Visualizing Transportation Flows with Mode Split using Glyphs

I. Perez-Messina[1] and E. Graells-Garrido[1,2]

[1]Data Science Institute, Universidad del Desarrollo, Chile
[2]Telefonica R&D, Chile

Abstract

The increasing trend of using unconventional data in urban planning environments has led to the need for developing systems that can visualize this data. Here we present a visualization for studying commuting flows within a city, with a particular focus on the distribution of mode of transportation usage. Our design, called ModalCell, uses a glyph-based flow map to show a city's flows considering mode split, direction, and distance range. We evaluate ModalCell with a pilot survey and a use case that shows the potential of the approach to make flows within a city visible and understandable.

CCS Concepts

• ***Human-centered computing*** $\rightarrow$ ***Information visualization;*** *Visual analytics;*

1. Introduction

In recent years, the importance of Big Data for urban mobility has been heavily noted. The availability of large amounts of data from non-traditional sources has allowed to study urban phenomena at previously unseen scales [ZCW*16], hinting that it could be possible to analyze the city at finer spatio-temporal granularity than before. One type of recently available data used to study fine-grained transportation and mobility is Data Detail Records (XDR). These records are generated by mobile operators with the purpose of billing customers, and have been used for a wide range of studies regarding mobility, transportation, and human behaviour in general [BDK15]. Current state of the art methods to infer commuting patterns go as far as disentangling mode of transportation usage [GGCP18].

However, domain experts do not work with tools that allow them to explore these vast data sets, and new inference methods based on Machine Learning are not always interpretable, nor follow the conventions of their disciplines. For instance, even though speed is a relevant feature for transportation experts, XDR-based inference does not rely on it, as speed is not reliable from XDR. As such, building trust in the results and making them transparent for domain experts is a challenge [B*01].

Visualization is a powerful tool to build trust [KPN16]. To contribute in bridging the gap between Data Science and Transportation, in this paper we introduce ModalCell, a visualization for XDR-inferred flows with mode split within a city. We propose a novel glyph design that encodes flow magnitude and mode of transport, with discretized direction and distance dimensions.

This paper is structured as follows. Section 2 shows the context and data set where we instantiate ModalCell. Section 3 defines the user tasks that ModalCell aims to. Section 4 discusses the related work and its influence on the proposed design space. The design space and the user interface is described in Section 5. Section 6 describes a brief evaluation of the system, through a use case and a pilot survey. Finally, Section 7 describes the conclusions and future lines of work.

2. Context and Data Set

In this paper we visualize trips in Santiago, the capital of Chile, a city with almost 8 million inhabitants in 35 administrative units denoted *municipalities*. A *trip* [Hal12] is a displacement from one point to another, with the following features: *origin* (which may be a geographical position, or a specific area, such as a block, neighbourhood, census tract, etc.), *destination* (*idem*), *waypoints* (*idem*) in some cases, *date*, *departure time*, *travel distance*, *travel purpose*, and *mode of transportation choice*.

Here, we work with a data set of inferred commuting trips from XDR, including purpose and mode(s) [GGCP18]. In total, it contains trips from approximately 600,000 devices, in a one-month period, August 2016. We restrict this paper to spatial analysis, so only one day of the dataset is showcased. Bus, subway (metro), and car trips were inferred for this data set. Pedestrian trips were left out of the analysis in the source data, as we focus on long distance transportation (>1 Km.).

In the next section we define the user tasks to be performed on this data set.

3. User Tasks

We target urban and transportation planners as users of our proposed system. Since transportation data contains geographical in-

DOI: 10.2312/evs.20191165

formation, domain experts traditionally work with Geographical Information Systems (GIS) that present information on tables, maps, and other visualizations [Thi00].

