Creating Choropleth Maps in R

Author: Jason Chen

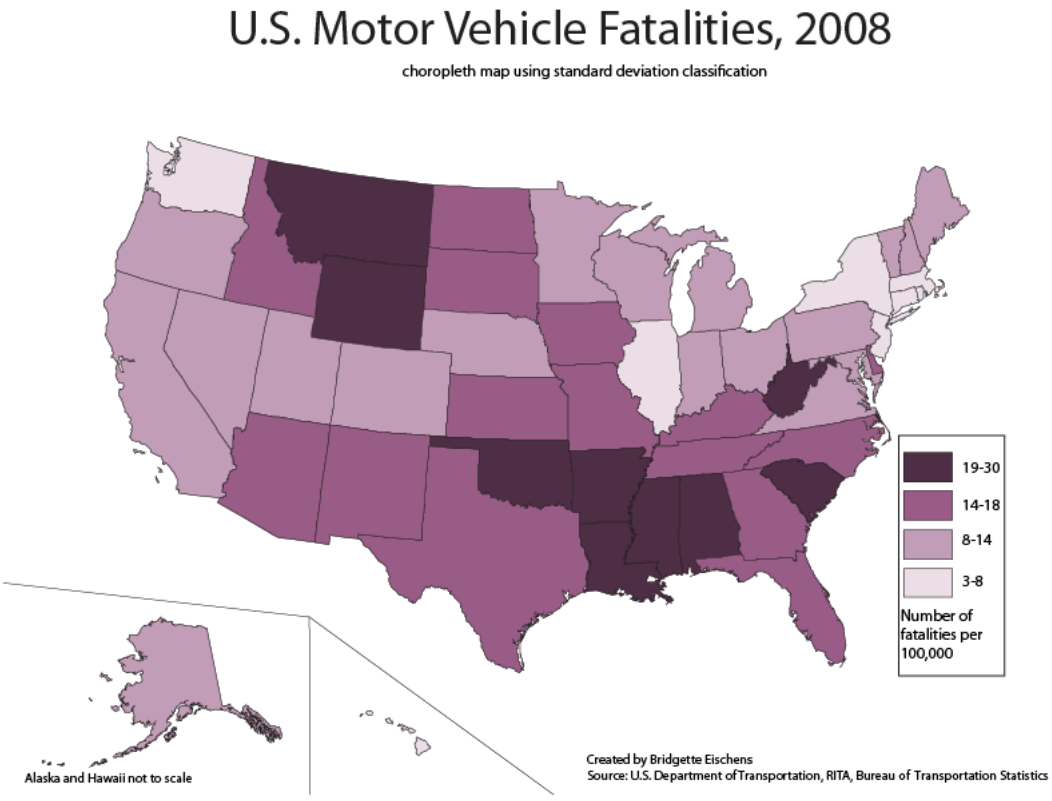
Theme: Data Visualization

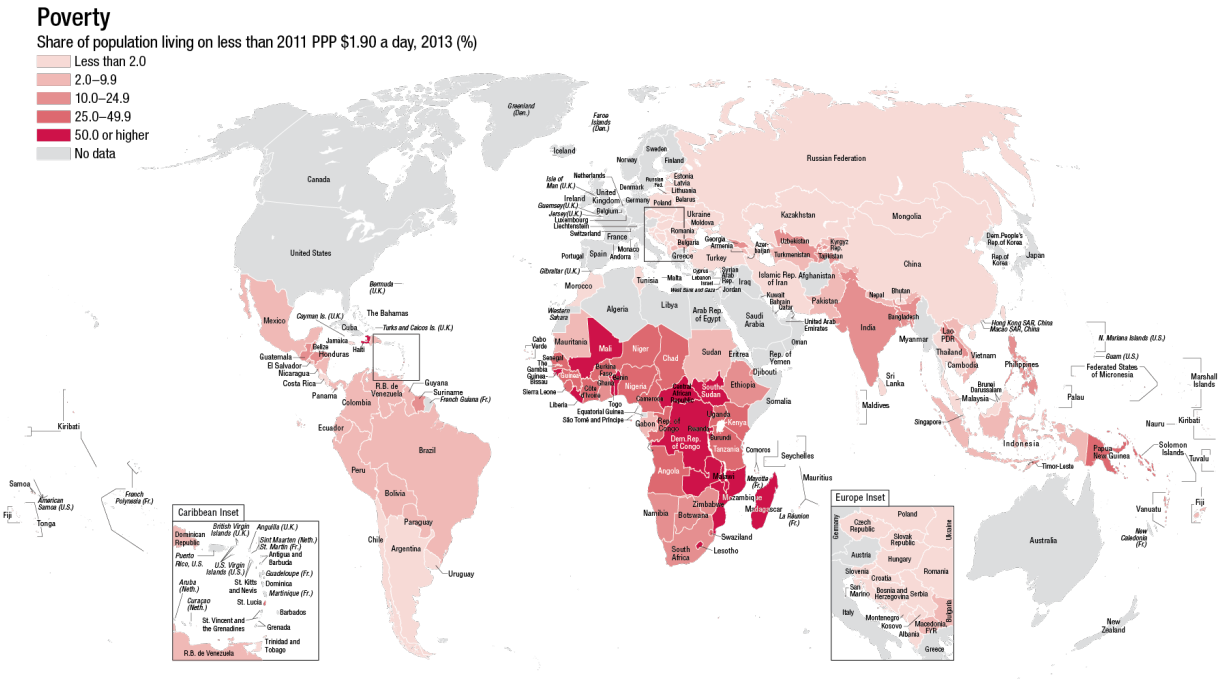
Introduction

So far in class we have learned a few ways to present data in R. We have learned how to display simple point plots,bar graphs, and various other simple graphs through both the ggplot package and native R methods. These types of graphs are very useful for displaying the relationship between numeric and categorical data. However, there are other forms of data that require more elaborate graphs to display correctly. Graphs with geographic data require maps to relate geographic location with some other form of data. Choropleth maps, in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map, are one form of map that incorporates geographic data. They are used to effectively convey geographically-specific data. Knowing how to create choropleth maps allows one to present data alongside its geographic context, something not possible with the graphing techniques we have so far learned in Stat133. Luckily, with the help of a couple packages, choropleth mapping is quite simple in R.

Examples

The following are two examples of choropleth maps. We can see that in each of them, the shading of each area represents the magnitude of the variable being plotted. Both of the maps elegantly convey this data, using shade and color instead of cluttering the map with unnecessary numbers and lines. The second map is from the World Bank, and shows that choropleth mapping is useful and powerful enough to be used by official policy-making institutions.





External Packages

