Quanteda

Susan Kamal for R-ladies meeting 07.06.2018

Background

- PhD in pharmacy- HIV medication adherence.
- Used quanteda in «Word Frequency and Content analysis of Antiretroviral Adherence Enhancing Interview Reports", published in PEC 2018, May.

Background

Motivational interviewing patient-pharmacist (MI)

- Face-to-face
- Patient-centered
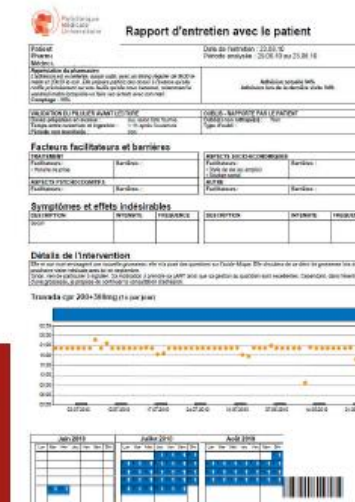- Directive & goal-oriented
- Short but repeated

Medication Event Monitoring System (MEMS®)

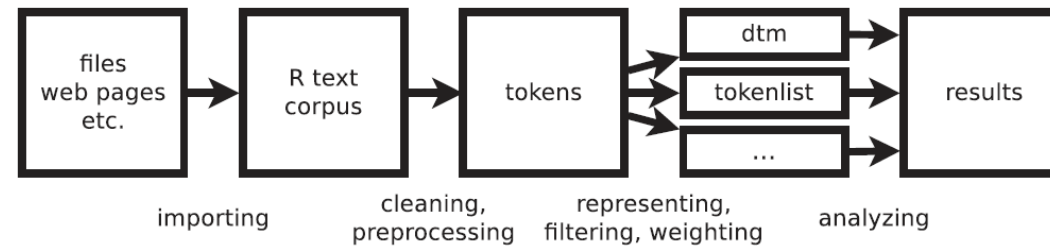- Objective and dynamic measure of daily medication intake

Medication adherence report

- Data collection and secure record
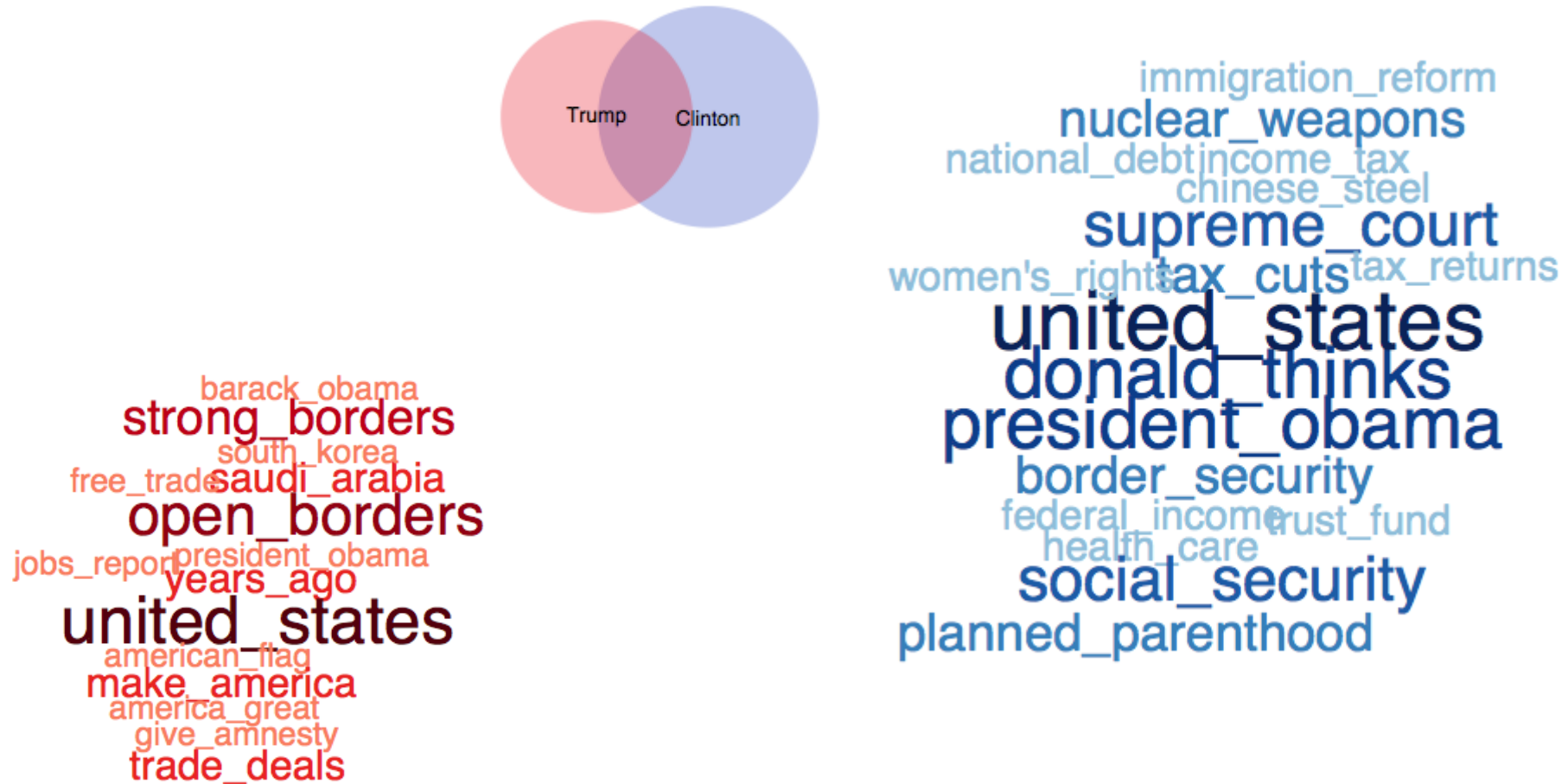- Feedback to the patient/physician
- Continuity of care

