

# Intro to R (Honors Seminar 2020)

Elena Barham

9/30/2020

## 1 Load Packages

```
### General use packages
# tidyverse # This package is actually a suite of packages
# raster    # Reading, writing, and manipulating raster form spatial data
# rgdal     # Provides access to spatial projection operations
# GISTools  # Choropleth plots and aggregation of point data to polygons
# rgeos     # Operations with spatial data (i.e. intersection, length of spatial forms)
# maptools   # More tools for manipulating geographic data
# ggplot2   # for graphing

# A function to see if we need to load
check.packages <- function(pkg){
  new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]
  if (length(new.pkg))
    install.packages(new.pkg, dependencies = TRUE,
                      repos = "https://cran.rstudio.com")
  sapply(pkg, require, character.only = TRUE)
}

pkgList <-
  c("raster", "rgdal", "GISTools", "rgeos",
    "maptools", "ggplot2", "haven", "stargazer", "sp", "knitr")

check.packages(pkgList)

##      raster      rgdal    GISTools      rgeos    maptools    ggplot2      haven stargazer
##      TRUE        TRUE       TRUE        TRUE      TRUE       TRUE      TRUE      TRUE
##      sp          knitr
##      TRUE        TRUE
```

Note that you first need to install packages and then save them to your library. Once you have installed them once you then just need to add them to the library for new sessions.

### 1.1 Learning Some Basic Commands

```
x <- 3 #this assigns an object
x == 3 #this is a logical question

## [1] TRUE
```

```

y <- "ABC" #this assigns a character string

x <- c(3,4,5,6) #this is a numeric vector
m <- c("panda", "octopus", "squirrel", "moose") #this is a character vector

panda <- c(x,m) # this combines x and m into one vector
panda <- rbind(x,m) # this binds x and m into a matrix by rows
panda <- cbind(x,m) # this binds x and m into a matrix by columns

panda <- as.data.frame(panda) # this makes y into a data frame

a <- panda[,1] # this subsets y into the first column
b <- panda[1,] # this subsets y into the first row

```

## 1.2 Loading and merging data - PANEL

```

# Load panel data
mexico <- read.csv("mexico.csv")

# Make sure our data is there
#View(mexico) # The 'View' function brings up the panel data

# Load a stata file
elections <- read_dta("mex_elections.dta")
# Citing this data: Blackman, Allen; Villalobos, Laura, 2020, "Replication data for Use Forests or Lose

#### Merging ####
# Merge two data frames - create a unique merge ID

#### Basic Cleaning Techniques ####
mexico$X <- NULL # Delete extraneous columns
mexico[is.na(mexico$GIS_AREA)] <- 0 # This sets NA observations of the variable attacks to '0'
mexico <- mexico[!is.na(mexico$GOV_TYPE),] # This drops NA observations of the specified variable

## Merging
merged <- merge(mexico, elections, by = "estado", all.x = TRUE)

```

## 1.3 Loading and merging data – Spatial

```

# Load spatial data
mex_shp <- readOGR(dsn = "mex_shp",
                     layer = "MEX_adm2")

## OGR data source with driver: ESRI Shapefile
## Source: "C:\Dropbox\Teaching\Honors\2020-21\Honors_2021\1 Resources and Readings\Materials - Monday 1
## with 1853 features
## It has 11 fields
## Integer64 fields read as strings: ID_0 ID_1 ID_2
mex_pa_shp <- readOGR(dsn = "mex_PAs_shp",
                        layer = "areas")

```

```

## OGR data source with driver: ESRI Shapefile
## Source: "C:\Dropbox\Teaching\Honors\2020-21\Honors_2021\1 Resources and Readings\Materials - Monday 16th October 2020\mex_pa.shp"
## with 1196 features
## It has 36 fields

# Subsetting Data
federal_PAs <- subset(mex_pa_shp, gv_cl == "federal")

# Make sure our data is there
plot(mex_shp) # The 'plot' function lets you look at version of the shape file with no attributes

