

Trabajar en investigación biosanitaria: Matemáticas y Estadística contra el Cáncer

Daniel Redondo Sánchez

Orientación Profesional Estudiantes Grado en Estadística (6^a edición)
21 de mayo 2021



UNIVERSIDAD
DE GRANADA

rcg registro de cáncer de granada

A Junta
de Andalucía | Consejería de Salud
y Familias | Escuela Andaluza
de Salud Pública

 **aecc**
Contra el Cáncer

ciberesp

¿Quién soy?

Daniel Redondo Sánchez



- 2014 - Licenciado en **Matemáticas** (UGR 2009-2014).
- 2015 - Máster de **Matemáticas** (UAL, UCA, UGR, UMA, UJA).



- 2015 - Prácticas del Máster de Matemáticas en la Escuela Andaluza de Salud Pública.

¿Quién soy?

Daniel Redondo Sánchez

- Diploma de Especialización en **Epidemiología e Investigación Clínica** (UGR, EASP, 2016).
- Máster en **Ciencia de Datos** e Ingeniería de Computadores (UGR, 2020).
- Técnico en 3 proyectos de investigación.

Y actualmente...

¿Quién soy?

Daniel Redondo Sánchez

- Diploma de Especialización en **Epidemiología e Investigación Clínica** (UGR, EASP, 2016).
- Máster en **Ciencia de Datos** e Ingeniería de Computadores (UGR, 2020).
- Técnico en 3 proyectos de investigación.

Y actualmente...

- Doctorando en Medicina Clínica y Salud Pública. Línea de investigación de Epidemiología y Salud Pública.
- Técnico de investigación de un proyecto de la **Asociación Española Contra el Cáncer**.



Índice

1. Epidemiología y cáncer
2. Herramientas
3. Series temporales
4. Machine learning
5. Análisis espacial

Índice

1. Epidemiología y cáncer
2. Herramientas
3. Series temporales
4. Machine learning
5. Análisis espacial

Epidemiología y cáncer

Epidemiología

La epidemiología es la ciencia que estudia la frecuencia y distribución de las enfermedades en las poblaciones humanas, así como las causas que los producen.

Cáncer

El cáncer es una enfermedad en la que se produce una división incontrolada de las células. No es una única enfermedad, sino un conjunto de enfermedades (+100 tipos distintos de cáncer).

¿Por qué es importante investigar en cáncer?

1. Es una enfermedad **muy frecuente**: 19 millones de casos anuales en todo el mundo, 282.000 en España.

1 de cada 3 mujeres y 1 de cada 2 hombres
desarrollará cáncer a lo largo de su vida.

2. Es una enfermedad con **alta mortalidad**: 10 millones de defunciones anuales por cáncer en el mundo, 113.000 en España. La supervivencia a 5 años está en torno al 60%, con diferencias por sexos y localizaciones.

Epidemiología y cáncer

¿Qué ciencias son útiles para investigar en cáncer?

- Medicina
- Enfermería
- Biología
- Nutrición
- Ciencias ambientales
- Psicología
- Matemáticas
- **Estadística**
- Y muchas más...

Equipos **interdisciplinares**.

Epidemiología y cáncer

Registro de Cáncer de Granada

- Escuela Andaluza de Salud Pública (Campus de Cartuja).
- Recoge todos los casos de cáncer de la provincia de Granada desde 1985: +130.000 casos.
- Equipo: 14 personas (2 matemáticos, 2 bioestadísticos).



<https://www.registrocancergranada.es>

Algunas tareas de un estadístico en un Registro de Cáncer

- Mantenimiento de base de datos.
- Control de calidad.
- Cálculo de tasas: brutas, estandarizadas, acumulativas, truncadas.
- Generación de tablas estadísticas.
- Automatización de procesos.
- Diseño epidemiológico.
- Elaboración de informes y artículos científicos.
- Divulgación científica.

Índice

1. Epidemiología y cáncer

2. **Herramientas**

3. Series temporales

4. Machine learning

5. Análisis espacial

Herramientas

Programación

- Es muy importante para **resolver problemas**.
- **Automatiza** procesos manuales, largos y tediosos.
- El lenguaje es lo de menos: **piensa en programación**.
- **Ordena y documenta** el código: para los demás, pero también para tu futuro yo.

```
gr.stats2 <- function(stat, n = 10, bp = .4, ny = 1, nx = 0, f = 1, partidos.mínimos = 5, breaks = NA){  
  stats2 <- subset(stats, Partidos > partidos.mínimos)  
  top10 <- top_n(stats2, n, stats2[stat] / Partidos)  
  g <- ggplot(data = stats, aes(Partidos, eval(as.name(stat)), label = Jugador)) +  
    geom_point(alpha = 0.3, shape = 16, col = "dodgerblue", size = 2) +  
    geom_point(data = top10, aes(Partidos, eval(as.name(stat))), alpha = 1, shape = 16, col = "darkblue", size = 3) +  
    geom_text_repel(data = top10, box.padding = bp, nudge_y = ny, force = f, nudge_x = nx, segment.alpha = 0.5, size = 3.5) +  
    ylab(stat) +  
    ylim(0, 20) +  
    scale_x_continuous(limits = c(0, 27), breaks = c(seq(0, 25, 5), 27)) +  
    theme_classic() +  
    theme(axis.text = element_text(color = "black"))  
  
  if(is.na(breaks[1]) == FALSE) g <- g + scale_y_continuous(breaks = breaks)  
  cbind(top10["Jugador"], top10["Partidos"], top10[stat]) %>% as.data.frame() %>% arrange(desc(eval(as.name(stat)))) %>% print  
  
  return(g)  
}
```

Herramientas

Programación

- Es muy importante para **resolver problemas**.
- **Automatiza** procesos manuales, largos y tediosos.
- El lenguaje es lo de menos: **piensa en programación**.
- **Ordena y documenta** el código: para los demás, pero también para tu futuro yo.

```
# Carpeta de datos
setwd("TFM/Analisis_cr/data/")

# ----- Carga de paquetes -----

# Instalación de KnowSeq: (es una versión fija de GitHub)
# devtools::install_github("CasedUgr/KnowSeq", ref = "f59cb9e1cb02702697c208cf2c61c45d6e0b7a08", force = TRUE)
# Si hay problemas del tipo "Error: (converted from warning)": Sys.setenv(R_REMOTES_NO_ERRORS_FROM_WARNINGS="true")
library(KnowSeq)      # Para trabajar con datos de transcriptómica de GDC Portal
library(dplyr)        # Para select, filter, pipes, ...
library(tictoc)       # Para medir tiempos con tic() y toc() a lo MATLAB
library(beepR)         # Para avisar con beeps cuando acaba un proceso
library(caret)        # Para machine learning
library(e1071)         # Para SVM
library(gplots)        # Para heatmaps
library(reshape2)      # Para melt
library(ggalluvial)    # Para diagrama de Sankey

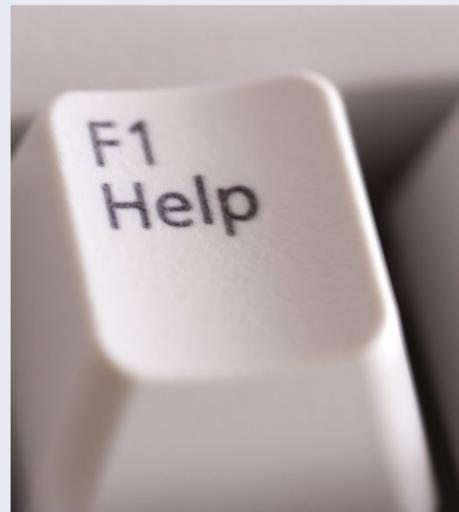
# ----- Sobreescibir la función dataPlot con una nueva función que pinta líneas discontinuas -----
```

Programación

- Hay mucha **ayuda...**

Google

stack**overflow**



Herramientas



+



Microsoft®
Visual Basic®
for Applications



Herramientas

TablaFinal_MujeresPeriodo4 - Access

Iniciar edición

Inicio

Crear

Datos externos

Herramientas de base de datos

Ayuda

Campos

Tabla

¿Qué desea hacer?

Ver

Cortar

Copiar

Pegar

Copiar formato

Filtrar

Ascendente

Descendente

Avanzadas

Quitar orden

Alternar filtro

Nuevo

Guardar

Actualizar todo

Eliminar

Buscar

Ajustar al formulario

Cambiar ventanas

Vistas

Portapapeles

Ordenar y filtrar

Registros

Buscar

Ventana

Todos los objetos

TablaFinal_AmbosSexosPeriodo3

TablaFinal_AmbosSexosPeriodo4

TablaFinal_HombresAño2011

TablaFinal_HombresAño2012

TablaFinal_HombresAño2013

TablaFinal_HombresAño2014

TablaFinal_HombresAño2015

TablaFinal_HombresPeriodo1

TablaFinal_HombresPeriodo2

TablaFinal_HombresPeriodo3

TablaFinal_HombresPeriodo4

TablaFinal_MujeresAño2011

TablaFinal_MujeresAño2012

periodo	sexo_txt	cie10_2n	Ncasos
2015-2015	Mujeres	TOTAL	2011
2015-2015	Mujeres	TOTAL, excepto piel no melanoma	1437
2015-2015	Mujeres	Piel no melanoma	574
2015-2015	Mujeres	Mama	422
2015-2015	Mujeres	ColonRecto	192
2015-2015	Mujeres	Colon	143
2015-2015	Mujeres	Cuerpo uterino	101
2015-2015	Mujeres	Melanoma de piel	59
2015-2015	Mujeres	Cerebro, sistema nervioso	56
2015-2015	Mujeres	Traquea, bronquios y pulmón	52
2015-2015	Mujeres	Vejiga	51
2015-2015	Mujeres	Tiroides	49
2015-2015	Mujeres	Recto	49
2015-2015	Mujeres	Otras y no especificadas.	42
2015-2015	Mujeres	Páncreas	42
2015-2015	Mujeres	Ovario	41

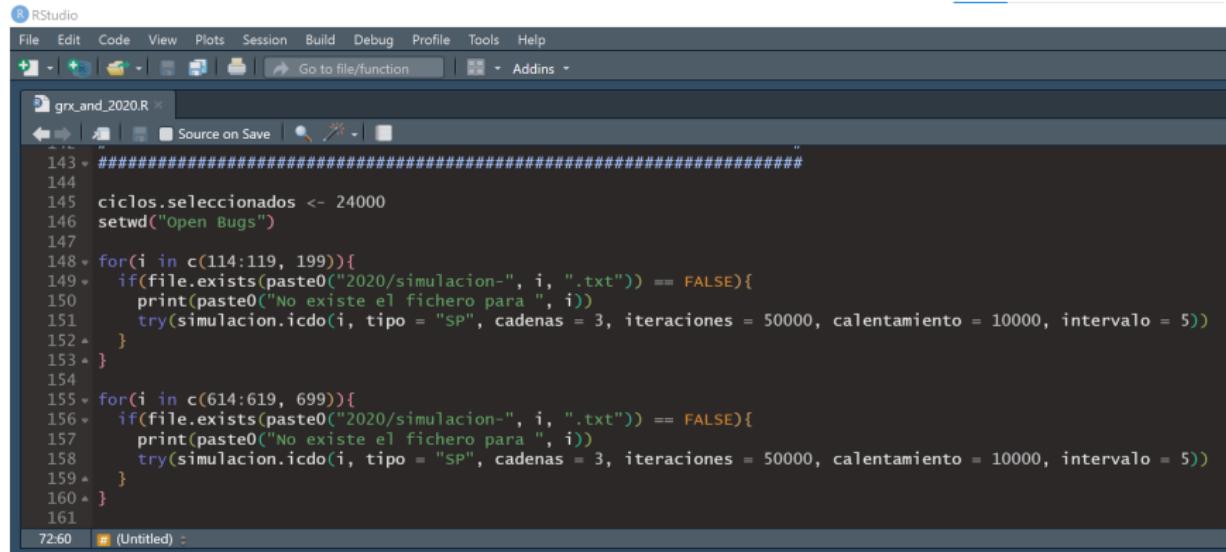
Herramientas

```
DoCmd.RunSQL "SELECT Z.cie10_1, Left([cie_10],3) as Expr1, 1 AS orden INTO agrupamoscie " _  
    & "FROM casos_RCG Z "  
    & "GROUP BY Z.cie10_1, Left([cie_10],3), 1 " _  
    & "ORDER BY Z.cie10_1, Left([cie_10],3);"  
With CurrentDb: With .OpenRecordset("agrupamoscie", dbOpenDynaset)  
    .MoveFirst  
    Do  
        i = !orden  
        a = !cie10_1:     .MoveNext: If .EOF Then Exit Do  
        b = !cie10_1  
        If a = b Then  
            .Edit: !orden = i + 1: .Update  
        Else  
            .Edit: !orden = 1:     .Update  
        End If  
    Loop While Not .EOF  
    .Close  
End With: End With
```

Herramientas



Herramientas



The screenshot shows the RStudio interface with a script file open. The menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The toolbar has icons for file operations like Open, Save, and Run. The code editor window displays a script named 'grx_and_2020.R' containing R code. The code uses loops to run simulations for different cycles and file paths.

```
143 #####
144
145 ciclos_seleccionados <- 24000
146 setwd("Open Bugs")
147
148 for(i in c(114:119, 199)){
149   if(file.exists(paste0("2020/simulacion-", i, ".txt")) == FALSE){
150     print(paste0("No existe el fichero para ", i))
151     try(simulacion.icdo(i, tipo = "SP", cadenas = 3, iteraciones = 50000, calentamiento = 10000, intervalo = 5))
152   }
153 }
154
155 for(i in c(614:619, 699)){
156   if(file.exists(paste0("2020/simulacion-", i, ".txt")) == FALSE){
157     print(paste0("No existe el fichero para ", i))
158     try(simulacion.icdo(i, tipo = "SP", cadenas = 3, iteraciones = 50000, calentamiento = 10000, intervalo = 5))
159   }
160 }
```

At the bottom left, the status bar shows '72:60' and '(Untitled)'. At the bottom right, there's a small orange icon.

Herramientas



Software específico de Registros de Cáncer:

- Control de calidad
- Conversión de codificaciones
- Análisis de tendencias

Herramientas

Herramientas

- Trabajo en equipo
- Inglés



- **Estadística pública:** Poblaciones, defunciones, encuestas, ...

