

First Year Exam: Bioinformatics Portion

Jack Reddan (PID: A59010543)

2022-07-18

Packages

Load all of the necessary packages

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.1.3
```

```
##
```

```
## 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(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 4.1.3
```

```
##
```

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

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##     date, intersect, setdiff, union
```

Data import

Read in the data from the .csv as a tibble

```
data <- as_tibble(read.csv(file = "covid19_variants.csv", # Specify File
                           sep = ",", # Make sure it reads "," as the separator
                           header = T)) # Indicate the first row is the column names
```

Look at the structure of the data

```
head(data)
```

```
## # A tibble: 6 x 8
```

```
##   date      area  area_type variant_name specimens percentage specimens_7d_avg
```

```
##   <chr>     <chr>  <chr>      <chr>          <dbl>      <dbl>          <dbl>
```

```
## 1 2021-01-01 Calif~ State Alpha 1 1.69 NA
## 2 2021-01-01 Calif~ State Beta 0 0 NA
## 3 2021-01-01 Calif~ State Mu 0 0 NA
## 4 2021-01-01 Calif~ State Gamma 0 0 NA
## 5 2021-01-01 Calif~ State Total 59 100 NA
## 6 2021-01-01 Calif~ State Omicron 1 1.69 NA
## # ... with 1 more variable: percentage_7d_avg <dbl>
```

Data processing

Make dates more workable

```
data <- data %>%
  mutate(date = ymd(date)) # Change the date column from a <chr> to a <date>
```

Subset the data to match the example

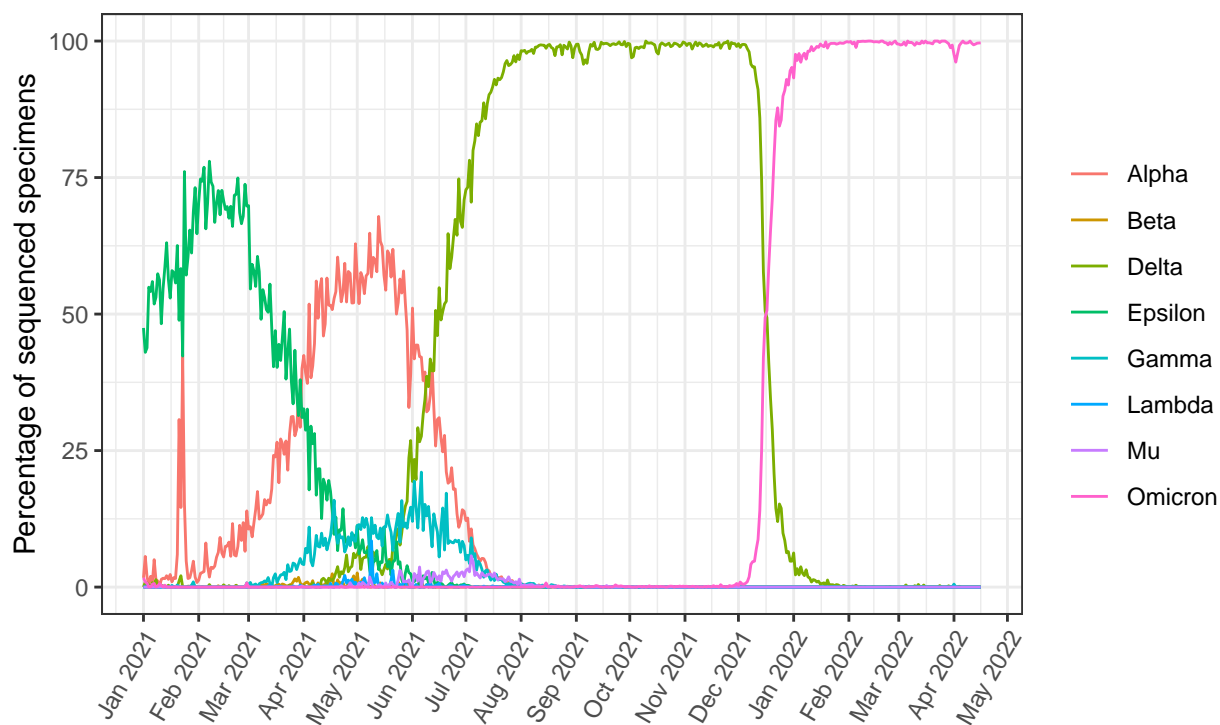
```
data <- data %>%
  filter(date < as.Date("2022-04-17")) %>% # drop late dates
  filter(variant_name != "Other", variant_name != "Total") # drop "other" & "total"
```

Data visulization

Plot the data

```
ggplot(data = data) + # Specify the data as 'data'
  geom_line(aes(x = date, # Plot date on the x-axis
                y = percentage, # Plot percentage for each variant on the y-axis
                color = variant_name)) + # Color and group by variant
  scale_x_date(date_breaks="1 month", # Specify the x-axis tick marks to go month-by-month
               date_labels="%b %Y") + # Specify the format for each month
  guides(x = guide_axis(angle = 60)) + # Angle the x-axis ticks to make more readable
  scale_color_discrete(name = "") + # Remove the legend title
  labs(title = "Covid-19 Variants in California", # Add a main title
       caption = "Data Source: <https://www.cdph.ca.gov>", # Add a caption
       x = "", # Remove the x-axis label
       y = "Percentage of sequenced specimens") + # Clean-up the y-axis label
  theme_bw() # Change theme to match example
```

Covid-19 Variants in California



Data Source: <<https://www.cdph.ca.gov>>

Session information

```
sessionInfo()
```

```
## R version 4.1.2 (2021-11-01)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19044)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] lubridate_1.8.0 ggplot2_3.3.6  dplyr_1.0.9
##
## loaded via a namespace (and not attached):
## [1] highr_0.9      pillar_1.7.0    compiler_4.1.2  tools_4.1.2
## [5] digest_0.6.29  evaluate_0.15    lifecycle_1.0.1  tibble_3.1.7
```

```
## [9] gtable_0.3.0      pkgconfig_2.0.3  rlang_1.0.2      cli_3.3.0
## [13] rstudioapi_0.13   yaml_2.3.5       xfun_0.31         fastmap_1.1.0
## [17] withr_2.5.0       stringr_1.4.0    knitr_1.39        generics_0.1.2
## [21] vctrs_0.4.1       grid_4.1.2       tidyselect_1.1.2 glue_1.6.2
## [25] R6_2.5.1          fansi_1.0.3      rmarkdown_2.14    purrr_0.3.4
## [29] farver_2.1.0      magrittr_2.0.3   scales_1.2.0      ellipsis_0.3.2
## [33] htmltools_0.5.2   colorspace_2.0-3 labeling_0.4.2     utf8_1.2.2
## [37] stringi_1.7.6     munsell_0.5.0    crayon_1.5.1
```