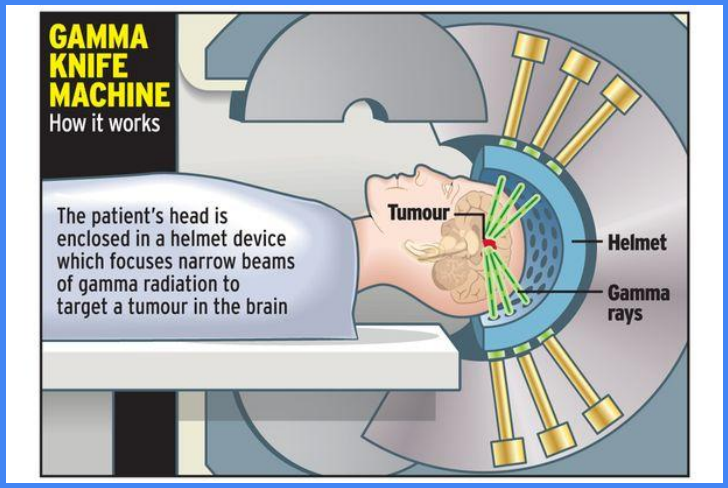
GIS (Geographical Information System)

By: Elliot Snitzer, Sareh Ghadersohi, Jacob Springborn

The Workshop



- Visited Roswell Park, UB CTRC, Woodward Institute
- Presentations on supercomputers at CCR, computer models of Buffalo, lectures on various branches of science
- Focused on R language
- Learned how to manage data sets
- Data Structures: Vectors and Data Frames, Descriptive Statistics, Writing Functions, Visualizing Datasets

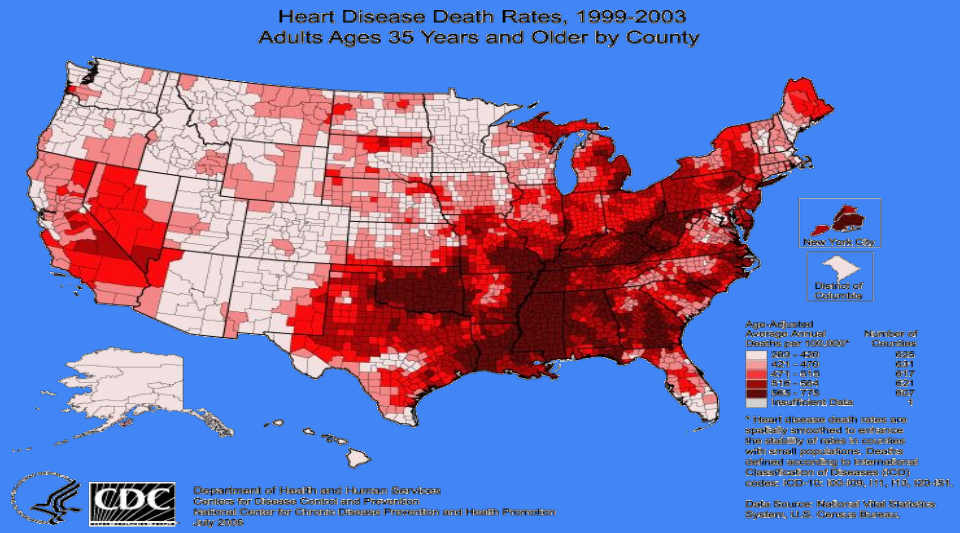
Sparcs Dataset

- De-identified Medical records from Hospital Admissions in NY State, 2011
- About 2.5 million records
- Records included reasons for hospitalization, age range, length of stay, severity of condition, geographical info, etc.
- Final Project: four groups

