

Introducción a R para manejo de bases de datos y operaciones geográficas

R es un lenguaje distribuido gratuitamente para realizar estadística y análisis computacional. Gracias a su penetración y su componente de código abierto, hoy podemos utilizarlo para realizar análisis geoespaciales utilizados en el estudio del transporte.

1 Características del curso

Dirigido a colaboradores de la Dirección de Consultoría: Gerencias de Planeación y Movilidad con conocimientos de las metodologías adoptadas por Cal y Mayor.

Duración: 10 sesiones de 90 minutos para el desarrollo de ejercicios teórico-prácticos relacionados con los procedimientos habitualmente utilizados.

2 Objetivos

Adquirir nociones básicas de manejo de software de programación con lenguaje R, para el manejo de bases de datos, análisis estadísticos y visualizaciones, para entender y formular modelos de movilidad y transporte.

Complementar el conocimiento y fomentar el uso de herramientas especializadas que permitan a los colaboradores desarrollar e implementar procesos y metodologías eficientes, reproducibles y mantener un mejor control de los mismos.

3 Contenido del curso

3.1 Estructura

Sesión	Temas	Contenido	Teoria	Practica	Q&A	Nivel
1	1, 2 y 3	Instalación, introducción y objetos	45 min	25 min	20 min	1
2	4	Funciones intermedias	30 min	60 min	15 min	1
3	5	Tidyverse	35 min	65 min	10 min	1
4	6	Importación y Union de datos	30 min	60 min	10 min	1
5	7	Visualización de datos	30 min	60 min	15 min	1
8	8	Limpieza de datos				2
8	9	Análisis Exploratorio de datos (EDA)				2
9	10	Estadística en R				2
10	11.1	Modelos de regresion				2
11	11.2	Modelos de regresion				2
12	12.1	Análisis espacial: visualización y analisis				3
13	12.2	Análisis espacial: estadística y mapas interactivos				3
14	13.1	Geocomputacion: introducción y zonificacion				3

Sesión	Temas	Contenido	Teoria	Practica	Q&A	Nivel
15	13.2	Geocomputacion: lineas de deseo y operaciones de ruteo				3

3.2 Descripción de los módulos

3.2.1 Instalación

Para su instalación y el mejor uso, tenemos que instalar el compilador, así como la interfaz de usuario más popular RStudio.

Compilador:

<https://cran.r-project.org>

RStudio

<https://www.rstudio.com/products/rstudio/download/>

3.2.2 Introducción a R

- Usos y funciones del lenguaje R
- Librerías
- Conceptos básicos de variables

3.2.3 Objetos

- Tipos de objetos:
 - Vectores
 - Matrices
 - Bases de datos (dataframes)
 - Listas

3.2.4 Funciones intermedias

- Condicionales
- Bucles (loops)
- Entendimiento y creación de funciones
- Familia “apply”
- Otras utilidades

3.2.5 Tidyverse

- ¿Qué es tidyverse?
- Manejo de datos
- Agrupación y síntesis
- Introducción a la visualización
- Tipos de visualización

3.2.6 Importación y Unión de datos

- Importación de datos
 - Importación con R base
 - Importación desde excel
 - Conexión de Excel con XLConnect
- Unión de bases de datos:
 - Unión interna
 - Uniones izquierda y derecha
 - Otros tipos de uniones

3.2.7 Visualización de datos

- Gramática de los gráficos
- Estéticas (aesthetics)
- Geometrías
- Temas

3.2.8 Limpieza de datos

- Problemas comunes en bases de datos
- Variables categóricas y de texto
- Problemas avanzados de limpieza de datos
- Enlaces de registros

3.2.9 Análisis Exploratorio de Datos (EDA)

- Variables categóricas
- Variables numéricas
- Caso de estudio: Encuesta OD INEGI 2017

3.2.10 Estadística en R

- Estadística descriptiva
- Números aleatorios y probabilidad
- Distribuciones y Teorema de Límite Central
- Correlación y diseño experimental

3.2.11 Modelos de regresión

- Regresión lineal simple
- Predicciones
- Evaluación de modelos
- Regresión logística simple

3.2.12 Análisis espacial

- Visualización de datos geoespaciales
- Análisis espacial
- Estadística espacial
- Mapas interactivos

3.2.13 Geocomputación

- Conceptos básicos de geocomputación
- Zonificación
- Líneas de deseo
- Operaciones de ruteo