final

Haoran Gu

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0

## Warning: package 'ggplot2' was built under R version 4.0.5

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

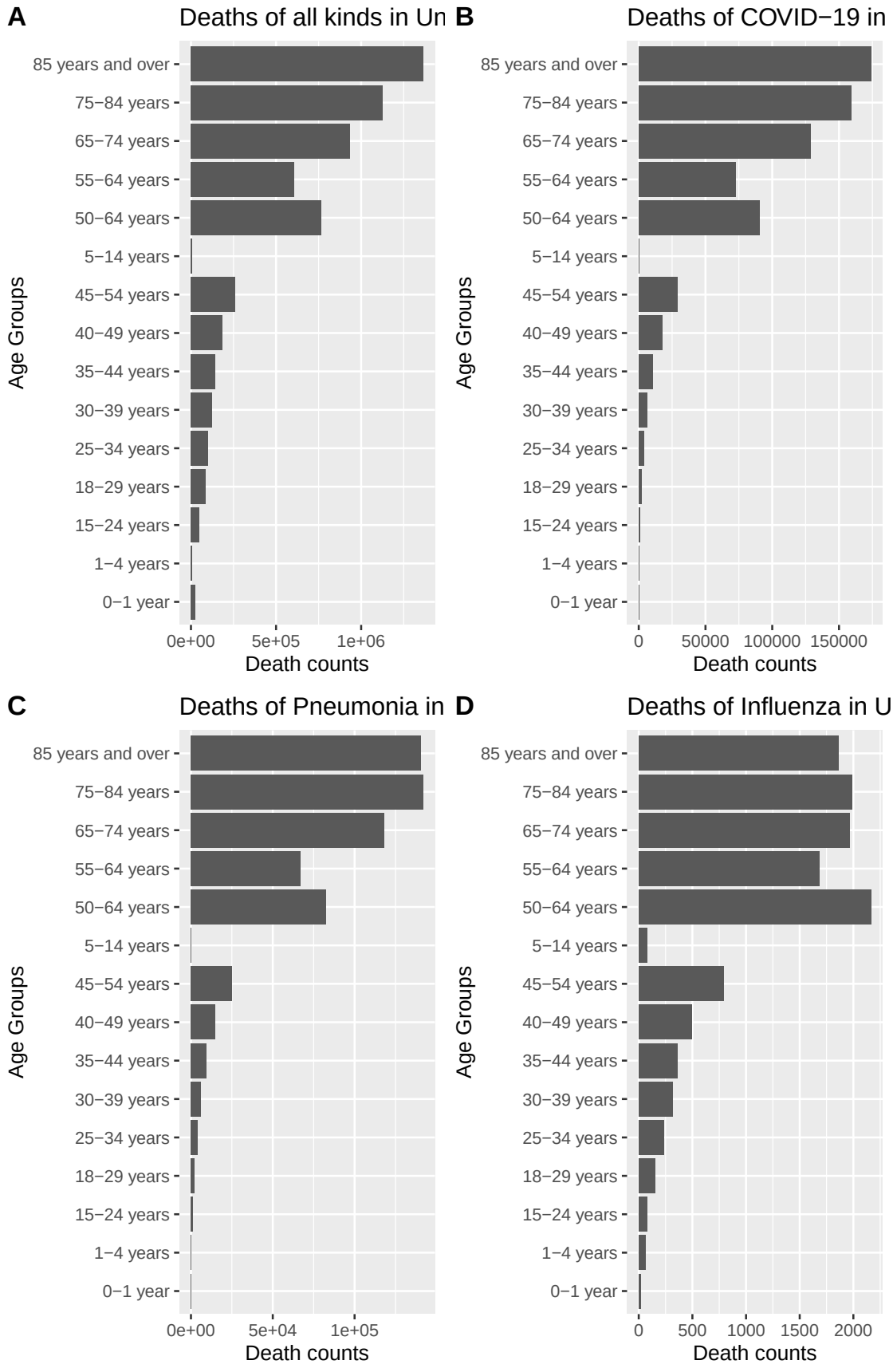
```
library(dplyr)
library(knitr)
library(questionr)
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 4.0.5
```

```
main <- read.csv("COVID_Main.csv")

race <- read.csv("COVID_Race.csv")
```

