



Ακρωνύμιο πρότασης:

DataPopEU

Τίτλος πρότασης:

Καινοτόμες μέθοδοι και δεδομένα υψηλής ποιότητας για τη μελέτη του λαϊκισμού και του Ευρωσκεπτικισμού

Π. 4.3 Έκθεση και κώδικας R για την εναρμόνιση των δεδομένων

D4.3 Harmonization R code report

Αύγουστος 2022

Εισαγωγή

Κατά τη διαδικασία συλλογής δεδομένων δοκιμάστηκαν διάφορες παράμετροι που θα μπορούσαν να βελτιώσουν την ποιότητα των δεδομένων και τα ποσοστά απόκρισης. Αυτό έγινε εφικτό με τη δημιουργία διαφορετικών εκδόσεων των ερωτηματολογίων και την αποστολή διαφορετικών προσκλήσεων σε διαφορετικές ομάδες συμμετεχόντων. Για το λόγο αυτό, χρειάστηκε να υλοποιηθεί κώδικας σε γλώσσα R που οδηγεί στην εναρμόνιση αυτών των δεδομένων. Η παρούσα έκθεση έχει ως στόχο την παρουσίαση του σχετικού κώδικα.

Κώδικας R για υποψήφιους

Τα δεδομένα για τους υποψηφίους συγκεντρώθηκαν από δύο διαφορετικές υλοποιήσεις του limesurvey (epolls και epoll) και από συνολικά 13 διαφορετικά ερωτηματολόγια στα οποία δοκιμάστηκαν διάφορες παράμετροι που θα μπορούσαν να βελτιώσουν την ποιότητα των δεδομένων και τα ποσοστά απόκρισης. Κάποιες πρώτες συγκρίσεις μεταξύ των παραμέτρων και ευρήματα για την αποτελεσματικότητά τους έχουν παρουσιαστεί σε διεθνή συνέδρια (Tsigkou et al., 2020)

Αρχικά χρειάστηκε κώδικας για την ανάγνωση των αρχείων:

```
library(tidyverse)
library(fs)
library(mokken)
library(psych)
select<-dplyr::select
filter<-dplyr::filter

# function to remove ] from names and replace [ with _
change_names<-function(tb){
  names(tb)<-str_remove(names(tb),"]")
  names(tb)<-str_replace(names(tb), "\\[", "_")
  return(tb)
}
data_dir <- "epoll"
epoll<-data_dir %>%
  dir_ls(regexp = "\\..csv$") %>%
  map_dfr(read_csv, col_types = cols(submitdate="?",
  startdate = "?",
  datestamp = "?",
  .default = col_character()), .id = "source")

epoll<- change_names(epoll)

data_dir <- "epolls"
epolls<-data_dir %>%
  dir_ls(regexp = "\\..csv$") %>%
  map_dfr(read_csv, col_types = cols(submitdate="?",
  startdate = "?",
  datestamp = "?",
  .default = col_character()), .id = "source")

epolls<- change_names(epolls)
```

Στη συνέχεια χρειάστηκε κώδικας για την ένωση όλων των αρχείων και την εναρμόνιση των δεδομένων καθώς οι διαφορετικές εκδόσεις των ερωτήσεων είχαν αναγκαστικά και διαφορετικά

