

Uso de la base de datos PostgreSQL y software estadístico R para análisis de trayectorias

[Using the PostgreSQL database and R statistical software for path analysis]

Gary Reyes Zambrano, Kevin Baque Puya, Ailyn Bones Flores, Kiara Cevallos Macias, Jully Estrada Rumiguano, and Katheryn Velepucha Iñaguazo

Facultad de Ciencias Matemáticas y Físicas, Universidad de Guayaquil, Guayaquil, Ecuador

Copyright © 2016 ISSR Journals. This is an open access article distributed under the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT: Despite the technological advances in the system for determining spatial coordinates in recent years, still they have minimal defects, therefore, the GPS not as evidenciable would be considered as the localization error is denoted in a few meters unlike what actually recorded. That's why, to analyze a vehicle trajectory, conducted the study of a set of records, downloaded from a website, which contain GPS trajectories 10,357 taxis during the period from 2 to 8 February 2008 in the city of Beijing. Fields containing the data set are: taxi identifier, date and time, longitude and latitude. a corresponding geographical map was also used to the city mentioned, which was unloaded in download.geofabrik.de, and was implemented by the Geographic Information System QGIS, the same which allows a geographical display of all vehicular paths. The resulting information will be integrated into the R statistical analysis tool for the user to visualize in statistical tables movements of taxis at certain times or points of the city, so it can be displayed in bar statistics frequency fields and time. These data obtained have greater relevance because, you can work and manage an accurate and concrete way.

KEYWORDS: dataset, beijing, database, taxis, geographical point.

RESUMEN: A pesar de los avances tecnológicos en el sistema de determinación de coordenadas espaciales de los últimos años, aún se tienen mínimos defectos, por lo tanto, se consideraría al GPS no tan evidenciable ya que el error de localización se denota en pocos metros a diferencia de lo que realmente registra. Es por eso que, para analizar una trayectoria vehicular, se llevó a cabo el estudio de un conjunto de registros, descargados desde una página web, los cuales contienen trayectorias GPS de 10.357 taxis durante el periodo del 2 al 8 de febrero del 2008 dentro de la ciudad de Beijing. Los campos que contiene el conjunto de datos son: el identificador del taxi, la fecha y hora, longitud y latitud. También se usó un mapa geográfico correspondiente a la ciudad mencionada, el cuál fue descargado en download.geofabrik.de, y fue implementado mediante el Sistema de Información Geográfico QGIS, el mismo que permite tener una visualización geográfica de todas las trayectorias vehiculares. La información resultante se integrará a la herramienta de análisis estadístico R para que el usuario visualice en cuadros estadísticos los movimientos de los taxis en determinadas horas o puntos de la ciudad, así mismo se podrá visualizar en barras estadísticas los campos de frecuencia y hora. Estos datos a obtenerse tendrán mayor relevancia, ya que, se podrá trabajar y manejar de una manera exacta y concreta.

PALABRAS CLAVES: conjunto de datos, beijing, base de datos, taxis, punto geográfico.

1 INTRODUCCIÓN

La presente investigación se refiere al estudio de un conjunto de datos representado por trayectorias vehiculares en un periodo de 6 días de la ciudad de Beijing.

Desde tiempos antiguos el hombre se incursionaba en viajes, ya sea por el medio en que se transportaba, es decir por la trayectoria que se guiaban por medio marítimo o por medio terrestre. Para poderse incursionar a diversas partes de las regiones geográficas sin llegar a perderse, incluyeron diversos instrumentos. Probablemente el método más antiguo y primitivo que utilizaron para no perderse fue el de colocar piedras en el camino o hacer marcas en los árboles, de forma tal que le permitieran regresar después sobre sus pasos. Esto quiere decir que desde hace tiempo se ha utilizado medios ya sea un poco ortodoxo, para poder encontrar su localización y no llegar a perderse preestableciendo puntos de referencia tanto de ida como de llegada, pero al final eran de una gran utilidad al hombre estos tipos de instrumentos que han ido evolucionando con el pasar del tiempo, de manera impresionante con la tecnología. En la antigüedad se utilizaba la piedra, luego vino lo que llamamos brújula y en la actualidad lo conocemos como GPS.

GPS es un sistema que tiene como objetivo la determinación de las coordenadas espaciales de puntos respecto de un sistema de referencia mundial. Los puntos pueden estar ubicados en cualquier lugar del planeta, pueden permanecer estáticos o en movimiento y las observaciones pueden realizarse en cualquier momento del día[1].