Visualización de datos

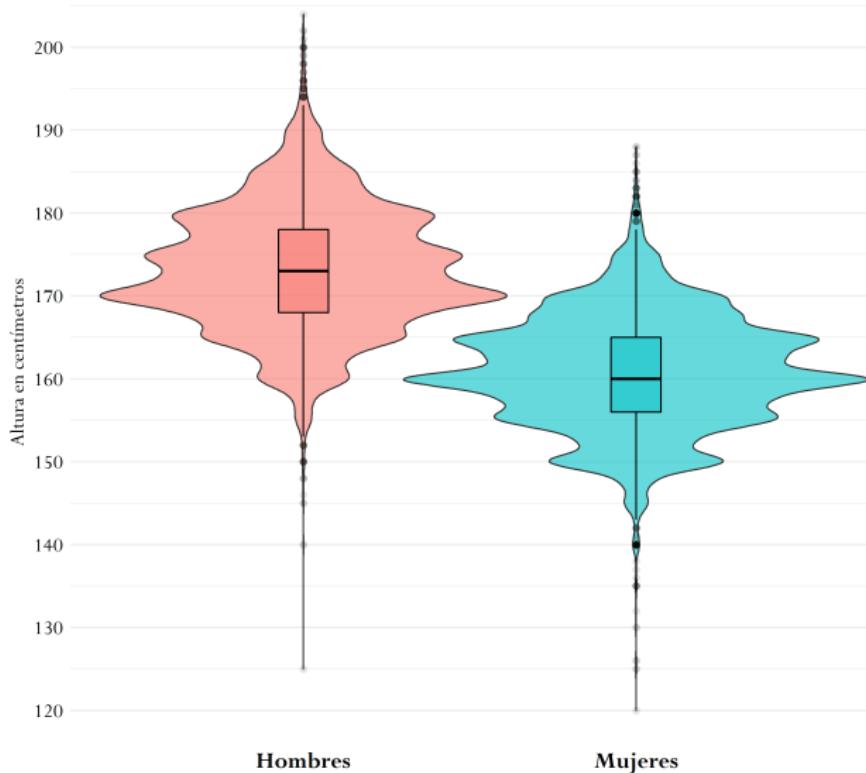
- Un gráfico vale más que mil palabras.
- Deben ser **claros** y **precisos**.
- Hay que saber **interpretar** y **crear** gráficos.
- A veces son convenientes **gráficos interactivos**:
<https://www.danielredondo.com/grafico2>
- Es uno de los puntos fuertes de R con el paquete {ggplot2}.



Herramientas

Altura por sexos en España, 2017

Respuesta a la pregunta "¿Podría decirme cuánto mide, aproximadamente, sin zapatos?"

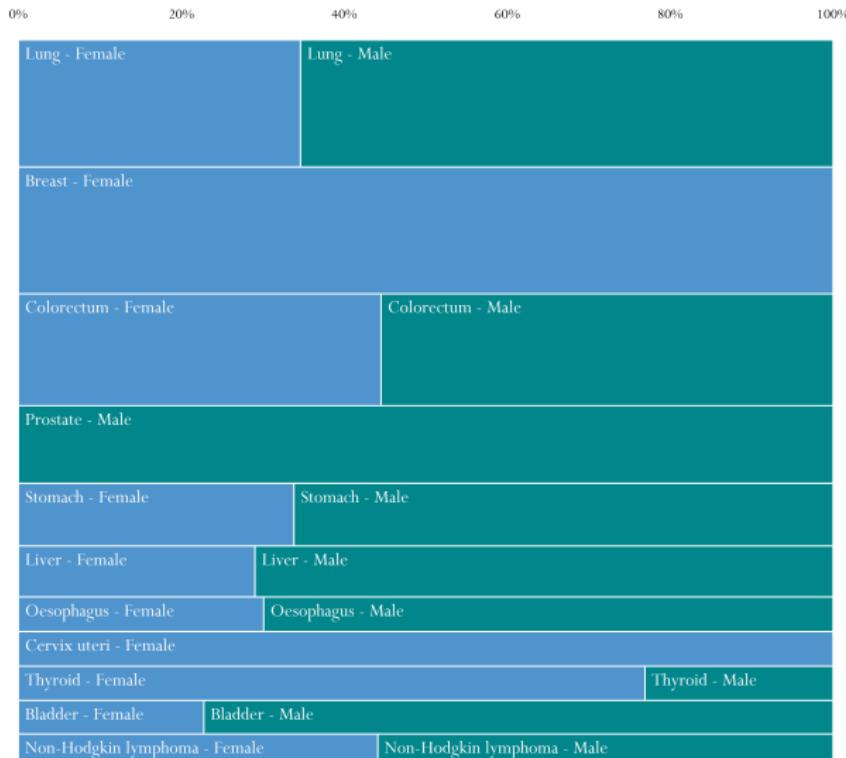


Herramientas

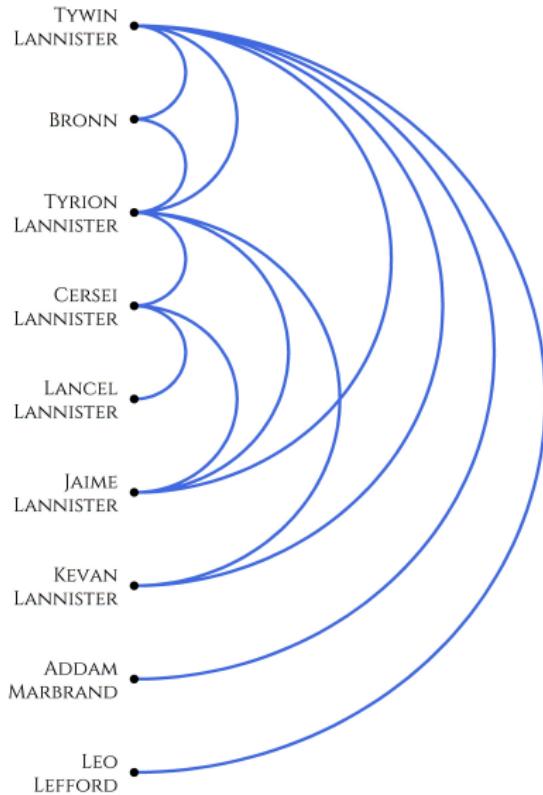
Cancer incidence, 2018

Distribution of cases by sex and anatomical site with +500,000 cases diagnosed.

Source: Global Cancer Observatory (World Health Organization).

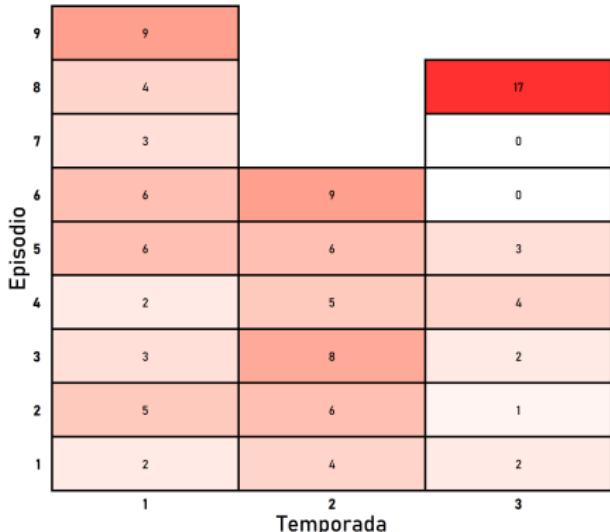


Herramientas

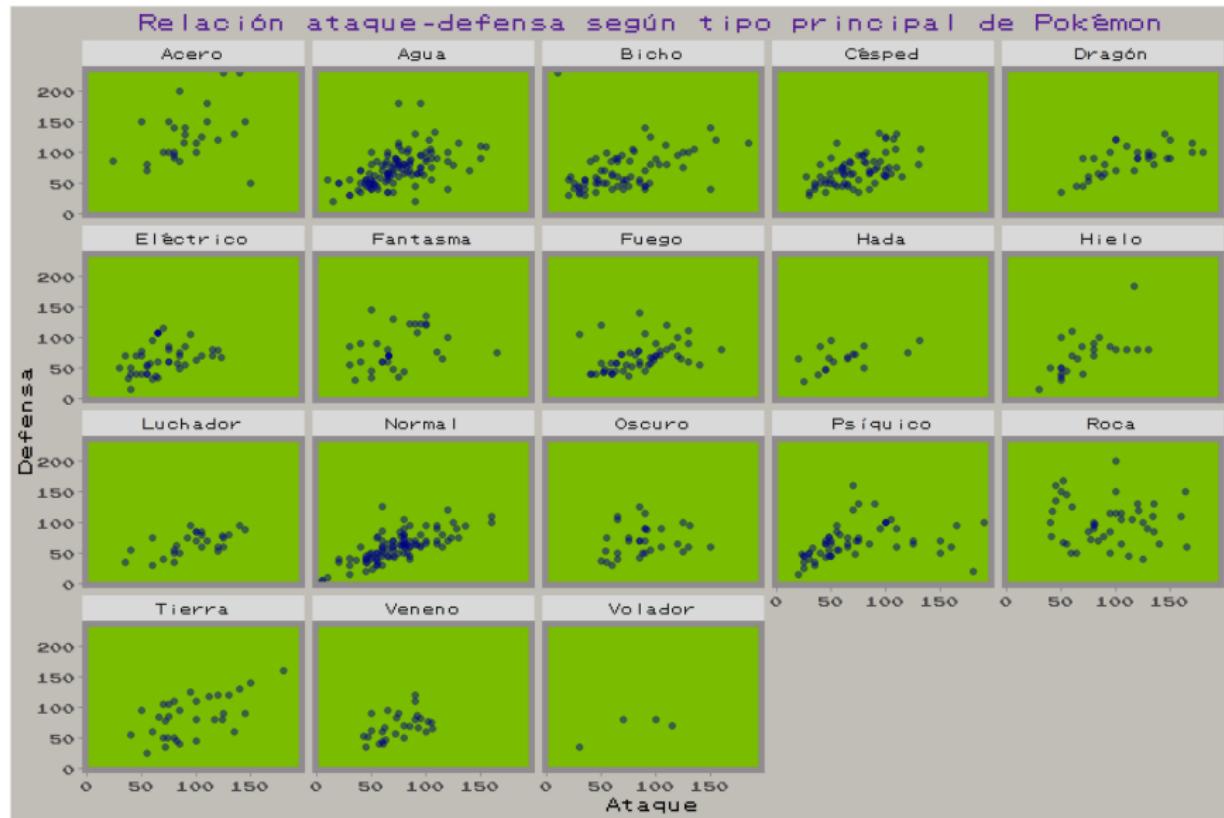


LA CASA DE PAPEL: NAIROBI

Número de veces que se menciona "Nairobi" por capítulo.
Gráficos por [@dredondosanchez](#)



Herramientas



Herramientas

Muchos aspectos a tener en cuenta...

R3



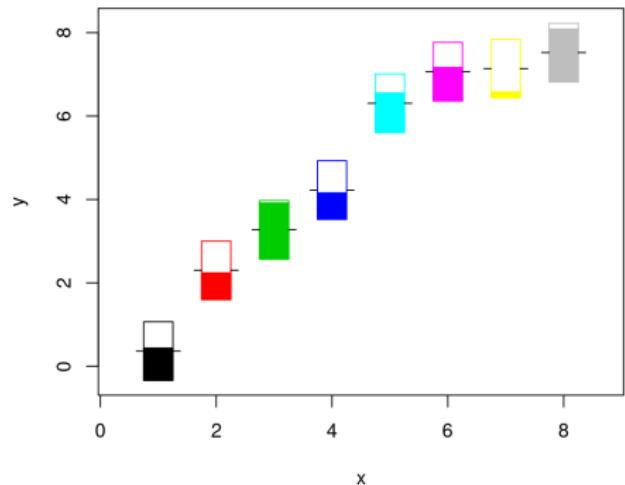
R4



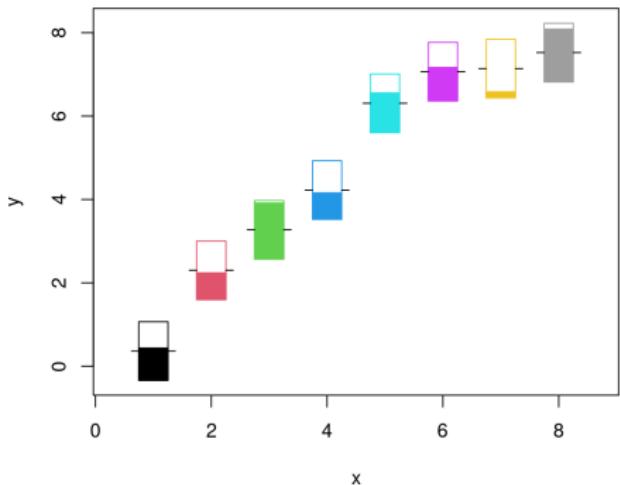
R va a cambiar, **por primera vez en 20 años**,
la paleta de colores por defecto. *¿Por qué?*

Herramientas

Paleta vieja



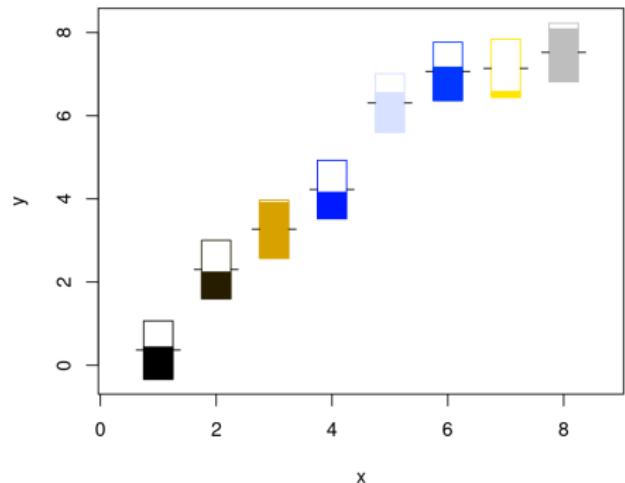
Paleta nueva



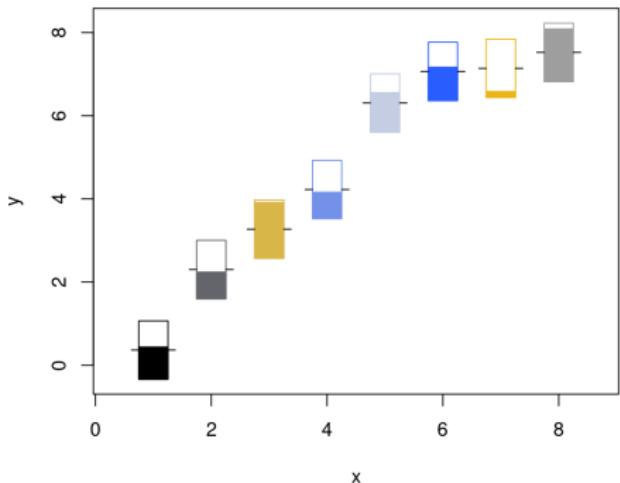
¿Por qué?

