

Web Scrapping: Datos de tipo de cambio

Yuri Plasencia

Estrategia Regulatoria

28 de agosto de 2021

¿Qué es?

- 1 Técnica para convertir data no estructurada (HTML) disponible en una página web a una data disponible en formato estructurado.

¿Por que usar Web Scraping?

- 1 ¿Es necesario y suficiente extraer información de la web de forma automática o semiautomática?

¿Por que R?

- ① R es libre y fácil de acceder: Descargas, Instalas, Lo usas de donde quieras y cuando quieras.
- ② R tiene millones de usuarios que van creando nuevas cosas y es usado en todas las disciplinas: Existe buena documentación.
- ③ R es open source. Puedes hacer cambios a algunas funciones con esfuerzo pequeño. Con esto el número de paquetes disponibles aumenta todos los días.

¿Por que R?

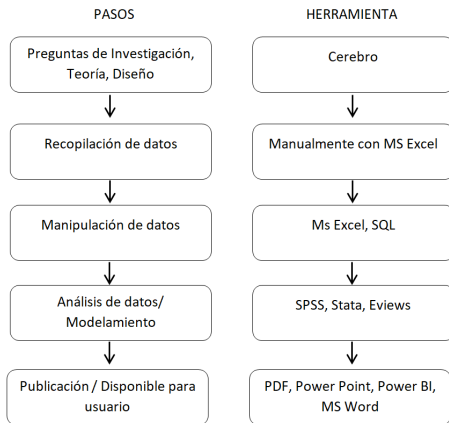


Figura: Proceso de Investigación sin R

¿Por que R?

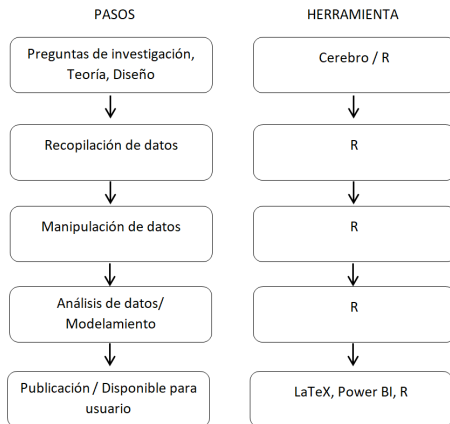


Figura: Proceso de Investigación con R

Usando la librería “rvest”

- 1 librería creada por Hadley Wickham
- 2 Podemos leer desde HTML, para poder acceder con mayor facilidad es bueno agregar una extensión si es que se está trabajando en Chrome: <https://www.analyticsvidhya.com/blog/2017/03/beginners-guide-on-web-scraping-in-r-using-rvest-with-hands-on-knowledge/>

Usando API's

- 1 Usando API's ("Application programming interface"), usamos librerías: jsonlite, ggplot, tidyverse.
- 2 Usar API's en lugar de descargar a mano, la probabilidad de cometer un error disminuye.

Yuri Plasencia

Lima, Perú

+51 961855268

yuplasencia@gmail.com