
## $ 1992 : num [1:36] 21950 6880 3410 2220 20 ...  
## $ 1993 : num [1:36] 22580 9270 3490 2060 30 ...  
## $ 1994 : num [1:36] 23480 11300 4080 2070 30 ...  
## $ 1995 : num [1:36] 25230 12720 4080 1930 30 ...  
## $ 1996 : num [1:36] 25390 15440 3910 1940 30 ...  
## $ 1997 : num [1:36] 25690 16160 3700 1470 30 ...  
## $ 1998 : num [1:36] 27890 16770 3560 1790 30 ...  
## $ 1999 : num [1:36] 27090 17610 3610 2970 30 ...  
## $ 2000 : num [1:36] 27080 16200 3670 3170 30 ...  
## $ 2001 : num [1:36] 27090 15960 3610 3230 30 ...  
## $ 2002 : num [1:36] 28610 15690 3560 3760 40 ...  
## $ 2003 : num [1:36] 30730 16760 3570 4070 40 ...  
## $ 2004 : num [1:36] 31620 17450 3500 4250 40 ...  
## $ 2005 : num [1:36] 33100 15810 3490 4030 40 ...  
## $ 2006 : num [1:36] 32590 16560 3440 4010 40 ...  
## $ 2007 : num [1:36] 35330 16970 3300 4140 50 ...  
## $ 2008 : num [1:36] 36400 19280 3130 4080 50 ...  
## $ 2009 : num [1:36] 35530 21150 3020 4220 50 ...  
## $ 2010 : num [1:36] 36900 22800 3030 4450 50 ...  
## $ 2011 : num [1:36] 36640 23870 3020 4850 50 ...  
## $ 2012 : num [1:36] 35770 23870 3060 4360 50 ...  
## $ 2013 : num [1:36] 35760 26960 3060 4440 50 ...  
## $ 2014 : num [1:36] 34900 29630 3090 4820 50 ...  
## $ 2015 : num [1:36] 37110 31650 3110 4620 50 ...  
## $ 2016 : num [1:36] 37430 29760 3090 4480 50 ...  
## $ 2017 : num [1:36] 37620 24250 3050 5140 50 ...  
## $ 2018 : num [1:36] 34000 23960 3010 5110 50 ...  
## $ 2019 : num [1:36] 35090 25210 2850 4830 50 ...  
## $ Indicator\_f: Factor w/ 2 levels "CO2","Methane": 2 1 2 1 2 1 2 1 2 1 ...

glimpse(Emissions)

