



De Hojas de Cálculo a Scripts: Razones para adoptar R sobre Excel

Expositora: Sherly Tarazona

14/12/2023



Cuéntanos algo sobre ti

- ¿De qué ciudad te conectas?
- ¿Para qué tipo de tareas usas excel?
- ¿Has usado R o es la primera vez?

Duración: 3 minutos

Caso: Ana



Ana es analista de datos en una cadena de colegios. Suele usar Excel en su trabajo para realizar informes para diferentes colegios, cada vez nota que maneja más información haciendo que su archivo Excel demore en abrir. El otro día, tuvo que ausentarse y pidió a un compañero que replique su proceso, pero no le fue bien porque su compañero no pudo rastrear sus fórmulas y tareas. A veces le pasa que el mismo cuadro tiene que recrear para los 20 colegios y ponerlo en una presentación. También tiene que actualizar de forma diaria sus informes y realiza muchos cálculos manuales. Para el 2024 le han pedido profundizar más en sus análisis como obtener pronósticos, pruebas de hipótesis.

¿Qué harían ustedes?



Herramientas que podría explorar:

- Excel con complementos

Open source

- R

Open source

- Python
- Power BI
- Tableau
- SQL



Artwork by Allison Horst

A ginger and white cat is shown from the chest up, looking upwards with its mouth slightly open. Its right front paw is raised towards its face. A white speech bubble with a blue outline is positioned to the right of the cat's head, containing the text '¿En qué casos usar R?'. The background is a solid dark brown.

**¿En qué
casos usar R?**

#rcatladies

1. Si manejas gran volumen de datos

Excel solo soporta 1,048,575 filas. Si procesas mayor volumen presentarás problemas de lentitud en la carga del archivo.

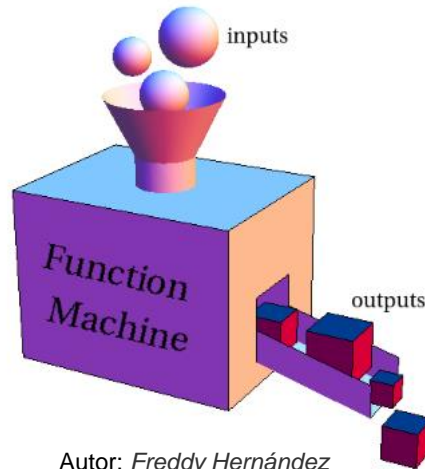
```
# Caso 1
{r}
#Leer

data<-read.csv("01volumen/data_ejemplo.csv", sep=";")
...
```

Environment	History	Connections	Tutorial
Import Dataset ▾ 354 MiB ▾			
R ▾ Global Environment ▾			
Data			
data		1048575 obs. of 5 variables	

2. Tareas repetitivas

Si necesitas realizar la misma tarea más de “n” veces, quiere decir que debe ser automatizable.



En R puedes crear funciones

```
{r}  
suma <- function(x, y) {  
  resultado <- x + y  
  return(resultado)  
}  
suma(5,2)
```

```
[1] 7
```


3. Deseas tener la trazabilidad del proceso

Un script es un conjunto de instrucciones, donde puedes hacer comentarios para facilitar la comprensión.

Es como una receta de cocina.

```

---
title: "Meetup"
author: "Sherry Tarazona"
date: "2023-12-14"
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)

Caso 1: Gran volumen de datos (>1Millón)
```{r}
#Leer

data<-read.csv("01volumen/data_ejemplo.csv", sep=";")

...

# Caso 2: Gráficos repetidos
```{r}
#Cargar
df_resumen<-read.csv("02iterativo/df_resumen_2.csv", header=T, sep=";")
head(df_resumen,10)
...
```{r, echo=FALSE}
library(ggplot2)
library(dplyr)
library(ggpubr)
library(gridExtra)
library(grid)
library(knitr)
library(scales)
library(gtable)

# Function - table
tb_bivariate<-function(data,variable){

  df_tb<-data %>%
    select(tramo,splits, records_count, default_rate, summary_name,feature,num) %>%
    filter(summary_name==variable & splits!="Special") %>%
    mutate(default_rate=as.numeric(default_rate))

  df_tb2<- df_tb %>%
    mutate(TasaMalos=scales::percent(df_tb$default_rate)) %>%
    mutate('Categoría' =df_tb$splits) %>%
    mutate(Tramo=df_tb$tramo) %>%
    mutate(Cantidad=scales::comma(df_tb$records_count)) %>%
    arrange(num) %>%
    select(Tramo,'Categoría',Cantidad,TasaMalos)
    #write_csv("15, Numer 1C, Round, 3)
  df_tb2
}

```

4. Si tienes fórmulas extensas

Si la fórmula es extensa en una celda puede prestarse al error. Más aún si depende de otras columnas y parámetros.

En R puedes apoyarte en objetos como vectores para reducir tareas manuales.

