

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

 registro de cáncer de granada

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


aecc
Contra el Cáncer



¿Quién soy?

Daniel Redondo Sánchez



- Licenciado en **Matemáticas** (UGR 2009-2014).
- Diploma de Especialización en **Epidemiología e Investigación Clínica** (UGR, EASP, 2016).
- Máster de **Matemáticas** (UAL, UCA, UGR, UMA, UJA, 2019).
- Máster en **Ciencia de Datos** e Ingeniería de Computadores (UGR, 2020).
- 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).



registro de cáncer de granada

www.registrationcancergranada.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



+



Herramientas





Software específico de Registros de Cáncer:

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

Trabajo en equipo

Inglés

Estadística pública: poblaciones, defunciones, censos, encuestas...

Visualización de datos

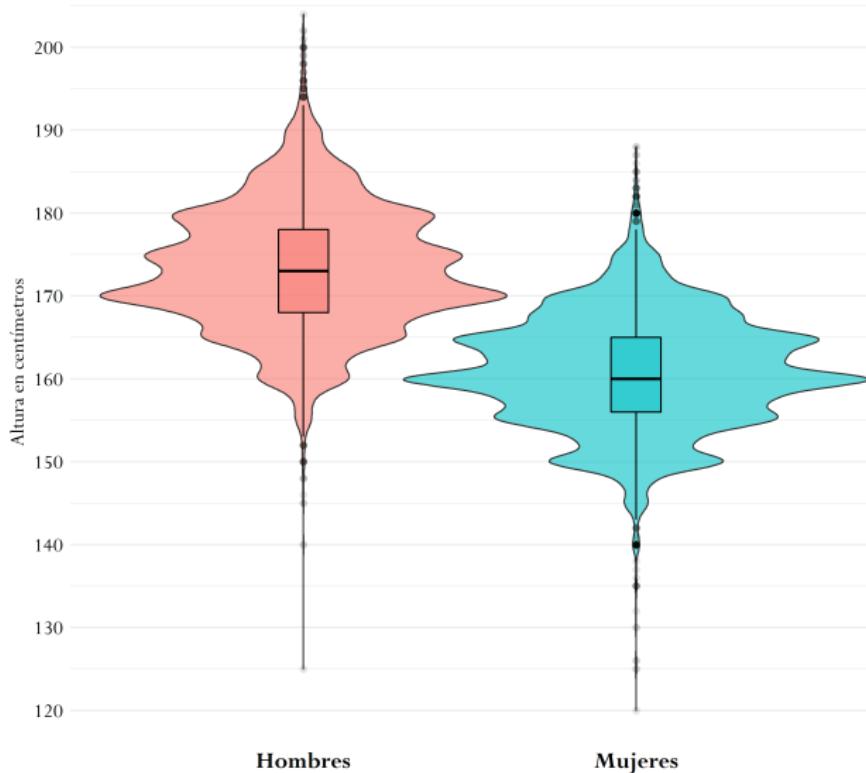
- Un gráfico vale más que mil palabras.
- Deben ser **claros** y **precisos**.
- Hay que saber **interpretar** y **crear** gráficos.
- A veces son convenientes **gráficos interactivos**:
<https://www.danielredondo.com/grafico2>
- Es uno de los puntos fuertes de R con el paquete {ggplot2}.



Herramientas

Altura por sexos en España, 2017

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

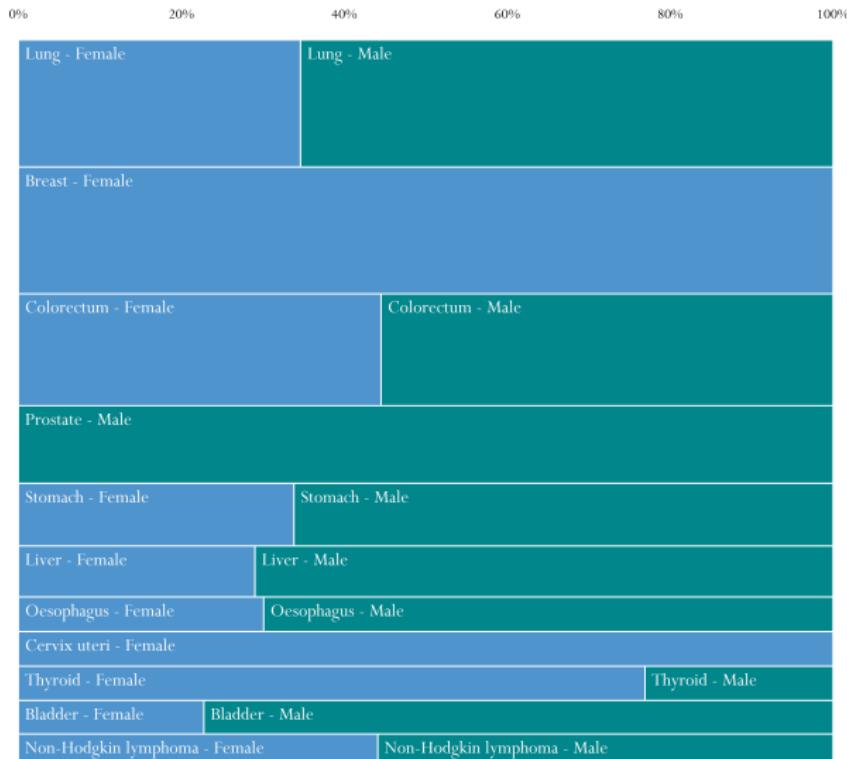


Herramientas

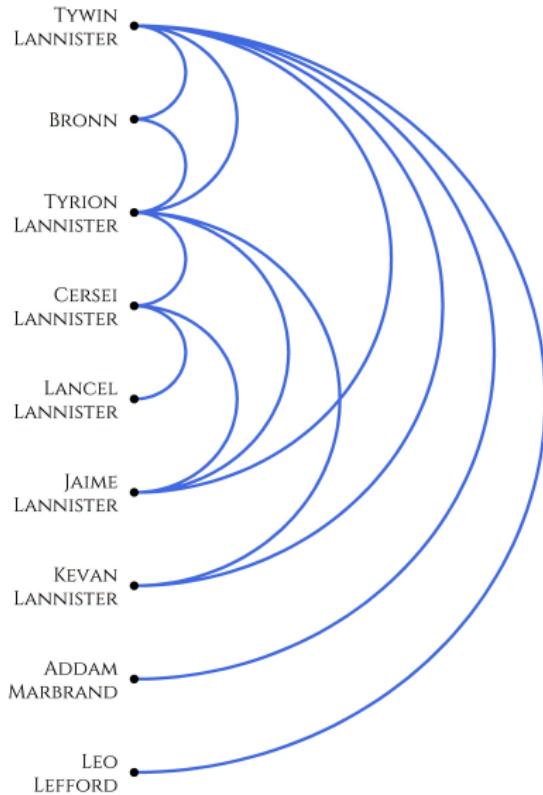
Cancer incidence, 2018

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

Source: Global Cancer Observatory (World Health Organization).

