RCons Polling Station Data Update

Luke Sonnet

3/3/2019

This document will present a preliminary report of the RCons data entry of polling station level data for the 2018 General Elections in Pakistan. First let me load some packages and create a helper function.

```
library(tidyverse)
library(haven)

find_non_coercable_chars <- function(x) {
  suppressWarnings(unique(x[!is.na(x) & is.na(as.numeric(x))]))
}
```

Data submitted

The raw data are in 5 files by province:

- Form 49 (the data on the candidate names and votes, candidate_id to merge with other variables, constituency totals, etc)
- Form 48 (the polling station votes per candidate, total votes, valid votes, and more)
- Form 45 (unofficial polling station level results, only male and female vote totals were recorded)
- Form 28 - Province (list of census block-polling station pairs for the provincial assemblies)
- Form 28 - National (list of census block-polling station pairs for the provincial assemblies)

I'll load each in turn and discuss any irregularities.

For RCons, please look at every line with the tag "RCONS". Please respond to each issue listed via email when you've had a chance to investiage and solve them. A report in email with any updated datasets would be perfect.

Form 49s

```
f49_files <- list.files(
  "source_data/rcons_data",
  pattern = "49",
  recursive = TRUE,
  full.names = TRUE
)
f49_df <- map_dfr(f49_files, read_dta)

# RCONS: However, party names aren't standardized, would be helpful to have them standardized
# Here are some examples:
sort(unique(f49_df$party_affiliation))[40:60]
```

```
##  [1] "Move on Pakistan"
##  [2] "Mustaqbil Pakistan"
##  [3] "Mutahida Majlis-e-Amal Pakistan"
##  [4] "Mutahidda Ulema Mashaikh Council of Pakistan"
##  [5] "Mutahiddia Qabail Party"
##  [6] "Muttahida MAJILID-E-AMAL PAKISTAN"
```

```
##  [7] "MUTTAHIDA MAJILIS-E-AMAL PAKISTAN"
##  [8] "MUTTAHIDA MAJLIS E AMAL PAKISTAN"
##  [9] "MUTTAHIDA MAJLIS-E- AMAL PAKISTAN"
## [10] "Muttahida Majlis-e-Amal Pakistan"
## [11] "MUTTAHIDA MAJLIS-E-AMAL PAKISTAN"
## [12] "Muttahida Qaumi Movement Pakistan"
## [13] "MUTTHIDA MAJLIS-E-AMAL PAKISTAN"
## [14] "National Party"
## [15] "National Peace Council Party"
## [16] "Pak Sarzameen Party"
## [17] "Pakhtoonkhwa Milli Awami Party"
## [18] "Pakistan Aman Party"
## [19] "Pakistan Aman Tehreek"
## [20] "Pakistan Awami Inqelabi League"
## [21] "Pakistan Awami League"
```

```
# What constituencies are there?
all_constituency_ids <- c(
  paste0("NA", 1:272),
  paste0("PS", 1:130),
  paste0("PP", 1:297),
  paste0("PK", 1:99),
  paste0("PB", 1:51)
)

# Quite a few constituencies missing from RCons data
rcons_missing <- setdiff(all_constituency_ids, f49_df$constituency_id)
rcons_missing
```

```
##  [1] "NA60"  "NA61"  "NA90"  "NA91"  "NA92"  "NA103" "NA108" "NA131"
##  [9] "NA271" "PS6"   "PS87"  "PP73"  "PP76"  "PP78"  "PP79"  "PP87"
## [17] "PP103" "PP118" "PP295" "PP296" "PK23"  "PK38"  "PK78"  "PK99"
## [25] "PB6"   "PB26"  "PB35"  "PB36"  "PB41"
```

```
# RCONS: How many of these constituencies had elections postponed, how many are missing
# for other reasons?

glimpse(f49_df)
```

```
## Observations: 28,049
## Variables: 13
## $ assembly_type              <chr> "National Assembly", "National Asse...
## $ constituency_id            <chr> "NA1", "NA1", "NA1", "NA1", "NA1", ...
## $ constituency_name          <chr> "CHITRAL", "CHITRAL", "CHITRAL", "C...
## $ registered_voters_male     <dbl> 151219, 151219, 151219, 151219, 151...
## $ registered_voters_female   <dbl> 118360, 118360, 118360, 118360, 118...
## $ registered_voters_total    <dbl> 269579, 269579, 269579, 269579, 269...
## $ candidate_id               <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, ...
## $ candidate_name             <chr> "IFTIKHAR U DIN", "TAQDIRA AJMAL", ...
## $ party_affiliation          <chr> "Pakistan Muslim League (N)", "Inde...
## $ valid_votes_polled         <dbl> 21127, 681, 3227, NA, 2416, 38819, ...
## $ total_valid_votes_polled   <dbl> 159973, 159973, 159973, 159973, 159...
## $ total_invalid_votes_polled <dbl> 5430, 5430, 5430, 5430, 5430, 5430,...
## $ comments                   <chr> "", "", "", "", "", "", "", "", "",...
```

