

# VI XORNADA DE USUARIOS DE EN GALICIA



| 10 de outubro de 2019  
| Santiago de Compostela

## LIBRO DE RESUMOS

```
boxplot(x, ...)  
boxplot.default(x, ..., range = 1.5, width = NULL,  
  varwidth = FALSE, notch = FALSE, names, boxwex = 0.8,  
  data = parent.frame(), plot = TRUE,  
  border = par("fg"), col = NULL, log = "", pars = NULL,  
  horizontal = FALSE, add = FALSE)  
boxplot.formula(formula, data = NULL, subset, na.action, ...)
```



> ORGANIZA



> COLABORAN



FACULTADE DE MATEMÁTICAS

> PATROCINAN



Agencia para a  
Modernización Tecnolóxica



XUNTA  
DE GALICIA



**VI XORNADA DE**  
**USUARIOS** <sup>DE</sup>  
EN GALICIA 

PROGRAMA E  
RESUMOS

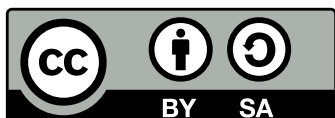
10 de outubro de 2019  
Santiago de Compostela

**Editora:** María José Ginzo Villamayor

**ISBN:** 978-84-09-15024-3

© 2019 Consorcio CIXUG

Obra baixo licenza Creative Commons Atribución-Compartir igual 4.0 Internacional



**Atribución - Compartir igual**

En calquera mención da obra debe citarse a autoría

Debe proveerse enlace á licenza e indicalo cando se introduzan cambios

A obra derivada debe licenciarse do mesmo xeito que a orixinal



## Presentación

A Oficina de Software Libre (OSL) do CIXUG comprácese en presentar a VI Xornada de Usuarios de R en Galicia.

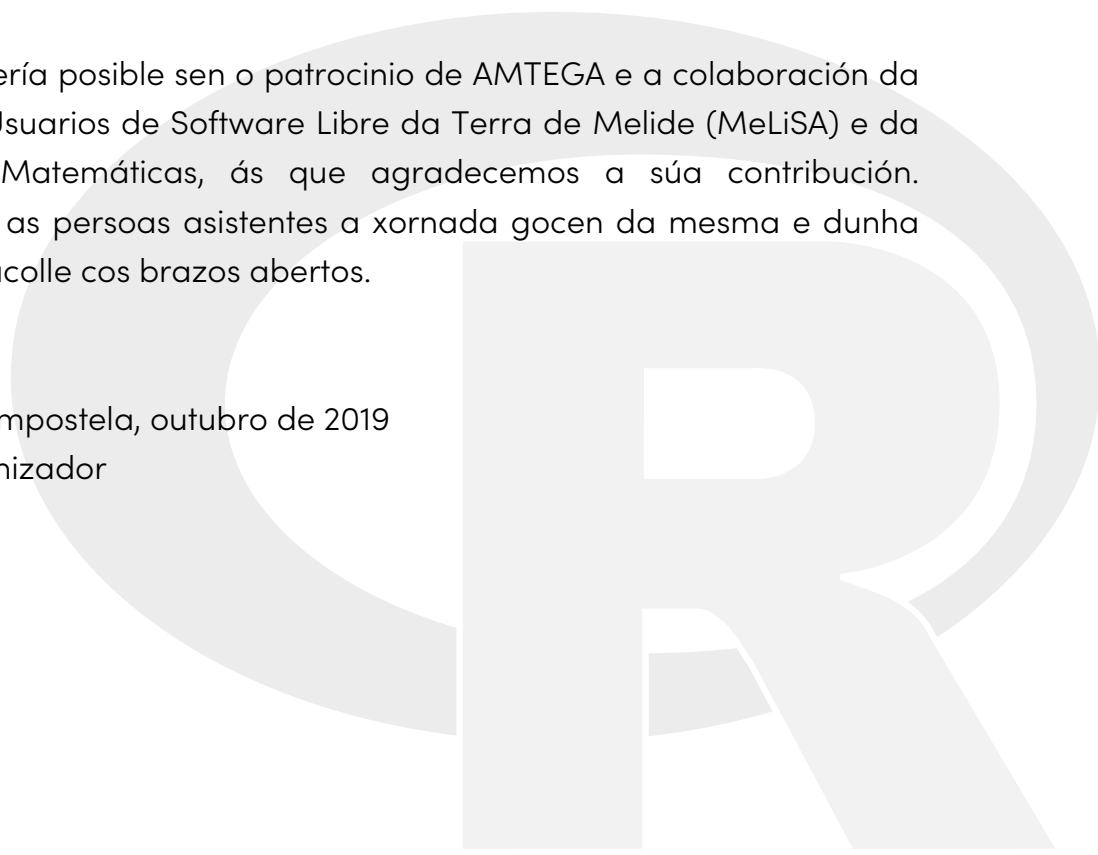
Pretende ser un punto de encontro para todas aquelas persoas interesadas en intercambiar as súas experiencias e atopar colaboracións co resto da comunidade, ademais trátase de promocionar e difundir o coñecemento libre da linguaxe estatística R e mostrar as súas aplicacións.

O programa contempla dezasete relatorios ao longo de todo o día, ademais de tres obradoiros: Introducción a R, Visualización de datos espaciais con R e Multivariate and Functional Data Classification in R.

Entre as persoas participantes figuran especialistas do Instituto Galego de Estatística, da Consellería de Sanidade e outros organismos da Xunta de Galicia, das tres universidades galegas, do Instituto Tecnolóxico de Matemática Industrial (ITMATI) e de empresas como Abanca ou Manserga y otros servicios, SL.

Todo isto non sería posible sen o patrocinio de AMTEGA e a colaboración da Asociación de Usuarios de Software Libre da Terra de Melide (MeLiSA) e da Facultade de Matemáticas, ás que agradecemos a súa contribución. Confiamos que as persoas asistentes a xornada gocen da mesma e dunha cidade que os acolle cos brazos abertos.

Santiago de Compostela, outubro de 2019  
O Comité Organizador



## **Comité organizador**

María José Ginzo Villamayor  
*Universidade de Santiago de Compostela*

Rafael Rodríguez Gayoso  
*Concello de Santiago de Compostela*

Xabier Sánchez Santos  
*Consortio Interuniversitario CIXUG*

## **Comité científico**

María José Ginzo Villamayor  
*Universidade de Santiago de Compostela*

Miguel Ángel Rodríguez Muíños  
*Dirección Xeral de Saúde Pública (Consellería de Sanidade)*



## **Información Xeral**

### **Sede**

Facultade de Matemáticas  
Universidade de Santiago de Compostela  
Rúa Lope Gómez de Marzoa s/n  
15782 - Santiago de Compostela

### **Data**

10 de outubro de 2019

### **Acceso WIFI na sede**

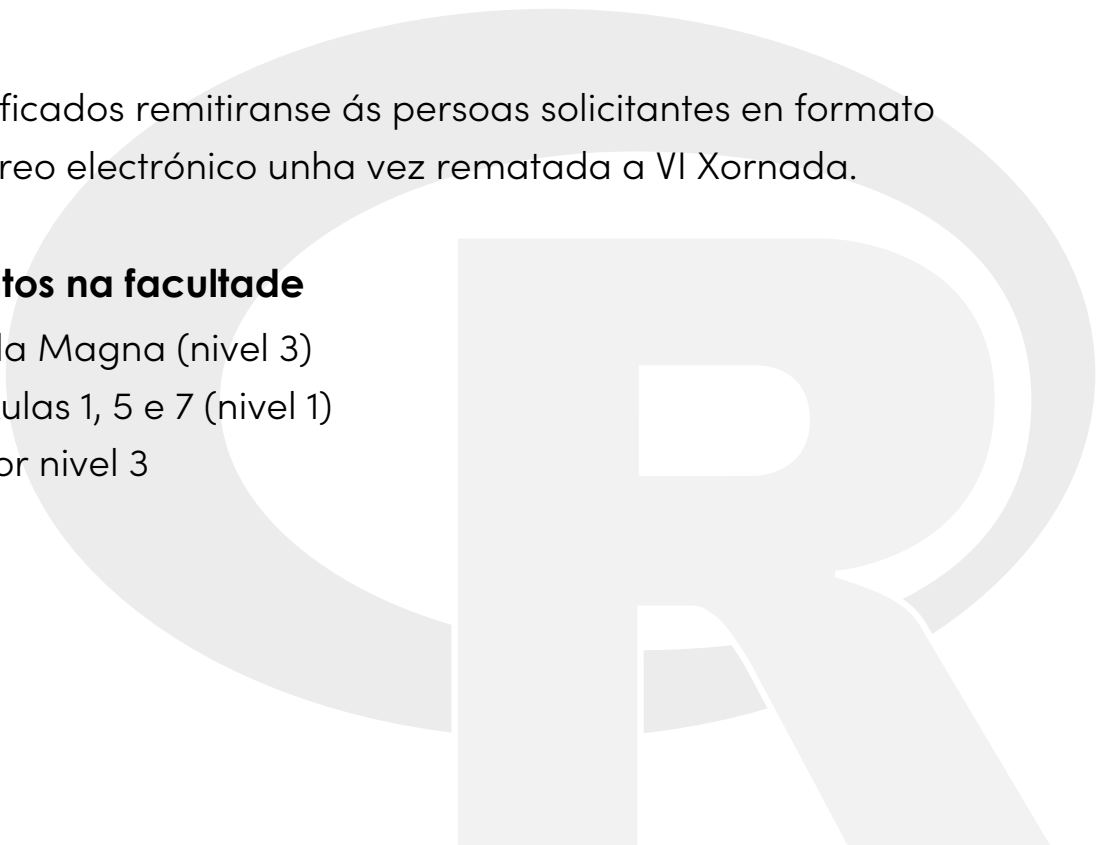
SSID: vixornadar  
Usuario: vixornadar  
Contrasinal: LzV#2ZgqXsYf

### **Certificados**

Todos os certificados remitiranse ás persoas solicitantes en formato dixital por correo electrónico unha vez rematada a VI Xornada.

### **Emprazamentos na facultade**

Relatorios: Aula Magna (nivel 3)  
Obradoiros: Aulas 1, 5 e 7 (nivel 1)  
Cafés: corredor nivel 3



# Programa da xornada

**08:45** Chegada de asistentes e recollida de material

**09:00** Sesión de apertura

**09:20** De Galicia a Ecuador. Ida e volta en R  
 Salvador Naya Fernández, Javier Tarrío Saavedra, Antonio Meneses, Miguel Flores e Yaroslava Robles  
 > Universidade da Coruña (UDC)

**09:40** Datos circulares con R  
 María Alonso Pena  
 > Universidade de Santiago de Compostela (USC)

**10:00** El nuevo estándar simple features para datos vectoriales en R: el paquete sf  
 Dominic Royé  
 > Universidade de Santiago de Compostela (USC)

**10:20** Implementación en R de algoritmos exactos y aproximados para el cálculo del core-center en problemas de aeropuerto y de bancarrota  
 Nadine Helena Espinoza Burgos e Iago Núñez Lugilde  
 > Universidade de Vigo (UVigo)

**10:40** R packages to study missing and longitudinal data  
 Susana Rafaela Martins  
 > Vigo University, Viana do Castelo Polytechnic Institute, and Higher Institute of Administration

**11:00** Pausa - café

**11:40** Escenarios de Stress Test de Riesgo de Crédito con Shiny  
 Sabela Rodríguez Calvo, Juan Anido Rodríguez e Víctor González Carro  
 > ABANCA

**12:00** O emprego de R na planificación estatística  
 María Martín Vila, Estefanía Villar Cheda e Carlos Iglesias Patiño  
 > Instituto Galego de Estatística (IGE)

**12:20** Cociña rápida de gráficos elegantes con R  
 Esther López Vizcaíno  
 > Instituto Galego de Estatística (IGE)

**12:40** Algúns exemplos de uso da librería ggplot2  
 Marcos Fernández Arias  
 > Axencia para a Modernización Tecnolóxica de Galicia (AMTEGA) - Xunta de Galicia

**13:00** Aplicación de R para mantenimiento predictivo en la industria 4.0  
 Manuel Vaamonde Rivas, Manuel Febrero Bande e Beatriz Pateiro López  
 > Instituto Tecnolóxico de Matemática Industrial (ITMATI)

**13:20** Detección de incendios con datos de voos  
 Manuel Antonio Novo Pérez, Ana Belén Buiide Carballosa e María José Ginzo Villamayor  
 > Instituto Tecnolóxico de Matemática Industrial (ITMATI)

**13:40** O proxecto shinyapps de MeLiSA  
 Miguel Ángel Rodríguez Muíños, María José Ginzo Villamayor e Rafael Rodríguez Gayoso  
 > Asociación de Usuarios de Software Libre da Terra de Melide (MeLiSA)

**14:00** Pausa

**16:00** Bdpur: un paquete en R para executar pipelines personalizables e preprocesado sobre fontes de datos heteroxéneos  
 Miguel Ferreiro Díaz, Tomás R. Cotos Yáñez e David Ruano Ordás  
 > Universidade de Vigo (UVigo)

**16:20** Análise de datos de Web of Science co paquete scimetr  
 Rubén Fernández Casal, Borja Lafuente Rego, María José Lombardía Cortiña, Javier Tarrío Saavedra e Julián Costa Bouzas  
 > Universidade da Coruña (UDC)

**16:40** Ferramentas para xestionar proxectos en R  
 José Luis Juncal Pérez  
 > Manserga y otros servicios, S.L.

**17:00** Confección de artigos científicos con R e Markdown  
 Aurora Baluja González  
 > Servizo de Anestesioloxía e Reanimación. Hospital Clínico Universitario. FIDIS. Santiago de Compostela

**17:20** Presentación do concurso MeLiSA con R  
 Miguel Ángel Rodríguez Muíños, María José Ginzo Villamayor e Rafael Rodríguez Gayoso  
 > Asociación de Usuarios de Software Libre da Terra de Melide (MeLiSA)

<b>17:40</b> a <b>19:40</b>	Obradoiro Aula 1	Obradoiro Aula 5	Obradoiro Aula 7
	<b>Introducción ao software R</b> María José Ginzo Villamayor Universidade de Santiago de Compostela (USC)	<b>Visualización de datos espaciais con R</b> María Isabel Borrajo García e Mercedes Conde Amboage Universidade de Santiago de Compostela (USC)	<b>Multivariate and Functional Data Classification in R</b> Manuel Oviedo de la Fuente Universidade de Santiago de Compostela (USC)

## Índice

De Galicia a Ecuador. Ida e volta en R. <i>Salvador Naya Fernández, Javier Tarrío Saavedra, Antonio Meneses, Miguel Flores e Yaroslava Robles. Universidade da Coruña (UDC)</i> .....	28
Datos circulares con R. <i>María Alonso Pena. Universidade de Santiago de Compostela (USC)</i> .....	3
El nuevo estándar simple features para datos vectoriales en R: el paquete sf. <i>Dominic Roye. Universidade de Santiago de Compostela (USC)</i> .....	39
Implementación en R de algoritmos exactos y aproximados para el cálculo del core-center en problemas del aeropuerto y de bancarrota. <i>Nadine Helena Espinoza Burgos e Iago Núñez Lugilde. Universidade de Vigo (UVigo)</i> .....	6
R packages to study missing and longitudinal data. <i>Susana Rafaela Martins. Vigo University, Viana do Castelo Polytechnic Institute, and Higher Institute of Administration</i> .....	21
Escenarios de Stress Test de Riesgo de Crédito con Shiny. <i>Sabela Rodríguez Calvo, Juan Anido Rodríguez y Víctor González Carro. ABANCA</i> .....	35
O emprego de R na planificación estatística. <i>María Martín Vila, Estefania Villar Cheda e Carlos Iglesias Patiño. Instituto Galego de Estatística (IGE)</i> .....	25
Cociña rápida de gráficos elegantes con R. <i>Esther López Vizcaíno. Instituto Galego de Estatística (IGE)</i> .....	17
Alguns exemplos de uso da librería ggplot2. <i>Marcos Fernández Arias. Xunta de Galicia</i> .....	9
Aplicación de R para mantenimiento predictivo en la industria 4.0. <i>Manuel Vaamonde Rivas, Manuel Febrero Bande e Beatriz Pateiro López. Instituto Tecnológico de Matemática Industrial (ITMATI)</i> .....	43
Detección de incendios con datos de voos. <i>Manuel Antonio Novo Pérez, Ana Belén Buide Carballosa e M<sup>a</sup> José Ginzo Villamayor. Instituto Tecnológico de Matemática Industrial (ITMATI)</i> .....	32
O proxecto shinyapps de MeLiSA. <i>Miguel Ángel Rodríguez Muíños, M<sup>a</sup> José Ginzo Villamayor e Rafael Rodríguez Gayoso. Asociación de Usuarios de Software Libre da Terra de Melide (MeLiSA)</i> .....	38
Bdpar: un paquete en R para executar pipelines personalizables de preprocesado sobre fontes de datos heteroxéneos. <i>Miguel Ferreiro Díaz, Tomás R. Cotos Yáñez e David Ruano Ordás Universidade de Vigo (UVigo)</i> .....	12
Análise de datos de Web of Science con el paquete scimetr. <i>Rubén Fernández Casal, Borja Lafuente Rego, María José Lombardía, Javier Tarrío Saavedra, Julián Costa Bouzas. Universidade da Coruña (UDC)</i> .....	11
Ferramentas para xestionar proxectos en R. <i>José Luis Juncal Pérez. Manserga y otros servicios, SL</i> .....	16
Confección de artigos científicos con R e Markdown. <i>Aurora Baluja González. Servizo de Anestesioloxía e Reanimación. Hospital Clínico Universitario. FIDIS. Santiago de Compostela</i> .....	5

Presentación do "concurso MeLiSA" con R. <i>Miguel Ángel Rodríguez Muíños, M<sup>a</sup> José Ginzo Villamayor e Rafael Rodríguez Gayoso. Asociación de Usuarios de Software Libre da Terra de Melide (MeLiSA)</i> .....	37
Obradoiro: Introducción ao software R. <i>M<sup>a</sup> José Ginzo Villamayor. Universidade de Santiago de Compostela (USC)</i> .....	48
Visualización de datos espaciais con R. <i>María Isabel Borrajo García e Mercedes Conde Amboage. Universidade de Santiago de Compostela (USC)</i> .....	46
Multivariate and Functional Data Classification in R. <i>Manuel Oviedo de la Fuente. Universidade de Santiago de Compostela (USC)</i> .....	49
AUTORES.....	51

## DATOS CIRCULARES CON R

María Alonso Pena<sup>1</sup>

<sup>1</sup>Departamento de Estatística, Análise Matemática e Optimización, Universidade de Santiago de Compostela

### RESUMO

Os datos circulares son aqueles que poden representarse como puntos na circunferencia unidade. Unha vez elixidas unha dirección inicial e un sentido de rotación, estes puntos poden ser expresados como ángulos. Os exemplos máis comúns deste tipo de datos son direccións ou observacións con compoñentes estacionarias (ocorrencias ao longo do tempo). Os datos circulares poden ser atopados en disciplinas moi diferentes: bioloxía (Scapini *et al.*, 2002), xeoloxía (SenGupta e Rao, 1966), medicina (Mooney *et al.*, 2003) ou na industria automobilística (Anderson-Cook, 1999).

