Final Project

Phuc Quang Vu pqv9968

2023-03-11

1.Introduction

Covid-19 had hit the US terribly and cost the US over 1.1 million citizens. Nevertheless, each state has been affected differently. This project aims to answer if GDP per capita and the ratio of people per healthcare worker of each state are associated with how Covid-19 spread among states.

To answer this question, this project looks at the incidence and mortality rate of Covid-19 in each state with different GDP per capita and the ratio of people per healthcare worker.

More specifically, the total of cases and deaths will be aggregated by each state for each month. Information regarding GDP per capita and the ratio of people per healthcare worker will be collected from public data sets. Finally, data will be joined via state name and transformed for data visualizations. From the data visualizations, readers can determine if there is any association between the mentioned factors and the incidence of Covid 19.

2.Packages Required

```
library(tidyverse)        # creating tidy data
library(dplyr)            # transforming data
library(ggplot2)          # data visualization
library(lubridate)        # group date by month, year
```

3.Data Preparation

3.1.Details of data for this project

3.1.1.us_states data:

Citation: The New York Times. (2021). Coronavirus (Covid-19) Data in the United States.

Source: https://github.com/nytimes/covid-19-data.

Variables: date, state, fips, cases, deaths

Details: number of Covid-19 cases and deaths by Covid-19 by state by date collected and published by NY Times

3.1.2.us_people_per_hospital

Citation: Number of hospitals and hospital employment in each state in 2019.: The Economics Daily: U.S. Bureau of Labor Statistics. (2020, April 6)

Source: https://www.bls.gov/opub/ted/2020/number-of-hospitals-and-hospital-employment-in-each-state-in-2019.htm

Variables: state, ratio of people per hospital, Hospital establishment in 2019, Population on July 1, 2019

Details: number of ratio of people per hospital, Hospital establishment, Population by state in 2019 published by Bureau of Labor Statistics

3.1.3.us_gdp_by_state

Citation: US Department of Commerce, BEA, Bureau of Economic Analysis. (n.d.). BEA.: Regional Economic Accounts: Download.

Source: https://apps.bea.gov/regional/downloadzip.cfm

Variables: GeoFIPS, GeoName, Region, TableName, LineCode, IndustryClassification, Description, Unit, GDP by Year

Details: GDP by states from 1997 to 2020 published by Bureau of Economic Analysis

3.1.4.us_health_care_by_state

Citation: Tables Created by BLS.: U.S. Bureau of Labor Statistics. (2022, December 19).

Sources: https://www.bls.gov/oes/tables.htm

Variables: state, number_of_healthcare_worker

Details: number of healthcare worker by state in 2019 collected by bureau of Labor Statistics

3.1.5.us_pop

Citation: US Census Bureau. (2021, October 28). 2019 National and State Population Estimates. Census.gov.

Sources: https://www.census.gov/newsroom/press-kits/2019/national-state-estimates.html

Variables: state, population by years

Details: US population by state from 2010 to 2019

3.2.Loading data

```
us_states=read_csv("us-states.csv")
us_gdp_by_state=read_csv("us_gdp_by_state.csv")
us_health_care_by_state=read_csv("us_healthcare_worker_2019.csv")
us_pop<-read_csv("us_pop.csv")
```

3.3.Clean data

3.3.1.us_states

```
summary(us_states)
```

```
##       date               state               fips               cases
##  Min.   :2020-01-21   Length:60430       Length:60430       Min.   :       1
##  1st Qu.:2020-12-02   Class :character   Class :character   1st Qu.:   68186
##  Median :2021-09-03   Mode  :character   Mode  :character   Median :  340269
##  Mean   :2021-08-31                                         Mean   :  865846
##  3rd Qu.:2022-05-31                                         3rd Qu.: 1017293
##  Max.   :2023-02-24                                         Max.   :12085063
##      deaths
##  Min.   :     0
##  1st Qu.:  1138
##  Median :  4932
##  Mean   : 11568
##  3rd Qu.: 14662
##  Max.   :103211
```

```
sum(is.na(us_states))
```

```
## [1] 0
```

```
sum(duplicated(us_states))
```

```
## [1] 0
```

```
# choosing necessary variables for analysis
us_states<- us_states %>%
  select(date, state,cases,deaths)
```

Since there is no NA, no duplicate and the values of cases and deaths lie in reasonable range(greater than 0), there is not much need for data cleaning besides choosing variables that are necessary for analysis.

