

Medio ambiente y salud pública

Métodos de análisis de datos complejos

Dariya Ordanovich

25-05-2018



This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

ESR9: Mapping Health Resources in Spain

- Catálogos web de diferentes recursos de información



Catálogo de Recursos Sanitarios en España



Catalogue of healthcare resources in Spain

Contextual data

Regional healthcare portals

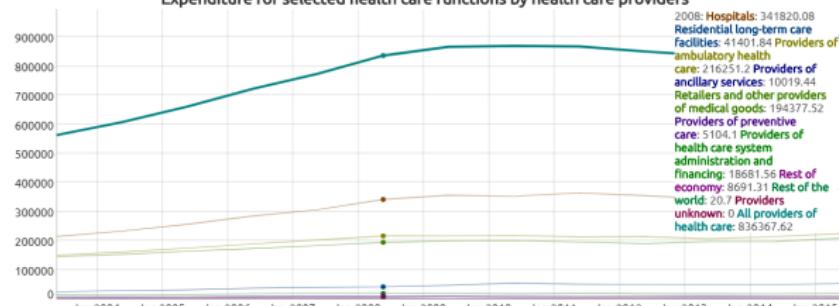
Plot

Table

Plot time-series by:

iche11_hp

Expenditure for selected health care functions by health care providers



Plot

Table

International providers

Eurostat

World Bank

World Health organization

Organisation for Economic Co-operation and Development (OECD)

Search pattern

health

Datasets available

Expenditure for selected health care functions by health care providers

National providers

https://longpop.shinyapps.io/catalogue_of_healthcare_resources

Catálogo de Registros Longitudinales de Población

Catalogue of Longitudinal Registers

Keywords:
Epidemiology Genealogy
[Reset keywords](#)

Date Range:
1820 2017
[Reset dates](#)

Continent:
Europe
[Reset continents](#)

Countries:
Switzerland
[Reset countries](#)

Databases available
Geneva Demographic Database
[Download metadata](#)

TIME COVERAGE
from 1800 to 1880

SAMPLE SIZE
35,592

WEBLINKS
[More info](#)
[Link to source](#)

Geneva Demographic Database



All countries involved in Geneva Demographic Database



Leaflet (Imagery provided by services from the Global Imagery Service (GIBS), operated by the NASA/GSFC Earth Science Data and Information System (ESDIS) with funding provided by NASA HQ.

Country	Database	Abbreviation	Summary
A	All	All	
Switzerland	Female Demographic Biographies	FDB-WALD	All baptisms, marriages and burials from 1880 to 1938; reconstructing female demographic life-courses (own baptism, baptisms of own children, own marriage, own burial) by record linkage, complete data entry, church books

https://longpop.shinyapps.io/catalogue_of_longitudinal_registers

Objetivos personales

- Catálogos web de diferentes recursos de información



Objetivos personales

- Catálogos web de diferentes recursos de información
- **Programa doctoral** en el Departamento de Geografía Humana de la Universidad Complutense de Madrid y en el CSIC



Objetivos personales

- Catálogos web de diferentes recursos de información
- Programa doctoral en el Departamento de Geografía Humana de la Universidad Complutense de Madrid y en el CSIC
 - Análisis espacial del impacto de las olas de calor en la salud en Andalucía



Resumen

- Líneas de investigación:

Resumen

- Líneas de investigación:
 - Relación entre clima, contaminación y mortalidad;

Resumen

- Líneas de investigación:
 - Relación entre clima, contaminación y mortalidad;
 - Accesibilidad a los recursos sanitarios, configuración de la red de transporte y salud;

Resumen

- Líneas de investigación:
 - Relación entre clima, contaminación y mortalidad;
 - Accesibilidad a los recursos sanitarios, configuración de la red de transporte y salud;
 - Olas de color en redes sociales - ¿predictor alternativo?

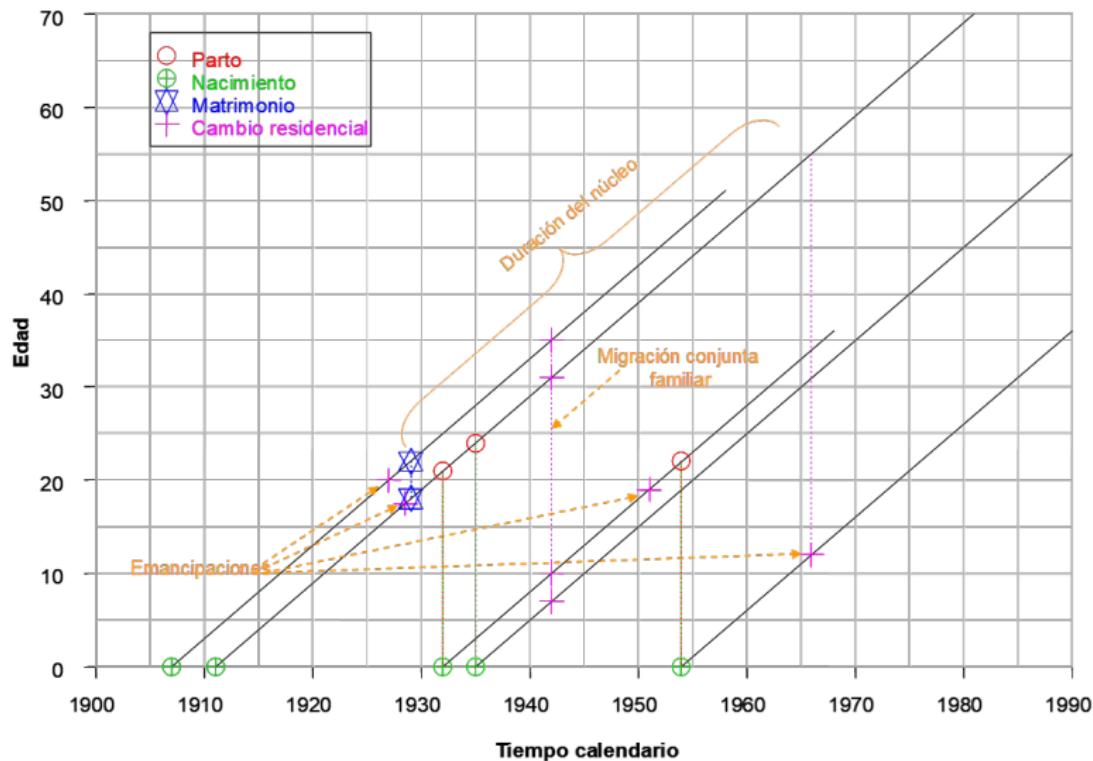
Resumen

- Líneas de investigación:
 - Relación entre clima, contaminación y mortalidad;
 - Accesibilidad a los recursos sanitarios, configuración de la red de transporte y salud;
 - Olas de color en redes sociales - ¿predictor alternativo?
- Colaboración con el Instituto de Estadística y Cartografía de Andalucía:



Instituto de Estadística y Cartografía de Andalucía
CONSEJERÍA DE ECONOMÍA Y CONOCIMIENTO

Registro Longitudinal de Población



Registro Longitudinal de Población



Instituto de Estadística y Cartografía de Andalucía
CONSEJERÍA DE ECONOMÍA Y CONOCIMIENTO

Calendario
Contacto
English



Está en: Inicio - Estadística - Población

Instituto | Estadística | Georreferenciación | Cartografía | Servicios

¿Qué información está buscando?

Buscar

Estadísticas Longitudinales de Supervivencia y Longevidad en Andalucía, 2002-2013 suministra los datos necesarios para la investigación de las funciones de supervivencia poblacional en el periodo 2002-2013, partiendo de la información recogida en el Censo de población y viviendas de 2001 y complementada con la proporcionada por los boletines estadísticos de defunción y las variaciones residenciales padronales, integrada en la Base de Datos Longitudinal de Población de Andalucía (BDLPA) para el seguimiento de la población censada en 2001 durante los doce años posteriores. Esta publicación incluye tabulaciones con causa de muerte así como mapas de indicadores de mortalidad utilizando celdas de 250 m para mortalidad general y otras de 1 km para mortalidad por causa.



UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional

- Organismo: Instituto de Estadística y Cartografía de Andalucía.
- Ámbito: Comunidad Autónoma, Unidad Territorial, Provincia, Municipio y Secciones Estadísticas.

