

Datos espaciales públicos, Python y QGIS al servicio de la evaluación de impacto en salud en Andalucía

REVISTA **MAPPING**

Vol.32, 210, 36-42

2023

ISSN: 1131-9100

Public spatial data, Python and QGIS at the service of health impact assessment in Andalusia

Cristina Fernández Álvaro, Luis Ángel Moya Ruano,
Francisco Javier Rodríguez Rasero, Miguel Sanz Pérez

Resumen

Mejores geodatos estadísticos implican mejores políticas y la experiencia de la Consejería de Salud de Andalucía es un ejemplo de ello.

La incertidumbre en la disponibilidad de información precisa sobre el número y características de las poblaciones potencialmente afectadas por una intervención suele ser el mayor factor limitante de la efectividad de toda evaluación de impacto. Una de las claves del éxito de la implementación de la Evaluación de Impacto en la Salud (EIS) en Andalucía ha sido la disponibilidad de datos de población georreferenciada en cuadrículas que difunde el Instituto de Estadística y Cartografía de Andalucía.

Estos datos permiten a la administración y a los agentes económicos y sociales afrontar los grandes retos que plantea la EIS: identificación y caracterización de la población vulnerable.

La tecnología y el *software* libre han permitido dar un paso más allá, facilitando herramientas a la ciudadanía que faciliten esa caracterización y representación espacial de los indicadores necesarios para la elaboración de los informes de evaluación. Es en este contexto donde se ha desarrollado un complemento en QGIS mediante el lenguaje de programación Python que facilita la accesibilidad a los datos geoespaciales indicados y su posterior tratamiento.

Abstract

Better statistical geodata mean better policies and the experience of the Andalusian Ministry of Health makes a good example about it.

Uncertainty in the availability of accurate information on the number and characteristics of populations potentially affected by an intervention is often the major limiting factor in the effectiveness of any impact assessment. One of the keys to the successful implementation of the Health Impact Assessment (HIA) in Andalusia has been the availability of georeferenced population data in grids disseminated by the Institute of Statistics and Cartography of Andalusia.

These data allow the administration and the economic and social agents to face the great challenges posed by the HIA: Identification and characterization of vulnerable population. Technology and free software enabled to go a step further, providing tools for citizens to facilitate the characterization and spatial representation of the indicators necessary for the preparation of assessment reports. It is in this context where a QGIS plugin has been developed using the Python programming language that facilitates accessibility to the geospatial data indicated and its subsequent processing.

Palabras claves: Evaluación de impacto, Salud, Interoperabilidad, Datos espaciales, Malla estadística, Software libre, Coordinación

Keywords: Impact assessment, Health, Interoperability, Spatial data, Statistical grid, Open source, Coordination

Cristina Fernández Álvaro. Instituto de Estadística y Cartografía de Andalucía.

cristina.fernandez@juntadeandalucia.es

Luis Ángel Moya Ruano. Consejería de Salud y Consumo

langel.moya@untadeandalucia.es

Francisco Javier Rodríguez Rasero. Consejería de Salud y Consumo.

francisco.rodriguez.rasero@juntadeandalucia.es

Miguel Sanz Pérez. Becario en prácticas Consejería de Salud

DOI: <https://doi.org/10.59192/mapping.385>

Recepción 19/12/2022

Aprobación 23/01/2023

1. INTRODUCCIÓN

La Agenda 2030 cuenta con un Objetivo de Desarrollo Sostenible (ODS) específico relacionado con la salud y el bienestar, el ODS3, que aspira a garantizar una vida sana y a promover el bienestar para todas las personas. El enfoque de salud en todas las políticas propuesto por la Organización Mundial de la Salud (OMS), insta a considerar las repercusiones de las políticas públicas en la salud de la población, por lo que facilitar la medición de los impactos en la salud de cualquier actuación pública es fundamental.

