Exam

Lizzie West (PID: 59010743)

5/29/2022

Steps

1. Make sure necessary packages are installed
2. Load necessary packages
3. Assign data to variable and look at data
4. Format dates using lubridate
5. Filter out entries that you do not want plotted
6. Generate plot using ggplot

Load Packages

```
#Load necessary packages.They were already installed using install.packages("package_name")
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.2
```

```
library(lubridate)
```

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

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

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.1.2

##
## 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
```

Import Data

```
#Import the data
dat <- read.csv("covid19_variants.csv")
```

Look at Data

```
#Look at the data
head(dat)
```

```
##         date       area area_type variant_name specimens percentage
## 1 2021-01-01 California     State        Alpha         1       1.69
## 2 2021-01-01 California     State           Mu         0       0.00
## 3 2021-01-01 California     State        Other        29      49.15
## 4 2021-01-01 California     State        Delta         0       0.00
## 5 2021-01-01 California     State         Beta         0       0.00
## 6 2021-01-01 California     State        Total        59     100.00
##   specimens_7d_avg percentage_7d_avg
## 1               NA                NA
## 2               NA                NA
## 3               NA                NA
## 4               NA                NA
## 5               NA                NA
## 6               NA                NA
```

Format Dates

```
#Use lubridate to format the date column as actual dates and not a string
dat$date <- ymd(dat$date)
```

Filter Entries

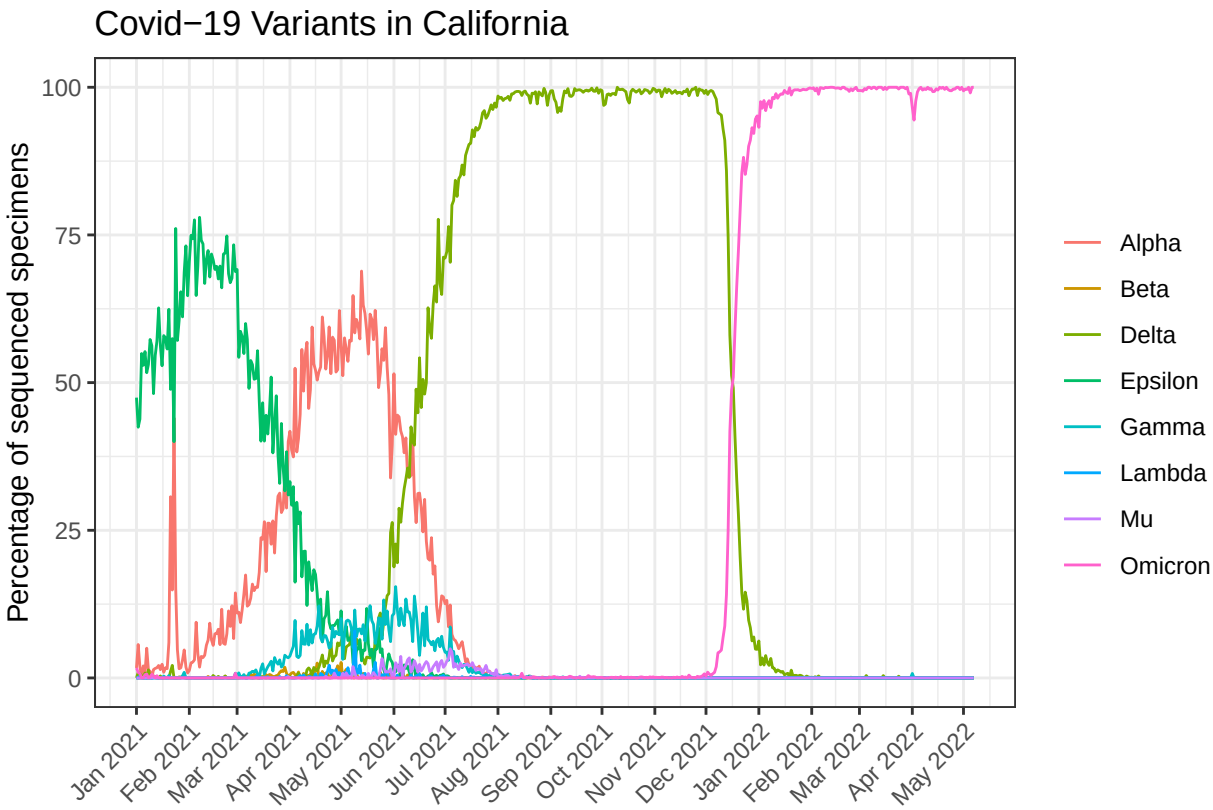
```
#Filter out the "Other" and "Total" entries
filter_dat <- filter(dat, variant_name != "Total" & variant_name != "Other")
```

Plot

Plot the filtered data using ggplot. The x-axis is date in lubridate format and y-axis is percentage. Geom_line allows for a line plot that can be colored by the variant. Labs enables labeling of axes and removal of label above variants by setting color="". scale_x_date allows you to set the x-axis tick marks, spacing by one month intervals, and labeling by abbreviated month name and year. Lastly, we can apply the black and white theme and adjust the axis labels.

```
#Plot!
vid_plot <- ggplot(filter_dat) + aes(date, percentage) +
  geom_line(aes(color = variant_name)) +
  labs(x="", y="Percentage of sequenced specimens",
       title= "Covid-19 Variants in California", color="") +
  scale_x_date(date_breaks="1 month", date_labels="%b %Y")

vid_plot + theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```


Covid−19 Variants in California
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
Alpha
Beta
Delta
Epsilon
Gamma
Lambda
Mu
Omicron