Practica 02

Paula Rozo Coy - Pablo Aguilera Capitan

29/2/2020

# Practica 2 - GitHub y Ciencia Reproducible

# OBJETIVO GENERAL

Evidenciar los conocimientos adquiridos relacionados principalmente con el funcionamiento de la libreria R Markdown.

# OBJETIVOS ESPECIFICOS

* Reconocer la importancia de la reproducibilidad de la ciencia y los elementos del flujo de trabajo reproducible a traves de la comprension del articulo **Ciencia reproducible: qué, por qué, cómo**.
* Hacer uso del repositorio GitHub y entender su importancia principalmente en el control de versiones y trabajo integrado por medio del articulo **¿Porqué usar GitHub’ Diez pasos para disfrutar de GitHub y no morir en el intento**.
* Desarrollar ligereza en el uso de los diferentes argumentos de R Markdown para la configuracion de lo que se desea mostrar tanto de texto como de codigo en el documento de salida final.

# DESARROLLO DE LA PRACTICA

Para el desarrollo de la presente practica se utiliza el lenguaje de programación R Core Team (2019), y el documento se genera haciendo uso de la librería R Markdown. Cabe resaltar que tanto la portada se ajusta en el documento final de word previo a exportarlo como PDF.

En primer lugar se crea un documento en word nombrado **docx\_template\_01** en el que se configura todo el estilo del documento de salida incluyendo encabezado y pie de pagina, adicional a esto se incluyen las referencias de los dos articulos propuestos por la guia en el archivo de texto **refs.bib**.

* **Articulo Ciencia reproducible: qué, por qué, cómo.**

En este apartado se dara respuesta a las preguntas planetadas en la guia en relación al articulo en mención.

* ¿ Que es la ciencia reproducible?

La ciencia reproducible hace referencia a la posibilidad de replicar el mismo estudio con nuevos datos, es decir, haciendo uso de los mismos modelos estadisticos, mismos calculos, y obteniendo en teoria la misma cantidad de resultados representados en el mismo numero de tablas y figuras. Sin embargo, hoy en dia la mayoria de los articulos y documentos cientificos no son reproducibles dado que trazar de una manera clara el proceso de obtencion de los resultados resulta bastante complicado si no se cuenta con con un codigo - **Scripts** - que permita la reproducibilidad de los mismos.

* ¿En que casos se consigue 100% de la reprodicibilidad?

Aunque dificilmente se pueda conseguir 100% de la roproducibilidad de un estudio, el texto menciona una serie de elementos que conforma la transicion hacia la reproducibilidad, la utilizacion inicial de un codigo (“scripts”) para el análisis de los datos, repositorios de datos en nube para almacenar, versionar y compratir los datos, la estructuración de un documento dinamico haciendo uso de herramientas como Rmarkdown o Ipython, y finalmenye la incorporacion a un repositorio para conservar un resgistro del desarrollo del proyecto y facilitar la participacion, en este prden respectivamente, puede llegar a ser el flujo dinamico antes mencionado para lograr una plena reproducibilidad de la ciencia. Lo mas complejo en todos estos casos es contar con el manejo de estas herramientas en los diferentes campos de la ciencia.

* ¿Cuales de los beneficios que se sugieren son considerados mas importantes para nosotros como equipo de trabajo?

Creemos que los beneficios enumerados por los autores del articulo estan todos interrelacionados de algun modo, sin embargo, hay dos de estos que consideramos resaltan y denotan la importania de realizar ciencia reproducible, el primero es que **“La utilización de código permite la automatización: ejecución de tareas repetitivas sin esfuerzo”**, este sin duda permite reproducir los mismos calculos y analisis no solo reduciendo el esfuerzo si no asegurando que los cambios que se dan en los resultados por el cambio de metodologias utilizadas incluso con los mismos datos se reduzca completamente, adicional a esto es importante mencionar que la automatizacion de la lectura de datos permite analizarlos incluso cuando no se tiene una comprension del 100% de la metodologia utilizada, lo que hace a la ciencia mas accesible y al alcance de todas las areas de estudio. En segundo lugar se considra dentro de estos beneficios la **“Reducción drástica del riesgo de errores”**, en primer lugar porque auqnue no se garaniza que los resultados obtenidos una vez se replica el estudio sean correctos si evita que se sumen mas errores a los ya existentes, es decir indiferente al numero de veces que se replique el estudio no se existen errores acumulados por otro lado como lo menciona el texto se abren nuevas lineas de colaboración y permite que los erreres existentes tengan la oportunidad de ser corregidos mas rapidamente por alguien que sea aaun mas experto en el tema en cuestión.

* ¿Porque no hace todo el mundo ciencia reproducible?