Herramientas

Paleta vieja



Paleta nueva



Simulación de daltonismo rojo-verde ($\sim 4\%$ población)

Índice

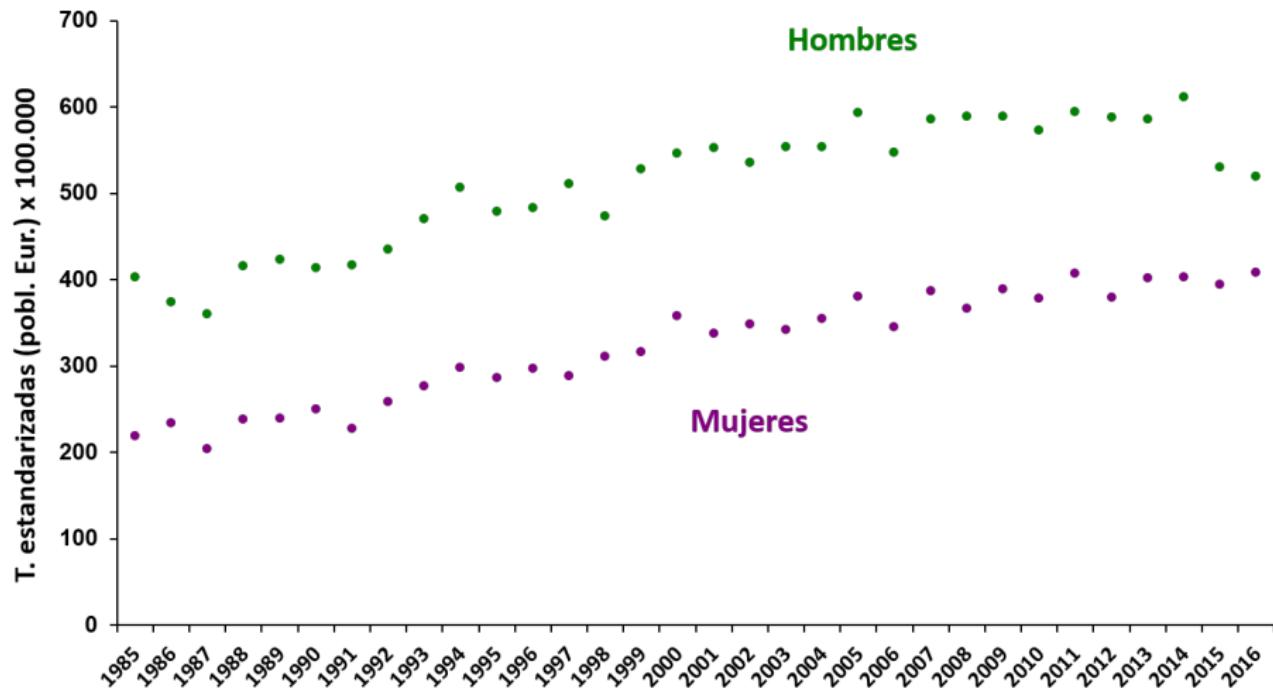
1. Epidemiología y cáncer
2. Herramientas
3. **Series temporales**
4. Machine learning
5. Análisis espacial

Series temporales

1. **Tendencias de la incidencia de cáncer**
2. Proyecciones de la incidencia de cáncer
3. Estimaciones de la incidencia de cáncer

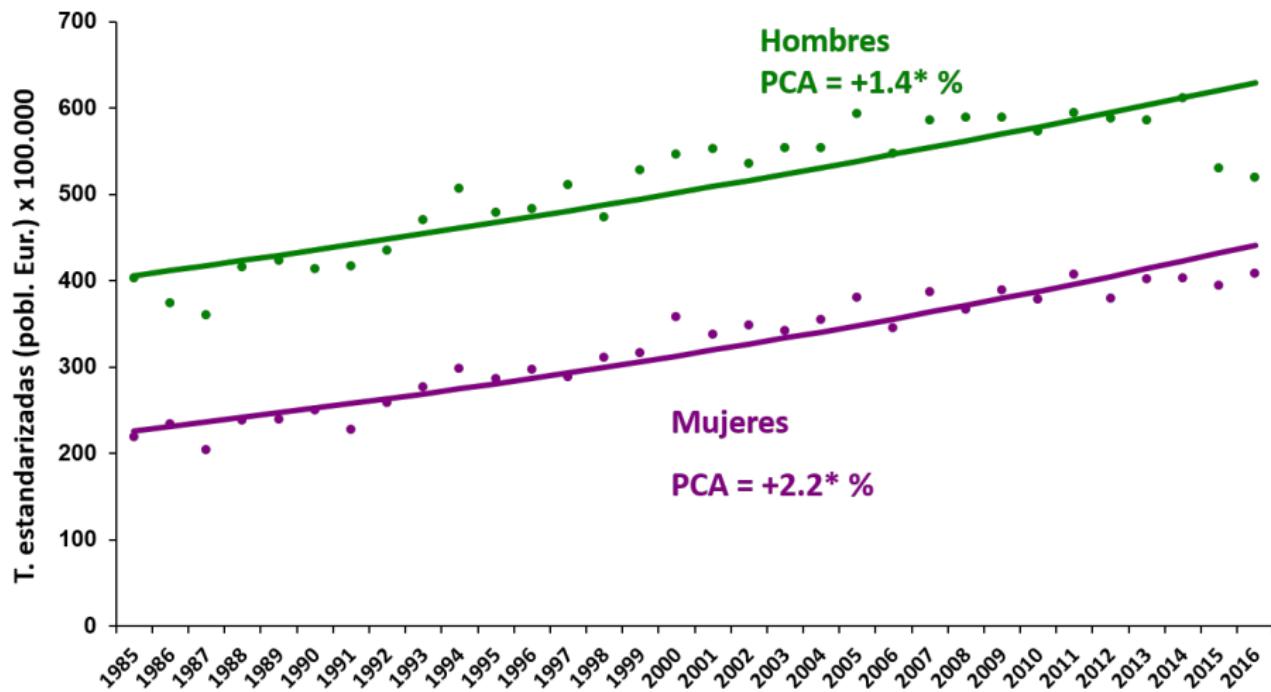
Series temporales - Tendencias

Incidencia del total del cáncer. Provincia de Granada, 1985-2016.



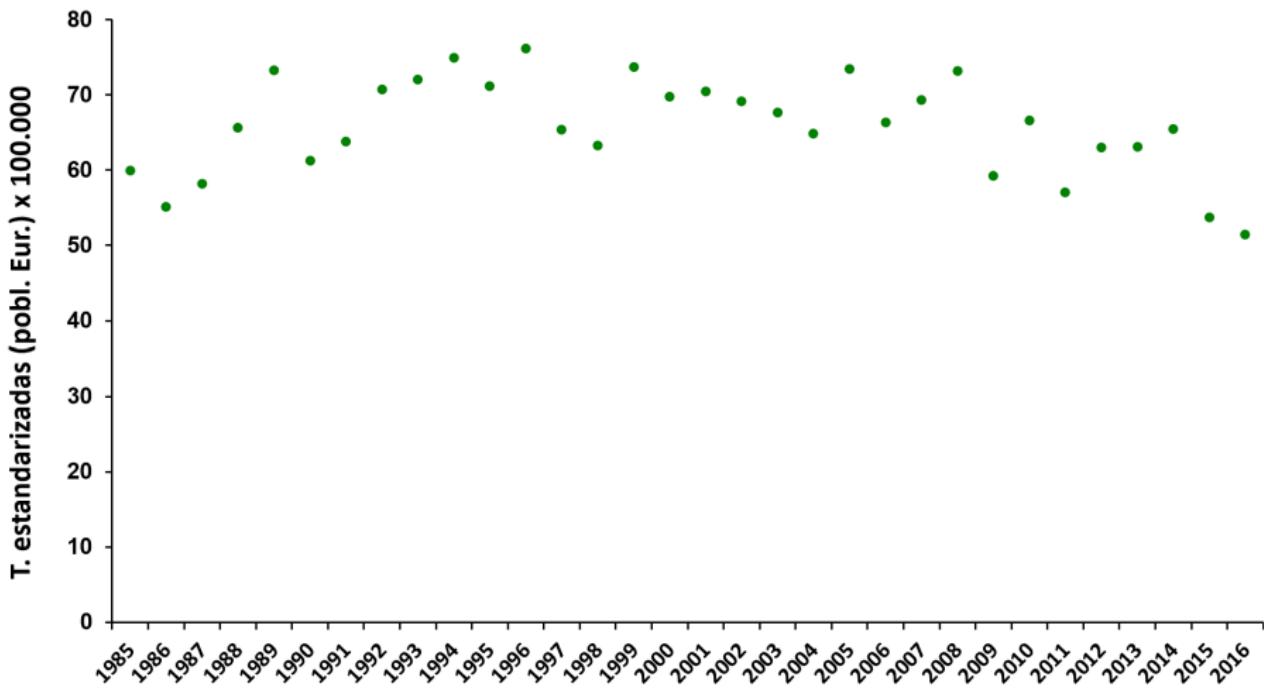
Series temporales - Tendencias

Incidencia del total del cáncer. Provincia de Granada, 1985-2016.



Series temporales - Tendencias

Incidencia de cáncer de pulmón en hombres. Provincia de Granada, 1985-2016.



Series temporales - Tendencias



JOINPOINT
REGRESSION PROGRAM

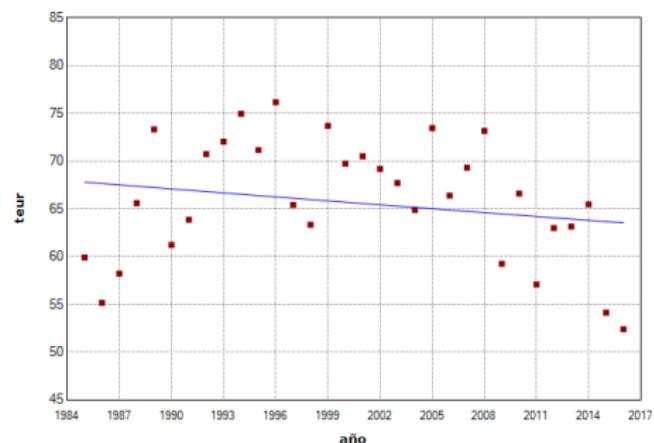
PRODUCED BY
Statistical Research and Applications Branch
Surveillance Research Program, Division of Cancer Control and Population Sciences
National Cancer Institute

Information Management Services, Inc.
Calverton, MD

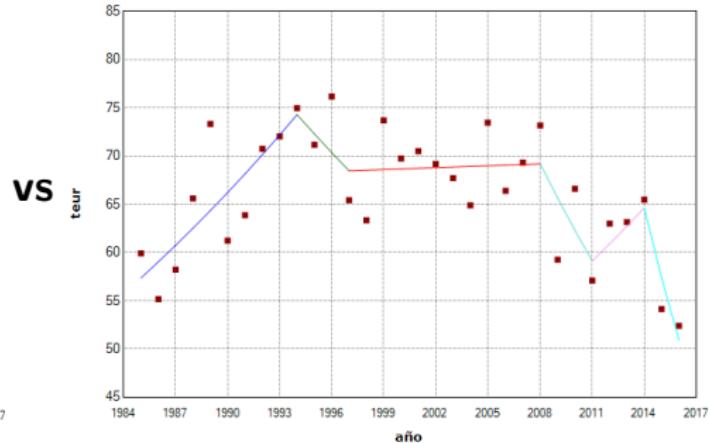
surveillance.cancer.gov/joinpoint
joinpoint@imsweb.com