## Rows: 36  
## Columns: 34  
## $ Country <chr> "Angola", "Angola", "Albania", "Albania", "Andorra", "Ando…  
## $ Code <chr> "AGO", "AGO", "ALB", "ALB", "AND", "AND", "ARB", "ARB", "A…  
## $ Indicator <chr> "Methane emissions (kt of CO2 equivalent)", "CO2 emissions…  
## $ `1990` <dbl> 21220, 6560, 3540, 5980, 20, 410, 407350, 636129, 16460, 5…  
## $ `1991` <dbl> 21490.0, 6670.0, 3500.0, 4060.0, 20.0, 410.0, 394510.0, 63…  
## $ `1992` <dbl> 21950.0, 6880.0, 3410.0, 2220.0, 20.0, 410.0, 409310.0, 69…  
## $ `1993` <dbl> 22580.0, 9270.0, 3490.0, 2060.0, 30.0, 410.0, 435070.0, 74…  
## $ `1994` <dbl> 23480.0, 11300.0, 4080.0, 2070.0, 30.0, 410.0, 457850.0, 7…  
## $ `1995` <dbl> 25230.0, 12720.0, 4080.0, 1930.0, 30.0, 430.0, 473230.0, 8…  
## $ `1996` <dbl> 25390.0, 15440.0, 3910.0, 1940.0, 30.0, 450.0, 484770.0, 8…  
## $ `1997` <dbl> 25690, 16160, 3700, 1470, 30, 470, 506850, 884777, 21520, …  
## $ `1998` <dbl> 27890.0, 16770.0, 3560.0, 1790.0, 30.0, 490.0, 530690.0, 8…  
## $ `1999` <dbl> 27090.0, 17610.0, 3610.0, 2970.0, 30.0, 510.0, 541430.0, 9…  
## $ `2000` <dbl> 27080.0, 16200.0, 3670.0, 3170.0, 30.0, 520.0, 565070.0, 9…  
## $ `2001` <dbl> 27090.0, 15960.0, 3610.0, 3230.0, 30.0, 520.0, 555350.0, 9…  
## $ `2002` <dbl> 28610, 15690, 3560, 3760, 40, 530, 552450, 1024787, 23720,…  
## $ `2003` <dbl> 30730, 16760, 3570, 4070, 40, 530, 569780, 1055756, 26970,…  
## $ `2004` <dbl> 31620, 17450, 3500, 4250, 40, 560, 589840, 1112585, 28540,…  
## $ `2005` <dbl> 33100, 15810, 3490, 4030, 40, 580, 610510, 1188504, 30890,…  
## $ `2006` <dbl> 32590, 16560, 3440, 4010, 40, 550, 621190, 1251800, 32700,…  
## $ `2007` <dbl> 35330, 16970, 3300, 4140, 50, 540, 625570, 1303116, 33440,…  
## $ `2008` <dbl> 36400, 19280, 3130, 4080, 50, 540, 640570, 1408053, 34840,…  
## $ `2009` <dbl> 35530, 21150, 3020, 4220, 50, 520, 628790, 1455347, 34070,…  
## $ `2010` <dbl> 36900, 22800, 3030, 4450, 50, 520, 647360, 1541777, 35720,…  
## $ `2011` <dbl> 36640, 23870, 3020, 4850, 50, 490, 629390, 1574878, 38360,…  
## $ `2012` <dbl> 35770, 23870, 3060, 4360, 50, 490, 684220, 1683788, 40310,…  
## $ `2013` <dbl> 35760, 26960, 3060, 4440, 50, 480, 669140, 1713792, 42440,…  
## $ `2014` <dbl> 34900, 29630, 3090, 4820, 50, 460, 660750, 1771948, 42870,…  
## $ `2015` <dbl> 37110, 31650, 3110, 4620, 50, 470, 671170, 1818105, 45710,…  
## $ `2016` <dbl> 37430, 29760, 3090, 4480, 50, 470, 696080, 1838603, 47360,…  
## $ `2017` <dbl> 37620, 24250, 3050, 5140, 50, 470, 715760, 1858498, 46900,…  
## $ `2018` <dbl> 34000, 23960, 3010, 5110, 50, 490, 733430, 1836883, 47460,…  
## $ `2019` <dbl> 35090, 25210, 2850, 4830, 50, 500, 747470, 1886219, 49800,…  
## $ Indicator\_f <fct> Methane, CO2, Methane, CO2, Methane, CO2, Methane, CO2, Me…

tibble(Emissions)

## # A tibble: 36 × 34  
## Country Code Indic…¹ `1990` `1991` `1992` `1993` `1994` `1995` `1996` `1997`  
## <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Angola AGO Methan… 2.12e4 2.15e4 2.20e4 2.26e4 2.35e4 2.52e4 2.54e4 2.57e4  
## 2 Angola AGO CO2 em… 6.56e3 6.67e3 6.88e3 9.27e3 1.13e4 1.27e4 1.54e4 1.62e4  
## 3 Albania ALB Methan… 3.54e3 3.5 e3 3.41e3 3.49e3 4.08e3 4.08e3 3.91e3 3.70e3  
## 4 Albania ALB CO2 em… 5.98e3 4.06e3 2.22e3 2.06e3 2.07e3 1.93e3 1.94e3 1.47e3  
## 5 Andorra AND Methan… 2.00e1 2.00e1 2.00e1 3.00e1 3.00e1 3.00e1 3.00e1 3.00e1  
## 6 Andorra AND CO2 em… 4.1 e2 4.1 e2 4.1 e2 4.1 e2 4.1 e2 4.3 e2 4.5 e2 4.7 e2  
## 7 Arab W… ARB Methan… 4.07e5 3.95e5 4.09e5 4.35e5 4.58e5 4.73e5 4.85e5 5.07e5  
## 8 Arab W… ARB CO2 em… 6.36e5 6.38e5 6.93e5 7.46e5 7.91e5 8.15e5 8.44e5 8.85e5  
## 9 United… ARE Methan… 1.65e4 1.79e4 1.78e4 1.83e4 1.97e4 2.04e4 2.13e4 2.15e4  
## 10 United… ARE CO2 em… 5.52e4 6.16e4 5.97e4 6.36e4 7.08e4 7.52e4 7.85e4 8.14e4  
## # … with 26 more rows, 23 more variables: `1998` <dbl>, `1999` <dbl>,  
## # `2000` <dbl>, `2001` <dbl>, `2002` <dbl>, `2003` <dbl>, `2004` <dbl>,  
## # `2005` <dbl>, `2006` <dbl>, `2007` <dbl>, `2008` <dbl>, `2009` <dbl>,  
## # `2010` <dbl>, `2011` <dbl>, `2012` <dbl>, `2013` <dbl>, `2014` <dbl>,  
## # `2015` <dbl>, `2016` <dbl>, `2017` <dbl>, `2018` <dbl>, `2019` <dbl>,  
## # Indicator\_f <fct>, and abbreviated variable name ¹​Indicator

