



AIRCyL

Dashboard Open Data

Calidad del Aire

En Castilla y León

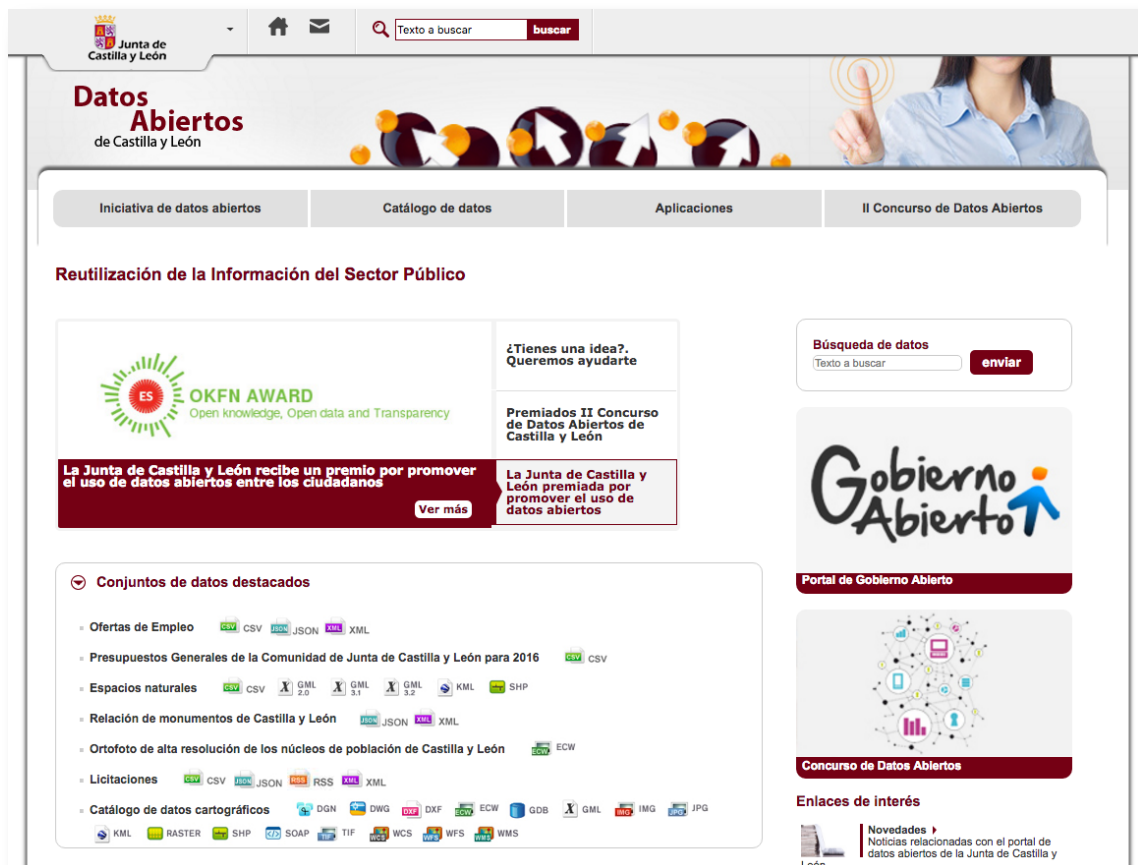


ÍNDICE

Introducción	3
Descripción de los datos de entrada	6
Calidad del Aire	6
Estaciones	7
Diccionario de datos	8
Calidad del Aire	8
Estaciones	9
Metodología	10
Estructura del proyecto	10
Breve detalle de cada fichero del proyecto	11
Metodología del proyecto	12
OpenAIR	12
Breve detalle de las funciones utilizadas de OpenAIR	13
Descripción, Mini-manual	16

INTRODUCCIÓN

La Junta de Castilla y León ha lanzado un programa para poner al alcance de los ciudadanos un extenso catálogo de datos abiertos referentes a distintos áreas de la Comunidad. El portal de dicho catálogo se encuentra en la URL: www.datosabiertos.jcyl.es



El catálogo de datos, cubre un amplio espectro tanto de tipo de datos, como calidad de los mismos y formatos ofrecidos para su descarga.

Encontramos formatos como:

- CSV
- GML2
- GML31
- GML32, etc.
- JSON
- XML, etc.

De entre todos los datos he seleccionado los referentes a **CALIDAD DEL AIRE** y **ESTACIONES DE MEDICIÓN**.

Las URLs de ambas:

Calidad del aire

Descripción:
Datos de concentración de contaminantes registrados en las estaciones de control de la calidad del aire de la Junta de Castilla y León

Publicación en JCYL.ES del conjunto de datos:

- **Inicio de la publicación:** 27 de marzo de 2012
- **Última actualización:** 23 de febrero de 2016
- **Frecuencia de actualización:** Anual

Formatos disponibles para descarga:

 CSV

Agente que publica el dato:

- Junta de Castilla y León

Ámbito geográfico del conjunto de datos:

- Castilla y León (España)

Etiquetas: Calidad del aire, medio ambiente, atmosfera, contaminación

Licencia: [Creative Commons Reconocimiento 3.0](#)

Información adicional:
Puedes encontrar más información relacionada con el contenido del dataset en: [Acceso a los datos de la Red de Control de la Calidad del Aire](#)

 Contactar

 Enviar contenido

 Imprimir contenido

 Escuchar

Contenidos Relacionados

▼ **DataSets:**

- Estaciones de control de la calidad del aire

Estaciones de control de la calidad del aire

Descripción:
Relación de estaciones de control de la calidad del aire en Castilla y León

Publicación en JCYL.ES del conjunto de datos:

- **Inicio de la publicación:** 27 de marzo de 2012
- **Frecuencia de actualización:** Anual

Formatos disponibles para descarga:

 CSV

Agente que publica el dato:

- Junta de Castilla y León

Ámbito geográfico del conjunto de datos:


- Castilla y León (España)

Etiquetas: Calidad del aire, medio ambiente, estaciones, atmosfera, contaminación


Licencia: [Creative Commons Reconocimiento 3.0](#)

Información adicional:
Puedes encontrar más información relacionada con el contenido del dataset en: [Red de control de calidad del aire](#)