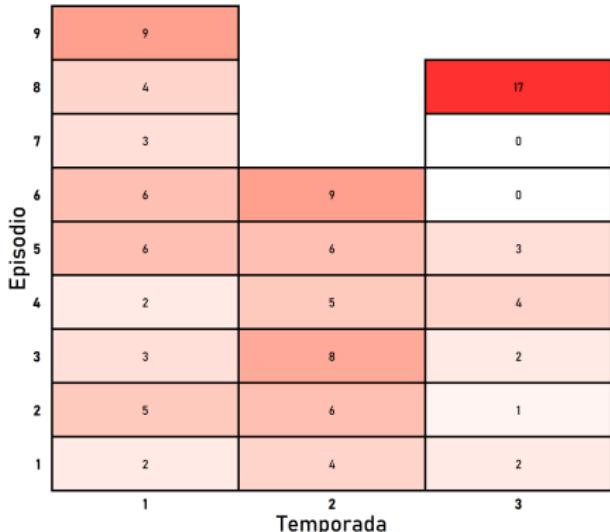


Herramientas

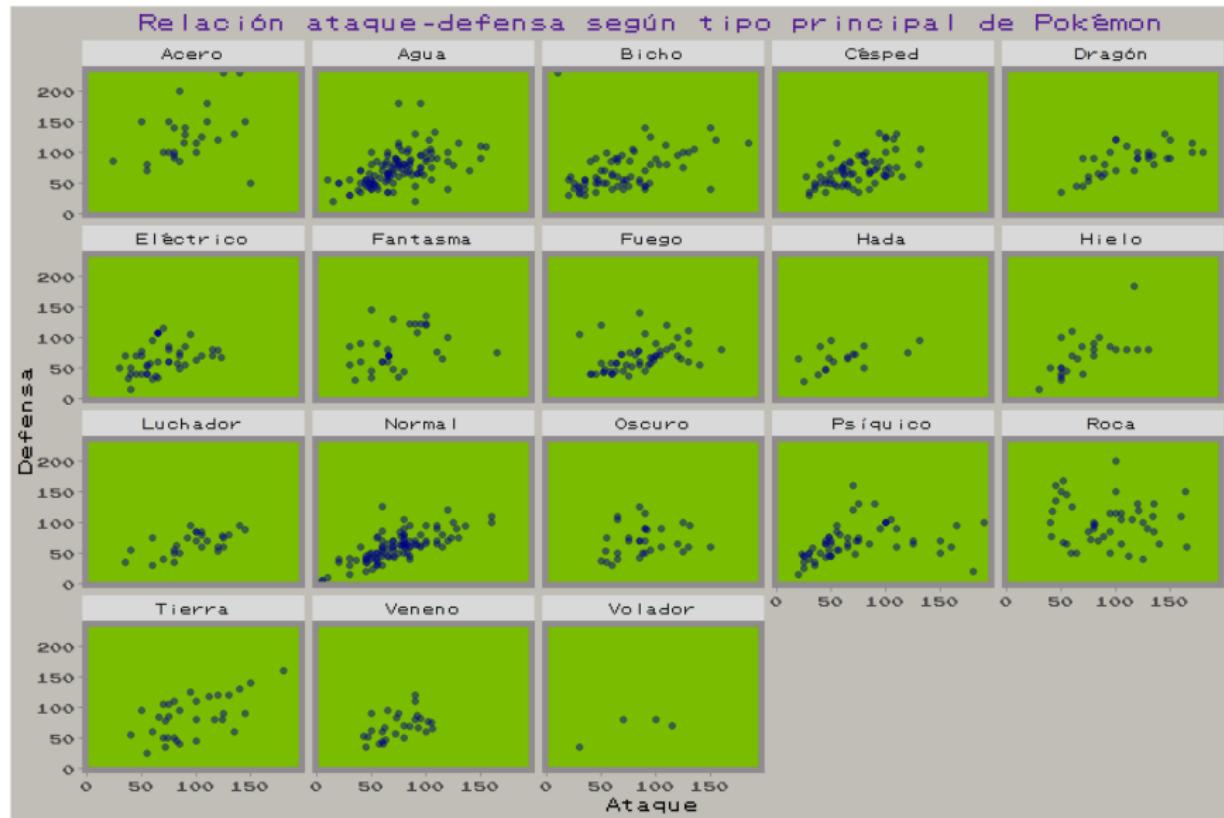


LA CASA DE PAPEL: NAIROBI

Número de veces que se menciona "Nairobi" por capítulo.
Gráficos por @dredondosanchez



Herramientas



Índice

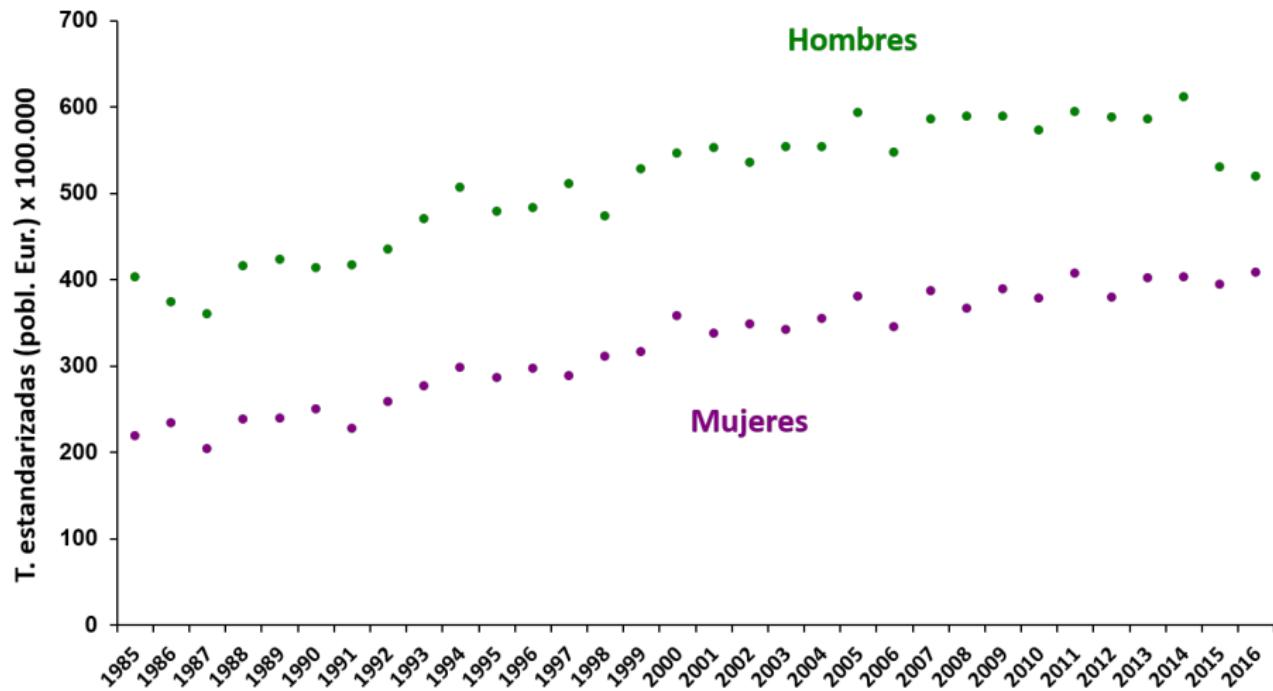
1. Epidemiología y cáncer
2. Herramientas
3. **Series temporales**
4. Machine learning
5. Análisis espacial

