



Universidad Católica
San Pablo

Análisis Visual de Datos Origen Destino (OD)

Jesús Erick Vera Callme

Orientador:

Tesis profesional presentada al Programa Profesional de Ciencia de la Computación como parte de los requisitos para obtener el Título Profesional de Lic. en Ciencia de la Computación.

UCSP- Universidad Católica San Pablo
Mayo de 2020

*A Dios, por todo lo que me ha dado y a
toda mi familia y amigos por acompa-
ñarme en este arduo proceso.*

UNIVERSIDAD CATÓLICA SAN PABLO
FACULTAD DE INGENIERÍA Y COMPUTACIÓN
PROGRAMA PROFESIONAL DE CIENCIA DE LA
COMPUTACIÓN

Análisis Visual de Datos Origen Destino (OD)

Tesis de graduación presentado por el bachiller Jesús Erick Vera Callme en el cumplimiento de los requisitos para obtener el título profesional de Licenciado en Ciencia de la Computación.

Arequipa, 15 de mayo de 2020

Aprobado por:

Abreviaturas

OD Origen Destino

ODM *Origin Destination Matrix*

NYC *New York City*

Agradecimientos

En primer lugar deseo agradecer a Dios por haberme guiado a lo largo de estos cinco años de estudio.

Agradezco a mis padres por el apoyo brindado para forjarme como un profesional.

Agradezco a la universidad, mi *alma matter*, por haberme cobijado y brindado la formación que ahora me permitirá ayudar a construir una mejor sociedad.

Agradezco de forma muy especial a mi orientador Dr. Erick Gomez Nieto por haberme guiado en esta tesis.

Resumen

En esta tesis abordamos el problema de la exploración visual de datos de origen-destino (OD), datos que pueden representar el movimiento de personas, energía, material o intercambio de información. Muchos datasets OD representan procesos globales como por ejemplo, flujos de refugiados y comercio entre países. Sin embargo, hasta ahora no se han desarrollado soluciones universalmente buenas para el análisis de los datos OD, especialmente cuando se trata del análisis de los cambios temporales y espaciales en ellos. La idea de esta tesis es que el uso de la visualización exploratoria interactiva es una clave para abordar los desafíos creados por la complejidad inherente de los datos OD. El verdadero desafío es diseñar visualizaciones interactivas que permitan la exploración de datos tan complejos y que ayuden a encontrar en estos datos información útil.

El trabajo descrito en esta tesis es un intento de obtener una mejor comprensión del proceso de visualización interactiva de datos OD como un todo y, con esta comprensión, facilitar el desarrollo de técnicas y herramientas que permitan la exploración visual de OD.

Estudiamos qué se puede aprender a partir de estos datos, cuáles son sus posibles representaciones y enfoques para su análisis visual, para qué tareas son más adecuadas las diferentes representaciones y cómo se usan realmente.

Este trabajo puede conducir a la aparición de soluciones más integrales y universales que ayuden a los analistas a dar sentido a sus datos y apoyar a los responsables de la toma de decisiones en la búsqueda e implementación de las políticas correctas.

Abstract

In this thesis we address the problem of the visual exploration of source-destination data (OD), data that can represent the movement of people, energy, material or exchange of information. Many ODsets represent global processes such as refugee flows and trade between countries. However, until now, universally good solutions for the analysis of OD data have not been developed, especially when dealing with the analysis of temporal and spatial changes in them. The idea of this thesis is that the use of interactive exploratory visualization is a key to address the challenges created by the inherent complexity of the OD data. The real challenge is to design interactive visualizations that allow the exploration of such complex data and help to find useful information in these data.

The work described in this thesis is an attempt to obtain a better understanding of the process of interactive visualization of data. OD as a whole and, with this understanding, facilitate the development of techniques and tools that allow the visual exploration of OD. We study what can be learned from these data, what are their possible representations and approaches for their visual analysis , for which tasks the different representations are better suited and how they are actually used.

This work can lead to the emergence of more integral and universal solutions that help analysts to make sense of their data and support those responsible for making decisions. decisions in the search and implementation of the correct policies.

Índice general

1. Introducción	2
1.1. Motivación y Contexto	2
1.2. Planteamiento del Problema	4
1.3. Objetivos	4
1.3.1. Objetivos Específicos	4
1.4. Organización de la tesis	5
2. Datos Origen Destino	6
2.1. Análisis de Datos	6
2.2. Visualización de datos espacio-temporales.	7
2.2.1. Agregación	7
2.2.2. Interactividad	8
2.2.3. Técnicas de Interacción	8
2.2.4. Estrategias de Exploración	9
2.2.5. Estrategias de Diseño	9
2.3. Datos OD temporales	10
3. Visualización de Datos Origen Destino	13
3.1. Visualizaciones con Grafos	13
3.2. Matrices OD y Mapas OD	14
3.3. Mapa de Fluxos	15
3.4. Mapa de coropleta	16

3.5.	Mapa de símbolo graduado	17
3.6.	Series de Tiempo	17
3.7.	Diagrama de Cuerdas	18
3.8.	TreeMap	19
3.9.	Consideraciones Finales	19
4.	Propuesta e Implementación	21
4.1.	Arquitectura de la Propuesta	21
4.1.1.	Entrada de datos OD	22
4.1.2.	Preprocesamiento de datos de entrada	22
4.1.3.	Visualización del Dataset	22
4.2.	Implementación	22
4.2.1.	Tratamiento del Dataset	23
4.2.2.	Visualización de Datos OD	23
4.3.	Definición de Pruebas	28
4.3.1.	Diseño de Experimentos	28
4.3.2.	Evaluación de Tareas	29
5.	Pruebas y Resultados	30
5.1.	Diseño del Estudio	30
5.2.	Las Condiciones	32
5.3.	Datos Colectados	33
5.4.	Análisis	33
5.5.	Codificación	33
5.6.	Pruebas	35
5.6.1.	Análisis Cualitativo	43
6.	Conclusiones y Trabajos Futuros	44
6.1.	Problemas encontrados	44

6.2. Recomendaciones	45
6.3. Trabajos futuros	45
Bibliografía	47

Índice de tablas

Índice de figuras

2.1.	Representación de matriz OD de datos OD.	10
2.2.	Datos OD temporales como una lista de eventos de flujo.	11
2.3.	Datos temporales de OD como flujos agregados por año.	11
2.4.	Datos sobre los orígenes y destinos necesarios para producir una visualización geográfica.	11
3.1.	Una secuencia de animación para agrupamiento de bordes.	14
3.2.	a) Matriz OD que muestra los flujos de condado a condado en Irlanda; filas = condados de origen, columnas = condados de destino. b) Los vectores de migración de condado a condado de EE. UU. Se muestran como un mapa OD. Cada celda de cuadrícula grande representa la ubicación de origen, dentro de la cual se muestra el mapa de densidades de destino que utilizan la misma cuadrícula	15
3.3.	Flujos de refugiados entre los países del mundo en 1996,2000	16
3.4.	Mapa de coropleta que muestra la migración fuera del condado de Tipperary en Irlanda (blanco)	17
3.5.	Mapa de símbolo graduado que muestra las entradas de migración a Manchester	18
3.6.	Gráfico de series de tiempo que muestra 52 precios de acciones semanales para 1430 acciones (eje x = semanas 52, eje y = precios de acciones). Arriba: consulta de un solo timebox para artículos que se intercambiaron entre 70 y 190 durante las semanas 1-5.	19
3.7.	Diagrama de migración circular interactivo de flujos de migración entre regiones en 2005-10	20
3.8.	Treemap que muestra el número de usuarios únicos de los 100 sitios web principales en enero de 2010.	20
4.1.	Arquitectura de la propuesta	21

- 4.2. Visualización *Flowstrates*. Visualización del *dataset* de Taxis de NYC de los 13 (CD) distritos comunitarios del burgo de *Manhattan* del año 2016. Las líneas de color más fuerte, comunican que se hizo una selección en el Origen 106 *Manhattan* y nos indican la posición de ese origen en el mapa de calor. En el mapa de calor también es posible ver que a mientras más ‘color oscuro’ presente un mes mayor flujo de movimientos tiene. 24
- 4.3. Visualización Circular con Cuerdas. Visualización del *dataset* de Taxis de NYC de los 13 distritos comunitarios de *Manhattan*. Usamos una visualización circular con cuerdas para indicar el número de emigraciones, en combinación con unos gráficos X e Y, para agregarle el componente temporal, y graficar los meses del 2016. Así mismo usamos un gráfico de barras para indicar el número de inmigraciones hacia cada distrito comunitarios. 25
- 4.4. Modificación de la Visualización Circular con Cuerdas inicial, ahora nos permite visualización **todos versus todos**. Visualización del *dataset* de Taxis de NYC de los 13 distritos comunitarios de *Manhattan*. En esta vista observamos que al pasar el mouse por el distrito 103 de Manhattan, nuestra vista superior nos muestra un gráfico temporal, en el cual podemos comparar dos distritos en simultaneo 26
- 4.5. Modificación de la Visualización Circular con Cuerdas inicial, ahora nos permite visualización **todos versus todos**. Visualización del *dataset* de Taxis de NYC de los 13 distritos comunitarios de *Manhattan*. En esta vista observamos que al pasar el mouse por el distrito 103 de Manhattan, nuestra vista superior nos muestra un gráfico temporal, en el cual podemos comparar todos los 13 distritos en simultaneo, así mismo se hizo un **zoom temporal** para ver a detalle 27
- 5.1. Resultados de la encuesta realizada a diferentes personas. Medimos el tiempo en minutos que demoró cada una de las personas realizando la Tarea 1: Consultas Espaciales. En la ultima fila del sector, se observa un campo TOTAL, el cual resume en sus columnas dos métricas (promedio y suma). Usamos el promedio para tener una medida intermedia de los tiempos, y suma para tener el conteo general de la penalidad. Se observa que las personas del sector C tienen un tiempo menor en la mayoría de las visualizaciones (Flowstrates con promedio de 4.70 y penalidad 4, Cuerdas con promedio 4.62 y penalidad de 0 y Cuerdas 2 con promedio 4.20 y penalidad de 4). Le siguen las personas del sector B, luego el sector D y finalizando con el sector A. De acuerdo a las medidas y promedios la Visualización con Cuerdas 2 es mejor que las anteriores. 36

Capítulo 1

Introducción

Los flujos de movimiento se definen como los desplazamientos de grupos humanos, de mercancías y servicios que impliquen movilidad como los taxis y migraciones. Estos flujos tiene la característica principal de tener un origen y un destino. Explicado de otra manera, los actores se mueven desde un punto inicial (origen) hacia un punto final (destino), todo esto ocurre en un periodo de tiempo determinado. En muchas aplicaciones es importante visualizar el flujo de movimiento de algún tipo de mercancía entre diferentes ubicaciones geográficas. Tales flujos incluyen, por ejemplo, movimiento de personas, movimiento de taxis, flujos de servicios, movimiento de bienes y patrones de migración.

1.1. Motivación y Contexto

En el mundo de hoy, muchas cosas que tienen lugar en la realidad se representan como datos. Gracias a la computación ubicua e Internet de las Cosas se generan a diario cantidades grandes de datos de tipo estructurado y no estructurado, estas cantidades aumenta a tasas porcentuales altas. La facilidad de la recopilación automatizada de datos, las instalaciones de almacenamiento baratas y, lo que es más importante, el impacto potencial del conocimiento que se puede extraer de los datos, hacen que el análisis de datos sea indispensable en muchos campos de la actividad humana.

Los datos Origen Destino Origen Destino (OD) son datos que representan movimientos, migraciones y flujos de personas, comercio, transporte, etc. Generalmente estos flujos se dan desde un punto de origen 'A' hacia un punto de destino 'B'. Estos datos pueden de por si representar componentes espaciales y temporales, y su análisis puede llevar al estudio de fenómenos como patrones de migraciones, patrones de movimientos y ocurrencia de eventos en determinados lugares y tiempos

Estos datos OD, tienen múltiples componentes como: un punto de origen, un punto de destino, tiempo, magnitud, descripciones de contexto, etc. Estos componentes en general son de 2 tipos, componentes espaciales (como origen y destino) los cuales pueden representar locaciones geográfica que permite visualizar relacionados a interacciones y *queries* de tipo espacial. Otro tipo de componente son los temporales los cuales representan

cambios de los datos para diferentes momentos en el tiempo.

La necesidad de representar un componentes espaciales o temporales introduce dificultades. Dichas dificultades también implican un desafío a la hora de representar los 2 tipos de componentes al mismo tiempo.

Lo anterior representa un problema de escalabilidad por el hecho de que las interacciones espaciales son conexiones entre locaciones geográficas y estas interacciones son usualmente mucho mas que las que las locaciones. Una visualización que intenta representar todas estas conexiones a la vez sin sacrificar el nivel de detalle probablemente sufra de desorden (conocido como *cluttering*) visual o muestre una cantidad abrumadora de información. En otras palabras, debido a la complejidad inherente de datos OD, para un *dataset* de considerable tamaño es difícilmente posible producir una sola visualización estática la cual represente los datos con todos sus detalles en una manera la cual no presente desorden, sea legible y no sea abrumadora para el usuario. Los siguientes enfoques pueden ser usados para manejar esos problemas:

- Reducir el tamaño del *dataset* usando: filtrado, agregación o resumen de los datos originales
- Enfocarse solo en algunos componentes de los datos y no tomando en cuenta los otros componentes
- Proveer al usuario la posibilidad de explorar los datos interactivamente.

El tercer enfoque es realmente la clave, de hecho este enfoque involucra los dos dos de manera interactiva. Esta combinación puede facilitar el análisis de datos en su total complejidad. Comparado con una visualización estática, el uso de interactividad hace mucho mas fácil proveer soporte para múltiples tareas en una solo herramienta de visualización.

La falta de herramientas fáciles de usar para obtener en la forma correcta y visualizar datos puede ser la razón principal por la que el análisis de datos es tan tarea difícil.

Entonces, el desafío real para permitir el análisis de los datos OD, como lo vemos, está en el diseño de herramientas de visualización interactivas que facilitan la exploración de estos datos complejos y ayudan a los analistas a encontrar en ellos fenómenos o eventos con información útil. Para ello, necesitamos obtener una mejor comprensión del proceso de visualización interactiva en su conjunto.

Un número significativo de los datos OD que tienen lugar en nuestro mundo son interacciones espaciales y temporales, es decir, movimiento de personas, energía, material, dinero e intercambio de información, todo esto sucede entre un origen y un destino. Las interacciones espaciales a menudo se capturan en forma de datos de origen-destino que es una forma de representarlos en una computadora que permite el uso de datos y hace uso de métodos que apoyan el análisis de datos. Algunos de los conjuntos de datos OD que consideramos representan fenómenos globales, por ejemplo, flujos de refugiados y comercio entre países. En muchos casos, analizar dichos datos es de gran importancia, ya que las decisiones basadas en su comprensión pueden tener un gran impacto en la vida de las personas.

Dar sentido a esta gran cantidad de datos y extraer conocimiento de ellos presenta un desafío y requiere un análisis adecuado y un uso de herramientas de exploración. Desde que las computadoras se convirtieron en un producto básico, y el uso de gráficos interactivos revolucionó el campo de la cartografía, se han desarrollado muchas herramientas de software para la exploración visual de datos espacio-temporales. Pero el potencial completo de los datos que representan las interacciones espaciales aún no se ha realizado en gran medida [Andrienko et al., 2010].

Actualmente no existen soluciones universalmente buenas para el análisis de este tipo de datos [Andrienko et al., 2010]. Las herramientas y los enfoques existentes para visualizar datos OD tienen deficiencias y sus capacidades limitadas requieren un estudio especial. Cuando se trata del análisis de los datos OD, la situación es más desafiante. Aún deben desarrollarse técnicas que proporcionen soporte integral para la exploración de interacciones espaciales y el análisis de cambios a lo largo del tiempo.

1.2. Planteamiento del Problema

Los abordajes actuales en política, impacto social, ciencias estadísticas, y los medios de comunicación, generalmente se centran en el análisis de los volúmenes de movimiento para exponer algún fenómeno o anomalía observada, y para luego proceder a tomar una decisión basada en la estadística y observación de dichos datos OD (origen-destino). Los flujos hacia y desde un lugar determinado se entienden si se los ubica dentro de un contexto particular. Visualizar estos datos se ha visto obstaculizado por la falta de técnicas de visualización que comuniquen con claridad patrones o fenómenos de los datos OD. Los datos medidos en un período de tiempo poseen información que no es posible interpretarla sin un adecuado gráfico, que enfoque las técnicas principalmente en visualizar la cantidad de información que proporciona el *dataset*. Sin embargo componentes como el tiempo, la densidad y eventos inesperados pasan desapercibidos en ese tipo de visualización.

