

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. #

```
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 valuable features from company names.  
linkedin_recent %>%  
  select(company_name,followers_count,industry) %>%  
  mutate(  
    name_length = nchar(company_name),  
    abbreviated_name = grepl("^[[:upper:]]+$", company_name), # does not include BLA Inc kind of stuff, or L-3  
    multi_worded_name = sapply(strsplit(company_name, " "), length) != 1  
  ) ->  
  name_role
```

```

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, 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



```
t.test(followers_count~multi_worded_name,data = name_role_plus)
```

```

##
##  Welch Two Sample t-test
##
## data:  followers_count by multi_worded_name
## t = 7.4801, df = 1259.9, p-value = 1.387e-13
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   79354.43 135778.24
## sample estimates:
## mean in group FALSE mean in group TRUE
##       143259.80       35693.47

```

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

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

```
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"
```