COVID\_economic\_analysis

Jesus Sanchez

2022-09-05

Let’s first load our packages

library(tidyverse)

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

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.6 v dplyr 1.0.7  
## v tidyr 1.1.4 v stringr 1.4.1  
## v readr 2.1.1 v forcats 0.5.1

## Warning: package 'stringr' was built under R version 4.1.3

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(readxl)

Then, using the “readx1” package we’ll load our data

covid\_2022 <- read.csv("covid\_data\_04\_09\_2022.csv")  
covid\_2020 <- read.csv("covid\_data\_20\_01\_2020.csv")  
gdp\_rate <- read.csv("GDP Growth Rate by Country.csv")  
gdp\_per\_capita <- read.csv("GDP Per Capita by Country.csv")  
infation <- read.csv("Inflation Rate by Country.csv")

#Procesing data

We want to analyze if there is a correlation between COVID and some economic metrics. In order to get it, we have to merge our 4 data frames in just one.

covid\_2020 <- covid\_2020 %>% rename(Country.Name=Country.Region)  
covid\_2022 <- covid\_2022 %>% rename(Country.Name=Country..Other)  
#We start by renaming one column to use the merge function

Before we merge our data, we have to make the rows names be the same, in order to get it we’ll change some abbreviate names

covid\_2020[covid\_2020=="USA"] <- "United States"  
covid\_2020[covid\_2020=="S. Korea"] <- "South Korea"  
covid\_2020[covid\_2020=="UK"] <- "United Kingdom"  
covid\_2022[covid\_2022=="USA"] <- "United States"  
covid\_2022[covid\_2022=="S. Korea"] <- "South Korea"  
covid\_2022[covid\_2022=="UK"] <- "United Kingdom"

all\_togheter\_2020 <- merge(x = covid\_2020,   
 y = gdp\_rate, by="Country.Name",   
 all=T)  
all\_togheter\_2020 <- merge(x=all\_togheter\_2020,  
 y=gdp\_per\_capita, by ="Country.Name",  
 all=T)  
all\_togheter\_2020 <- merge(x=all\_togheter\_2020,  
 y=infation, by="Country.Name",  
 all=T)  
  
#we do the same with data from 2022  
all\_togheter\_2022 <- merge(x = covid\_2022,   
 y = gdp\_rate, by="Country.Name",   
 all=T)  
all\_togheter\_2022 <- merge(x=all\_togheter\_2022,  
 y=gdp\_per\_capita, by ="Country.Name",  
 all=T)  
all\_togheter\_2022 <- merge(x=all\_togheter\_2022,  
 y=infation, by="Country.Name",  
 all=T)

#Cleaning data

We have merged our data, but now the column names are not readables, we don’t know what is the gdp\_rate, gdp\_per capita and the inflation rate. So, let’s change the column labels

colnames(all\_togheter\_2022) #first we visualize the colum names

## [1] "Country.Name" "X." "Total.Cases"   
## [4] "New.Cases" "Total.Deaths" "New.Deaths"   
## [7] "Total.Recovered" "New.Recovered" "Active.Cases"   
## [10] "Serious..Critical" "Tot.Cases..1M.pop" "Deaths..1M.pop"   
## [13] "Total.Tests" "Tests...1M.pop" "Population"   
## [16] "X2021.x" "X2020.x" "X2019.x"   
## [19] "X2018.x" "X2017.x" "X2021.y"   
## [22] "X2020.y" "X2019.y" "X2018.y"   
## [25] "X2017.y" "X2021" "X2020"   
## [28] "X2019" "X2018" "X2017"

