

Raw Data

```
library(tidyverse)
```

Warning: package 'tidyr' was built under R version 4.4.2

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.1      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.1
v purrr      1.0.2
```

```
-- Conflicts ----- tidyverse_conflicts() --
```

```
x dplyr::filter() masks stats::filter()
```

```
x dplyr::lag()     masks stats::lag()
```

```
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(ggplot2)
```

```
library(dplyr)
```

```
library(janitor)
```

Attaching package: 'janitor'

The following objects are masked from 'package:stats':

chisq.test, fisher.test

```
library(lubridate)
```

```
library(here)
```

here() starts at C:/Users/maria/OneDrive/Documents/GitHub/EPI-569-Final-Project

```

rff_2024 <- readRDS("C:/Users/maria/OneDrive/Documents/GitHub/EPI-569-Final-Project/data/rf
roster_2024 <- readRDS("~/GitHub/EPI-569-Final-Project/data/roster_2024.rds")

transmission_pairs_2024 <- readRDS("~/GitHub/EPI-569-Final-Project/data/transmission_pairs_2

setwd("C:/Users/maria/OneDrive/Documents/GitHub/EPI-569-Final-Project/data")

tabyl(rff_2024$`Date_of_Onset`)

```

rff_2024\$Date_of_Onset	n	percent	valid_percent
2024-09-09	1	0.01098901	0.01785714
2024-09-11	1	0.01098901	0.01785714
2024-09-12	3	0.03296703	0.05357143
2024-09-13	1	0.01098901	0.01785714
2024-09-14	3	0.03296703	0.05357143
2024-09-15	3	0.03296703	0.05357143
2024-09-16	1	0.01098901	0.01785714
2024-09-20	4	0.04395604	0.07142857
2024-09-21	3	0.03296703	0.05357143
2024-09-22	1	0.01098901	0.01785714
2024-09-23	4	0.04395604	0.07142857
2024-09-24	1	0.01098901	0.01785714
2024-09-25	2	0.02197802	0.03571429
2024-09-26	1	0.01098901	0.01785714
2024-09-28	2	0.02197802	0.03571429
2024-09-30	2	0.02197802	0.03571429
2024-10-01	6	0.06593407	0.10714286
2024-10-02	6	0.06593407	0.10714286
2024-10-03	5	0.05494505	0.08928571
2024-10-04	2	0.02197802	0.03571429
2024-10-05	1	0.01098901	0.01785714
2024-10-11	2	0.02197802	0.03571429
2024-10-12	1	0.01098901	0.01785714
<NA>	35	0.38461538	NA

```

class(rff_2024$`Date_of_Onset`)

```

```
[1] "Date"
```

```
epicurveonset <- rff_2024 %>%
  filter(!is.na(`Date_of_Onset`)) %>% arrange(`Date_of_Onset`)

epicurveonset$Date_of_Onset <- as.Date(epicurveonset$Date_of_Onset)

epicurveonset <- epicurveonset[order(epicurveonset$Date_of_Onset), ]

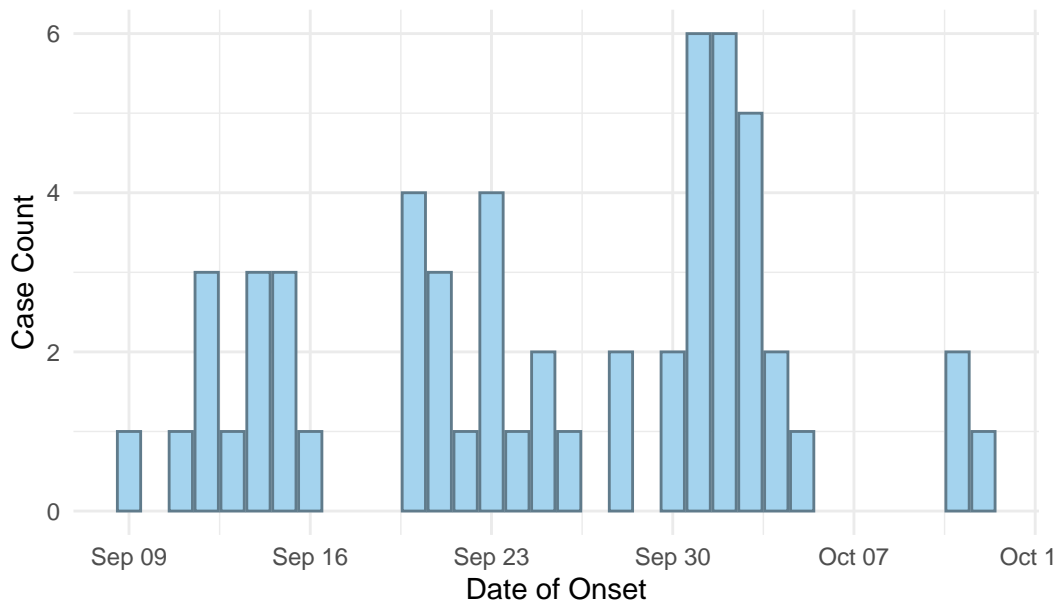
epicurveonset$cumulative_case_count <- cumsum(!is.na(epicurveonset$Date_of_Onset))

