

LabUrbano Criminalidad

Ariza B. Francisco, Manchola M. Jofre, Ruiz S. Jairo, Clavijo Andrés, Guerra G. John

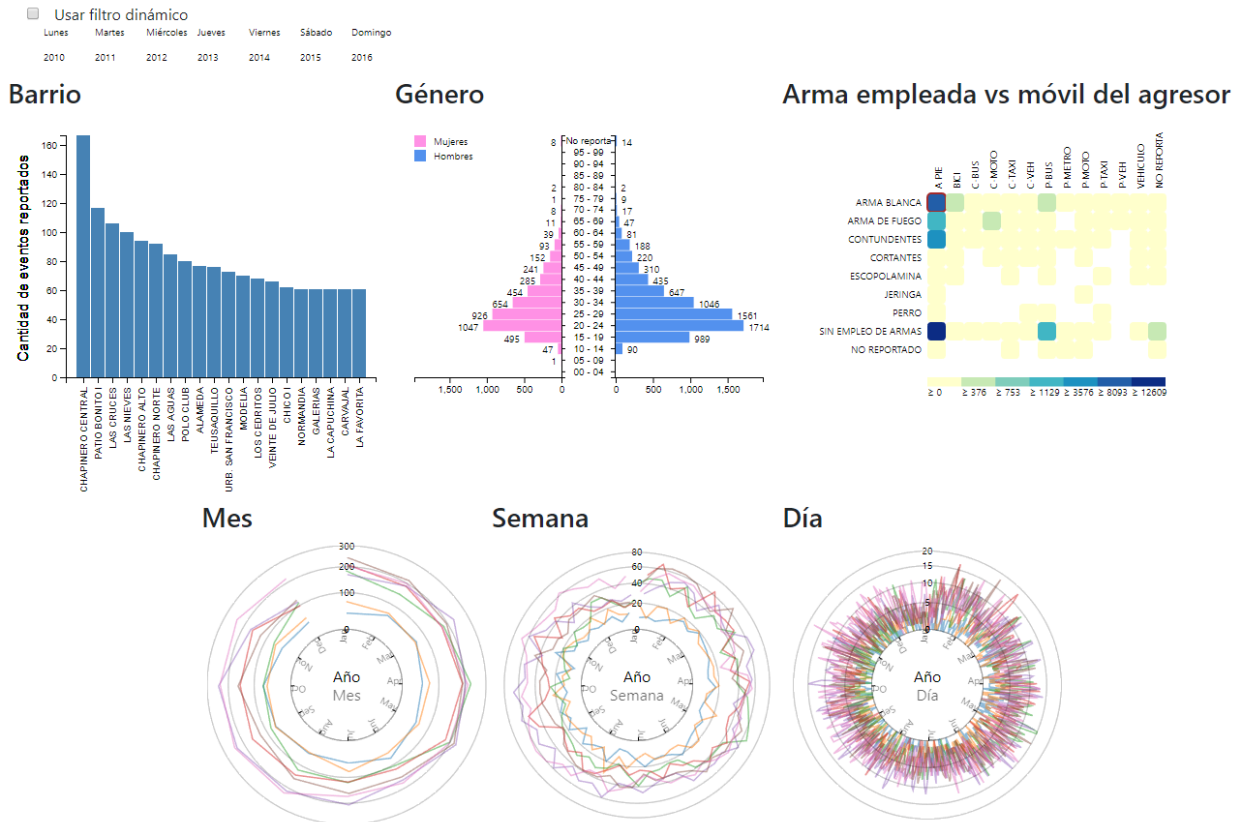


Fig. 1. Dashboard con variables interesantes para descubrir insights a partir de datos sobre el hurto de celulares.

Abstract—El hurto de celulares es un delito de alto impacto en la ciudad de Bogotá, el aumento en la cantidad de hurtos reportados, preocupa a las autoridades y a la población en general, sin embargo, la mayoría de la población desconoce las zonas en las que más se presenta este delito, pasando por alto variables de tiempo, modo y frecuencia de los eventos por el escaso conocimiento de la información y la poca divulgación por parte de las autoridades. Con datos oficiales de los hurtos a celular reportados entre los años 2010 y 2016, se propone implementar una visualización que permita presentar y descubrir *insights* para el uso y dominio público.

Index Terms—SIEDCO, hurto a celular, visualización, viz

1 INTRODUCCIÓN

De acuerdo a la última encuesta de Percepción y Victimización en Bogotá realizada en el mes septiembre de 2017 [2], el 77% de los encuestados manifestó haber sido víctima del delito de *hurto a persona* y al 40% de estos le *robaron el celular*.