 Contactar

 Enviar contenido

 Imprimir contenido

 Escuchar

Contenidos Relacionados

▼ **DataSets:**

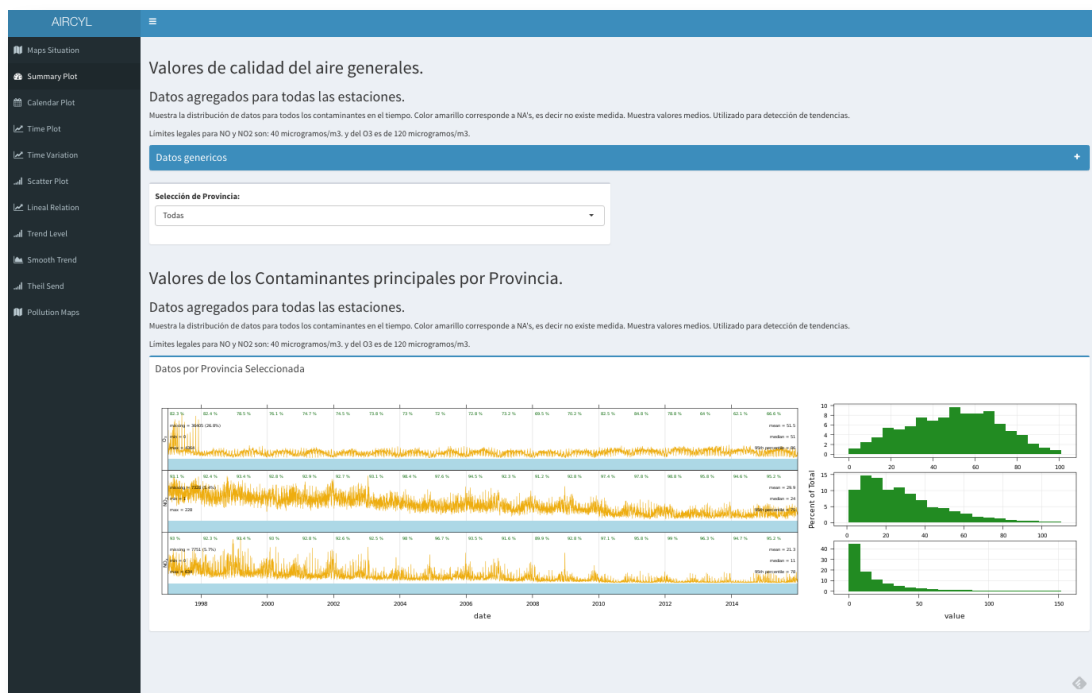
- Calidad del aire

Los datos en ambos casos son descargables en formato CSV. El problema a resolver en este caso es la propuesta como objetivo de la creación de un sistema de visualización y estudio de los datos que permita a los usuario de forma visual el acceso a los datos para el estudio, tanto de

los datos distribuidos por población y temporalidad, así como las posibles relaciones entre los distintos contaminantes y sus tendencias a lo largo del tiempo.

Como Trabajo Fin de Master se buscado un modelo de visualización en formato dashboard que contempla tanto el análisis descriptivo como el estudio de tendencias de las series temporales de dichos contaminantes.

El resultado queda reflejado en la siguiente imagen:



DESCRIPCIÓN

La base del proyecto son dos ficheros en formato CSV que contienen toda la información a explotar posteriormente en la visualización.

- Calidad del aire
- Estaciones de medición de calidad del aire.

CALIDAD DEL AIRE

Descripción:

Datos de concentración de contaminantes registrados en las estaciones de control de la calidad del aire de la Junta de Castilla y León

Publicación en JCYL.ES del conjunto de datos:

- Inicio de la publicación: 27 de marzo de 2012
- Última actualización: 23 de febrero de 2016
- Frecuencia de actualización: Anual

Formatos disponibles para descarga:

- [CSV](#)

Agente que publica el dato:

- Junta de Castilla y León

Ámbito geográfico del conjunto de datos:

- Castilla y León (España)

Etiquetas:

- Calidad del aire, medio ambiente, atmosfera, contaminación

Licencia:

- Creative Commons Reconocimiento 3.0

Información adicional:

Puedes encontrar más información relacionada con el contenido del dataset en: Acceso a los datos de la [Red de Control de la Calidad del Aire](#)

ESTACIONES DE MEDICIÓN CALIDAD DEL AIRE

Descripción:

Relación de estaciones de control de la calidad del aire en Castilla y León

Publicación en JCYL.ES del conjunto de datos:

- Inicio de la publicación: 27 de marzo de 2012
- Frecuencia de actualización: Anual

Formatos disponibles para descarga:

- [CSV](#)

Agente que publica el dato:

- Junta de Castilla y León

Ámbito geográfico del conjunto de datos:

- Castilla y León (España)

Etiquetas:

- Calidad del aire, medio ambiente, atmosfera, contaminación

Licencia:

- Creative Commons Reconocimiento 3.0

Información adicional:

Puedes encontrar más información relacionada con el contenido del dataset en: Acceso a los datos de la [Red de Control de la Calidad del Aire](#)

DICCIONARIO DE DATOS:

Se muestra la cabeza de los ficheros y los tipos de datos que se ofrecen. Además, comentaremos brevemente las correcciones que se han solicitado a la Entidad que ofrece los datos para su resolución.

CALIDAD DEL AIRE

Cabecera de fichero:

- DIA
- CO (mg/m3)
- NO (ug/m3)
- NO2 (ug/m3)
- O3 (ug/m3)
- PM10 (ug/m3)
- SH2 (ug/m3)
- PM25 (ug/m3)
- PST (ug/m3)
- SO2 (ug/m3)
- PROVINCIA
- ESTACIÓN

Tipología de dato:

- DIA: Formato texto.
- CO (mg/m3): Formato decimal.
- NO (ug/m3): Formato integer.
- NO2 (ug/m3): Formato integer.
- O3 (ug/m3): Formato integer.
- PM10 (ug/m3): Formato integer.
- SH2 (ug/m3): Campo vacío actualmente.
- PM25 (ug/m3): Campo vacío actualmente.
- PST (ug/m3): Campo vacío actualmente.
- SO2 (ug/m3): Formato integer.
- PROVINCIA: Formato texto.
- ESTACIÓN: Formato texto.

Errores detectados:

- Columna **DIA**, se ha solicitado la posibilidad de microdatos a nivel de hora para una profundidad en el análisis y visualización. De hecho, hay gráficas que se han dejado a dicho nivel para uso futuro sin necesidad de recodificar los algoritmos.