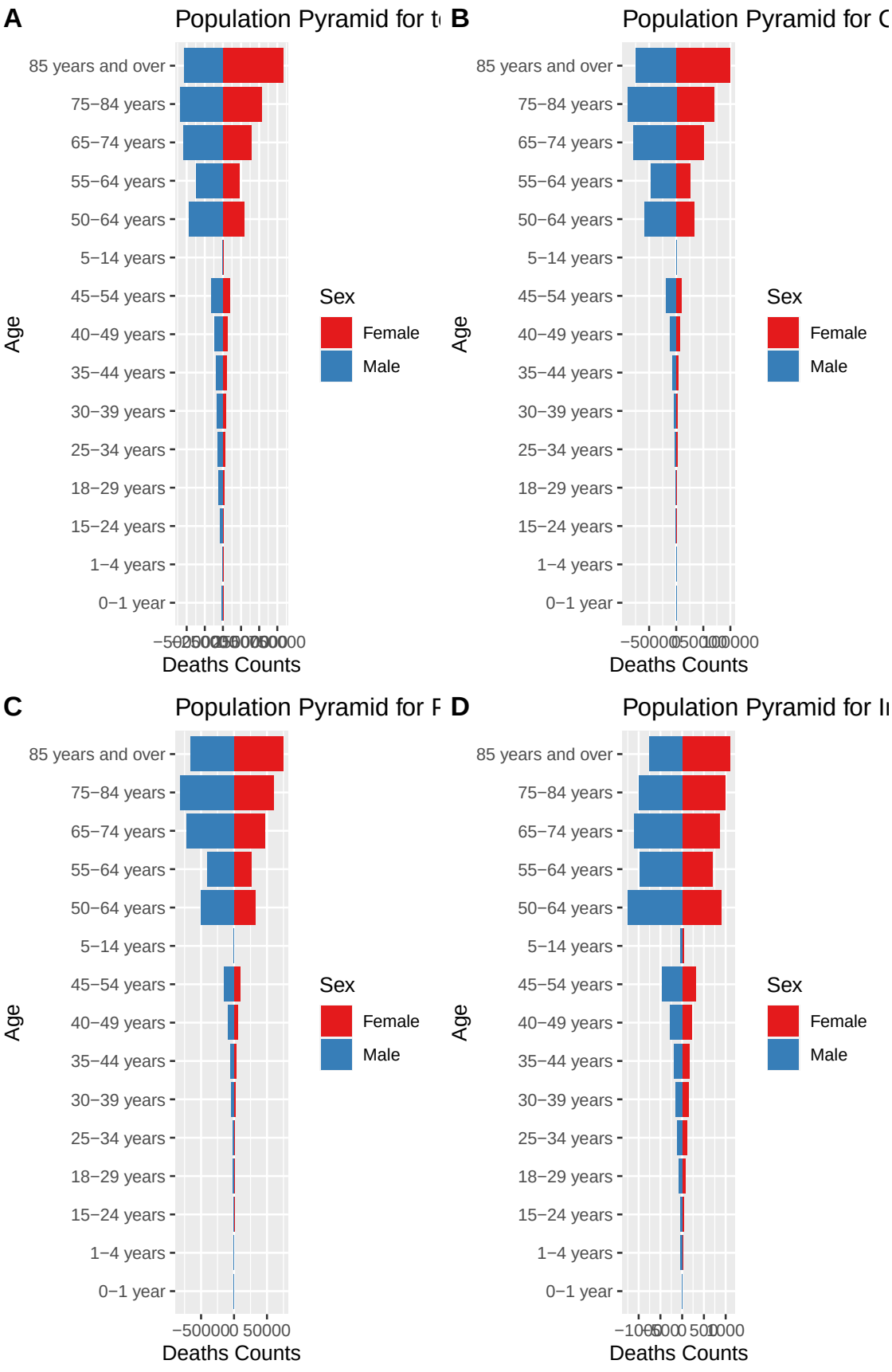
```
main.total = main %>% filter(State == "United States" & Group == "By Total" & Sex != "All Sexes")
main.total1 = main.total %>% filter(Age.Group != "All Ages" & Age.Group !="0-17 years")

tt <- ggplot(data = main.total1, aes(x = Total.Deaths, y = Age.Group)) +
  xlab("Death counts") + ylab("Age Groups") + ggtitle("Deaths of all kinds in United States from 2020/0
  geom_bar(stat="identity")
cv <-ggplot(data = main.total1, aes(x = COVID.19.Deaths, y = Age.Group)) +
  xlab("Death counts") + ylab("Age Groups") + ggtitle("Deaths of COVID-19 in United States from 2020/01
  geom_bar(stat="identity")
pn <- ggplot(data = main.total1, aes(x = Pneumonia.Deaths, y = Age.Group)) +
  xlab("Death counts") + ylab("Age Groups") + ggtitle("Deaths of Pneumonia in United States from 2020/0
  geom_bar(stat="identity")
In <- ggplot(data = main.total1, aes(x = Influenza.Deaths, y = Age.Group)) +
  xlab("Death counts") + ylab("Age Groups") + ggtitle("Deaths of Influenza in United States from 2020/0
  geom_bar(stat="identity")
```

```
ggarrange(tt,cv,pn,In,
                    labels = c("A", "B", "C", "D"),
                    ncol = 2, nrow = 2)
```