```
## Puntos de corte
vec_cortes_oro<-c(0.082,0.296,0.715,0.923)
vec_cortes_plata<-c(0.0648,0.295,0.526,0.804)
vec_cortes_bronce<-c(0.069,0.132,0.543)

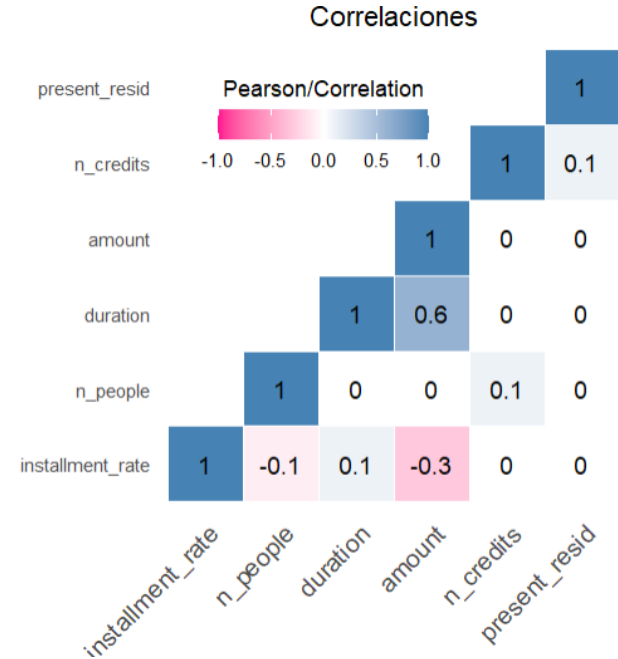
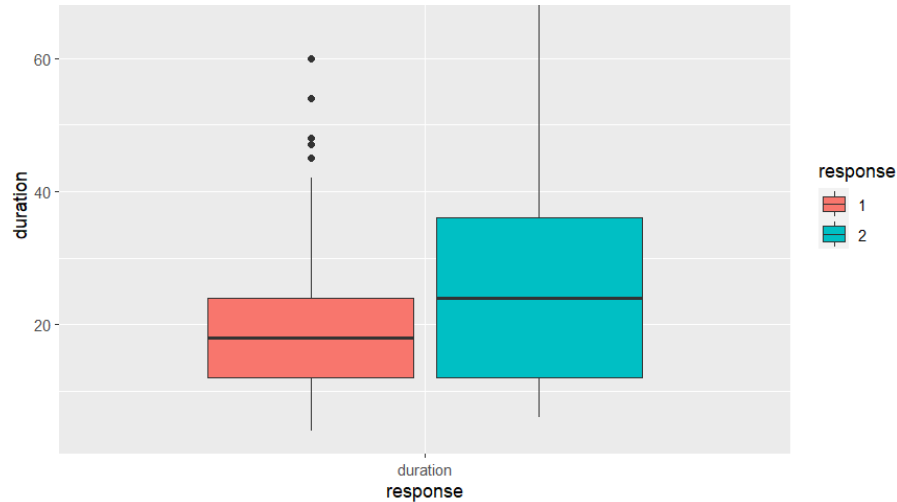
tb_resumen_3<-data_3 %>%
  select(Probabilidad,Tipo_cliente,Monto_deuda) %>%
  mutate(
    perfil_deuda=
      ifelse(Tipo_cliente=='Oro' & Probabilidad<=vec_cortes_oro[1],"A",
            ifelse(Tipo_cliente=='Oro' & Probabilidad<=vec_cortes_oro[2],"B",
            ifelse(Tipo_cliente=='Oro' & Probabilidad<=vec_cortes_oro[3],"C",
            ifelse(Tipo_cliente=='Oro' & Probabilidad<=vec_cortes_oro[4],"D",
            ifelse(Tipo_cliente=='Oro' & Probabilidad>vec_cortes_oro[4],"E",

            ifelse(Tipo_cliente=='Plata' & Probabilidad<=vec_cortes_plata[1],"A",
            ifelse(Tipo_cliente=='Plata' & Probabilidad<=vec_cortes_plata[2],"B",
            ifelse(Tipo_cliente=='Plata' & Probabilidad<=vec_cortes_plata[3],"C",
            ifelse(Tipo_cliente=='Plata' & Probabilidad<=vec_cortes_plata[4],"D",
            ifelse(Tipo_cliente=='Plata' & Probabilidad>vec_cortes_plata[4],"E",

