

final proj

Yudong Wang

5/13/2021

```
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
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
```

```
setwd("C:/Users/wyd98/Desktop/homework/STA 160")
```

```
mydata = read.csv("Conditions_Contributing_to_COVID-19_Deaths__by_State_and_Age.csv")
```

```
coviddata = mydata %>%
  filter(Condition == "COVID-19") %>%
  filter(Group == "By Total") %>%
  filter(State != "United States") %>%
  filter(State != "New York City") %>%
  filter(Age.Group != "All Ages") %>%
  filter(Age.Group != "Not stated")
```

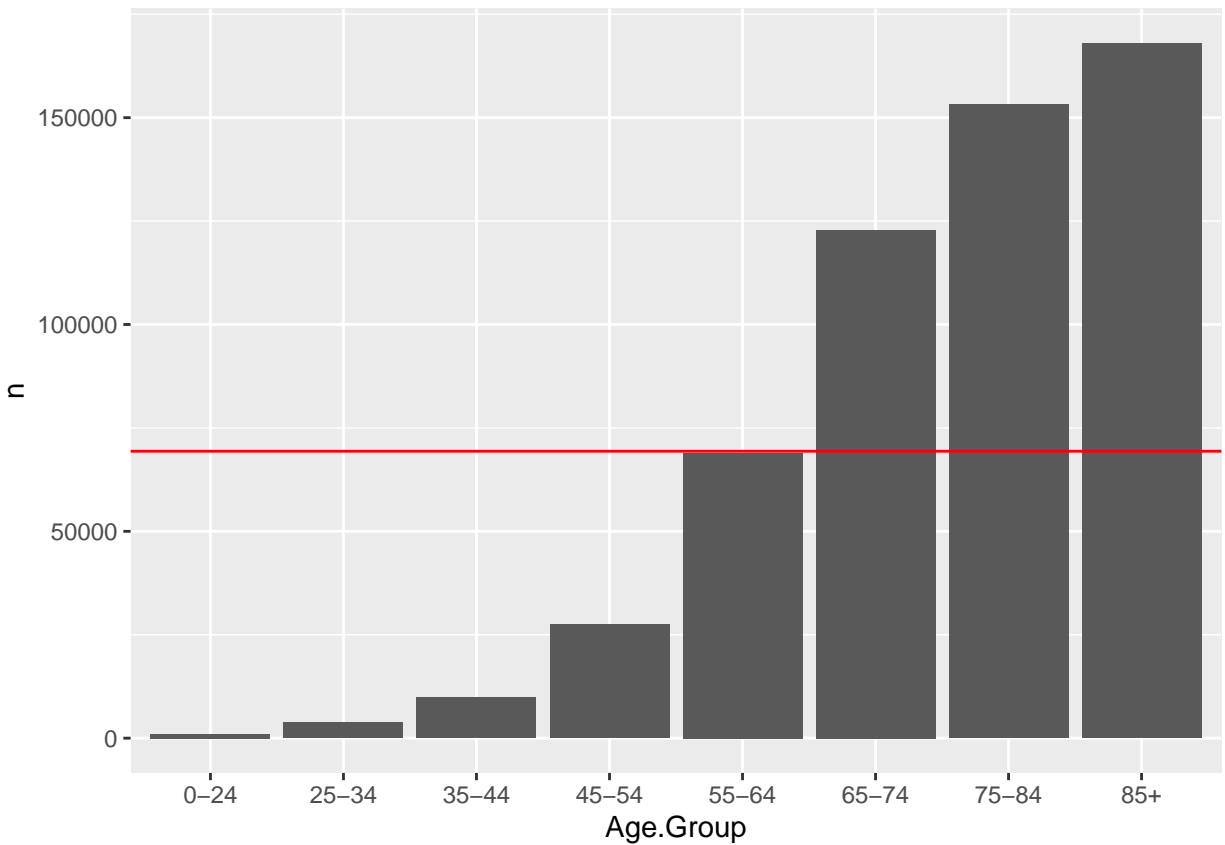
```
mydata = mydata %>%
  filter(Group != "By Total") %>%
  filter(Group != "By Year") %>%
  filter(State != "United States") %>%
  filter(State != "New York City") %>%
  filter(Age.Group != "All Ages") %>%
  filter(Age.Group != "Not stated")
```

```
# age group vs. death
```

```
age = coviddata %>%
  group_by(Age.Group) %>%
  summarize(n = sum(COVID.19.Deaths, na.rm = TRUE)) %>%