- Ariza B. Francisco estudiante de la Universidad de los Andes. E-mail: fa.arizab@uniandes.edu.co.
- Ruiz S. Jairo estudiante de la Universidad de los Andes. E-mail: ja.ruiz907@uniandes.edu.co.
- Manchola M. Jofre estudiante de la Universidad de los Andes. E-mail: ja.mancholam@uniandes.edu.co.
- Clavijo Andr Coordinator and Lead Researcher for Colombia — Data-Pop Alliance. E-mail: aclavijo@datapopalliance.org.
- Guerra G. John profesor de la Universidad de los Andes- E-mail: ja.guerrag@uniandes.edu.co.

Teniendo como punto de partida estas alarmantes cifras, que preocupan a los ciudadanos del común, autoridades, víctimas y expertos en el tema, surgieron los siguientes interrogantes: ¿Cuál es el barrio de Bogotá con mayor índice de hurto de celulares? ¿Cuáles podrían ser las posibles causas determinantes? ¿En qué zonas de la ciudad hurtan celulares con mayor frecuencia? ¿Cuáles son las causas probables del hurto de celulares? ¿El hurto de celulares es mayor en dispositivos de gama alta, o hurtan sin distinción? Son preguntas específicas que requieren respuestas soportadas en datos provistos por una entidad oficial como la Dirección General de la Policía Nacional.

Partiendo como base de estos interrogantes y con la participación de *Data-Pop Alliance*¹ como usuario experto y cliente principal, se revisó

¹Data-Pop Alliance es una coalición mundial de Big Data y desarrollo creada por la Iniciativa Humanitaria de Harvard, el MIT Media Lab y el Overseas Development Institute, que reúne a investigadores, expertos, practicantes y activistas para promover una revolución de Big Data centrada en la gente, a través de investigación colaborativa, el desarrollo de capacidades y el compromiso de

el estado del arte de la problemática propuesta y se definió como objetivo dar respuesta a estas y otras preguntas que se irán descubriendo con el análisis de los datos y las exploraciones con los usuarios expertos. Para lograr dicho objetivo, se propone en este paper como metodología de trabajo, realizar la limpieza y entendimiento de los datos, definir los usuarios y tareas, definir y desarrollar una visualización interactiva con opciones de filtrado que permita, desde una pregunta general, descubrir insights específicos, combinando dos o ms dimensiones de los datos.

2 ESTADO DEL ARTE

El hurto de celulares es uno de los fenómenos criminales más recurrentes en la ciudad de Bogotá y el cual se ha expandido por todo el territorio nacional, el alarmante crecimiento en las cifras de este delito ha servido para considerarlo como objeto de estudio social y económico, con el fin de lograr identificar posibles repercusiones como resultado de esta actividad ilícita.

A continuación, se analizan algunas visualizaciones de artículos e investigaciones:

- 1. Extinción de dominio como herramienta contra el hurto de celulares en la ciudad de Bogotá. [4]

Tarea 1: Presentar el impacto de la implementación de la extinción del dominio sobre las cifras históricas del hurto de celulares.

Visualizaciones utilizadas: Gráfico de barras (vertical y horizontal), líneas.

Análisis: las visualizaciones soportan la tarea propuesta, sin embargo se pueden incluir otro tipo de gráficos para realizar más tareas de comparación (que es el objetivo principal). Ver figuras 2 y 3.



Fig. 2. Celulares hurtados por localidades de Bogotá, 2011-2015.

la comunidad. A partir de febrero de 2016, la Fundación Flowminder se unió a Data-Pop Alliance como su cuarto miembro principal”.

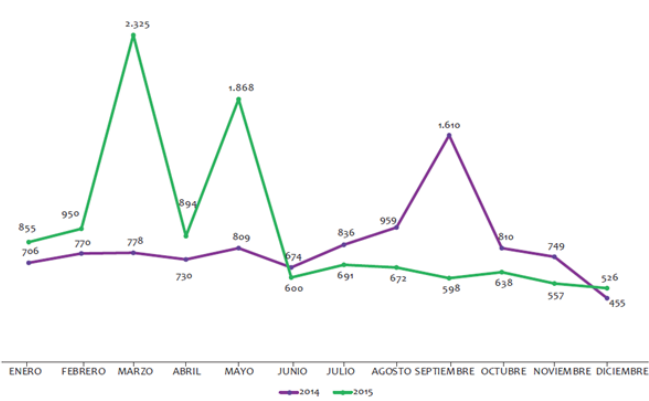


Fig. 3. Comportamiento mensual del hurto de celulares en Bogotá, 2014-2015.

