Visualising Origin-Destination Flows for the Public Transportation System, London, United Kingdom

Kuhu Dahiya[1]

[1]Department of Geography
University of Manchester

January, 2017

Summary

Motivated by the lack of web-based flow maps, this paper presents an online origin-destination map visualisation method created using journey data obtained through Transport for London's open data platform. Bezier curves of varying widths, opacity and colour are used to represent the flow of over 700,000 Oyster card journeys. The dynamic filtering of data allows users to analyse commuter behavior, understanding which could be useful to researchers, transportation system managers, transport planners or commuters themselves. This method is a demonstration of how web technology can be harnessed for flow visualisation tools that are open to the general public.

KEYWORDS: Visualisation, Origin-Destination, Big Data, Oyster Card, Web Maps

1. Introduction

Assessing demand for the services of public transportation has become an easier and more accurate process since the increase in use of Automated Fare Collection systems. While aggregated statistics regarding the number of commuters can provide a general idea of demand, data containing information related to individual journeys can be relied upon for analytical precision. For instance, public transport usage in London amounted to an average daily ridership of 3,614,000 for rail networks and 6,397,000 for bus networks, which effectively amount to billions of journeys made each year (LTA Academy, 2013). From the aggregated statistics it can be derived that the infrastructure needs to be able to support billions of journeys each year. However, it would be improper to assume that the journeys are evenly distributed across the network of the transport system. Individual journey data can provide better insight regarding the “when” and “where” of public transport usage and this knowledge can be used for assessing the requirements for additional infrastructure and frequency of services.

Accumulating individual journey data is no longer a challenge, however deriving meaning from large datasets continues to be exacting. While manually analysing high volumes of data can be impractical, it can be transformed into valuable knowledge by sorting, classifying or visualising it (Figure 1).

[1] kuhu.dahiya@manchester.ac.uk

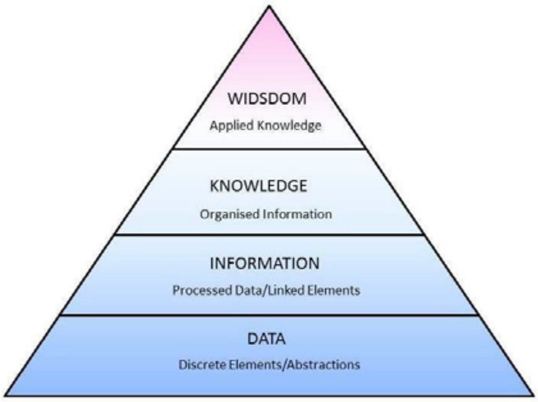

Figure 1: Data Information Knowledge Wisdom (DIKW) pyramid (adapted from Kitchin, 2014)

Visualisations of transport data have existed since the 19th century in the form of origin- destination maps (Robinson, 1955). While techniques for creating origin-destination maps have evolved over the years, further advancements can be made with the technologies available today.

The advancement and increased reliance on the internet has made it possible not only to easily obtain data but to be able to visualise it using web interfaces. While many desktop based applications and computer languages have been used to create origin-destination maps, there is still a lack of web based visualisation tools to map journey information data.

2. Requirements

There were three main objectives identified for this research design project:

(a) A database design which can be applied to all public transportation systems that utilize Automated Fare Collection systems.
(b) A web-based visualisation tool with the ability to handle data for millions of individual journeys and automatically categorize them into origin-destination vectors with minimal visual clutter.
(c) Incorporation of functionality that allows users to filter the visualisation on the basis of parameters such as time and day.

In order to create origin-destination flows, the minimum basic requirement for datasets is to include information regarding the geographical locations of a journey's origin and destination. To meet the objectives of this research project, the requirements for data can be met with information regarding geographical locations of origin and destination, times of entry and exit as well as the day on which a journey was made. Inclusion of time and day information is necessary to incorporate the functional requirement of filtering data. The data should be at the lowest level of aggregation, i.e. individual journeys, and the dataset should contain information regarding a minimum of 500,000 journeys. The minimum amount of journeys required is an important criterion to assess the effectiveness of design for the visualization and to test the ability of the database design.

3. Systems Development Life Cycle and Data Acquisition

A Systems Development Life Cycle (SDLC) divides the process of creating a system into different stages from inception to completion (Ruparelia, 2010).
The SDLC for the creation of this visualisation tool (Figure 2) is divided into ten stages where input from users is considered at three different stages between several stages of design and development.

