



Clab: Innovación en procesos de fabricación circular y abierta para entornos educativos

N.º Expediente: AEI-010500-2022b-177

PT2-E3: Herramientas de análisis de datos

Índice

Índice	2
Introducción	3
Herramientas	4
Herramienta web.....	4
Programación visual.....	7
Programación creativa.....	14
Análisis de datos avanzado.....	18

Introducción

Este entregable detalla las herramientas de visualización y análisis de datos ambientales desarrolladas durante el proyecto, que acompañan al kit de sensores descrito en PT2-E5 para facilitar el análisis de los datos provenientes de los mismos u otros sensores similares. En su conjunto, estas herramientas permiten introducir al alumnado a la visualización y al análisis de datos de forma progresiva, con diferentes enfoques, detallando posibles temáticas relacionadas, desde las plataformas online de datos, hasta la programación visual o creativa para la visualización de datos a través de métodos no formales. Por último, como herramienta más avanzada, se incluye un paquete de análisis escrito en *Python*, que permite la introducción a una herramienta real para niveles más avanzados.

La Tabla 1 resume las herramientas desarrolladas, incluyendo sus posibilidades, el nivel de complejidad y el entorno. En cuanto a las posibilidades de cada herramienta, se identifican las siguientes:

- Tiempo real: si la herramienta permite la visualización de datos en tiempo real (o casi tiempo real) sin intervención del usuario
- Visualización de datos: si es posible la visualización de datos, y en qué formato (*SI* indica cualquier formato)
- Comparativa entre dispositivos: si la herramienta permite realizar comparativas entre dispositivos de sensores
- Nivel del análisis de datos posibles: relacionado con la posibilidad de hacer análisis a diferentes niveles de complejidad

Herramienta	Posibilidades				Complejidad (1-3)	Entorno
	Tiempo real	Visualización	Comparativa entre dispositivos	Nivel del análisis de datos		
Dashboard	SI	Sólo, series temporales	NO	BÁSICO	1	Web
Orange	NO	SI	SI	INTERMEDIO	2	Programación visual
Processing	SI	SI	-	-	2	Programación creativa
sodata	NO	SI	SI	SI	3	Programación clásica

Tabla 1. Resumen de herramientas de datos

El resumen de la Tabla 1 muestra la complementariedad de las herramientas, permitiendo aquellas más sencillas hacer un análisis o visualización más básico, mientras que progresivamente, a medida que incrementa la complejidad, aumentan las posibilidades análisis de datos más complejos. En los siguientes apartados se describirán cada una de las herramientas.

Herramientas

Herramienta web

Palabras Clave

web, visualización, serie temporal, nivel básico

Descripción

Esta herramienta permite visualizar de forma básica los datos en series temporales de un dispositivo a través de un navegador web.

En relación al resto, esta herramienta se plantea como la introducción rápida a la visualización de los datos de forma sencilla. Al ser una herramienta web de código abierto, su código además permite tratar temas relacionados a las aplicaciones web, páginas dinámicas o plataformas de datos de forma básica.

Características

Esta herramienta cuenta con dos vistas: la primera permite una visualización rápida del estado de los dispositivos, mientras que la segunda es una visualización sencilla de los datos en una serie temporal. Además, permite descargar los datos en formato CSV para su análisis en otras aplicaciones a posteriori. Las vistas se detallan a continuación:

- *Resumen de dispositivos*: representada en la Figura 1, muestra una lista en forma de cuadrícula con diferentes dispositivos, el cual se puede filtrar por *ciudad*, o *etiquetas*. El color del recuadro indica si el dispositivo está enviando datos a la plataforma actualmente o no. El resumen de dispositivos, con los metadatos disponibles a través de la plataforma, se puede descargar a través del botón *Get this data*. Además, se pueden filtrar los dispositivos a través del cuadro de búsqueda superior, o presionando en los diferentes campos de los recuadros (para filtrar a través de la ciudad, o de las etiquetas del dispositivo). La vista de detalle se accede al presionar sobre el fondo del recuadro de un dispositivo.

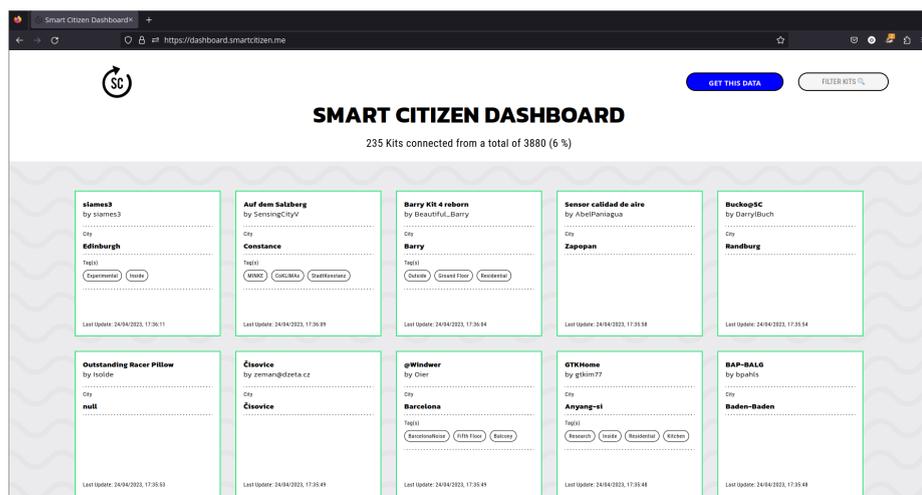


Figura 1. Resumen de dispositivos

- **Detalle de dispositivo:** Esta vista permite visualizar datos de forma básica en una serie temporal, o en una vista reducida con el último dato disponible. La vista es adaptable a diferentes tamaños de pantalla, y tiene un botón de configuración en la parte superior para filtrar las fechas, definir si la vista debe o no actualizarse con nuevos datos, o eliminar sensores de la visualización. Por último, los datos pueden descargarse en formato CSV a través del botón *Get this data*.

