

Covid19_data_Statistics_visualization.R

corn

2022-01-20

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
## filter, lag  
  
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
setwd("C:/Users/dongj/Desktop/Covid_Cal/Statewide_case_statistics_and_demographics")
```

```
data <- read.csv("./dataset/covid19cases_test_012022.csv")  
data_v1 <- read.csv("./dataset/covid19cases_test_012022.csv")
```

```
data_v1 <- data_v1[data_v1$area_type == 'County',]
```

```
data_v1 <- filter(data_v1,  
                  !area %in% c('Unknown', 'Out of state'))
```

```
unique(data_v1$area)
```

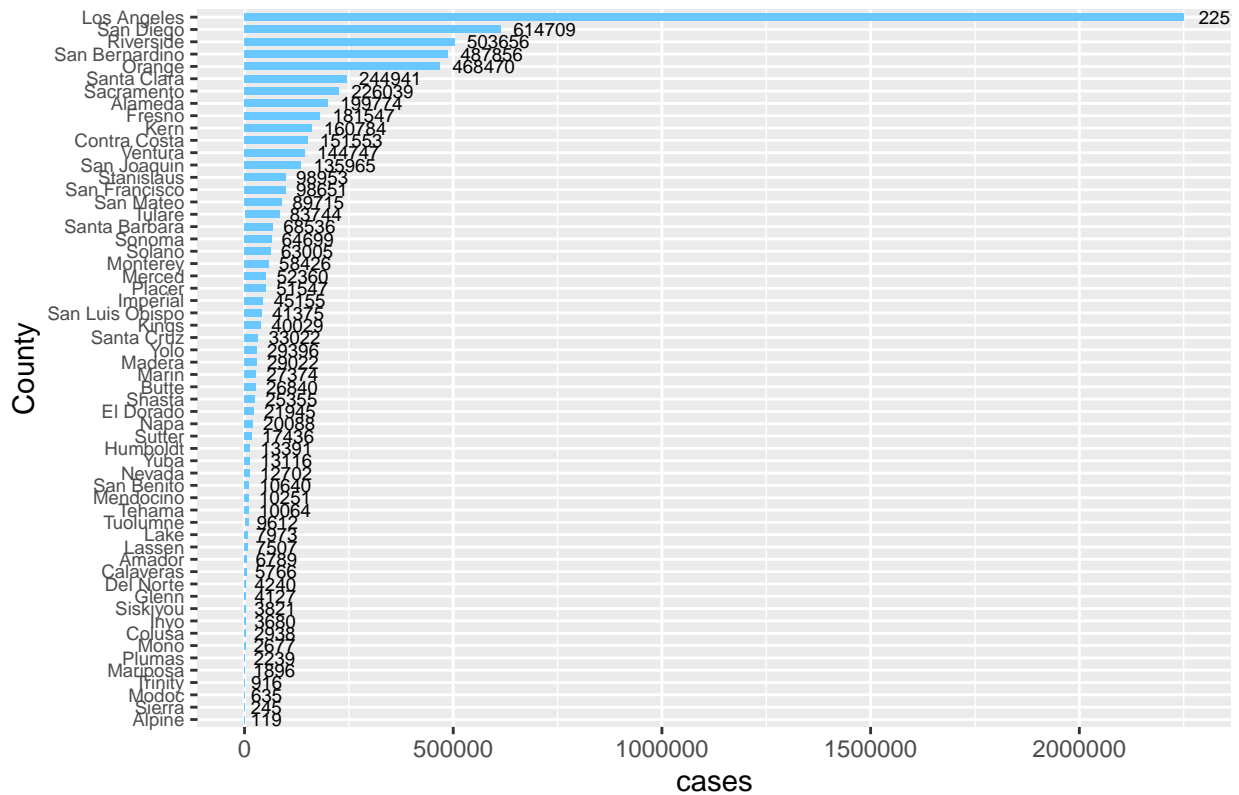
```
## [1] "Alameda"      "Alpine"      "Amador"      "Butte"  
## [5] "Calaveras"    "Colusa"      "Contra Costa" "Del Norte"  
## [9] "El Dorado"    "Fresno"      "Glenn"       "Humboldt"  
## [13] "Imperial"     "Inyo"        "Kern"        "Kings"  
## [17] "Lake"         "Lassen"      "Los Angeles" "Madera"  
## [21] "Marin"        "Mariposa"    "Mendocino"   "Merced"  
## [25] "Modoc"        "Mono"        "Monterey"    "Napa"  
## [29] "Nevada"       "Orange"      "Placer"      "Plumas"  
## [33] "Riverside"    "Sacramento"  "San Benito"  "San Bernardino"  
## [37] "San Diego"    "San Francisco" "San Joaquin" "San Luis Obispo"  
## [41] "San Mateo"    "Santa Barbara" "Santa Clara" "Santa Cruz"  
## [45] "Shasta"       "Sierra"      "Siskiyou"    "Solano"  
## [49] "Sonoma"       "Stanislaus"  "Sutter"      "Tehama"  
## [53] "Trinity"      "Tulare"      "Tuolumne"    "Ventura"  
## [57] "Yolo"         "Yuba"
```

```

#Total Cases Group by Area-----
total_cases_by_area <- aggregate(cases ~ area,
                                data_v1,
                                sum)
total_cases_by_area_graph <- ggplot(data=total_cases_by_area,
                                aes(x = cases,
                                    y = reorder(area,
                                                  cases,
                                                  sum)))
                                )+
  geom_bar(stat = "identity",
           width=.6,
           position = position_dodge(width = 0.5),
           fill = "#69c8ff")+
  labs(
    title="Total Cases by County",
    y = "County"
  )+
  theme(axis.text.y = element_text(size = 7)) +
  geom_text(aes(label=cases),
            hjust = -0.2,
            size = 2.5,
            position = position_dodge(width = 1),
            inherit.aes = TRUE)
options(repr.plot.width = 14, repr.plot.height = 8)
total_cases_by_area_graph

```

Total Cases by County



```
#Total Death Group by Area-----
total_deaths_by_area <- aggregate(deaths ~ area,
                                   data_v1,
                                   sum)

