

APLICACIÓN DE INTELIGENCIA DE NEGOCIOS ESPACIAL PARA VISUALIZACIÓN DE ENFERMEDADES DE LOS PACIENTES DE PAIPA.

Application of spatial business intelligence to display diseases in paipa's patients.

Julieth Paola Hurtado Ortiz¹, Gustavo Cáceres Castellanos²

1. Ingeniera de Sistemas y Computación, Magister (E) en Tecnología Informática, Asesora software CNT Sistemas de Información, grupo de Investigación GIMI, Universidad Pedagógica y Tecnológica de Colombia. Tunja. e-mail: paolitaho@yahoo.com
2. Ingeniero de Sistemas Magister en ciencias de la información y las comunicaciones, Especialista en diseño de Sistemas de Auditoría, Cursos de Doctorado en Informática e Inteligencia Artificial,. Docente. grupo de Investigación GIMI, Universidad Pedagógica y tecnológica de Colombia, e-mail: gcaceresc@gmail.com

Recibido:	20	01	2014	Revisado:	24	04	2014
Corregido:	26	05	2014	Aceptado:	30	06	2014

Estilo de referencias: Vancouver APA 6 Harvard ICONTEC

RESUMEN:

Este documento describe el proceso de inteligencia de negocios espacial aplicado en el sector salud a la E.S.E Hospital San Vicente de Paúl de Paipa (Colombia), en donde mes a mes, se deben analizar los sectores de presencia de enfermedades crónicas en los usuarios, pero, al contar con diferentes fuentes de datos se vuelve una tarea manual tediosa y con resultados no verídicos. Debido a lo anterior se planteó la aplicación de tecnología que permitiera la visualización geográfica del índice de enfermedades crónicas, que afectan a los usuarios del hospital con el fin de permitir a los funcionarios tener resultados reales para analizar posibles factores de riesgo, además, de aplicar campañas de promoción y prevención específicas que permitan a la comunidad mejorar su calidad de vida. Para el desarrollo del proceso se usaron herramientas libres que permitieron la unificación de fuentes de información, limpieza de datos, creación de la bodega espacial, aplicación del algoritmo k-means y visualización de información.

Dentro de los principales resultados se obtuvo el mapa del municipio de Paipa con círculos de colores según enfermedad crónica específica y se observa en que veredas hay presencia de estas patologías.

Palabras Clave: *inteligencia de negocios; inteligencia de negocios espacial; sistemas de información geográfica; salud; minería de datos espacial; algoritmo k-means.*

ABSTRACT

This document describes the process of spatial business intelligence applied in the health sector to the ESE Hospital San Vicente of Paul of Paipa (Colombia), in which month should be analyzed the presence of chronic diseases in users in different sectors, but having different sources of data becomes a tedious and manual task untrue results. Because of this the application of technology to enable the visualization of geographic rate of chronic diseases affecting hospital users in order to allow officials to have real results to analyze potential risk factors, and implementing campaigns raised promotion and specific prevention to enable the community to improve their quality of life. For free process development tools that allowed the unification of information sources, data cleansing, creating spatial datamart, application of k-means algorithm and visualization. the main results are presented in Paipa's Map with colored squares that represented the specific chronic diseases..

Keywords: *business intelligence; spatial business intelligence; geographic information system; health; spatial data minning; algorithm k-means..*

INTRODUCCION

La E.S.E Hospital San Vicente de Paúl, es una Institución prestadora de servicios de Salud de primer nivel de atención para el municipio de Paipa, Boyacá, Colombia, que cuenta con un software Institucional llamado CNT Pacientes, en donde, diariamente se registra la conducta medica de cada uno de los 50000 usuarios que tiene el Hospital, sin embargo, para el análisis de información es difícil combinar otras fuentes y usar otros tipos de datos que permitan un análisis más detallado del índice de enfermedades, que están afectando a los habitantes del municipio teniendo en cuenta su ubicación geográfica de residencia. Según información recibida de la Gerente de la Institución es de vital importancia tener una herramienta o método que permita unificar esta información para apoyar la toma de decisiones, ya que, la E.S.E cuenta con el desarrollo de programas de salud, que permiten asistencia técnica a los pacientes en sus hogares, promoción y prevención de enfermedades, análisis de riesgo entre otros; uno de estos programas se denomina APS (Atención primaria en salud) que busca llegar al hogar de cada usuario y tratar su enfermedad, es decir, llevar el hospital a la casa, sin embargo, este programa

se ha visto afectado por falta de análisis de información que le permita a los profesionales de la salud identificar antes de la visita las medidas a tomar, prevención y posibles riesgos de la enfermedad.