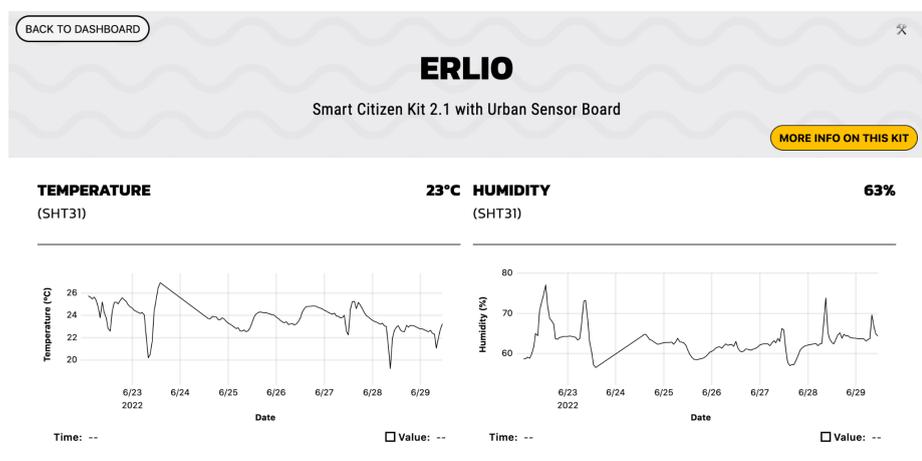


Figura 2. Detalle de dispositivo - ampliada

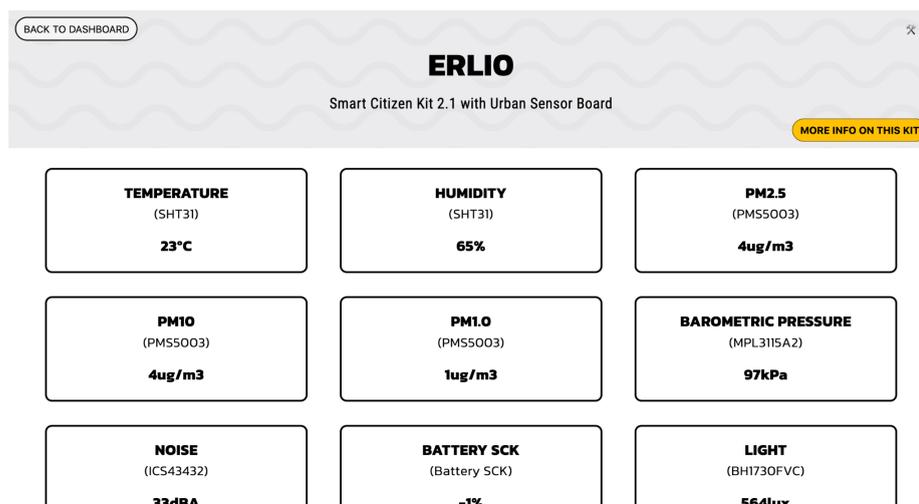


Figura 3. Detalle de dispositivo - reducida

Recursos

El código fuente de esta aplicación web se encuentra disponible en un repositorio abierto online¹. El dashboard es accesible a través de <https://dashboard.smartcitizen.me>

Una descripción detallada del uso de esta herramienta se encuentra en X.

¹ Repositorio abierto de la aplicación web: <https://github.com/fablabbcn/smartcitizen-js-dashboard/>

Programación visual

Palabras Clave

aplicación, visualización, serie temporal, nivel intermedio

Descripción

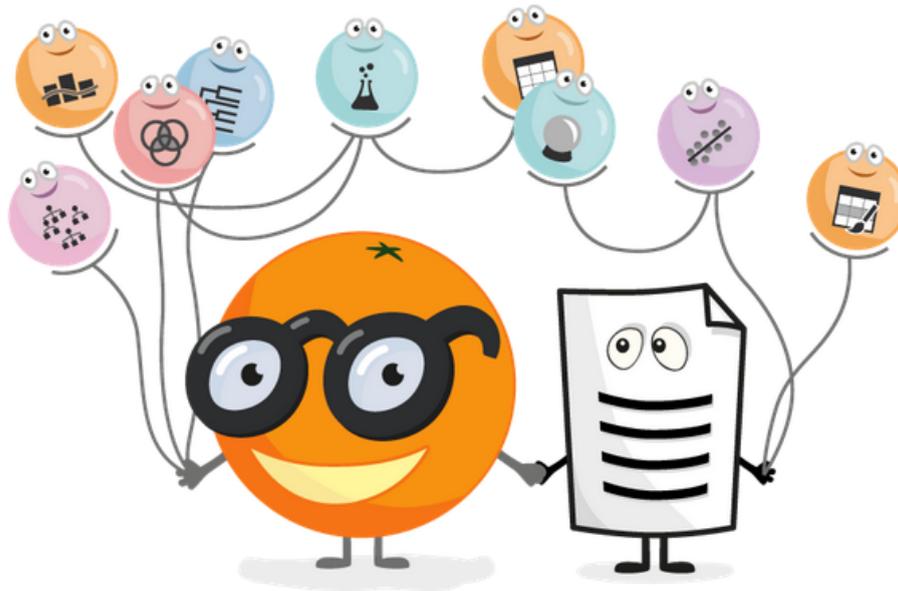


Figura 4. Ilustración del proyecto Orange Data Mining. Origen: Orange Data Mining

Esta herramienta es un módulo de análisis de datos enfocado a la interacción con la plataforma *Smart Citizen*², integrado en el entorno del proyecto *Orange Data Mining*. *Orange Data Mining* es un proyecto source para análisis de datos escrito en *Python*, basado en nodos que intercambian datos y realizan acciones relacionadas a la carga, procesado, análisis y visualización de los mismos. Adicionalmente a los principios abiertos del proyecto, es un software fácilmente instalable en múltiples plataformas, y no es necesario ningún otro complemento para su utilización y se puede descargar gratuitamente desde su página web³.

En concreto, el módulo programado es una integración en el repositorio *MECODA*⁴ (*Module for Citizen Observatory Data Analysis - Módulo para el análisis de datos de observatorios ciudadanos*), desarrollado en el contexto del proyecto *cos4cloud*⁵ (EU HORIZON 863463). Este repositorio pretende facilitar y analizar todo tipo de datos de ciencia ciudadana para el público general. Dada la presencia de diversos módulos existentes en este repositorio, se considera que la contribución a este a través de *CLAB* puede generar sinergias y una mayor versatilidad de las herramientas propuestas.

² Web de la plataforma Smart Citizen: <https://api.smartcitizen.me/>

³ Web de Orange Data Mining: <https://orangedatamining.com/>

⁴ Web del repositorio MECODA: <https://github.com/eosc-cos4cloud/mecoda-orange>

⁵ Web del proyecto cos4cloud: <https://cos4cloud-eosc.eu/>

Características

El módulo de *Smart Citizen* puede interactuar con la plataforma, para recuperar datos de cualquier *Smart Citizen Kit* que hayan sido subidos a la web. Adicionalmente, puede cargar datos desde una copia local en formato *.csv*.

