Question 15

Wanying Tian

7/16/2022

Wanying Tian A91028356

```
#read the CSV file and check how it is look like
covid<-read.csv("covid19_variants.csv")
head(covid)
```

```
##         date       area area_type variant_name specimens percentage
## 1 2021-01-01 California     State        Alpha         1       1.69
## 2 2021-01-01 California     State         Beta         0       0.00
## 3 2021-01-01 California     State           Mu         0       0.00
## 4 2021-01-01 California     State        Gamma         0       0.00
## 5 2021-01-01 California     State        Total        59     100.00
## 6 2021-01-01 California     State      Omicron         1       1.69
##   specimens_7d_avg percentage_7d_avg
## 1               NA                NA
## 2               NA                NA
## 3               NA                NA
## 4               NA                NA
## 5               NA                NA
## 6               NA                NA
```

```
#remove the "total" and "other" rows
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
```

```
nototal<-covid %>%
  filter(!variant_name=='Total'& !variant_name=='Other')
head(nototal)
```

```
##         date       area area_type variant_name specimens percentage
## 1 2021-01-01 California     State        Alpha         1       1.69
## 2 2021-01-01 California     State         Beta         0       0.00
## 3 2021-01-01 California     State           Mu         0       0.00
## 4 2021-01-01 California     State        Gamma         0       0.00
## 5 2021-01-01 California     State      Omicron         1       1.69
## 6 2021-01-01 California     State      Epsilon        28      47.46
##   specimens_7d_avg percentage_7d_avg
## 1               NA                NA
## 2               NA                NA
## 3               NA                NA
## 4               NA                NA
## 5               NA                NA
## 6               NA                NA
```

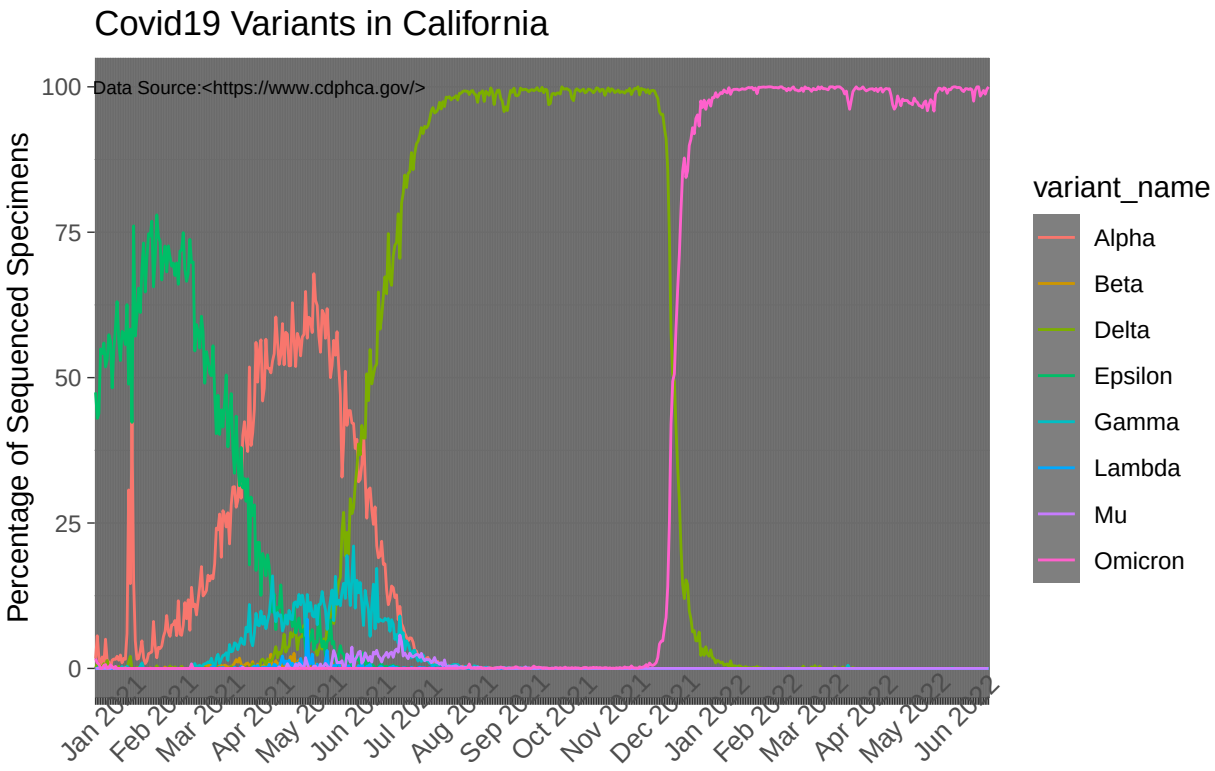
```
#to create the x axis label as month and year instead of date
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
date_list<-unique(nototal$date)
Yr <- format(as.Date(date_list), "%Y")
Month <- month(as.Date(date_list), label = TRUE, abbr = TRUE)
Month_Yr <- paste(Month, Yr, sep= " ")
Month_Yr[duplicated(Month_Yr)] <- ""
```

```
library(ggplot2)
ggplot(nototal,aes(date,percentage,group=variant_name,color=variant_name ))+
       geom_line()+
     labs(x = "", y = "Percentage of Sequenced Specimens",
              title = "Covid19 Variants in California")+
   scale_x_discrete(labels= Month_Yr)+
  theme_dark()+
  theme(axis.text.x = element_text( size=10, angle=45))+
   annotate('text', x = 100, y = 100, size=2.5,label = 'Data Source:<https://www.cdphca.gov/>')
```


Covid19 Variants in California
Data Source:<https://www.cdphca.gov/>
Percentage of Sequenced Specimens
100
75
50
25
0
Jan 2021
Feb 2021
Mar 2021
Apr 2021
May 2021
Jun 2021
Jul 2021
Aug 2021
Sep 2021
Oct 2021
Nov 2021
Dec 2021
Jan 2022
Feb 2022
Mar 2022
Apr 2022
May 2022
Jun 2022
variant_name
Alpha
Beta
Delta
Epsilon
Gamma
Lambda
Mu
Omicron