In the next table I summarised the mean levels, in kt, of “Methane” and “CO2” of each country from 1990 to 2019.

Emissions%>%  
 group\_by(`Country`,`Indicator\_f`)%>%  
 summarise(`Mean emissions from 1990 to 2019 (kt)` = mean(`1990`:`2019`),   
 `Sd` = sd(`1990`:`2019`))

## `summarise()` has grouped output by 'Country'. You can override using the  
## `.groups` argument.

## # A tibble: 36 × 4  
## # Groups: Country [18]  
## Country Indicator\_f Mean emissions from 1990 to 2019 (kt…¹ Sd  
## <chr> <fct> <dbl> <dbl>  
## 1 Albania CO2 5405 3.32e2  
## 2 Albania Methane 3195. 2.00e2  
## 3 Andorra CO2 455 2.64e1  
## 4 Andorra Methane 35.0 9.09e0  
## 5 Angola CO2 15884. 5.38e3  
## 6 Angola Methane 28155. 4.00e3  
## 7 Antigua and Barbuda CO2 364. 8.96e1  
## 8 Antigua and Barbuda Methane 165. 2.06e1  
## 9 Arab World CO2 1261173. 3.61e5  
## 10 Arab World Methane 577409. 9.82e4  
## # … with 26 more rows, and abbreviated variable name  
## # ¹​`Mean emissions from 1990 to 2019 (kt)`

In the next table I summarised the countries that polluted, in average, with more than 200000 kt of CO2 from 1990 to 2019

Emissions%>%  
 group\_by(`Country`,`Indicator\_f`)%>%  
 filter(mean(`1990`:`2019`)>200000)%>%  
 filter(Indicator\_f== "CO2")%>%  
 summarise(`Mean emissions from 1990 to 2019 (kt)` = mean(`1990`:`2019`))

## `summarise()` has grouped output by 'Country'. You can override using the  
## `.groups` argument.

## # A tibble: 2 × 3  
## # Groups: Country [2]  
## Country Indicator\_f `Mean emissions from 1990 to 2019 (kt)`  
## <chr> <fct> <dbl>  
## 1 Arab World CO2 1261173.  
## 2 Australia CO2 325080.

In the next table I summarised, in ascending order, the levels of CO2 of all the countries from 1990 to 2019. It is very clear that developed, populated and industrialized countries polluted more.

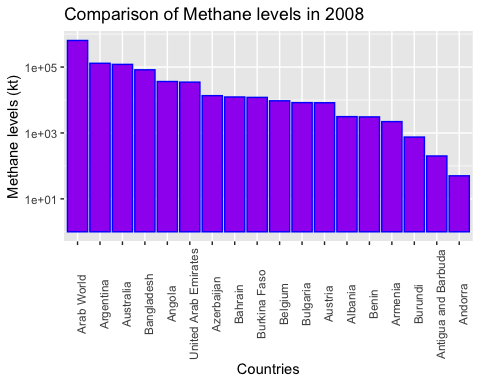
Emissions%>%  
 group\_by(`Country`,`Indicator\_f`)%>%  
 filter(Indicator\_f== "CO2")%>%  
 summarise(`Mean emissions from 1990 to 2019 (kt)` = mean(`1990`:`2019`))%>%  
 arrange(`Mean emissions from 1990 to 2019 (kt)`)