Para ello utiliza una serie de nodos de programación visual que se conectan entre sí, generando un *workflow*. Las líneas de conexión representan flujos de datos, haciendo que el conjunto sea intuitivo desde el punto de vista didáctico, un ejemplo se muestra en la Figura 5.



Figura 5. Ejemplo de workflow en Orange Data Mining

Al configurar diferentes conjuntos de widgets, se pueden procesar datos y generar visualizaciones en distintos formatos. Un ejemplo de visualización de datos en formato de serie temporal se muestra en la Figura 6:

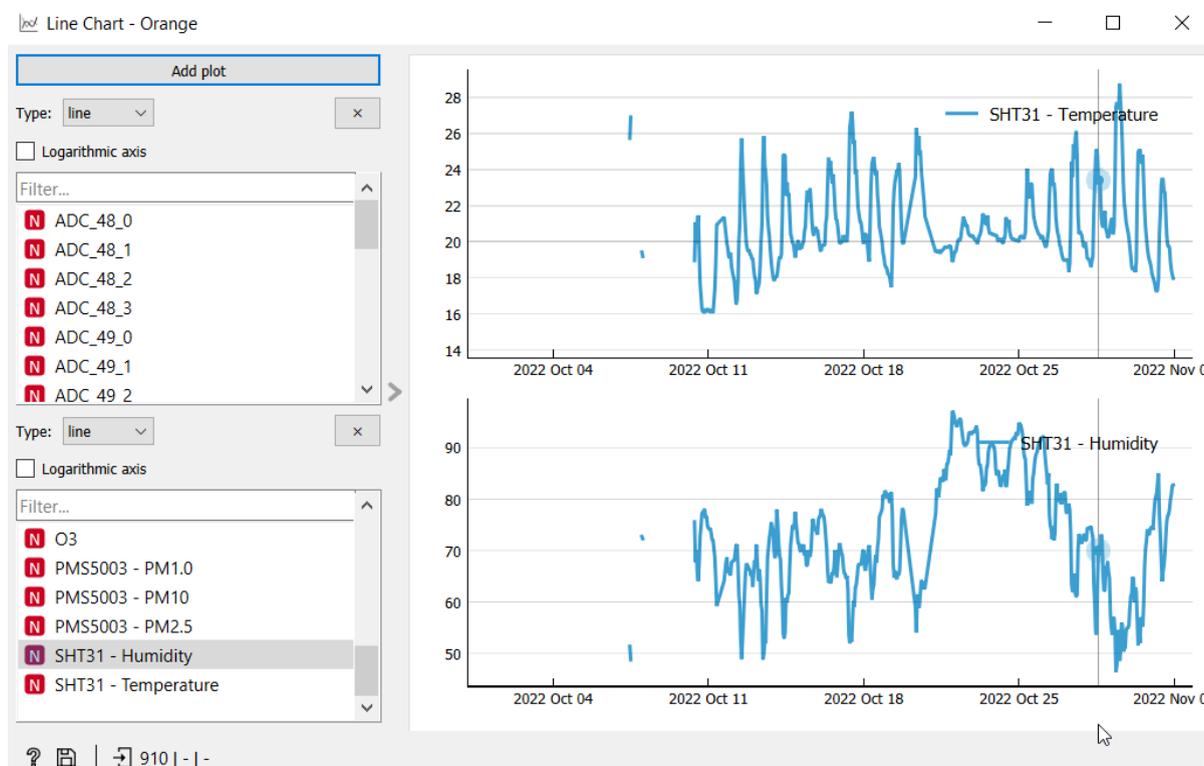


Figura 6. Ejemplo de visualización de series temporales en Orange Data Mining

Otro ejemplo, para datos geocalizados, se muestra en la Figura 7:

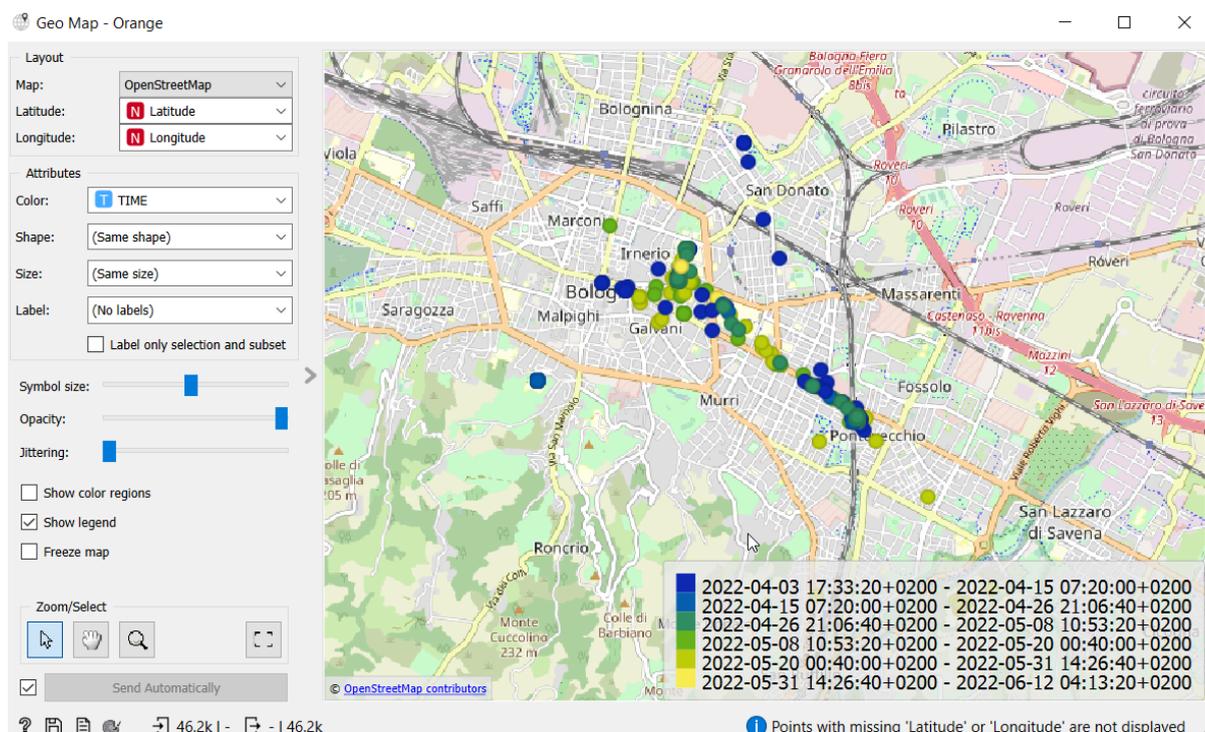


Figura 6. Ejemplo de visualización de datos geocalizados en Orange Data Mining

La contribución a este proyecto a nivel de módulo permite:

1. La integración en un sistema modular, con un *núcleo* de código abierto activamente mantenido, garantizando su soporte a largo plazo y la adaptación a nuevas plataformas y formatos.
2. La accesibilidad a nivel didáctico de la componente visual de la programación
3. La inclusión en el mismo entorno herramientas accesibles, pero con diferentes niveles de complejidad, para explorar los datos desde la perspectiva temporal o geocalizada
4. La integración de otras plataformas de datos de otros proyectos ambientales similares a Smart Citizen

En concreto, el módulo se compone fundamentalmente de dos nodos, representados en la Figura 7:

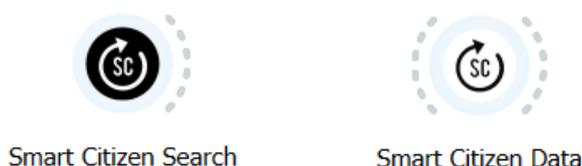


Figura 7. Nodos contribuidos para la integración de Smart Citizen en el proyecto MECODA / Orange

1. Nodo para la búsqueda de datos
2. Nodo para la descarga de datos, en función de los resultados del primero

El nodo para la búsqueda de datos, cuya interfaz se muestra en la Figura 8, permite realizar una búsqueda en la plataforma *Smart Citizen* a través de su *API*, y recuperar dispositivos que cumplan una serie de requisitos. Por ejemplo, permite recuperar todos los dispositivos que están ubicados en una determinada ciudad, o que han capturado datos durante un cierto periodo de tiempo, o un tipo particular de datos. Este nodo interactúa con la plataforma y retorna al usuario una tabla con los dispositivos que cumplen estas condiciones, así como sus metadatos.

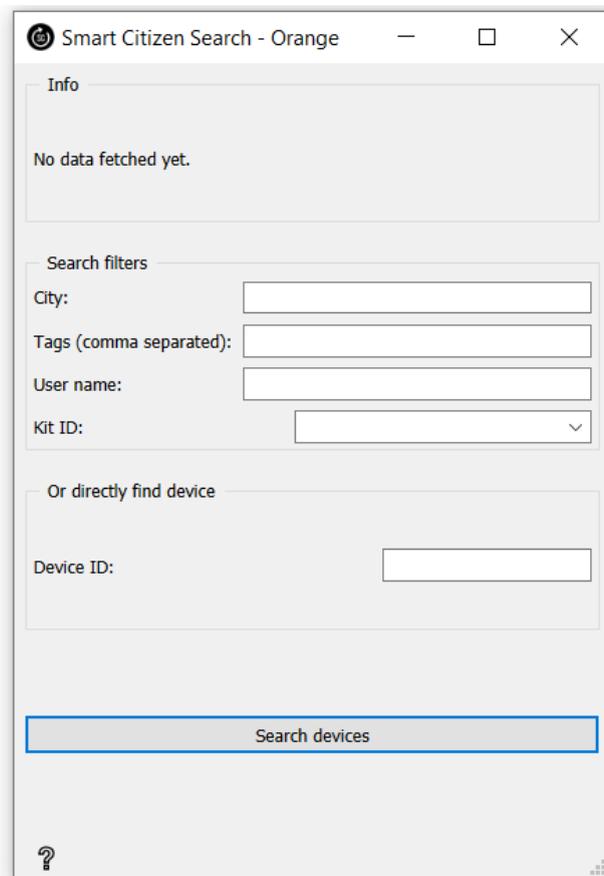


Figura 8. Interfaz principal del nodo de búsqueda de datos de Smart Citizen

El nodo para la la descarga de datos permite, a partir de un determinado dispositivo, descargar los datos asociados al mismo desde la plataforma *Smart Citizen* a través de su *API*, en formato tabular. Gracias a la modularidad de *Orange*, estos datos pueden, a posteriori, ser objeto de cualquier tipo de análisis ya implementado en *Orange* o en cualquier otro módulo, como una visualización de los datos, o su procesado en algoritmos de *machine learning*. La interfaz principal se muestra en la Figura 9, y el resultado de los datos en la Figura 10:

Smart C... — □ ×

Info

Device: 15618
Name: I-CHANGE SAN TEODORO
City: Genoa (IT)
By: ichangecima

Get data at a specific frequency

Rollup: 10

Rollup units: m

Filter by date (YYYY-MM-DD)

Initial Date:

End Date:

Resample data

Get data

?

Figura 9. Interfaz principal del nodo de descarga de datos de Smart Citizen

Data Table - Orange

Info
108 instances
19 features (0.1 % missing data)
No target variable.
No meta attributes.

Variables
 Show variable labels (if present)
 Visualize numeric values
 Color by instance classes

Selection
 Select full rows

Restore Original Order

Send Automatically

	tion	Horizontal Speed	ICS43432 - Noise	Latitude	Longitude	SA2 - Barometric P	PM ^
1	275	0.112	45.082	44.4813	11.3777	100.258	
2	589	3.81078	72.2372	44.487	11.3656	100.336	
3	6.31	1.98806	66.2389	44.4947	11.3544	100.415	
4	2.6	0.135	51.613	44.495	11.3528	100.339	
5	60.7	0.172	46.369	44.4949	11.3528	100.348	
6	60.5	0.3	36.564	44.4949	11.3529	100.334	
7	8.5	0.205	49.566	44.495	11.3528	100.349	
8	7.7	0.172	42.365	44.4949	11.3529	100.365	
9	220	0.306	50.449	44.4949	11.3529	100.373	
10	269	0.431	41.357	44.4949	11.3529	100.374	
11	5.5	0.431	38.055	44.495	11.3529	100.37	
12	333	0.343333	42.2256	44.4951	11.3529	100.377	
13	525	0.626875	46.9263	44.4951	11.3528	100.377	
14	1.19	0.527619	51.1733	44.4951	11.3528	100.376	
15	7.8	0.163	48.905	44.4948	11.3529	100.37	
16	11.3	0.138	48.519	44.4948	11.3529	100.37	
17	251	0.241	36.729	44.4949	11.3528	100.37	
18	11.7	0.295	37.796	44.4949	11.3528	100.373	
19	13.4	0.253	34.817	44.4949	11.3528	100.384	
20	4.1	0.175	34.922	44.4949	11.3528	100.396	
21	6.2	0.325	38.084	44.495	11.3528	100.39	

108 1 | 108

Figura 10. Resultado de la descarga de datos temporales de un dispositivo

Recomendaciones adicionales

Orange Data Mining, al ser una aplicación multiplataforma se puede instalar en casi cualquier dispositivo. Sin embargo, la instalación y la configuración del programa tiene una duración aproximada de media hora, por lo que se recomienda prever tiempo adicional a las actividades con el alumnado para instalarlo y configurarlo previamente.

