Google fit step data 2021-2024

Joel F. Swift

09/12/2022

```
steps.df <- read.csv("Daily activity metrics.csv", header = TRUE)
data.frame(steps.df$Step.count, steps.df$Date)[1:10,]
```

Load data

```
##    steps.df.Step.count steps.df.Date
## 1                  948    2021-04-26
## 2                 7183    2021-04-27
## 3                 1150    2021-04-28
## 4                  318    2021-04-29
## 5                 1160    2021-04-30
## 6                 4697    2021-05-01
## 7                 3264    2021-05-02
## 8                 8540    2021-05-03
## 9                 4966    2021-05-04
## 10                1349    2021-05-05
```

```
steps.df$Date <- ymd(steps.df$Date)
steps.df$day_of_week <- wday(steps.df$Date, label = TRUE)
steps.df$month <- month(steps.df$Date, label = TRUE)
steps.df$year <- year(steps.df$Date)

# check that things look okay
tapply(steps.df$month, steps.df$year, summary)
```

Formatting dates with lubridate

```
## $`2021`
## Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
##   0   0   0   5  31  30  31  31  30  31  30  31
##
## $`2022`
## Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
##  31  28  31  30  31  30  31  31  30  31  30  31
##
## $`2023`
## Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
##  31  28  31  30  31  30  31  31  30  31  30  31
##
## $`2024`
## Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
##  31  29   2   0   0   0   0   0   0   0   0   0
```

```
#tapply(steps.df$day_of_week, steps.df$month, summary)
```

```
levels(steps.df$day_of_week)
```

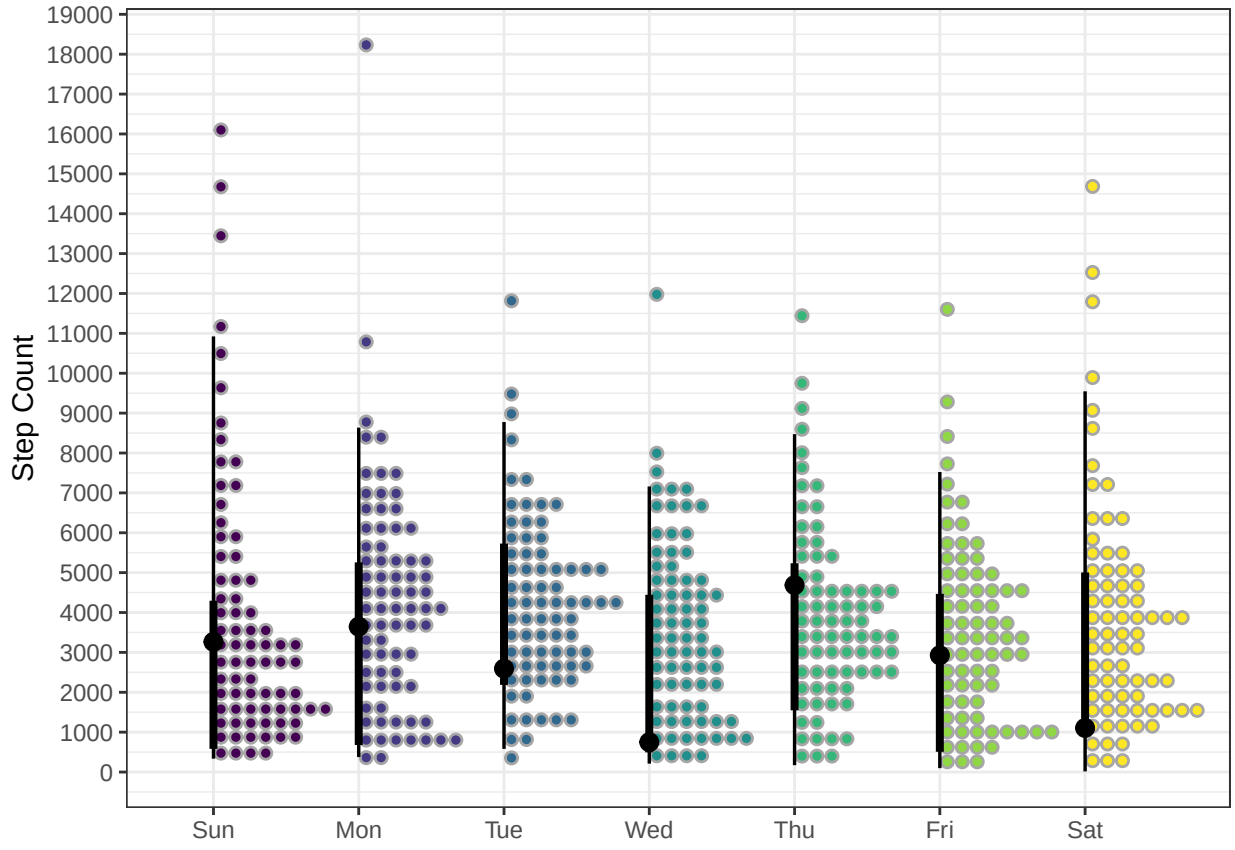
Quick plots

```
## [1] "Sun" "Mon" "Tue" "Wed" "Thu" "Fri" "Sat"
```