- 2. Impacto socio-económico, financiero y legal del comercio de celulares lícito e ilícito en Florencia Caquetá. [1]

Tarea 1: Presentar los resultados de la encuesta realizada a la población de Florencia.

Visualizaciones utilizadas: Gráfico de barras (vertical y horizontal) y gráfico de torta.

Análisis: Los gráficos de barras no están ordenados y en algunos casos son redundantes los valores que se presentan ej: cuenta de género y cuenta de edad. Ver figuras 4 y 5.

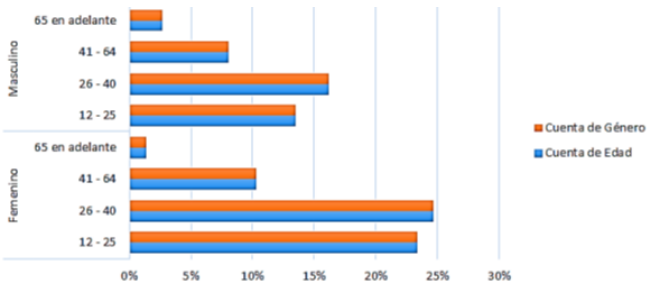


Fig. 4. Edad de las víctimas de hurto.

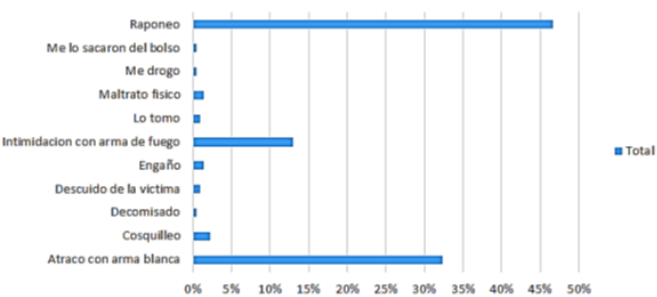


Fig. 5. Actividad de la víctima durante el hurto.

- 3. Mapa de calor, robo de celulares en Bogotá. [5]

Tarea 1: Identificar las localidades y las zonas críticas con mayor cantidad de hurtos.

Tarea 2: Presentar las cifras de hurto de celulares por localidad Visualizaciones utilizadas: Mapa interactivo.

Análisis: es una visualización orientada al usuario final en el cual se espera adquiera un seguro para su celular, por lo tanto los datos

están segmentados por localidad. Para realizar comparaciones y una mejor visualización de datos es mejor utilizar bar charts (Ver figura 6).

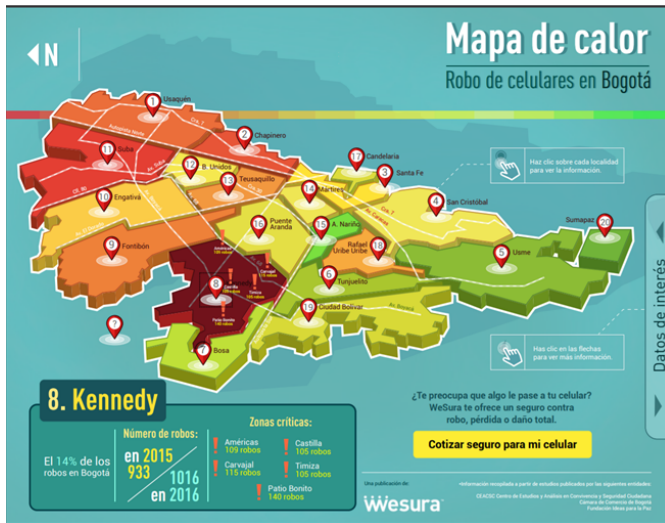


Fig. 6. Mapa de calor, robo de celulares en Bogotá.

3 CARACTERIZACIÓN

Se realizó el proceso de caracterización en base al *framework* de Tamara Munzner [3] que está compuesto por el What, Why y el How. A continuación se describe el What (abstracción de datos) y el Why (abstracción de tareas).

3.1 What

Dataset: Delito de hurto a celular en la zona urbana de Bogotá. (ver tabla 1).