Para ayudar a la E.S.E con la unificación de la información y posterior análisis, se hizo una investigación con respecto a herramientas tecnológicas que ayuden a sus funcionarios a tener información precisa y en tiempo real de sus pacientes; es por ello, que se desarrolló una investigación del uso de Inteligencia de negocios espacial en hospitales y se evaluó su aplicación en Paipa, dentro de los proyectos realizados a nivel mundial se encuentran:

Tabla 1. Proyectos de aplicación de inteligencia de negocios espacial en el sector salud.

NOMBRE DEL PROYECTO	PAÍS- INSTITUCION	HERRAMIENTA UTILIZADA	ÁREA EN QUE SE APLICA BI	CONCLUSIONES
GeoVisualAnalytics in the Public Health Sector, (Böhm, Mehler-bicher, & Fenchel, 2011)	Alemania	JasperSoft, Palo, Pentaho, Postgres, Postgis; Geomondrian.	Área Asistencial: prototipo de geo-Bi involucrando a la comunidad y al personal asistencial	
Spatial Data Integration and Analysis with Spatial Intelligence, (Bing She , LIEMARS, Wuhan Univ., Wuhan, China , Xinyan Zhu, 2010)	China	China Geo-Explorer, a web based platform for spatial intelligence	Área Asistencial:	Protótipo en la Web para permitir al usuario final generar el mapa con la información requerida según lo que desee saber en diferentes sectores incluyendo salud..
Easier surveillance of climate-related health vulnerabilities through a Web-based spatial OLAP application. (Bernier, Gosselin, Badard, & Bédard, 2009)	Canadá	Geokettle JMap Spatial	Área Asistencial: identificar las vulnerabilidades de salud por los cambio de clima	Se determinaron las causas de enfermedades específicas y aplicar medidas de prevención a los usuarios que vivían en los sectores estudiados
Spatial analysis of perinatal mortality rates with geographic information systems in Kocaeli, Turkey	Department of Geomatics, Engineering Faculty, Kocaeli University, Turquía	No definida	Área Asistencial: Análisis espacial de las tasas de mortalidad perinatal.	algunos factores demográficos y socio-económicos están asociados con las tazas de mortalidad

<p>(Arslana, Corresponding author contact information, E-mail the corresponding author, M.S. Çepnia, 2012)</p>			
<p>Using geo-business intelligence to improve quality of life. (Angelaccio, Dept. of Enterprise Eng., Univ. of Rome Tor Vergata, Rome, Basili, A. ; Buttarazzi, & Liguori, 2012)</p>	<p>Universidad de Roma</p>	<p>No definida</p>	<p>Área Asistencial: busca mejorar la calidad de vida de los pacientes.</p>

Fuente: los autores

En el área de la salud, son muy pocos proyectos desarrollados con la aplicación de inteligencia de negocios espacial enfocados al área asistencial, pero, es importante destacar que los resultados obtenidos han sido una contribución para mejorar la calidad de vida de las personas. Es por ello, que el objetivo principal de la aplicación de inteligencia de negocios espacial en el sector salud es combinar los datos de transacciones diarias sobre diagnósticos de los pacientes, con datos espaciales para permitir hacer análisis detallados a la información generada.

La aplicación de esta tecnología, permitirá a los funcionarios de la E.S.E conocer de manera anticipada las enfermedades crónicas que padecen los habitantes en los barrios y veredas que visitan a diario para hacer brigadas de salud y crear estrategias que permitan capacitar a los usuarios, llevar los medicamentos requeridos, además de crear conciencia por medio de campañas de prevención que informen a los ciudadanos sobre la enfermedad con mayor índice que se presenta en su sector.

