

Capacitación Básica en R

Bienvenidos a una introducción a R...!



Figure 1: Logo de R

Pero... ¿y que es R?

R es un entorno de software libre (licencia GNU GLP) y lenguaje de programación interpretado, es decir, ejecuta las instrucciones directamente, sin una previa compilación del programa a instrucciones en lenguaje máquina.

Este entorno es comúnmente utilizado para la computación estadística y gráfica, ya que dispone de una amplia variedad de técnicas estadísticas (modelos lineales y no lineales, pruebas estadísticas clásicas, análisis de series de tiempo, clasificación, agrupamiento, entre otras.) y gráficas, puede integrarse con distinto origen de datos; además que, existen bibliotecas que facilitan su utilización desde lenguajes de programación interpretados como Perl y Python.

Características de R

Algunas de las características sobresalientes de R, son:

- Manejo y almacenamiento efectivo de los datos.
- Un conjunto de operadores para la realización de cálculos con matrices.

- Una gran colección de herramientas para el análisis de datos.
- Utilidades gráficas para la visualización de datos.
- Un lenguaje de programación bien desarrollado que incluye saltos condicionales, bucles, funciones recursivas, utilidades para la entrada y salida de datos, entre otros.
- Tiene un formato de documentación basado en Latex, que se utiliza para proporcionar documentación completa tanto en formato físico como digital.

¿ Para qué se puede usar?

R es usado principalmente, para la manipulación, procesamiento y visualización gráfica de los datos. R permite:

- Crear visualizaciones de datos de alta calidad.
- Crear dashboards para visualizar y analizar datos.
- Crear informes automáticos.
- Disponer de herramientas de análisis estadístico para ahondar en el conocimiento de los datos.

Aunque R está diseñado para análisis estadístico, con el paso del tiempo los usuarios de este lenguaje han creado extensiones a R, llamadas paquetes, que han ampliado su funcionalidad. En la actualidad es posible realizar en R minería de textos, procesamiento de imagen, visualizaciones interactivas de datos y procesamiento de Big Data, entre muchas otras cosas.

¿En qué fases del análisis de datos se puede usar R?

Se utiliza en todas las fases de análisis de datos:

- *Adquisición de los datos de las fuentes disponibles:* bases de datos, archivos de texto, entre otros.
- *Preparación de los datos:* eliminación de duplicados, datos incorrectos, valores extremos, entre otros.
- *Análisis de los datos:* construcción de modelos predictivos, de clasificación, de agrupamiento...
- *Comunicación de los resultados:* realización de informes para presentación de los resultados y conclusiones.
- *Aplicación de los resultados obtenidos:* por ejemplo, utilización de modelos predictivos desarrollados para en función de una serie de datos históricos (datos de entrenamiento y test del modelo) predecir ciertas salidas.

RStudio

Es recomendable instalar y usar un entorno integrado de desarrollo (IDE, por sus siglas en inglés). Un IDE proporciona herramientas para escribir y revisar código, administrar los archivos que se estén usando, gestionar el entorno de trabajo y algunas otras herramientas de productividad. Tareas que serían difíciles o tediosas de realizar de otro modo, son fáciles a través de un IDE.

Existen varias opciones de IDE para R, y entre ellas mi preferido es R-Studio. Este entorno, además de incorporar las funciones esenciales de una IDE, es desarrollado por un equipo que ha contribuido de manera significativa para lograr que R sea lenguaje de programación más accesible, con un énfasis en la colaboración y la reproducción de los análisis.

Visual Studio Code

La extensión de R para Visual Studio Code admite el resaltado de sintaxis ampliado, la finalización de código, el formateo, la interacción con terminales de R, la visualización de datos, gráficos, variables de espacio de trabajo, páginas de ayuda, la gestión de paquetes y el trabajo con documentos de R Markdown.

Colab

Colaboratory, o “Colab” para abreviar, es un producto de Google Research. Permite a cualquier usuario escribir y ejecutar código arbitrario de Python en el navegador. Es especialmente adecuado para tareas de aprendizaje automático, análisis de datos y educación. Se puede usar también para lenguaje R de forma independiente o combinado con Python.

Referencias

- Unir (2022). Lenguaje R, ¿qué es y por qué es tan usado en Big Data?. Recuperado en agosto de 2022 de: <https://www.unir.net/ingenieria/revista/lenguaje-r-big-data/>
- Bhalla, D. (2016). Companies using R. Recuperado en agosto de 2022 de: <http://www.listendata.com/2016/12/companies-using-r.html>
- Microsoft (2014). Companies using R in 2014. Recuperado en agosto de 2022 de: <http://blog.revolutionanalytics.com/2014/05/companies-using-r-in-2014.html>
- Visual Studio Code. R in Visual Studio Code. Recuperado en agosto de 2022 de: <https://code.visualstudio.com/docs/languages/r>
- Google (2022). Te damos la bienvenida a Colab. Recuperado en agosto de 2022 de: <https://colab.research.google.com/?hl=es>