

one-worded-vs-multi-worded

Here I looked at how company's name effects its popularity on LinkedIn. I found that companies with a single word in their name have about 4 times more followers on average. Therefore, I suggest that companies with names like e.g. "Liberty Tax Service" use abbreviation: "LTS"; Companies with names like "Ashland Inc" should simply use Ashland.

```
linkedin<-read.csv("C:\\Users\\irakl\\Desktop\\temp_datalab_records_linkedin_company\\linkedin.csv", header = TRUE)
```

```
# work with most recent data
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
linkedin_recent <-
  linkedin %>%
  group_by(company_name) %>%
  filter(as_of_date == max(as_of_date))
```

```
# produce a column that indicates whether company name is one-worded or multi-worded.
linkedin_recent %>%
  select(company_name,followers_count,industry) %>%
  mutate(
    multi_worded_name = sapply(strsplit(company_name, " "), length) != 1 # e.g. Hewlett-Packard
    is considered one-worded with this code
  ) ->
  name_role
```

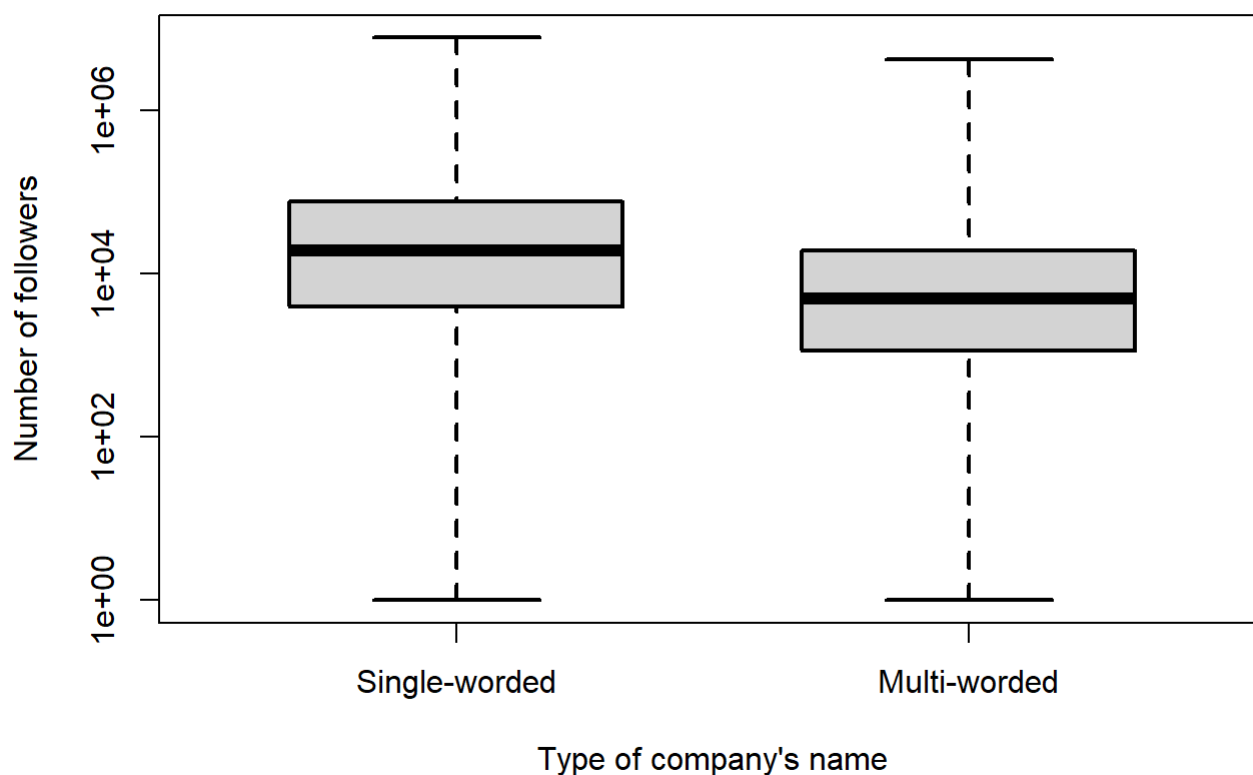
Following analysis shows that companies with single-worded name tend to attract few hundred percent more followers.

```

name_role_plus <- name_role
name_role_plus$followers_count <- name_role_plus$followers_count + 1 # added 1 follower to every
one to plot on log scale
boxplot(followers_count ~ multi_worded_name, data = name_role_plus, boxwex=0.65,lwd=2,
log = "y", range = 0, names=c("Single-worded","Multi-worded"), xlab = "Type of company's name",
ylab = "Number of followers", main = "Dependence of company popularity on its name")