1.3. Objetivos

El objetivo principal de esta tesis es el desarrollo de una herramienta de exploración interactiva para datos OD y lograr una mejor comprensión de los datos, lograr también una mejor comprensión de las tareas de análisis, y las formas en que se pueden utilizar las visualizaciones de datos OD, además de precisar al usuario de conocimiento explorativo que le permita un mejor entendimiento de los datos OD.

1.3.1. Objetivos Específicos

- Estudiar las técnicas disponibles para la visualización temporal de datos OD.
 - Desarrollar visualizaciones interactivas para satisfacer las necesidades del análisis de conjuntos de datos de OD del mundo real.
-

- Estudiar el conocimiento generado por la información que se puede obtener con el uso de diferentes visualizaciones de datos de OD.
- Implementar un *framework* de visualización de datos OD, en el cual estén resumidas las técnicas que aborde espacialidad y temporalidad
- Evaluar los resultados obtenidos y precisar la mejor compresión de los resultados.

1.4. Organización de la tesis

Este documento está estructurado en diferentes secciones: En el Capítulo 2, presentamos el abordaje teórico de los datos OD, la descripción de datos OD, algunas estrategias de visualización, el análisis de datos y las definiciones que nos ayudaran a comprender mejor la tesis.

En el capítulo 3 expone los trabajos relacionados de las diferentes técnicas de visualización de datos OD a través de los años, hasta llegar a la actualidad.

En el Capítulo 4, abordamos la Propuesta del trabajo, en donde desarrollamos las técnicas que aportaran a la ampliación de nuestros objetivos.

En el Capítulo 5, nos enfocamos en la definición de Pruebas Y Resultados, así como su implementación y desarrollo.

Como parte del Capítulo 6, nos centraremos en las Conclusiones, a las cuales llegaremos gracias al análisis de las pruebas y resultados obtenidos, en este capítulo también va los Trabajos Relacionados, y lo que haremos a futuro cercano con relación a nuestra investigación.

Capítulo 2

Datos Origen Destino

En este capítulo describimos los datos de origen y destino con mayor rigor que antes con el fin de especificar los modelos de datos a los que nos referiremos en el resto de la tesis. También describimos brevemente varios conjuntos de datos de OD que usaremos como ejemplos en la tesis.

2.1. Análisis de Datos

Visualizar los datos es crucial para que una persona pueda razonar sobre los datos, extraer información relevante y obtener conocimiento [Andrienko et al., 2010]. Se necesitan tres atributos principales para analizar los datos y sus interacciones en el tiempo y el espacio y extraer información valiosa de ellos: la información temática sobre el fenómeno, la ubicación de la observación y su temporalidad. Hoy en día, la adquisición y disponibilidad de datos ya no es el problema principal. Con el desarrollo de nuevas tecnologías, cantidades de datos complejos se han vuelto accesibles [Andrienko et al., 2010]. Por lo tanto es necesaria una investigación de los métodos para explorar estos datos con el fin de extraer información confiable y conocimiento de ellos [Liu, 2017]. Los métodos de análisis visual combinan tecnologías de base de datos, procesamiento computarizado de datos y métodos de análisis explorativo y análisis visual [Chen and Guo, 2015]. Una parte es el análisis computarizado de los datos, y otra es el conocimiento y la capacidad de inferencia del usuario, lo que ayuda a comprender y razonar sobre los conjuntos de datos a menudo muy grandes y complejos [Andrienko et al., 2010]. Las visualizaciones que se utilizan en el análisis visual incluyen características interactivas. Esto ayuda a los usuarios a analizar los datos de forma exploratoria, es decir, pueden probar diferentes escenarios y explorar diferentes posibilidades. Pueden variar parámetros específicos, seleccionar un subconjunto de los datos para mostrar o elegir una resolución espacial o temporal determinada [Dickens, 2009].

El desarrollo de un entorno o un *framework* para la exploración visual de datos OD, es decir, la investigación del movimiento de objetos o personas en el tiempo y el espacio, puede ser útil para la detección de patrones, la extracción de información y la creación de conocimiento [Wood et al., 2010]. Las capacidades humanas visuales y cognitivas jue-

gan un papel crucial en el análisis visual. Se adquiere un mejor conocimiento de manera más eficiente y efectiva en un entorno visual en comparación con el uso de texto o el trabajo en un entorno numérico [Zhou and Zheng, 2011]. Las conexiones espaciales entre objetos aumentan la conciencia de las personas en la búsqueda de objetos similares y, por lo tanto, pueden ayudar a memorizar mejor la información asociada. En el caso de datos espaciales y temporales, los métodos de geo-visualización apuntan a aprovechar esta capacidad visual y cognitiva para permitir al usuario detectar patrones o interpretar señales visuales [Heer et al., 2010]. Por lo discutido anteriormente, realzamos la importancia de crear representaciones que muestren la información relevante espacial y temporal, y al mismo tiempo mostrar grandes cantidades de datos y minimizar el desorden visual de información [Sander et al., 2014]

2.2. Visualización de datos espacio-temporales.

La representación visual de los flujos entre diferentes ubicaciones geográficas es importante para permitir su análisis. Los flujos entre zonas geográficas, incluyen flujos entre ubicaciones, por ejemplo, movimiento de personas o animales, enfermedades, bienes o conocimientos, comportamiento de transporte o tráfico urbano. Como se mencionó anteriormente, los flujos pueden consistir en flujos físicos así como flujos de información. Andrienko et al. representan la estructura de datos de Origen Destino formalmente como: $S \times T \Rightarrow A$, donde S = espacio / conjunto de lugares, T = tiempo / conjunto de momentos o intervalos, A = componente temático (atributo). Por lo tanto, S y T son variables independientes, mientras que los atributos de los datos dependen del espacio y el tiempo. Sobre la base de esta estructura, se pueden hacer preguntas sobre la situación espacial y la variación temporal, por ejemplo. cómo cambian los patrones espaciales de un atributo durante un cierto período de tiempo o cómo cambian los atributos a lo largo del tiempo en un lugar determinado. Andrienko et al. proponen tres clases de datos espacio-temporales en relación con los cambios que ocurren a lo largo del tiempo: 1) cambios existenciales (aparición y desaparición), 2) cambios en las propiedades espaciales (ubicación, forma, orientación, etc.) y 3) cambios de las propiedades temáticas (cambios de atributos cualitativos y aumento o disminución de las características ordinales o numéricas). Especialmente para conjuntos de datos grandes y complejos, generalmente es difícil representar adecuadamente los datos sin perder información relevante porque algunos de los elementos de datos se superponen con otros, por lo que es importante evaluar diferentes métodos para tareas específicas y conjuntos de datos [Zhao and Forer, 2008]. Los enfoques existentes para el análisis visual de datos grandes deben sistematizarse en un *framework* donde idealmente, las técnicas de visualización no solo deben permitir a los usuarios analizar patrones, sino también documentar y compartir esos resultados [Ward et al., 2010].

2.2.1. Agregación

Una fortaleza clave de la visualización de datos radica en la generalización y abstracción de los datos. Una forma de generalizar datos y fomentar la abstracción es la combinación de varios elementos de datos en una unidad, es decir, la agregación de elementos de datos. Esto reduce la cantidad de información mostrada y, por lo tanto, facilita

la visualización y la percepción, pero también conduce a una pérdida de información. Según el interés de un usuario, resaltar las características principales a costa de perder información detallada puede ser una ventaja o una desventaja [Jenny et al., 2016]. Especialmente en el caso de grandes conjuntos de datos, la representación visual de las estructuras temporales de espacio requiere una reducción de los datos, por ejemplo. a través de la selección de subconjuntos o agregación de datos [Rae, 2009]. Rae et al. sugirieron en un estudio sobre datos espacio-temporales, identifican tres tipos de agregación: 1) agregación espacial, 2) agregación temporal y 3) agregación categórica. En el caso de la agregación espacial, los elementos de datos se agregan a unidades espaciales adecuadas, por ejemplo. Unidades administrativas. Para la agregación temporal, el tiempo se divide en intervalos de tiempo y los elementos de datos individuales se agrupan de acuerdo con esos intervalos de tiempo. La agregación categórica agrupa los elementos de datos de acuerdo con sus atributos. Dependiendo de los datos que se estén analizando, dichos atributos categóricos puede, por ejemplo, ser de un tipo determinado. Estas agregaciones permiten al analista investigar los cambios de objetos a lo largo del tiempo en un lugar determinado [Andrienko and Andrienko, 2008].

2.2.2. Interactividad

Los métodos de análisis visual combinan visualizaciones interactivas con técnicas de análisis automatizadas. Esto permite al usuario decidir, por ejemplo. qué parte de los datos desea explorar con más detalle. Esto le permite al usuario definir a un cierto nivel lo que él o ella quiere ver y visualizar. Se especifican tres "niveles de lectura": El nivel elemental (que permite ver la información sobre un solo conjunto de datos), el nivel intermedio (que muestra información resumida sobre un grupo de *datasets*) y el nivel global (que proporciona una visión general de todos los elementos de datos). Los usuarios generalmente asumen que las pantallas dinámicas de computadora son superiores a las estáticas. Las técnicas interactivas mejoran la interpretación de mapas que muestran información incierta. Sin embargo no está claro en qué contextos la interactividad mejora el razonamiento con información incierta, pero se tiende a hacer la visualización de datos interactivo y en tiempo real, porque esto le da al usuario la capacidad de jugar con distintos tipos de configuraciones basadas en sus preferencias [Zeng, 2013].

2.2.3. Técnicas de Interacción

Algunas técnicas de visualización están destinados a la exploración interactiva. Para muchos conjuntos de datos es imposible visualizar todos los flujos simultáneamente sin filtrarlos o agregarlos [Yuan et al., 2010]. Si queremos que los usuarios sigan siendo capaces de explorar los datos en cada detalle, entonces debemos proporcionar medios de interacción para controlar el filtrado, el zoom y la agregación. Se puede proporcionar soporte con las siguientes técnicas: Zoom y panorámico, *query* visual y filtrado.

Zoom y panorámica En técnicas que combinan espacialidad y temporalidad como *Flowstrates*, las tres vistas (el mapa de origen, el mapa de calor y el mapa de destino)

se pueden ampliar y panoramizar de forma independiente. Por lo tanto, el usuario puede enfocarse en diferentes regiones para los orígenes y los destinos y seleccionar la parte más relevante del mapa de calor.

Consulta visual y filtrado El filtrado es la forma más sencilla de hacer que una visualización sea más comprensible al reducir la cantidad de datos representados en la vista.

Cuando se realiza una selección, el mapa de calor se actualiza, de modo que solo se muestran los flujos entre las ubicaciones seleccionadas. El usuario también puede seleccionar un período de tiempo y los mapas de flujos se actualizan con la nueva información.

2.2.4. Estrategias de Exploración

Las visualización de datos de migraciones admiten tres estrategias básicas de exploración. Las primeras dos estrategias están relacionadas con la observación de los patrones en el mapa de calor y difieren en el inicio desde un lugar a otro. El último es sobre la comparación de ubicaciones o períodos de tiempo.

Ubicación -> Patrón espacial o temporal : seleccione una ubicación o una región en el mapa de orígenes, luego descubra qué está sucediendo en el mapa de calor o en el mapa de destinos.

Patrón temporal -> Ubicación : encuentre algo interesante en el mapa de calor, concéntrese en el período de tiempo cuando estaba sucediendo y descubra en los mapas geográficos cuáles fueron los orígenes y los destinos.

Comparación de dos ubicaciones : seleccione dos ubicaciones o regiones en el mapa de orígenes y compare los cambios temporales de los flujos de estas ubicaciones en el mapa de calor y sus respectivos destinos en el mapa de destinos.

2.2.5. Estrategias de Diseño

Sugerencias selectivas El nombre de los arcos en las visualizaciones se puede especificar explícitamente. Sin embargo, no es una buena idea saturar la visualización con información detallada. Nuestra técnica solo muestra las etiquetas de los acordes cuando se acerca y cuando mueve sobre una cuerda. De esta forma, el usuario puede recibir la información sobre la cuerda particular que le interesa.

Colores Los colores pueden jugar un papel importante en la visualización. Se considera dos opciones de diseño con respecto a la coloración del diagrama. La primera decisión es

colorear los arcos de forma diferente para facilitar su reconocimiento por los usuarios. La segunda decisión es colorear una cuerda como el mismo color del arco en el que la cuerda tiene una longitud más amplia. En el caso de la misma longitud, coloreamos el acorde en blanco. De esta forma, se pueden visualizar más aspectos de las relaciones entre los nodos sin hacer que el diagrama sea ilegible.

2.3. Datos OD temporales

Los datos de origen-destino son una forma de representar interacciones espaciales, es decir, flujos de entidades entre pares de ubicaciones geográficas. En los datos OD se especifican los orígenes, los destinos y las magnitudes de los flujos, mientras que las rutas exactas que tomaron las interacciones entre las ubicaciones no se conocen. Este tipo de datos es bastante importante y muy a menudo se convierte en un objeto del análisis [Yue, 2009]. Los datos OD a menudo se representan y almacenan como una matriz OD. En la matriz OD, las filas y las columnas corresponden a los orígenes y destinos respectivamente (En la Fig. 2.1 se especifican con los códigos de país) y los valores en las celdas de la matriz representan las magnitudes de los flujos entre los respectivos orígenes y destinos. Esta es esencialmente la misma que la representación matricial de adyacencia de los gráficos dirigidos ponderados.

	ESP	PER	MOZ	DEU	ISR	HUN	ARG
SEN	3.0	3.0	2.0	2.0	2.0	2.0	2.0
CHN	1.0	2.0	2.0	2.0	2.0	2.0	
CHL		7.0	1.0				
ECU	6.0	0.0	8.0	1.0	3.0	0.0	4.0
SRB	8.0	1.0	3.0				
IRQ	8.0	4.0	2.0	5.0	2.0	3.0	6.0
URY	1.0	1.0	1.0			1.0	
...							

Figura 2.1: Representación de matriz OD de datos OD.

Las matrices OD que representan flujos entre un conjunto de ubicaciones a menudo son dispersas, es decir, muchos pares de ubicaciones no están conectados, por lo tanto, la mayoría de las magnitudes de flujo son ceros. En tales situaciones es más eficiente para almacenar solo los flujos con magnitudes no nulas, por ejemplo, como una lista de adyacencia. En esta tesis, consideraremos datos temporales de OD, lo que significa que cada flujo entre cada origen y destino está asociado con un punto en el tiempo o un período de tiempo. En el caso más simple, dichos datos se pueden representar como una lista de ‘eventos de flujo’.

Dichos datos detallados sobre cada flujo individual no siempre están disponibles en los conjuntos de datos OD. A menudo, los datos se agregan temporalmente en función de algún período de tiempo, por ejemplo, un día o un año. Dichos datos se pueden representar en forma de matriz. En esta matriz, las dos primeras columnas especifican el origen y el destino y todas las demás especifican el año o el período de tiempo. Los valores en las celdas representan la magnitud total agregada de todos los eventos de flujo que ocurrieron en el año o el período de tiempo correspondiente entre los respectivos orígenes y destinos.

Origin	Dest	Magnitude	Time
SEN	ESP	4.0	2010-01-15 13:43:32
CHN	PER	1.0	2010-01-15 14:05:10
SEN	ESP	4.0	2010-01-16 10:07:31
CHN	PER	1.0	2010-01-16 01:20:32
SEN	ESP	4.0	2010-01-17 03:30:16
CHN	PER	1.0	2010-01-17 22:01:55
IRQ	HUN	1.0	2010-01-17 23:14:07
...			

Figura 2.2: Datos OD temporales como una lista de eventos de flujo.

En la figura 2.3 algunas celdas estos valores están vacíos, lo que significa que faltan datos. Al mismo tiempo, también hay celdas que tienen valores cero. Por lo tanto, existe una distinción entre los datos faltantes y los flujos de magnitud cero. Esta distinción puede ser muy importante para el análisis y la visualización debe tenerla en cuenta.

Origin	Dest	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
SEN	ESP	4.0	3.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	2.0
CHN	PER	1.0	1.0	1.0	1.0	2.0	2.0	2.0	2.0	2.0	2.0
CHL	MOZ										
ECU	DEU			6.0	0.0	8.0	1.0	3.0	0.0	4.0	7.0
SRB	ISR										
IRQ	HUN	1.0	8.0	0.0	8.0	4.0	2.0	5.0	2.0	3.0	6.0
URY	ARG						1.0	1.0	1.0		
...											