Fuente de datos: Data-pop alliance ha proporcionado el dataset que ha consolidado sobre hurto a celular en Bogotá, cuya fuente, como productor de información, es la policía. Por medio de la *Fundación Ideas para la Paz* (en adelante FIP), fue posible obtener un dataset, que además de los atributos disponibles en el portal de datos abiertos de Colombia (<https://datos.gov.co/>), cuenta con la información geográfica (latitud y longitud) del evento hurto a celular. Esto permite enriquecer el dataset con otra información geográfica como barrio común, cuadrantes de policía o estaciones de policía. Vale la pena señalar que la información geográfica es de carácter confidencial por lo que el dataset procesado, análisis y visualizaciones no mostrarán específicamente esta información.

3.1.1 Limpieza de datos

Teniendo en cuenta que se desea realizar un análisis de los datos usando la fecha del hurto como principal variable, se eliminaron 5.003 registros, dado que no tienen información en el atributo fecha. Dado que se utilizarán únicamente los datos de la ciudad de Bogotá, las variables Código DANE, Departamento y Municipio dejan de ser relevantes, ya que se vuelven constantes, por tanto no se tendrán en cuenta. Como primer análisis de los datos, se realizó un perfilamiento de los mismos (ver anexo, archivo reporte en formato HTML), donde se evidenciaron errores por completitud de los datos, ya que utilizaron caracteres especiales -, o frases como NO REGISTRA, NO REGISTRARON, entre otros. Por tanto se buscó homogeneizar los datos y se usó NO REPORTADO como frase para indicar la falta de información. De igual manera se encontraron errores en los atributos hora, línea y edad, dado que utilizaron datos con diferentes tipos o formatos. Dado que el dataset original cuenta con un atributo hora y otro atributo fecha, se derivaron los datos generando una nueva variable denominada TIMES-TAMP, la cual registra la fecha y hora del hurto en un solo atributo, lo cual facilitará la generación de las visualizaciones.

Table 1. Relación de Atributos del dataset y Tipos

Atributo	Tipo
Arma empleada	Categorico
Barrio	Categorico
Cantidad	Ordenado, Cuantitativo
Clase sitio	Categorico
Código DANE	Categorico
Departamento	Categorico
Día	Ordenado, cíclico
Empleo	Categorico
Escolaridad	Categorico
Fecha	Ordenado, Ordinal, Secuencial
Género	Categorico
Hora	Ordenado, Ordinal, cíclico
Línea	Categorico
Marca	Categorico
Móvil Agresor	Categorico
Móvil Víctima	Categorico
Municipio	Categorico
País Persona	Categorico
Timestamp	Ordenado, Ordinal, Secuencial
Zona	Categorico

3.1.2 Conocimiento experto

Por medio de *data-pop-alliance* nos conectamos con la *Fundación Ideas para la Paz* (FIP) quienes cuentan con un dataset obtenido del sistema de información de la policía SIEDCO (Sistema de Información Estadística, Delincuencial, Contravencional y Operativa de la Policía Nacional), este dataset, además de los atributos presentados anteriormente, cuenta con atributos de posición, latitud y longitud, permitiendo mejorar la calidad del dato en cuanto a barrio u otra jurisdicción de interés, como localidades, unidades de planeamiento o cuadrantes de policía, realizando una intersección geográfica entre la capa de delitos (con geometría punto) y la capa de interés (con geometría polígono).

De acuerdo a la FIP, los datos registrados en SIEDCO tuvieron un cambio importante en cuanto a la forma de registrar la información en febrero de 2014. Antes de esta fecha, el policía registraba en SIEDCO la dirección en un campo de texto y posteriormente se obtenía su latitud y longitud por medio de un geocodificador, lo que generaba una baja confiabilidad de la posición obtenida y muchos registros no lograban ser geocodificados; después de esta fecha, la policía contó con un mapa, en el cual es posible ubicar el punto, eliminando el uso del geocodificador y obteniendo mayor confiabilidad en el atributo de geometría del dataset. Por esta razón, la FIP recomienda utilizar la serie a partir de 2014 si se realiza análisis geográfico y tomar la serie completa (a partir de 2010) si se realiza análisis alfanumérico.

3.2 Why

Para la definición del why, se ha caracterizado junto con *data-pop-alliance* el usuario objetivo del proyecto, el cual se ha definido con las siguientes características:

- Ciudadano analítico.
- Interesado en los problemas sociales.
- Ejemplo: Usuario lector de la silla vacía.
- Un usuario capaz de entender una visualización medianamente compleja sin llegar a ser un experto.

El objetivo de la visualización es permitir al usuario explorar los datos y **descubrir** características de su interés.