To do choropleth mapping, we will use the maptools package to assist in reading the geographic data. maptools also requires that rgeos be installed. We will then need to use ggplot2 to plot the geography.

Choropleth Map Tutorial

1. We first find the geographic data and corresponding numeric data we wish to plot. Geographic data will come in the form of a shp file, which contains information about the geography in the form of geometric shapes as well as textual labels for each area. For this tutorial, I obtained the shp of Canadian census data from Statistics Canada. Other data must contain a column for geographic location. For this tutorial, I use data for the numbers of cats and dogs in Toronto, taken from Toronto city data.
2. In R, we first initialize the required packages.

```
library(maptools)
```

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

```
## Checking rgeos availability: TRUE
```

```
library(ggplot2)
library(rgeos)
```

```
## Warning: package 'rgeos' was built under R version 3.4.2
```

```
## rgeos version: 0.3-25, (SVN revision 555)
##  GEOS runtime version: 3.6.1-CAPI-1.10.1 r0
##  Linking to sp version: 1.2-5
##  Polygon checking: TRUE
```

3. Use maptools::readShapeSpatial to import the shp file as a SpatialPolygonsDataFrame. Because the fsa ids of Toronto all start with 'M', use subsetting to make a new SpatialPolygonsDataFrame for only toronto. Use ggplot2::fortify to convert toronto into a regular data frame, and rename the id column to fsa.

```
#Load shapefile
f = maptools::readShapeSpatial("gfsa000b11a_e/gfsa000b11a_e.shp")
```

```
## Warning: use rgdal::readOGR or sf::st_read

## Warning: use rgdal::readOGR or sf::st_read
```