A periodicidade dos datos fai que as técnicas estatísticas clásicas non sexan axeitadas para tratar este tipo de observacións, o que implica a necesidade de desenvolver novos métodos que permitan unha análise apropiada. Os recentes avances tecnolóxicos permitiron obter métodos para rexistrar datos circulares de xeito preciso, o que impulsou o interese neste campo nos últimos anos (Pewsey *et al.*, 2013; Ley e Verdebout, 2017). Cando consideramos máis dunha variable, as ferramentas estatísticas pódense estender a outras variedades como o cilindro (unha variable circular e unha variable real) ou o toro (dúas variables circulares).

Coa fin de aplicar os distintos métodos estatísticos na práctica e poñelos a disposición dos usuarios, moitas destas técnicas foron implementadas no software estatístico R (R Core Team, 2018). O obxectivo desta charla é introducir ferramentas de R que permitan a análise de datos circulares, cilíndricos ou toroidais. En concreto, a librería `circular` (Agostellini e Lund, 2017) contén ferramentas para realizar unha análise básica de datos circulares, ademais de diferentes técnicas estatísticas, mentres que a librería `NPCirc` (Oliveira *et al.*, 2014) céntrase en métodos non paramétricos.

**Palabras e frases chave:** Datos circulares, estimación da densidade, regresión non paramétrica.

## Referencias

- [1] Agostinelli, C. e Lund, U. (2017). R package ‘circular’: Circular statistics (version 0.4-93). URL <https://r-forge.r-project.org/projects/circular/>.
- [2] Anderson-Cook, C.M. (1999). A tutorial on one-way analysis of circular-linear data. *Journal of Quality Technology*, 31, 109–119.
- [3] Ley, C. e Verdebout, T. (2017). *Modern Directional Statistics*. Chapman & Hall, Boca Raton.
- [4] Mooney, J.A., Helms, P.J. e Jolliffe, I.T. (2003). Fitting mixtures of von Mises distributions: a case study involving sudden infant death syndrome. *Computational Statistics & Data Analysis*, 41, 505–513.
- [5] Oliveira, M., Crujeiras, R.M. e Rodríguez-Casal, A. (2014). NPCirc : An R package for nonparametric circular methods. *Journal of Statistical Software*, 61, 1–26.

- [6] Pewsey, A., Neuhuser, M. e Ruxton, G.D. (2013). *Circular Statistics in R*. Oxford University Press, Oxford.
- [7] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [8] Scapini, F., Aloia, A., Bouslama, M.F., Chelazzi, L., Colombini, I., El Gtari, M., Fallaci, M., e Marchetti, G.M. (2002). Multiple regression analysis of the sources of variation in orientation of two sympatric sandhoppers, *Talitrus saltator* and *Talorchestia bito*, from an exposed Mediterranean beach. *Behavioural Ecology and Sociobiology*, 51, 403–414.
- [9] SenGupta, S. e Rao, J.S. (1966). Statistical analysis of cross-bedding azimuths from the Kamthi formation around Bheemaram, Pranhita: Godavari Valley. *Sankhya: The Indian Journal of Statistics*, Series B, 28, 165–174.



VI Xornada de Usuarios de R en Galicia  
Santiago de Compostela, 10 de outubro do 2019

## CONFECCION DE ARTICOS CIENTIFICOS CON R E MARKDOWN

Aurora Baluja González<sup>1</sup>

<sup>1</sup>Servizo de Anestesioloxía e Reanimación. Hospital Clínico Universitario. FIDIS. Santiago de Compostela, España

### RESUMO

Un dos obxectivos do xornalismo científico aplicado á contorna de R é que as análises poidan ser reproducibles. Para axudar nesta tarefa, é posible integrar nun só documento o noso código de R, os nosos resultados (sexan táboas ou gráficos) e a nosa bibliografía.

**Palabras y frases clave:** reproducibilidade, R-Studio, markdown, plantilla, documento, software libre.

### 1. INTRODUCCIÓN

A integración de análises de datos e bibliografía nun só documento é unha das vantaxes que nos ofrece o paquete `rmarkdown` de R. Veremos o proceso polo que poderemos crear este tipo de materiais, listos para o seu envío a unha revista científica.

### 2. MATERIAIS E MÉTODOS

Utilizando R e mais R-Studio, mostrarase a estrutura básica que debe ter un `rmarkdown` para se poder compilar coma un artigo aceptable para revisión. Faremos uso da configuración dispoñible no encabezamento `YAML` dos ficheiros `rmarkdown` (`Rmd`) para a súa compilación en formato Office, utilizando unha plantilla e diversos estilos bibliográficos coa axuda de `Zotero`. Asemade, verémo-lo xeito de modificar dita plantilla para nos achegar ao estilo de diferentes publicacións.

## Referencias

- [1] Allaire, J.J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., Iannone, R.. *rmarkdown: Dynamic Documents for R*. R package version 1.16. 2019.

VI Xornada de Usuarios de R en Galicia  
Santiago de Compostela, 10 de outubro do 2019

## IMPLEMENTACIÓN EN R DE ALGORITMOS EXACTOS Y APROXIMADOS PARA EL CÁLCULO DEL CORE-CENTER EN PROBLEMAS DEL AEROPUERTO Y DE BANCARROTA

Nadine Helena Espinoza Burgos<sup>1</sup> e Iago Núñez Lugilde<sup>1</sup>

<sup>1</sup>Grupo SiDOR. Técnico de apoyo a la Investigación. Universidade de Vigo

### RESUMEN

El core-center fue introducido por González-Díaz y Sánchez-Rodríguez en 2007 [7], como solución para juegos coalicionales de núcleo (core) no vacío. Las propiedades que verifica esta solución (véanse los trabajos [7], [5], [10] y [12]) justifican que sea de interés disponer de algoritmos, tanto exactos como aproximados, para su cálculo en problemas aplicados. Presentamos un paquete en R que dispone de diferentes algoritmos para su cálculo en dos problemas de interés: los problemas de bancarrota y los problemas del aeropuerto.

**Palabras y frases clave:** Juego coalicional, Core, Core-center, Problema del aeropuerto, Problema de bancarrota.

### 1. INTRODUCCIÓN

Los juegos coalicionales modelizan situaciones en las que la cooperación entre los agentes frecuentemente supone un beneficio para la coalición total que debe ser repartido. En determinados problemas se reparten beneficios y en otros son costes a sufragar entre los diferentes jugadores involucrados. Las diversas soluciones planteadas en juegos cooperativos, entre ellas el core-center, proponen repartos de los beneficios/costes obtenidos tras la cooperación.

Cuando existen repartos que dividen la utilidad total, en las que ningún grupo o coalición tiene incentivos a replicar objetando su disconformidad con el reparto propuesto, se dice que el core [7] del juego asociado es no vacío. Este conjunto de asignaciones es un politopo compacto y convexo contenido en un hiperplano, como ocurre en los problemas de estudio, juegos de aeropuerto, Figura 1a, y de bancarrota, Figura 1b. Por lo tanto, es razonable buscar soluciones que se encuentren en el core y, entre todos los posibles repartos, si queremos tener una solución ecuánime, el centro de gravedad del núcleo, el core-center, es aquella solución que trata por igual a todas las asignaciones seleccionadas por el core del juego, dado que promedia este conjunto considerando la distribución uniforme sobre el mismo.

## 2. PROBLEMAS DEL AEROPUERTO

El problema del aeropuerto y su juego asociado, Littlechild, S.C. and Owen, G (1973) [9], pertenece a la clase de juegos de costes. El objetivo es obtener un reparto del coste total del mantenimiento de una pista de aterrizaje compartida entre diversas compañías aéreas, en las que se considera el uso que hace cada aerolínea de la pista, por ejemplo, con el número de vuelos programados.

Existe una fórmula explícita para el cálculo del core-center en función del volumen del core en esta clase de juegos [6]. Sin embargo, abarcamos diferentes algoritmos para su estimación y proponemos nuevos métodos para su cálculo exacto abordando su cálculo cuando el número de jugadores es elevado. Así, estudiando en profundidad el cálculo del core-center en estos juegos podemos extender métodos de aproximación para juegos en los que el core-center no tiene formulación explícita y también simplificar los nuevos métodos exactos en determinados casos como son los juegos de aeropuerto con jugadores simétricos.

Los métodos de aproximación utilizados (ver [2] y [14]) se basan en el carácter ecuánime de esta solución, simulando una muestra uniformemente distribuida en nuestro core y tomando como aproximación del core-center la media muestral. Se utiliza Montecarlo para realizar una estimación con menos error. En el primer método exacto aplicamos resultados existentes para reformular los volúmenes del core [8], posteriormente utilizamos este mismo enfoque para obtener una nueva expresión de nuestra solución y, por último, implementamos un nuevo método basado en la teselación semiregular de nuestro núcleo.

## 3. PROBLEMAS DE BANCARROTA

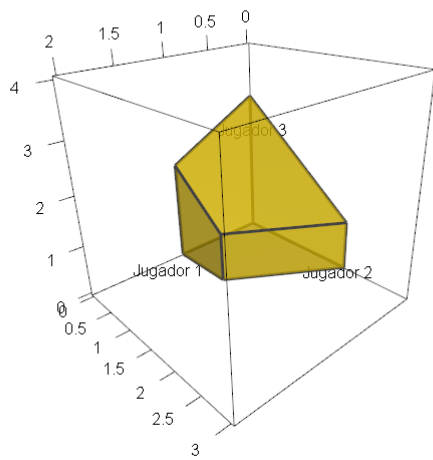
Un problema de bancarrota, [13], surge cuando un recurso no es suficiente para satisfacer todas las demandas de un conjunto de acreedores. En Aumann y Maschler (1985) [1] se presenta un análisis de este problema utilizando juegos cooperativos y se introducen diferentes soluciones a dicho problema. La teoría desarrollada para los problemas de bancarrota también es aplicable en otro contexto de gran interés: el reparto de un bien público, y en particular, se puede utilizar para determinar los impuestos que deben ser atribuidos a cada ciudadano en función de su salario bruto, lo que determinará su salario neto.

Una de las características de los juegos asociados a los problemas de bancarrota es que tienen core no vacío. Sin embargo, para esta clase de juegos no disponemos de una fórmula explícita del core-center, por lo que tendríamos que recurrir al uso de la integración para su cálculo. Por este motivo, hacemos uso de propiedades especiales del núcleo de los juegos de bancarrota que nos permiten poder computarlo mediante diferentes algoritmos.

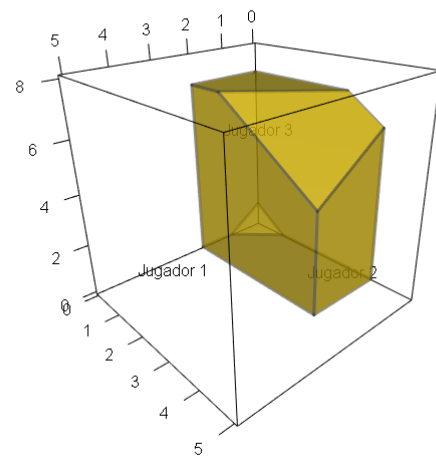
El núcleo de estos juegos está contenido en otro conjunto, denominado conjunto de imputaciones, del que podemos calcular fácilmente su centro de gravedad. Haciendo uso de este conjunto y, utilizando los conceptos arquimedianos del centro de masas, tendremos un primer método basado en la teselación de este conjunto de imputaciones [11]. Además de esta teselación exterior, implementamos otro algoritmo, en el que diseccionamos el núcleo de cualquier problema de bancarrota en diferentes núcleos de juegos de la misma clase, de los cuales conocemos su centro de gravedad, lo que nos dará una segunda forma de calcular esta solución. Combinaremos estos dos métodos para conseguir un algoritmo óptimo que vendrá en función de las restricciones que determinan el politopo. Por último, presentamos un cuarto algoritmo alternativo en función de las caras del del núcleo de juego siguiendo la formulación obtenida por Lasserre (1983) [8].

## AGRADECIMIENTOS

Este trabajo ha sido tutorizado por Estela Sánchez Rodríguez, a quien agradecemos que nos haya iniciado en líneas de investigación que han sido la base para éste y otros proyectos. Ha sido financiado por el grupo SIDOR de la Universidade de Vigo. Proyectos involucrados: the European Regional Development Fund (ERDF), Ministerio de Economía, Industria y Competitividad, proyecto MTM2017-87197-C3-2-P; Grupo de Referencia Competitiva, 2016-2019 (ED431C 2016/040), financiado pola Consellería de Cultura, Educación e Ordenación Universitaria da Xunta de Galicia.



(a) Juego del Aeropuerto



(b) Juego de Bancarrota

Figura 1: Core de juegos con 4 jugadores

## Referencias

- [1] Aumann R., Maschler M. (1985) Game theoretic analysis of a bankruptcy problem from the Talmud. *Journal of Economic Theory* 36, 195–213.
- [2] Cao Abad, R. (2002) Introducción a la simulación y a la teoría de colas. Netbiblo.
- [3] Emiris, I.Z. y Fisikopoulos, V. (2018). Practical polytope volume approximation. *ACM Transactions on Mathematical Software*, 44, n. 4.
- [4] Gillies, D.B. (1953) Some theorems on n-person games. PhD thesis, Princeton University.
- [5] González Díaz, J., Mirás Calvo, M. A., Quinteiro Sandomingo, C. and Sánchez Rodríguez, E. (2015) Monotonicity of the core-center of the airport game. *TOP*, 23(3), 773–798.
- [6] González Díaz, J., Mirás Calvo, M. A., Quinteiro Sandomingo, C. and Sánchez Rodríguez, E. (2016) Airport games: The core and its center. *Mathematical Social Sciences*, 82, 105–115.
- [7] González Díaz, J. and Sánchez Rodríguez, E. (2007) A natural selection from the core of a tu game: the core-center. *International Journal of Game Theory*, 36, 1, 27–46.
- [8] Lasserre, J. B. (1983) An analytical expression and an algorithm for the volume of a convex polyhedron. *Journal of optimization theory and applications*, 39, 3, 363–377.
- [9] Littlechild, S.C. and Owen, G (1973). A simple expression for the Shapley value in a special case. *Management Science*, 20, 370–372.
- [10] Mirás Calvo, M.A., Quinteiro Sandomingo, C. and Sánchez Rodríguez, E. (2016). Monotonicity implications for the ranking of rules for airport problems. *International Journal of Economic Theory*, 12, 379–400.
- [11] Mirás Calvo, M.A., Quinteiro Sandomingo, C. and Sánchez Rodríguez, E. (2019). Compensating utopical allocations in bankruptcy games
- [12] Mirás Calvo, M.A., Quinteiro Sandomingo, C. and Sánchez Rodríguez, E. (2019). The core-center of a bankruptcy problem. Preprint
- [13] O'Neill, B. (1982). A problem of rights arbitration from the Talmud. *Mathematical Social Sciences* 2, 345–371.
- [14] Smith, R.L. (1984). Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32, 6, 1296–1308.

VI Xornada de Usuarios de R en Galicia  
Santiago de Compostela, 10 de outubro do 2019

## ALGÚNS EXEMPLOS DE USO DA LIBRERÍA GGPLOT2

Marcos Fernández Arias<sup>1</sup>

<sup>1</sup>Axencia Galega de Modernización Tecnolóxica de Galicia, Xunta de Galicia

### RESUMO

Descubriremos varios casos prácticos do uso da librería gráfica declarativa `ggplot2`[1] e as librerías auxiliares `ggTimeSeries`[2] e `ggplotly`[3] para analizar series temporais. Explicaranse os exemplos mostrados nas Figuras 1-4.

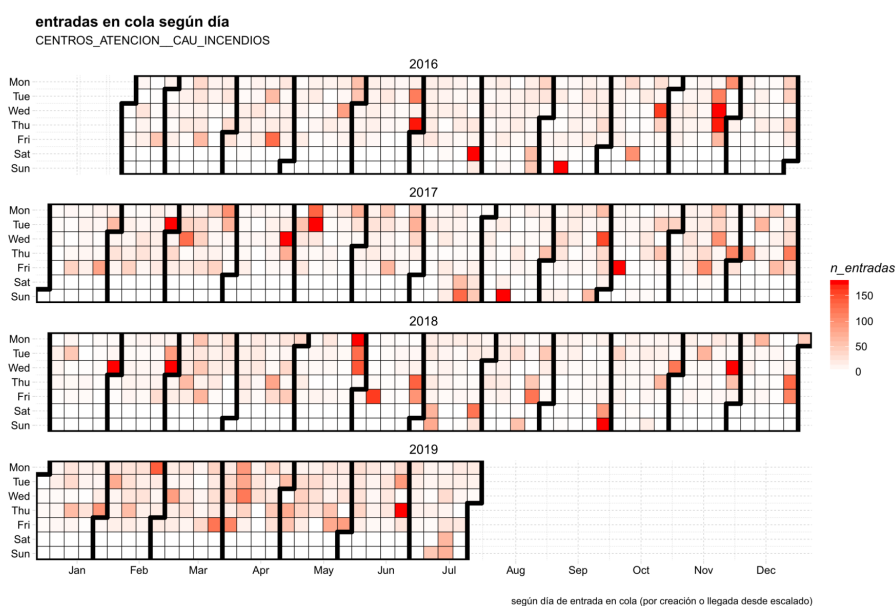


Figura 1: Exemplo.

