Suggest People Who Should Be Vaccinated

Carmen Li* Yibo Wang† Yihui Yang‡

Contents

1 Abstract **2**

2 Introduction **2**

2.1 Background 2

2.1.1 COVID-19 2

2.1.2 Vaccines 3

2.2 Objective 4

3 Data Summary and Analysis **4**

3.1 Data Summary 4

3.1.1 COVID-19 Data 4

3.1.2 Vaccine Data 5

3.2 Data Analysis 6

3.2.1 COVID-19 Data 6

3.2.2 Vaccine Data 10

4 Conclusion **12**

5 Appendix **13**

5.1 Vaccination rate for each state 13

5.2 Code 14

6 Supplementary material **23**

7 References **23**

*Department of Statistics, 917315244, bbyli@ucdavis.edu
†Department of Statistics, 916115267, yibwang@ucdavis.edu
‡Department of Statistics, 917297320, eyhyang@ucdavis.edu

1 Abstract

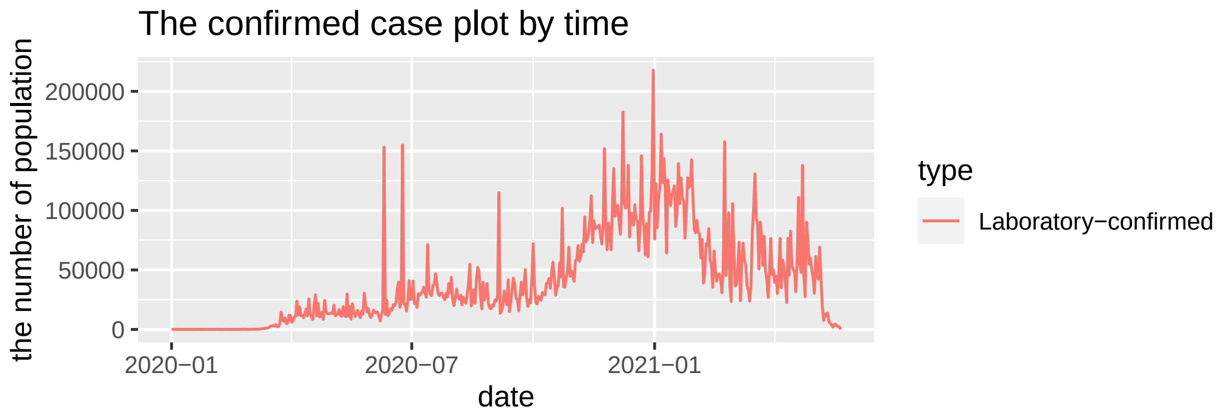
In this report, our main purpose is to summarize the susceptible population and the high-death population through data analysis, so as to provide suggestions for a certain group to be vaccinated as soon as possible. We use pie charts to visualize COVID-19 data to facilitate comparison of ratios. And then provide advice from two aspects of age and race. The report also added the most additional information about the current state of the vaccination process.This includes comparing the proportion of vaccinated people in different age groups, so that we can know about that the trend of vaccination for the different age.

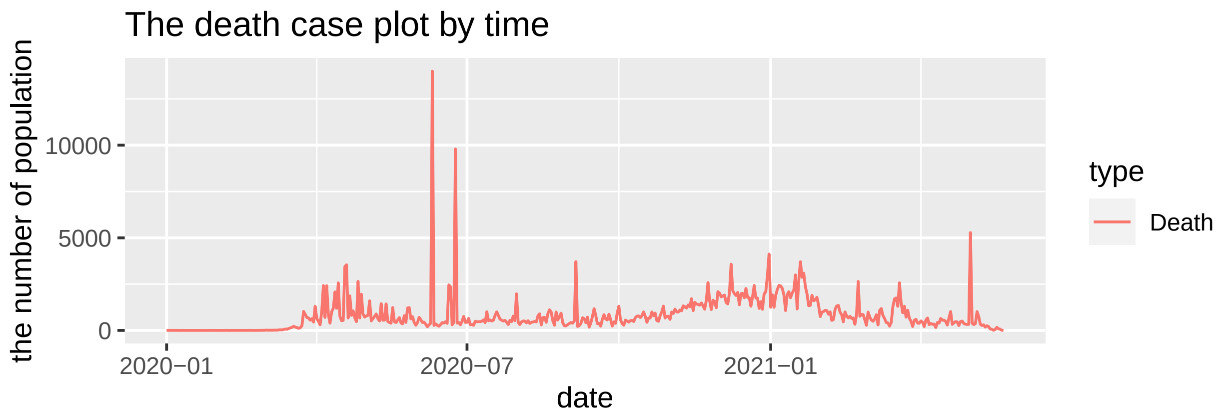
2 Introduction

2.1 Background

2.1.1 COVID-19

In late December 2019, investigation of a cluster of pneumonia cases of unknown origin in Wuhan, China resulted in identification of a novel coronavirus. The virus is distinct from both severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), although closely related. Early epidemiologic findings indicate COVID-19 may be less severe1 than SARS or MERS, but evidence suggests that the virus is more contagious than its predecessors. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a newly identified pathogen and it is assumed there is no pre-existing human immunity to the virus. Initial seroconversion, including neutralizing antibodies, has been documented and there is some evidence that immunity to SARS-CoV-2 re-challenge during early convalescence is likely. The extent of long-term immunity from anamnestic responses is unknown currently. At the beginning of the pandemic, everyone was assumed to be susceptible; it is now known that there are risk factors that increase an individual's illness severity.
According to Worldometer, up to now, there are 34,239,313 confirmed cases and 613,084 deaths from 490,263,839 tests in the US. Although the CDC proved that the death rate of COVID-19 in the United States is not as high as that of influenza, the virus is particularly contagious.





It can be seen that there have been two increases in the number of confirmed diagnoses in July 2020, but the main peak of the number of confirmed diagnoses is around January 2021. According to the analysis of the social situation at the time, January 2021 was the period of the presidential election. At that time, the society was very chaotic, with a large number of demonstrations and rallies.
The peak of the death toll only appeared in July 2020. It can be inferred that the hospital suddenly accepted a large number of patients, and the medical system was paralyzed for a while, resulting in a large number of patients not being treated in time.

2.1.2 Vaccines

Studies show that COVID-19 vaccines are effective at keeping you from getting COVID-19. Getting a COVID-19 vaccine will also help keep you from getting seriously ill even if you do get COVID-19.
COVID-19 vaccines teach our immune systems how to recognize and fight the virus that causes COVID-19. It typically takes 2 weeks after vaccination for the body to build protection (immunity) against the virus that causes COVID-19. That means it is possible a person could still get COVID-19 before or just after vaccination and then get sick because the vaccine did not have enough time to provide protection. People are considered fully vaccinated 2 weeks after their second dose of the Pfizer-BioNTech or Moderna COVID-19 Vaccine, or 2 weeks after the single-dose Johnson & Johnson's Janssen COVID-19 Vaccine.
Vaccines are now more widely accessible in the U.S. The federal government continues to work toward making vaccines widely available for everyone at no cost.
Many doctors' offices, retail pharmacies, hospitals, and clinics offer COVID-19 vaccinations. Your doctor's office or local pharmacy may contact you with information about their vaccination plans. According to Our World in Data, there are more than 140,000,000 people who are fully vaccinated.