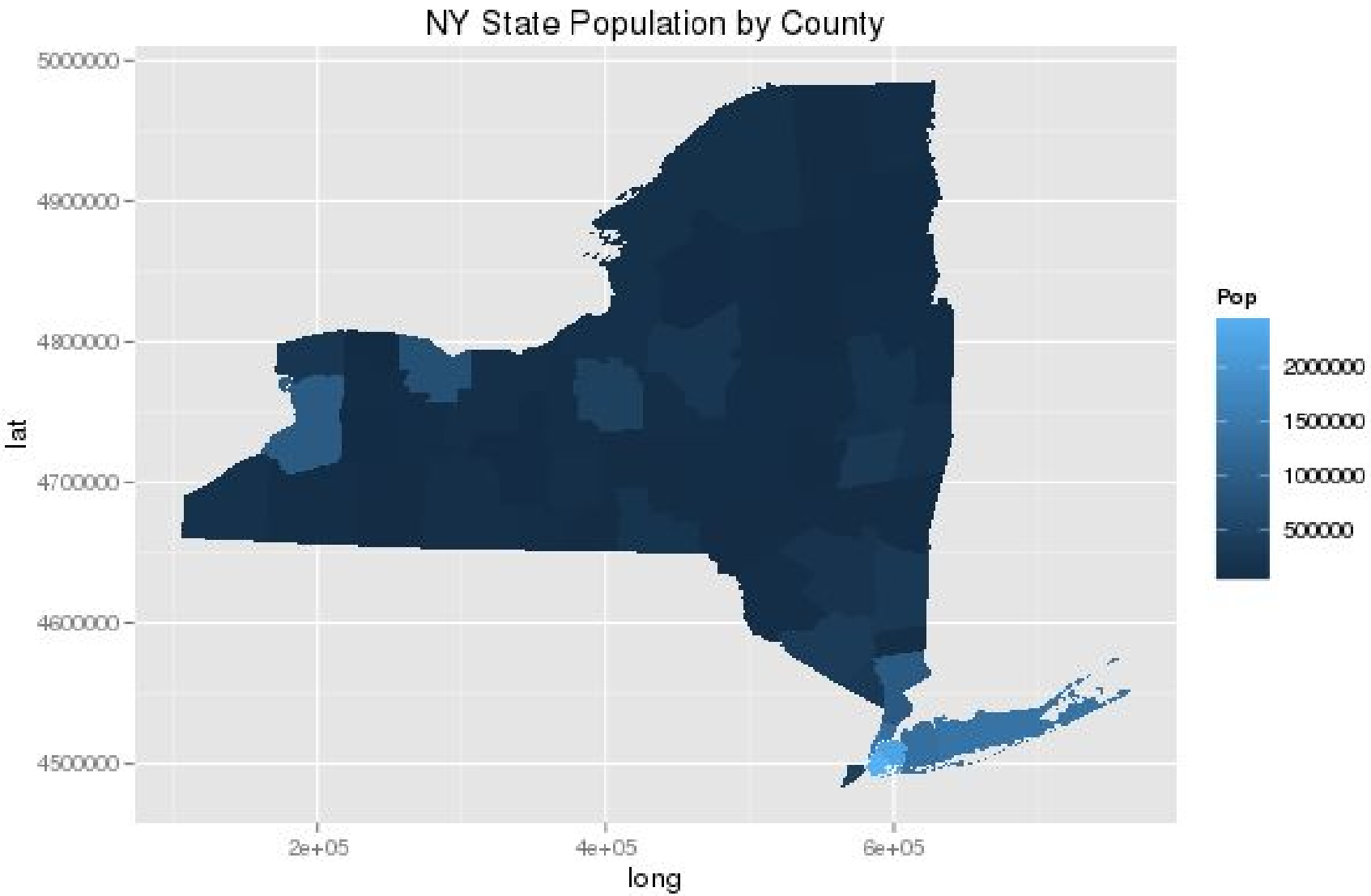
(GIS, Categorical, Regression, Survey)