Series temporales

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

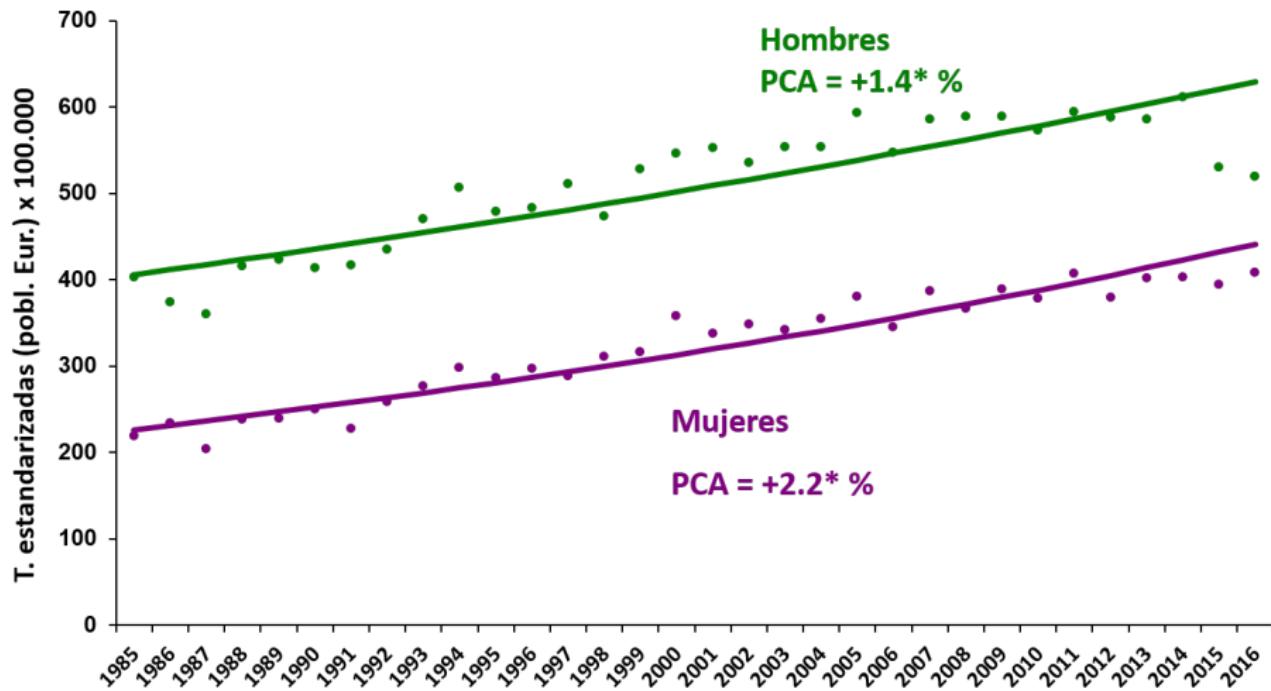
Series temporales - Tendencias

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



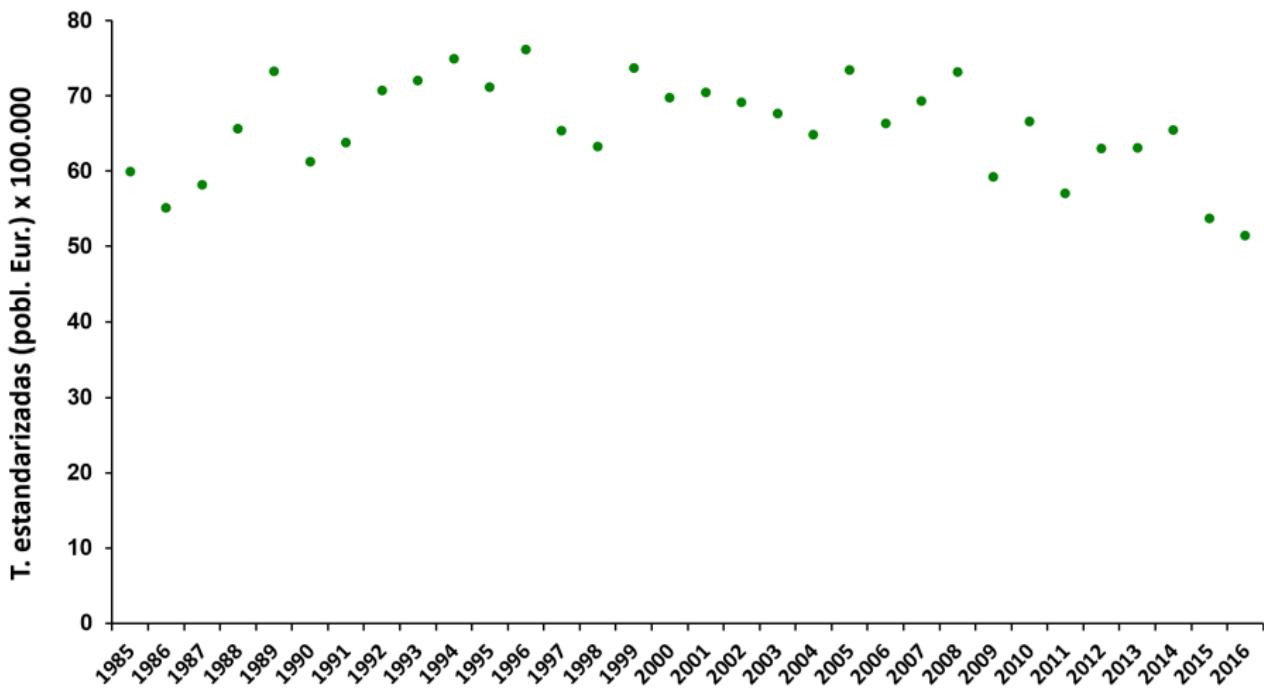
Series temporales - Tendencias

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



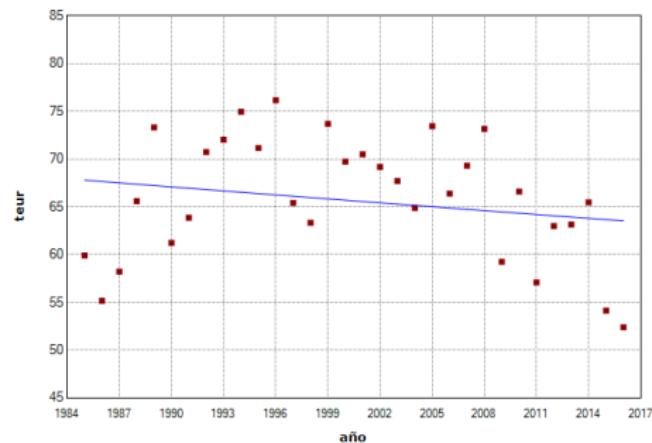
Series temporales - Tendencias

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

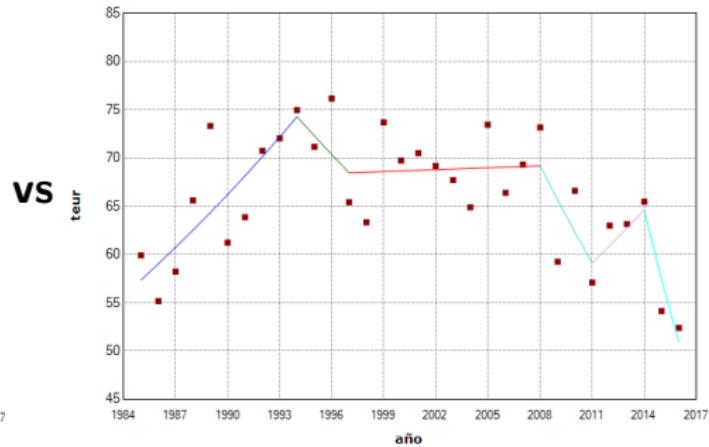


Series temporales - Tendencias

0 puntos de inflexión

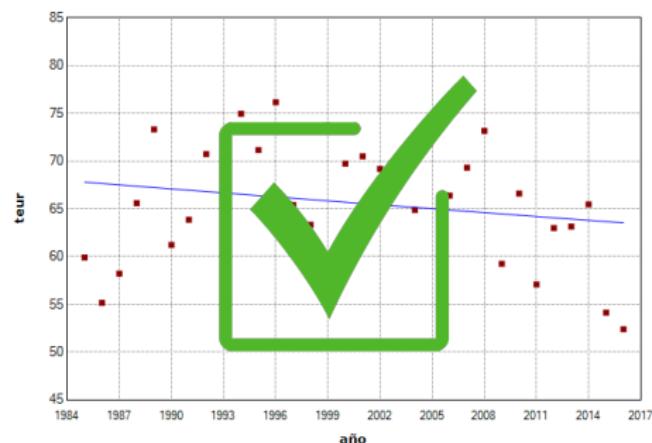


5 puntos de inflexión

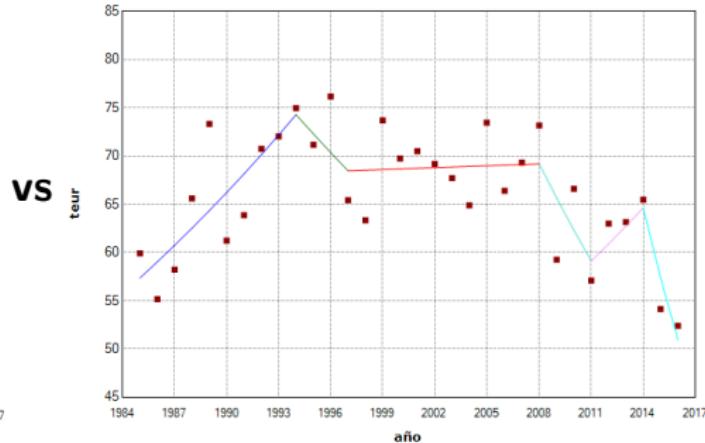


Series temporales - Tendencias

0 puntos de inflexión