ονόματα μεταβλητών. Για παράδειγμα το limesurvey δίνει διαφορετικό όνομα σε matrix και single questions

```
data<-epoll %>% bind_rows(epolls)
# correct different names between matrix and single questions
datac<-data %>% mutate(B5a=coalesce(B5a_B5a, B5a),
  B5b =coalesce( B5a_B5b, B5b),
  B5c =coalesce( B5a_B5c, B5c),
  B5d =coalesce(B5a_B5d, B5d),
  B5e=coalesce( B5b_B5e, B5e),
  B5f =coalesce( B5b_B5f, B5f),
  B5g =coalesce(B5b_B5g, B5g),
  B5h =coalesce( B5b_B5h, B5h),
  B8b=coalesce( B8_B8b, B8b),
  B8c=coalesce( B8_B8c, B8c),
  B8d=coalesce(B8_B8d, B8d),
  B8e=coalesce( B8_B8e, B8e),
  B8f=coalesce( B8_B8f, B8f),
  B8g=coalesce( B8_B8g, B8g ),
  B8h=coalesce( B8_B8h, B8h ),
  B9a=coalesce( B9_B9a, B9a),
  B9b=coalesce( B9_B9b, B9b),
  B9c=coalesce( B9_B9c, B9c),
  B9d=coalesce( B9_B9d, B9d),
  C2a=coalesce( C2a_C2a, C2a),
  C2b=coalesce( C2a_C2b, C2b),
  C2c=coalesce( C2a_C2c, C2c),
  C2d=coalesce( C2a_C2d, C2d),
  C2e=coalesce( C2a_C2e, C2e),
  C2f=coalesce( C2b_C2f, C2f),
  C2g=coalesce( C2b_C2g, C2g),
  C2h=coalesce( C2b_C2h, C2h),
  C2i=coalesce( C2b_C2i, C2i),
  C2j=coalesce(C2b_C2j, C2j),
  C3a=coalesce(C3_C3a, C3a ),
  C3b=coalesce( C3_C3b, C3b ),
  C3c=coalesce(C3_C3c, C3c),
  C3d=coalesce( C3_C3d, C3d),
  C4a=coalesce( C4_C4a, C4a),
  C4b=coalesce( C4_C4b, C4b),
  C4c=coalesce( C4_C4c, C4c),
  C4d=coalesce( C4_C4d, C4d),
  C4e=coalesce(C4_C4e, C4e),
  D6a=coalesce( D6_D6a, D6a),
  D6b=coalesce( D6_D6b, D6b),
  D6c=coalesce( D6_D6c, D6c),
  D7a=coalesce( D7ae_D7a, D7a),
  D7b=coalesce( D7ae_D7b, D7b),
  D7c=coalesce( D7ae_D7c, D7c),
  D7d=coalesce( D7ae_D7d, D7d),
  D7e=coalesce( D7ae_D7e, D7e),
  D7f=coalesce( D7fi_D7f, D7f),
  D7g=coalesce( D7fi_D7g, D7g),
  D7h=coalesce( D7fi_D7h, D7h),
  D7i=coalesce( D7_SQ001 , D7i ),
  D7j=coalesce( D7_SQ002, D7j),
  D7k=coalesce( D7_SQ003, D7k),
  D7l=coalesce( D7fi_D7i, D7l),
```

```
E13a=coalesce( E13a_E13a, E13a),
E13b=coalesce( E13a_E13b, E13b),
E13c=coalesce( E13a_E13c, E13c ),
E13d=coalesce( E13a_E13d, E13d ),
E13e=coalesce( E13a_E13e, E13e ),
E13f=coalesce( E13b_E13f, E13f ),
E13g=coalesce( E13b_E13g, E13g ),
E13h=coalesce(E13b_E13h, E13h ),
E13i=coalesce( E13b_E13i, E13i ),
E13j=coalesce( E13b_E13j, E13j ),
ECS31a=coalesce( ECS31_ECS31a,ECS31a ),
ECS31b=coalesce( ECS31_ECS31b, ECS31b ),
ECS31c=coalesce( ECS31_ECS31c, ECS31c ),
ECS31d=coalesce( ECS31_ECS31d, ECS31d ),
EU1=coalesce( ECS31_EU1, EU1 ),
EU3=coalesce(ECS31_EU3, EU3 ),
EU6a=coalesce(EU6_EU6a, EU6a ),
EU6b=coalesce( EU6_EU6b,EU6b ),
EU6c=coalesce( EU6_EU6c, EU6c ),
EU6d=coalesce( EU6_EU6d, EU6d ),
EU6e=coalesce( EU6_EU6e, EU6e ),
EU6f=coalesce( EU6_EU6f,EU6f ),
excl1=coalesce( IMexcl_Excl1,excl1 ),
excl2=coalesce( IMexcl_Excl2, excl2 ),
excl3=coalesce( IMexcl_Excl3, excl3 ),
excl4=coalesce( IMexcl_Excl4, excl4 ),
excl5=coalesce( IMexcl_Excl5, excl5 ),
IM1=coalesce( IMexcl_IM1, IM1 ),
IM2=coalesce( IMexcl_IM2, IM2 ),
NatId1=coalesce( NatId_NatId1, NatId1 ),
NatId2=coalesce( NatId_NatId2,NatId2 ),
NatId3=coalesce( NatId_NatId3,NatId3 ),
NatId4=coalesce( NatId_NatId4, NatId4 ),
P2=coalesce(P1_P2, P2 ),
P3=coalesce( P1_P3, P3 ),
P6=coalesce( P1_P6, P6 ),
TP1=coalesce( P1_TP1, TP1 ),
TP2=coalesce( P1_TP2, TP2 ),
TP3=coalesce( P1_TP3, TP3 ),
TP4=coalesce(P2_TP4, TP4 ),
TP5=coalesce(P2_TP5, TP5 ),
TP6=coalesce( P2_TP6, TP6 ),
TP7=coalesce( P2_TP7,TP7 ),
TP8=coalesce( P2_TP8, TP8 ),
TP9=coalesce( P2_TP9, TP9 ),
Pop1a=coalesce( Pop1_Pop1a, Pop1a ),
Pop1b=coalesce( Pop1_Pop1b, Pop1b ),
Pop1c=coalesce( Pop1_Pop1c, Pop1c ),
Q04a=coalesce( Q04_Q04a, Q04a ),
Q04b=coalesce( Q04_Q04b, Q04b ),
Q04c=coalesce( Q04_Q04c, Q04c ),
Q04d=coalesce( Q04_Q04d, Q04d ),
Q04e=coalesce( Q04_Q04e,Q04e ),
Q04f=coalesce(Q04_Q04f, Q04f ),
D7a=coalesce(QPop2_D7a,D7a),
D7b=coalesce(QPop1_D7b,D7b),
D7c=coalesce(QPop1_D7c,D7c),
D7e=coalesce(QPop2_D7e,D7e),
D7f=coalesce(QPop2_D7f,D7f),
```

```
D7g=coalesce(QPop2_D7g,D7g),
D7h=coalesce(QPop1_D7h,D7h),
D7i=coalesce(QPop1_D7i,D7i),
D7j=coalesce(QPop2_D7j,D7j),
D7k=coalesce(QPop1_D7k,D7k),
P2=coalesce(QPop2_P2,P2),
P3=coalesce(QPop1_P3,P3),
P6=coalesce(QPop1_P6,P6),
TP4=coalesce(QPop1_TP4,TP4),
TP5=coalesce(QPop1_TP5,TP5),
TP6=coalesce(QPop1_TP6,TP6),
Pop1a=coalesce(QPop1_Pop1a,Pop1a),
Pop1b=coalesce(QPop2_Pop1b,Pop1b),
Q04a=coalesce(QPop1_Q04a,Q04a),
Q04b=coalesce(QPop1_Q04b,Q04b),
Q04c=coalesce(QPop1_Q04c,Q04c),
Q04d=coalesce(QPop1_Q04d,Q04d),
Q04e=coalesce(QPop1_Q04e,Q04e),
Q04f=coalesce(QPop2_Q04f,Q04f),
C2b=coalesce(C2_C2b,C2b),
C2h=coalesce(C2_C2h,C2h),
C2a=coalesce(C2_C2a,C2a),
C2d=coalesce(C2_C2d,C2d),
C2f=coalesce(C2_C2f,C2f),
C2i=coalesce(C2_C2i,C2i),
C2j=coalesce(C2_C2j,C2j),
IM1=coalesce(C2_IM1,IM1),
IM2=coalesce(C2_IM2,IM2),
excl2=coalesce(C2_Excl2,excl2),
excl3=coalesce(C2_Excl3,excl3),
excl4=coalesce(C2_Excl4,excl4),
excl5=coalesce(C2_Excl5,excl5)
```