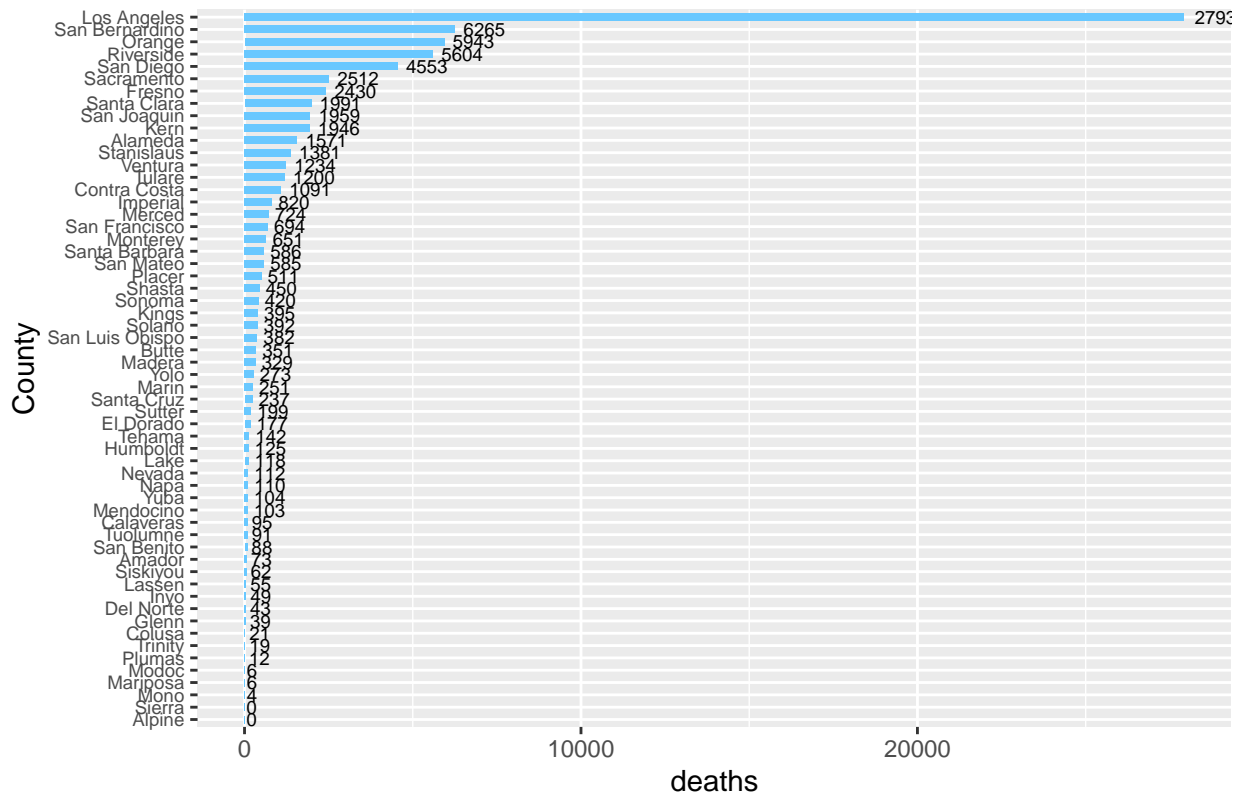
total_deaths_by_area_graph <- ggplot(data=total_deaths_by_area,
                                     aes(x = deaths,
                                          y = reorder(area,
                                                         deaths,
                                                         sum)))

  )+
  geom_bar(stat = "identity",
           width=.6,
           position = position_dodge(width = 0.5),
           fill = "#69c8ff")+

  labs(
    title="Total Deaths by County",
    y = "County"
  )+
  theme(axis.text.y = element_text(size = 7)) +
  geom_text(aes(label=deaths),
            hjust = -0.2,
            size = 2.5,
            position = position_dodge(width = 1),
            inherit.aes = TRUE)

options(repr.plot.width = 14, repr.plot.height = 8)
total_deaths_by_area_graph
```

Total Deaths by County



```
#Recent Date-----
RecentDate_end = data_v1$date[length(data_v1$date)-1]
RecentDate_start = data_v1$date[length(data_v1$date)-8]

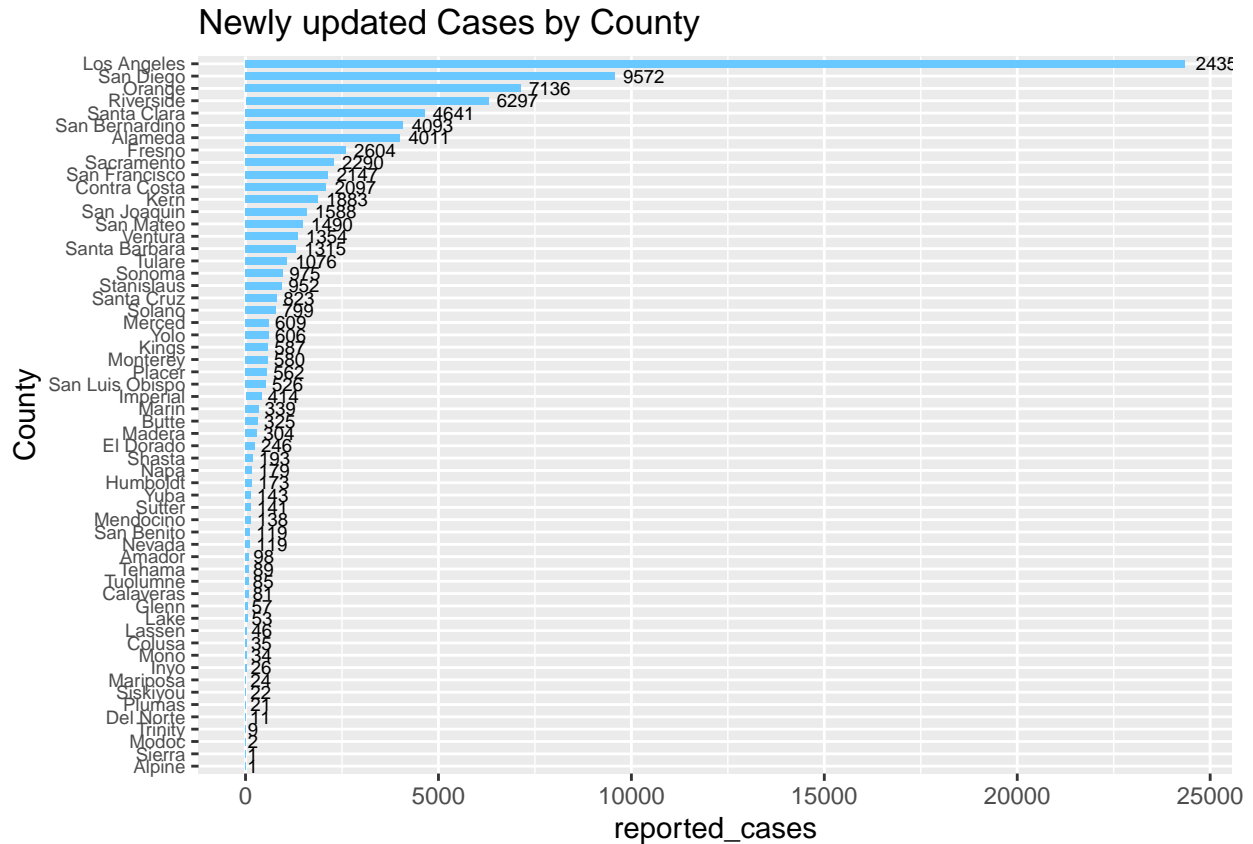
recentdate_data_v1 <- data_v1[data_v1$date == RecentDate_end,]

#Newly Updated Cases-----
newly_updated_cases_by_area <- aggregate(reported_cases ~ area,
                                          recentdate_data_v1,
                                          sum)
newly_updated_cases_by_area_graph <- ggplot(newly_updated_cases_by_area,
      aes(x = reported_cases,
          y = reorder(area,
                      reported_cases,
                      sum)))
  )+
  geom_bar(stat = "identity",
           width=.6,
           position = position_dodge(width = 0.5),
           fill = "#69c8ff")+
  labs(
    title="Newly updated Cases by County",
    y = "County"
  )+
  theme(axis.text.y = element_text(size = 7)) +
  geom_text(aes(label=reported_cases),
```

