

Covid19_data_Statistics_visualization.R

corn

2022-01-21

```
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_012122.csv")  
data_v1 <- read.csv("./dataset/covid19cases_test_012122.csv")
```

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

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

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

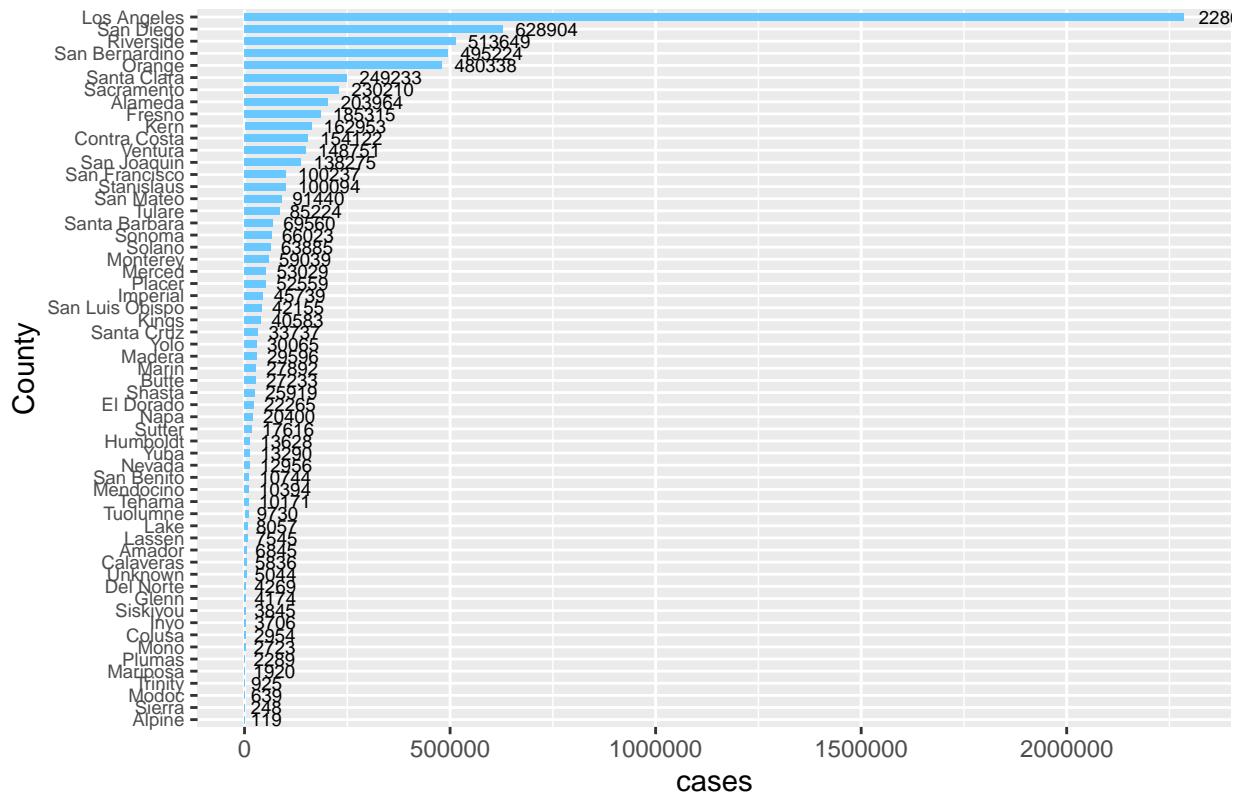
```
## [1] "Alameda"      "Alpine"      "Amador"      "Butte"  
## [5] "Calaveras"   "Colusa"      "Contra Costa" "Del Norte"  
## [9] "El Dorado"   "Fresno"      "Glenn"       "Humboldt"  
## [13] "Imperial"    "Inyo"        "Kern"        "Kings"  
## [17] "Lake"        "Lassen"      "Los Angeles" "Madera"  