La información utilizada para el análisis fue recolectada con ayuda del GPS, lo que permitió tener conocimiento acerca del estado de un vehículo es decir que pasa cuando un vehículo está en reposo o en marcha.

Los vehículos en estado de reposo, transmiten al suelo, solamente cargas gravitacionales, o sea, peso propio más carga. Cuando adquieren velocidad entran en estado dinámico, y como resultado, generan fuerzas inerciales, cuyas magnitudes dependen de la masa y la velocidad, es decir, ganan energía cinética con sus componentes dimensionales de Fuerza por Longitud. Estas nuevas fuerzas inerciales, dinámicas, importantes, se deben calcular, y considerar para el diseño de las vías y para el control del tránsito vehicular[2].

Concomitantemente, las trayectorias pueden constituirse en una de las fuentes esenciales para investigar la realidad. Incluir la lectura detenida de biografías, de relatos de vida, plasmados en aquellos registros escritos que reflejan una trayectoria humana o que dan noticia de la visión que los sujetos poseen de la realidad y de su propia existencia, ofrece elementos que contribuyen a una mayor comprensión del entramado del sujeto con lo social[3].

El objetivo principal de esta investigación es identificar la ubicación y la acogida de los taxis en los determinados lugares, momentos o días, a través de una justa combinación de cuatro herramientas, tres gestoras de datos y una estadística.

Recientemente, la usabilidad de los Sistemas de Información Geográfica ha ido ganando terreno en numerosas áreas de la sociedad. Si se analiza la definición dada por el NCGIA "un SIG es un sistema de hardware, software y procedimientos elaborados para facilitar la obtención, gestión, manipulación, análisis, modelado, representación y salida de datos espacialmente referenciados, para resolver problemas complejos de planificación y gestión"[4].

Los SIG son imprescindibles en Geografía de la cual depende de ellos como la Estadística depende de los paquetes estadísticos. Hoy en día no puede mantenerse al margen la localización geográfica al analizar datos. Además, esta implicación interactúa entre los diversos campos, de manera que el clima en un momento determinado lleva a unos asentamientos en los que se obtuvieron unos determinados yacimientos arqueológicos, o las rutas comerciales las cuales dependen de la topografía del terreno. Es por ello que se debe tener unos conocimientos básicos de Estadísticas y de R.

La aplicación de la herramienta estadística R en esta investigación es de gran importancia ya que permite analizar ampliamente los datos de las trayectorias vehiculares, con la amplia gama de métodos estadísticos que nos proporciona, una ventaja que a todos gusta es el hecho de ser un software libre, es decir, gratuito.

Desde el punto de vista estadístico es importante aprovechar y modelar esta nueva forma de generar información, donde se hacen relevantes los métodos descriptivos que nos ayudan a resumir y calificar la información y los métodos de análisis espacial, tanto del punto de vista inferencial como la aplicación de métodos de interpolación. Este trabajo pretende a través de un caso de estudio integrar de forma protocolar dichos procedimientos y verificar la eficiencia de los mismos[5].

2 RESULTADOS

La obtención de resultados en este análisis se dio uso de Postgresql con Postgis, los mismo que almacenan la información correspondiente al conjunto de datos de 10.357 taxis, estos datos son recibidos por el sistema QGIS, la misma que procede a visualizar un mapa geográfico de la ciudad de Beijing con los puntos de los taxis y las diferentes trayectorias. Con esta información continuamos con la integración de la herramienta estadística R, la misma que mediante sentencias declaradas nos lanza resultados interpretados por barras estadísticas. Esto facilita al usuario interpretar toda la información utilizada de una manera más sencilla, clara y específica.

Para visualizar en el mapa las trayectorias del conjunto de datos de taxis se crea un punto geométrico en la tabla, es decir, una vez cargado los registros en la tabla agregamos una columna con la sentencia:

```
Select addgeometrycolumn ('nombre de la tabla', 'nombre de la columna', '4326', 'point', 2, true)
```

El punto, como objeto o figura geométrica, se considera que no tiene dimensiones y se usa para indicar una posición en el espacio[6].