Recursos

Material adicional del proyecto *Orange Data Mining*:

- Tutoriales de *Orange Data Mining* (youtube oficial):
<https://www.youtube.com/channel/UCIKKWBe2SCAEyv7ZNGHle4g>
- Catálogo de *Widgets*:
<https://orangedatamining.com/widget-catalog/>

Guías específicas de *Smart Citizen*:

- Configuración de *Orange Data Mining* y *Add-ons* de *Smart Citizen Kit*:
<https://docs.smartcitizen.me/Resources/Tutorials/Configure%20Orange%20Data%20Analysis/>
- Guía de los *Widgets* de *Smart Citizen Kit*:
<https://docs.smartcitizen.me/Guides/data/Orange%20Data%20Widgets/>

- Tutorial de *workflow* de *Orange Data Mining* utilizando datos de *Smart Citizen Kit*:
<https://docs.smartcitizen.me/Resources/Tutorials/Simple%20Orange%20Workflow/>
- Tutorial de *workflow* de *Orange Data Mining* utilizando datos geolocalizados de *Smart Citizen Kit*:
<https://docs.smartcitizen.me/Resources/Tutorials/Orange%20and%20geolocated%20data/>
- Creación de medias y derivadas de datos de *Smart Citizen Kit* con *Orange Data Mining*:
<https://docs.smartcitizen.me/Resources/Tutorials/Averages-and-derivatives-in-Orange/#get-data>
- Repositorio de ejemplos:
<https://github.com/fablabbcn/smartcitizen-docs/tree/master/docs/assets/ows>

Programación creativa

Palabras Clave

visualización, tiempo real, programación creativa, Java, JavaScript

Descripción

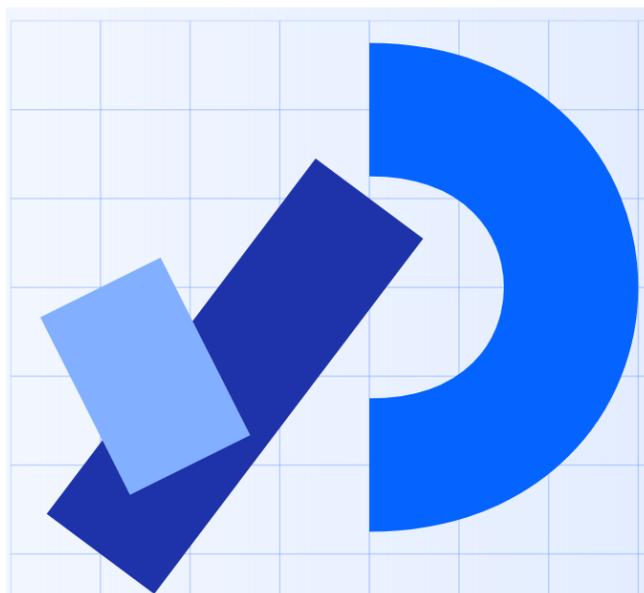


Figura 11. Ilustración del proyecto Processing. Origen: processing.org

Este apartado detalla un conjunto de ejemplos elaborados para la programación creativa a través de entorno del proyecto *Processing*⁶. *Processing* es un proyecto de código abierto que consiste en un lenguaje de programación y entorno de desarrollo integrado de código abierto basado en *Java*, de fácil utilización, y que sirve como medio para la enseñanza y producción de proyectos multimedia e interactivos de diseño digital. Similar a *Orange Data Mining*, se puede utilizar de manera gratuita y es muy fácil de configurar. Adicionalmente a *Processing*, una derivación del mismo entorno, pero para entornos web, es el proyecto *p5*⁷. *p5* es una librería de *JavaScript* que también implementa estos mismos principios y se puede utilizar desde el navegador.

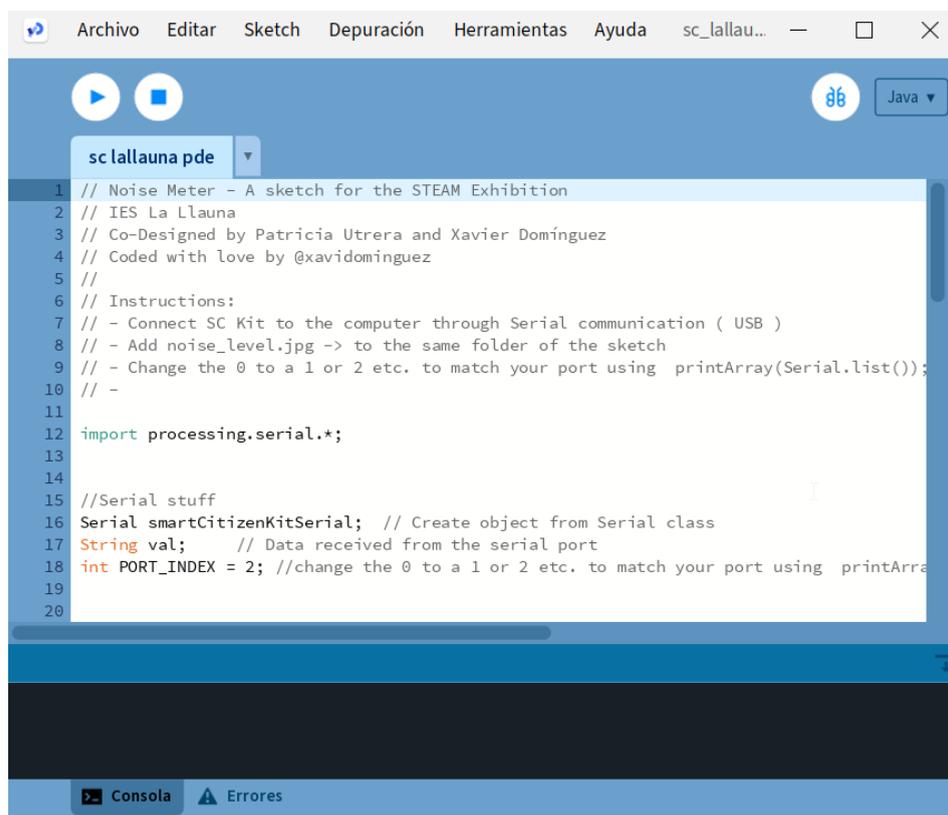
Ambos, tanto *Processing*, como *p5*, pueden utilizarse con múltiples objetivos y diferentes niveles de complejidad, tanto para generar programas sencillos, que permitan enseñar paradigmas de programación o características específicas de lecturas de los sensores, o proyectos más complejos que interactúan con un *API* o se comunican en tiempo real con el sensor.

