

## Notas de clase – Taller de Stata<sup>1</sup>

### Clase 12 – Gráficos III - Mapas

#### Contenido

1. Nociones de información geográfica
  - 1.1. Coordenadas
  - 1.2. Tipos de información geográfica
  - 1.3. Shapefiles
2. Dibujo de mapas en Stata
  - 2.1. Instalación y comandos
  - 2.2. Comando `spshape2dta` (`shp2dta`)
  - 2.3. Comando `grmap` (`spmap`)
  - 2.4. Opciones generales

1

---

#### Anexo: Programas UWS

- A. Introducción
- B. Búsqueda de comandos (paquetes)
- C. Instalación y actualización de comandos

[Estas notas se diseñaron con base en la versión 16 de *Stata*, conforme el software cambie las notas deben ser actualizadas]

---

<sup>1</sup> Estas notas están basadas en la guía desarrollada previamente para este curso por Rodrigo Azuero Melo, Nicolás de Roux, Luis Roberto Martínez, Román Andrés Zárate y Santiago Gómez Echeverry.

## 1. Nociones de información geográfica

Para utilizar comandos de mapas en *Stata* es necesario introducir los conceptos básicos más importantes sobre información geográfica. Esto es, entender qué es un dato georreferenciado y cómo usarlo, qué son las coordenadas del mapa y las unidades de observación que se quieren representar.

### 1.1. Coordenadas

Una coordenada es un dato que representa, a través de números, la posición geográfica de un lugar en el planeta. Normalmente una coordenada es un punto representado por dos números, uno que representa la latitud (positivo si es norte, negativo si es sur) y otro que representa la longitud (positivo si es oriente, negativo si es occidente). Estas coordenadas se calculan bajo sistemas de georreferenciación que aproximan la curvatura de la tierra para representar posiciones. Las coordenadas de los mapas tienen asociado un sistema de georreferenciación que se puede consultar en los metadatos. Este sistema se puede cambiar a través de un sistema de información geográfica como *ArcGIS* (de ESRI) o *QGIS* (Software libre). Entonces, construir un mapa a través de coordenadas se conoce como proyectar un mapa.

---

2

### 1.2. Tipos de información geográfica

Hay dos tipos de información geográfica, vectorial y ráster. La información *vectorial* se usa en los mapas de objetos que pueden ser mostrados geoméricamente. En ese sentido, estos mapas pueden contener a su vez tres tipos de geometría, puntos, líneas y polígonos. Los puntos pueden representar localizaciones, las líneas ríos o vías, y los polígonos pueden representar áreas como zonas administrativas o zonas geográficas. La información *ráster* contiene información geográfica a nivel de píxeles (un píxel representa un área cuadrada estándar), y es ampliamente utilizado para almacenar información ambiental como deforestación o intensidad de luz en la noche. Por simplicidad y su uso extendido, en estas notas solo vamos a referirnos a información geográfica vectorial de puntos y polígonos. En los ejemplos de la documentación y el material suplementario se puede encontrar el uso de otros tipos de datos geográficos en *Stata*.

Los mapas pueden mostrar información a diferentes niveles de unidades de observación, e.g., un mapa de municipios o departamentos contiene información de tipo vectorial de polígonos.

Por lo tanto, la información geográfica (coordenadas) está asociada con atributos de la unidad de observación, que son los datos de los departamentos o municipios en cuestión.

### 1.3. Shapefiles

La mayoría de los usuarios de información geográfica vectorial utilizan archivos con formato .shp (shapefile). Estos archivos son desarrollados por el Environmental Systems Research Institute (ESRI) y tienen tres componentes principales:

- Archivo shp: archivo que almacena las coordenadas de las entidades geométricas (unidades de observación). Como vimos, estas entidades pueden ser departamentos, municipios, países, ciudades, etc. Las coordenadas son los datos que representan cada uno de los puntos que permiten dibujar estas entidades geométricas. Entonces, si la entidad es un punto solo tiene asociada una coordenada. Si la entidad es un polígono puede tener asociada un gran número de coordenadas.
- Archivo dbf: contiene información de características (atributos) no geográficos de las entidades. En los sistemas de información geográfica, estas características se manejan a través de la tabla de atributos, que es una colección de datos de cada una de las entidades que está siendo proyectadas. La tabla de atributos tiene la misma estructura de información de una base de datos de *Stata*. En la siguiente sección veremos cómo importarla y trabajarla con lo que hemos aprendido hasta ahora en el curso.
- Archivo shx: archivo que almacena el índice o indicador de las entidades geométricas. Este índice permite relacionar los datos geográficos (archivo .shp) con los atributos de las entidades proyectadas (archivo .dbf)

