COVID-19 Survey - Data Cleaning

December 3, 2020

Import data

```
ds <- readxl::read_xlsx(
  "data/raw/Covid-19+Survey+-+baseline+-+subset_April+20%2C+2020_14.57.xlsx",
  skip = 2, col_names = FALSE)

# take trust pres from .csv format
pres <- read.csv("data/raw/Covid-19+Survey+-+baseline+-+subset_April+23,+2020_14.09_num.csv")[-c(1:2),]
```

Notes

- Not all respondents finished the survey, but we include incomplete surveys

```
table(ds$loc_country, ds$Finished)
```

```
##
##        0    1
##   1   43  604
##   2  111 1480
##   3   59  506
##   4    0   19
```

Consent

```
dat <- ds %>% mutate(
  consent = coalesce(consent_general,
                     consent_NG,
                     consent_KY,
                     consent_UG))

#Exclude non-consent
#REVIEW: leave incomplete surveys (Finished == 0)
dat <- dat %>% filter(consent == 1)
```

Location

```
dat %<>% mutate(
  #nigeria - zone, uganda - region
  loc_admin1 = coalesce(loc_zone_NG, loc_reg_UG),
  #nigeria - state, kenya - county, uganda - district
  loc_admin2 = coalesce(loc_state_nc_ng, loc_state_ne_ng, loc_state_nw_ng,
                        loc_state_se_ng, loc_state_ss_ng, loc_state_sw_ng,
```

```
                        loc_county_KY,
                        loc_dist_cent_ug, loc_dist_west_ug,
                        loc_dist_east_ug, loc_dist_north_ug),
  # uganda - division, nigeria - Lagos LGA
  loc_admin3 = coalesce(loc_kla_div, loc_LGA_lagos),
  # uganda - parish/subcounty
  loc_admin4 = coalesce(loc_kla_cent, loc_kla_kawempe,
                        loc_kla_rubaga, loc_kla_makindye,
                        loc_kla_nakawa, loc_wakiso_sc,
                        loc_mukono_sc,
                        loc_subc_ky))

# Location fixed effect
dat$loc_fe = sapply(1:nrow(dat), function(i)
  paste0(setdiff(as.character(dat[i,15:40]), c("NA", "", NA)), collapse = "_"))
```

Trust

```
dat %<>% mutate(trust_cov_moh = ifelse(!is.na(trust_cov_moh_ng),
                                       trust_cov_moh_ng, trust_cov_moh))

dat %<>%
  mutate(trust_gov_ng = (trust_cov_moh + trust_cov_hwork)/2,
         trust_gov = (trust_cov_moh + trust_cov_hwork + trust_cov_pres)/3) %>%
  group_by(country) %>%
  mutate(trust_gov = trust_gov - mean(trust_gov, na.rm = TRUE),
         trust_gov_ng = trust_gov_ng - mean(trust_gov_ng, na.rm = TRUE)) %>%
  ungroup()

dat$trust_gov <- ifelse(dat$country == "Nigeria", dat$trust_gov_ng, dat$trust_gov)
```

Multiple Choice Questions

```
vars_mchoice <- c("fact_trueforyou", "fact_truecovid",
                  "fact_reduce_spread", "fact_response",
                  "info2", "remit")

for(i in vars_mchoice){
  dat <- cbind(dat, expand_mchoice(i))
}
```

List Experiment

```
dat %<>% mutate(
  le_control = coalesce(le_kenya_control, le_nigeria_control,
                        le_uganda_control),
  le_treat = coalesce(le_kenya_treat, le_nigeria_treat,
                      le_uganda_treat),
  le_count = case_when(!is.na(le_control) ~ le_control,
                       !is.na(le_treat) ~ le_treat),
```

```
  le_condition = case_when(!is.na(le_control) ~ 0,
                           !is.na(le_treat) ~ 1))
```

Vignette Experiment

```
dat %<>% mutate(
  hypo_condition = case_when(!is.na(hypo_norm0) ~ "C",
                             !is.na(hypo_norm1) ~ "T1",
                             !is.na(hypo_norm2) ~ "T2"),
  hypo_answer = case_when(!is.na(hypo_norm0) ~ hypo_norm0,
                          !is.na(hypo_norm1) ~ hypo_norm1,
                          !is.na(hypo_norm2) ~ hypo_norm2))
```