⁶ Web del proyecto processing: <https://processing.org/>

⁷ Web del proyecto p5js: <https://p5js.org/>

Características

El entorno de programación (*IDE*) incluido en *Processing*, si bien no es tan complejo como un entorno de Java completo, requiere cierto nivel de tecnificación bien por el docente o por el alumnado que lo vaya a utilizar. Adicionalmente, el IDE es muy similar al *IDE* de *Arduino*⁸, por lo que docentes y alumnos familiarizados con este último encontrarán este entorno muy sencillo de utilizar.



```

1 // Noise Meter - A sketch for the STEAM Exhibition
2 // IES La Llauna
3 // Co-Designed by Patricia Utrera and Xavier Domínguez
4 // Coded with love by @xavidominguez
5 //
6 // Instructions:
7 // - Connect SC Kit to the computer through Serial communication ( USB )
8 // - Add noise_level.jpg -> to the same folder of the sketch
9 // - Change the 0 to a 1 or 2 etc. to match your port using printArray(Serial.list());
10 // -
11
12 import processing.serial.*;
13
14
15 //Serial stuff
16 Serial smartCitizenKitSerial; // Create object from Serial class
17 String val; // Data received from the serial port
18 int PORT_INDEX = 2; //change the 0 to a 1 or 2 etc. to match your port using printArra
19
20

```

Figura 12 Ejemplo del IDE de Processing

En concreto, este apartado proporciona librerías para la utilización de *Processing*, en el campo de la programación creativa, con datos ambientales. En concreto, se proporcionan herramientas y material didáctico para obtener datos:

- A través de puerto *USB*, obteniendo datos en tiempo real para representarlos en el momento.
- A través de la llamada a la *API* de Smart Citizen, lo cual permite recoger datos de la plataforma para luego procesarlos y dibujarlos.
- A través de un archivo *CSV* utilizando, utilizando datos offline descargados en la tarjeta *SD* del dispositivo

⁸ Web del IDE de Arduino: <https://www.arduino.cc/en/software>

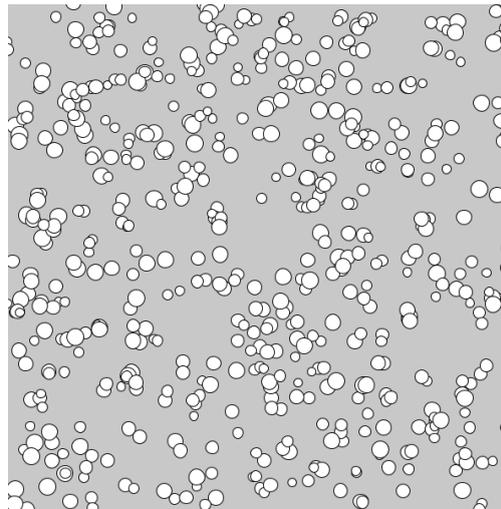


Figura 13. Ejemplo del resultado de un código que genera partículas que reaccionan el nivel de contaminación en tiempo real

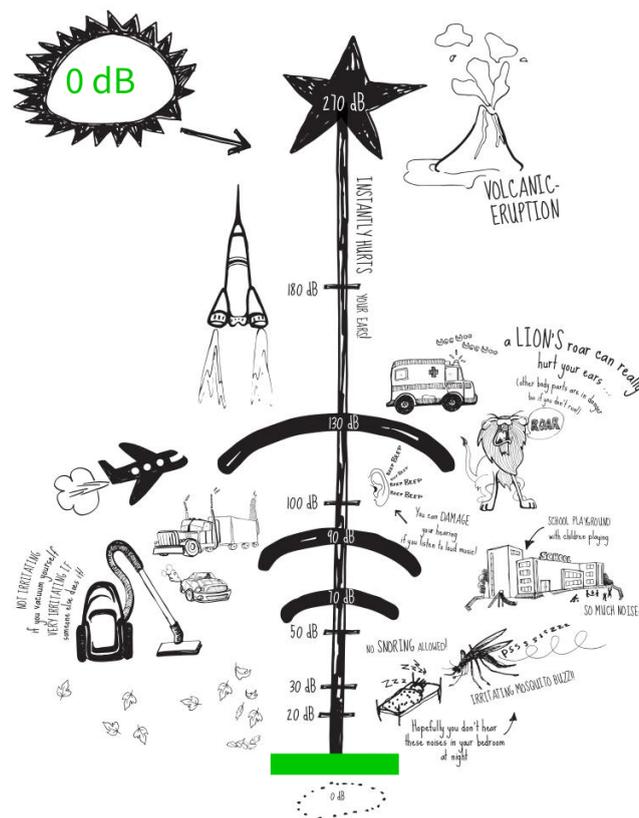


Figura 14. Ejemplo preparado con sistema que evalúa el sonido y lo pone en una escala de medida visual.

Recursos

- Tutoriales de *Processing*:
<https://happycoding.io/tutorials/processing/>
<https://processing.org/tutorials>
- Configuración de *Processing* y ejemplo utilizando datos en tiempo real:
<https://docs.smartcitizen.me/Resources/Tutorials/Setting-up-Processing-and-using-Serial/>
- Ejemplo de *Processing* utilizando *API*:
<https://docs.smartcitizen.me/Resources/Tutorials/How-to-call-the-Smart-Citizen-API-using-Processing/>
- Repositorio de ejemplos:
<https://github.com/fablabbcn/smartcitizen-docs/blob/master/docs/assets/pde/>

Análisis de datos avanzado

Palabras Clave

visualización, análisis de datos, python, R, plataformas de datos

Descripción

Esta herramienta son librerías que permiten la interacción con el API de *Smart Citizen*, así como otros orígenes de datos abiertos, de forma programática y en un entorno de programación *Python* o *R*. Gracias a su integración como librería de estos lenguajes, permite utilizar todas los recursos ya disponibles para análisis estadístico y visualización de datos dentro de los mismos, garantizando su escalabilidad. Ambas librerías son de código abierto y se encuentran disponibles en sus respectivos repositorios online⁹.