Figura 2.3: Datos temporales de OD como flujos agregados por año.

Para producir una visualización geográfica legible, generalmente se necesita más información sobre los orígenes y destinos, por ejemplo. Los nombres completos y las coordenadas geográficas de los lugares. Estos se pueden almacenar por separado y hacer referencia a ellos mediante un identificador (en el siguiente ejemplo, el identificador es el código del país):

Code	Name	Lat	Lon
LCA	Saint Lucia	13.903085	-60.9659
MDG	Madagascar	-18.054455	47.108621
UZB	Uzbekistan	41.447353	64.79929
LSO	Lesotho	-29.595733	28.244114
SLB	Solomon Islands	-8.910545	159.537743
MDV	Maldives	3.353159	73.260862
...			

Figura 2.4: Datos sobre los orígenes y destinos necesarios para producir una visualización geográfica.

El objetivo principal de los enfoques sobre datos OD es proporcionar información, es decir, hechos útiles o conocimiento sobre los datos OD, bajo análisis. Por lo tanto, las visualizaciones deben tener en cuenta ciertas exigencias para representar con precisión las características de los datos que son importantes para el análisis y comprensión del usuario. Las visualizaciones sobre datos OD generalmente se usan con uno de los tres siguientes propósitos:

- Exploración: consiste en analizar y encontrar patrones y relaciones en los datos.
- Confirmación: examinar y confirmar nuestra hipótesis sobre los datos.
- Comunicación: comunicar de la manera más apropiada un hecho conocido sobre los datos.

Capítulo 3

Visualización de Datos Origen Destino

El análisis de datos OD puede beneficiarse de las técnicas de *visual analytics*. En términos de estos, se reconocen las visualizaciones matriciales así como los diseños de grafos y los métodos de reducción de desorden como categorías importantes en la visualización de datos. Otros tipos de visualización de uso frecuente incluyen mapas de flujo y mapas OD . Las siguientes secciones analizan los tipos de visualización mencionados anteriormente, junto con otros métodos encontrados en la investigación sobre la visualización de datos espacio-temporales.

3.1. Visualizaciones con Grafos

La estructura de datos básica necesaria para los grafos es un conjunto dado de nodos y un conjunto dado de aristas, ya que a menudo se dibujan como diagramas de enlace de nodo. En el conjunto de datos OD, los nodos podrían representar los orígenes y destinos de los flujos, por ejemplo, Países, distritos, etc, según la escala espacial. Las aristas representan los flujos entre los respectivos orígenes y destinos [Wilkinson and Friendly, 2008] definen un grafo como una “red de nodos heterogéneos y enlaces”. Se utilizan grafos para organizar conceptos y relaciones como nodos y enlaces con la esperanza de descubrir tendencias, patrones e ideas clave. Un problema clave en la visualización de grafos es el tamaño del gráfico, es decir, el tamaño de los datos a visualizar [Zeng, 2013]. Los diferentes problemas al respecto, como las limitaciones dadas por las plataformas de visualización si el número de elementos es demasiado grande, o los problemas de visibilidad o usabilidad. Relacionan los problemas de usabilidad con las dificultades para diferenciar entre nodos y aristas si hay demasiados. Con una cantidad creciente de datos para mostrar, los grafos pueden llegar a ser demasiado complejos y sobrecargar la capacidad cognitiva del usuario. Por lo tanto, resulta difícil para el usuario realizar un análisis significativo. Debido a los problemas descritos anteriormente, la investigación a menudo se enfoca en maneras de resolver los problemas del desorden visual, por ejemplo, por agregación o técnicas de agrupamiento, que también es uno de los temas principales. Sander et al. investigan diagramas de enlace de nodo para visualización de datos OD. Su estudio se centra en el

problema de la saturación visual en las visualizaciones gráficas, ya que este es uno de los principales problemas en la representación de relaciones entre grandes datos. Herman et al. introducen un método para la agrupación de bordes basada en geometría para agrupar bordes en paquetes y, por lo tanto, reducir el desorden visual causado por el cruce del alto número de bordes (ver figura 3.1).

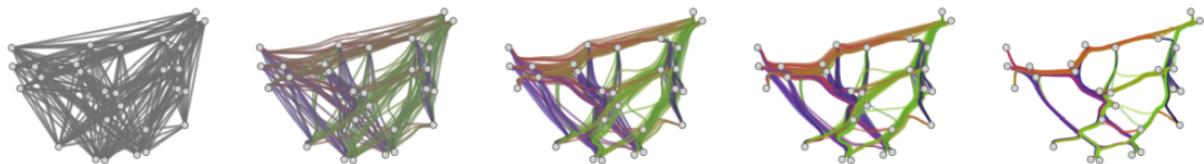


Figura 3.1: Una secuencia de animación para agrupamiento de bordes.

3.2. Matrices OD y Mapas OD

Como se mencionó anteriormente, otra forma frecuentemente utilizada para representar datos OD son las denominadas matrices de origen-destino (matrices *Origin Destination Matrix* (ODM) u OD). La matriz OD es el enfoque de matriz más básico, que representa cada origen como una fila (R) y cada destino como una columna (C) en la matriz. Cada celda (R, C) representa el flujo desde el origen hasta el destino (ver Figura 3.2 (a)). En cuanto a las visualizaciones gráficas, las posiciones abstractas de los nodos y enlaces en ODM (filas y columnas) no representan las ubicaciones geográficas reales [Yang et al., 2017]. Hay varios intentos para mejorar las matrices OD, por ej. en el desarrollo de matrices reordenables, en las que las filas y columnas pueden ser (re)ordenadas por ubicación (por ejemplo, de este a oeste o de norte a sur), jerarquía, tiempo, etc, [Guo and Zhu, 2014]. Otro enfoque se encuentra en los llamados mapas OD, con el objetivo de preservar la propiedades espaciales reales de los datos [Yang et al., 2017]. En el mencionado se usa mapas OD para visualizar la migración histórica dentro de Irlanda, Wood et al. usa esta representación para visualizar los flujos de migración de condado a estado para los Estados Unidos (ver figura 2.2 (b)). Wood et al. ordena las celdas de la matriz OD de acuerdo con su ubicación geográfica bidimensional original, evitando la superposición de celdas, también mapean vectores OD como celdas de una cuadrícula regular en lugar de líneas, y construyen un treemap espacial de dos niveles para preservar la distribución espacial de las ubicaciones de origen y destino. Los mapas OD son efectivos para estudiar los patrones espaciales de la migración interna.

Los mapas OD que proponen Wood et. al. intentan superar esta limitación a través de un diseño anidado de pequeños minimapas [Wood et al., 2010], como en la Fig. 2. Dicho enfoque proporciona información geográfica al dividir el mapa en una cuadrícula regular en función de las ubicaciones geográficas reales en el mapa utilizando una estructura *treemap* anidadas (mapa de jerarquías). Un segundo nivel de *treemaps* está integrado dentro del primero para presentar la información de OD utilizando sombreado de color. Como todas las ubicaciones tienen celdas de tamaño similar, esto permite la lectura de datos de una manera más gráfica y precisa para indicar la densidad de flujos y cantidades.

Una desventaja que observaron en este tipo de visualizaciones es que los países y sus divisiones son extensas y numerosas, los *treemaps* serán numerosos también, y la visualización se hará más difícil de leer, es lo que los autores llaman ‘distorsión de mapa geográfico’. Un enfoque similar es propuesto por Yang et. al., llamado MapTrix, el cual está diseñado para mostrar datos de densidades de flujo de múltiples fuentes junto con su información geográfica asociada [Yang et al., 2017]. Presenta tres componentes principales: (a) un mapa de origen, (b) un mapa de destino, (c) una matriz de origen-destino, (d) líneas que conectan la matriz con el mapa geográfico con cada columna y fila. MapTrix, explota los recursos mejor que otras visualizaciones basadas en Matrices, según Yang et. al. los usuarios aceptaron y se relacionaron mejor con su visualización. Sin embargo, este recurso no presenta el componente temporal. Ante la falta de un componente temporal, Boyandin et. al. proponen *FlowStrates*, que enfatiza el uso de indicadores como componente temporal representado por un mapa de calor [Boyandin et al., 2011]. *FlowStrates* hace uso de 3 componentes, uno en la izquierda que indica el origen, en el medio el mapa de calor, y a la derecha los destinos. La presencia de los 3 nos ayuda a la visualización a indicar: origen y destino, componente de tiempo y densidad de flujo. Este tipo de visualización aporta el componente temporal a diferencia de los anteriores enfoques, indicando la densidad de cada movimiento en relación con el tiempo. *FlowStrates* conecta un mapa de calor temporal con dos mapas que presentan las ubicaciones geográficas de origen y destino y muestra cómo el flujo cambia con el tiempo. *FlowStrates* se basa más en la interactividad con los usuarios al poder hacer las consultas mediante lazos para unir diversas regiones y compararlas entre sí.

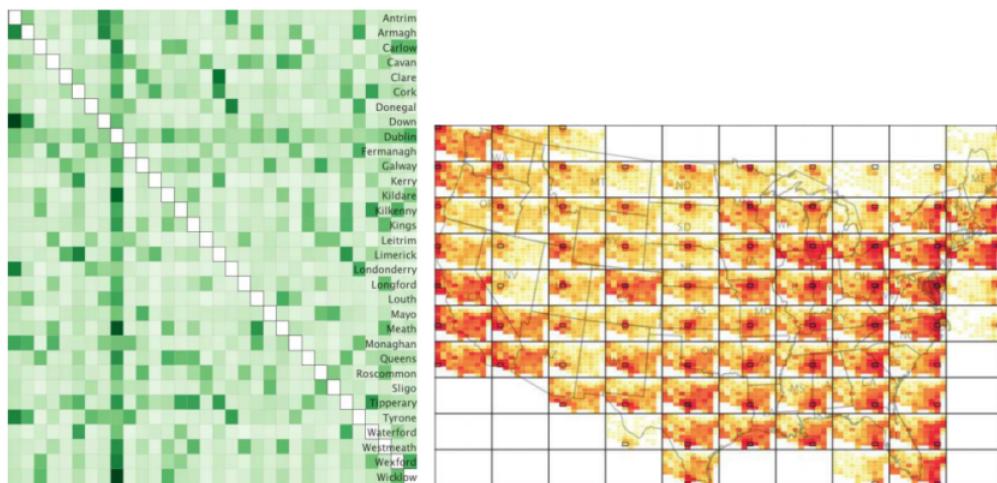


Figura 3.2: a) Matriz OD que muestra los flujos de condado a condado en Irlanda; filas = condados de origen, columnas = condados de destino. b) Los vectores de migración de condado a condado de EE. UU. Se muestran como un mapa OD. Cada celda de cuadrícula grande representa la ubicación de origen, dentro de la cual se muestra el mapa de densidades de destino que utilizan la misma cuadrícula

3.3. Mapa de Flujos

Los mapas de flujo se conocen desde hace mucho tiempo por representar flujos de varios tipos. Hasta hoy, los mapas de flujo siguen siendo populares para visualizar varios

tipos de flujos. Dichos mapas son muy intuitivos, lo que puede ser una de las razones de su popularidad. Sin embargo también se observa que los mapas de flujo son adecuados para mostrar flujos de una o muy pocas fuentes, pero no tanto para representar un gran número de fuentes (u objetivos). De nuevo, como se ha discutido para otros tipos de visualización, la saturación y la sobrecarga visual se convierten en un problema con grandes cantidades de datos [Yang et al., 2017].

Los ejemplos más recientes de mapas de flujo incluyen visualizaciones relacionadas con el movimiento de personas y flujos de comercio. Afirman que el mapeo de flujo ayudó a proporcionar descripciones estructurales de patrones, de lo contrario no es fácilmente observable. Boyandin presenta una Herramienta llamada JFlowMap que proporciona varios métodos de visualización para crear y analizar los mapas de flujo. Presentan la herramienta aplicándola a los datos de migración (ver figura 2.3).



Figura 3.3: Flujos de refugiados entre los países del mundo en 1996,2000

3.4. Mapa de coropleta

Los mapas de coropletas se utilizan con frecuencia para visualizar datos que se agregan a áreas administrativas como países, estados, distritos, etc (ver figura 3.5). Una aplicación de este tipo de visualización es mostrar cambios en ciertas áreas entre diferentes momentos en el tiempo, por ejemplo. mediante el uso de tonos de dos colores para mostrar la magnitud del aumento y la disminución del valor de un atributo, respectivamente [Andrienko et al., 2010]. Un error que se ve a menudo es que los valores de datos sin procesar se representan en lugar de los valores normalizados para crear un mapa de densidad, además, señalan que el tamaño de las áreas de la unidad puede afectar la percepción del analista de los valores sombreados [Heer et al., 2010]. Este problema también fue mencionado por Heer et al. quien mostró la migración interna histórica en Irlanda con mapas de coropletas, para asegurar una mayor visibilidad visual de cada condado. Sander et al. mencionan que los mapas de coropletas no son ideales para analizar conjuntos de datos con un gran número de interacciones espaciales, ya que es difícil visualizar grandes colecciones de datos con estos métodos básicos de visualización. En los nuevos enfoques, los datos se dividen en subconjuntos que permiten controlar los cambios en la variable dependiente que están relacionados con dos posibles variables explicativas.

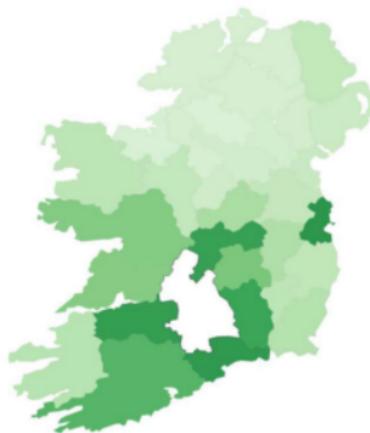


Figura 3.4: Mapa de coropleta que muestra la migración fuera del condado de Tipperary en Irlanda (blanco)

3.5. Mapa de símbolo graduado

Para mostrar valores absolutos se sugieren utilizar mapas de símbolos graduados en lugar de mapas de coropletas, que evitan confundir el área geográfica subyacente con los valores de los datos [Heer et al., 2010]. Además, los mapas de símbolos graduados permiten representar más de una dimensión utilizando varias variables visuales (forma, tamaño, color del símbolo). En su estudio sobre los métodos de agregación utilizados para la exploración visual de los datos de movimiento, Andrienko et al. recomiendan mapas de símbolos graduados para una vista del movimiento orientada a la situación (por ejemplo, el análisis del uso del espacio y la accesibilidad). Como métodos de agregación relevantes, señalan S (agregación espacial) y T (posición de los elementos en diferentes eventos discretos o instantáneas). También se utilizó mapas de símbolos graduados para representar los flujos migratorios hacia la ciudad de Manchester (ver figura 3.5). donde se convirtió las líneas de entrada en puntos y mostró los cambios en las cantidades utilizando las variables visuales ‘tamaño y color’ [Rae, 2009].