Los archivos tipo *shapefile* son de uso generalizado y por lo tanto son más fáciles de encontrar. Normalmente las entidades gubernamentales, universidades o centros de investigación alojan este tipo de datos en sus repositorios o geoportales<sup>2</sup>.

## 2. Dibujo de mapas en *Stata*

A continuación, describimos en detalle los pasos para generar un mapa.

---

<sup>2</sup> Como el geoportal del DANE: <https://geoportal.dane.gov.co/>  
Taller de Stata  
Clase 12 – Gráficos III - Mapas  
Miguel Garzón Ramírez, Cristhian Acosta Pardo  
Facultad de Economía, Universidad de los Andes

## 2.1. Comando **spshape2dta (shp2dta)**

Primero debemos convertir los datos espaciales a formato de *Stata*, para esto usamos el comando *spshape2dta*. En clase, el *shapefile* que manejamos está en un archivo comprimido llamado “co\_dep.zip”. Al interior de este archivo encontramos tres archivos: co\_dep.shp (coordenadas), co\_dep.shx (indicador) co\_dep.dbf (atributos del departamento). Es necesario asegurarse de descomprimir el archivo .zip. También, debe asegurarse de que los archivos estén en el directorio de trabajo o especificar en el comando el directorio donde se encuentran. Ejecutamos el comando *spshape2dta* de la siguiente manera:

```
unzipfile co_dep, replace // si el archivo zip está en el directorio de trabajo
```

```
spshape2dta co_dep, replace
```

Este comando crea dos archivos .dta con base en los archivos .shp y .dbf que hacen parte del shapefile. El archivo co\_dep\_shp.dta proviene del archivo co\_dep.shp, este contiene las coordenadas de cada punto que permite dibujar el mapa. El archivo co\_dep.dta proviene del archivo co\_dep.dbf, que contiene la tabla de atributos de las unidades espaciales que conforman el mapa (información de los departamentos, en este caso). Adicionalmente, en cada archivo .dta crea una variable llamada \_ID que relaciona las coordenadas con la información de cada unidad espacial del mapa, esta variable se construye con base en el archivo co\_. Por defecto, también crea dos variables adicionales en la tabla de atributos (co\_dep.dta) llamadas \_CX y \_CY, que son las coordenadas del centroide de cada polígono. Si quiere que los archivos .dta creados tengan un nombre distinto a los de su shapefile de base, puede usar la opción [saving()].

---

4

En versiones anteriores a *Stata* 15 se debe instalar y usar el comando desarrollado por usuarios (UWS) *shp2dta* (ver anexo), que se especifica de la siguiente manera:

```
shp2dta using co, database(col_dep) coordinates(col_dep_shp) geind(_ID)
```

Con “using co\_dep” le estamos señalando a *Stata* que los archivos que inician en co\_dep son los que contienen la información geográfica de interés. *Stata* genera 2 archivos en formato .dta, a los cuales debemos asignarles un nombre. En este caso, con la opción [database()] se crea el archivo col\_dep, que contiene la información almacenada en el archivo .dbf, que son los datos de cada unidad espacial. Con la opción [coordinates()] se crea el archivo col\_dep\_shp, que contiene las

coordenadas de las unidades espaciales, provenientes del archivo .shp. Con la opción `[genid()]` se crea la variable identificadora de la unidad espacial en las bases `co_dep.dta` y `co_dep_shp.dta`.

Normalmente requerimos agregar información adicional a la tabla de atributos `co_dep.dta`, para esto usaremos el comando *merge*. Claro está, hay que ser cuidadosos en que los indicadores usados para hacer usar *merge* sean consistentes entre bases de datos, en este caso, nuestra información está a nivel departamental, por lo cual vamos a usar el código de departamento del DANE. La variable que contiene este código debe llamarse igual en la tabla de atributos (`co_dep.dta`) y las bases de datos que vamos a unir a ella. Profundizamos en las operaciones y los comandos necesarios para unir bases de datos en las notas de la clase 5.