Una vez creada la columna de punto geométrico agregamos los datos generando la siguiente sentencia, recordemos que el punto geométrico trabaja con la columna longitud y latitud.

```
Update "nombre de la tabla" set geom = st_setsrid(st makepoint(a.longitude, a.latitude),4326) from(slect * from nombre de la tabla) as a where nombre de la tabla.unixtime = a.unixtime
```

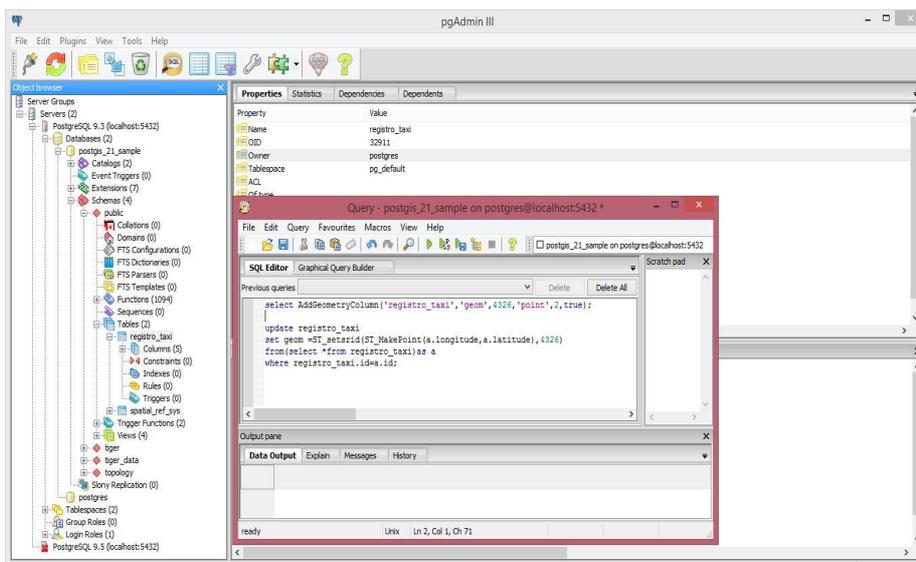


Figura 1: Ejecución de la sentencia para crear el punto geométrico.

Los primeros resultados que se obtiene al iniciar este análisis de trayectorias es en Postgresql, al momento que cargamos el conjunto de datos, como resultado se visualiza la tabla con las columnas establecidas por el dataset y la columna agregada.

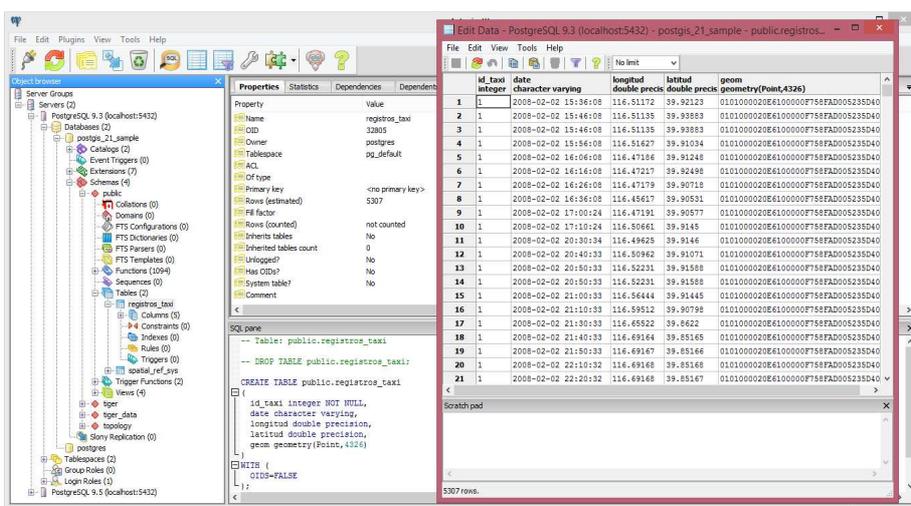


Figura 2: Conjunto de datos cargados con la columna del punto geométrico

QGIS utiliza la librería OGR para leer y escribir formatos de datos vectoriales, incluyendo formatos de archivos shape de ESRI, MapInfo y MicroStation, base de datos AutoCAD DXF, PostGIS, SpatiaLite, Oracle Spatial y MSSQL Spatial, y muchos más. La implementación de vector de GRASS y PostgreSQL es suministrado por complementos de proveedor de datos nativo de QGIS[7].

Para la visualización de las trayectorias en el sistema QGIS se debe conectar el gestor de datos con el sistema geográfico, esta conexión se la realiza en el QGIS mediante una configuración rápida y sencilla.