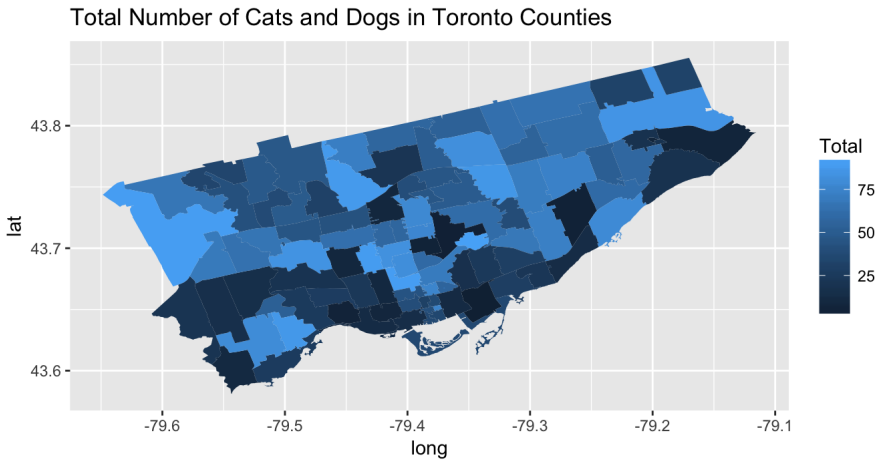
```
toronto = f[substr(f$CFSAUID, 1, 1) == 'M',]

#convert toronto into a regular data frame
data <- ggplot2::fortify(toronto, region = "CFSAUID")

# Create a fsa column instead of id in data.
data$fsa <- factor(data$id)
data$id <- NULL
```

4. Use read.csv to import pet data as a dataframe. Make sure each column of pets is type numeric rather than character. Use merge to merge data and pets by their fsa columns.

```
pets <- read.csv("pets.csv")
pets$DOG = as.numeric(pets$DOG)
pets$CAT = as.numeric(pets$CAT)
pets$Total = as.numeric(pets$Total)
merged = merge(data, pets, by = "fsa")
```

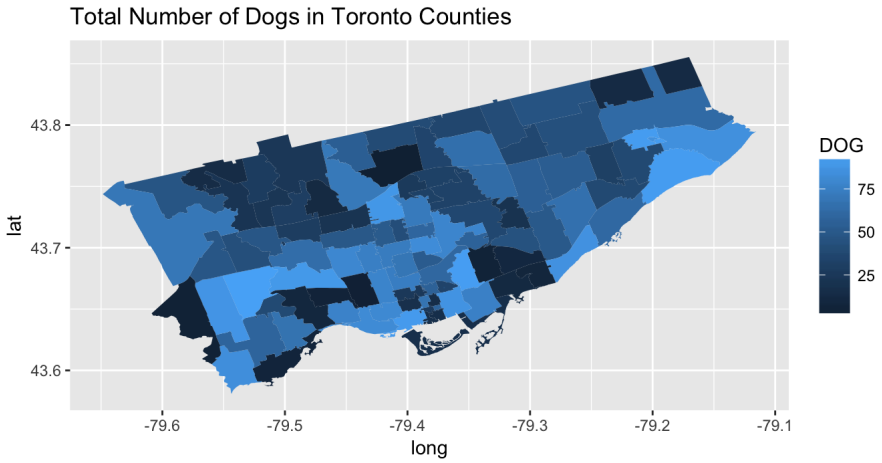
5. Use ggplot to plot the choropleth map. Set the dataframe as the merged one, set x and y to longitude and latitude. Fill by whichever variable you are plotting. Add to the ggplot initialization geom_polygon() and coord_equal() to finish the map. Add a title.

```
choro = ggplot(merged, aes(x = long, y = lat, group = group, fill = Total)) +
    geom_polygon() +
coord_equal() + ggtitle("Total Number of Cats and Dogs in Toronto Counties")
choro
```