head(all\_togheter\_2022) #display the first rows

## Country.Name X. Total.Cases New.Cases Total.Deaths New.Deaths  
## 1 NA 721 NA 15 NA  
## 2 Afghanistan 117 194163 NA 7782 NA  
## 3 Africa NA 12616085 NA 257418 NA  
## 4 Albania 101 330193 NA 3582 NA  
## 5 Algeria 105 270443 NA 6879 NA  
## 6 American Samoa NA NA NA NA NA  
## Total.Recovered New.Recovered Active.Cases Serious..Critical  
## 1 706 0 0  
## 2 172,168 14,213 1124  
## 3 11,943,257 415,410 1026  
## 4 322,849 3,762 NA  
## 5 182,067 81,497 6  
## 6 <NA> <NA> <NA> NA  
## Tot.Cases..1M.pop Deaths..1M.pop Total.Tests Tests...1M.pop Population  
## 1 NA NA NA NA NA  
## 2 4755 191 1073129 26282 40831974  
## 3 NA NA NA NA NA  
## 4 115011 1248 1941032 676087 2870979  
## 5 5935 151 230861 5066 45570311  
## 6 NA NA NA NA NA  
## X2021.x X2020.x X2019.x X2018.x X2017.x X2021.y X2020.y X2019.y X2018.y  
## 1 NA NA NA NA NA NA NA NA NA  
## 2 0.0000 -0.0235 0.0391 0.0119 0.0265 0 517 494 486  
## 3 NA NA NA NA NA NA NA NA NA  
## 4 0.0854 -0.0348 0.0209 0.0402 0.0380 6494 5332 5396 5288  
## 5 0.0385 -0.0510 0.0100 0.0110 0.0130 3765 3307 3990 4142  
## 6 0.0000 0.0392 -0.0049 0.0267 -0.0699 0 12845 11715 11522  
## X2017.y X2021 X2020 X2019 X2018 X2017  
## 1 NA NA NA NA NA NA  
## 2 517 0.0000 0.0000 0.0230 0.0063 0.0498  
## 3 NA NA NA NA NA NA  
## 4 4531 0.0204 0.0162 0.0141 0.0203 0.0206  
## 5 4110 0.0723 0.0242 0.0195 0.0427 0.0559  
## 6 11004 NA NA NA NA NA

filter(gdp\_per\_capita,Country.Name=="Afghanistan")

## Country.Name X2021 X2020 X2019 X2018 X2017  
## 1 Afghanistan 0 517 494 486 517

filter(gdp\_rate, Country.Name=="Afghanistan")

## Country.Name X2021 X2020 X2019 X2018 X2017  
## 1 Afghanistan 0 -0.0235 0.0391 0.0119 0.0265

filter(infation, Country.Name=="Afghanistan")

## Country.Name X2021 X2020 X2019 X2018 X2017  
## 1 Afghanistan 0 0 0.023 0.0063 0.0498

#filter the Afghanistan row to match the columns with their real name  
all\_togheter\_2020 <- rename(all\_togheter\_2020,   
 "gdp\_per\_capita\_2021"="X2021.y",   
 "gdp\_per\_capita\_2020"="X2020.y",   
 "gdp\_per\_capita\_2019"="X2019.y",   
 "gdp\_per\_capita\_2018"="X2018.y",   
 "gdp\_per\_capita\_2017"="X2017.y",   
 "gdp\_rate\_2021"="X2021.x",   
 "gdp\_rate\_2020"="X2020.x",  
 "gdp\_rate\_2019"="X2019.x",  
 "gdp\_rate\_2018"="X2018.x",   
 "gdp\_rate\_2017"="X2017.x",   
 "inflation\_rate\_2021"="X2021",   
 "inflation\_rate\_2020"="X2020",   
 "inflation\_rate\_2019"="X2019",   
 "inflation\_rate\_2018"="X2018",   
 "inflation\_rate\_2017"="X2017")  
#renaming columns from 2020 data  
  
all\_togheter\_2022 <- rename(all\_togheter\_2022,   
 "gdp\_per\_capita\_2021"="X2021.y",   
 "gdp\_per\_capita\_2020"="X2020.y",   
 "gdp\_per\_capita\_2019"="X2019.y",   
 "gdp\_per\_capita\_2018"="X2018.y",   
 "gdp\_per\_capita\_2017"="X2017.y",   
 "gdp\_rate\_2021"="X2021.x",   
 "gdp\_rate\_2020"="X2020.x",  
 "gdp\_rate\_2019"="X2019.x",  
 "gdp\_rate\_2018"="X2018.x",   
 "gdp\_rate\_2017"="X2017.x",   
 "inflation\_rate\_2021"="X2021",   
 "inflation\_rate\_2020"="X2020",   
 "inflation\_rate\_2019"="X2019",   
 "inflation\_rate\_2018"="X2018",   
 "inflation\_rate\_2017"="X2017")  
#renaming columns from 2022 data  
  
all\_togheter\_2022 %>%  
 filter(Country.Name %in% c("Europe", "Africa", "South America",  
 "Oceania", "Asia", "North America",  
 "Total:", "World")) #analizing what rows we will delete