3.3.2.us_gdp_by_state

The us_gdp_by_state table have data regarding multiple types of GDP and years but this project only focus on total GDP by state in 2009, therefore first project need to select from the table variables GeoName (state), Description (type of GDP), Unit(unit of GDP), and 2009 (the GDP value in 2009)

```
us_gdp_by_state<- us_gdp_by_state%>%
  select(GeoName,Description,Unit,"2019")
```

```
# checking for NA values
sum(is.na(unique(us_gdp_by_state$GeoName)))
```

```
## [1] 1
```

```
sum(is.na(unique(us_gdp_by_state$Description)))
```

```
## [1] 1
```

```
unique(us_gdp_by_state$Description)
```

```
## [1] "Real GDP (millions of chained 2012 dollars)"
## [2] "Chain-type quantity indexes for real GDP"
## [3] "Current-dollar GDP (millions of current dollars)"
## [4] "Compensation (millions of dollars)"
## [5] "Gross operating surplus (millions of dollars)"
## [6] "Taxes on production and imports (TOPI) less subsides (millions of dollars)"
## [7] "Taxes on production and imports (TOPI) (millions of dollars)"
## [8] "Subsidies (millions of dollars)"
## [9] NA
```

Taking a quick look at GeoName and Description variables, we found NA in both variables. Furthermore,we only need "Real GDP (millions of chained 2012 dollars)", so we drop NA values and filter the table to display only Real GDP

```
us_gdp_by_state<- us_gdp_by_state%>%
  drop_na()%>%
  filter(Description=="Real GDP (millions of chained 2012 dollars)")
head(us_gdp_by_state)
```

```
## # A tibble: 6 x 4
##   GeoName       Description                                 Unit          `2019`
##   <chr>         <chr>                                       <chr>          <dbl>
## 1 United States Real GDP (millions of chained 2012 dollars) Millions of ~ 1.90e7
## 2 Alabama       Real GDP (millions of chained 2012 dollars) Millions of ~ 2.03e5
## 3 Alaska        Real GDP (millions of chained 2012 dollars) Millions of ~ 5.34e4
## 4 Arizona       Real GDP (millions of chained 2012 dollars) Millions of ~ 3.25e5
## 5 Arkansas      Real GDP (millions of chained 2012 dollars) Millions of ~ 1.17e5
## 6 California    Real GDP (millions of chained 2012 dollars) Millions of ~ 2.73e6
```

We then change the name of variables to assist further coding and reselect variables that are important for the study (GeoName and 2019)

```
colnames(us_gdp_by_state)<-c("state","description","unit","GDP")
us_gdp_by_state<- us_gdp_by_state%>%
  select(state,GDP)
head(us_gdp_by_state)
```

```
## # A tibble: 6 x 2
##   state               GDP
##   <chr>             <dbl>
## 1 United States 19036052
## 2 Alabama         203433.
## 3 Alaska           53434.
## 4 Arizona         325395.
## 5 Arkansas        117126.
## 6 California     2729226.
```

3.3.3.us_health_care_by_state

```r
# checking for na and duplicate values
summary(us_health_care_by_state)
```

```
##     state           number_of_healthcare_worker
##  Length:54          Min.   :   34420
##  Class :character   1st Qu.:  707452
##  Mode  :character   Median : 1781000
##                     Mean   : 2737754
##                     3rd Qu.: 3257302
##                     Max.   :17382400
```

```r
sum(is.na(us_health_care_by_state))
```

```
## [1] 0
```

```r
sum(duplicated(us_health_care_by_state))
```

```
## [1] 0
```

```r
head(us_health_care_by_state)
```

```
## # A tibble: 6 x 2
##   state      number_of_healthcare_worker
##   <chr>                            <dbl>
## 1 Alabama                        1974170
## 2 Alaska                          317090
## 3 Arizona                        2866820
## 4 Arkansas                       1217420
## 5 California                    17382400
## 6 Colorado                       2678490
```

The table has no duplicate, NA and is already clean.

3.3.4.us_pop

Since the table have the population of US by state from 2010 to 2019, for this project, I first select variable state and 2019, then rename them for futher analysis

```r
us_pop<-us_pop%>%
  select(state,"2019")
colnames(us_pop)<-c("state","population")
head(us_pop)
```

```
## # A tibble: 6 x 2
##   state         population
##   <chr>              <dbl>
```

```
## 1 United States  328239523
## 2 Northeast       55982803
## 3 Midwest         68329004
## 4 South          125580448
## 5 West            78347268
## 6 Alabama          4903185
```

After selection, I check the summary, NA, and duplicate in the table

```
#checking for NA, duplicates and review summary
summary(us_pop)
```

```
##     state             population
##  Length:57          Min.   :   578759
##  Class :character   1st Qu.:  1934408
##  Mode  :character   Median :  4903185
##                     Mean   : 17331794
##                     3rd Qu.:  9986857
##                     Max.   :328239523
```

```
sum(is.na(us_pop))
```

```
## [1] 0
```

```
sum(duplicated(us_pop))
```

```
## [1] 0
```

```
head(us_pop)
```

```
## # A tibble: 6 x 2
##   state         population
##   <chr>              <dbl>
## 1 United States  328239523
## 2 Northeast       55982803
## 3 Midwest         68329004
## 4 South          125580448
## 5 West            78347268
## 6 Alabama          4903185
```

The table has no duplicate, NA and is already clean.

