Data Cleaning: Dutch Parliamentary Elections 2021 Wave 1

VU Political Communication Group: Mariken van der Velden (coordinator), Loes Aaldering, Wouter van...

Scripts

- Required Packages & Reproducibility
- Load Data
- Recode Block Background Variables
- Recode Block A Voting Behavior
- Recode Block B Performance Politics in the Media
- Recode Block H Other Issues
- Recode Block I News Consumption
- Merge & Save Data

Required Packages & Reproducibility

```r
rm(list=ls())

renv::snapshot()
```

```
## * The lockfile is already up to date.
```

```r
library(tidyverse)
library(sjlabelled)
```

Load Data

Load the data downloaded from Qualtrics, and only keep those that have given consent to participate in the panel study.

```r
f <- "../../data/raw-private/qualtrics_raw_wave1.csv"
col_names <- names(read_csv(f, n_max = 0))
d <- read_csv(f, col_names = col_names, skip = 3) %>%
  remove_all_labels() %>%
  tibble() %>%
  #only keep people that have given consent
  filter(consent1==1 & consent2==1)
rm(f, col_names)
```

Recode Block Background Variables

```r
BG <- d %>%
  mutate(wave = "wave 1",
         start_date = StartDate,
         end_date = EndDate,
         duration_min = round(Duration..in.seconds./60,2)) %>%
```

```r
  select(iisID, wave, start_date, end_date, duration_min,
         progress = Progress)
```

Recode Block A Voting Behavior

```r
A <- d %>%
  select(iisID, A1:A2, A2_otherparty = A2_14_TEXT,
         A2_DO_1:A2_DO_13, A3_DO_1:A3_DO_13) %>%
  mutate(A1 = recode(A1,
                     `2` = 0,
                     `3` = 998,
                     `4` = 999),
         A1_not = replace_na(A1_not, ""),
         A1_vote = replace_na(A1_vote, ""),
         A1_elab = paste0(A1_not, A1_vote),
         A2 = recode(A2,
                     `16` = 999),
         order_A2 = paste(A2_DO_1, A2_DO_2, A2_DO_3, A2_DO_4,
                          A2_DO_5, A2_DO_6, A2_DO_7, A2_DO_8,
                          A2_DO_9, A2_DO_10, A2_DO_11,
                          A2_DO_12, A2_DO_13, sep = "|"),
         order_A3 = paste(A3_DO_1, A3_DO_2, A3_DO_3, A3_DO_4,
                          A3_DO_5, A3_DO_6, A3_DO_7, A3_DO_8,
                          A3_DO_9, A3_DO_10, A3_DO_11,
                          A3_DO_12, A3_DO_13, sep = "|")) %>%
  select(iisID, A1, A1_elab, A2, A2_otherparty, order_A2, order_A3)

A3 <- d %>%
  select(iisID, A3_1:A3_13) %>%
  pivot_longer(cols = A3_1:A3_13,
               names_to = "variable") %>%
  drop_na(value) %>%
  separate(variable, c("variable", "party"), "_", extra = "merge")  %>%
  group_by(iisID) %>%
  summarise(n = row_number(),
            party = party) %>%
  ungroup() %>%
  mutate(n = paste("A3", n, sep="_"),
         n = factor(n),
         party = as.integer(party))

A3 <-  pivot_wider(A3, names_from = n, values_from = party,
                   values_fill = 0) %>%
  mutate(A3 = paste(A3_1, A3_2, A3_3, A3_4, A3_5,
                    A3_6, A3_7, A3_8, A3_9, A3_10,
                    A3_11, A3_12, sep = "|"))
A <- left_join(A, A3, by = "iisID") %>%
  select(iisID, A1:A2_otherparty, order_A2, A3, order_A3, A3_1:A3_12)
rm(A3)
```