## `summarise()` has grouped output by 'Country'. You can override using the  
## `.groups` argument.

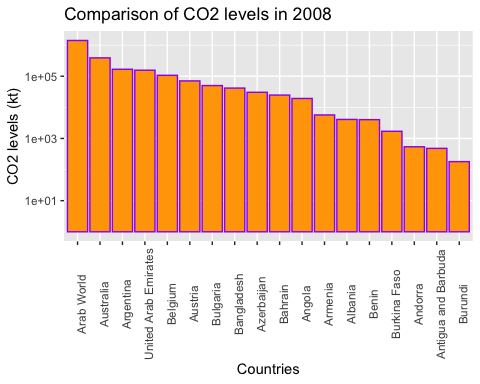
## # A tibble: 18 × 3  
## # Groups: Country [18]  
## Country Indicator\_f `Mean emissions from 1990 to 2019 (kt)`  
## <chr> <fct> <dbl>  
## 1 Antigua and Barbuda CO2 364.  
## 2 Burundi CO2 445   
## 3 Andorra CO2 455   
## 4 Burkina Faso CO2 2750   
## 5 Benin CO2 3815   
## 6 Albania CO2 5405   
## 7 Armenia CO2 13010.  
## 8 Angola CO2 15884.  
## 9 Bahrain CO2 22000.  
## 10 Azerbaijan CO2 44495   
## 11 Bangladesh CO2 51130.  
## 12 Bulgaria CO2 56375   
## 13 Austria CO2 61520.  
## 14 Belgium CO2 101160.  
## 15 United Arab Emirates CO2 122035   
## 16 Argentina CO2 134210   
## 17 Australia CO2 325080.  
## 18 Arab World CO2 1261173.

I decided to make a plot to view the relationship between CO2 and Methane levels against the countries in 2008 (a plot for each one).

Emissions\_CO2 <- Emissions%>%  
 filter(Indicator\_f == "CO2")  
  
Emissions\_Methane <- Emissions%>%  
 filter(Indicator\_f == "Methane")  
   
Emissions\_Methane%>%  
 ggplot(mapping = aes(x = reorder(Country, -`2008`), y = `2008`, col = Indicator\_f))+  
 geom\_col(fill = "purple", colour = "blue")+  
 scale\_y\_log10()+  
 theme(axis.text.x = element\_text(angle = 90))+  
 labs(title = "Comparison of Methane levels in 2008",  
 x = "Countries", y = "Methane levels (kt)")



Emissions\_CO2%>%  
 ggplot(mapping = aes(x = reorder(Country, -`2008`), y = `2008`, col = Indicator\_f))+   
 geom\_col(fill = "orange", colour = "purple", position = "dodge")+   
 scale\_y\_log10()+  
 theme(axis.text.x = element\_text(angle = 90))+  
 labs(title = "Comparison of CO2 levels in 2008",  
 x = "Countries", y = "CO2 levels (kt)")



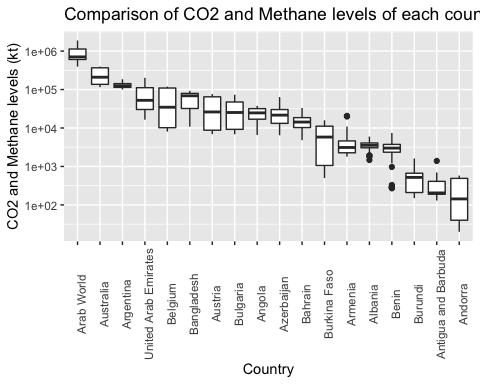
Then, I created Emissions\_long using the “pivot longer” to arrange the data and having a wide visualization of them.

Emissions\_long <- Emissions%>%  
 pivot\_longer(cols = 4:33,  
 names\_to = c("Year"),  
 values\_to = "Methane and CO2 Emissions (kt)")  
   
  
Emissions\_long%>%  
 filter(`Methane and CO2 Emissions (kt)` > 485700)

## # A tibble: 53 × 6  
## Country Code Indicator Indic…¹ Year Metha…²  
## <chr> <chr> <chr> <fct> <chr> <dbl>  
## 1 Arab World ARB Methane emissions (kt of CO2 equivale… Methane 1997 506850.  
## 2 Arab World ARB Methane emissions (kt of CO2 equivale… Methane 1998 530690.  
## 3 Arab World ARB Methane emissions (kt of CO2 equivale… Methane 1999 541430.  
## 4 Arab World ARB Methane emissions (kt of CO2 equivale… Methane 2000 565070.  
## 5 Arab World ARB Methane emissions (kt of CO2 equivale… Methane 2001 555350.  
## 6 Arab World ARB Methane emissions (kt of CO2 equivale… Methane 2002 552450.  
## 7 Arab World ARB Methane emissions (kt of CO2 equivale… Methane 2003 569780.  
## 8 Arab World ARB Methane emissions (kt of CO2 equivale… Methane 2004 589840.  
## 9 Arab World ARB Methane emissions (kt of CO2 equivale… Methane 2005 610510.  
## 10 Arab World ARB Methane emissions (kt of CO2 equivale… Methane 2006 621190.  
## # … with 43 more rows, and abbreviated variable names ¹​Indicator\_f,  
## # ²​`Methane and CO2 Emissions (kt)`