MARCO TEÓRICO

Inteligencia de negocios espacial

Para (Ponjavic, 2012) La inteligencia de negocios espacial se define como un componente que tiene por objetivo examinar, organizar y deducir procesos de negocio para el mejoramiento de la compañía, apoyándose del componente geográfico además de la información transaccional diaria de la misma. Esta necesidad surge debido a que muchos de los datos que encontramos hoy en día están relacionados a una ubicación geográfica. (T. D. Ndie, C. Tangha, 2012)

Almacenes de datos espaciales:



Los almacenes de datos contienen información relevante que puede estar relacionada con su ubicación geográfica por lo que han nacido los almacenes de datos espaciales, los cuales, tiene como objetivo agrupar información alfanumérica y geográfica relevante y le dan la posibilidad al usuario final de hacer análisis de dicho agrupamiento.(Viswanathan, 2012), esto con el fin de generar el cubo espacial del cual se desprenden los resultados requeridos; la cantidad de información que se maneja en estas bodegas es grande por lo que los sistemas que las contienen deben tener mayor procesamiento, memoria buscando mejorar su rendimiento en las consultas por lo que se han definido índices específicos para su procesamiento.(Siqueira, T.L., 2012)

Minería de datos espacial

La minería de datos espaciales es definida por (Ryu, 2005), como el proceso de análisis de datos de una bodega espacial con el uso de algoritmos y herramientas que permiten encontrar relaciones útiles en objetos geográficos. El principal objetivo de la minería de datos es la extracción de conocimiento relevante de los datos espaciales (M. Esther, 2009)

Existen diferentes técnicas para la minería de datos espaciales como se observa a continuación:

Tabla 2. Técnicas de minería de datos espacial

Técnica	Algoritmo
Generalización	Jerarquías
Agrupación (Clustering)	K-means, Fuzzy c-means, Cluster assigner
Asociación especial	A priori
Aproximación y agregación	
Reconocimiento de patrones	Los algoritmos utilizados hacen reconocimiento de imágenes y texto.

Fuente: autores.

Algoritmo K-means

Este algoritmo es definido como un algoritmo de agrupamiento, se eligen un numero deseado de centros (clúster) de los grupos denominados llamados R, y K-means mueve los centros iterativamente hasta minimizar la varianza dentro del cluster, los elementos dentro de un mismo clúster siempre serán más similares entre sí, comparándolos con los elementos que pertenecen a otros grupos. (Hastie, Tibshirani, & Friedman, 2008)

Después de determinar un centro inicial por cada grupo el algoritmo K-Means hace sus iteraciones específicamente en dos pasos:

- Por cada centro se identifica el conjunto de puntos. Cuando finaliza la asignación se tiene una nueva partición de datos.
- Se recalcula con la media aritmética de todos los puntos asignados, el nuevo centro de cada grupo.

El algoritmo termina cuando no hay cambio en los valores de los centros.

MATERIALES Y MÉTODO

Materiales. Los materiales que se usaron para el desarrollo de la Investigación fueron:

- Base de datos en SQL Server del software institucional de la E.S.E,
- Archivos planos y en hojas de cálculo de información básica de los pacientes
- Archivos en extensión shape de los mapas de barrios y veredas de Paipa

Para la limpieza de los datos se usó la herramienta libre Geokettle (OSGeo-Live, 2011), la bodega de datos se trabajó en Postgresql (Group, 2014) y para el llenado de la bodega de datos se usó Qgis (Gis, n.d.), Geokettle y Postgis(PostGis, 2012).

Para en análisis de la información se aplicó el algoritmo de minería de datos k-means, con la herramienta Knime, y para la visualización de datos se usó google fushion tables.

Método

A continuación se describe paso a paso del proceso realizado.

Se toma la información de las diferentes fuentes de datos que se encuentran en el hospital, secretaria de salud del municipio y secretaria de planeación (información geográfica)

Limpieza de datos de la información básica de los pacientes como dirección, edad, número de documento, entre otras, haciendo un cruce de la información que se encuentra registrada en la base de datos del hospital, con la información que posee la secretaria de salud de aseguramiento de cada ciudadano.