```

hjust = -0.2,
size = 2.5,
position = position_dodge(width = 1),
inherit.aes = TRUE)
options(repr.plot.width = 14, repr.plot.height = 8)
newly_updated_cases_by_area_graph

```



```

#Newly Updated Deaths-----
newly_updated_deaths_by_area <- aggregate(reported_deaths ~ area,
                                           recentdate_data_v1,
                                           sum)
newly_updated_deaths_by_area_graph <- ggplot(data=newly_updated_deaths_by_area,
                                              aes(x = reported_deaths,
                                                  y = reorder(area,
                                                                reported_deaths,
                                                                sum)))
+
geom_bar(stat = "identity",
         width=.6,
         position = position_dodge(width = 0.5),
         fill = "#69c8ff")+
labs(
  title="Newly Updated Deaths by County",
  y = "County"
)+

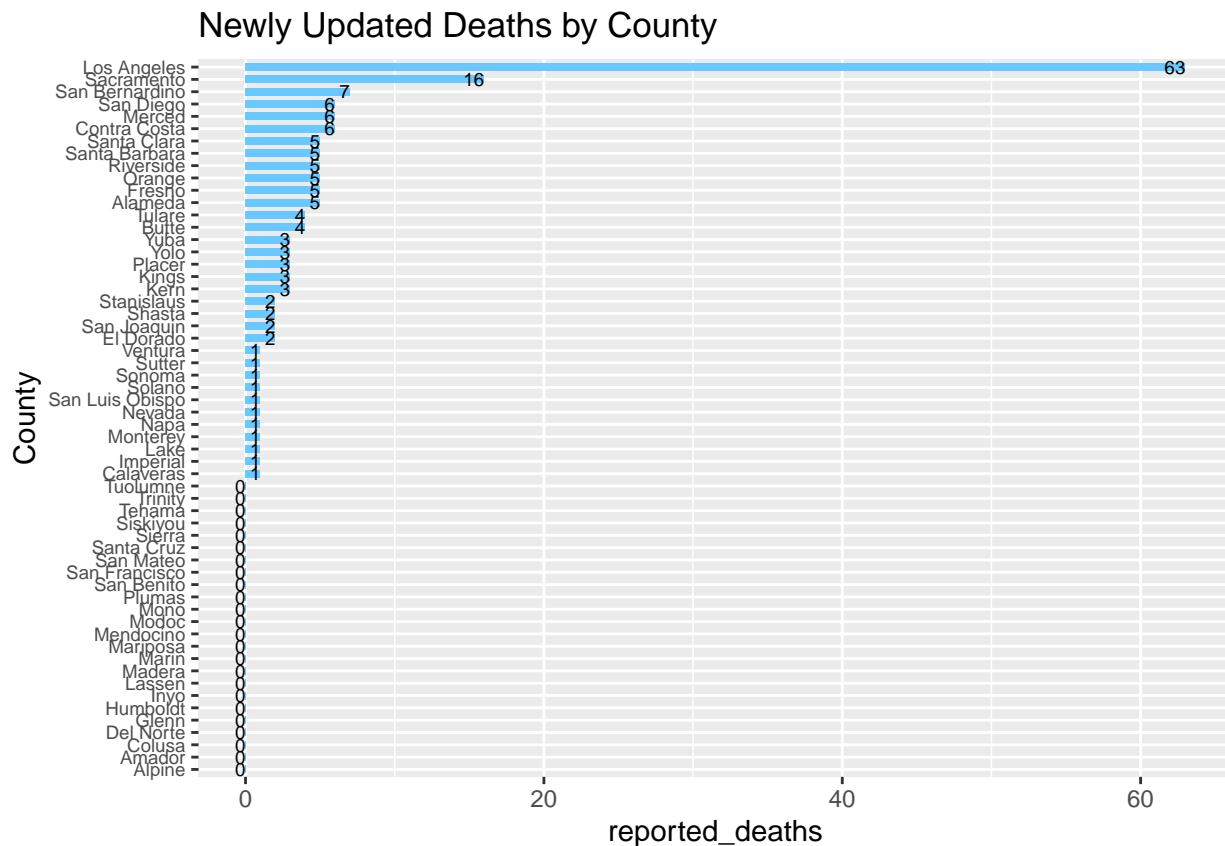
```

```

theme(axis.text.y = element_text(size = 7)) +
geom_text(aes(label=reported_deaths),
          hjust = 1,
          size = 2.5,
          position = position_dodge(width = 1),
          inherit.aes = TRUE)

options(repr.plot.width = 14, repr.plot.height = 8)
newly_updated_deaths_by_area_graph

```



```

newly_population_by_area <- aggregate(population ~ area,recentdate_data_v1,max)
#c(total_cases_by_area$area)
#c(total_cases_by_area$cases)
#c(total_deaths_by_area$deaths)
#c(newly_population_by_area$population)
#c((total_cases_by_area$cases/newly_population_by_area$population)* 100000)

area <- c(total_cases_by_area$area)
cumulative_cases_per_100k <- c(round((total_cases_by_area$cases/newly_population_by_area$population)* 100000))
cumulative_deaths_per_100k <- c(round((total_deaths_by_area$deaths/newly_population_by_area$population)* 100000))

cumulative_per_100k <- data.frame(area,
                                cumulative_cases_per_100k,
                                cumulative_deaths_per_100k)

```

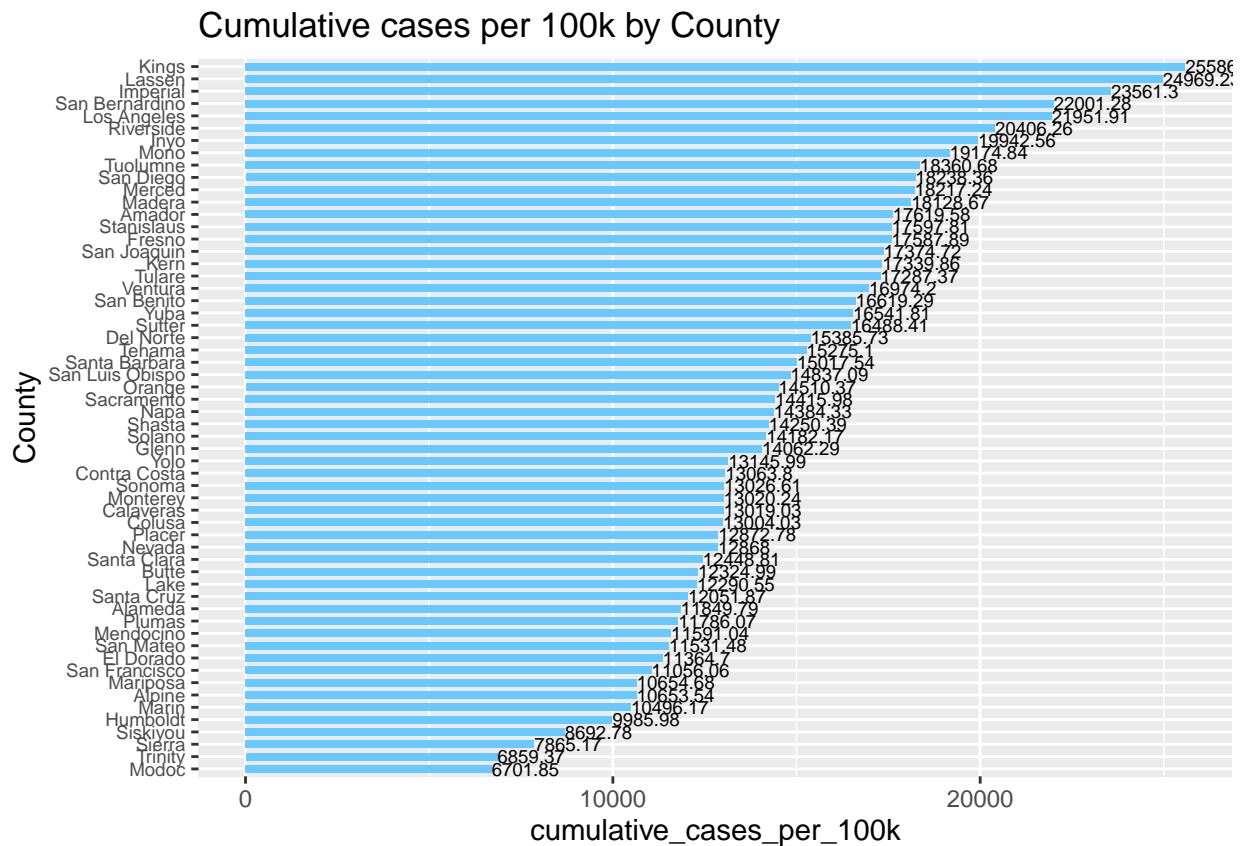
```