3.6. Series de Tiempo

El análisis de la dimensión temporal en los datos es importante en muchas áreas de investigación, pero presenta desafíos para la visualización y el análisis de los datos. Las representaciones visuales de los datos temporales se pueden dividir en categorías según las características de tiempo que muestran: 1) tiempo lineal frente a tiempo cíclico, 2) puntos de tiempo frente a intervalos de tiempo, 3) tiempo de orden vs. tiempo de ramificación frente a tiempo con múltiples perspectivas. Por ejemplo, los gráficos de líneas son prácticos para mostrar tendencias generales o valores atípicos, mientras que las visualizaciones en espiral o mapas de anillos pueden mostrar patrones cíclicos en los datos [Gu et al., 2014]. Andrienko et al. mencionan varios aspectos relacionados con la visualización de datos de series de tiempo para mejorar la efectividad de su análisis, por ejemplo, el acceso interacti-

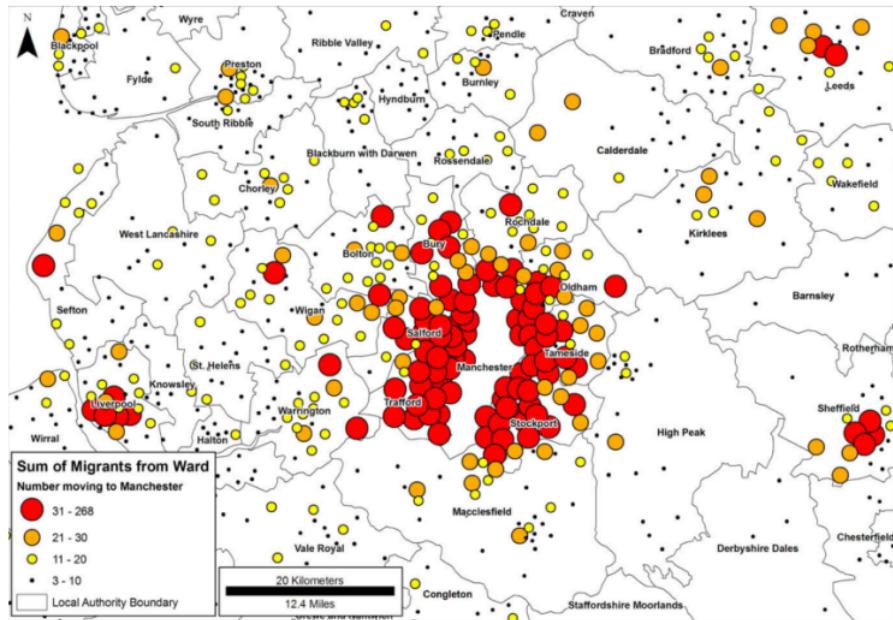


Figura 3.5: Mapa de símbolo graduado que muestra las entradas de migración a Manchester

vo a los valores reales al desplazarse sobre una parte de la línea de tiempo o la posibilidad de mostrar múltiples líneas simultáneamente, para compararlos en su estudio, evalúan varios métodos para explorar datos de series de tiempo que se distribuyen espacialmente. Proponen tres métodos diferentes para a) proporcionar una visión general de los cambios de valor en toda la región, b) detectar patrones espaciales de comportamiento temporal idéntico y c) detectar patrones espaciales con cambios similares a lo largo del tiempo. Para b) y c) utilizaron pantallas vinculadas dinámicamente para la visualización. Ellos tambien sugieren un enfoque diferente para mejorar la usabilidad de las visualizaciones de series de tiempo. Introducen las llamadas "cajas de tiempo" para consultar los datos y seleccionar una parte específica que sea de su interés (ver figura 3.6). El cuadro ayuda a seleccionar datos al especificar su extensión en el eje de tiempo (período de tiempo de interés), así como el eje y (rango de valores de interés en el período de tiempo dado). Por lo tanto, la caja de tiempo actúa como un filtro para elegir valores en un rango dado durante el intervalo de tiempo especificado por el ancho del cuadro [Ward et al., 2010].

3.7. Diagrama de Cuerdas

Los diagramas de cuerdas se han utilizado en estudios recientes para visualizar flujos, por ejemplo, personas, tráfico. Sander afirma que su objetivo fue principalmente visualizar flujos de datos complejos para una audiencia más amplia, incluidos usuarios no técnicos o no científicos. Mientras que encuentran que los patrones en el diagrama de migración circular (ver figura 3.7) mejoran no solo la capacidad de comunicación, sino también de analizar y comprender los patrones en los datos, afirman que una gran cantidad de tiempo fue necesaria para la transformación de datos con el fin de producir diagramas de cuerdas.

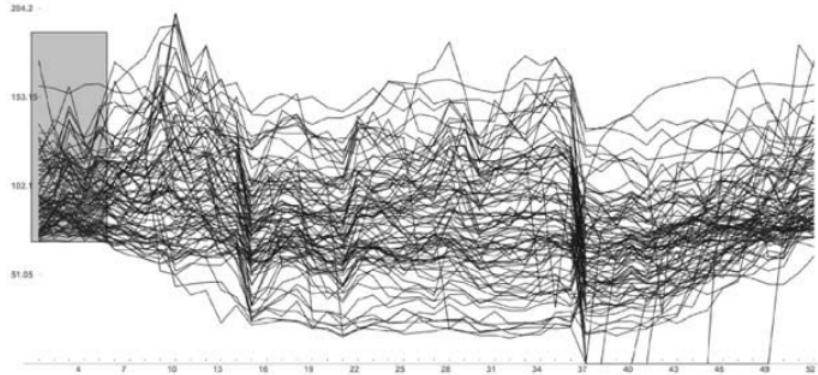


Figura 3.6: Gráfico de series de tiempo que muestra 52 precios de acciones semanales para 1430 acciones (eje x = semanas 52, eje y = precios de acciones). Arriba: consulta de un solo timebox para artículos que se intercambiaron entre 70 y 190 durante las semanas 1-5.

3.8. TreeMap

Un *treemap* es un diagrama de recinto de relleno de espacio (ver figura 3.8). Como sugiere su nombre, utiliza la contención en lugar de la adyacencia para visualizar la jerarquía en los datos (Heer, Bostock y Ogievetsky, 2010). Como afirman los autores, un *treemap* subdivide el área total en rectángulos de manera recursiva. Esto permite al analista recuperar rápidamente información sobre los datos que se visualizan.

3.9. Consideraciones Finales

En este capítulo, analizamos las diferentes técnicas para datos OD, de las cuales vemos su principal contribución y aporte al estado del arte actual. Vemos técnicas como los mapas de flujo, la representación más utilizada de los datos OD, así como también analizamos matrices OD y visualización circular con cuerdas (*chord*). Intentamos identificar y analizamos el estado actual de la investigación sobre la eficiencia de estas técnicas. Las técnicas que discutimos en este capítulo siempre se han entendido completamente. Conocer las limitaciones y las formas de lidiar con ellos, entender las tareas que estos abordan bien y no tan bien y tener una variedad de técnicas y representaciones alternativas que puedan abordar las debilidades abordando mejor ciertas tareas. En el siguiente capítulo, desarrollamos la implementación y la propuesta, así como la arquitectura que se seguirá para el desarrollo de la herramienta.



Figura 3.7: Diagrama de migración circular interactivo de flujos de migración entre regiones en 2005-10

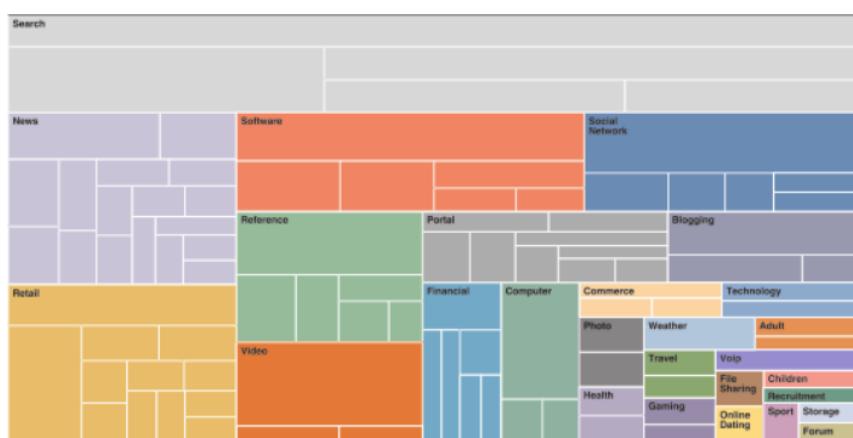


Figura 3.8: Treemap que muestra el número de usuarios únicos de los 100 sitios web principales en enero de 2010.

Capítulo 4

Propuesta e Implementación

En la presente sección definimos y exponemos los diferentes conceptos, conocimientos y detalles de la implementación de la propuesta. En esta sección también abordamos la parte de la implementación y la arquitectura que seguiremos para realizar nuestra tesis, ver Figura 4.1.

4.1. Arquitectura de la Propuesta

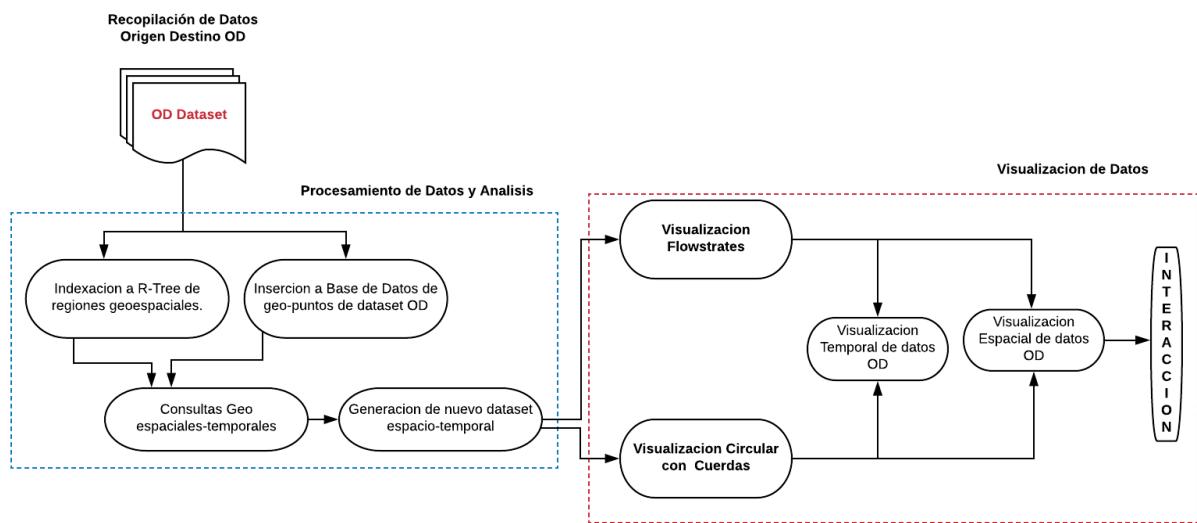


Figura 4.1: Arquitectura de la propuesta

En nuestra arquitectura, nosotros definimos 6 etapas, una primera etapa de recolección de *datasets*, una segunda etapa de evaluación de datos sin procesar, una siguiente etapa de procesamiento de datos, un cuarta etapa de analizado y filtrado de datos relevantes para nuestra evaluación, la siguiente etapa es la implementación del *framework*, la cual envuelve la etapa de visualización, la cual esta alimentada por diversas técnicos de visualización, entre ellas: diagrama de cuerdas y visualización *flowstrates*.

En el apartado actual, describiremos los aspectos a tener en cuenta en el desarrollo de nuestra propuesta. El primer paso es preprocessar un *dataset*, en nuestro caso el dataset de taxis de *New York City* [Yassky and NYC, 2016], estos datos están alojados en Internet y están disponibles para ser descargados [Whong, 2017]. Dicho preprocessamiento aborda etapas como limpieza de registros dañados o nulos, eliminando de filas innecesarias, para al final solo tener componentes espaciales y temporales del *dataset*. A continuación debemos ingresar el *dataset* preprocessado y poder reconocer los componentes espaciales y temporales por separado. Usamos una base de datos para la indexación y mejor manejo de datos. Realizamos consultas espaciales y *queries* sobre los datos indexados, en dichos datos también se realizará un proceso de *clustering* para un mejor análisis y manejo de datos. Una vez hecho el análisis e indexado de los datos, se le pasa los datos '*raw*' a cada una de las visualizaciones a mostrar.

4.1.1. Entrada de datos OD

Como input para nuestro *framework*, preprocessaremos el *dataset* de los taxis de *New York*. Dichos *datasets* cuentan con miles de viajes, y componentes de temporales (fecha y hora en formato GMT) y componentes espaciales (latitud y longitud en formato numérico) [Limousine Commission NYC, 2015]. El *dataset* usará extensión .csv.

4.1.2. Preprocesamiento de datos de entrada

El dataset de taxis de New York, será preprocessado usando una estructura de datos para indexar primero los 13 Community Districts de New York. Luego se procederá a realizar una operación de comparación con el bounding box de los Community Districts versus los puntos presentes en el dataset. Este proceso para saber cuantos viajes de taxis existen en cada uno de los Community Districts.

4.1.3. Visualización del Dataset

Una vez los datos están colectados y ubicados en un respectivo Community District, es posible visualizarlos para mostrar sus componentes temporales y espaciales, usando las técnicas de visualización Flowstrates y Circular con Cuerdas.

4.2. Implementación

En este apartado, describiremos de manera mas profunda la forma en como desarrollamos y explicamos la implementamos:

4.2.1. Tratamiento del Dataset

En este documento, analizamos y mapeamos el conjunto de datos de viajes de taxis del *New York City* (NYC) (New York City) Taxi Limousine Commission de Estados Unidos, del estado de New York del año 2016. El dataset esta compuesto por diversos campos, que la comisión de Taxis de New York colecta de cada viaje realizado por cada taxi o limusina del periodo 2016. Los campos más relevantes presentes son: vendorId, pickupDatetime, dropoffDate-time, passengerCount, pickupLongitude, pickupLatitude, dropoffLongitude, dropoffLatitude, paymentType, totalAmount, etc. Para trabajar con ese dataset en formato ‘raw’, se tuvo que investigar la manera de agruparlos a un lugar en específico del mapa de NYC.

Investigamos que el Estado de New York, tiene 5 Boroughs (burgo o ciudad), que son Manhattan, Bronx, Queens, Brooklyn y Staten Island, y cada burgo contiene los llamados community districts (distritos comunitarios), unos 70 en total. De los 70 community districts solo se trabajaran con 13 en total, todos ubicados en el burgo de Manhattan, debido a que en su mayoría los viajes se dan entre Manhattan. Para llegar a este análisis, se hizo uso de una herramienta desarrollada en Javascript para graficar los puntos de recojo pick up y los puntos de desembarco drop off. Los campos principales a tener en cuenta serán: pickupDatetime, pickupLongitude, pickupLatitude, dropoffLongitude, dropoffLatitude. En función a los campos geográficos, que son los 4, se los ubicará en los 13 distritos comunitarios del Estado de New York del burgo de Manhattan. Para aplicar nuestro enfoque, cada distrito comunitario se convierte en un punto geográfico dentro del mapa de New York.

4.2.2. Visualización de Datos OD

Debido a que el *dataset* se presenta como un conjunto de columnas y filas en un formato descrito anteriormente, es necesario hacer un pre-procesamiento de los datos. El *dataset* presenta coordenadas geográficas que se usarán para mapear e indicar a qué zona del Estado de *New York* pertenecen, en este trabajo se hace el procesamiento hacia los distritos comunitarios de *New York*, 13 de 70 en total. Para las implementaciones: Visualización *Flowstrates* y Diagrama Circular de Cuerdas, se hace una etapa de pre-procesamiento de datos, debido a que cada implementación, cada librería, y cada formato de lectura maneja su propia forma de leer y estructurar los datos. En la etapa de pre-procesamiento se usó herramientas de procesamiento para *python*.

Visualización Flowstrates Para cargar nuestro dataset con Visualización *Flowstrates*, es necesario configurar el archivo .jfmv (*flowmap* o *flowstrates*). Se necesita crear un archivo de configuración de vista (un archivo .jfmv) que luego se puede usar para abrir la visualización en la versión de *applet*. En el archivo .jfmv, se configuran los *datasets* necesarios para la visualización, como son los ‘*nodes*’, es decir las locaciones, con su respectiva coordenada geográfica, los ‘*flows*’ o los flujos desde el Origen A hacia el Destino B, se configuran además archivos *shapefile*, lo que ayudaran al dibujado del mapa del mundo, en nuestro caso del mapa de Estados Unidos, ver Figura 4.2.

