Wei's Netflix Circle

Ella Ngo

2/1/2023

Background

A brief data analyis of my boyfriend (Wei)'s Netflix watch history compared to his sister's, his parents', and his friend's.

Wei has always been into films, movies, shows, etc., so I aimed to put together a graph highlighting how much more things he watches on netflix on average than people in his circle.

I would like to thank my friend Taylor Blair, whose project on the difference in frequencies of watching Netflix between her sister and her inspired this project of mine.

Loading Libraries and Data

```
library(tidyverse)
library(openair)
library(readr)
```

```
wei_history <- read_csv("~/Desktop/R/NetflixViewingHistory.csv")
isa_history <- read_csv("~/Desktop/R/NetflixViewingHistory-2.csv")
parent_history <- read_csv("~/Desktop/R/ParentNetflixHistory.csv")
friend_history <- read_csv("~/Desktop/R/FriendNetflixHistory.csv")
```

Data Prep

As can be seen above, the initial dataset contains only two variables.

Date fixing

```
wei_history$Date <- as.Date(wei_history$Date,"%m/%d/%y")
isa_history$Date <- as.Date(isa_history$Date,"%m/%d/%y")
parent_history$Date <- as.Date(parent_history$Date,"%m/%d/%y")
friend_history$Date <- as.Date(friend_history$Date,"%m/%d/%y")
```

Movie or TV series

```
wei_history$type <- grepl(":", wei_history$Title)
wei_history$type <- wei_history$type %>%
  replace(wei_history$type==TRUE, "TV Series") %>%
  replace(wei_history$type==FALSE, "Movie")
```

```
isa_history$type <- grepl(":", isa_history$Title)
isa_history$type <- isa_history$type %>%
  replace(isa_history$type==TRUE, "TV Series") %>%
  replace(isa_history$type==FALSE, "Movie")
parent_history$type <- grepl(":", parent_history$Title)
parent_history$type <- parent_history$type %>%
  replace(parent_history$type==TRUE, "TV Series") %>%
  replace(parent_history$type==FALSE, "Movie")
friend_history$type <- grepl(":", friend_history$Title)
friend_history$type <- friend_history$type %>%
  replace(friend_history$type==TRUE, "TV Series") %>%
  replace(friend_history$type==FALSE, "Movie")
```

Series, Season and episode

```
wei_history <- wei_history %>%
  separate(Title,
           c("Series", "Season", "Episode"),
           ": ")
isa_history <- isa_history %>%
  separate(Title,
           c("Series", "Season", "Episode"),
           ": ")
parent_history <- parent_history %>%
  separate(Title,
           c("Series", "Season", "Episode"),
           ": ")
friend_history <- friend_history %>%
  separate(Title,
           c("Series", "Season", "Episode"),
           ": ")
```

Day of Week

```
day_of_week <- function(x){
  return (weekdays(x))
}
days <- c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday")
wei_history$day <- sapply(wei_history$Date, day_of_week)
wei_history$day <- factor(wei_history$day,
                          levels = days)
isa_history$day <- sapply(isa_history$Date, day_of_week)
isa_history$day <- factor(isa_history$day,
                                levels = days)
parent_history$day <- sapply(parent_history$Date, day_of_week)
parent_history$day <- factor(parent_history$day,
                          levels = days)
friend_history$day <- sapply(friend_history$Date, day_of_week)
friend_history$day <- factor(friend_history$day,
                                levels = days)
```

Outputted Tibble

```
sample_n(wei_history, 4)
```

```
## # A tibble: 4 x 6
##   Series          Season            Episode         Date       type      day
##   <chr>           <chr>             <chr>           <date>     <chr>     <fct>
## 1 BoJack Horseman Season 2          Still Broken    2019-02-01 TV Series Friday
## 2 DEATH NOTE      Death Note        Glare           2022-05-05 TV Series Thursd~
## 3 Paradise PD     Part 1            Ass on the Line 2018-09-03 TV Series Monday
## 4 Spartacus       Gods of the Arena The Bitter End  2018-01-05 TV Series Friday
```

#Compare dataset

```
wei_history$individual <- "Wei"
isa_history$individual <- "Isa"
friend_history$individual <- "Parent"
parent_history$individual <- "Friend"
merged <- rbind(wei_history, isa_history, parent_history, friend_history)
```