2.2 Objective

Since there are lots of people diagnosed with covid-19 every day, we suggest that people should be vaccinated as soon as possible to prevent the spread of more viruses. Based on the COVID-19 Case Surveillance Public Use Data and COVID-19 Vaccinations Data, we know the basic situation of the number of confirmed people, the number of deaths and so on, and we analyze the number of confirmed cases and the number of deaths from the aspects of age, race to indicate which groups of people are supposed to be vaccinated earlier.

3 Data Summary and Analysis

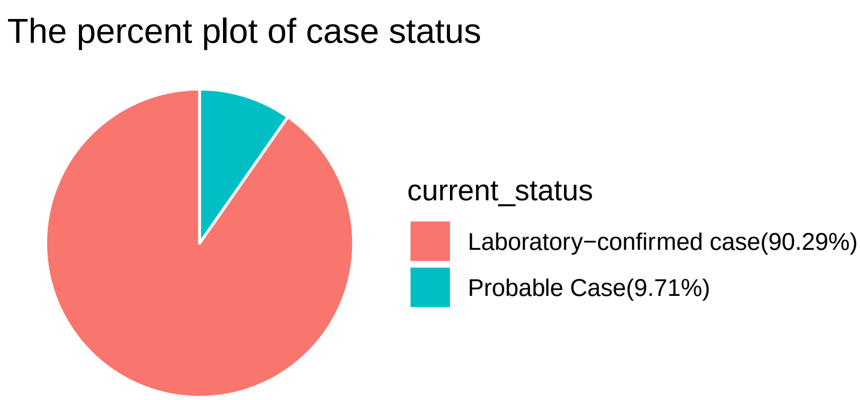
3.1 Data Summary

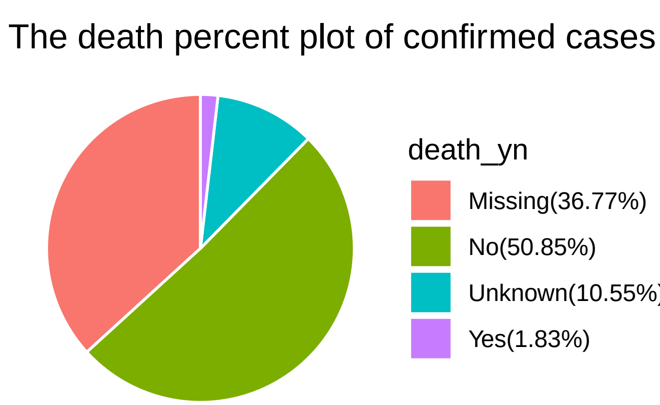
3.1.1 COVID-19 Data

The "COVID-19 Case Surveillance Public Use Data" dataset has 12 elements for all COVID-19 cases shared with CDC and includes demographics, any exposure history, disease severity indicators and outcomes, presence of any underlying medical conditions and risk behaviors, and no geographic data. The time span of this dataset is:

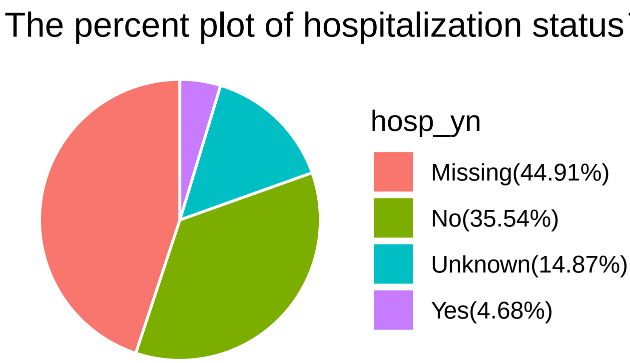
```
## [1] "2020/01/01" "2021/05/22"
```

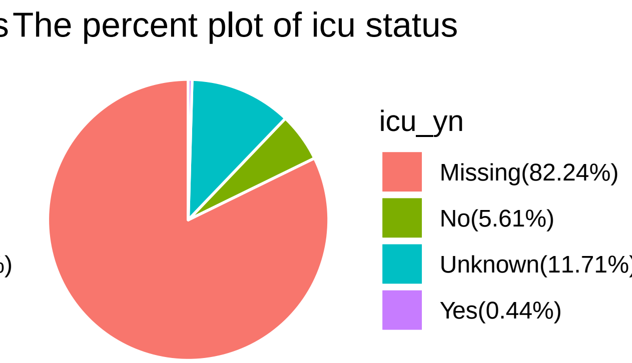
This data contains a total of 25,607,582 COVID-19 cases. We have drawn two pie charts to facilitate understanding of how many of these cases are confirmed cases and how many deaths.

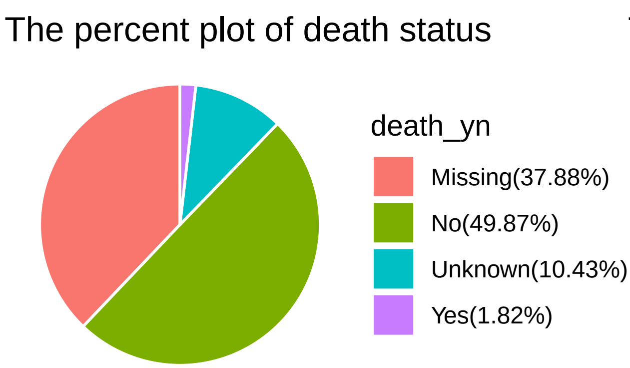


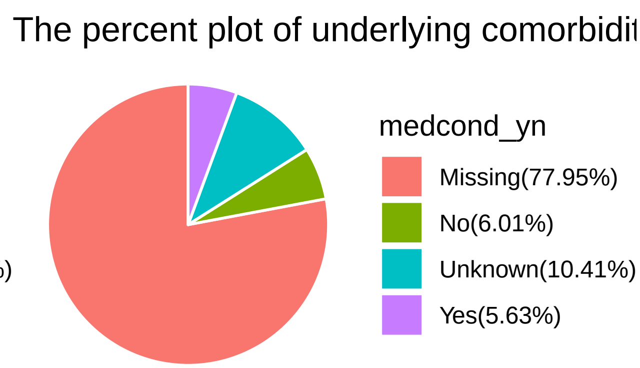


We can see that more than 90% of the cases in the data are confirmed cases. Among the confirmed cases, 1.83% of the cases have been confirmed dead. At the same time, we found that there are many missing and unknown values in the columns of Hospitalization, ICU, Presence of underlying comorbidity or disease and Death.