## [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"    "Unknown"  
## [57] "Ventura"    "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)

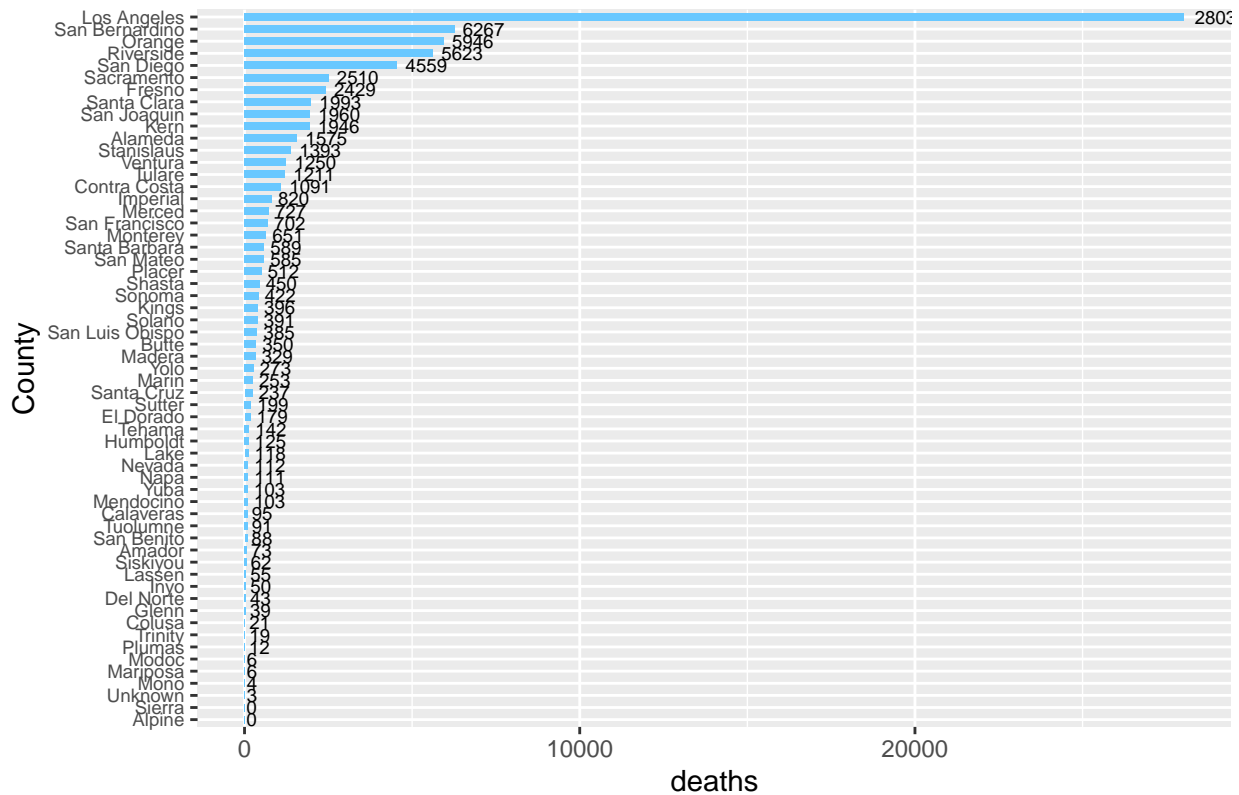
total_deaths_by_area_graph <- ggplot(data=total_deaths_by_area,
                                     aes(x = deaths,
                                          y = reorder(area,
                                                         deaths,
                                                         sum)))

  )+
  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]

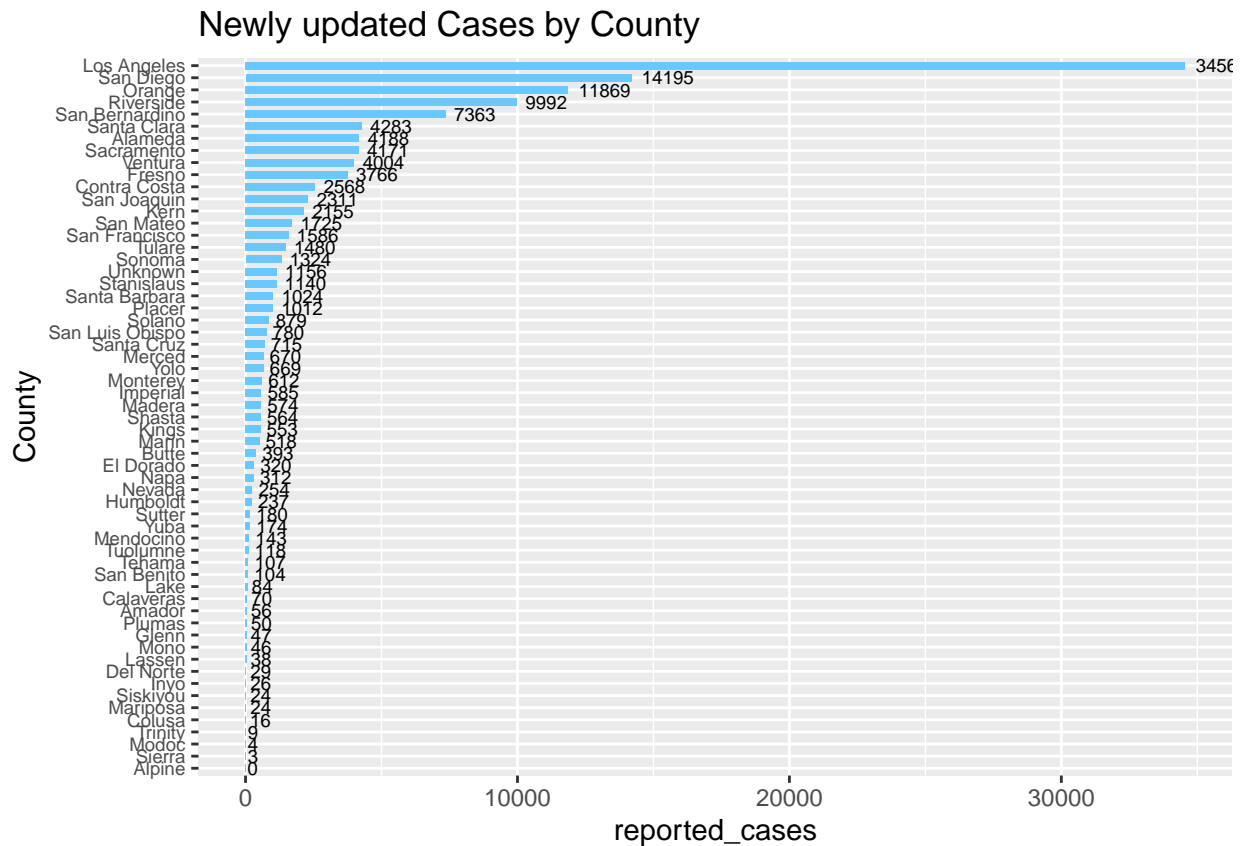
recentdate_data_v1 <- data_v1[data_v1$date == RecentDate_end,]

#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
```