## 2.2. Comando *grmap*<sup>3</sup>

La lógica del comando *grmap* para crear un mapa en *Stata* es la siguiente. Suponga que tenemos una base de datos con información de ingreso promedio del jefe del hogar por departamento en Colombia y que queremos hacer un mapa del país con colores según el nivel de ingreso, similar al que aparece en la Figura 1. Para hacerlo, *Stata* necesita dos bases de datos, creadas cuando usamos el comando *spshape2dta*. Una con la información de interés a nivel departamental (variables como ingreso promedio del jefe del hogar) y con una variable que identifica al departamento, que el comando *spshape2dta* llama `_ID`. Si usa el comando *shp2dta* se crea con las opciones `[database(col_dep)]` y `[gen(_ID)]`. La segunda base de datos tiene la información geográfica de las coordenadas de los departamentos dentro del mapa nacional y el identificador `_ID`. Si usa el comando *shp2dta*, esta base de datos se crea con la opción `[coordinates(col_dep_shp)]`. Por supuesto, el identificador `_ID` de un departamento toma el mismo valor en ambas bases. Para construir el mapa, el comando *grmap* hace el emparejamiento por medio de la variable identificadora `_ID` entre los valores de las variables de interés y las coordenadas.

El comando *grmap* es capaz de proyectar mapas de puntos, líneas o polígonos. En la ayuda del comando (escriba en la ventana de comandos “*help grmap*”) se puede consultar todas las opciones disponibles de edición para cada tipo de mapa. En estas notas solo vamos a referirnos a las opciones generales del mapa (*basemap*) como las opciones de color, formato y leyenda.

---

<sup>3</sup> En versiones anteriores a *Stata* 15 se debe instalar y usar el comando desarrollado por usuarios (UWS) *spmap* (ver anexo). Tiene las mismas especificaciones del comando *grmap*.

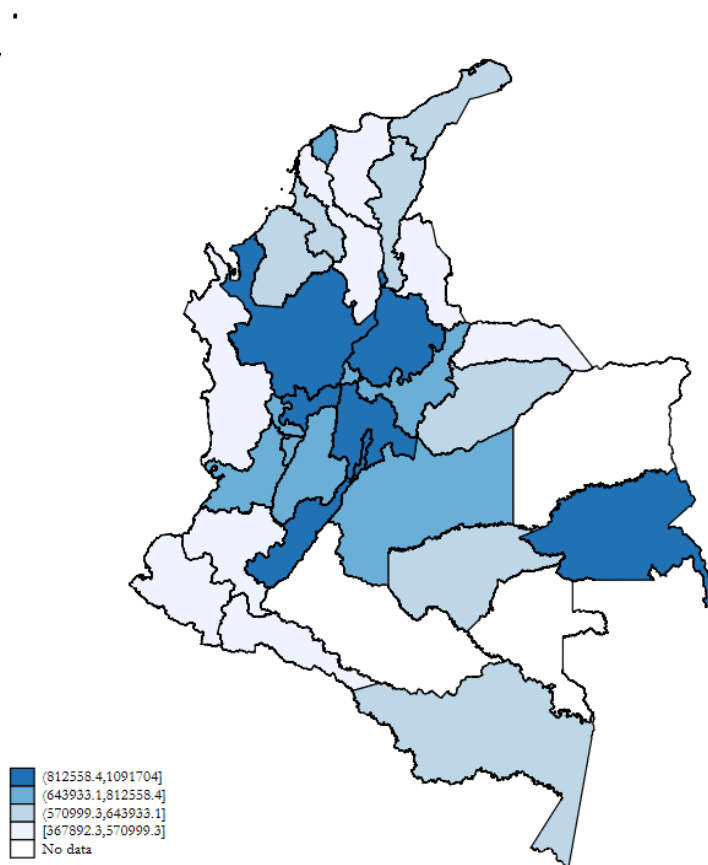
Suponga que tenemos la base `col_dep.dta` cargada en *Stata* con los datos de ingreso promedio del jefe del hogar por departamento en Colombia. El gráfico que aparece en la Figura 1 lo obtenemos ejecutando el siguiente comando.

```
grmap, activate
```

```
grmap ingresos_hogar_jefe using col_dep_shp.dta, id(_ID) name(mapa_1, replace)
```