**Palabras e frases chave:** gráficos, `ggplot2`, series temporais.

## Referencias

- [1] Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. <https://ggplot2.tidyverse.org/>
- [2] Kothari, A. *ggTimeSeries: Time Series Visualisations Using the Grammar of Graphics*. <https://github.com/Ather-Energy/ggTimeSeries>
- [3] Sievert, C. (2018) *plotly for R*. <https://plotly-book.cpsievert.me>

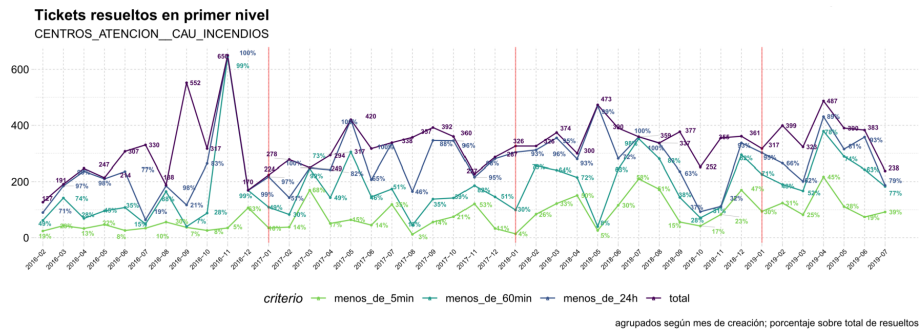


Figura 2: Ejemplo.

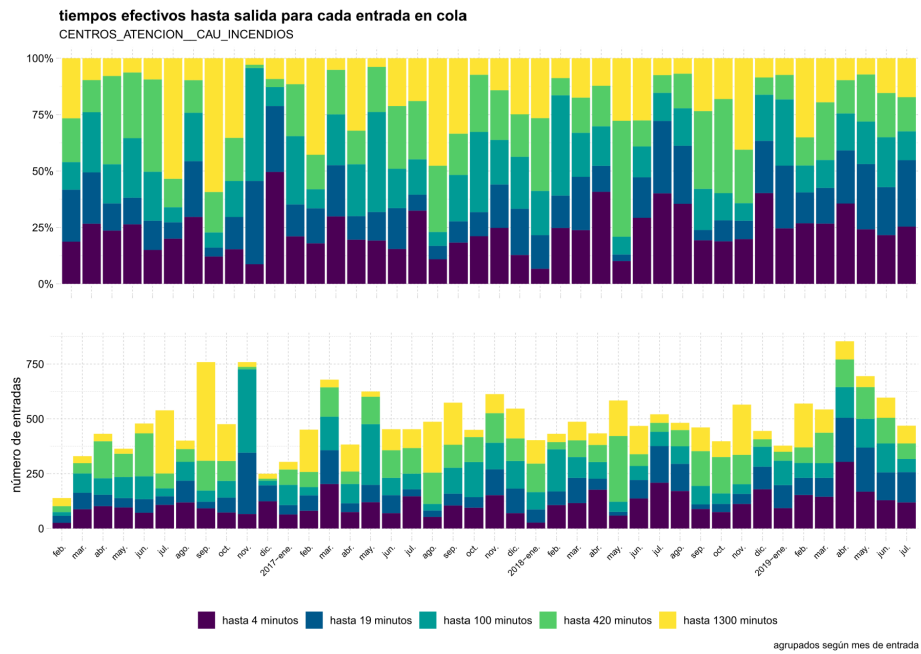


Figura 3: Ejemplo.

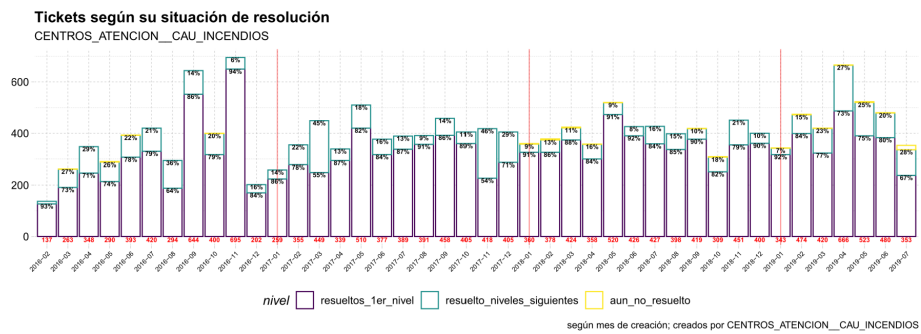


Figura 4: Ejemplo.

*VI Xornada de Usuarios de R en Galicia*  
*Santiago de Compostela, 10 de outubro do 2019*

## **ANÁLISE DE DATOS DE WEB OF SCIENCE CO PAQUETE `scimetr`**

Rubén Fernández Casal<sup>1</sup>, Borja Lafuente Rego<sup>1</sup>, María José Lombardía Cortiña<sup>1</sup>, Javier Tarrío Saavedra<sup>1</sup>, Julián Costa Bouzas<sup>1</sup>

<sup>1</sup>Universidade da Coruña

### **RESUMO**

O paquete `scimetr` implementa ferramentas para a investigación en cienciometría e bibliometría. Actualmente inclúe funcións para importar datos bibliográficos da Web of Science de Thomson Reuters (<http://www.webofknowledge.com>) e realizar análises bibliométricos. Para máis información visitar <https://rubenfcasal.github.io/scimetr/articles/scimetr.html>.

## **Bdpar: UN PAQUETE EN R PARA EXECUTAR PIPELINES PERSONALIZABLES DE PREPROCESADO SOBRE FONTES DE DATOS HETEROXÉNEOS**

Miguel Ferreiro-Díaz<sup>1</sup>, Tomás R. Cotos-Yáñez<sup>2</sup> and David Ruano-Ordás<sup>1</sup>

<sup>1</sup>Departamento de Informática. Universidade de Vigo

<sup>2</sup>Departamento de Estatística e Investigación Operativa. Universidade de Vigo

### **RESUMO**

Nos últimos anos, o Big Data convirteuse nun método útil para aproveitar múltiples fontes de información para atopar coñecemento en aplicacións reais (como por exemplo o deseño de campañas de marketing personalizadas ou axudar a paliar o efectos de diversas enfermidades mortais). As ferramentas e métodos de programación de big data evolucionaron co paso do tempo desde un MapReduce ata o paradigma baseado en pipeline. Concretamente, o uso de esquemas de pipeline converteuse no xeito máis fiable de analizar big data. Para iso, este traballo introduce BDPAR, un novo paquete baseado en pipeline altamente personalizable (usando o paradigma OOP proporcionado polas clases R6) capaz de executar varias tarefas de preprocesado en fontes de datos heteroxéneas. Ademais, para aumentar a flexibilidade e o rendemento, BDPAR ofrece funcións útiles como a definición dun operador de pipes mellorado (`%>I%`) ou a capacidade de deseñar e implementar facilmente novos analizadores de datos de entrada, tarefas and pipelines.

**Palabras e frases chave:** paquete de R, R6, pipeline, big data.

### **1. INTRODUCCIÓN**

As redes sociais e as aplicacións de mensaxería instantánea convertéronse, probablemente, nunha parte esencial da experiencia humana. De feito, hoxe en día máis do 60% da poboación procedente de países industrializados usa estes mecanismos para comunicar ou compartir información. Este fenómeno xurdiu debido a (i) a diminución dos custos dos ordenadores e sistemas de almacenamento por un factor de máis de 200 (National Academy of Engineering 1984), (ii) un aumento exponencial da velocidade de procesamento e das capacidades de hardware do ordenador (Iansiti and Khansa, 1995), (iii) a aparición de redes de comunicación de alto rendemento e totalmente dispoñibles (Dorogovtsev and Mendes, 2013), e (iv) determinadas necesidades humanas, como manter interconectadas e ter acceso permanente aos datos (Kabeer, 2005).

Este escenario promoveu un crecemento exponencial na cantidade de datos xerados e almacenados na última década. Concretamente, os últimos informes de 2018 demostraron que ao redor de 2.16EB (exabytes) de datos créanse todos os días (Domo-Data, 2019; VCloud, 2019), e as tendencias están demostrando que o crecemento da información dispoñible é catro veces superior á economía mundial (VCloud, 2019). De feito, o 90% dos datos mundiais creáronse nos dous últimos anos só (IBM, 2019).

Ademais da dispoñibilidade de fontes e ferramentas ilimitadas para xerar, intercambiar e manexar información, a falta dun xeito normalizado de representar datos, provocou un aumento masivo



de información non estruturada. De feito, aproximadamente o 80% dos datos existentes non está estruturado (VCloud (2019), IBM (2019)). Os datos obtidos dunha única fonte adoitan ser insuficientes para levar a cabo un proceso de decisión adecuado. Non obstante, a posibilidade de aproveitar a combinación de datos de múltiples fontes (e non estruturadas) require a execución de operacións preprocesadas que garantan un formato de datos unificado. A necesidade de facilitar a xestión e explotación de grandes cantidades de datos heteroxéneos (en termos de tipos e formatos de datos) nun tempo razoable e dun xeito rendible, provocou a aparición da era Big Data.

Este escenario, xunto coa ausencia de ferramentas dispoñibles capaces de crear pipelines Big Data personalizados en R, motivounos a deseñar e implementar (bdpar (2019)), un marco capaz de unificar datos heteroxéneos mediante o desenvolvemento e execución de pipelines personalizables. Para este fin, o noso paquete permite (i) usar ou desenvolver extractores de contido (como SMS ou analizadores de correo electrónico) e (ii) usar ou implementar novas tarefas de preprocesado (pipe).

Este artigo ofrece unha descrición completa das principais funcionalidades e recursos do paquete bdpar (Ferreiro-Díaz *et al.* (2019)). A versión actual é 1.0.0 e unha lista actualizada (con toda a colección de recursos) está dispoñible no documento da viñeta e no manual de referencia. A sección seguinte ofrece unha descrición completa da estrutura e funcionalidades do paquete. A continuación, descríbese o uso do paquete e finalmente, ofrécese un estudo de caso ilustrativo.

## 2. ESTRUCTURA DO PAQUETE

El paquete *bdpar* está centrado no deseño, implementación e implantación de fluxos de procesamento personalizados para o dominio de datos grandes. Para xestionar a información obtida das fontes de entrada, o paquete utiliza unha estrutura específica chamada INSTANCIA que é a encargada de almacenar as propiedades extraídas por cada tubo que comprende o fluxo de procesamento. A Figura 1 ofrece unha representación gráfica do seu funcionamento interno, que se divide en dúas etapas principais: (i) carga de datos e (ii) execución de pipeline.

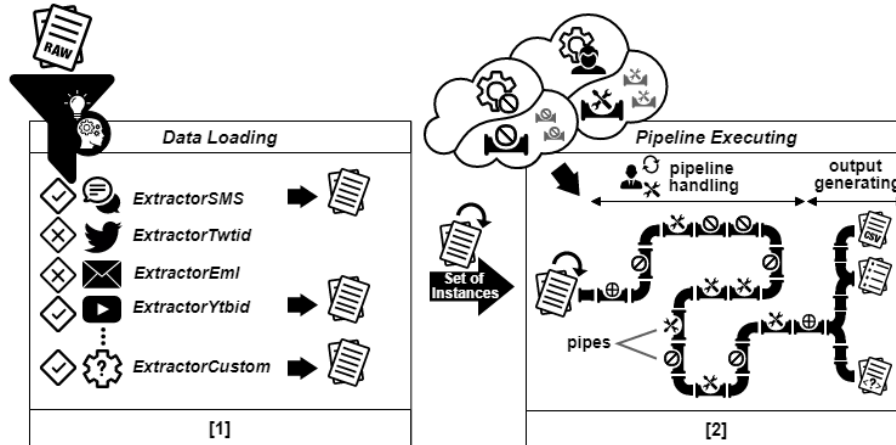


Figura 1: Descrición das dúas etapas do proceso de execución do paquete BDPAR.

Como se mostra na Figura 1, a primeira etapa comprende a carga dos extractores necesarios segundo o tipo de datos de entrada. De xeito predeterminado, BDPAR fornece catro tipos diferentes de extractores: (i) *EXTRACTORSMS* é capaz de extraer os contidos textuais intercambiados a través do servizo de mensaxes curtas (SMS); (ii) *EXTRACTORTWTID* é capaz de obter texto a partir de entradas de Twitter (tweets); (iii) *EXTRACTOREML* é capaz de reunir contido en bruto do corpo de mensaxes de correo electrónico; e finalmente (iv) *EXTRACTORYTBID* está ligado á extracción de comentarios publicados na plataforma YouTube. Ademais, para aumentar a compatibilidade de BDPAR con outros formatos de datos, o marco permite o fácil deseño e implantación de novos extractores personalizados (empregando unha simple relación de herdanza OOP).

Unha vez extraído o contido con éxito das fontes en bruto, a segunda etapa iníciase automaticamente por BDPAR. Esta etapa comprende a execución de dous pasos: (i) manexo de pipelines; e

(ii) xeración de saída. O primeiro é o encargado de realizar o procesamento de datos unificado executando un conxunto específico de tarefas (tamén chamado pipeline) sobre o contido extraído anteriormente. Cómpre destacar que cada tarefa que forma parte do pipeline (tamén chamada pipe) é a responsable de realizar unha operación sinxela sobre os datos de entrada. O segundo paso (chamado xerador de saída) transforma os datos preprocesados nun formato de saída específico (estrutura CSV). Ademais, BDPAR permítelle aos usuarios desenvolver novos métodos específicos de xeración de resultados para conseguir a saída desexada.

Como se pode concretar na fase de manexo de pipeline mostrada na Figura 1, a estrutura baseada en tubos proporciona unha gran flexibilidade e versatilidade aos usuarios xa que permite aos usuarios (i) modificar facilmente os pipelines existentes engadindo ou eliminando tarefas, (ii) desenvolver novas tarefas ou (iii) deseñar e implantar novos pipelines personalizados. Para axudar aos usuarios na creación e implantación de novos pipelines, BDPAR fornece un conxunto de 18 pipes combinables que implementan tarefas básicas de preprocesamento. As tarefas incluídas en BDPAR divídense en dúas categorías diferentes: (i) as que implementan funcionalidades básicas; e (ii) os que poidan acceder a ficheiros externos. Ademais, pódense implementar e con facilidade novas tarefas mediante unha simple relación de herdanza OOP (creando unha subclase de PIPEGENERIC).

### 3. EXEMPLO DE APLICACIÓN

Para ilustrar a funcionalidade do paquete BDPAR, colléronse 20 arquivos SMS (10 lexítimos + 10 spam) e 20 correos electrónicos (10 spam + 10 non spam) preprocesáronse e obtiveronse os seguintes wordclouds:

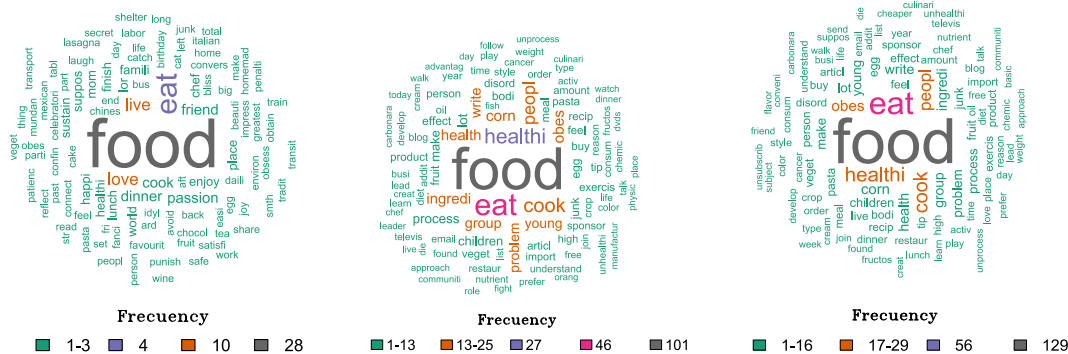


Figura 2: Wordcloud para SMS (esquerda), e-mail (centro) e para ambos conxuntamente (dereita).

### 4. CONCLUSIÓNS E TRABALLO FUTURO

Neste traballo, introducimos BDPAR (Ferreiro-Díaz *et al.* 2019), un paquete de R baseado en pipes para facilitar a creación de conxuntos de datos unificados a partir de fontes heteroxéneas. O noso marco permite aos usuarios (i) definir novos extractores de contido (anlizadores de datos), (ii) desenvolver e implementar novas tarefas de preprocesamento (pipes) e (iii) definir e construír fluxos de tarefas interconectadas personalizadas (pipeline). Ademais, para aforrar recursos computacionais e aumentar a velocidade de execución, BDPAR fornece un operador de pipes optimizado (anotado como  $\%>I\%$ ) capaz de abortar o procesamento dunha instancia se se detectase un erro. Finalmente, elaborouse un estudo de caso para demostrar a capacidade do marco para preprocesar e unificar datos heteroxéneos nun único ficheiro CSV.

O traballo futuro céntrase en dous aspectos principais: (i) o desenvolvemento de tarefas baseadas en semánticas capaces de explotar as relacións semánticas entre sinets e; (ii) o deseño de esquemas de execución para distribuír os requirimentos computacionais nun clúster de ordenadores.

### AGRADECIMENTOS

The second author's work has been partially supported by the Grant IN2017-84658-C2-1-R of the Spanish Ministry of Industry, the third author's work has been supported by a post-doctoral fellowship from Xunta de Galicia (ED481B 2017/018). SING group thanks CITI (Centro de Investigación, Transferencia e Innovación) from the University of Vigo for hosting its IT infrastructure.

## Referencias

- [1] Domo-Data (2019). Domo-data never sleeps 6.0. <https://www.domo.com/solution/data-never-sleeps-6>. Accessed: 2019-04-05.
- [2] Dorogovtsev, S., Mendes J. (2013). Evolution of Networks: From Biological Nets to the Internet and WWW. OUP Oxford. <https://books.google.es/books?id=FFL1AgAAQBAJ>.
- [3] Ferreiro-Díaz, M., Cotos-Yáñez, T.R., Ruano-Ordas, D. (2019). bdpar: Big Data Preprocessing Architecture. R package version 1.0.0. <https://CRAN.R-project.org/package=bdpar>.
- [4] Iansiti, M., Khansa, T. (1995). Technological Evolution, System Architecture and the Obsolescence of Firm Capabilities. *Industrial and Corporate Change*, 4(2):333–361. ISSN 0960–6491. <https://doi.org/10.1093/icc/4.2.333>.
- [5] IBM (2019). Ibm big data success stories. <http://public.dhe.ibm.com/software/data/sw-library/bigdata/ibm-big-data-success.pdf>. Accessed: 2019-04-8.
- [6] Kabeer, N. (2005) Introduction: The search for inclusive citizenship: Meanings and expressions in an interconnected world. In N Kabeer (ed.) *Inclusive Citizenship: Meanings and Expressions*.
- [7] National Academy of Engineering (1984). Cutting Edge Technologies. *The National Academies Press*, Washington, DC. <https://www.nap.edu/catalog/286/cuttingedge-technologies>.
- [8] VCloud (2019). Vcloud news. Accessed: 2019-04-17.