Estadísticas Longitudinales de Supervivencia y Longevidad en Andalucía, 2002-2013



- Nota divulgativa

DATOS

- Tabulación detallada. Serie 2002-2013
 - Resultados básicos
 - Resultados ampliados
- Descarga de microdatos

MAPAS

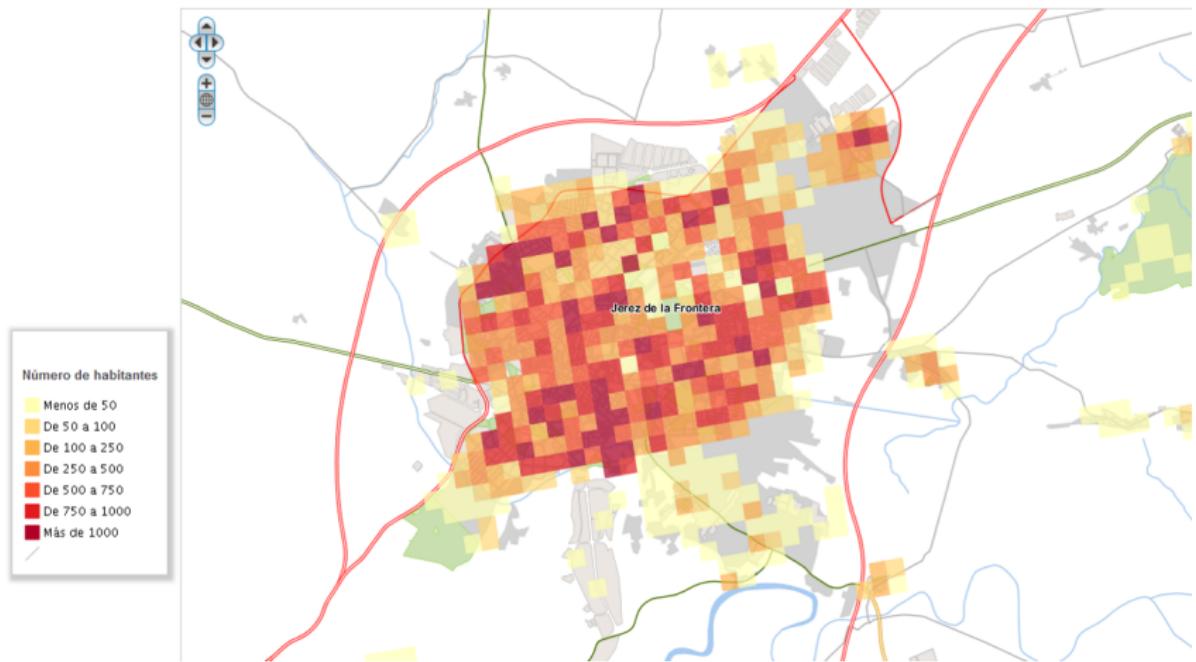
- Visualizador de mapas*:
 - Razón de mortalidad general. Celdas de 250 metros
 - Razón de mortalidad por causa. Celdas de 1 kilómetro

<http://www.juntadeandalucia.es/institutodeestadisticaycartografia/longevidad/>

Malla estadística

DISTRIBUCIÓN ESPACIAL DE LA POBLACIÓN EN ANDALUCÍA. AÑO 2013

Temas: [POBLACIÓN TOTAL](#) Población por nacionalidad Población por grupos de edad



Fuentes de datos

- Heterogeneidad y volúmenes brutales!



Fuentes de datos

- Heterogeneidad y volúmenes brutales!



Fuentes de datos

- Heterogeneidad y volúmenes brutales!



Fuentes de datos

- Heterogeneidad y volúmenes brutales!

Instituto de
Estadística
y Cartografía
de Andalucía



Agencia Estatal de Meteorología



Agencia de Medio Ambiente y Agua
**CONSEJERÍA DE MEDIO AMBIENTE
Y ORDENACIÓN DEL TERRITORIO**



Fuentes de datos

- Heterogeneidad y volúmenes brutales!

Instituto de
Estadística
y Cartografía
de Andalucía



Agencia Estatal de Meteorología



Agencia de Medio Ambiente y Agua
**CONSEJERÍA DE MEDIO AMBIENTE
Y ORDENACIÓN DEL TERRITORIO**



Fuentes de datos

- Heterogeneidad y volúmenes brutales!



Agencia de Medio Ambiente y Agua
CONSEJERÍA DE MEDIO AMBIENTE
Y ORDENACIÓN DEL TERRITORIO



- Obtener >>> Procesar >>> Normalizar >>> Analizar >>> Resumir

Fuentes de datos

- Heterogeneidad y volúmenes brutales!



- Obtener >>> Procesar >>> Normalizar >>> Analizar >>> Resumir
 - De multitud de ficheros de diferentes formatos a estructura única

Fuentes de datos

- Heterogeneidad y volúmenes brutales!



Agencia de Medio Ambiente y Agua
CONSEJERÍA DE MEDIO AMBIENTE
Y ORDENACIÓN DEL TERRITORIO



- Obtener >>> Procesar >>> Normalizar >>> Analizar >>> Resumir
 - De multitud de ficheros de diferentes formatos a estructura única
 - **magittr, dplyr, reshape, lubridate, stringi, data.table**

Fuentes de datos

- Heterogeneidad y volúmenes brutales!



Agencia de Medio Ambiente y Agua
CONSEJERÍA DE MEDIO AMBIENTE
Y ORDENACIÓN DEL TERRITORIO



- Obtener >>> Procesar >>> Normalizar >>> Analizar >>> Resumir
 - De multitud de ficheros de diferentes formatos a estructura única
 - magitr, dplyr, reshape, lubridate, stringi, data.table
 - RSQLite

Datos meteorológicos y calidad de aire

- Algunos paquetes para obtener y/o analizar datos:

Datos meteorológicos y calidad de aire

- Algunos paquetes para obtener y/o analizar datos:

- ALShum/rwunderground



Datos meteorológicos y calidad de aire

- Algunos paquetes para obtener y/o analizar datos:

- ALShum/rwunderground



- ropensci/rnoaa



Datos meteorológicos y calidad de aire

- Algunos paquetes para obtener y/o analizar datos:

- ALShum/rwunderground



- ropensci/rnoaa



- ropensci/rWBclimate



Datos meteorológicos y calidad de aire

- Algunos paquetes para obtener y/o analizar datos:

- ALShum/rwunderground



- ropensci/rnoaa



- ropensci/rWBclimate



- davidcarslaw/openair



Demo: rWBclimate

```
devtools::install_github("ropensci/rWBclimate", force = TRUE)
require(rWBclimate)
require(magrittr)
require(dplyr)

esp.dat <- get_model_temp("ESP", "mavg", 2020, 2100)

head(esp.dat)

# fromYear toYear      gcm      data scenario month locator
# 1    2020    2039 bccr_bcm2_0  9.404619     a2      1    ESP
# 2    2020    2039 bccr_bcm2_0  9.509619     a2      2    ESP
# 3    2020    2039 bccr_bcm2_0 10.937546     a2      3    ESP
# 4    2020    2039 bccr_bcm2_0 12.820676     a2      4    ESP
# 5    2020    2039 bccr_bcm2_0 15.364050     a2      5    ESP
# 6    2020    2039 bccr_bcm2_0 18.709639     a2      6    ESP
```

Demo: rWBclimate

```
esp.dat.bcc <- esp.dat %>%
  mutate(period = factor(paste0(fromYear, "-", toYear))) %>%
  select(period, scenario, gcm, month, data)

head(esp.dat.bcc)

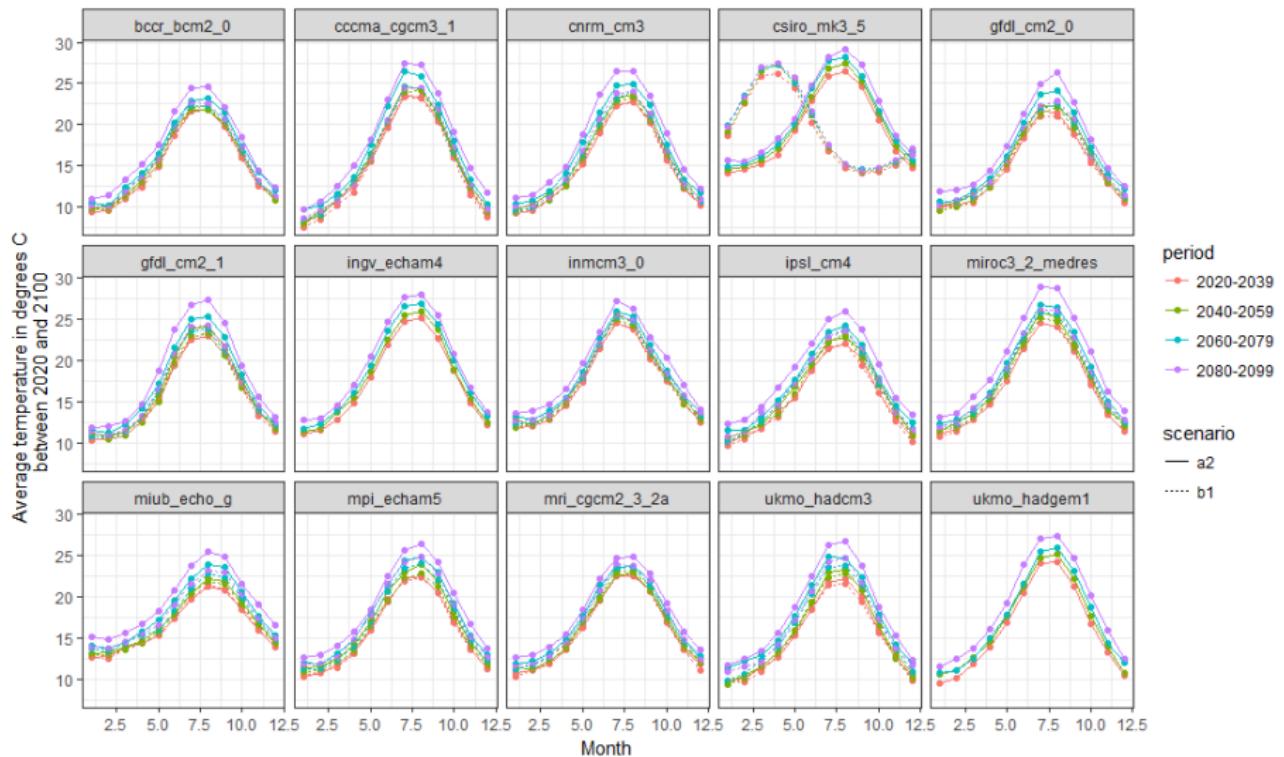
# period scenario      gcm month      data
# 1 2020-2039      a2 bccr_bcm2_0    1 9.404619
# 2 2020-2039      a2 bccr_bcm2_0    2 9.509619
# 3 2020-2039      a2 bccr_bcm2_0    3 10.937546
# 4 2020-2039      a2 bccr_bcm2_0    4 12.820676
# 5 2020-2039      a2 bccr_bcm2_0    5 15.364050
# 6 2020-2039      a2 bccr_bcm2_0    6 18.709639
```