#Cumulative Cases per 100k -----
cumulative_cases_per_100k_graph <- ggplot(data=cumulative_per_100k,
      aes(x = cumulative_cases_per_100k,
          y = reorder(area,
                      cumulative_cases_per_100k))
    )+
  geom_bar(stat = "identity",
    width=.6,
    position = position_dodge(width = 0.5),
    fill = "#69c8ff")+

  labs(
    title="Cumulative cases per 100k by County",
    y = "County"
  )+
  theme(axis.text.y = element_text(size = 7)) +
  geom_text(aes(label=cumulative_cases_per_100k),
    hjust = 0,
    size = 2.5,
    position = position_dodge(width = 1),
    inherit.aes = TRUE)

options(repr.plot.width = 14, repr.plot.height = 8)
cumulative_cases_per_100k_graph

```



```

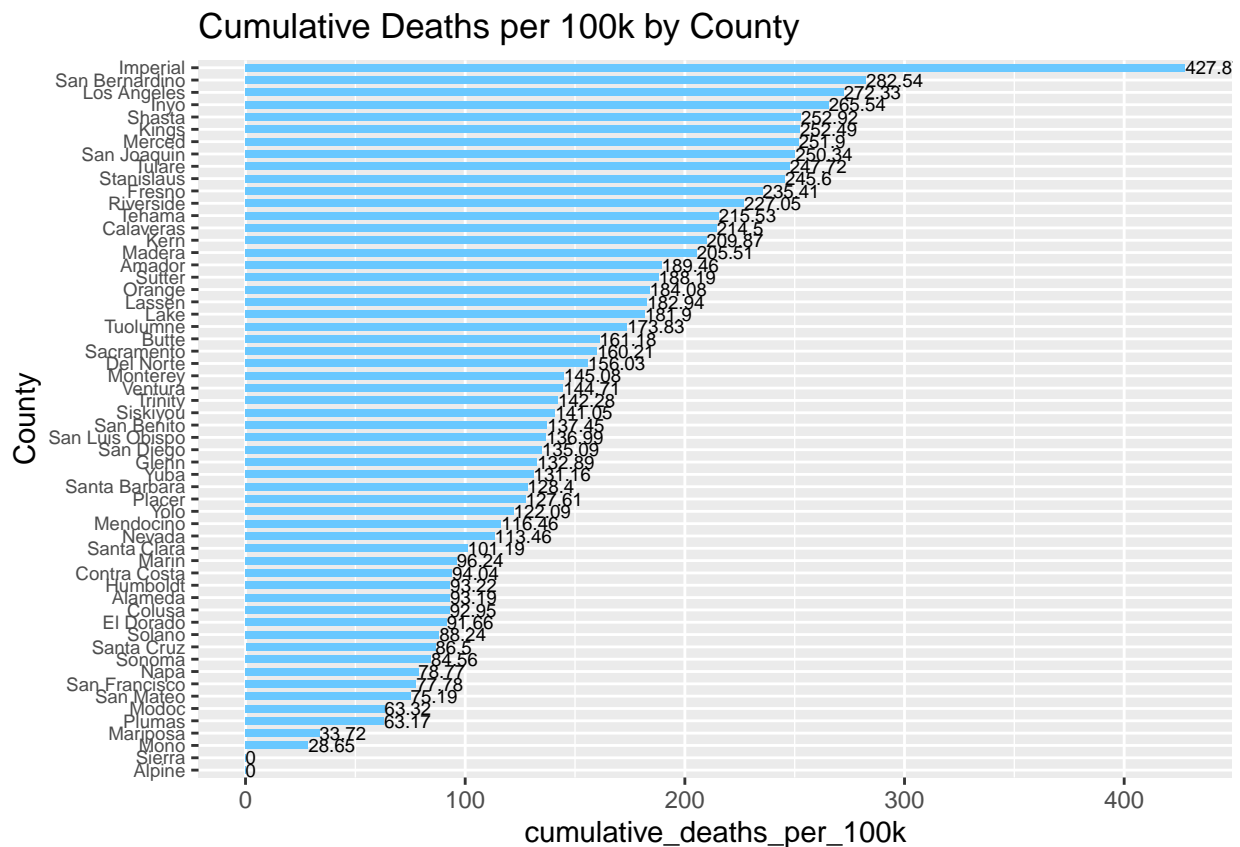
#Cumulative Deaths per 100k -----
cumulative_deaths_per_100k_graph <- ggplot(data=cumulative_per_100k,

```

```

aes(x = cumulative_deaths_per_100k,
    y = reorder(area,
                 cumulative_deaths_per_100k))
)+
geom_bar(stat = "identity",
         width=.6,
         position = position_dodge(width = 0.5),
         fill = "#69c8ff")+
labs(
  title = "Cumulative Deaths per 100k by County",
  y = "County"
)+
theme(axis.text.y = element_text(size = 7)) +
geom_text(aes(label=cumulative_deaths_per_100k),
          hjust = 0,
          size = 2.5,
          position = position_dodge(width = 1),
          inherit.aes = TRUE)
options(repr.plot.width = 14, repr.plot.height = 8)
cumulative_deaths_per_100k_graph

```



```

#Total by 7 days-----
recentdate_data_v2 <- data_v1[data_v1$date >=RecentDate_start & data_v1$date<=RecentDate_end,]
head(recentdate_data_v2,5)

```