The usual workflow in using GIS for transportation is aligned with a common paradigm to define visualization processes, known as the "Visual Information Seeking Mantra:" *overview first, zoom and filter, then details-on-demand* [Shn96]. An overview is an important part of the analytical process and a starting point for more detailed analysis. Andrienko et al. [AAFW17] proposed an analytical procedure targeted on revealing and exploring spatial and temporal patterns in flow data, a step of which considers a map-view with glyphs encoding flows used to perform spatial flow analysis tasks. We expand on those specific tasks by adding the mode of transportation dimension. Thus, we expect our system to assist in:

- **Task 1**: Getting an overview of the transportation mode spatial distribution of flows within a city and major spatial trends.
- **Task 2**: Inferring the transportation behavior of particular administrative zones of the city: identifying major hubs, high-density traffic areas and flow trends, and comparing flows of different modes, distance ranges and directions.

Next, we describe previous solutions found in the literature for traffic flow visualization.

4. Related Work

Visualization for traffic analysis is an active field of research, with unconventional data sources (e.g., mobile phone and social network data) as its predominant source [AAC*17]. There are three main approaches to work with the tasks we defined in the previous section: *graph-based*, *matrix-based*, and *glyph-based flow visualization*.

Graph-based visualization is a natural choice for showing geospatial relations, in which links connect related points of interest. However, when links between nodes become highly probable, which is the case with mobility data, readability of this scheme decreases abruptly, mainly due to the density of connections and the fixed positions of nodes. A series of techniques have been developed to cope with this: Edge aggregation [VLBR*16], flow simplification [BM14], edge-bundling [GSRB17, HVW09], and clustering algorithms [ZG14]. These methods have been applied in geospatial settings [CGW15] but are known for introducing artifacts and working well for special cases only [AAFW17]. For instance, the visual encoding of edges (bundled or not) is the same generally used for trajectory visualization, although edges may not represent routes.

Matrix-based visualizations make an efficient use of visual space but have a rigid layout which does not allow for geographic operations. Maptrix [YDGM17] enhances the traditional O-D Matrix with interactivity and a synchronized choropleth map. Flow tree [WDSR09] use another kind of layout where each square on the grid shows the flow to every other square. Tile maps have been developed to overcome square-grids while maintaining spatial relations [MH17]. In spite of particular improvements, matrices do not cope well with overviewing as they are not easily relatable to a map and rely on heavily abstracted space notions.

Glyphs allow to encode abstract flow information, confining it to a limited icon-sized space, and benefit from being visually independent from one another [BKC*13]. Following this line of research, Ma et al. [MLC*16] used the sunburst diagram as a model for their glyph, which encodes flow origin and direction. Zeng et al. [ZFAQ13] combined the glyph-based approach with explicit flows that attach to special parts of the glyphs to visualize subway commuting. In Andrienko et al. [AAFW17] radial diagrams are used to encode flow magnitude by direction and distance range. These glyph designs do not consider trip variables such as mode of transportation. Thus, it remains an open question how to make them multi-variate in that aspect.

To the extent of our knowledge, there is no solution that meets all the requirements of our tasks. Concluding, we consider a glyph-based design an appropriate choice to build upon, as glyphs produce less occlusion than graphs, and give a more intuitive spatial sense than matrices.

5. System Design

In this section we motivate the visual design of ModalCell. First, we describe the data structure that feeds ModalCell. Then we define the design space of the proposed glyphs that allow to perform the tasks under consideration. Finally, we describe the user interface containing the glyphs with interaction on a map. Figure 1 shows the graphical user interface for ModalCell, with three alternative glyph designs.

Data Structure An XDR-inferred trip is comprised by a sequence of cell towers from origin to destination, each one with a specific geolocation. Cell towers can be aggregated according to an arbitrary space partition; we use municipalities since our target users are transportation planners. Since a trip should only count once in the flow magnitude of the city, we produced three ways of accounting for them, which we call *flow orientations*: *outgoing*, where each trip is counted at its origin tower; *incoming*, where each trip is counted at its destination tower; and *passing*, where each trip's weight is divided equally into each tower of its sequence.

Trip features include mode(s) of transportation, direction (the origin-destination spatial vector) and distance (euclidean distance from origin-to-destination). We discretize the direction and distance dimensions of trips, and arrange the resulting flows in the data structure shown on Figure 2.

Visual Representation As discussed in Section 4, a glyph-based visualization seems the most appropriate choice to represent flows in the city as required by Section 3. Our main challenge is encoding mode of transportation along other variables.