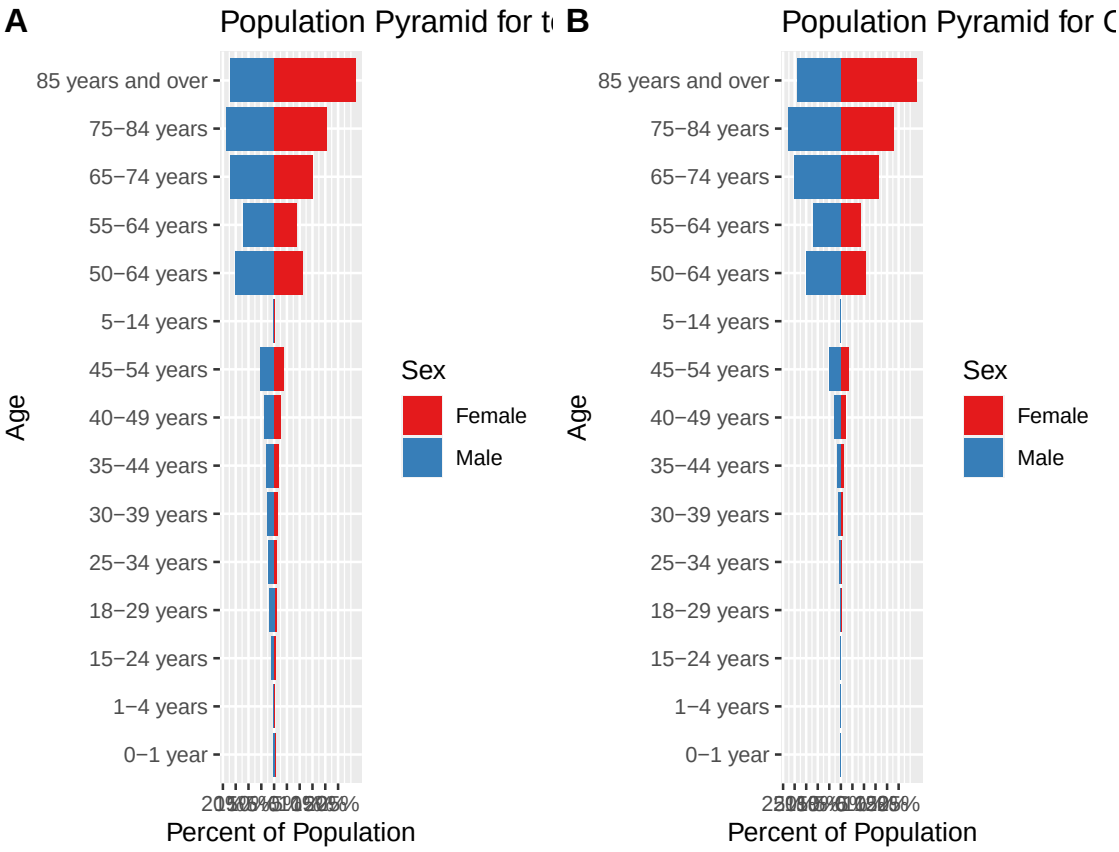

A
Deaths of all kinds in Un
B
Deaths of COVID−19 in
C
Deaths of Pneumonia in
D
Deaths of Influenza in U
Age Groups
85 years and over
75−84 years
65−74 years
55−64 years
50−64 years
5−14 years
45−54 years
40−49 years
35−44 years
30−39 years
25−34 years
18−29 years
15−24 years
1−4 years
0−1 year
0e+00
5e+05
1e+06
0
50000
100000
150000
5e+04
1e+05
500
1000
1500
2000
Death counts