- Petición de adición de columna de dirección del viento y velocidad.
- Columna **PROVINCIA** y **ESTACIÓN** poseen errores de codificación. No existe un estándar para una mejor detección de filas con el mismo valor (Ej.: Valladolid1, Valladolid 1 y Valladolid_1, etc.).
- Error en los datos de la columna **O3**. Los datos no se están recogiendo correctamente, muestra datos en un rango de 0 a 4364 (dato imposible). Quedó detectado en la provincia de Burgos, además va aumentando según aumentamos el día. Se envió información para que la Entidad pueda corregir el dato.

ESTACIONES DE CONTROL DE CALIDAD DEL AIRE

Cabecera de fichero:

- ESTACIÓN
- LOCALIZACIÓN
- PROVINCIA
- LONGITUD
- LATITUD
- ALTITUD
- OPERATIVA

Tipología de dato:

- ESTACIÓN: Formato texto.
- LOCALIZACIÓN: Formato texto.
- PROVINCIA: Formato texto.
- LONGITUD: Formato texto.
- LATITUD: Formato texto.
- ALTITUD: Formato integer.
- OPERATIVA: Formato texto.

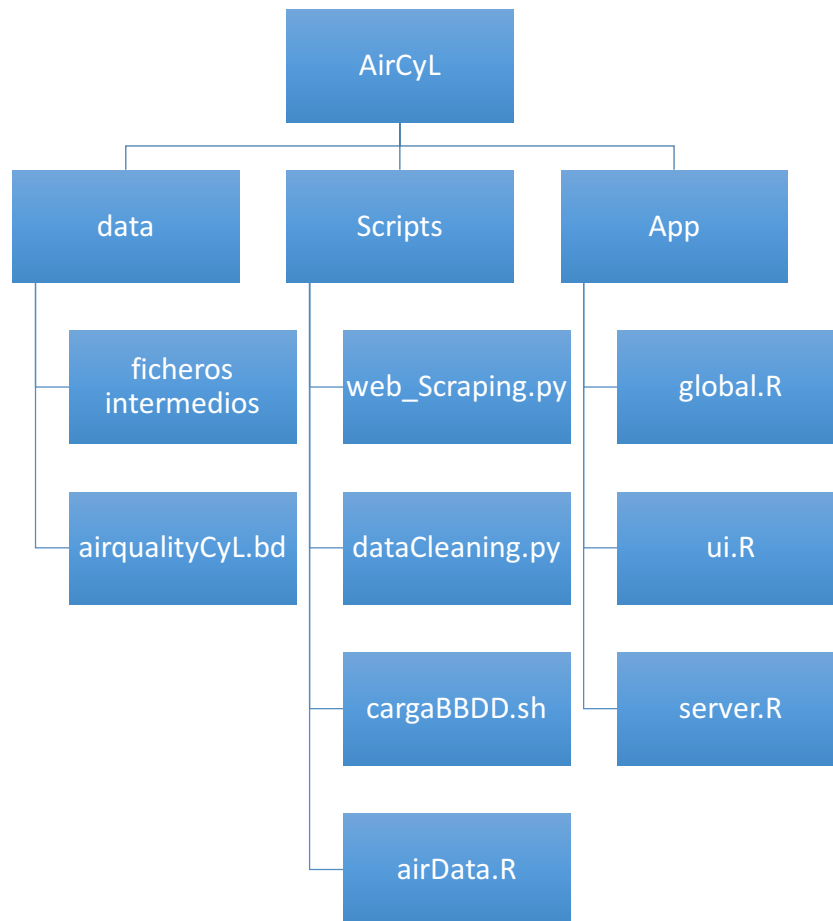
Errores detectados:

- Columna **PROVINCIA** y **ESTACIÓN** poseen errores de codificación. No existe un estándar para una mejor detección de filas con el mismo valor (Ej.: Valladolid1, Valladolid 1 y Valladolid_1, etc.).
- Actualmente, el fichero estaba muy desactualizado, pero la Entidad actualizó el fichero gracias a una petición para posibilitar un análisis más actual.

METODOLOGÍA:

ESTRUCTURA DEL PROYECTO:

El siguiente gráfico muestra la estructura final del proyecto:



BREVE DETALLE DEL PROYECTO GENERAL:

El proyecto lo he dividido en 3 bloques para su mejor mantenimiento posterior. La base son 3 carpetas o directorios:

- **SCRIPTS:** donde se almacenan los scripts que se lanzan desde el CRON de forma anual (que es cuando se actualizan los datos). Dichos scripts lo que hacen es recoger los datos desde la web, en formato CSV, limpiarlos para evitar errores y discrepancias posteriores, crear dentro de **DATA** las tablas en fichero y posteriormente desde dichas tablas se cargan los datos correctos a la base de datos. Un último script guarda una versión reducida de dichos datos de la BBDD para su posterior análisis en la **APP** de SHINY.

- **APP:** se almacenan exclusivamente los ficheros tanto de procesamiento de datos globales (variables y estructuras de datos que se utilizan durante el ciclo de vida de la Shiny App) como el front y el back de la App.
- **DATA:** se almacenan tanto los ficheros de los datos tanto raw como procesados, como la base de datos final. La base de datos se ha utilizado SQLite, base de datos local y de acceso rápido.

BREVE DETALLE DE CADA FICHERO DEL PROYECTO:

- **SCRIPTS/web_Scraping.py:** Recorre la página web y detecta ambos ficheros en formato CSV. Esto es necesario debido a que los nombres de los ficheros son una sucesión aleatoria de números que no se mantienen en el tiempo por lo que es imposible detectarlos de forma fija. Dicho script también se encuentra codificado en R obteniendo el mismo resultado. Crea las tablas Raw en DATA para su posterior tratamiento. Se lanza de forma anual mediante CRON.
- **SCRIPTS/dataCleaning.py:** Tratamiento de los ficheros RAW, creando los ficheros CSV ya tratados, eliminando incongruencias en la estandarización de la nomenclatura de los datos, así como recodificación de los nombres de columnas. Además añade los columnas a mayores (viento) para posteriores análisis cuando la Entidad facilite dichos datos y formatea de forma correcta latitud y longitud. Se lanza de forma anual mediante CRON.
- **SCRIPTS/cargaBBDD.sh:** Mediante CSVKIT, lo que hace el script es coger las tablas tratadas en el script anterior y crear la base de datos en SQLite. Se lanza de forma anual mediante CRON.
- **SCRIPTS/airData.R:** Script principal que lee los datos de la base de datos, cambia los formatos (tipos) de los datos y enriquece dicha tabla con datos finales de ambos ficheros (estaciones y medidas) para crear el fichero CSV principal para el posterior análisis. Elimina también variables y datos temporales. Se lanza de forma anual mediante CRON.
- **APP/global.R:** script R que lee las librerías de R necesarias, así como formatea correctamente los campos para la parte visual y lee la tabla de los datos creando una variable data.table que será el core de la app. De esta forma se gana velocidad en la lectura de los datos. Crea las listas que poblarán los componentes del user interface.
- **APP/ui.R:** script que crea la parte visual del dashboard (app) de shiny.
- **APP/server.R:** script base de la APP. Genera todas las llamadas a la librería para la posterior visualización de los distintos elementos.