)%>%

```
select(-c(
  B5a_B5a,
  B5a_B5b,
  B5a_B5c,
  B5a_B5d,
  B5b_B5e,
  B5b_B5f,
  B5b_B5g,
  B5b_B5h,
  B8_B8b,
  B8_B8c,
  B8_B8d,
  B8_B8e,
  B8_B8f,
  B8_B8g,
  B8_B8h,
  B9_B9a,
  B9_B9b,
  B9_B9c,
  B9_B9d,
  C2a_C2a,
  C2a_C2b,
  C2a_C2c,
  C2a_C2d,
  C2a_C2e,
  C2b_C2f,
```

C2b_C2g,
C2b_C2h,
C2b_C2i,
C2b_C2j,
C3_C3a,
C3_C3b,
C3_C3c,
C3_C3d,
C4_C4a,
C4_C4b,
C4_C4c,
C4_C4d,
C4_C4e,
D6_D6a,
D6_D6b,
D6_D6c,
D7ae_D7a,
D7ae_D7b,
D7ae_D7c,
D7ae_D7d,
D7ae_D7e,
D7fi_D7f,
D7fi_D7g,
D7fi_D7h,
D7_SQ001,
D7_SQ002,
D7_SQ003,
D7fi_D7i,
E13a_E13a,
E13a_E13b,
E13a_E13c,
E13a_E13d,
E13a_E13e,
E13b_E13f,
E13b_E13g,
E13b_E13h,
E13b_E13i,
E13b_E13j,
ECS31_ECS31a,
ECS31_ECS31b,
ECS31_ECS31c,
ECS31_ECS31d,
ECS31_EU1,
ECS31_EU3,
EU6_EU6a,
EU6_EU6b,
EU6_EU6c,
EU6_EU6d,
EU6_EU6e,
EU6_EU6f,
IMexcl_Excl1,
IMexcl_Excl2,
IMexcl_Excl3,
IMexcl_Excl4,
IMexcl_Excl5,
IMexcl_IM1,
IMexcl_IM2,
NatId_NatId1,
NatId_NatId2,

NatId_NatId3,
NatId_NatId4,
P1_P2,
P1_P3,
P1_P6,
P1_TP1,
P1_TP2,
P1_TP3,
P2_TP4,
P2_TP5,
P2_TP6,
P2_TP7,
P2_TP8,
P2_TP9,
Pop1_Pop1a,
Pop1_Pop1b,
Pop1_Pop1c,
Q04_Q04a,
Q04_Q04b,
Q04_Q04c,
Q04_Q04d,
Q04_Q04e,
Q04_Q04f,
QPop2_D7a,
QPop1_D7b,
QPop1_D7c,
QPop2_D7e,
QPop2_D7f,
QPop2_D7g,
QPop1_D7h,
QPop1_D7i,
QPop2_D7j,
QPop1_D7k,
QPop2_P2,
QPop1_P3,
QPop1_P6,
QPop1_TP4,
QPop1_TP5,
QPop1_TP6,
QPop1_Pop1a,
QPop2_Pop1b,
QPop1_Q04a,
QPop1_Q04b,
QPop1_Q04c,
QPop1_Q04d,
QPop1_Q04e,
QPop2_Q04f,
C2_C2b,
C2_C2h,
C2_C2a,
C2_C2d,
C2_C2f,
C2_C2i,
C2_C2j,
C2_IM1,
C2_IM2,
C2_Exc l2,
C2_Exc l3,
C2_Exc l4,

```
C2_Exc15
))
```