Como bien lo menciona Rodriguez-Sanchez, Pérez-Luque, Bartomeus, & Varela ([2016](#ref-rodriguez2016ciencia)), desarrollar estudios plenamente reproducibles implica un esfuerzo incial adicional de aprendizaje de diferentes areas de estudio, como lo son el manejo de lenguajes de programación, de bases de datos, de repositorios entre otras cosas, y este esfuerzo implica dedicar un tiempo considerable a aprender diciplinas que a la largar pueden no traer un beneficio economico para quien las aprende y muy por el contrario si requieren de una inversión; aunque generar ciencia reproducible lo vale, tambien es importante considerar que no todo el mundo tiene la misma facilidad de acceso a la herramientas necesarias para llevar a cabo este propósito, de ahi que otra parte fundamental para que la ciencia sea reproducible, sea no solo el conocimiento sino que estas herramientas sean de libre acceso.

* **Articulo ¿ Por qué usar GitHub? Diez pasos para disfrutar de GitHUb y no morir en el intento.**

En este apartado se tratara de comprobar que quedaron claros los conceptos mas importantes asociados al respositorio GitHub y como evidencia del uso y la correcta implementacion del mismo, el script de este documento mas los documentos anexos necesarios para su correcta ejecucion puede descargarse del siguiente link **[**[**https://github.com/PaulaRozo-C/Programacion\_avanzada/tree/master/Practica%2002**](https://github.com/PaulaRozo-C/Programacion_avanzada/tree/master/Practica%2002)**]**.

Por otro lado comentar a importancia de ciertos conceptos con respeto a GitHub:

* Repositorio: Un repositorio es un directorio donde van a estar los archivos que git puede monitorear.. Para crear un repositorio entramos en nuestra cuenta de GitHub y vamos a la pestaña Repositories y clicamos en New. El repositorio puede ser público o privado.
* Rama: Una rama se usará para hacer una prueba experimental. la rama principal es master, pero si queremos trabajar en una nueva, la crearemos a partir de create or find a new branch y la usaremos para experimentar, sin modificar la rama master. Esto va a provocar que si la experimentación ha sido óptima, se añadirá al código principal.
* Pull and push: Una vez que el archivo se ha realizado, para subirlo a la nube de GIT, habrá que darle a *commit/Push*, en cambio para bajarlo de la nube, solo se necesitará darle a *pull*.
* Fork: Los *‘forks’* nos sirven para hacer una copia de repositorios ajenos. Busquemos por GitHub algún repositorio que nos interese. En la parte derecha superior veremos un botón de *Fork*. Tras pulsarlo vamos a tener un repositorio igual a éste en nuestra cuenta de GitHub, sobre el que podemos hacer cuantos cambios queramos.
* **Las opciones de sintaxis básicas de Pandoc Markdown:**
* Texto plano.
* *Cursiva*.
* **Negrita**.
* Superindice3
* ~~Tachado~~
* Link: <https://uah.blackboard.com/ultra/courses/_19456_1/cl/outline>
* Ecuacion en linea de Texto.
* Ecuación en código

x = 2\*3^{2}  
print (x)

## [1] 18

* **Uso de argumentos para la configuración de texto y de código en R Markdown**
* Peguen las palabas “Hola” y “mundo” en una línea nueva:

a <- "hola"  
b <- "mundo"  
c <- paste (a,b)

* Peguen las palabas “Hola” y “mundo” en la misma línea de código: hola mundo

paste (a,b)

## [1] "hola mundo"

* Peguen las palabas “Hola” y “mundo” en una línea nueva de código y que no muestren el resultado de R:

paste ("hola", "mundo")

* Peguen las palabas “Hola” y “mundo” en una línea nueva de código y que no muestren el resultado de R y que no se muestre en el documento de Word:
* **Muestra una tabla y una gráfica en el archivo de salida, tomando alguno de los datos que vienen en R o alguno de sus paquetes.**

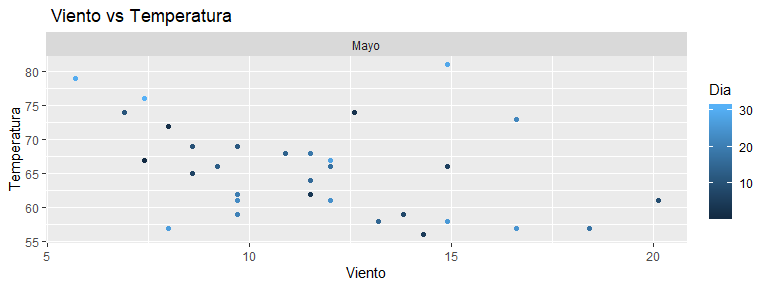
En este apartado se dejara en evidencia lo aprendido en la elaboración de tablas, graficos entre otras cosas haciendo uso de la data **airquality** de R.