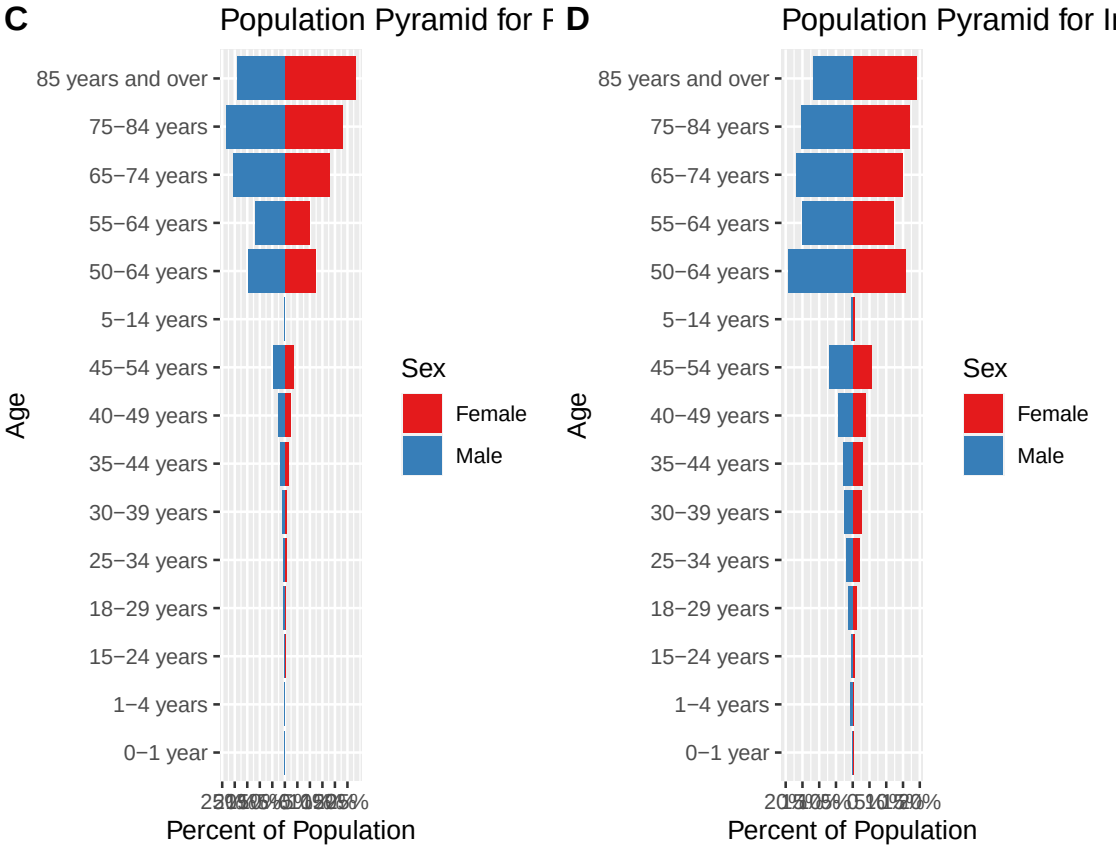
```r
main.age = main %>% filter(State == "United States" & Group == "By Total" & Sex != "All Sexes" & Age.Gr
  group_by(Sex) %>%
  count(COVID.19.Deaths,Pneumonia.Deaths,Influenza.Deaths, Total.Deaths, Age.Group) %>%
  mutate(Total.td = sum(Total.Deaths),
         Total.19 = sum(COVID.19.Deaths),
         Total.Pn = sum(Pneumonia.Deaths),
         Total.In = sum(Influenza.Deaths)) %>%
  mutate(Total.Deaths = ifelse(Sex == "Male", 0 - Total.Deaths, Total.Deaths),
         COVID.19.Deaths = ifelse(Sex == "Male", 0 - COVID.19.Deaths, COVID.19.Deaths),
         Pneumonia.Deaths = ifelse(Sex == "Male", 0 - Pneumonia.Deaths, Pneumonia.Deaths),
         Influenza.Deaths = ifelse(Sex == "Male", 0 - Influenza.Deaths, Influenza.Deaths))
```

```r
tts <- ggplot(main.age, aes(Age.Group, Total.Deaths, fill = Sex)) + geom_col() +
  coord_flip() + scale_fill_brewer(palette = "Set1") +
  scale_y_continuous(breaks = seq(-1000000, 1000000, 250000))+
  labs(x = "Age", y = "Deaths Counts", fill = "Sex", title = "Population Pyramid for total deaths")

cvs <- ggplot(main.age, aes(Age.Group, COVID.19.Deaths, fill = Sex)) + geom_col() +
  coord_flip() + scale_fill_brewer(palette = "Set1") +
  labs(x = "Age", y = "Deaths Counts", fill = "Sex", title = "Population Pyramid for COVID-19 deaths")

pns <- ggplot(main.age, aes(Age.Group, Pneumonia.Deaths, fill = Sex)) + geom_col() +
  coord_flip() + scale_fill_brewer(palette = "Set1") +
  labs(x = "Age", y = "Deaths Counts", fill = "Sex", title = "Population Pyramid for Pneumonia deaths")

Ins <- ggplot(main.age, aes(Age.Group, Influenza.Deaths, fill = Sex)) + geom_col() +
  coord_flip() + scale_fill_brewer(palette = "Set1") +
  labs(x = "Age", y = "Deaths Counts", fill = "Sex", title = "Population Pyramid for Influenza deaths")

ggarrange(tts,cvs,pns,Ins,
                    labels = c("A", "B", "C", "D"),
                    ncol = 2, nrow = 2)
```


