



# ESDA

(Exploratory Spatial Data Analysis)

---

## Análisis Exploratorio de Datos Espaciales Parte I

Dr. Luis Quintana Romero  
FES-Acatlán



# EDA

---

- Es una colección de técnicas para resumir propiedades de los datos (estadística descriptiva) pero también para detectar patrones en los datos, identificando características interesantes o inusuales, detectar errores, distinguir características accidentales de las que son relevantes y formular hipótesis a partir de los datos (Good, 1983).
- EDA también se usan para examinar resultados de modelos, proporcionar evidencia sobre los supuestos del modelo que se cumplen y determinar si existen efectos influyentes de los datos en el modelo estimado.



# ESDA

---

- Sintetiza las propiedades espaciales de los datos, detecta patrones espaciales, formula hipótesis de la geografía de los datos, identifica casos y sub conjuntos de casos inusuales dada su localización en un mapa.
- Al igual que EDA son técnicas visuales y numéricamente resistentes.
- A diferencia del EDA aquí el mapa asume un papel central y, por lo tanto, además de las técnicas del EDA incluye métodos adicionales para direccionar preguntas que son consecuencia de la georeferenciación de los datos.

# Modelos conceptuales de variación espacial

---

## ○ MODELO REGIONAL

- Regiones: Son las unidades espaciales utilizadas para análisis espacial de datos.
- Tres tipos de regiones:
  - 1) Formal (uniforme): Se construyen dividiendo el espacio en áreas cuasi-homogéneas. Sus fronteras representan cambios en los niveles de los atributos.
  - 2) Funcional (nodal): Se determinan utilizando datos de interacción. Sus fronteras usualmente se definen por características del territorio (carreteras, vías ferreas, etc.) que actúan como barreras a la interacción entre regiones.
  - 3) Administrativa: Se determinan por decisiones de política para administrar el espacio.

# Modelos conceptuales de variación espacial

---

- ESPACIO ÁSPERO Y SUAVE (rough and smooth)
  - Modelo de Tukey:  
Datos = suave + áspero
  - La parte suave es predecible (componente regular) en tanto que la áspera es impredecible (componente irregular). La áspera representa valores espaciales extremos, discontinuidad o áreas muy diferentes.
  - El ESDA utiliza la estructura suave de los datos:
    - Bloques o cúmulos de valores similares
    - Propensión de valores a estar juntos (autocorrelación espacial).
    - Contrastes entre diferentes segmentos de un mapa.
    - Tendencias o gradientes en el mapa.

# Modelos conceptuales de variación espacial

---

## ○ VARIACIÓN ESPACIAL DE ESCALAS

- Para datos de superficie continua se pueden representar como:

Datos espaciales = macro escala de variación  
+ media/pequeña (meso) escala de variación  
+ error

- Se analizan las áreas y sus tamaños.
- Si el componente meso domina el mapa de una variable se dice que despliega heterogeneidad regional y son más recomendables los estadísticos locales de análisis espacial.



# EDA

---

- No hace supuestos sobre la distribución de los datos y generalmente excluye las pruebas de hipótesis; no emplean teoría inferencial.
- Son técnicas principalmente visuales (tablas, gráficas, figuras) y/o cuantitativas como resúmenes de los datos.
- Utiliza estadísticos “resistentes”, es decir que no se afectan mucho por la presencia de valores extremos.
- Ejm. la mediana y el rango intercuartil.



# ESDA: MÉTODOS DE VISUALIZACIÓN

---

- Buscan detectar propiedades de los datos a través de visualización cartográfica.
- Es recomendable utilizar variables de relación para no distorsionar las comparaciones o utilizar alguna normalización de los datos.



# MAPAS CLOROPLÉTICOS


---

- CUANTILES: Los datos son ordenados y agrupados en categorías con igual número de observaciones o cuantiles y le asigna un color a cada una de ellas.
  - Ejm. Abra el proyecto: SIDS2 que contiene datos del síndrome de muerte infantil súbita en 100 condados de Carolina del Norte en 1974 (SIDS74) y 1979 (SIDS79). Utilice como identificador el código de los condados FIPSNO. Con el botón derecho abra el menú contextual y con la opción COLOR cambie el color al mapa (azul).

