It’s just some basic data manipulation with some heavy ggplot2 styling customization, so no real need for exposition beyond noting that there are many other ways to view the data. I just settled on centered segments early on and went from there. If you do a bit of gg-doodling yourself, drop a note in the comments with a link.

You can see a full-size version of the image via [this link](https://rud.is/b/wp-content/uploads/2018/07/hibp-lines.png).

library(hrbrthemes) # use github or gitlab version

library(tidyverse)

# get the data

dat\_url <- "https://raw.githubusercontent.com/themains/pwned/master/data/breaches.json"

jsonlite::fromJSON(dat\_url) %>%

mutate(BreachDate = as.Date(BreachDate)) %>%

tbl\_df() -> breaches

# selected breach labels df

group\_by(breaches, year = lubridate::year(BreachDate)) %>%

top\_n(1, wt=PwnCount) %>%

ungroup() %>%

filter(year %in% c(2008, 2015, 2016, 2017)) %>% # pick years where labels will fit nicely

mutate(

lab = sprintf("%s\n%sM accounts", Name, as.integer(PwnCount/1000000))

) %>%

arrange(year) -> labs

# num of known breaches in that year for labels

count(breaches, year = lubridate::year(BreachDate)) %>%

mutate(nlab = sprintf("n=%s", n)) %>%

mutate(lab\_x = as.Date(sprintf("%s-07-02", year))) -> year\_cts

mutate(breaches, p\_half = PwnCount/2) %>% # for centered segments

ggplot() +

geom\_segment( # centered segments

aes(BreachDate, p\_half, xend=BreachDate, yend=-p\_half),

color = ft\_cols$yellow, size = 0.3

) +

geom\_text( # selected breach labels

data = labs, aes(BreachDate, PwnCount/2, label = lab),

lineheight = 0.875, size = 3.25, family = font\_rc,

hjust = c(0, 1, 1, 0), vjust = 1, nudge\_x = c(25, -25, -25, 25),

nudge\_y = 0, color = ft\_cols$slate

) +

geom\_text( # top year labels

data = year\_cts, aes(lab\_x, Inf, label = year), family = font\_rc,

size = 4, vjust = 1, lineheight = 0.875, color = ft\_cols$gray

) +

geom\_text( # bottom known breach count totals

data = year\_cts, aes(lab\_x, -Inf, label = nlab, size = n),

vjust = 0, lineheight = 0.875, color = ft\_cols$peach,

family = font\_rc, show.legend = FALSE

) +

scale\_x\_date( # break on year

name = NULL, date\_breaks = "1 year", date\_labels = "%Y"

) +

scale\_y\_comma(name = NULL, limits = c(-450000000, 450000000)) + # make room for labels

scale\_size\_continuous(range = c(3, 4.5)) + # tolerable font sizes

labs(

title = "HIBP (Known) Breach Frequency & Size",

subtitle = "Segment length is number of accounts; n=# known account breaches that year",

caption = "Source: HIBP via "

) +

theme\_ft\_rc(grid="X") +

theme(axis.text.y = element\_blank()) +

theme(axis.text.x = element\_blank())