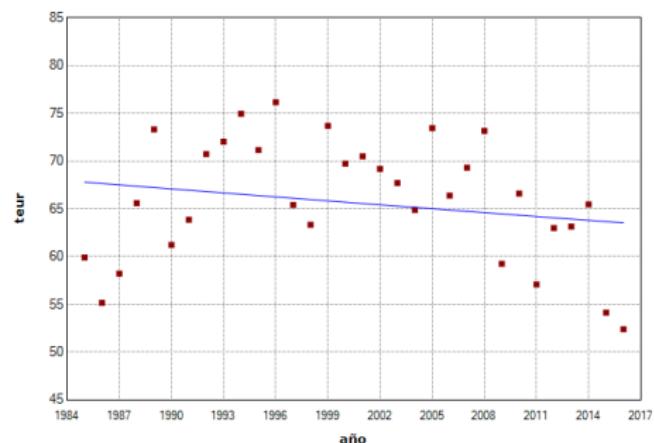


5 puntos de inflexión

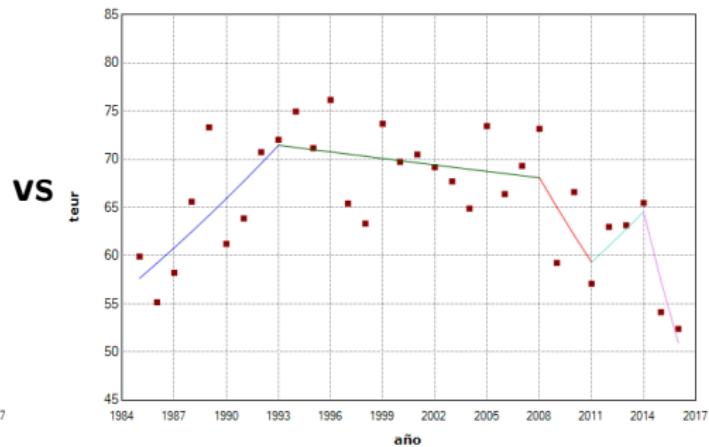


Series temporales - Tendencias

0 puntos de inflexión

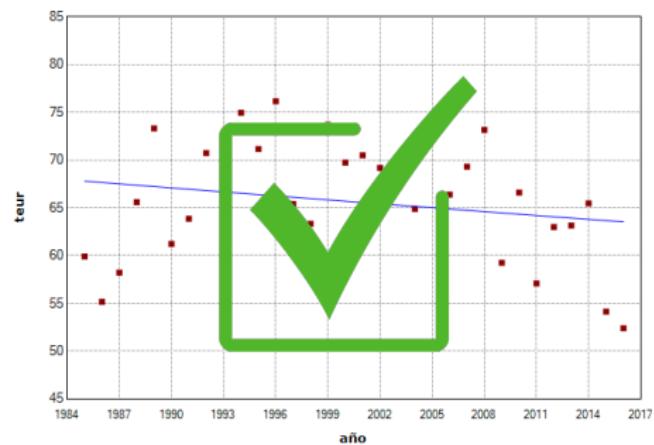


4 puntos de inflexión

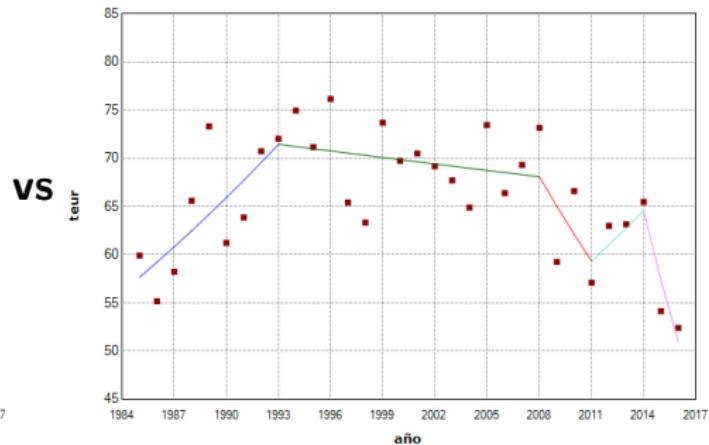


Series temporales - Tendencias

0 puntos de inflexión

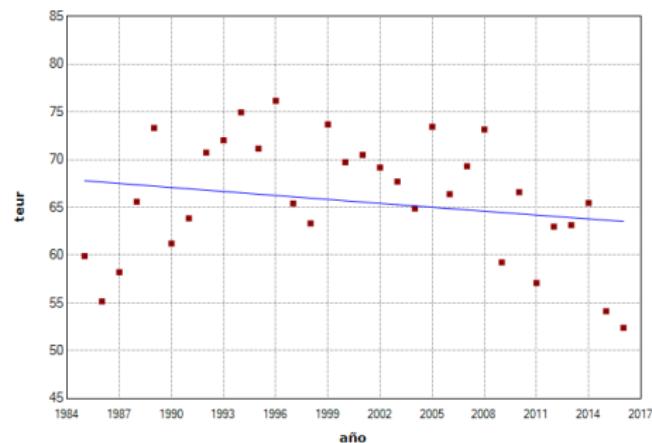


4 puntos de inflexión

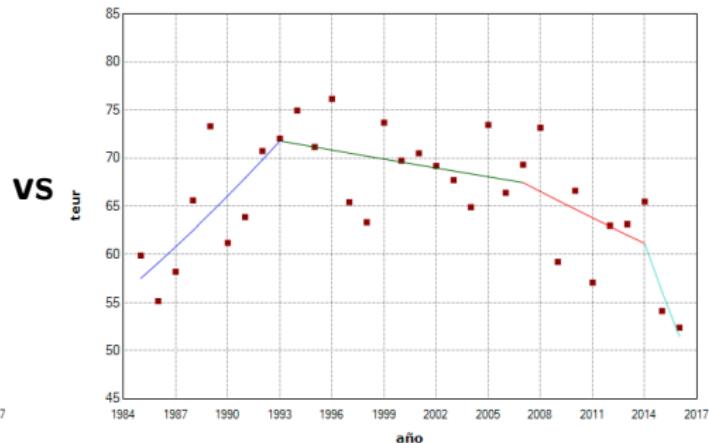


Series temporales - Tendencias

0 puntos de inflexión

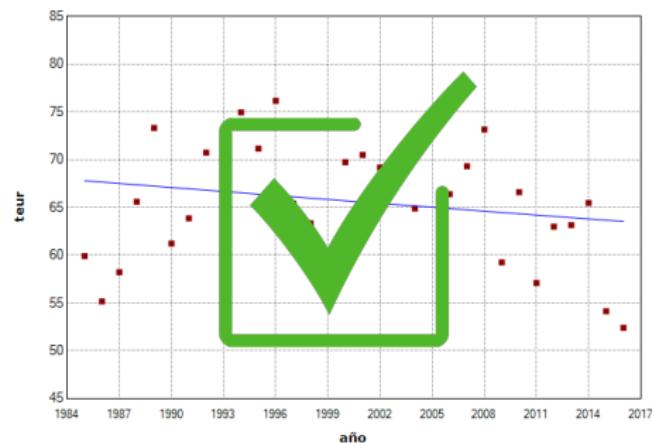


3 puntos de inflexión

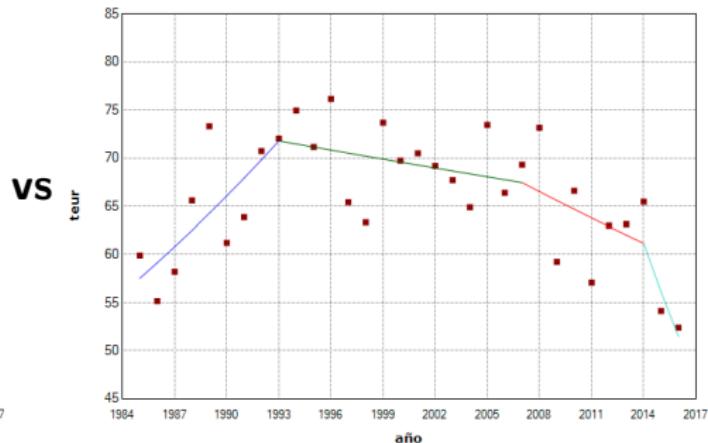


Series temporales - Tendencias

0 puntos de inflexión

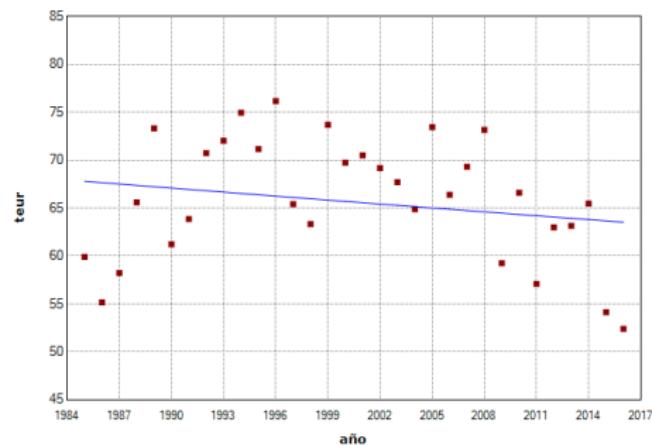