## Country.Name X. Total.Cases New.Cases Total.Deaths New.Deaths  
## 1 Africa NA 12616085 NA 257418 NA  
## 2 Asia NA 184743469 39257 1469343 78  
## 3 Europe NA 222253960 NA 1904758 NA  
## 4 North America NA 114689858 990 1526246 5  
## 5 Oceania NA 12191895 5539 19637 10  
## 6 South America NA 63731409 NA 1326129 NA  
## 7 Total: NA 12616085 NA 257418 NA  
## 8 Total: NA 12191895 5539 19637 10  
## 9 Total: NA 222253960 NA 1904758 NA  
## 10 Total: NA 184743469 39257 1469343 78  
## 11 Total: NA 610227397 45786 6503546 93  
## 12 Total: NA 114689858 990 1526246 5  
## 13 Total: NA 63731409 NA 1326129 NA  
## 14 Total: NA 721 NA 15 NA  
## 15 World NA 610227397 45786 6503546 93  
## Total.Recovered New.Recovered Active.Cases Serious..Critical  
## 1 11,943,257 415,410 1026  
## 2 176,662,607 +181,088 6,611,519 12443  
## 3 216,173,591 +29,100 4,175,611 9106  
## 4 108,543,243 +7,154 4,620,369 9090  
## 5 11,973,879 +1,693 198,379 144  
## 6 61,754,756 +5,400 650,524 10505  
## 7 11,943,257 415,410 1026  
## 8 11,973,879 +1,693 198,379 144  
## 9 216,173,591 4,175,611 9106  
## 10 176,662,607 +181,088 6,611,519 12443  
## 11 587,052,039 +224,435 16,671,812 42314  
## 12 108,543,243 +7,154 4,620,369 9090  
## 13 61,754,756 650,524 10505  
## 14 706 0 0  
## 15 587,052,039 +224,435 16,671,812 42314  
## Tot.Cases..1M.pop Deaths..1M.pop Total.Tests Tests...1M.pop Population  
## 1 NA NA NA NA NA  
## 2 NA NA NA NA NA  
## 3 NA NA NA NA NA  
## 4 NA NA NA NA NA  
## 5 NA NA NA NA NA  
## 6 NA NA NA NA NA  
## 7 NA NA NA NA NA  
## 8 NA NA NA NA NA  
## 9 NA NA NA NA NA  
## 10 NA NA NA NA NA  
## 11 78286 834.3 NA NA NA  
## 12 NA NA NA NA NA  
## 13 NA NA NA NA NA  
## 14 NA NA NA NA NA  
## 15 78286 834.3 NA NA NA  
## gdp\_rate\_2021 gdp\_rate\_2020 gdp\_rate\_2019 gdp\_rate\_2018 gdp\_rate\_2017  
## 1 NA NA NA NA NA  
## 2 NA NA NA NA NA  
## 3 NA NA NA NA NA  
## 4 NA NA NA NA NA  
## 5 NA NA NA NA NA  
## 6 NA NA NA NA NA  
## 7 NA NA NA NA NA  
## 8 NA NA NA NA NA  
## 9 NA NA NA NA NA  
## 10 NA NA NA NA NA  
## 11 NA NA NA NA NA  
## 12 NA NA NA NA NA  
## 13 NA NA NA NA NA  
## 14 NA NA NA NA NA  
## 15 NA NA NA NA NA  
## gdp\_per\_capita\_2021 gdp\_per\_capita\_2020 gdp\_per\_capita\_2019  
## 1 NA NA NA  
## 2 NA NA NA  
## 3 NA NA NA  
## 4 NA NA NA  
## 5 NA NA NA  
## 6 NA NA NA  
## 7 NA NA NA  
## 8 NA NA NA  
## 9 NA NA NA  
## 10 NA NA NA  
## 11 NA NA NA  
## 12 NA NA NA  
## 13 NA NA NA  
## 14 NA NA NA  
## 15 NA NA NA  
## gdp\_per\_capita\_2018 gdp\_per\_capita\_2017 inflation\_rate\_2021  
## 1 NA NA NA  
## 2 NA NA NA  
## 3 NA NA NA  
## 4 NA NA NA  
## 5 NA NA NA  
## 6 NA NA NA  
## 7 NA NA NA  
## 8 NA NA NA  
## 9 NA NA NA  
## 10 NA NA NA  
## 11 NA NA NA  
## 12 NA NA NA  
## 13 NA NA NA  
## 14 NA NA NA  
## 15 NA NA NA  
## inflation\_rate\_2020 inflation\_rate\_2019 inflation\_rate\_2018  
## 1 NA NA NA  
## 2 NA NA NA  
## 3 NA NA NA  
## 4 NA NA NA  
## 5 NA NA NA  
## 6 NA NA NA  
## 7 NA NA NA  
## 8 NA NA NA  
## 9 NA NA NA  
## 10 NA NA NA  
## 11 NA NA NA  
## 12 NA NA NA  
## 13 NA NA NA  
## 14 NA NA NA  
## 15 NA NA NA  
## inflation\_rate\_2017  
## 1 NA  
## 2 NA  
## 3 NA  
## 4 NA  
## 5 NA  
## 6 NA  
## 7 NA  
## 8 NA  
## 9 NA  
## 10 NA  
## 11 NA  
## 12 NA  
## 13 NA  
## 14 NA  
## 15 NA