3.2.1 En términos de Tamara

Se define como tarea principal *Analyze - Consume - Discover* y varias tareas secundarias asociadas así:

- **Descubrir** características interesantes sobre el hurto de celulares de acuerdo al lugar (barrio, clase de sitio), hora y día de la semana.
 - Realizar un análisis de series de tiempo del dataset y **Presentar** la distribución de hurto de celulares en función de variables temporales, ya sea hora, día de la semana, mes, o incluso a nivel anual.
 - Realizar un análisis geoespacial e **Identificar** si existe correlación entre las variables temporales y espaciales, que permitan identificar lugar y tiempo de mayor ocurrencia de los eventos de hurto.
 - **Comparar** las tendencias de hurto de celulares entre el año 2010 y 2017.

4 MÉTODO PROPUESTO

En el proceso de definición de la visualización, teniendo en cuenta las tareas y los datos, se realizó la propuesta del *HOW* (ver figura 7) la cual consta de 5 modismos descritos a continuación:

Modismo 1

- Modismo: mapa.
- Marca: Polígono (shape).
- Canal: Posición (latitud, longitud) y color (saturación).
- Atributo: Cantidad de hurtos por barrio.
- Descripción: Se representa la cantidad de hurtos a celular en el filtro seleccionado con la saturación del relleno del polígono.

Modismo 2

- Modismo: Bar Chart.
- Marca: Línea.
- Canal: Posición horizontal, color (tono).
- Atributo: Cantidad de hurtos a hombres y mujeres por rangos de edad.
- Descripción: Representa la distribución de hurto a celulares por género, discriminado por rango de edades.

Modismo 3

- Modismo: Matriz de adyacencia.
- Marca: Rectángulos (shape) y color (saturación).
- Canal: Posición horizontal, posición vertical, Color (saturación).
- Atributo: Móvil del agresor y arma empleada.
- Descripción: Representa la relación entre el móvil del agresor y el arma empleada, con la saturación se representa la cantidad de hurtos.

Modismo 4

- Modismo: Coordenadas paralelas. (aún se encuentra en discusión)
- Marca: Líneas.
- Canal: Posición vertical.
- Atributo: Hora, día de la semana, mes y año.
- Descripción: Representa la relación entre las variables del tiempo en las que se presentaron los hurtos.

Modismo 5

- Modismo: Multi-line chart.
- Marca: Líneas
- Canal: Posición horizontal, posición Vertical, Color (tono)
- Atributo: Fecha, cantidad de hurtos, tendencias
- Descripción: Representa el comportamiento (feature) de la cantidad de hurtos de celulares respecto a la serie de tiempo (fecha)

5 RESULTADOS Y EVALUACIÓN

Una vez definido el método se procedió a crear la que sería la primera versión del mockup (figura 7), posteriormente se sometió a pruebas de usuario sobre el cual se recibió retroalimentación de estudiantes de *Visual Analytics* y del cliente del proyecto, los principales resultados de la evaluación se listan a continuación para cada modismo.

Modismo 1

- Los usuarios tienen dificultad para buscar barrios específicos, se sugiere incluir textbox para búsqueda de ubicaciones por nombre.
- Usar la saturación del color para expresar información (posiblemente número de eventos de hurto).
- Puede llegar a ser complicado encontrar la información de barrios pequeños, incluir función de zoom en el mapa.

Modismo 2

- Algunos usuarios entienden la divergencia de los datos y la clasificación de información por rangos de edad, sin embargo, los diferentes colores utilizados confunden a los usuarios. Se debe usar un color por género.
- Incluir alguna interacción/objeto para presentar la suma de eventos por género (total hombres / total mujeres) y total general.

Modismo 3

- Consideran interesante presentar las variables de móvil de ladrón VS arma empleada, sin embargo se debe usar un *heatmap* y no una matriz de adyacencia.
- Usar solo un color para presentar la ocurrencia de eventos.
- Se podría considerar agregar *barcharts* para presentar las variables separadas.

Modismo 4

- Se presenta confusión por parte de los usuarios, dada la complejidad del gráfico. En caso de implementarse, las variables categóricas deben presentar los valores/categorías dentro de una misma columna o en varias columnas para poder expresar los valores numéricos de las mismas.
- Se sugiere implementar gráficos radiales para expresar la continuidad (cíclica) de las variables temporales. <https://bl.ocks.org/mbostock/6fead6d1378d6df5ae77bb6a719afcb2>

Modismo 5

- Los usuarios relacionan inmediatamente la gráfica con datos temporales, sin embargo, al momento de interactuar no es claro el propósito de la gráfica.
- La interacción de filtrar los datos no es intuitiva.
- Los checkboxes no son lo suficientemente visibles.