Επίσης έγινε μετατροπή και άλλων μεταβλητών έτσι ώστε να έχουν τη μορφή που απαιτεί το CCS

```
#D4 in epoll is of type ranking; it is converted to match epolls and the CCS
codebook
```

```
converted_ranking <- datac %>%
  select(source, id, D4_1:D4_5) %>%
  pivot_longer(-c(source, id)) %>%
  separate(name, into=c("var", "pos")) %>%
  filter(!is.na(value)) %>%
  unite(new_var, var, value) %>%
  pivot_wider(names_from = new_var, values_from=pos)
```

```
datac<- datac %>%
  left_join(converted_ranking) %>%
  select(-c(D4_1:D4_5)) %>%
  mutate(D4_D4a=coalesce(D4_A1,D4_D4a),
  D4_D4b=coalesce(D4_A2,D4_D4b),
  D4_D4c=coalesce(D4_A3,D4_D4c),
  D4_D4d=coalesce(D4_A4,D4_D4d),
  D4_D4e=coalesce(D4_A5,D4_D4e))%>%
  select(-c(D4_A1,
  D4_A2,
  D4_A3,
  D4_A4,
  D4_A5))
```

```
datac<-datac %>%
  rename_with(~ str_replace(.x, "_other", "other")) %>%
  rename_with(~ str_remove(.x, "_SQ00")) %>%
  rename_with(~ str_remove(.x, "A4a1_")) %>%
  rename_with(~ str_remove(.x, "A4b1_")) %>%
  rename_with(~ str_remove(.x, "A5_")) %>%
  rename_with(~ str_remove(.x, "EU8_")) %>%
  rename_with(~ str_remove(.x, "B7a_")) %>%
  rename_with(~ str_remove(.x, "B10_")) %>%
  rename_with(~ str_remove(.x, "B13_")) %>%
  rename_with(~ str_remove(.x, "D4_"))
```

```
datac<-datac %>%
  rename_with(~ str_remove(.x, "_"))
```

```
datai<-datac %>%
  mutate(across(integer_vars, ~str_remove(., "A"))) %>%
  mutate(across(integer_vars, as.integer))
```


Επόμενο βήμα ήταν η δημιουργία του δείκτη ευρωσκεπτικισμού:

```
reverse_item<-function(x, n_options=5) {
  out<-(n_options+1)-x
  return(out)
}

data1 %>% names() %>% str_subset("EUnewadd")
data1 %>%
  mutate(DPEU8=round((rowMeans(select(.,EU8b:EU8h), na.rm = TRUE)-2)*(-4)+1, 0)) %>%
  mutate(DPEU1 = case_when(C6 == 1 ~ 5,
    C6 == 2 ~ 1,
    C6 == 3 ~ 3,
    TRUE ~ NA_real_)) %>%
  mutate(DPEU2 = case_when(EU2 == 1 ~ 5,
    EU2 == 2 ~ 3,
    EU2 == 3 ~ 1,
    TRUE ~ NA_real_)) %>%
  mutate(DPEU4 = case_when(EU4 == 1 ~ 1,
    EU4 == 2 ~ 2,
    EU4 == 3 ~ 4,
    EU4 == 4 ~ 5,
    TRUE ~ NA_real_)) %>%
  mutate(DPEU7 = case_when(D2 == 1 ~ 5,
    D2 == 2 ~ 4,
    D2 == 3 ~ 2,
    D2 == 4 ~ 1,
    TRUE ~ NA_real_)) %>%
  #mutate(positiveEU = C6n + EU2n) %>%
  #select(C6n , EU2n) %>% cor(., use="complete.obs")
  mutate(DPEU3=round(1+4*(C7-1)/10, 0)) %>%
  select(DPEU1, DPEU2, DPEU3, DPEU4, ECS31a, EU6a:EU6d, DPEU7, DPEU8, ECS31b,
ECS31c, ECS31d, EU1, EU3) %>%
  select(!starts_with("ASK")) %>%
  select(!starts_with("RA")) %>%
  select(!starts_with("ST")) %>%
  mutate(across(c(ECS31b, ECS31d), reverse_item)) %>% # reversed
  #drop_na() %>%
  as.data.frame() %>%
  #aisp() %>%
  select(-c(ECS31b, DPEU3, DPEU4)) %>%
  #coefH(se = TRUE, nice.output = TRUE)
  #alpha(.,check.keys=TRUE)
  #fa.parallel(fa="fa")
  #fa()
  mutate(euroscepticism=6-rowMeans(., na.rm = T)) %>%
  pull(euroscepticism) -> data1$euroscepticism
```

Ενώ με παραπλήσια διαδικασία δημιουργήθηκε και ο δείκτης λαϊκισμού:

```
pop_vars<-c('D7i', 'D7j',
  #'D7k',
  'P2',
  'P3',
  #'P6',
  'TP1',
```

```
'TP2',  
'TP3',  
'TP4',  
'TP5',  
'TP6',  
'TP7',  
'TP8',  
'TP9',  
'Pop1a',  
'Pop1b',  
'Pop1c',  
'Q04a',  
'Q04b',  
'Q04c',  
'Q04d',  
'Q04e',  
'Q04f')
```

```
datai %>%  
  select(pop_vars) %>%  
  mutate(across(c(Q04b, TP2, TP5, TP8), reverse_item)) %>% # reversed  
  select(-c( TP1, TP2, TP7, TP8, TP9, Pop1c, Q04d)) %>% #remove irrelevant  
  select(-c(D7j, P2, TP3)) %>%  
  #select(c( TP1, TP2, Pop1c)) %>%  
  #drop_na() %>% #count(Pop1c)  
  as.data.frame() %>%  
  #aisp(search="ga")  
  #coefH(se = TRUE, nice.output = TRUE)  
  #alpha(., check.keys=TRUE)  
  #fa.parallel(fa="fa")  
  #fa()  
  mutate(populism=rowMeans(., na.rm = T)) %>%  
  pull(populism) -> datai$populism
```