Figura 1. Limpieza de información.

52639041	1103279	GAMBOA	SOLANO	ADAN	RURAL	VEREDA CRUZ DE BONSA
45364079	1103279	GAMBOA	SOLANO	ADAN	RURAL	VEREDA CRUZ DE BONSA
47780045	1103279	GAMBOA	SOLANO	ADAN	RURAL	VEREDA CRUZ DE BONSA
50336147	1103279	GAMBOA	SOLANO	ADAN	RURAL	VEREDA CRUZ DE BONSA
40597045	1103279	GAMBOA	SOLANO	ADAN	RURAL	VEREDA CRUZ DE BONSA
36270097	1103279	GAMBOA	SOLANO	ADAN	RURAL	VEREDA CRUZ DE BONSA
36270097	1103279	GAMBOA	SOLANO	ADAN	RURAL	VEREDA CRUZ DE BONSA
26649077	1103279	GAMBOA	SOLANO	ADAN	RURAL	VEREDA CRUZ DE BONSA
26649077	1103279	GAMBOA	SOLANO	ADAN	RURAL	VEREDA CRUZ DE BONSA
26649077	1103279	GAMBOA	SOLANO	ADAN	RURAL	VEREDA CRUZ DE BONSA
31207011	1103279	GAMBOA	SOLANO	ADAN	RURAL	VEREDA CRUZ DE BONSA
31207011	1103279	GAMBOA	SOLANO	ADAN	RURAL	VEREDA CRUZ DE BONSA
38436001	1103279	GAMBOA	SOLANO	ADAN	RURAL	VEREDA CRUZ DE BONSA
54323005	1103279	GAMBOA	SOLANO	ADAN	RURAL	VEREDA CRUZ DE BONSA
54323005	1103279	GAMBOA	SOLANO	ADAN	RURAL	VEREDA CRUZ DE BONSA
134042001	1103064	BRIJALDO	PEDRAZA	GONZALO	RURAL	BONSA
131783079	1103064	BRIJALDO	PEDRAZA	GONZALO	RURAL	BONSA
137661006	1103064	BRIJALDO	PEDRAZA	GONZALO	RURAL	BONSA
137661006	1103064	BRIJALDO	PEDRAZA	GONZALO	RURAL	BONSA
58313020	1103279	GAMBOA	SOLANO	ADAN	RURAL	BONSA

Fuente: Autores

Algunos de los errores más comunes que se presentaron era la manera como digitaban el lugar de residencia del paciente, ya sea el nombre del barrio o la vereda.

Creación de la bodega de datos espacial, teniendo en cuenta las dimensiones y la tabla de hechos. Para la creación de la bodega se tuvo en cuenta las siguientes dimensiones y tabla de hechos:

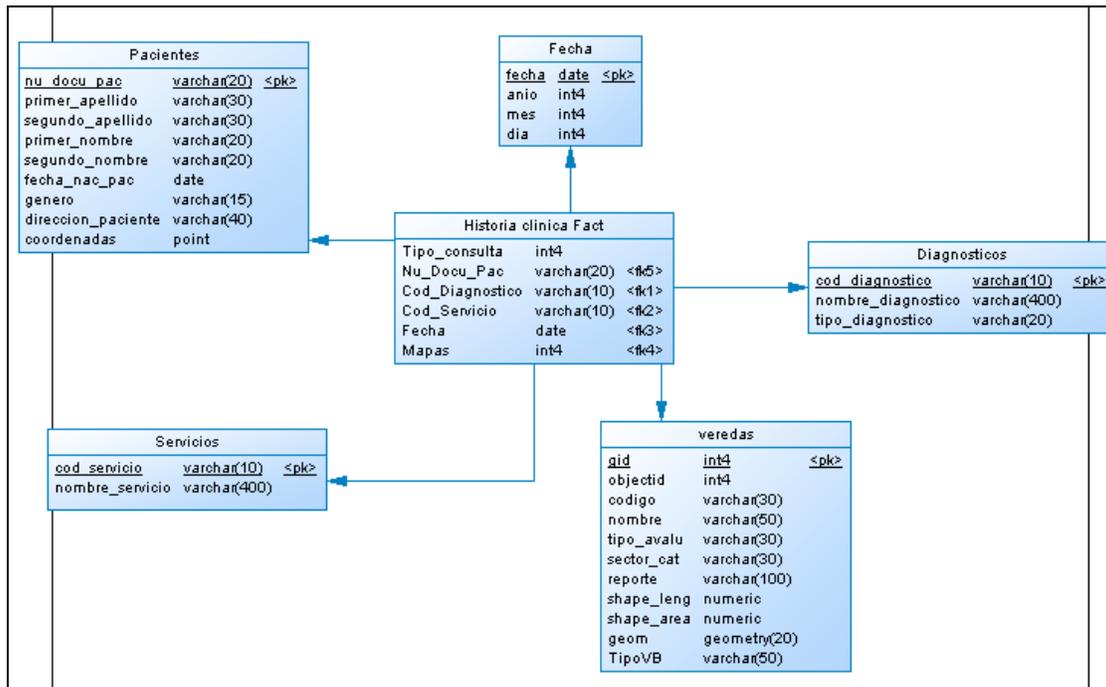
Tabla 3. Dimensiones y tabla de hechos

Dimensión	Descripción
Pacientes	Guarda los datos básicos del paciente como número de documento, nombres, apellidos, género, Fecha de nacimiento, dirección.
Diagnósticos	Guarda el código y la descripción de los diagnósticos del CIE10.
Servicios	Guarda el código y descripción de cada uno de los servicios de la E.S.E Hospital San Vicente de Paul de Paipa en donde se almacenan los servicios que son la vía de ingreso del paciente para ser evaluado por el profesional de la salud.
Fecha	Guarda la fecha, día, mes y año en la cual el paciente es atendido
Ubicación_Geografica	Es la dimensión espacial que guarda la información de los barrios y veredas del municipio esportadas desde el mapa (Archivo Shape)
Tabla de Hechos	Almacena el documento del Paciente, su diagnóstico, el servicio en el cual el profesional de la salud evalúa su estado y define diagnóstico, fecha de atención, código del barrio o vereda donde vive y un tipo consulta que se refiere a la evaluación del diagnóstico si es primera vez o de control.

Fuente: Autores

A continuación se presenta el modelo físico de la bodega de datos espacial.

Figura 2. Bodega de datos espacial

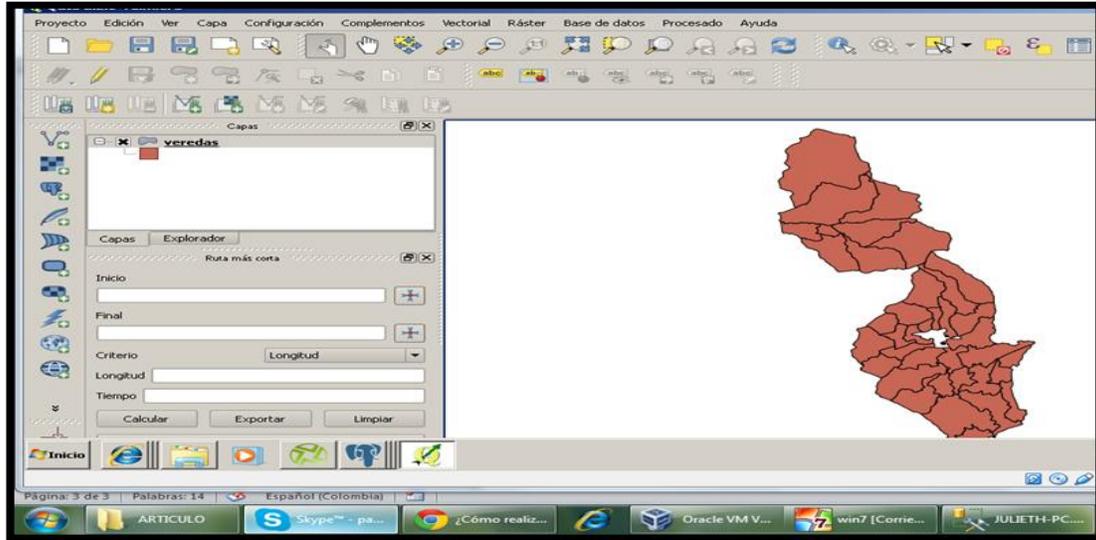


Fuente. Autores

Llenado de la bodega de datos por medio de las herramientas Geokettle, Postgis y Qgis.

