Educational Attainment in Honduras

Max H. McDonald

2019-03-18

This file contains background mapping and graphing work for our final project for Public Policy

599 Winter 2019

```
## Clear the workspace
rm(list = ls())

# Set working directory and bring in the
# data and libraries

library(MASS)
library(maps)
library(RColorBrewer)
library(maptools)
library(sp)
library(spdep)
library(ggplot2)
library(rgdal)
library(foreign)
library(haven)
library(expss)
library(histogram)
library(raster)
library(sf)


# We will use three data frames for this
# descriptive analysis: a csv file
# containing our variables of interest
# and two shapefiles for representing our
# descriptive variables graphically

hn_data <- read.csv("https://github.com/arsell/599-Project/blob/master/data/cleandata.csv?raw=true")
hn_data <- na.omit(hn_data)
# hn_spatial <- read_sf(dsn
# ='https://github.com/arsell/599-Project/blob/master/maps',
# layer = 'HNGE61FL.shp')
hn_spatial <- readOGR(dsn = "C:/users/arsell/documents/github/599-Project/maps/HNGE61FL.shp")
## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\arsell\documents\GitHub\599-Project\maps\HNGE61FL.shp", layer: "HNGE61FL"
## with 1148 features
## It has 20 fields
# SpatialPointsDataFrame

# These data are from the GADM country
```

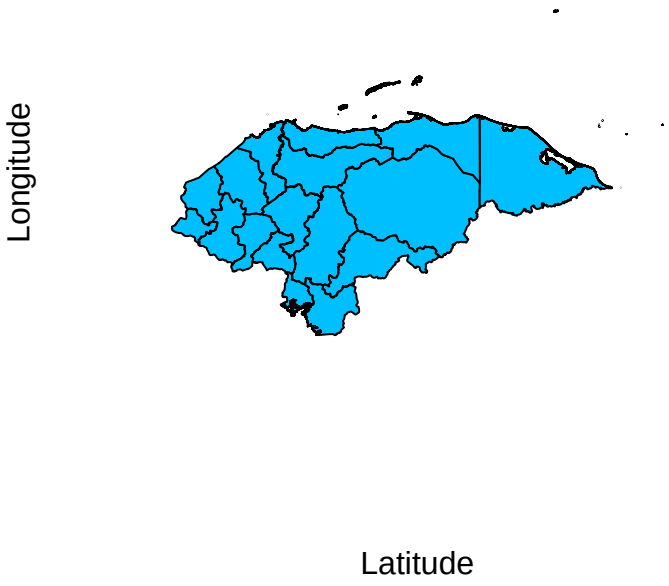


Figure 1: Map of SDRs for Honduras 2011

```
# codes, a set of codes for most
# countries which provides spatial
# polygon data. The data can be found at:
# https://gadm.org/download_country_v3.html
hn_map1 <- readOGR("C:/users/arsell/documents/github/599-project/maps/gadm36_HND_1.shp")  #SpatialPolyg
## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\arsell\documents\GitHub\599-project\maps\gadm36_HND_1.shp", layer: "gadm36_HND_1"
## with 18 features
## It has 10 fields
```

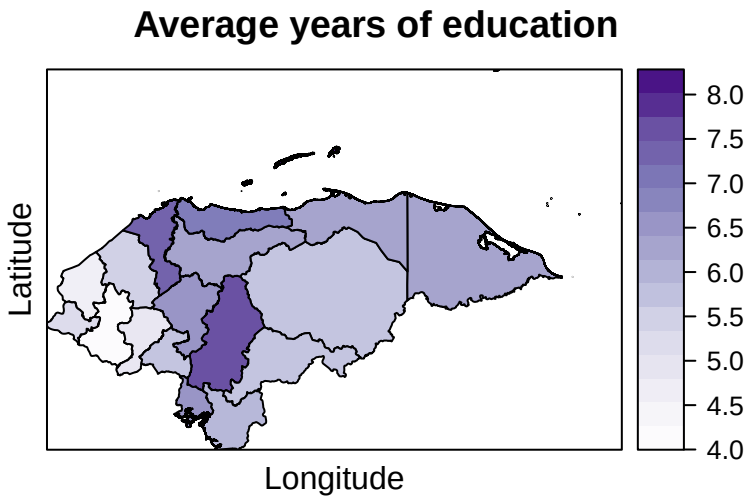
```
# Honduras has 18 administrative departments, outlined here
plot(hn_map1, ylab = "Longitude", xlab = "Latitude", col = "deepskyblue")
```