```



```

head(mex_shp, 5) # The 'head' function lets you quickly look at the variables in a shape file

##   ID_0 ISO NAME_0 ID_1           NAME_1 ID_2           NAME_2    TYPE_2
## 0 145 MEX Mexico      1 Aguascalientes      1 Aguascalientes Municipio
## 1 145 MEX Mexico      1 Aguascalientes      2 Asientos Municipio
## 2 145 MEX Mexico      1 Aguascalientes      3 Calvillo Municipio
## 3 145 MEX Mexico      1 Aguascalientes      4 CosÃ±o Municipio
## 4 145 MEX Mexico      1 Aguascalientes      5 JesÃ±s MarÃ±a Municipio
##          ENGTYPE_2 NL_NAME_2  VARNAME_2
## 0 Municipality        <NA>       <NA>
## 1 Municipality        <NA>       <NA>
## 2 Municipality        <NA>       <NA>
## 3 Municipality        <NA>       Cosio
## 4 Municipality        <NA>       Jesus Maria

```

## 1.4 Data Exploration

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00146	0.19869	2.38153	19.09800	18.14911	199.52657

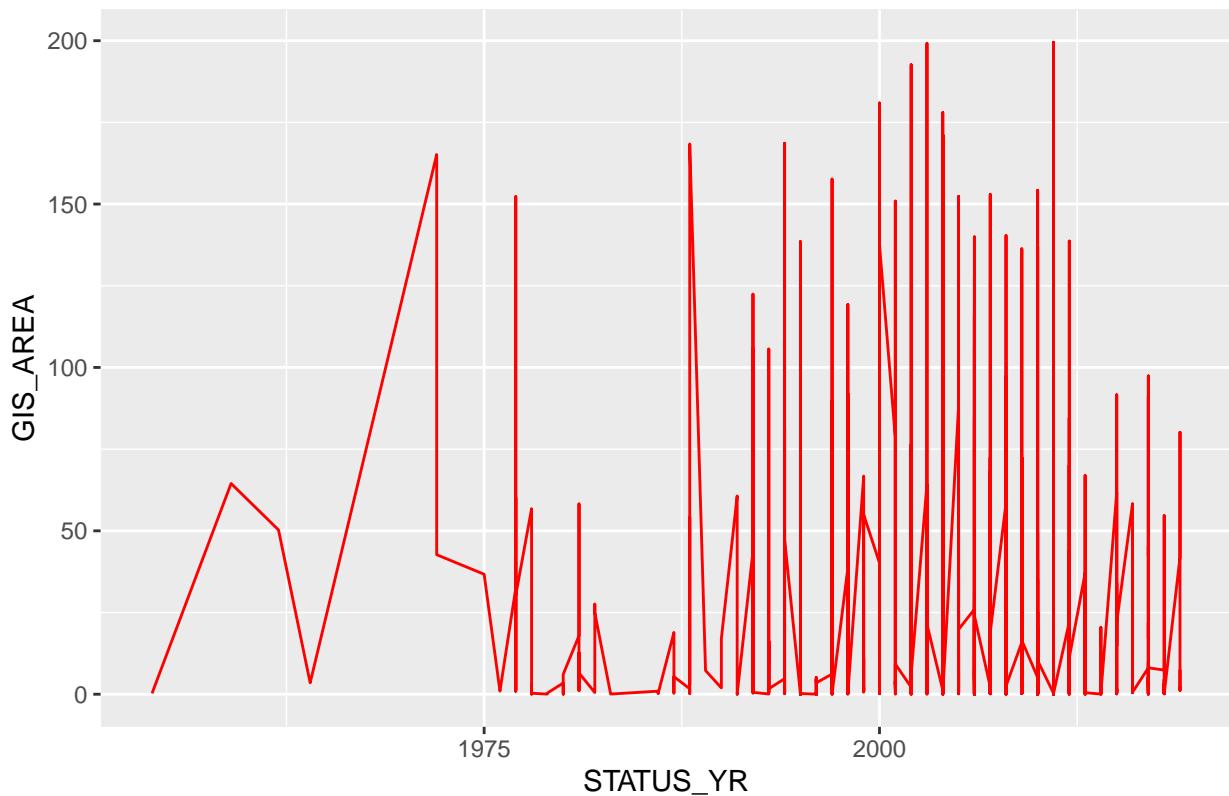
collaborative	federal	local	nonprofit	private
9	110	10	5	336
subnational				
390				