2. DATOS ESPACIALES Y POLÍTICAS PÚBLICAS

Los datos espaciales o geodatos son fuentes que aportan una gran cantidad de información sobre el territorio. Su uso se ha extendido en la medida en que su publicación en formatos estándares se va generalizando. Además el continuo desarrollo de los programas informáticos ha permitido la generalización de su uso a un público menos especializado.

Por su parte, las políticas públicas, como no podría ser de otra forma, van tomando cada vez un mayor grado de complejidad y profesionalización. Un ejemplo de ello es el Decreto 169/2014 que establece el procedimiento de la Evaluación del Impacto en la Salud de la comunidad autónoma de Andalucía. Según este decreto para cualquier proyecto perteneciente a sectores como la minería, instalaciones energéticas, industrias químicas o petroquímicas, papeleras o textiles entre muchas, proyectos de infraestructuras, de ingeniería hidráulica, tratamiento de residuos, etc, así como otros de planeamiento urbanístico en determinadas condiciones, y cuya ejecución se encuentre a menos de 1 km de asentamientos poblacionales, se tiene que realizar un informe de Valoración del Impacto en Salud (VIS). La entidad pública competente, la que ostente las competencias en salud, se encargará a su vez de elaborar la Evaluación de Impacto en Salud (EIS), es decir de valorar los impactos potenciales de la actuación sobre la salud, así como la distribución de tales efectos en la población.

La valoración de los impactos en salud (VIS) de una actuación y su distribución en la población pasa a ser por tanto responsabilidad del promotor de la actuación. Uno de los pasos necesarios en dicha valoración es caracterizar la población afectada por la actuación comparando una serie de indicadores demográficos y sociales de la población afectada con otra población de referencia.

Es lógico pensar que para poder llevar todo ello a cabo se necesitan datos. El principal factor limitante de la efectividad de toda evaluación de impacto suele ser la incertidumbre en la disponibilidad de información precisa sobre el número y características de las poblaciones potencialmente afectadas ante una actuación (ubicación de un polígono, una industria...). En este sentido, uno de los factores clave del éxito de la implementación de la EIS en Andalucía ha sido precisamente la disponibilidad de datos de cuadrículas de población georreferenciadas anualmente, incluidas las características demográficas y socioeconómicas. Estos datos permiten a la administración y a los agentes económicos y sociales afrontar los dos grandes retos que plantea la EIS: identificación y caracterización de la población vulnerable.

Desde 2011, el Instituto de Estadística y Cartografía de Andalucía (IECA) publica anualmente los datos espaciales de «Distribución espacial de la población de Andalucía», información sobre población georreferenciada a nivel de desagregación de celdas de 250 m x 250 m consultable a través de servicios interoperables. Junto con estos datos el IECA ha seguido publicando otra información con el mismo nivel de detalle, «Razones de Mortalidad», «Indicadores de fecundidad» o con temática urbanística y constructiva «Caracterización y distribución del espacio construido en Andalucía». Estos datos son de gran relevancia por el nivel de detalle de la información y por seguir siendo uno de los conjuntos de datos con mayor nivel de detalle de información en celdas regulares.

Tener la obligación de la elaboración de informes y tener los datos no es garantía de éxito a la hora de tener una buena Evaluación de Impacto en Salud. Podemos imaginar que este tipo de informe planteado puede ser realizado de muy diversas formas y basándose en muchos criterios, siempre dentro de la legalidad. Facilitar herramientas para elaborarlos y los mejores datos disponibles para ello es un gran paso con dos importantes consecuencias: por un lado homogeneizar la información analizada y utilizada, así como los procedimientos técnicos y hacer que las evaluaciones sean de este modo comparables en el caso de diversos informes simultáneos (conurrencia).

3. DATOS ESTADÍSTICOS GEORREFERENCIADOS

La cartografía es una disciplina que se apoya en dos componentes esenciales para representar la información. Trabaja con el concepto de que todos

los elementos tienen una localización en el espacio y unos atributos de carácter temático que los definen y los caracterizan. De este modo, y de forma general, es importante definir qué es lo que se desea representar en un mapa (la información) y cuál es la unidad territorial de observación que se va a utilizar como referencia.