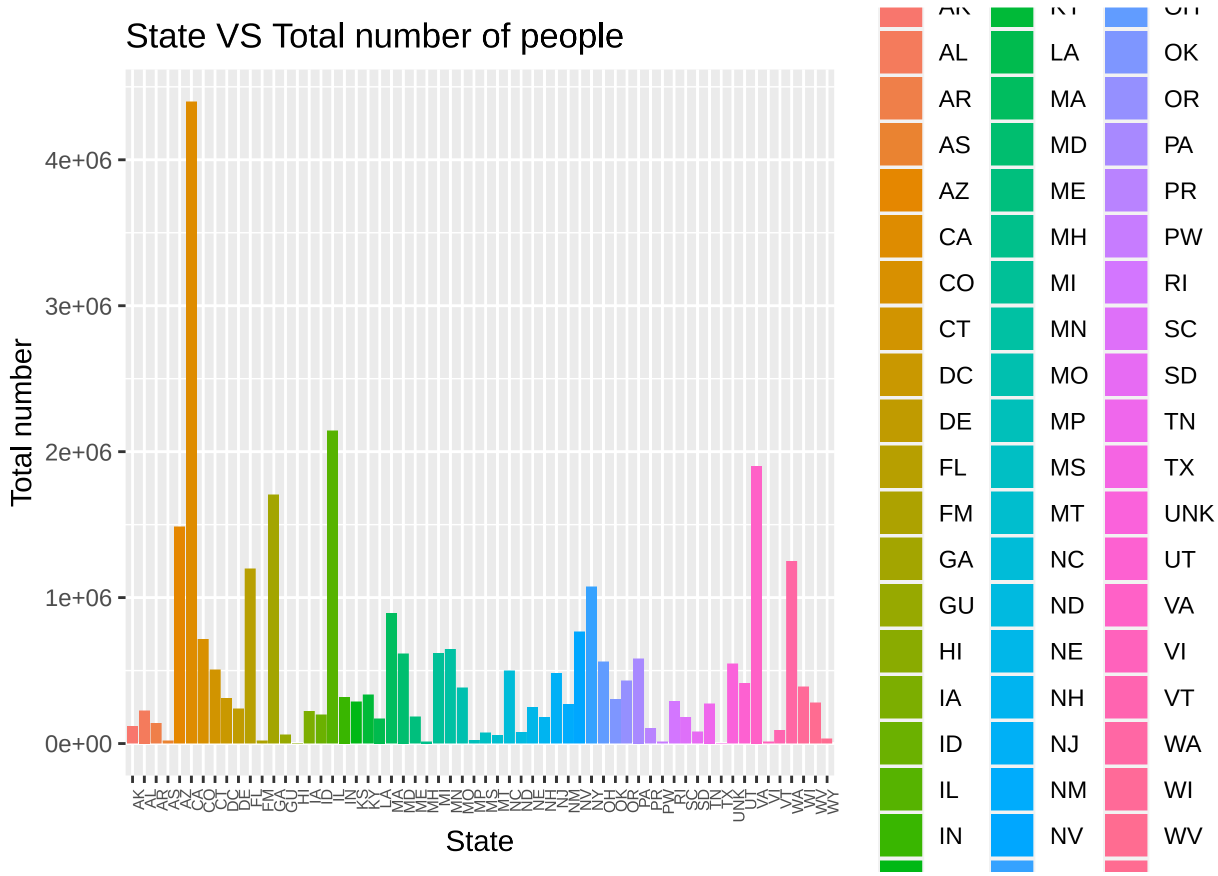




For the accuracy of the report, we will avoid using Hospitalization, ICU, Presence of underlying comorbidity or disease and Death as the main analysis objects.

3.1.2 Vaccine Data

The data is about the overall US COVID-19 Vaccine administration data at county level from 12/13/2020 to 06/02/2021. We can know about the situation of people getting vaccine of each state in Unite States. There will be saparated into 3 groups using the their age. The first group is 12 years old to 18 years old, the second group is 18 years old to 65 years old and the third group is over 65 years old. Also, it will show the Percent or total number of people who are fully vaccinated (have second dose of a two-dose vaccine or one dose of a single-dose vaccine) based on the jurisdiction and county where recipient lives.
As the below output, we can know that the basic distribution of different states of total number of people who are fully vaccinated (have second dose of a two-dose vaccine or one dose of a single-dose vaccine). Different colors represent one state. We can know that some states have high number of people.



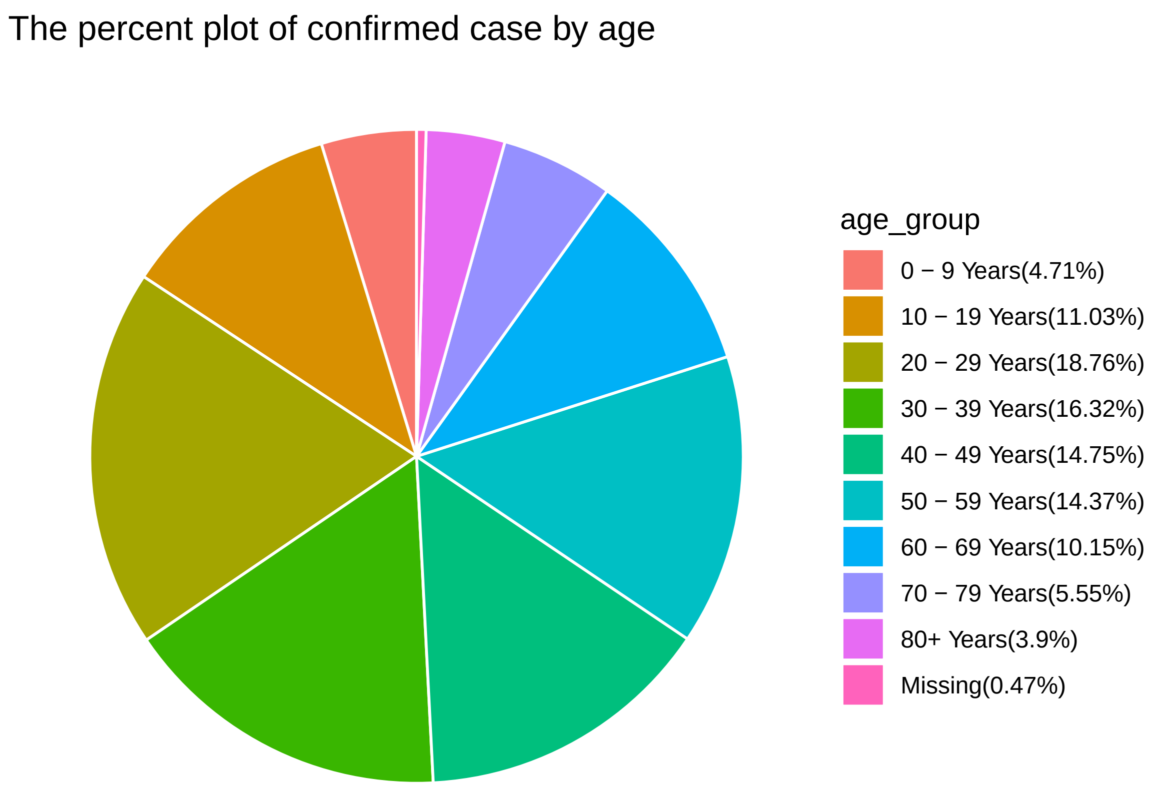
As below output, we are interested in the 6 most populated states of Unite States and want to know the trend of population who are fully vaccinated (have second dose of a two-dose vaccine or one dose of a single-dose vaccine). California is much more than others. These is no doubt that California is the most populous state in the U.S. Illinois even though do not much population as California, but It has a high vaccination population. Virginia and Georgia, Arizona and Washington have the similar vaccination population.

```
## # A tibble: 6 x 2
## # Groups:   Recip_State [6]
##   Recip_State Series_Complete_Yes
##   <chr>                     <int>
## 1 CA                      4397153
## 2 IL                      2142692
## 3 VA                      1901928
## 4 GA                      1705467
## 5 AZ                      1485670
## 6 WA                      1247940
```