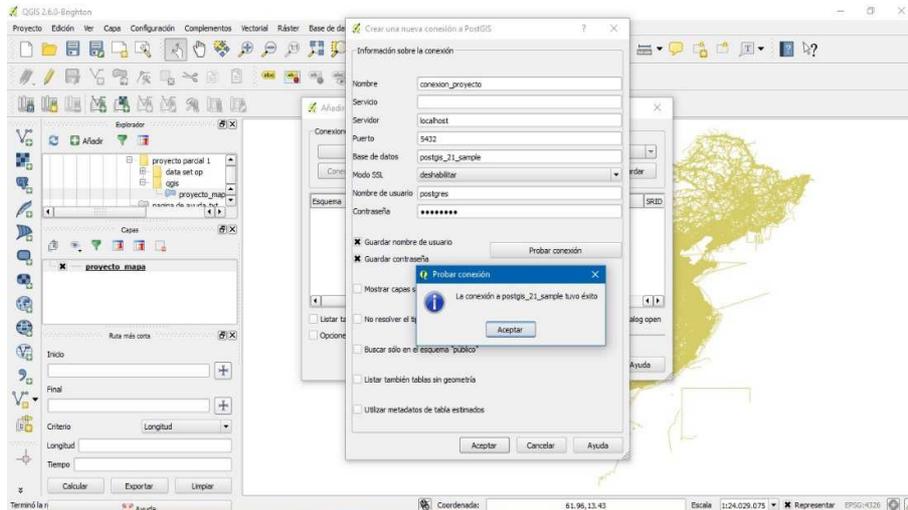


Figura 3: Conexión entre el gestor de datos y sistema geográfico.

Los datos pueden ser importados a PostgreSQL/PostGIS utilizando varias herramientas, incluyendo el complemento SPIT y la herramienta de línea de comandos shp2pgsql y ogr2ogr[7].

El sistema Geográfico QGIS también nos da un resultado, este se lo podrá conseguir una vez importado el mapa de la ciudad de Beijing y establecida la conexión con la tabla de registros en Postgresql.

