Assignment2 622

Sagnik Chakravarty

Find Subreddits

```
war_sub_df <- find_subreddits("Russia Ukraine")
```

```
parsing URLs on page 1...
parsing URLs on page 2...
parsing URLs on page 3...
```

```
war_sub_df <- data.frame(war_sub_df)
war_sub_df_clean <- war_sub_df %>%
  select(subreddit,
         title,
         description,
         subscribers,
         date_utc)
```

```
war_sub_df_clean$date_utc <- as.Date(war_sub_df_clean$date_utc)
rownames(war_sub_df_clean) <- 1:nrow(war_sub_df_clean)
str(head(war_sub_df_clean, 1))
```

```
'data.frame':   1 obs. of  5 variables:
 $ subreddit  : chr "IrelandPics"
 $ title      : chr "Poblacht na hÉireann"
 $ description: chr "A subreddit for photos of Ireland!\n\nMember of the /r/NationalPhotoSubs
 $ subscribers: num 9071
 $ date_utc   : Date, format: "2015-01-22"
```

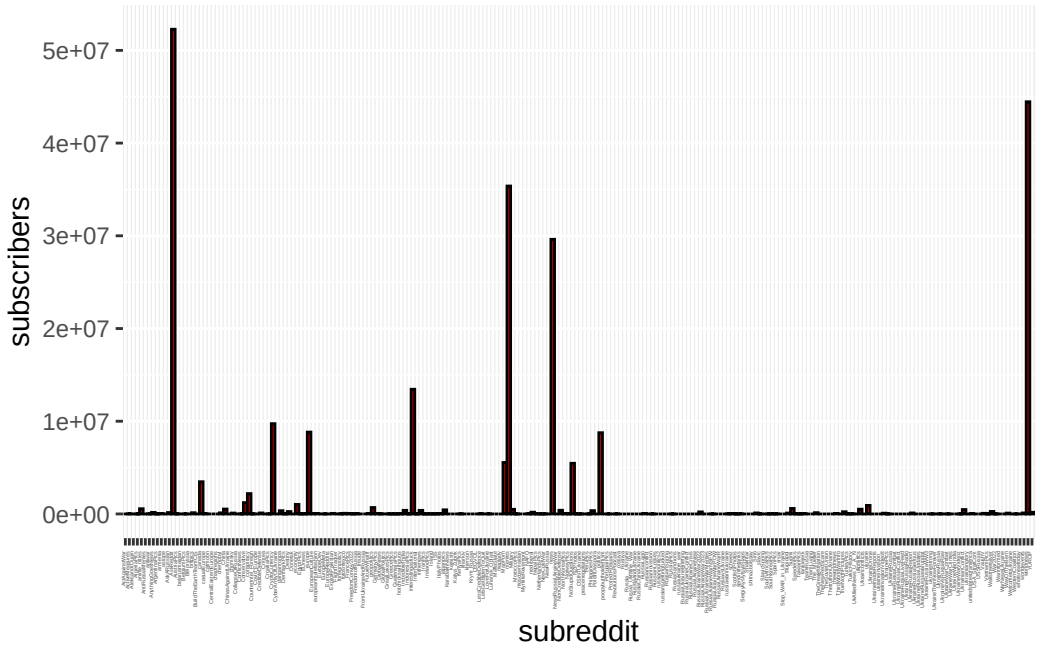
```
pander(dim(war_sub_df))
```

228 and *7*

```
pander(dim(war_sub_df_clean))
```

228 and *5*

```
war_sub_df_clean %>%ggplot(aes(x = subreddit, y = subscribers))+
  geom_bar(stat = 'identity',
           color = 'black',
           fill = 'red')+
  theme(axis.text.x = element_text(angle = 90,
                                   vjust = 0,
                                   hjust = 1,
                                   size = 2))+
  scale_y_continuous(
    breaks = pretty_breaks(n=5))
```



```r
desc_clean <- function(x){
  tk <- tibble(line = 1, text = x)
  tk <- tk %>%
    unnest_tokens(word, text) %>%
    anti_join(stop_words)
  return(tolower(str_c(tk$word, collapse = ' ')))
}

war_sub_df_clean <- war_sub_df_clean %>%
  rowwise() %>%
  mutate(cleaned_description = desc_clean(description)) %>%
  select(-description) %>%
  ungroup()
```

```r
check_war <- function(x){
  keywords <- c('Zelensky',
                'Putin',
                'Trump',
                'USA',
                'United States',
                'America',
                'Russia Ukriane',
                'War',
                'Minerals Deal',
                'Ceasefire',
                'Security'
                )
  pattern <- paste0("\\b(", paste(keywords, collapse = "|"), ")\\b")
  any(str_detect(x, pattern = pattern))
}

war_sub_df_clean <- war_sub_df_clean %>%
  rowwise()  %>%
  mutate(war_related = if_else(check_war(cleaned_description), TRUE, FALSE)) %>%
  ungroup()
```

```r
subreddit_df <- war_sub_df_clean %>%
  filter(war_related == TRUE) %>%
  arrange(desc(subscribers))  %>%
  head(n=4)
```

```
subreddit_df[c(-5,-6)] %>%
  pander(caption = "Subreddits Related to Ukraine Russia War")
```

Table 1: Subreddits Related to Ukraine Russia War

subreddit	title	subscribers	date_utc

```
subreddit_df %>% ggplot(aes(x = subscribers,
                            y = subreddit))+
  geom_bar(stat = 'identity',
           fill = 'black')+
  theme_classic()
```