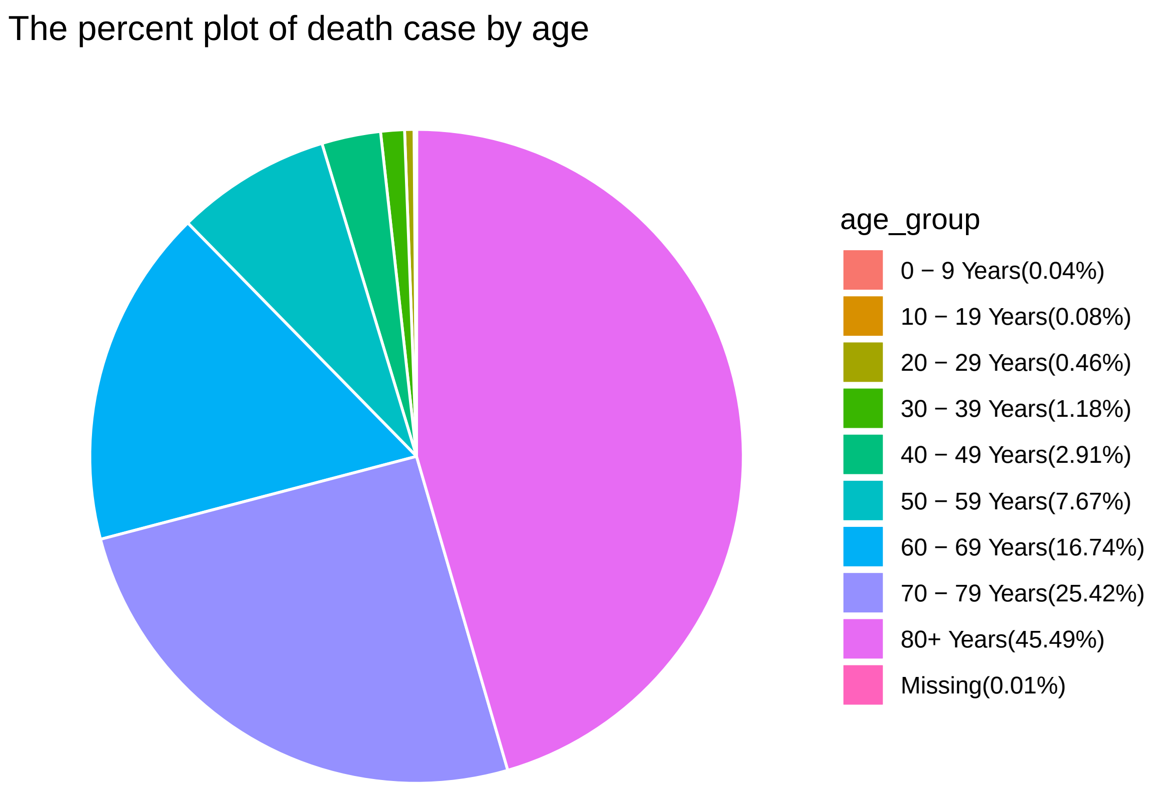
3.2 Data Analysis

3.2.1 COVID-19 Data

As we all know, patients with COVID-19 range from newborn babies to 80-year-olds, so we want to analyze the diagnosis rate and death rate of different age groups (every 10 years old).

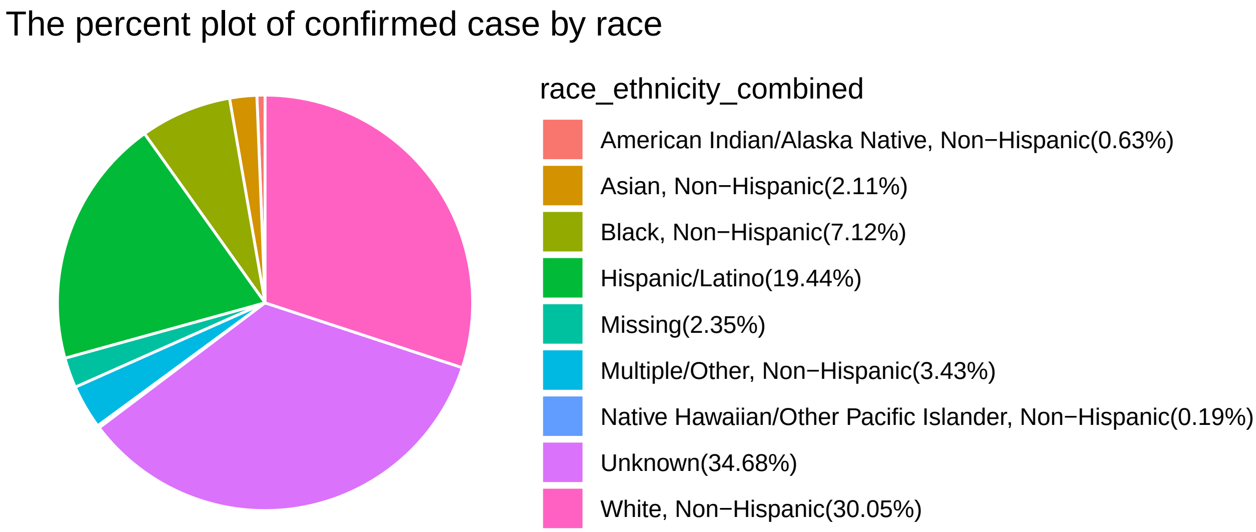


It can be seen that the main age group is 20 to 59 years old. People in this age group need frequent contact with others due to work and other reasons. The possibility of contact with patients is higher and the diagnosis rate is higher. Based on this, we recommend that people aged 20-59 get vaccinated as soon as possible to reduce their chances of getting sick.
At the same time, we analyze the population at risk from the mortality rate. So we drew the plot below.



