

# REDCap User Lifecycle

Philip Chase

2022-07-08

The report is an example of how you could characterize your REDCap users using data from the redcap\_user\_information. The data is easy to access using a few R packages that manage configuration data (dotenv), database connections (DBI and redcapcustodian), dates (lubridate), and data transformation (tidyverse)

```
library(dotenv)
library(DBI)
library(redcapcustodian)
library(lubridate)
library(tidyverse)

rc_conn <- connect_to_redcap_db()

# get the data
redcap_user_information <- tbl(rc_conn, "redcap_user_information")
```

We need to select the columns we care about and transform them to meet our needs:

```
user_life_cycle <- redcap_user_information %>%
  select(username, user_creation, user_lastlogin, user_suspended_time) %>%
  collect() %>%
  filter(!username %in% c("site_admin", "master")) %>% # these accounts are junk
  mutate(
    lifespan = if_else(is.na(user_suspended_time),
                      today() - as.Date(user_creation),
                      user_suspended_time - user_creation),
    dormancy = today() - as.Date(user_lastlogin),
    creation = as.Date(user_creation),
    lastlogin = as.Date(user_lastlogin),
    suspension = as.Date(user_suspended_time)
  ) %>%
  select(
    lifespan,
    dormancy,
    creation,
    lastlogin,
    suspension
  )
```