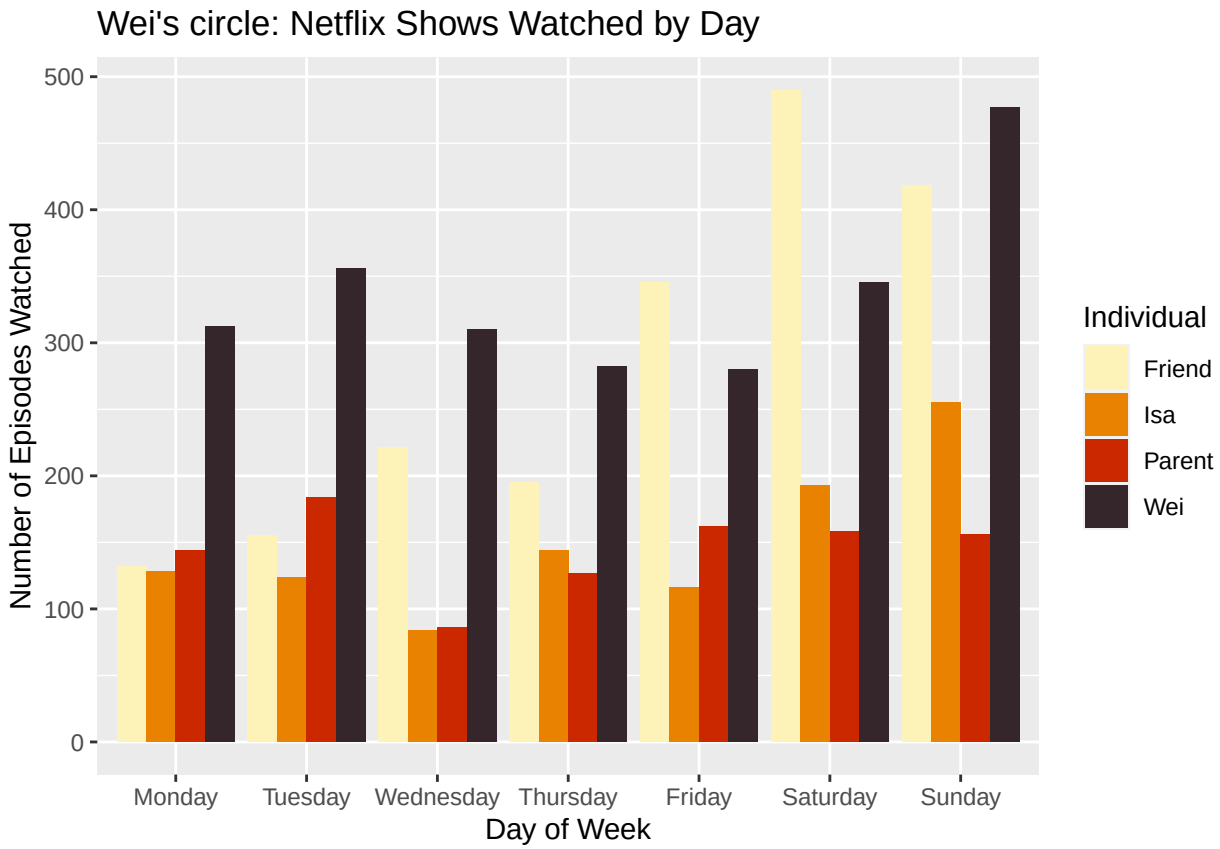
Numerical analysis

Binge Watching

Everyone has binged watched Netflix at some point during 2021, unfortunately :(.

