Bioinfo Exam Q10

Zijing Tan PID: A14534409

2023-07-02

```
#load packages
library('ggplot2')
library('lubridate')
```

```
## Loading required package: timechange
```

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

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

```
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
```

```
#load & inspect data
data <- read.csv('covid19_variants.csv')
head(data)
```

```
##         date       area area_type variant_name specimens percentage
## 1 2021-01-01 California     State        Alpha         1       1.67
## 2 2021-01-01 California     State        Other        29      48.33
## 3 2021-01-01 California     State        Delta         0       0.00
## 4 2021-01-01 California     State        Gamma         0       0.00
## 5 2021-01-01 California     State      Omicron         1       1.67
## 6 2021-01-01 California     State        Total        60     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
```

```
tail(data)
```

```
##            date       area area_type variant_name specimens percentage
## 8835 2023-06-03 California     State      Omicron         5        100
## 8836 2023-06-03 California     State        Other         0          0
## 8837 2023-06-03 California     State       Lambda         0          0
## 8838 2023-06-03 California     State           Mu         0          0
## 8839 2023-06-03 California     State         Beta         0          0
## 8840 2023-06-03 California     State        Alpha         0          0
##      specimens_7d_avg percentage_7d_avg
## 8835               23               100
## 8836                0                 0
## 8837                0                 0
## 8838                0                 0
## 8839                0                 0
## 8840                0                 0
```

```
#remove the 'Total' and 'Other' variant_name data from the set
data <- data%>%filter(variant_name!='Total')%>%filter(variant_name!='Other')
#change data type in date column to date
data$date <- ymd(data$date)
```

```
#plotting the data
ggplot(data,aes(date,percentage,group=variant_name))+
  geom_line(aes(color=variant_name))+
  labs(title='Covid-19 Variants in California',y='Percentage of sequenced specimens',color='Covid Varia
  scale_x_date(date_breaks = 'month',date_labels = '%b %Y')+
  theme_bw()+
  theme(axis.text.x = element_text(angle = 45,vjust=1,hjust=1,size=7))
```