

Tipo de artículo: Artículo original
Temática: Reconocimiento de patrones
Recibido: 15/04/2016 | Aceptado: 05/05/2016

Estratificación de territorios basada en indicadores de salud sobre el Sistema de Información Geográfica QGIS

Stratification of territories based on health indicators on the Geographic Information Systems QGis

Yadian Guillermo Pérez Betancourt^{1*}, Liset González Polanco¹, Rolando Morales Pérez¹, Yanislay Torres Vega¹

¹Facultad 3. Universidad de las Ciencias Informáticas. Carretera a San Antonio de los Baños, km 2 $\frac{1}{2}$, Torrens, Boyeros, La Habana, Cuba. CP.: 19370.

*Autor para correspondencia: ygbetancourt@uci.cu

Resumen

El desarrollo de las tecnologías de la información y las comunicaciones en la actualidad ha potenciado el uso de la información geográfica en disímiles ramas como la agricultura, la meteorología, el turismo y la salud pública, implicando la utilización de los mapas como factor importante en la toma de decisiones. El uso de los Sistemas de Información Geográfica en la rama de la salud ha aumentado considerablemente, sin embargo su utilización en el sector aún es limitada y no se explotan en su totalidad la componente espacial de los datos. El objetivo del presente trabajo es desarrollar una propuesta para la estratificación de territorios basada en indicadores de salud que integrada a QGIS facilite la incorporación del espacio en los estudios salubristas. Se obtuvo como resultado un sistema que permite integrar datos de variada naturaleza en el análisis y construcción de estratos; la solución tiene su base en las técnicas de agrupamiento de datos y contribuye a la identificación de riesgos de salud de los territorios cubanos y a la toma de decisiones en las entidades de salud. Se realizó un proceso de estratificación con indicadores de salud seleccionados del anuario estadístico del 2001 y se valoraron los resultados obtenidos.

Palabras claves: estratificación, riesgos de salud, sistemas de información geográfica, técnicas de agrupamiento

Abstract

The development of information technology and communications has boosted the use of geographic information in dissimilar branches such as agriculture, meteorology, tourism and public health, involving the use of maps as an important factor in the decision making. The use of Geographic Information Systems in the field of health has increased significantly, but its use in the field is still limited and not fully exploit the spatial component of data. The aim of this work is to develop a proposal for the stratification of territories based on health indicators. A system to integrate data of various nature for analysis and construction of strata was obtained as a result; the solution is based on data clustering techniques and contributes to the identification of health risks in Cuban territory and decision-making in health institutions. Stratification process was conducted with selected health indicators of the Statistical Yearbook 2001 and the results were evaluated.

Keywords: stratification, health risks, geographic information systems, clustering techniques

Introducción

En la actualidad, el progresivo avance de las tecnologías computacionales y la necesidad del manejo de información periódicamente en aumento, conlleva a la necesidad de un mayor uso de los sistemas informáticos en diversas actividades de la sociedad. En particular los Sistema de Información Geográfica (SIG) se han convertido en una herramienta de gran impacto por su aplicación en ramas como la agricultura ([Suprem et al., 2013](#)), la meteorología ([Wang, 2014](#)), el turismo ([Li et al., 2015](#)) y la salud pública ([Tanser and Le Sueur, 2002](#); [Shi and Kwan, 2015](#)), por solo mencionar algunas.

Para justificar su importancia y el papel que estos juegan, es necesario mencionar que se acepta por la comunidad científica que entre el 80 y 90 % de la información utilizada en la toma de decisiones tiene asociada una componente espacial ([Batista Moliner et al., 2001a](#); [Bivand et al., 2013](#)). El desarrollo de herramientas para el análisis de la información, considerando los datos relativos a su posición en el espacio, constituyen una oportunidad para las nuevas investigaciones. El uso de los SIG en la rama de la salud tiene cada día mayor utilidad, su empleo contribuye al fortalecimiento de la capacidad de análisis en materia de salud pública y epidemiológica, brindando información de utilidad para la toma de decisiones ([Rouhani and Hosseini, 2013](#)). Por otra parte, facilitan la identificación de la ubicación geográfica de establecimientos de salud y grupos de población que presentan mayor riesgo de enfermar o de morir prematuramente y que por tanto requieren de mayor atención preventiva, curativa o de promoción de la salud ([Hay et al., 2013](#)).