Figure 3 shows the design space of ModalCell, with three alternative glyph designs. As mode of transportation is the only categorical variable under consideration, with a limited number of values (in our case: bus, metro, and car), color was deemed the best visual channel to encode it. Shape, size, and position are left as possible channels for flow magnitude, distance and direction. A radial diagram naturally follows from the data structure, so elliptical forms immediately appear as a congruent option. As the central space in

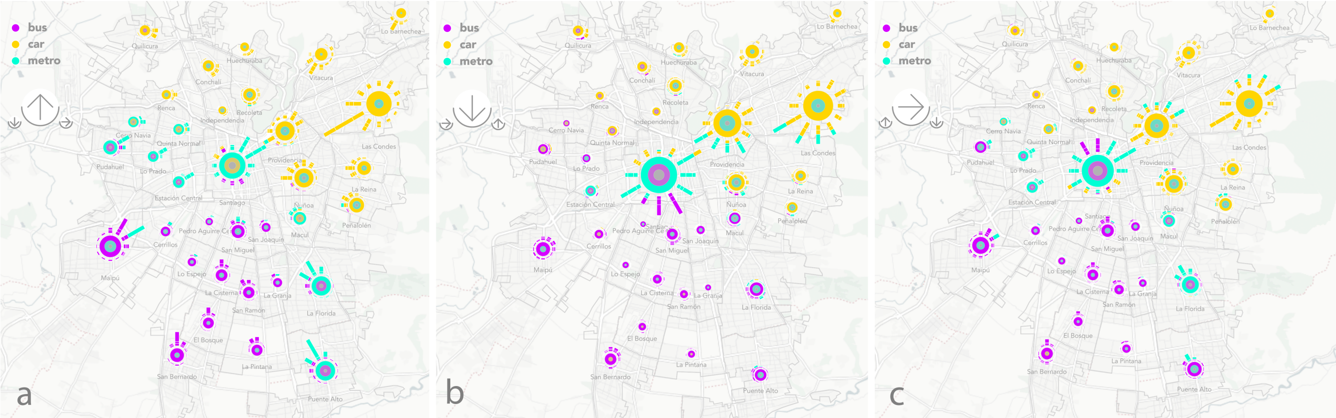

Figure 1: *ModalCell showing commute data of Santiago for the different flow orientations: (a) outgoing, (b) incoming and (c) passing. The legend at the top-left corner of the screen can be interacted with to filter by mode of transportation and switch flow orientation (by clicking on the arrow-shaped icon).*

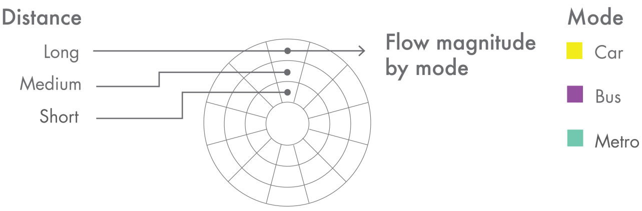

Figure 2: *Flow data structure diagram. Angular space is split between twelve bins, while the distance dimension is split into short (<5 km.), medium (<10 km.) and long (>=10 km.). Each bin has a flow magnitude associated to each mode of transportation equivalent to its total trip count.*

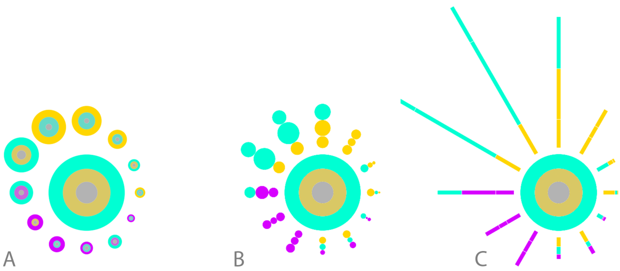

Figure 3: *Flow diagram alternatives, representing the same set of values. A Represents flow in each direction as a combination of transportation modes. B and C differentiate flows by distance (short to long distance range from the center out) but only show the predominant mode for each range. The embedded circles at the center, common to all, show the aggregated flow magnitude by mode.*