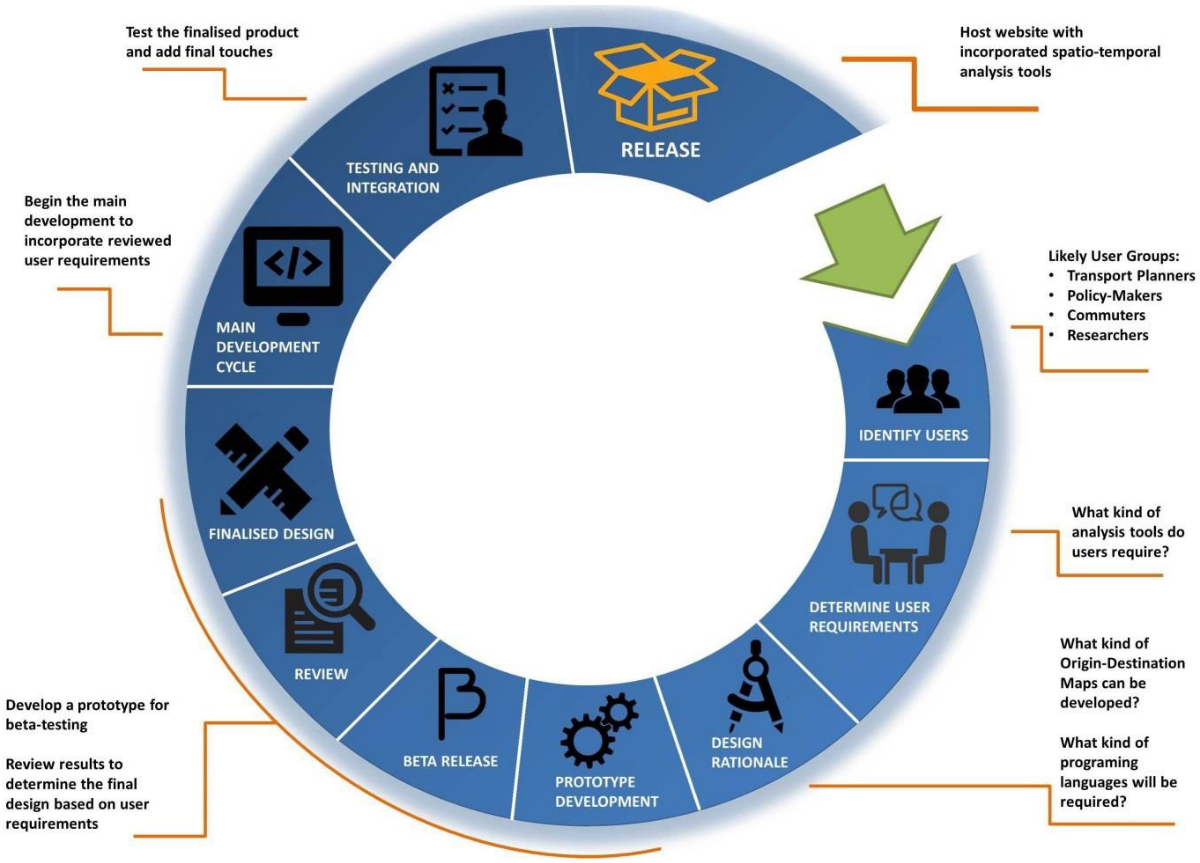

Figure 2: Systems Development Life Cycle for proposed Origin-Destination Visualisation Tool

The first dataset for creating the visualisation was downloaded from the open data platform of Transport for London (TfL) and is entitled "Oyster Card Journey Information" (TfL, 2014). The dataset consists of a 5% sample of individual journeys made using Oyster cards (smart cards for

AFC) during a 7 day week in November, 2009. The dataset was made available in the format of Comma Separated Values (CSV) and was comprised of 2,622,488 rows of individual journeys. The journeys were made using various modes of public transport, such as buses, trams and Mass Transit Railway (MRT) in London.

Several elements of the journeys were recorded:

- Day on which the journey was made
- Mode or modes of journey
- Names of stations where the journey started and ended
- Times of entry and exit
- Fare and type of ticket used for the journey
- Route numbers for buses

While the names of stations where entries and exits were made were present, the first dataset did not contain information regarding the geographical location of the stations. A second dataset containing geographical coordinates of station locations was downloaded through the TfL open data platform. Entitled “Station Locations” (TfL, 2016), this dataset was in the Keyhole Markup Language (KML) format. Data for bus and tram journeys had to be discounted due to lack of information regarding destination.

4. Visualisation

Visual clutter is an issue that arises when creating flow maps using large amounts of data. Such clutter can make visulisations incomprehensible, hide important patterns and leave little room for exploratory analysis. Two main concerns for graphical considerations of this project’s visualisation were choosing appropriate geometry for creating flow lines and appropriating data into different classes.

