assignment_4.R

jyiwu

2022-02-17

```
# # ================================================
# GBA464: Assignment 4
# Author: Yufeng Huang
# Description: write a function that plots crimes
#              incidence in Baltimore city
# Data: Baltimore crime data
# Source: https://data.baltimorecity.gov/
# ================================================
# DUE:  Sunday, Tuesday 12 at 11:59pm
# Send by email to r.programming.simon@gmail.com, one copy per team
#    please inidicate the set of team members


# clear everything
rm(list = ls())

# libraries
#   need to install.packages() these
#   let me know if installation does not work
library(maps)
library(maptools)
```

```
## Loading required package: sp
```

```
## Checking rgeos availability: TRUE
## Please note that 'maptools' will be retired by the end of 2023,
## plan transition at your earliest convenience;
## some functionality will be moved to 'sp'.
```

```
library(dplyr)
```

```
## Registered S3 methods overwritten by 'tibble':
##   method     from
##   format.tbl pillar
##   print.tbl  pillar
```

```
##
## 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
```

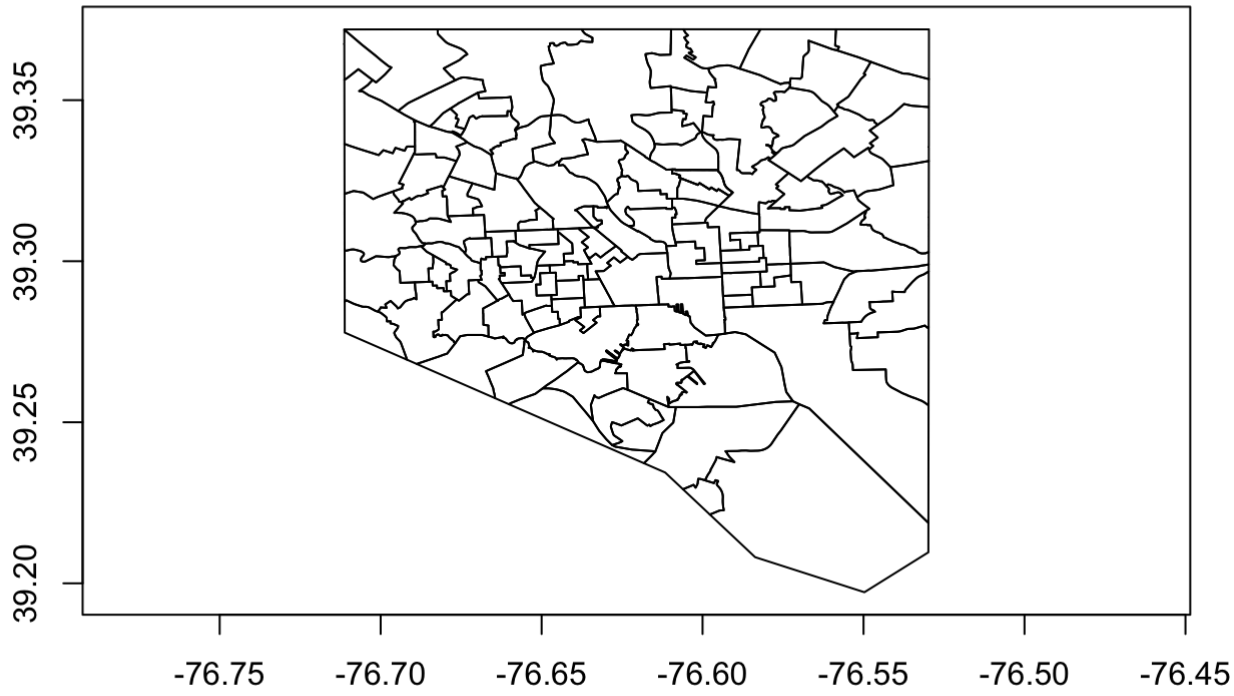
```
library(ggplot2)

# download, unzip and read the shape file
url_zip <- 'https://dl.dropboxusercontent.com/s/chyvmlrkkk4jcgb/school_distr.zip'
if(!file.exists('school_distr.zip')) download.file(url_zip, 'school_distr.zip')     # do
wnload file as zip
unzip('school_distr.zip')   # unzip in the default folder
schdstr_shp <- readShapePoly('school.shp')  # read shape file
```

```
## Warning: readShapePoly is deprecated; use rgdal::readOGR or sf::st_read
```

```
xlim <- schdstr_shp@bbox[1,]
ylim <- schdstr_shp@bbox[2,]

# example of how to use the shape file
#   if there are no error code reading the above, you can directly plot the map of Balti
more (lines within are school districts)
#   we'll be overlaying our plots of crime incidents on this map:
plot(schdstr_shp, axes = T)     # axes = T gives x and y axes
```




```
# ======= now let's follow instructions in the pdf file ======

# download and load the crime csv data
#   link is https://dl.dropboxusercontent.com/s/4hg5ffdds9n2nx3/baltimore_crime.csv

df = read.csv("https://dl.dropboxusercontent.com/s/4hg5ffdds9n2nx3/baltimore_crime.csv",
header = TRUE, stringsAsFactors = FALSE)

# transform dates and time variables depending on what you need

date <- as.Date(df$CrimeDate,format = "%m/%d/%Y")
df$month <- as.numeric(format(date, "%m"))
df$day <- as.numeric(format(date, "%d"))
time <- as.POSIXlt(strptime(df$CrimeTime,"%H:%M:%S"))
hour <- as.numeric(format(time, "%H"))
minute <- as.numeric(format(time, "%M"))/60
df$time <- hour + minute
head(df)
```

```
##    CrimeDate CrimeTime CrimeCode            Location       Description  Weapon
## 1 07/16/2016  19:51:00        4A      900 BENNETT PL      AGG. ASSAULT FIREARM
## 2 07/16/2016      1951        9S      900 BENNETT PL          SHOOTING FIREARM
## 3 07/16/2016  20:30:00        6D      6900 MOYER AVE LARCENY FROM AUTO
## 4 07/16/2016  21:00:00        6E ST & S CATHERINE ST           LARCENY
## 5 07/16/2016  22:00:00        7A         0 N PACA ST        AUTO THEFT
## 6 07/16/2016  22:20:00        4E     