Series temporales - Tendencias

0 puntos de inflexión

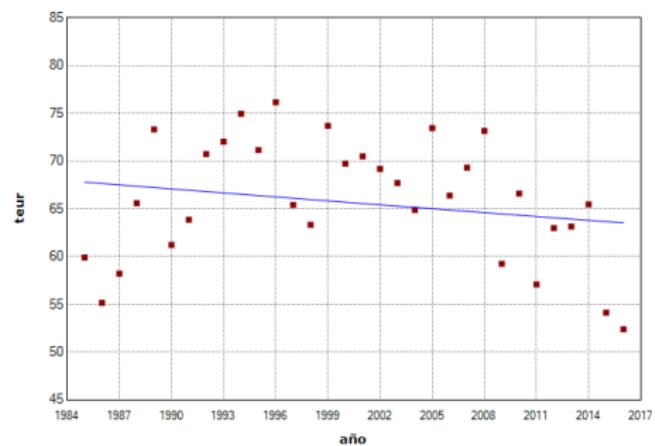


5 puntos de inflexión

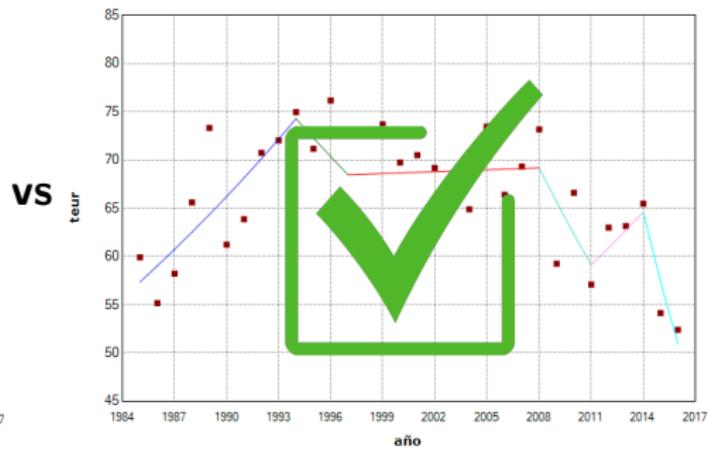


Series temporales - Tendencias

0 puntos de inflexión

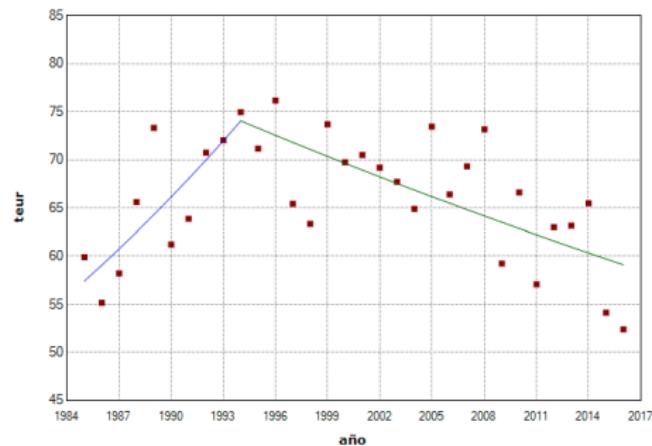


5 puntos de inflexión

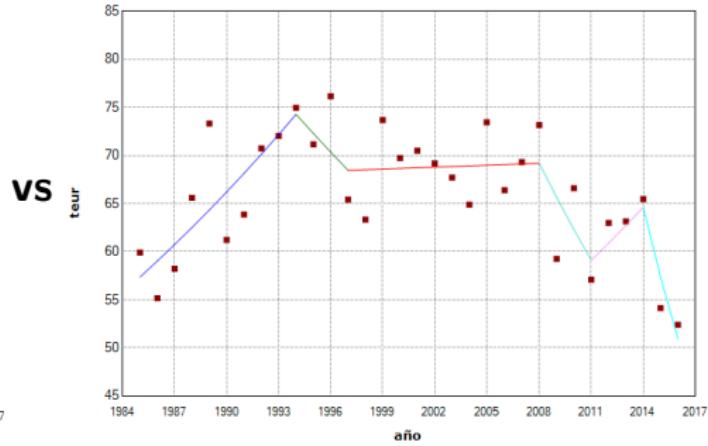


Series temporales - Tendencias

1 punto de inflexión

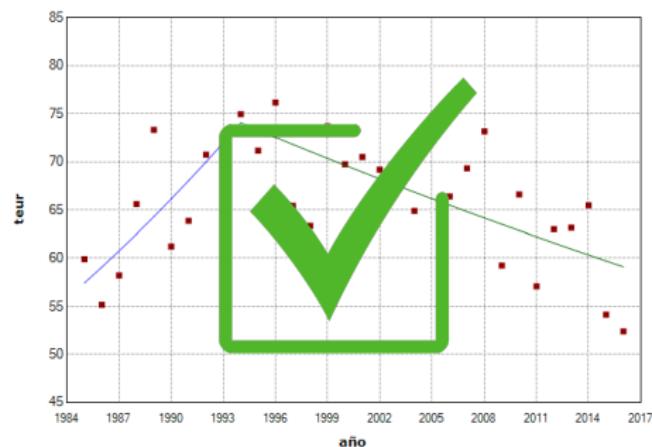


5 puntos de inflexión

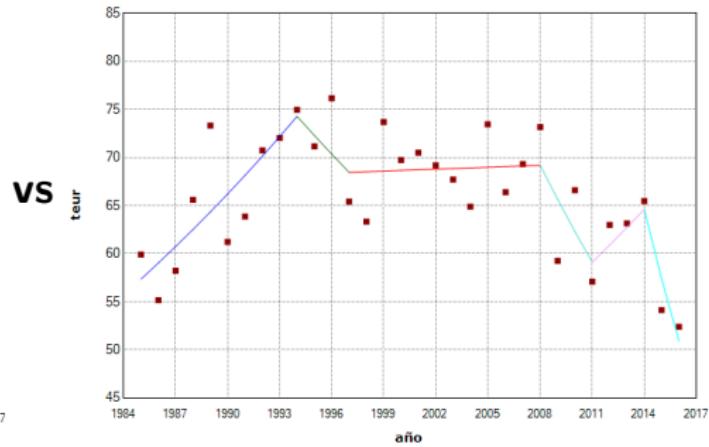


Series temporales - Tendencias

1 punto de inflexión

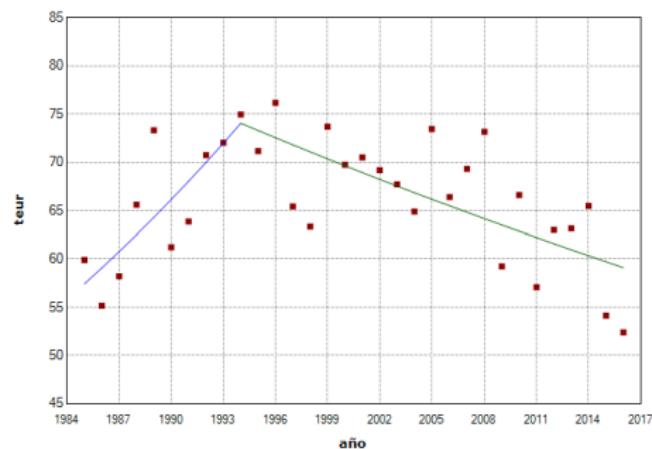


5 puntos de inflexión

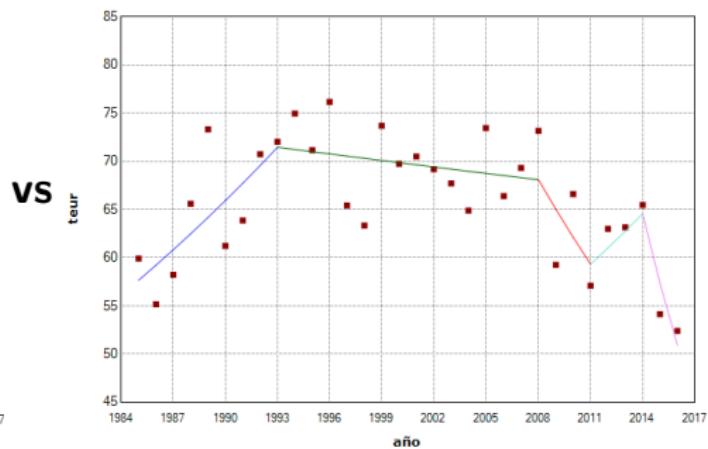


Series temporales - Tendencias

1 punto de inflexión

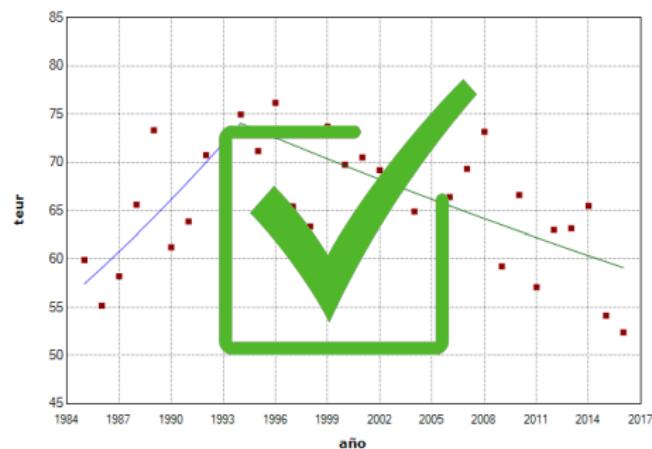


4 puntos de inflexión

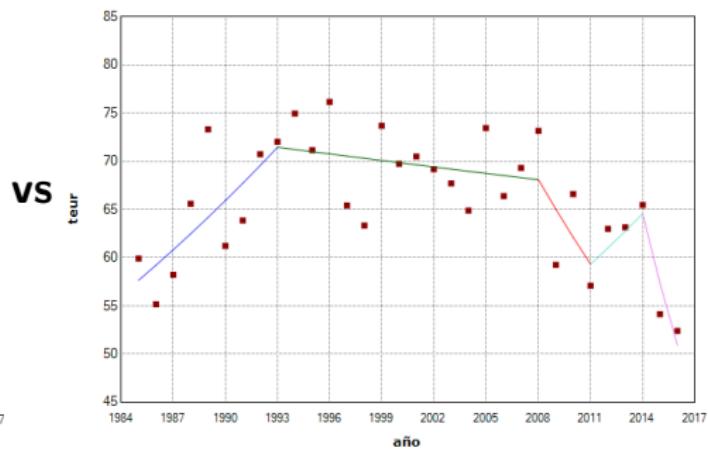


Series temporales - Tendencias

1 punto de inflexión

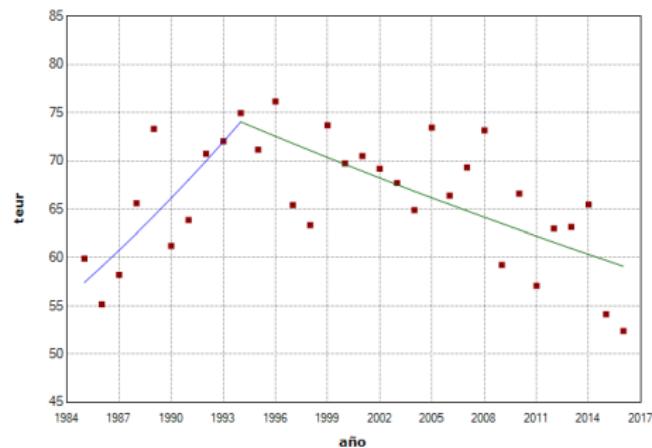


4 puntos de inflexión

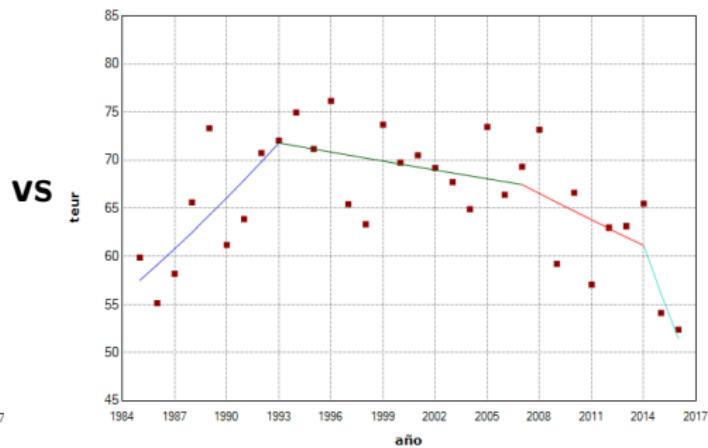


Series temporales - Tendencias

1 punto de inflexión

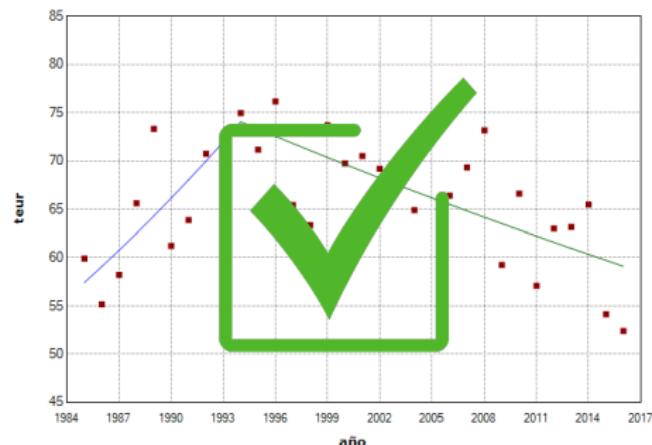


3 puntos de inflexión

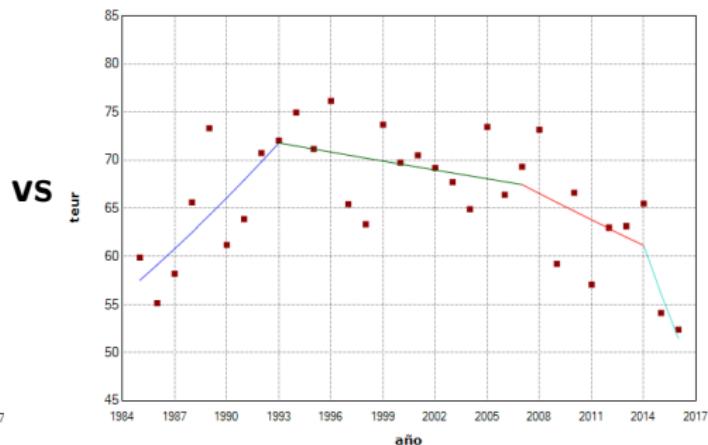


Series temporales - Tendencias

1 punto de inflexión

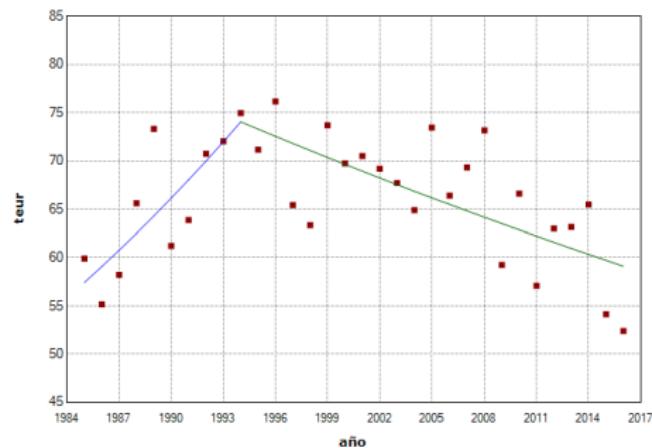


3 puntos de inflexión

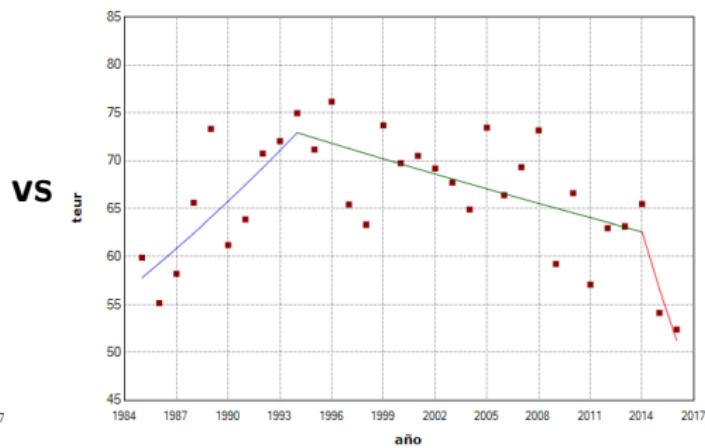