I would like to append a couple of vectors to the honduras_map shape file

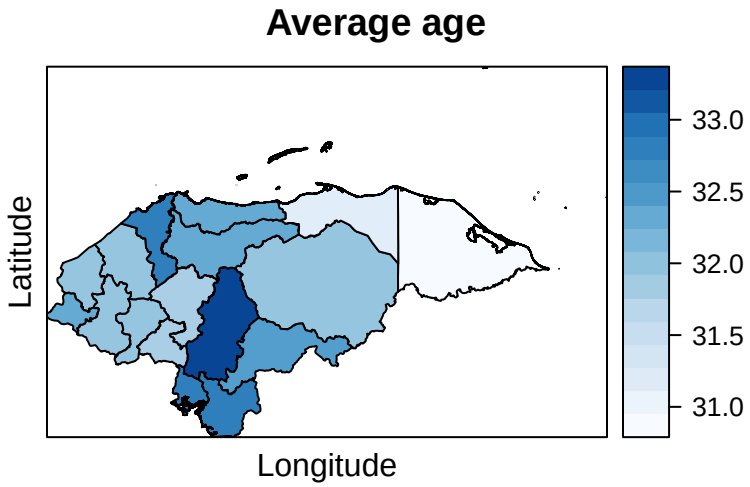
```
## Because our covariate data are at a
## different level of observation than the
## shapefile data we will aggregate to
## average data by region. Here we are
## interested in age, education, and
## wealth
avg_wealthDec <- aggregate(hn_data$wealthDec,
```

```
    list(hn_data$region), mean)
avg_age <- aggregate(hn_data$age, list(hn_data$region),
    mean)
avg_educ <- aggregate(hn_data$educ, list(hn_data$region),
    mean)


## Let's add each of these vectors to tht
## Spatial Poly data frame
hn_map1@data$avg_wealthDec <- avg_wealthDec[,
    2]
hn_map1@data$avg_age <- avg_age[, 2]
hn_map1@data$avg_educ <- avg_educ[, 2]

## Data are missing for average education
## for the region Francisco Morazan Have
## to figure out what's going on
## View(hn_map1@data)
```

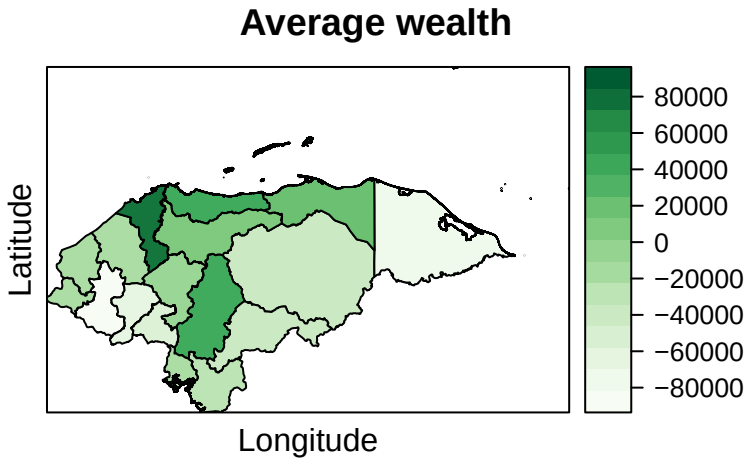
```
# Display education averages by region

spplot(hn_map1, "avg_educ", col.regions=colorRampPalette(brewer.pal(8, "Purples"))(50),
       main="Average years of education",
       xlab="Longitude", ylab="Latitude")
```




```
# Display average ages by region
spplot(hn_map1, "avg_age", col.regions=colorRampPalette(brewer.pal(8, "Blues"))(50),
       main="Average age",
       xlab="Longitude", ylab="Latitude")
```




```
# Display average wealth by region

spplot(hn_map1, "avg_wealthDec", col.regions=colorRampPalette(brewer.pal(8, "Greens"))(50),
       main="Average wealth",
       xlab="Longitude", ylab="Latitude")
```

