

tidy raw data

Gina

October 14, 2015

```
library(foreign)
library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(lubridate)

##
## Attaching package: 'lubridate'
##
## The following object is masked from 'package:base':
##
##   date

setwd("raw_data/version4.1_dta")

files=list.files(pattern = "dta")

file.info(files)

##                                     size isdir mode
## alliance_v4.1_by_directed_yearly.dta 16905309 FALSE 755
## alliance_v4.1_by_directed.dta          712538 FALSE 755
## alliance_v4.1_by_dyad_yearly.dta      8306345 FALSE 755
## alliance_v4.1_by_dyad.dta              409869 FALSE 755
## alliance_v4.1_by_member_yearly.dta    2029817 FALSE 755
## alliance_v4.1_by_member.dta           115804 FALSE 755
##                                     mtime
## alliance_v4.1_by_directed_yearly.dta 2013-04-10 19:09:34
## alliance_v4.1_by_directed.dta         2013-04-10 19:02:06
## alliance_v4.1_by_dyad_yearly.dta      2013-04-10 16:36:06
## alliance_v4.1_by_dyad.dta             2013-04-10 15:55:08
## alliance_v4.1_by_member_yearly.dta    2013-04-10 15:52:16
## alliance_v4.1_by_member.dta           2013-04-10 15:42:06
##                                     ctime
## alliance_v4.1_by_directed_yearly.dta 2016-05-31 01:10:49
## alliance_v4.1_by_directed.dta         2016-05-31 01:10:49
## alliance_v4.1_by_dyad_yearly.dta      2016-05-31 01:10:49
## alliance_v4.1_by_dyad.dta             2016-05-31 01:10:49
## alliance_v4.1_by_member_yearly.dta    2016-05-31 01:10:49
## alliance_v4.1_by_member.dta           2016-05-31 01:10:49
```

```
##                                     atime uid gid
## alliance_v4.1_by_directed_yearly.dta 2017-05-09 22:00:07 501 20
## alliance_v4.1_by_directed.dta        2016-06-02 23:13:23 501 20
## alliance_v4.1_by_dyad_yearly.dta     2016-06-02 23:13:24 501 20
## alliance_v4.1_by_dyad.dta            2017-05-09 14:36:59 501 20
## alliance_v4.1_by_member_yearly.dta   2017-05-09 23:32:58 501 20
## alliance_v4.1_by_member.dta          2018-04-29 16:33:08 501 20
##                                     uname grname
## alliance_v4.1_by_directed_yearly.dta evangelinereynolds staff
## alliance_v4.1_by_directed.dta        evangelinereynolds staff
## alliance_v4.1_by_dyad_yearly.dta     evangelinereynolds staff
## alliance_v4.1_by_dyad.dta            evangelinereynolds staff
## alliance_v4.1_by_member_yearly.dta   evangelinereynolds staff
## alliance_v4.1_by_member.dta          evangelinereynolds staff

alliance_v4.1_by_member_0 = read.dta("raw_data/version4.1_dta/alliance_v4.1_by_member.dta")
names(alliance_v4.1_by_member_0)
```

```
## [1] "version4id"      "ccode"           "state_name"      "all_st_day"
## [5] "all_st_month"    "all_st_year"     "all_end_day"     "all_end_month"
## [9] "all_end_year"    "ss_type"         "mem_st_day"      "mem_st_month"
## [13] "mem_st_year"     "mem_end_day"     "mem_end_month"   "mem_end_year"
## [17] "left_censor"     "right_censor"    "defense"         "neutrality"
## [21] "nonaggression"  "entente"         "version"

alliance_v4.1_by_member = alliance_v4.1_by_member_0 %>%
  rename(alliance_id = version4id) %>%
  mutate(alliance_start_date =
    as_date(paste(all_st_year, all_st_month, all_st_day, sep = "-"))) %>%
  mutate(alliance_end_date =
    as_date(paste(all_end_year, all_end_month, all_end_day, sep = "-"))) %>%
  mutate(member_start_date =
    as_date(paste(mem_st_year, mem_st_month, mem_st_day, sep = "-"))) %>%
  mutate(member_end_date =
    as_date(paste(mem_end_year, mem_end_month, mem_end_day, sep = "-"))) %>%
  dplyr::select(-all_st_year, -all_st_month, -all_st_day,
    -mem_st_year, -mem_st_month, -mem_st_day,
    -all_end_year, -all_end_month, -all_end_day,
    -mem_end_year, -mem_end_month, -mem_end_day)
```

```
## Warning: 373 failed to parse.
```

```
## Warning: 359 failed to parse.
```

```
names(alliance_v4.1_by_member)
```

```
## [1] "alliance_id"      "ccode"           "state_name"
## [4] "ss_type"          "left_censor"     "right_censor"
## [7] "defense"          "neutrality"      "nonaggression"
## [10] "entente"          "version"         "alliance_start_date"
## [13] "alliance_end_date" "member_start_date" "member_end_date"
```

```
# country information is from COW unique countries list
```

```
alliances <- alliance_v4.1_by_member %>%
  group_by(alliance_id, ss_type, alliance_start_date, alliance_end_date) %>%
```

```

    summarise() %>% ungroup()
dim(alliances)

## [1] 414    4

alliance_country_membership <- alliance_v4.1_by_member %>%
  group_by(alliance_id, ccode, member_start_date, member_end_date,
            defense, neutrality, nonaggression, entente) %>%
  summarise() %>% ungroup() %>%
  rename(cown = ccode)
dim(alliance_country_membership)

## [1] 1221    8

dim(alliance_v4.1_by_member)

## [1] 1221    15

save(list = c("alliances", "alliance_country_membership"),
     file = "data_products/tidy_alliance_v4.Rdata")

file.info("data_products/tidy_alliance_v4.Rdata")$size

## [1] 10161

file.info("raw_data/version4.1_dta/alliance_v4.1_by_member.dta")$size

## [1] 115804

```