Straight lines have often been used to create flow maps but there is little clarity about the origins and destinations without distinguishable terminals. This information is important, especially for analysing temporal flow. To display differences in origins and destinations, the Bezier curve (type of parametric curve) geometry was chosen. Quadratic Bezier curves with a 10% offset were considered most appropriate as one terminal could be distinguished by a flat end and the other by a curved end.

To quantify the data, 10 classes of Bezier curves were created with varying line thickness, opacity and colours. Since scalability and applicability are important factors, the method to classify the data has not been restricted to data available for this research and can be used with data of varying size. The finished design of the visualisation tool (figure 3) comprises of a web map with Bezier curves overlaid. The left panel on the tool's web page can be used to filter day and time parameters of the data (figures 4 & 5).

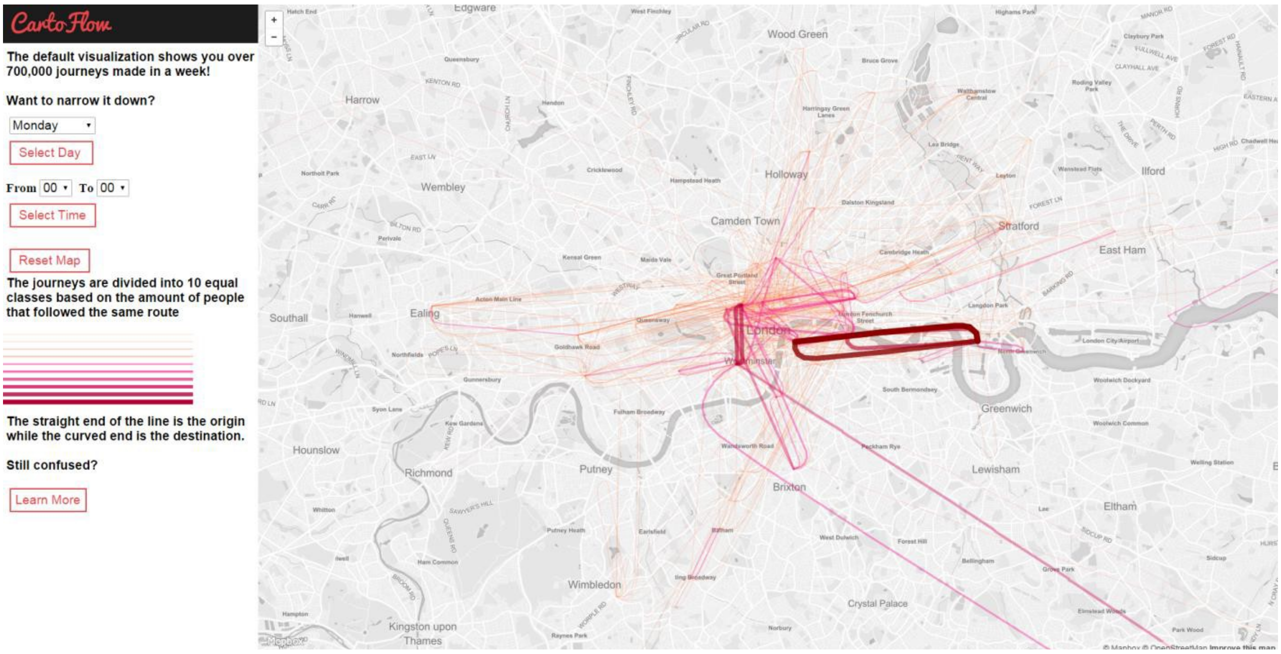

Figure 3: Origin-Destination Visualisation Tool

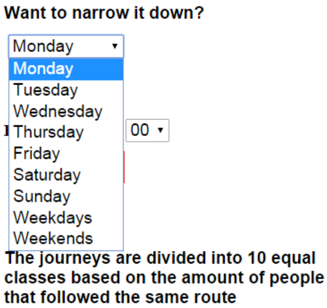

Figure 4: Origin-Destination Visualisation Tool - Day Filter

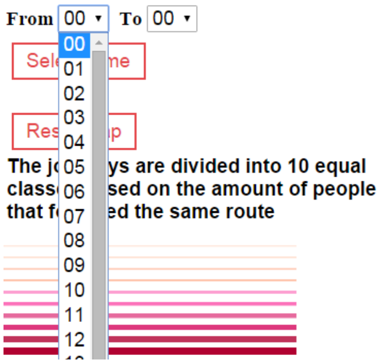

Figure 5: Origin-Destination Visualisation Tool – Time Filter

Filtering the data using time as a parameter, distinct patterns start to emerge. Using this sample dataset, filtering between 7am – 9am displays a prominent pattern where a large number of people appear to be exiting at the Canary Wharf station (figure 6). This could be indicative of a large working population around the Canary Wharf area.