A
Population Pyramid for t
B
Population Pyramid for C
C
Population Pyramid for F
D
Population Pyramid for I
85 years and over
75−84 years
65−74 years
55−64 years
50−64 years
5−14 years
45−54 years
40−49 years
35−44 years
30−39 years
25−34 years
18−29 years
15−24 years
1−4 years
0−1 year
Age
Deaths Counts
Sex
Female
Male

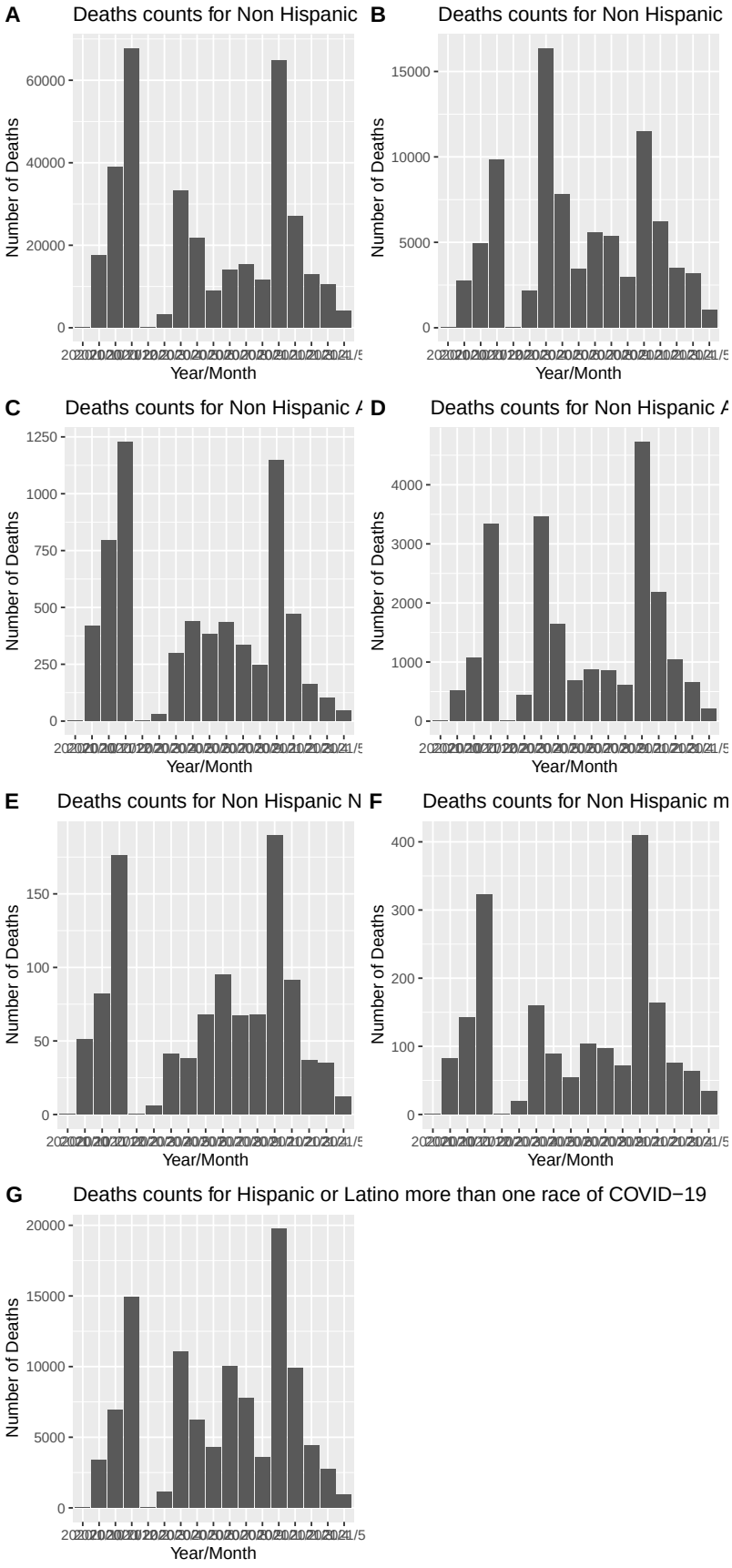
```
tts1 <- ggplot(main.age, aes(Age.Group, Total.Deaths/Total.td, fill = Sex)) + geom_col() +
  coord_flip() + scale_fill_brewer(palette = "Set1") +
  scale_y_continuous(breaks = seq(-.25, .25, .05),
                     labels = c("25%","20%","15%", "10%","5%", "0","5%", "10%","15%","20%", "25%")) +
  labs(x = "Age", y = "Percent of Population", fill = "Sex", title = "Population Pyramid for total deatl

cvs1 <- ggplot(main.age, aes(Age.Group, COVID.19.Deaths/Total.19, fill = Sex)) + geom_col() +
  coord_flip() + scale_fill_brewer(palette = "Set1") +
  scale_y_continuous(breaks = seq(-.25, .25, .05),
                     labels = c("25%","20%","15%", "10%","5%", "0","5%", "10%","15%","20%", "25%")) +
  labs(x = "Age", y = "Percent of Population", fill = "Sex", title = "Population Pyramid for COVID-19 d

