Try feather format

Julian Barg

March 16, 2019

Read feather

```
library(feather)
sample <- read_feather('sample 2019-03-16.feather')
```

```
## Warning: Coercing int64 to double
```

Explore data

How many year-executive observations do we have per company?

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------------- tidyverse 1.2.
## v ggplot2 3.1.0      v purrr   0.3.0
## v tibble  2.0.1      v dplyr   0.8.0.1
## v tidyr   0.8.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0

## -- Conflicts ------------------------------------------------------------------ tidyverse_conflicts(
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```
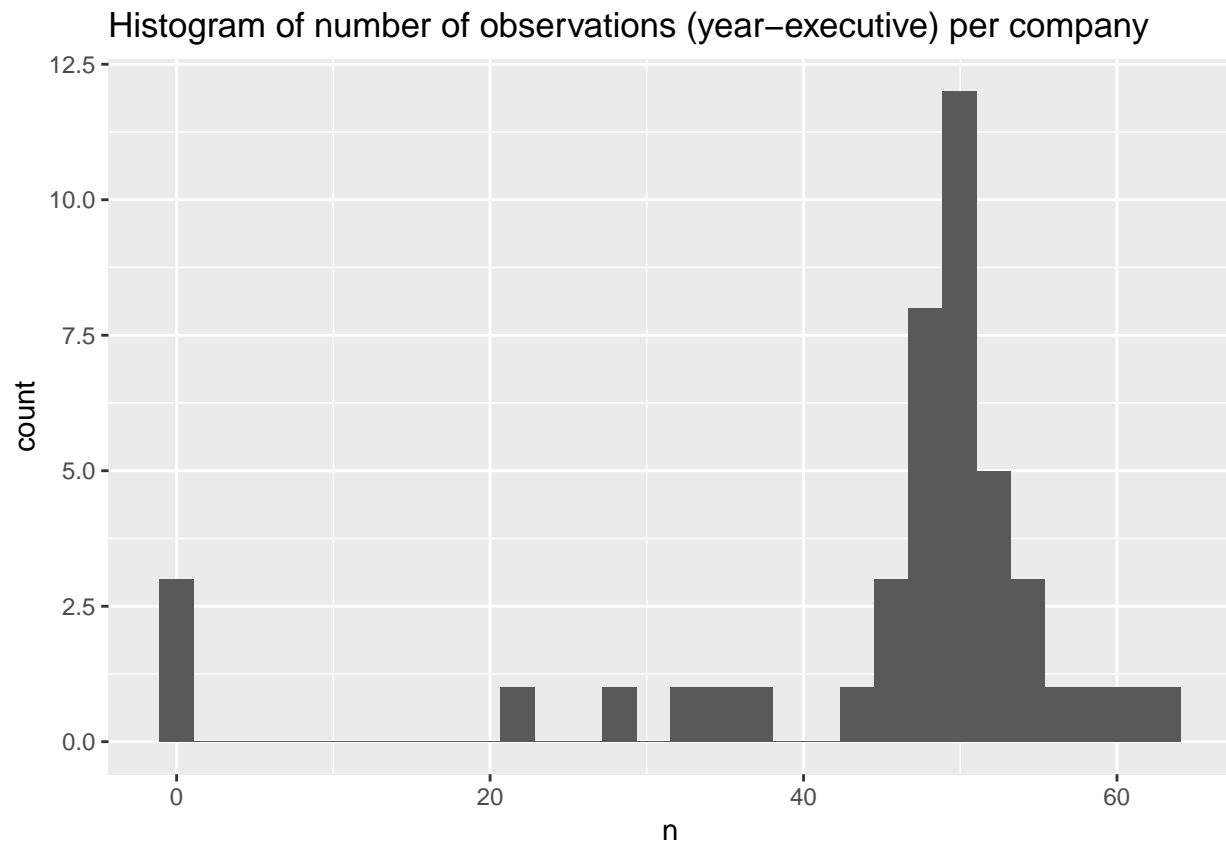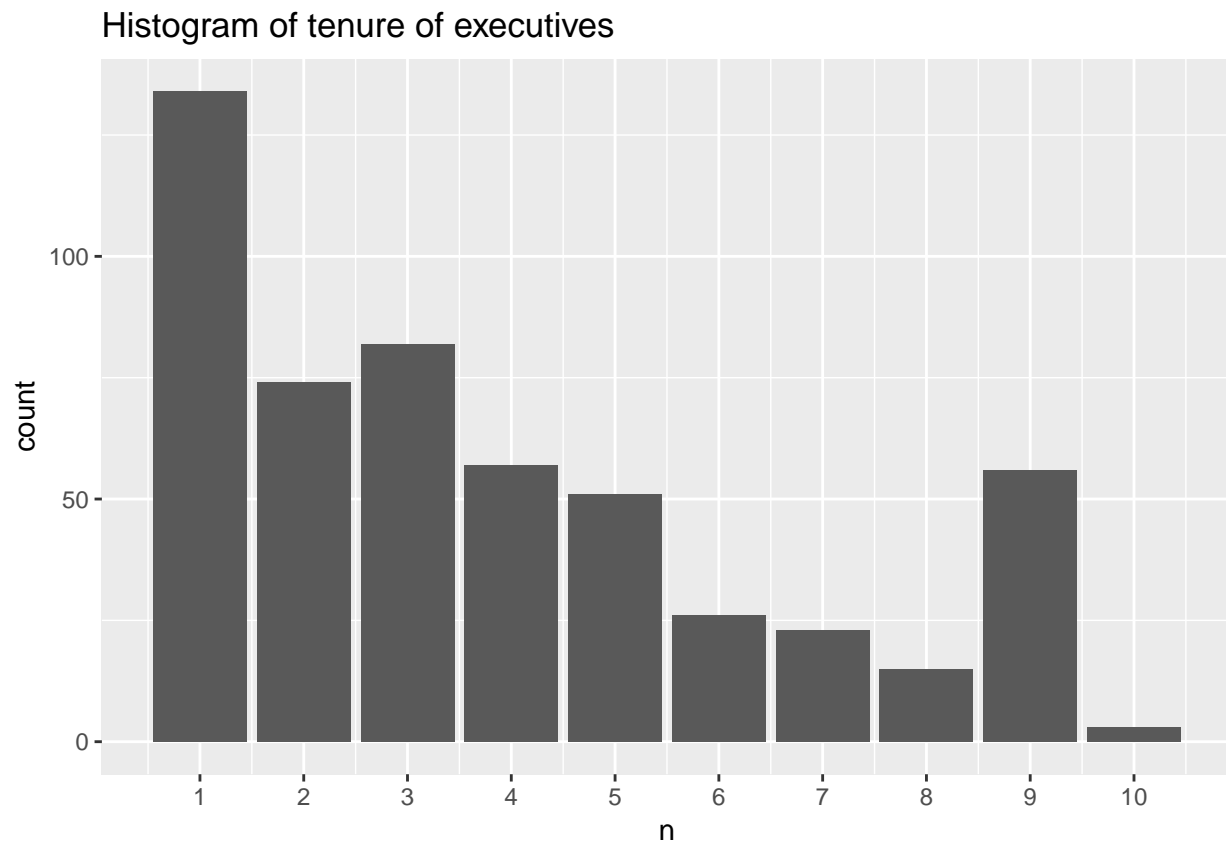
```
sample %>%
  group_by(gvkey) %>%
  summarize(n = n()) %>%
  ggplot(aes(x=n)) +
  geom_histogram() +
  labs(title='Histogram of number of observations (year-executive) per company')
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Histogram of number of observations (year−executive) per company

What's the tenure of the executives?

```r
sample %>%
  group_by(execid) %>%
  summarize(n = n()) %>%
  ggplot(aes(x=n)) +
  geom_bar() +
  labs(title='Histogram of tenure of executives') +
  scale_x_continuous(breaks = seq(0, 10, by=1))
```

Histogram of tenure of executives

How many observations are there for every year? Are observations more complete for recent years?

```
sample %>%
  ggplot(aes(x=year)) +
  geom_bar() +
    labs(title='Observations (executives) per year') +
  scale_x_continuous(breaks = seq(2009, 2018, by=1))
```

```
## Warning: Removed 3 rows containing non-finite values (stat_count).
```

Observations (executives) per year