



UNIVERSIDAD
DE MÁLAGA



UNIVERSIDAD
DE
CÓRDOBA

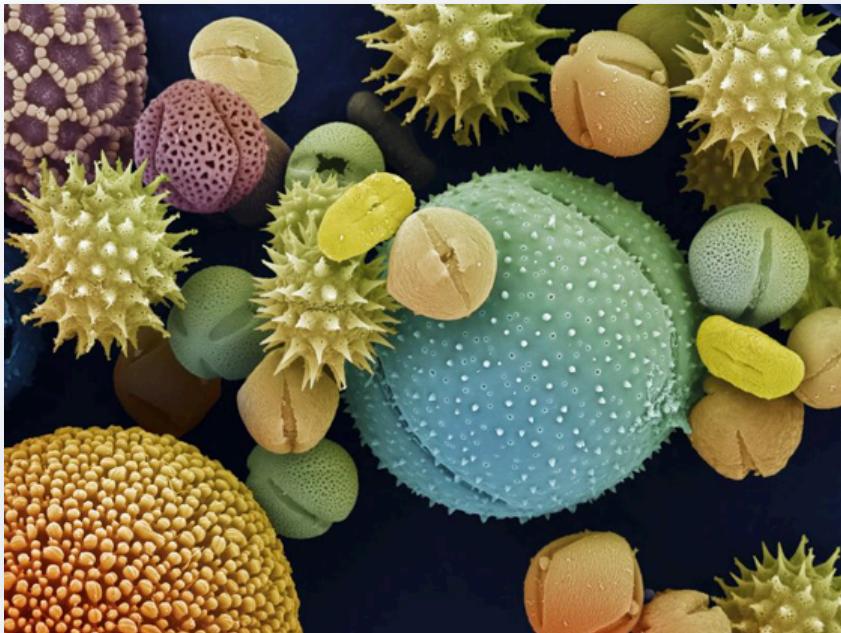
AeRobiology

**Análisis de datos biológicos
en el aire con R**

Jesús Rojo, Antonio Picornell, Jose Oteros

aerobiology.package@gmail.com

AEROBIOLOGÍA



Shakun, 2018. *Nature*

- Estudio del contenido en el aire de organismos o partículas de origen biológico
- Interés desde diferentes ámbitos (salud pública, agronomía, medio ambiente...)

La **gran cantidad de datos** registrados durante décadas y el creciente avance de **muestreadores automáticos** requieren el uso de nuevas **herramientas de análisis de mayor eficiencia** que permitan la automatización de tareas.

AeRobiology
cc) J. Rojo, A. Picornell & J. Oteros



Procesamiento

Raw data

Data processing

Interpretation

Results

Publication

AeRobiology

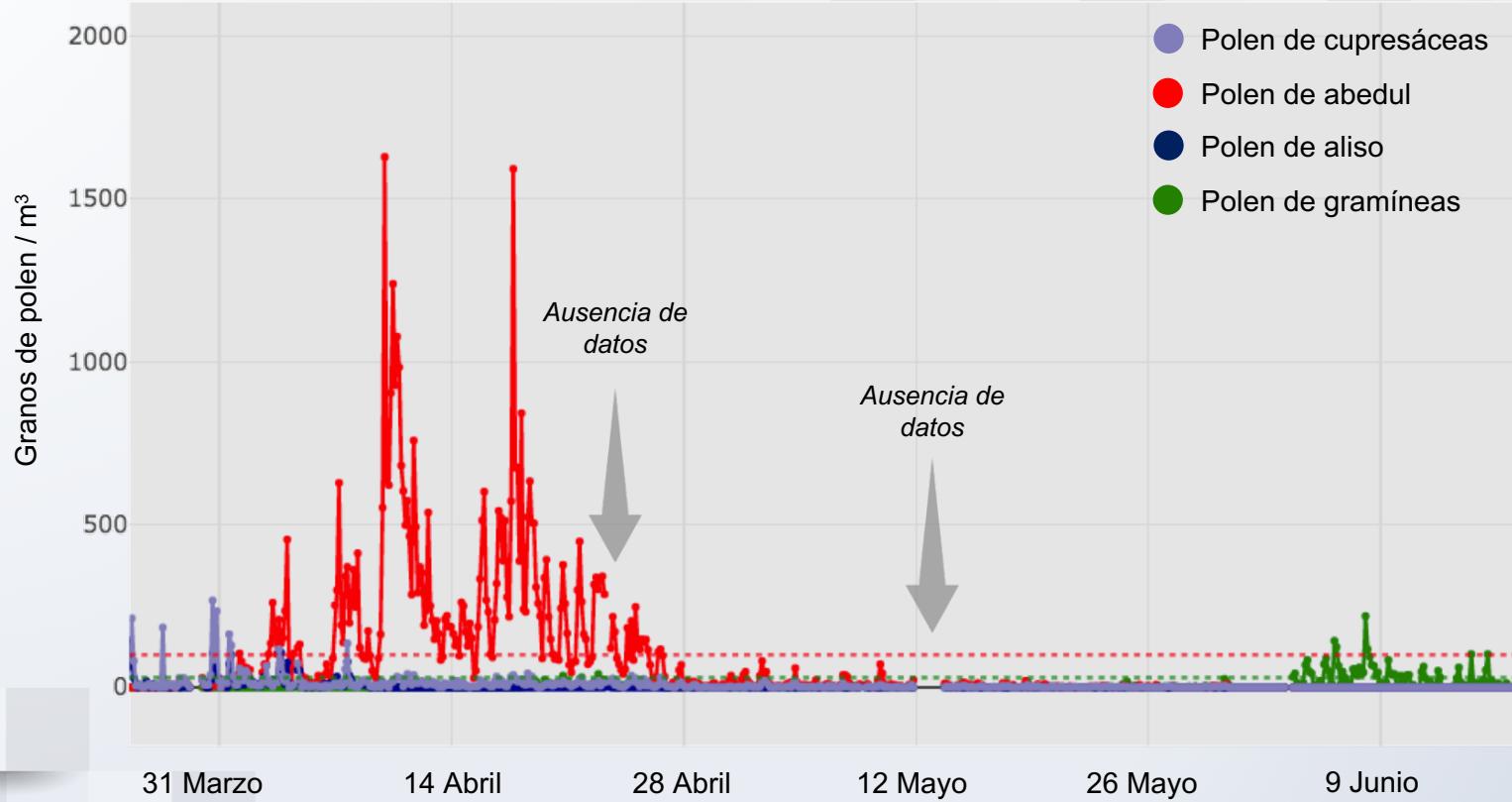


Visualización

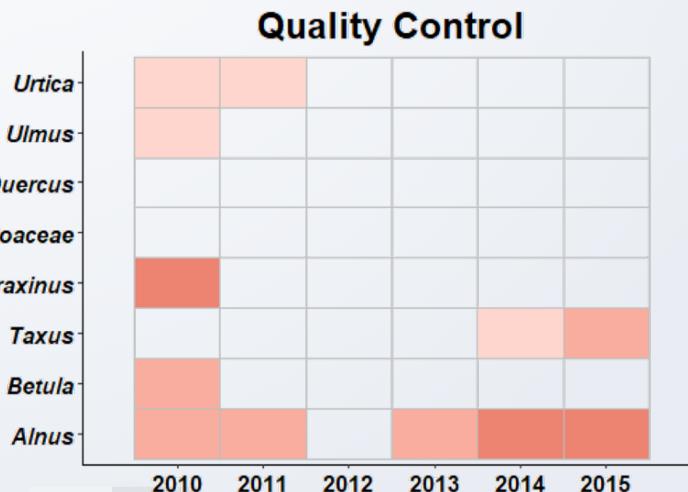
1.

PROCESAMIENTO DE DATOS

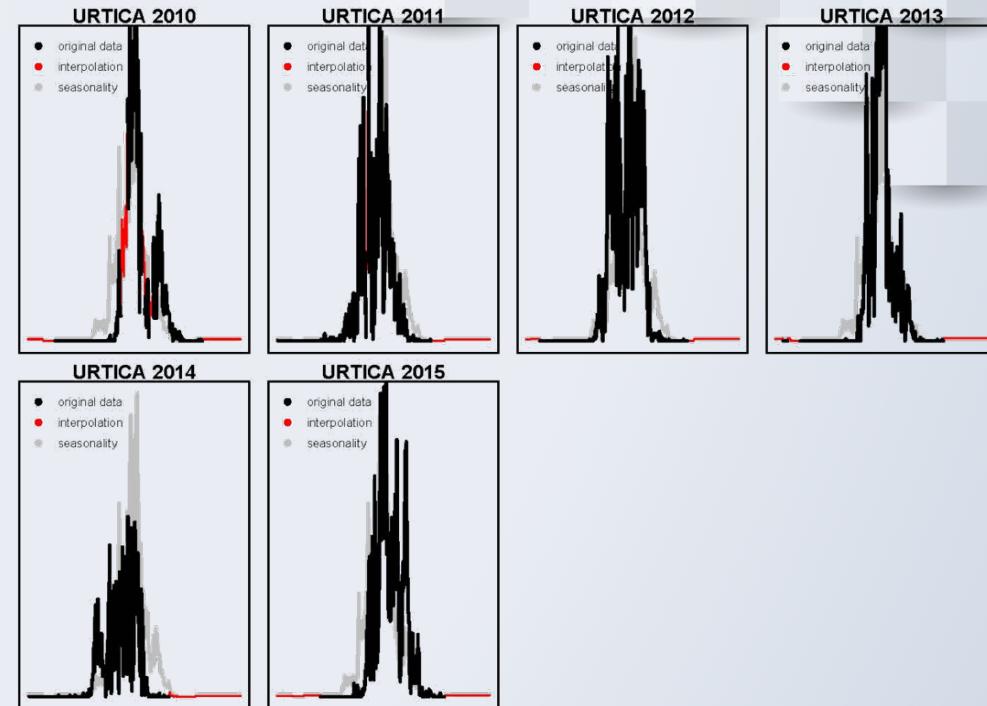
Funciones para evaluar la calidad de los
datos



quality_control()



Interpollen()

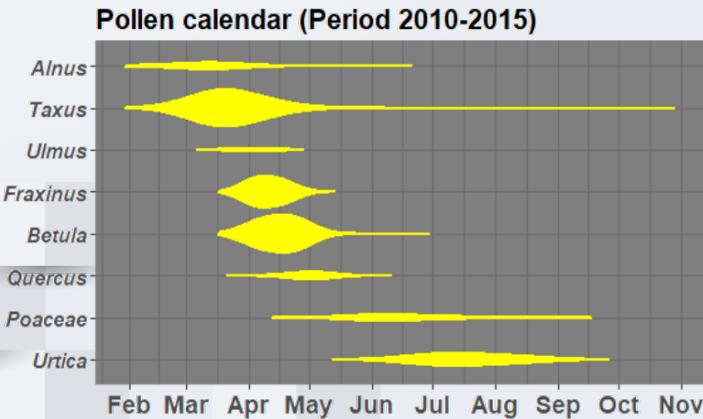
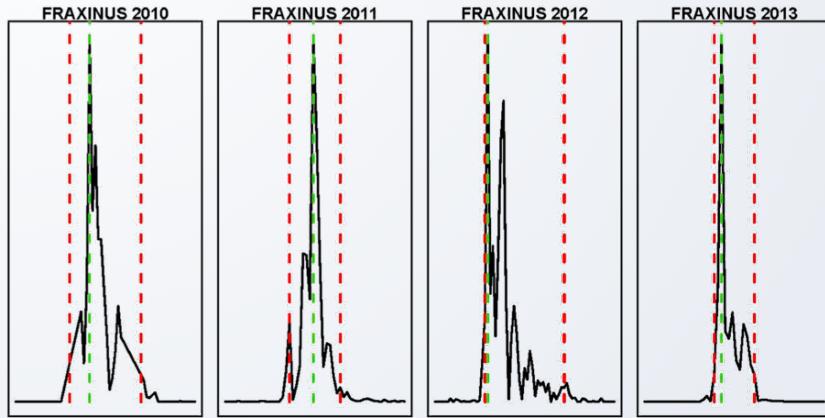


2.