```

Dependence of company popularity on its name



```

# produce follower's vector for companies with multi-worded names
name_role %>%
  filter(multi_worded_name == TRUE) %>%
  ungroup() %>%
  select(followers_count) ->
  multi_worded

# produce follower's vector for companies with single-worded names
name_role %>%
  filter(multi_worded_name == FALSE) %>%
  ungroup() %>%
  select(followers_count) ->
  single_worded

```

```

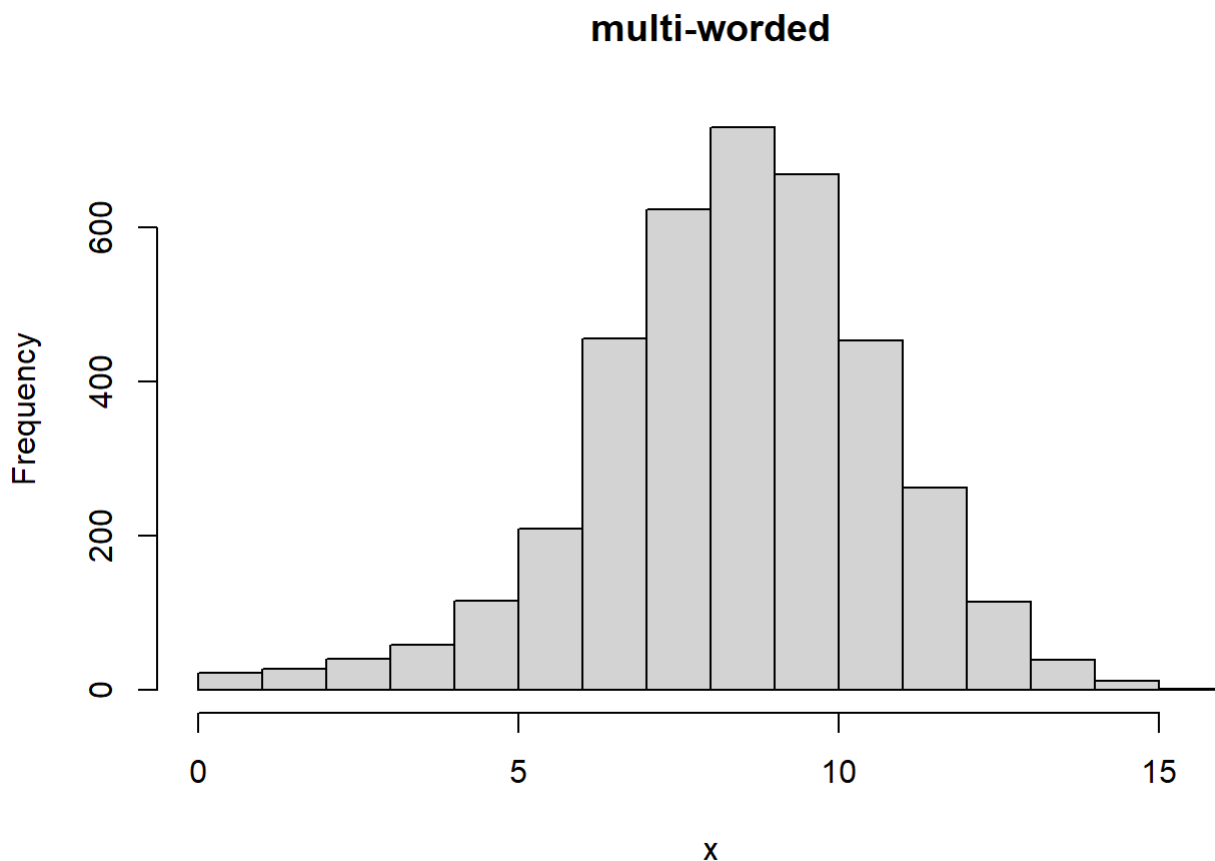
ratio <- sapply(single_worded, mean, na.rm = TRUE)/sapply(multi_worded, mean, na.rm = TRUE)
rt = toString(round(ratio))
paste("Companies with names consisting of a single word have about",rt," times more followers")