Para poder usar el comando *grmap* es necesario activarlo escribiendo una vez por sesión “*grmap*, activate”. En la sintaxis de *grmap* se escribe el nombre la variable a ilustrar en el mapa y se utiliza la sintaxis “using” para relacionar las características geográficas con los atributos (variables). La opción `id()` señala la variable identificadora de la entidad geográfica (`_ID`). Si no se hace ninguna especificación adicional, *Stata* por defecto divide los valores de la variable en cuartiles, cada uno de los cuales es un intervalo que se representa en el mapa con una tonalidad de color distinta y lo hace en escala de grises.

**Figura 1. Mapa base - Ingreso promedio del jefe del hogar por departamento**



## 2.3. Opciones generales

Los mapas tienen propiedades similares a los gráficos de *Stata*. Las opciones `[, title(,suboptions)]`, `[, subtitle(,suboptions)]`, `[, caption(,suboptions)]`, `[, note(,suboptions)]`, `[, name(,suboptions)]`, `[, saving(,suboptions)]` se usan de la misma manera. En seguida se explica el uso de algunas opciones particulares a los mapas en *Stata*.

### 2.3.1. Opciones de color

El mapa mostrado en la Figura 1 es un mapa coroplético, se utilizan para mostrar el comportamiento espacial de una variable. Normalmente, los valores más altos de la variable se representan con tonos más intensos de color. Por defecto, *Stata* utiliza colores de la escala de azules para las diferentes categorías en el mapa. Sin embargo, estos colores pueden ser modificados con la opción de tipo *basemap* `[, fcolor( )]`. Para ver las transiciones y combinaciones de colores disponibles puede consultar la ayuda del comando y luego haciendo clic en el enlace en azul *colorlist*.

Adicionalmente, como se muestra en la tabla *colorlist*, existen tres tipos de escalas que pueden ser utilizadas para rellenar los mapas. El primer tipo es secuencial y se utiliza cuando es necesario representar información con algún tipo de ordenamiento. Por ejemplo, ingresos de mayor a menor. El segundo tipo es cualitativo y se utiliza cuando no es posible realizar un ordenamiento numérico de la variable. Por ejemplo, el tipo de transporte que prima en el departamento, la enfermedad con mayor incidencia en la región, etc. El tercer tipo es el divergente y es utilizado cuando es posible establecer un tipo de ordenamiento según la diferencia entre el valor de la unidad geográfica y un punto de referencia. Por ejemplo, qué tanto difieren las transferencias a un municipio respecto al promedio nacional.

### 2.3.2. Opciones para el manejo del dato

Cuando la variable de interés es numérica se puede modificar el método con el que se construyen los intervalos y su cantidad con la opción `[, clmethod(method)]`. Además de los cuantiles, uno de los métodos más comunes es el *boxplot*, donde las categorías son las que toma un diagrama de caja y bigotes (o caja de distribución, o diagrama de vela), en donde hay una categoría superior y otra inferior para los valores extremos. También es posible introducir los intervalos manualmente con el método *custom*, seguido de la opción `[, clbreaks(numlist)]` para introducir los valores que definen el inicio o fin de un intervalo. La opción `[, cln(#)]` permite definir la cantidad

de intervalos, # señala el número de grupos en los que se desea dividir la información dentro del mapa.

### 2.3.3. Opciones de leyenda

Además de las opciones de leyenda que se utilizan para los gráficos estándar, hay cuatro tipos de leyenda que se pueden utilizar en los mapas. El tipo de leyenda utilizado se selecciona con la opción `[, legstyle(#)]`. El tipo de leyenda cero (0) elimina las etiquetas para los colores. El tipo uno (1) utilizan la notación matemática de conjuntos  $(\cdot, [ \cdot ])$  y es el utilizado en *Stata* por defecto, como se ve en la Figura 1. El tipo dos (2) utiliza la notación:  $li - ls$  donde  $li$  es límite inferior y  $ls$  es límite superior. Finalmente, el tipo tres (3) sólo pone etiquetas para el primer y el último elemento (tipo 3).