# Ejemplo

---

- Construya un mapa de cuantiles para comparar la distribución de nacimientos de niños no blancos (NWBIRTHS74) y el SIDS74: MAP MENU/QUANTILE. Utilice la opción de cuantiles (4 categorías).
- Utilice la opción DUPLICATE MAP para abrir una copia del mapa original y construya el mapa de cuantiles para la otra variable.
- Con la opción SELECTION SHAPE puede seleccionar partes de un mapa y se seleccionan las mismas en el mapa vinculado.
- Problemas en el mapa: Tipo de variables (extensivas) y distribución de los casos (muy sesgada)

- 
- 
- Abra el archivo SIDS, genere una nueva columna: TABLE/ADD COLUMN con nombre SIDR74 .
  - Genere una variable intensiva: TABLE/FIELD CALCULATION, con las opciones: Result: SIDR74, Methods: Raw rate, Event Variables: SID74, Base Variables: Birt74.
  - Reescale la variable a una tasa por cada 100,000 nacimientos usando FIELD...
  - Visualice en cuartiles su nuevo mapa.



# MAPAS CLOROPLÉTICOS

---

## ○ POLIGONOS DE THIESSEN

- Son polígonos creados a partir de un mapa de puntos. Cada polígono contiene los puntos del mapa que son más cercanos al centroide. Se utilizan para identificar área de mercado.
- Son más útiles cuando el patrón que siguen los puntos es difícil de identificar.
- Adicionalmente permiten calcular en mapas de puntos matrices de vecindad utilizando los polígonos definidos para los puntos.

# Ejm.

---

- Calculando centroides: Con el menú TOOLS/Polygons to points utilice como archivo input OHLUNG.SHP que muestra datos de cáncer de pulmón para los 88 condados de Ohio. Al archivo destino que va a llevar los resultados de los centroides nómbrelo centros. Oprima el botón CREATE y luego DONE.
- Abra el archivo original Ohlung y luego con EDIT añada como nueva capa el archivo centros.
- Con el botón derecho del ratón seleccione la opción de añadir centroides a la tabla, con coordenadas X,Y.



## Ejm.

---

- Convierta de puntos a polígonos el archivo oz9799 que tiene los observatorios de calidad del aire de la ciudad de los Ángeles.
- Compare los dos mapas y evalúe la distribución de los observatorios.

# EDA: Histogramas

---

- Histogramas vinculados: Abra el archivo grid100s que muestra datos de variables simuladas con un mapa de celdas.
- Genere dos mapas de cuantiles para zar09 y ranzar09. ¿Qué diferencia hay entre las variables?
- Obtenga los histogramas para cada una de las variables: Son iguales, pero los patrones espaciales no lo son.
- Seleccione una barra del histograma y observe como se distribuye regionalmente en el mapa.

# EDA: Box Plot

---

- Box-Plot: Muestra la mediana (al centro de la caja), primero y tercer cuartil de una distribución. Clasifica outliers tomando el rango intercuartil como umbral (1.5 y 3 veces dicho rango)
- Abra el archivo `stl_hom`. Con Explore seleccione Box-Plot para la variable `HR8893` que es la tasa de homicidio para esos años.
- El rectángulo que se forma es la distribución acumulada. La línea central muestra la mediana y el rectángulo sombreado es el rango intercuartil. La línea tenue es el umbral y arriba aparecen seis condados como outliers.
- Seleccione con un rectángulo los valores extremos del Box-Plot, visualícelos en el mapa e identifíquelos en la tabla.



# EDA: Diagrama de dispersión

---

- Realice con EXPLORE un diagrama de dispersión de la variable HR7984 y RDAC80 que es un índice de privación de recursos.
- Con el botón derecho en la gráfica de dispersión seleccione la opción de datos estandarizados la cual se interpreta como un coeficiente de correlación.
- Seleccione las observaciones extremas, elimínelas con la opción Exclude selected y observe lo que sucede con la línea de regresión



# EDA: Cepillado (Brushing)

---

- Seleccionar casos o subcasos en una gráfica y observar en dónde caen en otra gráfica se conoce como brushing.
- Al realizar diagramas de dispersión puede ir creando rectángulos y mirar el efecto de esos puntos excluidos en la pendiente de la recta de regresión. También lo puede mirar en un mapa.



# Tarea


---


- Realizar las prácticas de los capítulos 6 al 8 del Geoda work book

# EDA MULTIVARIADO

---

- Permite visualizar la asociación entre más de una variable.
- Diagramas de dispersión matriciales; Grafican pares de variables y se ordenan matricialmente considerando que en cada fila la variable en el eje y es la misma.
- Ejm. Abra el archivo Police, genere un mapa de quintiles para las variables police, crime. Genere diagramas de dispersión para las dos variables alternando su posición en los ejes y forme una matriz con las gráficas y los mapas. Observe sus resultados utilizando cepillado.

- 
- 
- Coordenadas Paralelas: Son gráficas en donde cada variable representa un eje que es paralelo al resto. Los valores en el eje van de menor a mayor (izq. a derecha). Permiten ubicar regiones en el mapa y clusters.
  - Ejm. Realice un PP con las tres variables (police, unemp, crime)

- 
- 
- Gráfica condicional:
  - Ejm. Genere centroides,