1st_year_final_CBrennan

Caitriona Brennan PID: A59011982

6/28/2022

Importing data and saving as a variable:

```
covid_data <- read.csv("covid-19-variant-data.csv")
head(covid_data)
```

```
##         date       area area_type variant_name specimens percentage
## 1 2021-01-01 California     State        Delta         0       0.00
## 2 2021-01-01 California     State        Alpha         1       1.69
## 3 2021-01-01 California     State        Other        29      49.15
## 4 2021-01-01 California     State        Total        59     100.00
## 5 2021-01-01 California     State         Beta         0       0.00
## 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
```

Loading packages required to make plot:

```
library(ggplot2)
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(lubridate)
```

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

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

```
library(scales)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(ggthemes)
```

Filtering out unwanted data:

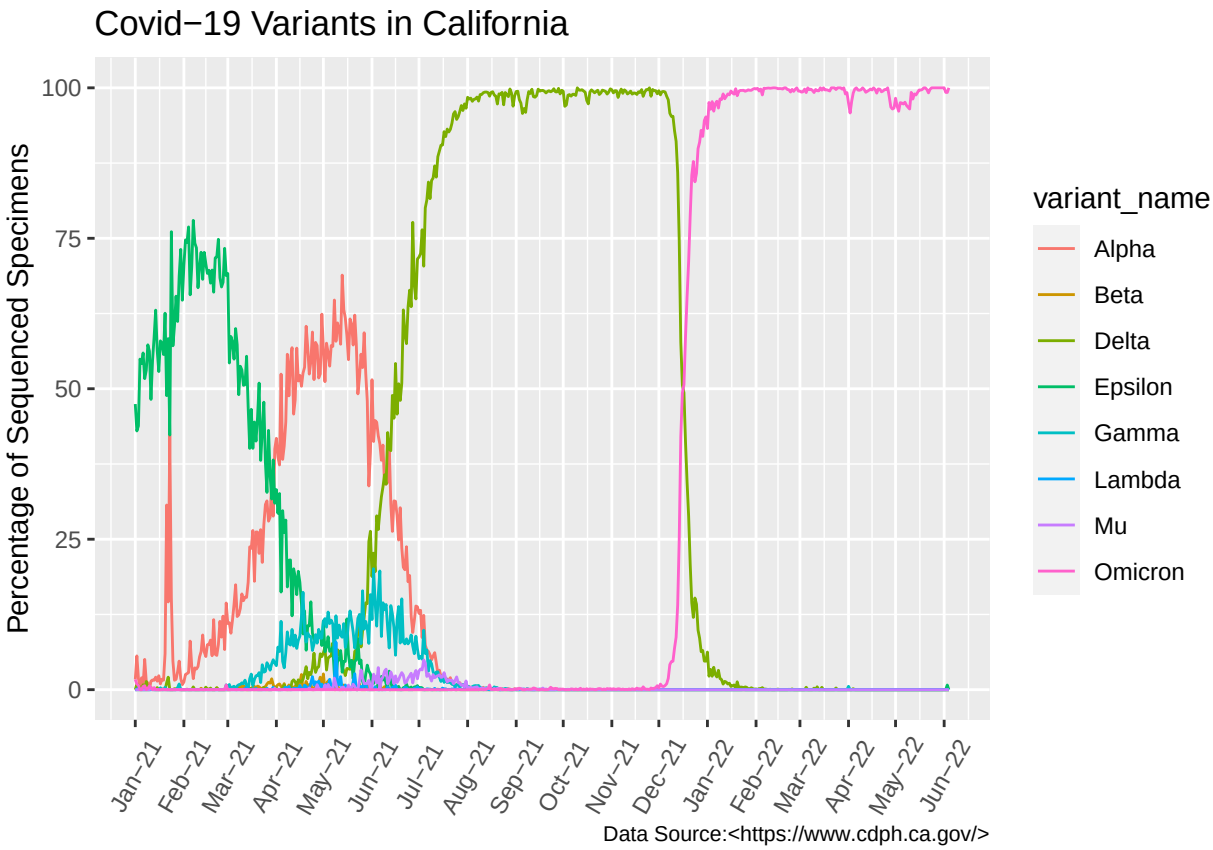
```
covid_data4 <- filter(covid_data, variant_name!="Total" & variant_name!='Other')
```

Converting "date" column to date format:

```
covid_data4$date <- as.Date(covid_data4$date)
```

Making plot which uses one tick per month on x-axis. X-axis format is in "month-year".

```
ggplot(covid_data4, aes(x = date, y=percentage, group=variant_name)) +
  geom_line(aes(color=variant_name)) +
  theme(axis.text.x = element_text(angle = 60, vjust = 0.5)) +
  scale_x_date(labels=date_format("%h-%y"), breaks=date_breaks("1 month")) +
  xlab("Data Source:<https://www.cdph.ca.gov/>") +
  ylab("Percentage of Sequenced Specimens") +
  ggtitle("Covid-19 Variants in California") +
  theme(axis.title.x=element_text(size=8, hjust = 1))
```


Covid−19 Variants in California
Percentage of Sequenced Specimens
100
75
50
25
0
Jan−21
Feb−21
Mar−21
Apr−21
May−21
Jun−21
Jul−21
Aug−21
Sep−21
Oct−21
Nov−21
Dec−21
Jan−22
Feb−22
Mar−22
Apr−22
May−22
Jun−22
Data Source:<https://www.cdph.ca.gov/>
variant_name
Alpha
Beta
Delta
Epsilon
Gamma
Lambda
Mu
Omicron