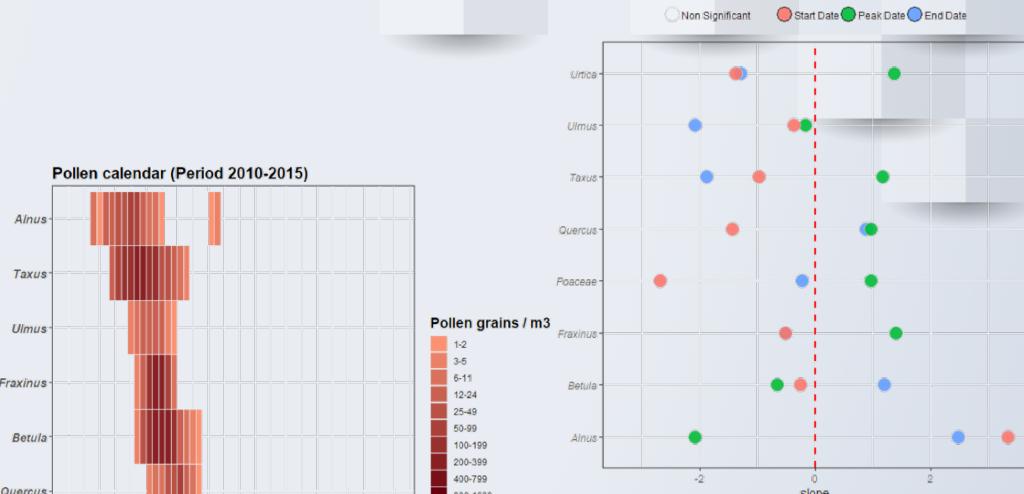
ANÁLISIS AEROBIOLÓGICO

Funciones para analizar los datos y
realizar los principales cálculos

calculate_ps()



analyse_trend()



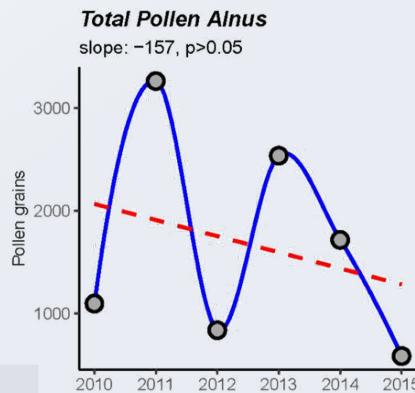
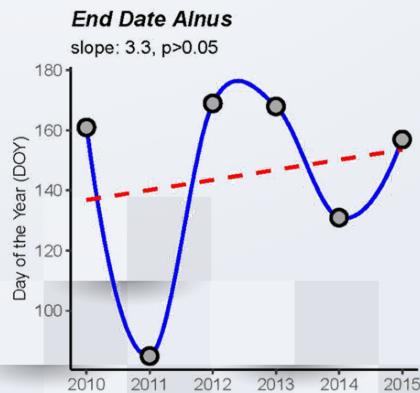
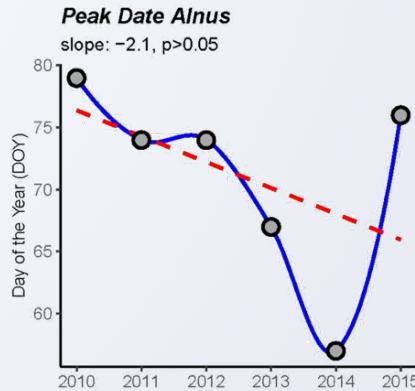
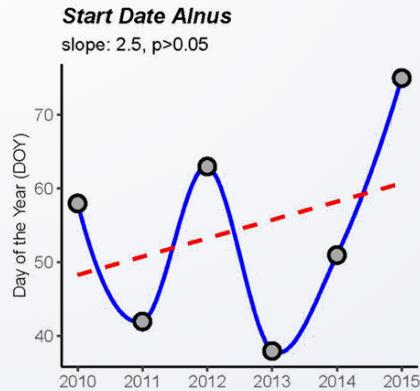
pollen_calendar()

3.

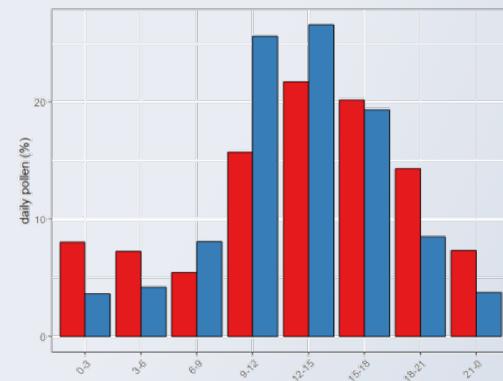
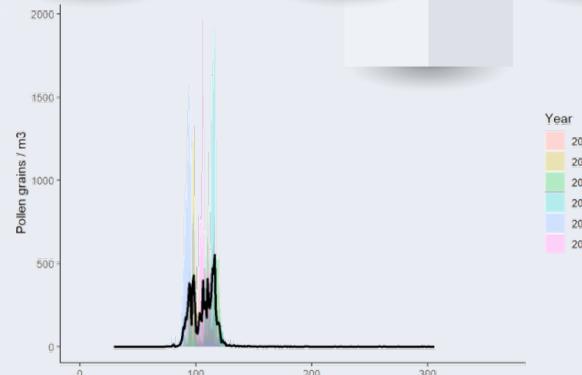
GRÁFICOS Y VISUALIZACIONES

Funciones especializadas en la
visualización de los resultados

plot_trend()

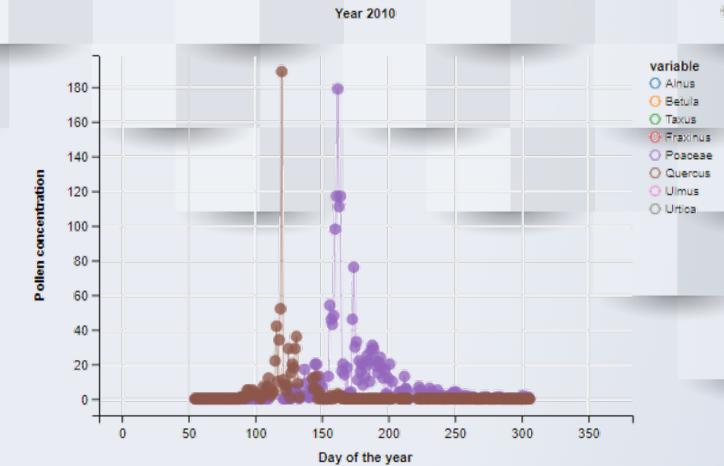
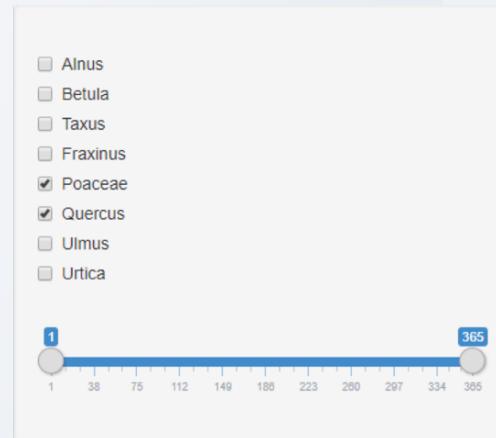


plot_summary()

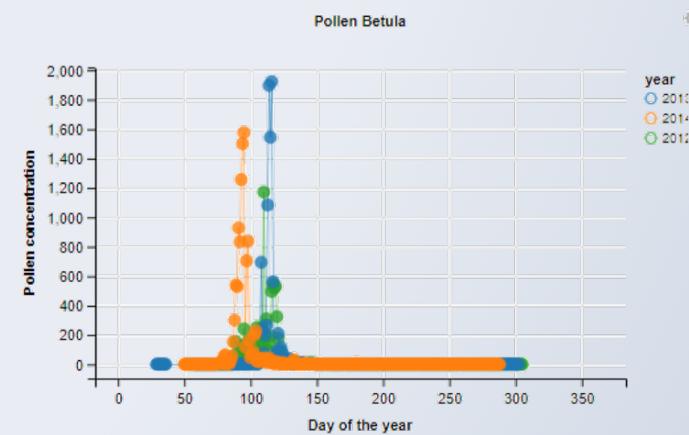
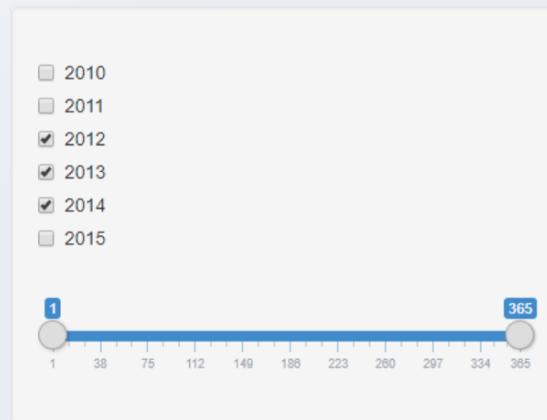


plot_hour()

iplot_pollen()



iplot_years()



CONTRIBUCIÓN CIENTÍFICA

DOI: 10.1111/2041-210X.13203

Methods Ecol Evol.
2019; 10: 1371–1376.

Factor de impacto (2018):
7,090 (Primer Decil)

Methods in Ecology and Evolution



AeRobiology: The computational tool for biological data in the air

Jesús Rojo^{1,2} | Antonio Picornell³ | Jose Oteros²

¹Institute of Environmental Sciences (Botany), University of Castilla-La Mancha, Toledo, Spain

²Center of Allergy & Environment (ZAUM), Member of the German Center for Lung Research (DZL), Technical University of Munich/Helmholtz Center, Munich, Germany

³Department of Plant Biology, University of Malaga, Malaga, Spain

Correspondence
Jesús Rojo
Email: jesus.rojo.ubeda@gmail.com

Funding information
University of Castilla-La Mancha [UCLM]; University of Malaga [UMA]; Zentrum Allergie und Umwelt [ZAUM]; Ministry of Education, Culture and Sport of Spain, Grant/Award Number: FPU15/01668; Helmholtz Zentrum Munich of Germany, Grant/Award Number: PFP II 2018-2020

Handling Editor: Samantha Price

Abstract

1. Aerobiological databases are constantly increasing. Many of them contain long and extensive time series of data which are very difficult and tedious to manage.
2. The development of new real-time automatic sampling devices also requires new tools to reduce time of calculations and data management. In this sense, the AeRobiology R package has been implemented to accelerate and facilitate these tasks.
3. This package was structured in three sections based on (a) the checking of the database, (b) calculation of the main aerobiological indexes and (c) visualization of the results.
4. The AeRobiology package contains numerous functions which, in conjunction, solve the main general tasks that scientists must assume for the analysis of the biological data.
5. The package is freely distributed under GNU General Public License and can directly be installed from CRAN (<http://cran.r-project.org/>). The reference manual is available at <https://cran.r-project.org/web/packages/AeRobiology>. Contact: aero-biology.package@gmail.com.

KEY WORDS

aerobiology, automation, bioaerosols, bioinformatics, computational biology, pollen diversity, R programming

REPERCUSIÓN

Cursos especializados, tutoriales,
vignettes, trípticos...



Visualise and manage your aerobiological data

Elaborate your graphs instantly and customize your results in real time with interactive plots.

Do you have any ideas or improvements for the next version?

Please, contact us: aerobiology_pkdckag@gmail.com

AeRobiology ©J. Rojo, A. Picornell & J. Oteros

A Computational Tool for Aerobiological Data

Phenological parameters

See this code for more information

+UCLM UNIVERSIDAD DE MÁLAGA CAUM

8 seasons "Aerobiology" based on R software

AeRobiology package for R

AeRobiology v. 1.0.0: A computational Tool for Aerobiological Data

The new R package developed by Jesus Rojo, Antonio Picornell and Jose Oteros. This is a very useful tool for managing aerobiological databases, elaborating the main calculations and visualization of results. In a first step, data are checked using tools for quality control and all missing gaps are completed. Then, the main parameters of the pollen season are calculated and represented graphically. Multiple graphical tools are available: pollen calendars, phenological plots, time series, tendencies, interactive plots, abundance plots and many more. Full description available in:

<https://cran.r-project.org/web/packages/AeRobiology/AeRobiology.pdf>
<https://www.rdocumentation.org/packages/AeRobiology/versions/1.0.0>



Spanish Association of Aerobiology

and the University of Castilla-La Mancha

IV SEMINAR ADVANCES IN AEROBIOLOGY

AeRobiology: a new R package to deal with aerobiological data

teachers

Antonio Picornell (University of Malaga)

Jesus Rojo (University of Castilla-La Mancha & ZAUM)

Jose Oteros (Zentrum Allergie und Umwelt, ZAUM)

MARCH 14, 2019 | 3:00 - 7:00 PM
TOLEDO (SPAIN)
UNIVERSITY OF CASTILLA-LA MANCHA

Two options of assistance

Classroom: University of Castilla-La Mancha
Avda. Carlos III s/n (Toledo, Spain)

Webinar by Skype



Registration & Information

Contact & Registration:
secretaria@aerobiologia.org
Deadline: 7th March
Members of AEA and IAA: Free

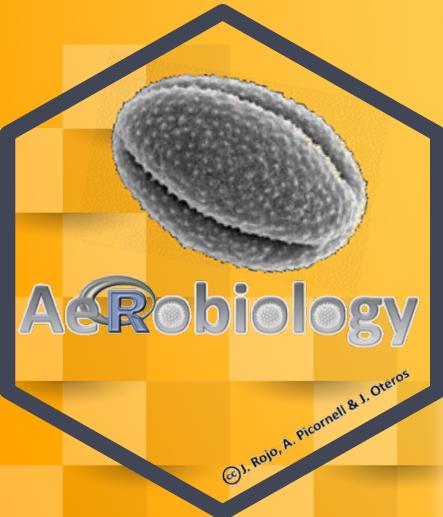




R Studio®



GRACIAS!



aerobiology.package@gmail.com



UNIVERSIDAD
DE MÁLAGA



UNIVERSIDAD
DE
CÓRDOBA

Reducción dimensional, clustering y visualización interactiva con looking4clusters: aplicación al análisis de expresión de célula única

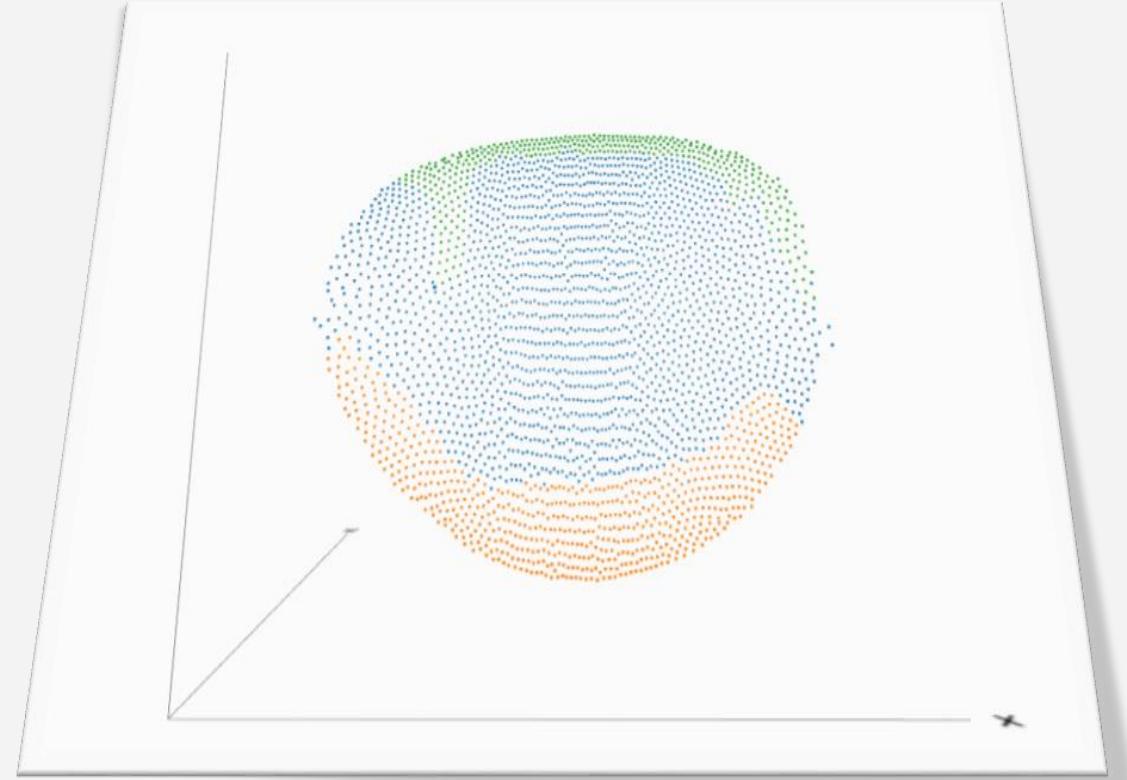


Ángela Villaverde-Ramiro (Servicio de Bioinformática, Nucleus, Universidad de Salamanca)
David Barrios (Servicio de Bioinformática, Nucleus, Universidad de Salamanca)
Carlos Prieto (Servicio de Bioinformática, Nucleus, Universidad de Salamanca)

Los estudios de **Single Cell** consisten en la secuenciación del ARN de cada célula de manera individual.