            ifelse(Tipo_cliente=='Bronce' & Probabilidad<=vec_cortes_bronce[1],"A",
            ifelse(Tipo_cliente=='Bronce' & Probabilidad<=vec_cortes_bronce[2],"B",
            ifelse(Tipo_cliente=='Bronce' & Probabilidad<=vec_cortes_bronce[3],"C",
            ifelse(Tipo_cliente=='Bronce' & Probabilidad>vec_cortes_bronce[3],"D","Otro"
            )))))))) %>%
  group_by(Tipo_cliente) %>%
  summarise(Prob_promedio=mean(Probabilidad)
            ,sum_mto_deuda=sum(Monto_deuda)
            ,conteo=n()) %>%
  arrange(Prob_promedio)
```

5. Deseas hacer análisis estadístico

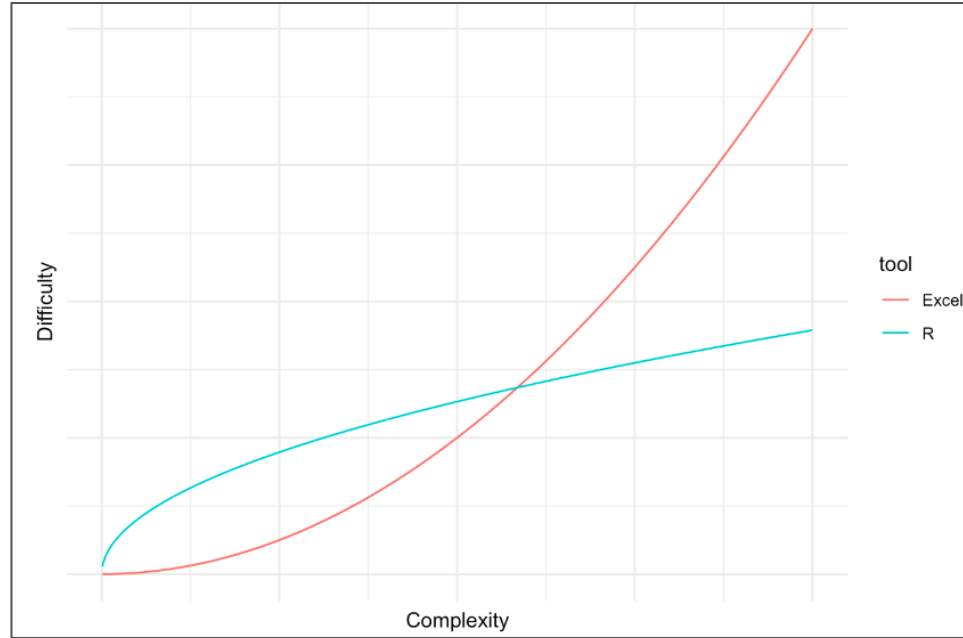
R está diseñado para el análisis estadístico.





¿Por dónde empiezo?

Aprendizaje: Excel vs. R



Autor: Gordon Shotwell

Al principio, cuando intentas lograr cosas simples, R es definitivamente más difícil de aprender que Excel. Sin embargo, a medida que la tarea se vuelve más compleja, se vuelve más fácil de realizar en R que en Excel. Esto no quiere decir que no puedas resolver muchos problemas complejos con Excel, es sólo que la herramienta no te lo pondrá fácil.

Recursos gratuitos

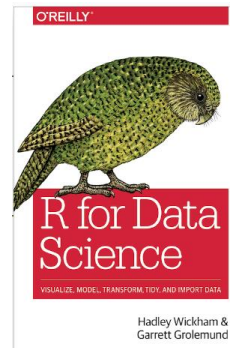


<https://rstudio-conf-2020.github.io/r-for-excel/>

Introducing RStudio and R Markdown

Gettin' giffy wit it.

<https://www.pipinghotdata.com/posts/2020-09-07-introducing-the-rstudio-ide-and-r-markdown/>



<https://es.r4ds.hadley.nz/>

STAT 545

Data wrangling, exploration, and analysis with R

Jenny Bryan

The STAT 545 TAs

<https://stat545.com/>



<https://fhernanb.github.io/Manual-de-R/>



The Turing Way

<https://the-turing-way.netlify.app/index.html>

Adopta buenas prácticas de trabajo



Objetivo:

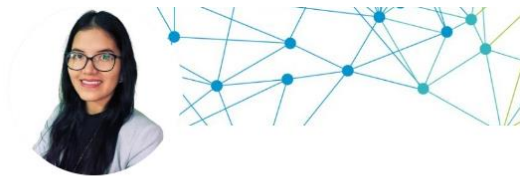
- Reproducibles
- Reusables

Fig. 1 The Turing Way project illustration by Scriberia. Zenodo.

Recomendaciones

- No te desanimes si no te sale al inicio, recuerda que todo se aprende con la práctica y constancia.
- Ponlo en práctica. Lleva algunas tareas de Excel a R de a pocos.
- Uno nunca aprende a conocer todo de un lenguaje, siempre hay cosas nuevas que salen.
- Participa en comunidades, es bonito aprender en grupo y conocer a personas que te inspiren a seguir aprendiendo.

Gracias



Sherly Tarazona Tocco

Advanced Analytics Lead | Co-organizer at R-Ladies Lima | Social Impact



Sherly Tarazona

<https://linktr.ee/sherlytarazona>