Series temporales - Tendencias

1 punto de inflexión



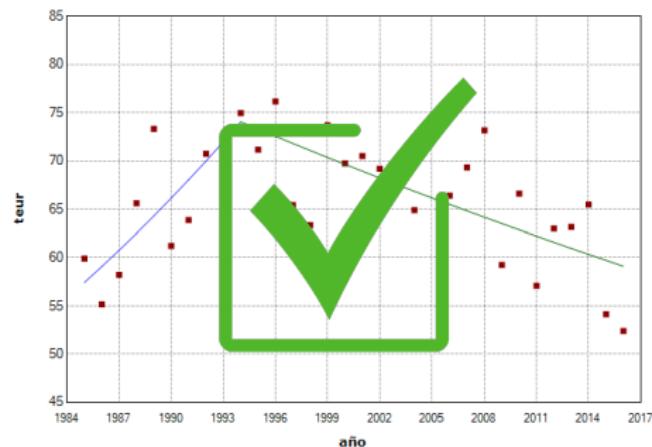
2 puntos de inflexión



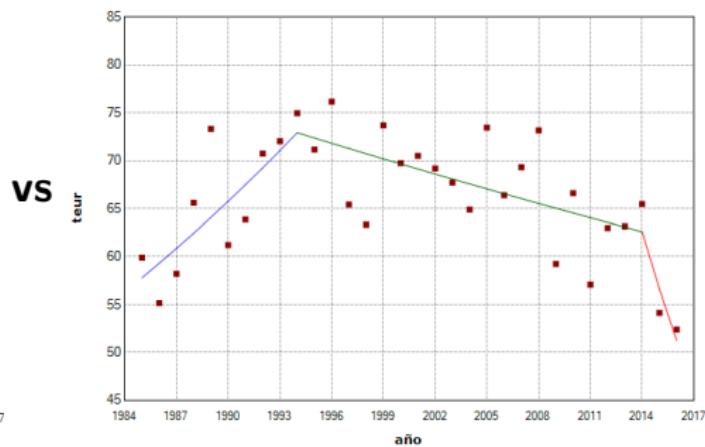
vs

Series temporales - Tendencias

1 punto de inflexión



2 puntos de inflexión



vs

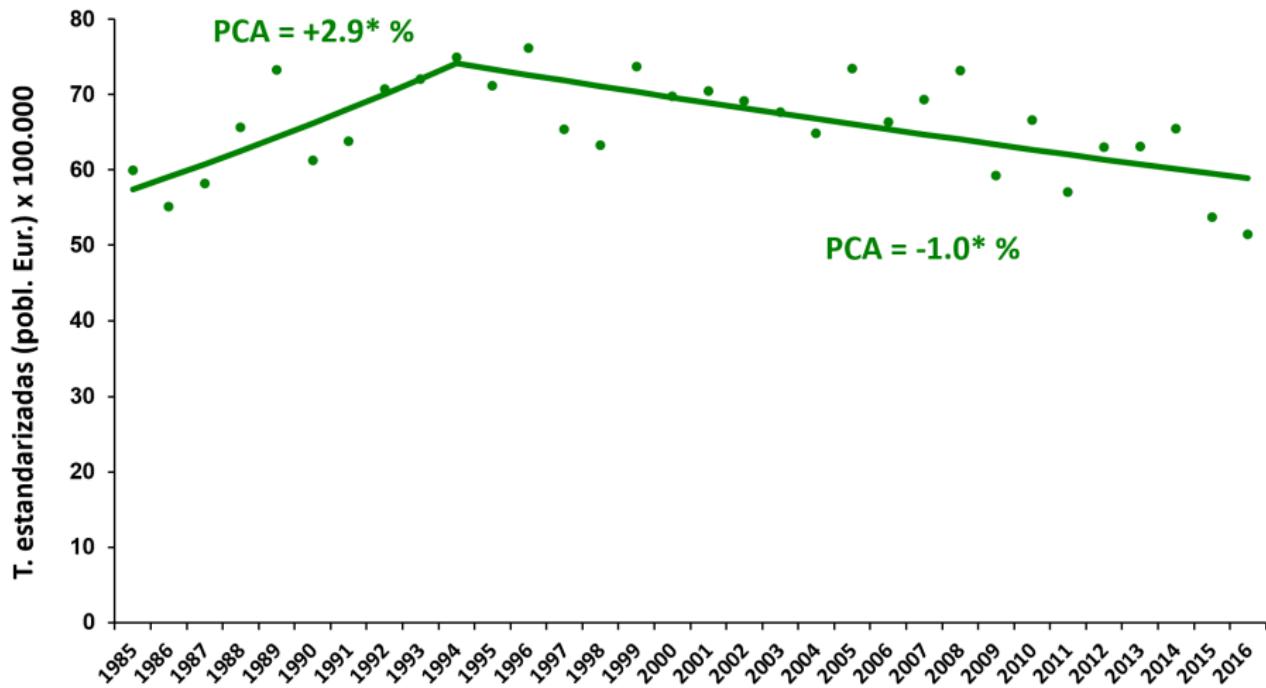
Series temporales - Tendencias

Cohort	Test Number	Null Hypothesis	Alternate Hypothesis	P-Value
1	#1	0 Joinpoint(s)	5 Joinpoint(s) ^	0.0008889
1	#2	1 Joinpoint(s) ^	5 Joinpoint(s)	0.1417778
1	#3	1 Joinpoint(s) ^	4 Joinpoint(s)	0.1217778
1	#4	1 Joinpoint(s) ^	3 Joinpoint(s)	0.1444444
1	#5	1 Joinpoint(s) ^	2 Joinpoint(s)	0.0328889

^ Selected Model

Series temporales - Tendencias

Incidencia de cáncer de pulmón en hombres. Provincia de Granada, 1985-2016.

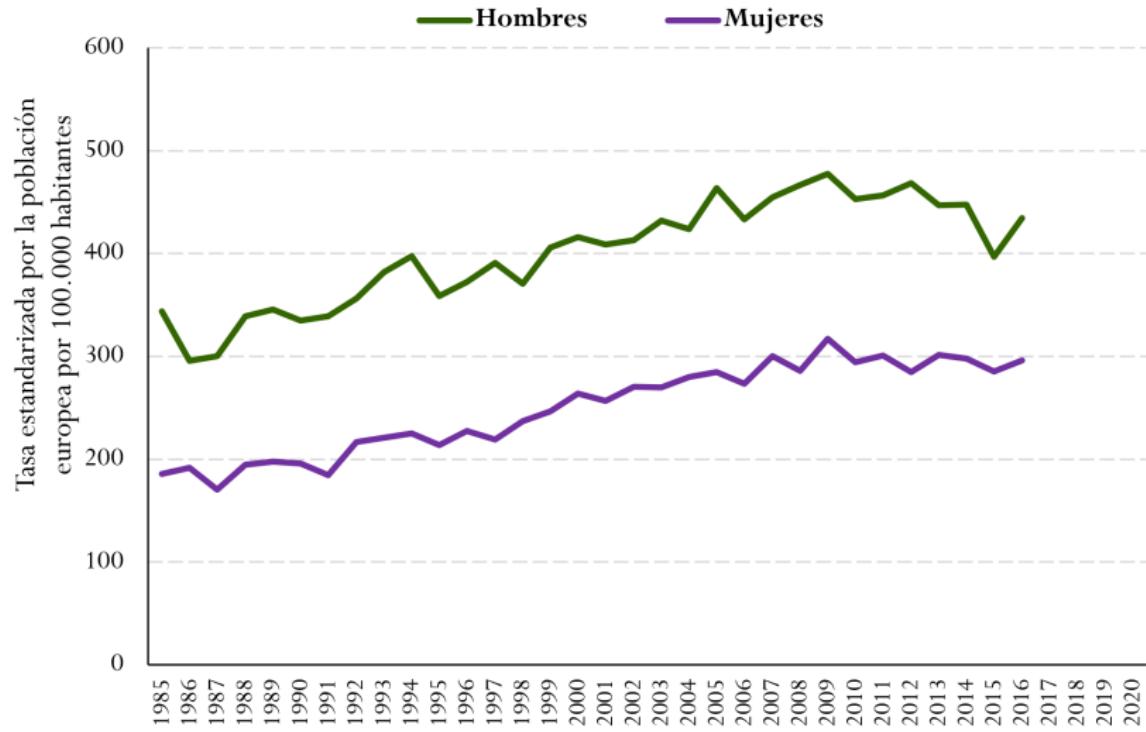


Series temporales

1. Tendencias de la incidencia de cáncer
2. **Proyecciones de la incidencia de cáncer**
3. Estimaciones de la incidencia de cáncer

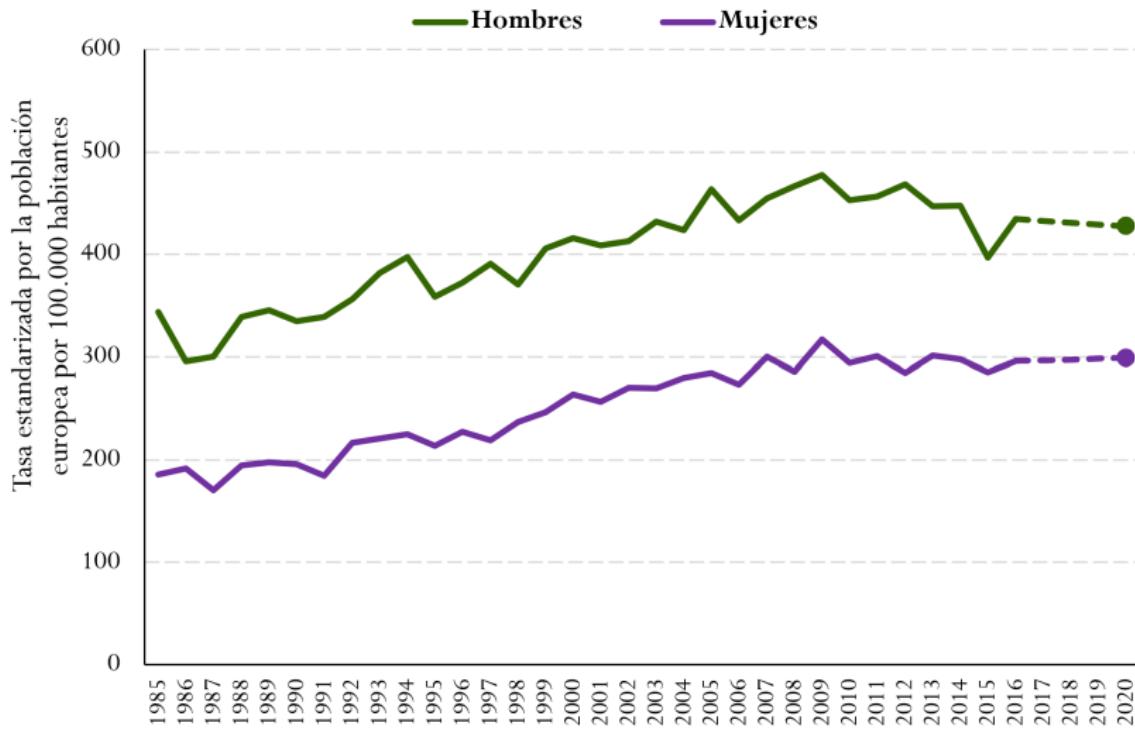
Series temporales - Proyecciones

Incidencia del total del cáncer excepto piel no melanoma. Provincia de Granada, 1985-2016.



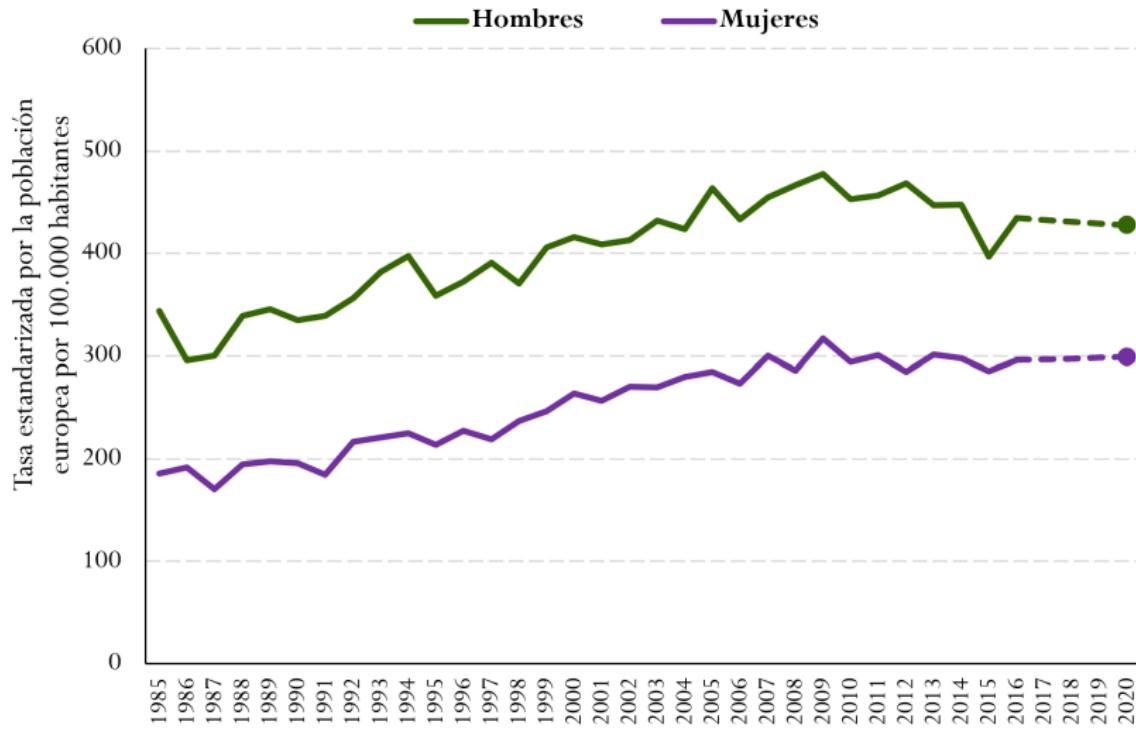
Series temporales - Proyecciones

$$\log(\text{CASOS}) = \alpha + \beta_0 \text{AÑO} + \sum_{i=1}^{18} \beta_i \text{EDAD}_i + \log(\text{POBLACIÓN})$$