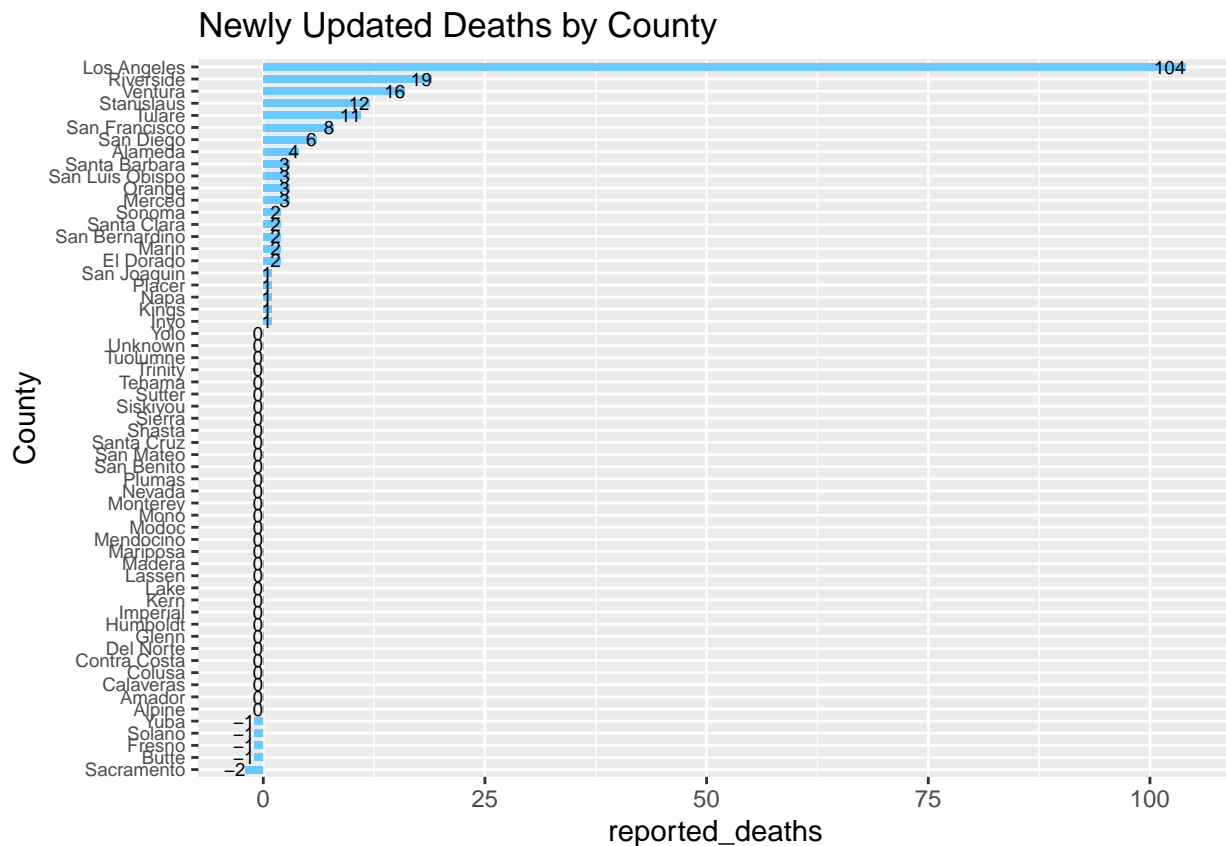
```
#Newly Updated Deaths-----
newly_updated_deaths_by_area <- aggregate(reported_deaths ~ area,
                                           recentdate_data_v1,
                                           sum)
newly_updated_deaths_by_area_graph <- ggplot(data=newly_updated_deaths_by_area,
                                              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))