```
##           date      area area_type population cases cumulative_cases deaths
## 712 2022-01-12 Alameda County   1685886  4214          189440      0
## 713 2022-01-13 Alameda County   1685886  3700          193140      0
## 714 2022-01-14 Alameda County   1685886  2755          195895      1
## 715 2022-01-15 Alameda County   1685886  1400          197295      0
## 716 2022-01-16 Alameda County   1685886  1028          198323      1
##      cumulative_deaths total_tests cumulative_total_tests positive_tests
## 712              1566      28557          5220490          5436
## 713              1566      27158          5247648          4744
## 714              1567      21072          5268720          3606
## 715              1567       9331          5278051          1774
## 716              1568       6615          5284666          1303
##      cumulative_positive_tests reported_cases cumulative_reported_cases
## 712              230109          3820          173001
## 713              234853          4767          177768
## 714              238459           0          177768
## 715              240233           0          177768
## 716              241536         12923          190691
##      reported_deaths cumulative_reported_deaths reported_tests
## 712              2          1563          22954
## 713             -2          1561          25440
## 714              0          1561           NA
## 715              0          1561           NA
## 716              3          1564          89440
```

```
recentdate_data_v3 <- aggregate(cases ~ date, recentdate_data_v2, sum)
recentdate_data_v4 <- aggregate(deaths ~ date, recentdate_data_v2, sum)

date <- c(recentdate_data_v3$date)
Total_Cases_7days <- c(recentdate_data_v3$cases)
Total_Deaths_7days <- c(recentdate_data_v4$deaths)

covid_dataset_newly_dataset <- data.frame(date,
                                           Total_Cases_7days,
                                           Total_Deaths_7days)
covid_dataset_newly_dataset
```

```
##           date Total_Cases_7days Total_Deaths_7days
## 1 2022-01-12          106329          49
## 2 2022-01-13           85779          45
## 3 2022-01-14           69760          32
## 4 2022-01-15           34046          29
## 5 2022-01-16           19049          30
## 6 2022-01-17           25077          16
## 7 2022-01-18            4526           0
## 8 2022-01-19              0           0
```

```
length(c(covid_dataset_newly_dataset$date))
```

```
## [1] 8
```

```

cases_7day <- c()
deaths_7day <- c()
for (i in 1:length(c(covid_dataset_newly_dataset$date))){
  if (i == 1){
    cases_7day <- append(cases_7day, covid_dataset_newly_dataset$Total_Cases_7days[1])
    deaths_7day <- append(deaths_7day, covid_dataset_newly_dataset$Total_Deaths_7days[1])
  }else{
    cases_window = covid_dataset_newly_dataset$Total_Cases_7days[1:i]
    deaths_window = covid_dataset_newly_dataset$Total_Deaths_7days[1:i]
    cases_window_sum= sum(cases_window)
    deaths_window_sum= sum(deaths_window)
    cases_7day <- append(cases_7day, cases_window_sum)
    deaths_7day <- append(deaths_7day, deaths_window_sum)
  }
}

covid_dataset_newly_dataset$Total_Cases_7days_sum <- cases_7day
covid_dataset_newly_dataset$Total_Deaths_7days_sum <- deaths_7day
covid_dataset_newly_dataset

```

```

##      date Total_Cases_7days Total_Deaths_7days Total_Cases_7days_sum
## 1 2022-01-12      106329           49      106329
## 2 2022-01-13      85779           45      192108
## 3 2022-01-14      69760           32      261868
## 4 2022-01-15      34046           29      295914
## 5 2022-01-16      19049           30      314963
## 6 2022-01-17      25077           16      340040
## 7 2022-01-18       4526            0      344566
## 8 2022-01-19         0            0      344566
##      Total_Deaths_7days_sum
## 1           49
## 2           94
## 3          126
## 4          155
## 5          185
## 6          201
## 7          201
## 8          201

```

```

cases_7day_avg <- c()
deaths_7day_avg <- c()
for (i in 1:length(c(covid_dataset_newly_dataset$date))){
  cases_window_avg = covid_dataset_newly_dataset$Total_Cases_7days_sum[i] / (i)
  deaths_window_avg = covid_dataset_newly_dataset$Total_Deaths_7days_sum[i] / (i)
  cases_7day_avg <- append(cases_7day_avg, cases_window_avg)
  deaths_7day_avg <- append(deaths_7day_avg, deaths_window_avg)
}

```

```

Cases_per_100k_7_day_average = round((sum(cases_7day_avg)/7) / sum(newly_population_by_area$population))
Deaths_per_100k_7_day_average = round((sum(deaths_7day_avg)/7) / sum(newly_population_by_area$population))

Cases_per_100k_7_day_average

```

```
## [1] 204.91
```

```
recentdate_data_v5 = subset(recentdate_data_v2, select = -c(area_type, cumulative_cases, cumulative_deaths,
                                                             cumulative_total_tests, positive_tests, cumulative_reported_deaths,
                                                             reported_cases, cumulative_reported_cases,
                                                             cumulative_reported_deaths, reported_tests))

head(recentdate_data_v5,5)
```

```
##           date      area population cases deaths
## 712 2022-01-12 Alameda   1685886  4214      0
## 713 2022-01-13 Alameda   1685886  3700      0
## 714 2022-01-14 Alameda   1685886  2755      1
## 715 2022-01-15 Alameda   1685886  1400      0
## 716 2022-01-16 Alameda   1685886  1028      1
```

```
extract_cases_7_days_average_per_100k <- function(area_name) {
  cases_7day <- c()
  cases_7day_avg <- c()
  county_toal_cases = recentdate_data_v5[recentdate_data_v5$area==area_name,]$cases
  population = max(recentdate_data_v5[recentdate_data_v5$area==area_name,]$population)
  for (i in 1:length(c(covid_dataset_newly_dataset$date))){
    if (i == 1){
      cases_7day <- append(cases_7day, county_toal_cases[1])
    }else{
      cases_window = county_toal_cases[1:i]
      cases_window_sum= sum(cases_window)
      cases_7day <- append(cases_7day, cases_window_sum)
    }
  }
  for (i in 1:7){
    cases_window_avg = cases_7day[i] / (i)
    cases_7day_avg <- append(cases_7day_avg, cases_window_avg)
  }
  Cases_per_100k_7_day_average = round((sum(cases_7day_avg)/7) / population*100000,2)
  return(Cases_per_100k_7_day_average)
}
```

```
extract_deaths_7_days_average_per_100k <- function(area_name) {
  deaths_7day <- c()
  deaths_7day_avg <- c()
  county_toal_deaths = recentdate_data_v5[recentdate_data_v5$area==area_name,]$deaths
  population = max(recentdate_data_v5[recentdate_data_v5$area==area_name,]$population)
  for (i in 1:length(c(covid_dataset_newly_dataset$date))){
    if (i == 1){
      deaths_7day <- append(deaths_7day, county_toal_deaths[1])
    }else{
      deaths_window = county_toal_deaths[1:i]
      deaths_window_sum= sum(deaths_window)
      deaths_7day <- append(deaths_7day, deaths_window_sum)
    }
  }
  for (i in 1:7){
    deaths_window_avg = deaths_7day[i] / (i)
  }
}
```

```

    deaths_7day_avg <- append(deaths_7day_avg, deaths_window_avg)
  }
  Deaths_per_100k_7_day_average = round((sum(deaths_7day_avg)/7) / population*100000,2)
  return(Deaths_per_100k_7_day_average )
}

```

