

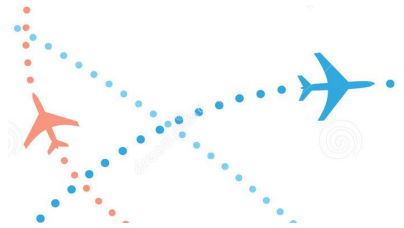
Análisis Visual de Datos Origen Destino (OD)

Jesús Erick Vera Callme

**Facultad de Ingeniería y Computación
Universidad Católica San Pablo**

- **Los flujos de movimiento:**

- Se definen como los desplazamientos de:
 - Grupos humanos
 - Mercancías y
 - Servicios como:



- **Los datos Origen-Destino (OD) poseen:**

- Un punto inicial (**O**) y van hacia un punto final (**D**)
- Ocurre en un periodo de tiempo determinado.

Problema

```
pickup_time,dropoff_time,pickup_long,pickup_lat,dropoff_long,dropoff_lat,id_taxi,distance,fare_amount,surcharge,mta_tax,tip_amount,tolls_amount,payment_type,pas  
sengers,field1,field2,field3,field4  
2013-01-01 15:11:48,2013-01-01 15:18:10,-73.978165,40.757977,-73.989838,40.751171,89D227B655E5C82AECF13C3F540D4CF4,1.00,6.5,0.0,0.5,0.0,CSH,4,0,0,0,0  
2013-01-06 00:18:35,2013-01-06 00:22:54,-74.006683,40.731781,-73.994499,40.75066,0BD7C8F5BA12B88E0B67BED28BEA73D8,1.50,6.0,0.5,0.0,0.5,0.0,CSH,1,0,0,0,0  
2013-01-05 18:49:41,2013-01-05 18:54:23,-74.004707,40.73777,-74.009834,40.726002,0BD7C8F5BA12B88E0B67BED28BEA73D8,1.10,5.5,1.0,0.5,0.0,CSH,1,0,0,0,0  
2013-01-07 23:54:15,2013-01-07 23:58:20,-73.974602,40.759945,-73.984734,40.759388,DFD2202EE08F7A8DC9A57B02ACB81FE2,2.70,5.0,0.5,0.0,0.5,0.0,CSH,2,0,0,0,0  
2013-01-07 23:25:03,2013-01-07 23:34:24,-73.97625,40.748528,-74.002586,40.747868,DFD2202EE08F7A8DC9A57B02ACB81FE2,2.10,9.5,0.5,0.0,0.5,0.0,CSH,1,0,0,0,0  
2013-01-07 15:27:48,2013-01-07 15:38:37,-73.966743,40.764252,-73.983322,40.743763,20D9ECB2CA0767CF7A01564DF2844A3E,1.70,9.5,0.0,0.5,0.0,CSH,1,0,0,0,0  
2013-01-08 11:01:15,2013-01-08 11:08:14,-73.995804,40.743977,-74.007416,40.744343,496644932DF3932605C22C7926FF0FE0,.80,6.0,0.5,0.0,0.5,0.0,CSH,1,0,0,0,0  
2013-01-07 12:39:18,2013-01-07 13:10:56,-73.989937,40.756775,-73.86525,40.77063,0B57B9633A2FEC3D3B1944AFC7471CF,10.70,34.0,0.0,0.5,0.0,4.8,CSH,3,0,0,0,0  
2013-01-07 18:15:47,2013-01-07 18:20:47,-73.980072,40.743137,-73.982712,40.735336,2C0E91FF20A856C891483ED63589F982,.80,5.5,1.0,0.5,0.0,CSH,1,0,0,0,0  
2013-01-07 15:33:28,2013-01-07 15:49:26,-73.977936,40.786983,-73.952919,40.80637,2D4B95E2FA7B2E85118EC5CA4570FA58,2.50,13.0,0.0,0.5,0.0,0.5,0.0,CSH,2,0,0,0,0  
2013-01-08 13:11:52,2013-01-08 13:19:50,-73.982452,40.773167,-73.964134,40.773815,E12F6AF991172EAC3553144A0AF75A19,1.30,7.5,0.0,0.5,0.0,CSH,1,0,0,0,0  
2013-01-08 09:50:05,2013-01-08 10:02:54,-73.99556,40.749294,-73.988686,40.759052,E12F6AF991172EAC3553144A0AF75A19,.70,9.0,0.0,0.5,0.0,CSH,1,0,0,0,0  
2013-01-10 12:07:08,2013-01-10 12:17:29,-73.971497,40.791321,-73.964478,40.775921,78FFD9CD0CA541F335EF8B38FB494D6,2.30,9.5,0.0,0.5,0.0,CSH,1,0,0,0,0
```

	id	vendor_id	pickup_datetime	dropoff_datetime	passenger_count	pickup_longitude	pickup_latitude	dropoff_long
0	id2875421	2	2016-03-14 17:24:55	2016-03-14 17:32:30	1	-73.982155	40.767937	-73.964630
1	id2377394	1	2016-06-12 00:43:35	2016-06-12 00:54:38	1	-73.980415	40.738564	-73.999481
2	id3858529	2	2016-01-19 11:35:24	2016-01-19 12:10:48	1	-73.979027	40.763939	-74.005333
3	id3504673	2	2016-04-06 19:32:31	2016-04-06 19:39:40	1	-74.010040	40.719971	-74.012268
4	id2181028	2	2016-03-26 13:30:55	2016-03-26 13:38:10	1	-73.973053	40.793209	-73.972923

Solo Datos !! , pero qué nos podrían decir?