3.3.5.cleaned_data

For the final step, I join data sets that I have cleaned by variable state and drop rows with NA value

```
# joining tables through columns state
cleaned_data<- us_states%>%
  left_join(us_gdp_by_state,by = "state")%>%
  left_join(us_health_care_by_state,by="state")%>%
  left_join(us_pop,by="state")%>%
  drop_na() # remove NA values
dim(cleaned_data)
```

```
## [1] 55611     7
```

```
head(cleaned_data)
```

```
## # A tibble: 6 x 7
##   date       state      cases deaths      GDP number_of_healthcare_wor~1 popul~2
##   <date>     <chr>      <dbl>  <dbl>    <dbl>                      <dbl>   <dbl>
## 1 2020-01-21 Washington     1      0  533150.                    3318510  7.61e6
## 2 2020-01-22 Washington     1      0  533150.                    3318510  7.61e6
## 3 2020-01-23 Washington     1      0  533150.                    3318510  7.61e6
## 4 2020-01-24 Illinois       1      0  775998.                    6025790  1.27e7
## 5 2020-01-24 Washington     1      0  533150.                    3318510  7.61e6
## 6 2020-01-25 California     1      0 2729226.                   17382400  3.95e7
## # ... with abbreviated variable names 1: number_of_healthcare_worker,
## #   2: population
```

4.Exploratory Data Analysis

4.1.Calculate required variables

```
data_pre_analyze<- cleaned_data%>%
  #extract month from date
  mutate(month = lubridate::floor_date(date, "month"),
         #calculate incidence per million
         cases_per_mil=cases/population*1e6,
         #calculate mortality per million
         death_per_mil=deaths/population*1e6,
         #calculate ratio of people per healthcare worker
         people_per_healthcare_worker=population/number_of_healthcare_worker,
         #calculate GDP per capita
         GDP_percapita=GDP/population*1e6)%>%
  # Selecting variables that are necessary for analysis
  select(state,
         cases,
         deaths,
         people_per_healthcare_worker,
         month,
         cases_per_mil,death_per_mil,
         GDP_percapita,
         population)
# Group data to view data at begining of each month
data_analyze<- data_pre_analyze%>%
  group_by(state,month)%>%
  summarize(cases_per_mil=max(cases_per_mil),
            death_per_mil=max(death_per_mil),
            people_per_healthcare_worker=max(people_per_healthcare_worker),
            GDP_percapita=max(GDP_percapita))%>%
  ungroup()%>%
  mutate(ratio_death_vs_case=death_per_mil/cases_per_mil)
head(data_analyze)
```

```
## # A tibble: 6 x 7
##   state   month      cases_per_mil death_per_mil people_per_he~1 GDP_p~2 ratio~3
##   <chr>   <date>             <dbl>         <dbl>           <dbl>   <dbl>   <dbl>
## 1 Alabama 2020-03-01          204.          2.86            2.48  41490.  0.0140
## 2 Alabama 2020-04-01         1442.         55.5             2.48  41490.  0.0385
## 3 Alabama 2020-05-01         3661.        128.              2.48  41490.  0.0351
## 4 Alabama 2020-06-01         7759.        194.              2.48  41490.  0.0250
## 5 Alabama 2020-07-01        17891.        322.              2.48  41490.  0.0180
## 6 Alabama 2020-08-01        25709.        445.              2.48  41490.  0.0173
## # ... with abbreviated variable names 1: people_per_healthcare_worker,
## #   2: GDP_percapita, 3: ratio_death_vs_case
```

```
dim(data_analyze)
```

```
## [1] 1850    7
```

4.2.GDP per capita

First, the GDP per capita column is categorized into 4 categories (first quartile: <25 quantile, second quartile: 25-50 quantile, third quartile: 50-75 quantile,and Fourth quartile: >75 quantile)

```
# Factor GDP_percapita to quantile
data_analyze$GDP_quantile=cut(data_analyze$GDP_percapita,
c(-Inf,quantile(unique(data_analyze$GDP_percapita),c(0.25,0.5,0.75,1))),
labels=c("First quartile GDP","Second quartile GDP","Third quartile GDP","Fourth quartile GDP"))
```

4.2.1.GDP per capita vs cases

```
df_GDP<- data_analyze%>%
  group_by(GDP_quantile,month)%>%
  #calculate average incidence and mortality for each group
  summarize(average_incidence_per_mil=round(mean(cases_per_mil),2),
            average_deaths_per_mil=round(mean(death_per_mil),2))%>%
  ungroup()
# display table
knitr::kable(df_GDP %>%
               select(month,GDP_quantile,average_incidence_per_mil)%>%
  pivot_wider(names_from=GDP_quantile,
              values_from = average_incidence_per_mil),
              digit=6)
```

