

RAPID Helping

Sara Weston

7/26/2021

Contents

Workspace	1
Packages	1
Data	2
Cleaning	2
Recode and rename variables.	2
Score parental mental health variables	3
Extend demographics	3
Well-being composite	3
Clean open-ended	4
Filter responses	4
Prepare for topic modeling	5
How many topics?	5

Workspace

Packages

```
library(here)
library(tidyverse)
library(stm)
library(tidytext)
library(rio)
#library(furrr)
library(lubridate)
```

Data

Here we load the raw data file.

```
master = import(here("data/MasterFile_groupings.sav"))
```

We select only those variables to be used in these analyses.

```
master = master %>%
  select(CaregiverID,           # id variable
         UserLanguage,         # language
         StartDate,            # date
         starts_with("DEMO.007"), # race
         GAD2.002.a, GAD2.002.b, # anxiety
         PHQ.002.a, PHQ.002.b,  # depression
         STRESS.002,            # stress
         LONE.001.b,            # loneliness
         CBCL.002.a,            # child fussiness
         CBCL.002.b,            # child fearfulness
         FPL.150,               # below 1.5 x Federal poverty level?
         OPEN.002               # response to open-ended Q
  )

save(master, file = here("data/master_subset.Rdata"))
```

```
load(here("data/master_subset.Rdata"))
```

Cleaning

Recode and rename variables.

First, we recode the race variables into a single categorical variable.

```
master = master %>%
  mutate(
    race = case_when(
      DEMO.007_3 == 1 ~ "Black",
      DEMO.007_5 == 1 ~ "White",
      DEMO.007_1 == 1 ~ "Other",
      DEMO.007_2 == 1 ~ "Other",
      DEMO.007_4 == 1 ~ "Other",
      DEMO.007_6 == 1 ~ "Other",
      TRUE ~ NA_character_)
  )
```

We rename single items to better communicate the construct assessed.

```
master = master %>%
  rename(
    stress = STRESS.002,
    lonely = LONE.001.b,
```

```
child_fussy = CBCL.002.a,
child_fearful = CBCL.002.b,
poverty = FPL.150)
```

We use the date variable to create two new variables, one which indicates the number of months since the pandemic started (for analysis purposes) and the other which changes the class to a Date format.

```
master = master %>%
  mutate(Date = as.Date(StartDate)) %>%
  mutate(month = interval(as.Date("2020-03-13"), Date),
         month = month %/% months(1))
```

For each caregiver, we also create a variable to indicate which response this row is (i.e., their first, 1; second, 2; third, 3; etc).

```
master = master %>%
  group_by(CaregiverID) %>%
  mutate(response = row_number()) %>%
  ungroup()
```

Score parental mental health variables

Next we calculate parental anxiety and depression by taking the averages of the two items assessed.

```
master$anxiety = rowMeans(master[,c("GAD2.002.a", "GAD2.002.b")], na.rm = T)
master$depression = rowMeans(master[,c("PHQ.002.a", "PHQ.002.b")], na.rm = T)
```

Extend demographics

Some variables were only assessed the first time a parent entered the survey. This includes race and financial status. To ensure these responses are appropriately tied to questions assessed weekly – including the open-ended questions – we use the `fill` function to carry forward variables.

```
master = master %>%
  fill(race, poverty, .direction = "downup")
```

Well-being composite

We standardize the mental health variables to the sample at hand.

```
master = master %>%
  mutate(across(c(anxiety, depression, stress, lonely,
                  child_fussy, child_fearful),
               .fns = list(z = scale)))
```

Next, we calculate for each observation the average standardized response these questions. We multiply this average by -1, so that higher scores indicate better well-being.

```

master = master %>%
  rowwise %>%
  mutate(parent_wellbeing = mean(c_across(anxiety_z:lonely_z), na.rm=T)*-1,
         child_wellbeing = mean(c_across(child_fussy_z:child_fearful_z), na.rm=T)*-1) %>%
  ungroup()

```

Clean open-ended

```

master = master %>%
  mutate(OPEN.002 = str_replace(OPEN.002, "\n", " ")) %>%
  filter(!(OPEN.002 %in% c("N/A", "N/A.", "N/a", "N/a.", "n/a", "n/a.",
                        "None", "NA", "none", "Na", "no", "nope")))

```

Filter responses

We select only responses which have valid responses to the open-ended question and for which the participant completed the questionnaire in English.

```

master = master %>%
  filter(!is.na(OPEN.002)) %>%
  filter(OPEN.002 != "") %>%
  filter(UserLanguage == "EN")

nrow(master)

```

```
## [1] 21921
```

Analyses also require full information, so we remove participants missing on the well-being variables.

```

master = master %>%
  filter(!is.na(child_wellbeing)) %>%
  filter(!is.na(parent_wellbeing))

nrow(master)

```

```
## [1] 21900
```

Finally, we select only the variables of interest to us.

```

data = master %>%
  select(CaregiverID, Date, month, OPEN.002,
         parent_wellbeing, child_wellbeing, race, poverty,
         stress, lonely, child_fussy, child_fearful, poverty)
rm(master)

```

Prepare for topic modeling

We create a tidied version of the data frame that extracts distinct words. In other words, each word in each response will receive its own row in the data.

```
tidy_data <- data %>%  
  unnest_tokens(word, OPEN.002, token = "words")
```

Next, we remove stop words and numbers.

```
tidy_data = tidy_data %>%  
  anti_join(get_stopwords()) %>%  
  filter(!str_detect(word, "[0-9]+"))
```

We count the number of times each word was used in the dataset. We filter out words with fewer than 20 uses.

```
tidy_data = tidy_data %>%  
  add_count(word) %>%  
  filter(n > 20) %>%  
  select(-n)
```

““

How many topics?