Figura 1. Dataset de viajes en taxi en la ciudad de Nueva York año 2016 [7]

- NYC Taxis and Limousine Commision 2018, [Online] <https://www1.nyc.gov/site/tlc/index.page>

- ¿Por qué **visualizar/interpretar** estos datos OD?
- ¿Existe algún **fenómeno/comportamiento** en los datos OD?
- ¿En qué **lugares/horas** se presentan un mayor flujo de movimiento?
- ¿Porque en ciertos **lugares/meses** existe mayor cantidad de flujos?

Posible Solución

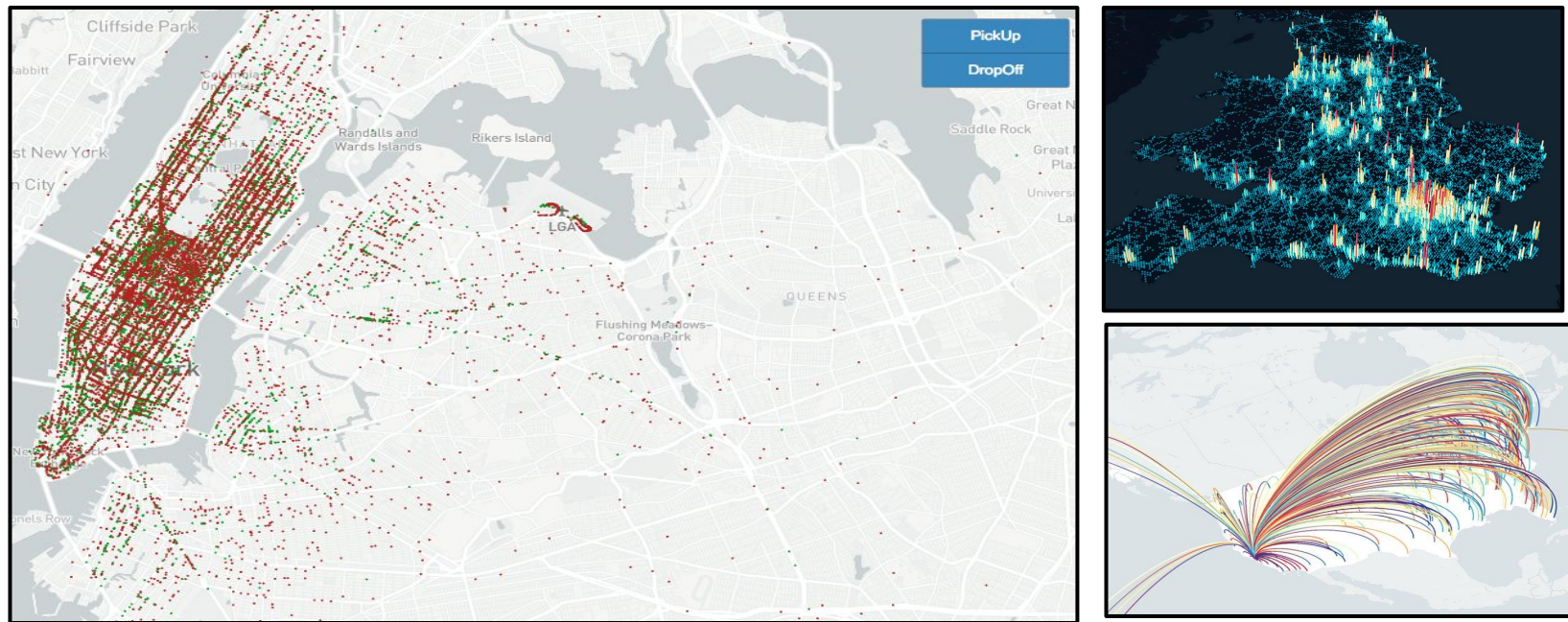
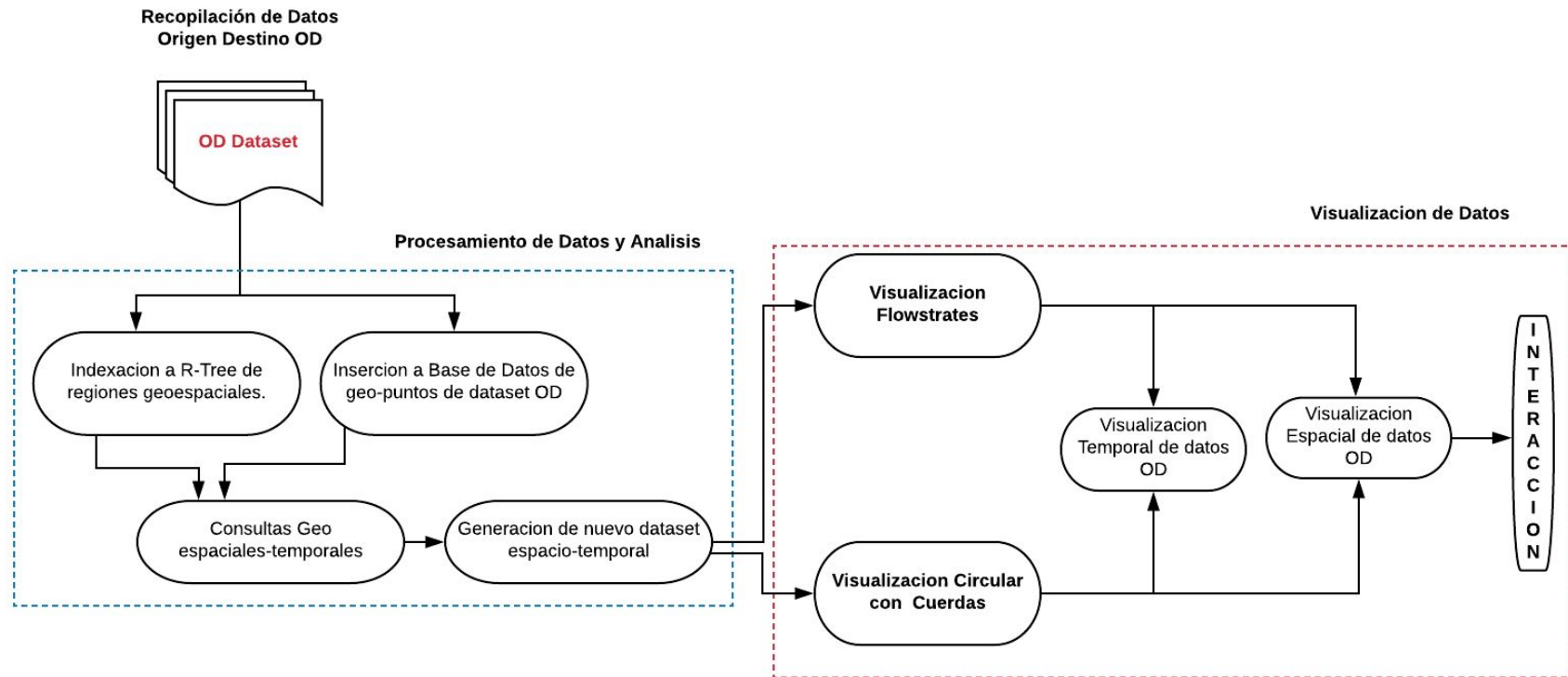