Generalmente las unidades que se utilizan se corresponden con divisiones de carácter ambiental o natural para representar la información de carácter físico, unidades bioclimáticas, comarcas naturales y otras que se utilizan en un caso u otro en función de la información temática que se desee representar. En el caso de la información socioeconómica, ésta se agrega generalmente en función de las distintas unidades administrativas, hablándose generalmente de países, regiones, provincias, comarcas, municipios o secciones censales.

En este sentido, lo más habitual siempre ha sido la representación de la información a nivel provincial o municipal. Para la homogeneización de las tareas de difusión y análisis de la información estadística, instituciones como Eurostat, la Oficina Estadística de la Unión Europea, definen una correspondencia entre unidades administrativas de los distintos estados, son los denominados NUTS.

A pesar de esa homogeneización lo que se busca cada vez más es una mayor desagregación territorial

de la información. Se trabaja con frecuencia en la representación de la información a nivel de sección censal, fundamentalmente con información de carácter demográfico. Esta tiene dos inconvenientes fundamentales. Uno es el de la frecuente modificación de los límites para adaptarse a las necesidades propias que definen la generación de las secciones censales. El otro es el de la heterogeneidad espacial que poseen estas divisiones, en lo que respecta a las amplias diferencias en cuanto a sus dimensiones superficiales, pero también en su estructura territorial, ya que en muchos casos una sección puede tener un espacio reducido con una estructura territorial urbana residencial, frente a un espacio proporcionalmente mucho mayor con una estructura urbana no residencial o de carácter rústico.

Dentro de esta tendencia, ha ido tomando fuerza la generación de unidades de observación de dimensiones más reducidas y de geometría regular. A partir del Foro Europeo para la Geografía y la Estadística (EFGS) y las experiencias en el proyecto Geostat (ESSnet project Geostat) se desarrolló la generación de una malla regular formada por celdas de 1 km de lado, utilizando un mismo sistema de referencia espacial para la totalidad de Europa.

La mayor desagregación espacial permite un conocimiento más exhaustivo del territorio, pudiéndose analizar de una forma más precisa los matices en los patrones de comportamiento de las distintas variables dentro del territorio.