METODOLOGÍA DEL PROYECTO:

La Shiny App, se divide principalmente en 3 partes. La parte correspondiente a situación geolocalizada (mapas de situación de Estaciones y comportamiento de contaminantes), parte correspondiente a descriptivos (Summary, series temporales, etc.) y la parte de tendencias de comportamiento de los contaminantes.

A parte de la librería de Shiny, se han utilizado las siguientes:

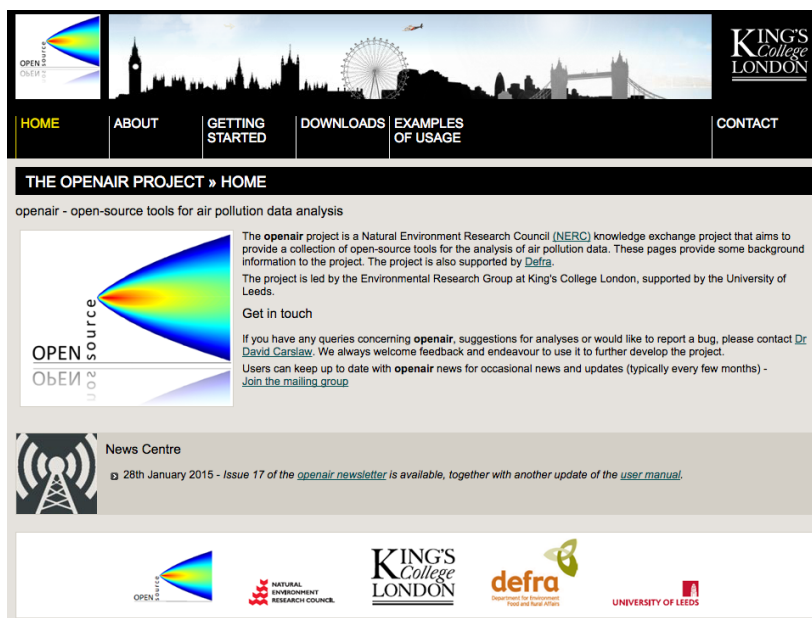
- Shinydashboards (Para el trabajo visual, extensión de shiny).
- Plyr: para trabajo con datos.
- Lubridate: manipulado de fechas.
- ggmap: para geolocalización.
- Openair: Librería base de la App. Especializada en estudio y visualización de contaminantes y estudio del aire en general.

OPENAIR:

La Shiny App, se divide principalmente en 3 partes. La parte correspondiente a situación geolocalizada tanto de estaciones como de contaminantes, la parte descriptiva de comportamiento de contaminantes y las series temporales de los mismos con posibilidad de filtros (provincia, etc.) y la parte de tendencias y correlaciones entre dichos contaminantes.

Se ha seleccionado OPENAIR por ser una librería R cuyo objetivo es facilitar el estudio del aire y la influencia de los contaminantes en el mismo.

La URL de la librería es: <http://www.openair-project.org/Default.aspx> :



BREVE DETALLE DE LAS FUNCIONES UTILIZADAS DE OPENAIR:

- **Summary Plot:** Simplemente es un “añadido” al summary de R, lo que hace es mostrar de una forma más rápida y visual los datos que se obtendrían con dicha función de R.
- **Calendar Plot:** Muestra las concentraciones en formato calendario, por mes y día. Dichas concentraciones aparecen, además, en color para detectar patrones. Está abierto a futuras actualizaciones con la fuerza y dirección del viento.
- **Time Plot:** Muestra series temporales de uno o varios contaminantes para buscar patrones de comportamiento.
- **Time Variation:** Produce cuatro gráficos: variación a lo largo de las horas de cada día, otro con las horas, otro con los meses y otro con los días de la semana. También se muestra el intervalo de confianza del 95% en la media. Estas límites de incertidumbre pueden ser útiles cuando se trata de determinar si un candidato fuente es diferente de otro. Los intervalos de incertidumbre se calculan a través bootstrap re-muestreo, que proporcionará mejores estimaciones que la aplicación de suposiciones basadas en la normalidad, sobre todo cuando hay pocos datos disponibles. La función puede considerar una o dos variables de entrada. Además, existe la opción de normalización en las concentraciones (u otras cantidades). La normalización es muy útil para la comparación de los patrones de dos contaminantes diferentes, que a menudo cubren muy diferente rangos de concentración. La normalización se consigue dividiendo la concentración de un contaminante por su valor medio.
- **Scatter Plot:** Se proporciona para buscar de forma rápida y visual relaciones entre dos contaminantes.
 - Un ajuste suave se agrega automáticamente para ayudar a revelar la relación subyacente entre dos variables junto con los intervalos de confianza del 95% estimado del ajuste. Esto es, en general, una cosa muy útil, ya que ayuda a mostrar la (posiblemente) la relación no lineal entre las variables de una manera muy robusta - o incluso si la relación es lineal.
 - Es fácil de agregar una línea de regresión lineal. La ecuación resultante se muestra en el gráfico junto con el valor de R².
 - Para grandes conjuntos de datos, existe la posibilidad de 'bin' de los datos utilizando las estimaciones de intervalos de densidad kernel o hexagonales. Este enfoque es muy útil cuando hay considerable sobre-trazado.
 - Es fácil demostrar cómo dos variables están relacionadas entre sí depende de los niveles de una tercera variable. Esta capacidad es muy útil para explorar cómo las diferentes variables que dependen unos de otros y pueden ayudar a revelar las relaciones importantes subyacentes.
 - Posibilidad de añadir una tercera variable mediante una escala de colores.
 - Puede manejar / hora formatos de fecha del eje x para proporcionar una forma alternativa de mostrar series de tiempo, que a su vez puede ser un código de colores por una tercera variable