## FERRAMENTAS PARA XESTIONAR PROXECTOS EN R

José Luis Juncal Pérez<sup>1</sup>

<sup>1</sup>Manserga y otros servicios, SL

### RESUMO

Introdución ó uso de ferramentas para a xestión do ciclo de vida dun proxecto complexo en R, no que hai que traballar con distintos desenroladores, equipos e configuracións durante un período longo de tempo. Mostrarase a necesidade que cobren e casos de uso.

**Palabras e frases chave:** desenrolo de software, xestión de proxectos.

### 1. INTRODUCCIÓN

Ante o aumento da complexidade nos proxectos en R, necesitamos utilizar distintas ferramentas que nos axuden a xestionar as modificacións que facemos nos 3 grandes compoñentes dun proxecto de análise de datos:

- código → `Git` - rexistro de cambios no código,
- datos → `Drake`[2]- rexistro de cambios e actualización nos datos e
- paquetes → `Packrat`[1] - rexistro dos paquetes utilizados.

Con estas tres ferramentas conseguimos que a xestión do proxecto sexa moito máis sinxela, podendo retomalo estado (snapshot) do mesmo meses despois sen ter que configurar o noso entorno de novo. Permiten que os distintos colaboradores poidan traballar inmediatamente sobre un mesmo entorno dun xeito cómodo e rápido, ou mesmo pedir axuda a un colaborador externo de forma sinxela. Tamén facer alusión ó contedor `rocker`[1], unha forma de conseguir ter un entorno común e homoxéneo entre tódolos compoñentes do equipo, e poder ter versións de R distintas segundo as necesidades.

### Referencias

- [1] Upadhyay, N., (2016) *Introduction to packrat package*. [https://rstudio-pubsstatic.s3.amazonaws.com/221948\\_fb7215fecb0d49ac903f701fd8d45132.html](https://rstudio-pubsstatic.s3.amazonaws.com/221948_fb7215fecb0d49ac903f701fd8d45132.html)
- [2] Landau, W. (2018) *The prequel to the drake R package*. <https://github.com/ropensci/drake>

## COCIÑA RÁPIDA DE GRÁFICOS ELEGANTES CON R

M<sup>a</sup> Esther López Vizcaíno<sup>1</sup>

<sup>1</sup>Instituto Galego de Estatística

### RESUMO

Neste relatorio exporase como de forma sinxela e rápida se poden elaborar gráficos moi elegantes coa librería `ggplot2` de R. A elaboración pódese facer con poucos ingredientes e permitirá unha mellor interpretación da información estatística.

**Palabras e frases chave:** visualización da información, gráficos, `ggplot2`, R, `tidyverse`.

### 1. INTRODUCCIÓN

A análise estatística de datos ten a misión de descubrir patróns de comportamento dos fenómenos que se presentan a partir do estudo da información dispoñible. Unha das etapas iniciais do proceso de análise constitúe a representación gráfica da información, o que, na maioría dos casos, se presenta como a pista fundamental para proceder a un mellor tratamento dos datos. Unha boa visualización mostrarache detalles que non esperabas ou fará xurdir novas preguntas acerca dos datos. Tamén pode dar pistas acerca de se estás facendo as preguntas equivocadas ou se necesitas datos diferentes.

Existen múltiples softwares para facer representación de gráficos estatísticos, este relatorio está centrado nos gráficos que se poden elaborar co software libre R. Dos moitos sistemas que posúe R para facer gráficos, `ggplot2` é un dos máis elegantes e versátiles. A librería `ggplot2` permite de maneira sinxela crear gráficos complexos a partir dun conxunto de datos. Este paquete ofrece unha estrutura para que poidamos especificar que variables imos pintar, como se poden presentar e outras propiedades visuais xerais. Desta maneira, só necesitamos realizar cambios mínimos se os datos sofren algunha modificación ou se decidimos pasar, por exemplo, dun gráfico de barras a un diagrama de dispersión.

### 2. A LIBRERÍA `ggplot2`

A librería `ggplot2` baséase nun sistema coherente para describir e construír gráficos, coñecido como a gramática de gráficos, unha idea que expón que calquera gráfico pode expresarse a partir da combinación destes compoñentes: un conxunto de datos, un sistema de coordenadas e un conxunto de geoms (que determinan a representación visual dos datos).

A clave para entender `ggplot2` é pensar nunha figura como un conxunto de capas. Isto quere dicir que os gráficos de `ggplot2` constrúense paso a paso agregando novos ingredientes ou capas, resultando nunha gran flexibilidade para a personalización dos mesmos.

Para construír un `ggplot`, empregaremos esta plantilla básica que pode usarse para distintos tipos de gráficos:

```
ggplot(data = datos, mapping = aes())+geom_funcion()
```

Emprégase a función `ggplot()` para vincular o gráfico a un conxunto de datos a través do argumento `data` (neste caso o arquivo chámase `datos`). Para traballar coas funcións de `ggplot2` é recomendable ter unha columna para cada variable e unha fila para cada observación.

Agora é necesario definir como se queren representar as variables a través do argumento `mapping`. Defínese empregando a función `aes()` dentro de `ggplot()`. Para seleccionar as variables que queremos representar e especificar como as queremos presentar, por exemplo, nos eixos `x/y` ou as características tales como tamaño, forma, color, etc. empregamos `aes()`. `aes()` dille a `ggplot` cómo se relaciona cada unha das variables nos datos coas propiedades `aesthetic` (estéticas) da figura. Neste caso, dicímoslle a `ggplot` que queremos representar a columna "Idade.Media" do data frame "datos" no eixo `x`, e a columna "pens\_hab" no eixo `y`.

```
ggplot(data = datos, mapping = aes(x=Idade.Media,y=pens_hab))
```

Agora temos que agregar `geoms` para indicarlle como se representan os datos no gráfico (puntos, liñas, barras, ...). `ggplot2` ofrece moitos `geoms` diferentes, algúns habituais son: `geom_point()` para diagramas de dispersión, `geom_boxplot()` para diagramas de caixa, `geom_bar()` para diagramas de barras e `geom_line()` para liñas de tendencias, series de tempo, ...

Para engadir un `geom`, emprégase o operador `+`. Pódense agregar varios `geoms` unidos a través de múltiples operadores `+`. Neste caso imos agregar puntos á nosa gráfica empregando o `geom_point()`

```
ggplot(data = datos, mapping = aes(x=Idade.Media,y=pens_hab))+geom_point()
```

### 3. GRÁFICOS DE LIÑAS: a sinxeleza de `ggplot2` fronte a R base

O gráfico lineal emprégase para representar, mediante unha ou varias liñas, a evolución dunha variable ao longo do tempo. No eixo vertical sitúanse os datos que se queren representar e no eixo horizontal sitúase a variable temporal, os anos do período que se analiza. Imos elaborar o gráfico de evolución do Índice de envellecemento nos 7 grandes concellos galegos e no período 2002–2018. Emprégaremos o arquivo `datosSerie` que ten o seguinte aspecto:

```
head(datosSerie)
## Espazo Nome Ano Poboacion Pob_20_64 Indice.Envellecemento
## 1 15030 A Coruña 2006 243320 65.27 130.21
## 2 15030 A Coruña 2018 244850 59.51 146.28
## 3 15030 A Coruña 2012 246146 62.70 139.69
## 4 15030 A Coruña 2008 245164 64.94 132.53
```

No eixe das `x` poremos a variable "Ano" e no eixe das `y` a variable "Indice.Envellecemento". Queremos unha liña por concello, polo que empregaremos o argumento `color`. Como o gráfico que queremos é lineal empregaremos o `geom_line()`. A sintaxe é a seguinte:

```
ggplot(datosSerie,aes(x=Ano,y=Indice.Envellecemento,color=Nome))+geom_line()
```

Se agora este gráfico o queremos facer con R base necesitaríamos as seguinte sentencias:

```
xrange <- range(datosSerie$Ano)
yrange <- range(datosSerie$Indice.Envellecemento)
n=length(unique(datosSerie$Espazo))

plot(xrange, yrange, type="n", xlab="Ano",
     ylab="Indice" )
colors <- rainbow(n)
linetype <- c(1:n)
plotchar <- seq(18,18+n,1)
j=1
for (i in unique(datosSerie$Espazo)) {
  un <- subset(datosSerie, Espazo==i)
  lines(un$Ano, un$Indice.Envellecemento, type="p", lwd=1.5,
        lty=linetype[j], col=colors[j], pch=plotchar[j])
}
```

```

j=j+1
}
title("Índice de Envelhecimento en Galicia")
legend(xrange[1], yrange[2], 1:n, cex=0.8, col=colors,
pch=plotchar, lty=linetype, title="Concello")

```

Con este exemplo queda patente a sinxeleza da librería `ggplot2`.

#### 4. ENGADINDO MÁIS INGREDIENTES

Na sección anterior representamos a evolución do Índice de envelhecimento nos concellos das 7 grandes cidades galegas, ¿que teríamos que facer se queremos engadir puntos ao gráfico de liñas?, pois necesitamos engadir un ingrediente máis, necesitamos dicirlle a `ggplot2` que nos engada puntos, con `geom_point()` do seguinte xeito:

```

ggplot(datosSerie, aes(x=Ano, y=Indice.Envelhecimento, color=Nome))+geom_line()
+geom_point()

```

Se agora se necesita facer un gráfico para cada un dos concellos de Galicia, podemos pedirlle que engada a capa `facet_wrap()`, o que nos permitirá ver o gráfico anterior en múltiples paneis. Esta función toma unha fórmula como argumento, o que se indica co símbolo `~`. Isto dille a R que debe debuxar un panel por cada valor único da columna "Nome" do data frame "datosSerie". Dentro da capa `facet_wrap()` tamén se pode controlar a configuración dos paneis cos argumentos do número de filas (`nrow`) e número de columnas (`ncol`).

```

ggplot(datosSerie, aes(x=Ano, y=Indice.Envelhecimento, color=Nome))+geom_line()
+geom_point()+facet_wrap(~Nome, nrow=2)

```

E se a este gráfico necesitamos engadirlle un título e cambiarlle os nomes aos eixes, pois engadimos un novo ingrediente máis, `labs`, do seguinte xeito:

```

ggplot(datosSerie, aes(x=Ano, y=Indice.Envelhecimento, color=Nome))+geom_line()
+geom_point()+labs(title="Evolución do Índice de envelhecimento 2002-
2018", subtitle="7 grandes concellos de Galicia", x="Periodo", y="Índice de
envelhecimento", color="Concello")

```

#### 5. MÁIS GRAFICOS

Con `ggplot2` pódense facer, ademais, gráficos de barras, boxplots, histogramas, ..., dun xeito moi parecido a como se elaboran os gráficos de liñas. Non se exporán neste relatorio porque se poden consultar en calquera dos manuais básicos de `ggplot2`. Uns gráficos máis particulares que non se encontran de xeito habitual e que se poden elaborar con `ggplot2` son os mapas de coropletas. Os mapas de coropletas son unha representación onde con distintas cores ou intensidades, texturas, etc., indícase a distribución espacial dun fenómeno (variable). A variable que se estuda sóese

categorizar en intervalos, e o valor de cada intervalo aplícase a unha unidade territorial concreta, normalmente determinada por límites administrativos, como unha provincia, unha comarca ou un concello. A cada intervalo asignaráselle unha tonalidade de cor, máis suave para os valores máis baixos e máis intensa canto máis alto sexa o valor. Neste caso imos representar nun mapa de concellos de Galicia o número de pensións por 1000 habitantes (`pens_hab`) que teremos no ficheiro datos que ten esta estrutura:

```

names(datos)
## [1] "Nome" "Espazo"
## [3] "cod_provincia" "provincia"
## [5] "Ano" "Poboacion"
## [7] "parados" "graao"
## [9] "pens_hab"

```

As sentencias son as seguintes :

```

library(rgdal)
concellos=readOGR("Concellos_314_F_region.shp")
proj4string(concellos) <- CRS("+proj=utm +zone=29 +ellps=GRS80 +units=m
+no_defs")
map=spTransform(concellos,CRS("+proj=longlat"))
map_df <- fortify(map,region="COD_MUN2")
datos2=datos[,c(2,15)]
names(datos2)[1]="id"
datos2$id=as.character(datos2$id)
map_df=left_join(map_df,datos2)
p.gal <- ggplot(map_df, aes(x = long, y = lat, group = group)) +
geom_polygon(aes(fill = pens_hab), colour = "white", size =
0.1)+scale_fill_continuous(name="Pensións por habitante",low = "#fff7ec", high
= "#7F0000")
p.gal

```

E obteríamos o mapa da Figura 1.

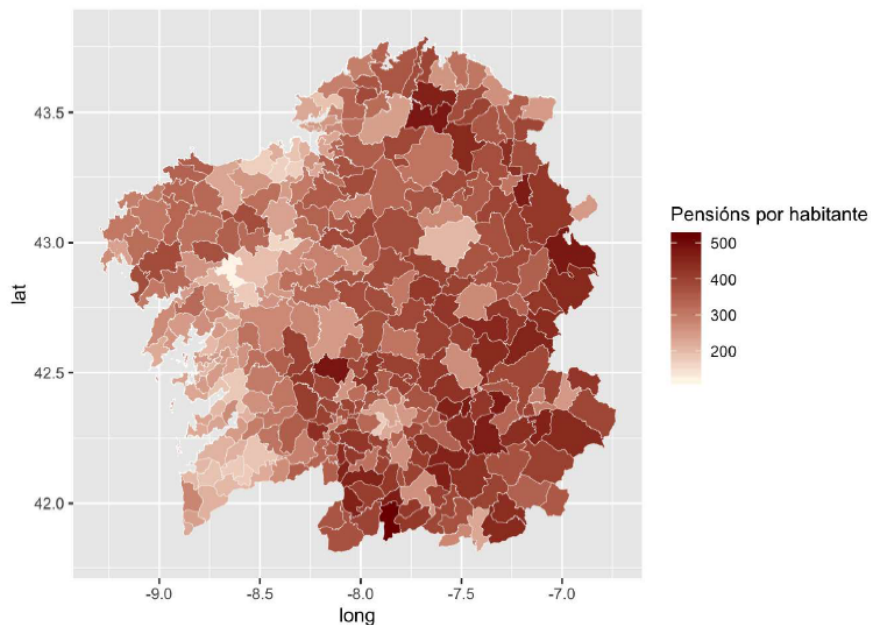


Figura 1: Mapa do número de pensións por 1000 habitantes nos concellos de Galicia.

## 6. CONCLUSIÓNS

En definitiva, con `ggplot2` podes facer máis cousas en menos tempo, aprendendo un único sistema e aplicándoo en diferentes ámbitos.



*VI Xornada de Usuarios de R en Galicia*  
*Santiago de Compostela, 10 de outubro do 2019*

## **R PACKAGES TO STUDY MISSING AND LONGITUDINAL DATA**

Susana R. G. Martins<sup>1</sup>

<sup>1</sup>University of Vigo , Institute Polytechnic of Viana do Castelo and ISAG-European Business School

### **ABSTRACT**

Missing data is a cross-sectional issue with studies involving multiple records over time, and especially longitudinal studies are known for having these issues. Thus, the analysis of missing data, particularly missing data in longitudinal data, is a topic of extraordinary importance. In this text, a brief review of existing methodologies, particularly R-packages for the analysis of missing longitudinal data, will be done. A brief introduction mentioning the work motivation is made in the first section. In the second and the third sections the methodology used for this review and the main results are presented. In the last section there is a brief conclusion on the subject.

**Keywords:** R package, missing values, longitudinal data.

### **1. INTRODUCTION**

The application of statistics in the most diverse areas is a reality that we face on a daily basis. Data collection can be single or regular, which implies studies of different natures. If data is collected regularly, we can have a longitudinal study.

Longitudinal studies are of great interest to medicine, sports and other areas, and they often present a problem that conditions them - the lack of data. The cause for this lack may be diverse, and it can cause constraints on the conclusions drawn.

Nowadays, some techniques that allow the existence of missing data to be filled have been explored. However, the study of missing data in longitudinal databases is not yet explored. This way, the aim of this work is to state the progress of computational tools, namely the R-software.

### **2. METHODOLOGY**

In order to have an overview of the methods of analysing missing values in longitudinal data, particularly in software resources, a global survey using the methodology traditionally known as the "Scoping Review" was made.

For this, the reference base "Web of Science" and the keywords "R package missing longitudinal data" were used. We found 27 results of which only 11 were open access articles. From these articles we selected those that referred to at least one R package, leaving a total of 8 articles for analysis.

For each article, a short summary highlighting the package used was elaborated, as well as its main functionality.

### 3. RESULTS

- MEASURING THE IMPACT OF NONIGNORABLE MISSINGNESS USING THE R PACKAGE ISNI

The article presents a package that implements the ISNI - index of local sensitivity to non-ignorability - methodology. This methodology assesses the dependence of inferences on the ignorability assumption by measuring its sensitivity to violation. The ISNI methodology has been little used due to the lack of software support for its application, which is now being addressed. The ISNI package implements the index of local sensitivity to nonignorability methodology and is a good tool for a systematic and efficient analysis of the reliability of empirical inferences that derive from incomplete data [5].

- IDENTIFYING PATTERNS OF ITEM MISSING SURVEY DATA USING LATENT GROUPS: AN OBSERVATIONAL STUDY

The article is based on a study on health and physical activity and its main objective is to understand if individuals can be grouped according to unanswered questions. The article does not explore any specific packages, but uses them to find results. The packages mentioned are MI and Rmixmod. MI is an imputation package that allows a global view of patterns in missing data. Rmixmod is a package that allows data modeling [1].

- TIME-COURSE GENE SET ANALYSIS FOR LONGITUDINAL GENE EXPRESSION DATA

The article introduces a new TeGSA package, which implements the method of the same name: Time-Course Gene Set Analysis. This package implements the extension of existing gene set analysis methods to longitudinal data [4].