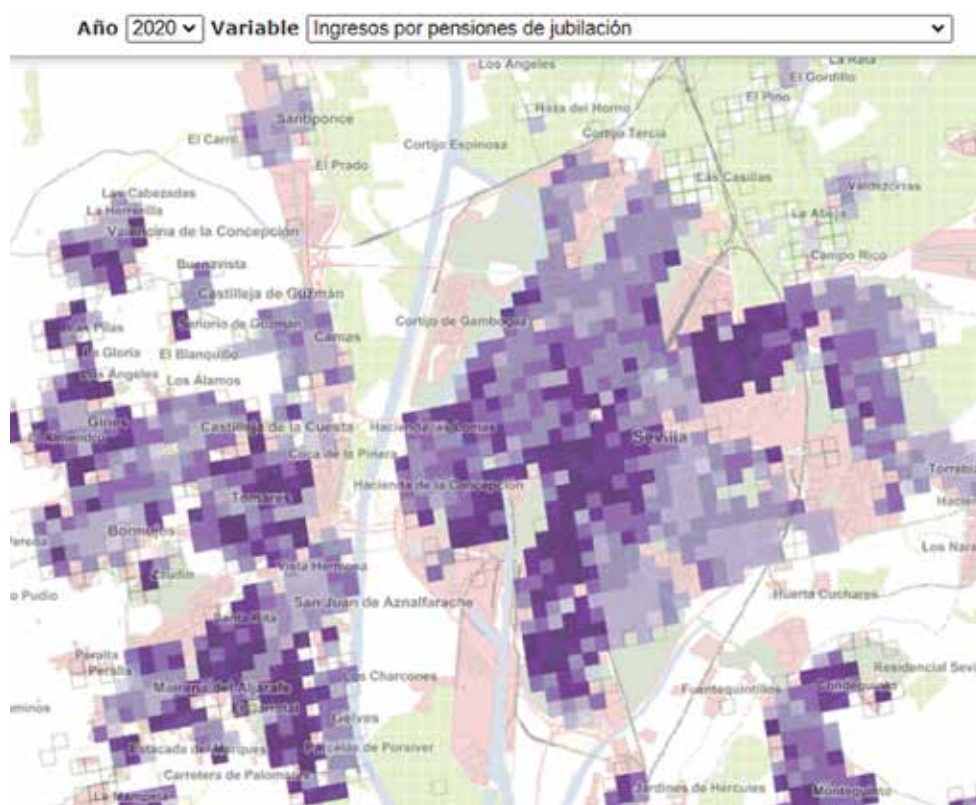


Figura 1. Celdas de 250 x 250 m de publicación «Distribución Espacial de la Población en Andalucía»

El objetivo perseguido por el IECA es representar la distribución de la población de Andalucía según diversas características sociodemográficas en un mapa, buscando una amplia desagregación territorial y homogeneidad espacial. Para ello, se optó por una malla regular formada por celdas de 250 m de lado en la que se representa una estimación de la población andaluza obtenida a partir de la explotación de la Base de Datos Longitudinal de Población de Andalucía (BDLPA).

A la información individual de la BDLPA se le ha incorporado una serie de características presentes en

otros ficheros administrativos, como el de afiliaciones en alta laboral y cuentas de cotización a la Seguridad Social, prestaciones de la Seguridad Social o personas demandantes de empleo del Servicio Andaluz de Empleo.

Por otro lado, se geocodifican todas las direcciones postales que según la BDLPA tienen los habitantes en una fecha dada, de este modo queda toda la información estadística geocodificada. Se utiliza como fuente de referencia para la geocodificación el portalero del Callejero Digital de Andalucía Unificado y el Censo de Edificios de 2011. Este proceso de geocodificación se lleva a cabo por medio de la herramienta de fusión de ficheros aLink, aplicación de *software* libre desarrollada por el Instituto de Estadística y Cartografía de Andalucía (IECA). El sistema de referencia espacial utilizado para la geocodificación es el ETRS 1989 UTM Zona 30N.

El resultado es el mostrado en la Figura 1 para datos sobre ingresos por pensiones de jubilación.

4. HERRAMIENTAS DE SOFTWARE LIBRE PARA LA EIS

El mejor análisis que puede llevarse a cabo en una EIS es de tipo espacial y requiere el uso de Sistemas de Información Geográfica. Hasta hace relativamente poco era un tipo de herramienta cuya utilización estaba limitada a especialistas, debido a la existencia de pocos *softwares* y además sometidos a importantes licencias.

La aparición de herramientas SIG de *software* libre como GVSIG o QGIS ha democratizado la utilización de estos sistemas de análisis espacial. En el caso concreto de la elaboración de los informes de EIS por la administración, pero especialmente por el promotor de la actuación, disponer de *software* libre es fundamental pues facilita su elaboración y reduce sus costes.

En referencia a estos informes la Consejería de Salud ha dado un paso más allá y ha creado utilizando Python un complemento (*plugin*) para QGIS que facilita la caracterización y representación espacial de los indicadores sociales para la elaboración de estos informes de evaluación. Con ello se integra en el propio proceso de análisis espacial la incorporación de los mejores datos disponibles y homogeneiza en cierto modo la propia evaluación.

El complemento se denomina «*Andalusian population*» y utiliza gran parte de los datos espaciales del IECA generados en celdillas de 250 x 250 m para la elaboración de mapas e indicadores base del análisis

espacial que es necesario realizar en todo EIS.

El complemento puede encontrarse en <https://plugins.qgis.org/plugins/andalusianpopulation/>

El *plugin* permite comparar de una forma sencilla e intuitiva diversas variables socioeconómicas de las celdas de la Distribución Espacial de la Población en Andalucía con los niveles medios regionales, provinciales o de su entorno más inmediato. El principal objetivo es comparar indicadores sociales que permitan caracterizar la población.