Aplicaciones:

- ✓ Identificación de tipos celulares.
- ✓ Determinación de la etapa celular.
- ✓ Resistencia a medicamentos.
- ✓ Diferenciación de células cancerígenas.
- ✓ ...



Uno de los retos en los estudios de Single Cell es la **clasificación**, objetivo de esta herramienta.

Son necesarios dos tipos de técnicas:

Reducción de dimensiones

- Representación bidimensional de los datos.

Métodos de clustering

- Visualización de los posibles agrupamientos no supervisados.

Otros ámbitos de aplicación:

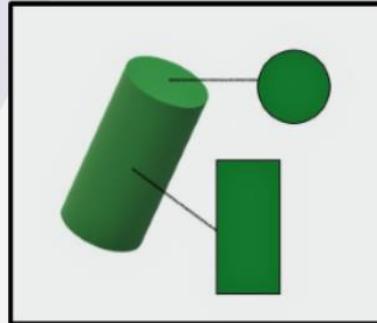
- *Marketing*: Patrones de consumo
- *Medicina*: Evaluación de pacientes
- *Política*: Intención de voto
- *Biología*: Caracterización de especies
- *Sociología*: Estratificación de población
- ...



Datos

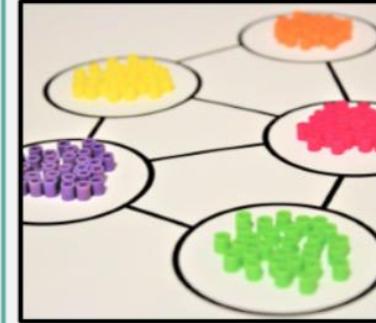


Reducción de Dimensiones



- PCA
- t-SNE
- MDS
- UMAP
- NMF

Técnicas de Cluster



- k-means
- PAM
- Jerárquico

Visualización



- D3
- JavaScript
- SVG

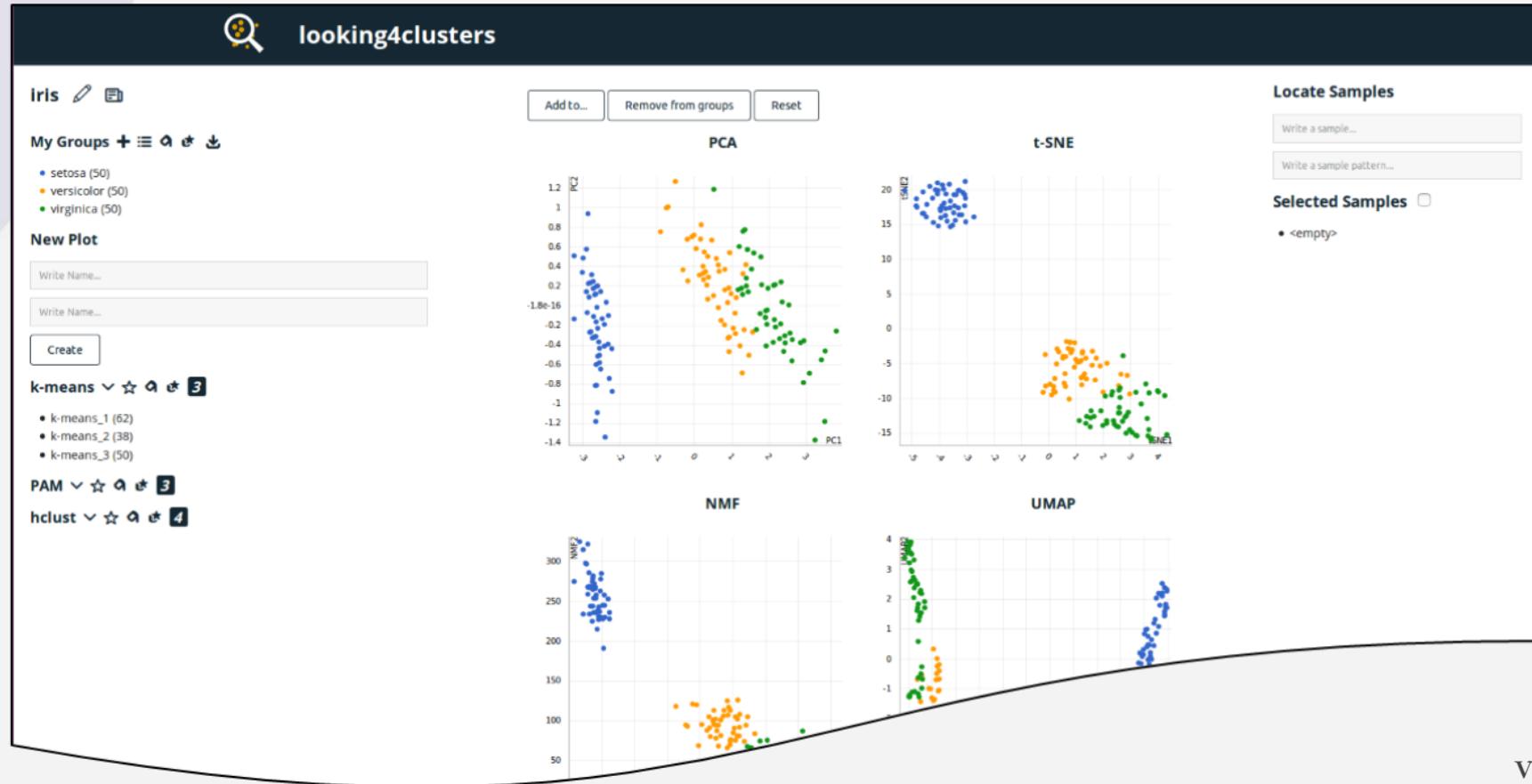
Usuario



Looking4clusters



```
library(looking4clusters)  
getClusters(iris[,1:4], groups=iris[,5], directory="iris")
```



Servicio de Bioinformática:

Responsable: Carlos Prieto

Técnico en desarrollo de aplicaciones: David Barrios

Técnico en análisis de datos: Ángela Villaverde

Contacto: angelavr@usal.es

¡ GRACIAS POR SU ATENCIÓN !





XI JORNADAS DE USUARIOS DE R

MADRID, 14-16 noviembre 2019

Control de la Producción en fábricas de queso del grupo Lactalis

MARTA ALONSO

INES GONZALEZ

PABLO DOMINGUEZ

CARLOS MENENDEZ

MIGUEL ANGEL MORENO



Uso de R en Lactalis división quesos España

- Herramienta de análisis:

- Estandarización manejo y preparación de datos → tidyverse
- Seguimiento variables técnicas (KPI)
 - Gráficos de control → qicharts2, qcc
 - Gráficos generales de presentación → ggplot2
 - Análisis de series básico → time series, decompose(), plot()
 - Forecast → fbprophet
 - Visualización: Plotly → ggplotly()
- Análisis estadísticos básicos

- Herramienta de comunicación:

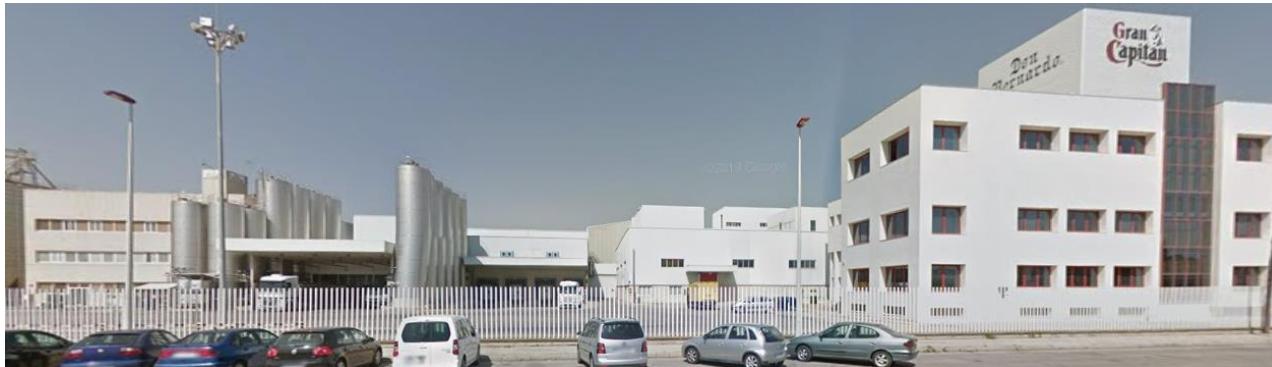
- Gráficos básicos: plot(), boxplot()
- ggplot2
- Heatmaps
- Reporting con Rmarkdown
- Integración con Microsoft PowerBI



Nuestros datos:

- Recogidos en base de datos de variables clave por cuba de fabricación de queso
- Producción de aprox. 30 cubas diarias sobre las que se miden más de 80 parámetros a lo largo de la fabricación
- En un mes, casi 1.000 líneas a analizar → tampoco es realmente “BIG” data ☺
- En una cultura “Excel” es posible hacerlo con tablas dinámicas, pero R proporciona mucha más flexibilidad y mejores herramientas gráficas
- La dificultad de uso de R hace que a veces hasta el equipo técnico tiene ansias de “libertad” ... aunque ahí seguimos

R que R



Resumen de las ventajas / inconvenientes R:



Competencia estratégica clave de los ingenieros y técnicos (competitividad en el mercado laboral)

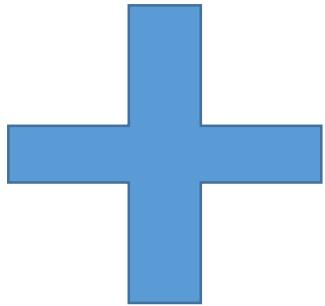
Extensión de su uso en universidad/escuelas técnicas

Lenguaje común, compartido, disponibilidad de recursos

Falta de método en el almacenamiento de información técnica

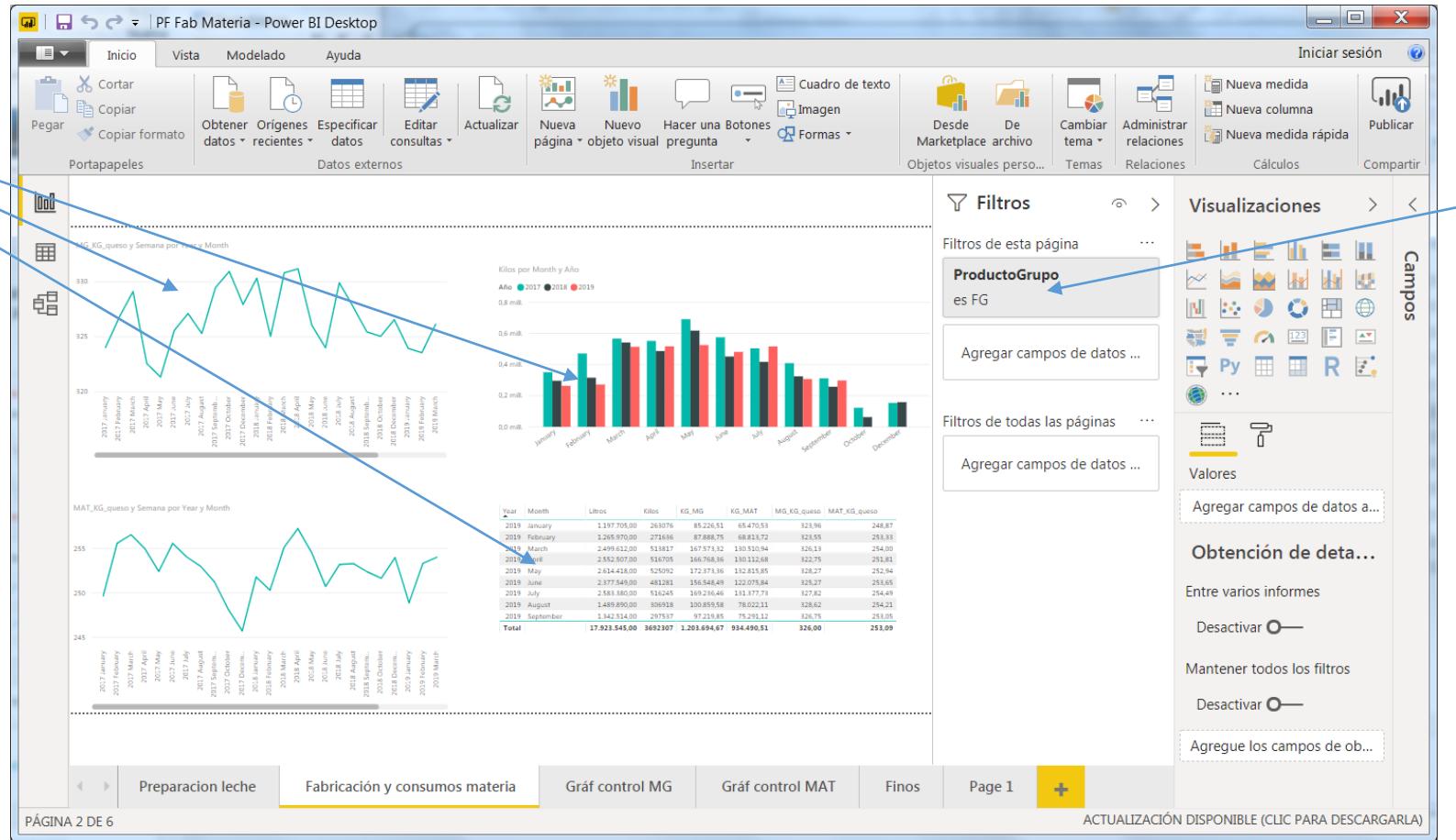
Base teórica (programación, estadística)