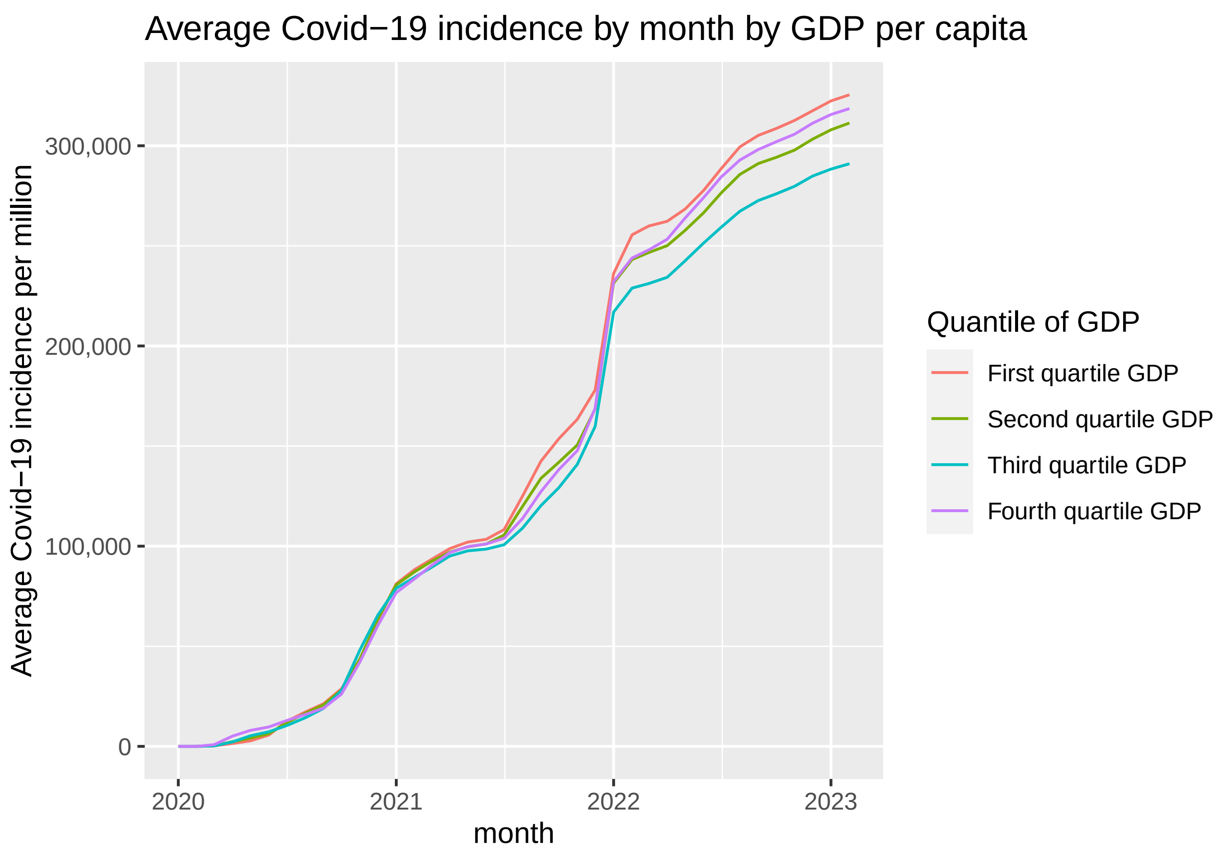
month	First quartile GDP	Second quartile GDP	Third quartile GDP	Fourth quartile GDP
2020-01-01	0.14	NA	0.16	0.10
2020-02-01	0.14	0.20	2.42	0.64
2020-03-01	249.97	347.81	219.31	846.21
2020-04-01	1418.16	2331.09	2211.44	5136.47
2020-05-01	2730.40	4132.44	5285.74	7981.81
2020-06-01	5668.94	6315.42	7275.66	9686.24

month	First quartile GDP	Second quartile GDP	Third quartile GDP	Fourth quartile GDP
2020-07-01	12513.79	11798.80	10370.90	12902.06
2020-08-01	17126.94	16463.63	14278.50	15878.45
2020-09-01	21314.30	20748.21	18977.55	19105.00
2020-10-01	28780.84	28097.73	27914.81	26097.40
2020-11-01	43482.72	43671.39	48200.89	42200.82
2020-12-01	62794.68	63567.65	65534.25	60277.72
2021-01-01	81248.32	80719.91	78975.79	76774.05
2021-02-01	88340.53	87226.81	84501.56	83733.03
2021-03-01	93358.87	92390.63	89364.38	90343.89
2021-04-01	98814.12	97021.70	95085.10	96779.23
2021-05-01	102043.68	99688.96	97648.15	99751.41
2021-06-01	103422.70	101081.17	98524.80	101066.82
2021-07-01	108256.16	105537.92	100688.22	103958.73
2021-08-01	124974.18	119838.97	108994.38	113781.89
2021-09-01	142474.75	133893.93	120275.10	127276.85
2021-10-01	153664.85	141896.39	129203.45	138257.81
2021-11-01	163352.75	150534.54	140853.49	147687.77
2021-12-01	177991.90	168541.86	159836.26	168874.18
2022-01-01	236158.04	231387.08	216936.58	232088.65
2022-02-01	255520.53	243163.03	228905.53	243956.32
2022-03-01	259908.81	246667.00	231177.58	247975.32
2022-04-01	262285.87	250104.39	234282.53	253305.20
2022-05-01	268374.68	257733.18	242520.03	263864.19
2022-06-01	277612.32	266489.82	251367.92	274064.37
2022-07-01	288697.40	276534.90	259387.58	284496.92
2022-08-01	299434.01	285668.88	267280.27	292859.65
2022-09-01	305189.58	291114.41	272640.37	298127.00
2022-10-01	308603.66	294207.67	275972.59	301981.51
2022-11-01	312669.09	297855.12	279769.76	305802.72
2022-12-01	317474.49	303279.75	284843.43	311254.71
2023-01-01	322400.09	307967.90	288336.73	315598.96
2023-02-01	325464.35	311369.37	291024.17	318568.80

From the table, It appears that the fourth quantile (the highest GDP per capita) states have a high incidence rate at the first 6 months but this trend quickly disappeared after this time.