La utilización de QGIS, Python y los datos públicos del IECA permite crear una aplicación que facilita que entidades públicas y privadas elaboren las EIS, sin necesidad de adquirir ninguna licencia. Además, al tratarse de código abierto, también existe la posibilidad de que este trabajo sea ampliado y personalizado para otras aplicaciones que puedan surgir o que sea completado por los agentes involucrados en la elaboración de estos análisis.

El código fuente de esta aplicación puede verse en Github <https://github.com/devmemi/AndalusianPopulation> Su única limitación es tener instalada al menos la versión de QGIS 3.16.

5. IMPLEMENTACIÓN DE TODOS LOS ELEMENTOS: CASO DE USO DE SOFTWARE LIBRE PARA LA EIS

En el proceso de implementación de los tres elementos vistos: QGIS, Python y los datos espaciales es necesario distinguir cuatro etapas:

1. Obtención de la información
2. Tratamiento de datos
3. Almacenamiento de datos postprocesados
4. Diseño de la interfaz y algoritmia

Posiblemente el pilar principal de todo este proceso sea la obtención de datos. Para poder llevar a cabo un trabajo tan intensivo en datos georreferenciados es necesario tener el soporte de un organismo cuyos datos sean fiables, públicos y, sobre todo, sigan una estructura fija a partir de la cual se pueda obtener toda la información necesaria.

Todos los datos empleados en la elaboración de indicadores sociales se han obtenido a partir de los datos públicos del IECA (Instituto de Estadística y Cartografía de Andalucía). Este organismo provee anualmente de un gran conjunto de indicadores socioeconómicos de la región de Andalucía distribuidos espacialmente en mallas de distintos tamaños (250 x 250 metros y 1x1 km) como ya se ha mencionado

Para este trabajo se decidió emplear la malla de 250 x 250 metros, ya que permite conocer con mayor precisión las características principales de una determinada región de Andalucía, principalmente a pequeña escala.

Sin embargo, el gran nivel de detalle de la información también tiene algunos lados «perversos». Uno de ellos es la aparición de censura en la información debido a la vulneración del secreto estadístico. Cuando una variable toma un valor por debajo del límite del valor establecido como secreto estadístico en la celda, aparece con el valor -1, algo que suele ocurrir con frecuencia en zonas rurales y despobladas. La decisión tomada ante estas situaciones ha consistido en descartar las celdas censuradas para esa variable.

El otro lado «perverso» tiene que ver con el hecho de emplear una malla 250 x 250 metros en un territorio tan grande como Andalucía, lo provoca que la cantidad de datos e información se vea incrementada y por lo tanto, las necesidades computacionales.

Es necesario por tanto, emplear etapas intermedias para tratar los datos obtenidos, incorporando la identificación del municipio y la provincia, seleccionando y generando los principales indicadores relativos que permitan la comparabilidad y reduciendo la cantidad de información a la mínima necesaria para poder trabajar posteriormente con los datos ya procesados. La salida de esta etapa es una tabla en Excel donde cada fila es una de

las celdas de la malla y las columnas son los indicadores sociales.

Esta información generada es conveniente tenerla almacenada para etapas posteriores y no calcularla al vuelo cada vez que se inicie el proceso, puesto que las etapas iniciales del proceso son de por sí costosas computacionalmente hablando. El propio instalador del *plugin* ha sido diseñado para que cuando se inicia por primera vez recoja todos los indicadores sociales a partir de los datos del IECA y los almacene en un fichero Excel.

Una vez se han visto las etapas relacionadas más puramente con los datos, se analiza el diseño de la interfaz, es decir la parte más visible del trabajo desarrollado y la cual permite satisfacer los principales objetivos del proyecto.