  head(8)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

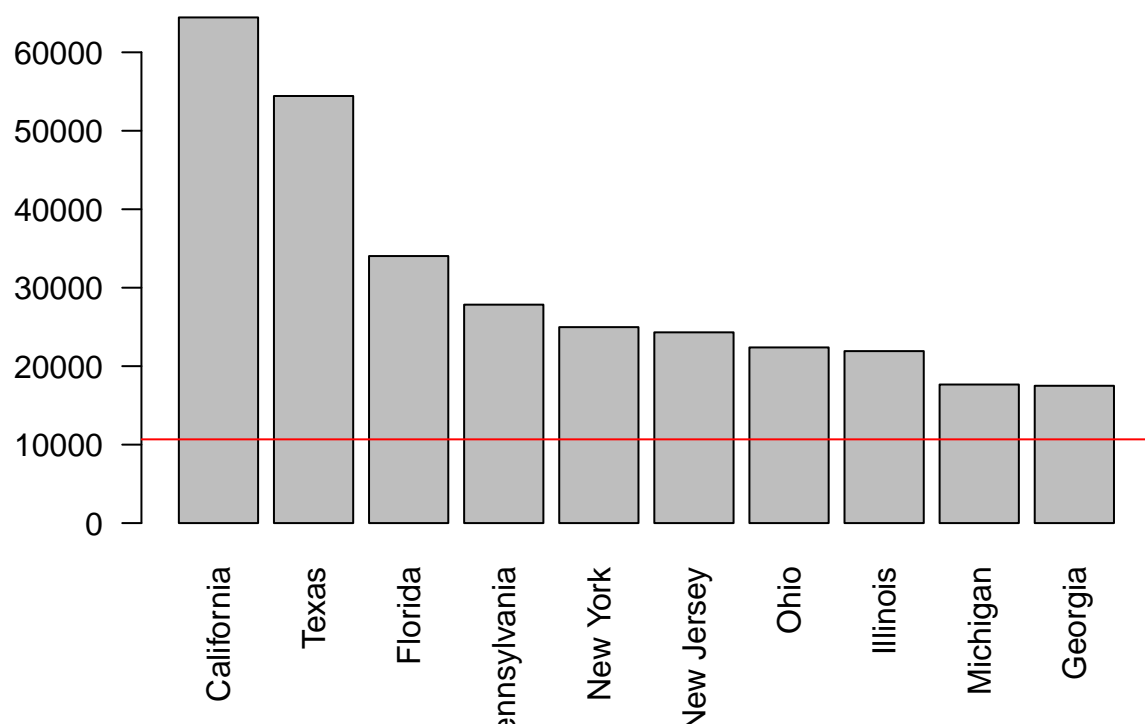
```
ggplot(age, aes(x = Age.Group, y = n)) + geom_bar(stat = "identity") + geom_hline(yintercept = mean(age
```



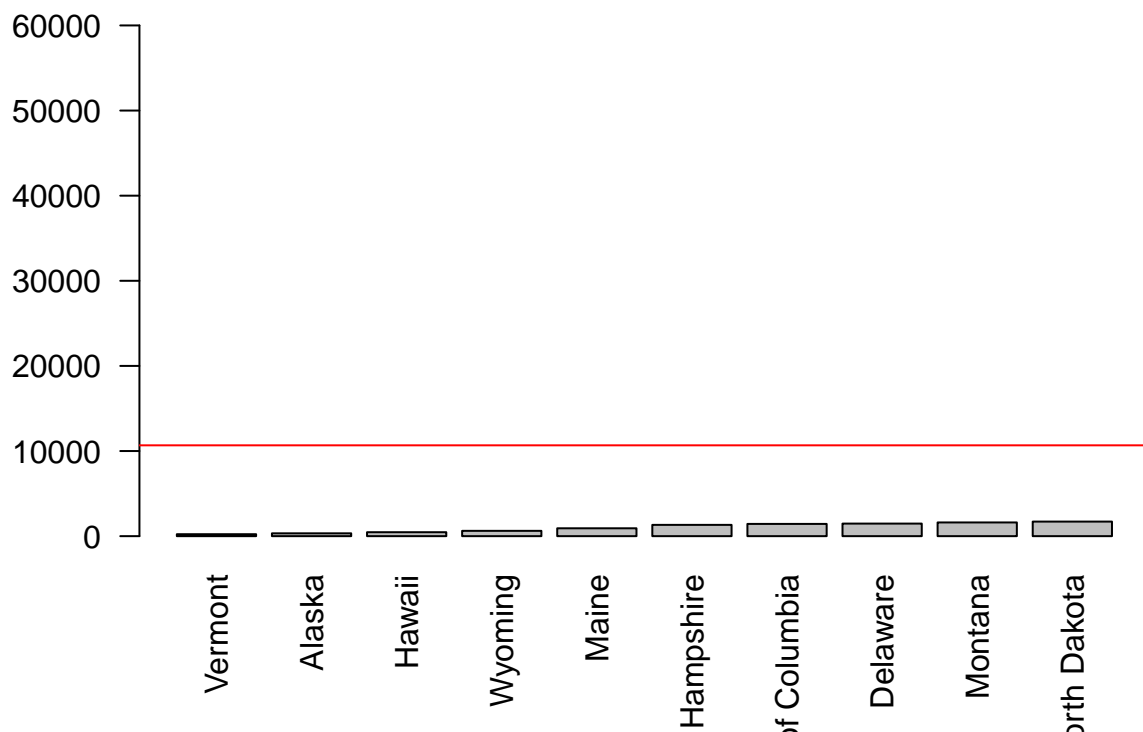
```
# 10 states with the most death cases
state = coviddata %>%
  group_by(State) %>%
  summarize(n = sum(COVID.19.Deaths, na.rm = TRUE)) %>%
  arrange(desc(n)) %>%
  filter(State != "United States")