Demo: rWBclimate

```
require(ggplot2)

ggplot(data = esp.dat.bcc,
       aes(x = month, y = data, colour = period, linetype = scenario))+
  geom_point() +
  geom_line() +
  ylab("Average temperature in degrees C \n between 2020 and 2100") +
  xlab("Month") +
  theme_bw()+
  facet_wrap(~ gcm, ncol = 5)
```

Demo: rWBclimate



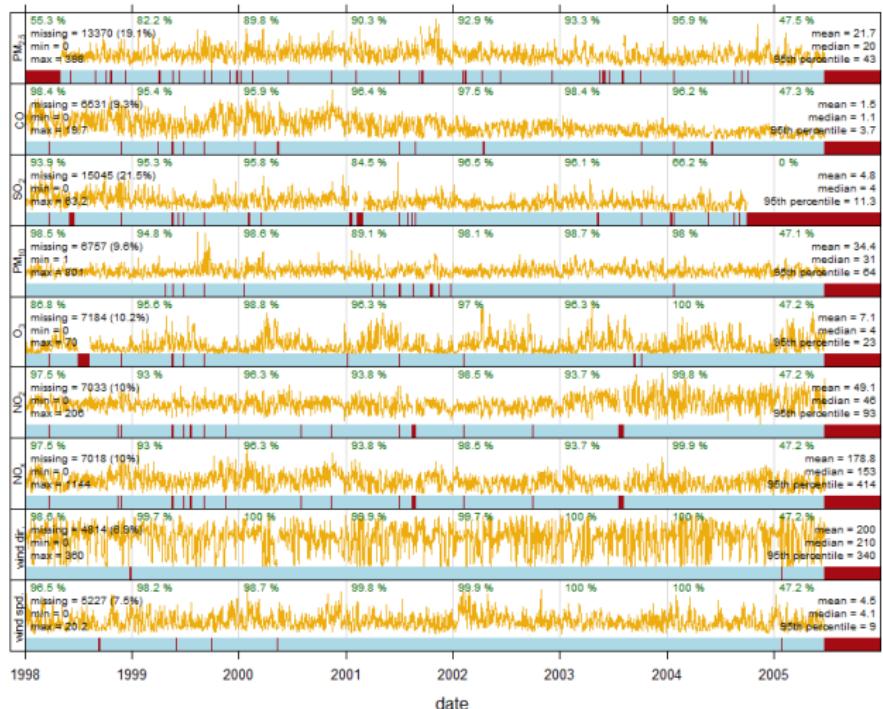
Demo: openair

```
install_github('davidcarslaw/openair')
require(openair)

head(data(mydata))

# A tibble: 6 x 10
#>   date           ws    wd    nox    no2     o3    pm10    so2    co    pm25
#>   <dttm>      <dbl> <int> <int> <int> <dbl> <dbl> <dbl> <dbl>
#> 1 1998-01-01 00:00:00 0.600  280    285    39     1    29  4.72  3.37    NA
#> 2 1998-01-01 01:00:00 2.16   230     NA     NA     NA    37    NA     NA     NA
#> 3 1998-01-01 02:00:00 2.76   190     NA     NA     3     34  6.83  9.60     NA
#> 4 1998-01-01 03:00:00 2.16   170    493    52     3     35  7.66 10.2     NA
#> 5 1998-01-01 04:00:00 2.40   180    468    78     2     34  8.07  8.91     NA
#> 6 1998-01-01 05:00:00 3.00   190    264    42     0     16  5.50  3.05     NA
```

Demo: openair



Medio ambiente: análisis exploratorio

- Cómo definir una ola de calor

Medio ambiente: análisis exploratorio

- Cómo definir una ola de calor
 - Estados Unidos: Un período de tiempo anormal e inusualmente caliente y húmedo que dura 2 o más días.

Medio ambiente: análisis exploratorio

- Cómo definir una ola de calor
 - Estados Unidos: Un período de tiempo anormal e inusualmente caliente y húmedo que dura 2 o más días.
 - Suiza: Un período de estrés térmico extremo, que puede poner en peligro la salud humana: índice de calor superior a 90, duración mínima de 3 días.

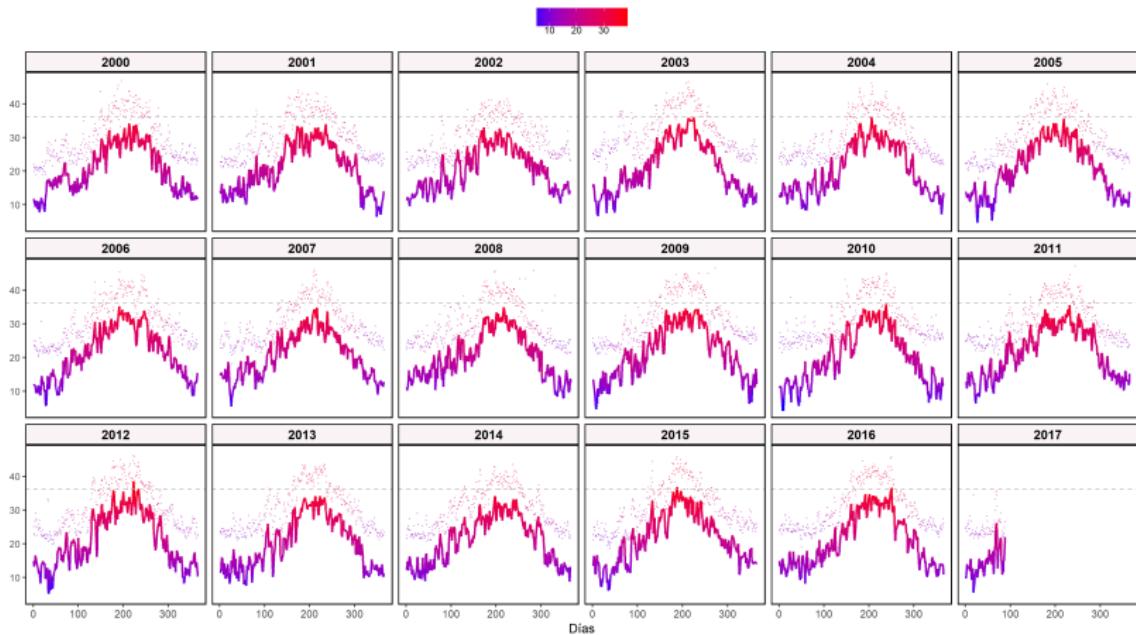
Medio ambiente: análisis exploratorio

- Cómo definir una ola de calor
 - Estados Unidos: Un período de tiempo anormal e inusualmente caliente y húmedo que dura 2 o más días.
 - Suiza: Un período de estrés térmico extremo, que puede poner en peligro la salud humana: índice de calor superior a 90, duración mínima de 3 días.
 - España: Un episodio de al menos tres días consecutivos, en que como mínimo el 10% de las estaciones consideradas registran máximas por encima del percentil del 95% de su serie de temperaturas máximas diarias de los meses de julio y agosto del período 1971-2000.