## Warning in total_cases_by_area$cases/newly_population_by_area$population: longer
## object length is not a multiple of shorter object length

cumulative_deaths_per_100k <- c(round((total_deaths_by_area$deaths/newly_population_by_area$population)* 100000))

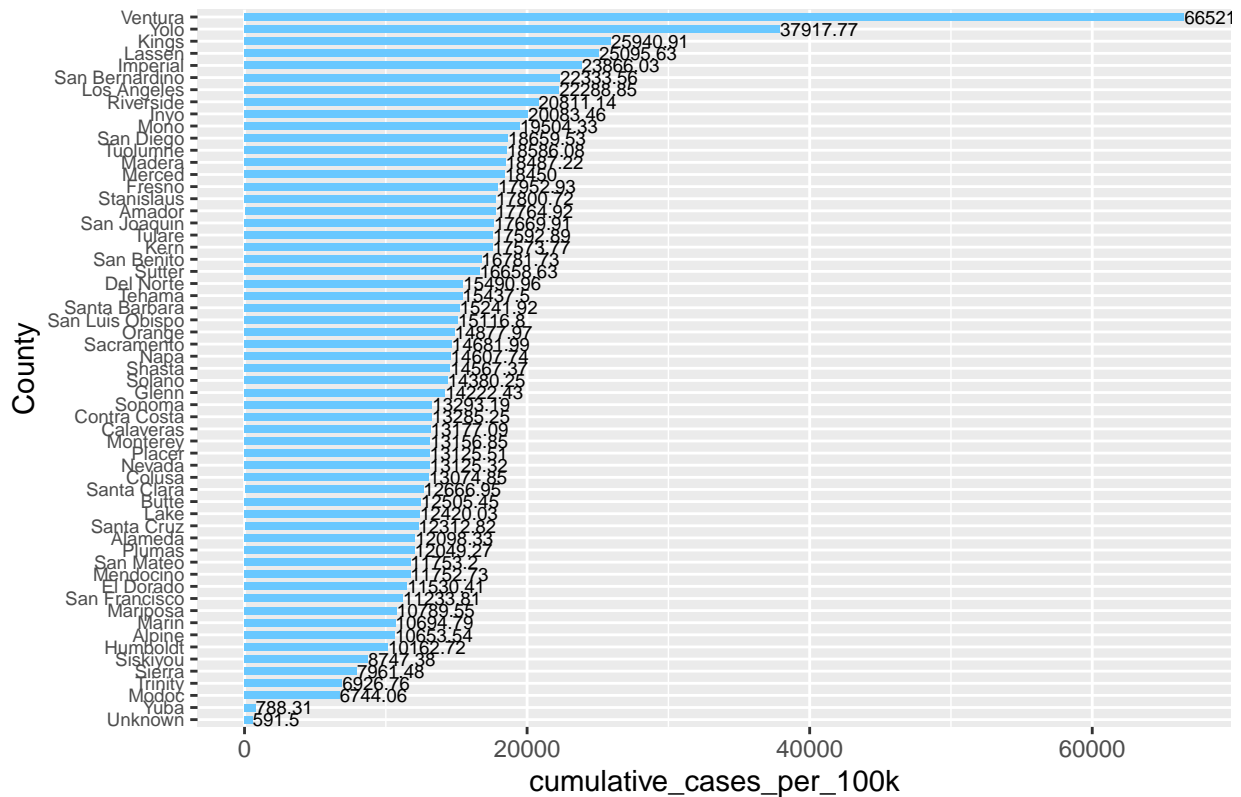
```

```
## Warning in total_deaths_by_area$deaths/newly_population_by_area$population:
## longer object length is not a multiple of shorter object length
```