Formación, curva de aprendizaje



Power BI. Ejemplo de pantalla

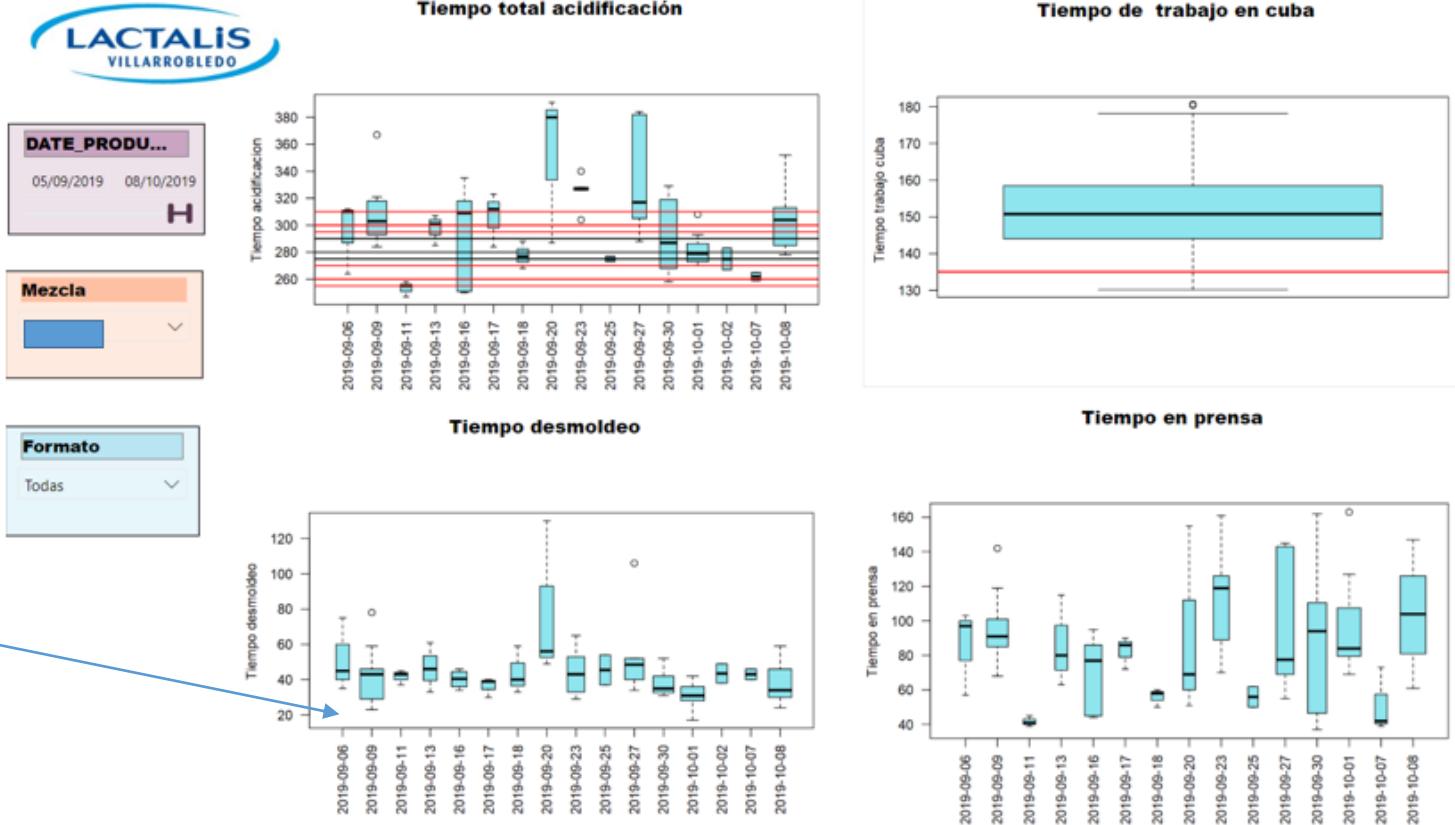
Power BI permite una gama muy grande de visualizaciones de gran calidad para las que no se necesita R



El uso de filtros es muy visual y sencillo para las personas no formadas en R

Uso de boxplot (1)

El uso de R con Power BI permite que personas no formadas en R puedan utilizar con facilidad elementos de R (presentaciones o análisis)



Uso de qicharts2 (1)

The screenshot shows the Power BI Desktop interface with two charts displayed. The left chart is titled 'Fechafab y MG_KG_Quotient' and the right chart is titled 'MR Chart of MG_KG_Quotient'. Both charts show a time series with 'Value' on the Y-axis and 'Subgroup' on the X-axis. The right chart also includes a 'Producto' field with a dropdown menu. On the right side of the screen, there is a 'Visualizaciones' pane showing various chart icons, a 'Filters' section, and a 'Valores' section where 'FechaFab' and 'MG_KG' are listed. Below these are sections for 'Obtención de datos', 'Entre varios informes', 'Desactivar', 'Mantener todos los filtros', and 'Activar'. At the bottom, there is an 'Editor de script R' containing the following code:

```
1 library(qicharts2)
2 dataset$FechaFab<- as.Date(dataset$FechaFab)
3 qic(FechaFab, MG_KG) = dataset, chart = "mr", decimals = 2, point.size=3)
```

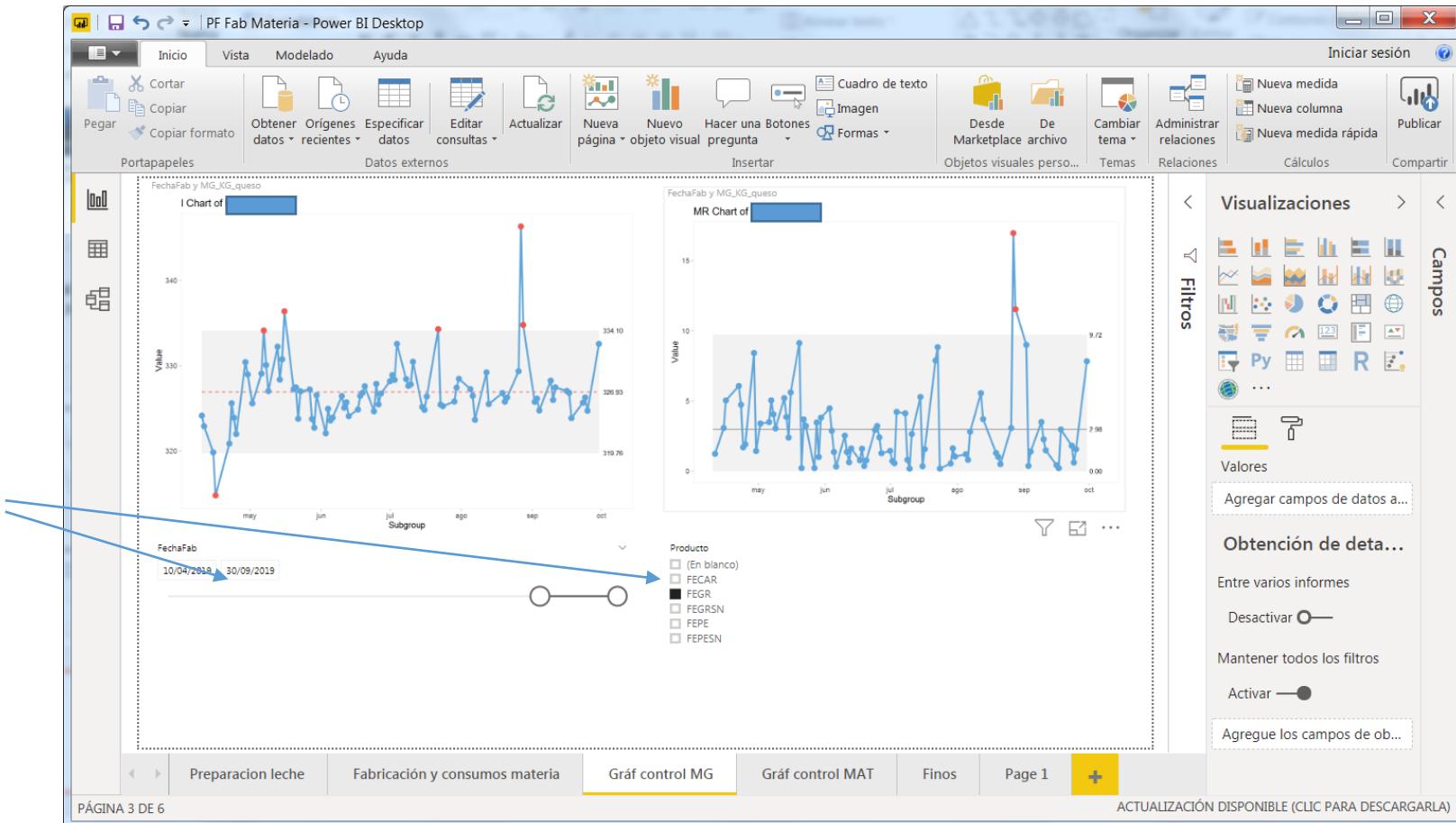
A blue arrow points from the text 'Script muy sencillo, en este caso sólo tres líneas' to the R code in the editor. Another blue arrow points from the text 'Las variables de análisis se puede actualizar simplemente cambiando en la lista de valores la variable seleccionada' to the 'Valores' section in the Power BI interface.

Script muy
sencillo, en
este caso sólo
tres líneas

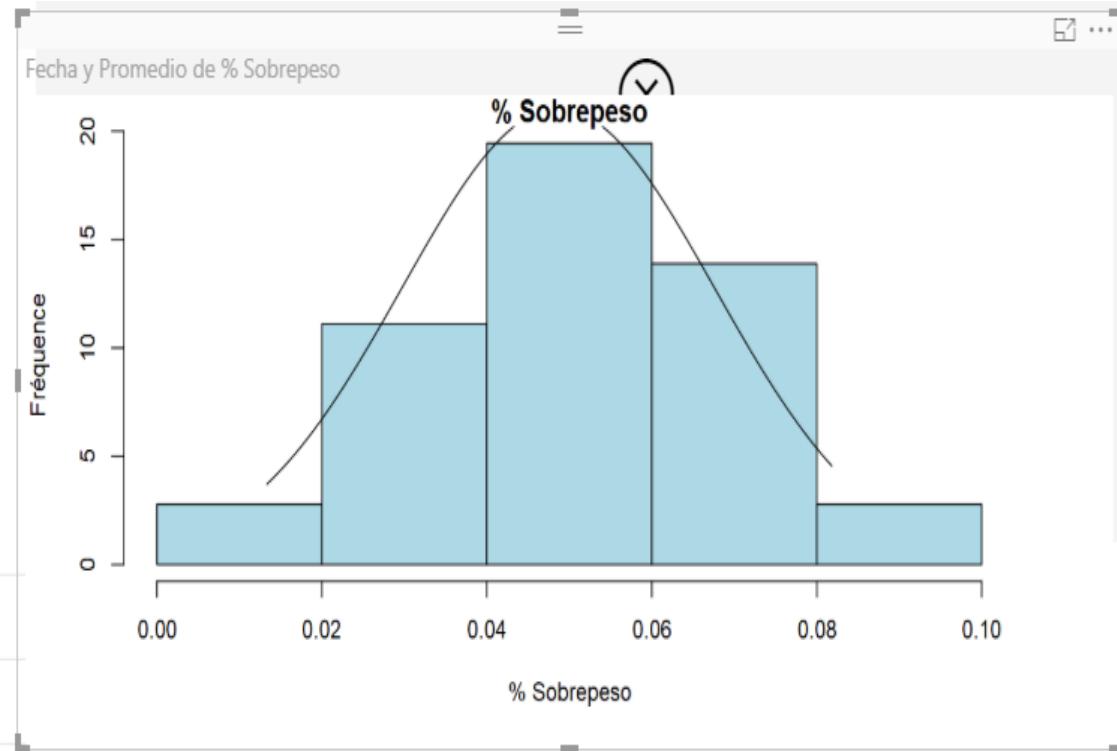
Las variables de
análisis se
puede
actualizar
simplemente
cambiando en
la lista de
valores la
variable
seleccionada

Uso de qicharts2 (2)

Los selectores de fecha y tipo de producto permiten una selección de un subset de datos automática que son pasados al script



Otras implementaciones en Power BI: Uso de hist (1) Uso de qcc (1)



R-Markdown. Uso de qic (1)

Informes recurrentes con una estructura definida

Útil para seguimiento y análisis de variables clave de forma rutinaria

Seguimiento semanal parametros clave

Lactalis Villarrobledo

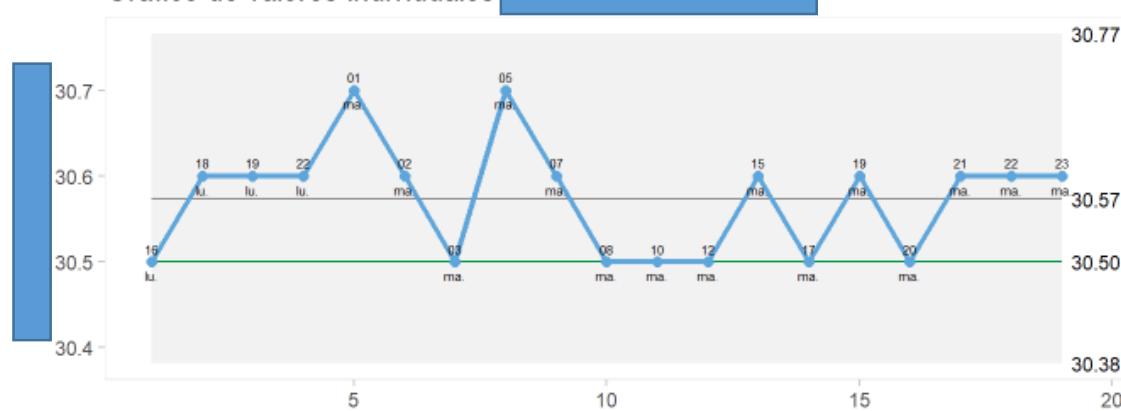
2019-10-09

Control semanal de los parámetros clave

Revisión de la producción de la semana pasada a través de cinco parámetros: saladero, HQD.