Para el llenado de la bodega de datos espacial se usó Qgis y Postgis para pasar los datos de los mapas con extensión shape a la base de datos, las demás dimensiones se llenaron por medio de la herramienta Geokettle.

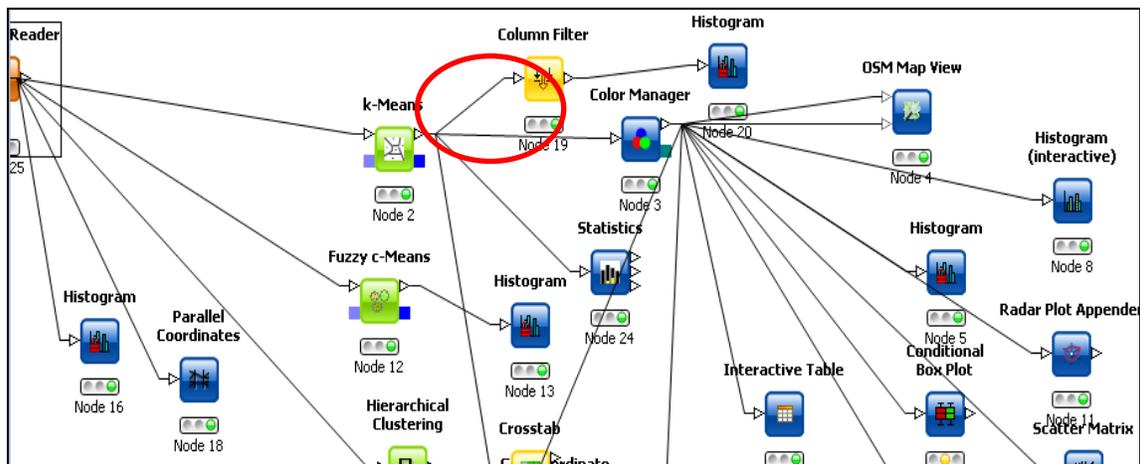
Figura 3. Mapa de veredas de Paipa en Archivo Shape.



Fuente. Autores

Aplicación del algoritmo K-means en la bodega de datos espacial y análisis de resultados.

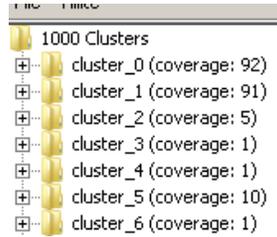
Figura 5. Programación del algoritmo K-means en la herramienta Knime



Fuente. Autores

Para la aplicación del algoritmo K-means se tuvo en cuenta variables como latitud, longitud y diagnóstico médico; el algoritmo redefine centros hasta que encuentre valores diferentes dentro de éstos.

Figura 6. Resultado de los clúster luego de aplicar los algoritmos



Fuente. Autores

Según los resultados obtenidos por el algoritmo K-means se pudo observar que se crean diferentes núcleos llamados clúster, los cuales, contenían información que se encuentre repetida entre los datos que se evaluaron.

Figura 7. Visualización de clústeres luego de la aplicación del algoritmo k-means

1,73117763	7/30/2009	PRINCIPAL	VEREDA EL TEJAR PAIPA	ROMERO	CELY	MARTHA	SIN OTRA ESPECIFICACION	NO_CRONICA	cluster_106
1,73117763	12/8/2009	PRINCIPAL	VEREDA EL TEJAR PAIPA	RODRIGUEZ	MATEUS	FILEMON	PARASTOSIS INTESTINAL SIN OTRA ESPECIFICACION	NO_CRONICA	cluster_106
1,73117763	2/21/2008	PRINCIPAL	VEREDA EL TEJAR PAIPA	RODRIGUEZ	MATEUS	FILEMON	PARASTOSIS INTESTINAL SIN OTRA ESPECIFICACION	NO_CRONICA	cluster_106

Visualización de las enfermedades crónicas en el mapa de Paipa.

