Problem Set 5

Shisham Adhikari

Math 345

Due: Before class on Thursday, 03/11/2021

Collaborators: Insert names of anyone you talked about this assignment with

Goals of this lab

1. Take the work you've done in Problem Set 4 and add some covariates Load any packages you need:

```
pacman::p_load(tidyverse, spdep, spatstat, raster, glm, sf, tidycensus, sp)
```

Chooses two covariates that you'd like to explore related to your power plant points. You may not select population density. Select one that you think explains the distribution of the power plants and another that you think does not. Load them here:

```
census_api_key("e20f7e545ee9474d9d353d401cc0e1d00d31deeb")
pa_pop <- get_acs(geography = "county",
      variables = c(pop="B01003_001"),
      state = "PA",
       year = 2019,
       geometry = TRUE)
```

```
##   |                                                                          |
```

```
pa_pop <- pa_pop %>% dplyr::select(estimate, geometry)
pa_inc <- get_acs(geography = "county",
      variables = c(medincome = "B19013_001"),
      state = "PA",
       year = 2019,
       geometry = TRUE)
pa_inc <- pa_inc %>% dplyr::select(estimate, geometry)
```

Problem 1

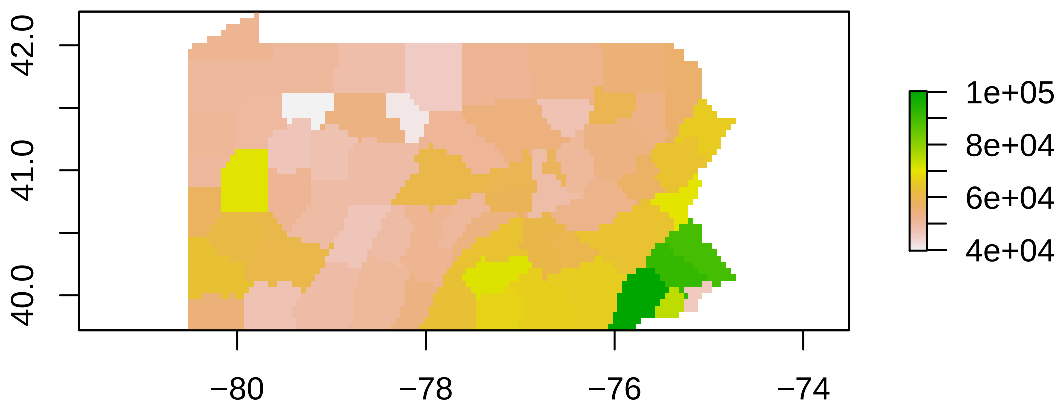
Create a tesslated quadrat map with the variable that you think should be related to power plant spatial distribution.

```
# Load an PApolygon shapefile
pa_shape <- spData::us_states %>% filter(NAME=="Pennsylvania")
pashape_sf <- pa_shape$geometry
pa_transform <- sf::st_transform(pashape_sf, crs = 6345)
pa_transform <- as(pa_transform, "Spatial")
w <- as(pa_transform, "owin")
```

```r
#A ppp point layer of Coal power plant in PA
coal <- as.data.frame(read_csv("coal.csv"))
coals_points <-as.ppp(coal, c(min(coal$X), max(coal$X), min(coal$Y), max(coal$Y)))

#An im raster layer of income distribution
pa_sp <- as(pa_inc, 'Spatial')
r <- raster(pa_sp)
res(r) <- 0.05
g <- rasterize(pa_sp, r, pa_inc$estimate)
plot(g, main="Income Distribution Across PA")
```

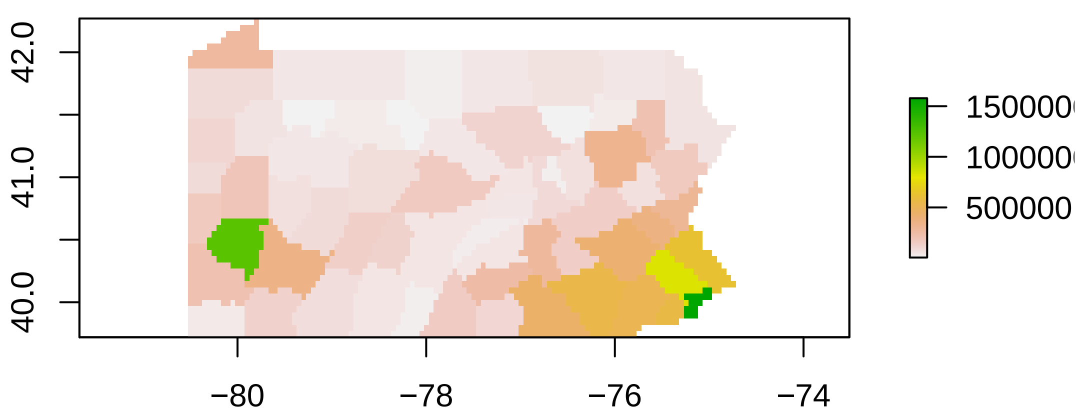
Income Distribution Across PA




```r
inc <- as.im(g)

#An im raster layer of population distribution
papop_sp <- as(pa_pop, 'Spatial')
r <- raster(papop_sp)
res(r) <- 0.05
gg <- rasterize(papop_sp, r, pa_pop$estimate)
plot(gg, main="Population Distribution Across PA")
```

