one-worded-vs-multi-worded

Following data analysis demonstrates that companies that are registered with a one-worded name (including abbreviations) on linkedin have more followers.

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

```r
# 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
```

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

```r
# 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", xlab = "one vs mu
lti worded",
    ylab = "Mean followers", main = "one vs multi worded")
```

one vs multi worded
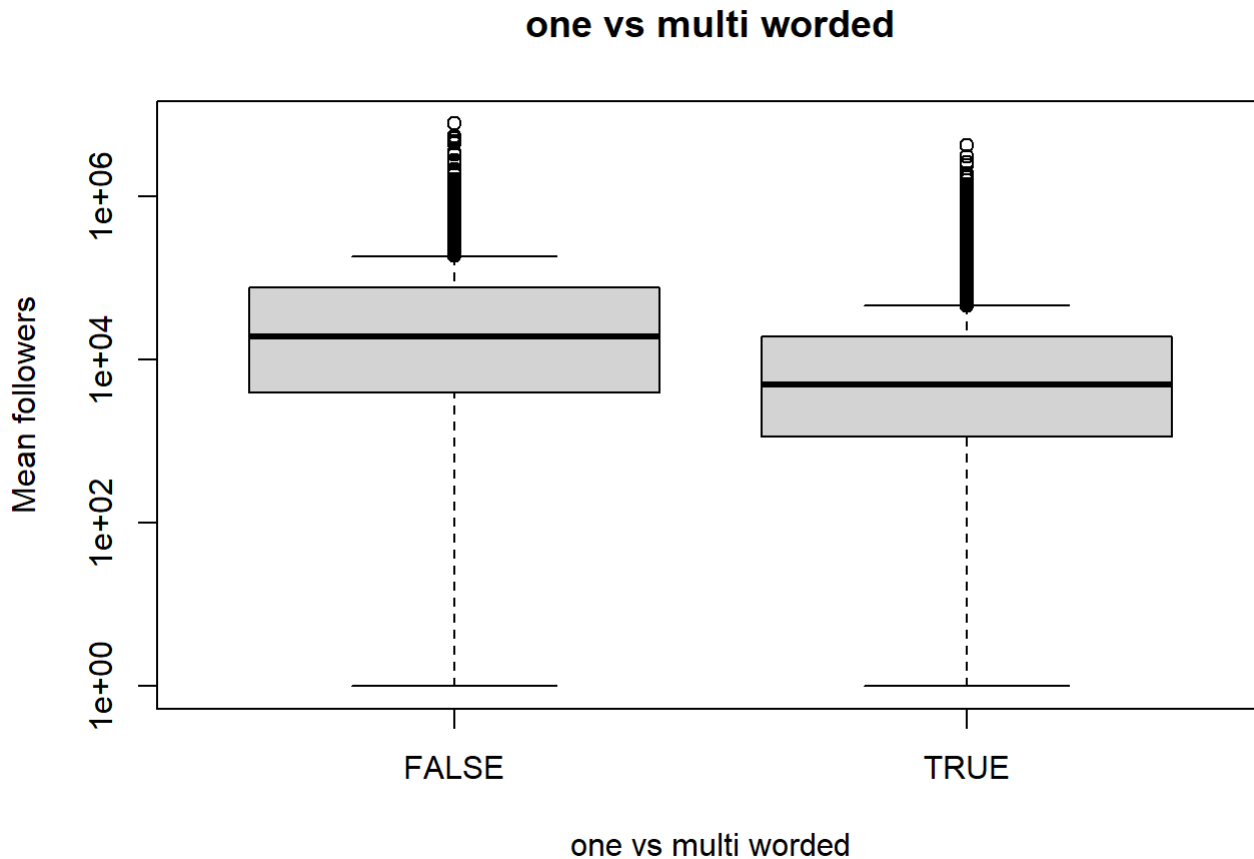
one vs multi worded

```
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
```