TIERNO

Gráfico de valores individuales



```
---
```

```
title: "Seguimiento de la variabilidad"
author: "Lactalis Villarrobledo"
date: ``r Sys.Date()``
output:
  html_document:
    df_print: paged
  word_document: default
params:
  fin: '2019-09-30'
  inicio: '2019-01-01'
```

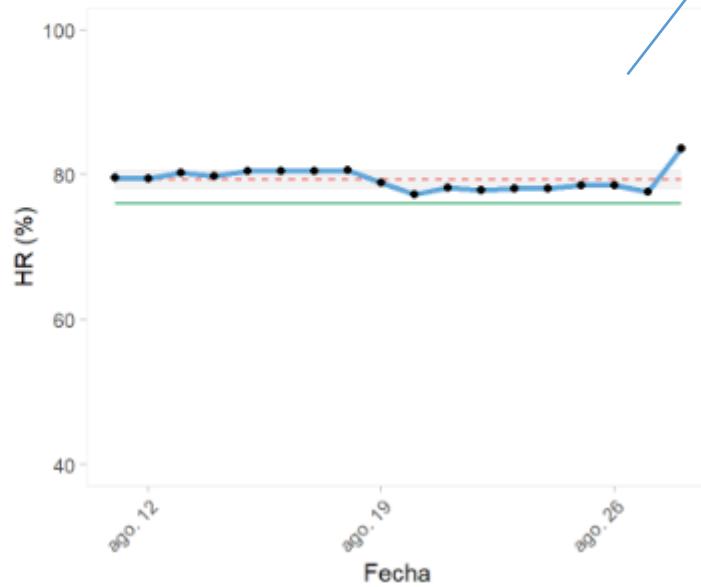
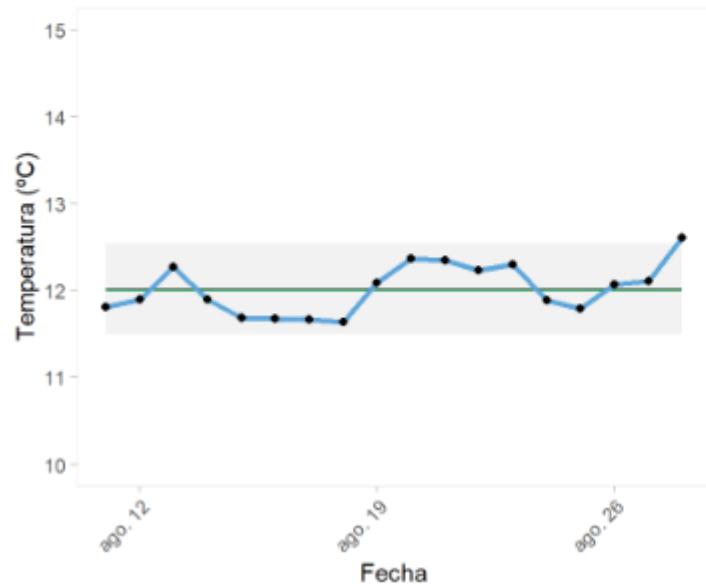
```
---
```

Estructura clara al inicio del script que permite la modificación del período a visual por parte de personas que tengan un conocimiento mínimo de R

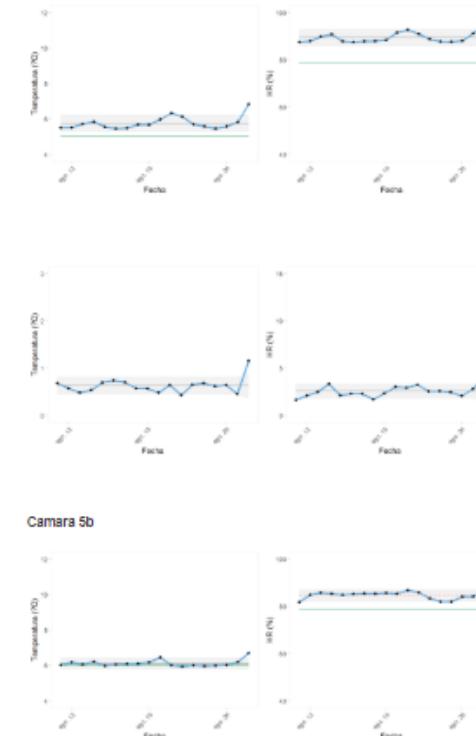
Uso de qic (2)

Graficos de control camaras

Camara 4

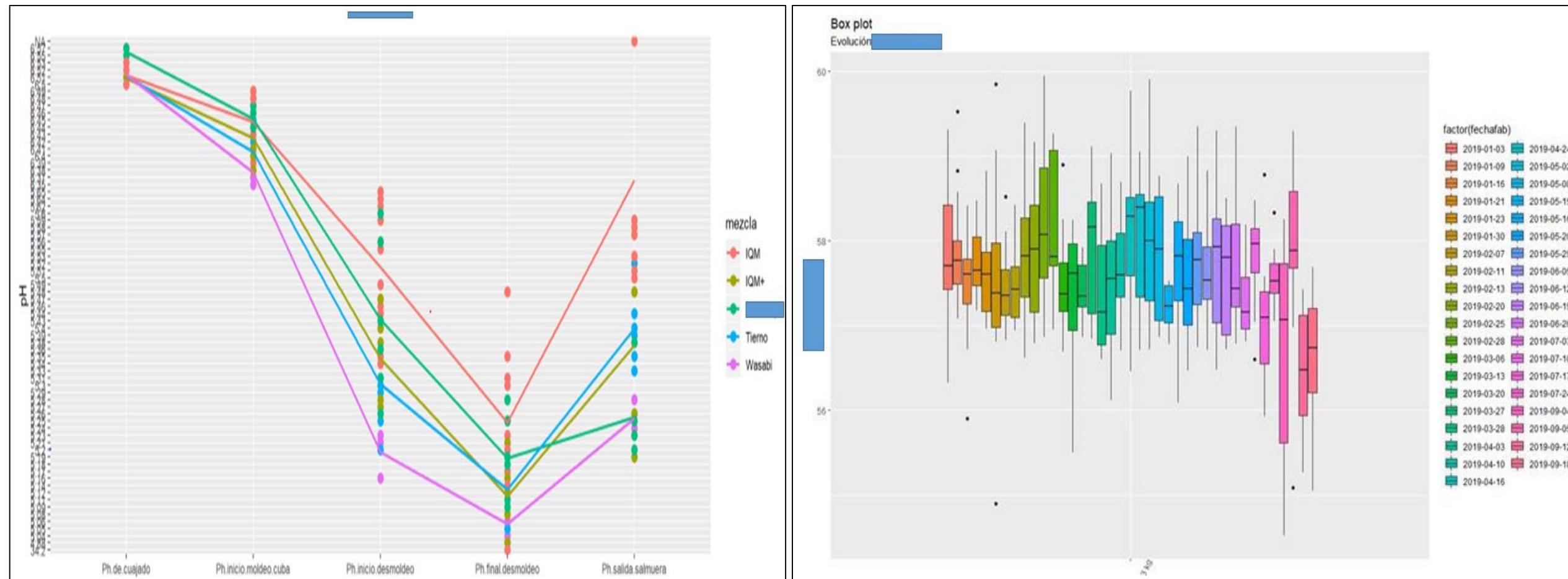


Camara 5a



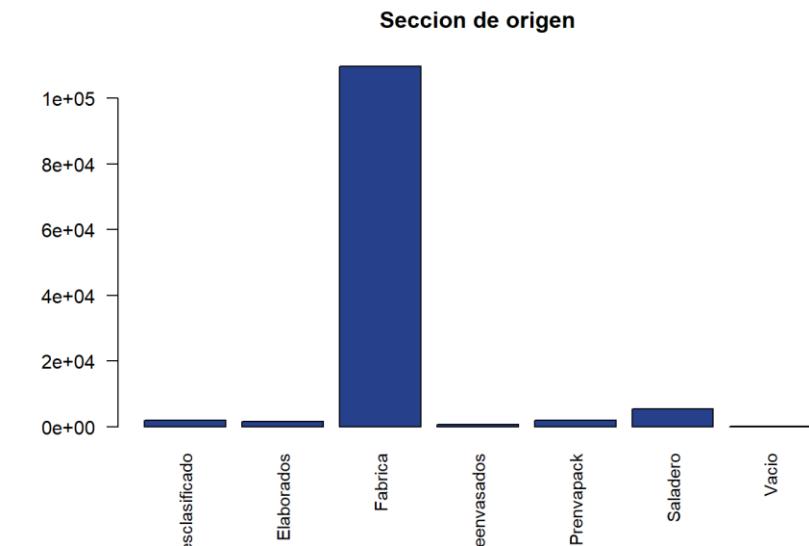
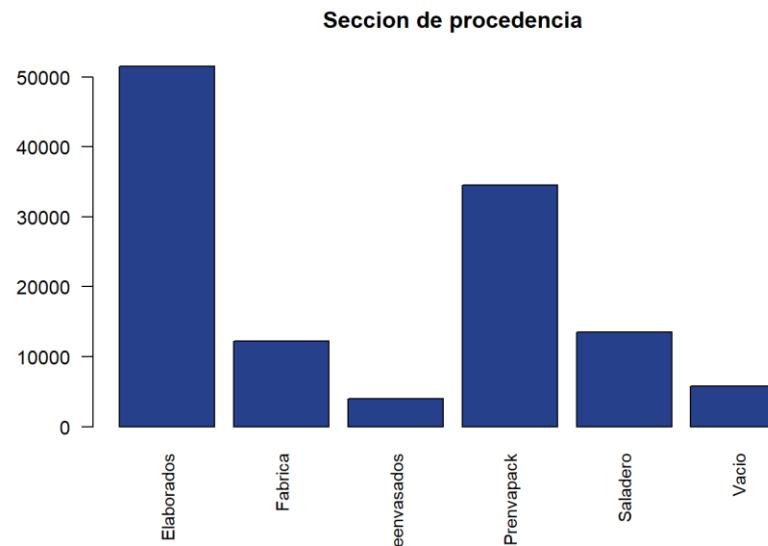
Camara 5b

Script de R. Uso de ggplot y ggthemes



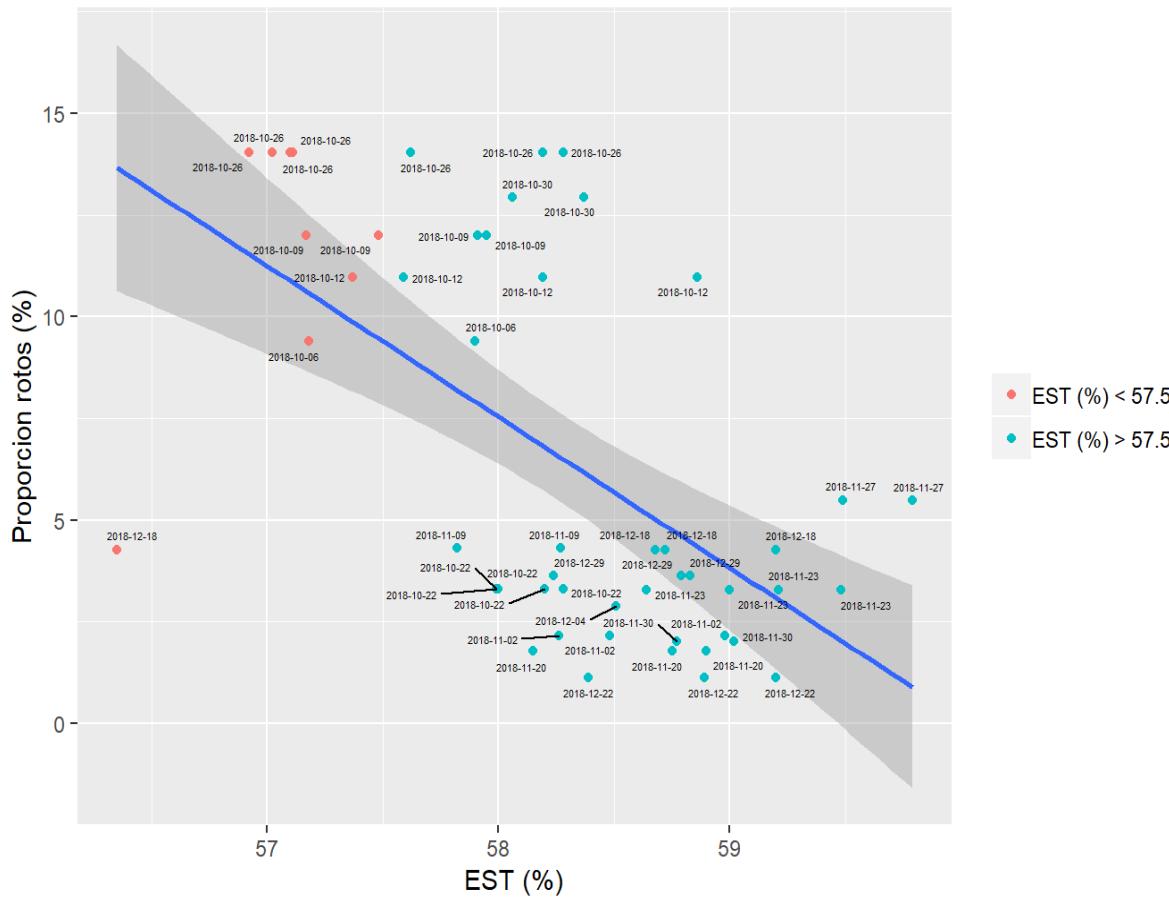
Otros usos técnicos:

- Sólo por las personas formadas → ingenieros o científicos con formación estadística de base
- ESTUDIO RELACIONADO CON ROTURAS DE QUESO
- ANALISIS DE LA VARIABILIDAD DE TEMPERATURA
- ESTUDIOS RELACIONADOS CON EL PESO
- ANALISIS ESTADISTICOS
- ...