all\_togheter\_2022 <-all\_togheter\_2022 %>%  
 filter(!Country.Name %in% c("Europe", "Africa", "South America",  
 "Oceania", "Asia", "North America",  
 "Total:", "World"))  
#finally, let's remove some columns that have general metric by region

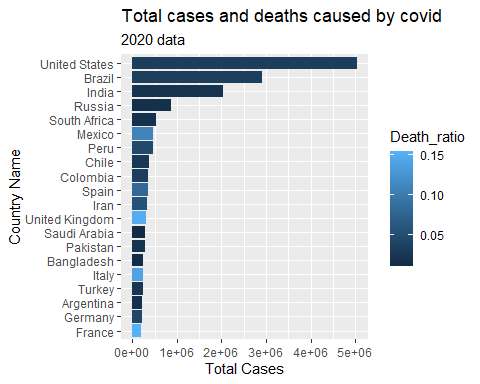
Now that we have our data, it’s time to make our analysis

#Analysis

data\_2020 <- all\_togheter\_2020 %>%   
 arrange(desc(TotalCases)) %>%   
 slice(1:20) # Top 20 highest TotalCases by country 2020  
  
  
data\_2022 <- all\_togheter\_2022 %>%   
 arrange(desc(Total.Cases)) %>%   
 slice(1:20) # Top 20 highest TotalCases by country 2022

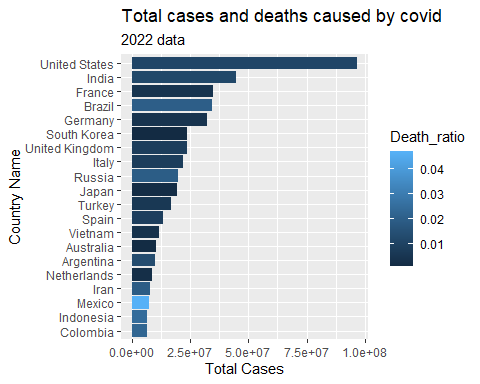
#Analizing Now it’s time to analize our data. First we will compare the population’s countries with their Death\_ratio, which is just a division, Population/Total Death.

data\_2020 <- mutate(data\_2020, Death\_ratio=TotalDeaths/TotalCases)   
#we created in a first step our new Death\_ratio column  
data\_2020 %>% ggplot(aes(x=reorder(Country.Name, +TotalCases), TotalCases, fill=Death\_ratio)) +  
 geom\_bar(stat="identity", position="stack")+  
 labs(x="Country Name", y="Total Cases",  
 title="Total cases and deaths caused by covid",  
 subtitle="2020 data") + coord\_flip()



Now we analize the same but with data from 2022

data\_2022 <- mutate(data\_2022, Death\_ratio=Total.Deaths/Total.Cases)  
#we created in a first step our new Death\_ratio column  
data\_2022 %>% ggplot(aes(x=reorder(Country.Name, +Total.Cases), Total.Cases, fill=Death\_ratio)) +  
 geom\_bar(stat="identity", position="stack")+  
 labs(x="Country Name", y="Total Cases",  
 title="Total cases and deaths caused by covid",  
 subtitle="2022 data") + coord\_flip()