	collaborative	federal	local	nonprofit	private	subnational
1954	0	1	0	0	0	0
1959	0	1	0	0	0	0
1962	0	1	0	0	0	0
1964	0	1	0	0	0	0
1972	0	0	0	0	0	2
1975	0	0	0	0	0	1
1976	0	0	0	0	0	1
1977	0	0	0	0	0	7
1978	0	0	0	0	0	5
1979	0	0	0	0	0	2
1980	0	1	0	0	0	4
1981	0	4	0	0	0	2
1982	0	2	0	0	0	1
1983	0	0	0	0	0	1
1986	0	0	0	0	0	2
1987	0	1	0	0	0	2
1988	0	3	0	0	0	3
1989	0	0	0	0	0	1
1990	0	0	0	0	0	2
1991	0	1	0	0	0	3
1992	0	3	0	0	0	5
1993	0	0	0	0	0	8
1994	0	4	0	0	0	8
1995	0	1	0	0	0	37
1996	0	1	0	0	0	45
1997	0	2	0	0	0	11
1998	0	5	0	0	0	8
1999	0	1	0	0	0	9
2000	0	15	0	0	0	28
2001	0	2	0	0	0	8
2002	0	20	0	0	1	11
2003	0	3	0	0	1	15
2004	2	3	0	2	5	29
2005	1	5	1	1	5	20
2006	0	1	0	1	43	25
2007	0	2	0	0	15	19
2008	4	8	7	1	9	16
2009	0	5	1	0	14	14
2010	0	4	0	0	53	13
2011	1	1	0	0	58	6
2012	0	4	0	0	35	3
2013	1	2	1	0	19	4
2014	0	0	0	0	18	3

	collaborative	federal	local	nonprofit	private	subnational
2015	0	1	0	0	9	2
2016	0	0	0	0	8	4
2017	0	1	0	0	26	0
2018	0	0	0	0	11	0
2019	0	0	0	0	6	0

## 1.5 Data Visualization - Graphs!

Mexico: Annual Land Conserved



## 1.6 Data Visualization - Spatial

```
mex_shp <- spTransform(mex_shp, CRS("+proj=longlat +datum=WGS84 +no_defs")) # Project
mex_pa_shp <- spTransform(mex_pa_shp, CRS("+proj=longlat +datum=WGS84 +no_defs")) # Project

#Make a map!
plot(mex_shp, col='light gray', border='gray')
plot(subset(mex_pa_shp, gv_cl == "private"), add=TRUE, col="purple", border = "purple")
```



## 1.7 Regression and Displaying Results

```
# Basic Regression - these are nonsense specifications but
# will help you understand the form of how to specify a regression in r
x_1 <- lm(federal ~ participation, data = merged)
# summary(x_1)

# Fixed Effects
x_2 <- lm(federal ~ participation + as.factor(STATUS_YR), data = merged)
# summary(x_2)

# Interaction Effects
x_3 <- lm(federal ~ participation*povertyt1 + as.factor(STATUS_YR), data = merged)

stargazer(x_1, x_2, x_3, omit = "STATUS_YR",
           add.lines = list(c("Fixed effects?", "No", "Yes", "Yes"))),
           header=FALSE) # This produces a table that can be typeset in LaTeX.
```

Table 2:

	<i>Dependent variable:</i>		
	federal		
	(1)	(2)	(3)
participation	-0.003* (0.002)	-0.0003 (0.002)	-0.018*** (0.006)
povertyt1			-0.033*** (0.008)
participation:povertyt1			0.0005*** (0.0001)
Constant	0.281*** (0.093)	1.016*** (0.290)	2.228*** (0.418)
Fixed effects?	No	Yes	Yes
Observations	746	746	746
R <sup>2</sup>	0.004	0.337	0.370
Adjusted R <sup>2</sup>	0.003	0.293	0.325
Residual Std. Error	0.324 (df = 744)	0.273 (df = 698)	0.266 (df = 696)
F Statistic	3.072* (df = 1; 744)	7.555*** (df = 47; 698)	8.335*** (df = 49; 696)

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01