Krummenacher, Cavassini, Bugnon, Schneider. AIDS Care 2011; 23(5):550-61
Krummenacher, Cavassini, Bugnon, Spirig, Schneider. PWS 2010; 32:776–786
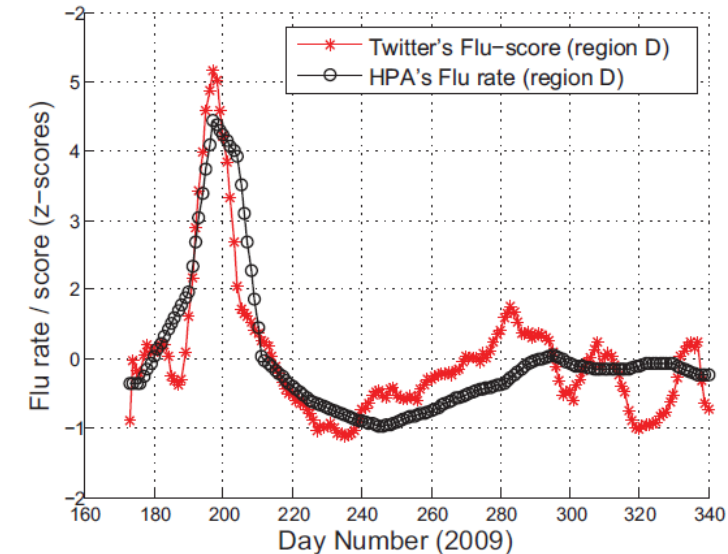Lelubre, et al. BioMed research international 2015 (2015)

Method

files
web pages
etc.

R text
corpus

tokens

dtm

tokenlist

...

results

importing

cleaning,
preprocessing

representing,
filtering, weighting

analyzing

Method
Presidential Debate of 2016

Method
Tracking the flu pandemic by monitoring the social web

lung	unwel	temperatur	like	headach	season
unusu	chronic	child	dai	appetit	stai
symptom	spread	diarrhoea	start	muscl	weaken
immun	feel	liver	plenti	antivir	follow
sore	peopl	nation	small	pandem	pregnant
thermomet	bed	loss	heart	mention	condit
high	group	tired	import	risk	carefulli
work	short	stage	page	diseas	recognis
servic	wors	case	similar	term	home
increas	exist	ill	sens	counter	better
cough	vomit	earli	neurolog	catch	onlin
fever	concern	check	drink	long	far
consid	ach	breath	flu	member	kidnei
mild	number	sick	throat	famili	water
read	includ	swine	confirm	need	nose
medic	phone	cancer	disord	unsur	suddenli
runni					

Lampos, Vasileios, and Nello Cristianini. "Tracking the flu pandemic by monitoring the social web." . IEEE, 2010.

Code

```r
library("quanteda") library("magrittr") library("dplyr") library("glmnet") library("ggplot2") library("xtable")

runModel <- function(triPatMat, combDf, type, a) { if (type == "logit", 0.01)

wordWeights$Positive <- ifelse(wordWeights$estimate > 0, wordWeights$estimate, 0)
wordWeights$Negative <- ifelse(wordWeights$estimate < 0, wordWeights$estimate, 0)

# some analysis just on the non-text data

# context search dv <- select(comb, c(-text)) fullCorpus <- corpus(comb$text, docvars = dv)
kwic(fullCorpus, "adhésion selon")

wordcloud::comparison.cloud((as.matrix(wct)), title.size = 1.2, random.order = FALSE, rot.per = 0,
scale = c(1, 2))

# Extra stopwords here (French default stopword list is already used) extra <- c("apparement", "vrac",
"apres", "suivregestion", "mems.je", "sans", "vas", "aucun")

thisCondition <- intervisitAll thisCondition$id <- as.factor(thisCondition$id)
```

Results I

Words [within context]** associated with high adherence (>=90%)
"have disappeared " [the side effects…]
"excellent motivation"
"social support"
"mutually" [patient and others support the medication intake mutually]
"very regular hours" [...of treatment intake]
"patient is happy" [...to see good adherence results as shown by EM]
"MEMS* as memory aid"
"women" [female patient discusses her HIV status with other HIV+ women]
"partner"
"husband"

Words [within context] associated with low adherence (<90%)
"DOT" [Pharmacist suggested switching to directly observed therapy]
" TEL" [pharmacist telephoned the treating ID physician]
"cry"
"prepares in advance" [patient…their doses]
"discrete" [patient want to be ...about their HIV status]
"small sachets" [patient keeps their medication in]
"name of ID physician"
"garbage bag" [patient uses a garbage bag to hide their medication]
"homeless" [patient is]
""other appointments" [...with several doctors]
"bitter taste" [of medication]
"baggage" [travel]
"name of psychosocial nurse"

Results II