00_Leukemia_in_NY

Frances Lin

3/23/2022

Load packages

```
# Load packages
library(spdep)
```

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

```
## Loading required package: spData
```

```
## Loading required package: sf
```

```
## Linking to GEOS 3.9.1, GDAL 3.4.0, PROJ 8.1.1; sf_use_s2() is TRUE
```

```
library(DClusterm) # data
```

```
## Loading required package: parallel
```

```
## Loading required package: spacetime
```

```
## Loading required package: DCluster
```

```
## Loading required package: boot
```

```
## Loading required package: MASS
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x dplyr::select() masks MASS::select()
```

```
library(pander)
library(ggplot2)
```

The NY8 data

The NY8 data set contains the number of leukemia cases in an eight-country region of upstate New York from 1978-1982.

```
# Load data
data(NY8)

# View data
#head(NY8)
NY8
```

```
## class       : SpatialPolygonsDataFrame
## features    : 281
## extent      : 358241.9, 480393.1, 4649755, 4808545  (xmin, xmax, ymin, ymax)
## crs         :  +proj=utm +zone=18 +ellps=WGS84 +units=m +no_defs
## variables   : 17
## names       :    AREANAME,     AREAKEY,        X,        Y,  POP8, TRACTCAS,  PROPCAS, PCTOWNHOME,  I
## min values  : Auburn city, 36007000100, -55.4823, -75.2907,     9,        0,        0, 0.00082237, 0
## max values  : Vestal town, 36109992300,  53.5086, 56.41013, 13015,     9.29, 0.006993,          1, 0
```

```
# Check class
class(NY8)
```

```
## [1] "SpatialPolygonsDataFrame"
## attr(,"package")
## [1] "sp"
```

```
# Convert it to a df?
# https://www.paulamoraga.com/book-geospatial/sec-spatialdataandCRS.html
head(NY8@data) %>% pander
```

Table 1: Table continues below

	AREANAME	AREAKEY	X	Y	POP8	TRACTCAS
0	Binghamton city	36007000100	4.069	-67.35	3540	3.08
1	Binghamton city	36007000200	4.639	-66.86	3560	4.08
2	Binghamton city	36007000300	5.709	-66.98	3739	1.09
3	Binghamton city	36007000400	7.614	-66	2784	1.07
4	Binghamton city	36007000500	7.316	-67.32	2571	3.06
5	Binghamton city	36007000600	8.559	-66.93	2729	1.06

Table 2: Table continues below

	PROPCAS	PCTOWNHOME	PCTAGE65P	Z	AVGIDIST	PEXPOSURE
0	0.00087	0.3277	0.1466	0.142	0.2374	3.167
1	0.001146	0.4268	0.2351	0.3555	0.2087	3.039
2	0.000292	0.3377	0.138	-0.5817	0.1709	2.838
3	0.000384	0.4616	0.1189	-0.2963	0.1406	2.643
4	0.00119	0.1924	0.1416	0.4569	0.1578	2.759
5	0.000388	0.3652	0.1411	-0.2812	0.1726	2.848

	Cases	Xm	Ym	Xshift	Yshift
0	3.083	4069	-67353	423391	4661502
1	4.083	4639	-66862	423961	4661993
2	1.087	5709	-66978	425031	4661878
3	1.065	7614	-65996	426935	4662859
4	3.06	7316	-67318	426638	4661537
5	1.064	8559	-66934	427880	4661921

```
# # Plot it
# plot(NY8) # Just the map now.
```

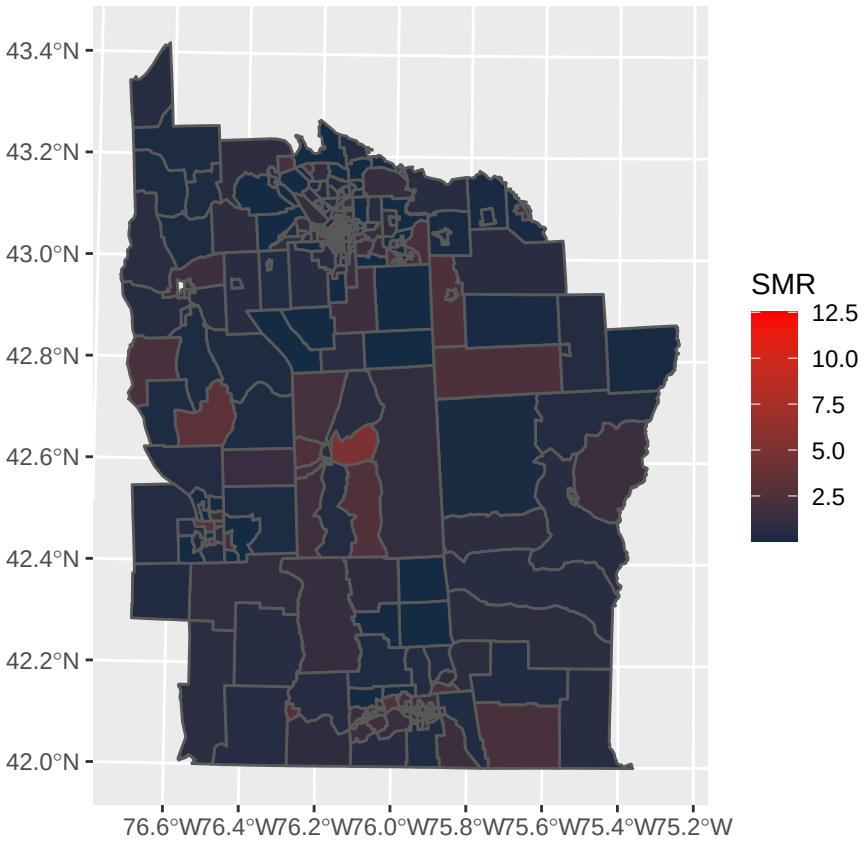
Plotting

```
# Convert to sf
library(sf)
NY8_sf <- st_as_sf(NY8)
```

```
# Create the standardized mortality ratio (SMR) variable
# https://www.r-bloggers.com/2019/11/spatial-data-analysis-with-inla/
rate <- sum(NY8_sf$Cases) / sum(NY8_sf$POP8)

NY8_sf <- NY8_sf %>% mutate(
  Expected = POP8 * rate,
  SMR = Cases / Expected
)
```