Para el análisis de salud, se hace necesario conocer con el mayor grado de detalle posible, las características de cada una de las unidades territoriales, así como sus grupos poblacionales([Batista Moliner et al., 2001b](#)), a partir de diferentes indicadores, que pueden ser: demográficos, socio-económicos y ambientales, solo por mencionar algunos. Todos estos elementos tienen un impacto determinante en la caracterización de un territorio y constituyen la base en el establecimiento de la estratificación territorial.

La estratificación territorial es un proceso que permite dimensionar espacialmente los eventos a través de un proceso de agregación y desagregación de los territorios a evaluar, a partir de variables seleccionadas para dichos territorios que permitan agregaciones (por homologías de las características) o desagregaciones (por heterogeneidades de estas) ([Batista Moliner et al., 2001b](#)). La utilización de los SIG en el análisis de la distribución espacial de enfermedades ha aumentado considerablemente, sustentado en las herramientas de análisis existentes que posibilitan resolver problemas asociados a la distribución espacial ([Fotheringham and Rogerson, 2013](#)). Sin embargo estas no son extensibles, su utilización se limita a llevar información a la cartografía, y la componente espacial de los datos no es explotada en su totalidad. Si bien el espacio es un

elemento importante en estos estudios, no siempre se le da la importancia requerida, motivado por: i) acceso limitado a los SIG por los costos que ellos implican, ii) poco conocimiento de las herramientas y iii) el tiempo de formación en el área de los SIG es elevado ([González Polanco and Pérez Betancourt, 2013](#)).

El empleo de la estratificación territorial en las áreas de la salud basado en algún indicador se presenta como el de más utilización en los estudios, los métodos aplicados son mayoritariamente estadísticos y no se explotan otras técnicas que tienen en cuenta la naturaleza espacial de los datos. En Cuba este proceso se desarrolla mediante soluciones informáticas por separado, primeramente se realiza el análisis estadístico y luego se presentan los resultados en mapas temáticos utilizando los SIG ([Pérez Martínez et al., 2003](#)), lo que reduce la eficiencia del trabajo. Estos elementos afectan cuando se requiere el análisis de la relación espacial de indicadores en diferentes áreas e influyen negativamente en la capacidad de gestión de las entidades de salud. Por otro lado los SIG cuentan con licencias comerciales y no están alineados a la estrategia del país sobre la migración a software libre.

En este sentido el SIG QGIS está ganando protagonismo dentro del mundo del software libre y cuenta con una comunidad activa para su soporte, sin embargo aún presenta muchas limitaciones para realizar análisis de máximo interés para los estudios de salud. En el presente trabajo se presenta una propuesta para realizar el proceso de estratificación utilizando QGIS. Se realiza una estratificación a partir de otros trabajos realizados y se muestran los resultados.

Materiales y métodos

En el presente trabajo el proceso de estratificación se desglosa en las fases siguientes:

Selección: se eligen los factores de estratificación (territorios, indicadores).

Pre-procesamiento: se obtiene el aporte informacional y se normalizan los datos de los indicadores seleccionados.

Agrupamiento: se clasifican los territorios en grupos homogéneos.

Visualización: se representa en un mapa temático cada grupo homogéneo de territorios.