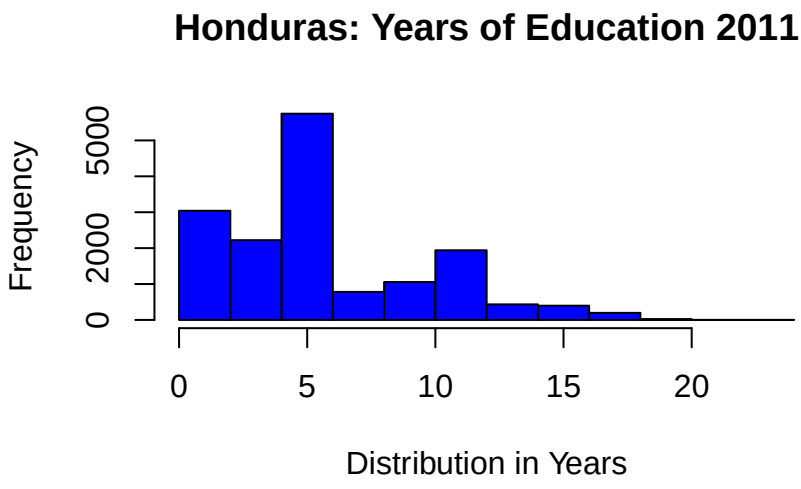



```
## Let's send the plots to a pdf in the working directory, titled "avgs_by_region"

pdf("avgs_by_region1.pdf", height=6, width=6)
```

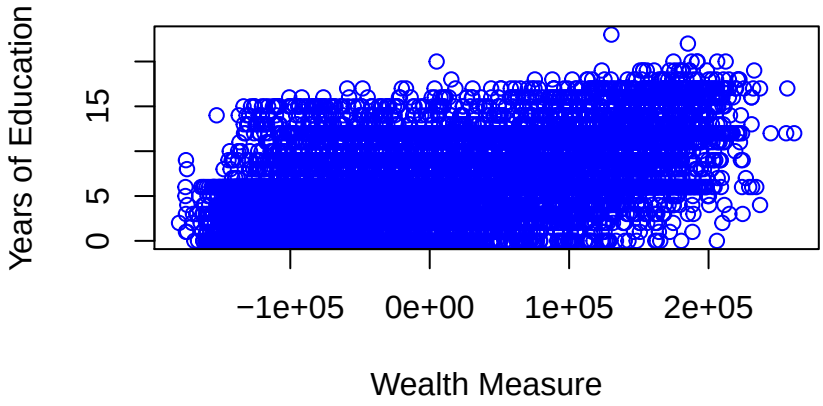
```
spplot(hn_map1, "avg_educ", col.regions=colorRampPalette(brewer.pal(8, "Purples"))(50), main="Average ed
       xlab="Longitude", ylab="Latitude")

dev.off()
## pdf
##   2

pdf("avgs_by_region2.pdf", height=6, width=6)

spplot(hn_map1, "avg_age", col.regions=colorRampPalette(brewer.pal(8, "Blues"))(50),
       main="Average age",
       xlab="Longitude", ylab="Latitude")

dev.off()
## pdf
##   2

pdf("avgs_by_region3.pdf", height=6, width=6)

spplot(hn_map1, "avg_wealthDec", col.regions=colorRampPalette(brewer.pal(8, "Greens"))(50),
       main="Average wealth",
       xlab="Longitude", ylab="Latitude")



dev.off()
## pdf
##   2
```

```
## Tables and Summary Statistics
## This is more for internal analysis rather than for displaying to an audience
cro(hn_data$region, hn_data$educ)
cro(hn_data$region, hn_data$religion)

## Summaries of key variables
summary(hn_data)
summary(hn_data$region)
summary(hn_data$wealthCat)
summary(hn_data$educ)
```

```
hist(hn_data$educ,main="Honduras: Years of Education 2011",xlab ="Distribution in Years", col = "blue")
```

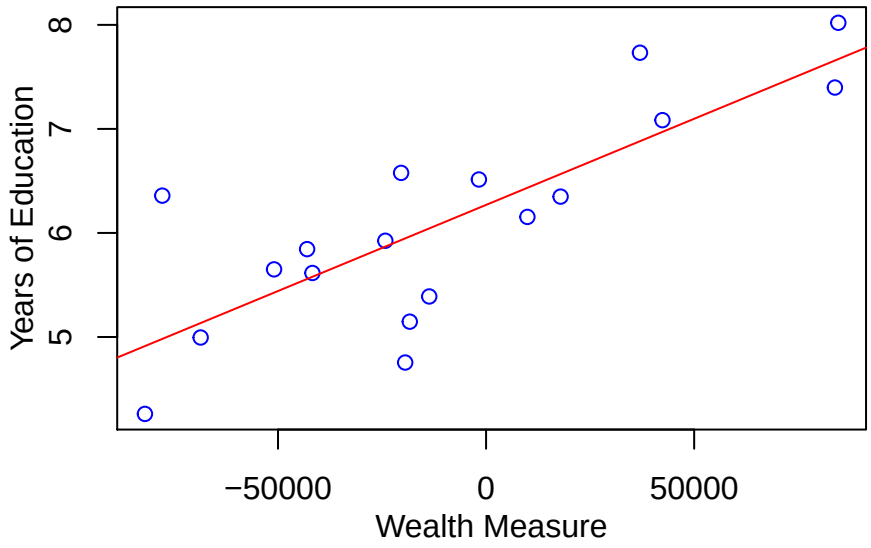



```
## This plot uses the full data of approximately 15,800 observations. It is difficult to
## get a sense of the possible relationship. We'll look at another plot aggregated to the
## department level
plot(hn_data$educ ~ hn_data$wealthDec , ylab="Years of Education",xlab="Wealth Measure",col="blue")
```