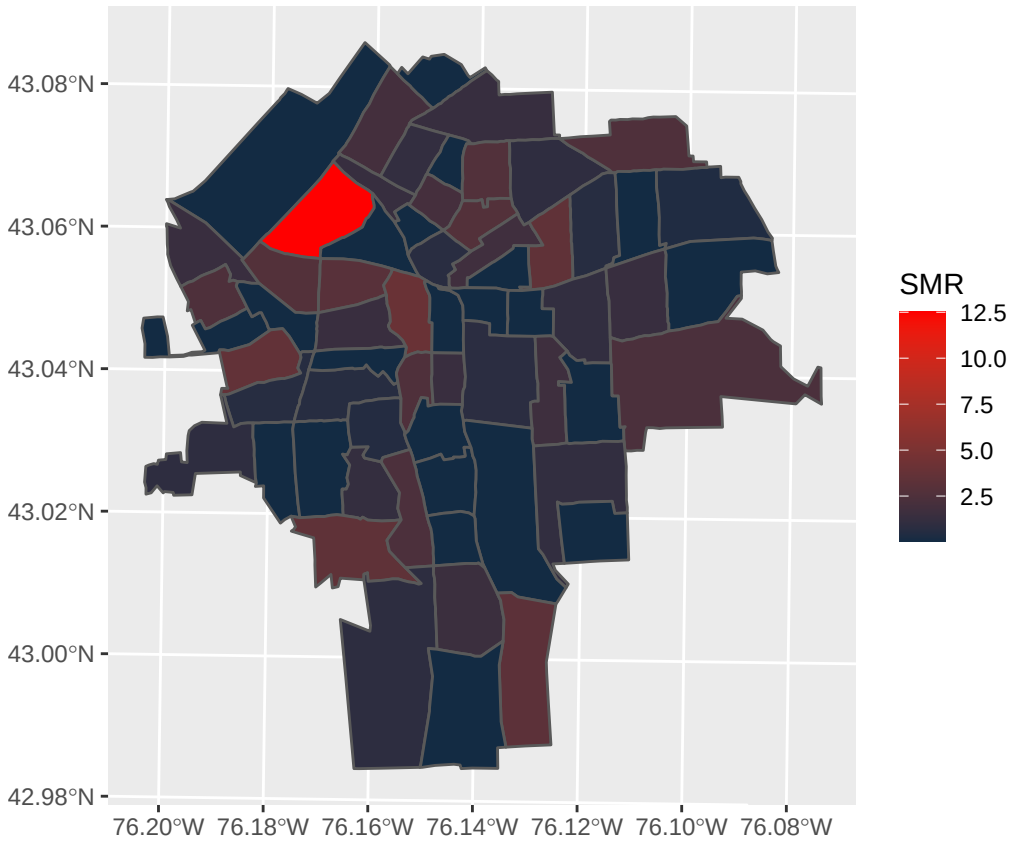
```
# Plot SMR
ggplot(NY8_sf) + geom_sf(aes(fill = SMR)) + # Look nice!
  scale_fill_gradient(high="red")
```



Subsetting then plotting

```
# Subset to include Syracuse city only
syracuse <- which(NY8$AREANAME == "Syracuse city")

# Plot it
ggplot(NY8_sf[syracuse, ]) + geom_sf(aes(fill = SMR)) +
  scale_fill_gradient(high="red")
```



Fitting a mixed-effects (Poisson regression) model

```
#install.packages("INLA") # run once
#not available for this R version...
#install.packages("INLA", repos=c(getOption("repos"), INLA="https://inla.r-inla-download.org/R/stable")
library(INLA) # now it works?
```

```
## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack

## Loading required package: foreach

##
## Attaching package: 'foreach'

## The following objects are masked from 'package:purrr':
##
##     accumulate, when
```

```
## This is INLA_22.03.16 built 2022-03-16 13:24:07 UTC.
##  - See www.r-inla.org/contact-us for how to get help.
```

Let's work on some toy examples first before coming to fix the issue.

```
# # Fit a Poisson regression model
# m1 <- inla(Cases ~ 1 + AVGIDIST,
#            data = NY8_sf,
#            family = "poisson",
#            E = NY8_sf$Expected,
#            control.predictor = list(compute = TRUE),
#            control.compute = list(dic = TRUE, waic = TRUE))
```

```
# # Fit a different model (random-effects model)
# NY8_sf <- NY8_sf %>% mutate(
#   ID = 1:nrow(NY8))
#
# m2 <- inla(Cases ~ 1 + AVGIDIST + f(ID, model = "iid"),
#   data = as.data.frame(NY8), family = "poisson",
#   E = NY8$Expected,
#   control.predictor = list(compute = TRUE),
#   control.compute = list(dic = TRUE, waic = TRUE))
```

Reference

Gómez-Rubio, V. (2019). R-bloggers. Spatial Data Analysis with INLA. https://www.r-bloggers.com/2019/11/spatial-data-analysis-with-inla/.