En la Figura 1 se presenta el modelo conceptual de la propuesta de solución. Para realizar el proceso de estratificación se propone la utilización de indicadores estadísticos seleccionados por el usuario y el empleo de la naturaleza espacial de los datos a través de los indicadores cartográficos siguientes:

- Cantidad de fuentes contaminantes
- Cantidad de ríos que presentan contaminación

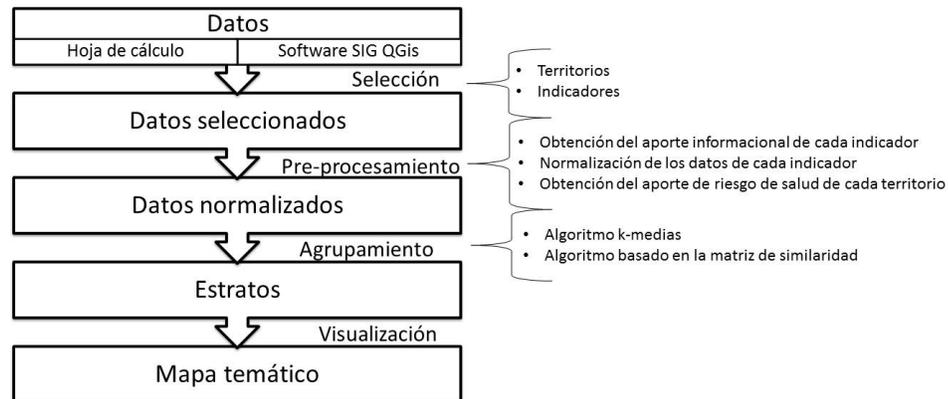


Figura 1. Modelo conceptual de la propuesta para la estratificación.

Para la incorporación de estos dos indicadores se tuvo en consideración:

- El problema de contaminación de las aguas, se encuentra entre los principales problemas ambientales a los que se expone la sociedad, vinculado principalmente al impacto producido en áreas densamente pobladas y las alteraciones a la salud y a la calidad de vida de la población (Barceló and de Alda, 2008; Gil et al., 2012).
- El problema de la contaminación del aire, afecta a la sociedad y a la salud del ser humano. Según datos de la Oficina Nacional de Estadísticas se evidencia que cada año más del 30 por ciento de los cubanos sufren enfermedades respiratorias y otras asociadas con la contaminación del aire.
- La utilización de los de datos espaciales puntos y polilíneas permite servir de nivel de partida para incorporar el desarrollo de futuros trabajos, debido a que las distintos indicadores estudiados en la literatura están asociados a este tipo de datos.

Pre-procesamiento de los datos

Obtención del aporte informacional de los indicadores

El aporte informacional de los indicadores se obtiene mediante el coeficiente de variación a partir de la ecuación

(1) en función del dominio de cada indicador.

$$\sigma^2 = \sqrt{\frac{1}{t} \sum_{i=1}^t (V_i - \bar{x})^2} \quad (1)$$

Donde t es el total de la población finita de los datos, \bar{x} promedio de los datos, V_i valor i , donde $i = 1, 2, \dots, t$. Posteriormente se obtiene la desviación estándar σ que es la raíz cuadrada de la varianza de los datos. El coeficiente de variación obtiene la dispersión de los datos en función de su promedio. El coeficiente de variación se determina de la siguiente manera:

$$w_k = \frac{\sigma}{\bar{x}} \quad (2)$$

Siendo w_k el aporte informacional del indicador k , donde $k = 1, 2, \dots, t$.

Normalización de los datos seleccionados

A partir de los diferentes dominios en los que se presentan los valores de los indicadores es necesario normalizarlos (Molina and García, 2006). Para normalizar los valores se utiliza la siguiente ecuación (Valpreda, 2007).

$$x_{i,f} = \frac{X_{i,f} - X_{Min_f}}{Max_f - Min_f} \quad (3)$$

Donde $X_{i,f}$ es el valor i del indicador f con $i = 1, 2, \dots, t$; Min_f , Max_f mínimo y máximo valor del indicador k respectivamente.