pns1 <- ggplot(main.age, aes(Age.Group, Pneumonia.Deaths/Total.Pn, fill = Sex)) + geom_col() +
  coord_flip() + scale_fill_brewer(palette = "Set1") +
   scale_y_continuous(breaks = seq(-.25, .25, .05),
                     labels = c("25%","20%","15%", "10%","5%", "0","5%", "10%","15%","20%", "25%")) +
  labs(x = "Age", y = "Percent of Population", fill = "Sex", title = "Population Pyramid for Pneumonia

Ins1 <- ggplot(main.age, aes(Age.Group, Influenza.Deaths/Total.In, fill = Sex)) + geom_col() +
  coord_flip() + scale_fill_brewer(palette = "Set1") +
   scale_y_continuous(breaks = seq(-.25, .25, .05),
                     labels = c("25%","20%","15%", "10%","5%", "0","5%", "10%","15%","20%", "25%")) +
  labs(x = "Age", y = "Percent of Population", fill = "Sex", title = "Population Pyramid for Influenza

ggarrange(tts1,cvs1,pns1,Ins1,
                     labels = c("A", "B", "C", "D"),
                     ncol = 2, nrow = 2)
```


A
Population Pyramid for t
B
Population Pyramid for C
85 years and over
75–84 years
65–74 years
55–64 years
50–64 years
5–14 years
45–54 years
40–49 years
35–44 years
30–39 years
25–34 years
18–29 years
15–24 years
1–4 years
0–1 year
Age
Sex
Female
Male
Percent of Population