- **Lineal Relation:** Esta función considera las relaciones lineales entre dos contaminantes. Las relaciones se calculan sobre diferentes bases de tiempo utilizando un modelo lineal. La pendiente y el intervalo de confianza del 95% en las relaciones de pendiente por unidad de tiempo se representan gráficamente en varias maneras diferentes. Una regresión lineal es útil para probar la fuerza de la relación. Sin embargo, mucha más información puede ser obtenida considerando diferentes períodos de tiempo, como por ejemplo cómo la relación entre dos contaminantes varían con el tiempo, por día de la semana, diurnas y así sucesivamente. La función `linearRelation` hace exactamente eso - se ajusta a una relación lineal entre dos contaminantes en un amplio rango de períodos de tiempo determinados por período.
- **Trend Level:** proporciona una forma de mostrar rápidamente una gran cantidad de datos de una manera condensada. Es particularmente útil para el trazado del nivel de un valor en contra de dos variables categóricas. Estas variables categóricas pueden pre-existir en un conjunto de datos o estar elaborados sobre la marcha utilizando `openair`. Por defecto se mostrará el valor medio de una variable contra dos variables categóricas, pero también se puede considerar una gama más amplia de estadísticas, por ejemplo, el máximo, la frecuencia, o incluso una función definida por el usuario. La función es mucho más flexible que esto mostrando datos temporales y puede trazar mapas de calor de muchas maneras flexibles. Ambas escalas de color continuas y escalas categóricas definidos por el usuario pueden ser utilizados. La función `trendLevel` muestra cómo el valor de una variable varía en función de intervalos de otras dos variables. El variables `x` e `y` pueden ser categóricos (factor o carácter) o numérico. La tercera variable (`z`) debe ser numérico y es de color de acuerdo a su valor. A pesar de ser llamado `trendLevel` la función es lo suficientemente flexible como para considerar una amplia gama de variables de trazado.
- **Smooth Trend:** calcula tendencias suaves en las concentraciones medias mensuales de contaminantes. En su uso básico va a generar un gráfico de las concentraciones mensuales y ajustar una línea suave a los datos y mostrar los intervalos de confianza del 95% del ajuste. La línea suave se determina esencialmente mediante modelización de aditivos generalizados con el paquete `mgcv`. Este paquete proporciona un conjunto completo y potente de los métodos para el modelado de datos. En este caso, sin embargo, el modelo es una relación entre el tiempo y la concentración de contaminantes es decir, una tendencia. Una de las principales ventajas de este enfoque es que la cantidad de suavidad en la tendencia se ha optimizado en el sentido de que no es ni demasiado suave (por lo tanto faltan características importantes) ni ('ruido' quizá apropiado en lugar de efectos reales) demasiado variable. Parte de la información de antecedentes sobre el uso de este enfoque en un ajuste de calidad de aire se puede encontrar en Carslaw et al. (2007).
- **Theil Sen:** Un método utilizado comúnmente para cálculo de la tendencia en los estudios de la contaminación del aire es el enfoque de Mann-Kendall no paramétrico (Hirsch et al. 1982). Wilcox (2010) ofrece un excelente caso para el uso de métodos modernos para la regresión incluyendo los beneficios de los enfoques no paramétricos y

simulaciones de arranque. Se debe tener en cuenta también que los todos los parámetros de regresión se estiman a través de bootstrap resampling.

El método de Theil-Sen se remonta a 1950, Es uno de los métodos que necesitan gran carga de cómputo. La idea básica es la siguiente. Dado un conjunto de n , pares de xy , se calculan las pendientes entre todos los pares de puntos. (Nota, el número de pistas puede aumentar por $\approx n^2$ de modo que el número de pistas puede aumentar rápidamente a medida que la longitud del conjunto de datos aumenta). La estimación Theil-Sen de la pendiente es la mediana de todas estas pistas. La ventaja del uso del estimador de Theil-Sen es que tiende a producir intervalos de confianza exactos incluso con datos que no son normales y heteroescasticidad (varianza del error no constante). También es resistente a los valores atípicos - ambas características pueden ser importantes en la contaminación del aire. Como se mencionó anteriormente, las estimaciones de estos parámetros se pueden hacer más robusta a través de bootstrap resampling, lo que aumenta aún más la carga computacional, pero no es un problema para la mayoría de series de tiempo que se expresa como medios mensuales o anuales. Bootstrap resampling también proporciona la estimación de p para la pendiente.

Un tema que puede ser muy importante para la serie de tiempo es la dependencia o autocorrelación en los datos. “Normal” asumen que los datos son independientes, pero en series de tiempo esto es raramente el caso. La cuestión es que puntos vecinos de datos son similares a los otros (correlación) y por lo tanto no es independiente. Haciendo caso omiso de esta dependencia tendería a dar una impresión excesivamente optimista de incertidumbres. En openair se sigue la sugerencia de Kunsch (1989) de ajuste de la longitud del bloque de $n/3$, donde n es la longitud de la serie temporal.

DESCRIPCIÓN/MANUAL APP AIRCYL

La aplicación está accesible desde las siguientes URL's:

- vps.valdeolmillos.eu
- <https://jmvaldeolmillos.shinyapps.io/AirCyL/>

A continuación mostramos un breve manual sobre el uso de la Aplicación AIRCyL.

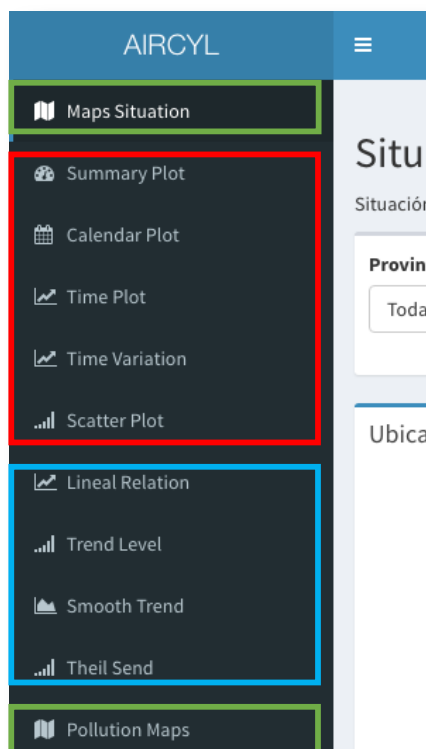
La aplicación se divide en dos partes principales:

- Zona de menú.
- Zona de visualización.