Comparing data from 2020 and 2022 Changing the colnames to make data\_2022 and data\_2020 can match. And merge their rows

colnames(data\_2020)

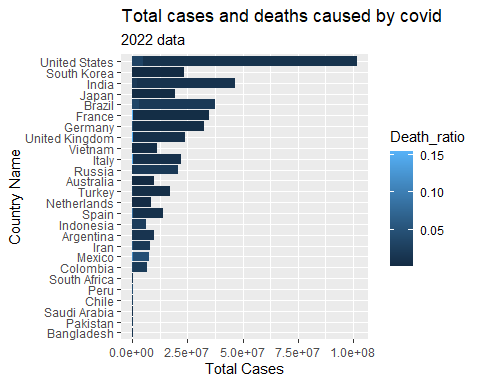
## [1] "Country.Name" "Continent" "Population"   
## [4] "TotalCases" "NewCases" "TotalDeaths"   
## [7] "NewDeaths" "TotalRecovered" "NewRecovered"   
## [10] "ActiveCases" "Serious.Critical" "Tot.Cases.1M.pop"   
## [13] "Deaths.1M.pop" "TotalTests" "Tests.1M.pop"   
## [16] "WHO.Region" "gdp\_rate\_2021" "gdp\_rate\_2020"   
## [19] "gdp\_rate\_2019" "gdp\_rate\_2018" "gdp\_rate\_2017"   
## [22] "gdp\_per\_capita\_2021" "gdp\_per\_capita\_2020" "gdp\_per\_capita\_2019"  
## [25] "gdp\_per\_capita\_2018" "gdp\_per\_capita\_2017" "inflation\_rate\_2021"  
## [28] "inflation\_rate\_2020" "inflation\_rate\_2019" "inflation\_rate\_2018"  
## [31] "inflation\_rate\_2017" "Death\_ratio"

colnames(data\_2022)

## [1] "Country.Name" "X." "Total.Cases"   
## [4] "New.Cases" "Total.Deaths" "New.Deaths"   
## [7] "Total.Recovered" "New.Recovered" "Active.Cases"   
## [10] "Serious..Critical" "Tot.Cases..1M.pop" "Deaths..1M.pop"   
## [13] "Total.Tests" "Tests...1M.pop" "Population"   
## [16] "gdp\_rate\_2021" "gdp\_rate\_2020" "gdp\_rate\_2019"   
## [19] "gdp\_rate\_2018" "gdp\_rate\_2017" "gdp\_per\_capita\_2021"  
## [22] "gdp\_per\_capita\_2020" "gdp\_per\_capita\_2019" "gdp\_per\_capita\_2018"  
## [25] "gdp\_per\_capita\_2017" "inflation\_rate\_2021" "inflation\_rate\_2020"  
## [28] "inflation\_rate\_2019" "inflation\_rate\_2018" "inflation\_rate\_2017"  
## [31] "Death\_ratio"

data\_2020 <- select(data\_2020, -Continent, WHO.Region,-NewRecovered, -NewDeaths, -TotalRecovered,  
 NewCases, -ActiveCases, -ActiveCases)  
data\_2022 <- select(data\_2022, -New.Recovered, -New.Deaths, -Total.Recovered,  
 New.Cases, -Active.Cases, -Active.Cases)  
data\_2022 <- data\_2022[,-2]  
  
data\_2020 <-data\_2020 %>% rename(Total.Cases=TotalCases,  
 Total.Deaths=TotalDeaths,  
 Serious..Critical=Serious.Critical,  
 Tot.Cases..1M.pop=Tot.Cases.1M.pop,  
 Deaths..1M.pop=Deaths.1M.pop,  
 Total.Tests=TotalTests,  
 Tests...1M.pop=Tests.1M.pop,  
 Deaths..1M.pop=Deaths.1M.pop  
 )  
  
all\_data <- bind\_rows(data\_2020, data\_2022)

all\_data %>% ggplot(aes(x=reorder(Country.Name, +Total.Cases), Total.Cases, fill=Death\_ratio)) +  
 geom\_bar(stat="identity", position="stack")+  
 labs(x="Country Name", y="Total Cases",  
 title="Total cases and deaths caused by covid",  
 subtitle="2022 data") + coord\_flip()