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

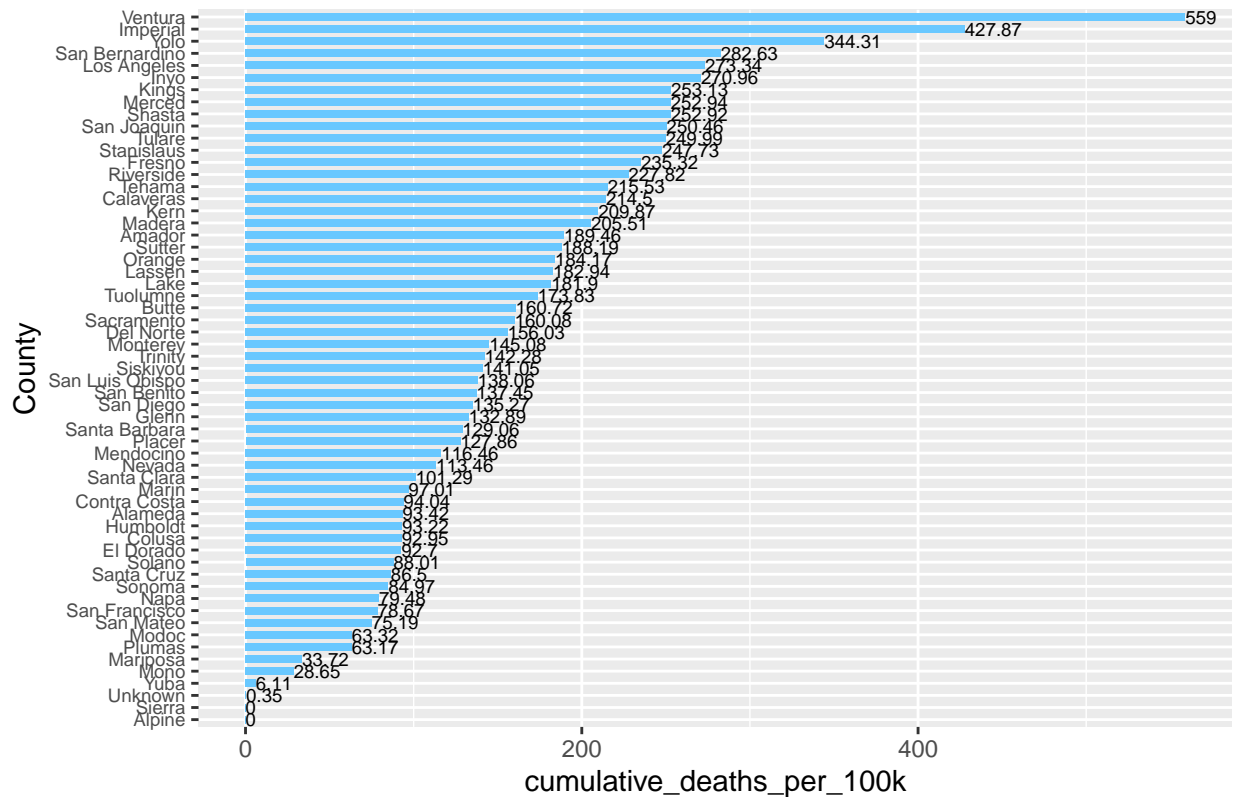
#Cumulative Cases per 100k -----
cumulative_cases_per_100k_graph <- ggplot(data=cumulative_per_100k,
    aes(x = cumulative_cases_per_100k,
        y = reorder(area,
                    cumulative_cases_per_100k))
    )+
  geom_bar(stat = "identity",
    width=.6,
    position = position_dodge(width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



```
#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
```


Cumulative Deaths per 100k by County



#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
## 713	2022-01-13	Alameda	County	1685886	4367	194475	0
## 714	2022-01-14	Alameda	County	1685886	3338	197813	1
## 715	2022-01-15	Alameda	County	1685886	1479	199292	1
## 716	2022-01-16	Alameda	County	1685886	1126	200418	1
## 717	2022-01-17	Alameda	County	1685886	2098	202516	0

##	cumulative_deaths	total_tests	cumulative_total_tests	positive_tests
## 713	1569	30569		5255142
## 714	1570	24856		5279998
## 715	1571	9895		5289893
## 716	1572	7509		5297402
## 717	1572	13653		5311055

##	cumulative_positive_tests	reported_cases	cumulative_reported_cases
## 713	236502	4767	177768
## 714	240900	0	177768
## 715	242784	0	177768
## 716	244208	12923	190691
## 717	246929	2585	193276

##	reported_deaths	cumulative_reported_deaths	reported_tests
## 713	-2	1561	25440
## 714	0	1561	NA

```
## 715          0          1561          NA
## 716          3          1564        89440
## 717          1          1565        18702
```

```
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 2022-01-13          101645             61
## 2 2022-01-14           84954             43
## 3 2022-01-15           39720             47
## 4 2022-01-16           24151             43
## 5 2022-01-17           44865             33
## 6 2022-01-18           33286             22
## 7 2022-01-19           3858              7
## 8 2022-01-20              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 2022-01-13      101645      61      101645
## 2 2022-01-14      84954      43      186599
## 3 2022-01-15      39720      47      226319
## 4 2022-01-16      24151      43      250470
## 5 2022-01-17      44865      33      295335
## 6 2022-01-18      33286      22      328621
## 7 2022-01-19       3858       7      332479
## 8 2022-01-20         0       0      332479
##   Total_Deaths_7days_sum
## 1          61
## 2         104
## 3         151
## 4         194
## 5         227
## 6         249
## 7         256
## 8         256
```

```
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] 190.78
```

```
recentdate_data_v5 = subset(recentdate_data_v2, select = -c(area_type, cumulative_cases, cumulative_deaths,
  cumulative_total_tests, positive_tests, cumulative_reported_cases, cumulative_reported_deaths, reported_tests))
head(recentdate_data_v5,5)
```

```
##           date      area population cases deaths
## 713 2022-01-13 Alameda   1685886   4367      0
## 714 2022-01-14 Alameda   1685886   3338      1
## 715 2022-01-15 Alameda   1685886   1479      1
## 716 2022-01-16 Alameda   1685886   1126      1
## 717 2022-01-17 Alameda   1685886   2098      0
```

```
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] 174.59
```

