ragis_data_wrangle.R

Tom Brailey

2019-12-03

```
# Regional Autonomy GIS data wrangling

# Setup
rm(list=ls())

library(magrittr)
library(ggplot2)

# Load data
epr <- rio::import(paste0(here::here(), "/data/EPR-2018.1.1.csv"))
geo_epr <- rio::import(paste0(here::here(), "/data/GeoEPR-2018.1.1.csv"))
geom <- sf::st_read(paste0(here::here(), "/data/GeoEPR-2018.1.1/GeoEPR.shp")) %>%
  dplyr::select(group, geometry)
```

```
## Reading layer `GeoEPR' from data source `C:\Users\Tom Brailey\Dropbox\github_private\RegionalAutonomy
## replacing null geometries with empty geometries
## Simple feature collection with 1470 features and 10 fields (with 134 geometries empty)
## geometry type:  GEOMETRY
## dimension:      XY
## bbox:           xmin: -180 ymin: -55.31195 xmax: 180 ymax: 76.99887
## epsg (SRID):    4326
## proj4string:    +proj=longlat +ellps=WGS84 +no_defs
```

```
mali_conflict <- rio::import(paste0(here::here(), "/data/1997-01-01-2003-12-31-Mali.csv")) %>%
  dplyr::mutate(time = ifelse(year < 2000, 0, 1))
write.csv(mali_conflict, paste0(here::here(), "/data/mali_conflict.csv"))

health <- rio::import(paste0(here::here(), "/data/health.xlsx")) %>%
  dplyr::mutate(`DATE OUVERTURE` = stringr::str_extract(`DATE OUVERTURE`, "^.{4}"))
write.csv(health, paste0(here::here(), "/data/health.csv"))

arc_conf <- dplyr::as_data_frame(rio::import(paste0(here::here(), "/data/final_conflict.xlsx")))
```

```
## Warning: `as_data_frame()` is deprecated, use `as_tibble()` (but mind the new semantics).
## This warning is displayed once per session.
```

```
xtable::xtable(arc_conf)
```

```
## % latex table generated in R 3.5.3 by xtable 1.8-4 package
## % Tue Dec 03 15:45:49 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{rlrrr}
##   \hline
##  & Autonomy & Total Number of Conflicts & Conflict in Autonomous Region & Conflict in AR as Percentag
##   \hline
## 1 & Pre-Autonomy & 89.00 & 57.00 & 64.04 \\
##   2 & Post-Autonomy & 36.00 & 3.00 & 8.33 \\
##    \hline
```

```
## \end{tabular}
## \end{table}
```

```
# Subset data (Explore Variables)
group_t <- epr %>%
  dplyr::filter(reg_aut == "TRUE") %>%
  dplyr::distinct(group) %>%
  dplyr::pull()

group_f <- epr %>%
  dplyr::filter(reg_aut == "FALSE") %>%
  dplyr::distinct(group) %>%
  dplyr::pull()

group_t <- dplyr::as_tibble(group_t)
```

```
## Warning: Calling `as_tibble()` on a vector is discouraged, because the behavior is likely to change i
## This warning is displayed once per session.
```

```
group_f <- dplyr::as_tibble(group_f)

group_aut <- group_t %>%
  dplyr::filter(value %in% unique(group_f$value))

epr_sub <- epr %>%
  dplyr::filter(group %in% unique(group_aut$value))

geo_epr_sub <- geo_epr %>%
  dplyr::filter(group %in% unique(group_aut$value)) %>%
  dplyr::select(-from, -to)

epr_join <- dplyr::left_join(epr_sub, geo_epr_sub)
```

```
## Joining, by = c("gwid", "statename", "group", "groupid", "gwgroupid", "umbrella")
```

```
unique(epr_join$statename)
```

```
##   [1] "Canada"
##   [2] "Mexico"
##   [3] "Belize"
##   [4] "Guatemala"
##   [5] "Honduras"
##   [6] "El Salvador"
##   [7] "Nicaragua"
##   [8] "Costa Rica"
##   [9] "Panama"
##  [10] "Colombia"
##  [11] "Venezuela"
##  [12] "Guyana (British Guiana)"
##  [13] "Suriname"
##  [14] "Ecuador"
##  [15] "Peru"
##  [16] "Brazil"
##  [17] "Bolivia"
##  [18] "Paraguay"
##  [19] "Chile"
```

```