La interfaz de usuario ha sido desarrollada en el lenguaje de programación Python, al igual que el resto del *plugin*. Su principal característica es la capacidad de abstraer al usuario de la complejidad que conlleva realizar mapas elaborados en QGIS a través de un sencillo formulario, que una vez rellenado ofrece como resultado un mapa listo para su inclusión en cualquier informe VIS o EIS. El tiempo empleado en ello es mínimo y se dedica en su mayor parte a la toma de decisiones sobre contra qué comparar y el indicador a seleccionar, que con la elaboración del mapa en sí.

Las principales decisiones que es necesario tomar frente a la interfaz son las siguientes:

- Cómo se van a llamar nuestras imágenes y dónde queremos guardarlas (sección Archivo). Con la intención de organizar la información resultante, el *plugin* permite darle a nuestras futuras imágenes un prefijo que servirá para identificarlas así como el lugar donde se desean guardar.
- Qué vamos a comparar y con qué (sección Parámetros espaciales). Como se ha mencionado anteriormente, uno de los objetivos es poder comparar una zona de interés (zona donde se quiere iniciar un proyecto) respecto a los valores del municipio en el que se encuentra, la provincia o la comunidad autónoma. Concretamente, en la interfaz es necesario introducir las coordenadas en UTM Huso 30 del lugar de la actuación y la región con la que se quiere comparar: Andalucía, cualquier provincia, cualquier municipio u otro *buffer* o región entorno y sus coordenadas introducidas (en metros).
- Otro parámetro a introducir es la distancia de visualización del mapa. Esto permite alejar o acercar la vista del mapa para que englobe a una mayor región o por el contrario, se centre en las coordenadas introducidas y las celdas de su entorno.
- En siguiente lugar se escoge el indicador que se desea mostrar en el mapa, que se generará al final de la ejecución.



Figura 2. Cuadro de diálogo con el plugin

- Una vez seleccionado el indicador, se podrán modificar colores y los rangos en los que los indicadores se mostrarán por defecto, pero en principio no es obligatorio.
- Por último y antes de la salida final, se ofrece la opción de mantener las capas generadas en el proyecto de QGIS, por si una persona con la formación necesaria para manipular capas en QGIS desea realizar mapas más complejos o elaborados que el *plugin* diseñado no permite.
- Finalmente, existe una pestaña de configuración que permite a los usuarios modificar la dirección desde donde se toman las capas WFS y WMS, permitiendo actualizar estas capas cuando el IECA las actualiza en su página web.

6. INTERPRETACIÓN DE LOS RESULTADOS

Tomadas todas esas decisiones, el resultado que se obtiene es una imagen en formato png donde pueden verse, para un *buffer* de 1 km a partir del punto de

intervención, los valores en cada celda del indicador seleccionado en la interfaz y con un código de colores que la caracteriza. Además la imagen va acompañada de una leyenda que reflejará la siguiente información:

- Indicador social
- Coordenadas del punto de interés
- Distancia de visualización (*buffer*)
- Región con la que se comparan las celdas de la imagen
- Código y rangos de colores (simbología)

Esta imagen queda guardada con el prefijo que le hayamos propuesto, en la carpeta donde hayamos indicado.

El código de color de las celdas hace referencia a la parte de la comparativa con otros ámbitos territoriales.

Por defecto se establecen 5 intervalos que representan los quintiles en los que se mueven las celdas de la variable seleccionada en el ámbito territorial de comparación. Los colores van de verde a rojo de modo que el 20 % de las celdas de menor valor en el ámbito de comparación estarán representadas en verde y el 20 % de mayor valor, en rojo.