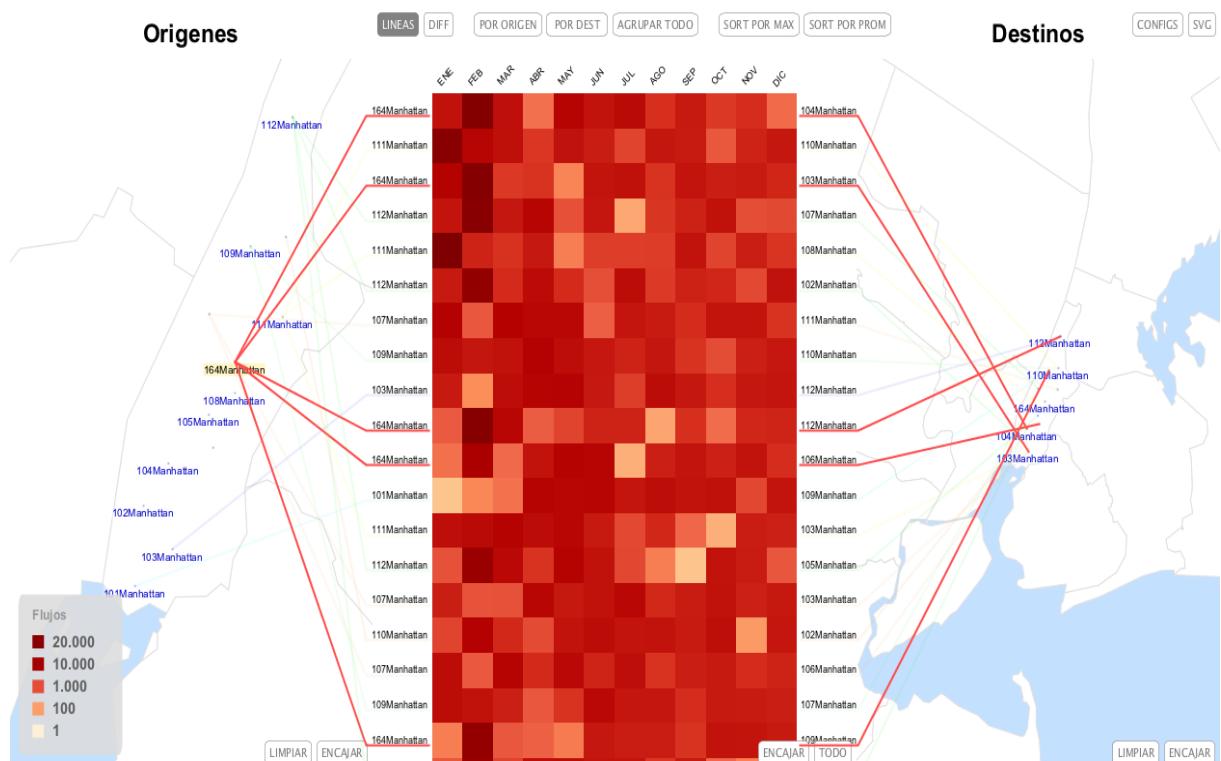


Figura 4.2: Visualización *Flowstrates*. Visualización del *dataset* de Taxis de NYC de los 13 (CD) distritos comunitarios del burgo de *Manhattan* del año 2016. Las líneas de color más fuerte, comunican que se hizo una selección en el Origen 106 *Manhattan* y nos indican la posición de ese origen en el mapa de calor. En el mapa de calor también es posible ver que a mientras más ‘color oscuro’ presente un mes mayor flujo de movimientos tiene.

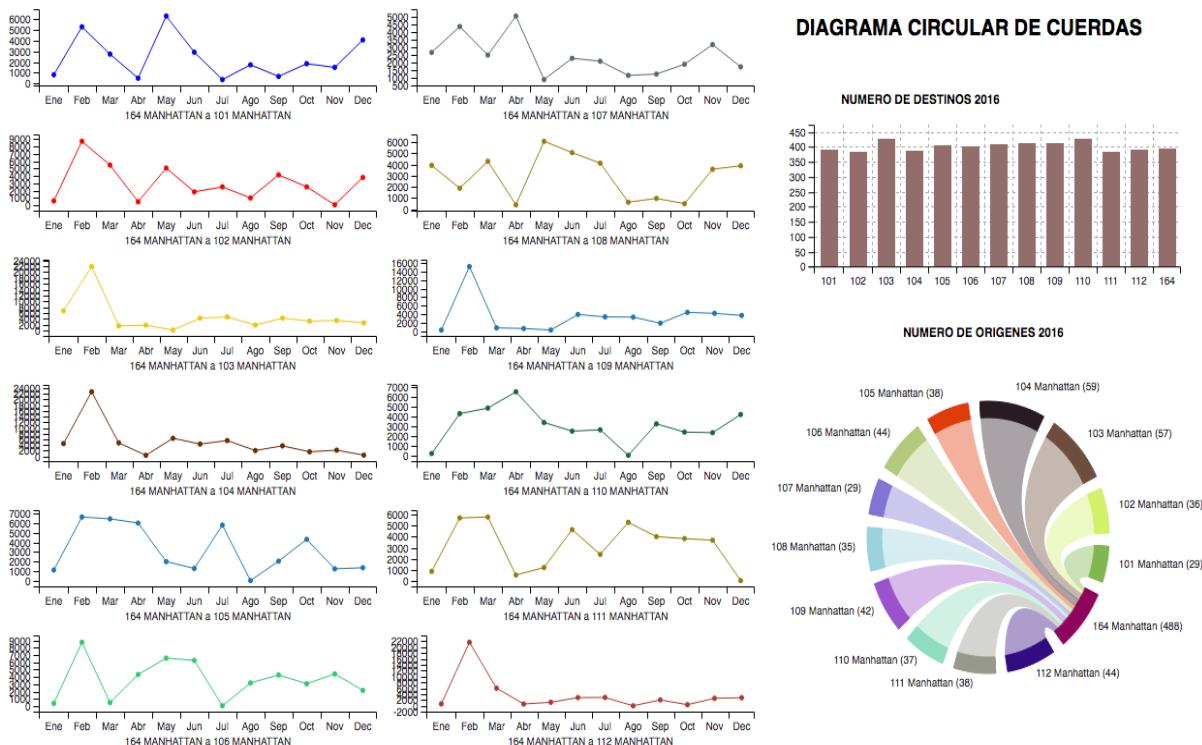


Figura 4.3: Visualización Circular con Cuerdas. Visualización del *dataset* de Taxis de NYC de los 13 distritos comunitarios de *Manhattan*. Usamos una visualización circular con cuerdas para indicar el número de emigraciones, en combinación con unos gráficos X e Y, para agregarle el componente temporal, y graficar los meses del 2016. Así mismo usamos un gráfico de barras para indicar el número de inmigraciones hacia cada distrito comunitario.

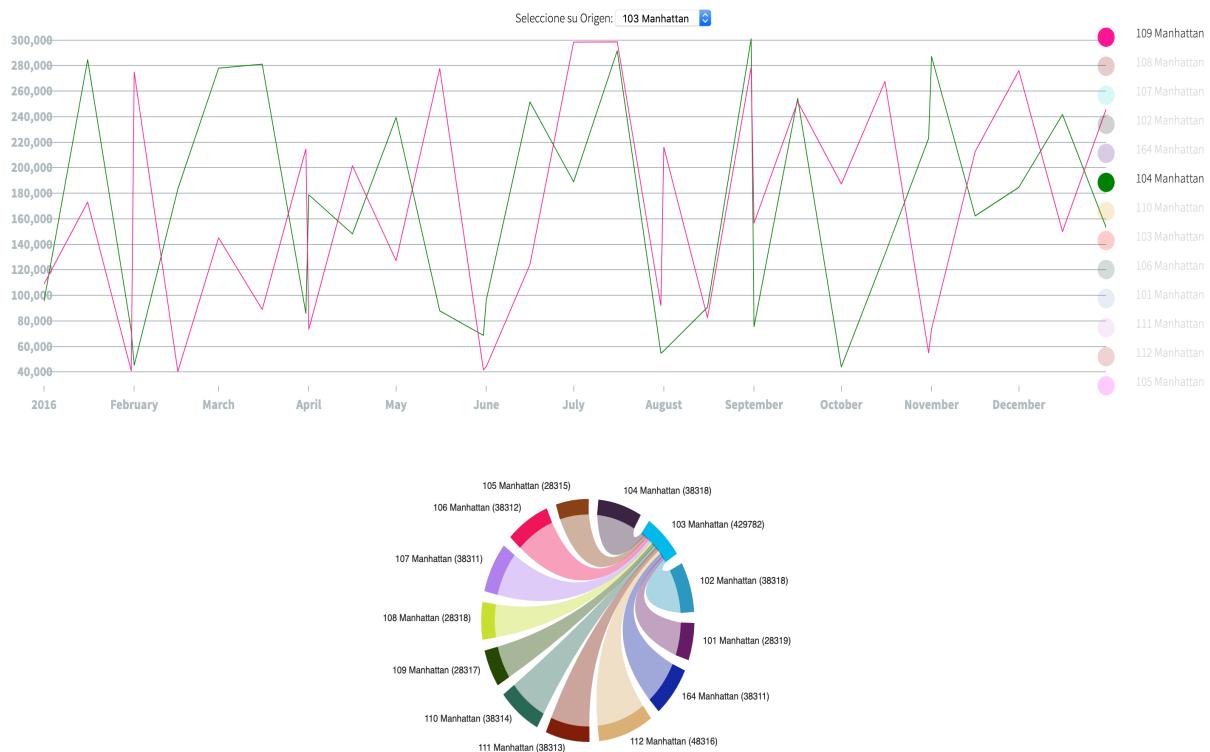


Figura 4.4: Modificación de la Visualización Circular con Cuerdas inicial, ahora nos permite visualización **todos versus todos**. Visualización del *dataset* de Taxis de NYC de los 13 distritos comunitarios de *Manhattan*. En esta vista observamos que al pasar el mouse por el distrito 103 de Manhattan, nuestra vista superior nos muestra un gráfico temporal, en el cual podemos comparar dos distritos en simultaneo

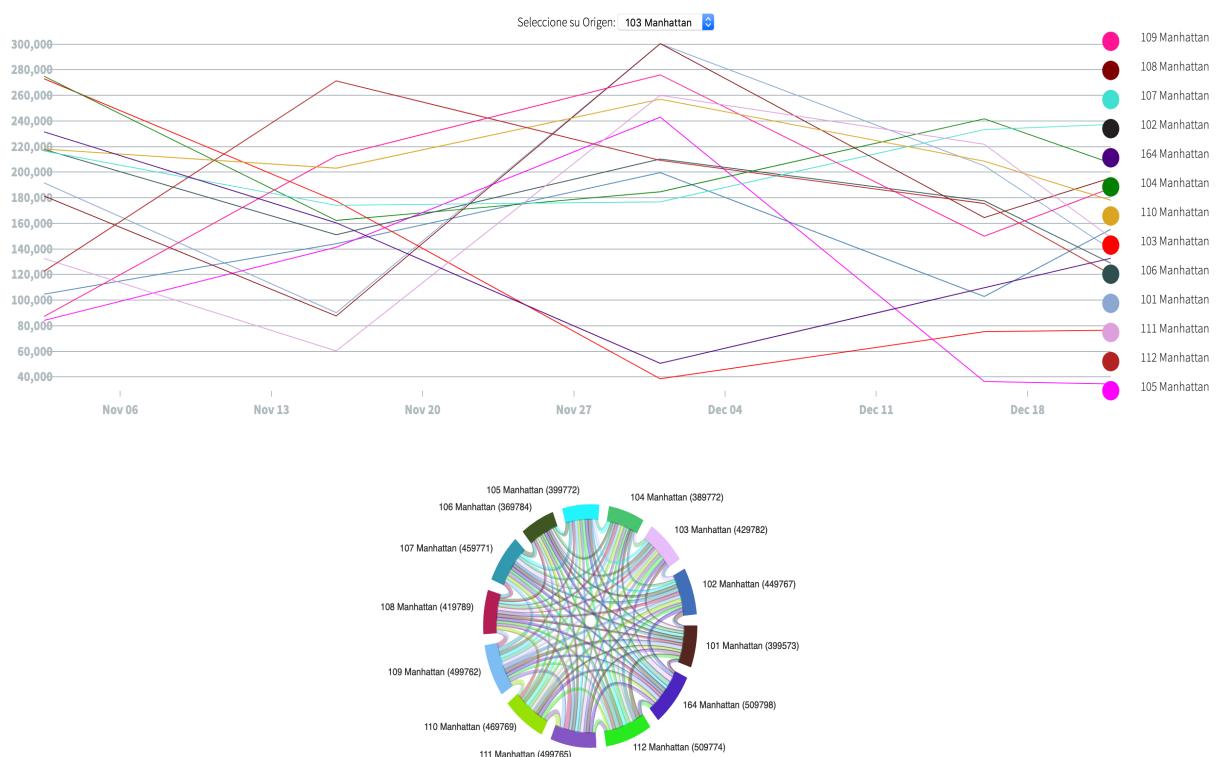


Figura 4.5: Modificación de la Visualización Circular con Cuerdas inicial, ahora nos permite visualización **todos versus todos**. Visualización del *dataset* de Taxis de NYC de los 13 distritos comunitarios de *Manhattan*. En esta vista observamos que al pasar el mouse por el distrito 103 de Manhattan, nuestra vista superior nos muestra un gráfico temporal, en el cual podemos comparar todos los 13 distritos en simultaneo, así mismo se hizo un **zoom temporal** para ver a detalle

Visualización Circular de Cuerdas Para cargar nuestro *dataset* con Visualización Circular con Cuerdas, se usó el mismo *dataset* de Taxis de NYC del año 2016, pero esta vez tomando solo en cuenta los 13 distritos comunitarios de New York del burgo de Manhattan. Una vez preprocesado y construido el *dataset*, es necesario hacer un conteo de los números de viajes desde un origen A hacia una destino B, el producto final de ese conteo se guarda en un *dataset* que contiene el punto de origen, el punto final y el número de viajes que se hizo desde A hacia B. Todo lo anterior ha sido hecho para plasmarlo y visualizarlo con la Visualización Circular de Cuerdas. Para los 12 gráficos X e Y presentes también en la visualización, se hizo un pre-procesamiento esta vez tomando en cuenta el mes en el que se hizo el viaje del Taxi, tomamos 12 meses del año 2016, y los pre-procesamos del *dataset* inicial, pero tomando en cuenta los meses en los que se efectuaron, una vez preprocesado guardamos ese *dataset* con el punto de Origen A, el punto de Destino B, los 12 meses del año 2016, y su correspondiente número de viajes hechos en cada uno de los meses, ver figura 4.3.

4.3. Definición de Pruebas

En esta sección, presentamos a detalle cómo evaluamos nuestras técnicas, para ello desarrollamos experimentos plasmados en tareas dadas al usuario. En base a esas tareas, evaluamos y definimos un modelo de entrevista y recolectamos los datos para cada tarea de usuario, para obtener resultados y de esa manera poder comparar qué técnica es mejor que otra.

4.3.1. Diseño de Experimentos

Para la evaluación de nuestras técnicas descritas anteriormente, se propone al usuario una serie de tareas que tienen el trasfondo espacial y temporal, a fin de evaluar el desempeño de cada una de las técnicas :

- Tarea 1: Consultas espaciales
- Tarea 2: Consultas temporales
- Tarea 3: criterio usuario

Tarea 1

Para la Tarea 1: de identificación de Regiones de mayor emigración/inmigración, el usuario tendrá que identificar espacialmente que distritos comunitarios de *New York*, presentan un mayor cantidad de flujos, tanto en la acción de emigración como inmigración. Dicho conocimiento espacial, será presentado en igualdad de condiciones por las técnicas implementadas.

Tarea 2

Para la Tarea 2: de identificación de meses o mes de mayor emigración/inmigración, el usuario tendrá que identificar temporalmente que mes o meses, presentan una mayor cantidad de flujos, tanto en la acción de emigración como inmigración. Dicho conocimiento temporal, será presentado en igualdad de condiciones por las técnicas implementada.

Tarea 3

Para la tarea 3, los usuarios evaluaran la interactividad y el conocimiento que obtiene de las visualización de las técnicas mostradas. La interactividad engloba aspectos como: usabilidad de la visualización y experiencia de usuario. El conocimiento abarca la calidad y cantidad de las conclusiones que el usuario genera gracias a la visualización, ejemplo (inferir eventos, fenómenos, cambios drásticos).

4.3.2. Evaluación de Tareas

En función a la definición de Tareas, podemos obtener métricas que nos ayudarán a medir y comparar cada una de las técnicas propuestas.

Se presentaran ambas técnicas a usuarios potenciales y personas cercanas, se les explicará el uso de ambas técnicas, y se le acompañará al usuario en el momento del uso de cada técnica.

Diseño de la Entrevista

1. Presentación de la motivación del trabajo.
2. Demostración del uso de las técnicas.
3. Exploración por parte del usuario.
4. Aclaración y proposición de Tareas.
5. Planteamiento de Tareas 1 y 2.
6. Explicación detallada de la Tarea 3.
7. Toma de tiempo y evaluación a criterio

Capítulo 5

Pruebas y Resultados

Evaluamos las dos técnicas implementadas en función a las tareas definidas. Presentamos nuestras técnicas a Sesenta y un (61) personas diferentes, a las cuales se les explico siguiendo el Diseño de la Entrevista que proponemos en la anterior sección.

Como hemos visto el diseño de visualizaciones para el análisis de cambios temporales en datos OD es una tarea difícil. La naturaleza compleja de los datos hace que sea difícil encontrar la representación más adecuada que muestre cómo cambian las relaciones espaciales a lo largo del tiempo mientras se mantiene intacta la metáfora geográfica.