Dashboard Análisis Hurto Celulares - Ciudad de Bogotá

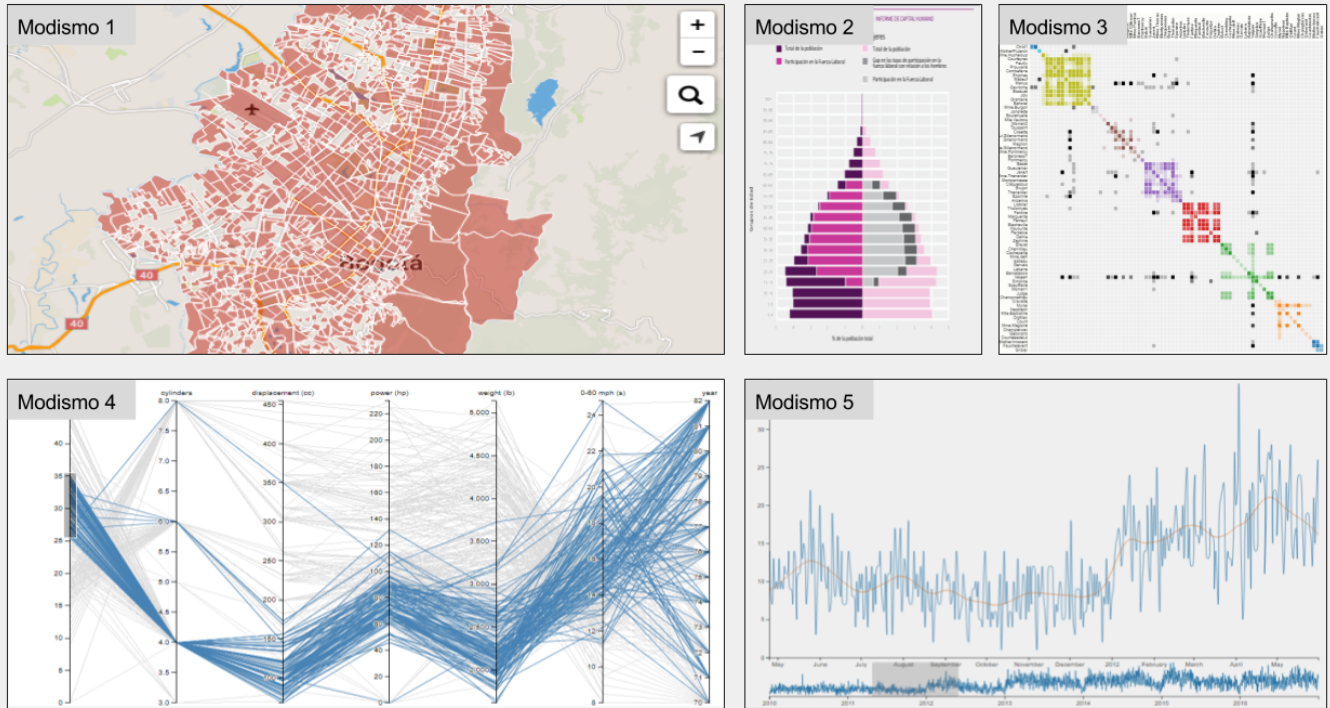


Fig. 7. Mocup versión 1 de la visualización propuesta.

