

Dinamic Documents

Reproducible Research

Estudiante(s): Pedro J. Vargas Barrios

Universidad Distrital Francisco José de Caldas

7 de junio de 2016

"Investigación Reproducible"

Índice

- 1 Contexto
 - Elementos Investigación Reproducible
- 2 Hacia la Investigación Reproducible
 - Etapas destacadas
- 3 Buenas y malas prácticas en Investigación Reproducible
 - Lista de buenas y malas prácticas
- 4 Barreras
 - Lista de obstáculos
- 5 Bibliografía

Contexto

Reproducibilidad - Programación literaria

Investigación Reproducible

Jon Claerbout - Universidad de Stanford - 2009

"El producto final de la investigación no es sólo el papel en sí , sino que también es el entorno computacional completo utilizado para producir los resultados en el documento"

Elementos para garantizar la reproducibilidad

- El artículo o informe técnico.
- El programa de cómputo o paquetería de trabajo.
- El experimento numérico o flujo de trabajo, incluyendo los códigos y la secuencia de instrucciones para generar los resultados.
- Los datos empleados.
- Los resultados del experimento, como figuras y datos.

Etapas destacadas

Hacia la Investigación Reproducible

Etapas destacadas

- (1983) - primera implementación de herramienta llamada WEB para programación literaria, poco adecuada para análisis de datos.
- (1983) - Knuth plantea el paradigma de programación literaria, muy relacionado con la investigación reproducible.
- (1994) - implementación de herramienta Noweb para programación literaria, poco adecuada para análisis de datos.
- (2002) - Sweave fue una de las primeras implementaciones para tratar los documentos dinámicos en R.
- (2009) - Roger Peng publica en revista de Bioestadística criterios de reproducibilidad de su investigación.
- (2015) - Herramienta roxygen2, implementación de R Doxygen
- (2012 -2015) - El paquete knitr se basa en las ideas de las herramientas anteriores con un framework rediseñado.

Recomendaciones

Buenas y malas prácticas en Investigación Reproducible

Lista de buenas y malas prácticas

En el capítulo se menciona que la clave a tener en cuenta para la investigación reproducible es que otras personas puedan ser capaces de reproducir nuestros resultados, por lo tanto, se debe hacer todo lo posible para hacer que la computación sea portable.

- Gestionar todos los archivos de origen en el mismo directorio y utilizar rutas relativas siempre que sea posible.
- No cambiar el directorio de trabajo después de que la ejecución ha comenzado.
- Compilar los documentos en una sesión de R limpia.
- Evitar los comandos que requieren la interacción humana.
- Evitar las variables de entorno para el análisis de datos.
- Adjuntar `sessioninfo()` y las instrucciones de cómo compilar el documento.

Barreras

Barreras

Lista de obstáculos

Se mencionan en el documento algunas barreras prácticas que existen a pesar de todas las ventajas de la Reproducible Research, y se menciona a continuación una lista no exhaustiva de ellas.

- Los datos pueden ser enormes
- Confidencialidad de los datos.
- Versión del software y configuración.
- La competencia.

Bibliografía



Yihui Xie (2013).

Dynamic Documents with R and knitr.

CRC Press 2016. 2 Ed.