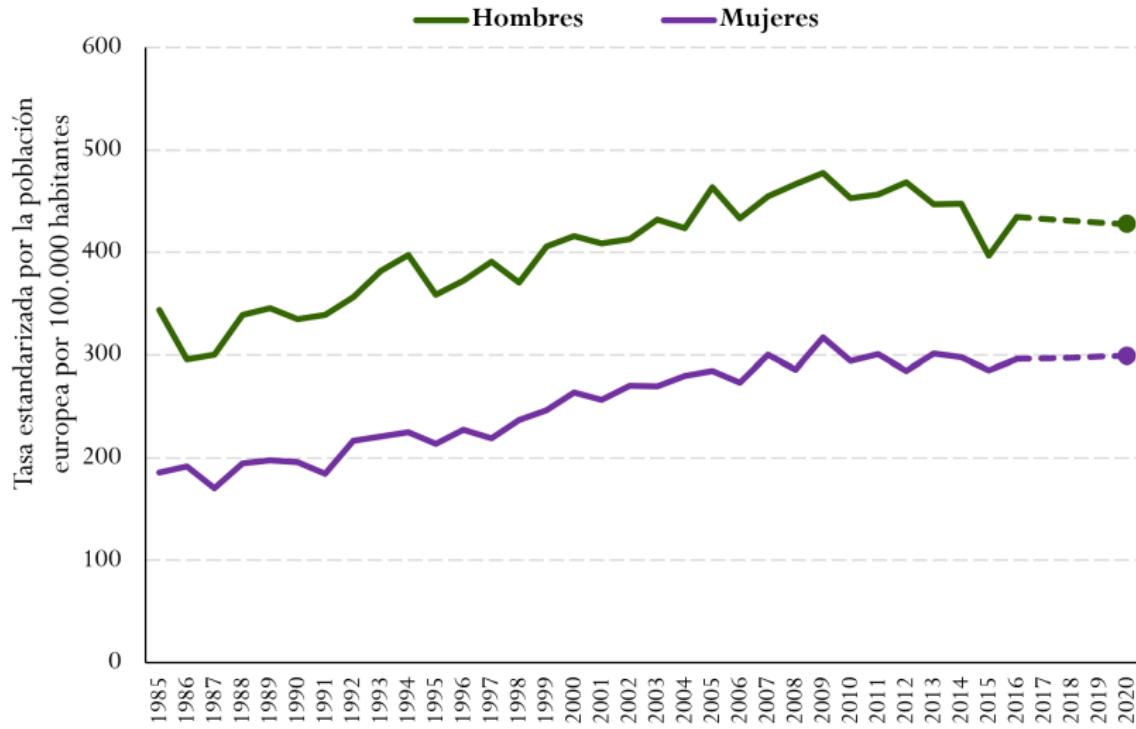
Series temporales - Proyecciones

$$\log(\text{CASOS}) - \log(\text{POBLACIÓN}) = \alpha + \beta_0 \tilde{\text{AÑO}} + \sum_{i=1}^{18} \beta_i \text{EDAD}_i$$



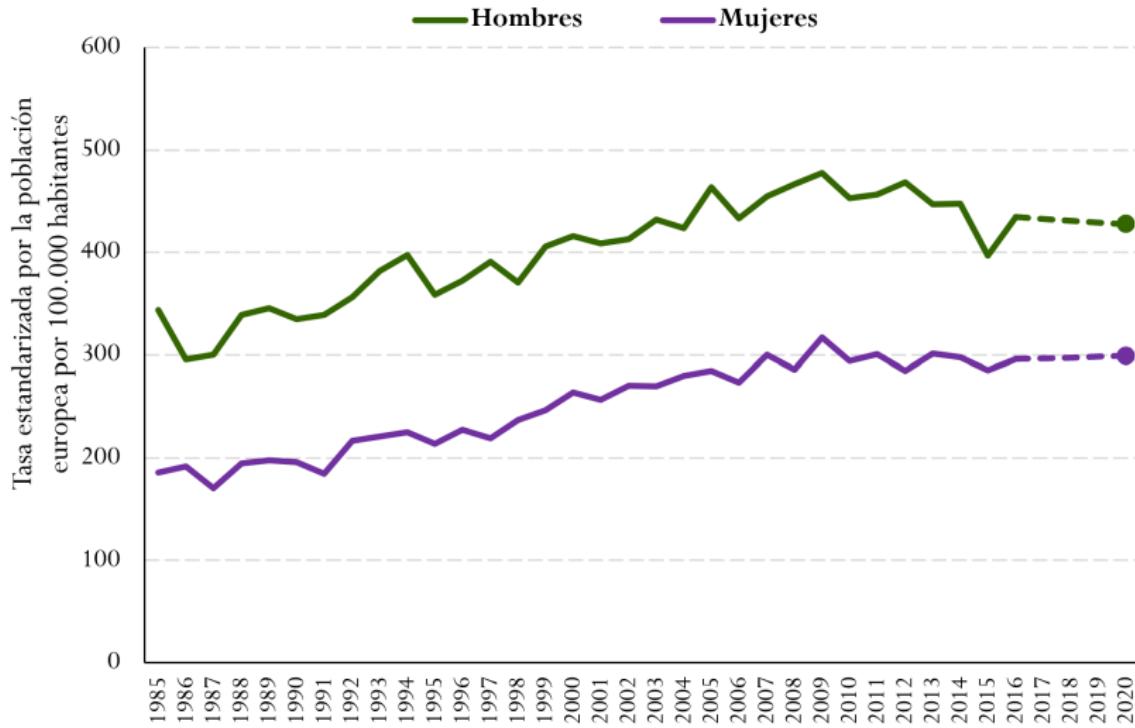
Series temporales - Proyecciones

$$\log\left(\frac{\text{CASOS}}{\text{POBLACIÓN}}\right) = \alpha + \beta_0 \text{AÑO} + \sum_{i=1}^{18} \beta_i \text{EDAD}_i$$



Series temporales - Proyecciones

$$\log(\text{TASA}) = \alpha + \beta_0 \text{AÑO} + \sum_{i=1}^{18} \beta_i \text{EDAD}_i$$



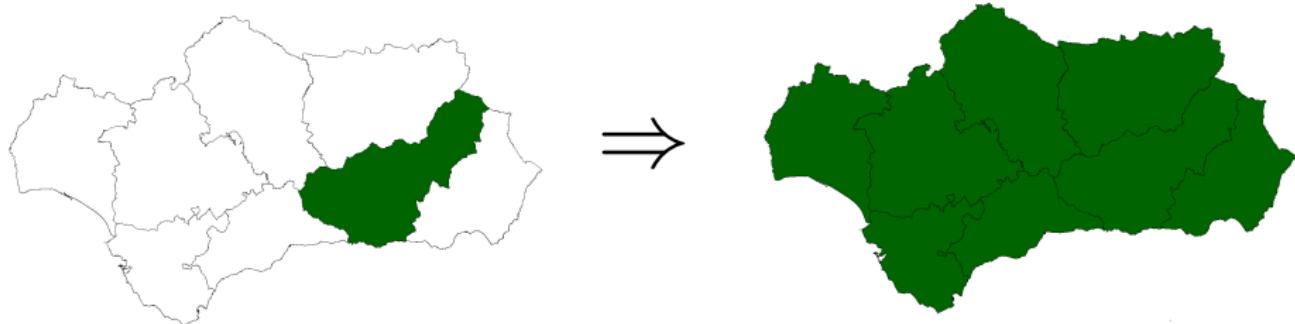
Series temporales

1. Tendencias de la incidencia de cáncer
2. Proyecciones de la incidencia de cáncer
3. **Estimaciones de la incidencia de cáncer**

Series temporales - Estimaciones



Series temporales - Estimaciones



Se estima la incidencia **actual** en Andalucía usando **los datos de Granada** y varios métodos estadísticos (**cadenas de Markov-Montecarlo, modelos edad-periodo-cohorte, modelos lineales generalizados mixtos, suavizado exponencial ...**)

Series temporales - Estimaciones

Gráfico 2. Estimaciones de incidencia de cáncer en Andalucía en hombres.

Tasa estandarizada por la población europea de 1976 (ASR-E) por 100.000 hombres.

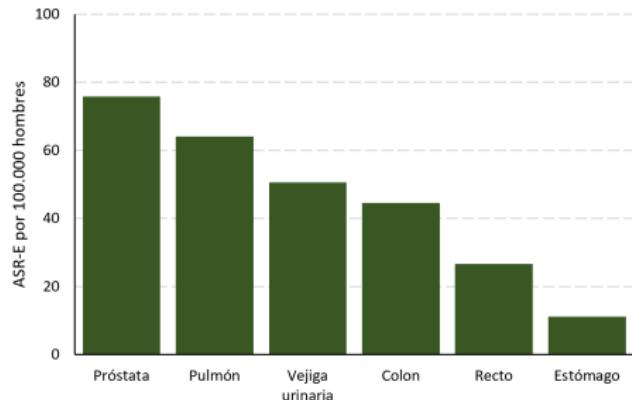
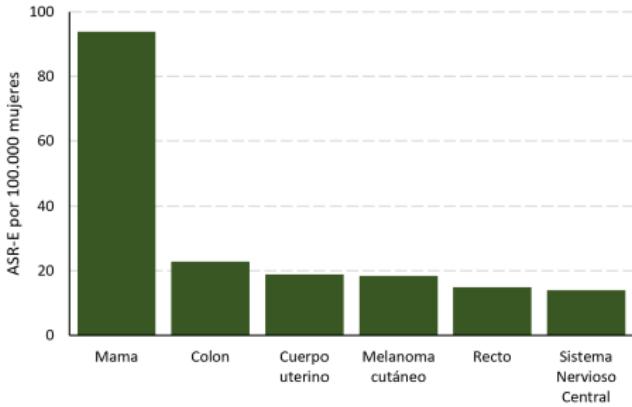


Gráfico 3. Estimaciones de incidencia de cáncer en Andalucía en mujeres.

Tasa estandarizada por la población europea de 1976 (ASR-E) por 100.000 mujeres.



Series temporales - Estimaciones

Gráfico 6. Estimaciones de incidencia del cáncer de próstata en Andalucía por provincias.

Tasa estandarizada por la población europea de 1976 (ASR-E) por 100.000 hombres.

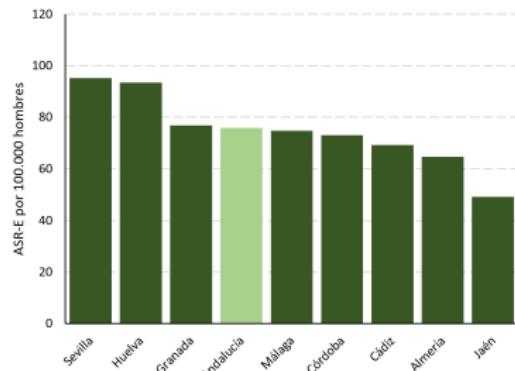
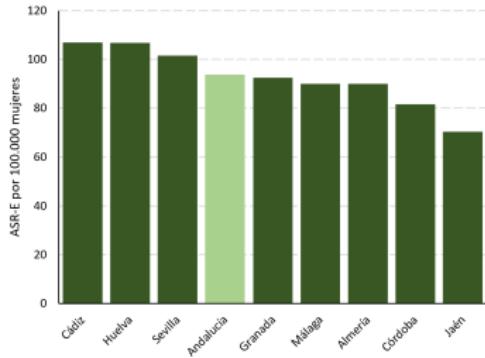


Gráfico 7. Estimaciones de incidencia del cáncer de mama en mujeres en Andalucía por provincias.

Tasa estandarizada por la población europea de 1976 (ASR-E) por 100.000 mujeres.



Series temporales - Estimaciones



TRABAJO FIN DE MÁSTER
MÁSTER DE MATEMÁTICAS

Modelización Matemática de la Estimación de Incidencia de Cáncer

Autor:
Daniel Redondo Sánchez

Tutores:
Francisco Javier Alonso Morales
DEPARTAMENTO DE ESTADÍSTICA

Miguel Rodríguez Barranco
REGISTRO DE CÁNCER DE GRANADA

Redondo-Sánchez et al. *Population Health Metrics* (2021) 19:18
<https://doi.org/10.1186/s12963-021-00248-1>

Population Health Metrics

RESEARCH

Open Access

Cancer incidence estimation from mortality data: a validation study within a population-based cancer registry

Daniel Redondo-Sánchez^{1,2,3}, Miguel Rodríguez-Barranco^{1,2,3*}, Alberto Ameijide⁴, Francisco Javier Alonso⁵, Pablo Fernández-Navarro^{3,6}, Jose Juan Jiménez-Moleón^{2,3,7} and María-José Sánchez^{1,2,3,7}



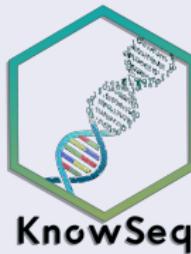
Índice

1. Epidemiología y cáncer
2. Herramientas
3. Series temporales
4. **Machine learning**
5. Análisis espacial

Machine learning

Machine learning

- Campo híbrido: Estadística + Informática + Matemáticas...
- Algoritmos de **selección de características**: elección de variables relevantes.
- Algoritmos de **regresión** y **clasificación**.
- Dos ejemplos con R + {KnowSeq}...

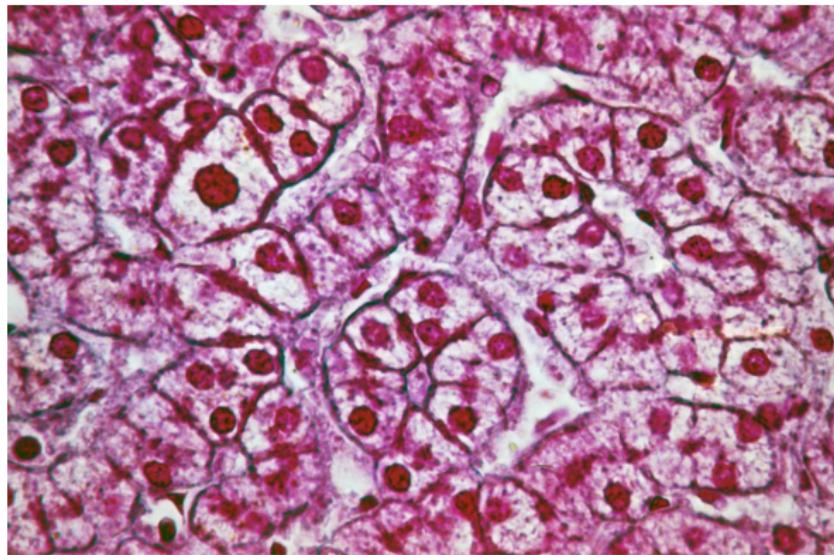


<https://github.com/CasedUgr/KnowSeq>

Machine learning

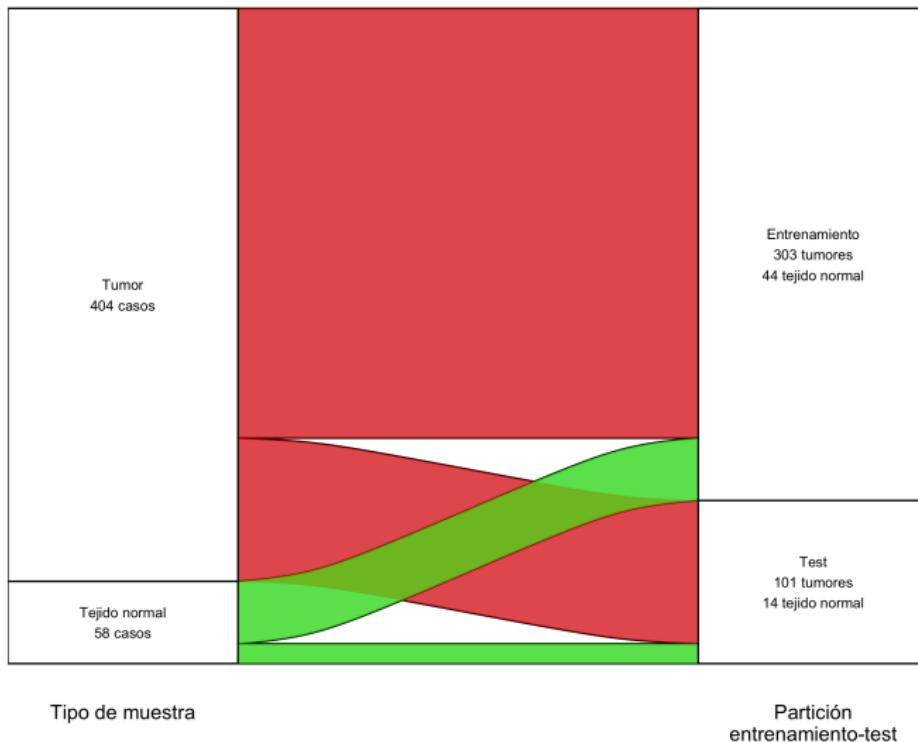
- 462 muestras de tejido de hígado.
- 404 identificados como tumores, 58 como tejido sano.
- +24.500 genes en cada muestra.