3 puntos de inflexión

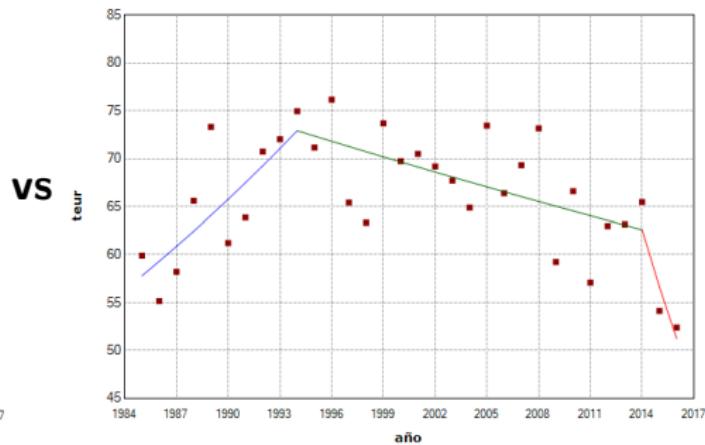


Series temporales - Tendencias

0 puntos de inflexión



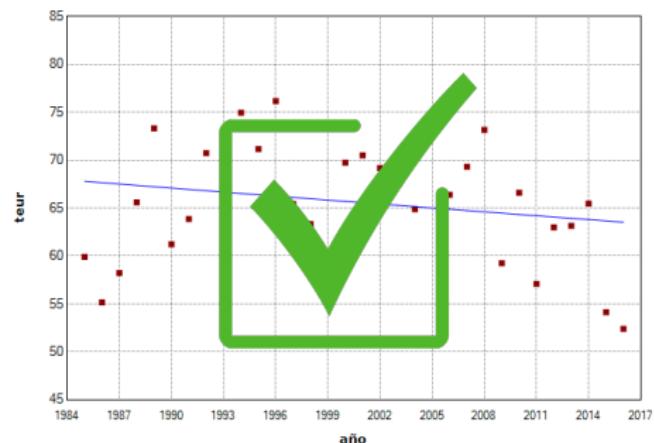
2 puntos de inflexión



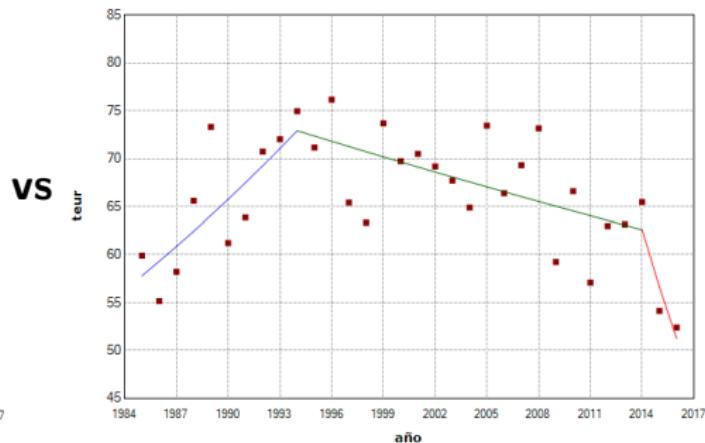
vs

Series temporales - Tendencias

0 puntos de inflexión

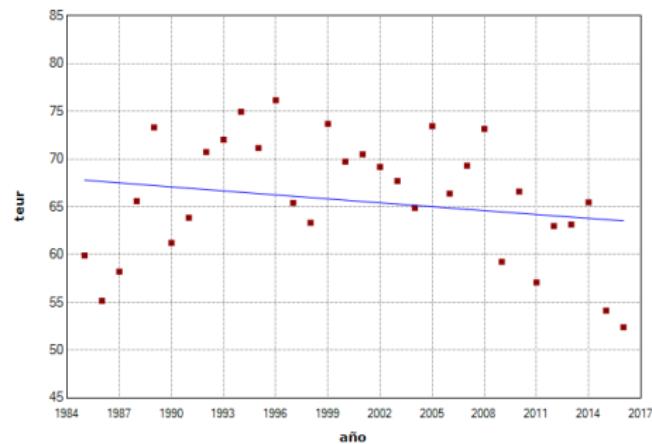


2 puntos de inflexión

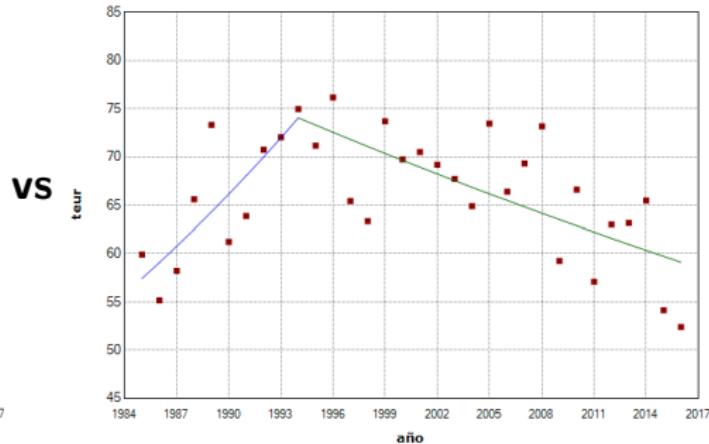


Series temporales - Tendencias

0 puntos de inflexión

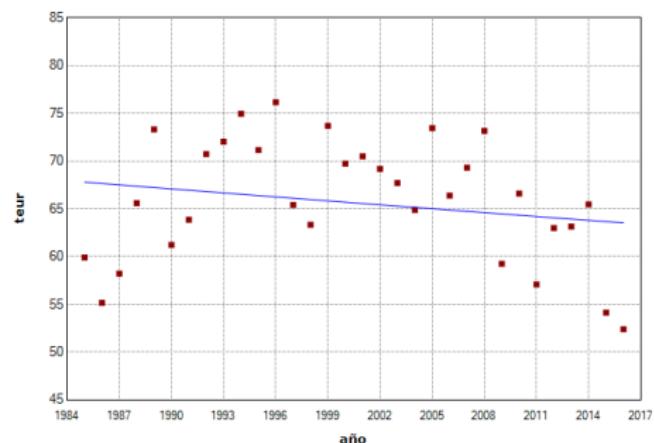


1 punto de inflexión

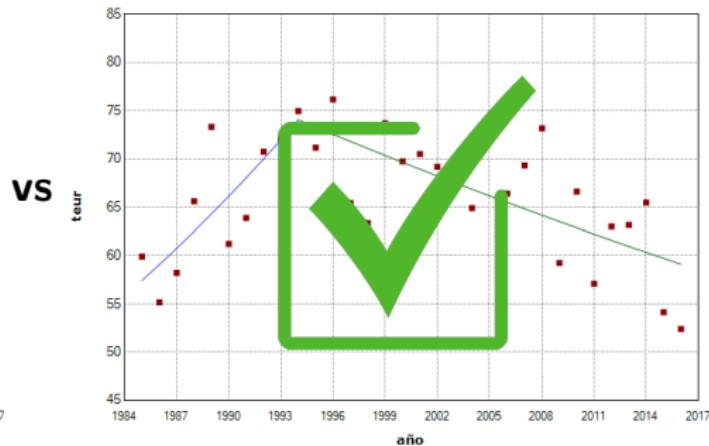


Series temporales - Tendencias

0 puntos de inflexión



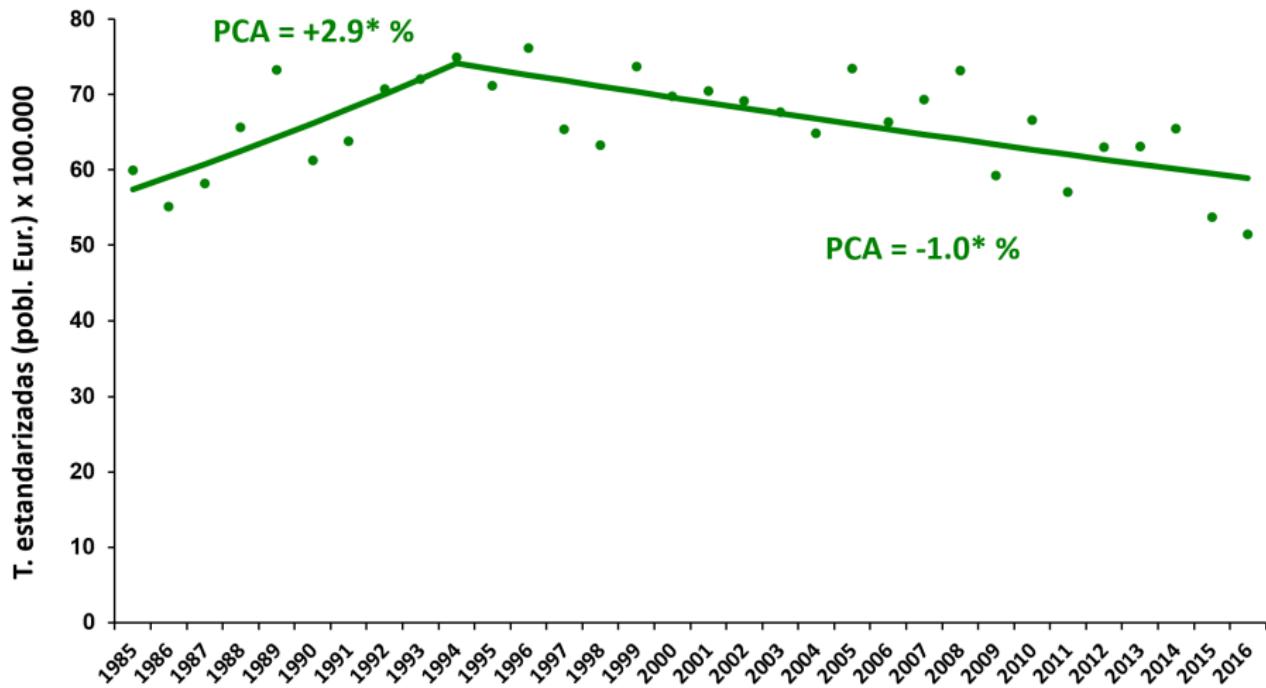
1 punto de inflexión



vs

Series temporales - Tendencias