Population Distribution Across PA

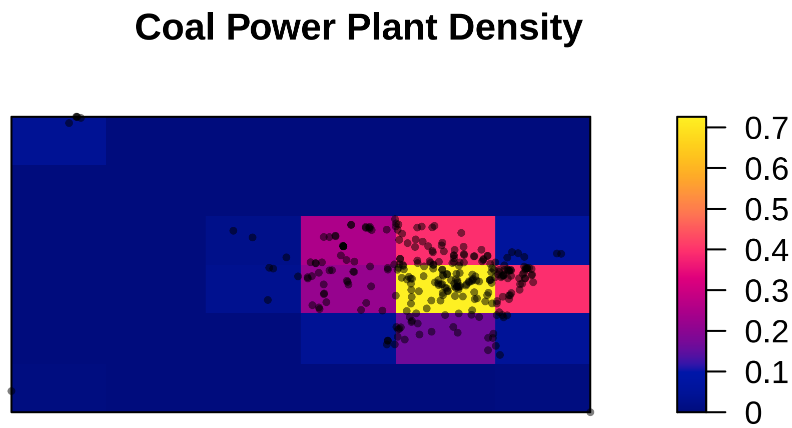



```r
pop <- as.im(gg)
```

```r
#inc=pop, coals_points=starbucks, w=ma
marks(coals_points) <- NULL
```

```r
# Compute the density for each quadrat (in counts per km2)
Q   <- quadratcount(coals_points, nx= 6, ny=6)
Q.d <- intensity(Q)

# Plot the density
plot(intensity(Q, image=TRUE), main="Coal Power Plant Density", las=1)  # Plot density raster
plot(coals_points, pch=20, cex=0.6, col=rgb(0,0,0,.5), add=TRUE)  # Add points
```



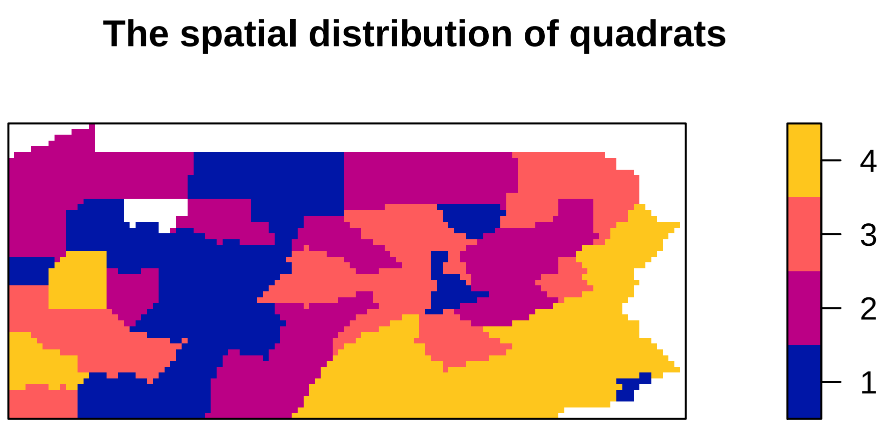
Looking at the normal density map, looks like the coal plants are densed around south-east PA.

#Quadrat density on a tessellated surface

```r
brk  <- c(quantile(pa_inc$estimate)[1],
         quantile(pa_inc$estimate)[2],
         quantile(pa_inc$estimate)[3],
         quantile(pa_inc$estimate)[4],
         quantile(pa_inc$estimate)[5])# Define the breaks
Zcut <- cut(inc, breaks=brk, labels=1:4)  # Classify the raster
E    <- tess(image=Zcut)  # Create a tesselated surface
```

The tessellated object can be mapped to view the spatial distribution of quadrats.

```r
plot(E, main="The spatial distribution of quadrats", las=1)
```