Figura 2. Ploteo del dataset de taxis de New York, usando solo coordenadas geográficas.

- Map-App Demo, [Online] <https://jesusveca.github.io/map-app/#>, B) deck.gl Uber, [Online] uber.github.io/deck.gl/, C deck.gl Uber, [Online] uber.github.io/deck.gl/
- Uber Visualization team, VIS.GL, [Online] <https://vis.gl>, 2017

Propuesta

Arquitectura



Arquitectura

- **Recopilación de datos OD**
 - **Recolección de dataset NYC taxis**
- **Procesamiento de Datos y análisis**
 - **Indexación a R-Tree**
 - **Inserción a Base de datos**
 - **Consultas geo espaciales-temporales**
 - **Generación de nuevo dataset espacio-temporal.**
- **Visualización de datos OD**
 - **Visualizacion Flowstrates**
 - **Visualización Circular con cuerdas**
 - **Interacción**

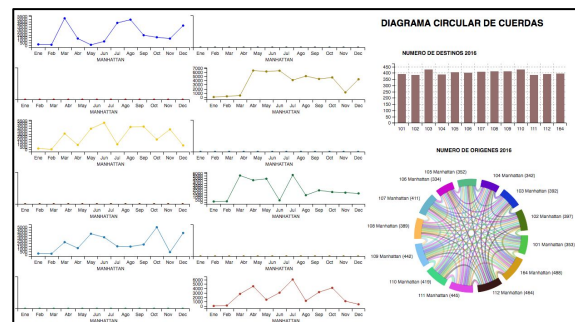
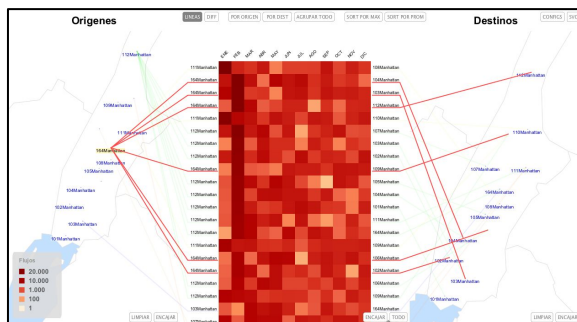
Implementación

