Twitter reproducible methods

2-27-22

Introduction to procedurals

The following code below shows procedures to make the initial data frame. It is used to inform the data, but it not run. In the grouped code below, we pick up the code from the storage device on the computer

```
library(magrittr)

a <-
readRDS("C:\\Users\\mario\\OneDrive\\Desktop\\twitter_shiny\\twittter\\data\\
tweetschool.rds")

o <- readxl::read_excel("C:\\Users\\mario\\OneDrive\\Desktop\\twitter -11-25-
21\\data\\latest data 11-28-21\\teaching online 12-3-21 - Copy.xlsx")

a1 <- setdiff(a$content, o$content) %>% data.frame()
o1 <- setdiff(o$content,a$content) %>% data.frame()

ts <- sample(a1$., 2000, replace = FALSE, prob = NULL)

to <- sample(o1$., 2000, replace = FALSE, prob = NULL)

saveRDS(to, "c:/users/anonymous/downloads/to.rds")
saveRDS(ts, "c:/users/mario/downloads/ts.rds")
```

Here, we pick up the code with the data from the storage device on the computer. Quants methods wilcox tests

```
ts <- readRDS("C:\\Users\\mario\\OneDrive\\Desktop\\twitter -11-25-
21\\data\\Latest-data-12-4-21\\ts.rds")
to <- readRDS("C:\\Users\\mario\\OneDrive\\Desktop\\twitter -11-25-
21\\data\\Latest-data-12-4-21\\to.rds")

ts1 <- syuzhet::get_nrc_sentiment(ts)

data <- data.frame(cbind(ts1,ts))

to1 <- syuzhet::get_nrc_sentiment(to)

data1 <- data.frame(cbind(to1,to))

data$kind <- rep("ts", 2000)
data1$kind <- rep("to", 2000)
```

```r
options(scipen=999)

#

dataTeachschoolpos <-data[,c("kind", "positive")]

shapiro.test(dataTeachschoolpos$positive)

##
##  Shapiro-Wilk normality test
##
## data:  dataTeachschoolpos$positive
## W = 0.90671, p-value < 0.00000000000000022

mean(dataTeachschoolpos$positive)

## [1] 2.772

dataTeachonlinepos <- data1[,c("kind", "positive")]

shapiro.test(dataTeachonlinepos$positive)

##
##  Shapiro-Wilk normality test
##
## data:  dataTeachonlinepos$positive
## W = 0.91227, p-value < 0.00000000000000022

mean(dataTeachonlinepos$positive)

## [1] 2.883

now_data_pos <- data.frame(rbind(dataTeachonlinepos,dataTeachschoolpos))



(rpositive <- wilcox.test(now_data_pos$positive~now_data_pos$kind))

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  now_data_pos$positive by now_data_pos$kind
## W = 2089647, p-value = 0.01169
## alternative hypothesis: true location shift is not equal to 0

#

dataTeachschoolAnger <- data[, c("kind", "anger")]
```

```r
shapiro.test(dataTeachschoolAnger$anger)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  dataTeachschoolAnger$anger
## W = 0.60172, p-value < 0.00000000000000022
```

```r
mean(dataTeachschoolAnger$anger)
```

```
## [1] 0.3865
```

```r
dataTeachonlineAnger <- data1[,c("kind", "anger")]

shapiro.test(dataTeachonlineAnger$anger)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  dataTeachonlineAnger$anger
## W = 0.52225, p-value < 0.00000000000000022
```

```r
mean(dataTeachonlineAnger$anger)
```

```
## [1] 0.255
```

```r
now_data_anger <-
data.frame(rbind(dataTeachschoolAnger,dataTeachonlineAnger))

(ranger <- wilcox.test(now_data_anger$anger~now_data_anger$kind))
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  now_data_anger$anger by now_data_anger$kind
## W = 1818346, p-value = 0.00000000006812
## alternative hypothesis: true location shift is not equal to 0
```

```r
#

dataTeachschoolAnticipation <- data[, c("kind", "anticipation")]

shapiro.test(dataTeachschoolAnticipation$anticipation)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  dataTeachschoolAnticipation$anticipation
## W = 0.78649, p-value < 0.00000000000000022
```

```r
mean(dataTeachschoolAnticipation$anticipation)
