BST 6200 Spatial Statistics and Disease Mapping
Homework 4

Miao Cai

Overall Goal: Perform an ecological study of the relationship between smoking and pancreatic cancer rates in the state of Minnesota.

1 Homework description

Obtain the shape file for Minnesota through the tigris package.

Obtain the smoking data by selecting Minnesota and Adult Smoking here:

https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors/health-behaviors/tobacco-use

You can probably copy the data and paste it into Excel.

Obtain the number of cases and the population size here: https://data.web.health.state.mn.us/cancer_query.

Use shift-click to select all counties. Then select "Pancreas" under Indicator, "2012-2016" under Year, and "All" under Sex. Then click "Submit" and finally "Download" at the bottom of the data. Two of the columns in the resulting file are "count" and "population". You will need both of these. You might want to clean up this file using Excel.

As part of your report, give choropleth maps of smoking rates and pancreatic cancer rates. Compute Moran's I for both and assess their significance.

Run a regression model of the form:

$$Y_i \sim \text{Poisson}(P_i \eta_i)$$

where

$$\eta_i = \exp(\beta_0 + \beta_1 x_i + u_i + \nu_i)$$

$$P_i = \text{population of county } i$$

u_i is correlated heterogeneity of county i

v_i is uncorrelated heterogeity of county i

x_i is the smoking rate of county i

Write a report addressing the question of smoking and pancreatic cancer. The project shouldn't be too long. I'm thinking something like 3 to 6 pages, counting figures.

[Note: This project is much like the term project for the course, except I have pointed you to the data and given you a specific model with just one predictor variable (smoking). You should look for more than one predictor variable in your term project model.]

2 Data

```r
pacman::p_load(tigris, sp, ggplot2, nimble,
               tmap, dplyr, sf, data.table, rgeos)

MN = counties("Minnesota", cb = TRUE)

##   |                                                                    |
adult_smoke = fread("data/Minnesota_adult_smoking.csv") %>%
  select(FIPS, smoking_rate = `% Smokers`)
pancreas = fread("data/Minnesota_Pancreas_cancer.csv") %>%
  select(FIPS = fips,
         year,
         N_pancreas = count,
         population,
         rate_pancreas = rate) %>%
  mutate(rate_pancreas = gsub(" \\(UR\\)", "", rate_pancreas) %>%
           as.numeric())

MN_sf = st_as_sf(MN) %>%
  mutate(FIPS = as.integer(paste0(STATEFP, COUNTYFP))) %>%
  select(NAME, FIPS) %>%
  arrange(NAME) %>%
  left_join(pancreas, by = 'FIPS') %>%
  left_join(adult_smoke, by = 'FIPS')

head(MN_sf)
```

```
## Simple feature collection with 6 features and 7 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: -96.83545 ymin: 45.03536 xmax: -93.01841 ymax: 48.54035
## CRS:            +proj=longlat +datum=NAD83 +no_defs +ellps=GRS80 +towgs84=0,0,0
##        NAME  FIPS      year N_pancreas population rate_pancreas smoking_rate
## 1    Aitkin 27001 2012-2016         24      79419          12.9           15
## 2     Anoka 27003 2012-2016        220    1709067          12.4           15
## 3    Becker 27005 2012-2016         30     166637          13.0           17
## 4  Beltrami 27007 2012-2016         37     227854          14.8           19
## 5    Benton 27009 2012-2016         28     195535          14.3           17
## 6 Big Stone 27011 2012-2016          6      25301          12.3           15
##                         geometry
## 1 MULTIPOLYGON ((((-93.81217 4...
## 2 MULTIPOLYGON ((((-93.51251 4...
## 3 MULTIPOLYGON ((((-96.19467 4...
## 4 MULTIPOLYGON ((((-95.60175 4...
## 5 MULTIPOLYGON ((((-94.35295 4...
## 6 MULTIPOLYGON ((((-96.83063 4...
```

3 Choropleth maps

```r
map_pancreas = tm_shape(MN_sf) +
  tm_fill(title = "Pancreas cancer\nrates",
          col = "rate_pancreas",
```

```
        palette = "Reds") +
  tm_borders(col = "black") +
  tm_layout(main.title = "Pancreas cancer rates in Minnesota",
            main.title.size = 1.2, frame = FALSE) +
  tm_legend(legend.position = c("right", "center"))

map_smoking = tm_shape(MN_sf) +
  tm_fill(title = "Adult smoking\nrates",
          col = "smoking_rate",
          palette = "Reds") +
  tm_borders(col = "black") +
  tm_layout(main.title = "Adult smoking rates in Minnesota",
            main.title.size = 1.2, frame = FALSE) +
  tm_legend(legend.position = c("right", "center"))

tmap_arrange(map_pancreas, map_smoking, ncol = 2)
```
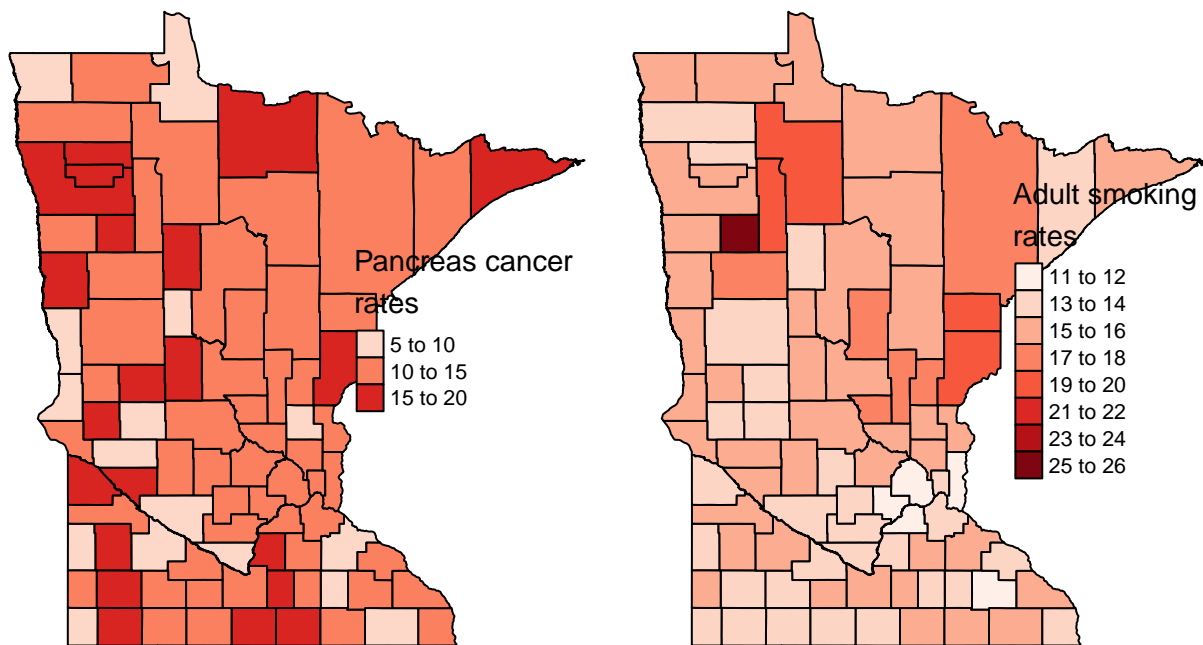

Figure 1: Choropleth maps of pancreas cancer rates (left) and adult smoking rates (right) in Minnesota