Obtención del riesgo de salud por territorio

Para obtener el riesgo por cada territorio se dividen los indicadores en dos grupos, identificados como: los que mitigan los riesgos de salud y los que aumentan el riesgo. A continuación se presenta la ecuación para el cálculo del aporte al riesgo para cada grupo en correspondencia con otros trabajos de la literatura. (Ruiz-Shulcloper et al., 1999; López Caviedes, 2004):

Grupo 1: indicadores que mayor valor implican mayor aporte al riesgo.

$$Ap_{x_{ik}} = \frac{x_{ik}}{Max_{x_k}} \quad (4)$$

Donde $Ap_{x_{ik}}$ es el aporte de riesgo y Max_{x_k} es el mayor valor que toma el indicador.

Grupo 2: indicadores que mayor valor implican menor aporte al riesgo.

$$Ap_{x_{ik}} = 1 - \frac{x_{ik}}{Max_{x_k}} \quad (5)$$

Finalmente para obtener el riesgo por territorio a partir de los indicadores seleccionados se plantea la ecuación:

$$\gamma(O) = \frac{1}{\sum_{i=1}^n w_k} \sum_{i \in \{p | X_p(O) \neq ?\}} w_k Ap_i \quad (6)$$

Donde ? denota la ausencia de información.

Agrupamiento de los datos

Para clasificar los territorios en grupos homogéneos (estratos), se seleccionó el algoritmo de agrupamiento K-Means. A partir del análisis de los trabajos realizados sobre estratificación, se identificó que este algoritmo es uno de los más utilizados y con mejores resultados (López Caviedes, 2004; Bombino Companioni, 2006; Bock, 2007; Perreard et al., 2006). En la solución se implementaron una gran variedad de algoritmos de agrupamiento de los encontrados en la literatura.

Medida de similitud empleada

Para determinar la semejanza entre los territorios se utiliza la métrica distancia euclidiana ponderada como se muestra en la ecuación (4). Esta medida de similitud se identifica como una de las más utilizadas y sencillas (Edna, 2006). En esta métrica cuando los valores son numéricos, se obtienen resultados satisfactorios en la clasificación (Rodríguez et al., 2013).

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^p (w_k (x_{ik} - x_{jk}))^2} \quad (7)$$

Donde x_i es el territorio i , x_j es el territorio j , w_k es el aporte informacional del indicador k , p total de indicadores, x_{ik} valor de x_i para el indicador k , para $k = 1, 2, \dots, p$ e $i, j = 1, 2, \dots, t$

Resultados y discusión

Para realizar el proceso de estratificación teniendo en cuenta los elementos descritos anteriormente se desarrolló un complemento para el SIG QGIS. Se definieron los siguientes requisitos funcionales:

- RF 1: Importar indicadores estadísticos desde una hoja de cálculo.
- RF 2: Obtener características cartográficas a través de QGIS.
- RF 3: Construir estratificación.
- RF 3.1: Construir estratos.
- RF 3.2: Visualizar estratos construidos en mapa temático
- RF 4: Gestionar las estratificaciones.
- RF 4.1: Adicionar estratificación.
- RF 4.2: Mostrar estratificación.
- RF 4.3: Eliminar estratificación.
- RF 5: Exportar mapa temático de una estratificación como imagen.
- RF 6: Exportar estratificación hacia una hoja de cálculo.

Para el almacenamiento de los estudios y estratificaciones se diseñó la base de datos. La Figura 2 muestra el diagrama entidad relación.

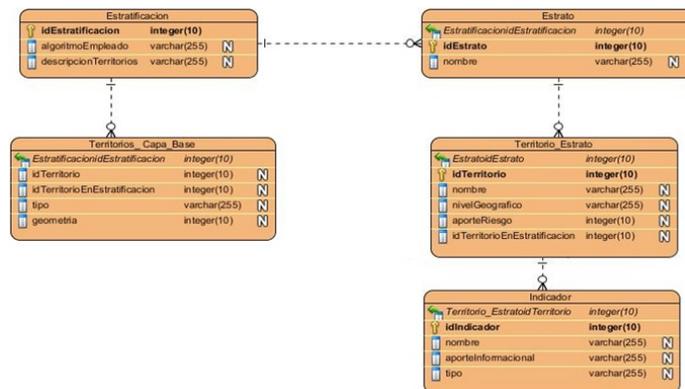


Figura 2. Diagrama entidad relación.