Behavioral Questions

```
#REVIEW: Validate letter content - some 15 people wrote "None"
# grep("^No", dat$behav_letter, value = TRUE, ignore.case = TRUE)
NAletter <- grepl("^No", dat$behav_letter,
                  ignore.case = TRUE) & nchar(dat$behav_letter) < 15

dat <- dat %>% mutate(
  #recode behav_write = 1-yes, 0-no
  behav_write = case_when(behav_write == 4 ~ 1,
                          behav_write == 5 ~ 0),
  #new var if behav_letter content does not include "None"-type messages
  behav_wrote = case_when(!behav_letter %in% NAletter & !is.na(behav_letter) ~ 1,
                          NAletter | behav_write == 0 ~ 0))

dat <- dat %>% mutate(att_lock_yes = 1*(att_lock > 0)) %>%
  group_by(country) %>%
  mutate(att_lock_pct = mean(att_lock_yes, na.rm = TRUE)) %>%
  ungroup() %>%
  mutate(att_lock_correct = round(att_lock_pct, 1),
         att_lock_gap = (att_lock_guess/10)-att_lock_correct,
         att_lock_gap_abs = abs(att_lock_guess/10-att_lock_correct))
```

Demographic

```
dat <- dat %>% mutate(
  female = dem_sex,

  #create combined country_ethnicity var
  dem_eth = coalesce(dem_eth_ke, dem_eth_ng, dem_eth_ug),
  dem_eth = ifelse(!is.na(dem_eth),
                   paste(loc_country, dem_eth, sep = "_"),
                   dem_eth),

  # code urban as dummy
  loc_urbrur = case_when(loc_urbrur == 2 ~ 0,
                         loc_urbrur == 1 ~ 1),
```

```
  #code binary var for five main occupations
  occ_student = ifelse(dem_occ == 5, 1, 0),
  occ_midlev = ifelse(dem_occ == 14, 1, 0),
  occ_uprlev = ifelse(dem_occ == 15, 1, 0),
  occ_never = ifelse(dem_occ == 1, 1, 0),
  occ_manual = ifelse(dem_occ == 11, 1, 0),

  #code binary vars for main religions
  rel_catholic = ifelse(dem_rel == 1, 1, 0),
  rel_protestant = ifelse(dem_rel == 2, 1, 0),
  rel_evangelical = ifelse(dem_rel == 3, 1, 0),
  rel_muslim = ifelse(dem_rel == 4, 1, 0),
  rel_other = ifelse(dem_rel %in% c(0,5:7), 1, 0),

  #create var whether voted previous election
  voted = case_when(
    dem_vote_ke == 0 | dem_vote_ng == 0 | dem_vote_ug == 0 ~ 0,
    dem_vote_ke  > 0 | dem_vote_ng  > 0 | dem_vote_ug > 0 ~ 1),

  #create var whether voted for incumbent
  #1-voted for incumbent, 0-did not vote or did not vote for incumbent
  voted_incumbent = case_when(
    dem_vote_ke == 1 | dem_vote_ng == 1 | dem_vote_ug == 1 ~ 1,
    dem_vote_ke != 1 | dem_vote_ng != 1 | dem_vote_ug != 1 ~ 0),

  #create var for co-partisanship with incumbent president
  copartisan = case_when(
    dem_party_ke == 6 | dem_party_ng == 1 | dem_party_ug == 1 ~ 1,
    dem_party_ke != 6 | dem_party_ng != 1 | dem_party_ug != 1 ~ 0),

  #create var for vote choice or copartisan
  voted_copartisan = ifelse(voted_incumbent == 1 |
                              (!is.na(copartisan) & copartisan == 1), 1, 0)
)
```