```

```
## [1] "Companies with names consisting of a single word have about 4 times more followers"
```

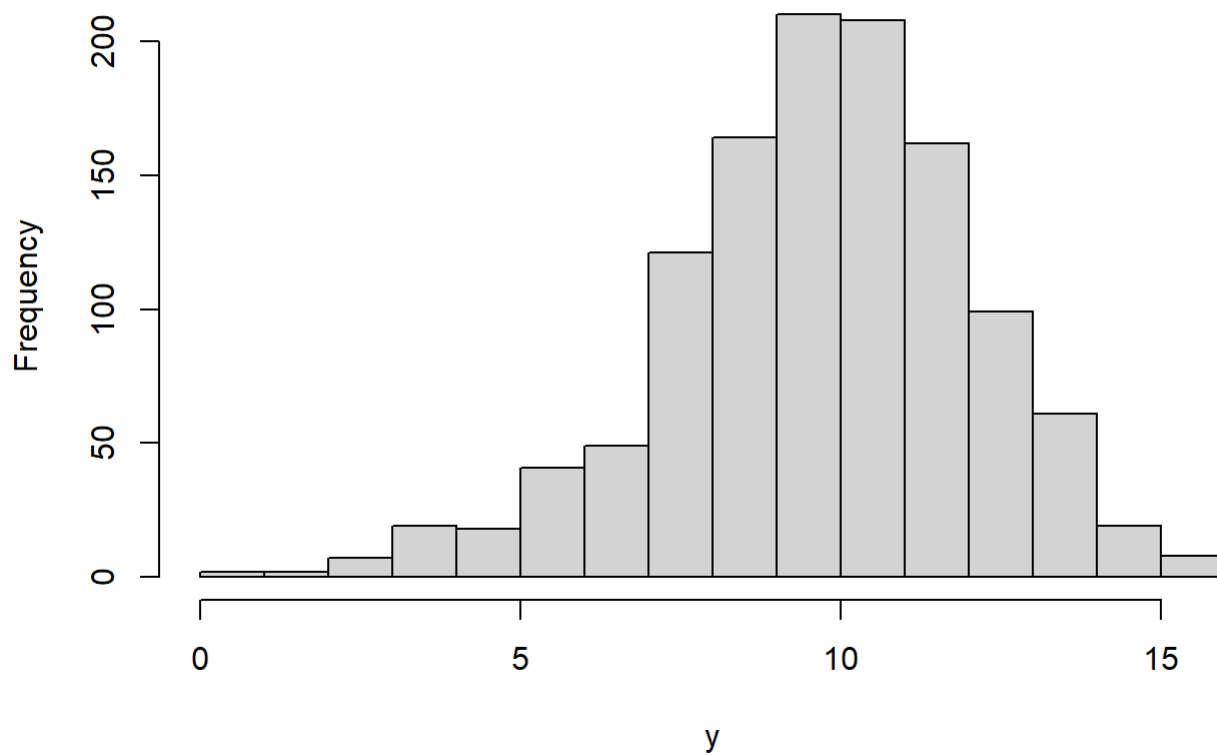
Following histograms demonstrate that number of followers for both, multi-worded and single-worded company names are Log-normally distributed. That (among other things) justifies comparison of their logs' means by T-test.

```
# adding 1 follower to every company to resolve the problem of log0
x <- log(multi_worded$followers_count+1);
y <- log(single_worded$followers_count+1);
# Plotting histograms
hist(x,xlim=c(0,16),breaks=16,main = "multi-worded");
```



```
hist(y,xlim=c(0,16),breaks=16, main = "single-worded");
```

single-worded



```
# comparing follower count for single-worded and multi-worded company names
t.test(x,y)
```