Para valorar los resultados de la solución propuesta se decide aplicar un caso de estudio, en correspondencia con el trabajo realizado por (Bombino Companioni, 2006), donde se realiza un proceso de estratificación de

las catorce provincias de Cuba definidas en la división política-administrativa de 1976. Se utilizó como fuente de información el Anuario Estadístico del año 2001 (INFOMED, 2001) y se seleccionaron los indicadores mortalidad infantil por cada 1000 nacidos vivos, mortalidad infantil de los niños menores de cinco años por cada 1000 nacidos vivos, mortalidad por enfermedades del corazón por cada 100 000 habitantes, mortalidad por tumores malignos por cada 100 000 habitantes, mortalidad por enfermedades cerebrovasculares por cada 100 000 habitantes, mortalidad por influenza y neumonía por cada 100 000 habitantes, mortalidad por accidentes por cada 100 000 habitantes, mortalidad perinatal por cada 1000 nacidos vivos, mortalidad por enfermedades infecciosas y parasitarias por cada 100 000 habitantes, mortalidad materna por cada 100 000 nacidos vivos, incidencia de tuberculosis por cada 100 000 habitantes, incidencia de hepatitis por cada 100 000 habitantes, incidencia de diabetes por cada 100 000 habitantes. Incidencia de hipertensión por cada 100 000, incidencia de asma por cada 100 000 habitantes, incidencia de bajo peso al nacer, consultas médicas por habitante, ingresos por cada 100 habitantes, camas de asistencia por cada 1000 habitantes, consultas de puericultura por habitante, consultas de pediatría por habitante, densidad poblacional, población mayor de 60 años, población menor de 1 año, población menor de 15 años, natalidad por cada 1000 habitantes.

Para realizar la clasificación de cada una de las provincias de Cuba se utilizó la herramienta desarrollada, se empleó el algoritmo de agrupamiento K-Means y el número de estratos se fijó en 4. La Figura 3 muestra parte del proceso realizado.

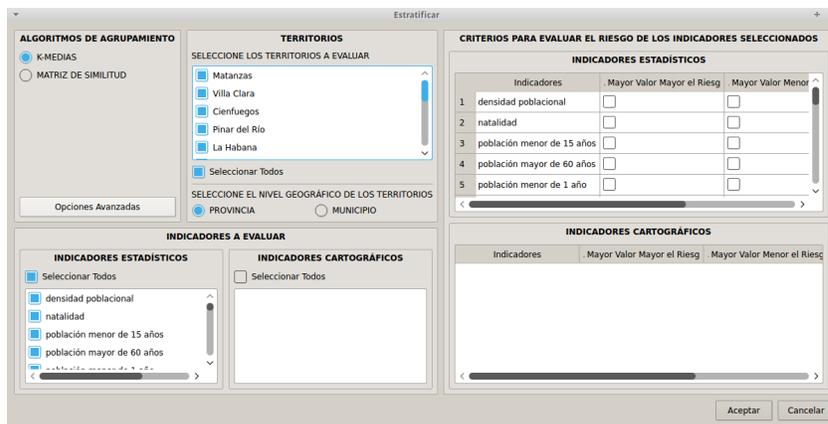


Figura 3. Interfaz de usuario Vista estratificación.

A continuación se muestran los resultados de la estratificación de las provincias de Cuba según las variables escogidas. En la Figura 4 se presentan los resultados en un mapa y la solución brinda la posibilidad de interactuar con los estratos como se muestra en la Figura 5. En la Tabla 1 se muestran los resultados por



Figura 4. Mapa estratificado.