- ACCOUNTING FOR INTERACTIONS AND COMPLEX INTER-SUBJECT DEPENDENCY IN ESTIMATING TREATMENT EFFECT IN CLUSTER-RANDOMIZED TRIALS WITH MISSING OUTCOMES

The article focuses on the study of semiparametric methods for estimating the effects of randomly correlated clinical trial results. When results are randomly absent, weighted inverse probability methods are used to incorporate missing covariates. However, there are often interactions between covariates and treatment and there is no method that, alone, can produce consistent estimates. In this article, double robust estimator, DR, is proposed, which allows to provide correct estimates for missing data as well as allows a correctly specified result model. This estimator is implemented in the CRTgeeDR package [7].

- CRTGEE DR: AN R PACKAGE FOR DOUBLY ROBUST GENERALIZED ESTIMATING EQUATIONS ESTIMATIONS IN CLUSTER RANDOMIZED TRIALS WITH MISSING DATA

The article presents the CRTgeeDR package, developed for cluster randomized clinical trials with missing data. This package has the advantage of addressing the failure of other software that generates biased estimates of the correlation structure when independence is not verified.

When compared to an already existing geePack, this package has better results associated with the use of DR to analyse a binary result when using logistic regression [6].

- JOINT MODELS FOR PREDICTING TRANSPLANT-RELATED MORTALITY FROM QUALITY OF LIFE DATA

The article aims to test whether health-related longitudinal measures related to life quality cause changes related to paediatric hematopoietic stem cell transplantation. In this article, where longitudinal data is analysed, we use the package JM, whose main use is modelling [8].

- **KML AND KML3D: R PACKAGES TO CLUSTER LONGITUDINAL DATA**

The article is based on the presentation of two packages `kml` and `kml3d`. These packages provide an implementation of the k-means method and provide tools for longitudinal data analysis, such as path imputation methods, methods for defining k-means starting conditions, and quality criteria for choosing the best number of clusters. In addition, they allow to produce graphs on the trajectories of the variables, which can be unique or articulated [2].

- **LMEST: AN R PACKAGE FOR LATENT MARKOV MODELS FOR LONGITUDINAL CATEGORICAL DATA**

The article is based on the presentation of the `LMest` package that allows the implementation of latent Markov models. These models are very useful in longitudinal data analysis, particularly when the response variables are categorical in nature. The presented package is appropriate to implement the traditional model, as well as some formulations to implement the model using covariates as well as missing data [3].

#### 4. CONCLUSION

Package	Missing data	Longitudinal data
ISNI	X	–
MI	X	–
Rmixmod	–	–
TcGSA	–	X
CRTgeeDR	X	X
JM	–	X
Kml	X	X
kml3D	X	X
LMest	X	X

Table 1: Summary of presented packages.

The table 1 show that are several packages available for missing data and longitudinal data studies, but only packages `CRTgeeDR`, `Kml`, `kml3D` and are suitable for missing and longitudinal data.

As future work we propose the experimentation of the identified packages and the verification of their suitability for the analysis of the data under study.

## References

- [1] Barnett, Adrian G and McElwee, Paul and Nathan, Andrea and Burton, Nicola W and Turrell, Gavin (2017). Identifying patterns of item missing survey data using latent groups: an observational study. *BMJ Open* 7.
- [2] Christophe Genolini and Xavier Alacoque and Mariane Sentenac and Catherine Arnaud (2015). `kml` and `kml3d`: R Packages to Cluster Longitudinal Data. *Journal of Statistical Software, Articles* 65,1–34.
- [3] Francesco Bartolucci and Silvia Pandolfi and Fulvia Pennoni(2017). `LMest`: An R Package for Latent Markov Models for Longitudinal Categorical Data. *Journal of Statistical Software, Articles* 81, 1–38.
- [4] Hejblum, Boris P. AND Skinner, Jason AND Thiébaud, Rodolphe (2015). Time-Course Gene Set Analysis for Longitudinal Gene Expression Data. *PLoS Computational Biology* 11, 1–21.
- [5] Hui Xie and Weihua Gao and Baodong Xing and Daniel F. Heitjan and Donald Hedeker and Chengbo Yuan (2018). Measuring the Impact of Nonignorable Missingness Using the R Package `isni`. *Computer Methods and Programs in Biomedicine* 164, 207–220.

- [6] Melanie Prague and Rui Wang and Victor De Gruttola (2017). CRTgeeDR: an R Package for Doubly Robust Generalized Estimating Equations Estimations in Cluster Randomized Trials with Missing Data. *The R Journal* 9, 105–115.
- [7] Prague, Melanie and Wang, Rui and Stephens, Alisa and Tchetgen Tchetgen, Eric and De-Gruttola, Victor (2016). Accounting for interactions and complex inter-subject dependency in estimating treatment effect in cluster-randomized trials with missing outcomes. *Biometrics* 72, 1066-1077.
- [8] Terrin, Norma and Rodday, Angie Mae and Parsons, Susan K (2015). Joint models for predicting transplant-related mortality from quality of life data. *Quality of Life Research* 24,31–39.

## O EMPREGO DE R NA PLANIFICACIÓN ESTATÍSTICA

María Martín Vila<sup>1</sup>, Estefanía Vilar Cheda<sup>1</sup> e Carlos Luis Iglesias Patiño<sup>1</sup>

<sup>1</sup>Instituto Galego de Estatística

### RESUMO

O IGE, como un dos principais provedores de información estatística, ten tamén o cometido de planificar e supervisar a actividade estatística dentro da comunidade autónoma. R[1] máis aló de ser unha ferramenta de cálculo permítenos elaborar informes de planificación, seguimento e avaliación que resumen a evolución desta actividade. Para estas tarefas mestúranse as distintas opcións que proporciona este software, desde a potencia de cálculo do R ata as ferramentas de visualización como *Shiny* ou a xeración de informes mediante *R-Markdown*.

**Palabras e frases chave:** Planificación, programación, seguimento, avaliación, estatística pública.

### 1. INTRODUCCIÓN

Un organismo de estatística pública como o IGE ten múltiples cometidos e entre eles inclúese a planificación e seguimento da actividade estatística a realizar no marco da comunidade autónoma[2].

O elemento fundamental da planificación e seguimento da actividade estatística é o inventario de operacións e actividades estatísticas, que pretende recoller de xeito sistemático, estruturado e homoxéneo as principais características das operacións (OE) e actividades estatísticas (AE), que abreviaremos conxuntamente por OAE, realizadas pola administración da comunidade autónoma, permitíndonos coñecer a actividade estatística desenvolta polo sistema estatístico galego. Con este propósito, a base de datos que comprende o inventario recolle, por unha parte, toda a información técnica das OAE (periodicidade, forma de recollida, . . .) e pola outra, toda a información necesaria para a xestión dos programas estatísticos anuais (data de publicación dos resultados, estado da operación estatística, . . .) o seu seguimento e a avaliación do plan. A partir do inventario e coas diversas ferramentas que nos proporciona o software R elabóranse polo tanto o Programa anual e o seu seguimento. O inventario almacénase nunha base de datos en MySQL, que pode consultarse parcialmente mediante a web do IGE.

### 2. PROGRAMACIÓN

O programa anual consiste, entre outras actuacións, nunha listaxe de OAE por realizar no ano. Para cada OAE detállase ademais das características propias da OAE, como poden ser os obxectivos, a periodicidade etc, a forma e prazos de difusión e o orzamento propio do ano.

Para obter esta listaxe emprégase un programa que traslada as OAE programadas e as non programadas no ano anterior e as inclúe en función da periodicidade. Ademais este programa xa reflexa os cambios de estado para as operacións que se programaron o ano anterior como novas. Trátase polo tanto dunha lectura de datos na base de datos, transformación en R dos rexistros e volta a gravar na base de datos (incluímos novos rexistros). Neste sinxelo programa emprégase os paquetes básicos *RODBC* e *dplyr*.

O programa anual (PEA) é un informe[3] que se eleva ao CGE e se publica no DOG polo que unha vez actualizada a base de datos do inventario cómpre darlle forma de informe e incluír todos os campos que necesitamos. Consta dun corpo que se fai de xeito manual e uns anexos que son listaxes de OAE segundo certos parámetros e con distinta información. Para xerar estes anexos de forma automática empregamos **R-Markdown**, de xeito que ligado coa base de datos do inventario xera os anexos do informe:

- ANEXO I. Metas de información coas operacións e actividades estatísticas que as satisfán
- ANEXO II. Operacións e actividades estatísticas por organismo responsable
- ANEXO III. Descrición das operacións e das actividades estatísticas
- ANEXO IV. Actividades de interese estatístico
- ANEXO V. Relación entre os programas actual e precedente

Ademais, con anterioridade á elaboración do PEA definitivo elabórase un ANEXO III por consellería que se envía aos Órganos Estatísticos Sectoriais (Unidades encargadas da planificación e coordinación estatística dos distintos organismos) para que modifiquen e/ou confirmen os datos a incluír das OAE baixo a súa responsabilidade. Este programa ten a mesma estrutura que o ANEXO III pero filtrase por consellería para que cada Órgano valore de xeito independente a súa programación.

Dáse a casuística de que se emprega unha única base de datos, a do inventario, para nutrir o inventario web e para a planificación polo que, no caso de que se requira modificar características das OAE non ligadas ao ano de programación é preciso modificalo na base de datos unha vez aprobado o programa e introducila no anexo de xeito manual, de aí que os documentos **R-Markdown** xeren os informes en **MS-Word**.

### 3. SEGUIMIENTO

O software R emprégase tamén nas distintas fases do seguimento, por unha banda da realízase un seguimento mensual das OAE que teñen previsto publicar datos no mes en curso, para o que se emprega un programa R que facilita un arquivo .csv coa data de publicación prevista e a referencia do resultado a avaliar no mes correspondente, filtrando as OAE con difusión prevista no mes. Como parámetros de entrada unicamente precisa o mes de seguimento. Así mesmo, mensualmente se almacenan datos de varias operacións que publican resultados provisionais na web do IGE. Temos un programa que das operacións que seguimos comproba se o dato provisional mudou e de ser así o almacena. Por motivos de aforro de espazo na base de datos unicamente almacenamos os datos que mudan, desde xeito temos controlado o dato que mudou, cando e as súas diferentes versións.

Por outra banda, ao rematar o ano realízase o seguimento do PEA que consiste en calcular os indicadores de seguimento para cada OAE así como os indicadores globais por consellería. Esta tarefa require dunha batería de programas en R que a partir da táboa da base de datos que actualizamos mensualmente calcula os indicadores de calidade e os agregados por consellería empregando unha base de datos de apoio para o seguimento.

Ademais o seguimento elévase ao CGE polo que faise necesario redactar un informe no que consten diversos gráficos e táboas que reflectan tanto a execución como os diversos indicadores calculados, tanto por OAE como por consellería. O corpo do informe elabórase de xeito manual nun procesador de texto mentres que para os anexos emprégase a ferramenta **Sweave**[4] para compilar arquivos .Rnw, e xunto co MikTeX xeramos os arquivos pdf con toda a información dos anexos 1,2,3 e 5:

- Anexo 1. Informe resumo
- Anexo 2. Informe detallado por consellería
- Anexo 3. Análise por actividades
- Anexo 5. Satisfacción das metas de información

A combinación das distintas ferramentas mencionadas permítenos a partir da base de datos xerar os anexos formatados, coas táboas que precisamos, logos e notas, . . . , tal e como se pode ver na seguinte ligazón: <http://www.ige.eu/estatico/pdfs/s6/lexislacion/InformeSeguimentoPEA2018.pdf>.

#### 4. AVALIACIÓN

Ademais do seguimento dos programas anuais, cada cinco anos faise unha avaliación do plan. O último avaliado foi o plan 2012–2016[5] e para a presentación da súa avaliación elaborouse una aplicación Shiny de visualización. A avaliación clásica consiste nun informe cunha serie de táboas e gráficos que mostran a evolución dos indicadores de seguimento e outros a maiores propios do plan, como a avaliación da perspectiva do xénero ou a difusión da información nos sites dos distintos organismos. Ademais nesta última avaliación, gracias ás posibilidades do Shiny, fíxose unha aplicación de consulta das táboas e gráficas resumo de todos os programas que comprende o plan segundo consellería, actividade e ano, así como o acceso directo aos informes de seguimento e ao propio informe de avaliación do plan: <http://www.ige.eu/Shiny/AvaliacionPlan20122016/>.

### Referencias

- [1] R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- [2] *LEI 9/1988, do 19 de xullo, de Estatística de Galicia*. URL [http://www.ige.eu/estatico/pdfs/s6/lexislacion/lei\\_estadistica\\_galicia.pdf](http://www.ige.eu/estatico/pdfs/s6/lexislacion/lei_estadistica_galicia.pdf)
- [3] *DECRETO 165/2018, do 27 de decembro, polo que se aproba o Programa estatístico anual da Comunidade Autónoma de Galicia para o ano 2019*. URL [http://www.ige.eu/estatico/pdfs/s6/lexislacion/programa\\_estadistico\\_2019\\_gl.pdf](http://www.ige.eu/estatico/pdfs/s6/lexislacion/programa_estadistico_2019_gl.pdf)
- [4] Morales Rivera, M.A. *Generación automática de reportes con R y LATEX* (URL [https://cran.r-project.org/doc/contrib/Rivera-Tutorial\\_Sweave.pdf](https://cran.r-project.org/doc/contrib/Rivera-Tutorial_Sweave.pdf) )
- [5] *LEI 8/2011, do 9 de novembro, do Plan galego de estatística 2012–2016*. URL [http://www.ige.eu/estatico/pdfs/s6/lexislacion/plan\\_galego\\_estadistica\\_2012\\_2016\\_gl.pdf](http://www.ige.eu/estatico/pdfs/s6/lexislacion/plan_galego_estadistica_2012_2016_gl.pdf)

VI Xornada de Usuarios de R en Galicia  
Santiago de Compostela, 10 de outubro do 2019

## De Galicia a Ecuador. Un viaxe de ida y vuelta en R

Salvador Naya<sup>1</sup>, Javier Tarrío-Saavedra, Antonio Meneses<sup>2</sup>, Miguel Flores<sup>3</sup> y Yaroslava Robles-Bykbaev<sup>4</sup>

<sup>1</sup>Escola Politécnica Superior. Grupo Modes, CITIC e ITMATI. Departamento de Matemáticas. Universidade da Coruña.

<sup>2</sup>Facultad de Ingeniería, Universidad Nacional de Chimborazo. Riobamba Ecuador.

<sup>3</sup>Departamento de Matemáticas. Escuela Politécnica Nacional. Grupo Modes. Lab. ADA. Quito. Ecuador.

<sup>4</sup>Universidad Politécnica Salesiana. Cuenca. Ecuador.

### RESUMEN

In this work we present the collaborations that we have been carrying out between Galicia and Ecuador. Specifically those recording the application the R in the frame work the PhD developed out by 3 Ecuadorian students who recently studied at the University of A Coruña.

**Palabras y frases clave:** modelización, degradación, paquetes FCGR TTS, qcr, ILS.

## 1. INTRODUCCIÓN

Hace unos años, en el marco del plan estratégico del ex presidente de Ecuador, Rafael Correa, que quería colocar al país andino en el centro del mapa mundial de la ciencia, y a través de la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación, llegaron a Galicia los estudiantes Antonio Meneses, Miguel Flores y Yaroslava Robles para estudiar el Máster en Técnicas Estadísticas, y posteriormente continuar sus estudios en el Programa de Doctorado en Estadística e Investigación los dos primeros y en el de Ciencias de la Salud la tercera [3, 7, 8]. Pasado el tiempo todos ellos son ya nuevos doctores por nuestra universidad y en el desarrollo de su trabajo de investigación el uso del Software R ha sido de capital importancia.

La interacción entre nuestro grupo de investigación MODES y los doctorandos ecuatorianos ha propiciado un enriquecimiento en los dos lados del charco. Esta colaboración está siendo de doble sentido, generando frutos que pueden ser medidos en publicaciones conjuntas en revistas de impacto y que esperamos que siga con mucha vida. Además se ha propiciado también una red de usuarios de R entre ambos países. Como hito destacable cabe señalar que en el año 2012, en A Coruña, nace por iniciativa de Miguel Flores, el R Users Group -Ecuador (RUG-E), con la finalidad de compartir lo aprendido en Galicia, a través de cursos, eventos como Data y Bielas (Data Beers), charlas, entre otros, con la sociedad ecuatoriana de estadística, incluyendo academia, estudiantado y empresa.

Actualmente, la comunidad ecuatoriana de R tiene 3000 seguidores, que constantemente están interactuando en redes sociales, así como en un grupo de whatsapp donde los miembros realizan consultas. En este año 2019 se tiene planificado constituir legalmente a RUG-E como una Sociedad sin fines de lucro.

En este trabajo se hará una presentación de las principales aportaciones a la comunidad R de estos tres doctorandos. En primer lugar se presentarán las aportaciones realizadas con Antonio



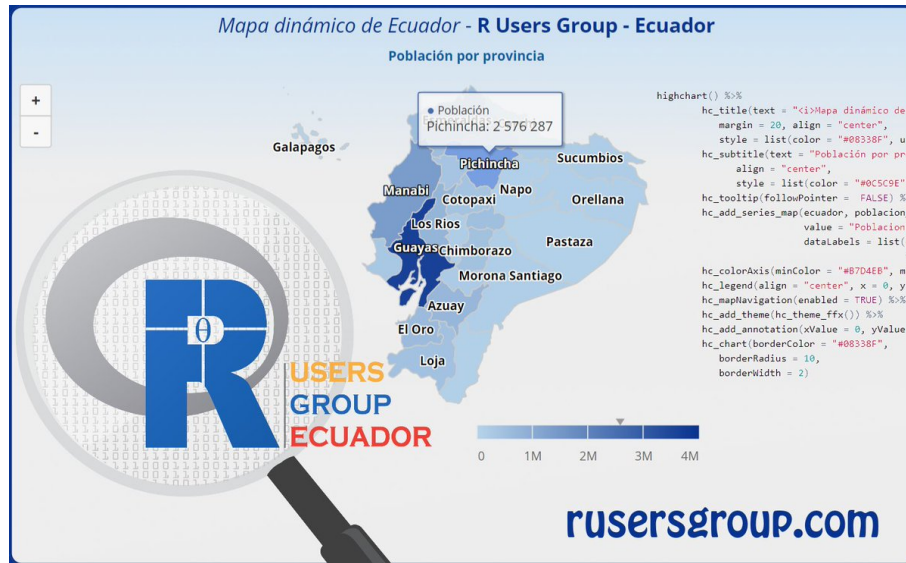


Figura 1: Una de las múltiples aplicaciones del R Users Group-Ecuador.