Και τέλος η δημιουργία του τελικού ενιαίου αρχείου για τους υποψηφίους:

```
data_out<-datai %>%  
  select(all_of(first_vars), all_of(eu_vars),  
         all_of(pop_vars))  
  
library(haven)  
  
data_out %>% write_sav("candidates2019.sav")  
  
#keep all items that have been found by MSA in voters  
datai %>%  
  mutate(across(c(Q04b, TP2, TP5, TP8), reverse_item)) %>%  
  select(A1f,  
         Q04a=D7i,  
         Q04b=Q04a,  
         Q04c=Q04b,  
         Q04d=Q04c,  
         Q04f,  
         Q04g=Q04e,  
         AMZ_POP1=D7j,  
         AMZ_POP3=P2,  
         AMZ_POP4=P3,
```

```

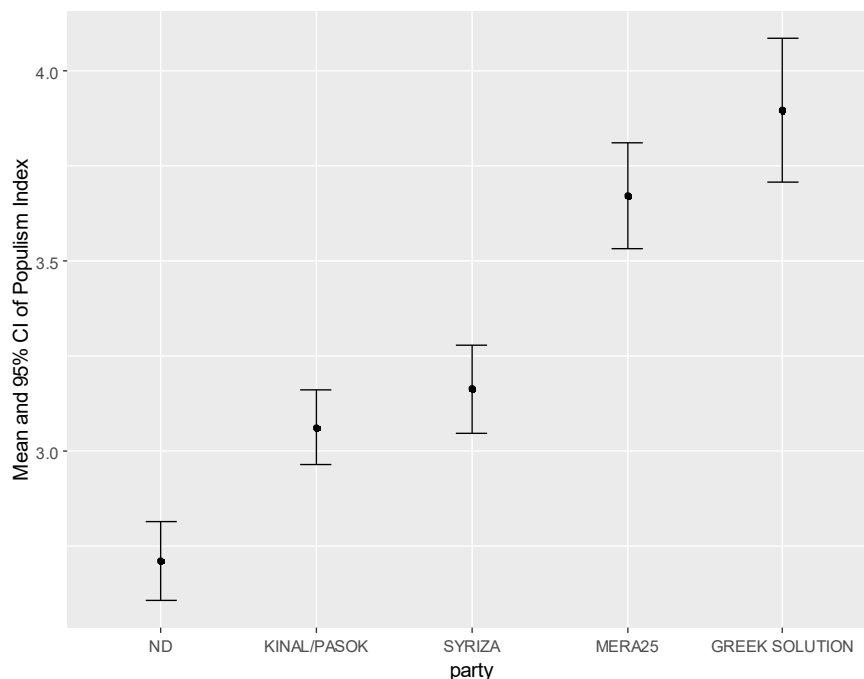
    AMZ_POP5=Pop1a,
    TP3,
    TP4,
    TP5,
    TP6,
    SAK_POP7=Pop1b) %>%
mutate(populism=rowMeans(select(.,Q04a:SAK_POP7), na.rm = T)) -> CCS_populism

save(CCS_populism, file="CCS-populism.Rdata")

CCS_populism %>%
mutate(party=case_when(
  A1f== 1~ "ND",
  A1f== 2~ "SYRIZA",
  A1f== 3~ "KINAL/PASOK",
  A1f== 5~ "GREEK SOLUTION",
  A1f== 6~ "MERA25" ,
  TRUE ~ NA_character_)) %>%
mutate(party=factor(party,
                    levels = c("ND", "KINAL/PASOK", "SYRIZA", "MERA25",
"GREEK SOLUTION"))) %>%
#dplyr::count(party)
dplyr::filter(!is.na(party)) %>%
ggplot(aes(x=party, y=populism)) +
  geom_point(stat="summary", fun="mean") +
  geom_errorbar(stat="summary", fun.data="mean_se", fun.args = list(mult =
1.96), width=0.2)+
  ylab("Mean and 95% CI of Populism Index")

```

Με τον οποίο εκτός από τη δημιουργία των αρχείων έχουμε τη δυνατότητα να παρατηρήσουμε πως τοποθετούνται οι υποψήφιοι των κομμάτων του Ελληνικού Κοινοβουλίου:



Κώδικας για δεδομένα ψηφοφόρων

Αρχικά γίνεται ανάγνωση των δεδομένων

```
library(tidyverse)
library(lubridate)
# download all survey responses and read it

survey112674<-read_csv("local_only/surveys/results-survey112674.csv",
                      col_types = cols(.default = col_character()))

survey115963<-read_csv("local_only/surveys/results-survey115963.csv",
                      col_types = cols(.default = col_character()))

surveys<-survey115963 %>% bind_rows(survey112674)

survey339727<-read_csv("local_only/surveys/results-survey339727.csv",
                      col_types = cols(.default = col_character()))

surveys<-surveys %>% bind_rows(survey339727)

permitted<-surveys %>% filter(Permission=="Y")

eligible<-permitted %>% filter(eligibility==1 | eligibility=="Y")

names(eligible)<-str_replace(names(eligible), "\\[", "_")
names(eligible)<-str_replace(names(eligible), "\\]", "")
```

Μετά το καθάρισμα των δεδομένων (Andreadis, 2021; Andreadis & Andreadis, 2022b, 2022a) δημιουργούμε τον δείκτη λαϊκισμού:

```
final %>%
  mutate(across(Q04_Q04a:Pop1_CCSD7j, as.integer)) %>%
  mutate(across(c(Q04_Q04a, Q04_Q04b, Q04_Q04d:Q04_Q04g, P1_TP2, P2_TP5, P2_TP8),
reverse_item)) %>% # reversed
  select(Q04_Q04a:Pop1_CCSD7j) %>%
  select(-P1_P6) %>% # included in the questionnaire by mistake
  select(Q04a=Q04_Q04a,
        Q04b=Q04_Q04b,
        Q04c=Q04_Q04c,
        Q04d=Q04_Q04d,
        Q04e=Q04_Q04e,
        Q04f=Q04_Q04f,
        Q04g=Q04_Q04g,
        AMZ_POP1=Pop1_CCSD7j,
        AMZ_POP3=P1_P2,
        AMZ_POP4=P1_P3,
        AMZ_POP5=Pop1_Pop1a,
        TP1=P1_TP1,
        TP2=P1_TP2,
        TP3=P1_TP3,
        TP4=P2_TP4,
        TP5=P2_TP5,
        TP6=P2_TP6,
```

```

        TP7=P2_TP7,
        TP8=P2_TP8,
        TP9=P2_TP9,
        SAK_POP7=Pop1_Pop1b,
        SAK_POP8=Pop1_Pop1c) %>%
select(-c(Q04e, TP7, TP8, TP9, TP1, TP2, SAK_POP8)) %>% #remove irrelevant
#drop_na() %>%
as.data.frame() %>%
#polychoric()
#aisp()
#coefH(se = TRUE, nice.output = TRUE)
#alpha(., check.keys=TRUE)
#fa.parallel(fa="fa")
#fa()
mutate(populism=rowMeans(., na.rm = T)) %>%
pull(populism) -> final$populism

final %>%
mutate(across(Q04_Q04a:Pop1_CCSD7j, as.integer)) %>%
mutate(across(c(Q04_Q04a, Q04_Q04b, Q04_Q04d:Q04_Q04g, P1_TP2, P2_TP5, P2_TP8),
reverse_item)) %>%
select(Q12LHb,
Q12LHbf,
Q04a=Q04_Q04a,
Q04b=Q04_Q04b,
Q04c=Q04_Q04c,
Q04d=Q04_Q04d,
Q04f=Q04_Q04f,
Q04g=Q04_Q04g,
AMZ_POP1=Pop1_CCSD7j,
AMZ_POP3=P1_P2,
AMZ_POP4=P1_P3,
AMZ_POP5=Pop1_Pop1a,
TP3=P1_TP3,
TP4=P2_TP4,
TP5=P2_TP5,
TP6=P2_TP6,
SAK_POP7=Pop1_Pop1b) %>%
filter(Q12LHb<7) %>%
filter(Q12LHb!="-oth-")%>%
mutate(across(Q04a:SAK_POP7, as.integer)) %>%
mutate(populism=rowMeans(select(., Q04a:SAK_POP7), na.rm = T)) ->voter_populism

save(voter_populism, file="voter-populism.Rdata")

```

και τον δείκτη του ευρωσκεπτικισμού

```

final %>%
mutate(across(
c(EU2,
EU5,
EU6_EU6a,
EU6_EU6b,
EU6_EU6c,
EU6_EU6d,