Then, we use lineplot to see how Covid-19 Incidence advanced in each category

```
ggplot()+geom_line(data=df_GDP,
                   mapping=aes(x=month,
                            y=average_incidence_per_mil,
                            group=GDP_quantile,
                            color=GDP_quantile))+
  scale_y_continuous(name="Average Covid-19 incidence per million",
                     labels = scales::comma)+
  ggtitle("Average Covid-19 incidence by month by GDP per capita")+
  labs(color="Quantile of GDP")
```

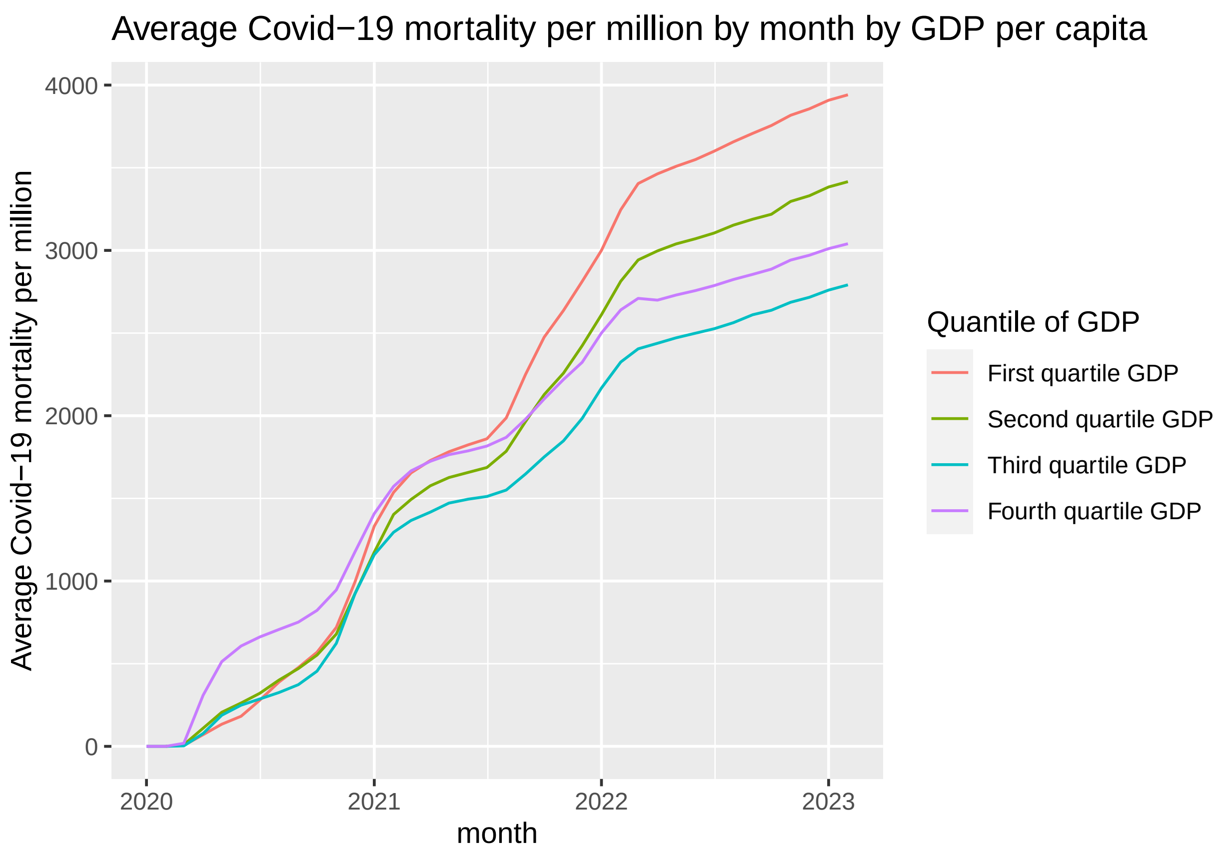


We can see that all four categories appear to follow the same pattern.

4.2.2.GDP per capita vs deaths

Now, let us examine the relationship between GDP per capita and Covid-19 mortality

```
ggplot()+geom_line(data=df_GDP,mapping=aes(x=month,
                                           y=average_deaths_per_mil,
                                           group=GDP_quantile,
                                           color=GDP_quantile))+
  ggtitle("Average Covid-19 mortality per million by month by GDP per capita")+
  ylab("Average Covid-19 mortality per million")+
  labs(color="Quantile of GDP")
```



For the first year, the states with the highest GDP_per_capita have the highest mortality per million. This can possibly be due to the high incidence rate and lack of treatments at the time. However, in the end, states with high GDP have lower mortality rates compared with states with low GDP.