Cada celda del *buffer* de estudio indica en qué quintil de la distribución de la variable en el ámbito

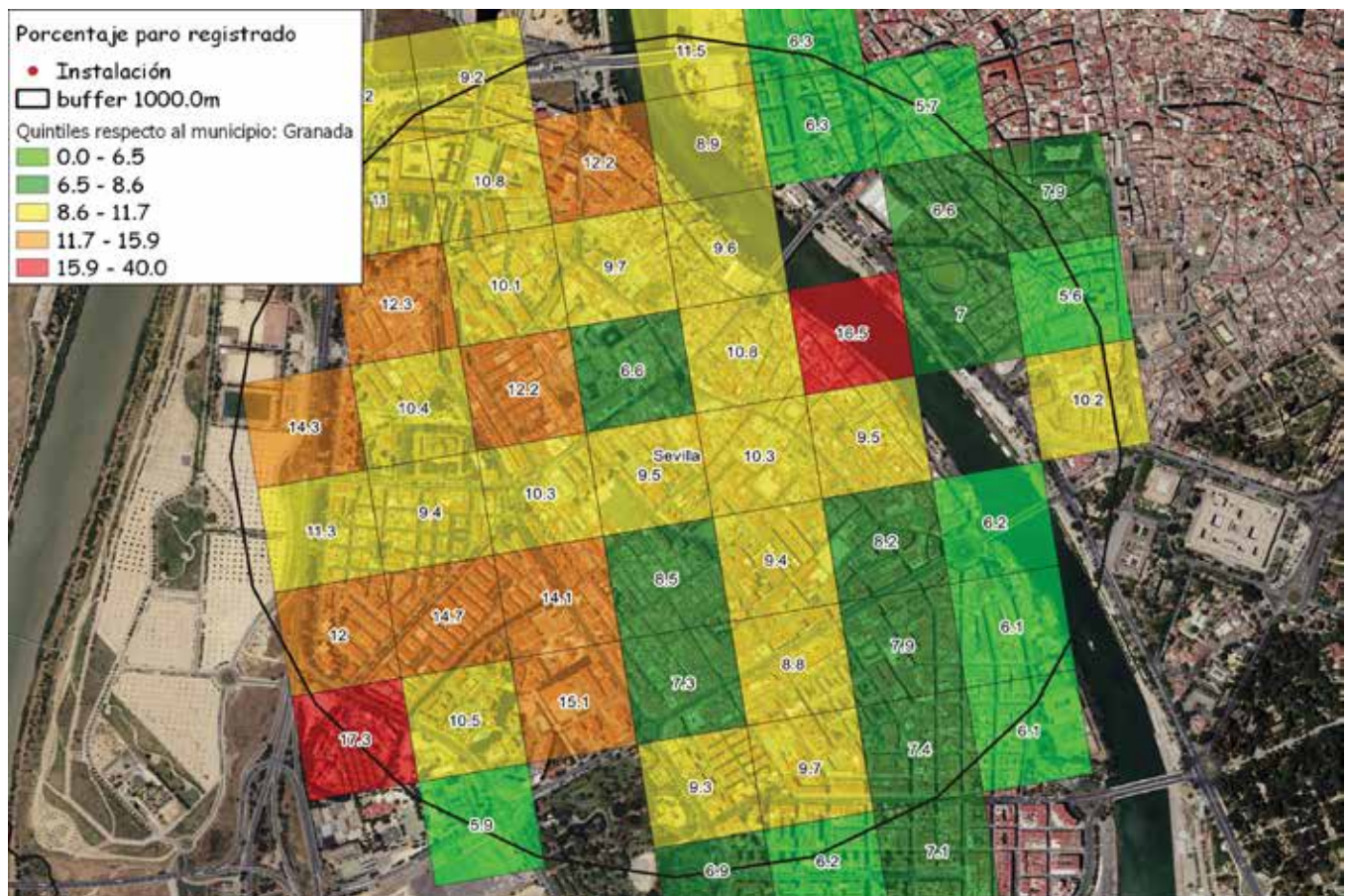


Figura 3. Mapa generado

comparado se encuentra. Por ejemplo, en la Figura 3 el valor de la variable de estudio donde se ubica el epicentro del *buffer* vale 9,5 % de paro registrado, se sabe que ese valor está en el tercer quintil de los valores de Granada ciudad.