Medio ambiente: análisis exploratorio

- Series temporales

Temperaturas máximas alcanzadas en Andalucía en 2000-2017
Valores diarios máximos absolutos (-) y medios máximos (·) para el conjunto de la comunidad autónoma



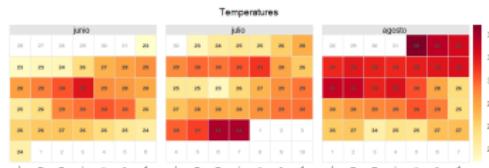
Fuente de datos: AEMET

Medio ambiente: análisis exploratorio

- Cálculos de indicadores: [geanders/weathermetrics](#)

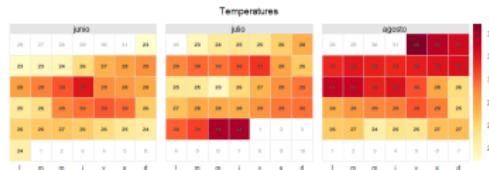
Medio ambiente: análisis exploratorio

- Cálculos de indicadores: [geanders/weathermetrics](#)
- temperatura

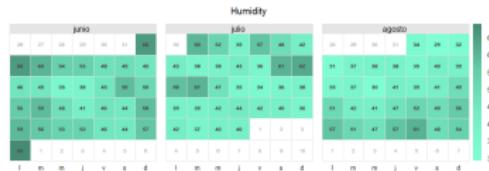


Medio ambiente: análisis exploratorio

- Cálculos de indicadores: [geanders/weathermetrics](#)
- temperatura

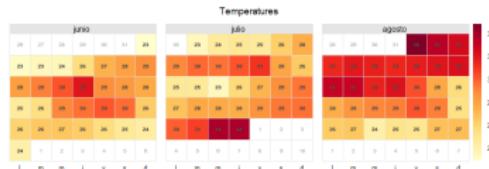


- humedad

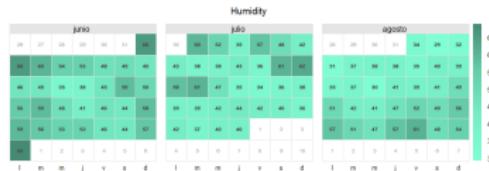


Medio ambiente: análisis exploratorio

- Cálculos de indicadores: [geanders/weathermetrics](#)
- temperatura



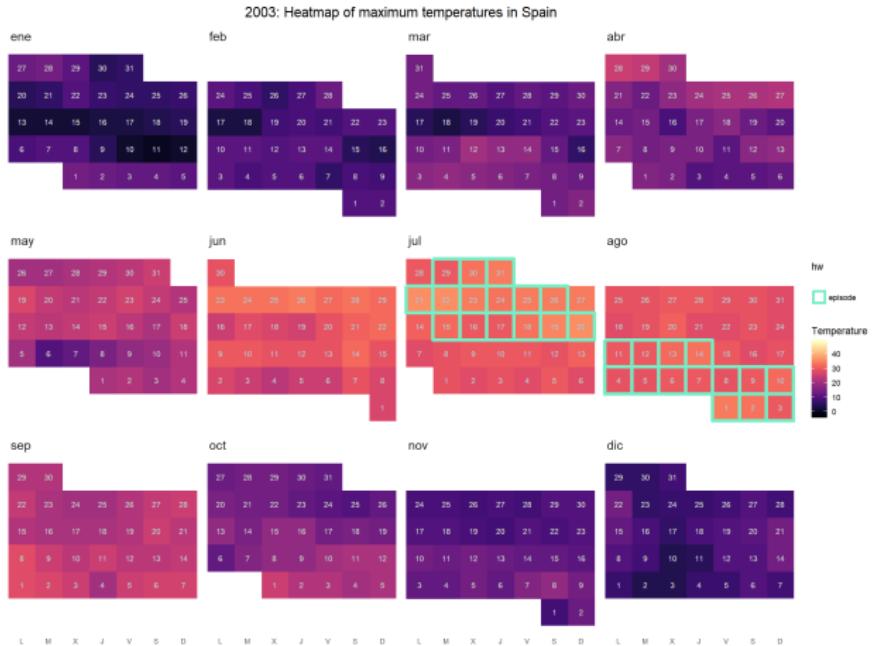
- humedad



- heat index



Medio ambiente: análisis exploratorio



Geoestadística y análisis espacial

- Operaciones básicas espaciales con datos vectoriales:

Geoestadística y análisis espacial

- Operaciones básicas espaciales con datos vectoriales:
 - rgdal, maptools, sp, sf

Geoestadística y análisis espacial

- Operaciones básicas espaciales con datos vectoriales:
 - rgdal, maptools, sp, sf
- Operaciones con datos tipo raster:

Geoestadística y análisis espacial

- Operaciones básicas espaciales con datos vectoriales:
 - rgdal, maptools, sp, sf
- Operaciones con datos tipo raster:
 - raster, rasterVis

Geoestadística y análisis espacial

- Operaciones básicas espaciales con datos vectoriales:
 - rgdal, maptools, sp, sf
- Operaciones con datos tipo raster:
 - raster, rasterVis
- Análisis espacial y geoestadístico:

Geoestadística y análisis espacial

- Operaciones básicas espaciales con datos vectoriales:
 - rgdal, maptools, sp, sf
- Operaciones con datos tipo raster:
 - raster, rasterVis
- Análisis espacial y geoestadístico:
 - gstat, rgeos, automap

Geoestadística y análisis espacial

- Operaciones básicas espaciales con datos vectoriales:
 - rgdal, maptools, sp, sf
- Operaciones con datos tipo raster:
 - raster, rasterVis
- Análisis espacial y geoestadístico:
 - gstat, rgeos, automap
- Puente entre el entorno ArcGIS y R:

Geoestadística y análisis espacial

- Operaciones básicas espaciales con datos vectoriales:
 - rgdal, maptools, sp, sf
- Operaciones con datos tipo raster:
 - raster, rasterVis
- Análisis espacial y geoestadístico:
 - gstat, rgeos, automap
- Puente entre el entorno ArcGIS y R:
 - **arcgisbinding**

Geoestadística: métodos de interpolación

```
estaciones <- rgdal::readOGR(dsn = ".", layer = "meteodatos")

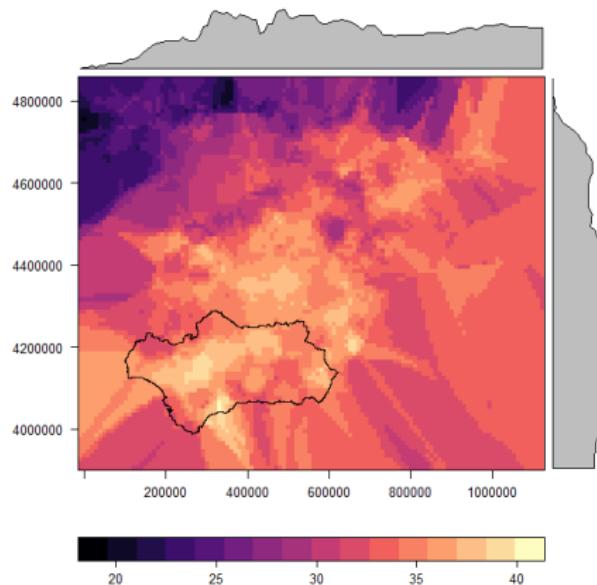
r <- raster::raster(xmn = -13503, xmx = 1126011, ymn = 3903695, ymx= 4858797,
                     res=10000, crs = estaciones@proj4string)

gs <- gstat::gstat(formula= temperatura ~ 1, locations = estaciones,
                    nmax=5, set=list(idp = 0))

nn <- raster::interpolate(r, gs)

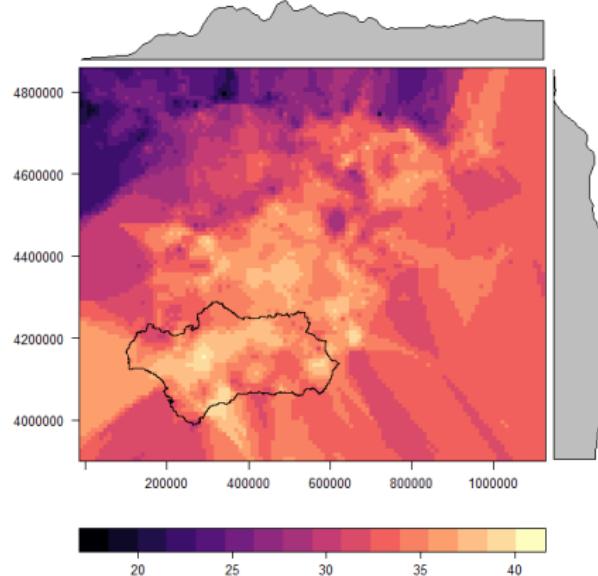
rasterVis::levelplot(nn, margin = list(FUN = 'median'),
                      main=list(label=paste0("Nearest Neighborhood Interpolation (",
                                             unique(dta$FECHA), ")\n",
                                             as.character(deparse(gs$call))),
                      side=1, line=0.5, cex=0.7),
                      contour=F, par.settings = magmaTheme)
```

Geoestadística: métodos de interpolación



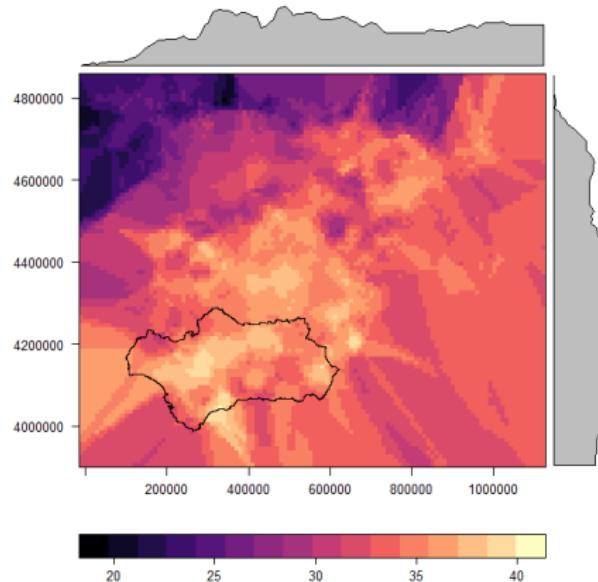
Nearest Neighborhood?