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

```
## [1] 0.03
```

```

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             174.59             0.03
## 2   Alpine              20.37             0.00
## 3  Amador              87.00             0.96
## 4   Butte              84.82             0.00
## 5 Calaveras           118.00             0.00

```

```

extract_7_days_average_per_100k <- filter(extract_7_days_average_per_100k,
                                           !area %in% c('Unknown'))

```

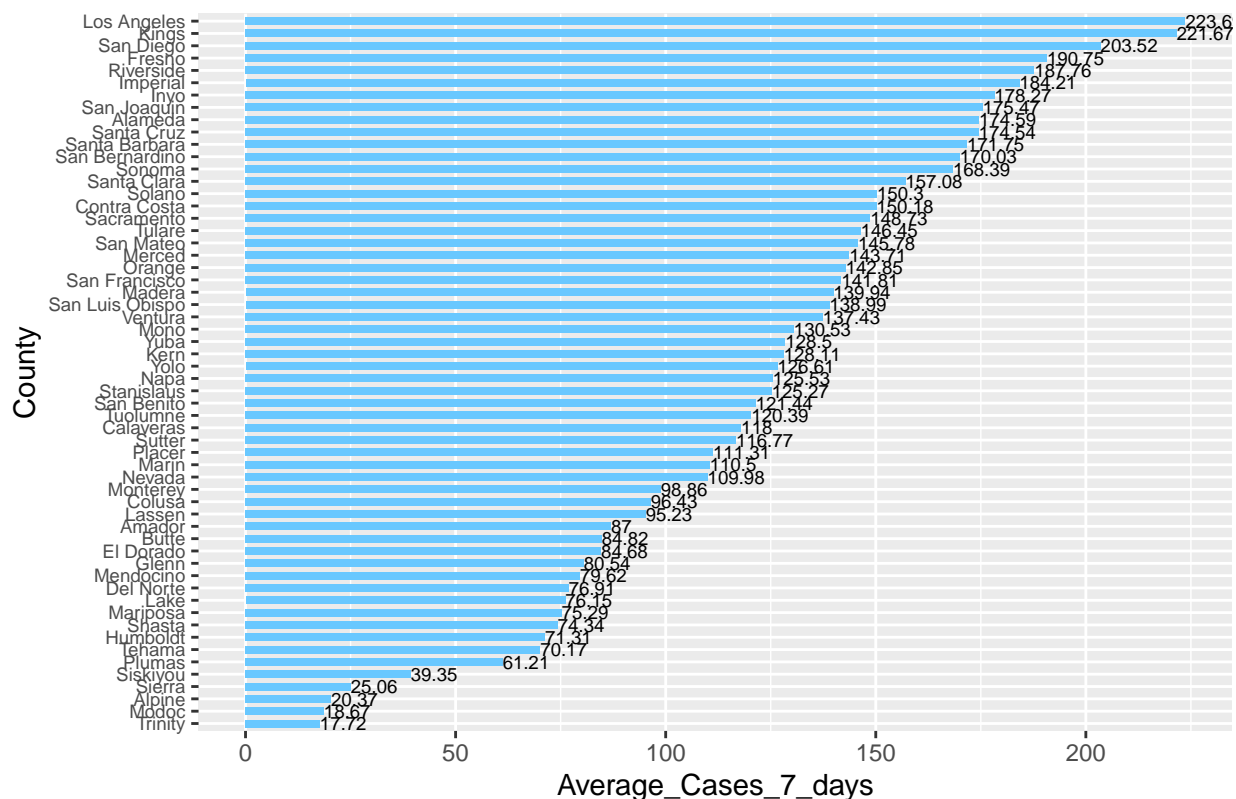
```

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