Esto permite de forma sencilla y rápida, conocer qué zonas de las representadas en el mapa se encuentran en un quintil o en otro respecto a la referencia que se haya escogido previamente en la interfaz.

Aunque por defecto la distribución es en quintiles esta puede ser modificada por el usuario, cambiando tanto el rango como la cantidad de tramos en los que se quiere subdividir la categorización del mapa de una forma muy intuitiva. De igual forma, el color asignado a cada uno de los tramos en los que se subdivide la categorización también puede ser modificado, así como la opacidad de cada color.

7. CONCLUSIONES

La implementación de este *plugin* ha cumplido el objetivo inicial de ayudar en las tareas de elaboración de informes EIS en dos aspectos muy importantes: la reducción del tiempo empleado en elaborar los mapas y en la posibilidad de realizar análisis más profundos.

Los resultados reales han sido tan satisfactorios pues se ha promovido y se sigue promoviendo su uso en otros campos de aplicación por su fácil acceso, la casi nula necesidad de conocimientos de SIG y la cantidad de información que es capaz de mostrar con muy poco tiempo de trabajo.

El uso de QGIS, Python y los datos públicos del IECA para elaborar esta aplicación facilita su uso por parte de entidades, públicas y privadas, sin necesidad de adquirir ninguna licencia, permitiendo, al tratarse de código abierto, la posibilidad de que este trabajo sea ampliado y personalizado para otras aplicaciones que puedan surgir.

El proyecto se ha presentado en distintos organismos públicos nacionales e internacionales con una muy buena acogida, aunque la necesidad de que existan datos bien estructurados y mallados como los ofrecidos por el IECA en Andalucía limitan mucho la exportación del proyecto en otros ámbitos.

REFERENCIAS

Distribución Espacial de la Población.
<https://www.juntadeandalucia.es/institutodeestadisticaycartografia/distribucionpob/index.htm>

Servicios Interoperables "Distribución Espacial de la Población"

<https://www.juntadeandalucia.es/institutodeestadisticaycartografia/servicios.htm>

Complemento de QGIS (3.16 y superior)

<https://plugins.qgis.org/plugins/andalusianpopulation/>

Código fuente del plugin. Github

<https://github.com/devmemi/AndalusianPopulation>

Sobre los autores

Cristina Fernández Álvaro

Cristina Fernández Álvaro es Doctora en Economía y licenciada en Ciencias y Técnicas Estadísticas por la Universidad de Sevilla. Funcionaria de carrera del Cuerpo Superior Facultativo, desarrolla su actividad en el servicio de Estudios y Métodos del Instituto de Estadística y Cartografía de Andalucía en la que participa en las tareas de divulgación y puesta en valor de los trabajos relacionados con la "Distribución Espacial de la Población de Andalucía".

Luis Ángel Moya Ruano

Luis Ángel Moya Ruano es ingeniero industrial por la Universidad de Sevilla y licenciado en Geografía e Historia por la UNED. Trabaja desde hace años en el servicio de Salud Ambiental de la Consejería de Salud y Consumo y la última década se ha especializado en el desarrollo y puesta en marcha de la Evaluación de Impacto en Salud. Además de una dilatada experiencia realizando evaluaciones, su principal interés radica en la aplicabilidad de los sistemas de información geográfica en la identificación y evaluación de impactos sobre la población.

Francisco Javier Rodríguez Raseo

Francisco Javier Rodríguez Raseo es licenciado y doctor en Farmacia por la Universidad de Sevilla y licenciado en Ciencias y Tecnología de los Alimentos por la Universidad de Córdoba. Ha trabajado más de una década en Salud Ambiental con especial interés en las aplicaciones SIG para el estudio de poblaciones y salud humana en el marco de las evaluaciones de impacto. Actualmente trabaja en análisis de datos de salud pública en el Centro Común de Investigación (JRC) de la Comisión Europea.

Miguel Sanz Pérez

Ingeniero de electrónica, robótica y mecatrónica por la Universidad de Sevilla especializado en la programación de sistemas de control industrial y de automatismos.