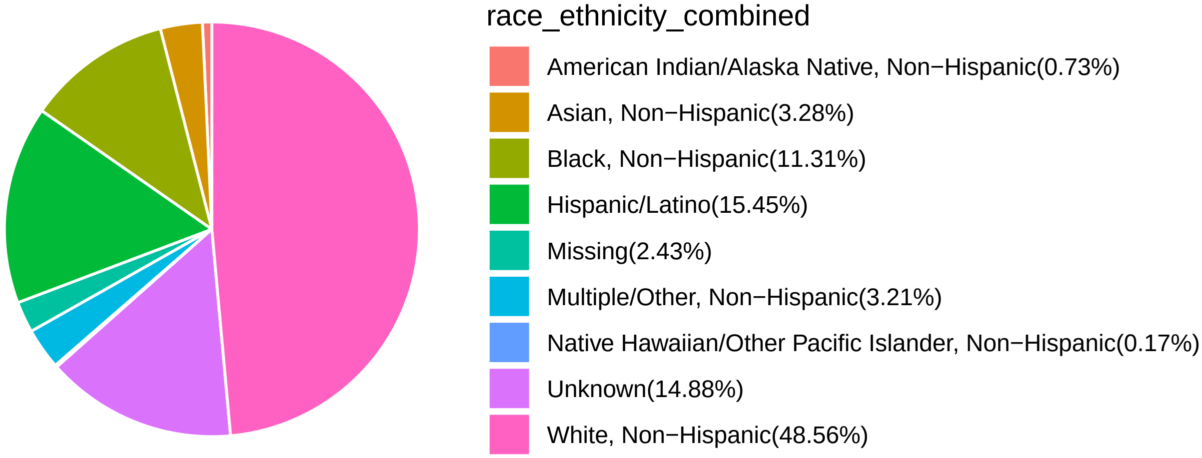
It can be seen that the death rate of people over 60 years old is higher, which may be related to the body's immune capacity. With age, the body's immune capacity will gradually decline. Therefore, we recommend that people over 60 years of age should be vaccinated as soon as possible on the premise of choosing a safe and suitable vaccine.

As a country of immigrants, the United States has gathered people of all races. The number of white people in the entire American society is the largest, reaching 77% of the total population. We will discuss the diagnosis rate and death rate of COVID-19 from the aspect of race. The plot below is the proportion of each race in the confirmed cases.



It can be seen that although the race is not indicated in 34% of the cases, whites are still the race with the largest number of diagnoses. At the same time, the most populous ethnic minority in the United States, Hispanics and Latinos have a very high diagnosis rate. We think this has something to do with their social personality. Social interaction increases the probability of contact with others. Therefore, we recommend that people of these two races get vaccinated as soon as possible.
Next is a plot of the mortality rate of each race.

The percent plot of death case by race



According to the ratio chart, it can be seen that the races with the highest mortality rates are whites, Hispanics, and Latinos. This result is not surprising. These races have a high diagnosis rate. Therefore, according to the death ratio chart, we still recommend that people of these two races get vaccinated as soon as possible.

3.2.2 Vaccine Data

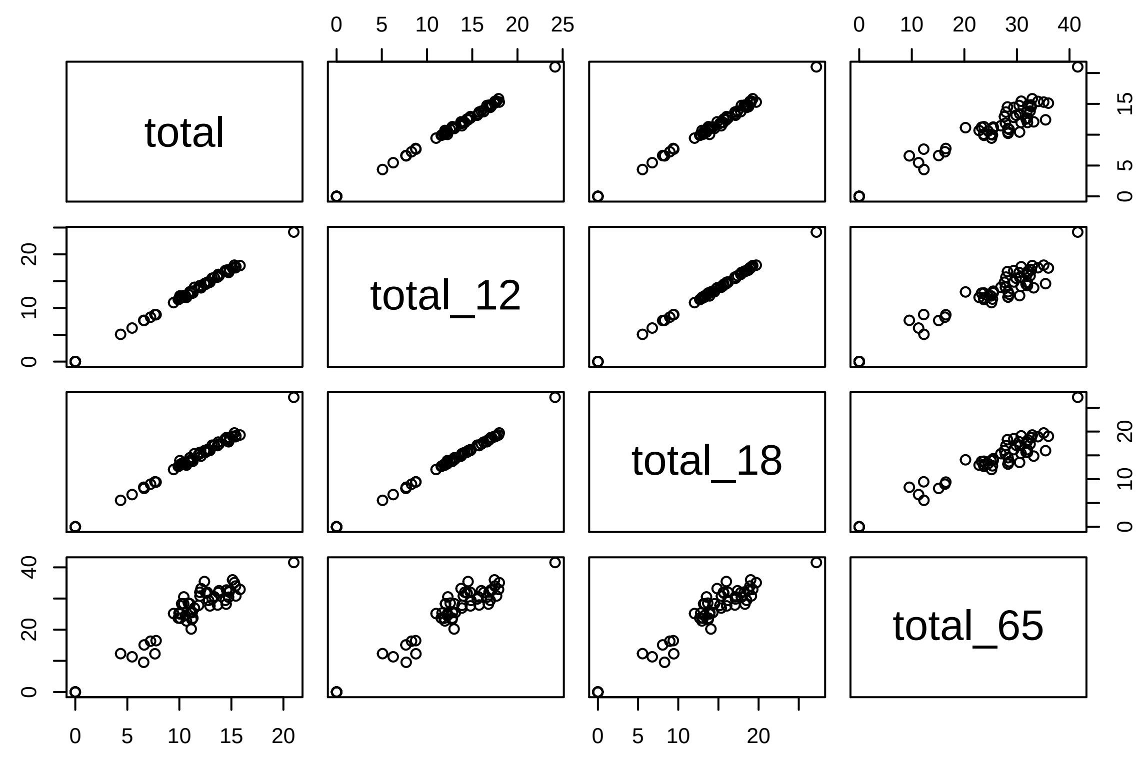
As the below output showing, for the 6 top states, we can see that different percentage of age are fully vaccinated (have second dose of a two-dose vaccine or one dose of a single-dose vaccine) based on the jurisdiction where recipient lives. For example, AZ(Arizona), there are total 14.872% getting the fully vaccinated or have second dose of a two-dose vaccine or one dose of a single-dose vaccine. For age of 12 to 18 years old is 17.517%, 18 to 65 is 19.057% and 65 or above is 29.888%. AZ has the highest vaccination rate in these states. On the other hand, CA has the most population but has a lower vaccination rate. We can know that vaccine penetration rate of CA is relatively low from 12/13/2020 to 06/02/2021.

```
## # A tibble: 6 x 5
##   Recip_State total total_12 total_18 total_65
##   <chr>       <dbl>    <dbl>    <dbl>    <dbl>
## 1 AZ          14.9     17.5     19.1      29.9
## 2 CA          11.2     13.1     14.3      25.6
## 3 GA           4.42     5.17     5.64     12.4
## 4 IL          12.9     15.0     16.4      29.6
## 5 VA           5.28     6.04     6.53     11.0
## 6 WA          13.6     15.7     17.1      32.1
```