Graphical Analysis

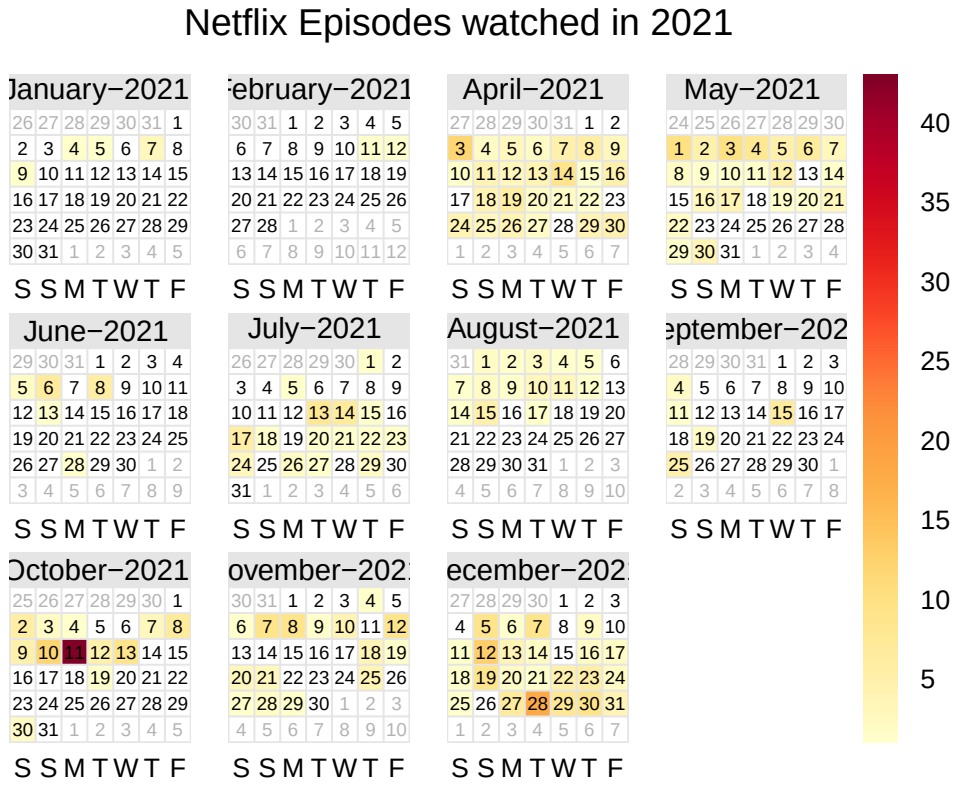
```
ggplot(merged, aes(x = day, fill=individual) ) +
  geom_bar(position = "dodge") +
  labs(title = "Wei's circle: Netflix Shows Watched by Day") +
  scale_fill_manual("Individual", values=c("#fdf2b8","#e88200", "#cb2800", "#34262b")) +
  xlab("Day of Week") +
  ylab("Number of Episodes Watched")
```