## `summarise()` ungrouping output (override with `.groups` argument)

state10 = state %>%
  head(10)
# 10 states with the least death cases
statelow10 = state %>%
  arrange(n) %>%
  head(10)
# graph of most 10
barplot(state10$n, names.arg = state10$State, las = 2)
abline(h = mean(state$n), col = "red")
```



```
# graph of least 10
barplot(state_low10$n, names.arg = state_low10$State, ylim = c(0, 60000), las = 2)
abline(h = mean(state$n), col = "red")
```



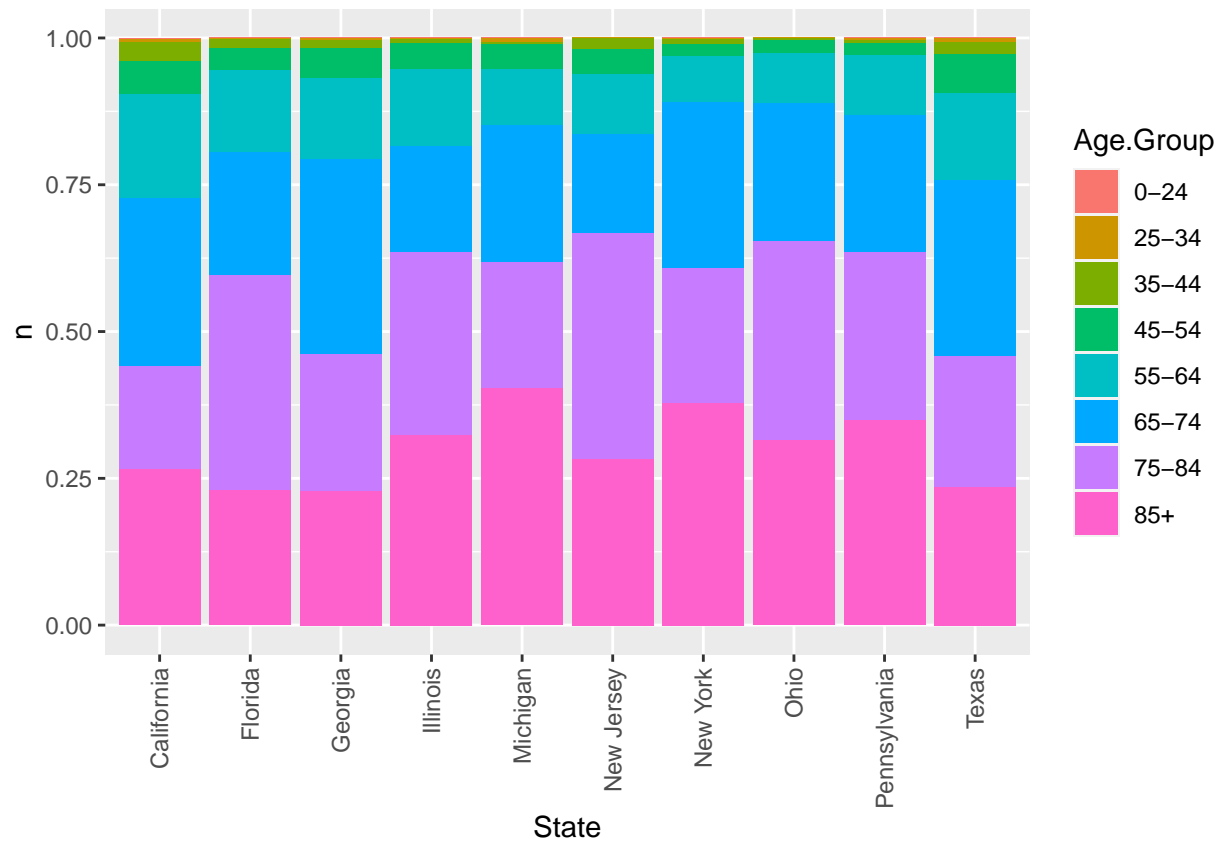
```
# get 10 states from original data
state10data = mydata %>%
  filter(State == state10$State) %>%
  filter(Age.Group != "Not stated") %>%
  group_by(Age.Group, State) %>%
  summarize(n = sum(COVID.19.Deaths, na.rm = TRUE))
```

```
## Warning in State == state10$State: longer object length is not a multiple of
## shorter object length
```

```
## `summarise()` regrouping output by 'Age.Group' (override with ` .groups ` argument)
```