As the output below showing, they are the percentage of total number people who are fully vaccinated (have second dose of a two-dose vaccine or one dose of a single-dose vaccine). They have the similar vaccination rate between the age of 12 years old and 65 years old. The age of 18 years old to 65 years old has a slightly higher vaccination rate since some parents are worried about the side effects of the vaccine will have a more profound impact on the child. But the age over 65 years old have much high vaccination rate. These no doubt that these people have more different disease than the young people. They need more protection.

```
## # A tibble: 6 x 4
##   Recip_State total_12 total_18 total_65
##   <chr>          <dbl>    <dbl>    <dbl>
## 1 AZ             17.5     19.1      29.9
## 2 CA             13.1     14.3      25.6
## 3 GA              5.17     5.64     12.4
## 4 IL             15.0     16.4      29.6
## 5 VA              6.04     6.53     11.0
## 6 WA             15.7     17.1      32.1
```

As the output below showing, we can see that people are getting vaccination increasingly from 12/13/2020 to 06/02/2021 since the graphs are increasing. Also, people are getting vaccination increasingly with the increasing age. That means that the vaccination rate will increase with increasing age. This result is same with comparing 6 top states as the table of above.



4 Conclusion

To sum up, according to the analysis of COVID-19 Data and Vaccine data, we learn about specific numbers of confirmed cases and dead cases among people with distinctive races and ages. From angle of age, we can know the highest diagnosis rate is in people who are 20 to 59 years old, since they are the majority of workers in the US so that they may get closer contact with each other. In addition, the highest mortality rate is from the people who are over 60 years old, the possible reason is that they have lower resistance ability to disease or virus than younger generation. From angle of race, we can know that Whites, Hispanics, and Latinos have the fairly high mortality rate because of the high diagnosis rate. Therefore, these groups of people are supposed to get vaccinated as soon as possible to reduce the transmission of COVID-19 and death rate.
In terms of vaccine data, the population over 65 years old in each state accounted for most of the proportion of vaccinated, while the 18-year-old and 12-year-old declined. Combined with the confirmed COVID-19 data, we believe that the 18-year-old vaccination rate needs to be increased to facilitate the daily social and work of young people.

5 Appendix

5.1 Vaccination rate for each state

```
##    Recip_State     total  total_12  total_18  total_65
## 1           AK 20.995656 24.158883 27.193464 41.560609
## 2           AL  9.447494 11.001018 12.028233 25.170116
## 3           AR  9.898500 11.570595 12.679914 23.725174
## 4           AS  0.000000  0.000000  0.000000  0.000000
## 5           AZ 14.424782 16.985262 18.506041 29.415575
## 6           CA 11.234648 13.134342 14.299694 25.497201
## 7           CO  7.654979  8.772372  9.444484 12.266592
## 8           CT 15.398317 17.504725 18.918252 33.993592
## 9           DC  6.618879  7.661947  8.058997 15.109440
## 10          DE 10.588012 12.247661 13.257602 24.522807
## 11          FL 12.083657 13.763576 14.838433 33.188703
## 12          FM  0.000000  0.000000  0.000000  0.000000
## 13          GA  4.355173  5.082015  5.553603 12.309125
## 14          GU  6.573373  7.697337  8.301479  9.533728
## 15          HI  0.000000  0.000000  0.000000  0.000000
## 16          IA 14.541225 17.062753 18.748692 32.734648
## 17          ID 11.014101 13.031265 14.505403 28.308104
## 18          IL 12.809959 14.855360 16.205612 29.410032
## 19          IN 12.410730 14.538013 15.985105 35.446573
## 20          KS 13.155242 15.551972 17.133068 29.841725
## 21          KY 11.849762 13.871416 15.194839 27.819203
## 22          LA 10.427727 12.301843 13.518708 30.507301
## 23          MA 11.204695 12.765153 13.724951 23.267249
## 24          MD 14.827561 17.180912 18.684101 32.649209
## 25          ME 15.828826 17.909346 19.273708 32.913899
## 26          MH  0.000000  0.000000  0.000000  0.000000
## 27          MI 14.782741 16.858809 18.229913 32.090834
## 28          MN 15.288872 17.987075 19.704242 35.092070
## 29          MO  9.971006 11.662820 12.798431 25.345952
## 30          MP  0.000000  0.000000  0.000000  0.000000
## 31          MS 10.221502 12.037953 13.276971 28.322336
## 32          MT 13.361582 15.668276 17.107204 30.546684
## 33          NC 12.544556 14.432483 15.671276 31.737676
## 34          ND 13.735767 16.262100 17.783522 31.900635
## 35          NE 10.084874 11.891203 13.096882 23.668052
## 36          NH  7.765643  8.743886  9.386183 16.475913
## 37          NJ 14.499101 16.791510 18.298043 28.198783
## 38          NM 11.140291 12.992370 14.064277 20.219076
## 39          NV 11.061494 12.817556 13.937698 25.492887
## 40          NY 15.431496 17.691960 19.099575 30.820216
## 41          OH 12.684520 14.792348 16.197367 31.981848
## 42          OK 11.975114 14.177253 15.623008 32.002602
## 43          OR 13.671467 15.762239 17.051879 27.915076
## 44          PA 12.939408 14.801625 16.026494 27.623775
## 45          PR 10.683568 11.978451 12.954590 22.804506
## 46          PW  0.000000  0.000000  0.000000  0.000000
## 47          RI 14.734791 16.621283 17.855879 30.418465
## 48          SC 10.868383 12.559151 13.668816 28.516186
## 49          SD 11.448416 13.855169 15.351141 26.945543
```

```
## 50          TN 10.418924 12.049037 13.157263 28.301775
## 51          TX  0.000000  0.000000  0.000000  0.000000
## 52         UNK  0.000000  0.000000  0.000000  0.000000
## 53          UT 10.051202 12.256323 13.895559 25.051241
## 54          VA  5.461593  6.259753  6.762560 11.294684
## 55          VI  0.000000  0.000000  0.000000  0.000000
## 56          VT 11.298719 12.799418 13.759061 23.722740
## 57          WA 13.816928 16.008406 17.378970 32.504494
## 58          WI 15.113989 17.457253 19.010595 35.984291
## 59          WV  7.233458  8.285177  8.949205 16.308310
## 60          WY 11.982881 14.015131 15.375364 30.791319
```