Geoestadística: métodos de interpolación



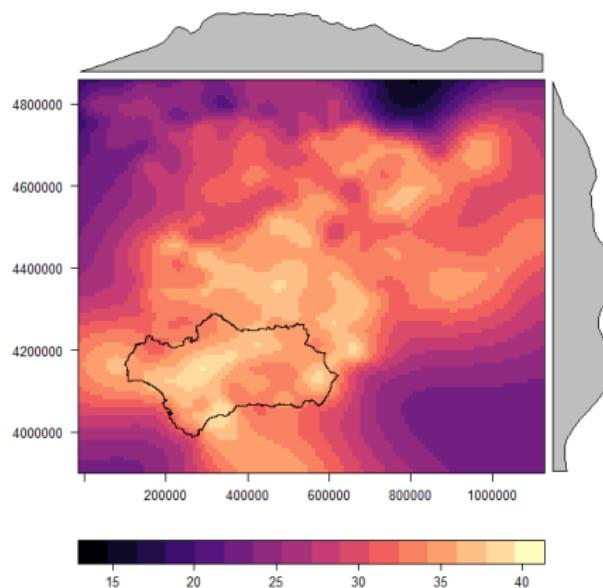
Inverse Distance Weighted?

Geoestadística: métodos de interpolación



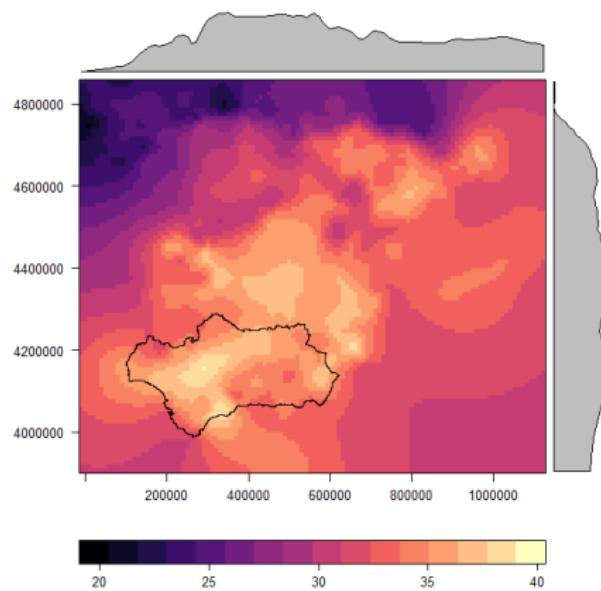
Inverse Distance Weighted optimized?

Geoestadística: métodos de interpolación



Thin plate spline regression?

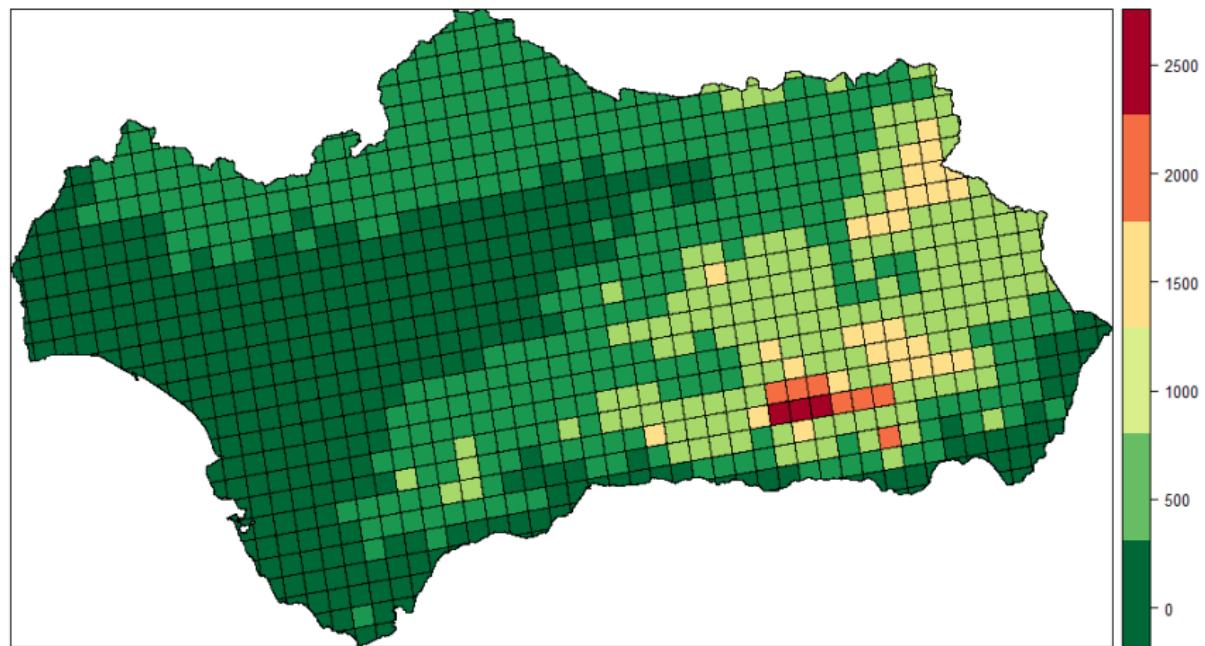
Geoestadística: métodos de interpolación



Ordinary Kriging?

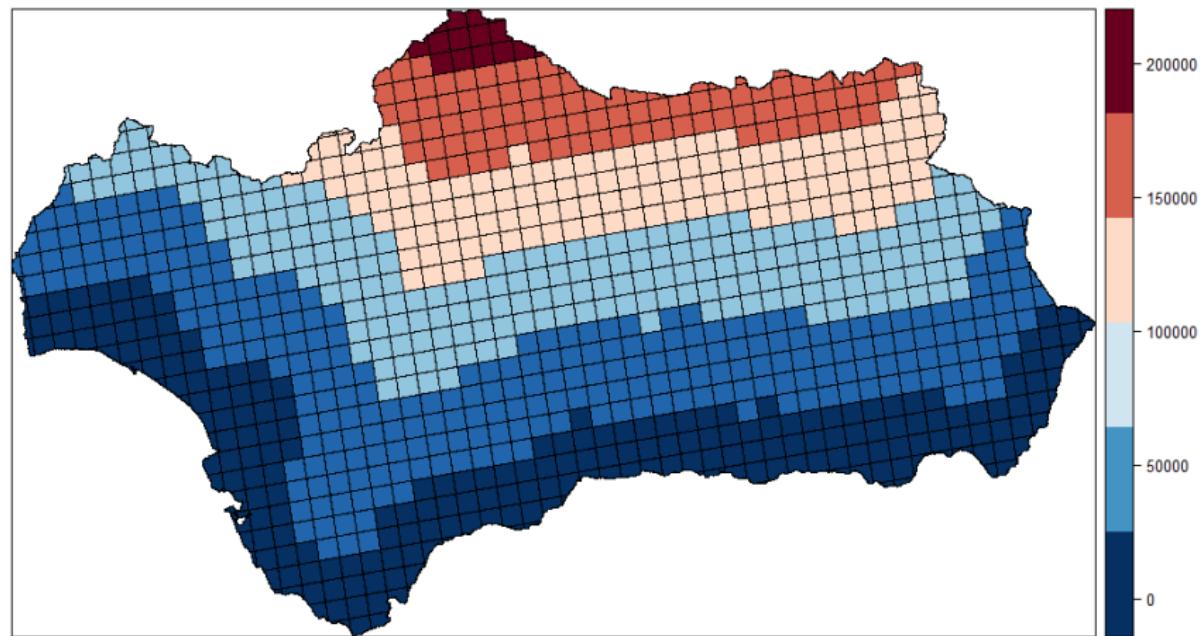
Geoestadística: métodos de interpolación

¿Y si queremos considerar las variaciones en elevación?



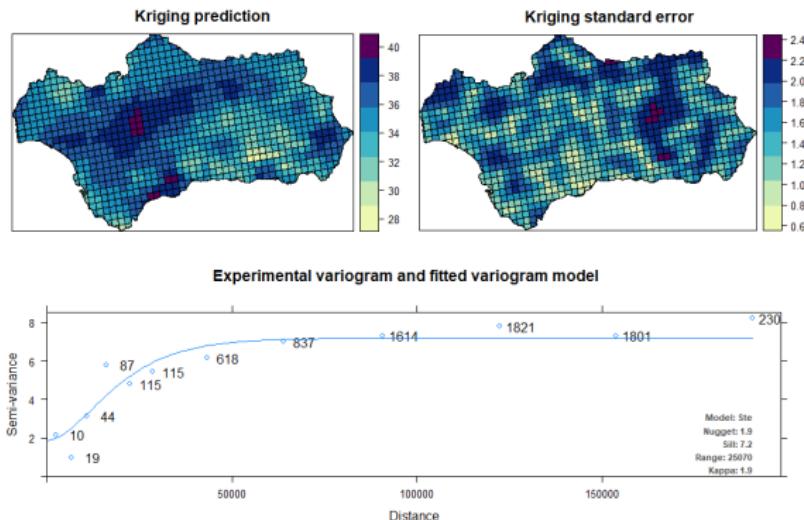
Geoestadística: métodos de interpolación

¿Y si queremos considerar las variaciones en elevación..y proximidad a la costa?