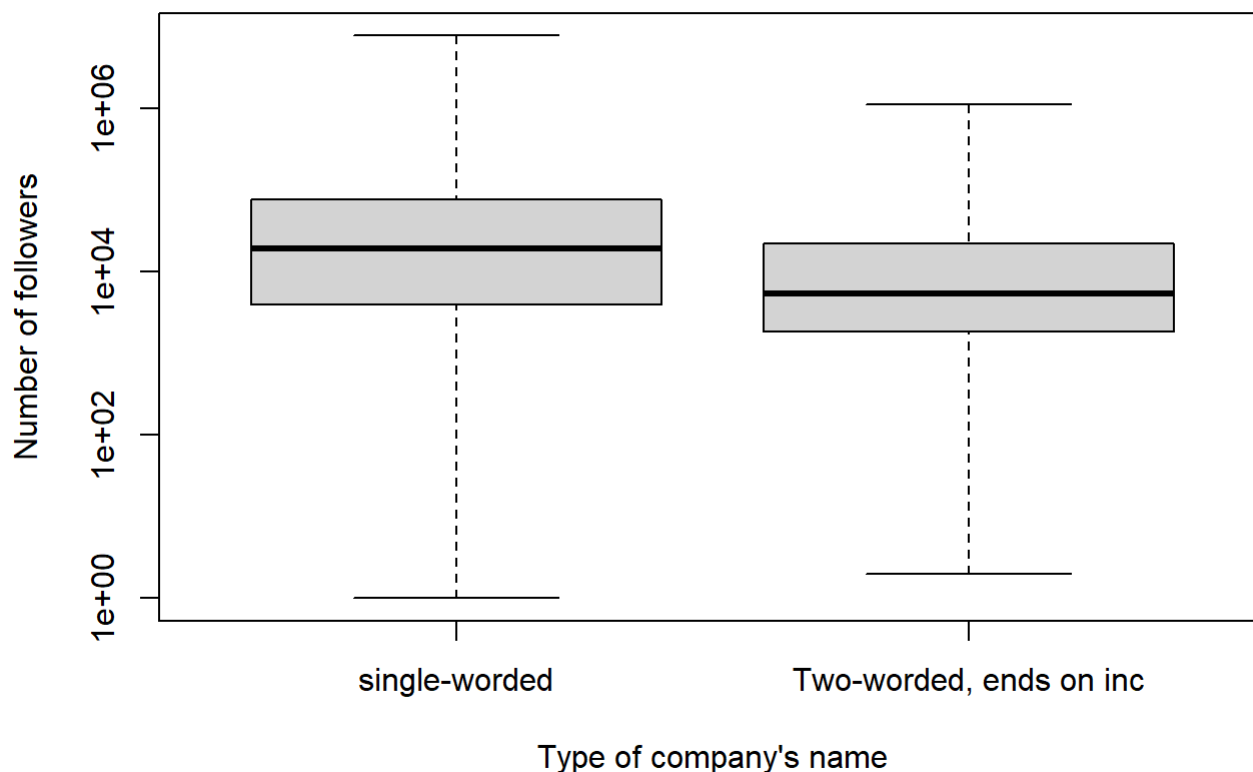
```
##
## Welch Two Sample t-test
##
## data: x and y
## t = -16.53, df = 1899.1, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.449985 -1.142410
## sample estimates:
## mean of x mean of y
## 8.378750 9.674947
```

```
# produce a column that indicates whether company's name is one-worded or one word +Inc./Inc
#Other companies are neglected
linkedin_recent %>%
  select(company_name, followers_count, industry) %>%
  filter(sapply(strsplit(company_name, " "), length) == 1 |
         (sapply(strsplit(company_name, " "), length) == 2) & grepl(" Inc$| Inc.$", company_name))
  %>%
  mutate(
    ends_with_inc = sapply(strsplit(company_name, " "), length) != 1
  ) ->
  inc_role
```

Following analysis clearly demonstrates the importance of one simple change companies can make to increase their number of followers on LinkedIn: removing Inc at the end of their name. For example, GGP Inc. would become GGP.

```
inc_role_plus <- inc_role
inc_role_plus$followers_count <- inc_role_plus$followers_count + 1 # added 1 follower to everyone
# to plot on log scale
boxplot(followers_count ~ ends_with_inc, data = inc_role_plus, log = "y", range = 0, names=c("single-worded",
"Two-worded, ends on inc"), xlab = "Type of company's name",
        ylab = "Number of followers", main = "Dependence of company popularity on its name")
```

Dependence of company popularity on its name



```
t.test(log(followers_count)~ends_with_inc,data = inc_role_plus)
```

```
##
## Welch Two Sample t-test
##
## data: log(followers_count) by ends_with_inc
## t = 6.3853, df = 292.88, p-value = 6.692e-10
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.7327925 1.3857867
## sample estimates:
## mean in group FALSE mean in group TRUE
##           9.674947           8.615658
```