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

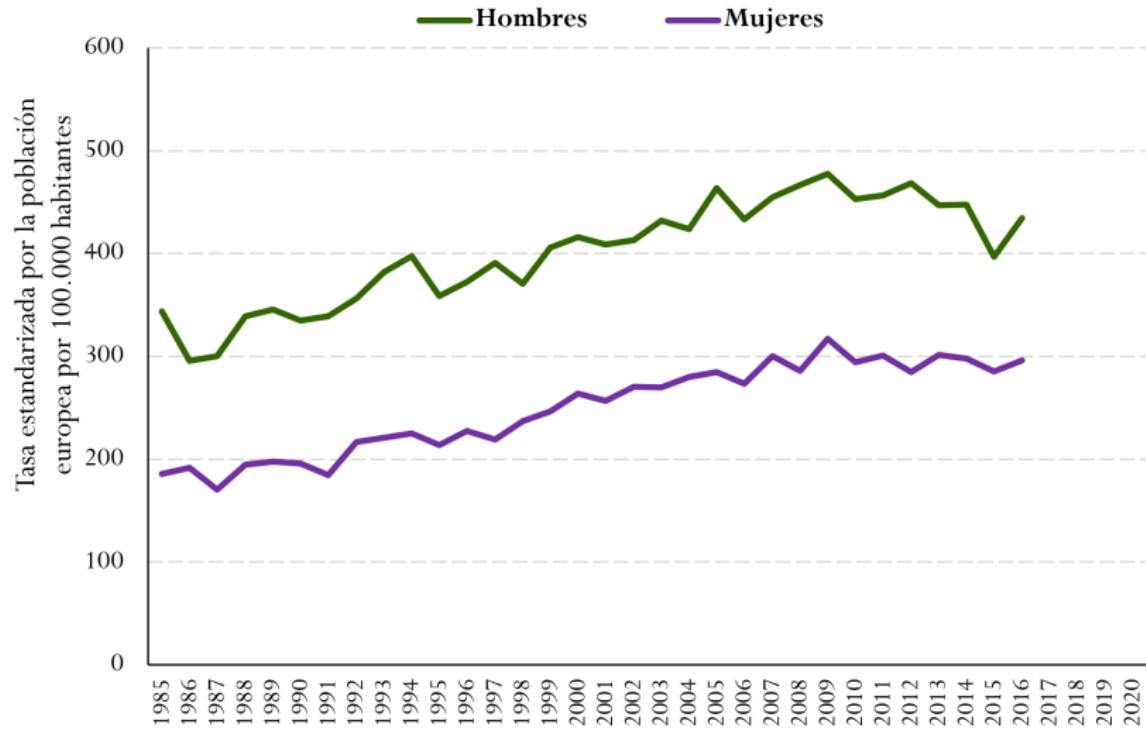


Series temporales

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

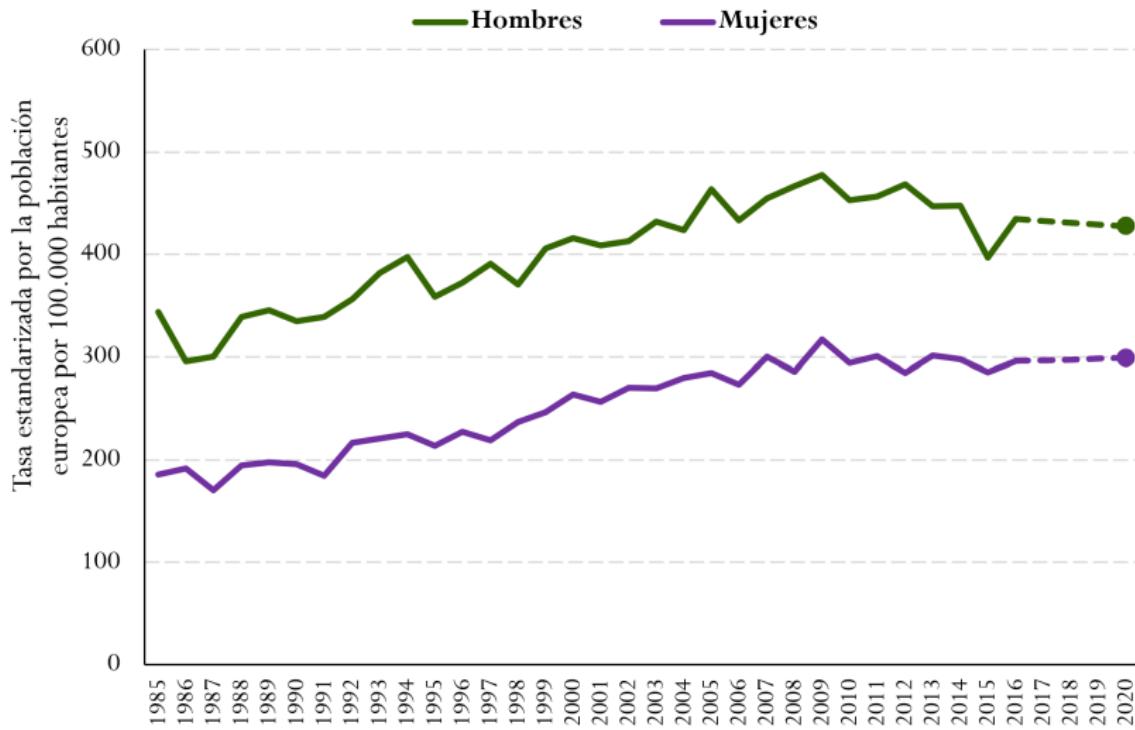
Series temporales - Proyecciones

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



Series temporales - Proyecciones

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



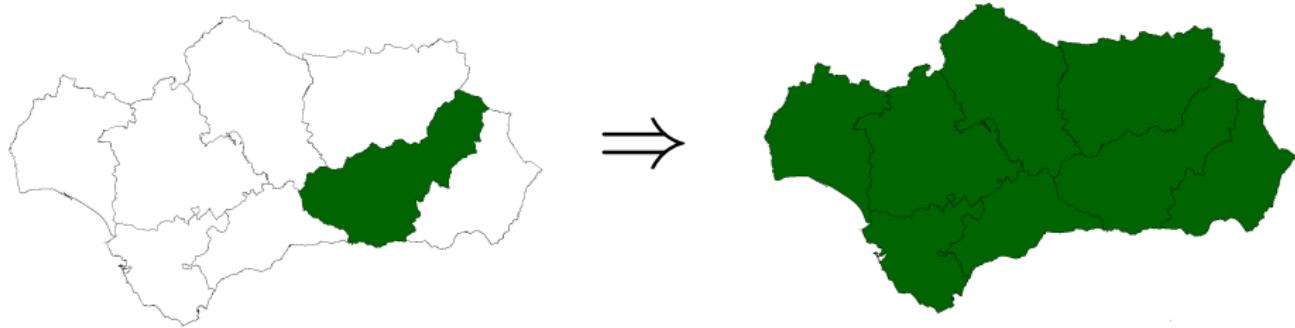
Series temporales

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 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.

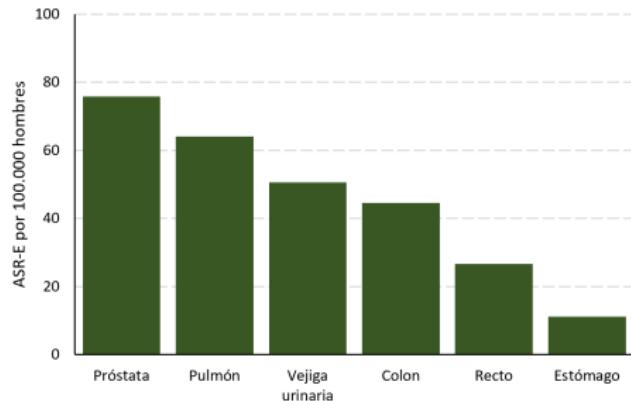
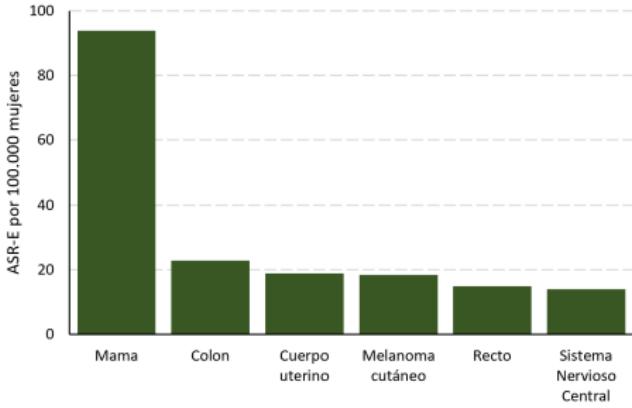


Gráfico 3. Estimaciones de incidencia de cáncer en Andalucía en mujeres.