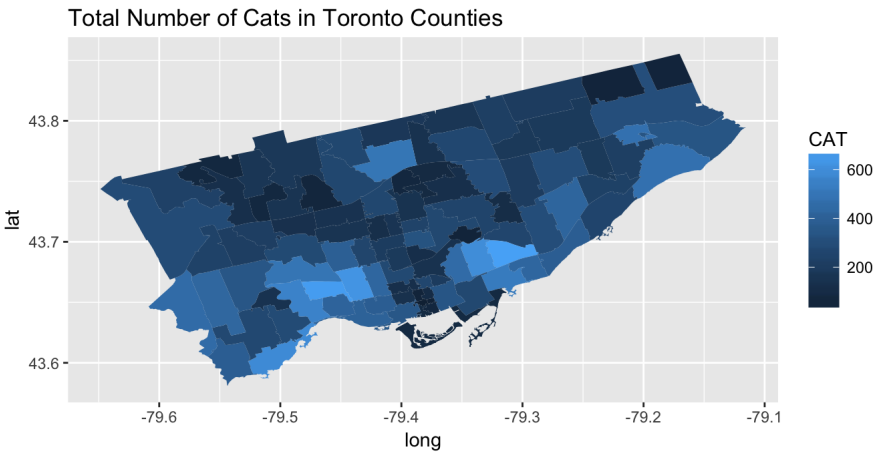


6. We can also easily make new maps for our other 2 variables.

```
#dogs
ggplot(merged, aes(x = long, y = lat, group = group, fill = DOG)) +
    geom_polygon() +
coord_equal() + ggtitle("Total Number of Dogs in Toronto Counties")
```

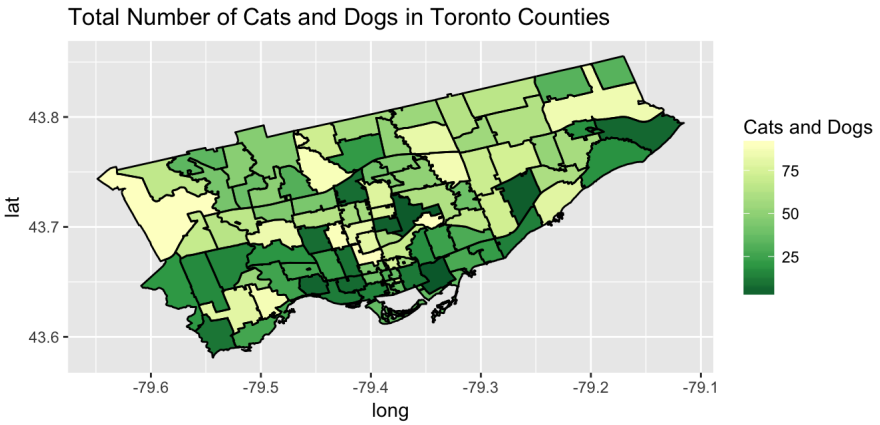



```
#cats
ggplot(merged, aes(x = long, y = lat, group = group, fill = CAT)) +
    geom_polygon() +
coord_equal() + ggtitle("Total Number of Cats in Toronto Counties")
```



7. We can now fiddle with the exact presentation of the map. For example, we can make the borders of each area better defined. We could also change the color.

```
choro = choro + geom_path(aes(x=long, y=lat, group=group), color='black') + scale_fill_distiller(name="Cats a
nd Dogs",    palette = "YlGn")
choro
```



Conclusion

Through this exercise, we can see that choropleth graphs are an effective way to visualize data alongside geography. The use of shading in these graphs allows for data to be elegantly displayed alongside their corresponding geographic location. The code itself is relatively easy to implement, and the resulting map is versatile and can be adjusted and modified. There are limits to this method of data visualization. The use of color to convey magnitude means that only one variable can be plotted on one map. Other limits of this particular method of generating choropleth maps include the necessity of shp files, which is an uncommon format and have to be specifically generated or found, and that both the shp file and the csv file must have a common column of geographic identifiers. *Nonetheless, choropleth mapping is a useful technique that is extremely handy for data visualization in a geographic context.*

References

R Bloggers - https://www.r-bloggers.com/shapefiles-in-r/

Wikipedia - Choropleth Map - https://en.wikipedia.org/wiki/Choropleth_map

PennState - Geog486 Data Visualization - https://www.e-education.psu.edu/geog486/node/1864

Statistics Canada - Census 2011 - http://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2011-eng.cfm

Toronto Licensed Cats and Dogs - https://www1.toronto.ca/wps/portal/contentonly?vgnextoid=0a7e1f46f71fb310VgnVCM10000071d60f89RCRD&vgnextchannel=75d6e03bb8d1e310VgnVCM10000071d60f89RCRD

R Tutorials for Choropleth Mapping - http://bl.ocks.org/prabhasp/raw/5030005/

- http://www.milanor.net/blog/maps-in-r-choropleth-maps/

Image Sources: University of Minnesota Duluth - http://www.d.umn.edu/~eisch032/ChoroplethLab5.png

World Bank Databank - http://databank.worldbank.org/data/download/site-content/wdi/maps/2017/maps-wdi-2017-sec-1-poverty.png