

Calculate Frequency of Citations in R

YIN YI

4/11/2019

```
library(tidyverse)

ref_test <- read.csv("reference_extract.csv", sep = "\t",
                    row.names=1, colClasses="character")
```

Calculate the most frequency of cited citation

```
cites <- ref_test %>% select(file_name, citations) %>%
  mutate(citations = gsub("\\[|\\]", "", citations)) %>%
  splitstackshape::cSplit("citations", ",", 2) %>%
  reshape2::melt(id.vars = "file_name") %>%
  select(-variable) %>% rename(cite=value) %>%
  filter(is.na(cite) == F)

cites %>% group_by(cite) %>%
  summarise(freq = n()) %>%
  arrange(desc(freq)) %>%
  top_n(n=5, wt = freq)
```

```
## # A tibble: 5 x 2
##   cite          freq
##   <chr>        <int>
## 1 '415 F.3d 1303'    30
## 2 '383 U.S. 1'      18
## 3 '550 U.S. 398'    16
## 4 '138 F.3d 1448'   13
## 5 '203 F.3d 1305'   13
```

Calculate the most frequency of cited regulations

```
regs <- ref_test %>% select(file_name, regulations) %>%
  mutate(regulations = gsub("\\[|\\]", "", regulations)) %>%
  splitstackshape::cSplit("regulations", ",", 2) %>%
  reshape2::melt(id.vars = "file_name") %>%
  select(-variable) %>% rename(reg=value) %>%
  filter(is.na(reg) == F)

regs %>% group_by(reg) %>%
  summarise(freq = n()) %>%
  arrange(desc(freq)) %>%
  top_n(n=5, wt = freq)
```

```
## # A tibble: 5 x 2
##   reg          freq
##   <chr>        <int>
```

## 1	'28 USC § 1295'	119
## 2	'35 USC § 103'	30
## 3	'35 USC § 112'	26
## 4	'35 USC § 271'	20
## 5	'125 Stat. 284'	17