

Exploración y visualización de datos para lo socioeconómico

Miguel Andrés Garzón Ramírez

14 de agosto de 2025

Clase 1: Introducción al uso de datos

Objetivo

Esta práctica tiene como objetivo explorar conceptos esenciales de limpieza y organización de datos para realizar análisis de datos. Trabjará en equipos conformados en clase a partir de una pregunta de indagación.

Condiciones

- Puede usar las herramientas que tenga a la mano para realizar la actividad: Excel, Stata, R, Python, Power BI, Tableau u otro.
- Se debe mostrar un proceso de acuerdo con la pregunta de indagación planteada. Puede tener en cuenta los siguientes pasos:
 - Carga de datos
 - Transformación por medio de operaciones con datos: esto implica plantear una estrategia de pasos con cada tabla involucrada
 - Elaboración de cálculos requeridos para responder la pregunta
 - Exploración del comportamiento de la variables existentes y creadas, respondiendo a las preguntas
- Cada integrante del equipo debe interactuar y contribuir con sus capacidades y conocimientos en el trabajo. Sugerimos iniciar el trabajo con una charla que les permita conocerse para definir los roles de trabajo
- Cuenta con dos tablas que debe procesar:
 - Población municipal de los últimos 4 censos (.xlsx)
 - Área y año de creación de los municipios de Colombia (.csv)
- Debe entregar:
 - Código o proyecto (Stata, R, Python, Excel/Power Query o Power BI) con pasos reproducibles.
 - Documento de una página de texto donde responda las preguntas según lo solicitado. Adicional a la página de texto de respuesta, puede colocar gráficos de apoyo para el análisis.

Entendimiento de los datos

1. Identifique y justifique la unidad de observación en cada tabla de datos suministrados
2. Clasifique cada variable como identificación o medición; indique tipo y formato.
3. Proponga al menos 3 reglas mínimas de consistencia para asegurar la calidad de los datos de las variables disponibles y que constituyen parte de la documentación del ejercicio. (e.g., AÑO \in {1985,1993,2005,2018}).

Preparación de los datos

1. Construya la llave única para la tabla de población (sugerencia: DPMP + AÑO).
2. Verifique duplicados; reporte conteos y la decisión de tratamiento.
3. Explique qué tipo de unión tendrían las tablas disponibles y por qué.
4. Asegure tipos de datos correctos de acuerdo con las reglas que identificó previamente (e.g. DP, DPMP caracteres que pueden iniciar con "0"; Población entera).
5. Detecte datos faltantes y decida su tratamiento, si es necesario en el contexto de este ejercicio.
6. Estandarice nombres (mayúsculas/acentos) en DPNOM y MPIO, si es necesario.
7. A partir de la estructura larga de la tabla población (AÑO en columna), cree una versión ancha con columnas Pob_1985, Pob_1993, Pob_2005, Pob_2018 y guárdela en un nuevo archivo.
8. Explique cuándo trabajar en largo vs. en ancho para las operaciones siguientes.

Integración de tablas (join), validaciones y cálculos

1. Realice la unión de las tablas población (largo) con mun_area.csv
2. Reporte registros sin correspondencia en ambas direcciones.
3. Documente la cardinalidad (muchos-a-uno) y cómo evita relaciones muchos-a-muchos.

Cálculos de interés (medición)

Con la tabla integrada:

1. Densidad poblacional = Población / MPIO_NAREA.
2. Crecimiento absoluto 1985→2018 y crecimiento porcentual 1985→2018 por municipio.
3. Agregación por departamento (DP) para densidad y crecimiento (aclare la definición usada).

Preguntas de indagación

1. ¿Cuáles son los 20 municipios de Colombia con mayor crecimiento poblacional desde 1985? Absoluto y porcentual.
2. ¿Cuáles son los 10 municipios de Colombia con mayor densidad en 1993 y en 2018?
3. ¿Cuáles son los 5 departamentos de Colombia con mayor densidad en 1993 y en 2018?
4. ¿Cómo ha cambiado la distribución de la población municipal colombiana desde 1985? Puede realizar un análisis por cuartiles o percentiles con gráficos.