##  [20] "Argentina"
##  [21] "Uruguay"
##  [22] "United Kingdom"
##  [23] "Belgium"
##  [24] "France"
##  [25] "Spain"
##  [26] "Germany"
##  [27] "Germany Democratic Republic"
##  [28] "Poland"
##  [29] "Austria"
##  [30] "Hungary"
##  [31] "Czechoslovakia"
##  [32] "Czech Republic"
##  [33] "Slovakia"
##  [34] "Italy"
##  [35] "Albania"
##  [36] "Serbia"
##  [37] "Montenegro"
##  [38] "Macedonia"
##  [39] "Croatia"
##  [40] "Serbia and Montenegro"
##  [41] "Bosnia and Herzegovina"
##  [42] "Kosovo"
##  [43] "Slovenia"
##  [44] "Greece"
##  [45] "Cyprus"
##  [46] "Bulgaria"
##  [47] "Moldova"
##  [48] "Romania"
##  [49] "Russia"
##  [50] "Estonia"
##  [51] "Latvia"
##  [52] "Lithuania"
##  [53] "Ukraine"
##  [54] "Belarus"
##  [55] "Armenia"
##  [56] "Georgia"
##  [57] "Azerbaijan"
##  [58] "Finland"
##  [59] "Mali"
##  [60] "Niger"
##  [61] "Nigeria"
##  [62] "Central African Republic"
##  [63] "Chad"
##  [64] "Congo"
##  [65] "Congo, DRC"
##  [66] "Uganda"
##  [67] "Tanzania"
##  [68] "Zanzibar"
##  [69] "Ethiopia"
##  [70] "Angola"
##  [71] "Zambia"
##  [72] "Comoros"
##  [73] "Libya (Tripolitania, Cyrenaica, Fezzan)"
```

```
##  [74] "Sudan"
##  [75] "South Sudan"
##  [76] "Iran"
##  [77] "Turkey"
##  [78] "Iraq"
##  [79] "Egypt"
##  [80] "Syria"
##  [81] "Lebanon"
##  [82] "Afghanistan"
##  [83] "Turkmenistan"
##  [84] "Tajikistan"
##  [85] "Kyrgyzstan"
##  [86] "Uzbekistan"
##  [87] "Kazakhstan"
##  [88] "China"
##  [89] "Mongolia"
##  [90] "Democratic People's Republic of Korea"
##  [91] "Republic of Korea"
##  [92] "Japan"
##  [93] "India"
##  [94] "Pakistan"
##  [95] "Myanmar"
##  [96] "Sri Lanka"
##  [97] "Thailand"
##  [98] "Laos"
##  [99] "Vietnam"
## [100] "Republic of Vietnam"
## [101] "Philippines"
## [102] "Indonesia"
## [103] "Papua New Guinea"
```

```
af_sub <- c("Mali", "Nigeria", "Chad", "Congo, DRC", "Comoros", "Sudan", "Zambia", "Libya (Tripolitania
            "Angola", "Ethiopia", "Congo", "Central African Republic", "Niger")

epr_join_af <- epr_join %>%
  dplyr::filter(statename %in% c(af_sub)) %>%
  dplyr::group_by(statename, from, to, group) %>%
  dplyr::summarise_all(dplyr::funs(dplyr::first(na.omit(.)))) %>%
  dplyr::filter(status != "SELF-EXCLUSION" &
                  !is.na(reg_aut) &
                  type != "Urban") %>%
  dplyr::group_by(statename, group) %>%
  tidyr::complete(group, from = min(from):2019) %>%
  tidyr::fill(gwid, groupid, gwgroupid, umbrella, size, status, reg_aut, sqkm, type, the_geom) %>%
  dplyr::select(-to, -the_geom) %>%
  dplyr::rename(year = from)
```

```
## Warning: funs() is soft deprecated as of dplyr 0.8.0
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
```

```
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once per session.
```

```
# Create final subsetted data for Mali
final <- sf::st_sf(dplyr::left_join(epr_join_af, geom)) %>%
  dplyr::filter(statename %in% c("Mali"))
```

```
## Joining, by = "group"
```

```
## Warning: Column `group` joining character vector and factor, coercing into
## character vector
```

```
write.csv(final, paste0(here::here(), "/data/epr_af.csv"))
sf::st_write(final, paste0(here::here(), "/data/epr_af.shp"), delete_dsn = TRUE)
```

```
## Deleting source `C:/Users/Tom Brailey/Dropbox/github_private/RegionalAutonomyGIS/data/epr_af.shp' us
## Writing layer `epr_af' to data source `C:/Users/Tom Brailey/Dropbox/github_private/RegionalAutonomyGI
## Writing 240 features with 12 fields and geometry type Unknown (any).
```

```
#rm(epr, epr_join, epr_sub, geo_epr, geo_epr_sub, group_aut, group_f, group_t, af_sub)
```