```

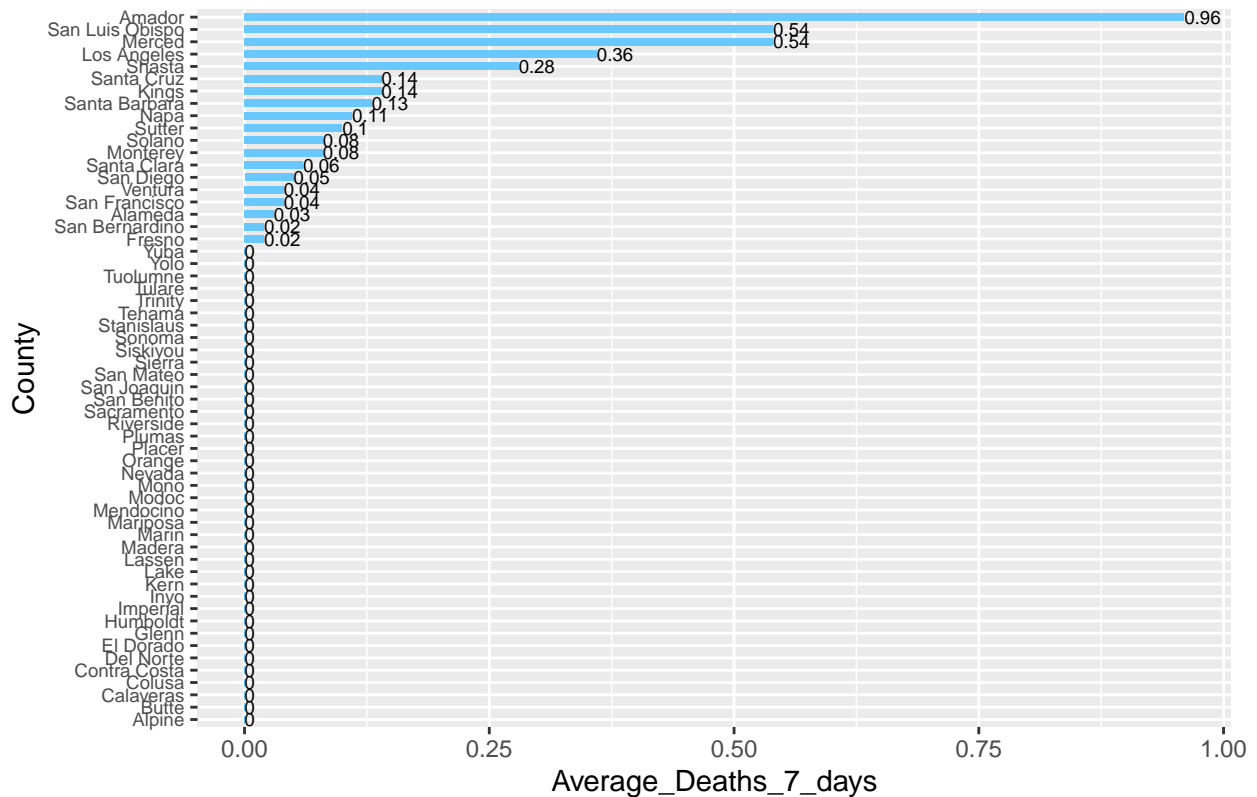
Cases 7 day average (per 100k) by County



```
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
```

Deaths 7 day average (per 100k) by County



```
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                9.000000
## 5 2020-02-05                 3                0                9.428571
## Deaths_Moving_average_7days
## 1                0.0000000
## 2                0.1428571
## 3                0.1428571
## 4                0.1428571
## 5                0.1428571

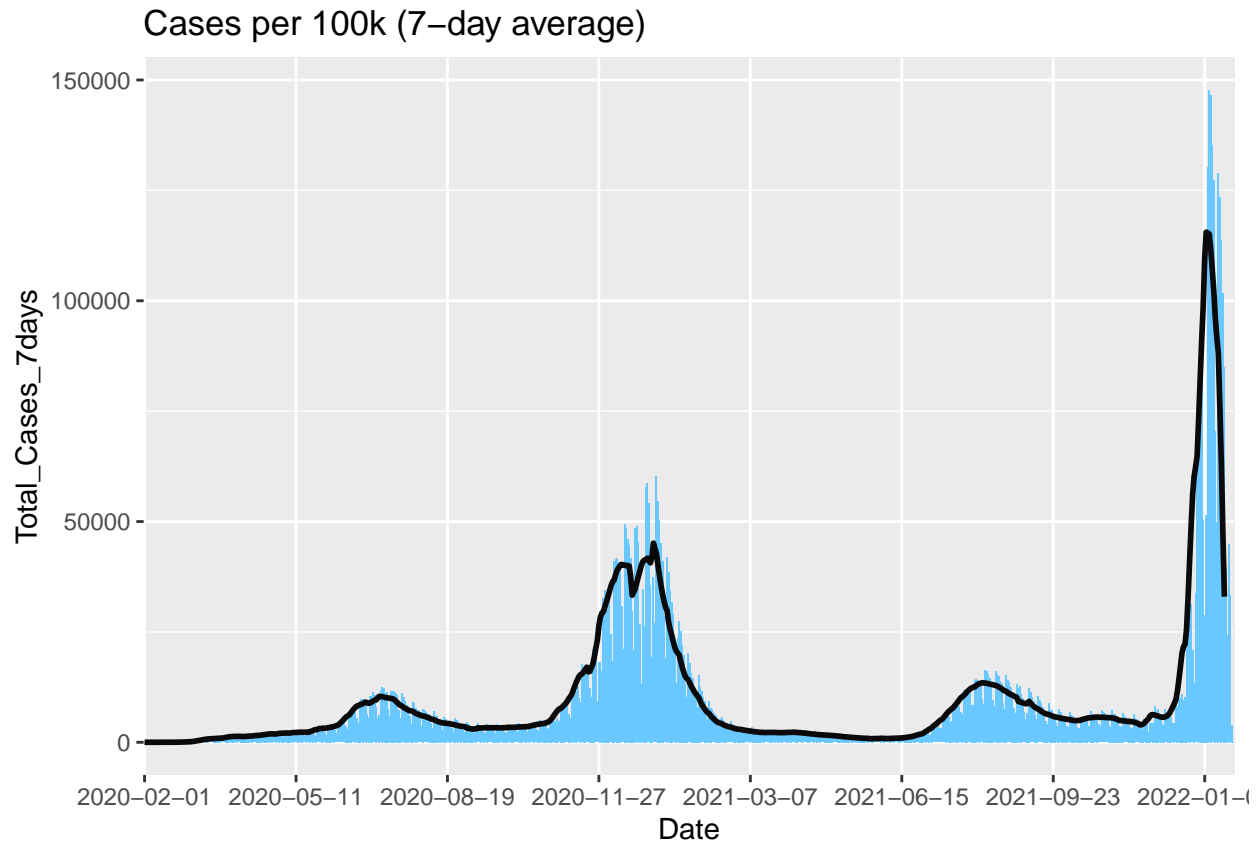
```

```

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