ZONA DE MENÚS

Nos encontramos 3 tipos de apartados principales:

- Zona de Geolocalización (Verde)
- Zona descriptiva (Rojo)
- Zona de tendencias (Azul)



ZONA DE VISUALIZACIÓN

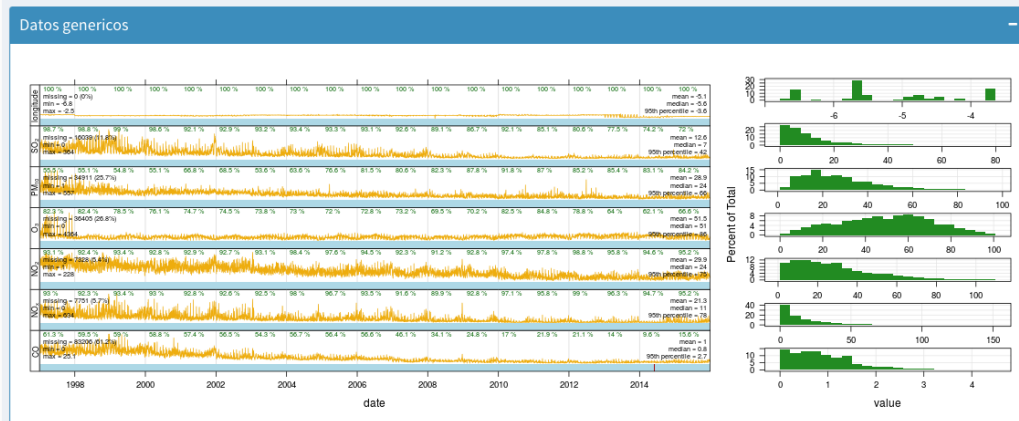
Es la parte central de la página. En ella encontramos tanto los selectores de selección de variables para filtrar las visualizaciones.

Valores de calidad del aire generales.

Datos agregados para todas las estaciones.

Muestra la distribución de datos para todos los contaminantes en el tiempo. Color amarillo corresponde a NA's, es decir no existe medida. Muestra valores medios. Utilizado para detección de tendencias.

Límites legales para NO y NO2 son: 40 microgramos/m3. y del O3 es de 120 microgramos/m3.



Selección de Provincia:

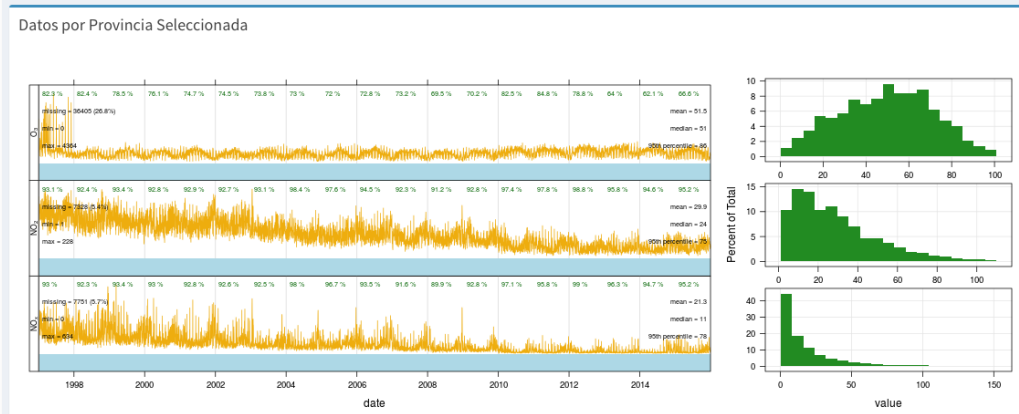
Todas

Valores de los Contaminantes principales por Provincia.

Datos agregados para todas las estaciones.

Muestra la distribución de datos para todos los contaminantes en el tiempo. Color amarillo corresponde a NA's, es decir no existe medida. Muestra valores medios. Utilizado para detección de tendencias.

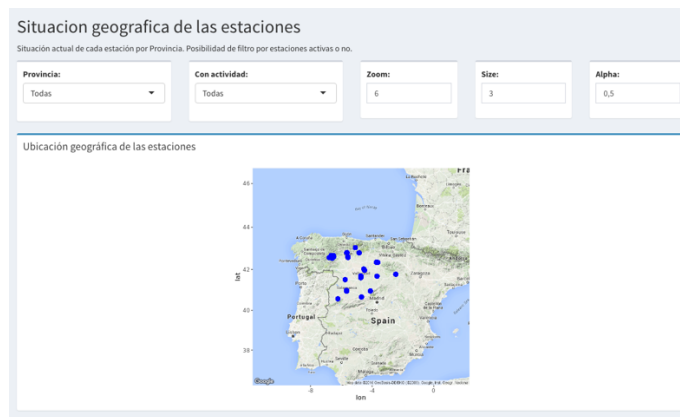
Límites legales para NO y NO2 son: 40 microgramos/m3. y del O3 es de 120 microgramos/m3.



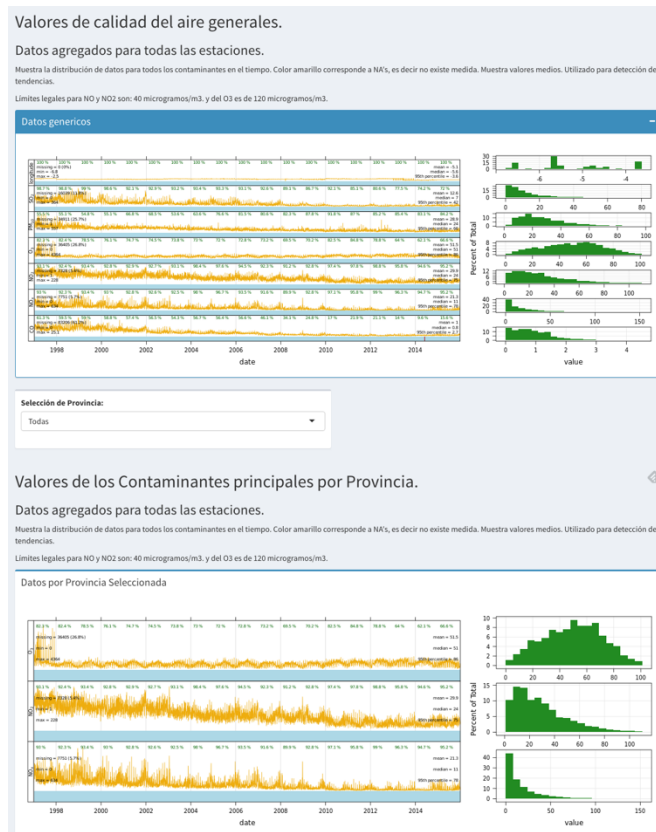
DESCRIPCIÓN DE CADA OPCIÓN DEL MENÚ

Las posibles opciones de selección son las siguientes:

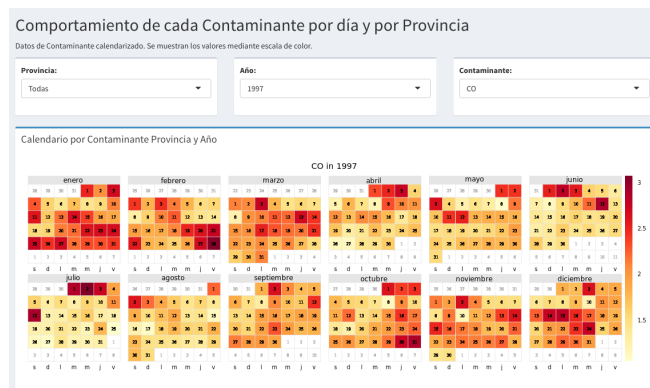
- **MAP SITUATION:** Muestra en un mapa la situación de cada Estación. Se puede filtrar por: Provincia y si poseen actividad o no. También hay opciones de cambiar el zoom, el tamaño de los puntos y la transparencia de los mismos de las estaciones en el mapa.



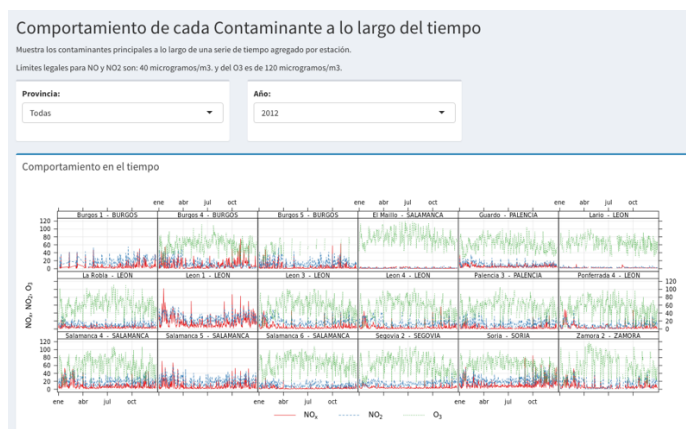
- **SUMMARY PLOT:** Muestra estadísticos típicos: máxima, mínimo, medianas, percentil 95, etc. Además añade la visualización de la serie temporal de los datos. En la parte superior nos muestra los datos de forma agregada. En la parte inferior se puede seleccionar la provincia y nos muestra los contaminantes principales a tener en cuenta en un análisis de calidad de los datos. A la derecha de cada gráfica temporal aparecen los histogramas de distribución de cada uno de los contaminantes.



- **CALENDAR PLOT:** Muestra las concentraciones en formato calendario, por mes y día. Dichas concentraciones aparecen, además, en color para detectar patrones. Está abierto a futuras actualizaciones con la fuerza y dirección del viento. Nos permite filtrar por provincia, año y contaminante.



- **TIME PLOT:** Este gráfico está pensado para trazar rápidamente series temporales de datos, mostrando los 3 contaminantes principales de análisis. Esto es, o debería ser, una tarea muy común en el análisis de la contaminación del aire. De este modo, es útil ser capaz de trazar varios contaminantes al mismo tiempo. Es esta pantalla, el filtrado de datos para estudio es por provincia y por año, mostrándonos cada una de las estaciones de medición.

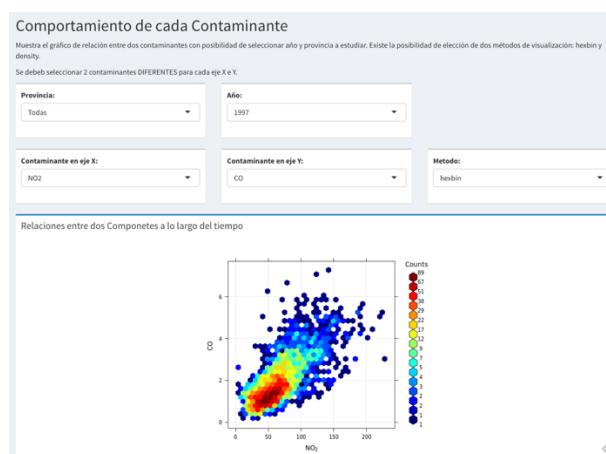


- **TIME VARIATION:** Muestra cuatro gráficos. Dos de ellos se encuentran libres para uso futuro, pues analiza los datos a nivel de hora del día (actualmente no hay datos, pero se ha solicitado dicha granularidad para posteriores actualizaciones de microdatos). Los gráficos nos muestran: Contaminantes a nivel de hora por día de la semana; contaminantes a nivel de hora; contaminantes a nivel de mes y contaminantes a nivel de día de la semana. Esta visualización se puede utilizar para comparar una parte de una serie de tiempo con otra. Esto es a menudo una cosa muy poderosa para hacerlo, sobre todo si las concentraciones son normalizadas. Por ejemplo, a menudo hay interés en

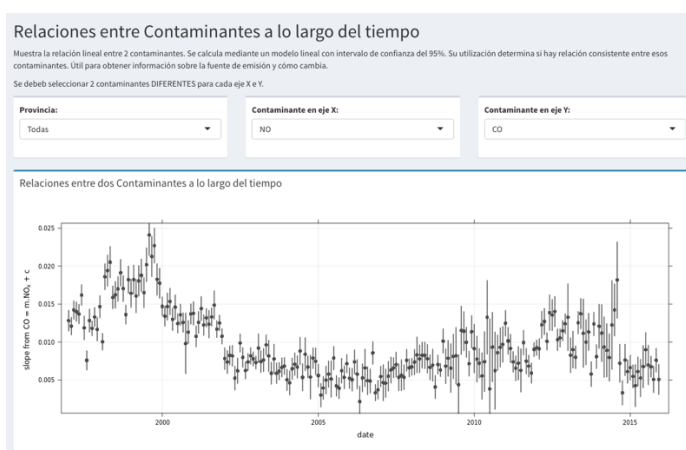
saber cómo los patrones estacionales varían con el tiempo. Si un contaminante mostró signos de un aumento en los últimos años, a continuación, dividir el conjunto de datos y la comparación de cada parte en conjunto pueden proporcionar información sobre lo que está impulsando el cambio. ¿Hay, por ejemplo, la evidencia de que las concentraciones de hora punta de la mañana han vuelto más importantes, o los domingos se han vuelto relativamente más importantes? También se muestra el intervalo de confianza del 95% en la media. Se pueden filtrar los datos a nivel de provincia. COMENTAR QUE DICHO GRÁFICO POSEE UN RETARDO EN CARGAR pues implica un alto procesamiento de todas las series temporales.