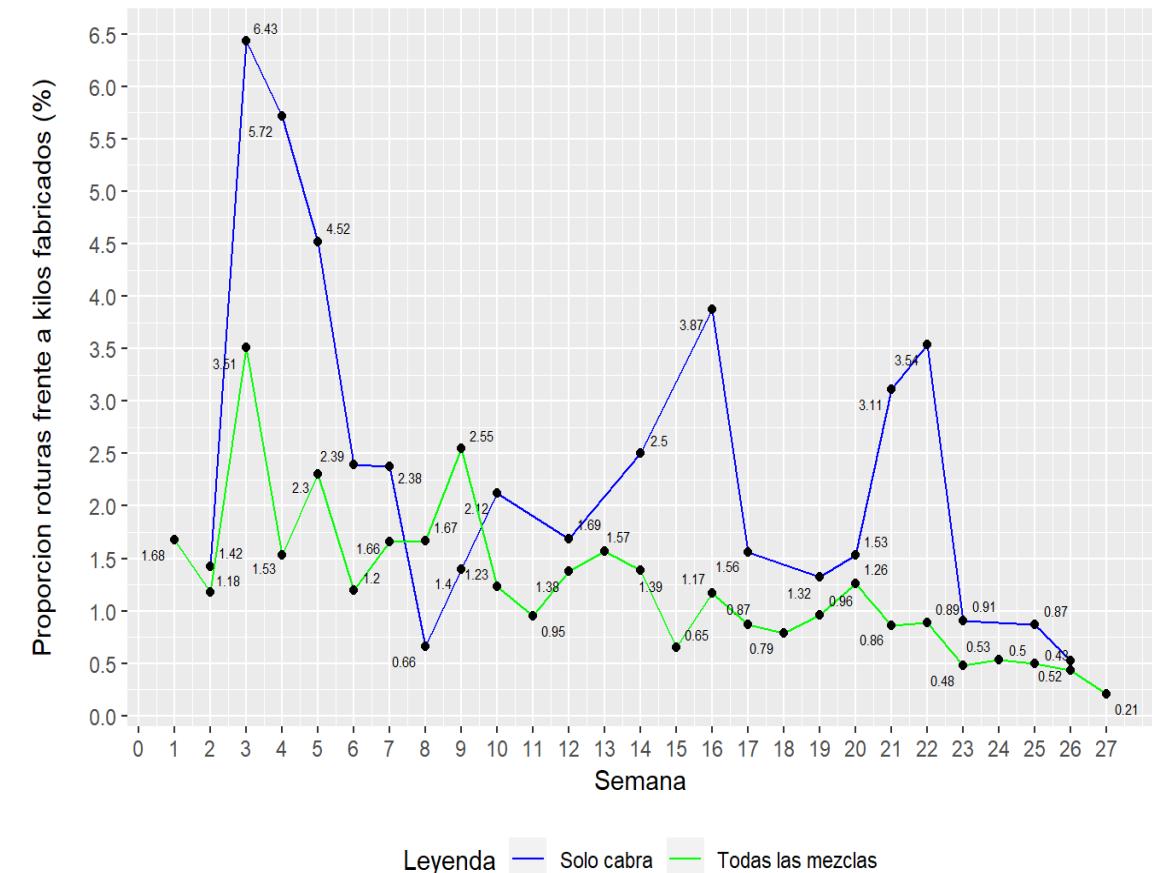


Estudio de Roturas (1)

Rotos vs ES barra cabra 4,2 año 2018-ggplot

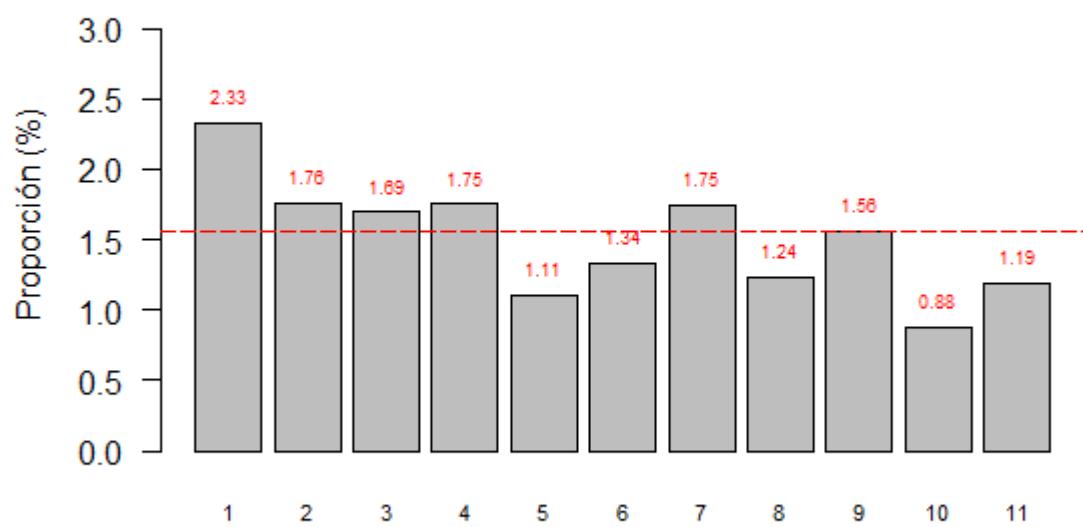


Proporción roturas barra 4,2 por semana 2019-ggplot

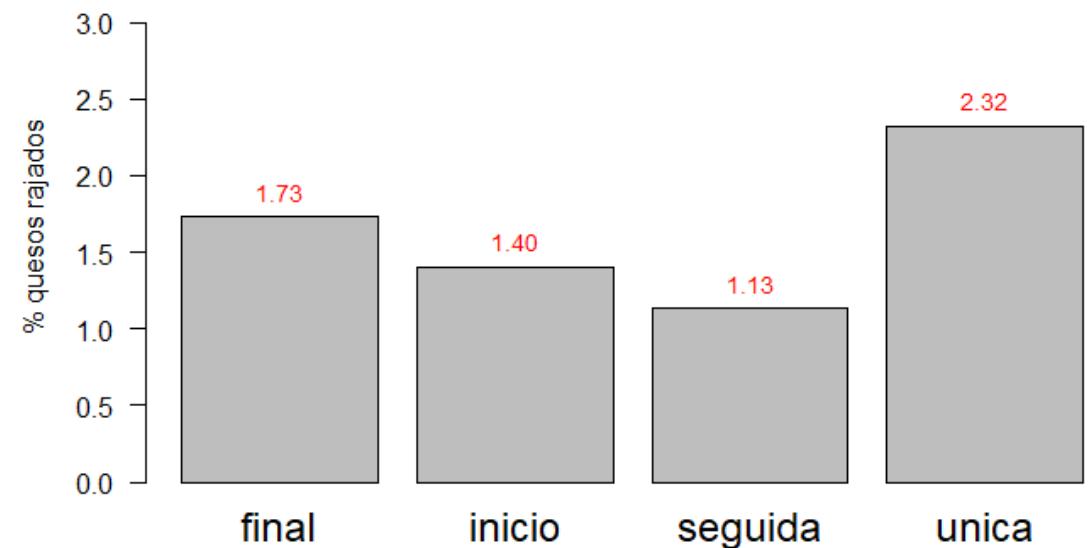


Estudio de Roturas (2)

Rotos vs número de cuba física-barplot

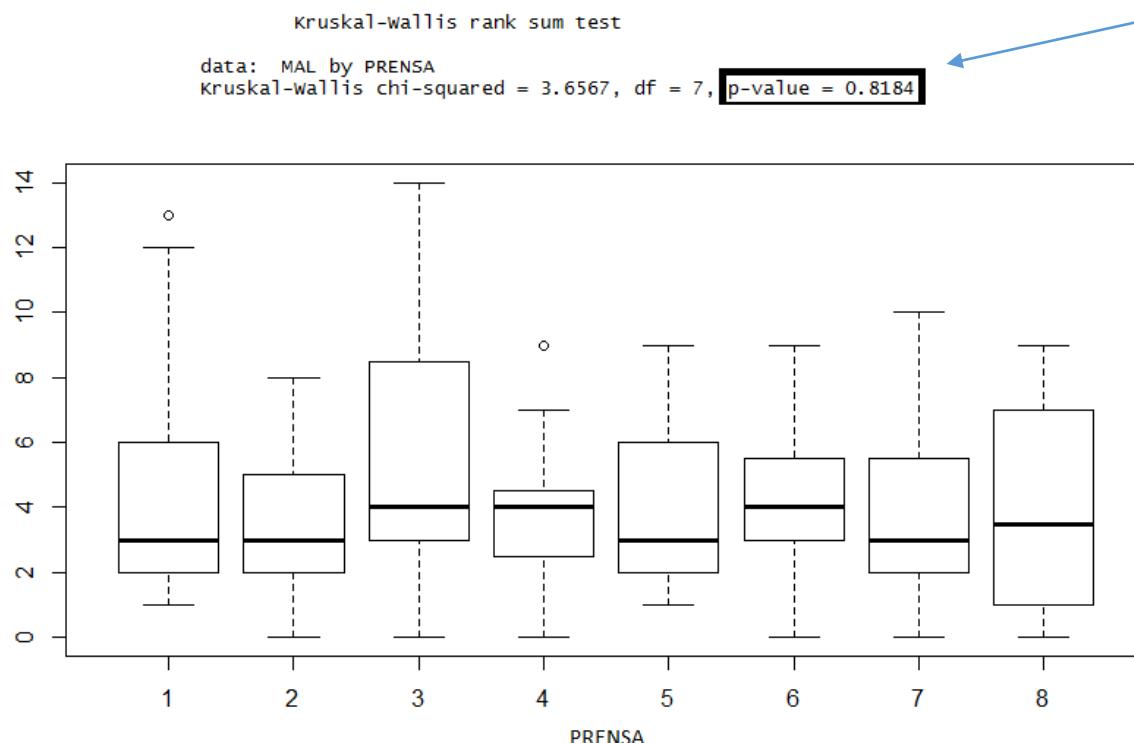


Rotos vs secuencia de llenado-barplot



Estudio de Roturas (3)

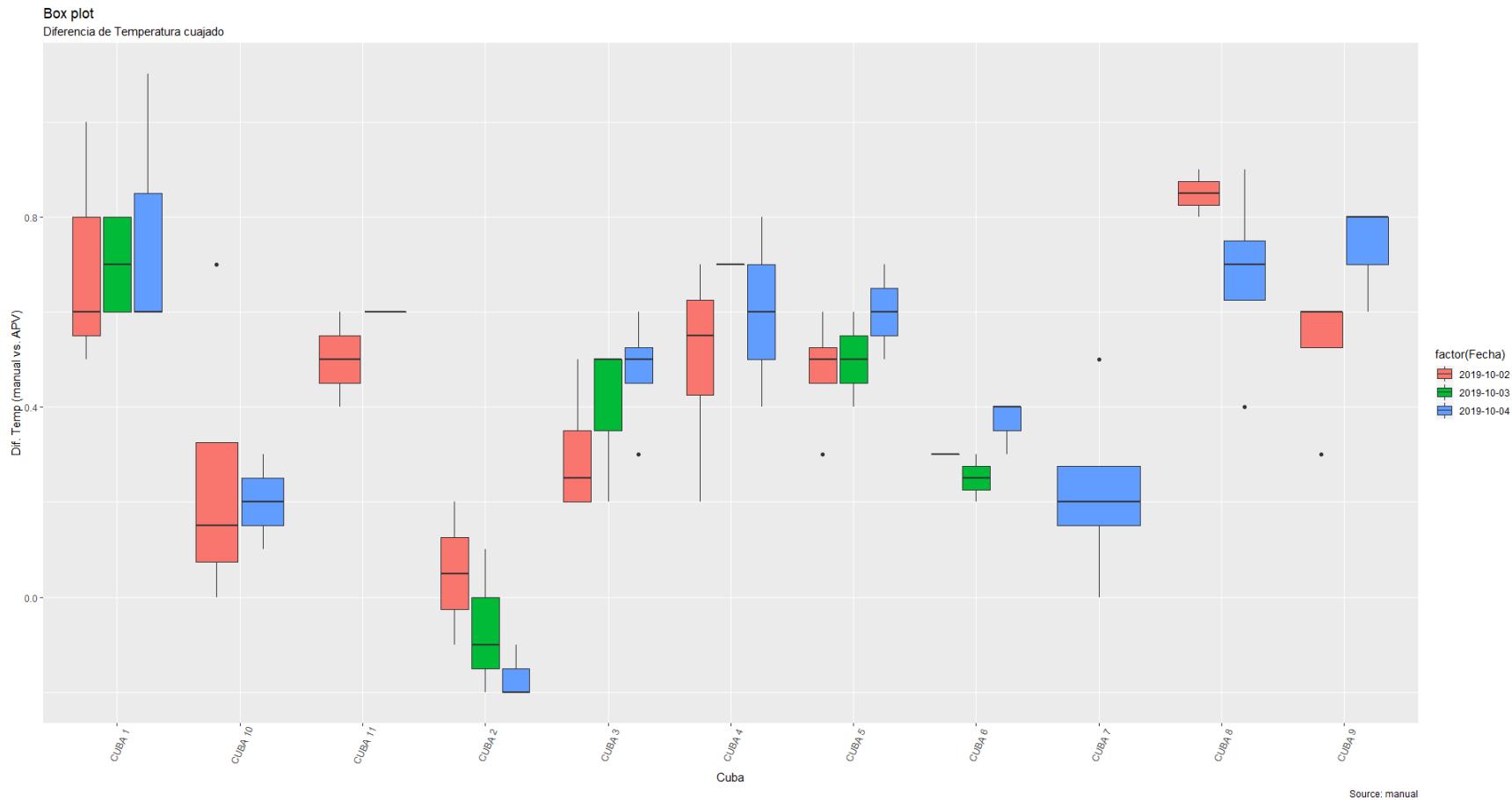
Mal prensados por prensa en LD2-boxplot y test estadístico