```

```

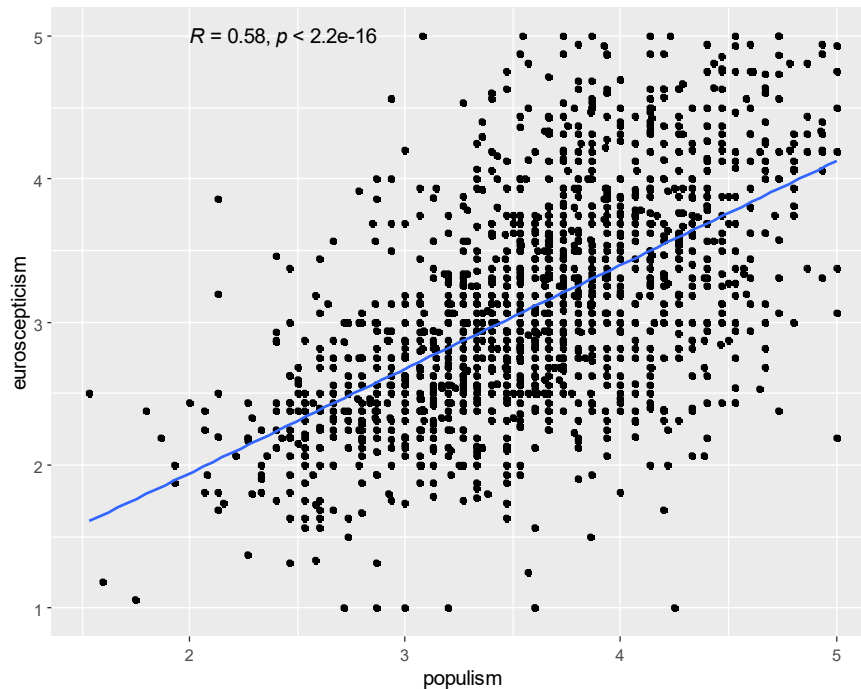
EU6_EU6e,
EU6_EU6f,
  EU8_EU8b,
  EU8_EU8d,
  EU8_EU8e,
  EU8_EU8f,
  EU8_EU8g,
  EU8_EU8h,
  EU8_EU8i),
  ~ str_replace(., "A", "")
)) %>%
mutate(across(EU8_EU8b:EU8_EU8i, ~ str_replace(., "3", "2"))) %>% #correct A3 in
339727
#select(C6:ECS31_ECS31d) %>%
mutate(across(everything(), as.integer)) %>%
mutate(DPEU8=round((rowMeans(select(.,EU8_EU8b:EU8_EU8h), na.rm = TRUE)-2)*(-
4)+1, 0)) %>%
mutate(DPEU1 = case_when(C6 == 1 ~ 5,
  C6 == 2 ~ 1,
  C6 == 3 ~ 3,
  TRUE ~ NA_real_)) %>%
mutate(DPEU2 = case_when(EU2 == 1 ~ 5,
  EU2 == 2 ~ 3,
  EU2 == 3 ~ 1,
  TRUE ~ NA_real_)) %>%
mutate(DPEU4 = case_when(EU4 == 1 ~ 1,
  EU4 == 2 ~ 2,
  EU4 == 3 ~ 4,
  EU4 == 4 ~ 5,
  TRUE ~ NA_real_)) %>%
mutate(DPEU7 = case_when(CSES2 == 1 ~ 5,
  CSES2 == 2 ~ 4,
  CSES2 == 3 ~ 2,
  CSES2 == 4 ~ 1,
  TRUE ~ NA_real_)) %>%
#mutate(positiveEU = C6n + EU2n) %>%
#select(C6n , EU2n) %>% cor(., use="complete.obs")
mutate(DPEU3=round(1+4*(C7-1)/10, 0)) %>% #count(DPEU3)
select(DPEU1 , DPEU2, DPEU3, DPEU4, ECS31_ECS31a, EU6_EU6a:EU6_EU6f, DPEU7,
DPEU8, ECS31_ECS31b:ECS31_ECS31d, EU1_EU1, EU1_EU3 ) %>%
mutate(across(c(ECS31_ECS31b, ECS31_ECS31d), reverse_item)) %>% # reversed
# select(-c(C6, C6n, EU2, EU2n, C7)) %>%

select(-c(DPEU3, ECS31_ECS31b)) %>%
#drop_na() %>%
as.data.frame() %>%
#aisp()
#coefH(se = TRUE, nice.output = TRUE)
#alpha(.,check.keys=TRUE)
#fa.parallel(fa="fa")
#fa()
mutate(euroscepticism=6-rowMeans(., na.rm = T)) %>%
pull(euroscepticism) -> final$euroscepticism

final %>%
ggplot(aes(y=euroscepticism, x=populism))+
geom_point()+
geom_smooth(method="lm", na.rm=TRUE)+
stat_cor(method = "pearson", label.x = 2, label.y = 5)

```

δίνοντας τη δυνατότητα για την αποτύπωση και των δύο δεικτών σε ένα διάγραμμα διασποράς.