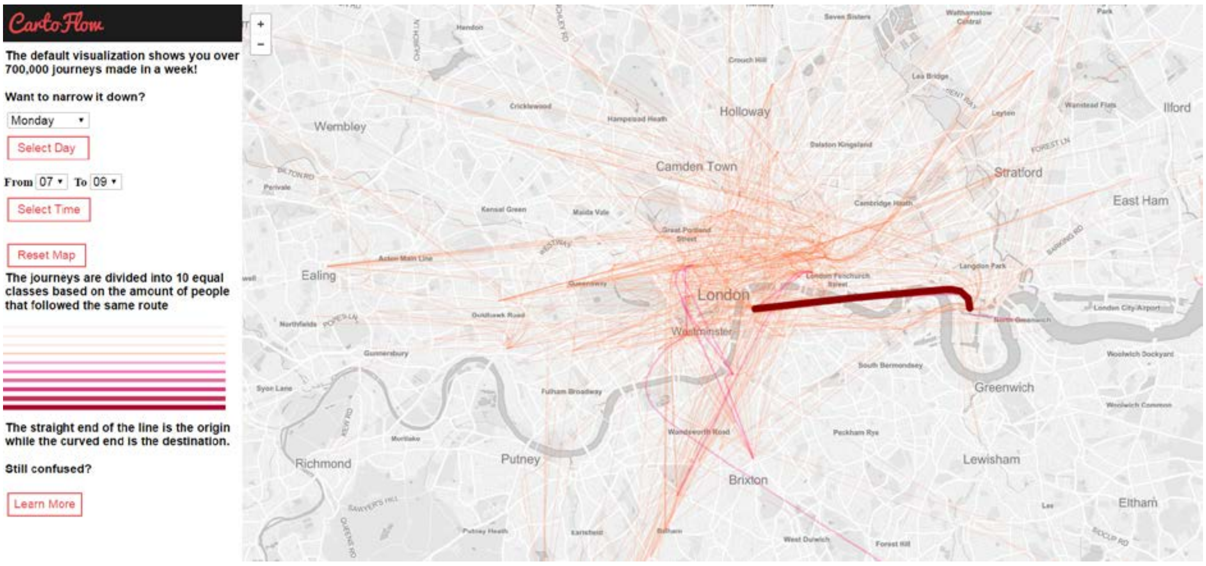

Figure 6: Origin-Destination Visualisation Tool – Filtered from 7am to 9am

This tool can be considered a prototype for origin-destination flow maps and the use case can extend beyond public transportation. To improve the analytical capabilities of this tool in the context of public transportation, additional elements can be added to filter data based on modes of transportation used, type of ticket used and points of interchange. Additionally, public transportation authorities could incorporate this tool with a database which gets updated instantaneously allowing them to view near-real time flows for managing resources.

Some interesting applications of this tool could include origin-destination maps using data related to trade and telecommunications. Since the tool was built using open source materials, it can be replicated without significant cost or effort. The code can be downloaded and modified to create iterations of origin-destination visualisation tools for mapping different kinds of flows. Using libraries such as Turf.js, other types of spatial analysis capabilities can also be incorporated to increase dynamic map content.

5. References

LTA Academy (2013). Journeys: Sharing Urban Transport Solutions, *Land Transport Authority*, Singapore, pp. 105 – 110

Robinson, A.H. (1955). The 1937 Maps of Henry Drury Harness, *The Geographical Journal*, Vol. 121, pp. 440-450

Ruparelia, N.B. (2010). Software Development Lifecycle Models, *ACM SIGSOFT Software Engineering Notes*, Vol. 35, No. 3, pp. 8-13

Transport for London (2014). Oyster Card Journey Information, Created: 13:17:02, 22nd January, 2014.Update Frequency: None. Database ID: 8ae76354-8951-4c8b-87a2- 862fc126ebc8

Transport for London (2016). Station Locations, Created 15:35:43, 2nd February, 2010. Update Frequency: Monthly. Database ID: 5585e6a9-2115-49a0-8300-b242a835b990

6. Acknowledgements

I would like to start by expressing my gratitude for my dissertation supervisor, Dr Jonathan Huck who helped me transform a vague idea into an ambitious project through constant support and guidance.

My appreciation extends to all the developers who have created the various programs, applications and code libraries used for this research design project. Their efforts and contribution to the open source community deserve admiration.

7. Biography

Kuhu Dahiya is a GIS analyst and a recent graduate of the MSc GIS program at the University of Manchester. She is passionate about data visualisation in the spatial context and hopes to continue learning as the field progresses with development in technology.