Form 48s

```
f48_files <- list.files(
  "source_data/rcons_data",
  pattern = "48",
  recursive = TRUE,
  full.names = TRUE
)

f48_temp_df <- map_dfr(
  f48_files,
  ~ {
    read_dta(.x) %>%
      mutate_at(
        vars(contains("_votes")),
        funs(as.character)
      ) %>%
      mutate(file = gsub("source_data/rcons_data/", "", .x))
  }
)

# RCONS: error in some of the votes data, + symbols, duplication of -99s, etc.
f48_temp_df %>%
  filter_at(
    vars(contains("_votes")),
    any_vars(grepl("\\+|/|\\-99\\-99$", .))
  ) %>%
  select(constituency_id,
         starts_with("ps_id"),
         contains("_votes"),
         -matches("can_votes_(19|[234][0-9]+)"),
         file) %>%
  as.data.frame()
```

```
##    constituency_id ps_id can_votes_1 can_votes_2 can_votes_3 can_votes_4
## 1            NA262   149         -99         -99      -99-99         -99
## 2            NA270   221          10           3           1           3
## 3             NA29   182          32           1          20         123
## 4             NA34   168           1           0           0          16
## 5             PK49   121           1           0          56          58
## 6             PK49   130           2           2         122         -99
## 7            NA105    38           0          13           1         217
## 8            NA107   194         348         268           8           3
## 9            NA111   129         220          32           0         464
## 10           NA152    76         319          23         148           1
## 11            NA69   104           1           1         545           0
## 12            NA70   145          19         379           1           1
## 13            NA94   136           1           1          19          12
## 14            NA95    59           2           2          10         110
## 15            NA98   303          10          10          10           0
## 16          PP115   143         218          14          17         323
## 17            PP2    63           3         120          11          36
## 18          PP248   122         466          80          11         166
## 19            PP8   231           4         112         148          57
```

```
## 20          NA198    70          83           0           0           3
## 21          NA199   132           5         181           7         98+
## 22          NA201   222          31         244          6+           2
## 23          NA205   178         167           1           4          12
## 24          NA241    12          37           0          43         134
## 25          NA243    71           0           0           0         187
## 26          NA245   115           0           1          13           0
## 27          NA245   124           0           0         28/           1
## 28          NA252    46          46           0         196           0
##    can_votes_5 can_votes_6 can_votes_7 can_votes_8 can_votes_9
## 1          -99         -99         -99         -99         -99
## 2            0           4          26          3+           0
## 3          13+                    <NA>        <NA>        <NA>
## 4            0         56+           0           0         214
## 5            7          3+         161          11          85
## 6            3          33         -99           6         167
## 7            0           0           6           0         397
## 8            1           0          23        <NA>        <NA>
## 9            2          46           0           0           3
## 10           5           1        <NA>        <NA>        <NA>
## 11           2          31          21           0           0
## 12          14          85           0           0           0
## 13           8         120         475         329        <NA>
## 14         548          10           0          48           0
## 15           8           1           9         714           4
## 16          12           1           1          18        <NA>
## 17           1         164           0        <NA>        <NA>
## 18           3           0           1        <NA>        <NA>
## 19           0           5           0           0          11
## 20           2         631           2           0         111
## 21           0           0        <NA>                    <NA>
## 22           0           6           0           1        <NA>
## 23          24         536           0          3+           2
## 24           2           4           1         198           4
## 25           0          12           6          21           0
## 26           2          25          13         128           1
## 27           1          79          20         445           0
## 28           0           0           8           0         242
##    can_votes_10 can_votes_11 can_votes_12 can_votes_13 can_votes_14
## 1           -99          -99          -99          -99          -99
## 2             3            1            1            1            0
## 3                        <NA>         <NA>         <NA>         <NA>
## 4            12           22            2            0            1
## 5           -99            5         <NA>         <NA>         <NA>
## 6          121+            0         <NA>         <NA>         <NA>
## 7            99           23            7            0         <NA>
## 8          <NA>         <NA>         <NA>         <NA>         <NA>
## 9             0            1         <NA>         <NA>         <NA>
## 10         <NA>         <NA>         <NA>         <NA>         <NA>
## 11            0            2            1           10         <NA>
## 12         <NA>         <NA>         <NA>         <NA>         <NA>
## 13         <NA>         <NA>         <NA>         <NA>         <NA>
## 14            0            3         <NA>         <NA>         <NA>
## 15         <NA>         <NA>         <NA>         <NA>         <NA>
```

