Ben F orgers

Ian D. Gow

28 April 2024

```r
library(tidyverse)
```

```r
options(HTTPUserAgent = "iandgow@gmail.com")

t <- "FirmFilings.zip"
url <- "https://pcaobus.org/assets/PCAOBFiles/FirmFilings.zip"
path <- file.path("data", t)

if (!file.exists(path)) {
  download.file(url, path)
}
```

```r
fix_names <- function(names) {
  names <- tolower(gsub("\\s+", "_", names))
  names <- gsub("<", "lt", names)
  names <- gsub(">", "gt", names)
  names <- gsub("[)(_]+", "_", names)
  names
}

form_aps <-
  read_csv(path, guess_max = Inf, show_col_types = FALSE) |>
  rename_with(fix_names)

combine_names <- function(names) {
  names <- str_replace_na(names, "")
  names <- str_c(names, collapse = " ")
  names <- str_replace_all(names, "\\s+", " ")
  names
}
```

```r
most_common_name <-
  form_aps |>
  count(engagement_partner_id,
        engagement_partner_last_name,
        engagement_partner_middle_name,
        engagement_partner_first_name,
        name = "n_forms") |>
    group_by(engagement_partner_id) |>
    arrange(desc(n_forms)) |>
  filter(row_number() == 1) |>
  rowwise() |>
  mutate(engagement_partner_name =
           combine_names(c(engagement_partner_first_name,
                           engagement_partner_middle_name,
                           engagement_partner_last_name))) |>
  select(engagement_partner_id, engagement_partner_name)

names_df <-
  form_aps |>
  distinct(engagement_partner_id,
           engagement_partner_last_name,
           engagement_partner_middle_name,
           engagement_partner_first_name) |>
  count(engagement_partner_id, name = "n_names") |>
  inner_join(most_common_name, by = "engagement_partner_id")

names_df |>
  filter(n_names >= 6) |>
  arrange(desc(n_names))
```

```
# A tibble: 13 x 3
   engagement_partner_id n_names engagement_partner_name
   <chr>                   <int> <chr>
 1 0504100001                 14 Ben F Borgers
 2 0596811101                  9 OLAYINKA TEMITOPE OYEBOLA
 3 0271600002                  8 Yong Yun Lee
 4 0616700001                  8 Thomas Michael O'Neal
 5 0053600002                  7 Louis V Esposito
 6 0625500001                  7 kian kok wong
 7 0028700001                  6 Derek Webb
 8 0045763853                  6 Richard J Fleischman
 9 0057200001                  6 Corey Eric Fischer
10 0117100608                  6 Patrick Wong
```

```
11 0611700001               6 Kristofer Heaton
12 0619707588               6 Jaslyn Huynh
13 0626800001               6 SAU JONG LIM
```

```
names_df |>
  count(n_names, name = "count") |>
  ggplot(aes(x = n_names, y = count)) +
  geom_col()
```
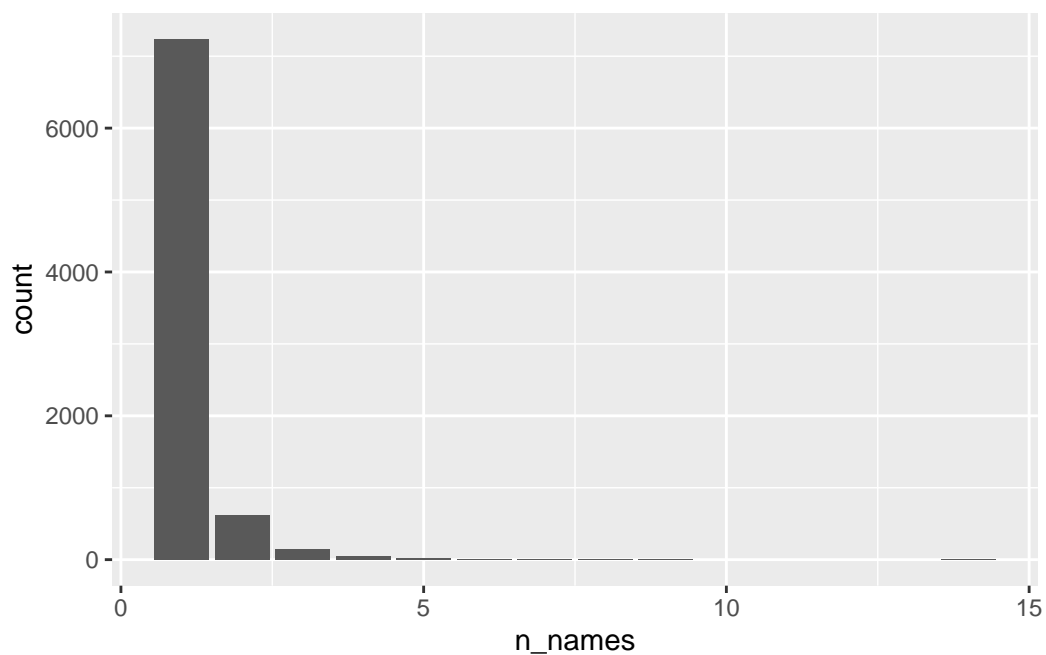