● Tratamiento del Dataset

Code	Name	Lat	Lon	region1	region2
101	101Manhattan	40.708702	-74.010391	Americas	Northern America
102	102Manhattan	40.731781	-74.006683	Americas	Northern America
103	103Manhattan	40.719303	-73.994026	Americas	Northern America
104	104Manhattan	40.743977	-73.995804	Americas	Northern America
105	105Manhattan	40.757977	-73.978165	Americas	Northern America
106	106Manhattan	40.748528	-73.97625	Americas	Northern America
107	107Manhattan	40.786983	-73.977936	Americas	Northern America
108	108Manhattan	40.764252	-73.966743	Americas	Northern America
109	109Manhattan	40.806396	-73.959908	Americas	Northern America
110	110Manhattan	40.809158	-73.944519	Americas	Northern America
111	111Manhattan	40.786156	-73.945915	Americas	Northern America
112	112Manhattan	40.843525	-73.941628	Americas	Northern America
164	164Manhattan	40.773125	-73.96666	Americas	Northern America

Origin	Dest	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC
101	102	175	108	5125	1287	94	738	4253	4870	1919	1529	1278	3751
101	103	12	222	1684	5749	359	99	3634	1033	3898	3501	1159	3169
101	104	210	67	3035	884	3968	5087	1015	4294	4364	1918	3838	796
101	105	105	9	532	4805	1440	5742	6196	3949	4274	2888	2975	2996
101	106	152	117	2223	1141	3757	3116	1469	1422	1799	4947	413	3900
101	107	212	176	4899	2024	6521	1503	3955	3044	2088	3643	1448	1796
101	108	228	125	1663	5346	841	4045	2230	1297	3933	186	3040	3731
101	109	26	195	392	6371	6146	6337	4082	5045	4350	4726	1138	4282
101	110	237	153	1040	1621	2986	4021	2977	2601	1172	789	1230	1643
101	111	109	149	5537	4562	4861	360	5645	1422	2438	2091	1954	1801
101	112	57	133	5586	6562	4254	1039	2039	4015	1920	1908	2497	4211
101	164	49	146	2761	4487	1459	3035	6019	1206	3165	4127	1131	423

● Visualización de Datos OD



- NYC Taxis and Limousine Commission 2018, [Online] <https://www1.nyc.gov/site/tlc/index.page>

Visualización Circular de Cuerdas

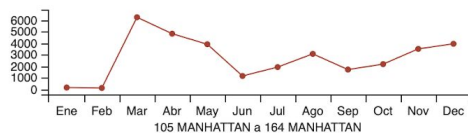
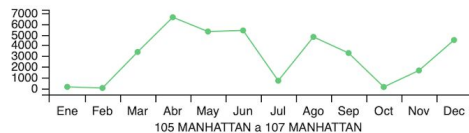
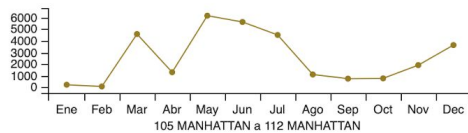
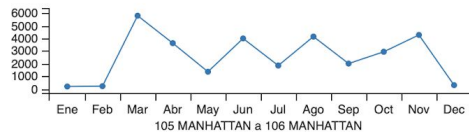
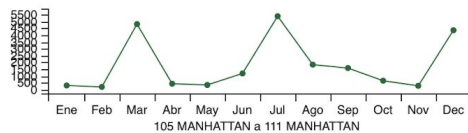
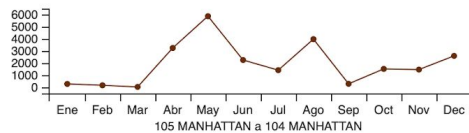
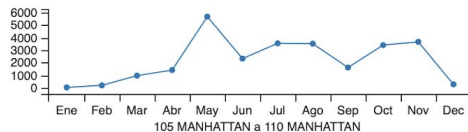
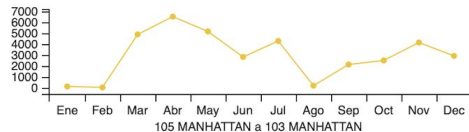
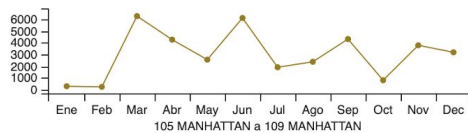
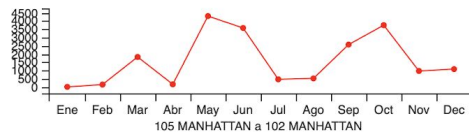
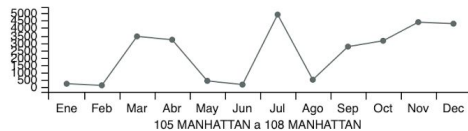
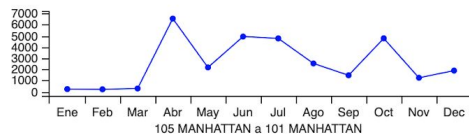
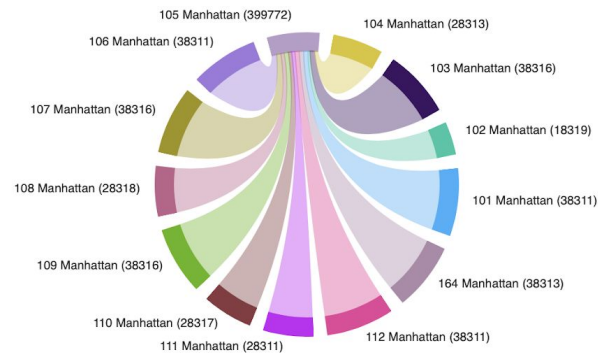