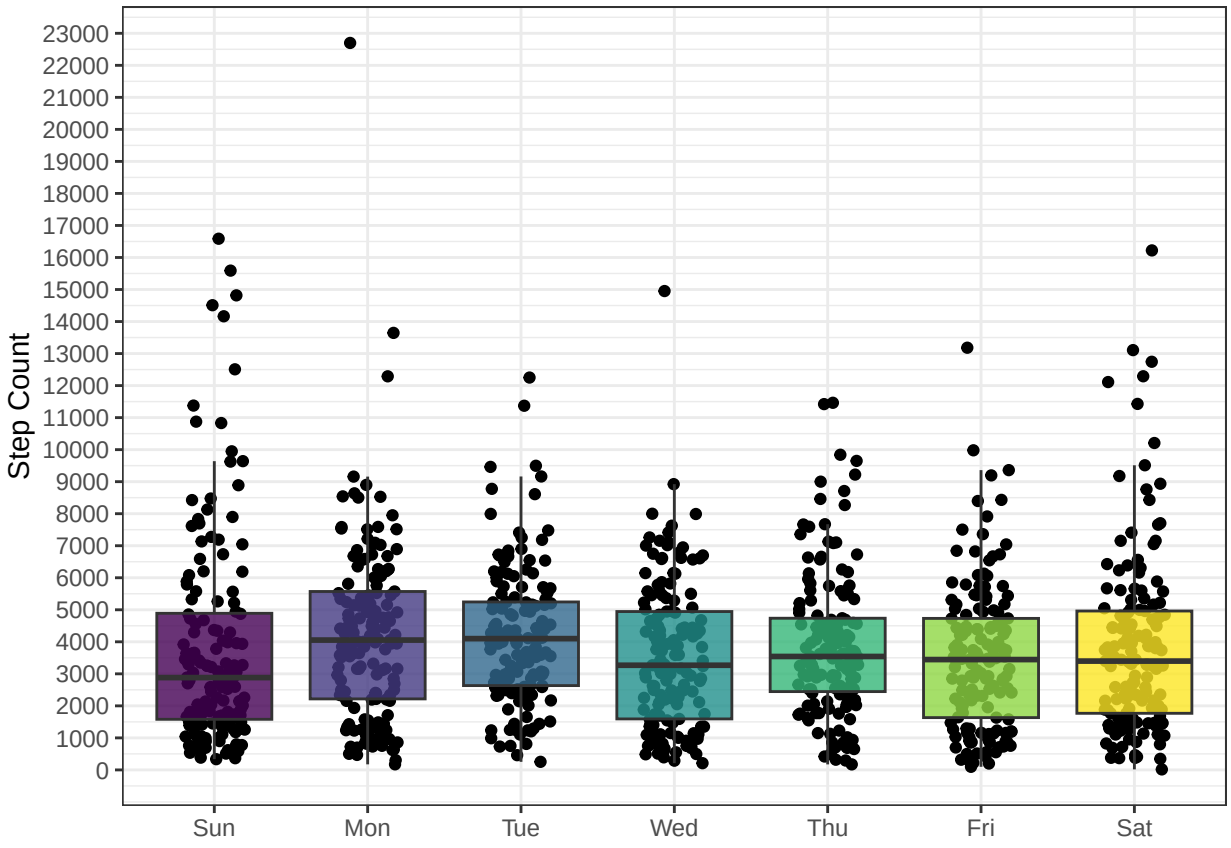
```
a <- ggplot(steps.df,aes(x= day_of_week, y = Step.count, fill=day_of_week)) +
  ggdist::stat_dotsinterval(quantiles = 75, point_interval = "mode_hdci") +
  ylab("Step Count") + xlab(NULL) + theme(legend.position = "none") +
  scale_y_continuous(breaks = seq(0, 20000, by = 1000))
b <- ggplot(steps.df, aes(x = day_of_week , y = Step.count, fill = day_of_week)) +
  geom_jitter(width = 0.2) + geom_boxplot(outlier.shape = NA, alpha = 0.8) +
  scale_y_continuous(breaks = seq(0, 25000, by = 1000)) + ylab("Step Count") +
  xlab(NULL) + theme(legend.position = "none")
c <- ggplot(steps.df, aes(x = Date , y = Step.count))  + geom_point() +
  geom_smooth(color = "green") + scale_y_continuous(breaks = seq(0, 25000, by = 1000)) +
  ylab("Step Count") + xlab(NULL)
d <- ggplot(steps.df, aes(x = month , y = Step.count)) + geom_jitter(width = 0.2) +
  geom_boxplot(outlier.shape = NA, alpha = 0.8) + scale_y_continuous(breaks = seq(0, 25000, by = 1000)) +
  ylab("Step Count")
e <- ggarrange(a,b,c,d, nrow = 2, ncol = 2, align = 'hv')
```

```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