```
extract_cases_7_days_average_per_100k("Alameda")
```

```
## [1] 185.03
```

```
extract_deaths_7_days_average_per_100k("Alameda")
```

```
## [1] 0.01
```

```

Area <- c()
Average_Cases_7_days <- c()
Average_Deaths_7_days <- c()
for (i in c(unique(recentdate_data_v5$area))){
  Area <- append(Area, i)
  Average_Cases_7_days <- append(Average_Cases_7_days, extract_cases_7_days_average_per_100k(i))
  Average_Deaths_7_days <- append(Average_Deaths_7_days, extract_deaths_7_days_average_per_100k(i))
}

```

```

extract_7_days_average_per_100k<- data.frame(Area,
                                              Average_Cases_7_days,
                                              Average_Deaths_7_days)
head(extract_7_days_average_per_100k,5)

```

```

##      Area Average_Cases_7_days Average_Deaths_7_days
## 1  Alameda             185.03             0.01
## 2   Alpine              13.98             0.00
## 3  Amador             117.22             0.59
## 4   Butte              96.25             0.00
## 5 Calaveras           190.12             0.00

```

```

extract_cases_7_days_average_per_100k_graph <- ggplot(data=extract_7_days_average_per_100k,
              aes(x = Average_Cases_7_days,
                  y = reorder(Area,
                              Average_Cases_7_days)))
  )+
  geom_bar(stat = "identity",
           width=.6,
           position = position_dodge(width = 0.5),
           fill = "#69c8ff")+
  labs(
    title="Cases 7 day average (per 100k) by County",
    y = "County"
  )+
  theme(axis.text.y = element_text(size = 7)) +
  geom_text(aes(label=Average_Cases_7_days),
            hjust = 0,

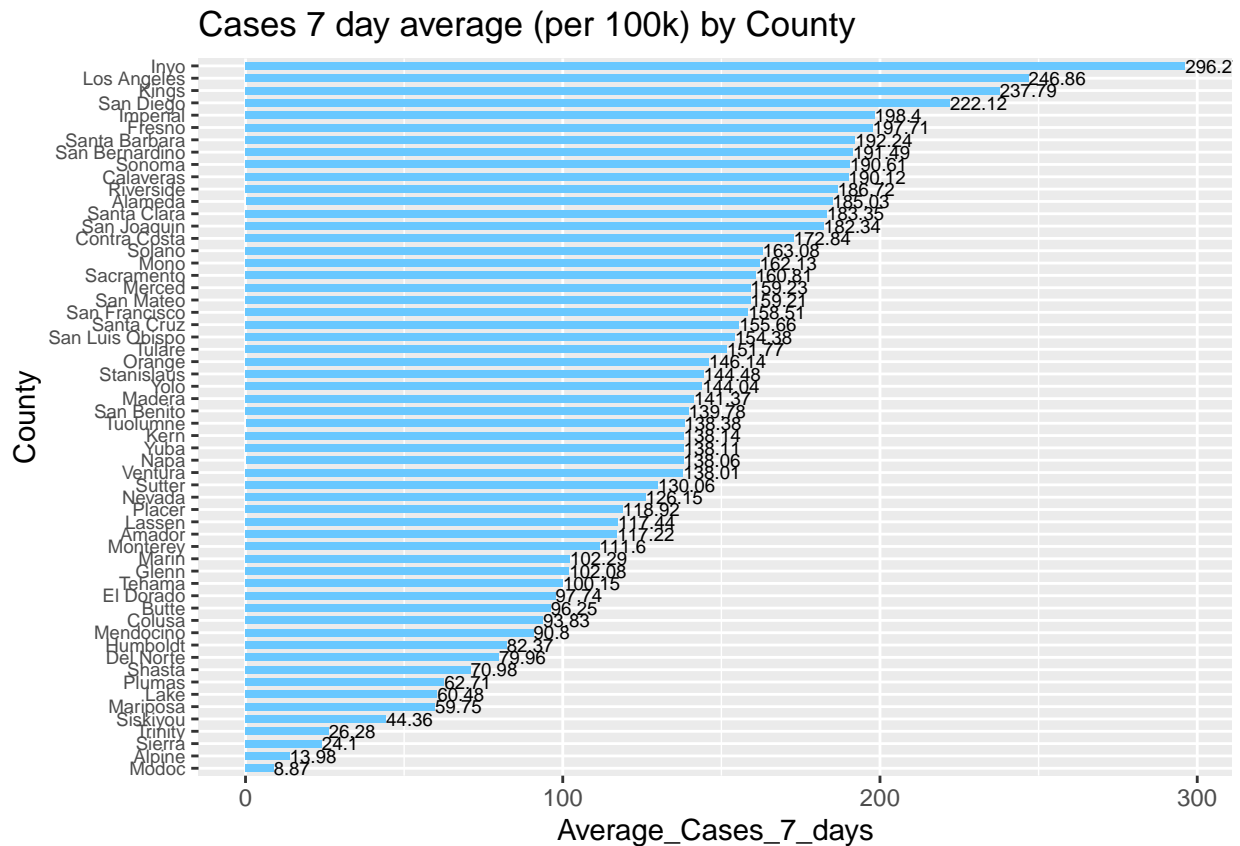
```

```

size = 2.5,
position = position_dodge(width = 1),
inherit.aes = TRUE)

options(repr.plot.width = 14, repr.plot.height = 8)
extract_cases_7_days_average_per_100k_graph

```

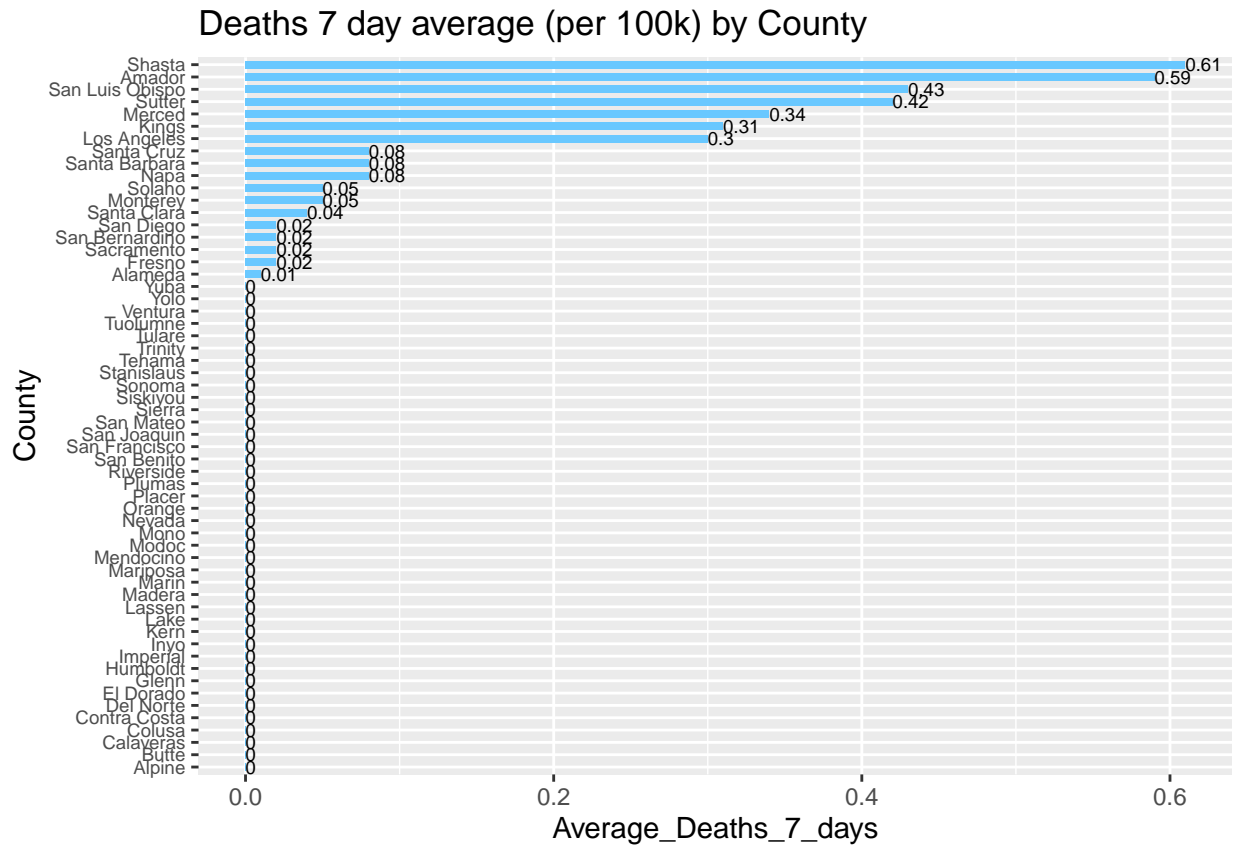