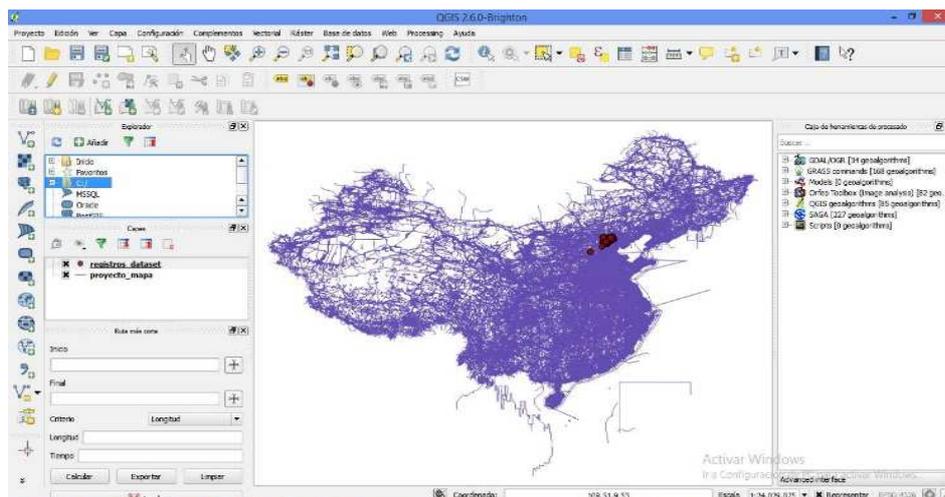


Figura 4: Trayectorias en el mapa con el sistema QGIS

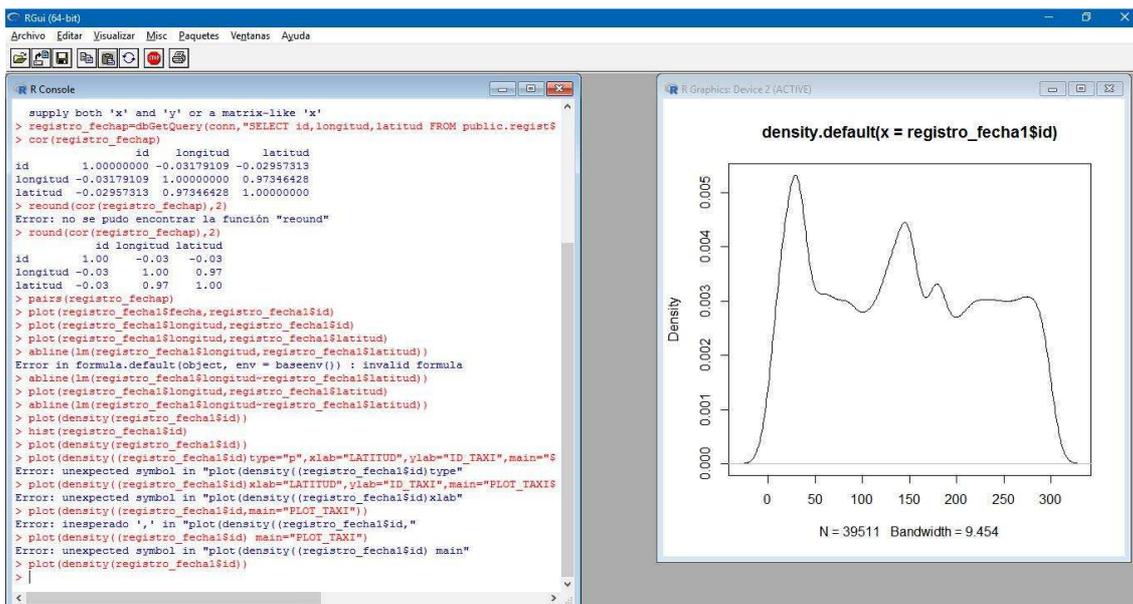


Figura 7: Resultado estadísticos de las trayectorias vehiculares en R

En las siguientes imágenes podremos visualizar la cantidad de taxistas en determinadas horas, como por ejemplo en el registro de taxi en la fecha: 2008-02-02 podemos deducir que en la hora 18 hay una cantidad de taxista de aproximadamente 3700, y así se deduciría las otras fechas. Esto daría al cliente una idea de la demanda de taxistas en determinadas horas.

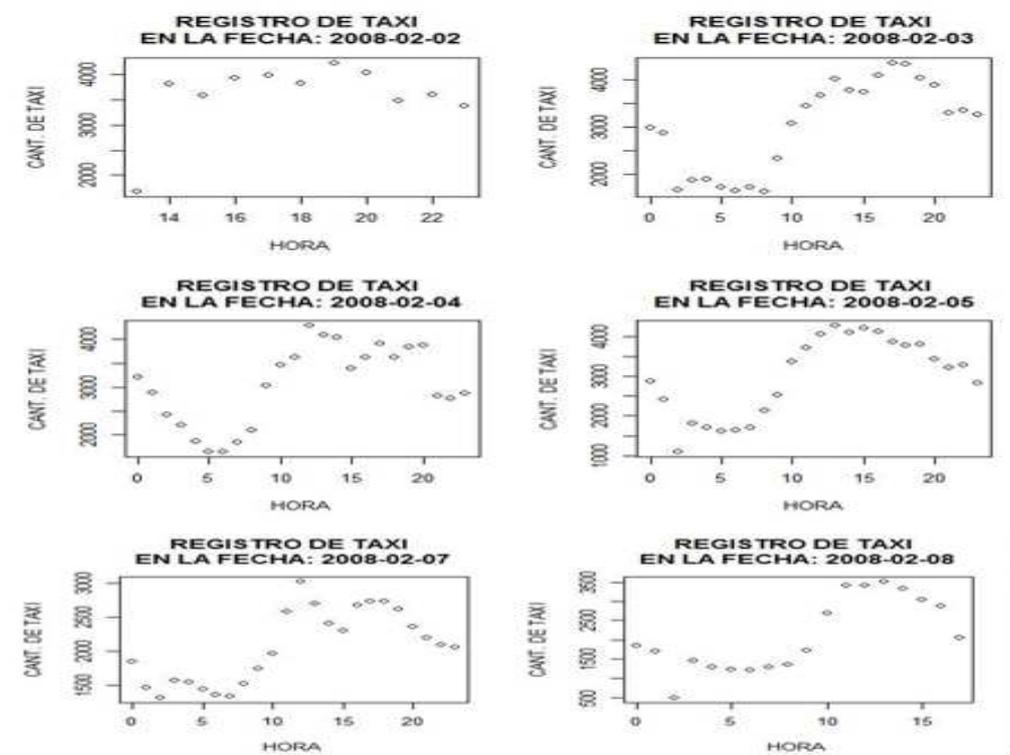


Figura 8: Resultado estadísticos de las trayectorias vehiculares en R

3 DISCUSIÓN

Existen numerosas aplicaciones que permiten el manejo de información de entidades en movimiento, especialmente para el control de flotas. El análisis se limita a la representación de información de la flota en mapas y desarrollo de pequeños escenarios de toma de decisiones. En este trabajo se establece una correspondencia entre el control de flotas y la investigación asociada a la representación de objetos móviles y procesamiento de su información desde una perspectiva espacio – temporal y semántica. El objetivo principal es reflexionar sobre la importancia de aplicar el conocimiento existente en esta área al control de flotas e identificar líneas abiertas para investigación futura[4].