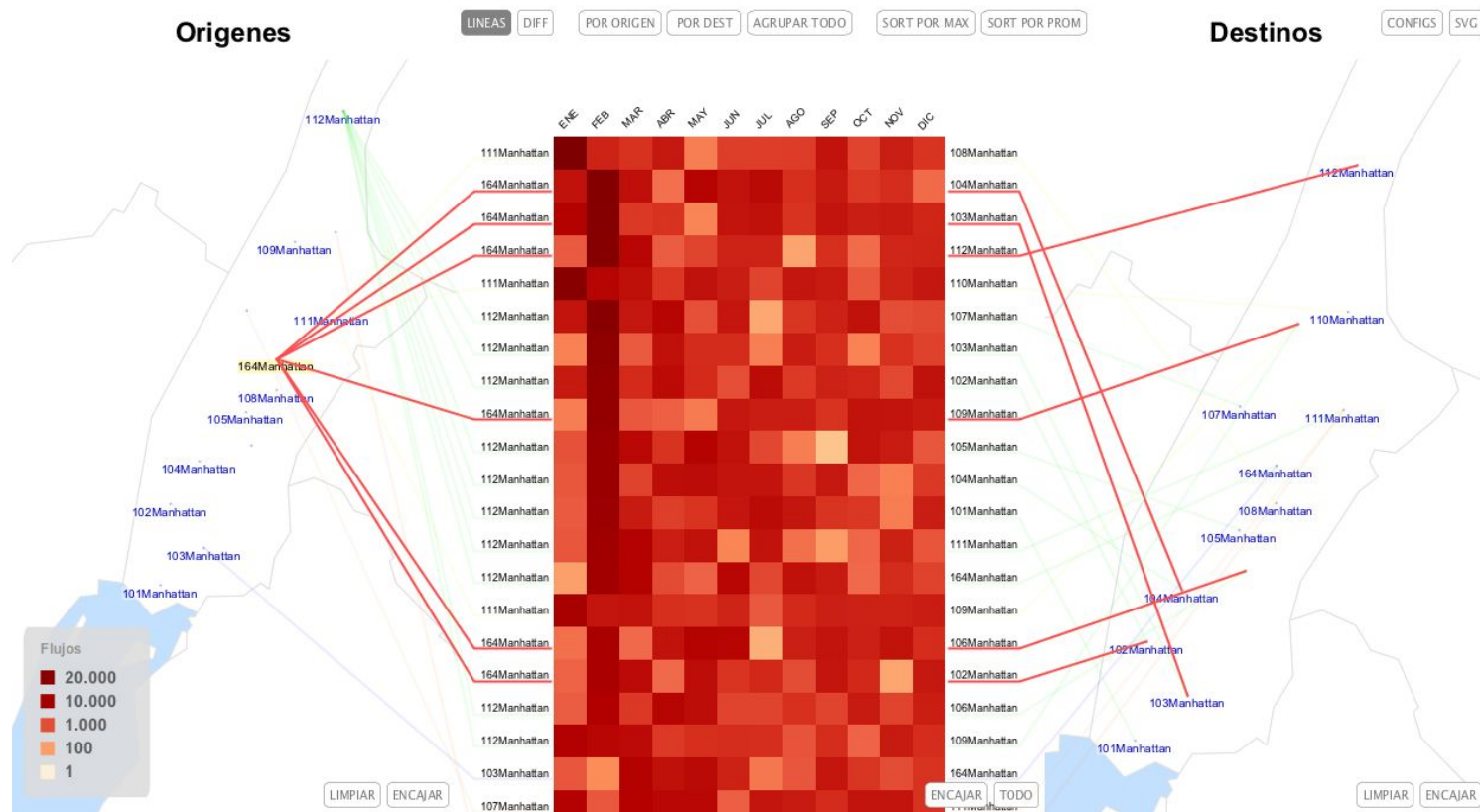
DIAGRAMA CIRCULAR DE CUERDAS





Comparación

Visualización Flowstrates



- Boyandin, Ilya, et al. "Flowstrates: An approach for visual exploration of temporal origin-destination data." *Computer Graphics Forum*. Vol. 30. No. 3. Oxford, UK: Blackwell Publishing Ltd, 2011.