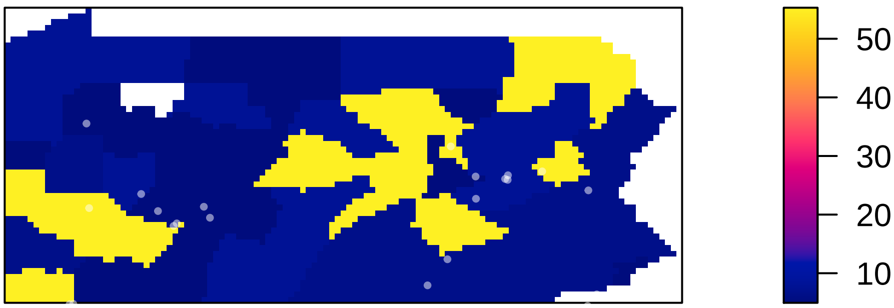

```r
Q   <- quadratcount(coals_points, tess = E)  # Tally counts
Q.d <- intensity(Q)  # Compute density
```

```
Q.d
```

```
## tile
##         1         2         3         4
##  4.946414  7.899934 55.299539  6.402696
```

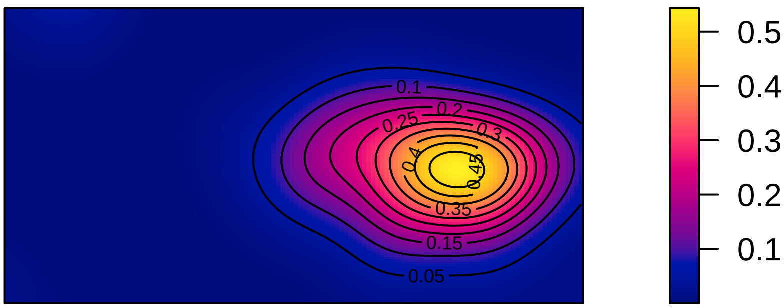
```
plot(intensity(Q, image=TRUE), las=1, main="The coal density values across each tessellated region")
plot(coals_points, pch=20, cex=0.6, col=rgb(1,1,1,.5), add=TRUE)
```

The coal density values across each tessellated region




```
K1 <- density(coals_points) # Using the default bandwidth
plot(K1, main="Kernel density raster", las=1)
contour(K1, add=TRUE)
```

Kernel density raster



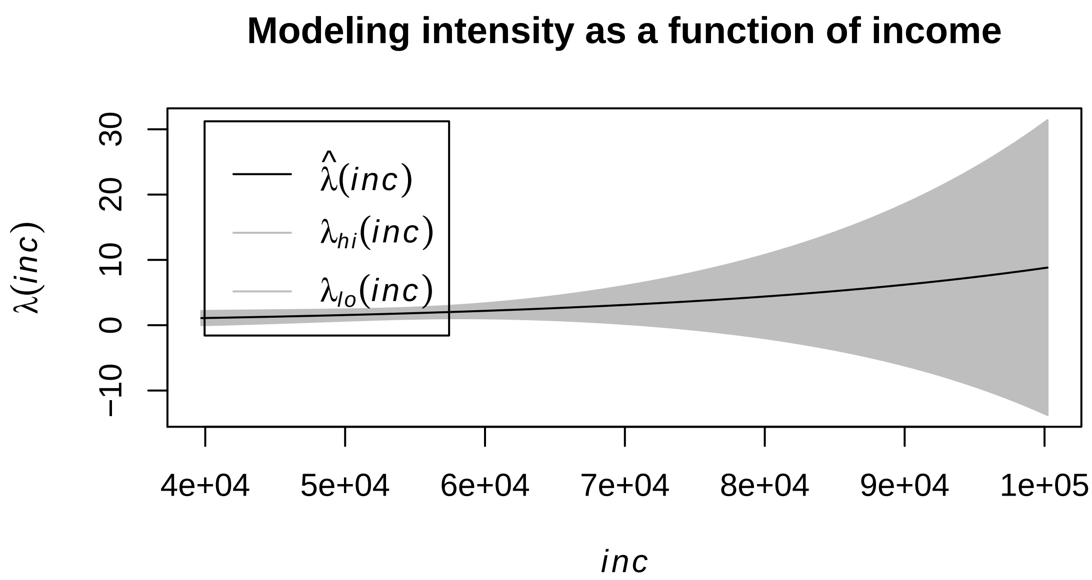
- Yes the covariate appear to be related. Based on our tessalated map, there seem to be more coal power plants in counties with lower median income than those places with higher median income. This in fact matches my hypothesis.

Problem 2

a) Model each intensity as a function of each covariate. Plot the relationships.