Additional cleaning

```
# Check that _TEXT vars are do not include systematically excluded categories

txt_var <- dat %>% dplyr::select(ends_with("_TEXT"))
# apply(txt_var, 2, table)

#----- loc_country_4_TEXT

# Manually input loc values for Nigeria residents
# table(dat$loc_country_4_TEXT)
dat$loc_country[dat$loc_country_4_TEXT %in%
                  c("Nigeria","nigeria", "Abuja")] <- 2

dat$country <- case_when(dat$loc_country == 1 ~ "Kenya",
                         dat$loc_country == 2 ~ "Nigeria",
                         dat$loc_country == 3 ~ "Uganda")
```

```r
dat$loc_admin1[dat$loc_fe == "4_Nigeria_Calabar"] <- "5"
dat$loc_admin1[dat$loc_fe == "4_Nigeria_Alagbado"] <- "6"
dat$loc_admin1[dat$loc_fe == "4_Nigeria_Ikeja"] <- "6"
dat$loc_admin1[dat$loc_fe == "4_Nigeria_Ibadan"] <- "6"
dat$loc_admin1[dat$loc_fe == "4_Nigeria_Warri Delta State"] <- "5"
# dat$loc_admin1[dat$loc_fe == "4_Nigeria_Benin Republic"]
dat$loc_admin1[dat$loc_fe == "4_Nigeria_Benin"] <- "5"
dat$loc_admin1[dat$loc_fe == "4_Nigeria_Kaduna"] <- "3"
dat$loc_admin1[dat$loc_fe == "4_Nigeria_Ogbomosho"] <- "6"
dat$loc_admin1[dat$loc_fe == "4_Abuja_Keffi"] <- "1"
# dat$loc_admin1[dat$loc_fe == "4_nigeria_aw"]

dat$loc_fe[dat$loc_fe == "4_Nigeria_Calabar"] <- "2_5_24"
dat$loc_fe[dat$loc_fe == "4_Nigeria_Alagbado"] <- "2_6_28_5"
dat$loc_fe[dat$loc_fe == "4_Nigeria_Ikeja"] <- "2_6_28_13"
dat$loc_fe[dat$loc_fe == "4_Nigeria_Ibadan"] <- "2_6_32"
dat$loc_fe[dat$loc_fe == "4_Nigeria_Warri Delta State"] <- "2_5_26"
# dat$loc_fe[dat$loc_fe == "4_Nigeria_Benin Republic"]
dat$loc_fe[dat$loc_fe == "4_Nigeria_Benin"] <- "2_5_27"
dat$loc_fe[dat$loc_fe == "4_Nigeria_Kaduna"] <- "2_3_15"
dat$loc_fe[dat$loc_fe == "4_Nigeria_Ogbomosho"] <- "2_6_32"
dat$loc_fe[dat$loc_fe == "4_Abuja_Keffi"] <- "2_1_4"
# dat$loc_fe[dat$loc_fe == "4_nigeria_aw"]

#----- fact_response_7_TEXT

#REVIEW: Manually input these?

# 45 `Other, specify` responses for `fact_response_7_TEXT`
# dat$fact_response_7_TEXT[!is.na(dat$fact_response_7_TEXT)]

# 34 mention "isolate" or "quarantine" or "avoid contact"
# grep("(isol)|(quarant)|(contact)", a$fact_response_7_TEXT,
# value = TRUE, ignore.case = TRUE)

# 11 remaining
# setdiff(grep("(isol)|(quarant)|(contact)", a$fact_response_7_TEXT,
# value = TRUE, invert = TRUE, ignore.case = TRUE), NA)

#----- info1_9_TEXT
# 23 unique responses
# unique(dat$info1_9_TEXT)

#----- att_self_concern_8_TEXT

# 38 unique responses
# unique(dat$att_self_concern_8_TEXT)

# 10 mention "health"
# grep("(health)", a$att_self_concern_8_TEXT, value = TRUE, ignore.case = TRUE)

# 12 mention "food" or "income", "welfare", "low"
# grep("(food)|(feed)|(income)|(welfare)|(low)", a$att_self_concern_8_TEXT,
```

```
# value = TRUE, ignore.case = TRUE)

# 7 mention "economy"
# grep("(econom)", a$att_self_concern_8_TEXT, value = TRUE, ignore.case = TRUE)

# 13 remaining
# setdiff(grep("(food)|(feed)|(income)|(welfare)|(low)|(health)|(econom)", a$att_self_concern_8_TEXT, v

#----- att_concern_pol_own_7_TEXT
# 15 non-missing values
# unique(dat$att_concern_pol_own_7_TEXT)

#----- att_fam_concern_8_TEXT
# 12 non-missing values
# unique(dat$att_fam_concern_8_TEXT)

#----- att_concern_pol_they_7_TEXT
# 13 non-missing values
# unique(dat$att_concern_pol_they_7_TEXT)

#----- att_socmed_concern_8_TEXT
# 13 non-missing values
# unique(dat$att_socmed_concern_8_TEXT)

#----- dem_occ_16_TEXT
# REVIEW: 171 non-missing values
# unique(dat$dem_occ_16_TEXT)

#----- dem_rel_7_TEXT
# 30 non-missing values
# unique(dat$dem_rel_7_TEXT)

#----- dem_eth_ug_21_TEXT
# Manually recode ethnicities coded as "Other"
# table(dat$dem_eth_ug_21_TEXT)
dat$dem_eth[dat$dem_eth_ug_21_TEXT %in% c("Muteso")] <- "3_8"
dat$dem_eth[dat$dem_eth_ug_21_TEXT %in% c("Madi")] <- "3_13"
dat$dem_eth[dat$dem_eth_ug_21_TEXT %in% c("Samya","Samia")] <- "3_14"

#----- dem_eth_ng_28_TEXT
# table(dat$dem_eth_ng_28_TEXT)

#----- dem_eth_ke_16_TEXT
# table(dat$dem_eth_ke_16_TEXT)
```