```

extract_deaths_7_days_average_per_100k_graph <- ggplot(data=extract_7_days_average_per_100k,
aes(x = Average_Deaths_7_days,
y = reorder(Area,
Average_Deaths_7_days)))
)+
geom_bar(stat = "identity",
width=.6,
position = position_dodge(width = 0.5),
fill = "#69c8ff")+
labs(
title="Deaths 7 day average (per 100k) by County",
y = "County")
)+
theme(axis.text.y = element_text(size = 7)) +
geom_text(aes(label=Average_Deaths_7_days),
hjust = 0,
size = 2.5,
position = position_dodge(width = 1),
inherit.aes = TRUE)

```

```
options(repr.plot.width = 14, repr.plot.height = 8)
extract_deaths_7_days_average_per_100k_graph
```



```
Total_Cases <- aggregate(cases ~ date, data_v1,sum)
Total_Cases <- Total_Cases[-1,]
row.names(Total_Cases) <- NULL
Total_Deaths <- aggregate(deaths ~ date, data_v1,sum)
Total_Deaths <- Total_Deaths[-1,]
row.names(Total_Deaths) <- NULL

date <- c(unique(data_v1$date))
date <- head(date, -1)
Total_Cases_7days <- c(Total_Cases$cases)
Total_Deaths_7days <- c(Total_Deaths$deaths)

covid_moving_average_dataset<- data.frame(date,
                                           Total_Cases_7days,
                                           Total_Deaths_7days)

cases_7day <- c()
deaths_7day <- c()
window_size =7
for (i in 1:length(c(covid_moving_average_dataset$date))){
  if (i ==1){
    cases_7day <- append(cases_7day, covid_moving_average_dataset$Total_Cases_7days[1])
```

```

    deaths_7day <- append(deaths_7day, covid_moving_average_dataset$Total_Deaths_7days[1])
  }else{
    cases_window = covid_moving_average_dataset$Total_Cases_7days[i:(window_size+i-1)]
    deaths_window = covid_moving_average_dataset$Total_Deaths_7days[i:(window_size+i-1)]
    cases_window_sum= sum(cases_window) / window_size
    deaths_window_sum= sum(deaths_window) / window_size
    cases_7day <- append(cases_7day, cases_window_sum)
    deaths_7day <- append(deaths_7day, deaths_window_sum)
  }
}

covid_moving_average_dataset$Cases_Moving_average_7days <- cases_7day
covid_moving_average_dataset$Deaths_Moving_average_7days <- deaths_7day
head(covid_moving_average_dataset, 5)

```

```

##           date Total_Cases_7days Total_Deaths_7days Cases_Moving_average_7days
## 1 2020-02-01                23                0                23.000000
## 2 2020-02-02                 7                0                8.857143
## 3 2020-02-03                 5                0                8.714286
## 4 2020-02-04                 1                0                8.857143
## 5 2020-02-05                 3                0                9.285714
## Deaths_Moving_average_7days
## 1                0.0000000
## 2                0.1428571
## 3                0.1428571
## 4                0.1428571
## 5                0.1428571