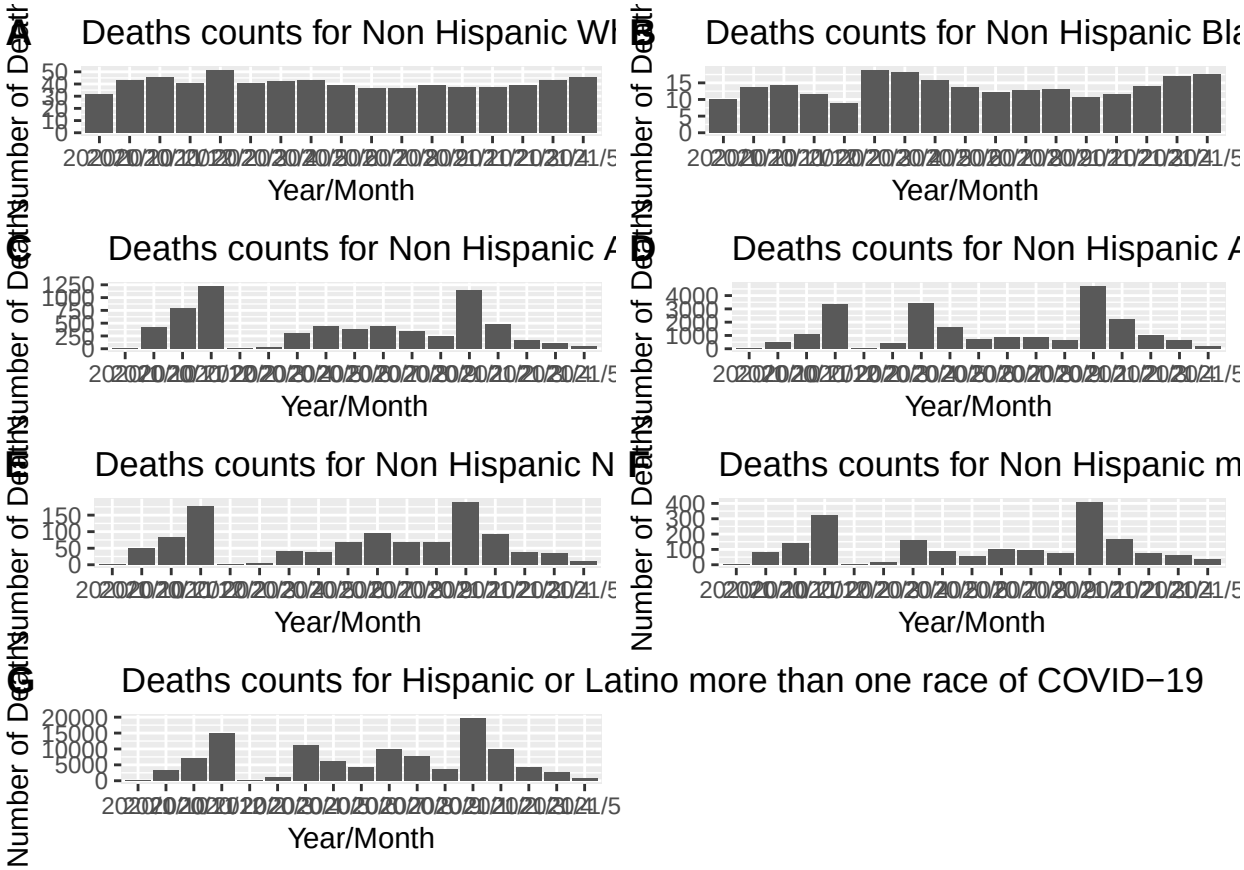

C
Population Pyramid for F
D
Population Pyramid for I
85 years and over
75–84 years
65–74 years
55–64 years
50–64 years
5–14 years
45–54 years
40–49 years
35–44 years
30–39 years
25–34 years
18–29 years
15–24 years
1–4 years
0–1 year
Age
Sex
Female
Male
Percent of Population

```r
# Race
race.count = race %>% filter(Group == "By Month" & State == "United States" & Indicator == "Count of CO
                      select(-Data.as.of, -Start.Date, -End.Date) %>%
                      mutate(Month.Year = paste(Year, Month, sep = "/")) %>%
                      arrange(Year, Month)


race.dist = race %>% filter(Group == "By Month" & State == "United States" & Indicator == "Weighted dis
                      select(-Data.as.of, -Start.Date, -End.Date) %>%
                      mutate(Month.Year = paste(Year, Month, sep = "/")) %>%
                      arrange(Year, Month)
```

```r
nhw <- ggplot(data = race.count, aes(Month.Year, Non.Hispanic.White)) +
  labs(x = "Year/Month", y = "Number of Deaths", title = "Deaths counts for Non Hispanic White of COVID-
  geom_col()

nhb <- ggplot(data = race.count, aes(Month.Year, Non.Hispanic.Black.or.African.American)) +
  labs(x = "Year/Month", y = "Number of Deaths", title = "Deaths counts for Non Hispanic Black or Africa
  geom_col()

nhaa <- ggplot(data = race.count, aes(Month.Year, Non.Hispanic.American.Indian.or.Alaska.Native)) +
  labs(x = "Year/Month", y = "Number of Deaths", title = "Deaths counts for Non Hispanic American Indian
  geom_col()

nha <- ggplot(data = race.count, aes(Month.Year, Non.Hispanic.Asian)) +
  labs(x = "Year/Month", y = "Number of Deaths", title = "Deaths counts for Non Hispanic Asian of COVID-
  geom_col()

nhn <- ggplot(data = race.count, aes(Month.Year, Non.Hispanic.Native.Hawaiian.or.Other.Pacific.Islander)
  labs(x = "Year/Month", y = "Number of Deaths", title = "Deaths counts for Non Hispanic Native Hawaiian
  geom_col()

nhm <- ggplot(data = race.count, aes(Month.Year, Non.Hispanic.more.than.one.race)) +
  labs(x = "Year/Month", y = "Number of Deaths", title = "Deaths counts for Non Hispanic more than one r
  geom_col()

hl <- ggplot(data = race.count, aes(Month.Year, Hispanic.or.Latino)) +
  labs(x = "Year/Month", y = "Number of Deaths", title = "Deaths counts for Hispanic or Latino more than
  geom_col()

ggarrange(nhw,nhb,nhaa,nha,nhn,nhm, hl,
                    labels = c("A", "B", "C", "D", "E", "F", "G"),
                    ncol = 2, nrow = 4)
```