Dashboard Análisis Hurto Celulares - Ciudad de Bogotá

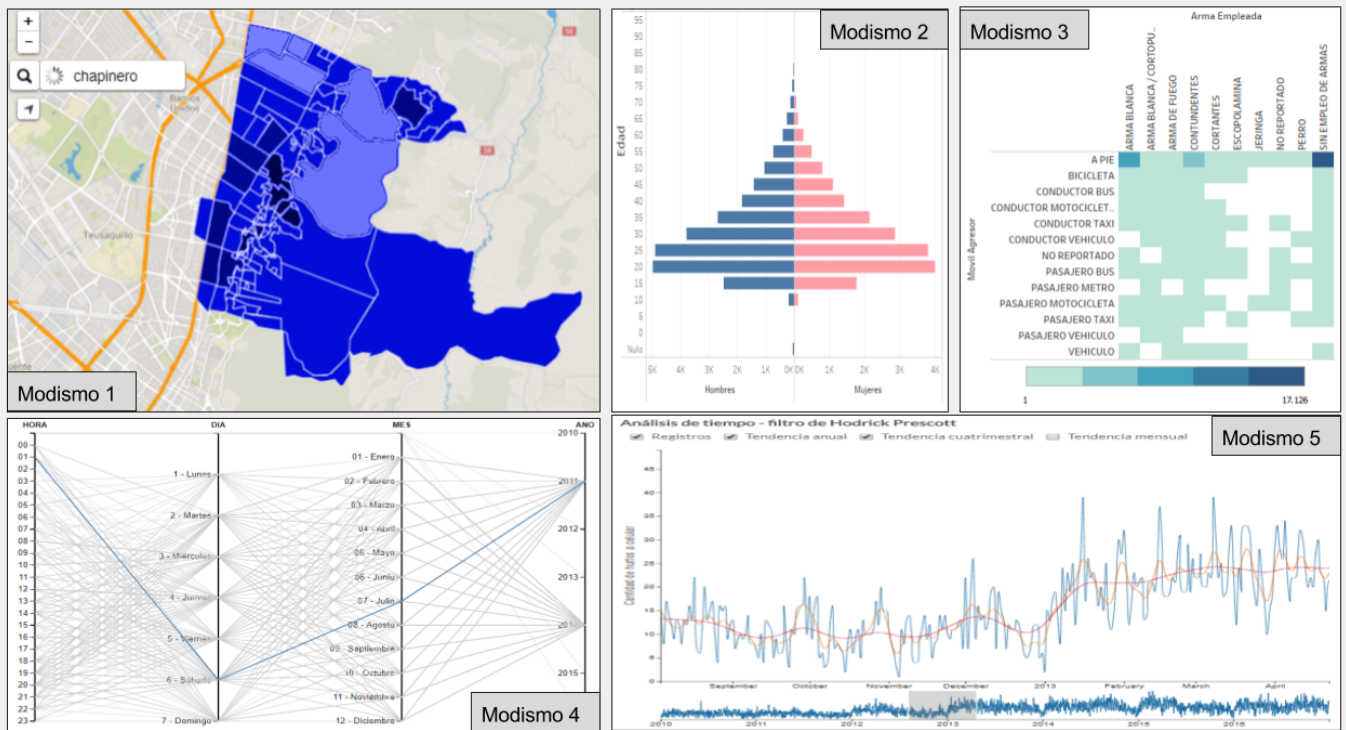


Fig. 8. Mocup versión 2 de la visualización propuesta.

- La gráfica funciona correctamente para presentar las diferentes tendencias, sin embargo, no facilita tareas como encontrar valores máximos y su respectivo periodo, o tareas de comparación.

- Validar si se cambia por otra gráfica.

Comentarios generales

- Indispensable incluir los títulos de los ejes en las gráficas y sus respectivas leyendas de las nomenclaturas usadas.

Comentarios del cliente

- Es importante que se pueda responder fácilmente a la pregunta: ¿Cuál es la probabilidad de robo de celular en un barrio dado a una hora XX y día de la semana YY?
- Para lo anterior sugiero que el dashboard tenga un recuadro donde este un reloj o calendario.
- Con el fin de articular mejor el dashboard es importante tener una columna con controles para seleccionar las principales variables de cruces para el análisis. Si se deja solo los cuadros actuales el usuario se puede perder un poco.
- Me gusta el mapa, la pirámide, el mapa de calor y el análisis de serie de tiempo. La gráfica de barras paralelas no tanto, es una buena gráfica para ver correlaciones, sin embargo, creo usarla para solo variables de tiempo no es recomendable.

Finalmente, después de evaluar los comentarios y realizar los ajustes a la visualización, la propuesta fue desarrollar un tablero de control utilizando *cross-filter*, las dimensiones se definieron en conjunto con data-pop-alliance durante una reunión de revisión del análisis descriptivo del dataset.

Como acercamiento y pensando en los atributos discutidos, se elabora una segunda propuesta (ver Fig. 8).

6 DISCUSIÓN Y CONCLUSIÓN

La segunda versión del mockup fue sometida a pruebas de usabilidad, como resultado de las pruebas se concluyó que los modismos uno, cuatro y cinco presentaba los siguientes inconvenientes:

- **Modismo 1:** Finalmente no se cuenta con los datos de las coordenadas x e y de cada evento, como tampoco la descripción del barrio homologada con la capa de barrios de Bogotá.
- **Modismo 4:** La visualización requiere de un conocimiento específico por parte del usuario para su correcta interpretación. Tiene problemas de oclusión por la cantidad de conexiones que hay entre puntos paralelos.
- **Modismo 5:** No es claro para el usuario su interpretación respecto a los datos filtrados en los demás modismos, principalmente porque se muestran líneas de tendencia en un rango de tiempo que imposibilitan combinar variables como año-mes, día-hora.