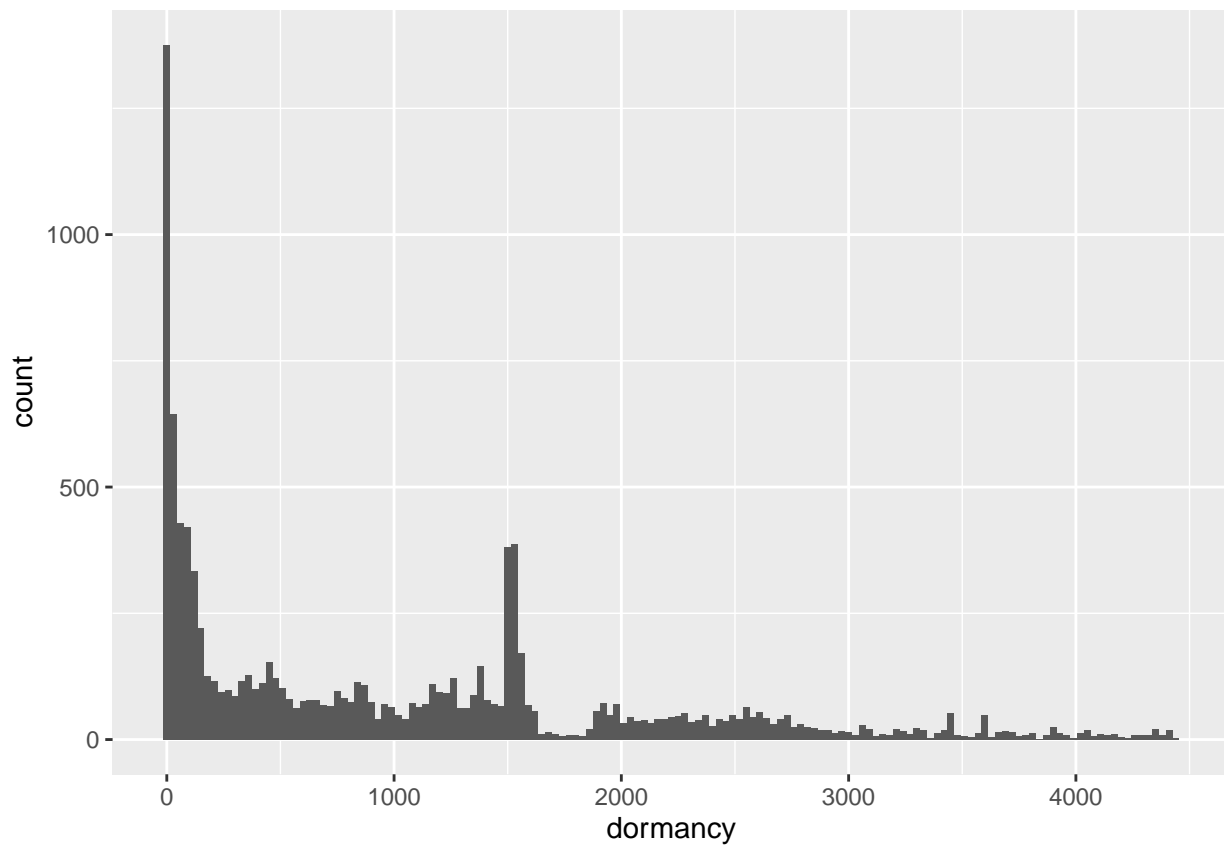


Figure 1: This plot shows how long user accounts have remained dormant with no login activity. With the binwidth set to 30 days, each bar represents about one month. Note the large cohort of active accounts represented by the first 6 bars AKA 180 days. The spikes about 1500 days ago were caused when we implemented a REDCap patch to the our single-sign-on system and updated a lot of empty values.

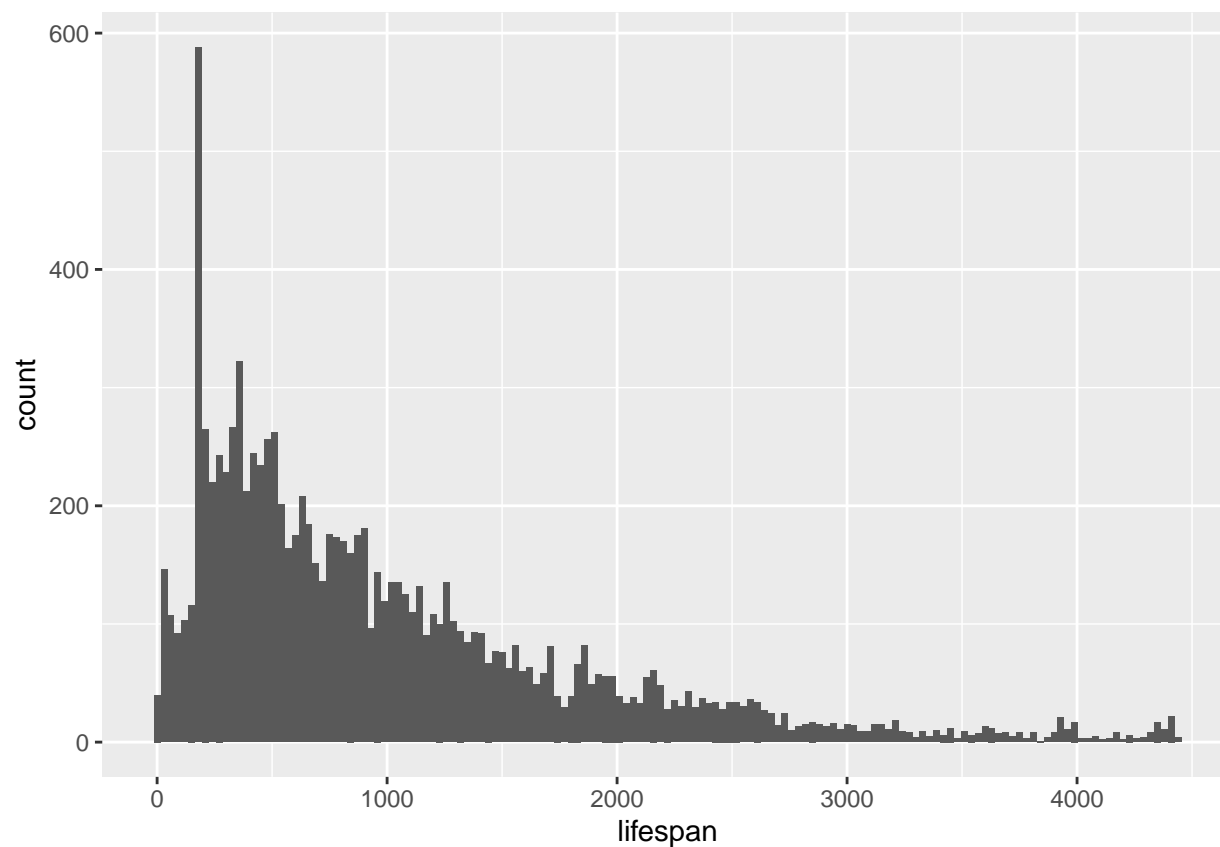


Figure 2: This plot shows the lifespan of the redcap accounts on the system where lifespan is expressed in days. Note: the spike in the 7th month. We have a policy of expiring REDCap accounts after 180 days of inactivity. The spike is probably showing REDCap users that signed up, played for a few days and then left never to return.