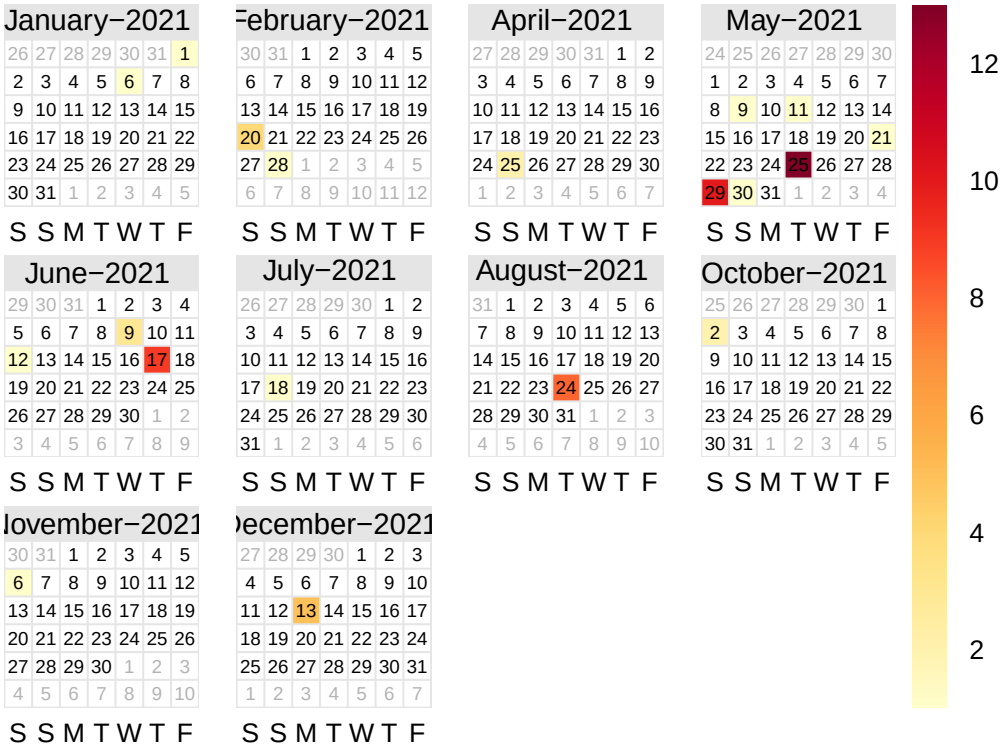
A Calander of Events

2021 was a unique year, with Wei going to college in September, so it's weird that he *might* have watched more Netflix then....

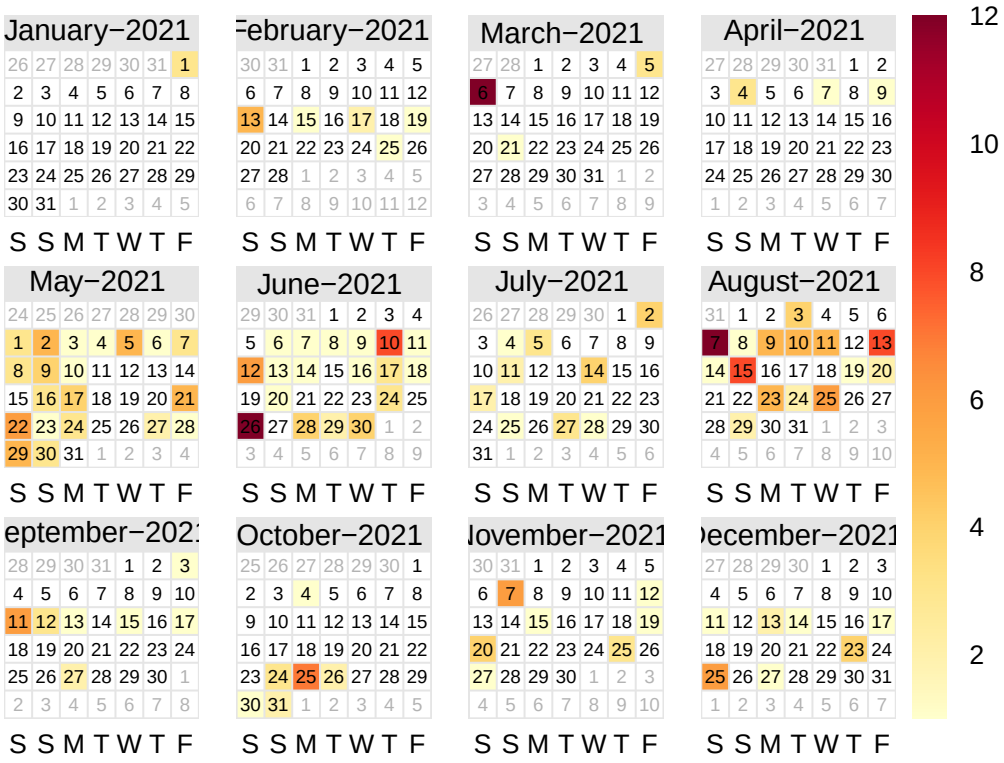
```
wei_history %>%
  group_by(Date) %>%
  summarise(count = n()) %>%
  rename(date=Date) %>%
  calendarPlot(pollutant = "count",
               year = 2021, main="Netflix Episodes watched in 2021")
```




```
isa_history %>%
  group_by(Date) %>%
  summarise(count = n()) %>%
  rename(date=Date) %>%
  calendarPlot(pollutant = "count",
               year = 2021,
               month = c(1:12))
```




```
parent_history %>%
  group_by(Date) %>%
  summarise(count = n()) %>%
  rename(date=Date) %>%
  calendarPlot(pollutant = "count",
               year = 2021,
               month = c(1:12))
```




```
friend_history %>%
  group_by(Date) %>%
  summarise(count = n()) %>%
  rename(date=Date) %>%
  calendarPlot(pollutant = "count",
               year = 2021,
               month = c(1:12))
```