La opción `[, legend (suboptions)]` permite modificar el tamaño, la ubicación y las etiquetas de los intervalos a través de subopciones. El tamaño de la leyenda se modifica con la subopción `[, legend(size(*#))]`, donde # toma un número entre cero y uno si se quiere reducir el tamaño o mayor a uno si se quiere aumentar. De la misma manera que con los gráficos estándar, la ubicación de la leyenda se puede modificar con una combinación de las subopciones `[, legend(ring(#) position(#))]`. Las etiquetas de la leyenda se pueden introducir manualmente con la opción `[, legend(order(1 “etiqueta uno” 2 “etiqueta dos” 3 “etiqueta tres”))]`, de tal manera que el primer intervalo se denota con el numero uno y en seguida se coloca la etiqueta asignada entre comillas.

Como aclaración de sintaxis, para introducir opciones y subopciones aquí tratadas, el código se escribiría así:

```
[, legstyle(#) legend(size(*#) ring(#) position(#) order(1 “etiqueta uno” )) ]
```

### 2.3.4. Opciones líneas divisorias

Las líneas divisoras de las unidades geográficas también se pueden modificar. Para esto se debe especificar la opción `[, mopattern(tipo de línea)]`. Los diferentes patrones para las líneas se presentan en la table *linepatternstyle* en la ayuda del comando.

De manera similar, es posible escoger el color de la línea de la lista en la tabla *colorstyle* utilizando la opción `[, ocolor(colorstyle)]`. Si se quiere aplicar para todos los grupos, se debe utilizar la opción de la siguiente manera: `[, ocolor(colorstyle ...)]`



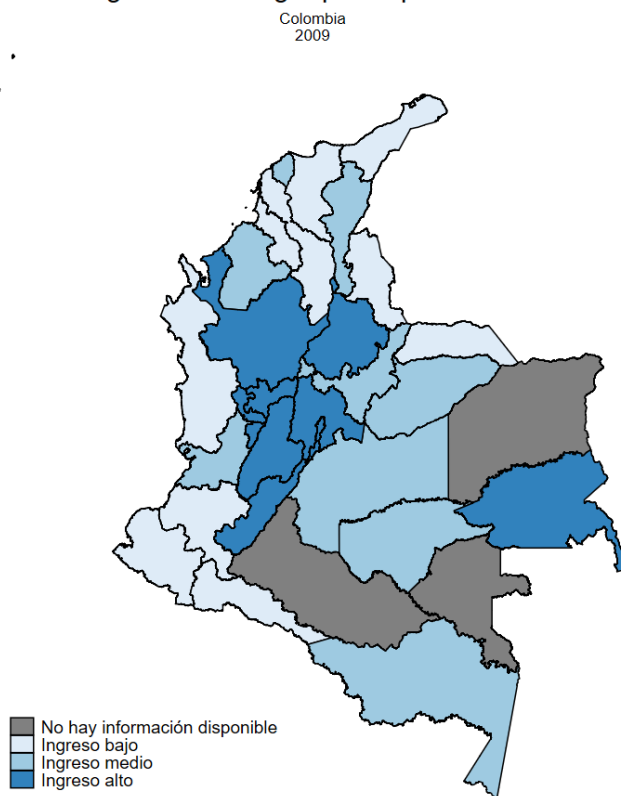
### 2.3.5. Opciones *missing values*

Si el dato no está disponible para alguna entidad, por defecto se crea la categoría “No data”. *Stata* grafica en color blanco las unidades geográficas con *missing value*. Sin embargo, esto se puede modificar con la opción [, ndf(color)] seleccionando un color de la tabla *colorstyle* disponible en la ayuda del comando. Los *missing values* son la primera categoría en la leyenda.