1100 COOKSIE ST    COMMON ASSAULT   HANDS
##   Post     District       Neighborhood                       Location1
## 1  713      WESTERN        Harlem Park (39.2950500000, -76.6325000000)
## 2  713      WESTERN        Harlem Park (39.2950500000, -76.6325000000)
## 3  424 NORTHEASTERN North Harford Road (39.3630100000, -76.5421400000)
## 4  841 SOUTHWESTERN   Carrollton Ridge (39.2809000000, -76.6541500000)
## 5  111      CENTRAL           Downtown (39.2893200000, -76.6222400000)
## 6  943     SOUTHERN       Locust Point (39.2728600000, -76.5907400000)
##   TotalIncidents month day     time
## 1              1     7  16 19.85000
## 2              1     7  16       NA
## 3              1     7  16 20.50000
## 4              1     7  16 21.00000
## 5              1     7  16 22.00000
## 6              1     7  16 22.33333
```

```
# split coordinates into longitude and latitude, both as numeric
# note: no for/while/repeat loop, and no substr() function

df$longitude <- gsub(".*,|\\s|\\)","",df$Location1)
df$latitude <- gsub(",\\s.*|\\(","",df$Location1)
df <- select(df,c("Location","District","CrimeDate","month","day","CrimeTime","time","la
titude","longitude","Description"))
head(df)
```

```
##              Location     District  CrimeDate month day CrimeTime     time
## 1      900 BENNETT PL      WESTERN 07/16/2016     7  16  19:51:00 19.85000
## 2      900 BENNETT PL      WESTERN 07/16/2016     7  16      1951       NA
## 3      6900 MOYER AVE NORTHEASTERN 07/16/2016     7  16  20:30:00 20.50000
## 4 ST & S CATHERINE ST SOUTHWESTERN 07/16/2016     7  16  21:00:00 21.00000
## 5         0 N PACA ST      CENTRAL 07/16/2016     7  16  22:00:00 22.00000
## 6     1100 COOKSIE ST     SOUTHERN 07/16/2016     7  16  22:20:00 22.33333
##        latitude      longitude       Description
## 1 39.2950500000 -76.6325000000      AGG. ASSAULT
## 2 39.2950500000 -76.6325000000          SHOOTING
## 3 39.3630100000 -76.5421400000 LARCENY FROM AUTO
## 4 39.2809000000 -76.6541500000           LARCENY
## 5 39.2893200000 -76.6222400000        AUTO THEFT
## 6 39.2728600000 -76.5907400000    COMMON ASSAULT
```

```
# generate geographic and time patterns for crimes with keyword "ASSAULT"
# note: no copy and paste of the same/similar command many times
assult <- grep("ASSAULT", df$Description)
df_assault <- data.frame(df$longitude[assult],df$latitude[assult],df$time[assult])
colnames(df_assault) <- c("longitude","latitude", "time")
head(df_assault)
```

```
##        longitude      latitude      time
## 1 -76.6325000000 39.2950500000 19.850000
## 2 -76.5907400000 39.2728600000 22.333333
## 3 -76.5487500000 39.3271300000 23.666667
## 4 -76.6454700000 39.2837700000  1.250000
## 5 -76.5843400000 39.3008600000  1.666667
## 6                               2.166667
```

```
par(mfrow = c(2,2))
pdf("plot.pdf")
for (i in 0:3) {
  i = i * 6
  u = i + 6
  plot(schdstr_shp, axes = T, main = paste0("hour: ", i, "-", u) )
  points(df_assault$longitude[df_assault$time > i & df_assault$time < u],
         df_assault$latitude[df_assault$time > i & df_assault$time < u],
         pch = 16, cex = 0.5, col=rgb(1,0,0.05,0.1))
  if (i > 24) break
}
dev.off()
```

```
## quartz_off_screen
##                 2
```