```
## 16         <NA>         <NA>         <NA>         <NA>         <NA>
## 17         <NA>         <NA>         <NA>         <NA>         <NA>
## 18         <NA>         <NA>         <NA>         <NA>         <NA>
## 19            0         <NA>         <NA>         <NA>         <NA>
## 20            3          78+            3            1
## 21         <NA>                      <NA>         <NA>
## 22         <NA>                      <NA>         <NA>
## 23         <NA>                      <NA>         <NA>
## 24           95            1            6           32           15
## 25          581           74            0            0          19+
## 26            0           12           37            0            0
## 27            0           68          113            0            1
## 28           20            0            0           12            0
##    can_votes_15 can_votes_16 can_votes_17 can_votes_18 valid_votes
## 1           -99         <NA>         <NA>         <NA>         -99
## 2             0            2         <NA>         <NA>          58
## 3                       <NA>         <NA>         <NA>         189
## 4             0            2           54           20         400
## 5                       <NA>         <NA>         <NA>         455
## 6                       <NA>         <NA>         <NA>         701
## 7          <NA>         <NA>         <NA>         <NA>         763
## 8          <NA>         <NA>         <NA>         <NA>         651
## 9          <NA>         <NA>         <NA>         <NA>        768/
## 10         <NA>         <NA>         <NA>         <NA>         497
## 11         <NA>         <NA>         <NA>         <NA>         614
## 12         <NA>         <NA>         <NA>         <NA>         499
## 13         <NA>         <NA>         <NA>         <NA>         965
## 14         <NA>         <NA>         <NA>         <NA>         733
## 15         <NA>         <NA>         <NA>         <NA>         766
## 16         <NA>         <NA>         <NA>         <NA>         604
## 17         <NA>         <NA>         <NA>         <NA>         335
## 18         <NA>         <NA>         <NA>         <NA>         727
## 19         <NA>         <NA>         <NA>         <NA>         337
## 20                                      <NA>         <NA>         917
## 21                                      <NA>         <NA>         291
## 22                                      <NA>         <NA>         290
## 23                                      <NA>         <NA>         749
## 24           1+                         <NA>         <NA>         573
## 25            0                         <NA>         <NA>         900
## 26          215                         <NA>         <NA>         447
## 27           54                         <NA>         <NA>         810
## 28          -99           0+            <NA>         <NA>         629
##    invalid_votes total_votes                                     file
## 1            -99         -99    Balochistan/Form_48_Result_Form_Data.dta
## 2              7          65    Balochistan/Form_48_Result_Form_Data.dta
## 3              3         192 KP/Election_Result_Form48_Data_11072018.dta
## 4              5         405 KP/Election_Result_Form48_Data_11072018.dta
## 5             26         481 KP/Election_Result_Form48_Data_11072018.dta
## 6             24         725 KP/Election_Result_Form48_Data_11072018.dta
## 7             46        809+  Punjab/Form_48_Result_Form_Data_Punjab.dta
## 8              3        654+  Punjab/Form_48_Result_Form_Data_Punjab.dta
## 9             44         812  Punjab/Form_48_Result_Form_Data_Punjab.dta
## 10             9        506+  Punjab/Form_48_Result_Form_Data_Punjab.dta
## 11             5        619+  Punjab/Form_48_Result_Form_Data_Punjab.dta
```

```
## 12            20        519+  Punjab/Form_48_Result_Form_Data_Punjab.dta
## 13            26        991+  Punjab/Form_48_Result_Form_Data_Punjab.dta
## 14            23        756+  Punjab/Form_48_Result_Form_Data_Punjab.dta
## 15            43        809+  Punjab/Form_48_Result_Form_Data_Punjab.dta
## 16            5+         609  Punjab/Form_48_Result_Form_Data_Punjab.dta
## 17             9        344+  Punjab/Form_48_Result_Form_Data_Punjab.dta
## 18           56+         783  Punjab/Form_48_Result_Form_Data_Punjab.dta
## 19            6+         343  Punjab/Form_48_Result_Form_Data_Punjab.dta
## 20            48         965          Sindh/Form_48_Result_Form_Data.dta
## 21            27         318          Sindh/Form_48_Result_Form_Data.dta
## 22            40         330          Sindh/Form_48_Result_Form_Data.dta
## 23            25         774          Sindh/Form_48_Result_Form_Data.dta
## 24            16         589          Sindh/Form_48_Result_Form_Data.dta
## 25            18         918          Sindh/Form_48_Result_Form_Data.dta
## 26            10        457+          Sindh/Form_48_Result_Form_Data.dta
## 27            22         832          Sindh/Form_48_Result_Form_Data.dta
## 28            18         647          Sindh/Form_48_Result_Form_Data.dta
```

```
# All constituency_id, ps_id combos are unique!
any(duplicated(select(f48_temp_df, constituency_id, ps_id)))
```

```
## [1] FALSE
```