Conclusion: There is not a clear association between GDP per capita and Covid-19 incidence rate but there seem to be a clear association between GDP per capita and Covid-19 mortality rate.

4.3.Ratio of people per healthcare worker

First the ratio between population and healthcare worker variable is categorized into 4 categories (first quartile: <25 quantile, second quartile: 25-50 quantile, third quartile: 50-75 quantile,and Fourth quartile: >75 quantile)

```
# Factor people_per_healthcare_worker to quantile
data_analyze$health_worker_quantile=cut(data_analyze$people_per_healthcare_worker,
       c(-Inf,quantile(unique(data_analyze$people_per_healthcare_worker),
c(0.25,0.5,0.75,1))),labels=c("First quartile health worker",
"Second quartile health worker","Third quartile health worker",
```

4.3.1.Ratio of people per healthcare worker vs cases

```
df_worker<- data_analyze%>%
  group_by(health_worker_quantile,month)%>%
```

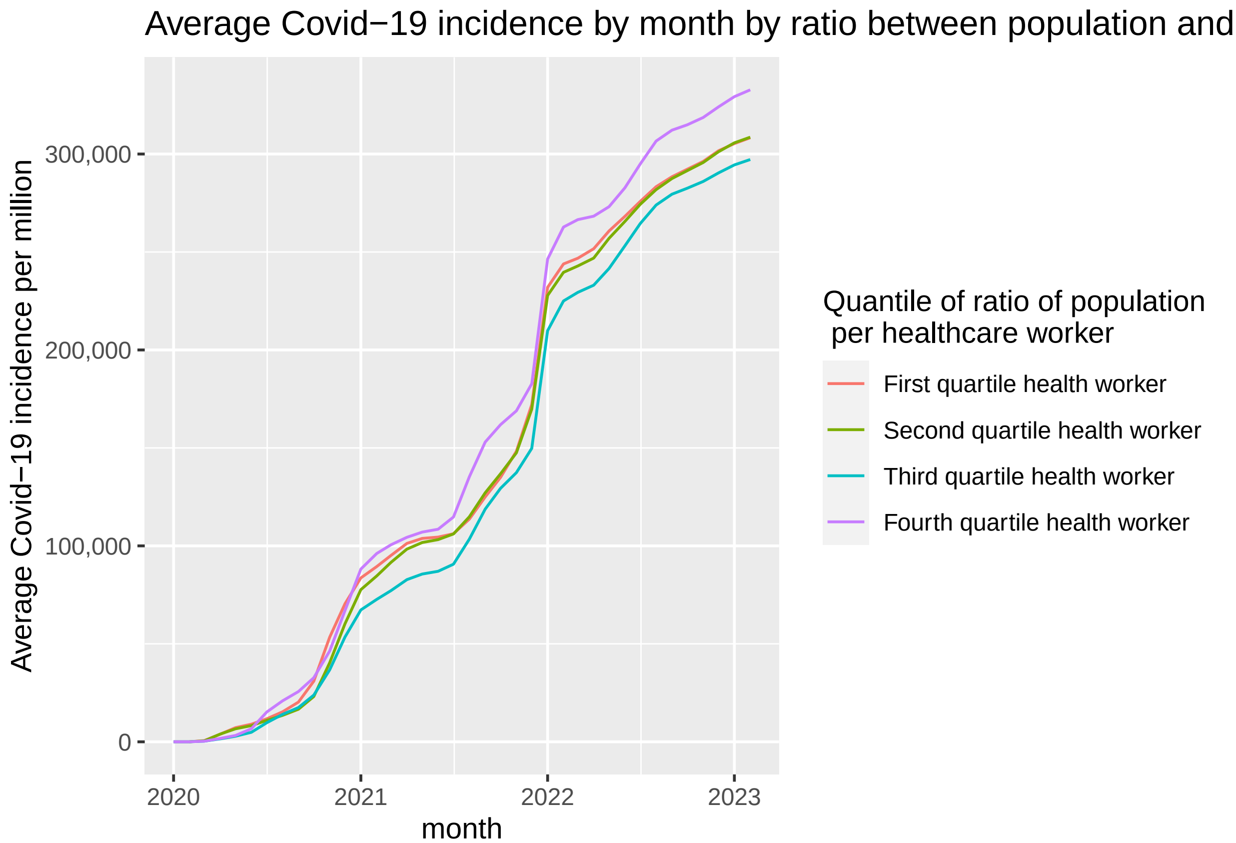
```
  summarize(average_incidence_per_mil=round(mean(cases_per_mil),2),
            average_deaths_per_mil=round(mean(death_per_mil),2))%>%
  ungroup()

knitr::kable(df_worker %>%
               select(month,health_worker_quantile,average_incidence_per_mil)%>%
  pivot_wider(names_from=health_worker_quantile,
              values_from = average_incidence_per_mil),
              digit=6)
```