Meneses, en especial los paquetes FTGC y TTS. En el caso de las contribuciones con Miguel Flores se presentarán las librerías qcr e ILS para terminar con las aplicaciones en modelización de la degradación de biomateriales que se han desarrollado en la tesis doctoral de Yaroslava Robles.

## 2. FCGR y TTS (Riobamba)

En el primer trabajo correspondiente a la tesis doctoral de Antonio Meneses se propusieron nuevas técnicas y metodologías estadísticas para el estudio de la fiabilidad de materiales sometidos a esfuerzos térmicos y de fatiga mecánica. Concretamente se proponen nuevos métodos para estimar los parámetros básicos de los modelos TTS (Temperatura-Tiempo-Superposición) y cómo estimar el crecimiento de grieta por fatiga en experimentos de pruebas aceleradas [1]. Como aportación fundamental de esta investigación se han creado dos librerías de R. Para desarrollar la metodología TTS se propuso la librería con el mismo nombre, mientras que para que el estudio del crecimiento de grieta se propone la librería FCGR (Fatigue Crack Growth in Reliability). Esta librería permite un nuevo enfoque flexible para modelizar la degradación de materiales por crecimiento de grietas a fatiga [2].

Entre las funciones de esta librería FCGR está la de estimar la distribución de tiempos de fallo a fatiga de materiales y la posibilidad de simular distintas situaciones sin necesidad de la realización de experimentos de laboratorio. Concretamente esta librería permite la estimación del nivel de degradación y de la distribución de tiempos de vida de materiales sometidos a fatiga, mediante la aplicación de técnicas de suavizado B-spline, regresión lineal y no lineal de efectos mixtos, estimación tipo núcleo de la función de distribución y herramientas para la estimación de máxima verosimilitud y simulación de Monte Carlo.

Para el caso del método de Superposición Tiempo/Temperatura (TTS) la librería creada permite el estudio de las propiedades físicas de los materiales como la deformación, módulo de elasticidad, cizalladura, etc. Para ello se ha desarrollado una metodología propia que denominamos MNAT (Meneses, Naya And Tarrío), que se basa en aplicar las propiedades de las derivadas de las curvas de fatiga para estimar los tiempos de vida. La idea es que los efectos del tiempo y la temperatura sobre las propiedades mecánicas de los materiales son generalmente equivalentes [2]. Este método permite la estimación de la llamada curva maestra (Master Curve) que es uno de los objetivos de los estudios TTS (Time-Temperature Superposition) [11, 12].

## 3. qcr e ILS (Quito)

En esta sección se recogen las aportaciones de librerías de R realizadas por Miguel Flores en su tesis doctoral [7]. Se proponen nuevas metodologías en el ámbito del control estadístico de la

calidad (CEC), en concreto se aportan técnicas adaptadas a datos obtenidos mediante sensores, monitorizados continuamente respecto al tiempo, y que se engloban dentro del análisis de datos funcionales (FDA).

Todas las metodologías y técnicas propuestas han sido desarrolladas para aportar posibles soluciones a problemas reales en la industria y en los laboratorios de análisis. Tienen su punto de partida en la necesidad de encontrar soluciones para la detección de anomalías relacionadas con el consumo eléctrico, confort térmico y calidad del aire en edificios comerciales y públicos, por un lado, y la detección de sensores o laboratorios atípicos en el marco de los estudios interlaboratorio o ILS [4, 5, 6].

Dada la naturaleza funcional de los datos, se propuso una nueva metodología de gráficos de control FDA basada en el concepto de profundidad de datos. Por otro lado, el germen del desarrollo de nuevas aproximaciones FDA a los estadísticos  $h$  y  $k$  de Mandel para la detección de laboratorios atípicos dentro de un estudio interlaboratorio llamados también estudios ILS (Inter-Laboratory-Study, véase [6, 13]). Es importante destacar que una parte significativa de esta tesis consistió en el desarrollo de dos librerías de R, la `qcr` y la `ILS`, que proporcionen a los profesionales de la industria y servicios, comunidad científica y usuarios en general una forma sencilla y eficaz de aplicar las metodologías propuestas en este trabajo. Para ello, todas las rutinas programadas han sido incluidas en dos nuevas librerías: por un lado la librería `Interlaboratory Study` o `ILS` y, por el otro, la librería `qcr` (Quality Control Review) [13, 14].

#### 4. Modelizando la degradación de biomateriales (Cuenca)

En esta sección se presentará cómo resolver un problema de modelación de la degradación de biomateriales usados en regeneración ósea. En concreto se presentará el empleo de modelos estadísticos programados en R que se han usado en la tesis defendida en julio de este año 2019, por la doctora Yaroslava Robles [8].

Como resultado de este trabajo de investigación se han publicado dos artículos en revistas de Open Access, uno en la revista *PLOS One* [9] y otro en la revista *PeerJ* [10]. El nexo de todos estos trabajos es el modelado estadístico de la degradación de scaffolds con aplicaciones en odontología e ingeniería tisular. De este modo, se han identificado las variables críticas para la degradación de los scaffolds, detectando las variables más influyentes en la degradación midiendo su importancia, realizando la modelización del denominado camino de degradación.

Para este propósito, se aplicaron distintas técnicas estadísticas, como Modelos Aditivos Generalizados (GAM) y regresión no lineal de efectos mixtos basada en las funciones logística y asintótica. Para estas aplicaciones se han usado distintas librerías de R como las librerías `nls`.

Otro de los temas tratado en esta investigación ha sido el estudio mediante tratamiento de imágenes y el uso de técnicas de machine y statistical learning para modelar y estimar el grado de degradación del colágeno tipo I (usado como andamio en regeneración ósea) que podría usarse como metodología en problemas de medicina regenerativa.

#### 5. CONCLUSIONES

Este trabajo pretende ser una puesta en valor por la colaboración internacional en el ámbito de la estadística y presentando como nexo común el empleo y difusión del R a través de la relación que hemos mantenido profesores del grupo MODES en estos últimos años con doctorandos/profesores de Ecuador. Dentro de esta interesante interacción, se hace una mención especial a las aportaciones relacionadas con la creación y uso de paquetes de R para resolver problemas de investigación y transferencia, destacando a su vez la creación de una incipiente e enriquecedora red de colaboración en investigación entre las dos partes.

#### AGRADECIMIENTOS

Los dos primeros autores agradecen los proyectos: MTM2014-52876-R y MTM2017-82724-R, y por la Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2016-015 y Centro Singular de Investigación de Galicia ED431G/01 2016-19), todos ellos a través de los ERDF. Los autores ecuatorianos agradecen las ayudas de la Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) de Ecuador que han posibilitado sus estudios en España.

## Referencias

- [1] Naya, S., Meneses, A., Tarrío-Saavedra, J., Artiaga, R., Lopez-Beceiro, J. y Gracia-Fernández, C. (2013). New method for estimating shift factors in time-temperature superposition models. *Journal of Thermal Analysis and Calorimetry*, 113, 453-460.
- [2] Meneses, A., Naya S., López-de-Ullibarri I. y Tarrío-Saavedra J. (2016). Nonparametric Method for Estimating the Distribution of Time to Failure of Engineering Materials. *Nonparametric Statistics. Springer Proceedings in Mathematics Statistics*, vol 175. 211-224, Springer.
- [3] Meneses, A. (2017). Nuevas aportaciones estadísticas al estudio de la fiabilidad de materiales. Tesis doctoral. Universidade da Coruña.
- [4] Flores, M., Naya, S., Tarrío-Saavedra, J. y Fernández-Casal, R. (2017). *Functional data analysis approach of Mandelâsh and k statistics in Interlaboratory Studies. In Functional Statistics and Related Fields*. A Springer. A Coruña.
- [5] Flores, M., Tarrío-Saavedra, J., Fernández-Casal, R. y Naya, S. (2018). Functional extensions of Mandel's h and k statistics for outlier detection in interlaboratory studies. *Chemometrics and Intelligent Laboratory Systems*, 176, 134–148.
- [6] Flores, M., Tarrío-Saavedra, J., Fernández-Casal, R., Bossano, R. y Naya, S. (2018). ILS: An R package for statistical analysis in Interlaboratory Studies. *Chemometrics and Intelligent Laboratory Systems*, 181, 11–20.
- [7] Flores, M. (2018). Nuevas aportaciones del análisis de datos funcionales en el control estadístico de procesos. Tesis doctoral. Universidade da Coruña.
- [8] Robles-Bykbaev, Y. (2019). Modeling and Analysis of the Degradation of Smart Biomaterials for Dental Use, and the Development and Differentiation of Bone Cells from the Culture of Mesenchymal Cells Inoculated in Dental Scaffolds. Tesis Doctoral. Universidade da Coruña.
- [9] Robles-Bykbaev, Y. Tarrío-Saavedra, J. Quintana-Pita, S. Díaz-Prado, S. Sabán, F.J. y Naya, S. Statistical degradation modelling of poly (d, l-lactide-co-glycolide) copolymers for bioscaffold applications. *PLoS One*, 13, 10, 2018.
- [10] Robles-Bykbaev, Y., Naya, S., Díaz-Prado, S., Calle-López, D., Robles-Bykbaev, V., Garzón, L., Sanjurjo-Rodríguez, C. y Tarrío-Saavedra J. (2019). An artificial-vision-and statistical-learning-based method for studying the biodegradation of type I collagen scaffolds in bone regeneration systems. *PeerJ*. 7, 7233.
- [11] <https://cran.r-project.org/web/packages/FCGR/index.html>
- [12] <https://cran.r-project.org/web/packages/TTS/index.html>
- [13] <https://cran.r-project.org/web/packages/ILS/index.html>
- [14] <https://cran.r-project.org/web/packages/qcr/index.html>

VI Xornada de Usuarios de R en Galicia  
Santiago de Compostela, 10 de outubro do 2019

## DETECCIÓN DE INCENDIOS CON DATOS DE VOOS

Manuel Antonio Novo Pérez<sup>1</sup>, Ana Belén Buide Carballosa<sup>1</sup> e María José Ginzo Villamayor<sup>2</sup>

<sup>1</sup>Instituto Tecnológico de Matemática Industrial (ITMATI)

<sup>2</sup>Departamento de Estatística, Análise Matemática e Optimización. Universidade de Santiago de Compostela (USC)

### RESUMO

Neste traballo amósase un algoritmo desenvolvido no marco do proxecto Civil UAVs Initiative (CUI), programado en R, o *Algoritmo de Detección de Incendios*. Este algoritmo calcula a localización dos incendios ocorridos nunha zona e data determinadas, a partir de datos históricos de voos das aeronaves de extinción. Para iso, o algoritmo debe realizar peticións (en formato JSON) a varios servizos web REST para obter os datos, para o que utiliza os paquetes `httr`[3] e `jsonlite`[2]. Finalmente, o algoritmo procesa os datos para obter a localización, a data e a área dos incendios acontecidos.

**Palabras e frases chave:** Incendios forestais, servizos web REST, aeronaves.

### 1. ALGORITMO DE DETECCIÓN DE INCENDIOS

Os incendios forestais son un dos principais problemas medioambientais en Galicia durante o que vai de século, sendo ademais un dos territorios máis afectados de Europa. Debido á dificultade que entraña combatalos, nos últimos anos xorden proxectos de investigación como o proxecto ENJAMBRE (2014-2018) ou a Civil UAVs Initiative (CUI). ITMATI, que traballa na CUI mediante un contrato coa compañía aeronáutica Babcock, desenvolve algoritmos orientados a mellorar a xestión das aeronaves de extincións pertencentes á compañía aeronáutica, que ofrece os seus servizos a distintos clientes (entre eles a Consellería de Medio Rural da Xunta de Galicia).

A flota de aeronaves de Babcock conta coas tecnoloxías adaptadas ás necesidades e peculiaridades das distintas operacións nas que pode participar (salvamento marítimo, prevención e extinción de incendios forestais...). Nos servizos nos que participan capturan e transmiten distinta información acerca do voo. Estes datos son almacenados e é preciso un tratamento dos mesmos para recopilar información de interese, como por exemplo, cantas e cales son as aeronaves que participan nun determinado incendio. Así, o obxectivo do *Algoritmo de Detección de Incendios* é obter a localización dos incendios ocorridos nun determinado día a partir da información capturada e transmitida a través de sensores embarcados nas aeronaves de extinción. A información fai referencia ao estado en cada instante de tempo (lonxitude, latitude, velocidade...) e aos eventos que realiza, como pode ser o momento no que realizan unha descarga de auga. Estes datos son almacenados nun servidor, e son obtidos realizando peticións HTTP de tipo POST contra dous servizos web REST<sup>1</sup>, un de busca de voos por data e rexión nun mapa, e outro de descarga. JSON<sup>2</sup> é o formato de intercambio de

<sup>1</sup>REST: refírese a calquera interfaz entre sistemas que use HTTP para obter datos ou xerar operacións sobre eses datos.

<sup>2</sup>JSON: é un formato de texto sinxelo para o intercambio de datos.

información cos servizos.

O algoritmo consta de dúas partes diferenciadas: Primeiramente, estrutúrase a petición aos servizos web para obter os voos realizados polas aeronaves de extinción nunha data e rexión definida polo usuario, podendo restrinxir a busca a un cliente ou a determinadas matrículas. Para isto, faise uso da función `toJSON()` da librería `jsonlite`[2]. Dita petición é enviada ao servizo de busca coa función `POST()` da librería `httr`[3]. A resposta da mesma emprégase para realizar a petición ao servizo de descarga, cuxa resposta en formato `JSON` é convertida a `data.frame` coa función `fromJSON()` da librería `jsonlite`[2]. Logo, realízase o procesado desta resposta, filtrando as posicións das descargas de auga realizadas polas aeronaves. Estas, agrúpanse utilizando o algoritmo `DBSCAN` (do paquete `dbscan`[1]), correspondéndose cada grupo cun incendio. Para cada *cluster*, é dicir, para cada incendio, calcúlase o seu centroide (xeorreferenciado) mediante o cal se identifica, e a envoltura convexa das descargas que o forman, que aportan unha estimación da área do mesmo. Esta información, xunto coas matrículas das aeronaves que realizaron as descargas e a data da primeira e última descarga, é proporcionada ao usuario nun arquivo en formato `csv`. Ademais, tamén se proporciona ao usuario dúas capas vectoriais en formato `ESRI Shapefile`, unha coas estimacións dos incendios e outra coas descargas realizadas polas aeronaves.

## 2. EXEMPLO DE EXECUCIÓN

A continuación amósase un exemplo de execución do algoritmo. A partir dos voos das aeronaves, rexistrados o 26 de marzo de 2019 en Galicia, quérese saber se ocorreu algún incendio e a súa localización.

Neste caso, indícaselle ao algoritmo as horas, por exemplo desde as 00:00 do 26 de marzo de 2019 ata as 23:00 dese mesmo día; a extensión de Galicia e filtrase a busca de voos aos pertencentes a un determinado cliente. Tras executar o algoritmo obtense un arquivo en formato `CSV` cos incendios detectados (Táboa 1). Na Figura 1 represéntanse as estimacións dos tres incendios obtidos, xunto coas descargas realizadas en cada un.

	area (ha)	centroide	1ª descarga	última descarga
incendio 1	894.529	-8, 761716338; 42, 67863735	26/03/2019 9:42	26/03/2019 17:32
incendio 2	0.832	-7, 854724848; 41, 95187052	26/03/2019 13:45	26/03/2019 14:10
incendio 3	16.449	-8, 101246964; 41, 92189327	26/03/2019 14:26	26/03/2019 17:21

Táboa 1: Resultados da execución do algoritmo para un determinado día en Galicia.

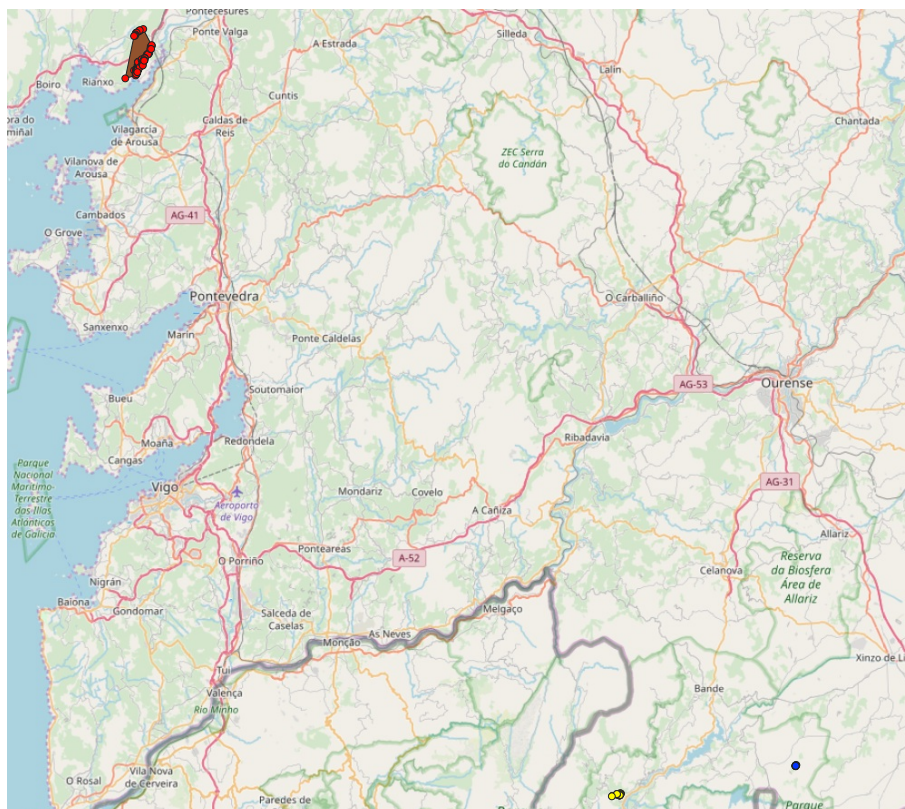


Figura 1: Estimacións das áreas dos 3 incendios detectados o día 26 de marzo de 2019 en Galicia xunto coas descargas de auga realizadas en cada un (representadas en cor vermello, amarelo e azul).

### 3. CONCLUSIÓNS

As probas experimentais amosan que o algoritmo detecta de forma razoábel os incendios ocorridos. A versión actual está pensada para executarse diariamente, polo que é preciso restrinxir a busca a un día.