A continuación se presentan diferentes vistas del mapa de Paipa con las enfermedades que se presentan en cada una de las veredas y barrios por medio de cuadros de colores.

Figura 8. Vista 1 del mapa de Paipa con las enfermedades que se presentan en cada sector.

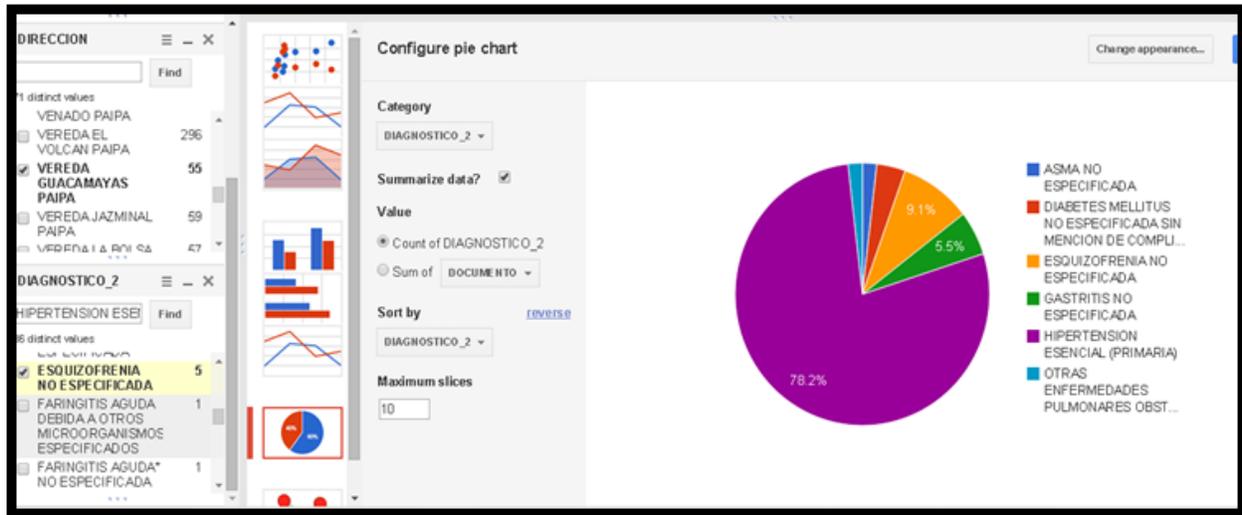


Fuente. Autores

RESULTADOS

A continuación se presentan los resultados generados en el mapa.

Figura 9. Vista general del mapa de Paipa con los sectores donde se presenta la diabetes.



Fuente. Autores

Los colores en el mapa representan cada una de las enfermedades crónicas que se presentan en los diferentes sectores y en la gráfica se presenta un porcentaje de incidencia teniendo en cuenta el total de habitantes de cada barrio o vereda.

DISCUSION DE LOS RESULTADOS

A continuación se hace un análisis con el apoyo del secretario de salud del municipio de los resultados más relevantes obtenidos:

En la vereda Caños es donde más se presenta Hipertensión arterial primaria con un número de 85 pacientes que la padecen.

La hipertensión arterial secundaria se presenta en mayor grado en el barrio centro de Paipa.

En las veredas en donde más se presenta la diabetes mellitus es en la vereda sativa seguida de las veredas pastoreos y río arriba.

Diagnósticos como leucemia se presentan con mayor índice en la vereda el retiro.

Las veredas que presentan menos índice de enfermedades crónicas son Peña Amarilla, Peña Blanca y Quebrada Honda, además de los barrios Corinto, la Paradera y las Quintas.

CONCLUSIONES



Por medio de esta herramienta tecnológica, se puede saber de antemano, las enfermedades que más afectan a cada una de las veredas del municipio, para tomar medidas respectivas en cuanto a la prevención de la enfermedad. En algunas veredas se hace urgente, el análisis de planes de acción, para aplicarlo a sus ciudadanos y el manejo de las enfermedades crónicas en tiempo real.