Recode Block B Performance Politics in the Media

```r
d <-d %>%
mutate(B2_1 = recode(B2_1,`1` = 999, `2` = 1, `24` = 2,`25` = 3),
       B2_2 = recode(B2_2,`1` = 999, `2` = 1, `24` = 2,`25` = 3),
       B2_3 = recode(B2_3,`1` = 999, `2` = 1, `24` = 2,`25` = 3),
       B2_4 = recode(B2_4,`1` = 999, `2` = 1, `24` = 2,`25` = 3),
       B2_5 = recode(B2_5,`1` = 999, `2` = 1, `24` = 2,`25` = 3),
       B2_6 = recode(B2_6,`1` = 999, `2` = 1, `24` = 2,`25` = 3),
       B2_7 = recode(B2_7,`1` = 999, `2` = 1, `24` = 2,`25` = 3),
       B2_8 = recode(B2_8,`1` = 999, `2` = 1, `24` = 2,`25` = 3),
       B2_9 = recode(B2_9,`1` = 999, `2` = 1, `24` = 2,`25` = 3),
       B2_10 = recode(B2_10,`1` = 999, `2` = 1, `24` = 2,`25` = 3),
       B2_11 = recode(B2_11,`1` = 999, `2` = 1, `24` = 2,`25` = 3),
       B2_12 = recode(B2_12,`1` = 999, `2` = 1, `24` = 2,`25` = 3),
       B2_13 = recode(B2_13,`1` = 999, `2` = 1, `24` = 2,`25` = 3),
       order_B2 = paste(B2_DO_1, B2_DO_2, B2_DO_3, B2_DO_4,
                        B2_DO_5, B2_DO_6, B2_DO_7, B2_DO_8,
                        B2_DO_9, B2_DO_10, B2_DO_11, B2_DO_12,
                        B2_DO_13, sep = "|"))

B3 <- d %>%
  select(iisID, X1_B3_3, X2_B3_3, X3_B3_3,
         X4_B3_3, X5_B3_3, X6_B3_3, X7_B3_3,
         X8_B3_3, X9_B3_3, X10_B3_3, X11_B3_3,
         X12_B3_3, X13_B3_3,
         X1_B3_4, X2_B3_4, X3_B3_4,
         X4_B3_4, X5_B3_4, X6_B3_4, X7_B3_4,
         X8_B3_4, X9_B3_4, X10_B3_4, X11_B3_4,
         X12_B3_4, X13_B3_4) %>%
  pivot_longer(cols = X1_B3_3:X13_B3_4,
               names_to = "variable") %>%
  separate(variable, c("variable", "question"), "_", extra = "merge")  %>%
  group_by(variable) %>%
  mutate(row = row_number()) %>%
  pivot_wider(names_from = question, values_from = value) %>%
  unite("B3", B3_3:B3_4, remove = T, na.rm = T) %>%
  mutate(B3 = ifelse(B3 == "", "999", B3),
         B3 = as.numeric(B3))

B3 <- B3 %>%
  group_by(variable) %>%
  mutate(row = row_number()) %>%
  pivot_wider(names_from = variable, values_from = B3) %>%
  select(iisID, B3_1 = X1,
         B3_2 = X1, B3_3 = X3, B3_4 = X4, B3_5 = X5,
         B3_6 = X6, B3_7 = X7, B3_8 = X1, B3_9 = X9,
         B3_10 = X10, B3_11 = X11, B3_12 = X12, B3_13 = X13) %>%
  distinct(iisID, .keep_all = TRUE)

B <- d %>%
  select(iisID, B2_1:B2_13, order_B2)
B <- left_join(B, B3, by = "iisID")
rm(B3)
```

Recode Block H Other Issues

```r
H <- d %>%
  select(iisID, H1, H2, H7, H7a, H7b:H10) %>%
  mutate(H7 = recode(H7, `3` = 2, `4` = 3, `5` = 4))
```

Recode Block I News Consumption

```r
I <- d %>%
  select(iisID, I1_1:I1_9, I2_1:I2_13, I2_other = I2_13_TEXT,
         I3_1:I3_14, I3_15 = I3_17,I3_16 = I3_15, I3_other = I3_15_TEXT,
         I4_1:I4_8, I4_9 = I4_15, I4_other_blogs = I4_8_TEXT,
         I4_other_sites = I4_15_TEXT,
         I5_1:I5_11, I5_other = I5_11_TEXT,
         I6_1:I6_6, I6_other = I6_6_TEXT,
         I7_1:I7_9, I7_10 = I7_11, I7_other = I7_11_TEXT)
```

Merge & Save Data

```r
df <- left_join(BG, A, by = "iisID")
df <- left_join(df, B, by = "iisID")
df <- left_join(df, H, by = "iisID")
df <- left_join(df, I, by = "iisID") %>%
  distinct(iisID, .keep_all = T) %>%
  drop_na(A1) %>%
  filter(iisID !="007",
         iisID !="009") %>%
  mutate(iisID = as_numeric(iisID)) %>%
  drop_na(iisID)

write_csv(df, "../../data/intermediate/wave1.csv")
rm(A, B, H, I)
```