Objetivo: buscar genes que permitan “clasificar” una muestra nueva como tumor o tejido sano.

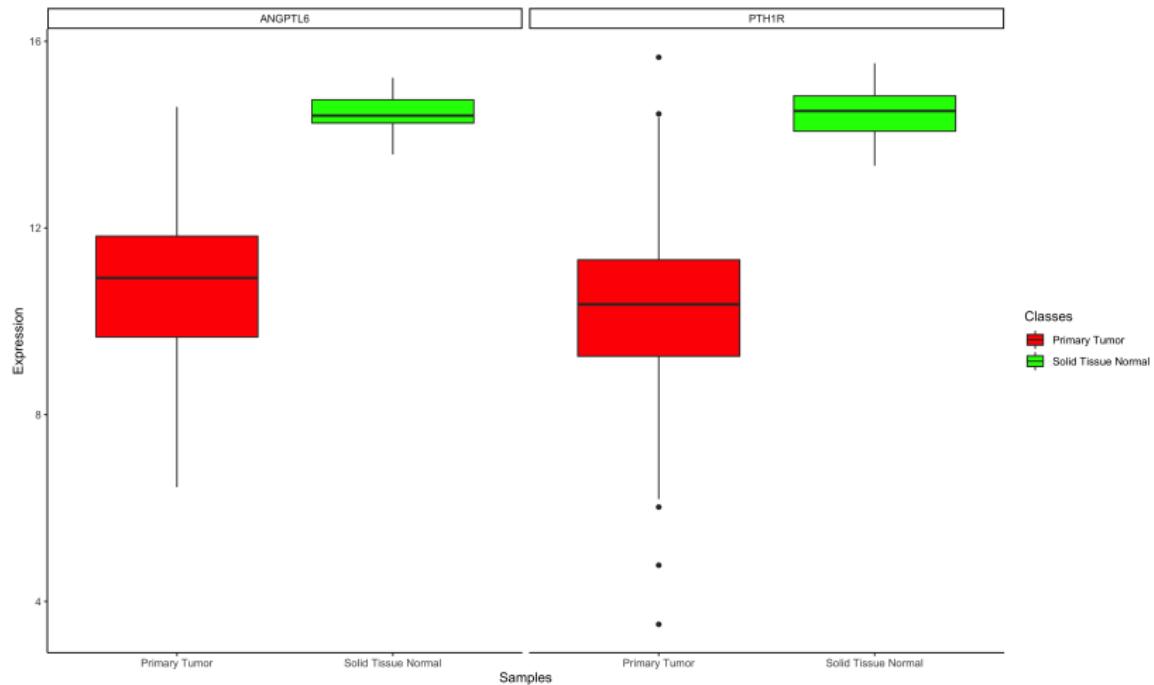


Machine learning

Partición en conjuntos de entrenamiento y test
Reparto 75% - 25% con equilibrio de clases



Machine learning



Machine learning

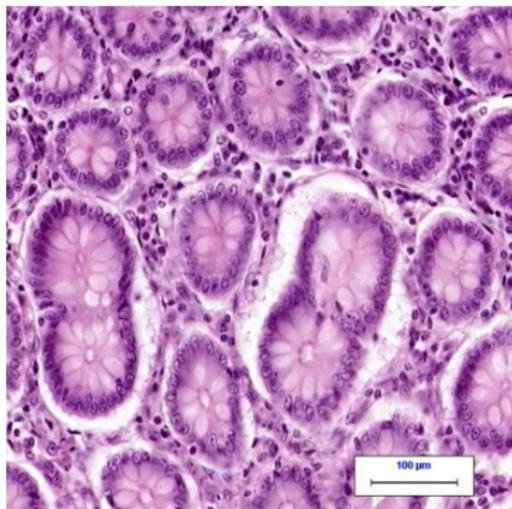
Se entrena un modelo con 2 genes y se aplica al conjunto de test:



Machine learning

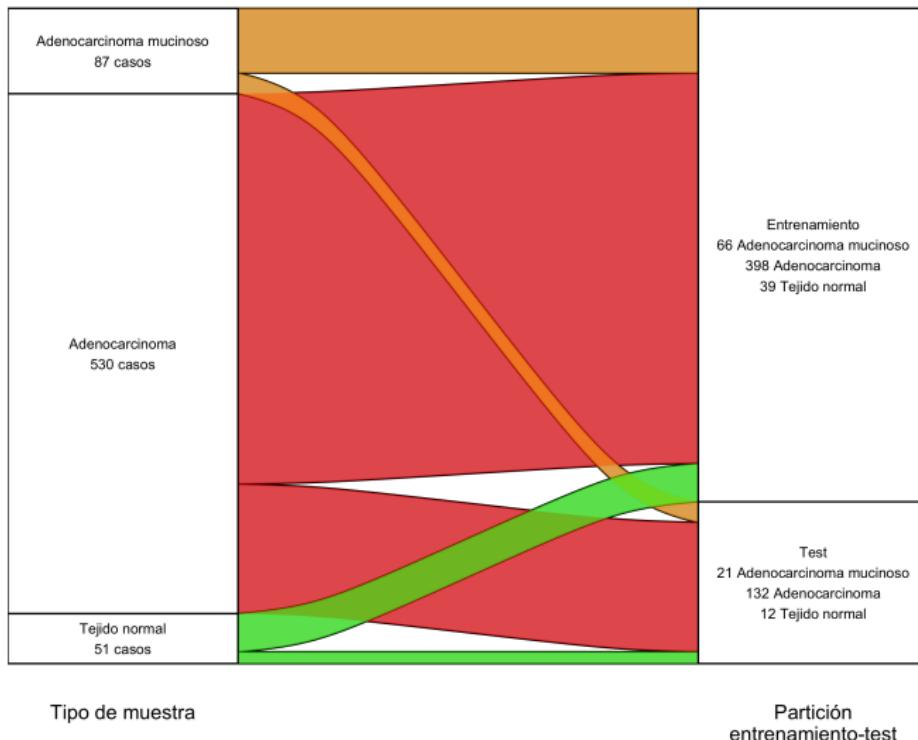
- 668 muestras de tejido de colon-recto.
- 530 identificados como adenocarcinoma, 87 como adenocarcinoma mucinoso.
- 27.000 genes en cada muestra.

Objetivo: buscar genes que permitan “clasificar” una muestra nueva como adenocarcinoma, adenocarcinoma mucinoso o tejido sano.

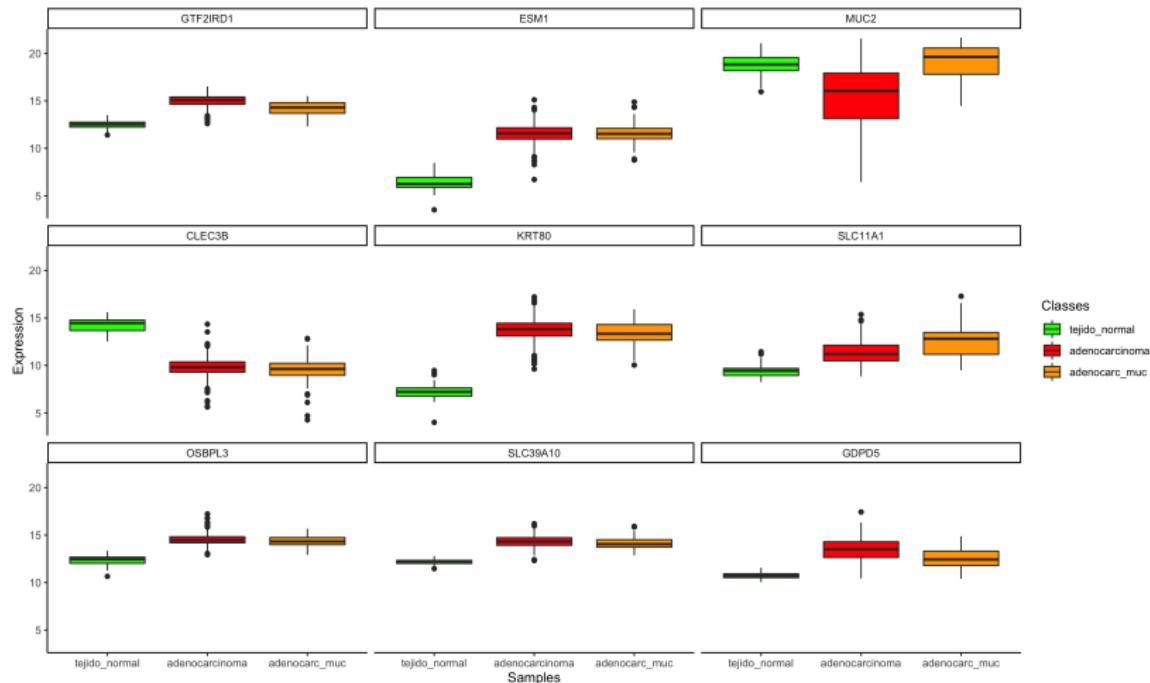


Machine learning

Partición en conjuntos de entrenamiento y test
Reparto 75% - 25% con equilibrio de clases

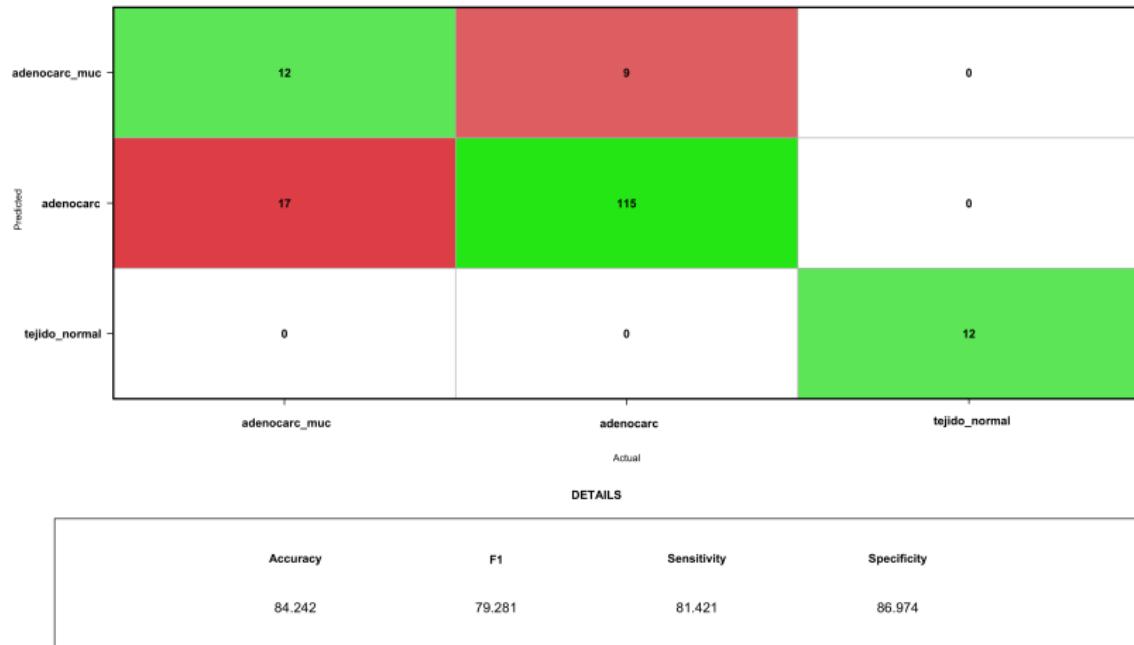


Machine learning



Machine learning

Se entrena un modelo con 9 genes y se aplica al conjunto de test:

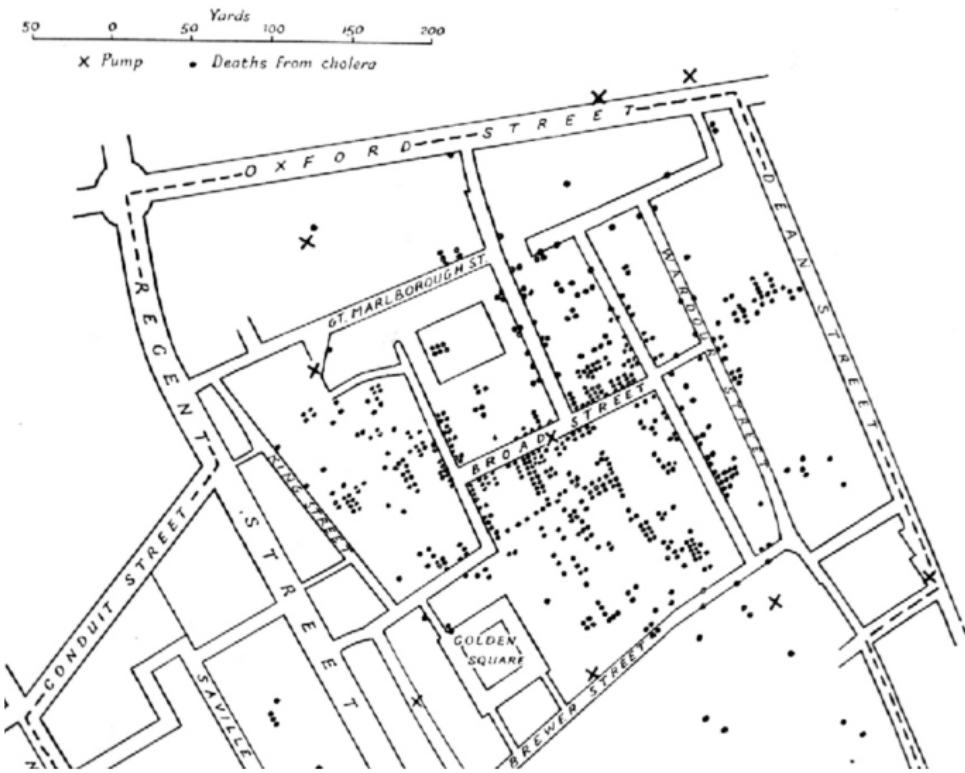


Índice

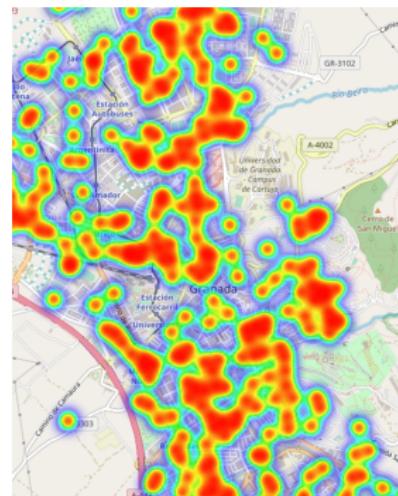
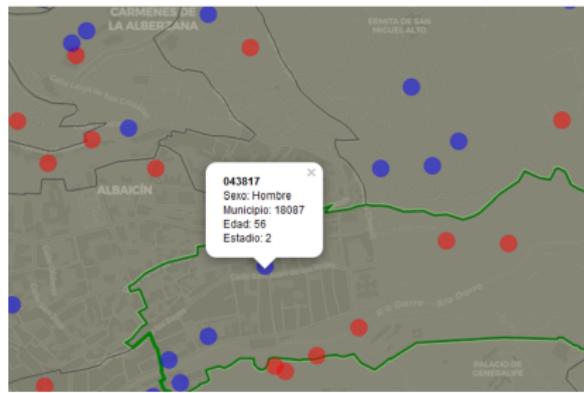
1. Epidemiología y cáncer
2. Herramientas
3. Series temporales
4. Machine learning
5. **Análisis espacial**

Análisis espacial

Estudio de John Snow sobre cólera.