NOTAS BIBLIOGRAFICAS.

- 1) Angelaccio, M., Dept. of Enterprise Eng., Univ. of Rome Tor Vergata, Rome, I., Basili, A. ; Buttarazzi, B., & Liguori, W. (2012). Using Geo-Business Intelligence to improve Quality of Life. *Satellite Telecommunications (ESTEL), 2012 IEEE First AESS European Conference on*, 1–6. Retrieved from <http://ieeexplore.ieee.org/xpl/abstractKeywords.jsp?arnumber=6400196>
- 2) Arslana, Corresponding author contact information, E-mail the corresponding author, M.S. Çepnia, N. E. (2012). Spatial analysis of perinatal mortality rates with geographic information systems in Kocaeli, Turkey. *To Determine the Spatial Patterns of Perinatal Mortality in Kocaeli, Turkey Using Geographic Information Systems (GIS); to Examine Whether Regional Differences Exist for the Period Selected; and Whether These Differences Are Linked to Regional Risk Factor*, 127, 369–379. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0033350612004581>
- 3) Bernier, E., Gosselin, P., Badard, T., & Bédard, Y. (2009). Easier surveillance of climate-related health vulnerabilities through a Web-based spatial OLAP application. *International Journal of Health Geographics*, 8(1), 18. doi:10.1186/1476-072X-8-18
- 4) Bing She , LIEMARS, Wuhan Univ., Wuhan, China , Xinyan Zhu, S. B. (2010). Spatial Data Integration and Analysis with Spatial Intelligence. In *International Conference on Geoinformatics* (pp. 1–6).
- 5) Böhm, K., Mehler-bicher, A., & Fenchel, D. (2011). GeoVisualAnalytics in the Public Health Sector, 291–294.
- 6) Gis, Q. (n.d.). Documentación QGIS.
- 7) Group, T. P. G. D. (2014). Postgresql. Retrieved from <http://www.postgresql.org/>
- 8) Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning Data Mining, Inference, and Prediction* (pp. 479–498). New York.
- 9) M. Esther, H. K. and J. S. (2009). Algorithm for Characterization and Trend Detection in Spatial Databases. In *4th Int. Conference on Knowledge Discovery and Data Mining (KDD 98)*.
- 10) OSGeo-Live. (2011). GeoKettle.
- 11) Ponjavic, A. K. and M. (2012). Geoportal as decision support system with spatial data warehouse. In *Federated Conference in Computer Science and Information Systems (FedCSIS)* (pp. 915–918).
- 12) PostGis. (2012). PostGis Spatial and Geographic objects for PostgreSQL. Retrieved from <http://postgis.net/>
- 13) Ryu, C. E. and T. (2005). A database clustering methodology and tool. *Information Sciences*, 17, 29–59.
- 14) Siqueira, T.L., et al. (2012). The SB-index and the HSB-index: efficient indices for spatial data warehouses. *Geoinformatica. (ACM)*, 165–205.
- 15) T. D. Ndie, C. Tangha, and B. J. (2012). GIAPSCart: A geo-intelligence application based on semantic cartography. In *6th International Conference in Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)* (pp. 106–113).
- 16) Viswanathan, G. and M. S. (2012). User-centric spatial data warehousing: a survey of requirements and approaches. *International Journal of Data Mining, Modelling and Management*.

COMO CITAR ESTE ARTICULO:

Hurtado Ortiz JP, Caceres Castellanos GC. Aplicación de inteligencia de negocios espacial para visualización de enfermedades de los pacientes de Paipa. Rev.salud.hist.sanid.on-line 2014; 9(1): 37-50 (Enero-junio). Disponible en <http://www.shs.histosaluduptc.org/> Fecha de consulta ()

*Los textos publicados en esta revista pueden ser reproducidos citando las fuentes.
Todos los contenidos de los artículos publicados, son responsabilidad de sus autores.*

Copyright. Revista Salud Historia y Sanidad ©

Grupo de Investigación en Salud Pública GISP-UPTC
Grupo de investigación Historia de la salud de Boyacá.

Tunja 2014