For the last plot we want to see three events in a REDCap account's lifecycle over time. To do that, we need to pivot the data, combining the dates of those events into a single column `event_date`. We use the column names to populate the `event` column which will serve as a grouping variable and label for the dates in `event_date`.

```
lifecycle <- user_life_cycle %>%
  pivot_longer(
    cols = c("creation",
             "lastlogin",
             "suspension"
            ),
    names_to = "event",
    values_to = "event_date"
  )
```

Now the data can be easily plotted.

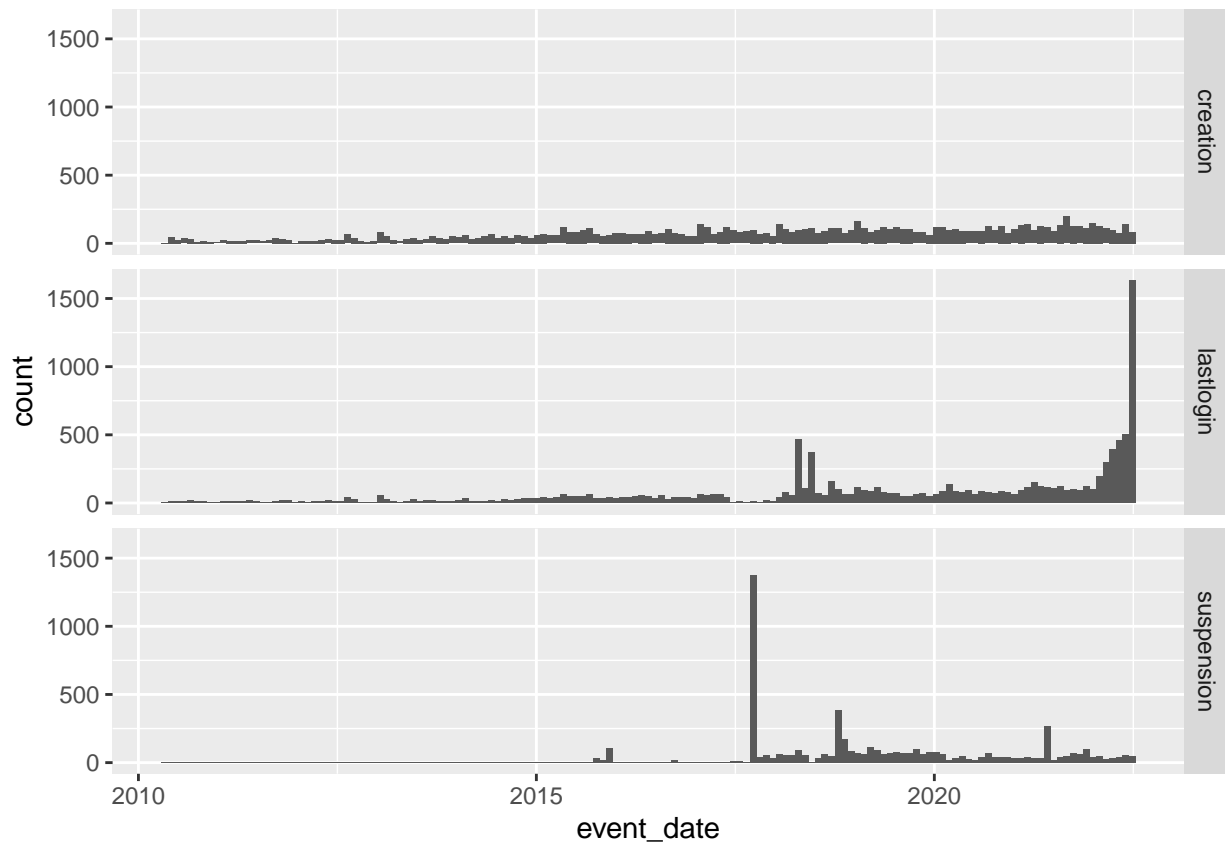


Figure 3: Note the account creation rate is steadily growing since system creation in 2010. The last login graph tells a similar story as the dormancy graph. The suspension graph shows a large spike of activity when we implemented our suspension-for-inactivity policy. Smaller spikes show the result of occasional account cleanup activity.