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.

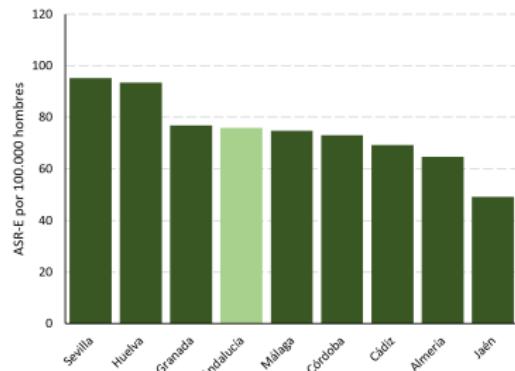
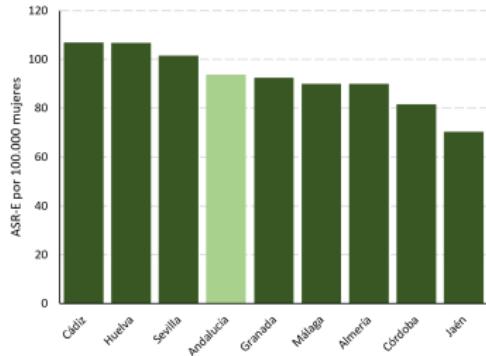


Gráfico 7. Estimaciones de incidencia del cáncer de mama en mujeres en Andalucía por provincias.

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}

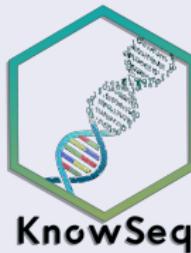
Índice

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**.
- Un ejemplo con R+{KnowSeq}...

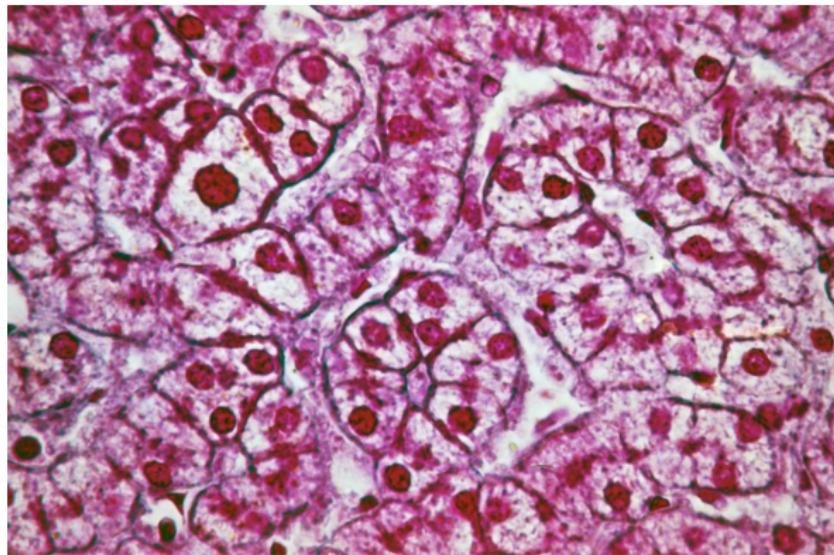


<https://github.com/CasedUgr/KnowSeq>

Machine learning

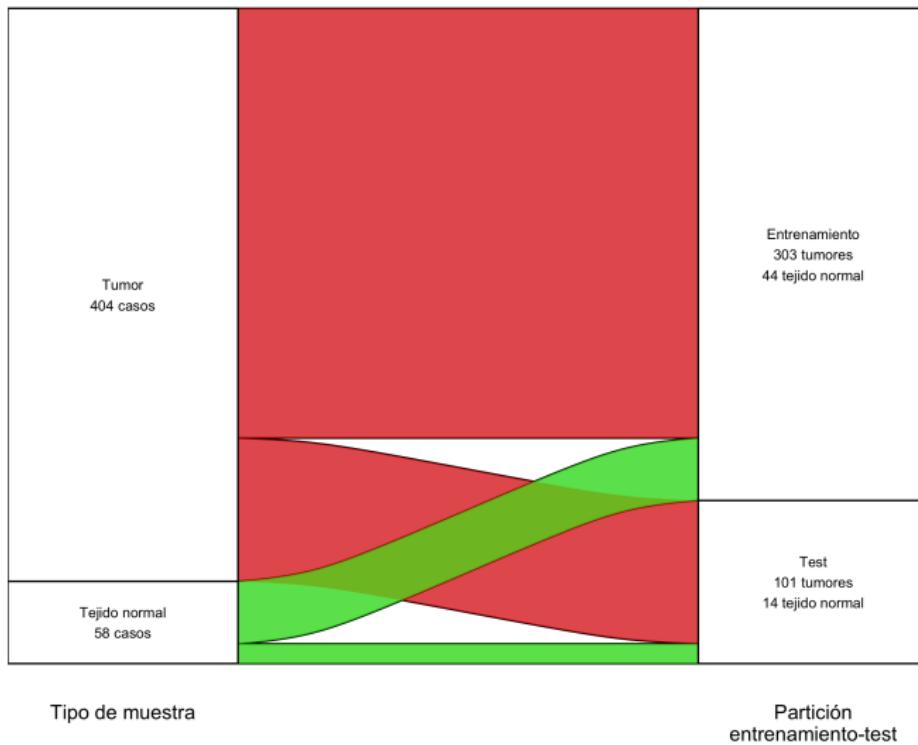
- 462 muestras de tejido de hígado.
- 404 identificados como tumores, 58 como tejido sano.
- +24.500 genes en cada muestra.