Visualización Circular de Cuerdas

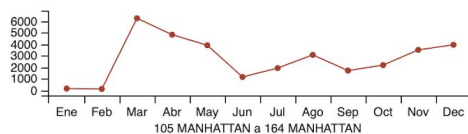
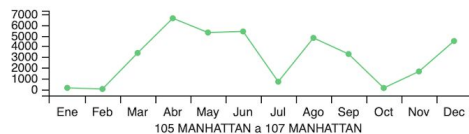
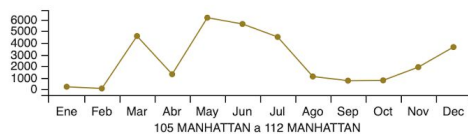
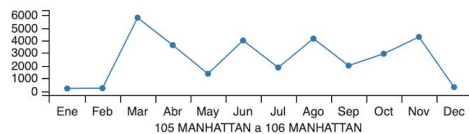
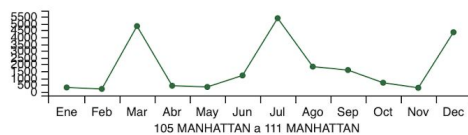
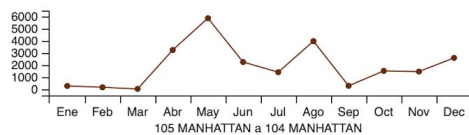
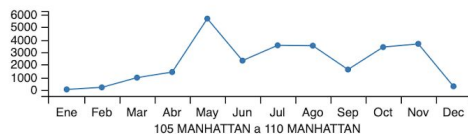
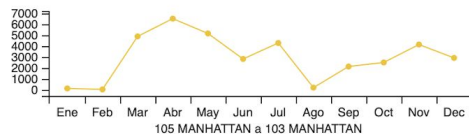
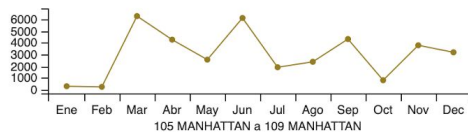
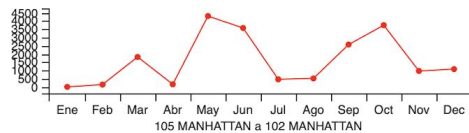
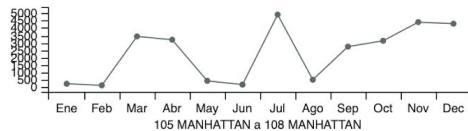
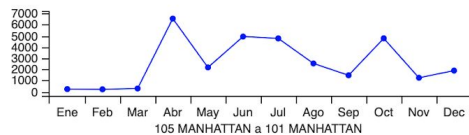
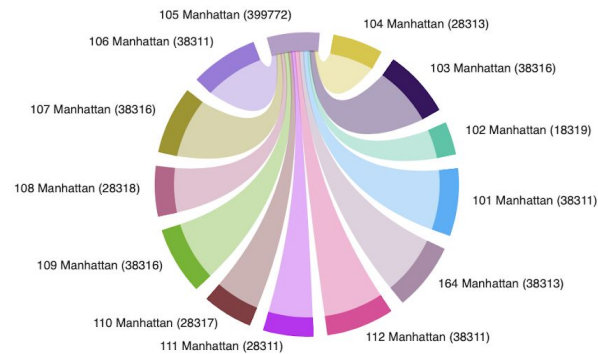


DIAGRAMA CIRCULAR DE CUERDAS



Experimentación y Resultados

Diseño de Entrevista

- En cuanto a la definición de tareas, son las siguientes:
 - Tarea 1: Consultas relacionadas a la **espacialidad**
 - Identificar “espacialmente” qué zonas presentan una $>$ o $<$ cantidad de flujos.
 - Tarea 2: Consultas relacionadas a la **temporalidad**
 - Identificar “temporalmente” que meses presentan una $>$ o $<$ cantidad de flujos.
 - Tarea 3: Consultas relacionadas al **criterio del usuario**
 - Interactividad y el conocimiento obtenida de las visualización y su interacción.
 - La interactividad se da por:
 - **Usabilidad y**
 - **Experiencia del usuario**

Esquema 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 propósito de las tareas**
- 5. Planteamiento de Tareas 1 y 2**
- 6. Explicación detallada de Tarea 3**
- 7. Toma de Tiempo y evaluación a criterio**

Diseño de Experimentos

- **Tarea 1:**
 - Consultas Espaciales
 - ¿De qué lugar o lugares las personas migran más?
- **Tarea 2:**
 - Consultas Temporales
 - ¿En qué mes o meses es mayor el número de flujos de migraciones?
- **Tarea 3:**
 - Criterio Usuario
 - Interactividad
 - Conocimiento