5.2 Code

```
## ----setup, include=FALSE------------------------------------------------------------
knitr::opts_chunk$set(echo = TRUE)


## ----include=FALSE-------------------------------------------------------------------
library(data.table)
library(tidyverse)
library(lubridate)
library(ggplot2)
library(tidyr)
library(gridExtra)
library(dplyr)


## ----message=FALSE, warning=FALSE, include=FALSE------------------------------------
# Load Data Part
public.data <- read.csv(
  "COVID-19_Case_Surveillance_Public_Use_Data.csv", header = T
  )
vaccination.data <- read.csv(
  "COVID-19_Vaccinations_in_the_United_States_County.csv", header = T
  )


## ----include=FALSE-------------------------------------------------------------------
# Summary Part
# Public Data
# Time range
time.line <- public.data %>%
  group_by(cdc_report_dt) %>%
  filter(cdc_report_dt != "")

time.bound <- c(min(time.line$cdc_report_dt), max(time.line$cdc_report_dt))


# The confirm and probable percent plot
percent_stat <- public.data %>%
  group_by(current_status) %>%
```

```
  summarise(confirm_count = n()) %>%
  na.omit()

lable.value.stat <- paste(
  "(",
  round(percent_stat$confirm_count/sum(percent_stat$confirm_count) * 100, 2),
  "%)", sep = ""
  )
lable.stat <- paste(percent_stat$current_status, lable.value.stat, sep = "")

percent.stat <- ggplot(percent_stat,
                       aes(x = "", y = confirm_count, fill = current_status)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y", start = 0) +
  ggtitle("The percent plot of case status") +
  scale_fill_discrete(labels = lable.stat) +
  theme_void()


# The death percent plot
# The confirmed case
percent_death_confirm <- public.data %>%
  filter(current_status == "Laboratory-confirmed case") %>%
  group_by(death_yn) %>%
  summarise(confirm_count = n()) %>%
  na.omit()

lable.value.death.con <- paste(
  "(",
  round(
    percent_death_confirm$confirm_count/sum(percent_death_confirm$confirm_count
                                            ) * 100, 2),
  "%)", sep = "")
lable.death.con <- paste(
  percent_death_confirm$death_yn,
  lable.value.death.con, sep = ""
  )

percent.death.con <- ggplot(percent_death_confirm,
                            aes(x = "", y = confirm_count, fill = death_yn)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y", start = 0) +
  ggtitle("The death percent plot of confirmed cases") +
  scale_fill_discrete(labels = lable.death.con) +
  theme_void()


# The missing and unknown percent plot of hosp, icu, death, and medcont

# hosp
percent_hosp <- public.data %>%
  group_by(hosp_yn) %>%
  summarise(confirm_count = n()) %>%
```

```
  na.omit()

lable.value.hosp <- paste(
  "(",
  round(
    percent_hosp$confirm_count/sum(percent_hosp$confirm_count) * 100, 2
    ),
  "%)", sep = ""
  )
lable.hosp <- paste(percent_hosp$hosp_yn, lable.value.hosp, sep = "")

percent.hosp <- ggplot(percent_hosp,
                       aes(x = "", y = confirm_count, fill = hosp_yn)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y", start = 0) +
  ggtitle("The percent plot of hospitalization status") +
  scale_fill_discrete(labels = lable.hosp) +
  theme_void()


# icu
percent_icu <- public.data %>%
  group_by(icu_yn) %>%
  summarise(confirm_count = n()) %>%
  na.omit()

lable.value.icu <- paste(
  "(",
  round(
    percent_icu$confirm_count/sum(percent_icu$confirm_count) * 100, 2
    ),
  "%)", sep = ""
  )
lable.icu <- paste(percent_icu$icu_yn, lable.value.icu, sep = "")

percent.icu <- ggplot(percent_icu,
                      aes(x = "", y = confirm_count, fill = icu_yn)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y", start = 0) +
  ggtitle("The percent plot of icu status") +
  scale_fill_discrete(labels = lable.icu) +
  theme_void()


# death
percent_death <- public.data %>%
  group_by(death_yn) %>%
  summarise(confirm_count = n()) %>%
  na.omit()

lable.value.death <- paste(
  "(",
  round(
```

```r
    percent_death$confirm_count/sum(percent_death$confirm_count) * 100, 2
    ),
  "%)", sep = ""
  )
lable.death <- paste(
  percent_death$death_yn, lable.value.death, sep = ""
  )

percent.death <- ggplot(percent_death,
                        aes(x = "", y = confirm_count, fill = death_yn)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y", start = 0) +
  ggtitle("The percent plot of death status") +
  scale_fill_discrete(labels = lable.death) +
  theme_void()


# medcont
percent_med <- public.data %>%
  group_by(medcond_yn) %>%
  summarise(confirm_count = n()) %>%
  na.omit()

lable.value.med <- paste(
  "(",
  round(
    percent_med$confirm_count/sum(percent_med$confirm_count) * 100, 2
    ),
  "%)", sep = ""
  )
lable.med <- paste(percent_med$medcond_yn, lable.value.med, sep = "")

percent.med <- ggplot(percent_med,
                      aes(x = "", y = confirm_count, fill = medcond_yn)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y", start = 0) +
  ggtitle("The percent plot of underlying comorbidity or disease") +
  scale_fill_discrete(labels = lable.med) +
  theme_void()


## ----include=FALSE-------------------------------------------------------------------------------
# Summary Part
# Vaccination Data
# State plot
state_total <- vaccination.data %>%
  select(Recip_State, Series_Complete_Yes) %>%
  group_by(Recip_State) %>%
  slice(which.max(Series_Complete_Yes))

state.plot <- ggplot(state_total,
                     aes(x = Recip_State, y = Series_Complete_Yes)) +
  geom_bar(stat = "identity", aes(fill = Recip_State)) +
```

```r
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 6)) +
  ggtitle("State VS Total number of people") +
  xlab("State") +
  ylab("Total number")


top.6 <- state_total %>%
  arrange(-Series_Complete_Yes) %>%
  head()


## ----include=FALSE--------------------------------------------------------------------------------
# Analysis part
# Public Data
# The confirm time-line plot
confirm_stat <- public.data %>%
  group_by(cdc_report_dt) %>%
  summarise(confirm_count = sum(
    current_status == "Laboratory-confirmed case")
    ) %>%
  filter(cdc_report_dt != "") %>%
  gather("type", "count_value", -cdc_report_dt) %>%
  mutate(type = if_else(type == "confirm_count",
                        "Laboratory-confirmed", "Death"),
         cdc_report_dt = ymd(cdc_report_dt)) %>%
  arrange(cdc_report_dt)

confirm.plot <- ggplot(confirm_stat) +
  geom_line(aes(x = cdc_report_dt,
                y = count_value, group = type, color = type)) +
  xlab("date") +
  ylab("the number of population") +
  ggtitle("The confirmed case plot by time")


# The death time-line plot
death_stat <- public.data %>%
  group_by(cdc_report_dt) %>%
  summarise(death_count = sum(death_yn == "Yes")) %>%
  filter(cdc_report_dt != "") %>%
  gather("type", "count_value", -cdc_report_dt) %>%
  mutate(type = if_else(type == "confirm_count",
                        "Laboratory-confirmed", "Death"),
         cdc_report_dt = ymd(cdc_report_dt)) %>%
  arrange(cdc_report_dt)

death.plot <- ggplot(death_stat) +