5.1. Diseño del Estudio

El objetivo de nuestro estudio es encontrar cómo se comparan las diferentes técnicas de visualización de flujo animados y pequeños múltiples en términos de los tipos de observaciones que las personas hacen con ellos. Por lo tanto, analizamos los hallazgos realizados por los participantes del estudio en estas dos condiciones. Realizamos un experimento con sujetos que eran estudiantes de Ciencia de la Computación de la Universidad Católica San Pablo y con personas sin conocimiento experto en flujos migratorios o visualización geográfica. Las tareas fueron divididos en 3 categorías: tareas espaciales, temporales y basadas en el criterio del usuario. La primera ronda del estudio se diseñó como un experimento entre sujetos. A cada grupo se le asignó una de las dos vistas, ya sea FlowStrates y Visualización circular con cuerdas en sus variación 1 y 2. Se pidió a los participantes que exploraran los datos interactuando con la vista y que documentaran sus hallazgos.

Los hallazgos se recopilaron en una base de datos y, posteriormente, se clasificaron manualmente. Luego realizamos una comparación detallada de los tipos de hallazgos de la primera ronda entre las vistas. Hubo una segunda ronda adicional en nuestro experimento. Se pidió a los sujetos que continuaran haciendo observaciones con el mismo conjunto de datos, pero en la vista que no usaron en la primera ronda. Nuestro objetivo fue ver si el cambio de una vista a otra mientras se analiza el mismo conjunto de datos induciría a los sujetos a realizar hallazgos de diferentes tipos en comparación con los que obtuvieron en la primera vista. Por lo tanto, no comparamos los hallazgos hechos por los dos grupos

de sujetos en la segunda ronda. En su lugar, comparamos los tipos de hallazgos que los sujetos realizaron en la segunda ronda con los que los mismos sujetos hicieron en la primera ronda. Preferimos este enfoque sobre el diseño dentro del sujeto, ya que con un experimento dentro del sujeto, sería mucho más difícil analizar los efectos de cambiar de vista mientras se sigue explorando el mismo conjunto de datos sobre los tipos de hallazgos que hacen los sujetos. La razón principal para diseñar la primera ronda como un experimento entre dos sujetos fue también nuestro deseo de analizar los efectos del cambio de vista en la segunda ronda.

Al comienzo de cada sesión los sujetos fueron orientados. Las visualización y sus interacciones fueron presentados y explicados. Los sujetos podrían interactuar con las vistas por un tiempo y hacer preguntas sobre ellos. La orientación se realizó con un conjunto de datos similar al utilizado. Después de la fase de orientación que duró entre 5 y 10 minutos, comenzó la primera ronda. Se pidió a los sujetos que hicieran hallazgos en los datos utilizando la vista seleccionada automáticamente según el grupo en el que se encontraban. Más precisamente, se asignó a los sujetos la siguiente tarea: *Las vistas que verán representan los flujos migratorios de taxis de New York City. Explore estas vistas y escriba los hallazgos importantes que haga*

También les dimos a los sujetos una idea de lo que es un hallazgo importante: *Cuando decida qué hallazgos son importantes, imagine que tiene que usarlos para presentar a otra persona lo que ha aprendido sobre estos datos.*

Decidimos no dar ejemplos de hallazgos para asegurarnos de que los sujetos no estén sesgados por sus tipos particulares. El objetivo era ver qué tipo de hallazgos obtendrían los sujetos por su cuenta.

El sujeto tenía que describirnos que de manera corta en una frase, que hallazgos encontró o que patrones identificó. Después de que un sujeto nos comentó un hallazgo, el hallazgo se almacenó en la base de datos, de modo que se pudo enviar un nuevo hallazgo. Si un sujeto considerara que no se podrían hacer hallazgos más importantes en él, se termina la etapa de orientación y encuentro de hallazgos.

Después de cada ronda, los sujetos calificaron los hallazgos que acaban de hacer por su importancia en una escala con cuatro opciones entre "No importante" "Muy importante". También tuvieron la posibilidad de marcar un resultado como "Incorrecto" si descubrieron un error, pero no se les permitió cambiar los resultados sobre sus hallazgos.

En la segunda ronda, se pidió a los usuarios que continuaran la exploración del mismo conjunto de datos durante 10 minutos adicionales, pero utilizando la otra visualización. Para evitar demasiada fatiga, y teniendo en cuenta el hecho de que los usuarios ya estarían familiarizados con los datos, decidimos acortar la segunda ronda. Durante el análisis no comparamos el número absoluto de hallazgos realizados por los sujetos, pero los porcentajes promedio de los tipos de hallazgos realizados en cada una de las rondas.

Finalmente, después de completar la segunda ronda y calificar los hallazgos, a los sujetos se les hicieron estas preguntas:

-
- ¿Cuál de las técnicas prefieres en general?

- ¿Cuál de las técnicas fue más fácil de usar?
- ¿Cuál de las técnicas permite un mayor número de descubrimientos?
- ¿Con cuál de las técnicas es más probable que se pierda información relevante?

Las posibles respuestas fueron: “animación”, “pequeños múltiplos”, “no hay diferencia”, “No sé”. También se pidió a los participantes que calificaran sus impresiones generales de las dos condiciones en una escala con cinco opciones y que describieran las fortalezas y debilidades de las condiciones. El cuestionario terminó con una pregunta abierta para cualquier comentario general o sugerencia. El estudio se realizó en un ambiente universitario y en algunos casos en casa, también en las oficinas de HolaCliente, frente a la misma computadora. El autor estuvo sentado al lado del sujeto como orientador durante toda la sesión.

5.2. Las Condiciones

Las condiciones que utilizamos en el experimento se basaron en las técnicas de Flowstrates y Visualización Circular con Cuerdas. En esta representación, los flujos de personas que migran entre los países del mundo se muestran con líneas rectas que conectan a los países en un mapa geográfico. Los anchos y los colores de cada línea de flujo representan el número de personas que migran. Decidimos no mostrar las direcciones de los flujos para simplificar las vistas y evitar el desorden adicional. Para este experimento, estábamos más interesados en analizar los hallazgos relativos a los flujos que cambiaron sus magnitudes a lo largo del tiempo, no en las direcciones del flujo.

Los sujetos tuvieron la posibilidad de resaltar un flujo al pasar sobre él con el mouse. Cuando se destacó un flujo, se mostró información detallada sobre él, a saber, el origen y los destinos y el número de personas que emigran entre ellos. En ambas condiciones, FlowStrates y Visualización Circular con Cuerdas, fue posible hacer interactuar con el mouse. La interacción en FlowStrates se aplicaron simultáneamente a cada uno de los pequeños múltiplos. La vista animada proporcionó a los usuarios controles de animación que permitió seleccionar cualquiera de los meses presentados en el conjunto de datos. La animación interpola suavemente los datos entre los meses de esos años.

En ambas vistas, Flowstrates y Visualización Circular con Cuerdas, los participantes tuvieron la posibilidad de cambiar a la ”vista de diferencia” que mostró solo las diferencias entre lo seleccionado y el elemento a comparar. Los flujos que tuvieron un mayor número de personas en movimiento en comparación con el elemento a comparar se resaltaron. El ancho de estos flujos representó los valores absolutos de las diferencias. De esta manera, los participantes pudieron ver qué cambió exactamente en comparación con el mes anterior. Esto les facilitó la tarea de hacer y documentar los hallazgos relacionados con los cambios entre los meses subsiguientes. El conjunto de datos principal que utilizamos para el experimento representó los flujos de migración durante los 12 meses. Los datos de cada mes contenían unos 13 nodos y unos pocos cientos de flujos. Tuvimos que filtrar los flujos, mostrando solo unos pocos cientos de los más grandes, para garantizar que la animación

se desarrollara sin problemas. El conjunto de datos que utilizamos para la orientación era similar: representaba a los los flujos de taxis en NYC pero de algunos meses.

5.3. Datos Colectados

Durante el experimento recogimos los siguientes datos:

- Breves descripciones textuales de los hallazgos presentados por los sujetos.
- La importancia de los hallazgos según la calificación de los sujetos.
- Capturas de pantalla de las vistas tomadas automáticamente cuando se presentaron los resultados.
- Se registraron todas las acciones de los usuarios compatibles con las vistas, por ejemplo, resaltado, comparación, etc)
- Calificación de cuestionario
- Feedback de usuario y comentarios de mejora.

Tener datos tan abundantes nos ayudó durante el análisis. No solo nos permitió discernir varios aspectos del proceso de hacer hallazgos, sino que también fue útil para aclarar los significados de esos hallazgos, que no estaban debidamente formulados.

5.4. Análisis

El objetivo principal del análisis de los hallazgos recopilados fue averiguar si existían diferencias cualitativas entre los tipos de hallazgos realizados en la vista dinámica de Flowstrates y en la variación de Visualización con Cuerdas. Partimos del análisis de los hallazgos y los registros de interacción, desarrollando una teoría a partir de estos datos. Para lograr este objetivo, identificamos las categorías de codificación, realizamos la codificación manual de los hallazgos y analizamos cuidadosamente su distribución entre las categorías. En el resto de esta sección discutimos este proceso en detalle.

5.5. Codificación

Como los hallazgos se referían a los flujos de taxis entre ubicaciones geográficas que cambian con el tiempo, elegimos el “ámbito geográfico espacial” y el “ámbito temporal” como las principales propiedades de la codificación. Aquí están las definiciones de las propiedades que usamos:

- Ámbito temporal: el intervalo de tiempo al que se refiere el hallazgo.

- Alcance geográfico - El alcance de las entidades geográficas mencionadas en el hallazgo.
- Validez: si la declaración del hallazgo puede interpretarse como un hallazgo válido.

Primero, aplicamos un enfoque de arriba hacia abajo e intentamos identificar las categorías de codificación para cada una de las propiedades antes de la codificación. Luego, la codificación manual se realizó con las categorías predefinidas. Cada uno de los codificadores tuvo que revisar la lista completa de hallazgos y asignar las categorías de propiedades al elegir uno de los valores predefinidos que se enumeran para cada uno de ellos. Después de eso, calculamos las tasas de acuerdo entre anotadores para garantizar la confiabilidad de la codificación. Eran los siguientes:

Las categorías que finalmente utilizamos para el “alcance temporal” fueron las siguientes (el enfoque de abajo hacia arriba):

- Un mes: describe lo que estaba sucediendo en un mes específico (por ejemplo, “Flujo importante de Febrero en Manhattan”).
- Hasta o después: describe un patrón que fue aparente durante un período de tiempo antes o después de un mes específico (por ejemplo, “en Marzo comenzaron los movimientos desde o hacia Bronx”).
- Intervalo: describe lo que estaba sucediendo en un lapso de tiempo de varios años.
- Todo el tiempo: se aplica a todo el período de tiempo durante el cual los datos estuvieron disponibles (por ejemplo, “la migración implica cada vez más zonas a lo largo del tiempo”).

Y para el “ámbito geográfico”(el enfoque de arriba hacia abajo):

- Distrito Comunitarios: describe flujos que especifican solo el distrito en el que se originan o que tienen como destino, no ambos (por ejemplo, “Grandes flujos de Manhattan en Diciembre”).
- Distrito - Distrito: describe un flujo entre dos distritos específicos (por ejemplo, “Gran flujo de Manhattan a Bronx en Julio”).
- Burgo o Ciudad: describe los flujos que se originan en una ciudad específica o que tienen la otra ciudad como destino.
- Distrito - Ciudad - Describe los flujos entre un burgo y una ciudad.
- Ciudad - Ciudad - Describe flujos entre dos ciudades.
- Global: describe un patrón global (geográficamente) (por ejemplo, “cuando va lejos, los distritos cercanos al agua son destinos más populares”).

Los valores de las propiedades de **razonamiento y validez** eran simplemente **sí o no**. Al final, las categorías que creamos resultaron ser útiles para lograr nuestro objetivo: identificar las diferencias entre los tipos de hallazgos realizados en la vista animada.

5.6. Pruebas

Se categorizó a las sesenta y un personas en cuatro grupos distintos: un grupo compuesto por 15 personas Administrativas **A** o usuarios sin ninguna similitud con el entrevistador y con conocimiento mínimo de migraciones en su mayoría, otro grupo compuesto por 20 personas universitarias **B** las cuales son lejanas al entrevistador, otro grupo compuesto por 14 personas compañeros de ‘computación’ **C** personas del área de computación, estudiantes en su mayoría, y otro grupo de 12 personas familiares **D**.

Presentamos los resultados en tres tablas diferentes, cada una para una tarea diferente.

Evaluación Tarea 1

Para la Tarea 1: Consultas Espaciales, evaluamos a las personas en función a una serie de preguntas que se detallan a continuación:

- ¿De qué lugar o lugares las personas **emigran** más?
- ¿A qué lugar o lugares las personas **inmigran** más?

En función a las preguntas planteadas al usuario, se le toma el tiempo al usuario. Como lo mostramos en la Figura 5.1, Consulta Espacial (en minutos). Para evaluar un aspecto oculto de los tiempos tomados, se uso un recurso: la penalidad. Dicha penalidad es una medida que nos indica si el tiempo del usuario es mayor o menor de acuerdo al promedio general de su sector. Entonces esta penalidad sera positiva si el tiempo es menor al promedio, y negativa si el tiempo es mayor.

De las mediciones de la Tarea 1, podemos inferir que la técnica visualización *Circular con Cuerdas* demora menos que la técnica visualización Flowstrates.

Para evaluar la consulta espacial de la tarea 1, se uso un gráfico de barras, dicho gráfico se basa en el promedio de las 3 visualización por cada sector de personas (A,B,C y D). Así mismo se uso un gráfico radial para evaluar la tendencia de los sectores de las personas basadas en la penalidad (si el tiempo de una persona es mayor al promedio su penalidad sera negativa, si es el tiempo es menor al promedio de su sector, entonces la penalidad sera positiva). Dichos gráficos los observamos en la Figura 5.4.

Podemos inferir también que presentan un menor tiempo y un mejor desempeño las personas del grupo de C, debido a que se les es fácil usar una computadora, y este caso nuestra visualización. Un tiempo más alto lo presentan las personas de la categoría NORMAL, las cuales se demoraron más que las personas de la categoría A y B debido a que era la primera vez que experimentaban con las visualizaciones, en un lugar intermedio se presentan las personas de la categoría D, presentan un tiempo intermedio debido a que han estado en contacto directo con las técnicas.