Ademais, este algoritmo pode empregarse para dispoñer dunha base de datos dos incendios e coñecer que aeronaves de extinción traballaron en cada un. Isto é preciso para desenvolver outros algoritmos, por exemplo un que proporcione cales son as zonas de maior incidencia de incendios ou cal é o risco de que nunha zona se produza un incendio.

### AGRADECEMENTOS

Os investigadores Manuel Antonio Novo, Ana Belén Buide e María José Ginzo agradecen o apoio do proxecto CUI da Axencia Galega de Innovación (GAIN) da Xunta de Galicia.

### Referencias

- [1] Ooms, J. (2014). The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. arXiv:1403.2805 [stat.CO] URL <https://arxiv.org/abs/1403.2805>.
- [2] Hahsler, M. and Piekenbrock, M. (2018). dbscan: Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms. R package version 1.1-3. <https://CRAN.R-project.org/package=dbscan>
- [3] Wickham, H. (2018). httr: Tools for Working with URLs and HTTP. R package version 1.4.0. <https://CRAN.R-project.org/package=httr>

VI Xornada de Usuarios de R en Galicia  
Santiago de Compostela, 10 de outubro do 2019

## ESCENARIOS DE STRESS TEST DE RIESGO DE CRÉDITO CON SHINY

Juan Anido Rodríguez<sup>1</sup>, Víctor González Carro<sup>2</sup> y Sabela Rodríguez Calvo<sup>3</sup>

<sup>1</sup>Abanca

### RESUMEN

La Dirección General de Control Corporativo y Riesgos está formada por un equipo multidisciplinar que desempeña tareas de diversa índole y de relevancia para la Entidad en el mundo del Data Science. Entre dichas tareas se encuentra la simulación de escenarios de pérdidas crediticias o pérdida esperada (Credit Risk Stress Testing o Impairments Projection) que las Entidades Financieras necesitan para realizar sus Planes de Negocio, llevar a cabo ejercicios regulatorios (Stress Test, . . .), así como trasladar a los órganos de control del Banco cuales son los potenciales escenarios de pérdidas en función de los escenarios macroeconómicos que se planteen.

Para la proyección de la pérdida esperada se parten de las series históricas de entrada en default de clientes y de recobro de la Entidad y de las series históricas de los principales indicadores macroeconómicos y se generan modelos que cumplen los requisitos estadísticos mínimos exigidos y que relacionan las series históricas internas con las series de las variables macroeconómicas.

A partir de los modelos seleccionados, y utilizando las proyecciones de las variables macro proporcionadas por el área de Planificación Estratégica, se obtienen las proyecciones de nuevos dudosos y del coste de crédito.

Estos modelos se han desarrollado con el software estadístico R. Además, se ha implementado una aplicación que permite a los usuarios interactuar de manera sencilla permitiendo seleccionar el parámetro a modelizar, los segmentos sobre los que se quiere realizar la estimación, introducción manual o búsqueda automatizada de las variables macroeconómicas a considerar en la modelización y obtención de proyecciones en diversos escenarios.

**Palabras y frases clave:** series temporales, modelos proyección, Shiny.

### 1. GENERACIÓN DE MODELOS DE PROYECCIÓN

Una de las tareas del área de Control Corporativo y Riesgos consiste en la proyección de pérdida esperada. Para realizar esto se han desarrollado diferentes utilidades en R haciendo uso de diversos paquetes y librerías, que permiten:

- Modelizar las series de entrada en default y las series de recobro por segmento a partir de las series de los principales indicadores macroeconómicos.
- Obtener las proyecciones a tres años de la entrada en default y del coste de crédito en varios escenarios (escenario base, escenario adverso, . . .)

- A menudo se obtienen numerosos modelos que cumplen los requisitos estadísticos mínimos exigidos, por lo que se realiza la selección del modelo teniendo en cuenta otros aspectos como el sentido económico de las variables que intervienen en él. Para facilitar esta selección, se ha desarrollado un mecanismo que permite exportar a un fichero excel, entre otra información, las variables macroeconómicas del modelo, los términos autorregresivos, principales estadísticos como son el coeficiente de correlación, Durbin-Watson, . . . , y proyecciones de los parámetros modelizados en diversos escenarios.
- Para los usuarios que no conocen el lenguaje R se ha desarrollado una aplicación con la librería Shiny en la que se incluyen las funcionalidades descritas anteriormente. Las motivaciones y prestaciones por las que se ha optado por esta vía son, entre otras:
  - Facilitar la visualización y análisis de los resultados de una manera más intuitiva mediante tablas y gráficos dinámicos.
  - A partir de la modelización de las series históricas de los parámetros permite conocer las proyecciones de estos parámetros a tres años vista en función de las variables macroeconómicas utilizadas en la modelización.
  - Permite a los usuarios ver la sensibilidad de los parámetros a cambios en las proyecciones de las variables macroeconómicas.
  - Proporciona cierta autonomía a los usuarios que no están habituados a trabajar con código R, permitiéndoles la interacción y selección de distintas opciones en los resultados.
  - Garantiza la trazabilidad, consistencia y calidad de la información proporcionada.

## Referencias

- [1] Chang, W. Cheng, J., Allaire, JJ., Xie, Y. and McPherson, J. (2018). shiny: Web Application Framework for R. R package version 1.1.0. URL <https://CRAN.R-project.org/package=shiny>.
- [2] Golemund, G., Wickham, H. (2011). Dates and Times Made Easy with lubridate. *Journal of Statistical Software*, 40(3), 1-25. URL <http://www.jstatsoft.org/v40/i03/>.
- [3] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.



*VI Xornada de Usuarios de R en Galicia*  
*Santiago de Compostela, 10 de outubro do 2019*

## PRESENTACIÓN DO «CONCURSO MELISA» DE R

Miguel Ángel Rodríguez Muñíos<sup>1</sup>, María José Ginzo Villamayor<sup>1</sup> e Rafael Rodríguez Gayoso<sup>1</sup>

<sup>1</sup>MELISA. Asociación de Usuarios de Software Libre da Terra de Melide

### RESUMO

A Asociación de Usuarios de Software Libre da Terra de Melide (MeLiSa) ven desenvolvendo, durante os últimos anos, unha labor de apoio e divulgación do software libre R (<https://www.r-project.org/>) colaborando na organización das xornadas de usuarios de R en Galicia (<https://www.r-users.gal/>), nas xornadas de software libre científico (<http://xornada.melisa.gal/>) e en proxectos relacionados con R como son EpiLinux (<https://www.sergas.es/Saude-publica/Epilinux>) e BioStatFLOSS (<https://www.sergas.es/Saude-publica/BioStatFLOSS>).

Este ano, na VI Xornada, se presenta por primeira vez a convocatoria do «Concurso Melisa de R» no que se proporá aos interesados entregar un proxecto de análise ou visualización sobre un conxunto de datos común.

Neste relatorio presentarase dito concurso así como as bases para a participación no mesmo e os premios aos gañadores.

**Palabras e frases chave:** R, concurso.

### AGRADECEMENTOS

Á Asociación Melisa polo financiamento e organización do premio, a traveso do convenio de colaboración en materia de software libre coa AMTEGA.

*VI Xornada de Usuarios de R en Galicia  
Santiago de Compostela, 10 de outubro do 2019*

## O PROXECTO SHINYAPPS DE MELISA

Miguel Ángel Rodríguez Muñíos<sup>1</sup>, María José Ginzo Villamayor<sup>1</sup> e Rafael Rodríguez Gayoso<sup>1</sup>

<sup>1</sup>MELISA. Asociación de Usuarios de Software Libre da Terra de Melide

### RESUMO

A Asociación de Usuarios de Software Libre da Terra de Melide (MeLiSa) ven desenvolvendo, durante os últimos anos, unha labor de apoio e divulgación do software libre R.

Neste relatorio se presentará o novo proxecto de MELISA baseado en ShinyApps no que se porá a disposición dos interesados un acceso «Basic» a <https://www.shinyapps.io/> para poder aloxar e desenvolver proxectos baseados en RStudio e Shiny. Explicarase o procedemento de solicitude de acceso así como o sistema de xestión de tokens para poder aloxar aplicacións no espazo de MELISA.

\* A diferenza do acceso «Free» o «Basic» soporta máis memoria RAM, capacidade de proceso, instancias simultáneas e horas activas.

**Palabras e frases chave:** R, Shiny, ShinyApps.

### AGRADECEMENTOS

Á Asociación Melisa polo financiamento deste proxecto, a traveso do convenio de colaboración en materia de software libre coa AMTEGA.

## EL NUEVO ESTÁNDAR SIMPLE FEATURES PARA DATOS VECTORIALES EN R: EL PAQUETE `sf`

Dominic Royé<sup>1</sup>

<sup>1</sup>Departamento de Geografía, Universidade de Santiago de Compostela. Grupo de Epidemiología y Salud Pública

### RESUMEN

En este breve relatorio se expone el nuevo estándar simple features para trabajar con datos vectoriales. Se muestran algunas funciones básicas y el gran potencial cuando se emplean funciones de la colección de tidyverse. Además se incluye una tabla para migrar entre el paquete de `sf` y `sp`.

**Palabras y frases clave:** datos espaciales, vectorial, simple features, tidyverse.

### 1. INTRODUCCIÓN

El paquete `sf` (*simple features*) se ha convertido en un nuevo estándar para la gestión y manipulación de datos espaciales. Anteriormente, se hizo uso fundamentalmente de las funcionalidades de paquetes `sp` y `rgdal`. El cambio en `sf` está basado en convenciones del *Open Geospatial Consortium* (OGC) y en el estándar ISO 19125, que especifica un modelo de acceso y almacenamiento común de geometrías en su mayoría bidimensionales (punto, línea, polígono, multipunto, multilínea, etc.). La estructura en R y la sintaxis de las funciones se caracterizan por un estilo más conocido de la colección de paquetes tidyverse. `sf` requiere de una forma de escribir el código distinta, más intuitiva, pero utilizando estructuras de datos nativas de R (matrix, vector, list, etc.). Algunas de las características fundamentales de `sf` son: 1) todos los objetos son data.frames con las coordenadas espaciales que le corresponden en una columna propia, 2) todas las funciones empiezan por `st_*` (por ejemplo, `st_read()`), 3) las funciones son compatibles con el operador pipe (`%>%`) y con muchas funciones de la colección tidyverse, 4) `sf` incluye directamente librerías como GDAL, GEOS y Proj.4.

### 2. LAS FUNCIONES BASICAS DE SF

La lectura de archivos, igual que con `rgdal`, es muy sencilla. Podemos importar un shapefile de ESRI o cualquier otro archivo de datos vectoriales (geojson, geodatabase, etc.) con la función `st_read()`.

Los metadatos de los objetos `sf` nos resumen toda la información relevante del objeto espacial. En la cabecera vemos el tipo de clase, el número de filas (features), el número de columnas (fields), el tipo de geometría, las dimensiones, la extensión y la proyección. Cuando imprimimos un objeto `sf`, R también resume los atributos.

```
# paquetes
library(tidyverse)
library(sf)
## Linking to GEOS 3.6.1, GDAL 2.2.3, PROJ 4.9.3
library(rnaturalearth)
library(maps)
```

```

## Simple feature collection with 255 features and 3 fields
## geometry type: MULTIPOLYGON
## dimension: XY
## bbox: xmin: -180 ymin: -90 xmax: 180 ymax: 83.6341
## epsg (SRID): 4326
## proj4string: +proj=longlat +datum=WGS84 +no_defs
## First 10 features:
##   featurecla scalerank labelrank geometry
## 1 Admin-0 country 5 2 MULTIPOLYGON (((117.7036 4...
## 2 Admin-0 country 5 3 MULTIPOLYGON (((117.7036 4...
## 3 Admin-0 country 6 2 MULTIPOLYGON (((-69.51009 -...
## 4 Admin-0 country 0 3 MULTIPOLYGON (((-69.51009 -...
## 5 Admin-0 country 0 2 MULTIPOLYGON (((-69.51009 -...
## 6 Admin-0 country 0 2 MULTIPOLYGON (((-67.28475 -...

```

Figura 1: Estructura de un Simple Feature.

Cargamos un dataset de ciudades del mundo del paquete maps con columnas de longitud y latitud.

```

data(world.cities)
str(world.cities)

## 'data.frame': 43645 obs. of 6 variables:
## $ name : chr "'Abasan al-Jadidah" "'Abasan al-Kabirah" "'Abdul Hakim" ...
## $ country.etc: chr "Palestine" "Palestine" "Pakistan" "Kuwait" ...
## $ pop : int 5629 18999 47788 21817 2456 3434 9198 5492 22706 41731 ...
## $ lat : num 31.3 31.3 30.6 29.4 32 ...
## $ long : num 34.3 34.4 72.1 48 35.1 ...
## $ capital : int 0 0 0 0 0 0 0 0 0 0 ...

```

Podemos convertir esta tabla a un objeto sf muy fácilmente.

```

cities <- st_as_sf(world.cities,
                  coords = c("long", "lat"),
                  crs = 4326)
cities

```

Es posible usar funciones de la colección tidyverse.

```

# ordenar de mayor a menor
arrange(cities, desc(pop))

## Simple feature collection with 43645 features and 4 fields
## geometry type: POINT
## dimension: XY
## bbox: xmin: -178.8 ymin: -54.79 xmax: 179.81 ymax: 78.93
## epsg (SRID): 4326
## proj4string: +proj=longlat +datum=WGS84 +no_defs
## First 10 features:
##   name country.etc pop capital geometry
## 1 Shanghai China 15017783 2 POINT (121.47 31.23)
## 2 Bombay India 12883645 0 POINT (72.82 18.96)
## 3 Karachi Pakistan 11969284 0 POINT (67.01 24.86)
## 4 Buenos Aires Argentina 11595183 1 POINT (-58.37 -34.61)

```

```
## 5      Delhi      India 11215130      0 POINT (77.21 28.67)
## 6      Manila Philippines 10546511      1 POINT (120.97 14.62)
## 7      Moscow    Russia 10472629      1 POINT (37.62 55.75)
## 8      Seoul Korea South 10409345      1 POINT (126.99 37.56)
## 9      Sao Paulo  Brazil 10059502      0 POINT (-46.63 -23.53)
## 10     Istanbul  Turkey 10034830      0 POINT (29 41.1)
```

```
# filtrar grandes ciudades > 5 millones
big_cities <- filter(cities, pop > 5000000) %>%
  mutate(pop2 = pop/1000000)
```

Podemos encontrar una lista de funciones compatibles con objetos `sf` y la lista total de funciones `st_*` en <https://r-spatial.github.io/sf/reference/index.html>.

Una ventaja muy importante de `sf` es la existencia de una geometría específica dentro del paquete de `ggplot2`. La función `geom_sf()` y `coord_sf()` nos ayuda a visualizar datos espaciales.

```
world <- ne_countries(returnclass = "sf")
ggplot() +
  geom_sf(data = world,
          fill = NA) +
  geom_sf(data = big_cities,
          aes(size = pop2),
          alpha = 0.7,
          colour = "#005a32",
          show.legend = 'point') +
  labs(size = "Población (millones)") +
  coord_sf(crs = "+proj=robin +lon_0=0 +x_0=0 +y_0=0 +ellps=WGS84
+datum=WGS84 +units=m +no_defs") +
  theme_void()
```



Figura 2: Ejemplo de un mapa con GGLOT2 a partir de un objeto `sf`.

### 3. MIGRACIÓN ENTRE SP Y SF

Para facilitar la migración, podemos encontrar un resumen de las funciones más importantes entre `sp` y `sf` en la Tabla 1.

<code>sp</code>	<code>sf</code>
<code>DISAGGREGATE</code>	<code>st_cast()</code>
<code>PLOT(X)</code>	<code>plot(st_geometry(x))</code>
<code>COORDINATES(X)</code>	<code>st_coordinates(x)</code>
<code>COORDINATES(X) &lt;- ~ X + Y</code>	<code>st_as_sf(x, coords = c("x", "y"))</code>
<code>OVER(X, Y, RETURNLIST = TRUE)</code>	<code>st_intersects(x, y)</code>
<code>SPDISTS(X, Y)</code>	<code>st_distance(x, y)</code>
<code>AGGREGATE(X, BY, MEAN, AREAWEIGHTED = TRUE)</code>	<code>st_interpolate_aw(x, by, mean)</code>
<code>CRS(PROJ4ARGS)</code>	<code>st_crs(proj4string)</code>
<code>PROJ4STRING(X)</code>	<code>st_crs(x)</code>
<code>PROJ4STRING(X) &lt;- P4S</code>	<code>st_crs(x) &lt;- p4s</code> ó <code>st_set_crs(x, p4s)</code>
<code>SPTRANSFORM(X, CRSOBJ)</code>	<code>st_transform(x, crs)</code>
<code>BBOX(X)</code>	<code>st_bbox(x)</code> ó <code>matrix(st_bbox(x), 2)</code>

Tabla 1: Migración entre los paquetes `sp` y `sf`.

Podemos obtener más correspondencias para `rgeos` y `rgdal` en <https://github.com/r-spatial/sf/wiki/migrating>.

VI Xornada de Usuarios de R en Galicia  
Santiago de Compostela, 10 de outubro do 2019

## APLICACIÓN DE R PARA MANTENIMIENTO PREDICTIVO EN LA INDUSTRIA 4.0

Manuel Vaamonde Rivas<sup>1</sup>, Manuel Febrero Bande<sup>2</sup> y Beatriz Pateiro López<sup>2</sup>

<sup>1</sup>Instituto Tecnológico de Matemática Industrial (ITMATI)

<sup>2</sup>Departamento de Estatística, Análise Matemática e Optimización. Universidade de Santiago de Compostela

### RESUMEN

Actualmente se está produciendo la cuarta revolución industrial o Industria 4.0, la cual busca mejorar los procesos industriales incorporando a estos las nuevas tecnologías, mediante una digitalización de los mismos. Uno de los aspectos fundamentales de este proceso es el flujo de grandes cantidades de datos, los cuales deben ser correctamente tratados e interpretados. Por ello, las técnicas estadísticas toman un papel de gran importancia en la Industria 4.0.

Dentro de este contexto surge el proyecto europeo PreCoM (EU-H2020-FOF-09, 2017-2020), centrado en el mantenimiento predictivo (vs reactivo o preventivo) en entornos industriales. El mantenimiento predictivo consiste en, mediante el acceso a datos de equipos y herramientas, así como de otras fuentes externas de relevancia, aplicar técnicas que permitan detectar posibles fallas y defectos de maquinaria. De este modo se pueden evitar paros de emergencia, mejorar la calidad de los productos y, en general, perfeccionar el proceso productivo.