month	First quartile health worker	Second quartile health worker	Third quartile health worker	Fourth quartile health worker
2020-01-01	0.16	NA	0.10	0.14
2020-02-01	1.82	0.31	0.66	0.14
2020-03-01	602.43	482.23	297.56	287.10
2020-04-01	3882.46	3930.17	1507.25	1723.11
2020-05-01	7223.04	6640.50	2845.55	3233.60
2020-06-01	8994.30	8258.75	4848.64	6657.88
2020-07-01	11725.68	10919.36	9708.88	15180.70
2020-08-01	15322.17	13496.26	13999.33	20908.28
2020-09-01	20272.30	16625.07	17489.14	25644.06
2020-10-01	30947.60	23100.26	23834.57	32694.48
2020-11-01	53349.63	40302.80	36697.17	46321.32
2020-12-01	70479.69	60256.87	53500.34	67011.71
2021-01-01	83662.80	77675.60	67315.68	88167.04
2021-02-01	89441.04	84657.15	72687.95	96107.05
2021-03-01	95149.00	91604.66	77185.46	100581.81
2021-04-01	101257.82	98346.39	82779.25	104370.09
2021-05-01	103818.14	101717.50	85628.29	107043.67
2021-06-01	104480.64	103198.76	87036.85	108495.55
2021-07-01	106193.30	106093.01	90655.30	114727.66
2021-08-01	113646.52	114839.24	103433.04	135242.83

month	First quartile health worker	Second quartile health worker	Third quartile health worker	Fourth quartile health worker
2021-09-01	124982.50	127089.74	118758.45	152973.27
2021-10-01	134976.10	136809.56	129377.94	161872.32
2021-11-01	148333.86	147506.55	137389.52	168932.17
2021-12-01	172147.01	169801.90	149781.02	182740.78
2022-01-01	231972.76	227814.34	209852.30	246386.00
2022-02-01	243906.25	239614.72	224996.73	262727.05
2022-03-01	246808.85	242860.89	229393.09	266528.62
2022-04-01	251665.72	246883.42	233056.39	268278.14
2022-05-01	260727.18	256981.24	241593.67	273118.72
2022-06-01	268178.61	265600.89	253136.73	282754.27
2022-07-01	275808.71	274304.82	264494.64	294901.47
2022-08-01	283320.47	281840.94	273971.10	306624.98
2022-09-01	288406.09	287498.96	279502.63	312191.54
2022-10-01	292262.43	291529.90	282559.19	314920.57
2022-11-01	296240.62	295680.00	285999.20	318656.06
2022-12-01	301674.25	301166.02	290370.95	324066.35
2023-01-01	305327.43	305715.90	294453.16	329277.68
2023-02-01	308324.37	308609.73	297197.18	332770.25

Then, we use lineplot to see how Covid-19 Incidence advanced in each category

```
ggplot()+geom_line(data=df_worker,mapping=aes(x=month,
                                              y=average_incidence_per_mil,
                                              group=health_worker_quantile,
                                              color=health_worker_quantile))+
  scale_y_continuous(name="Average Covid-19 incidence per million", labels = scales::comma)+
  ggtitle("Average Covid-19 incidence by month by ratio between population and healthcare worker")+
  labs(color="Quantile of ratio of population \n per healthcare worker")
```