No se observan diferencias significativas entre el porcentaje de mal prensados y el número de prensa

Análisis de la variabilidad (1)

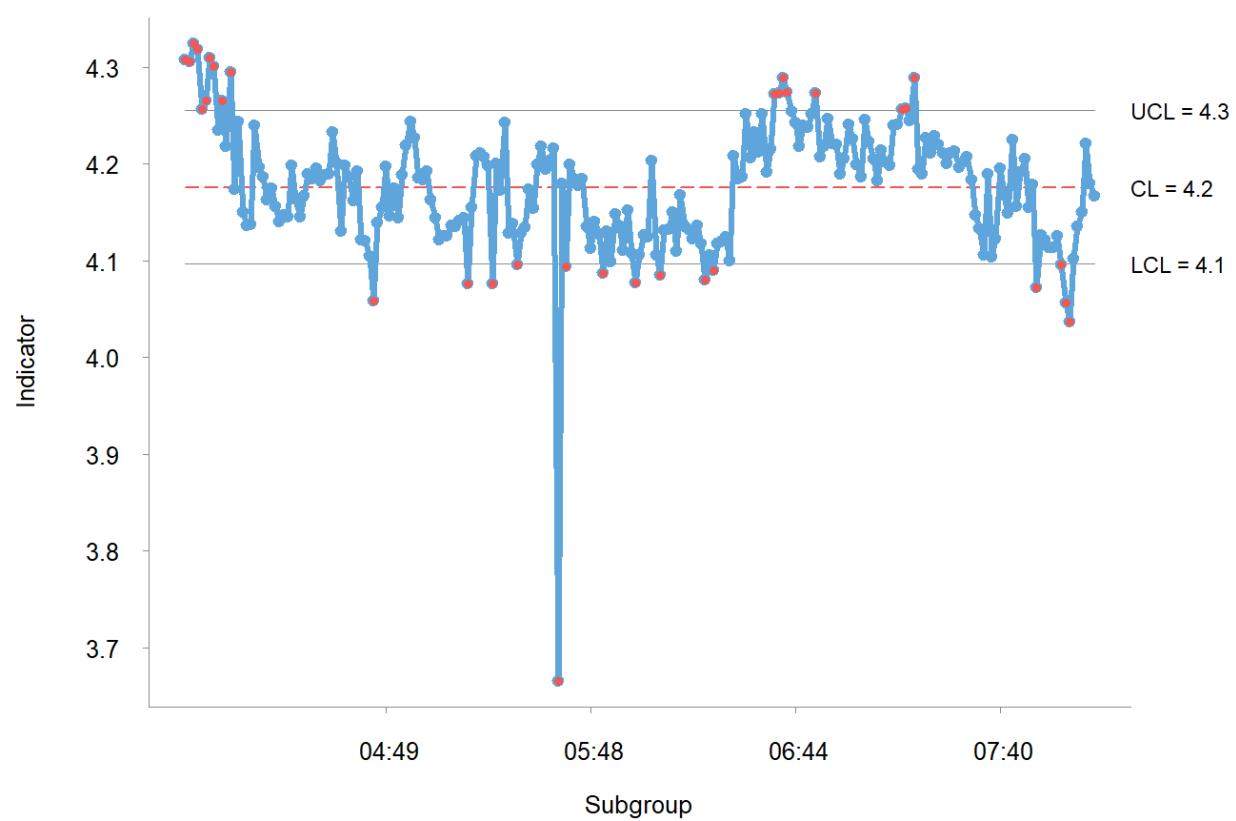
Diferencia temperatura entre sonda y termómetro manual por cuba- ggthemes



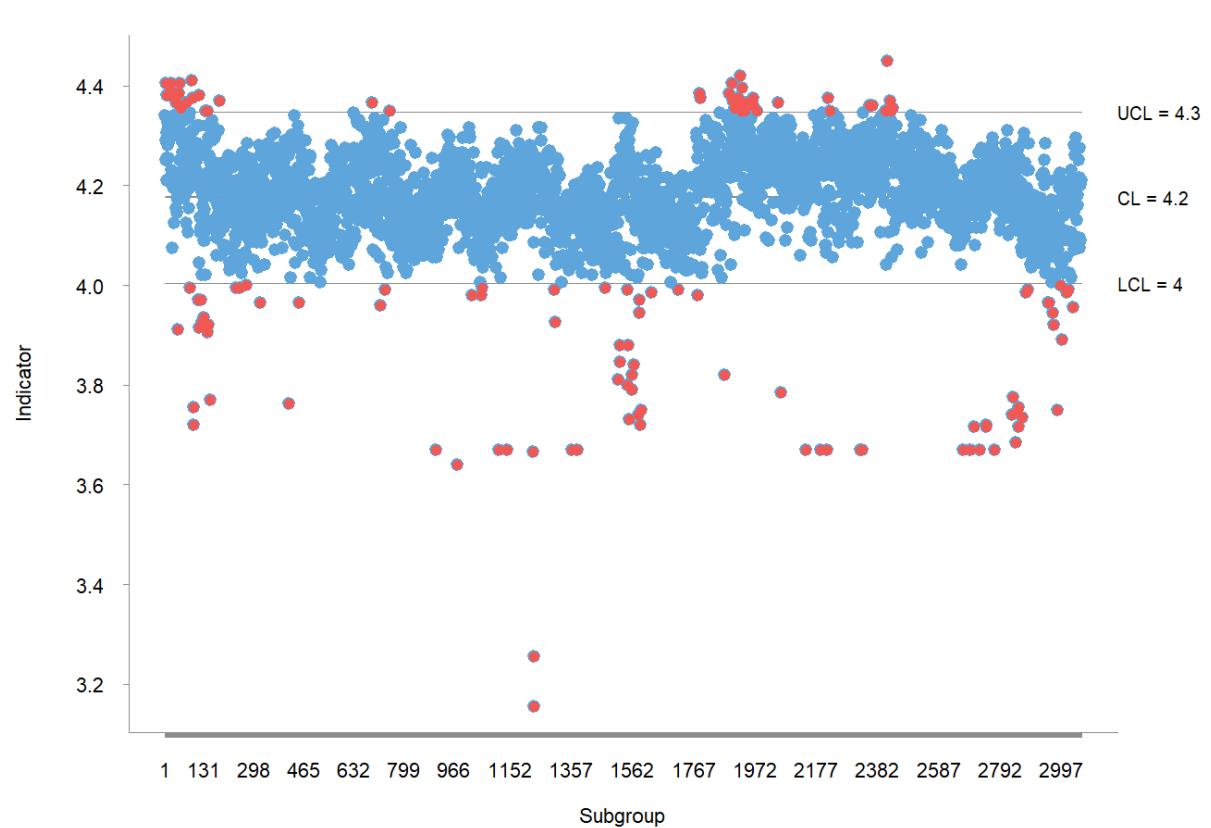
Estudio de peso (1)

Gráficas de control pesos queso A formato B-qic

I Chart of PESO



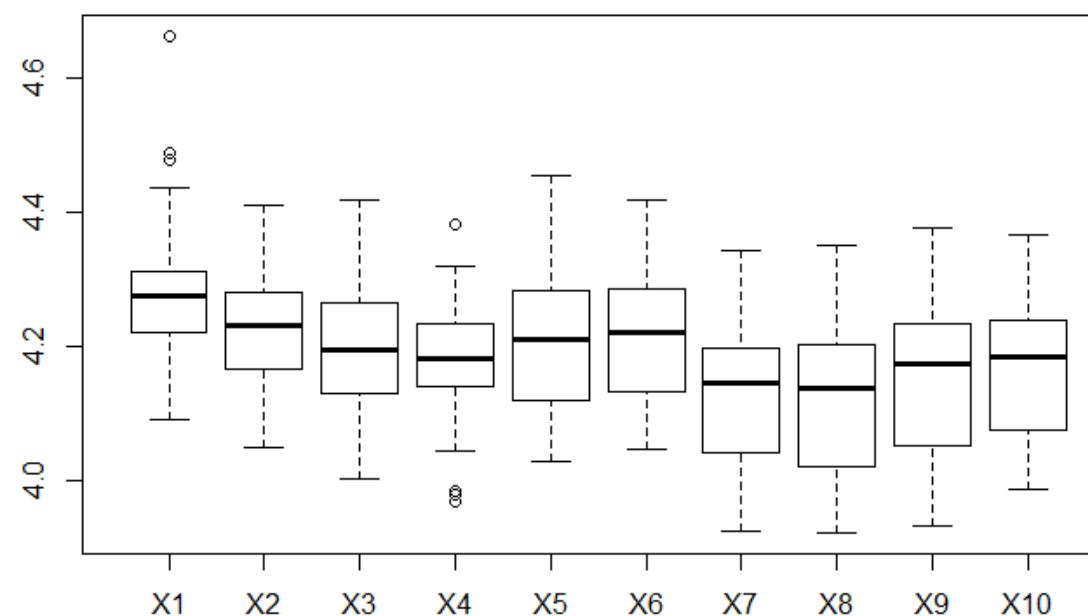
I Chart of PESO



Estudio de peso (2)

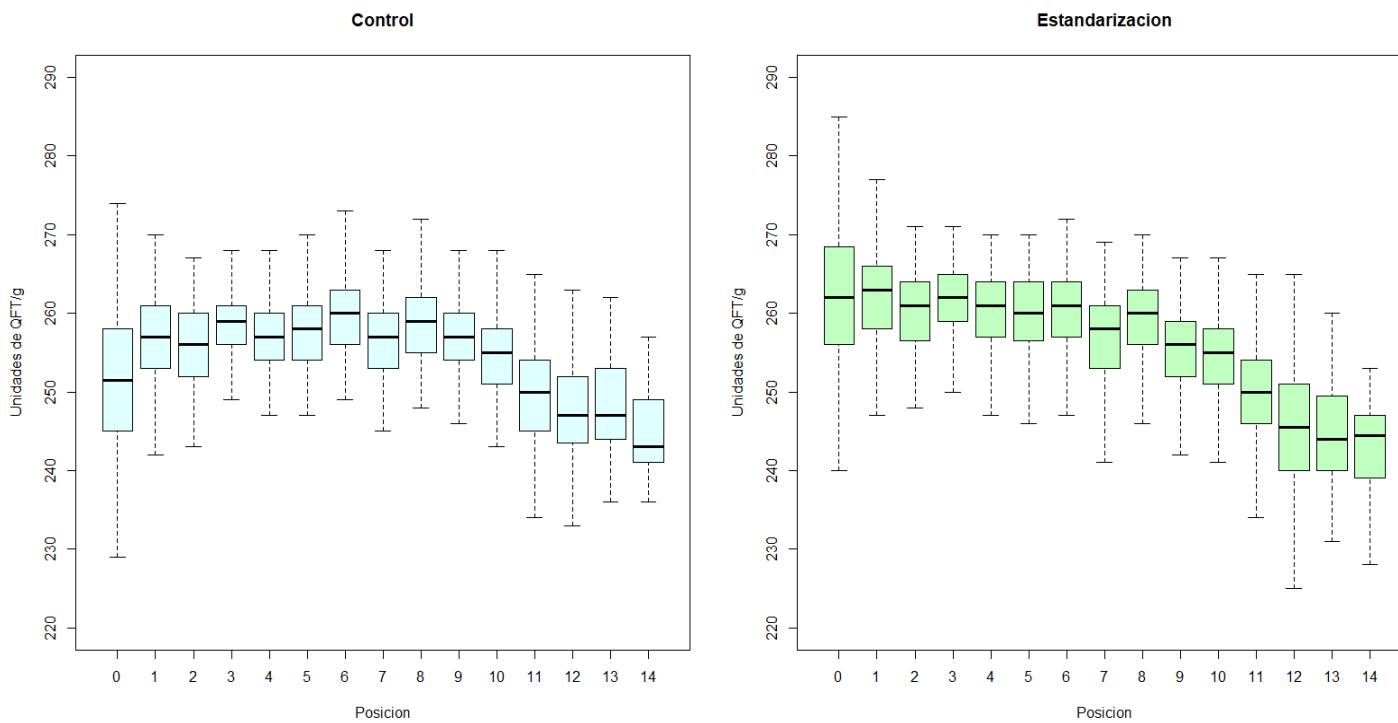
Pesos desmoldeo formato X mezcla Y-boxplot y test estadístico

```
Kruskal-wallis rank sum test  
data: DESMOLDEO  
Kruskal-wallis chi-squared = 59.017, df = 9, p-value = 2.073e-09
```



Estudio de peso (3)

Distribución de peso por unidad-boxplot



Distribución de peso por unidad-test snk

Posición	Control			Estandarización			
	μ (g)	σ (g)	SNK	μ (g)	σ (g)	SNK	
0	251.2	9.51	b	262.8	9.07	a	
1	256.8	5.71	a	262.1	5.74	ab	
2	256.0	5.26	a	260.4	5.14	bc	
3	258.5	4.22	a	261.8	4.92	ab	
4	257.2	4.74	a	260.3	5.04	bc	
5	257.5	4.82	a	260.0	5.23	bc	
6	259.4	4.90	a	260.8	5.09	bc	
7	256.5	4.99	a	257.0	5.36	d	
8	258.8	4.68	a	259.1	5.07	c	
9	256.7	4.78	a	255.4	4.81	e	
10	254.6	5.27	a	254.2	5.48	e	
11	249.5	6.36	b	249.9	6.05	f	
12	247.8	7.02	bc	245.6	7.81	g	
13	248.1	5.38	bc	245.1	6.64	g	
14	244.7	5.27	c	241.8	7.16	h	
<i>p</i> -valor			< 2e ⁻¹⁶ ***			< 2e ⁻¹⁶ ***	