GIS

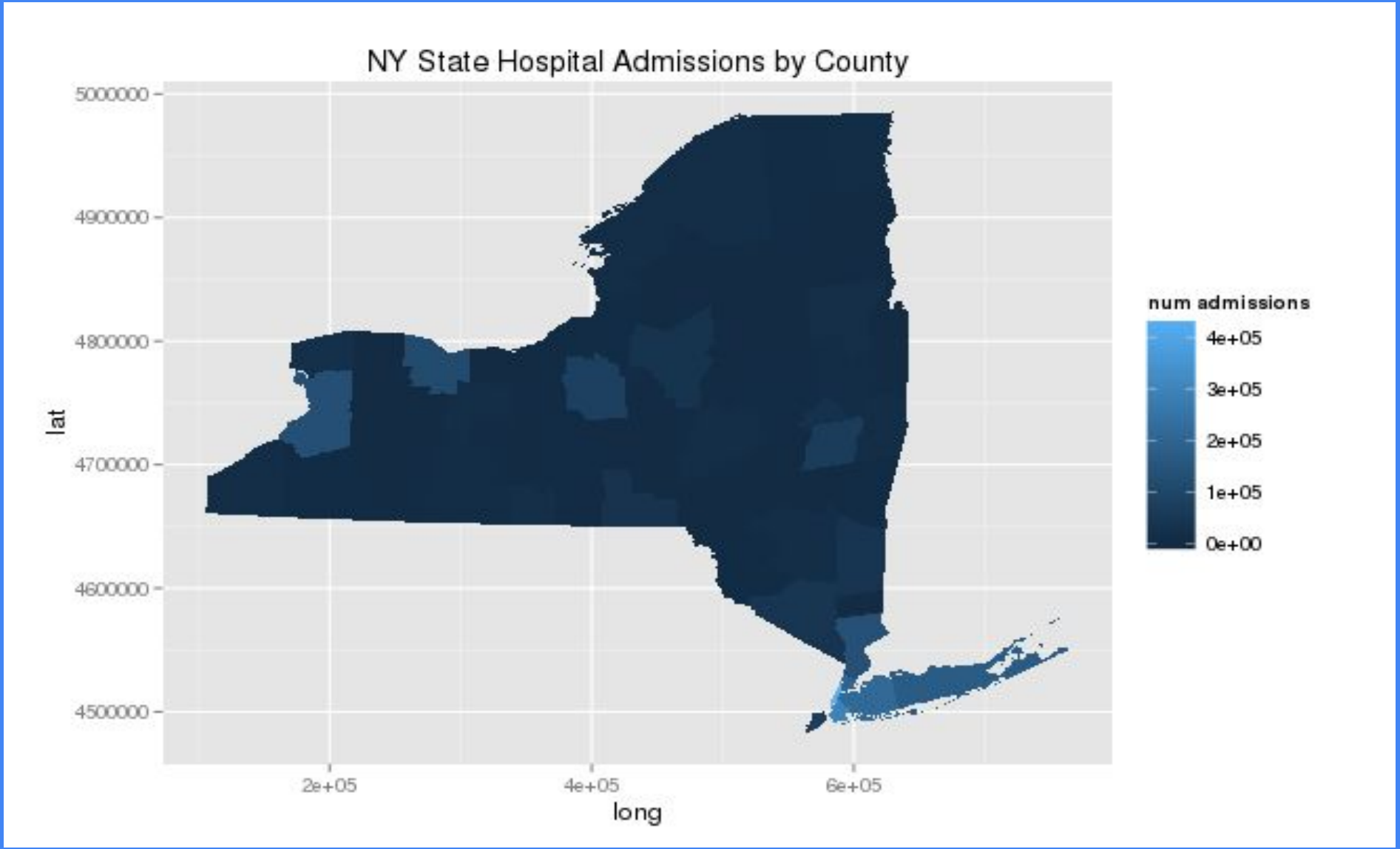
- Geographical Information System
- Helps display trends in data based on locations, for example based on states or counties
- We used the Sparcs 2011 dataset to create a GIS plot of NY state based on counties