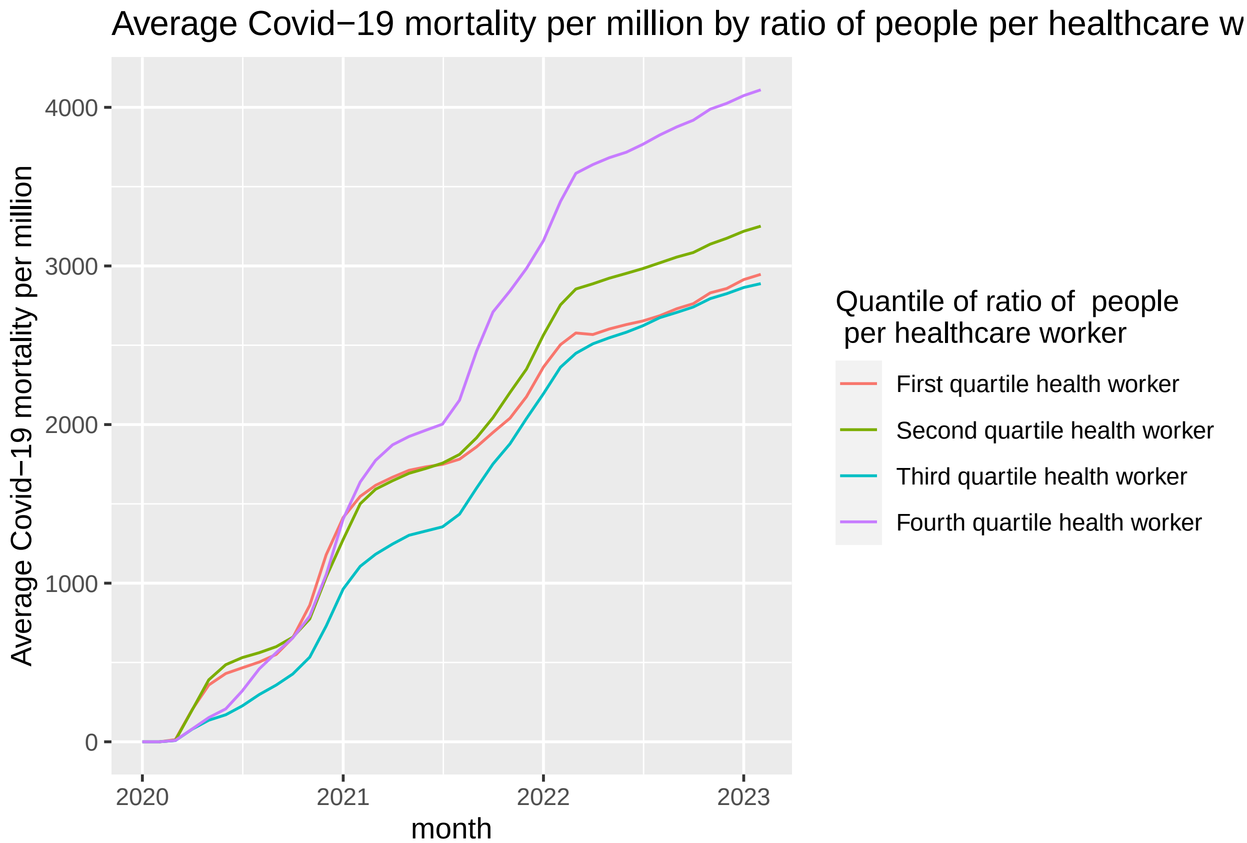


We can see that all four categories appear to follow the same pattern.

4.3.2.Ratio of people per healthcare worker vs deaths

Now, let us examine the relationship between GDP per capita and Covid-19 mortality

```
ggplot()+geom_line(data=df_worker,mapping=aes(x=month,
                                              y=average_deaths_per_mil,
                                              group=health_worker_quantile,
                                              color=health_worker_quantile))+
  ggtitle("Average Covid-19 mortality per million by ratio of people per healthcare worker")+
  ylab("Average Covid-19 mortality per million")+
  labs(color="Quantile of ratio of  people \n per healthcare worker")
```



States with low ratio of people per healthcare workers have lower mortality rate compare to states with high ratio of people per healthcare workers.

Conclusion: The ratio of people per healthcare workers is not associated with high Covid-19 incidence rate but there is a negative association between The ratio of people per healthcare and the Covid-19 mortality rate.

4.4.Relationship between GDP per capita and ratio of people per healthcare worker

```
new_data<-data_analyze[!duplicated(data_analyze$state),1:9]

GDP_healthcare<- table(new_data$GDP_quantile,
                       new_data$health_worker_quantile)
knitr::kable(GDP_healthcare)
```

	First quartile health worker	Second quartile health worker	Third quartile health worker	Fourth quartile health worker
First quartile GDP	0	1	2	10
Second quartile GDP	2	4	4	3

	First quartile health worker	Second quartile health worker	Third quartile health worker	Fourth quartile health worker
Third quartile GDP	7	3	2	0
Fourth quartile GDP	4	5	4	0

From the summary, we can see that there is a possible relationship between GDP per capita and the ratio of people per healthcare worker. Specifically, we have neither states with low GPD (first quartile) and a low ratio of people per healthcare worker (first quartile) nor states with high GDP(fourth quartile) and a high ratio of people per healthcare worker (fourth quartile). Another interesting point is that third-quartile GDP states are also more likely to have the lowest ratio of people per healthcare worker (7 states). This could also explain why third-quartile GDP states performed better than Fourth quartile GDP states.

5.Summary

The project aims to find out how GDP per capita and the ratio of people per healthcare worker affect how Covid-19 advanced in each state.

We then use a line plot to display how Covid-19 Incidence and Covid-19 Mortality change over time for states with different ranges of GDP and the ratio of people per healthcare worker.

Overall, the difference between GDP per capita and the ratio of people per healthcare worker appears to have an association with how Covid-19 spread but on Covid-19 mortality. Specifically, states with the lowest GDP per capita have the highest average Covid-19 mortality while states with relatively high GDP per capita (Third quartile) have the lowest average Covid-19 mortality. In addition, states with the highest ratio of people per healthcare worker also have the highest Covid-19 mortality. It is also worth noticing that Third quartile GDP per capita states are the most likely to be the ones with the lowest ratio of people per healthcare worker(first quartile).

These findings could mean that states with a low ratio of people per healthcare worker are more prepared to fight against Covid-19 and legislators may want to conduct more research to understand the mechanism behind this.

However, the project did not consider other health relating factors including spending on the health of a state, the ratio of people per doctor, the ratio of people per nurse, etc. Since a high GDP per capita does not necessarily mean high spending on health, further studies can take a deeper look at how states spending on health affect Covid-19 mortality.