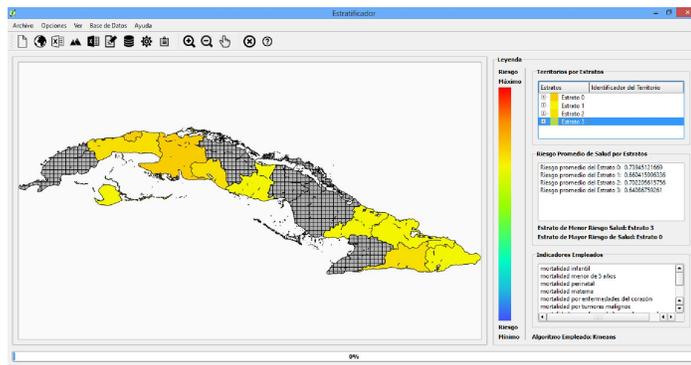


Figura 5. Vista principal para interactuar con los estratos.

estratos de manera más detallada. Los resultados muestran la utilidad del empleo de la propuesta, permitiendo

Tabla 1. Estratos y riesgo promedio asociado a cada estrato.

Estrato	Provincias	Riesgo
Estrato1	Ciudad de la Habana	0.76665
Estrato2	Holguín, Las Tunas, Villa Clara, Sancti Spíritus	0.47898
Estrato3	Matanzas, Cienfuegos, La Habana, Camagüey, Guantánamo	0.51907
Estrato4	Pinar del Río, Santiago de Cuba, Ciego de Ávila, Granma	0.46165

obtener información de valor añadido para la toma de decisiones en los estudios salubristas.

Conclusiones

Como resultados de la presente investigación se obtuvo una propuesta de solución para la estratificación de territorios utilizando SIG que contribuye al mejoramiento de la capacidad de gestión de las entidades de salud. En función de los resultados obtenidos se arribó a las siguientes conclusiones:

- La solución desarrollada posibilitó realizar el proceso de estratificación de las provincias de Cuba, identificando cuales de estas presentan mayor riesgo de salud.
- La integración de la solución propuesta al sistema QGIS permite realizar el proceso de estratificación de territorios utilizando indicadores de variada naturaleza representada como objetos geográficos.
- Luego de realizar el proceso de estratificación con los indicadores definidos para el estudio se comprobó que la provincia Ciudad de la Habana es la que mayor riesgo de salud presenta y el estrato formado por las provincias Pinar del Río, Santiago de Cuba, Ciego de Ávila y Granma obtiene el riesgo promedio más bajo en correspondencia con otros estudios realizados.
- Como líneas de trabajo futuro se pretende incorporar temas asociados al tratamiento de datos heterogéneos, de la incertidumbre, el ruido y ausencia de información, además de la integración de otras técnicas de análisis espacial.

Referencias

- Barceló, D. and de Alda, M. J. L. (2008). Contaminación y calidad química del agua: el problema de los contaminantes emergentes. *Panel científicotécnico de seguimiento de la política de aguas. Fundación nueva Cultura del Agua. España.*
- Batista Moliner, R., Coutin Marie, G., Feal Cañizares, P., González Cruz, R., and Rodríguez Milord, D. (2001a). Determinación de estratos para priorizar intervenciones y evaluación en Salud Pública. *Revista Cubana de Higiene y Epidemiología*, 39(1):32–41.
- Batista Moliner, R., Feal Cañizares, P., Coutin Marie, G., Rodríguez Milord, D., and González Cruz, R. (2001b). Guía para la realización del proceso de estratificación epidemiológica. *La Habana: MINSAP.*
- Bivand, R. S., Pebesma, E., and Gómez-Rubio, V. (2013). *Applied Spatial Data Analysis with R*, chapter Hello World: Introducing Spatial Data, pages 1–16. Springer New York, New York, NY.
- Bock, H.-H. (2007). Clustering methods: a history of k-means algorithms. In *Selected contributions in data analysis and classification*, pages 161–172. Springer.