Teniendo en cuenta lo anterior se modifican los tres modismos mencionados anteriormente, por los siguientes:

- **Modismo 1,** permite visualizar la cantidad de hurtos reportados por barrio, se filtra por uno o más barrios. (ver Fig. 9)
- **Modismo 4 y 5,** permiten al usuario visualizar y seleccionar de forma independiente y fácil las diferentes magnitudes de tiempo. (ver Fig 10 y Fig 11)

6.1 Conclusiones

El *dashboard* construido con sus 6 modismos permitió descubrir varios insights interesantes entre los que se resaltan los que a consideración de los autores son significativos:

- En el año 2010 hurtaban más celulares a hombres que a mujeres, sin embargo se identifica que año tras año las cifras de hurto a mujeres han aumentado y para el año 2016 la cantidad de hurtos son similares entre ambos géneros.
- Los hombres son más propensos a ser atacados con arma blanca en el evento de hurto a celular.
- La mayoría de personas a las que les hurtan su celular utilizando un arma de fuego son hombres.
- Los barrios de la zona de chapinero y zona centro son los que presentan mayor cantidad y frecuencia en el hurto a celular.

Barrio

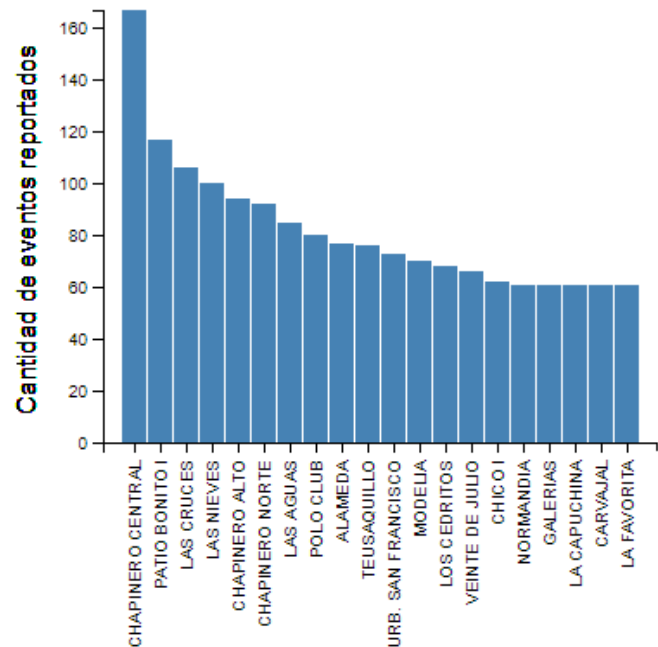


Fig. 9. Modismo 1 versin final.



Fig. 10. Modismo 4 y 5 versin final.

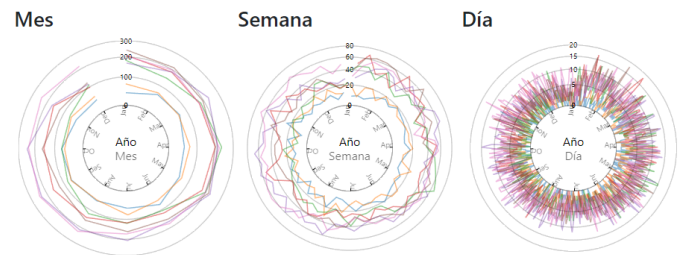


Fig. 11. Modismo 4 y 5 versin final.

REFERENCES

- [1] F. de Ciencias Económicas y Empresariales (Face). Universidad de Pamplona. Impacto socio-económico, financiero y legal del comercio de celulares lícito e ilícito en Florencia Caquetá. 16-1, 2016.
- [2] C. de Comercio de Bogotá. Encuesta de percepción y victimización en bogotá. *The Visual Computer*, 1, Sept. 2017.
- [3] T. Munzner. *Visualization Analysis and Design*. Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, 1st ed., 2014. doi: ISBN 978-1-4665-0891-0
- [4] L. Roa. Extinción de dominio como herramienta contra el hurto de celulares en la ciudad de Bogotá. *Revista Criminalidad*, 58-2, 2016.
- [5] Wesura. Mapa de calor, robo de celulares en Bogotá. Wesura, Octubre 2017.