```

```
## [1] 0.8025

dataTeachonlineAnticipation <- data1[,c("kind", "anticipation")]

shapiro.test(dataTeachonlineAnticipation$anticipation)

##
##  Shapiro-Wilk normality test
##
## data:  dataTeachonlineAnticipation$anticipation
## W = 0.79491, p-value < 0.00000000000000022

mean(dataTeachonlineAnticipation$anticipation)

## [1] 0.8375

now_data_anticipation <-
data.frame(rbind(dataTeachschoolAnticipation,dataTeachonlineAnticipation))

(ranticipation <-
wilcox.test(now_data_anticipation$anticipation~now_data_anticipation$kind))

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  now_data_anticipation$anticipation by now_data_anticipation$kind
## W = 2037048, p-value = 0.2747
## alternative hypothesis: true location shift is not equal to 0

#


dataTeachschoolDisgust <- data[, c("kind", "disgust")]

shapiro.test(dataTeachschoolDisgust$disgust)

##
##  Shapiro-Wilk normality test
##
## data:  dataTeachschoolDisgust$disgust
## W = 0.57492, p-value < 0.00000000000000022

mean(dataTeachschoolDisgust$disgust)

## [1] 0.319

dataTeachonlineDisgust <- data1[,c("kind", "disgust")]

shapiro.test(dataTeachonlineDisgust$disgust)

##
##  Shapiro-Wilk normality test
```

```
##
## data:  dataTeachonlineDisgust$disgust
## W = 0.44581, p-value < 0.00000000000000022

mean(dataTeachonlineDisgust$disgust)

## [1] 0.1805

now_data_disgust <-
data.frame(rbind(dataTeachschoolDisgust,dataTeachonlineDisgust))

(rdisgust <- wilcox.test(now_data_disgust$disgust~now_data_disgust$kind))

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  now_data_disgust$disgust by now_data_disgust$kind
## W = 1792446, p-value = 0.0000000000000007639
## alternative hypothesis: true location shift is not equal to 0

#

dataTeachschoolFear <- data[, c("kind", "fear")]

shapiro.test(dataTeachschoolFear$fear)

##
##  Shapiro-Wilk normality test
##
## data:  dataTeachschoolFear$fear
## W = 0.75481, p-value < 0.00000000000000022

mean(dataTeachschoolFear$fear)

## [1] 0.7245

dataTeachonlineFear <- data1[,c("kind", "fear")]

shapiro.test(dataTeachonlineFear$fear)

##
##  Shapiro-Wilk normality test
##
## data:  dataTeachonlineFear$fear
## W = 0.65595, p-value < 0.00000000000000022

mean(dataTeachonlineFear$fear)

## [1] 0.4685
```

```r
now_data_fear <- data.frame(rbind(dataTeachschoolFear,dataTeachonlineFear))

(rdisgust <- wilcox.test(now_data_fear$fear~now_data_fear$kind))

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  now_data_fear$fear by now_data_fear$kind
## W = 1682637, p-value < 0.00000000000000022
## alternative hypothesis: true location shift is not equal to 0

#



dataTeachschoolJoy <- data[, c("kind", "joy")]

shapiro.test(dataTeachschoolJoy$joy)

##
##  Shapiro-Wilk normality test
##
## data:  dataTeachschoolJoy$joy
## W = 0.71771, p-value < 0.00000000000000022

mean(dataTeachschoolJoy$joy)

## [1] 1.537

dataTeachonlineJoy <- data1[,c("kind", "joy")]

shapiro.test(dataTeachonlineJoy$joy)

##
##  Shapiro-Wilk normality test
##
## data:  dataTeachonlineJoy$joy
## W = 0.68757, p-value < 0.00000000000000022

mean(dataTeachonlineJoy$joy)

## [1] 1.512

now_data_joy <- data.frame(rbind(dataTeachschoolJoy,dataTeachonlineJoy))

(rjoy <- wilcox.test(now_data_joy$joy~now_data_joy$kind))

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  now_data_joy$joy by now_data_joy$kind
```

```
## W = 1952444, p-value = 0.1338
## alternative hypothesis: true location shift is not equal to 0

#

dataTeachschoolSadness <- data[, c("kind", "sadness")]

shapiro.test(dataTeachschoolSadness$sadness)