CONSULTA ESPACIAL (Tiempo en minutos)								
Categoria	Persona	Flowstrates		Cuerdas		Cuerdas 2		
		Tiempo	Penalidad	Tiempo	Penalidad	Tiempo	Penalidad	
A	1	6,06	-1	4,12	1	7,34	-1	
	2	3,05	1	6,02	-1	3,58	1	
	3	3,02	1	6,04	-1	5,43	-1	
	4	6,06	-1	5,04	-1	7,02	-1	
	5	4,04	1	4,12	1	4,46	-1	
	6	4,07	1	4,12	1	4,08	1	
	7	6,01	-1	6,14	-1	3,14	1	
	8	5,01	-1	3,04	1	7,24	-1	
	9	6,06	-1	3,14	1	3,14	1	
	10	3,04	1	5,02	-1	2,1	1	
	11	3,03	1	5,02	-1	2,04	1	
	12	4,06	1	3,12	1	3,1	1	
	13	3,05	1	5,06	-1	2,04	1	
	14	6,03	-1	4,02	1	3,48	1	
	15	3,02	1	4,14	1	7,06	-1	
B	TOTAL	4,37	3	4,54	1	4,35	3	
	16	5,02	-1	6,13	-1	6,33	-1	
	17	4,02	1	4,27	1	3,03	1	
	18	3,08	1	3,33	1	3,30	1	
	19	5,14	-1	5,17	-1	3,10	1	
	20	3,08	1	5,03	-1	6,03	-1	
	21	3,02	1	4,23	1	2,30	1	
	22	5,02	-1	3,33	1	6,03	-1	
	23	5,02	-1	5,23	-1	6,03	-1	
	24	4,04	1	4,10	1	4,03	-1	
	25	6,06	-1	4,03	1	3,07	1	
	26	5,02	-1	4,13	1	4,20	1	
	27	4,12	1	4,20	1	6,27	-1	
	28	3,04	1	3,20	1	6,20	-1	
C	29	3,12	1	5,10	-1	5,10	-1	
	30	5,14	-1	5,33	-1	4,10	1	
	31	6,08	-1	5,33	-1	3,20	1	
	32	3,14	1	4,17	1	2,23	1	
	33	5,02	-1	6,30	-1	2,07	1	
	34	5,12	-1	6,20	-1	2,33	1	
	35	3,08	1	3,17	1	6,03	-1	
	TOTAL	4,32	0	4,60	2	4,25	2	
D	TOTAL	4,70	4	4,54	1	4,35	3	
	36	4,26	1	3,26	1	3,42	1	
	37	3,05	1	3,53	1	6,47	-1	
	38	5,21	-1	5,26	-1	6,21	-1	
	39	4,53	1	4,37	1	3,53	1	
	40	4,21	1	5,21	-1	5,42	-1	
	41	4,16	1	6,42	-1	5,37	-1	
	42	6,21	-1	5,05	-1	3,47	1	
	43	5,53	-1	3,47	1	2,05	1	
	44	4,42	1	5,21	-1	3,53	1	
	45	5,16	-1	3,37	1	3,05	1	
	46	4,47	1	4,21	1	3,42	1	
	47	4,26	1	6,53	-1	3,21	1	
	48	4,11	1	5,47	-1	6,11	-1	
	49	6,16	-1	3,37	1	3,53	1	
	TOTAL	4,70	4	4,62	0	4,20	4	
D	50	4,11	1	3,47	1	6,26	-1	
	51	6,26	-1	6,11	-1	3,16	1	
	52	6,37	-1	4,42	1	4,11	1	
	53	3,16	1	3,53	1	6,53	-1	
	54	5,47	1	3,53	1	2,32	1	
	55	5,37	1	5,11	-1	3,16	1	
	56	4,16	1	6,47	-1	3,05	1	
	57	6,53	-1	4,26	1	3,37	1	
	58	5,42	-1	3,21	1	3,42	1	
	59	6,42	-1	4,37	1	2,32	1	
	60	6,05	-1	5,21	-1	7,11	-1	
	61	5,42	-1	5,05	-1	6,53	-1	
	TOTAL	5,39	-2	4,56	2	4,28	4	

Figura 5.1: Resultados de la encuesta realizada a diferentes personas. Medimos el tiempo en minutos que demoró cada una de las personas realizando la Tarea 1: Consultas Espaciales. En la ultima fila del sector, se observa un campo TOTAL, el cual resume en sus columnas dos métricas (promedio y suma). Usamos el promedio para tener una medida intermedia de los tiempos, y suma para tener el conteo general de la penalidad. Se observa que las personas del sector C tienen un tiempo menor en la mayoría de las visualizaciones (Flowstrates con promedio de 4.70 y penalidad 4, Cuerdas con promedio 4.62 y penalidad de 0 y Cuerdas 2 con promedio 4.20 y penalidad de 4). Le siguen las personas del sector B, luego el sector D y finalizando con el sector A. De acuerdo a las medidas y promedios la Visualización con Cuerdas 2 es mejor que las anteriores.

CONSULTA TEMPORAL (Tiempo en minutos)								
Categoria	Persona	Flowstrates		Cuerdas		Cuerdas 2		
		Tiempo	Penalidad	Tiempo	Penalidad	Tiempo	Penalidad	
A	1	4,37	1	5,26	-1	4,05	1	
	2	4,32	1	5,11	-1	4,53	1	
	3	5,32	-1	4,37	1	3,16	1	
	4	6,32	-1	6,16	-1	5,05	-1	
	5	5,26	-1	4,47	1	5,75	-1	
	6	4,11	1	3,21	1	3,32	1	
	7	3,16	1	4,21	1	4,42	1	
	8	3,26	1	5,47	-1	4,05	1	
	9	3,47	1	4,42	1	5,16	-1	
	10	6,32	-1	5,16	-1	4,53	1	
	11	4,32	1	4,42	1	3,11	1	
	12	5,42	-1	4,42	1	5,37	-1	
	13	5,26	-1	5,42	-1	5,16	-1	
	14	4,53	1	5,21	-1	5,32	-1	
	15	4,42	1	3,21	1	6,42	-1	
	TOTAL	4,66	3	4,70	1	4,63	1	
B	16	4,37	1	4,16	1	4,21	1	
	17	4,16	1	6,05	-1	6,37	-1	
	18	5,42	-1	5,42	-1	4,21	1	
	19	6,21	-1	4,26	1	6,11	-1	
	20	6,53	-1	4,16	1	4,42	1	
	21	4,42	1	4,21	1	4,32	1	
	22	6,53	-1	6,21	-1	2,47	1	
	23	4,16	1	4,32	1	5,27	-1	
	24	6,53	-1	5,47	-1	5,47	-1	
	25	4,47	1	6,53	-1	5,37	-1	
	26	6,47	-1	5,47	-1	4,53	1	
	27	6,53	-1	6,11	-1	3,32	1	
	28	4,42	1	5,37	1	4,21	1	
	29	6,37	-1	6,26	-1	6,47	-1	
	30	6,53	-1	5,47	-1	6,32	-1	
	31	4,21	1	5,47	-1	4,05	1	
	32	4,37	1	4,16	1	6,53	-1	
	33	4,42	1	6,37	-1	4,21	1	
	34	4,37	1	4,21	1	4,47	1	
	35	6,42	-1	3,37	1	5,11	-1	
	TOTAL	5,34	0	5,15	-2	4,87	2	
CONSULTA TEMPORAL (Tiempo en minutos)								
C	Persona	Flowstrates		Cuerdas		Cuerdas 2		
		Tiempo	Penalidad	Tiempo	Penalidad	Tiempo	Penalidad	
	36	6,16	-1	3,47	1	3,02	1	
	37	4,26	1	4,05	1	6,11	-1	
	38	5,37	-1	3,53	1	5,17	-1	
	39	3,16	1	6,21	-1	6,37	-1	
	40	5,42	-1	6,05	-1	4,47	1	
	41	4,37	1	3,53	1	6,07	-1	
	42	5,11	-1	3,53	1	4,05	1	
	43	6,32	-1	3,05	1	5,21	-1	
	44	5,05	-1	6,32	-1	2,12	1	
	45	4,47	1	4,32	1	4,05	1	
	46	4,42	1	6,37	-1	6,42	-1	
	47	5,05	-1	6,37	-1	4,21	1	
	48	6,16	-1	5,37	-1	3,32	1	
	49	3,37	1	5,05	-1	3,16	1	
	TOTAL	4,91	-2	4,80	0	4,55	2	
CONSULTA TEMPORAL (Tiempo en minutos)								
D	Persona	Flowstrates		Cuerdas		Cuerdas 2		
		Tiempo	Penalidad	Tiempo	Penalidad	Tiempo	Penalidad	
	50	6,26	-1	3,32	1	4,19	1	
	51	5,47	1	6,47	-1	4,15	1	
	52	6,37	-1	3,32	1	4,37	1	
	53	4,26	1	3,32	1	4,26	1	
	54	6,32	-1	3,16	1	5,04	-1	
	55	4,42	1	3,21	1	4,30	1	
	56	6,16	-1	3,37	1	4,19	1	
	57	6,26	-1	6,05	-1	4,37	1	
	58	5,47	1	4,21	1	3,19	1	
	59	6,26	-1	6,53	-1	4,15	1	
	60	5,37	1	5,16	-1	7,30	-1	
	61	5,05	1	6,37	-1	6,26	-1	
	TOTAL	5,64	0	4,54	2	4,65	6	

Figura 5.2: Resultados de la encuesta realizada a diferentes personas. Medimos el tiempo en minutos que demoró cada una de las personas realizando la Tarea 2: Consultas Temporales. En la ultima fila del sector, se observa un campo TOTAL, el cual resume en sus columnas dos métricas (promedio y suma). Usamos el promedio para tener una medida intermedia de los tiempos, y suma para tener el conteo general de la penalidad. Se observa que las personas del sector C tienen un tiempo menor en la mayoría de las visualizaciones (Flowstrates con promedio de 4.91 y penalidad -2, Cuerdas con promedio 4.80 y penalidad de 0 y Cuerdas 2 con promedio 4.55 y penalidad de 2). Le siguen las personas del sector A, luego el sector D y finalizando con el sector B. De acuerdo a las medidas y promedios la Visualización con Cuerdas 2 es mejor que las anteriores.

		CRITERIO					
Categoria	Persona	Flowstrates		Cuerdas		Cuerdas 2	
		Int.	Con.	Int.	Con.	Int.	Con.
		Puntos	Puntos	Puntos	Puntos	Puntos	Puntos
A	1	5	4	4	3	8	5
	2	8	6	6	7	9	5
	3	7	7	10	4	8	7
	4	6	7	9	3	8	5
	5	6	8	9	2	5	4
	6	7	8	7	2	9	8
	7	7	5	8	10	8	8
	8	5	6	8	2	8	10
	9	4	6	7	8	3	5
	10	4	8	9	7	9	5
	11	7	7	4	2	6	7
	12	7	9	9	6	7	7
	13	7	7	4	7	8	5
	14	7	7	7	3	7	5
	15	2	5	7	4	7	5
Total		5,93	6,67	7,20	4,67	7,33	6,07
B	16	4	5	7	7	6	8
	17	7	4	4	7	5	7
	18	5	5	4	8	10	5
	19	3	5	7	3	6	8
	20	6	6	6	7	9	9
	21	5	4	7	8	5	7
	22	8	4	6	3	3	4
	23	9	9	5	10	9	8
	24	8	6	8	4	4	9
	25	3	10	8	10	4	4
	26	6	5	3	3	4	6
	27	7	9	8	5	8	7
	28	4	5	4	4	8	4
	29	7	10	4	9	7	4
	30	8	7	9	5	3	9
	31	7	4	6	4	9	5
	32	8	7	4	5	6	7
	33	6	5	4	10	8	4
	34	6	9	7	6	6	5
	35	6	10	9	5	9	8
TOTAL		6,15	6,45	6,00	6,15	6,45	6,40
C	36	6	5	3	3	8	4
	37	6	9	8	7	9	9
	38	5	5	4	3	6	6
	39	4	10	10	5	4	3
	40	5	8	6	5	8	4
	41	3	7	6	10	6	6
	42	9	7	6	4	5	3
	43	4	6	8	10	7	10
	44	9	8	7	7	7	10
	45	8	4	9	5	7	6
	46	8	5	10	3	4	9
	47	9	5	6	3	5	9
	48	9	8	5	8	10	4
	49	10	5	9	7	4	7
TOTAL		6,40	6,50	6,37	5,98	6,44	6,41
D	50	7	4	5	5	9	10
	51	4	6	5	9	9	6
	52	9	4	5	8	3	5
	53	10	9	5	4	6	5
	54	4	6	4	7	5	5
	55	3	7	7	4	6	6
	56	7	9	6	5	4	3
	57	9	7	9	7	8	7
	58	8	7	6	3	10	10
	59	5	9	4	10	3	6
	60	6	4	6	6	8	7
	61	3	4	9	7	9	9
TOTAL		6,25	6,33	5,92	6,25	6,67	6,58

Figura 5.3: Resultados de la encuesta realizada a diferentes personas. Medimos mediante el criterio de usuarios: La (Int.) interactividad y el (Con.) conocimiento que lograron obtener de acuerdo a las visualización de las técnicas presentadas. En la ultima fila del sector, se observa un campo TOTAL, el cual resume en sus columnas una métrica (promedio). Usamos el promedio para tener una medida intermedia de los puntos. Por ejemplo se observa que las personas del sector C (6,40 puntos) en cuanto al criterio de Interactividad y en cuanto a la visualización Flowstrates evalúan con mas alto puntaje. Seguida del sector D (6,25 puntos), luego B (6,15 puntos)y finalizando con A (5,93 puntos).

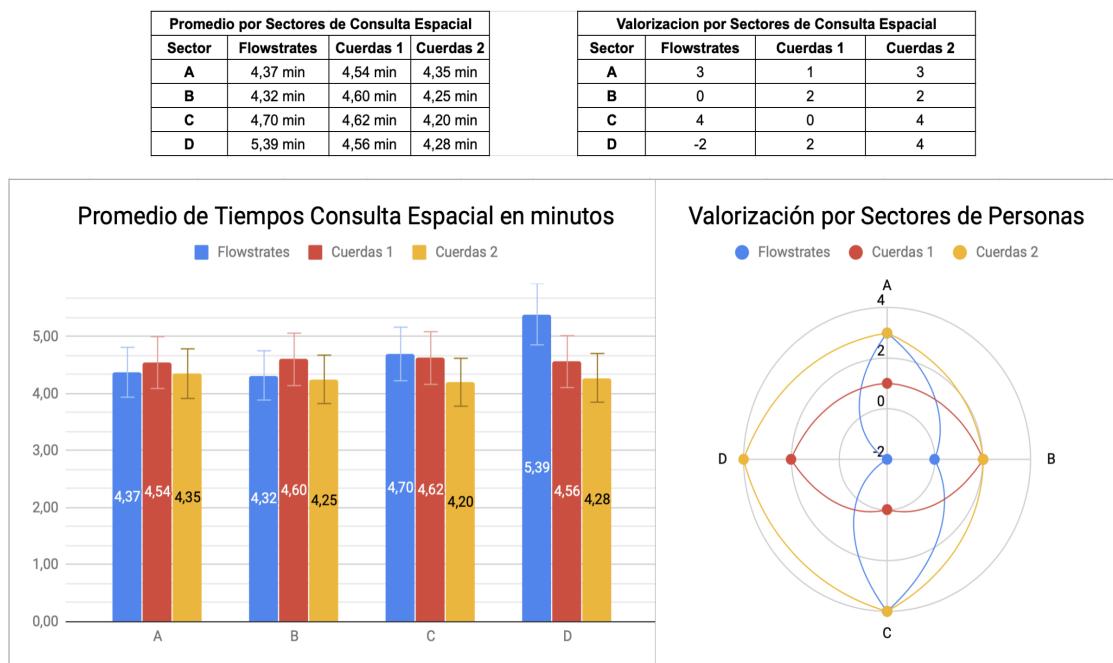


Figura 5.4: En esta figura observamos en la parte superior, dos tablas, la primera tabla de la izquierda, nos resume los promedios de la consulta espacial medida en tiempo, la otra tabla nos muestra la valoración basada en la penalidad para la consulta espacial. En la parte inferior a la izquierda se observa un gráfico de barras, para ver con mas claridad que visualización es mejor por sectores. En la parte derecha se observa un gráfico radial, el cual mide tendencias, orientado en esta caso por tendencias por sectores hacia cada visualización. En resumen Visualización Circular con Cuerdas es mayormente aceptada por todos los sectores y tiene un mejor desempeño en tiempo (menor numero de minutos es igual a que el usuario realizo la actividad de manera rápida).

Evaluación Tarea 2

Para la Tarea 2: Consultas Temporales, evaluamos a las personas en función a una pregunta que se detalla a continuación:

- ¿En qué mes o meses es mayor el número de flujos de migraciones?

En función a la pregunta planteada al usuario, se le toma el tiempo al usuario. Como lo plasmamos en la Figura 5.2, Consulta Temporal (en minutos).

De las mediciones de la Tarea 2, podemos inferir que la técnica Visualización *Circular con Cuerdas*. demora menos que la técnica Flowstrates.

Para evaluar la consulta temporal de la tarea 2, se uso un gráfico de barras, dicho gráfico se basa en el promedio de las 3 visualización por cada sector de personas (A,B,C y D). Así mismo se uso un gráfico radial para evaluar la tendencia de los sectores de las personas basadas en la penalidad (si el tiempo de una persona es mayor al promedio su penalidad sera negativa, si es el tiempo es menor al promedio de su sector, entonces la penalidad sera positiva). Dichos gráficos los observamos en la Figura 5.5.

De la Figura 5.2 inferimos que las personas que tuvieron menos tiempo y mejor desempeño fueron las personas de C, seguidos de las personas de A, y seguidas por las personas D y B, deducimos las mismas razones que la Tarea 1, la experiencia y el contacto directo con las técnicas.

Evaluación Tarea 3

Para la Tarea 3: Criterio Usuario, se le pregunta a la persona sobre dos criterios: la interactividad y el conocimiento que obtuvo de la visualización. En la parte de interactividad el usuario evalúa en función a su experiencia con la visualización, cuán útil fue para él y si presento problemas al momento de la interacción. En el conocimiento obtenido, el usuario comenta las conclusiones que obtuvo de las visualizaciones y la interactividad con las tareas anteriores. Como lo mostramos en la Figura 5.3.

De las mediciones de la Tarea 3, podemos inferir que la técnica visualización *Circular con Cuerdas* en el criterio de interactividad es mejor en puntos que la técnica visualización Circular con Cuerdas. Podemos inferir también que en el criterio de conocimiento Flowstrates es mejor que Circular con Cuerdas.

