Bioinformatic_FY_exam

San Luc (PID: A59010657)

2022-06-20

```
library(bio3d)
library(ggplot2)
library(tidyr)
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
```

```
#load csv file of COVID-19 Variant Data
Variant <- read.csv("/Users/sanluc/Desktop/Lab Folder/MISC/First Year Exam/covid19_variants.csv")
View(Variant)
```

```
#changing the format of the date column
Variant$date <- ymd(Variant$date)
str(Variant$date)
```

```
##  Date[1:5130], format: "2021-01-01" "2021-01-01" "2021-01-01" "2021-01-01" "2021-01-01" ...
```

```
#excluding rows containing Total or Other variant

new_variant <- filter(Variant, variant_name != "Other" & variant_name != "Total")
```

```
#graph the plot
ggplot(new_variant) +
  geom_line(aes(x = date, y = percentage, colour = variant_name))+
  scale_x_date(date_break = "1 month", date_labels = "%b %Y")+
  theme(axis.text.x = element_text(angle = 45, size = 5)) +
  xlab("") +
  ylab("Percentage of sequenced specimens") +
  ggtitle("Covid-19 Variants in California")
```