##
##  Shapiro-Wilk normality test
##
## data:  dataTeachschoolSadness$sadness
## W = 0.71777, p-value < 0.00000000000000022

mean(dataTeachschoolSadness$sadness)

## [1] 0.587

dataTeachonlineSadness <- data1[,c("kind", "sadness")]

shapiro.test(dataTeachonlineSadness$sadness)

##
##  Shapiro-Wilk normality test
##
## data:  dataTeachonlineSadness$sadness
## W = 0.646, p-value < 0.00000000000000022

mean(dataTeachonlineSadness$sadness)

## [1] 0.449

now_data_sadness <-
data.frame(rbind(dataTeachschoolSadness,dataTeachonlineSadness))

(rsadness <- wilcox.test(now_data_sadness$sadness~now_data_sadness$kind))

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  now_data_sadness$sadness by now_data_sadness$kind
## W = 1817233, p-value = 0.000000006632
## alternative hypothesis: true location shift is not equal to 0

#

dataTeachschoolSurprise <- data[, c("kind", "surprise")]

shapiro.test(dataTeachschoolSurprise$surprise)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  dataTeachschoolSurprise$surprise
## W = 0.59249, p-value < 0.00000000000000022

mean(dataTeachschoolSurprise$surprise)

## [1] 1.283

dataTeachonlineSurprise <- data1[,c("kind", "surprise")]

shapiro.test(dataTeachonlineSurprise$surprise)

##
##  Shapiro-Wilk normality test
##
## data:  dataTeachonlineSurprise$surprise
## W = 0.52073, p-value < 0.00000000000000022

mean(dataTeachonlineSurprise$surprise)

## [1] 1.2395

now_data_surprise <-
data.frame(rbind(dataTeachschoolSurprise,dataTeachonlineSurprise))

(rsurprise <- wilcox.test(now_data_surprise$surprise~now_data_surprise$kind))

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  now_data_surprise$surprise by now_data_surprise$kind
## W = 1926513, p-value = 0.00612
## alternative hypothesis: true location shift is not equal to 0

#

dataTeachschoolNegative <- data[, c("kind", "negative")]

shapiro.test(dataTeachschoolNegative$negative)

##
##  Shapiro-Wilk normality test
##
## data:  dataTeachschoolNegative$negative
## W = 0.84461, p-value < 0.00000000000000022

mean(dataTeachschoolNegative$negative)

## [1] 1.204
```

```r
dataTeachonlineNegative <- data1[,c("kind", "negative")]

shapiro.test(dataTeachonlineNegative$negative)

##
##  Shapiro-Wilk normality test
##
## data:  dataTeachonlineNegative$negative
## W = 0.77244, p-value < 0.00000000000000022

mean(dataTeachonlineNegative$negative)

## [1] 0.851

now_data_negative <-
data.frame(rbind(dataTeachschoolNegative,dataTeachonlineNegative))

(rnegative <- wilcox.test(now_data_negative$negative~now_data_negative$kind))

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  now_data_negative$negative by now_data_negative$kind
## W = 1663646, p-value < 0.00000000000000022
## alternative hypothesis: true location shift is not equal to 0

#

dataTeachschoolTrust <- data[, c("kind", "trust")]

shapiro.test(dataTeachschoolTrust$trust)

##
##  Shapiro-Wilk normality test
##
## data:  dataTeachschoolTrust$trust
## W = 0.90301, p-value < 0.00000000000000022

mean(dataTeachschoolTrust$trust)

## [1] 2.792

dataTeachonlineTrust <- data1[,c("kind", "trust")]

shapiro.test(dataTeachonlineTrust$trust)

##
##  Shapiro-Wilk normality test
##
## data:  dataTeachonlineTrust$trust
## W = 0.85411, p-value < 0.00000000000000022
```

```
mean(dataTeachonlineTrust$trust)

## [1] 2.1335

now_data_trust <-
data.frame(rbind(dataTeachschoolTrust,dataTeachonlineTrust))

(rtrust <- wilcox.test(now_data_trust$trust~now_data_trust$kind))

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  now_data_trust$trust by now_data_trust$kind
## W = 1340221, p-value < 0.00000000000000022
## alternative hypothesis: true location shift is not equal to 0
```