Figure 1: Distribution of auditors by number of reported spellings

```
names_df |>
  count(n_names, name = "count") |>
  ggplot(aes(x = n_names, y = count)) +
  geom_col() +
  scale_y_continuous(transform = "log1p")
```
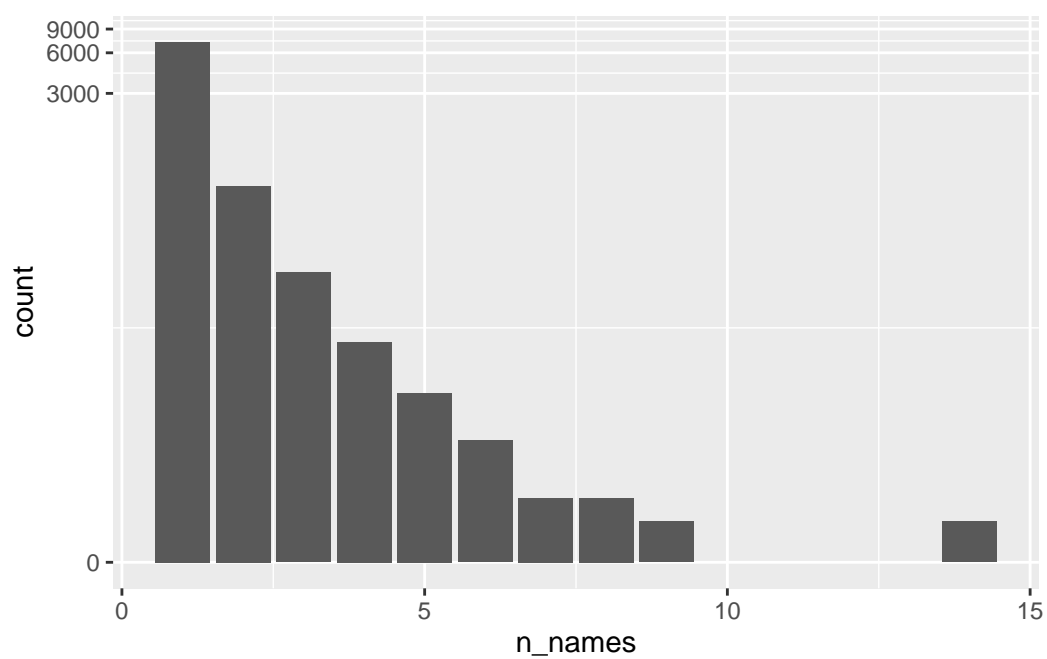
Figure 2: Distribution of auditors by number of reported spellings $(log(1 + y))$