Rescale variables and create weights

```
dat <- dat %>%
  group_by(country) %>%
  mutate(hypo_answer_scaled = scale(hypo_answer)) %>%
  # mutate(hypo_answer_bsize = n(),
  #        wt = 1/3*(1/hypo_answer_bsize)) %>%
```

```
  ungroup()
  # mutate(wt = wt/sum(wt))
```

Code text variable

```
# Dictionary created using 10% sample of responses
collective_terms <- c("\\<us\\>", "\\<we\\>","\\<our\\>", "\\<ourselves\\>")
externality_terms <- c("(collab)","(together)","(others)", "(us all)",
                       "(let us)","(let's)","(unit)", "(union)",
                       "(one another)")
religious_terms <- c("\\<god\\>", "(pray)", "(bless)", "(Allah)", "(Lord)", "(relig)")
civic_terms <- c("(citizen)", "(our country)", "(our nation)")

dat$letter_col <- code_terms(collective_terms, dat$behav_letter)
dat$letter_ext <- code_terms(externality_terms, dat$behav_letter, perl = TRUE)
dat$letter_rel <- code_terms(religious_terms, dat$behav_letter, perl = TRUE)
dat$letter_civ <- code_terms(civic_terms, dat$behav_letter, perl = TRUE)
```

Code lockdown

```
# Akwa Ibom, Kwara, Anambra, Niger, Lagos, Ogun, FCT, Ekiti, Delta,
# Osun (suspended for a day on April 15). Niger?
# Uganda nationwide

locked <- c("2_5_1", grep("2_6_28", dat$loc_fe, fixed = TRUE, value = TRUE),
            "2_6_29", "2_1_7", "2_1_3", "2_4_22", "2_1_5", "2_6_1", "2_5_26",
            "2_6_31")

dat %<>% mutate(lockdown = ifelse((loc_fe %in% locked)| country == "Uganda", 1, 0))
```

Remove incomplete responses

```
# Remove respondents who attrited before completing the survey
# Last answer that was required was `dem_religiosity`, which 0.9131810193 of the
# starting sample completed

dat %<>% filter(!is.na(country) & Finished==1)
```

Export Data

```
to_export <- dat %>%
  dplyr::select(StartDate:UserLanguage, consent, loc_country, country,
         loc_admin1, loc_admin2, loc_admin3, loc_admin4, loc_fe,
         loc_urbrur, fact_trueforyou, fact_trueforyou_1:fact_trueforyou_12,
         fact_truecovid, fact_truecovid_1:fact_truecovid_6,
         fact_tested, fact_tf_older:fact_reduce_spread,
         fact_reduce_spread_1:fact_reduce_spread_7,
         fact_response, fact_response_1:fact_response_7,
```

```
        fact_numb_cases:info1, info2, info2_1:info2_4,
        trust_cov_pres, trust_cov_moh, trust_cov_hwork, trust_cov_media,
        trust_gov,
        le_condition, le_count, le_control, le_treat,
        at_lock_time:att_self_concern, att_concern_pol_own,
        att_risk_ownhealth:att_fam_concern, att_concern_pol_they,
        att_risk_ffhealth:att_socmed_concern,
        att_lock_correct:att_lock_gap_abs,
        hypo_condition, hypo_answer, hypo_answer_scaled,
        hypo_norm0:hypo_norm2,
        letter_col, letter_ext, letter_rel, letter_civ,
        behav_write, behav_wrote, behav_letter,
        remit, remit_1:remit_4, remit_rcv_amount_abr:remit_rcv_amt_dom,
        female, dem_age:dem_occ, dem_vote_ug:dem_party_ug,
        voted, voted_incumbent, copartisan, voted_copartisan,
        dem_rel, dem_religiosity,
        dem_eth, dem_eth_ug, dem_eth_ng, dem_eth_ke,
        click_WHOlink, click_WHOphone, lockdown, occ_student:occ_manual,
        rel_catholic:rel_other)

write_csv(to_export, file = "data/clean/covid_survey.csv")
saveRDS(to_export, file = "data/clean/covid_survey.RDS")
```