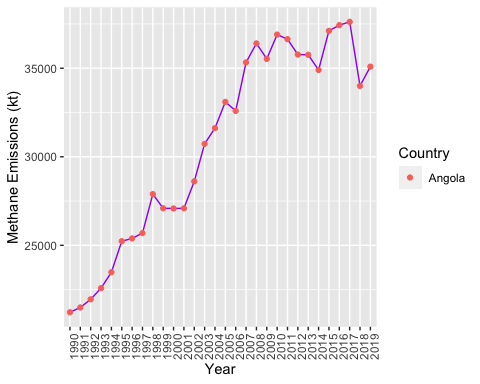
I created the next Box plot comparing the two indicators, CO2 and Methane, versus the whole countries.By visualizing the plot we see that countries like Belgium, Burkina Faso and Andorra had a considerable variation of polluting levels in that period of time. In countries like Benin of Albania, they were consistent with their pollution levels, but in some years they were polluting less.

Emissions\_long\_CO2 <- Emissions\_long%>%  
 filter(Indicator\_f == "CO2")  
  
Emissions\_long\_Methane <- Emissions\_long%>%  
 filter(Indicator\_f == "Methane")  
  
  
ggplot(data = Emissions\_long, mapping = aes(x = reorder(Country, -`Methane and CO2 Emissions (kt)`), y = `Methane and CO2 Emissions (kt)`))+   
 geom\_boxplot()+  
 theme(axis.text.x = element\_text(angle = 90))+  
 scale\_y\_log10()+  
 labs(title = "Comparison of CO2 and Methane levels of each country from 1990 to 2019.",  
 x = "Country", y = "CO2 and Methane levels (kt)")



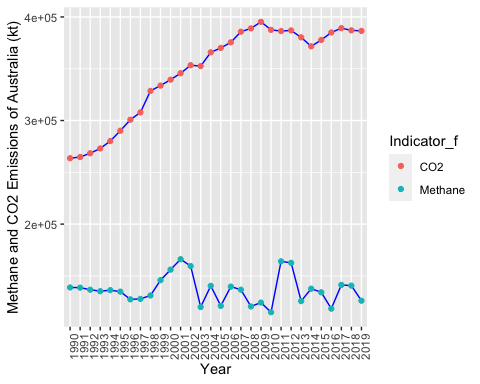
And I also learned how to make a subset of data and make a plot of it.

Angola <- Emissions\_long %>%  
 filter(Country == "Angola", Indicator\_f == "Methane") %>%  
 rename(`Methane Emissions (kt)` = "Methane and CO2 Emissions (kt)")  
  
  
ggplot(data = Angola, mapping = aes(x = Year, y = `Methane Emissions (kt)`, group = 1, col = Country)) +   
 geom\_line(colour = "purple")+  
 geom\_point()+  
 theme(axis.text.x = element\_text(angle = 90))



I decided to compare the levels of CO2 and Methane of Australia from 1990 to 2019.

Australia <- Emissions\_long %>%  
 filter(Country == "Australia")%>%  
 rename(`Methane and CO2 Emissions of Australia (kt)` = "Methane and CO2 Emissions (kt)")  
  
ggplot(data = Australia, mapping = aes(x = Year, y = `Methane and CO2 Emissions of Australia (kt)`,colour = Indicator\_f, fill = Indicator\_f, group = Indicator\_f)) +   
 geom\_line(colour = "blue")+  
 geom\_point()+  
 theme(axis.text.x = element\_text(angle = 90))



And comparing just the CO2 levels of two countries, in this case, Austria and Belgium.

`Austria and Belgium` <- Emissions\_long %>%  
 filter(Country == "Austria"| Country == "Belgium", Indicator\_f == "CO2")%>%  
 rename(`CO2 levels (kt)` = "Methane and CO2 Emissions (kt)")  
  
  
ggplot(data = `Austria and Belgium`, mapping = aes(x = Year, y = `CO2 levels (kt)`, group = Country, colour = Country)) +  
 geom\_line()+  
 geom\_point(size = 2.5, shape = 7.8)+  
 scale\_y\_log10()+  
 theme(axis.text.x = element\_text(angle = 90))