NY State Population by County

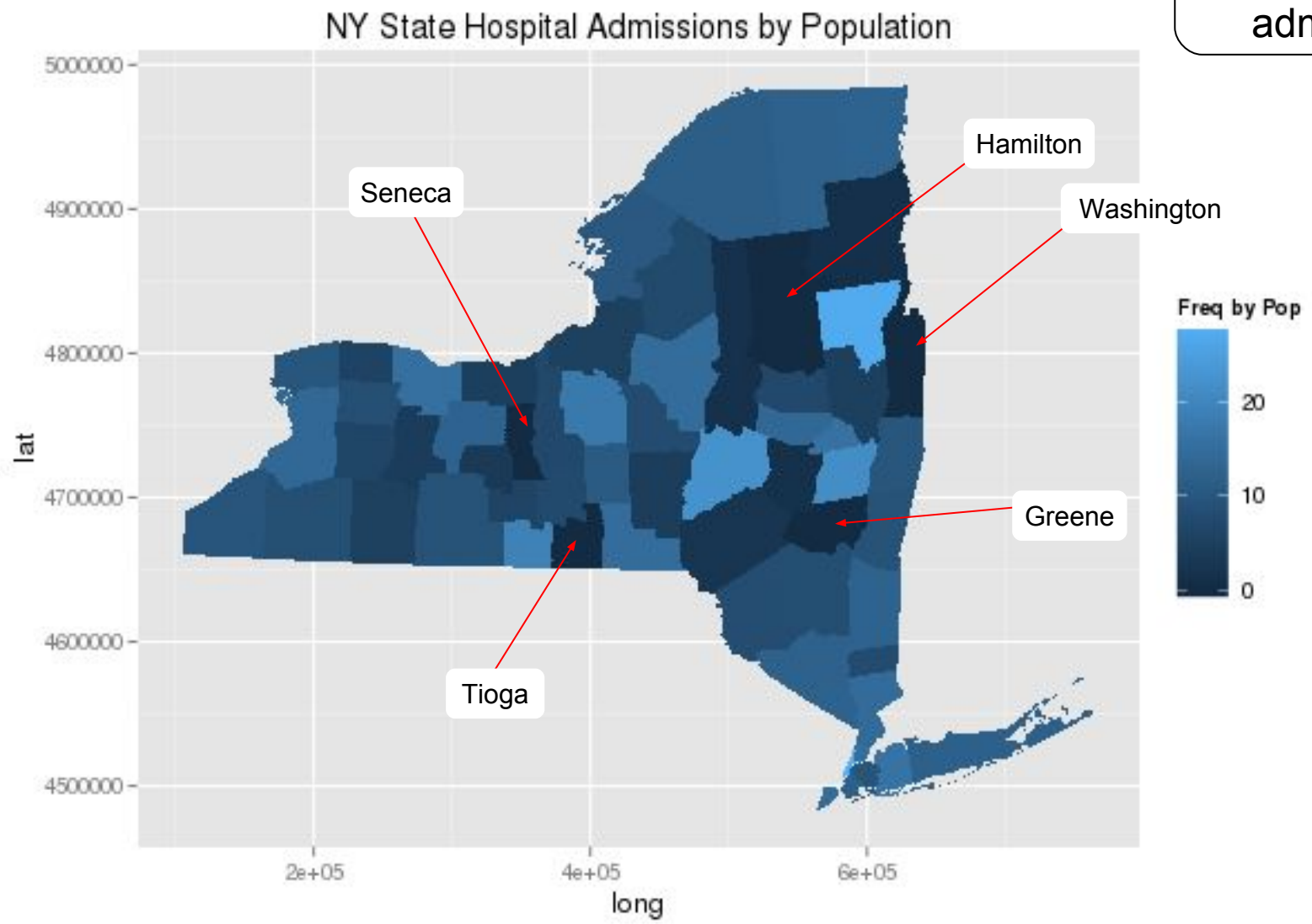


NY State Hospital Admissions by County



NY State Hospital Admissions by Population

These counties had zero hospital admissions



Some Code

gisIncludes.R | Final_Project.R* | t

Source on Save | Run

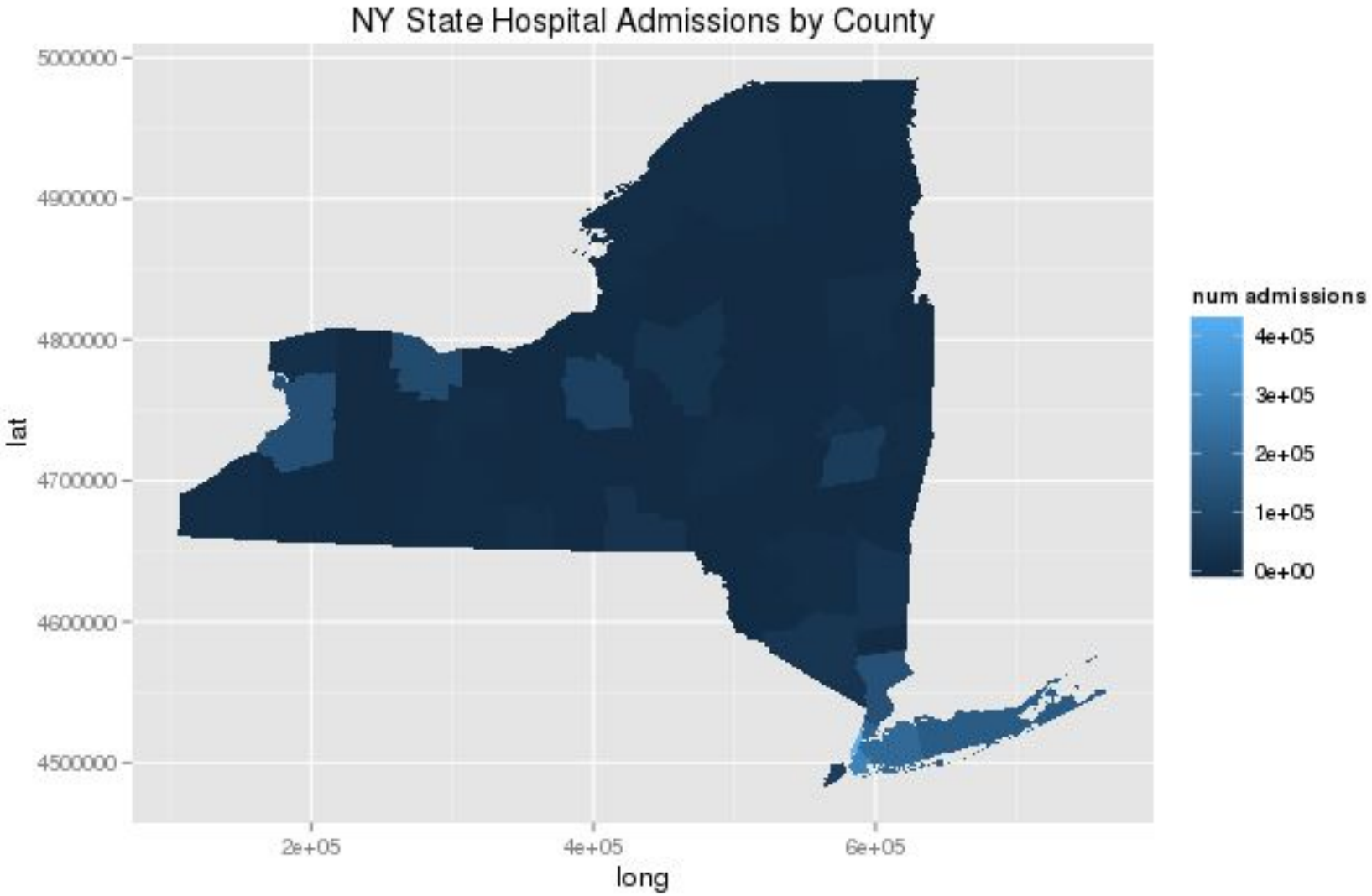
```
# and the names from the output of the table() operator.
t = data.frame(table(s$hospital_county))
b = data.frame(Var1 = c("Greene", "Washington", "Hamilton", "Seneca", "Tioga"), Freq = c(0,0,0,0,0))
tb = merge(x = t,
           y = b,
           by.d = "county",
           by.f = "facility",
           all = TRUE)
tb = na.omit(tb)
library(plyr)
tb = rename(tb, c("Var1" = "id"))

# Check the classes, check for NAs. Does it look like the right computation?


counties.h=merge(x=counties.m,
                 y=tb,
                 by.xx="county",
                 by.yy="facility",
                 all=TRUE)
# ------------------------------------------

# Do the data merge and plot the result.
P = as.numeric(counties.h$POP2000)
Fr = counties.h$Freq
vec = c(Fr/P)
vec = vec*100
# Merge the hospital admissions by county with the counties data frame (as we
# did above), and create the GIS plot from this data frame (also as we did above).
Map <- ggplot(counties.h, aes(x=long, y=lat, group=group, fill=Freq)) + geom_polygon() +
  coord_equal() + labs(x = "long", y = "lat", fill = "num admissions") +
  ggtitle("NY State Hospital Admissions by County")
Map
```

Result


NY State Hospital Admissions by County
5000000
4900000
4800000
4700000
4600000
4500000
lat
2e+05
4e+05
6e+05
long
num admissions
4e+05
3e+05
2e+05
1e+05
0e+00

Thank You For Your Attention!

Questions?