the data structure is empty, it is used as a summary of the data, allowing for a better overview (Task 1), while the radial part of the diagram shows the details of the flows' distance and direction (Task 2). For the central part, an embedded circles glyph was used, with the predominant mode of transportation as the outermost, to enhance color preattentiveness. In our first design (Fig. 3 A) the same embedded circles were used to represent the flow disaggregated by direction radially, but does not incorporate distance dimension. For our second design (Fig. 3 B) circles were juxtaposed to differentiate by distance range, with color only showing the predominant mode for each distance range to avoid information saturation. Short, medium and long trips are arranged center-out, in that order. Juxtaposed bars were employed instead of circles in the last design (Fig. 3 C), as circular area is not well suited for magnitude comparison.

Colors to encode mode of transportation were generated using the HSL model by splitting the hue dimension into n equal parts, with n the number of mode of transportation classes. As the design produces embedded circles of different classes, the following method is used to avoid optical vibration and diminished readability due to high contrast: the first circle to be embedded into another has its saturation reduced to half (thus reducing optical vibration while retaining class association), then, the remaining classes are aggregated into a gray (i.e., non saturated) circle. This achieves maintaining the preattentive attributes of color for the predominant mode of transportation while leaving the rest of the information available for closer inspection.

To avoid overlapping glyphs, glyph positions are precalculated through an iterative algorithm, which takes into account each element's size and distance to its respective area. In each iteration, overlapping glyphs repel each other and then are attracted to their respective area's centroid, so that they do not exceed its boundaries.

User Interface The glyph design is embedded into an interactive tiled Web map, implemented in *p5.js* [p5]. The map contains a legend menu on the top-right corner. Rather than using a legend and a separate widget for options, a legend menu serves both functions and saves screen space. In Figure 1 the Legend Menu is shown: an interactive legend that serves as a widget for filtering options and global parameters (for a study on interactive legends, see [RLP10]). Available options are *mode of transportation filtering*, and the *flow orientation switch*, which induces the glyphs to display either outgoing, passing or incoming flows.

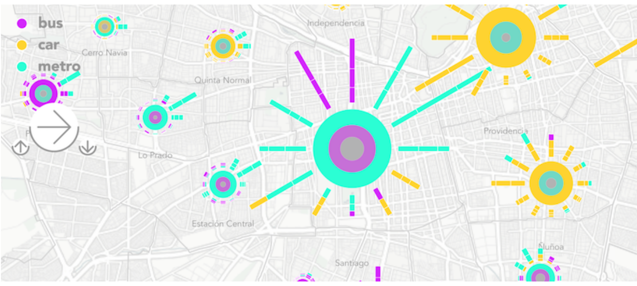

Figure 4: *ModalCell switched to display passing flows and zoomed in on Santiago downtown. A mixed mode distribution can be observed, with metro and bus as the most numerous. Important flow variations are associated to the incidence angle.*

6. Evaluation

In this section we describe a preliminary two-fold evaluation of ModalCell: a use case, showing how the defined tasks can be solved with the system, and a pilot survey, where we asked domain experts to evaluate the glyph designs as overviews of the city.

Use case: Commuting Flows We conduct a use case in which the user wants to get a general understanding of the flows in Santiago (Task 1) and see the transportation behavior of a particular zone of interest (Task 2).

Task 1 Figure 1a shows outgoing flows at municipality granularity in Santiago. By looking at the overall shape and color of the glyphs the user can get two insights: the distribution of mode of transportation predominance in space (based on color hue), and the possible commuting attractors of the city (based on glyph size and tilt). Switching the flow orientation to *incoming* (Fig. 1b) reveals the heavy intake of Santiago (downtown), Providencia and Las Condes, confirming their centric role.

Task 2 Next, the user wants to know more about the center of the city. Looking at the radial part of Santiago downtown through the different flow orientations (Fig. 1) it can be observed that this municipality shows a mixed behavior in terms of mode distribution, but also that subway is the most important in each case. Bus flow comes mostly from the south and is the largest mode in the long distance range (Fig. 1b). Switching flow orientation to *passing* and zooming in (Fig. 4) reveals that the south-east and south-west directions show predominantly car flow. Then, the user can repeat this process for other zones of interest, and with this knowledge perform more detailed questions over the data.

