

Introducción al análisis geoestadístico básico con R



Dr. Juan Pedro Pérez Alcántara

Programa de Doctorado en Geografía

21 - 25 de octubre de 2024

¿Por qué usar un lenguaje de programación?

- 1 **Objetivo (cumplido) fundamental de R:** acercar la programación a los científicos sin tener que tener ingentes conocimientos de ingeniería de software.
- 2 **Flexibilidad:** la interacción basada en herramientas gráficas (botones) es limitada en sus posibilidades.
- 3 **Reproducibilidad:** es la mejor forma de hacer computación reproducible y portable.
- 4 **Integrabilidad y escalabilidad:** se puede integrar con otros sistemas y escalar en volumen de datos.

El lenguaje de programación R

Características básicas:

- 1 Orígenes en el lenguaje S de los Laboratorios Bell en 1978.
- 2 Lanzado en 1995 con licencia GNU GPL, versión 1.0.0 estable en 2000.
- 3 R es un lenguaje de programación de muy alto nivel orientado al análisis de datos, la estadística, la visualización y el Machine Learning.
- 4 Es un lenguaje interpretado (no compilado), con diversos modelos de ejecución, como scripts o REPL (Read-Eval-Print Loop).
- 5 Es multiplataforma.
- 6 Posee una buena integración con C, C++ y Fortran, lo que le permite la ejecución de código compilado de alto rendimiento.

El lenguaje de programación R

Características básicas:

- 1 Es software libre y de código abierto, de altísima calidad.
- 2 Posee una gran comunidad de usuarios y de recursos de aprendizaje.
- 3 Posee una gran cantidad de paquetes reutilizables (CRAN) orientados a los más diversos fines.
- 4 Permite programar con estilo imperativo, estructurado, funcional y orientado a objetos.

¿R o Python?

- 1 **No son excluyentes:** integración de microservicios.
- 2 **Competidores / colaboradores:** entre ambos dominan el sector del análisis de datos.
- 3 **Orientados a hacer ciencia:** ambos son lenguajes de alto nivel altamente productivos y expresivos.
- 4 **Enfoque:** Python es más generalista, mientras que R es más específico en sus objetivos, que son análisis y visualización de datos.
- 5 **Curva de aprendizaje:** en R es más suave en el contexto de programadores nóveles que se quieran circunscribir al análisis de datos.

Consideraciones de diseño y limitaciones

- 1 **Limitado a la memoria:** R sólo trabaja con los datos que caben en memoria. No hace *Big Data* por sí solo.
- 2 **Duck Typing:** débilmente tipado, aproximación laxa a los tipos de datos.
- 3 **Altamente coercitivo:** R trata de ajustar tipos de datos automáticamente en sus estructuras de datos.
- 4 **Alta capacidad de broadcasting en estructuras de datos matriciales:** operaciones vectorizadas.

En qué consiste la programación

- 1 Escribir un **programa de ordenador** en un lenguaje determinado consiste en dar una serie de instrucciones, que en su forma más básica son **secuenciales**, que van **operando y transformando** los datos que almacenamos en la memoria del ordenador hasta alcanzar un **resultado** determinado.
- 2 **Herramientas del lenguaje:**
 - 1 **Estructuras de datos:** vectores, matrices, listas, data frames.
 - 2 **Estructuras de control de flujo:** secuenciales, condicionales, iterativas (bucles).
 - 3 **Funciones:** bloques de código reutilizables que realizan pequeñas tareas.