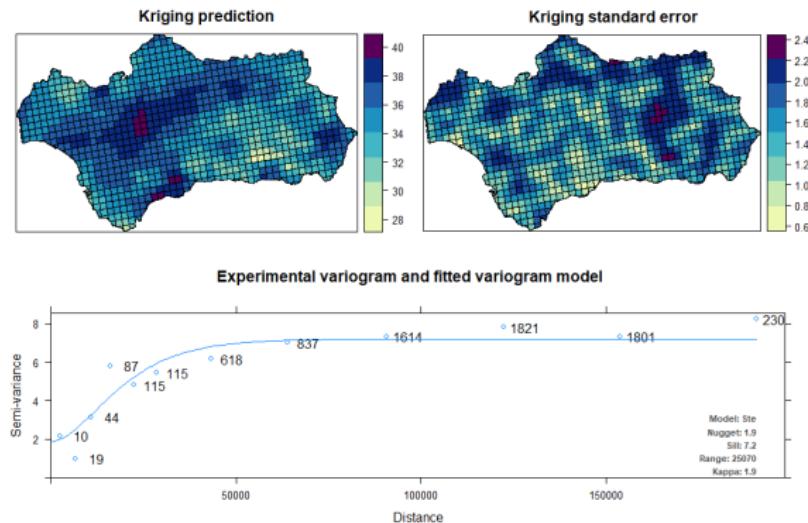
Geoestadística: métodos de interpolación

Y NO queremos ni tocar los variogramas!



Geoestadística: métodos de interpolación

Y NO queremos ni tocar los variogramas!



..pero nos gustaría algo más. . .

Geoestadística: métodos de interpolación

... más sofisticado cómo, quizás...



Empirical Bayesian Kriging Regression Prediction?

Geoestadística: métodos de interpolación

- R script como herramienta de geoprocесamiento de ArcGIS



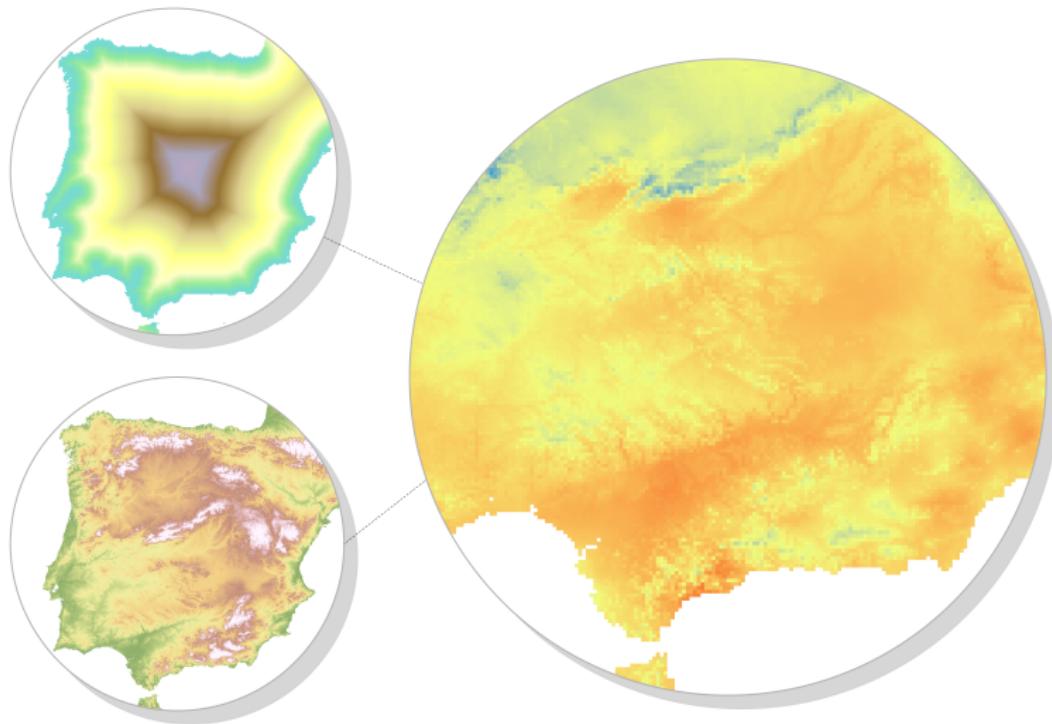
Geoestadística: métodos de interpolación

- R script como herramienta de geoprocесamiento de ArcGIS

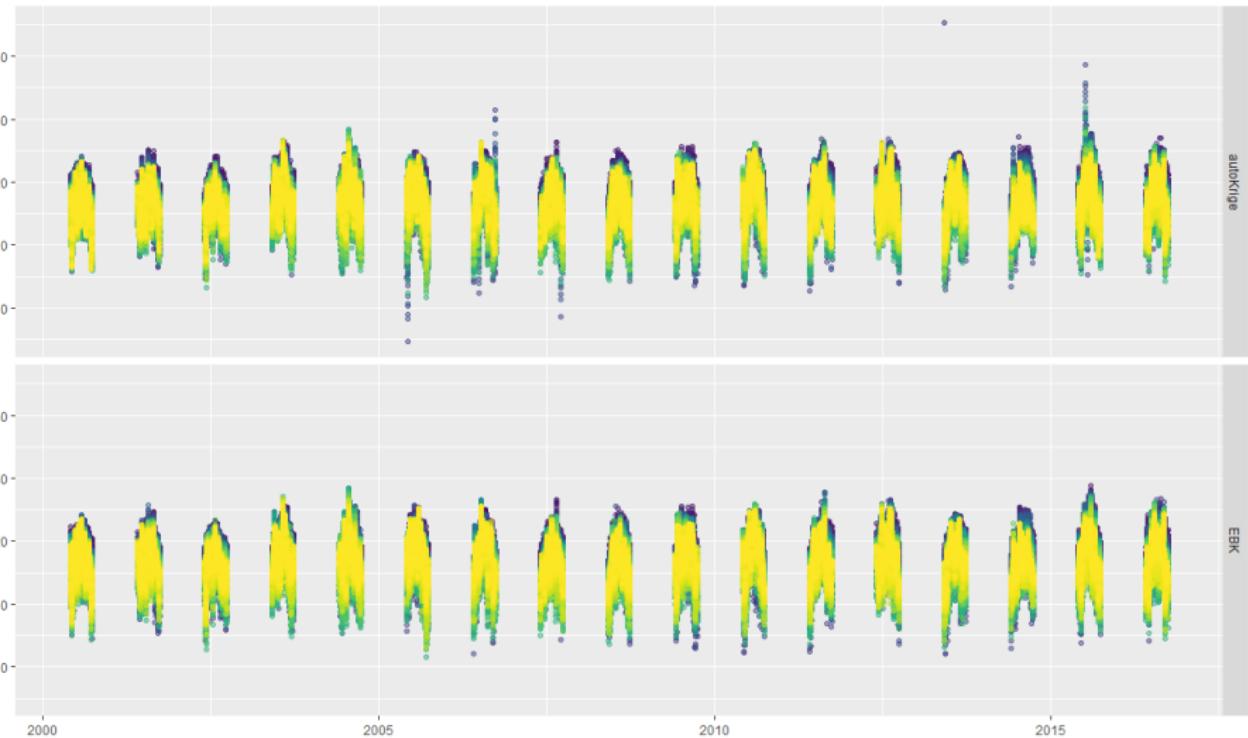


- <https://github.com/R-ArcGIS/>

Geoestadística: métodos de interpolación



Geoestadística: métodos de interpolación



ERA-Interim

- Re-análisis global atmosférico desde 1979, actualizado continuamente en tiempo real



ERA-Interim

- Re-análisis global atmosférico desde 1979, actualizado continuamente en tiempo real



- <http://apps.ecmwf.int/datasets/data/interim-full-daily/>

ERA-Interim

- Re-análisis global atmosférico desde 1979, actualizado continuamente en tiempo real