Pilot Survey We designed a small survey to grasp how domain experts perceive the glyph designs (c.f. Fig. 3) applied to XDR data from Santiago. We showed the three outflow overviews of the city (one per glyph), and for each one we asked task-related questions (e.g., "Which municipality has the biggest outgoing flow?", "Which municipality has more long trips?"), and asked them to rate the legibility of the visualization with a 5-point Likert scale. Then, we asked global questions, such as which design was easier to understand, which one was more functional according to their needs, and which one was more aesthetically pleasing.

In total, we surveyed 13 domain experts (7 from Transportation, 4 from Data Science, 2 from Urbanism). As a summary of the results, we observed that the first design (Fig. 3, A) was the more legible, functional, and liked by users in terms of aesthetics. In terms of tasks, all designs allowed users to give satisfactory answers, however, this is only a validation in terms of coherence—these users were experts and probably had prior knowledge about the context.

Some users claimed that the geometry of the glyph was useful, but that the color encoding was confusing. They explicitly suggested to use one glyph per mode of transportation. This suggests that having the option of coordinated-views in ModalCell would help to target experts used to different interaction paradigms.

7. Discussion and Future Work

In this paper we contributed a work-in-progress visualization tool, ModalCell, that follows a glyph-based overview of the movement flows within a city, with a focus on mode split. We observed that the design allows to identify city-level commuting patterns as a starting point for more detailed questions and tasks to be performed, evaluated with a use case and a pilot survey.

In addition to conduct a full user study with ModalCell, we devise three lines of future work. First, as suggested from our pilot survey, we may include coordinated views in the system, through either using the same visual encoding, or by using different visualizations (which would enable to show other details-on-demand, available from other data sets), extending the capabilities of our system in-line with the analytical procedure performed by domain experts. Second, we have not explored the full possibilities of the design space in ModalCell. For instance, a composite glyph, where the inner part shows another radial visualization instead of concentring circles, could be useful to depict the summary distribution. A pie chart could be a reasonable choice, given the small number of categories depicted. Finally, time-dependency analysis should be addressed. This aspect was left out this work, but as our system fits into the analytical procedure for revealing spatial and temporal flow patterns proposed by Andrienko et al. [AAFW17], the same techniques could be applied to expand this work into the temporal dimension, adapting it to accommodate mode of transportation into their clustering algorithms.

A limitation of our design is that it does not consider intermodality, i.e., the usage of more than one mode of transportation in a trip. This is relevant in cities such as Santiago, where trips combining bus and metro are common. The solution may depend on the task at hand. For instance, if not all combinations of modes are relevant, then a combination could be considered as an additional mode in our current encoding.

To conclude, understanding the city with new data sources and new ways of analyzing data has the potential to improve cities [CMS*16]. We believe our work-in-progress contributes to bridging this gap through visualization design.

Acknowledgements The authors were partially funded by Fondecyt de Iniciación project #11180913. We also thank the comments by the anonymous reviewers, as they helped to improve this paper.

References

[AAC*17] Andrienko G., Andrienko N., Chen W., Maciejewski R., Zhao Y.: Visual analytics of mobility and transportation: State of the art and further research directions. *IEEE Transactions on Intelligent Transportation Systems 18*, 8 (2017), 2232–2249. 2

[AAFW17] Andrienko G., Andrienko N., Fuchs G., Wood J.: Revealing patterns and trends of mass mobility through spatial and temporal abstraction of origin-destination movement data. *IEEE Transactions on Visualization & Computer Graphics*, 1 (2017), 1–1. 2, 4

[B*01] Breiman L., et al.: Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science 16*, 3 (2001), 199–231. 1

[BDK15] Blondel V. D., Decuyper A., Krings G.: A survey of results on mobile phone datasets analysis. *EPJ Data Science 4*, 1 (2015), 10. 1