Uno de los módulos de PreCoM, desarrollado por el Instituto Tecnológico de Matemática Industrial (ITMATI), consiste en la utilización de técnicas de series de tiempo para el mantenimiento predictivo. Dicho módulo es denominado TSA (Time Series Analysis), y fue desarrollado en R. En este trabajo se introducirán brevemente las técnicas de series de tiempo utilizadas, para, posteriormente, presentar su implementación práctica. Concretamente, el trabajo se centrará en la aplicación web que incorpora el módulo, la cual fue construida a partir del paquete Shiny.

**Palabras y frases clave:** Industria 4.0., Mantenimiento predictivo, PreCoM, Series de Tiempo, Shiny.

### 1. INTRODUCCIÓN

PreCoM (EU-H2020-FOF-09, 2017-2020) es un proyecto de 3 años, financiado por el Programa Horizonte 2020, que tiene como objetivo implementar y validar un sistema de decisión-asistencia para mantenimiento predictivo. En la actualidad, sensores más potentes y baratos, junto con análisis de Big Data, ofrecen una oportunidad única para monitorizar el estado y rendimiento de máquinas y herramientas. El sistema desarrollado en PreCoM busca identificar y localizar daños, determinar la gravedad de estos, predecir su evolución, estimar tiempos de vida útil, reducir la probabilidad de falsas alarmas, proporcionar una mejor detección de fallos, emitir avisos para realizar tareas de mantenimiento preventivo y mejorar la eficiencia de actividad de las máquinas. Los objetivos específicos son:

- Incrementar disponibilidad y mantenibilidad
- Incrementar el tiempo invertido en mantenimiento predictivo.
- Reducir accidentes debidas a averías.
- Reducir el consumo de energía.
- Reducir el consumo de materiales.

PreCoM contiene más de 20 módulos distintos, y su integración constituirá el sistema descrito anteriormente. En este trabajo es presentado el módulo TSA, el cual proporciona predicciones de vibración para componentes críticas, con el objetivo de evaluar el estado de los equipos y predecir fallos o deterioros, y así, junto con otros módulos, ofrecer recomendaciones para el mantenimiento.

## 2. ANÁLISIS DE SERIES DE TIEMPO

Las señales de vibración son secuencias tomadas de forma sucesiva en el tiempo. Por tanto el uso de técnicas de series de tiempo sobre estas señales puede permitir identificar patrones en los datos y realizar predicciones en base a ellos. Los modelos autorregresivos integrados de media móvil (ARIMA) son una clase de modelos dinámicos de series de tiempo. Sea  $\{Y_t\}$  una serie de tiempo, donde  $Y_t$  representa la observación en el instante  $t$ . En primer lugar, el modelo ARMA(p,q), donde  $p$  representa el orden de la parte autorregresiva y  $q$  el de la de media móvil, se representa de la siguiente forma:

$$Y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (1)$$

Los parámetros determinan las componentes determinista ( $\mu$ ), autorregresiva ( $\phi$ ), de media móvil ( $\theta$ ) y el error ( $e$ ), donde  $e_t$  se asume que es ruido blanco.

El modelo ARIMA se puede definir a partir del anterior.  $\{Y_t\}$  sigue un proceso ARIMA(p,d,q) si la  $d$ -ésima diferenciación regular de  $Y_t$  sigue un proceso ARMA(p,q), invertible y estacionario. Entonces, sea  $B$  el operador retardo,  $BY_t = Y_{t-1}$ , y sean los polinomios  $\Phi_p(B) = (1 - \sum_{i=1}^p \phi_i B^i)$  y  $\Theta_q(B) = (1 - \sum_{i=1}^q \theta_i B^i)$ . El modelo ARIMA(p,d,q) puede ser representado del siguiente modo:

$$\Phi_p(B)(1 - B)^d Y_t = \Theta_q(B)e_t \quad (2)$$

El módulo TSA del proyecto PreCoM utiliza una ampliación del modelo ARIMA, que permite incorporar variables regresoras que puedan ofrecer información relevante sobre la señal objetivo. Este modelo es denominado ARIMAX:

$$\Phi_p(B)(1 - B)^d Y_t = \beta X_t + \Theta_q(B)e_t \quad (3)$$

## 3. APLICACIÓN WEB (SHINY)

Los modelos anteriores se utilizan en dos etapas. La primera, *offline*, consiste en ajustar el modelo a partir de datos históricos de la señal objetivo y sus covariables. Los modelos ajustados son utilizados en la segunda etapa, *online*, para ofrecer predicciones sobre la evolución futura de la señal.

Mediante el uso del paquete de R Shiny, se construyó una aplicación web que actúa como interfaz del módulo TSA. En las Figura 1 se muestran capturas de ambas partes (offline y online) de la aplicación. Las dos están compuestas por un panel a la izquierda en el que se pueden determinar parámetros de interés como la fecha o la señal. La parte derecha muestra las salidas, que en el caso offline es un resumen del modelo ajustado, junto a una representación de los datos utilizados, y en el online una tabla y un gráfico con las predicciones e intervalos de confianza.

Para que la aplicación funcione correctamente debe haber conexión a internet, ya que al introducir la fecha deseada, esta lanza una petición a la base de datos para descargar los datos correspondientes. El módulo TSA se conecta con la base de datos de PreCoM, que está proporcionada por



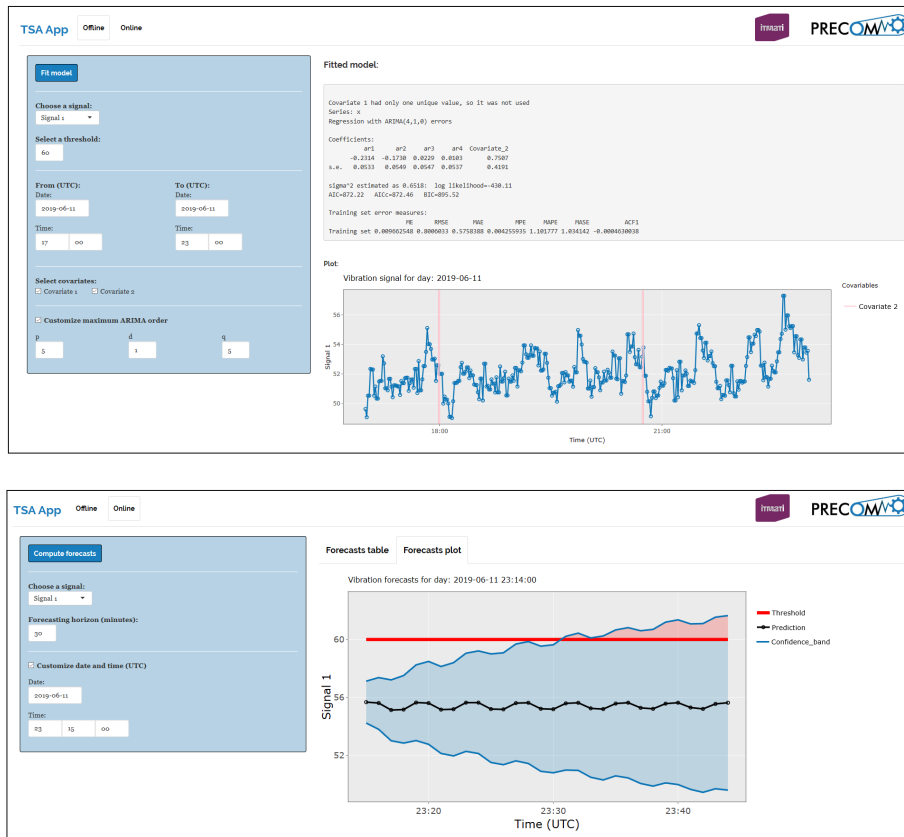


Figura 1: Aplicación web del módulo TSA de PreCoM. Arriba la parte offline, abajo la online.

una empresa de monitorización avanzada y Big Data para la industria, a través de una API tipo REST. Los datos son descargados en formato *json* y procesados en R para poder ser utilizados posteriormente. Actualmente la aplicación actúa sobre varios equipos, correspondientes a los casos de uso del proyecto PreCoM.

### 3. CONCLUSIONES

El software estadístico R ha permitido dar una solución a un problema de gran relevancia y actualidad en la industria. La correcta integración con el entorno de Big Data permite aprovechar al máximo, en tiempo real, las capacidades de los modelos estadísticos. Además, la aplicación web ofrece una interfaz agradable e intuitiva que aporta gran accesibilidad y comodidad al módulo TSA de PreCoM.

### AGRADECIMIENTOS

Proyecto PreCoM (EU-H2020-FOF-08, 2017-2019), financiado por el programa de investigación e innovación de la Unión Europea Horizonte 2020, bajo el Grant Agreement N° 768575.

### Referencias

- [1] Box, G., Jenkins, G., Reinsel, G. & Ljung, G. (2016). *Time Series Analysis. Forecasting and Control (Fifth Edition ed.)*. John Wiley & Sons.
- [2] Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2019). *shiny: Web Application Framework for R*. R package version 1.3.2. <https://CRAN.R-project.org/package=shiny>.

VI Xornada de Usuarios de R en Galicia  
Santiago de Compostela, 10 de outubro do 2019

## OBRADOIRO: VISUALIZACIÓN DE DATOS ESPACIAIS CON R

María Isabel Borrajo García<sup>1</sup> e Mercedes Conde Amboage<sup>1</sup>

<sup>1</sup>Departamento de Estatística, Análise Matemática e Optimización. Universidade de Santiago de Compostela

### RESUMO

A Estatística Espacial é unha rama da Estatística con aplicacións en numerosos ámbitos (como a Ecoloxía, a Epidemioloxía, o Medio Ambiente ou a Xeoloxía, entre outros) na que ten especial importancia a visualización dos datos a empregar. Ao longo deste obradoiro traballaranse aspectos relacionados coa lectura e o preprocesamento de diferentes formatos de datos, a súa visualización así como o emprego de diferentes ferramentas de Estatística Descritiva neste eido dispoñibles en R.

**Palabras e frases chave:** estatística espacial, lectura e visualización de datos, representacións gráficas, estatística descritiva.

### 1. LECTURA DE DATOS

A primeira tarefa que será necesario afrontar ao analizar calquera conxunto de datos, sexa espacial ou non, é a lectura dos mesmos dende R. Comezarase o obradoiro facendo un repaso sobre como levar a cabo a lectura de datos en formatos clásicos (.txt, .csv, .xlsx, .RData, . . .), para logo estender as ideas aos formatos nos que se almacenan habitualmente os datos espaciais.

Dependendo do tipo de dato espacial a tratar (puntos, grellas de valores ou datos de área) os formatos soen variar. Comunmente os valores de localizacións almacénanse en arquivos .txt ou .csv, mentres que os datos máis complexos soen atoparse en formatos como *shapefiles*, .nc ou .tif. Ilustrarase como importar dende R cada un deles, así como o procesamento inicial que precisan para que o seu correcto tratamento. Tamén se abordará brevemente a cuestión das proxeccións xeográficas.

### 2. VISUALIZACIÓN DE DATOS

As primeiras e máis sinxelas representacións que podemos facer con R abranguen a análise de variables aleatorias unidimensionais (datos non espaciais). Farase un breve repaso de ditas representacións destacando os diagramas barras ou sectores, os histogramas, os diagramas de caixas e os diagramas de dispersión.

En canto á visualización de datos espaciais, dividirase esta sección en varias partes segundo o tipo de datos dos que se dispoña:

**Puntos:** para representar puntos, entendendo puntos como coordenadas no plano, empregaranse principalmente ferramentas dos paquetes *sp* e *spatstat*.

**Polígonos:** neste caso necesitarase tamén o paquete *sp* e algunhas outras dependencias, que permitan representar de xeito adecuado este tipo de datos. Tamén se verá como personalizar estas representacións.

**Raster ou arrays:** presentaranse diferentes posibilidades que ofrece o paquete *raster*, así como algúns consellos e ideas sobre a xestión de memoria cando se traballa con ficheiros moi pesados (algo moi habitual, por exemplo, no contexto das Ciencias Medioambientais).

### 3. FERRAMENTAS PARA O RESUMO DE DATOS

Para rematar este obradoiro, ilustrárase a aplicación de medidas resumo de Estatística Descritiva con R. Ao igual que na sección anterior, comezase cun repaso de como obter medidas características de posición, dispersión e forma asociadas a datos non espaciais empregando o *software* R. Finalmente, e sempre e cando as restricións temporais o permitan, presentaranse algunhas ferramentas sinxelas de Estatística Descritiva que permitan caracterizar numericamente os datos espaciais.

### AGRADECEMENTOS

Proxecto (MTM2016-76969P) financiado polo Ministerio de Economía y Competitividad e cofinanciado por European Regional Development Fund (ERDF).

#### **Requisitos:**

Requírense coñecementos básicos de R. Poderase seguir o curso tanto con R como con RStudio. Ademais, recoméndase ter instalados os seguintes paquetes: *sp*, *raster*, *rgdal*, *rgeos*, *spatstat*, *maps*, *maptools* e *rworldmap*.

*VI Xornada de Usuarios de R en Galicia*  
*Santiago de Compostela, 10 de outubro do 2019*

## **OBRADOIRO: INTRODUCCIÓN AO SOFTWARE R**

M<sup>a</sup> José Ginzo Villamayor<sup>1</sup>

<sup>1</sup>Departamento de Estatística, Análise Matemática e Optimización. Universidade de Santiago de Compostela (USC)

### **RESUMO**

Neste obradoiro mostrarase unha introdución a linguaxe de programación R (importar datos, obxectos en R, operacións básicas, vectores, matrices, data frame, as funcións `outer` e `apply` entre outras, . . .), así como os diferentes tipos de representacións gráficas que se poden facer, principais liñas de código ou comandos para facer unha sinxela análise estatística e presentación do paquete R Commander (`rcmdr`).

**Requisitos**, ter descargado e se é posible instalado no PC:

- R e
- Rstudio.

## COURSE: MULTIVARIATE AND FUNCTIONAL DATA CLASSIFICATION IN R

Manuel Oviedo de la Fuente<sup>1</sup>

<sup>1</sup> Departamento de Estatística, Análise Matemática e Optimización, Universidade de Santiago de Compostela

### ABSTRACT

This course offers an introduction to multivariate and functional data classification in R using the `fda.usc` package [3]. This package carries out exploratory and descriptive analysis of functional data exploring its most important features such as depth measurements or functional outliers detection, among others. It also helps to explain and model the relationship between a dependent variable and independent (regression models) and make predictions. Methods for supervised classification of a set of multivariate and functional data regarding a feature of the data are also included. In particular, this course focuses on the following topics:

1. Overview of Functional Data Analysis
  - `fdata` class definition
  - Fixed and data-driven basis
  - Smoothing
2. Classifiers:
  - (Linear and Additive) Logistic Models [5]
  - Non-parametric Models (kNN, Kernel) [4] and based on Depth (DD<sup>G</sup> classifier) [2]
  - Machine Learning Methods (NNET, SVM, RPART,...) [1]
3. Special features
  - Inverse probability weighting and cutoff probability selection
  - Parameter Tuning: K-fold cross validation
  - Multiclass Classification: Majority Voting and One vs Rest schemes
  - Ensemble Methods: Adaboost [5] and Random Forest [1]
4. Image Data Classification
  - Classification of Hyperspectral and LiDAR Data

### FURTHER INFORMATION

*Course material:* <http://eamo.usc.es/pub/moviedo/index.php/software>

*Requirements:* Install `fda.usc`, `e1071` and `randomForest` packages.

**Keywords:** Functional Data Classification, Machine Learning, Hyperspectral and LiDAR Data

### ACKNOWLEDGMENT

This work has been supported by Project MTM2016-76969-P from Ministerio de Economía y Competitividad - Agencia Estatal de Investigación.

## References

- [1] Breiman, L. (2001), Random Forests. *Machine Learning* 45(1), 5-32.
- [2] Cuesta-Albertos, J. A., Febrero-Bande, M., & Oviedo de la Fuente, M. (2017). The  $DD^G$ -classifier in the functional setting. *TEST*, 26(1), 119-142.
- [3] Febrero-Bande, M. & Oviedo de la Fuente, M. (2012). Statistical computing in functional data analysis: the R package fda. usc. *Journal of Statistical Software*, 51(4):1-28.
- [4] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*. Springer.
- [5] Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics*, 28(2), 337-407.

## AUTORES

Alonso Pena, M. ....	3
Borrajo García, M.I.....	46
Buide Carballosa, A.B.....	32
Conde Amboage, M.....	46
Costa Bouzas, J. ....	11
Cotos Yáñez, T.R.....	12
Espinoza Burgos, N.H. ....	6
Febrero Bande, M. ....	43
Fernández Arias, M. ....	9
Fernández Casal, R. ....	11
Ferreiro Díaz, M.....	12
Flores, M. ....	28
Ginzo Villamayor, M.J.....	32, 37, 38, 48
González Carro, V.....	35
Iglesias Patiño, C. ....	25
Juan Anido Rodríguez, J.....	35
Juncal Pérez, J.L. ....	16
Lafuente Rego, B. ....	11
Lombardía Cortiña, M.J. ....	11
López Vizcaíno, E. ....	17
Martín Vila, M. ....	25
Martins, S.R. ....	21
Meneses, A. 27.....	28
Naya Fernández, S. ....	28
Novo Pérez, M.A. ....	32
Núñez Lugilde, I. ....	6
Oviedo de la Fuente, M. ....	49
Pateiro López, B. ....	43
Robles, Y. ....	28
Rodríguez Gayoso, R. ....	37, 38
Rodríguez Muíños, M.A.....	37, 38
Roye, D.....	39

Ruano Ordás, D. ....	12
Sabela Rodríguez Calvo, S. ....	35
Tarrío Saavedra, J.....	11, 28
Vaamonde Rivas, M. ....	43
Villar Cheda, E. ....	25





# VI XORNADA DE USUARIOS DE EN GALICIA



```
boxplot(x, ...)  
boxplot.default(x, ..., range = 1.5, width = NULL,  
  varwidth = FALSE, notch = FALSE, names, boxwex = 0.8,  
  data = parent.frame(), plot = TRUE,  
  border = par("fg"), col = NULL, log = "", pars = NULL,  
  horizontal = FALSE, add = FALSE)  
boxplot.formula(formula, data = NULL, subset, na.action, ...)
```



## > ORGANIZA



## > COLABORAN



## > PATROCINAN



ISBN 9 7884091150243