Formato de la Entrevista

Formato de la Entrevista Proyecto de Tesis CCOMP. 9-1					
Motivación					
-El motivo de esta entrevista es recopilar datos de usuarios -Los usuarios serán evaluados en el uso de diferentes técnicas de visualización de datos Origen Destino, datos que presentan una serie de informaciones geográficas (espaciales) así como información temporal. -Se le permite al usuario la exploración de datos que ayuden a encontrar en estos datos información útil. -El principal motivo del trabajo es 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.					
Preguntas					
- (T1) Tarea 1: Consultas relacionadas a la espacialidad - (T2) Tarea 2: Consultas relacionadas a la temporalidad - (T3) Tarea 3: Consultas relacionadas al criterio del usuario.					
Tarea	Numero	Pregunta	Tiempo	Puntaje	Calidad (*)
T1	1	Indique el lugar donde las personas migran más?			
		Comentarios:			
T1	2	Indique el lugar donde las personas migran menos?			
		Comentarios			
T2	3	Indique el mes en el cual es mayor el número de migraciones?			
		Comentarios			
T2	4	Indique el mes en el cual es menor el número de migraciones?			
		Comentarios			
T3	5	Interactividad			
		Comentarios			
T3	6	Conocimientos			
		Comentarios			
			SCORE		

Resultados

- Evaluamos las dos técnicas implementadas en función a las tareas definidas.
- Presentamos nuestras técnicas a nueve personas diferentes:
 - 3 personas NORMAL
 - A, B y C
 - 2 personas FAMILIA
 - D y E
 - 4 personas COMPUTACIÓN
 - F, G, H e I

Evaluación Tarea 1

TAREA 1: CONSULTA ESPACIAL (en minutos)		
Persona	Flowstrates	Circular Cuerdas
A (NORMAL)	5 min	4 min
B (NORMAL)	3 min	5 min
C (NORMAL)	4 min	3 min
D (FAM)	4 min	5 min
E (FAM)	3 min	3 min
F (COMP)	3 min	4 min
G (COMP)	2 min	3 min
H (COMP)	3 min	2 min
I (COMP)	2 min	2 min
TOTAL	29 min	31 min
PROMEDIO	3,22 min	3,44 min

Tabla 1. Tiempo que demoró cada persona realizando la Tarea 1: Consultas Espaciales

Penalidad Tarea 1

TAREA 1: CALCULO DE PENALIDAD		
(+) Mejor que el promedio		
(-) Peor que el promedio		
Persona	Flowstrates	Circular con Cuerdas
A (NORMAL)	- 1	- 1
B (NORMAL)	+ 1	- 1
C (NORMAL)	- 1	+ 1
D (FAM)	- 1	- 1
E (FAM)	+ 1	+ 1
F (FAM)	+ 1	- 1
G (CCOMP)	+ 1	+ 1
H (CCOMP)	+ 1	+ 1
I (CCOMP)	+ 1	+ 1
TOTAL	3	1

Tabla 2. Medimos la penalidad (+ O -). Definimos un umbral (promedio del tiempo de Flowstrates 3.22 min. y Circular con Cuerdas 3.44 min.).

Evaluación Tarea 2

TAREA 2: CONSULTA TEMPORAL (en minutos)		
Persona	Flowstrates	Circular Cuerdas
A (NORMAL)	4 min	3 min
B (NORMAL)	2 min	3 min
C (NORMAL)	2 min	2 min
D (FAM)	4 min	2 min
E (FAM)	1 min	2 min
F (COMP)	4 min	4 min
G (COMP)	4 min	3 min
H (COMP)	1 min	2 min
I (COMP)	2 min	5 min
TOTAL	24 min	26 min
PROMEDIO	2,67 min	2,89 min

Tabla 3. Tiempo en minutos que demoró cada persona realizando la Tarea 2: Consultas Temporales

Penalidad Tarea 2

TAREA 2: CALCULO DE PENALIDAD		
(+) Mejor que el promedio		
(-) Peor que el promedio		
Persona	Flowstrates	Circular con Cuerdas
A (NORMAL)	- 1	- 1
B (NORMAL)	+1	- 1
C (NORMAL)	+ 1	+ 1
D (FAM)	- 1	+ 1
E (FAM)	+ 1	+ 1
F (FAM)	- 1	- 1
G (CCOMP)	- 1	- 1
H (CCOMP)	+ 1	+ 1
I (CCOMP)	+ 1	- 1
TOTAL	1	-1

Tabla 4. Definimos un umbral (el promedio del tiempo de Flowstrates 2.67 min. y Circular con Cuerdas 2.89 min.)

Evaluación Tarea 3

TAREA 3: CRITERIO USUARIO (en puntos)				
Persona	FlowStrates		Circular Cuerdas	
	Int.	Con.	Int.	Con.
A (NORMAL)	8 pts	6 pts	6 pts	8 pts
B (NORMAL)	7 pts	8 pts	7 pts	7 pts
C (NORMAL)	5 pts	6 pts	9 pts	8 pts
D (FAM)	9 pts	8 pts	4 pts	7 pts
E (FAM)	7 pts	7 pts	6 pts	8 pts
F (COMP)	8 pts	9 pts	8 pts	6 pts
G (COMP)	6 pts	8 pts	8 pts	9 pts
H (COMP)	7 pts	9 pts	7 pts	7 pts
I (COMP)	8 pts	8 pts	9 pts	8 pts
TOTAL	65 pts	69 pts	64 pts	68 pts
PROMEDIO	7,22 pts	7,67 pts	7,11 pts	7,56 pts