Hoy en día, existen diversos estudios para la medición de la movilidad en distintos entornos, utilizando desde encuestas escritas hasta dispositivos RFID. Sin embargo, estos estudios resultan ser costosos en cuanto a equipamiento, implantación, o recolección, ocasionando que su puesta en marcha no escale más allá del caso de estudio, o que la frecuencia de aplicación sea de varios años, ocasionando pérdida de información entre estos[10].

Con la combinación de los gestores de base de datos, el sistema geográfico y la herramienta estadística el usuario conseguirá resultados óptimos que ayudará en su búsqueda de información y visualización de lo requerido.

4 CONCLUSIÓN

Del desarrollo del análisis de trayectorias vehiculares de taxis en la ciudad de Beijín se puede concluir lo siguiente:

- ✓ Como primera instancia de la investigación el uso de las herramientas existentes necesarias para realizar el análisis de trayectorias, se llega a la conclusión de que se necesitan recursos tanto hardware como software para la visualización de los puntos en el sistema geográfico QGIS ya que el uso de esta herramienta es fundamental en el análisis.
- ✓ Por medio de este artículo se llega a la conclusión que podría beneficiar a las cooperativas que presta su servicio de la trayectoria o localización de una unidad de taxi, es decir, con los planteamiento que se dieron al sistematizar un proyecto en buscar los beneficio del cliente y del sueño de la unidad, que vean una manera de localización de la persona y del vehículo de una manera exacta sin necesidad de mínimos errores, con los que había en las anteriores sistema que solo verificaban el objeto o vehículo de en otras forma, ya sea por su velocidad que andaba o si el vehículo estaba en inercia pero aun así no daban con la latitud y la longitud exacta porque se manejaba de puntos de lugares menos de los relieves que se indicaban.
- ✓ Con la visualización de los diagramas estadísticos podemos concluir que los taxis en el día 1 no trabajan de 00:00H a 06:00H y el resto de días laboran con normalidad. Las horas en todos los seis días que tienen un tope máximo esta entre 18:00H a 20:00H y las horas que menos se laboran van desde las 05:00H hasta las 10:00H.

REFERENCIAS

- [1] Huerta Eduardo; Mangiaterra Aldo; Noguera Gustavo., GPS Posicionamiento Satelital. 2005.
- [2] H. Augusto, S. García, I. Civil, and A. Dinámico, "VEHICULAR," 2006.
- [3] C. Lera, Al. Genolet, V. Rocha, Z. Schoenfeld, and S. Bolcatto, "Trayectorias : un concepto que posibilita pensar y trazar otros caminos en las intervenciones profesionales del Trabajo Social," Cátedra Paralela, vol. 4, pp. 33–39, 2007.
- [4] C. Paper and I. Sciences, "Análisis de la Semántica del Movimiento en el Control de Flotas usando Sistemas de Información Geográfica," no. January 2011, pp. 0–19, 2016.
- [5] B. Teresa, R. Gabriel, and M. Vanín, "Análisis estadístico integral de información obtenida en un lote bajo agricultura de precisión," XVI Reun. Científica del GAB, pp. 1–8, 2011.
- [6] J. D. Godino and F. Ruiz, Geometría Y Su Didáctica Para Maestros. 2002.
- [7] L. Name, F. Name, O. Training, P. Training, C. Darin, R. O. Training, M. Kimberly, G. Deepa, E. Board, E. Principal, I. Primary, F. Systems, E. B. Study, and N. Co-investigator, "QGIS User Guide," Igarss 2014, no. 1, pp. 1–5, 2014.
- [8] A. Castillo, "Métodos Estadísticos con R y R Commander," Cran.Cs.Wwu.Edu, p. 138, 2012.
- [9] J. Contreras, E. Molina, and P. Arteaga, Introducción a la programación estadística con R para profesores. 2010.
- [10] M. Nuñez and R. Isela, "COMPUTACIÓN."