R: es un lenguaje de programación opensource

leng de programación: la forma de comunicarnos con una computadora

open source: de código abierto. podemos ver el código

github: aquí la gente sube código open source, suben el repositorio. para contribuir al código fuente de R. te dejan un mirror

lenguaje de programación que se definió en 1996… legado de s for statistics, pero este era private software, había que pagar.

en la universidad de Auckland nació en el 2000, tiene 20 en realidad. 29 de febrero de 200

tuvo 4 iteraciones

estamos en la R 4.0.3

palabras reservadas:

<-

data.frame()

function

la computadora entiende unos y ceros

que la computadora interprete estas palabras reservadas para que pueda correr secuencia de instrucciones.

podemos ver el código fuente de R

nodos

impementacion: gnu

gnu r: está escrito en C (lenguaje de programación de menor nivel): es el más popular. en su mayoría está escrito en c y fortran.

riposte: escrito en c++

renjin: otra implementación, hecho en java.

pqr github: está en c en fortran, es una implementacion

Rstudio: es una compañía como tal. Se dedica a hacer software open source.

fundada en 2008.

con lo que nosotros escogemos cómo programar.

rstudio ide: integrated developement ecosystem. Facilita la escritura del codigo. Su repositorio está en github. En JAVA, c++

1. palabras reservadas
2. implementacion: en un lenguaje de programacion menor
3. el ambiente de desarrollo (Rstudio)

sitio web en el que se deposita el codigo fuente de proyectos: github: es la pagina más popular para que te hosten tu open source.

github artic code vault

el algoritmo de bitcoin es open source, linux.

python tambien es open source. esta escrito principalmente en c.

jython en java y en python.

project jupyter github

Buenas practicas de software.

excel tiene 1 millon 485 mil filas, a r le puedes meter mas volumen.

visualizacion

lo fundamental que ganamos con un lenguaje de programacion estructurada es la reproducibilidad del codigo. hay un paper en el cual los investigadores se equivocaron porque promediaron mal en el excel: the reinhart rogof error

el que sea evidente y comunicable, tambien para ti en el futuro.

para que podamos tener algo reproducible necesitamos buenas practicas de software.

lo mas facil es hacer pasta, lo mas dificil es hacer una buena pasta

vicki boykis tiene un articulo que se llama: data science is different now

https://www.dropbox.com/sh/ax00fl4bmtrniqo/AAANAdsx9rw80TamsjZRHbRwa?dl=0

https://theconversation.com/the-reinhart-rogoff-error-or-how-not-to-excel-at-economics-13646 http://veekaybee.github.io/2019/02/13/data-science-is-different/

https://r4ds.had.co.nz/ https://www.tidymodels.org/ https://adv-r.hadley.nz/ https://ggplot2-book.org/

1. creamos una carpeta

tres libros:

R for Data Science

tidymodels

advanced R welcome

ggplot2-book.org

imss tiene una base de datos que se llama cubos

incidencia delictiva mexico

1. guardar el script

alt -: da la flechita : <-

cookiecutter <https://drivendata.github.io/cookiecutter-data-science/>

BUENAS PRACTICAS

1. Hacer una estructura de carpetas

R project: lo que me va a garantizar que mi analisis sea reproducible

getwd()

un path relativo a la carpeta de proyectos, lo dijo jenny bryan: project oriented workflow

<https://www.tidyverse.org/blog/2017/12/workflow-vs-script/>

sharla gelfand

las librerias exitienden la funcionalidad basica de R.

no integran las librerias al lenguaje de programacion base porque puede “romperse algo”.

es invasivo que te carguen paquetes que no vas a usar, te ocupan memorias.

library(tidyverse)

se necesita un comentario y 4 caracteres

#lectura ====

#lectura ----

**15 diciembre 2020**

librería: conjunto de codigo que extiende las capacidades basicas del lenguaje de programacion

API : forma en la que nos comunicamos con el lenguaje de programación, forma mediante la cual le pedimos que haga cosas.

Nos vamos a enfocar en API de tydiverse, es moderna, y es facil hacer calculos

Hadley Wickham: el mejor en ciencia de datos ahora, explica como poner el workshop para que no te lo esté guardando

tydiverse viene de sql. SQL: es un lenguaje de programación mas antiguo, principalmente para hacer selección de datos.