Características

Ambas librerías cuentan con una interacción por defecto con el API de Smart Citizen para la descarga de datos de la misma para su posterior procesado. Adicionalmente, la librería en *Python*, proporciona herramientas avanzadas para el análisis de datos desde el punto de vista de modelado, calibración, visualización en el entorno de desarrollo *Jupyter*¹⁰, el cual ofrece una interfaz más amigable para la programación en estos lenguajes.

Dada la extensa documentación existente al respecto de estas herramientas, se referencian algunos ejemplos en la parte inferior para mostrar las posibilidades de la misma, pero referido al lector la documentación en la parte de recursos.

id	name	description	owner_id	owner_username	latitude	longitude	city	country_code	kit_id	state	system_tags	user_tags
4365	Giant Spark Eggnog	Making Sense Pilot #1	6066	Marabales	41.396570	2.194473	Barcelona	ES	11.0	has_published	[offline, outdoor]	[Barcelona, MakingSense, Inside, First Floor, ...]
4140	UA_FA	Departamento I+D+i Geomática (s)	5696	sck.fablab.atacama	-23.701574	-70.420252	Antofagasta	CL	3.0	has_published	[offline, outdoor]	[Research, Experimental]
3618	ISP01	None	5452	ISP01	41.374510	2.100678	Esplugues de Llobregat	ES	3.0	has_published	[offline, outdoor]	[Barcelona]
2451	ZKM-Citizen-KIT	Smart Citizen Kit located in Karlsruhe inside ...	3921	ZKM4Karlsruhe	49.006890	8.403653	Karlsruhe	DE	3.0	has_published	[indoor, offline]	[]
4135	Carlos FA	Casa feliz de Carlos: D	5696	sck.fablab.atacama	-23.656478	-70.402443	Antofagasta	CL	3.0	has_published	[offline, outdoor]	[Research, Experimental]
12514	Solar-Flash test	Smart Citizen Kit 2.1 with Urban Sensor Board	5105	vico	41.438206	2.172849	Barcelona	ES	26.0	has_published	[online, outdoor]	[firmware_tests]
15267	KU Zemynos gimnazija	Smart Citizen Kit 2.1 with Urban Sensor Board	8364	ikt	55.726706	21.130517	Klaipėda	LT	26.0	has_published	[online, outdoor]	[]
16191	Eibingweg	Smart Citizen Kit 2.1 with Urban Sensor Board	8503	SensingCityV	47.676318	9.195248	Constance	DE	26.0	has_published	[online, outdoor]	[Outside, Terrace, MINKE, CoKLIMAX, Stadtkonst...]
16308	324F	Smart Citizen Kit 2.1 with Urban Sensor Board	8400	Unibo	44.496232	11.354157	Bologna	IT	37.0	has_published	[new, offline, outdoor]	[Bologna, LABTEST, I-CHANGE]
16250	Sensor AD&D 4D	Smart Citizen Kit 2.1 with Urban Sensor Board	8651	sensor AD&D 4D	39.988855	-0.034919	Castelló de la Plana	ES	26.0	has_published	[online, outdoor]	[Research]

3912 rows x 15 columns

Figura 15. Ejemplo de dataset recuperado del API de Smart Citizen

⁹ Repositorio online de análisis en python: <https://github.com/fablabbcn/smartcitizen-data/> y en R: <https://github.com/fablabbcn/smartcitizen-R-data/>