[BKC*13] Borgo R., Kehrer J., Chung D. H., Maguire E., Laramee R. S., Hauser H., Ward M., Chen M.: Glyph-based visualization: Foundations, design guidelines, techniques and applications. In *Eurographics (STARs)* (2013), pp. 39–63. 2

[BM14] Bahbouh K., Morency C.: Encapsulating and visualizing disaggregated origin-destination desire lines to identify demand corridors. *Transportation Research Record: Journal of the Transportation Research Board*, 2430 (2014), 162–169. 2

[CGW15] Chen W., Guo F., Wang F.-Y.: A survey of traffic data visualization. *IEEE Transactions on Intelligent Transportation Systems 16*, 6 (2015), 2970–2984. 2

[CMS*16] Chen C., Ma J., Susilo Y., Liu Y., Wang M.: The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation research part C: emerging technologies 68* (2016), 285–299. 4

[GGCP18] Graells-Garrido E., Caro D., Parra D.: Inferring modes of transportation using mobile phone data. *EPJ Data Science 7*, 1 (2018), 49. 1

[GSRB17] Graser A., Schmidt J., Roth F., Brändle N.: Untangling origin-destination flows in geographic information systems. *Information Visualization* (2017), 1473871617738122. 2

[Hal12] Hall R.: *Handbook of transportation science*, vol. 23. Springer Science & Business Media, 2012. 1

[HVW09] Holten D., Van Wijk J. J.: Force-directed edge bundling for graph visualization. In *Computer graphics forum* (2009), vol. 28, Wiley Online Library, pp. 983–990. 2

[KPN16] Krause J., Perer A., Ng K.: Interacting with predictions: Visual inspection of black-box machine learning models. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), ACM, pp. 5686–5697. 1

[MH17] McNeill G., Hale S. A.: Generating tile maps. In *Computer Graphics Forum* (2017), vol. 36, Wiley Online Library, pp. 435–445. 2

[MLC*16] Ma Y., Lin T., Cao Z., Li C., Wang F., Chen W.: Mobility viewer: An eulerian approach for studying urban crowd flow. *IEEE Transactions on Intelligent Transportation Systems 17*, 9 (2016), 2627–2636. 2

[p5] p5.js. https://p5js.org. Accessed: 2019-02-27. 3

[RLP10] Riche N. H., Lee B., Plaisant C.: Understanding interactive legends: a comparative evaluation with standard widgets. In *Computer graphics forum* (2010), vol. 29, Wiley Online Library, pp. 1193–1202. 3

[Shn96] Shneiderman B.: The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on* (1996), IEEE, pp. 336–343. 2

[Thi00] Thill J.-C.: Geographic information systems for transportation in perspective. *Transportation Research Part C: Emerging Technologies 8*, 1-6 (2000), 3–12. 2

[VLBR*16] Von Landesberger T., Brodkorb F., Roskosch P., Andrienko N., Andrienko G., Kerren A.: Mobilitygraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering. *IEEE transactions on visualization and computer graphics 22*, 1 (2016), 11–20. 2

[WDSR09] Wood J., Dykes J., Slingsby A., Radburn R.: Flow trees for exploring spatial trajectories. 2

[YDGM17] Yang Y., Dwyer T., Goodwin S., Marriott K.: Many-to-many geographically-embedded flow visualisation: an evaluation. *IEEE transactions on visualization and computer graphics 23*, 1 (2017), 411–420. 2

[ZCW*16] Zheng X., Chen W., Wang P., Shen D., Chen S., Wang X., Zhang Q., Yang L.: Big data for social transportation. *IEEE Transactions on Intelligent Transportation Systems 17*, 3 (2016), 620–630. 1

[ZFAQ13] Zeng W., Fu C.-W., Arisona S. M., Qu H.: Visualizing interchange patterns in massive movement data. In *Computer Graphics Forum* (2013), vol. 32, Wiley Online Library, pp. 271–280. 2

[ZG14] Zhu X., Guo D.: Mapping large spatial flow data with hierarchical clustering. *Transactions in GIS 18*, 3 (2014), 421–435. 2