PresidentSpeech

Steven Tran
Matthew Ciaramitaro

April 4, 2018

```r
purl("bush41 first state of the union.txt", output = "Bush.R")
```

```
##
##
## processing file: bush41 first state of the union.txt
```

```
##
  |
  |                                                              |   0%
  |
  |..............................................................| 100%
```

```
## output file: Bush.R
```

```
## [1] "Bush.R"
```

```r
purl("JFK first state of the union.txt", output = "JFK.R")
```

```
##
##
## processing file: JFK first state of the union.txt
```

```
##
  |
  |                                                              |   0%
  |
  |..............................................................| 100%
```

```
## output file: JFK.R
```

```
## [1] "JFK.R"
```

```r
source("Bush.R")
source("JFK.R")
nrc <- get_sentiments("nrc")

JFK <- data.frame(readLines("JFK.R"))
BUS <- data.frame(readLines("Bush.R"))

colnames(JFK) <- c("text")
colnames(BUS) <- c("text")

JFK <- JFK %>% mutate(President = "JFK")
BUS <- BUS %>% mutate(President = "Bush")

JFK[, 1] <- as.character(JFK[, 1])
BUS[, 1] <- as.character(BUS[, 1])

DATA <- JFK %>% # TOTAL COUNT
  rbind(BUS) %>%
```

```r
  unnest_tokens(word, text) %>%
  inner_join(nrc) %>%
  count(word, sort = TRUE)
```

```
## Joining, by = "word"
```

```r
DATAJ <- JFK %>% # JFK COUNT
  unnest_tokens(word, text) %>%
  inner_join(nrc) %>%
  count(word, sort = TRUE)
```

```
## Joining, by = "word"
```

```r
DATAB <- BUS %>% # BUSH COUNT
  unnest_tokens(word, text) %>%
  inner_join(nrc) %>%
  count(word, sort = TRUE)
```

```
## Joining, by = "word"
```

```r
# View(DATA)
# View(DATAJ)
# View(DATAB)

# We should use these 3 tables to determine % share JFK and BUSH have of total
joined <- DATA %>% full_join(DATAJ, by="word") %>% full_join(DATAB, by="word")
#puts words and counts for total, bush, and kennedy in seperate columns
colnames(joined) <- c("word","Total", "JFK", "Bush") #name columns
nums <- tail(colnames(joined), -1)
joined <- mutate_at(joined, .vars=vars(nums), .funs=funs(replace(.,is.na(.), 0)) ) #replace na with 0
print(joined)
```

```
## # A tibble: 0 x 4
## # ... with 4 variables: word <chr>, Total <dbl>, JFK <dbl>, Bush <dbl>
```

```r
#View(joined)
```

TODO: DISCUSS RESULTS OF TABLE

```r
#now we must divide the JFK and BUS cols by the total row
#percents <- joined %>% mutate_at(.vars=vars("JFK", "Bush"), . = . / Total)
percents <- joined %>% mutate(JFK = JFK / Total * 100) %>% mutate(Bush = Bush / Total * 100) %>% select
colnames(percents) <- c("word", "JFK (%combined)", "Bush (%combined)") #name columns
#View(percents)
print(percents)
```

```
## # A tibble: 0 x 3
## # ... with 3 variables: word <chr>, `JFK (%combined)` <dbl>, `Bush
## #   (%combined)` <dbl>
```

```r
# We should count differences in top 10 words
top10jfk <- head(DATAJ, 10)[1] # top 10 words for jfk
top10bus <- head(DATAB, 10)[1] # top 10 words for bush
compare <- top10jfk %>%
  mutate(inTop10Bus = word %in% top10bus[[1]])
# View(top10jfk)
top10merged <- top10jfk %>% rbind(top10bus) %>% distinct()
insct <- top10jfk %>% intersect(top10bus)
```

```
#View(insct)
numsame <- nrow(insct)
print(paste("The number of words that did not change in the top 10 is ", numsame))
```

[1] "The number of words that did not change in the top 10 is 0"

```
print("The following words remained in the top 10 most-used for both speeches:")
```

[1] "The following words remained in the top 10 most-used for both speeches:"

```
print(insct$word)
```

character(0)