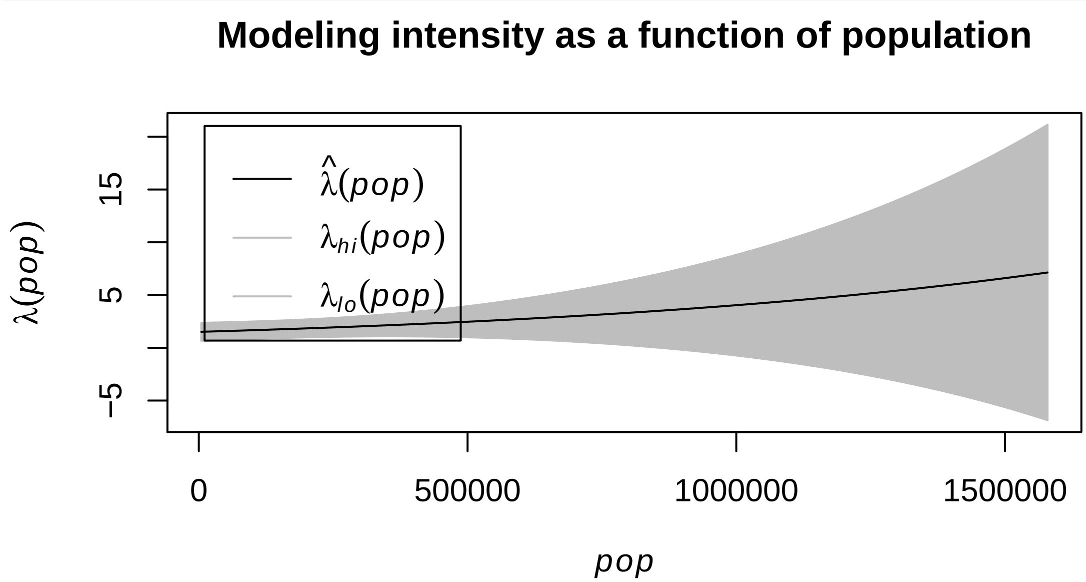
With Income Distribution

```
# Create the Poisson point process model
ppm1 <- ppm(coals_points ~ inc)
# Plot the relationship
plot(effectfun(ppm1,se.fit=TRUE), main="Modeling intensity as a function of income")
```



With Population Distribution

```r
# Create the Poisson point process model
ppm2 <- ppm(coals_points ~ pop)
# Plot the relationship
plot(effectfun(ppm2,se.fit=TRUE), main="Modeling intensity as a function of population")
```



b) Run a likelihood ratio test with each covariate. Can you reject the homogeneous model?

```r
ppm0 <- ppm(coals_points ~ 1)
#For Income Distribution
anova(ppm0, ppm1, test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: ~1   Poisson
## Model 2: ~inc     Poisson
##   Npar Df Deviance Pr(>Chi)
```

```
## 1 1871
## 2 1872  1   1.3467   0.2458
```

The value under Pr(>Chi) is the p-value which gives us the probability that we would be wrong in rejecting the null. Here p>0.05 suggests that there is close to a 62% chance that we would be wrong in rejecting the base model in favor of the alternate model–put another way, the alternate model (that the income distribution can help explain the distribution of coal power plants) is a significant improvement over the null. Thus, there is a higher chances that we can't reject the null of homogenous model.

```r
#For Population Distribution
anova(ppm0, ppm2, test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: ~1   Poisson
## Model 2: ~pop     Poisson
##   Npar Df Deviance Pr(>Chi)
## 1 1871
## 2 1872  1   1.5376    0.215
```

Surprisingly, the p-value is smaller than the income variable. The Pr(>Chi)=0.3674 suggests that there is close to a 37% chance that we would be wrong in rejecting the base model in favor of the alternate model–put another way, the alternate model (that the population distribution can help explain the distribution of coal power plants) is a significant improvement over the null.Thus, there is a higher chance that we can't reject the null of homogenous model but the chances is lower than using the income covariate.

c) Compare the two non-homogeneous models.

```r
PPPM1 <- ppm(coals_points ~ pop)
PPPM2 <- ppm(coals_points ~ pop + inc)
anova(PPPM1, PPPM2, test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: ~pop     Poisson
## Model 2: ~pop + inc   Poisson
##   Npar Df Deviance Pr(>Chi)
## 1    2
## 2    3  1  0.28041   0.5964
```

Again, the p-value greater than 0.05 suggests that we can't reject the null more than 95% of the time. In fact, Pr(>Chi)=0.4058 suggests that there is close to a 41% chance that we would be wrong in rejecting the base model in favor of the alternate model.

Were your priors about the two covariates correct?

- No, my priors were not correct about the two covariates. I selected income as the variable that would affect coal power plant density and population as the variable that wouldn't. But, based on the analysis, it looks like population affects coal plant density more than income does. Honestly, it makes sense now that I think about it; the coal plant density are less likely to be in densely populated areas. And also, unlike my prior hypothesis, coal power plant might not be always in neighborhoods with lower median income; in fact, people near coal power plants have potential to be richer because of all the money they get from the power plant and we mightn't be able to tell much about coal power density just based on income distribution. Cool stuff!