TAREA 3: CALCULO DE PENALIDAD				
(+) Mejor que el promedio				
(-) Peor que el promedio				
2*Persona	Flowstrates		Circular con Cuerdas	
	Int	Con.	Int	Con.
A (NORMAL)	- 1	+1	+ 1	- 1
B (NORMAL)	+1	- 1	+ 1	+ 1
C (NORMAL)	+ 1	+1	- 1	- 1
D (FAM)	- 1	- 1	+ 1	+ 1
E (FAM)	+ 1	+1	+ 1	- 1
F (FAM)	- 1	+ 1	- 1	+ 1
G (CCOMP)	+ 1	- 1	- 1	- 1
H (CCOMP)	+ 1	+ 1	+ 1	+ 1
I (CCOMP)	- 1	-1	- 1	-1
TOTAL	1	1	1	-1

Tabla 5. 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.

Calidad de Respuestas Tarea 3

TAREA 3: CALIDAD DE LA RESPUESTA (en puntos)		
Persona	Flowstrates	Circular con Cuerdas
A (NORMAL)	7 pts	6 pts
B (NORMAL)	6 pts	6 pts
C (NORMAL)	7 pts	7 pts
D (FAM)	5 pts	6 pts
E (FAM)	5 pts	5 pts
F (FAM)	6 pts	5 pts
G (CCOMP)	8 pts	7 pts
H (CCOMP)	8 pts	8 pts
I (CCOMP)	7 pts	8 pts
TOTAL	59 pts	50 pts
PROMEDIO	6.556 pts	5.556 pts

Tabla 7. De la Tarea 3, medimos la calidad de la respuesta dada por el usuario, es decir si acertó en su respuesta o si se aproximó. Este puntaje es puesto sobre todas las preguntas presentadas en las Tareas 1, 2 y 3

Análisis Cuantitativo

- Medimos **tiempo, puntaje, calidad y penalidad** que obtuvimos de la realización de tareas.
- Con las medidas anteriores afirmamos que la visualización Flowstrates
 - Es más eficiente y es de
 - Mejor manejo por parte de los usuarios.
- Al presentar un **menor tiempo y mejor puntaje**, inferimos que su contenido y la forma como presenta la información son las más adecuadas para la visualización de datos OD.

Análisis Cualitativo

- Flowstrates al **presentar un menor penalidad** (como una métrica que relaciona la ejecución de la tarea con su métrica temporal-puntaje) es la mas adecuada visualización OD.
- Flowstrates es también mejor que Circular Con cuerdas en calidad de respuestas, pues aborda de mejor manera el acierto de las respuestas al compararlas con las respuestas correctas.

Conclusiones (I)

- 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 movimiento de un lugar a otro.

Conclusiones (II)

- Analizando nuestros resultados, podemos concluir que :
 - En 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 Flowstrates.
 - 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 Flowstrates.

Trabajos Futuros

- En un futuro cercano, se planea recolectar más **retroalimentación del usuario y ejecutar** una evaluación del usuario con diversas tareas.
- 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.

Referencias

- [1] A. Rae. From spatial interaction data to spatial interaction information, Geovisualisation and spatial structures of migration from the 2001 UK census. Computers, Environment and Urban Systems.
- [2] Jenny, B., Stephen, D. M., Muehlenhaus, I., Marston, B. E., Sharma, R., Zhang, E., Jenny, H. (2018). Design principles for origin-destination flow maps. Cartography and Geographic Information Science, 45(1).
- [3] D. Guo and X. Zhu, Origin-Destination Flow Data Smoothing and Mapping in IEEE Transactions on Visualization and Computer Graphics, vol. 20, no. 12, pp. 2043-2052, Dec. 31 2014.
- [4] J. Wood, J. Dykes, and A. Slingsby. Visualisation of Origins, Destinations and Flows with OD Maps. The Cartographic Journal, 47(2):117–129, May 2010.
- [5] Y. Yang, T. Dwyer, S. Goodwin and K. Marriott, "Many-to-Many Geographically-Embedded Flow Visualisation: An Evaluation, in IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 1, pp. 411-420, Jan. 2017. doi: 10.1109/TVCG.2016.2598885
- [6] I. Boyandin, E. Bertini, P. Bak, and D. Lalanne. Flowstrates: An Approach for Visual Exploration of Temporal Origin-Destination Data. Computer Graphics Forum, 30(3):971–980, 2011.
- [7] Jesus Vera, Map-App Demo, [Online] https://github.com/iesusveca/map-app/blob/master/map-app/public/sample_merged_1.csv
- [8] Wei Qi, Guy J Abel, Raya Muttarak and Shenghe Liu, Circular visualization of China's internal migration flows 2010–2015, 2016, Environment and Planning A.
- [9] Heer, Jeffrey, Bostock, Michael, Ogievetsky, Vadim. 2010. A tour through the visualization zoo. Commun. ACM, 53 (6), 59–67
- [10] N. Sander, G. Abel, R. Bauer and J. Schmidt, "Visualising Migration Flow Data with Circular Plots". VID Austrian Academy of Sciences. 2/2014