Παράδειγμα ένωσης αρχείων ψηφοφόρων και υποψηφίων

Ενώ με τον παρακάτω κώδικα δίνεται παράδειγμα ένωσης των δύο αρχείων ώστε να γίνει η σύγκριση μεταξύ ψηφοφόρων και υποψηφίων όσον αφορά στο θέμα του λαϊκισμού.

```
library(tidyverse)
library(knitr)
library(forcats)
# dgof includes new ks.test
library(dgof)
#Matching includes ks.boot test
library(Matching)
select<-dplyr::select

load("voter-populism.Rdata")
load("CCS-populism.Rdata")

voter_populism %>%
  group_by(Q12LHbf) %>%
  summarize(n=n(), populism=mean(populism, na.rm=TRUE))

CCS_populism %>%
  group_by(A1f) %>%
  summarize(n=n(), populism=mean(populism, na.rm=TRUE))

candidates<-CCS_populism %>%
  select(party_code=A1f, populism) %>%
  filter(party_code<7)%>%
```

```

mutate(Group=2) %>%
mutate(party=case_when(
party_code== 1~ "NEW DEMOCRACY (ND)",
party_code== 2~ "COALITION OF THE RADICAL LEFT (SYRIZA)",
party_code== 3~ "MOVEMENT FOR CHANGE (KINAL)",
party_code== 4~ "COMMUNIST PARTY OF GREECE (KKE)",
party_code== 5~ "GREEK SOLUTION",
party_code== 6~ "EUROPEAN REALISTIC DISOBEDIENCE FRONT (MERA25)" )
) %>%
mutate(party=factor(party, levels=levels(voter_populism$Q12LHbf)))%>%
select( party_code, party, populism, Group) %>%
filter(!is.na(populism))

candidates %>% pull(populism) %>% summary()

voters<-voter_populism %>%
mutate(party_code=as.integer(Q12LHb)) %>%
select(party_code, party=Q12LHbf, populism) %>%
mutate(Group=1) %>%
filter(!is.na(populism))
voters %>% pull(populism) %>% summary()

data<-bind_rows(voters, candidates)
data$Group<- as.factor(data$Group)
levels(data$Group)
data$Group<-data$Group %>%
fct_recode("Voters" = "1", "Candidates" = "2")

data<-data %>% filter(party_code!=4)

#Clean L/R self-placement
table(data$populism, useNA="always")
table(data$Group, useNA="always")

data<-data%>%mutate(party=case_when(
party_code== 1~ "ND",
party_code== 2~ "SYRIZA",
party_code== 3~ "KINAL/PASOK",
party_code== 4~ "KKE",
party_code== 5~ "GREEK SOLUTION",
party_code== 6~ "MERA25" )
) %>%
mutate(party=factor(party,
levels = c("ND", "KINAL/PASOK", "SYRIZA", "MERA25", "GREEK
SOLUTION")))

data %>% count(party)

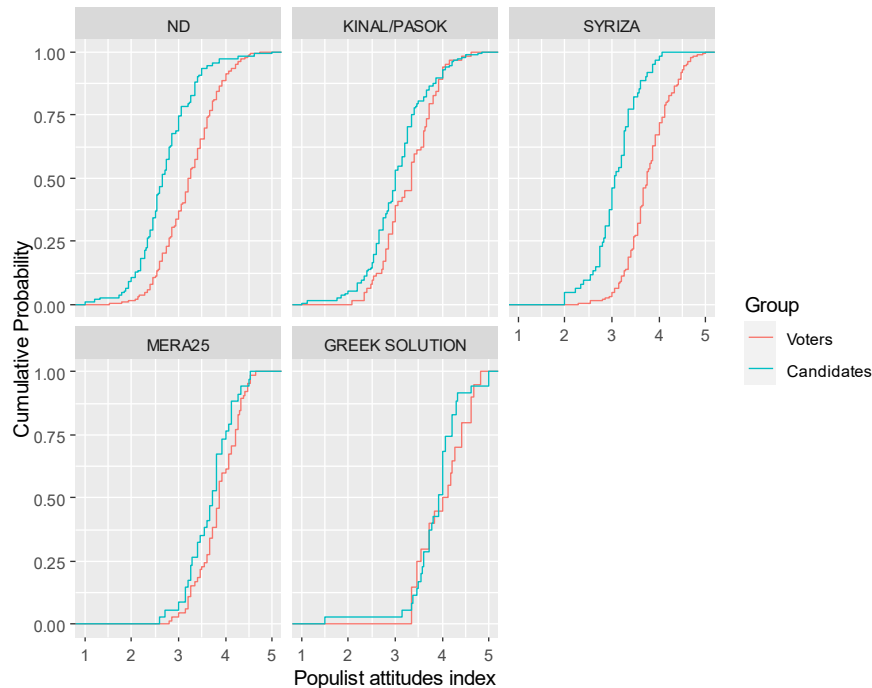
#we need to set the seed to get stable simulation results
set.seed(1234)
data %>%
group_by(party) %>%
#summarize(ks=ks.test(lrsp[Group=="Voters"], lrsp[Group=="Candidates"],
simulate.p.value=TRUE, B=10000)$p.value)
summarize(ks=ks.boot(populism[Group=="Voters"], populism[Group=="Candidates"],
nboots=1000)$ks.boot.pvalue)

```



```
data %>%
  ggplot(aes(populism, colour = Group)) + stat_ecdf()+
  labs(y = "Cumulative Probability", x="Populist attitudes index")+
  facet_wrap(~party)
```

που δίνει τη δυνατότητα της δημιουργίας του παρακάτω διαγράμματος:



Αναφορές

Andreadis, I. (2021). Web Survey Data Quality Assessment. *AAPOR's 76th Annual Conference*.

<https://www.youtube.com/watch?v=q6vK0bI5fJ0>

Andreadis, I., & Andreadis, A. (2022a). Survey Data Quality. *JSM Proceedings, Survey Research Methods Section, 1999–2006*.

Andreadis, I., & Andreadis, A. (2022b). Survey Data Quality—A Workshop. *Southwestern Social Science Association Annual Meeting, 2022 Annual Meeting - April 20-23, San Antonio, Texas*.

<https://tinyurl.com/y4ykg7vn>

Tsigkou, M., Kartsounidou, E., & Andreadis, I. (2020). Using web surveys to collect data from candidates: Experience from the 2019 Greek candidate survey. *ECPR General Conference 2020*. <https://ecpr.eu/Events/Event/PaperDetails/54500>