```
# bar plot showing proportion of death number
```

```
ggplot(state10data, aes(x = State, y = n)) + geom_bar(position = "fill", stat = "identity", aes(fill = ...))
```

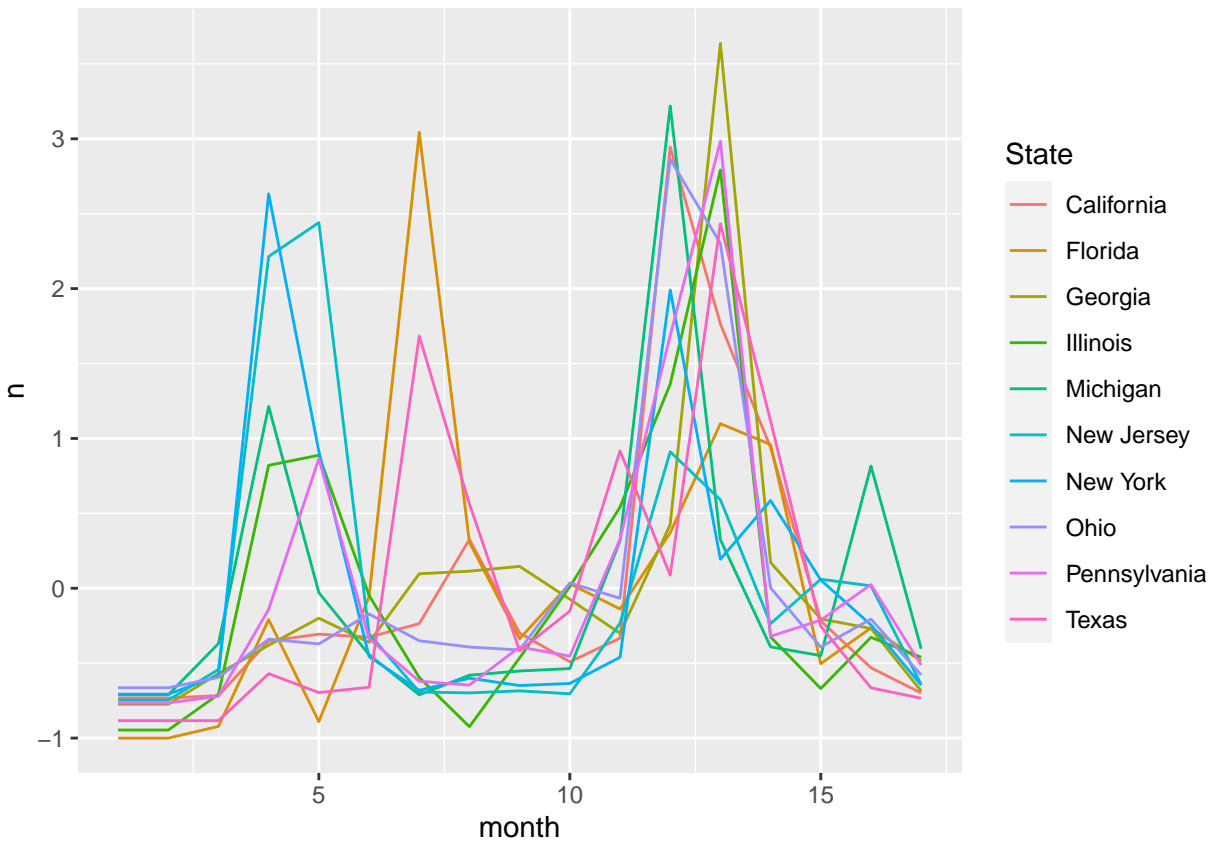


```
# get monthly death in each state
statetrend = mydata %>%
  filter(State == state10$State) %>%
  group_by(End.Date, State) %>%
  summarise(n = sum(COVID.19.Deaths, na.rm = TRUE)) %>%
  arrange(as.Date(End.Date, format = "%m/%d/%Y"))