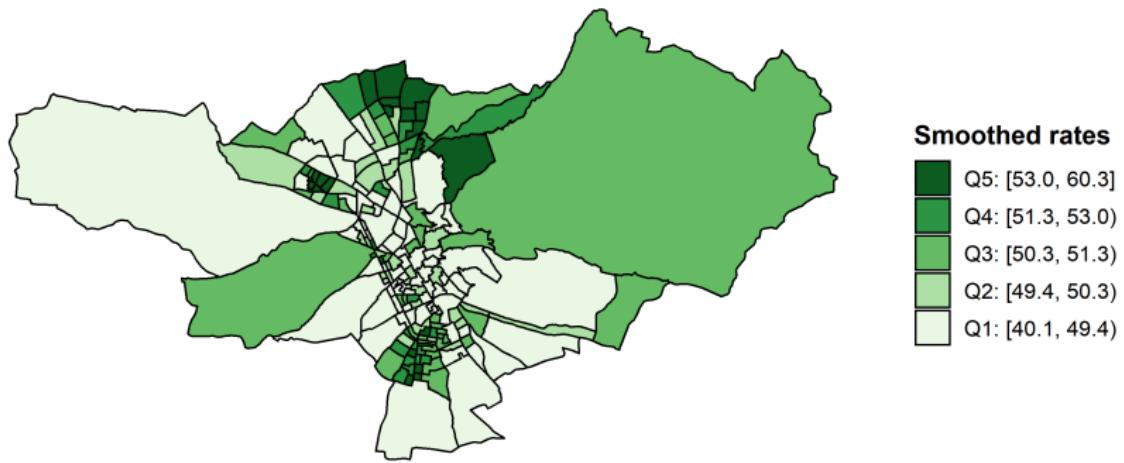


Análisis espacial



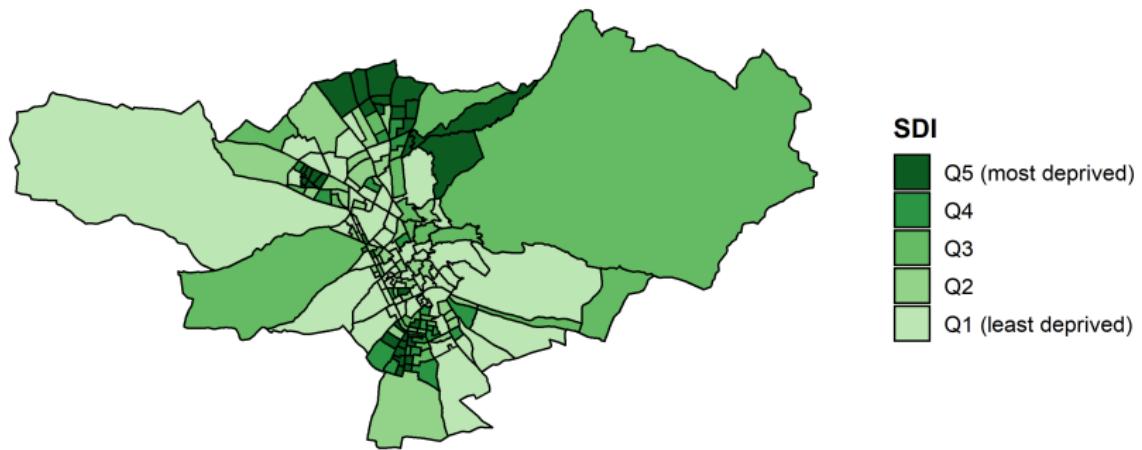
Análisis espacial

Incidencia de cáncer de pulmón en Granada capital.

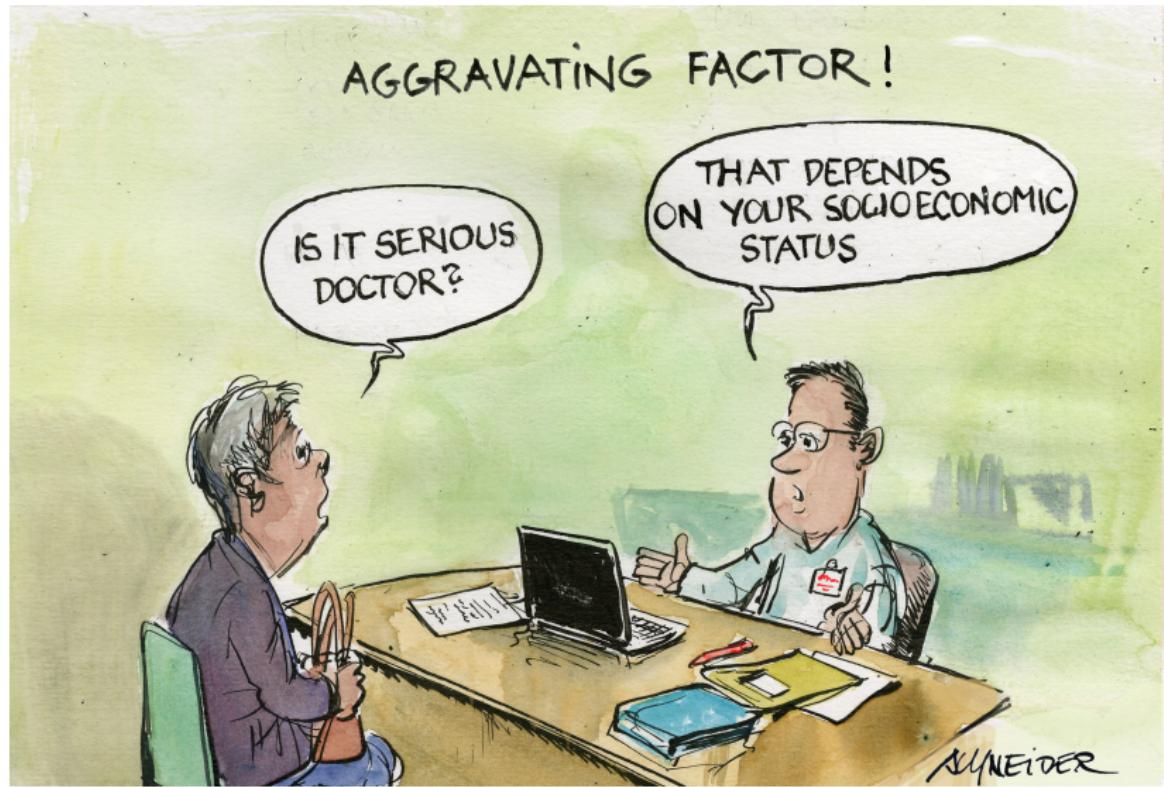


Análisis espacial

Índice de privación en Granada capital.



Análisis espacial



Y mucho más...

- Análisis espacial con **distancia a focos**.
- **Epidemiología ambiental.** Factores de riesgo físicos (calor, ruido, radiaciones...), químicos (plaguicidas, metales...) y biológicos (virus, bacterias...).
- Impacto de **comorbilidades**.
- Estadísticas de cáncer **por municipios**.
- **Retrasos** diagnósticos.
- **Salud mental** en cáncer.
- Efectividad de **programas de cribado**.
- **Evaluación económica** en cáncer.
- **Nutrición** y cáncer.

No todo es cáncer...



The spread of SARS-CoV-2 in Spain: Hygiene habits, sociodemographic profile, mobility patterns and comorbidities

Miguel Rodríguez-Barranco^{a, b, c, 1}, Lorenzo Rivas-García^{d, e, 1}, José L. Quiles^{d, f}, Daniel Redondo-Sánchez^{a, b, c}, Pilar Aranda-Ramírez^{d, e}, Juan Llopis-González^{d, e}, María José Sánchez Pérez^{a, b, c}, Cristina Sánchez-González^{d, e}✉

✉

Highlights

- Living with a COVID-19 patient increased the risk of contagion by 60 times.
- Walking the dog increases the risk of contagion of COVID-19 by 78%.
- The most effective hygiene measure was disinfecting products purchased.
- Working on site at the workplace increased the risk of contagion by 76%.
- Obtaining basic products using home delivery service raised the risk of contagion.

Epidemiología y Prevención de precisión

DEL 3 AL 6 DE SEPTIEMBRE DE 2019

FACULTAD DE MEDICINA Y CIENCIAS DE LA SALUD
OVIEDO 2019



SE SOCIEDAD
ESPAÑOLA DE
EPIDEMIOLOGÍA

SESPAS
Sociedad Española de
Prevención y Salud Pública

PORTUGAL
EPIDEMIOLOGIA

XXXVII REUNIÓN ANUAL DE LA SEE

XIV CONGRESO DA APE

XVIII CONGRESO SESPAS

Consumo de homeopatía en España

Jesús Henares-Montiel ^{1,2}, Dafina Petrova ^{2,3,4}, Daniel Redondo-Sánchez ^{2,3,4}

1. Servicio de Medicina Preventiva y Salud Pública. Hospital Universitario San Cecilio, Granada.

2. Escuela Andaluza de Salud Pública, Granada.

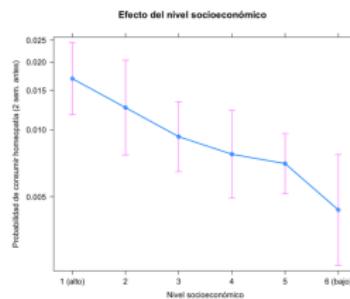
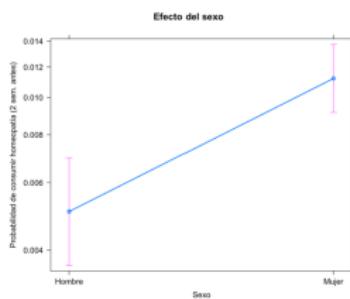
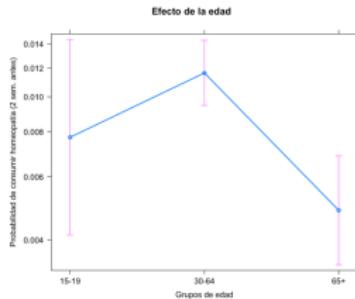
3. Instituto de Investigación Biosanitaria ibs.GRANADA, Granada.

4. CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid.

Resultados



- **15.863** personas (69% del total) habían tomado algún **medicamento** en las últimas dos semanas.
- **154** (0,67% del total) habían tomado algún **producto homeopático** en las últimas dos semanas y **70** de ellas (45%) declararon que el producto homeopático le fue **recetado por un médico**.
- El **consumo de homeopatía** es más común entre adultos de entre **30 y 64 años** (**OR = 1,51**), **mujeres** (**OR = 2,23**) y personas de **muy alto nivel socioeconómico** (**OR = 3,92**), con un gradiente claro.



Descarga de datos del Instituto Nacional de Estadística con R usando el servicio API JSON

Daniel Redondo-Sánchez, Miguel Ángel Luque Fernández, Miguel Rodríguez Barranco, Pablo Fernández-Navarro, María José Sánchez Pérez

Instituto de Investigación Biosanitaria de Granada (ibs.GRANADA),
Universidad de Granada

Registro de Cáncer de Granada, Escuela Andaluza de Salud Pública
CIBER de Epidemiología y Salud Pública



Fortaleza: Descarga masiva de información

Haciendo uso de **156 URLs diferentes** (52 provincias \times 3 años) realizamos una descarga automática de **+6.000.000 de filas** (población por edad simple por secciones censales de los años 2010 , 2011 y 2012).

```
[1] "Se han descargado 17556 filas de Araba/Álava"  
[1] "se ha anexado la información de Araba/Álava"  
[1] "Se han descargado 19866 filas de Albacete"  
[1] "Se ha anexado la información de Albacete"  
[1] "Se han descargado 80982 filas de Alicante/Alacant"  
[1] "se ha anexado la información de Alicante/Alacant"  
[1] "Se han descargado 27984 filas de Almería"  
[1] "Se ha anexado la información de Almería"  
[1] "Se han descargado 20724 filas de Ávila"  
[1] "Se ha anexado la información de Ávila"  
[1] "Se han descargado 36498 filas de Badajoz"  
[1] "se ha anexado la información de Badajoz"
```



INE_R



Tutorial para descargar datos del Instituto Nacional de Estadística (INE) con R.



HTML



7



2

https://github.com/danielredondo/INE_R

Mensajes clave

- **Programar** es muy importante y ayuda a **resolver problemas** y **automatizar procesos**.
- Es difícil encontrar **personal de estadística especializado en investigación biomédica**. Se valora especialmente el grado de doctor.
- **La investigación biosanitaria necesita estadísticos y estadísticas.**



Daniel Redondo Sánchez

Contacto

```
email      <- "daniel.redondo.easp@juntadeandalucia.es"  
web        <- "danielredondo.com"  
github     <- "github.com/danielredondo"  
twitter    <- "@dredondosanchez"  
telegram   <- "@danielredondo"
```