- Bombino Companioni, Y. (2006). *Construcción de tipologías: metodología de análisis para la estratificación según indicadores de salud. Rep Téc Vigilanc [serie en Internet]. 2005 [citado 30 Dic 2009]; 9 (6):[aprox. 14 p.]*.
- Edna, H. V. (2006). Algoritmo de clustering basado en entropía para descubrir grupos en atributos de tipo mixto.
- Fotheringham, S. and Rogerson, P. (2013). *Spatial analysis and GIS*. CRC Press.
- Gil, M. J., Soto, A. M., Usma, J. I., and Gutiérrez, O. D. (2012). Contaminantes emergentes en aguas, efectos y posibles tratamientos. *Producción + Limpia*, 7:52 – 73.
- González Polanco, L. and Pérez Betancourt, Y. G. (2013). La minería de datos espaciales y su aplicación en los estudios de salud y epidemiología. *Revista Cubana de Información en Ciencias de la Salud*, 24(4):482–489.
- Hay, S. I., Battle, K. E., Pigott, D. M., Smith, D. L., Moyes, C. L., Bhatt, S., Brownstein, J. S., Collier, N., Myers, M. F., George, D. B., and Gething, P. W. (2013). Global mapping of infectious disease. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 368(1614).
- INFOMED (2001). Anuario Estadístico de Cuba | Biblioteca Virtual en Salud de Cuba.
- Li, M., Fang, L., Huang, X., and Goh, C. (2015). A spatial-temporal analysis of hotels in urban tourism destination. *International Journal of Hospitality Management*, 45:34 – 43.
- López Caviedes, M. A. (2004). Herramienta para la estratificación de municipios en zonas de riesgo para la salud.
- Molina, J. and García, J. (2006). Técnicas de análisis de datos: aplicaciones prácticas utilizando microsoft excel y weka.
- Pérez Martínez, T. T., Íñiguez Rojas, L., Sánchez Valdés, L., and Remond Noa, R. (2003). Vulnerabilidad espacial al dengue: Una aplicación de los sistemas de información geográfica en el municipio Playa de Ciudad de La Habana. *Revista Cubana de Salud Pública*, 29:0 – 0.
- Perreard, L., Fan, C., Quackenbush, J. F., Mullins, M., Gauthier, N. P., Nelson, E., Mone, M., Hansen, H., Buys, S. S., Rasmussen, K., and others (2006). Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay. *Breast Cancer Res*, 8(2):R23.
- Rodríguez, J. E. R., Blanco, E. A. R., and Camacho, R. O. F. (2013). Clasificación de datos usando el método k-nn. *Vínculos*, 4(1):4–18.

- Rouhani, S. and Hosseini, S. M. (2013). Creating a GIS based data bank of health facilities in Mazandarn province. *Life Science Journal*, 10(9):381–386.
- Ruiz-Shulcloper, J., Guzmán Arenas, A., and Martínez-Trinidad, J. F. (1999). Enfoque lógico combinatorio al reconocimiento de patrones. i. *Selección de Variables y Clasificación Supervisada, Primera edición, Ed. IPN*.
- Shi, X. and Kwan, M.-P. (2015). Introduction: geospatial health research and gis. *Annals of GIS*, 21(2):93–95.
- Suprem, A., Mahalik, N., and Kim, K. (2013). A review on application of technology systems, standards and interfaces for agriculture and food sector. *Computer Standards & Interfaces*, 35(4):355 – 364.
- Tanser, F. C. and Le Sueur, D. (2002). The application of geographical information systems to important public health problems in Africa. *International journal of health geographics*, 1(1):1.
- Valpreda, C. (2007). Sistema de información geográfica (sig)-teledetección y evaluación multicriterio (emc) en un estudio de evaluación de impacto ambiental (eia). In *Memorias XI Conferencia Iberoamericana de Sistemas de Información Geográfica*, pages 29–31.
- Wang, Y. Q. (2014). Meteoinfo: Gis software for meteorological data visualization and analysis. *Meteorological Applications*, 21(2):360–368.