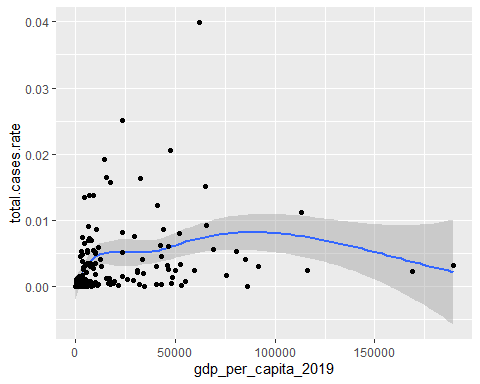
But, there’s a relationship between COVID and the gdp per capita in 2020? Let’s make a linear regression analysis to have an answer.

all\_togheter\_2020 <- mutate(all\_togheter\_2020, total.cases.rate=TotalCases/Population)  
#first we standarize our variable, having a relation between total cases and population by country, in order to make a fair analysis  
  
all\_togheter\_2020 %>% ggplot() +   
 geom\_smooth(aes(gdp\_per\_capita\_2019, total.cases.rate))+  
 geom\_point(aes(gdp\_per\_capita\_2019, total.cases.rate))

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 81 rows containing non-finite values (stat\_smooth).

## Warning: Removed 81 rows containing missing values (geom\_point).



all\_togheter\_2020 %>% lm(formula=total.cases.rate ~gdp\_per\_capita\_2019) %>%   
 summary()

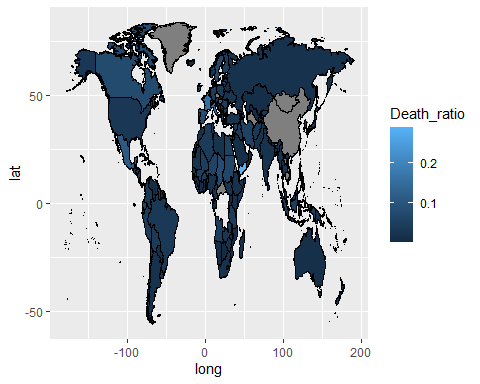
##   
## Call:  
## lm(formula = total.cases.rate ~ gdp\_per\_capita\_2019, data = .)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.007329 -0.002458 -0.001863 0.000711 0.034830   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.450e-03 4.746e-04 5.161 6.95e-07 \*\*\*  
## gdp\_per\_capita\_2019 4.255e-08 1.355e-08 3.139 0.00201 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.005111 on 166 degrees of freedom  
## (81 observations deleted due to missingness)  
## Multiple R-squared: 0.05603, Adjusted R-squared: 0.05034   
## F-statistic: 9.853 on 1 and 166 DF, p-value: 0.002007

all\_togheter\_2020 %>% lm(formula=gdp\_per\_capita\_2019~total.cases.rate) %>%   
 summary()

##   
## Call:  
## lm(formula = gdp\_per\_capita\_2019 ~ total.cases.rate, data = .)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -28259 -14655 -11770 3614 170124   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 15172 2589 5.859 2.44e-08 \*\*\*  
## total.cases.rate 1316858 419524 3.139 0.00201 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 28440 on 166 degrees of freedom  
## (81 observations deleted due to missingness)  
## Multiple R-squared: 0.05603, Adjusted R-squared: 0.05034   
## F-statistic: 9.853 on 1 and 166 DF, p-value: 0.002007

To interpret this model, we will not focus on the r-squared, we will just focus on the significance of the variables in the model. In this case we have two statistical significant variables in both models. The first one is the GDP per capita, which is significant, explains that the GDP per capita of a country determine the covid cases in a country.

mapdata <- map\_data("world")  
mapdata <- mapdata %>% rename("Country.Name" = "region")  
mapdata[mapdata=="USA"] <- "United States"  
all\_togheter\_2020 <- mutate(all\_togheter\_2020,  
 Death\_ratio=TotalDeaths/TotalCases)  
  
map\_data <- right\_join(mapdata, all\_togheter\_2020, by="Country.Name")  
  
ggplot(map\_data, aes(long, lat, group=group))+  
 geom\_polygon(aes(fill=Death\_ratio), color="black")



ggplot(map\_data, aes(long, lat, group=group))+  
 geom\_polygon(aes(fill=gdp\_per\_capita\_2020), color="black")