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



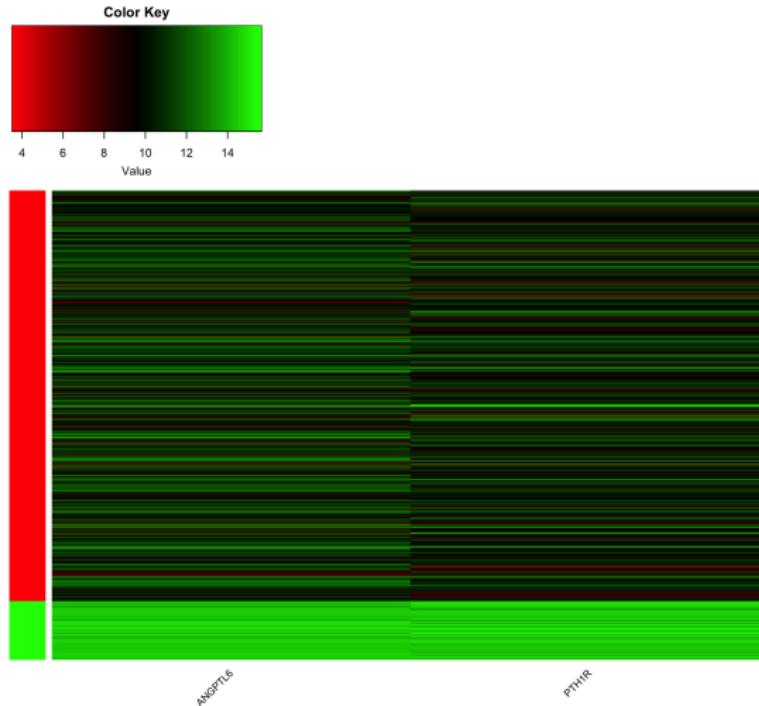
Machine learning

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

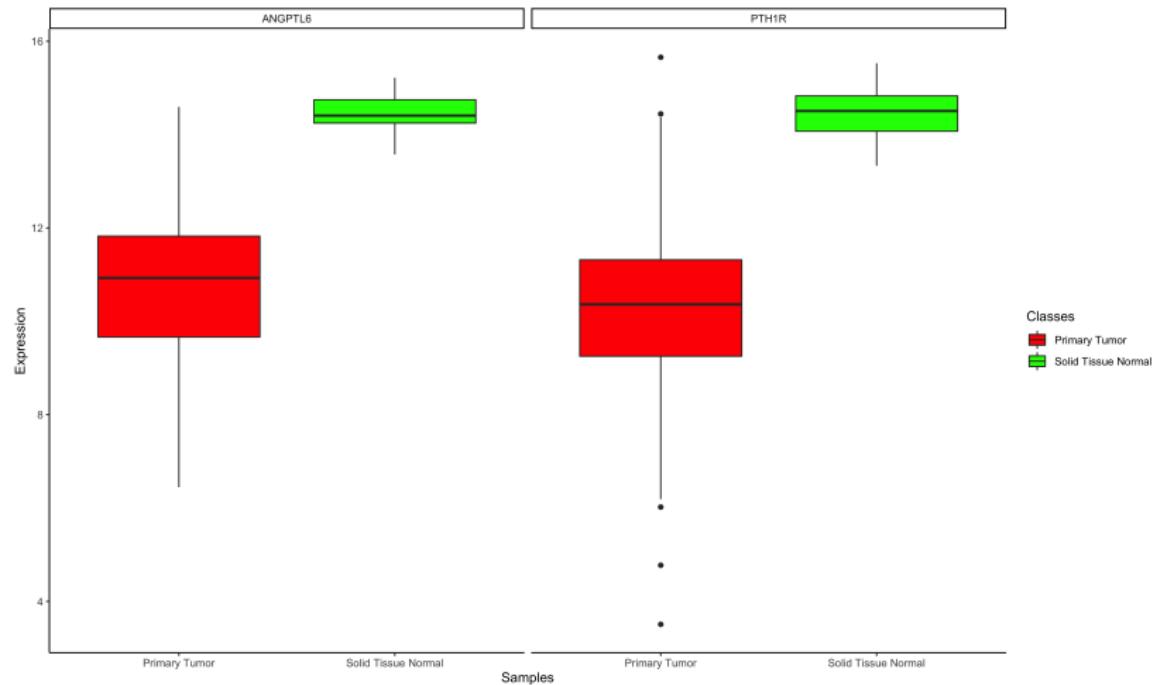


Machine learning

Con el conjunto de entrenamiento:

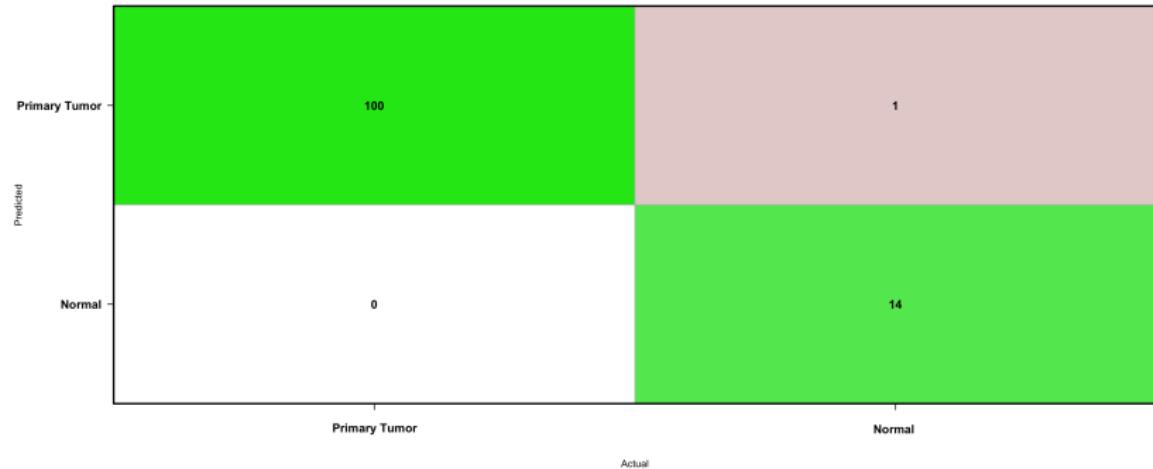


Machine learning



Machine learning

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



Índice

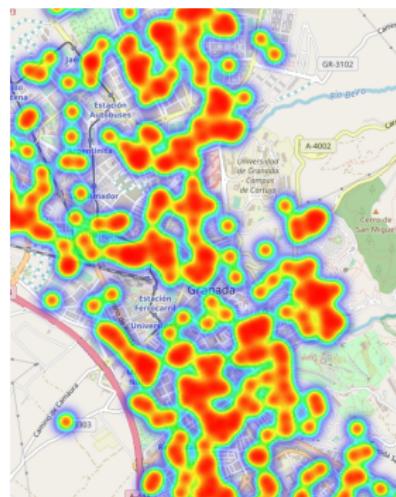
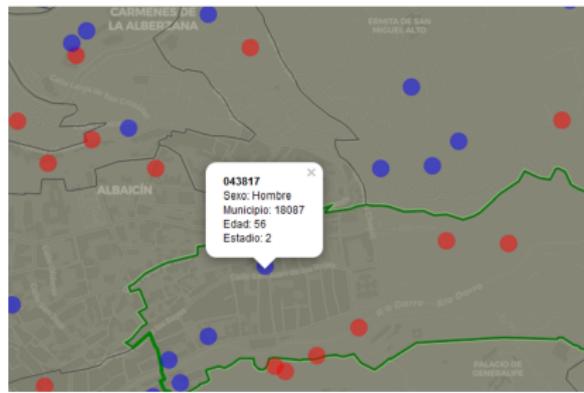
1. Epidemiología y cáncer
2. Herramientas
3. Series temporales
4. Machine learning
5. **Análisis espacial**

Análisis espacial

Estudio de John Snow sobre cólera.

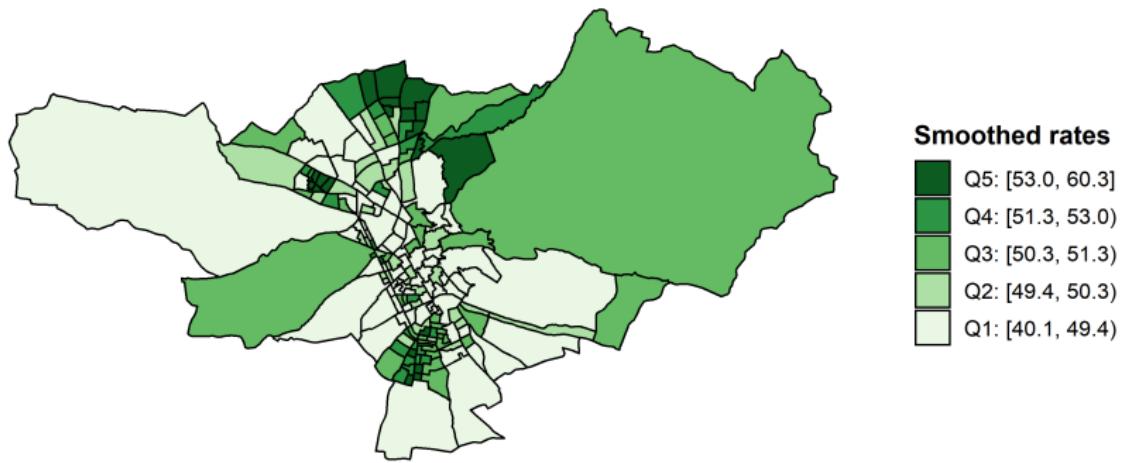


Análisis espacial



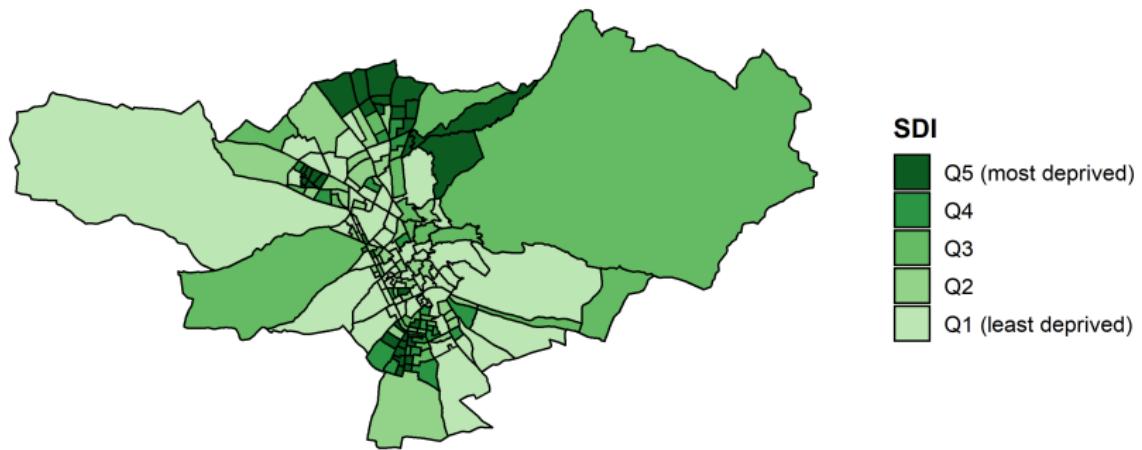
Análisis espacial

Incidencia de cáncer de pulmón en Granada capital.



Análisis espacial

Índice de privación en Granada capital.



Y mucho más...

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

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"
```