- <http://apps.ecmwf.int/datasets/data/interim-full-daily/>
- NetCDF

Accesibilidad a recursos sanitarios

```
require(ggmap)

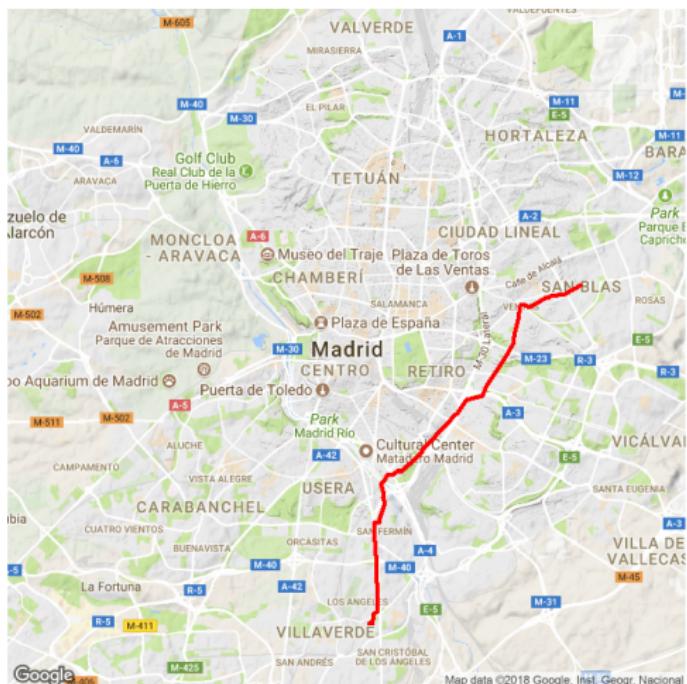
from <- "Emilio Munoz, Madrid"
to <- "Calle Cifuentes, Madrid"

route_df <- route(from, to, mode = "walking", structure = "route")

qmap("Madrid, Spain", zoom = 12) +
  geom_path(aes(x = lon, y = lat), colour = "red", size = 1.5,
            data = route_df, lineend = "round")

routeQueryCheck()
# 2481 routing queries remaining
```

Accesibilidad a recursos sanitarios



- Un destino + un origen + un momento en tiempo

Accesibilidad a recursos sanitarios

¿ Y qué hacer si tenemos..



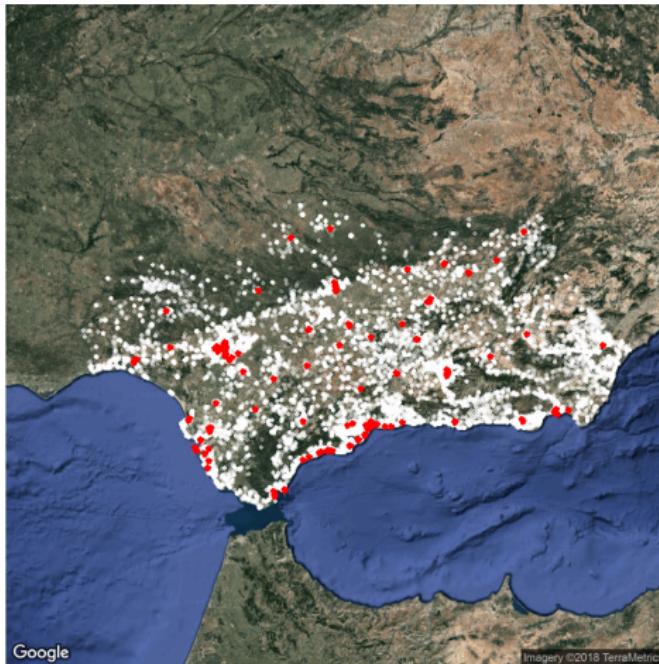
Accesibilidad a recursos sanitarios

..más de 40.000 orígenes?



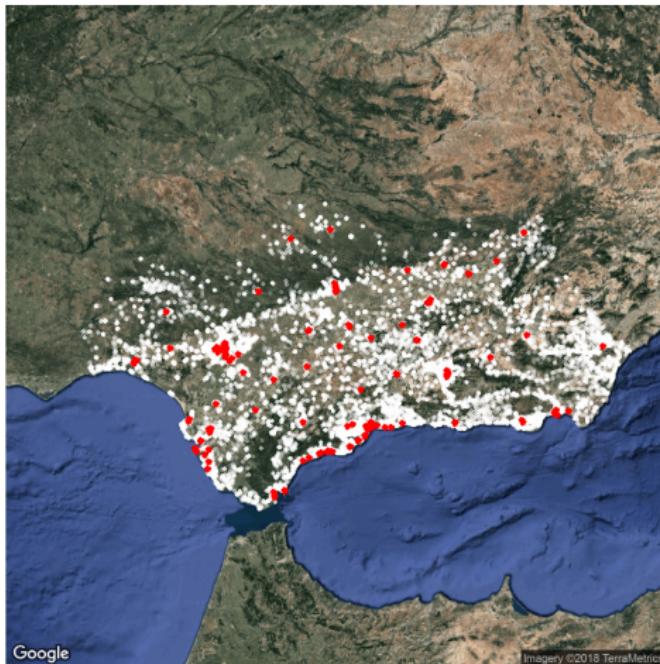
Accesibilidad a recursos sanitarios

..mas de 40.000 orígenes? y 200 destinos?



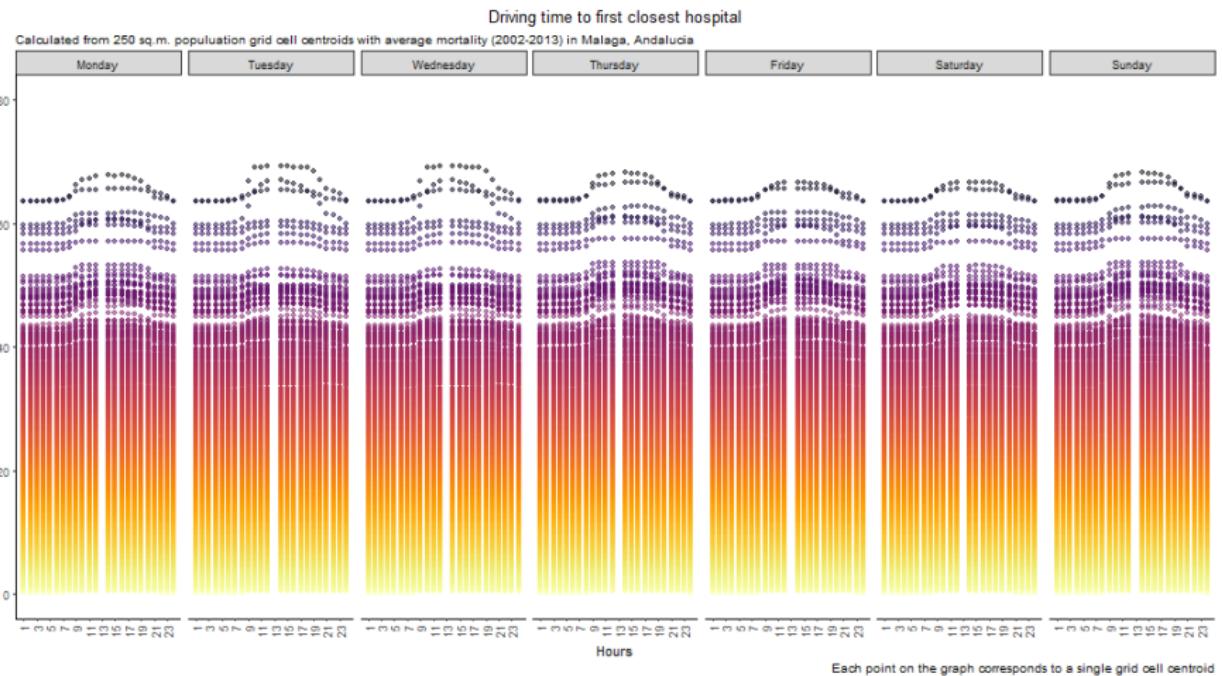
Accesibilidad a recursos sanitarios

..mas de 40.000 orígenes? y 200 destinos?



...y queremos calcular tiempos de viaje para cada hora de cada día de la semana?