A
Deaths counts for Non Hispanic
Number of Deaths
Year/Month
B
Deaths counts for Non Hispanic
Number of Deaths
Year/Month
C
Deaths counts for Non Hispanic A
Number of Deaths
Year/Month
D
Deaths counts for Non Hispanic A
Number of Deaths
Year/Month
E
Deaths counts for Non Hispanic N
Number of Deaths
Year/Month
F
Deaths counts for Non Hispanic m
Number of Deaths
Year/Month
G
Deaths counts for Hispanic or Latino more than one race of COVID−19
Number of Deaths
Year/Month

```r
nhw <- ggplot(data = race.dist, aes(Month.Year, Non.Hispanic.White)) +
  labs(x = "Year/Month", y = "Number of Deaths", title = "Deaths counts for Non Hispanic White of COVID-
  geom_col()

nhb <- ggplot(data = race.dist, aes(Month.Year, Non.Hispanic.Black.or.African.American)) +
  labs(x = "Year/Month", y = "Number of Deaths", title = "Deaths counts for Non Hispanic Black or Africa
  geom_col()

nhaa <- ggplot(data = race.count, aes(Month.Year, Non.Hispanic.American.Indian.or.Alaska.Native)) +
  labs(x = "Year/Month", y = "Number of Deaths", title = "Deaths counts for Non Hispanic American Indian
  geom_col()

nha <- ggplot(data = race.count, aes(Month.Year, Non.Hispanic.Asian)) +
  labs(x = "Year/Month", y = "Number of Deaths", title = "Deaths counts for Non Hispanic Asian of COVID-
  geom_col()

nhn <- ggplot(data = race.count, aes(Month.Year, Non.Hispanic.Native.Hawaiian.or.Other.Pacific.Islander)
  labs(x = "Year/Month", y = "Number of Deaths", title = "Deaths counts for Non Hispanic Native Hawaiian
  geom_col()

nhm <- ggplot(data = race.count, aes(Month.Year, Non.Hispanic.more.than.one.race)) +
  labs(x = "Year/Month", y = "Number of Deaths", title = "Deaths counts for Non Hispanic more than one r
  geom_col()

hl <- ggplot(data = race.count, aes(Month.Year, Hispanic.or.Latino)) +
  labs(x = "Year/Month", y = "Number of Deaths", title = "Deaths counts for Hispanic or Latino more than
  geom_col()

ggarrange(nhw,nhb,nhaa,nha,nhn,nhm, hl,
                    labels = c("A", "B", "C", "D", "E", "F", "G"),
                    ncol = 2, nrow = 4)
```


A
Deaths counts for Non Hispanic Wh
Number of Deaths
Year/Month
B
Deaths counts for Non Hispanic Bla
Number of Deaths
Year/Month
C
Deaths counts for Non Hispanic A
Number of Deaths
Year/Month
D
Deaths counts for Non Hispanic A
Number of Deaths
Year/Month
E
Deaths counts for Non Hispanic N
Number of Deaths
Year/Month
F
Deaths counts for Non Hispanic m
Number of Deaths
Year/Month
G
Deaths counts for Hispanic or Latino more than one race of COVID−19
Number of Deaths
Year/Month