

Utilizar R como un Sistema de Información Geográfica y hacer análisis reproducible.

Los Sistemas de Información Geográfica (SIG) permiten almacenar, organizar, manipular y hacer análisis de datos que posean una referencia espacial (coordenadas). Son utilizados principalmente para realizar cruces de información de diferente tipo (social, ecológica, política, física) para mejorar la toma de decisiones en diferentes áreas, tales como: ciencias de la tierra, ciencias sociales, ingeniería en transporte y aplicaciones más recientes en marketing y ventas, entre otras.

Existen conocidos SIG, como ArcGis (comercial) y Qgis (libre) que son los más utilizados, pero que no siempre permiten almacenar el proceso en un código reproducible, ya que: 1) se centran más en la interfaz y 2) la ejecución de código es vista como herramientas para usuarios más avanzados (ambos permiten código Python en una interfaz de código básica). En el último tiempo, las librerías que permiten realizar análisis espacial en R han tenido una gran mejoría y se encuentran en constante desarrollo, tanto así, que ahora es posible realizar la mayoría de las acciones que permiten los SIG convencionales, con la ventaja de poder generar un código reproducible y realizar labores iterativas en bucle de forma fácil sin tener que aprender nuevos lenguajes ni instalar (o crackear) un software comercial. Además, se están desarrollando librerías que permiten conectar R con los SIG tradicionales para utilizar herramientas más específicas.

Utilizar R tiene otras ventajas, como poder hacer mapas y figuras uniformes (colecciones de mapas), realizar mapas interactivos con leaflet y visualización mediante Shiny.

El objetivo de la presentación será mostrar las diferentes tareas que se pueden hacer con R en este ámbito, mostrando su uso en casos aplicados con entidades vectoriales y raster. También, hacer hincapié en las ventajas de realizar todos los procesos con líneas de código para hacer investigación reproducible, poder realizar cambios “más fáciles” cuando se proponen nuevas metodologías, actualizaciones más rápidas y trazabilidad de errores.

Y mostrar que R no es sólo un software estadístico o para hacer los gráficos, sino que es eso y ¡mucho más!

Específicamente hablaré sobre los paquetes más conocidos para llevar a cabo un flujo típico de análisis reproducible:

- 1) Leer datos (paquetes: raster, sf)
- 2) Manipular datos (paquetes: sp, rgdal, gdalcubes)
- 3) Analizar datos, generar estadísticas, gráficos o mapas (paquetes: dismo, ggplot, rastervis, leaflet, etc.)
- 4) Comunicar resultados (Shiny, Rmarkdown)

Se mostrarán dos aplicaciones, la primera es utilizando el paquete dismo, el cual permite realizar modelamiento de nicho ecológico con base en variables físicas/climáticas y a datos de ocurrencia de especies. Mostraré cómo desde del mismo paquete se pueden descargar datos desde gbif (Global Biodiversity Information Facility <https://www.gbif.org/>) para realizar este tipo de estudios.

También mostraré cómo se ha facilitado el proceso de análisis de imágenes satelitales con la nueva librería “gdalcubes”, mostraré un pequeño ejemplo de análisis de temperatura superficial con datos del sensor MODIS (Moderate-Resolution Imaging Spectroradiometer) los cuales se pueden descargar de forma libre desde la web de la NASA.

Dejaré los códigos de ambos análisis en github para que los participantes puedan acceder a ellos.