  geom_line(aes(x = cdc_report_dt,
                y = count_value, group = type, color = type)) +
  xlab("date") +
  ylab("the number of population") +
  ggtitle("The death case plot by time")
```

```r
# The confirm plot by age
confirm_age_stat <- public.data %>%
  filter(current_status == "Laboratory-confirmed case") %>%
  group_by(age_group) %>%
  summarise(confirm_count = n()) %>%
  na.omit()

lable.value.con.age <- paste(
  "(",
  round(
    confirm_age_stat$confirm_count/sum(confirm_age_stat$confirm_count) * 100, 2
    ),
  "%)", sep = ""
  )
lable.con.age <- paste(
  confirm_age_stat$age_group, lable.value.con.age, sep = ""
  )

confirm.age <- ggplot(confirm_age_stat,
                      aes(x = "", y = confirm_count, fill = age_group)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y", start = 0) +
  ggtitle("The percent plot of confirmed case by age") +
  scale_fill_discrete(labels = lable.con.age) +
  theme_void()


# The confirm plot by race
confirm_race_stat <- public.data %>%
  filter(current_status == "Laboratory-confirmed case") %>%
  group_by(race_ethnicity_combined) %>%
  summarise(confirm_count = n()) %>%
  na.omit()

lable.value.con.race <- paste(
  "(",
  round(
    confirm_race_stat$confirm_count/sum(
      confirm_race_stat$confirm_count
      ) * 100, 2
    ),
  "%)", sep = ""
  )
lable.con.race <- paste(confirm_race_stat$race_ethnicity_combined,
                        lable.value.con.race, sep = "")

confirm.race <- ggplot(confirm_race_stat,
                       aes(x = "", y = confirm_count,
                           fill = race_ethnicity_combined)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y", start = 0) +
  ggtitle("The percent plot of confirmed case by race") +
  scale_fill_discrete(labels = lable.con.race) +
```

```
  theme_void()


# The death plot by age
death_age_stat <- public.data %>%
  filter(death_yn == "Yes") %>%
  group_by(age_group) %>%
  summarise(death_count = n()) %>%
  na.omit()

lable.value.death.age <- paste(
  "(",
  round(death_age_stat$death_count/sum(death_age_stat$death_count) * 100, 2
        ),
  "%)", sep = ""
  )
lable.death.age <- paste(death_age_stat$age_group,
                         lable.value.death.age, sep = "")

death.age <- ggplot(death_age_stat,
                    aes(x = "", y = death_count, fill = age_group)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y", start = 0) +
  ggtitle("The percent plot of death case by age") +
  scale_fill_discrete(labels = lable.death.age) +
  theme_void()


# The death plot by race
death_race_stat <- public.data %>%
  filter(death_yn == "Yes") %>%
  group_by(race_ethnicity_combined) %>%
  summarise(death_count = n()) %>%
  na.omit()

lable.value.death.race <- paste(
  "(",
  round(death_race_stat$death_count/sum(death_race_stat$death_count) * 100, 2),
  "%)", sep = ""
  )
lable.death.race <- paste(death_race_stat$race_ethnicity_combined,
                          lable.value.death.race, sep = "")

death.race <- ggplot(death_race_stat,
                     aes(x = "", y = death_count,
                         fill = race_ethnicity_combined)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y", start = 0) +
  ggtitle("The percent plot of death case by race") +
  scale_fill_discrete(labels = lable.death.race) +
  theme_void()
```

```r
## ----message=FALSE, warning=FALSE, include=FALSE------------------------------------------
# Analysis part
# Vaccination Data
total_perctg <- vaccination.data %>%
  group_by(Recip_State) %>%
  filter(Recip_State == c("CA", "IL", "VA", "GA", "AZ", "WA")) %>%
  summarise(total = mean(Series_Complete_Pop_Pct),
            total_12 = mean(Series_Complete_12PlusPop_Pct),
            total_18 = mean(Series_Complete_18PlusPop_Pct),
            total_65 = mean(Series_Complete_65PlusPop_Pct))

total_perctg_all <- vaccination.data %>%
  group_by(Recip_State) %>%
  summarise(total = mean(Series_Complete_Pop_Pct),
            total_12 = mean(Series_Complete_12PlusPop_Pct),
            total_18 = mean(Series_Complete_18PlusPop_Pct),
            total_65 = mean(Series_Complete_65PlusPop_Pct))

per_data <- total_perctg_all[-1]


## ----echo=FALSE, message=FALSE, warning=FALSE-------------------------------------------
grid.arrange(confirm.plot, death.plot, ncol = 1)


## ----echo=FALSE---------------------------------------------------------------------------
time.bound


## ----echo=FALSE, message=FALSE, warning=FALSE-------------------------------------------
grid.arrange(percent.stat, percent.death.con, ncol = 1)


## ----echo=FALSE, message=FALSE, warning=FALSE-------------------------------------------
grid.arrange(percent.hosp, percent.icu, percent.death, percent.med, ncol = 2)


## ----echo=FALSE, message=FALSE, warning=FALSE-------------------------------------------
state.plot


## ----echo=FALSE, message=FALSE, warning=FALSE-------------------------------------------
top.6


## ----echo=FALSE, message=FALSE, warning=FALSE-------------------------------------------
confirm.age


## ----echo=FALSE, message=FALSE, warning=FALSE-------------------------------------------
death.age
```

```
## ----echo=FALSE, message=FALSE, warning=FALSE-----------------------------------------------------------
confirm.race


## ----echo=FALSE, message=FALSE, warning=FALSE-----------------------------------------------------------
death.race


## ----echo=FALSE, message=FALSE, warning=FALSE-----------------------------------------------------------
total_perctg


## ----echo=FALSE, message=FALSE, warning=FALSE-----------------------------------------------------------
total_perctg[-2]


## ----echo=FALSE, message=FALSE, warning=FALSE-----------------------------------------------------------
plot(per_data)


## ----echo=FALSE, message=FALSE, warning=FALSE-----------------------------------------------------------
print.data.frame(as.data.frame(total_perctg_all))
```

6 Supplementary material

The COVID-19 data is "COVID-19_Case_Surveillance_Public_Use_Data.csv", and the vaccine data is "COVID-19_Vaccinations_in_the_United_States_County.csv". Both of these data come from CDC and are analyzed through "Final Project.rmd".The links of datasets are: https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data/vbim-akqf and https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh
The "Final Project.Rmd" file includes all codes and the text part, and the "Final Project.R" file includes only codes.
We include two data files, one ".pdf" file, one ".rmd" file and one ".r" file in the "Final Project.zip" file.

7 References

https://www.cdc.gov/coronavirus/2019-ncov/vaccines/keythingstoknow.html?s_cid=10496:covid%2019%20vaccine%20information:sem.ga:p:RG:GM:gen:PTN:FY21

https://wwwn.cdc.gov/nndss/conditions/coronavirus-disease-2019-covid-19/case-definition/2020/08/05/

https://www.worldometers.info/coronavirus/

https://ourworldindata.org/covid-vaccinations?country=~USA

https://en.wikipedia.org/wiki/Race_and_ethnicity_in_the_United_States