Esta data alamcena datos de radiacion solar, tempratura, viento y ozono pra cada dia de los meses de mayo a septiembre. Estos meses estan en el data frame original representados con los numeros del 5 al 9, por eso, en primer lugar se actualiza la tabla asignandole el nombre correspondiente al mes para mejor comprension con la realización de un factor y posteriormente para la creación de la tabla, se hace un resumen de la descripcion de una de las variables que componen la data, en este caso se selecconó la temperatura y se obtiene, la mediana, la media, la minima, la mazima y la desviación estandar y se almacena en la siguiente tabla:

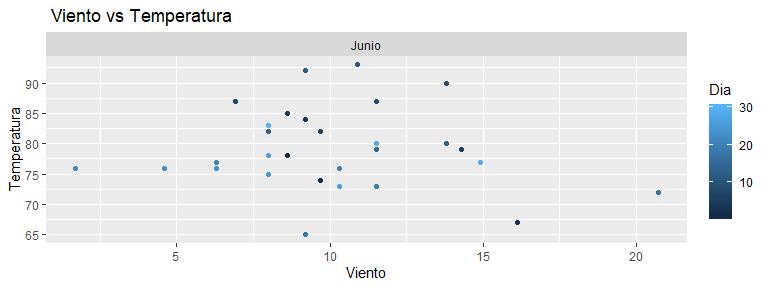
| Month | Mediana | Media | Desviacion\_estandar | Minima | Maximo |
| --- | --- | --- | --- | --- | --- |
| 1.0 | 66.0 | 65.5 | 6.9 | 56.0 | 81.0 |
| 2.0 | 78.0 | 79.1 | 6.6 | 65.0 | 93.0 |
| 3.0 | 84.0 | 83.9 | 4.3 | 73.0 | 92.0 |
| 4.0 | 82.0 | 84.0 | 6.6 | 72.0 | 97.0 |
| 5.0 | 76.0 | 76.9 | 8.4 | 63.0 | 93.0 |

Por otro lado, de este mismo data frame se realizan unos graficos de puntos para ver el comportamiento de la temperatura en relación al viento y determinar si existe correlación entre las dos variables, para esto con el fin de evitar que el codigo sea largo y repetitivo se divide el data frame con la funcion **split.data.frame** y se obiene una lista de dataframes, seguido a esto se crea una funcion que grafique las favirables y finalmente con la funcion **lappy** se obtienen los graficos correpondientes para mes como se puede ver en las siguientes imagenes:

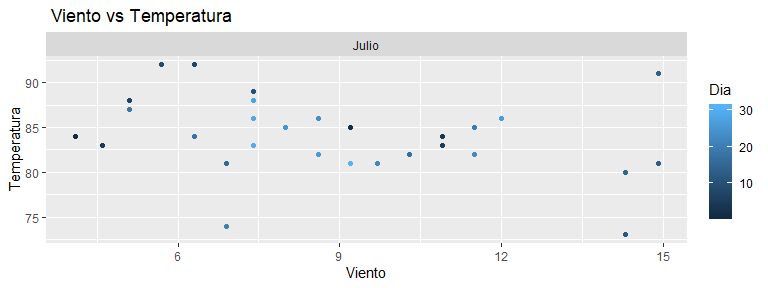
## $Mayo



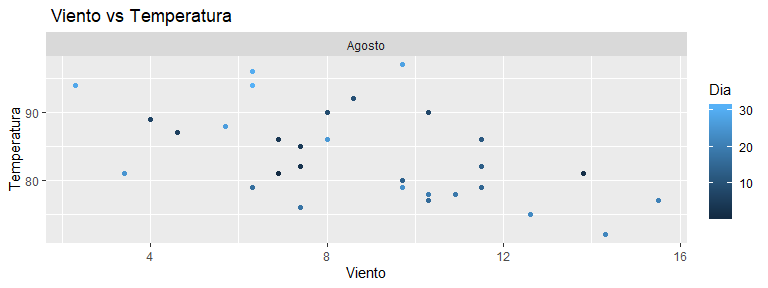
##   
## $Junio



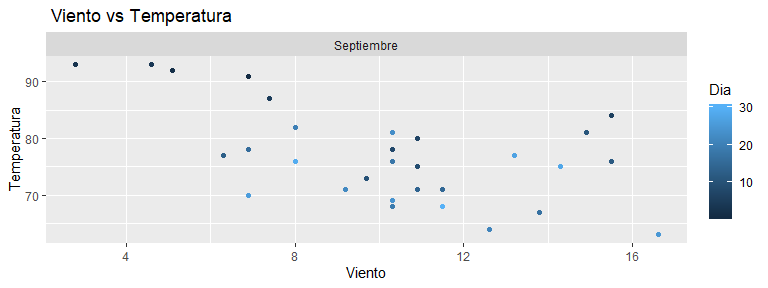
##   
## $Julio



##   
## $Agosto



##   
## $Septiembre



Otra forma de determinar si existe correlación de las variables viento y temperatura es analizar los valores de una menera global, es decir, todos los meses a la vez, obtenemos entonces que la correlación entre las variables es de -0.4579879, lo que nos dice que existe una correlación negativa, es decir a mayor temperatura, menor es el viento.

Rodriguez-Sanchez, F., Pérez-Luque, A. J., Bartomeus, I., & Varela, S. (2016). Ciencia reproducible: Qué, por qué, cómo. *Revista Ecosistemas*, *25*, 83–92.