¹⁰ Web oficial de proyecto *Jupyter*: <https://jupyter.org/>

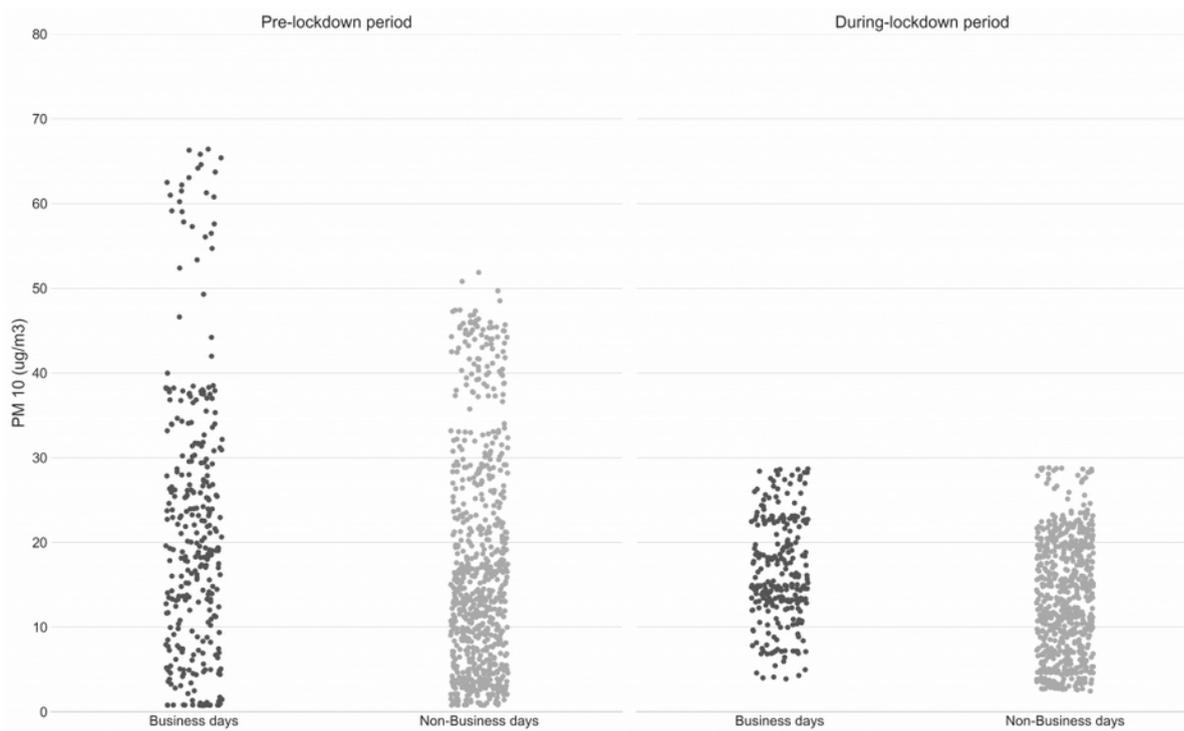


Figura 16. Ejemplo de visualización realizada con scdata

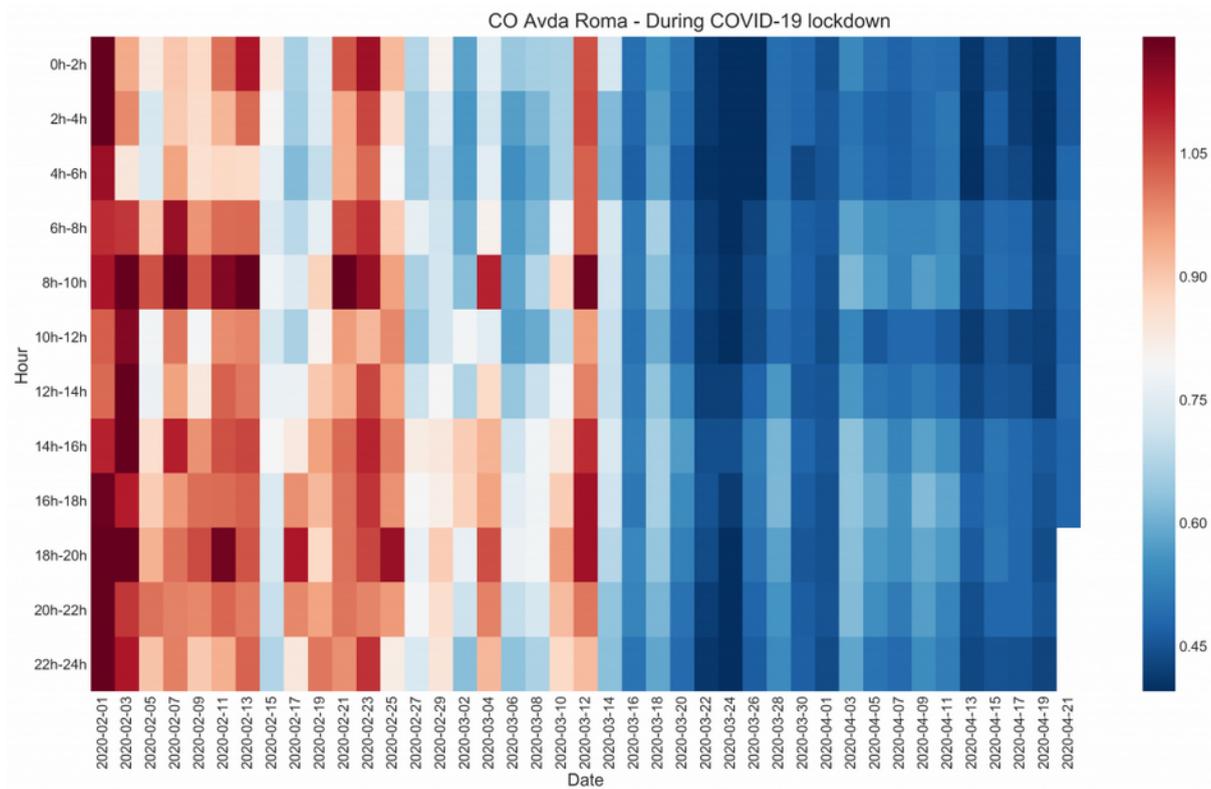


Figura 17. Ejemplo de visualización realizada con scdata (2)

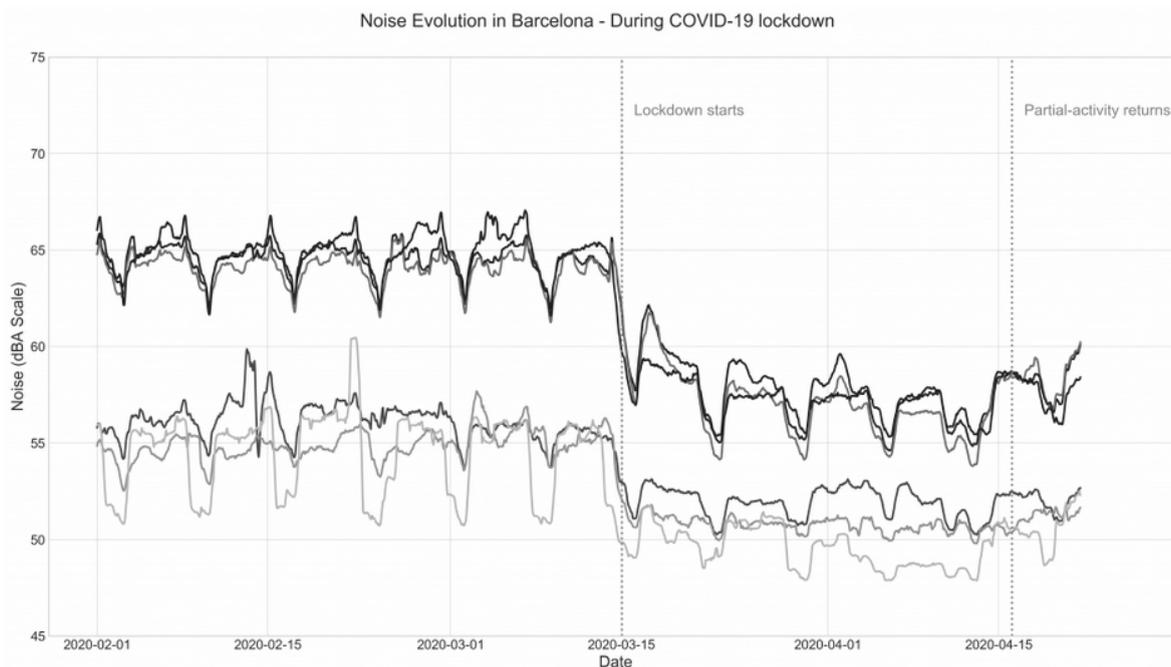


Figura 18. Ejemplo de visualización realizada con sdata (3)

Recursos

- Documentación del entorno de análisis de datos:
<https://docs.smartcitizen.me/Data/Data%20Analysis/>
- Guía de instalación: <https://docs.smartcitizen.me/Guides/data/Install%20the%20framework/>
- Guía para la preparación de informes:
<https://docs.smartcitizen.me/Guides/data/Make%20reports%20of%20your%20data/>
- Guía para la creación de algoritmos:
<https://docs.smartcitizen.me/Guides/data/Custom%20data%20processing/>
- Ejemplos de utilización:
<https://github.com/fablabbcn/smartcitizen-data/tree/master/examples/>