Form 45s

```
f45_files <- list.files(
  "source_data/rcons_data",
  pattern = "45",
  recursive = TRUE,
  full.names = TRUE
)

f45_temp_df <- map_dfr(
  f45_files,
  ~ {
    read_dta(.x) %>%
      mutate_at(
        vars(contains("_turnout"), contains("total_votes")),
        funs(as.character)
      ) %>%
      mutate(file = gsub("source_data/rcons_data/", "", .x))
  }
)

# RCONS: error in some of the turnout and votes data, + symbols
f45_temp_df %>%
  filter_at(
    vars(contains("_turnout"), contains("total_votes")),
    any_vars(grepl("\\+", .))
  ) %>%
  select(constituency_id, starts_with("ps_id"), contains("turnout"), total_votes) %>%
  as.data.frame()
```

```
##    constituency_id ps_id total_male_turnout total_female_turnout
```

```
## 1            NA258   113               338                    0
## 2            NA258   353              268+9                   9
## 3            NA162   261               376                  260
## 4            NA209   185               265                 245+
## 5            NA231   198               357                  200
## 6            NA245   115               302                  155
## 7            NA258   113               338                    0
## 8            NA258   353              268+9                   9
## 9             NA94   136               550                  441
## 10            NA98   303               473                  336
## 11           NA209   185               265                 245+
## 12           NA231   198               357                  200
## 13           NA236   116               517                  411
## 14           NA245   115               302                  155
## 15           NA258   113               338                    0
## 16           NA258   353              268+9                   9
##    total_turnout total_votes
## 1           338+         338
## 2            277         277
## 3           636+         636
## 4            510         510
## 5           557+         557
## 6            457        457+
## 7           338+         338
## 8            277         277
## 9            991        991+
## 10           809        809+
## 11           510         510
## 12          557+         557
## 13          928+         928
## 14           457        457+
## 15          338+         338
## 16           277         277
```

```
# RCONS: also some duplicates of rows, is this expected?
any(duplicated(select(f45_temp_df, constituency_id, ps_id)))
```

```
## [1] TRUE
```

```
f45_temp_df %>%
  group_by(constituency_id, ps_id) %>%
  filter(n() > 1) %>%
  select(constituency_id, ps_id, file) %>%
  arrange(constituency_id, ps_id)
```

```
## # A tibble: 42,852 x 3
## # Groups:   constituency_id, ps_id [18,336]
##    constituency_id ps_id file
##    <chr>           <dbl> <chr>
##  1 NA1                 1 KP/Election_Result_Form45_Data_11072018.dta
##  2 NA1                 1 Punjab/Form_45_Combined.dta
##  3 NA1                 1 Sindh/Form_45_Male_Female_Turnout.dta
##  4 NA1                 2 KP/Election_Result_Form45_Data_11072018.dta
##  5 NA1                 2 Punjab/Form_45_Combined.dta
##  6 NA1                 2 Sindh/Form_45_Male_Female_Turnout.dta
```

```
##  7 NA1                 3 KP/Election_Result_Form45_Data_11072018.dta
##  8 NA1                 3 Punjab/Form_45_Combined.dta
##  9 NA1                 3 Sindh/Form_45_Male_Female_Turnout.dta
## 10 NA1                 4 KP/Election_Result_Form45_Data_11072018.dta
## # ... with 42,842 more rows
```