## Warning in State == state10$State: longer object length is not a multiple of
## shorter object length

## `summarise()` regrouping output by 'End.Date' (override with `.groups` argument)

month = rep(1:17, each = 10)
trend = data.frame(statetrend, month) %>%
  group_by(State) %>%
  mutate(n = (n - mean(n))/sd(n))
ggplot(trend, aes(month, n)) + geom_line(aes(col = State))
```

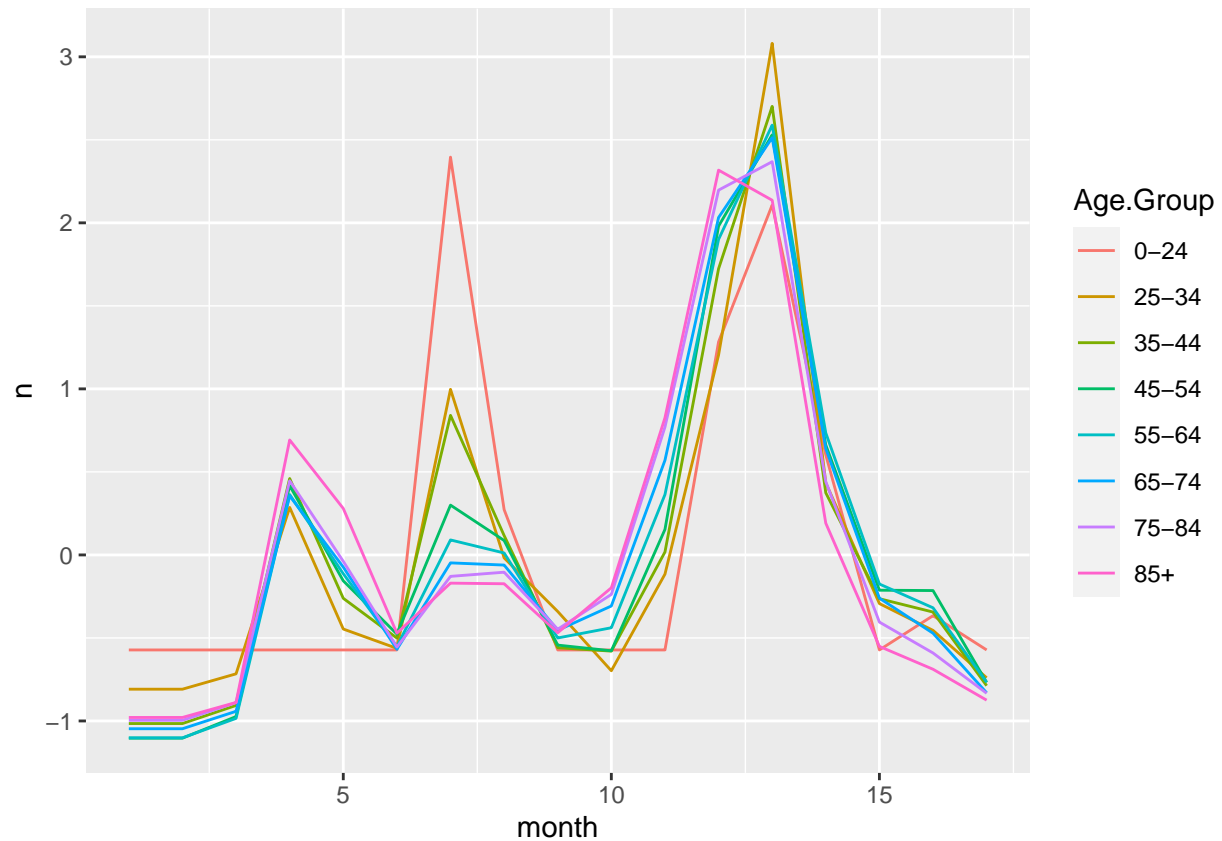


```

agetrend = mydata %>%
  group_by(Age.Group, End.Date) %>%
  summarise(n = sum(COVID.19.Deaths, na.rm = TRUE)) %>%
  arrange(as.Date(End.Date, format = "%m/%d/%Y"))

## `summarise()` regrouping output by 'Age.Group' (override with `.groups` argument)
month = rep(1:17, each = 8)
agetrend = data.frame(agetrend, month) %>%
  group_by(Age.Group) %>%
  mutate(n = (n - mean(n))/sd(n))
ggplot(agetrend, aes(month, n)) + geom_line(aes(col = Age.Group))

```



```
# get monthly death caused by each condition
```

```
cond = mydata %>%
  group_by(End.Date, Condition.Group) %>%
  summarise(n = sum(COVID.19.Deaths, na.rm = TRUE)) %>%
  arrange(as.Date(End.Date, format = "%m/%d/%Y")) %>%
  head(-12)
```

```
## `summarise()` regrouping output by 'End.Date' (override with `.groups` argument)
```

```
month = rep(1:16, each = 12)
cond = data.frame(cond, month) %>%
  group_by(Condition.Group) %>%
  mutate(n = (n - mean(n))/sd(n))
ggplot(cond, aes(month, n)) + geom_line(aes(col = Condition.Group)) + theme(legend.text = element_text(
  legend.title = element_text(size = 5))
```

