Violations flag/select by dates

Jim Sheehan

April 30, 2019

## 0. Load packages

library(RMySQL)  
library(data.table)  
library(dplyr)

## I. Get data

kr\_name <- "my\_keyring"  
kr\_service <- "my\_database"  
kr\_username <- "my\_username"  
kr\_password <- "my\_password"  
  
db\_name <- "safe\_water"  
db\_host <- "127.0.0.1" # for local access  
db\_port <- 3306  
  
mydb <- dbConnect(MySQL(),   
 user = keyring::backend\_file$new()$get(service = kr\_service,  
 user = kr\_username,  
 keyring = kr\_name),   
 password = keyring::backend\_file$new()$get(service = kr\_service,  
 user = kr\_password,  
 keyring = kr\_name),   
 dbname = db\_name, host = db\_host, port = db\_port)

## Please enter password in TK window (Alt+Tab)

# dbListTables(mydb)  
  
db\_table <- 'VIOLATION'  
s <- paste0("SELECT \* FROM ", db\_table, " WHERE PWS\_ACTIVITY\_CODE = 'A'")  
rs <- dbSendQuery(mydb, s)  
VIOLATION <- fetch(rs, n = -1)  
  
dim(VIOLATION)

## [1] 1664916 35

dbDisconnect(mydb)

## Warning: Closing open result sets

## [1] TRUE

## II. Date type conversion

VIOLATION <- VIOLATION %>% mutate(COMPL\_PER\_BEGIN\_DATE = as.Date(COMPL\_PER\_BEGIN\_DATE),   
 COMPL\_PER\_END\_DATE = as.Date(COMPL\_PER\_END\_DATE),  
 RTC\_DATE = as.Date(RTC\_DATE))

## III. Flag rows meeting conditions

VIOLATION <- VIOLATION %>%  
 mutate(CPBD\_RTCD =   
 case\_when(COMPL\_PER\_BEGIN\_DATE < as.Date("2008-01-01") &   
 between(RTC\_DATE, as.Date("2008-01-01"),   
 as.Date("2015-12-31")) ~ "CPBD\_prior"),  
 RTCD\_CPBD =   
 case\_when(RTC\_DATE > as.Date("2015-12-31") &  
 between(COMPL\_PER\_BEGIN\_DATE, as.Date("2008-01-01"),   
 as.Date("2015-12-31")) ~ "RTCD\_post"))  
  
# take a look  
VIOLATION %>% select(PWSID, COMPL\_PER\_BEGIN\_DATE, RTC\_DATE, CPBD\_RTCD) %>%   
 filter(!is.na(CPBD\_RTCD)) %>% slice(1:10)

## PWSID COMPL\_PER\_BEGIN\_DATE RTC\_DATE CPBD\_RTCD  
## 1 IN5279014 1993-08-01 2013-04-22 CPBD\_prior  
## 2 090400049 2004-01-01 2010-10-28 CPBD\_prior  
## 3 IN2441054 2000-10-01 2013-04-24 CPBD\_prior  
## 4 IN5216003 1999-01-01 2010-05-07 CPBD\_prior  
## 5 NN0400380 2006-01-01 2012-10-16 CPBD\_prior  
## 6 CO0115288 2004-12-01 2012-04-30 CPBD\_prior  
## 7 TX1110006 2004-01-01 2010-07-01 CPBD\_prior  
## 8 IN2090865 2001-07-01 2012-04-10 CPBD\_prior  
## 9 NC0190516 2003-01-01 2010-06-11 CPBD\_prior  
## 10 105300027 2007-11-01 2009-04-01 CPBD\_prior

VIOLATION %>% select(PWSID, COMPL\_PER\_BEGIN\_DATE, RTC\_DATE, RTCD\_CPBD) %>%   
 filter(!is.na(RTCD\_CPBD)) %>% slice(1:10)

## PWSID COMPL\_PER\_BEGIN\_DATE RTC\_DATE RTCD\_CPBD  
## 1 TX1520225 2014-01-01 2018-10-11 RTCD\_post  
## 2 NY3830021 2010-01-01 2016-03-18 RTCD\_post  
## 3 NY3820553 2010-01-01 2016-03-18 RTCD\_post  
## 4 PA2590051 2013-10-01 2016-12-09 RTCD\_post  
## 5 AK2291952 2015-11-01 2016-08-02 RTCD\_post  
## 6 UTAH19002 2014-01-01 2017-11-03 RTCD\_post  
## 7 CA1503662 2012-01-01 2018-03-01 RTCD\_post  
## 8 CT0969223 2011-01-01 2016-12-31 RTCD\_post  
## 9 PA2590317 2014-10-01 2016-12-13 RTCD\_post  
## 10 CO0139434 2014-01-01 2017-04-12 RTCD\_post