#### Ejemplo

```
grmap ingresos_hogar_jefe using "col_dep_shp", id(_ID) ///
    fcolor(Blues) ndf(gs8) /// Opciones de colores
    clmethod(custom) clb(367892.3 590846.4 715309 1091704) /// Opciones de intervalos
    legstyle(3) legend (position(1) ring(0) size(*1.2)) /// Opciones de leyenda
    legend(order( 1 "No hay información disponible" ///
        2 "Ingreso bajo" ///
        3 "Ingreso medio" ///
        4 "Ingreso alto") c(1) position(7)) /// Opciones de la leyenda
    title ("Ingresos del hogar por departamento", size(*0.8)) ///
    subtitle("Colombia" "2009", size (*0.6)) ///
    name(ejemplo, replace)
```

**Figura 2. Mapa con opciones generales**  
Ingresos del hogar por departamento



## Anexo: Programas UWS

Muchos de los comandos avanzados son creados por usuarios y el paquete básico de *Stata* no los incluye. Este es el caso de los comandos para crear mapas. Por esta razón, estas notas de clase inician con una introducción a cómo buscar e instalar este tipo de programas, llamados UWS (User-Written-Software). Ya se había hecho referencia a este tipo de comandos en clases anteriores, pero aquí entraremos en más detalles.

### A. Introducción

La mayor parte de los comandos UWS se encuentran en la página web de RePEc<sup>4</sup> en la sección de *Statistical Software Components* (SSC, <http://ideas.repec.org/i/c.html>). Si bien existen otros sitios en internet en los que se encuentran programas desarrollados por usuarios, estos no cuentan con una plena garantía de la fiabilidad del archivo. Estos paquetes, por lo general, incluyen un archivo de ayuda (.hlp) y un archivo .ado, es decir, un *do-file* destinado a ser ejecutado como comando. Aunque *Stata* no incluye estos comandos en su paquete básico, sí incluye un comando (**ssc**) que permite acceder a esta base de datos para buscar e instalar comandos desarrollados por usuarios.

10

### B. Búsqueda de comandos (paquetes)

El comando *ssc new* presenta una lista de los últimos paquetes desarrollados. El comando *ssc hot* presenta los 10 comandos UWS más populares en los últimos tres meses. La opción [, n(x)] presenta los x primeros comandos UWS más populares y la opción [,author(nombre)] presenta los comandos UWS más populares del autor especificado dentro de la opción. Si solo se conoce algún procedimiento realizado por el comando, lo más útil es usar el comando *findit*. Al ejecutar el comando:

**findit** nombre

*Stata* presenta una lista de programas, incluidos los paquetes UWS que contienen la palabra nombre en su archivo de ayuda o en su descripción. Para una búsqueda más detallada puede ir

---

<sup>4</sup> Repec es el nombre abreviado de *Research Papers in Economics*. El objetivo de esta página de internet es divulgar conocimiento útil para la investigación económica. Para esto, cuenta con una base de datos de documentos de trabajo o *working papers*, artículos publicados y programas específicos para diferentes *softwares*, como *Stata*. El acceso a la mayoría de los archivos es gratis. Sin embargo, en el caso de algunos artículos publicados es necesaria la suscripción a otras bases de datos de acceso restringido como *J-stor*.

directamente a la página de los SSC del Boston College Department of Economics (<http://ideas.repec.org/s/boc/bocode.html>), el instituto que almacena la mayor parte de los UWS.

### C. Instalación y actualización de comandos

El comando *ssc install* instala el conjunto de archivos pertenecientes al paquete del comando. Estos comandos son constantemente modificados por sus autores y se recomienda mantenerlos actualizados. El comando *ssc describe* presenta una descripción del comando en su versión actual y, si una vez instalado se requiere actualizarlo, es necesario volver a instalarlo usando la opción `[replace]`:

**ssc install** comando, replace.

Por ejemplo, en versiones anteriores a *Stata* 15, los dos comandos más utilizados para manejar datos espaciales y crear mapas en *Stata* son *shp2dta* y *spmap* (homólogos a *spshape2dta* y *grmap*). El comando *shp2dta* es utilizado para transformar la información de las coordenadas de los mapas a un archivo de formato *Stata*, el comando *spmap* es utilizado para proyectar mapas como gráficos de *Stata*. Puede instalar los comandos con los siguientes códigos:

**ssc install** spmap, replace

**ssc install** shp2dta, replace