Futuro:

- Machine learning: interacciones entre KPI técnicos, explicar indicadores de calidad (organolépticos) → ¿wine dataset?
- Automatizar tareas, uso de shiny en servidor para pilotaje de gráficos de control en línea
- Diseño de experimentos, teoría y práctica. Muy bien implementado en Minitab, no tanto en R
- Posiblemente conocer Python para complementar la fase analítica y de manejo de datos.
- Y seguir

R que R



Proposiciones

- Proyectos de colaboración
- Formación continua
- Necesidades Lactalis:
 - Personas formadas en la utilización de R
 - Seguir acercando R al usuario final  
 - (Interfaz más amigable)
 - La regla de las 3 "

```
> print("¡Muchas gracias por  
vuestra atención!")  
> [1] "¡Muchas gracias por  
vuestra atención!"
```



Alonso Abalde, Marta
Responsable Tecnología Quesera



Agregar

[CONTACTO](#) | [ORGANIZACIÓN](#) | [NOVEDADES](#) | [PERTENENCIA A](#)

Calendario

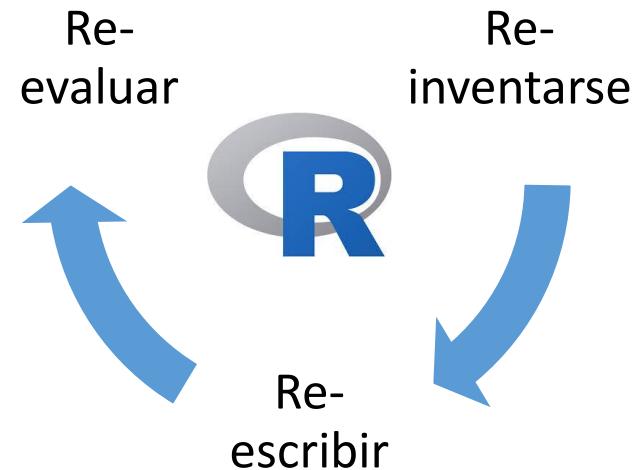
Disponible para las 8 horas
siguientes

[Programar una reunión](#)

Enviar correo electrónico
Marta.ALONSO@lactalis.es

Oficina
Villarrobledo

Compañía
LACTALIS VILLARROBLEDO, S.L.





UNIVERSIDAD DE CÓRDOBA

FSinR: Un paquete extenso para selección de características

Francisco Aragón¹, Alfonso Jiménez², Antonio Arauzo², José M. Benítez¹

¹ *Depto. Ciencias de la Computación e Inteligencia Artificial, DECSAI, DICITS*

Universidad de Granada

² *Área de Proyectos de Ingeniería*

Universidad de Córdoba

The logo for madRid features the word "madRid" in a bold, sans-serif font. The letter "R" is particularly large and stylized, with a blue gradient. A light blue oval shape surrounds the letters "mad" and "R".

XI Jornadas de Usuarios de R

INTRODUCCIÓN

¿Qué es selección de características?

- Es el proceso de seleccionar un subconjunto de características relevantes



- Se ha convertido en parte clave del Machine Learning



Feature Selection

¿Por qué es importante?

- Su aplicación presenta un impacto directo en el rendimiento de los modelos
 - Elimina información redundante, irrelevante, ruido, ...
 - Facilita la interpretación
 - Reduce el tiempo de cómputo
 - Reduce la complejidad del modelo
 - ...
- Mejora los resultados de los modelos



MOTIVACIÓN

Es un problema vigente

- La Selección de Características es un proceso bastante complejo
- Esto hace que aún no se haya encontrado una solución definitiva



- Por lo tanto, siguen siendo interesantes las nuevas propuestas y aportaciones

¿Qué hay en R sobre FS?

- Paquetes dedicados con varios métodos:
 - FSelector
 - MXM
- Paquetes dedicados a un método:
 - Boruta
 - spSRF
 - varSelRF
 - ...
- Paquetes no dedicados:
 - caret
 - CORElearn
 - ...



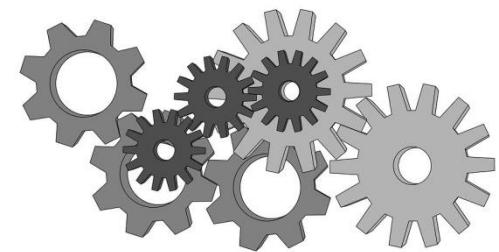
¿Qué vemos que falta?

- Los paquete actuales abordan el problema de una forma poco extensa:
 - Tienen pocos métodos...
 - ... ó solo un método concreto
 - ... ó no están ni siquiera dedicados
- No contienen una gran variedad de métodos, ni un gran número de métodos conocidos en el estado del arte, ni dan facilidades de uso...
- Falta un gran paquete dedicado a la Selección de Características

PRINCIPAL PROBLEMA

¿Qué aporta nuestro paquete?

- Gran recopilación de métodos de filtro y wrapper muy usados en la literatura
 - + de 15 métodos de filtro
 - Para los métodos de wrapper se usa el paquete caret
 - + de 200 modelos de regresión y clasificación
- Estrategias de búsqueda
 - Combinación con filtro y wrapper
- Fácil de usar, amplia documentación, vignettes de uso, web con más detalle...

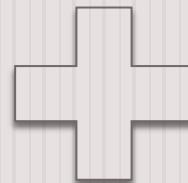


PAQUETE FSinR

Paquete FSinR

Es un paquete R con herramientas
para realizar el
Proceso de Selección de Características

Algoritmo de búsqueda



Método de Wrapper

Método de Filtro

Métodos de Wrapper

Métodos de Filtro

Métodos de
corte

Métodos de
Wrapper

Métodos de Filtro

Métodos de búsqueda

- Guían la búsqueda de la mejor solución en el espacio formado por todas las combinaciones de características
 - Secuenciales (SFS, SFFS, SBS, SFBS)
 - Anchura / profundidad
 - Local (Hill-Climbing, Taboo Search)
 - Probabilísticas (Las Vegas)
 - Heurísticas (GA, SA, WOA, ACO)
- Se aplica con métodos de filtro y wrapper



Métodos de filtro

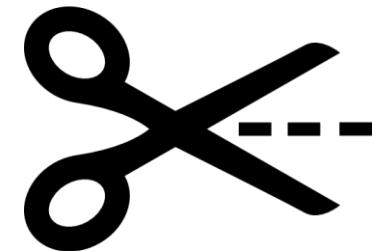
- Devuelven la valoración de un subconjunto de características basándose en test estadísticos
 - Individuales
 - Chi-Squared, Cramer V, F-score, Relief
 - Conjunto
 - Rough Sets Consistency, Binary Consistency, IE Consistency, IEP Consistency, Determination Coefficient, Mutual Information, Gain Ratio, Symmetrical Uncertain, Gini Index, Jd Evaluation, MDLC Evaluation, RFSM Evaluation
 - Híbridas
 - LCC

Métodos de wrapper

- Devuelven una evaluación de un subconjunto de características por medio de un algoritmo de aprendizaje
- Se puede hacer uso de los 238 modelos disponibles del paquete caret
- Por medio de la función `wrapperGenerator`, que define:
 - Los parámetros de remuestreo (`trainControl` de caret)
 - Los parámetros de ajuste (`train` de caret)
 - Nombre del método de aprendizaje (de los disponibles en caret)

Métodos de corte

- Las medidas de corte eligen un subconjunto de características según una serie de criterios
 - Select k-best
 - Select percentile
 - Select Threshold
 - Select Threshold range
 - Select difference
 - Select Slope
- Se pueden usar con métodos de filtro y de wrapper (evaluación individual)



EJEMPLOS DE USO

Búsqueda con wrapper (clas.)

```
1 library(FSinR)
2
3 # Values for trainControl function
4 resamplingParams <- list(method = "cv", number = 10)
5 # Values for train function (x, y, method and trainControl not necessary)
6 fittingParams <- list(preProc = c("center", "scale"), metric="Accuracy", tuneGrid = expand.grid(k = c(1:20)))
7
8 # wrapper method
9 wrapper <- wrapperGenerator("knn", resamplingParams, fittingParams)
10
11 # search method (sfs) + wrapper method
12 sfs(iris, 'Species', wrapper)
13 # search method (ga) + wrapper method
14 ga(iris, 'Species', wrapper, popSize = 8, pcrossover = 0.8, pmutation = 0.1, maxiter=4, verbose=TRUE)
15 # search method (hc) + wrapper method
16 hc(iris, 'Species', wrapper, verbose=TRUE)
17
18 # ...
```

Salida por consola

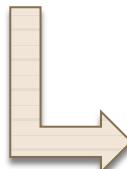


```
HC | InitialVector = 0010 | InitialFitness = 0.9533333
HC | Vector = 0011 | Fitness = 0.9666667 | End? = no
HC | Vector = 1011 | Fitness = 0.9733333 | End? = no
HC | Vector = 0011 | Fitness = 0.9666667 | End? = yes
```

Búsqueda con wrapper (reg.)

```
1 library(FSinR)
2
3 # Values for trainControl function
4 resamplingParams <- list(method = "repeatedcv", repeats = 3)
5 # Values for train function (x, y, method and trainControl not necessary)
6 fittingParams <- list(preProc = c("center", "scale"), metric="RMSE",
7                         tuneGrid = expand.grid(size = seq(1,12,by=2), decay=0), trace=FALSE)
8
9 # wrapper method
10 wrapper <- wrapperGenerator("nnet",resamplingParams, fittingParams)
11
12 # search method (sbs) + wrapper method
13 sbs(mtcars, 'mpg', wrapper)
14 # search method (sa) + wrapper method
15 sa(mtcars, 'mpg', wrapper, temperature = 5, temperature_min=0.01, reduction=0.6, innerIter=1, verbose=TRUE)
16 # search method (ts) + wrapper method
17 ts(mtcars, 'mpg', wrapper, numNeigh = 4, tamTabuList = 4, iter = 5, intensification=1, iterIntensification=5,
18     diversification=1, iterDiversification=5, verbose=TRUE)
19
20 # ...
```

Salida por consola



```
TS | InitialVector = 0011011111 | InitialFitness = 19.7355644
TS | Iter = 1 | Vector = 0011010111 | Fitness = 19.6899545 | BestFitness = 19.6899545
TS | Iter = 2 | Vector = 0011010011 | Fitness = 19.6207992 | BestFitness = 19.6207992
TS | Iter = 3 | Vector = 1011010011 | Fitness = 19.6620354 | BestFitness = 19.6207992
TS | Iter = 4 | Vector = 1011000011 | Fitness = 19.6026519 | BestFitness = 19.6026519
TS | Iter = 5 | Vector = 1011001011 | Fitness = 19.4186876 | BestFitness = 19.4186876
```

Búsqueda con filtro (clas.)

```
1 library(FSInR)
2
3 # search method (sffs) + filter method
4 sffs(iris, 'Species', giniIndex)
5
6 # ...
```

Resultados

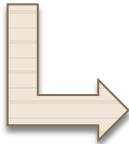


	\$bestFeatures	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
[1,]		1		1	1
	\$bestValue				0
		[1]	1		

Filtro / Wrapper individual

```
1 library(FSInR)
2
3 chiSquared(iris, 'Species', c("Sepal.Length","Sepal.Width","Petal.Length","Petal.Width"))
```

Resultados



Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
156.26667	89.54629	271.80000	271.75000

```
1 library(FSInR)
2
3 # Values for trainControl function
4 resamplingParams <- list(method = "cv", number = 10)
5 # Values for train function (x, y, method and trainControl not necessary)
6 fittingParams <- list(preProc = c("center", "scale"), metric="Accuracy", tuneGrid = expand.grid(k = c(1:20)))
7
8 # wrapper method
9 wrapper <- wrapperGenerator("knn", resamplingParams, fittingParams)
10
11 wrapper(iris,"Species",c("Sepal.Length","Sepal.Width"))
```

Resultados



[1] 0.8133333

Métodos de corte

```
1 library(FSinR)
2
3 # Values for trainControl function
4 resamplingParams <- list(method = "cv", number = 10)
5 # Values for train function
6 fittingParams <- list(preProc = c("center", "scale"), metric="Accuracy", tuneGrid = expand.grid(k = c(1:20)))
7
8 # wrapper method
9 wrapper <- wrapperGenerator("knn", resamplingParams, fittingParams)
10
11 # cutoff with wrapper
12 selectKBest(iris, 'Species', wrapper, 2)
13
14 # cutoff with filter
15 selectThreshold(iris, 'Species', cramer, 0.70)
```

Resultados

```
$bestFeatures
Sepal.Length Sepal.Width Petal.Length Petal.Width
[1,]          0          0          1          1

$featuresSelected
[1] "Petal.Width"  "Petal.Length"

$valuePerFeature
[1] 0.9600000 0.9533333
```

Resultados

```
$bestFeatures
Sepal.Length Sepal.Width Petal.Length Petal.Width
[1,]          1          0          1          1

$featuresSelected
[1] "Petal.Length" "Petal.Width"  "Sepal.Length"

$valuePerFeature
[1] 0.9518403 0.9517528 0.7217263
```

CONCLUSIÓN

Conclusiones

- La Selección de Características es importante dentro del Machine Learning
- Existe una carencia de paquetes extensos dedicados en R
- En este sentido hemos presentado el paquete FSinR que destaca por:
 - Métodos de filtro y wrapper más comunes en la literatura
 - Combinación con métodos de búsqueda y corte
 - La herramienta más extensa y completa
 - Fácil de usar, muy documentada

FSinR



GRACIAS