This part harnesses the qualitative part of the data:

get the visual wheel with term frequencies at 10 minimum.

```
library(magrittr)
library(quanteda)

## Package version: 3.1.0
## Unicode version: 13.0
## ICU version: 69.1

## Parallel computing: 6 of 6 threads used.

## See https://quanteda.io for tutorials and examples.

ts <- readRDS("C:\\Users\\mario\\OneDrive\\Desktop\\twitter -11-25-
21\\data\\Latest-data-12-4-21\\ts.rds")
to <- readRDS("C:\\Users\\mario\\OneDrive\\Desktop\\twitter -11-25-
21\\data\\Latest-data-12-4-21\\to.rds")
ts1 <- syuzhet::get_nrc_sentiment(ts)

data <- data.frame(cbind(ts1,ts))

to1 <- syuzhet::get_nrc_sentiment(to)


data1 <- data.frame(cbind(to1,to))
data$kind <- rep("ts", 2000)
data1$kind <- rep("to", 2000)

names(data1)[11] <- "content"
names(data)[11] <- "content"

bigdata <- data.frame(rbind(data,data1))
```

```
bigdata$content <- gsub("@[A-Za-z0-9]+", "", bigdata$content)


a <- quanteda::corpus(bigdata, text_field = "content")

library(quanteda.textplots)
this <- quanteda::corpus_subset(a,
                kind %in% c("to", "ts")) %>%
  quanteda::tokens(remove_punct = TRUE) %>%
  quanteda::tokens_remove(stopwords("english")) %>%
  quanteda::dfm() %>%
  quanteda::dfm_group(groups = kind) %>%
  quanteda::dfm_trim(min_termfreq = 10, verbose = FALSE) %>%
  quanteda.textplots::textplot_wordcloud(comparison = TRUE,
color = c("blue", "purple"))
```

to

ts

```
library(magrittr)
ts <- readRDS("C:\\Users\\mario\\OneDrive\\Desktop\\twitter -11-25-
21\\data\\Latest-data-12-4-21\\ts.rds")
to <- readRDS("C:\\Users\\mario\\OneDrive\\Desktop\\twitter -11-25-
21\\data\\Latest-data-12-4-21\\to.rds")

ts1 <- syuzhet::get_nrc_sentiment(ts)
to1 <- syuzhet::get_nrc_sentiment(to)

data_to <- data.frame(cbind(to1,to))
```

```
data_ts <- data.frame(cbind(ts1,ts))

# Read in data once it has been saved and stored
to <- readRDS("C:\\Users\\mario\\OneDrive\\Desktop\\twitter -11-25-
21\\data\\Latest-data-12-4-21\\qualitative tables\\data_to.rds")
ts <- readRDS("C:\\Users\\mario\\OneDrive\\Desktop\\twitter -11-25-
21\\data\\Latest-data-12-4-21\\qualitative tables\\data_ts.rds")

library(magrittr)

tsc <- data.frame(ts) %>% quanteda::corpus(., text_field = "Content",
docid_field = "Num")

toc <- data.frame(to) %>% quanteda::corpus(., text_field = "Content",
docid_field = "Num")

# save the dfm matrix to the text by copy pasting

tstoks1 <- quanteda::tokens(tsc, remove_numbers = TRUE, remove_punct = TRUE)
c1 <- quanteda::dfm(tstoks1, remove = stopwords::stopwords("english"))

totoks1 <- quanteda::tokens(toc, remove_numbers = TRUE, remove_punct = TRUE)
c2 <- quanteda::dfm(totoks1, remove = stopwords::stopwords("english"))




result <- ldatuning::FindTopicsNumber(
  c1,
  topics = seq(from = 2, to = 200, by = 5),
  metrics = c("Griffiths2004", "CaoJuan2009", "Arun2010", "Deveaud2014"),
  method = "Gibbs",
  control = list(seed = 77),
  mc.cores = 2L,
  verbose = TRUE
)


#=== NOTE adding topic modeling

# Note that for each of the topics you must change the K and first argument
of "stm" function


ldatuning::FindTopicsNumber_plot(result)

this <- topicmodels::LDA(c1,k =22) # use the c1 and topic number for best
results
```

```
that <- topicmodels::LDA(c2, k = 57)
(ap_topics <- tidytext::tidy(this, matrix = "beta"))

(ap_topics <- tidytext::tidy(that, matrix = "beta"))



library(ggplot2)
library(dplyr)
library(tidytext)

(ap_top_terms <- ap_topics %>%
    group_by(topic) %>%
    top_n(10, beta) %>%
    ungroup() %>%
    arrange(topic, -beta))

(ap_top_terms %>%
  mutate(term = tidytext::reorder_within(term, beta, topic)) %>%
  ggplot(aes(term, beta, fill = factor(topic))) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ topic, scales = "free") +
  coord_flip() +
  tidytext::scale_x_reordered())
```