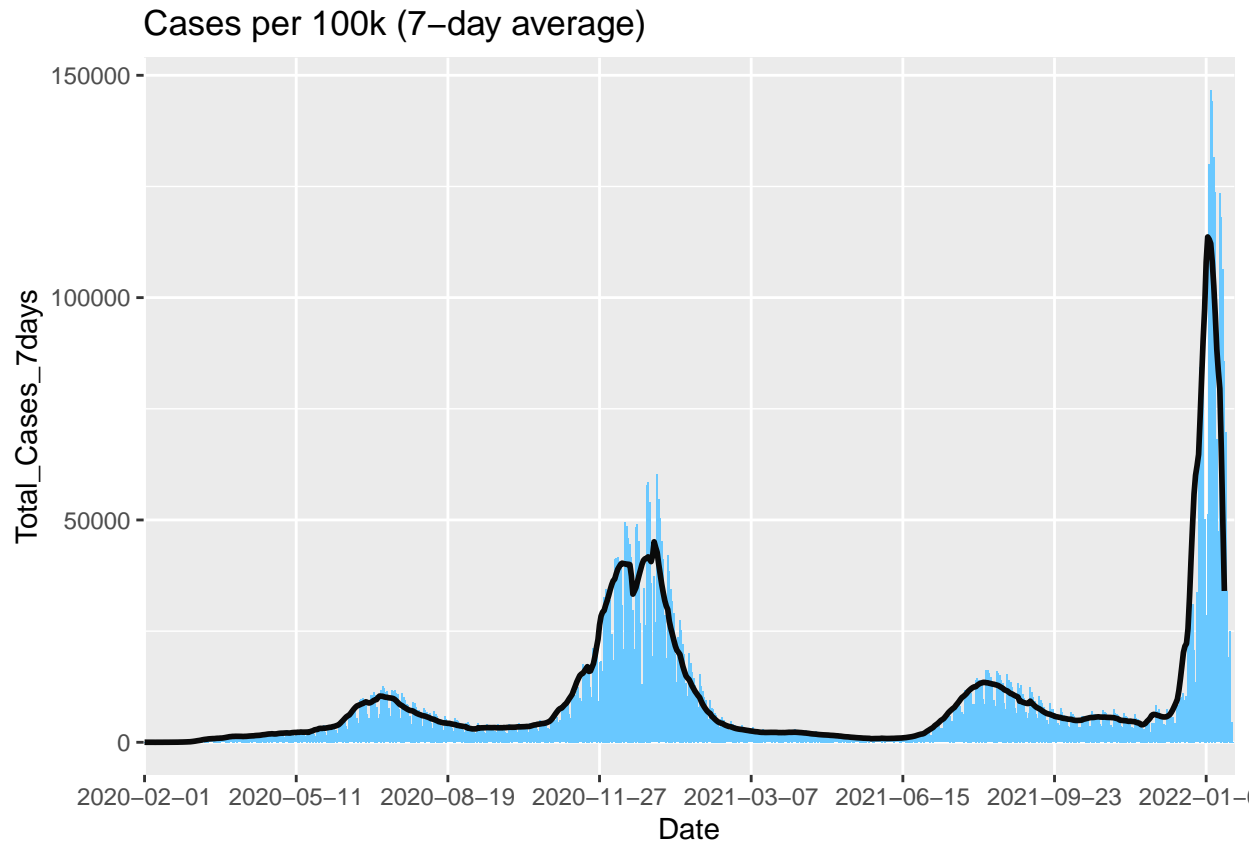
```

```

covid_moving_average_cases_per_100k_graph <- ggplot(covid_moving_average_dataset)+
  geom_bar(aes(x=date,
               y=Total_Cases_7days),
           stat="identity",
           fill= "#69c8ff")+
  geom_line(aes(date,
                Cases_Moving_average_7days,
                group = 1),
            col = "#0a0a0a",
            size=1)+
  labs(
    title="Cases per 100k (7-day average)",
    x = "Date"
  )+
  scale_x_discrete(breaks = function(x) x[seq(1, length(x),
options(repr.plot.width = 14, repr.plot.height = 8)
covid_moving_average_cases_per_100k_graph

```

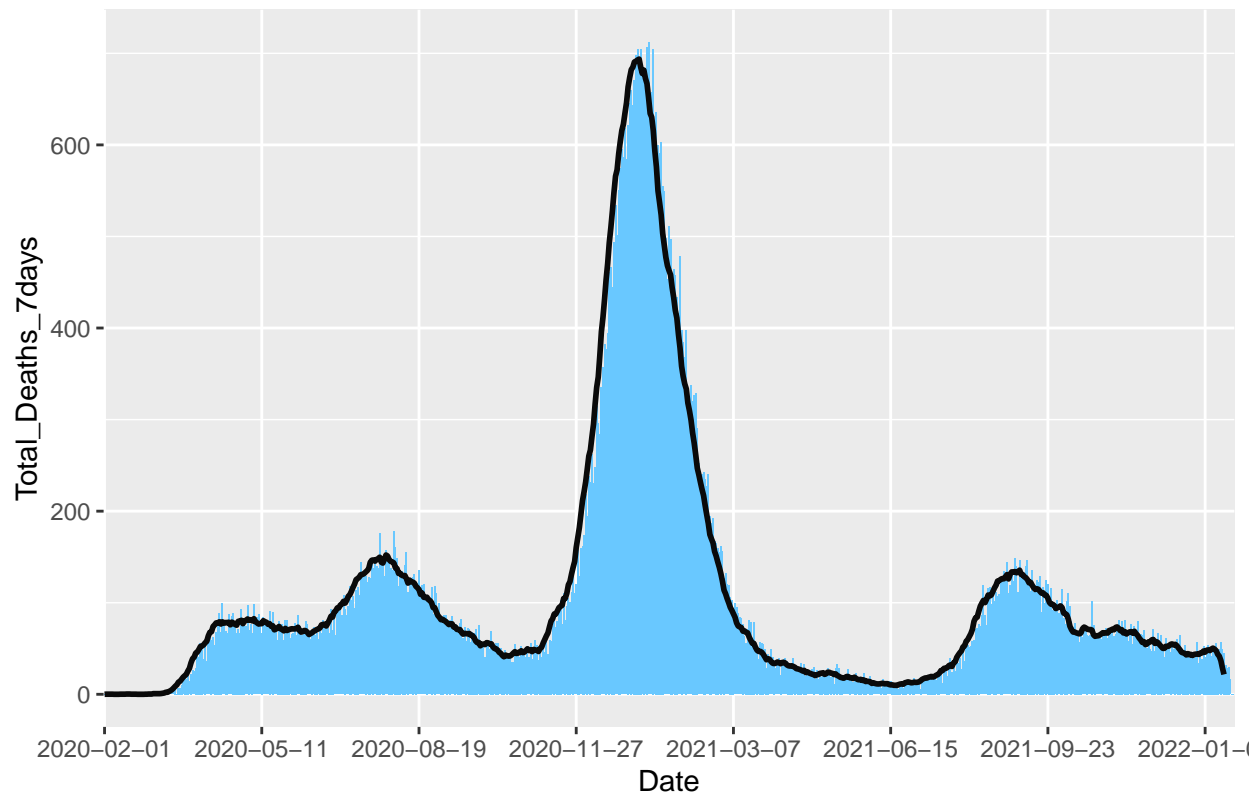
```
## Warning: Removed 6 row(s) containing missing values (geom_path).
```



```
covid_moving_average_deaths_per_100k_graph <- ggplot(covid_moving_average_dataset)+
  geom_bar(aes(x=date,
               y=Total_Deaths_7days),
           stat="identity",
           fill= "#69c8ff")+
  geom_line(aes(date,
                 Deaths_Moving_average_7days,
                 group = 1),
            col = "#0a0a0a",
            size=1)+
  labs(
    title ="Deaths per 100k (7-day average)",
    x = "Date"
  )+
  scale_x_discrete(breaks = function(x) x[seq(1, length(x),
options(repr.plot.width = 14, repr.plot.height = 8)
covid_moving_average_deaths_per_100k_graph
```

```
## Warning: Removed 6 row(s) containing missing values (geom_path).
```


Deaths per 100k (7-day average)



```
Total_CA_Cases = sum(total_cases_by_area$cases)
Total_CA_Deaths = sum(total_deaths_by_area$deaths)
RecentCases = sum(data_v1[data_v1$date== RecentDate_end ,]$reported_cases)
RecentDeaths = sum(data_v1[data_v1$date==RecentDate_end, ]$reported_deaths)
Cases_percent = round((RecentCases/Total_CA_Cases)*100,4)
Deaths_percent = round((RecentDeaths/Total_CA_Deaths)*100,4)

cat("Cases (Statewide)", "\n",
    "Total US Cases : ", Total_CA_Cases, "\n",
    "Covid19 cases in (", RecentDate_end ,"): ",RecentCases, "(+", Cases_percent, "%)", "\n",
    "Cases per 100k 7-day average (7 period ending", RecentDate_end ,"): ", Cases_per_100k_7_day_averag
    "-----", "\n",
    "Deaths (Statewide)", "\n",
    "Total US Deaths : ", Total_CA_Deaths, "\n",
    "Covid19 Deaths in (", RecentDate_end ,"): ",RecentDeaths, "(+", Deaths_percent, "%)", "\n",
    "Deaths per 100k 7-day average (7 period ending", RecentDate_end ,"): ", Deaths_per_100k_7_day_aver
    )
```

```
## Cases (Statewide)
## Total US Cases : 6993788
## Covid19 cases in ( 2022-01-19 ): 88499 (+ 1.2654 %)
## Cases per 100k 7-day average (7 period ending 2022-01-19 ): 204.91
## -----
## Deaths (Statewide)
## Total US Deaths : 77518
## Covid19 Deaths in ( 2022-01-19 ): 176 (+ 0.227 %)
```

Deaths per 100k 7-day average (7 period ending 2022-01-19): 0.11