En cuanto a las mejores puntuaciones las personas de C tienen cierta tendencia a evaluar de una mejor manera, debido al conocimiento que tienen de cómo ‘implementar la técnica’, un menor puntuación la dan las personas de A, debido a que son más críticas y exigentes al momento de presentarles un demo funcional, en lugar intermedio se ubican las personas de categoría D y B.

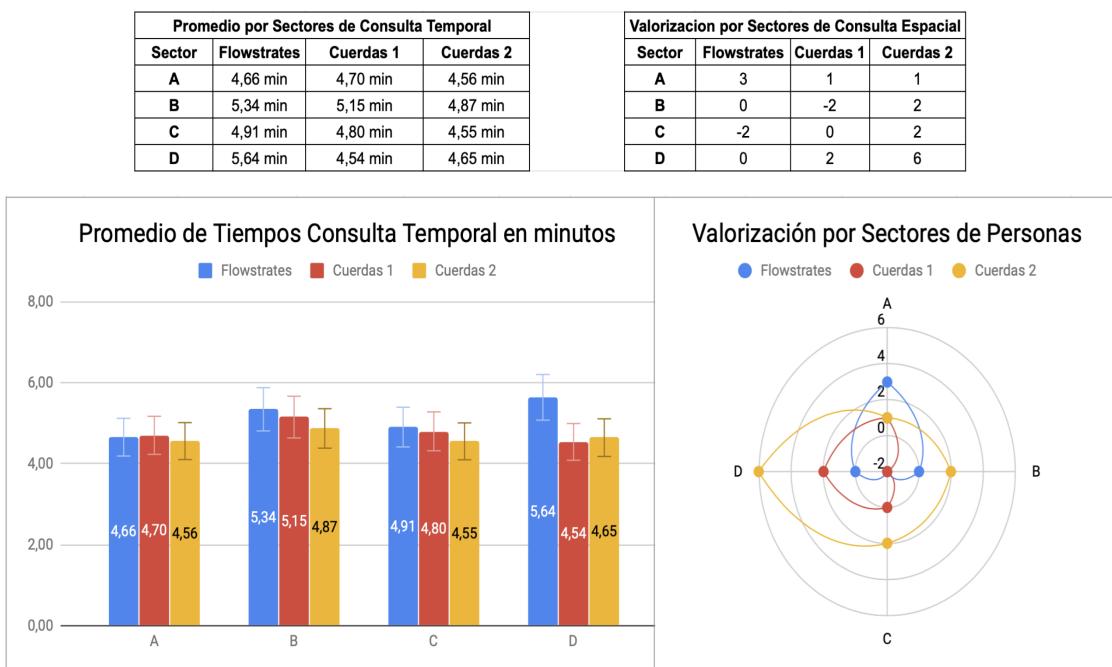


Figura 5.5: En esta figura observamos en la parte superior, dos tablas, la primera tabla la de la izquierda, nos resume los promedios de la consulta temporal medida en tiempo, la otra tabla nos muestra la valoración basada en la penalidad para la consulta temporal. En la parte inferior a la izquierda se observa un gráfico de barras, para ver con mas claridad que visualización es mejor por sectores. En la parte derecha se observa un gráfico radial, el cual mide tendencias, orientado en este caso por tendencias por sectores hacia cada visualización. En resumen Visualización Circular con Cuerdas es mayormente aceptada por todos los sectores y tiene un mejor desempeño en tiempo (menor numero de minutos es igual a que el usuario realizo la actividad de manera rápida).

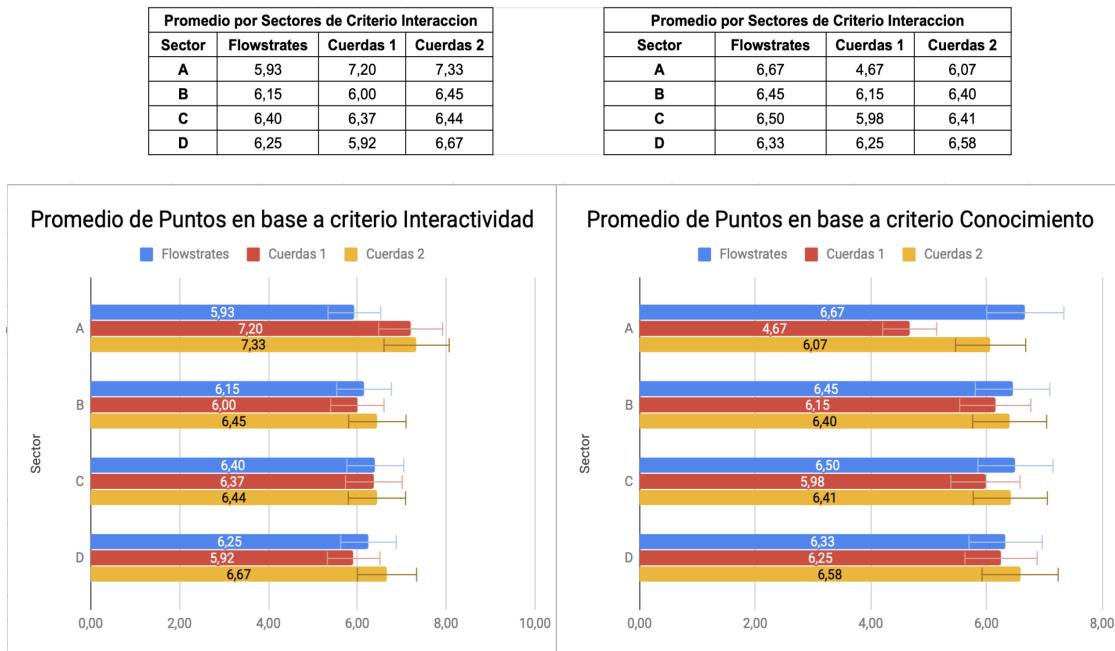


Figura 5.6: En esta figura observamos en la parte superior, dos tablas, la primera tabla la de la izquierda, nos resume los promedios del criterio basado en la interactividad medida en puntos del 1 al 10, la otra tabla nos muestra el criterio basado en conocimiento con puntos del 1 al 10. En la parte inferior a la izquierda se observa dos gráficos de barras, para ver con mas claridad que visualización es mejor por sectores. En resumen Visualización Circular con Cuerdas es mayormente aceptada por todos los sectores y tiene un mejor desempeño en tiempo (mayor numero de puntos es igual a que el usuario valoró la visualización de mejor manera).

5.6.1. Análisis Cualitativo

En esta etapa, describimos e identificamos las limitaciones, mejoras, ventajas y desventajas de cada una de las técnicas desarrolladas.

Limitaciones y ventajas de FlowStrates

La visualización *FlowStrates* se enfoca en representar información temporal y espacial del *dataset* de taxis de NYC. *Flowstrates* combina la presencia de 2 mapas de flujos y un mapa de calor en el centro el cual refleja el componente temporal. Otra ventaja es la posibilidad de hacer consultas directamente señalando el lugar geográfico y también es posible señalar el mapa de calor y analizar el componente temporal y ver que sucedió entre el origen y destino. Esta visualización sin embargo no provee una información cuantitativa adecuada de la densidad del flujo de un lugar a otro, como si lo hace la visualización circular con cuerdas. Otra información que no provee la visualización es la demarcación de cada distrito comunitario en su 2 mapas de orígenes y Destinos.

Mejoras, limitaciones y ventajas de Visualización circular con cuerdas

La visualización Circular con cuerdas, en un principio solo presentaba el número de viajes realizados entre un destino y otro, sin embargo para hacer posible la comparación con *Flowstrates*, se realizó modificaciones como ‘agregarle un componente temporal’ mostrado en los 12 gráficos X e Y, y el gráfico de temporalidad de los 13 distritos, a la visualización también se le añadió un gráfico de barras que nos muestra los flujos de destino que llegaron a un determinado distrito comunitario de los 13. En un inicio esta Visualización tenía una desventaja con respecto a *Flowstrates*, pero con el incremento de elementos visuales, se mejora el aspecto temporal de la visualización.

Esta visualización, a diferencia de *Flowstrates*, no provee un mapa para indicarle y hacerle más familiar al usuario de que lugar se está señalando, el usuario tiene que tener un conocimiento previo de los lugares y sitios más importantes en el mapa. Otro inconveniente en el momento de la implementación, fue el aspecto aglomerativo de la visualización, a una mayor cantidad de datos, el diagrama circular de cuerdas, se satura mucho, y hace imposible leer algunos lugares. Una posible solución que se pensó a ese problema, fue el incremento del tamaño del diagrama, en cierto punto lo soluciona. La solución por la que se optó fue la de utilizar la información del burgo *Manhattan*, y sus 13 distritos comunitarios para que la visualización no se sature y sea de fácil identificación.

Capítulo 6

Conclusiones y Trabajos Futuros

A continuación detallamos las conclusiones de nuestro trabajo, conclusiones a las que hemos llegado gracias al desarrollo de nuestra implementación y su posterior fase de evaluación.

- Presentamos diversas técnicas para la visualización de datos OD.
- Detallamos las similitudes y categorizamos las distintas técnicas en tres: visualización de flujos de mapas, matrices de OD, y visualización circular con cuerdas.
- La visualización circular con cuerdas nos muestra la relación que existen entre OD, y la cantidad de flujos, así como el aspecto temporal añadido en la etapa de la implementación. La visualización de *FlowStrates* con mapa de calor presenta el componente espacial y temporal.
- Analizando nuestros resultados, podemos concluir que la visualización *Circular con Cuerdas* tiene un mejor desempeño en cuanto a la realización de tareas. En las tareas 1 y 2, los usuarios mostraron un mejor reconocimiento del aspecto espacial y temporal que se reflejó en el menor tiempo en el uso de la Visualización *Circular con Cuerdas*. En la tarea 3, se le hizo una evaluación a criterio del usuario, de cuán interactiva es la visualización y si esta le aportaba conocimiento a la hora de generar conclusiones, esta tarea se midió en puntos, siendo la mejor puntuada la visualización *Circular con Cuerdas*.
- Cada una de estas técnicas presentan muchas ventajas como la simplicidad, la fácil interacción con el usuario y la habilidad de comunicar con claridad el flujo de movimiento. También poseen desventajas una respecto a otra, sin embargo estas técnicas podrían ser complementarias.

6.1. Problemas encontrados

Detallaremos cuales fueron los problemas encontramos a lo largo de nuestro trabajo:

- Es necesario realizar una etapa de procesamiento y limpieza de datos, esta etapa es obligatoria, porque hay algunos registros de datos OD, que son vacíos, o en su defecto nulos.
- Algunas visualizaciones como las circulares con cuerdas (*chord*) o los *treeMaps* presentan dificultades al momento de mostrar las visualizaciones y los datos presentes en el *dataset*, esto debido a que los gráficos con los que cuentan, no soportan demasiados datos espaciales y temporales.
- Una vez hecho el pre-procesamiento de datos, es necesaria una etapa de “adaptación del *dataset*”, en la cual tendremos que adaptar el *dataset* preprocesado para los tipos de visualizaciones que presentamos, una adaptación para la visualización ‘*Flowstrates*’ y otra para la visualización de ‘*Chord*’ o circulares con cuerdas.

6.2. Recomendaciones

A lo largo del desarrollo de esta tesis, nos ha permitido adquirir conocimientos de los cuales podemos inferir y recomendar lo siguientes:

- En las visualizaciones OD, es necesario tomar en cuenta los aspectos temporales y espaciales de un *dataset*.
- Una visualización adecuada y eficiente para el usuario, es aquella que le permite al usuario interactuar en tiempo real, tanto en sus dimensiones espaciales como temporales.
- La combinación de los aspectos temporales y espaciales de la visualización, le otorgan la posibilidad al usuario de poder hacer *queries*, generar un conocimiento , e inferir datos y opiniones de la visualización.

6.3. Trabajos futuros

- En un futuro cercano, se planea recolectar más retroalimentación del usuario y ejecutar una evaluación del usuario con diversas tareas.
- Otra mejora que se pretende realizar es la de adaptar el *dataset* de los transportes y migraciones de Perú, para la identificaciones de fenómenos migratorios.
- Planeamos también la identificación de tareas claves abocadas a nuestro contexto con el uso de un *dataset* de Perú.
- En la técnica de visualización circular con cuerdas, se planea adaptar los gráficos y reducir su complejidad, usando menos elementos gráficos implementando *forms*, *buttons*, *selectors* y *vistas*.
- En la técnica de *flowstrates* se pretende mejorar la adaptación del ‘*shapefile*’, para permitir una mejor demarcación de los lugares geográficos.

Bibliografía

- [Andrienko and Andrienko, 2008] Andrienko, G. and Andrienko, N. (2008). Spatio-temporal aggregation for visual analysis of movements. pages 13–25.
- [Andrienko et al., 2010] Andrienko, G., N., D. Dransch, J. D., Fabrikant, S. I., Jern, M., Kraak, M., Schumann, H., and Tominski, C. (2010). Space, time and visual analytics. international journal of geographical information science. page 1577–1600.
- [Boyandin et al., 2011] Boyandin, I., Bertini, E., Bak, P., , and Lalanne, D. (2011). Flows-trates: An approach for visual exploration of temporal origin-destination data. page 971–980.
- [Chen and Guo, 2015] Chen, W. and Guo, F. (2015). A survey of traffic data visualization”. in: Ieee transactions on intelligent transportation systems. page 2970–2984.
- [Dickens, 2009] Dickens, M. (2009). The benefits of public transportation: Relieving tracking congestion. pages 13–25.
- [Gu et al., 2014] Gu, Z., Gu, L., and Eil, R. (2014). Flowstrates: An approach for visual exploration of temporal origin-destination data. page 45–63.
- [Guo and Zhu, 2014] Guo, D. and Zhu, X. (2014). Origin-destination flow data smoothing and mapping. pages 243–252.
- [Heer et al., 2010] Heer, J., Bostock, M., and Ogievetsky, V. (2010). *A tour through the visualization zoo*. Association for Computing Machinery.
- [Jenny et al., 2016] Jenny, B., Stephen, D. M., Muehlenhaus, I., Marston, B. E., Sharma, R., and Zhang, E. (2016). Design principles for origin-destination flow maps. pages 62–75.
- [Limousine Commission NYC, 2015] Limousine Commission NYC, C. (2015). Nyc traffic advisory aug 2-4. pages 86–89.
- [Liu, 2017] Liu, Y. (2017). Visualization of multivariate data. page 2970–2984.
- [Rae, 2009] Rae, A. (2009). From spatial interaction data to spatial interaction information, geovisualisation and spatial structures of migration from the 2001 uk census. page 161–178.
- [Sander et al., 2014] Sander, N., Abel, G., Bauer, R., and Schmidt, J. (2014). Visualising migration flow data with circular plots.

- [Ward et al., 2010] Ward, M., Grinstein, G., and Keim, D. (2010). Interactive data visualization: Foundations, techniques, and applications. page 1–5.
- [Whong, 2017] Whong, C. (2017). Foiling nyc’s taxi trip data.
- [Wilkinson and Friendly, 2008] Wilkinson, L. and Friendly, M. (2008). The history of the cluster heat map. page 179–184.
- [Wood et al., 2010] Wood, J., Dykes, J., and Slingsby, A. (2010). Visualisation of origins, destinations and flows with od maps. page 117–129.
- [Yang et al., 2017] Yang, Y., Dwyer, T., Goodwin, S., and Marriott, K. (2017). Many-to-many geographically-embedded flow visualisation: An evaluation. pages 411–420.
- [Yassky and NYC, 2016] Yassky, D. and NYC, T. . L. C. (2016). Hail market analysis. pages 123–134.
- [Yuan et al., 2010] Yuan, J., Zheng, Y., Zhang, C., Xie, G., Sun, G., and Huan, Y. (2010). T-drive: driving directions based on taxi trajectories. springer. page 99–108.
- [Yue, 2009] Yue, Y. (2009). Mining time-dependent attractive areas and movement patterns from taxi trajectory data. page 1–6.
- [Zeng, 2013] Zeng, W. (2013). Visualizing interchange patterns in massive movement data. page 271– 280.
- [Zhao and Forer, 2008] Zhao, P. and Forer, P. (2008). Activities, ringmaps and geovisualization of large human movement fields. information visualization. springer. page 198–209.
- [Zhou and Zheng, 2011] Zhou, X. and Zheng, Y. (2011). Computing with spatial trajectories. springer. page 271– 280.