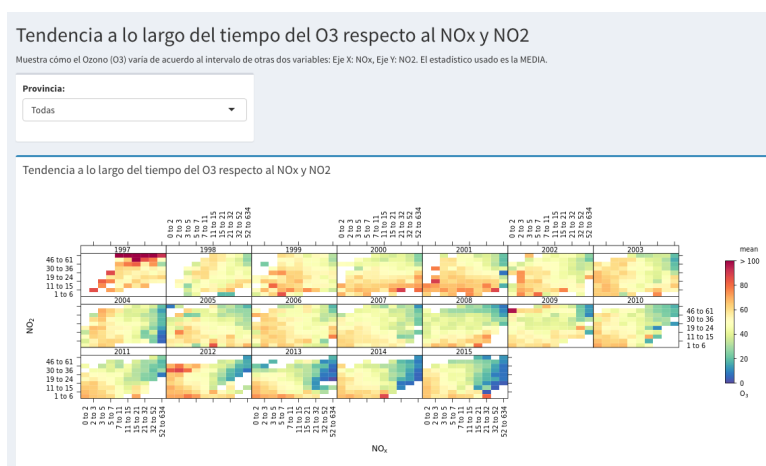
- **SCATTER PLOT:** Los gráficos de dispersión son extremadamente útiles y una técnica de análisis muy comúnmente usado para la consideración de cómo las variables se relacionan entre sí. El propósito de este gráfico es hacer que sea sencillo considerar cómo las variables están relacionadas entre sí de una manera consistente. Los filtros que se pueden aplicar son provincia y año de estudio. Se debe elegir los dos contaminantes a estudiar, no pudiendo ser el mismo a la vez. También comentar que hay dos formas de tratar los datos (sobre todo con grandes datasets): hexbin y density (para más información se debe consultar manual de Openair, pags. 187-191).



- LINEAL RELATION:** Muestra la relación lineal entre dos contaminantes. Las relaciones se calculan sobre diferentes bases de tiempo utilizando un modelo lineal. La pendiente y el intervalo de confianza del 95% en las relaciones de pendiente por unidad de tiempo se representan gráficamente en varias maneras diferentes. Una regresión lineal es útil para probar la fuerza de la relación. El gráfico muestra un ajuste a una relación lineal entre dos contaminantes en un amplio rango de períodos de tiempo determinados por período. Como filtros tenemos la posibilidad de desagregar los datos por provincia y también seleccionar los contaminantes que queremos estudiar. Teniendo en cuenta que no se puede seleccionar el mismo para ambos ejes.



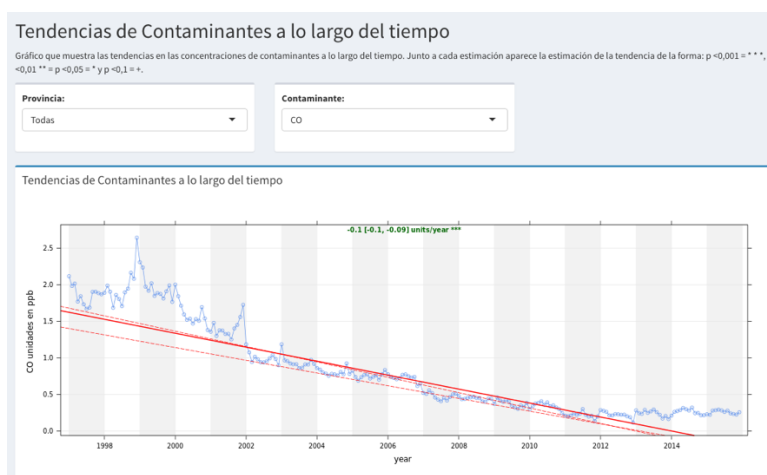
- TREND LEVEL:** proporciona una forma de mostrar rápidamente una gran cantidad de datos de una manera condensada. Es particularmente útil para el trazado del nivel de un valor en contra de dos variables categóricas. Por defecto se mostrará el valor medio del Ozono O3 contra dos variables de contaminantes (NOx y NO2). Lo que muestra el gráfico es cómo el valor de una variable varía en función de intervalos de otras dos variables. El variables x e y son los contaminantes a seleccionar. En este gráfico existe la posibilidad de filtrar por provincia. Dando los datos agregados para todas las estaciones de medición.



- **SMOOTH TREND:** Esta opción calcula tendencias suaves en las concentraciones medias mensuales de contaminantes. Se genera un gráfico de las concentraciones mensuales y ajusta una línea suave a los datos y muestra los intervalos de confianza del 95% del ajuste. Podemos realizar un filtro por provincia a estudiar.



- **THEIL SEN:** Esta opción se utiliza normalmente para determinar las tendencias en las concentraciones de contaminantes durante varios años. Se calcula los valores medios mensuales de todos los días, cada hora o los datos de mayor resolución de tiempo, además de trabajar directamente con medios mensuales. Si tiene sentido para calcular tendencias durante determinados periodos de tiempo más cortos (por ejemplo 2 años) depende en gran medida de los datos. Es muy posible que las tendencias estadísticamente significativas se pueden detectar durante períodos relativamente cortos, pero es otro asunto si tiene importancia. Debido a los efectos estacionales pueden ser importantes para los datos mensuales. Tenga en cuenta también los símbolos que aparecen junto a cada estimación de la relación de tendencia. La estimación de la tendencia es: $p < 0,001 = ***$, $p < 0,01 = **$, $p < 0,05 = *$ y $p < 0,1 = +$. Las opciones de filtro para esta opción son provincia y contaminante a estudiar.



- **POLLUTION MAPS:** Mediante la selección del contaminante, se ofrecen tres tipos de mapas que muestran de forma agregada en toda la Comunidad los siguientes datos:
 - Medición Anual de Contaminante por Provincia.
 - Media de Contaminante por Provincia y estación del año.
 - Máxima de Contaminante por Provincia y Estación del año.