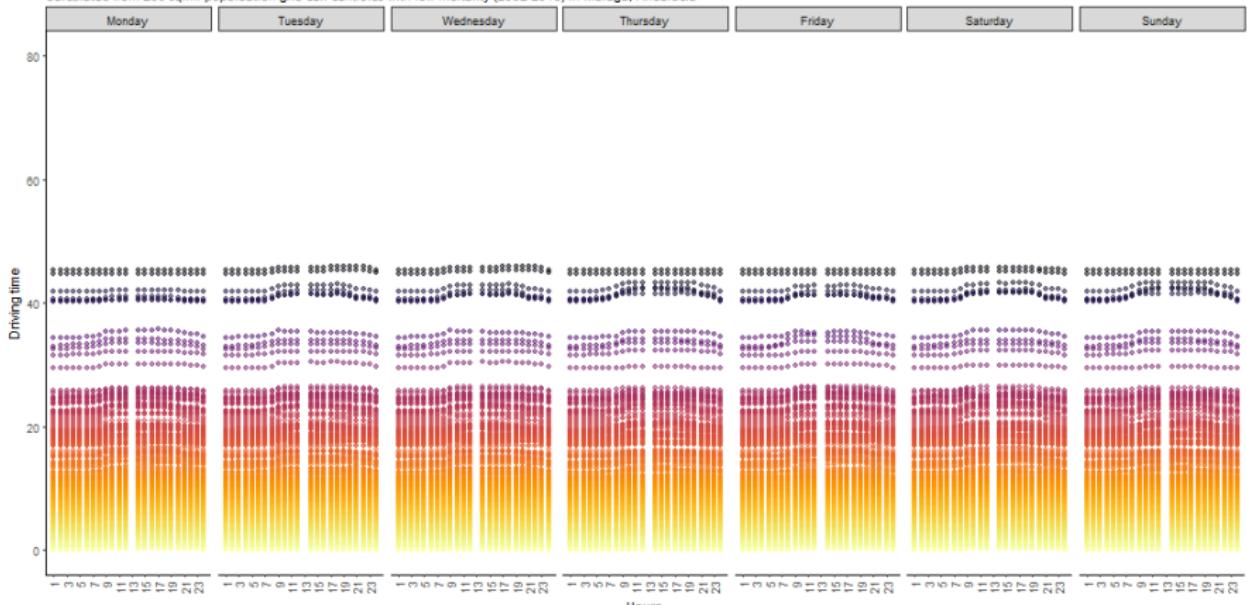
Accesibilidad: mortalidad media



Accesibilidad: mortalidad baja

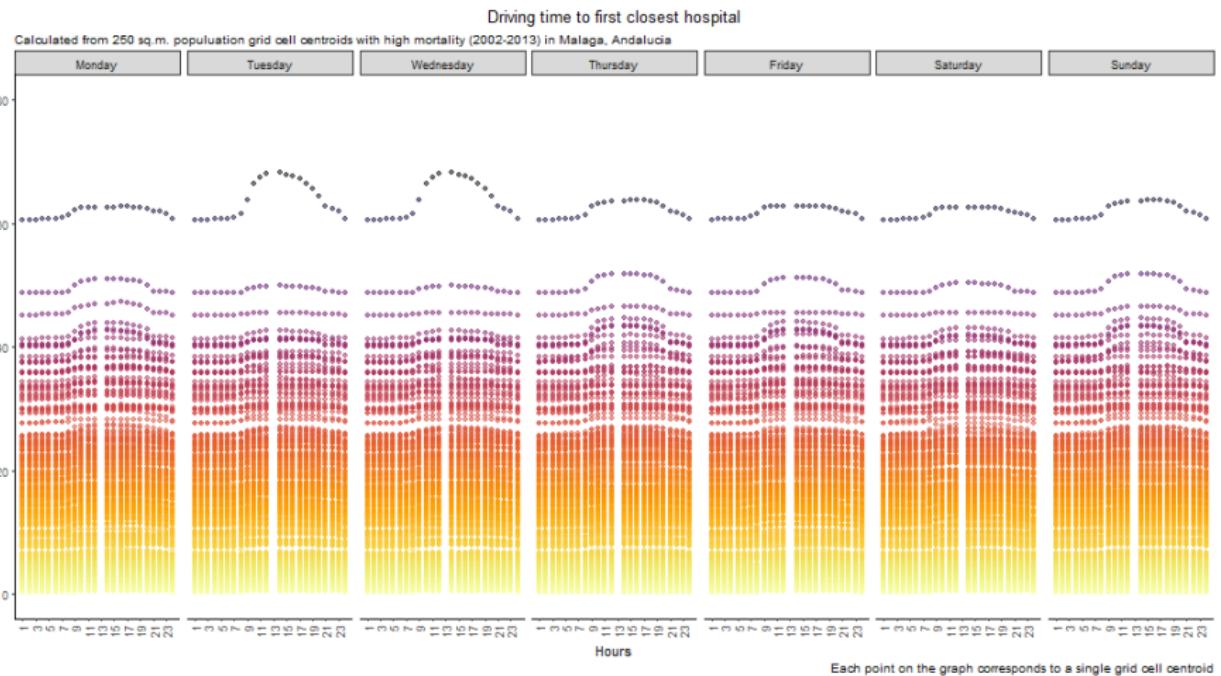
Driving time to first closest hospital

Calculated from 250 sq.m. population grid cell centroids with low mortality (2002-2013) in Málaga, Andalucía



Each point on the graph corresponds to a single grid cell centroid

Accesibilidad: mortalidad alta



Redes sociales como predictor alternativo

- Obtener datos a través de streaming API <https://apps.twitter.com/>

```
devtools::install_github("mkearney/rtweet")

appname <- "..."
key <- "..."
secret <- "..."

twitter_token <- create_token(
  app = appname,
  consumer_key = key,
  consumer_secret = secret)

stream_tweets(lookup_coords("Seville, Spain"), timeout = FALSE,
              file_name = "SevillaTweets.json", parse = FALSE)

parse_stream("SevillaTweets.json")
```

Redes sociales como predictor alternativo

- Minería de texto en R:

Redes sociales como predictor alternativo

- Minería de texto en R:
 - Procesamiento de texto: stringi, stringr, rjson, RTextTools, textclean, qdap, tau

Redes sociales como predictor alternativo

- Minería de texto en R:
 - Procesamiento de texto: stringi, stringr, rjson, RTextTools, textclean, qdap, tau
 - Matrices de términos y documentos: tm, SnowballC

Redes sociales como predictor alternativo

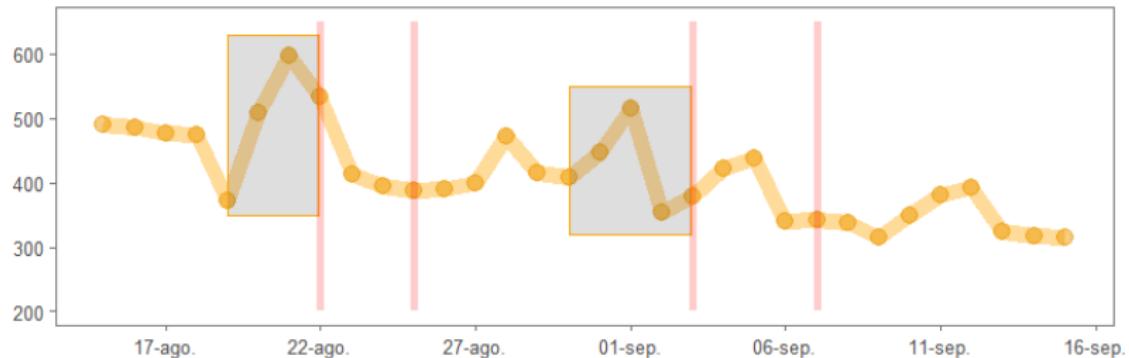
- Minería de texto en R:
 - Procesamiento de texto: stringi, stringr, rjson, RTextTools, textclean, qdap, tau
 - Matrices de términos y documentos: tm, SnowballC
 - Análisis de sentimientos: tidytext, RSentiment, sentiment

Redes sociales como predictor alternativo

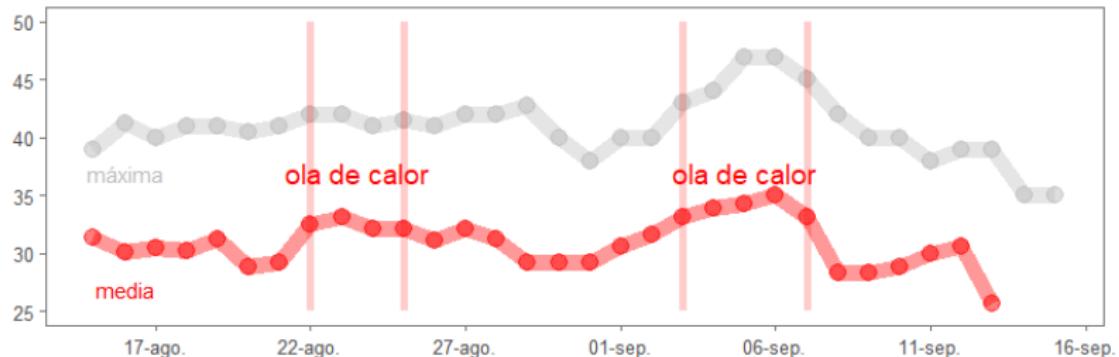
- Minería de texto en R:
 - Procesamiento de texto: stringi, stringr, rjson, RTextTools, textclean, qdap, tau
 - Matrices de términos y documentos: tm, SnowballC
 - Análisis de sentimientos: tidytext, RSentiment, sentiment
 - Modelación de temas: spatialEco, topicmodels

Redes sociales como predictor alternativo

Tweets relacionados con calor

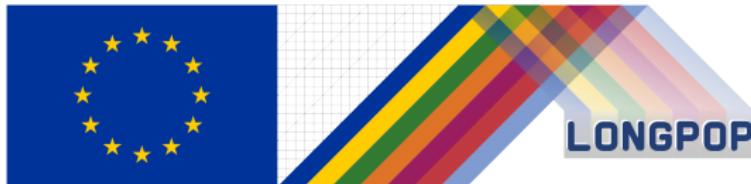


Temperaturas



¡Muchas gracias!

Datos de contacto
[@dordanovich](https://twitter.com/dordanovich)
dariya.ordanovich@esri.es



This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060