```

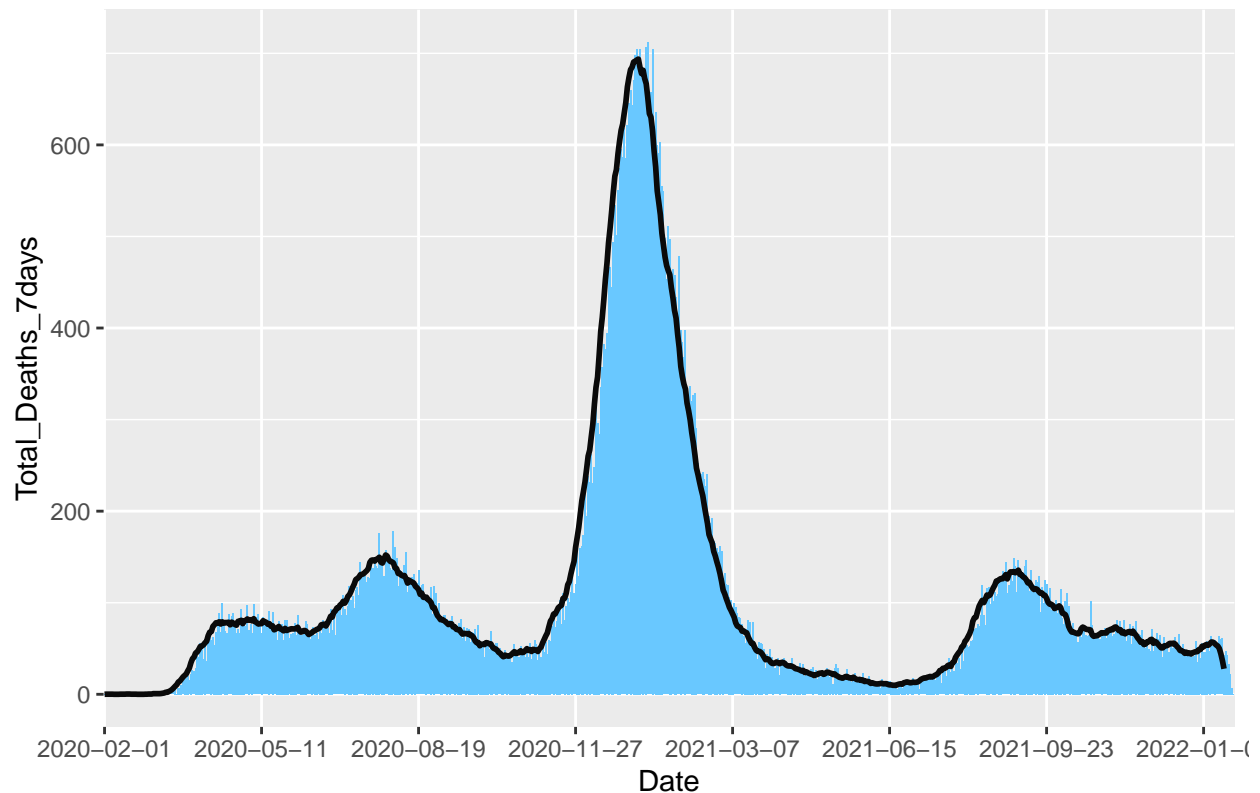
```
## Warning: Removed 6 row(s) containing missing values (geom_path).
```

```
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 : 7123571
## Covid19 cases in ( 2022-01-20 ): 125861 (+ 1.7668 %)
## Cases per 100k 7-day average (7 period ending 2022-01-20 ): 190.78
## -----
## Deaths (Statewide)
## Total US Deaths : 77722
## Covid19 Deaths in ( 2022-01-20 ): 201 (+ 0.2586 %)
```

Deaths per 100k 7-day average (7 period ending 2022-01-20): 0.13