Form 28s

```
f28_files <- list.files(
  "source_data/rcons_data",
  pattern = "Polling_Station_List|28",
  recursive = TRUE,
  full.names = TRUE
)

# Just read in to list
f28s <- map(f28_files, read_dta)

# Errors in `block_code_rural` and `block_code_urban`
# 1) RCONS: The the rural and urban block codes in Sindh have the dashes in them. Is this an error or i
map(f28s, ~{find_non_coercable_chars(.x$block_code)[1:10]})[7:8]
```

```
## [[1]]
##  [1] "332030301-06-07"       "332030205-12-13-14-15"
##  [3] "332030302-08-09"       "332030404-05-09"
##  [5] "332030305-13"          "332030402-07"
##  [7] "332030403-08"          "332030401-06"
##  [9] "332030202-06-08-09"    "332030204-10-11"
##
## [[2]]
##  [1] "332030301-06-07"       "332030205-12-13-14-15"
##  [3] "332030302-08-09"       "332030404-05-09"
##  [5] "332030305-13"          "332030402-07"
##  [7] "332030403-08"          "332030401-06"
##  [9] "332030202-06-08-09"    "332030204-10-11"
```

```
# 2) RCONS: Typo in  block_code_rural in the Punjab PA data
f28s[[3]] %>%
  filter(block_code_rural %in% c("26030907 (Part\n2)", "m   81021103")) %>%
  select(constituency_id, starts_with("ps_id"), starts_with("block_code"))
```

```
## # A tibble: 3 x 5
##   constituency_id ps_id block_code_rural     block_code_urban block_code
##   <chr>           <dbl> <chr>                           <dbl>      <dbl>
## 1 NA17              320 "26030907 (Part\n2)"               NA   26030907
## 2 NA41               58 m   81021103                       NA   81021103
## 3 NA41               59 m   81021103                       NA   81021103
```

```
# RCONS: `male_booths`, `female_booths`, and `total_booths` have some errors
# in the KP provincial data
f28s[[4]] %>%
  filter(male_booths %in% c("4\n4", "Male")) %>%
  select(constituency_id, starts_with("ps_id"), ends_with("_booths"))
```

```
## # A tibble: 3 x 6
##   constituency_id ps_id_NA ps_id male_booths female_booths total_booths
##   <chr>              <dbl> <dbl> <chr>       <chr>         <chr>
## 1 PK31                 243    57 "4\n4"      "0\n0"        "4\n4"
## 2 PK31                 244    58 "4\n4"      "0\n0"        "4\n4"
## 3 PK50                   1     1 Male        Female        Total
```

```
# RCONS: Plus signs in male_voters codes for Punjab PA and NA;
# there are plus signs in lots of the data later on
f28s[[5]] %>%
  filter(grepl("\\+", male_voters)) %>%
  select(constituency_id, starts_with("ps_id"), male_voters)
```

```
## # A tibble: 1 x 3
##   constituency_id ps_id male_voters
##   <chr>           <dbl> <chr>
## 1 NA104             284 267+
```

```
# If we hide the above problems we can get the data anyways
f28_df <- map_dfr(
  f28_files,
  ~ {
    read_dta(.x) %>%
      mutate_at(
        vars(starts_with("block_code")), funs(gsub("null", NA, as.character(.)))
      ) %>%
      mutate(male_voters = gsub("\\+", "", male_voters)) %>%
      mutate_at(vars(ends_with("_booths")), funs(as.numeric))
  }
)

glimpse(f28_df)
```

```
## Observations: 461,334
## Variables: 20
## $ constituency_id      <chr> "NA257", "NA257", "NA257", "NA257", "NA25...
## $ constituency_area    <chr> "KILLA SAIFULLAH-CUM-ZHOB-CUM-SHERANI", "...
## $ ps_id                <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2,...
## $ ps_name              <chr> "GOVT GIRLS HIGH SCHOOL KILLA SAFIULLAH",...
## $ name_ea_rural        <chr> "", "", "", "", "", "", "", "", "", "", "...
## $ block_code_rural     <chr> "0", "0", "0", "0", "0", "0", "0", "0", "...
## $ name_ea_urban        <chr> "CIRCUIT HOUSE-CIVIL HOSPITAL QUARTERS-KI...
## $ block_code_urban     <chr> "461080103", "461080201", "461080202", "4...
## $ block_code           <chr> "461080103", "461080201", "461080202", "4...
## $ no_voters_on_ea      <chr> "0", "0", "0", "0", "0", "0", "0", "0", "...
## $ male_voters          <chr> "0", "0", "0", "0", "0", "0", "0", "0", "...
## $ female_voters        <dbl> 15, 192, 179, 174, 89, 52, 180, 79, 59, 3...
## $ total_voters         <dbl> 15, 192, 179, 174, 89, 52, 180, 79, 59, 3...
## $ male_booths          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ female_booths        <dbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,...
## $ total_booths         <dbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,...
## $ assembly_type        <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ constituency_no_NA   <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ constituency_area_NA <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ ps_id_NA             <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
```