```
ggsave("Step count by month, day, and date.png", e , height = 8, width = 12)
a
```

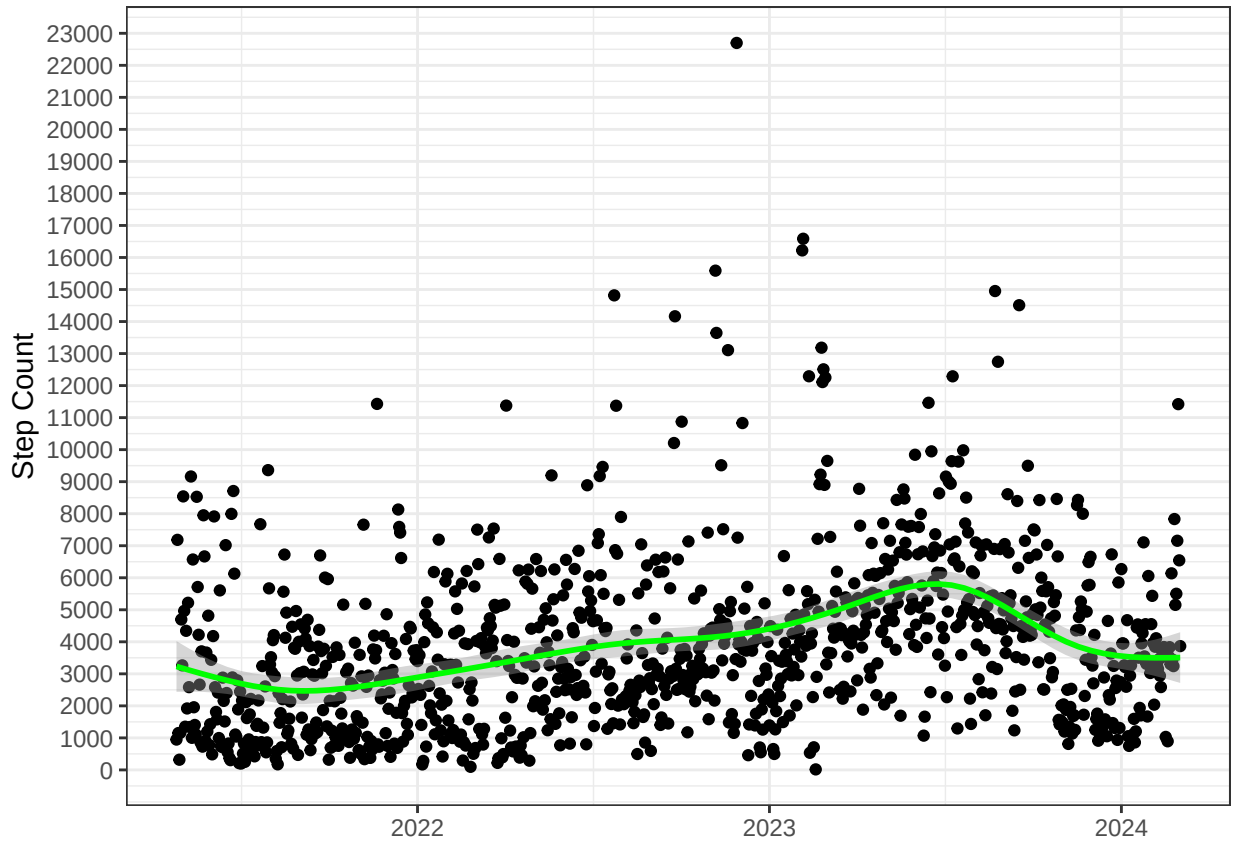

Step Count
19000
18000
17000
16000
15000
14000
13000
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Sun
Mon
Tue
Wed
Thu
Fri
Sat

b

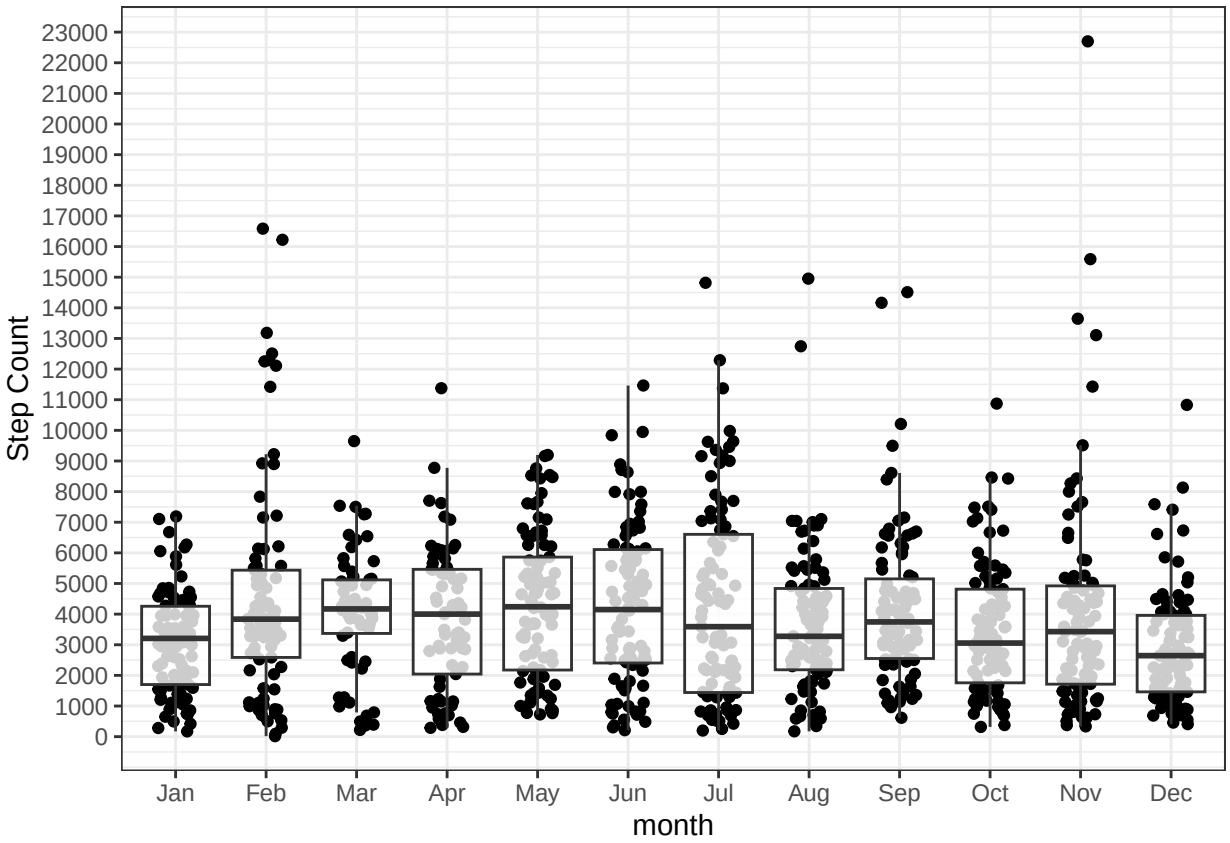



```
c
```

```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```


Step Count
23000
22000
21000
20000
19000
18000
17000
16000
15000
14000
13000
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
2022
2023
2024

```
d
```


Step Count
23000
22000
21000
20000
19000
18000
17000
16000
15000
14000
13000
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
month