head(epicurveonset)
```

```
# A tibble: 6 x 35
  caseID Infectedby Gender Course      If you are in EPI 56~1 `Date of Exposure`
  <dbl>   <dbl> <chr>  <chr>      <chr>                                <date>
1     1      NA Female EPI 569 (C~ Jessica Rothman      2024-09-09
2     2       1 Male   EPI 569 (C~ Jessica Rothman      2024-09-10
3     5       2 Female EPI 569 (C~ Sara Kim                2024-09-12
4     6       5 Female EPI 569 (C~ Charlotte Doran      2024-09-12
5     7       5 Female EPI 569 (C~ Charlotte Doran      2024-09-12
6     4       2 Male   EPI 569 (C~ Jessica Rothman      2024-09-12
# i abbreviated name: 1: `If you are in EPI 569, which TA group are you in?`
# i 29 more variables: `Time of Exposure` <chr>, Date_of_Onset <date>,
#   Date_Time_Onset <chr>, Date_Time_Exposure <chr>,
#   `Did you have symptoms?` <chr>,
#   `How many hours after exposure did you develop symptoms?` <dbl>,
#   `How many hours after your symptom onset did you feel better?` <dbl>,
#   Date_Time_Recovery <dtm>, Date_of_Recovery <date>, Severe <chr>, ...
```

```
ggplot(data = epicurveonset, aes(x = Date_of_Onset)) +
  geom_bar(fill = "lightskyblue2", color = "lightskyblue4") +
  labs(
    x = "Date of Onset",
    y = "Case Count",
    title = "Temporal Distribution of Cases by Date of Onset"
  ) +
  theme_minimal()
```

Temporal Distribution of Cases by Date of Onset



```
sdate <- as.Date(min(rff_2024$Date_of_Onset, na.rm = TRUE))
edate <- as.Date(max(rff_2024$Date_of_Onset, na.rm = TRUE))

sequenciald <- seq.Date(sdate, edate, by = "day")

numb_suceptible <- integer(length = length(sequenciald))
numb_infectious <- integer(length = length(sequenciald))
numb_recovered <- integer(length = length(sequenciald))

total_pop <- nrow(rff_2024)

for (i in seq_along(sequenciald)) {
  current_date <- sequenciald[i]

  numb_infectious[i] <- sum(rff_2024$Date_of_Onset <= current_date &
    (is.na(rff_2024$Date_of_Recovery) | rff_2024$Date_of_Recovery > current_date))

  numb_recovered[i] <- sum(rff_2024$Date_of_Recovery <= current_date, na.rm = TRUE)

  numb_suceptible[i] <- total_pop - numb_infectious[i] - numb_recovered[i]
}
```

```

state_no <- data.frame(
  Date = sequentiald,
  Susceptible = numb_suceptible,
  Infectious = numb_infectious,
  Recovered = numb_recovered
)

state_no_long <- data.frame(
  Date = rep(sequentiald, times = 3),
  State = rep(c("Susceptible", "Infectious", "Recovered"), each = length(sequentiald)),
  Count = c(numb_suceptible, numb_infectious, numb_recovered)
)

library(ggplot2)
ggplot(state_no_long, aes(x = Date, y = Count, color = State, group = State)) +
  geom_line(size = 1) +
  labs(
    title = "Number of Susceptible, Infectious, and Recovered Individuals by Day",
    x = "Date", y = "Count of individuals "
  ) +
  scale_color_manual(
    values = c("Susceptible" = "chocolate3", "Infectious" = "darkslategray", "Recovered" = "darkred")
  ) +
  scale_x_date(
    breaks = scales::date_breaks("2 days"),
    labels = scales::date_format("%b %d")
  ) +
  theme_minimal() +
  theme(
    legend.title = element_blank(),
    axis.text.x = element_text(angle = 45, hjust = 1)
  )

```

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
 i Please use `linewidth` instead.

Number of Susceptible, Infectious, and Recovered Individuals b