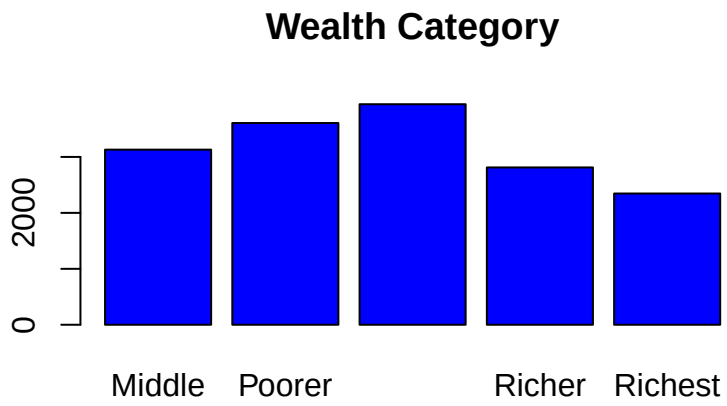



```
## Let's display the relationship between avg years of education and avg wealth

par(mgp=c(2,1,0), mar=c(3,3,1,1))

# Fit regression line
reg <- lm(avg_educ ~ avg_wealthDec, data = hn_map1)
coeff=coefficients(reg)

# equation of the line :
equation = paste0("y = ", round(coeff[2],1), "*x ", round(coeff[1],1))
```

```r
# plot
plot(hn_map1$avg_educ ~ hn_map1$avg_wealthDec , ylab="Years of Education",xlab="Wealth Measure",col="blu
abline(reg, col="red")
```

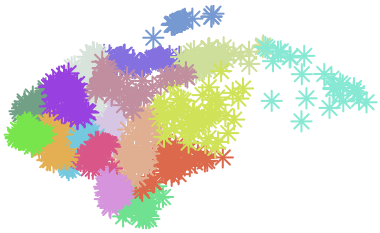



```r
#plot(hn_data$educ ~ hn_data$wealthDec , ylab="Years of Education",xlab="Wealth Measure",col="blue")

barplot(table(hn_data$wealthCat), main = "Wealth Category", col="blue")
```




```r
## This chunk contains exploratory code that I'll work on over the week coming after
## discussing with the team.


## Subsets

hn_urban <- hn_data[which(hn_data$urban==1), ]
hn_rural <- hn_data[which(hn_data$urban==0), ]
```

```r
hn_dense <- hn_data[which(hn_data$region == 'Cortes' |hn_data$region == 'Francisco Morazan'), ]
summary(hn_dense$age)

library(randomcoloR)
library(reshape)
##
## Attaching package: 'reshape'
## The following object is masked from 'package:Matrix':
##
##     expand

## Merging the two data frames by the unique cluster value

# region boundaries
hn_region <- hn_data[, c('region', 'DHSCLUST')]
hn_region <- unique(hn_region)

# wealth categories
hn_wc <- hn_data[, c('DHSCLUST', 'wealthCat')]
hn_wc <- unique(hn_wc)


DFW <- merge(x=hn_spatial, y=hn_region, by.x = "DHSCLUST", by.y = "DHSCLUST")
DFW

temp_which <-which(DFW$LATNUM != 0 & DFW$LONGNUM != 0)
DFW <- DFW[temp_which, ]

n_regions <- unique(as.character(DFW@data$region))
n_regions

n_colors <- distinctColorPalette(k=length(n_regions))

plot(DFW,
     col=(n_colors)[DFW$region],
     pch=8)
```



```
View(DFW@data)


## Want to aggregate honduras_data by cluster

agg_hn <- aggregate(list(category_count = hn_data$wealthCat), by = list(cluster = hn_data$DHSCLUST, weal

agg_reshape <- reshape(data = agg_hn, direction = 'wide', idvar = 'cluster', timevar = 'wealthcat')

new_hn <- merge(x=DFW, y=agg_reshape, by.x = "DHSCLUST", by.y = "cluster")

plot(new_hn,
     col=(n_colors)[DFW$wealthcat],
     pch=8)
```