

Visual Exploration of Utah Trajectory Data and their Applications in Transportation

Seth Miller
Zachary Vander Laan
Yinhu Wang
Nikola Markovic, Ph.D.



Visual Exploration of Utah Trajectory Data and their Applications in Transportation

Final Report

NITC-SS-1264

by

Seth Miller
Zachary Vander Laan
Yinhu Wang
Nikola Markovic
University of Utah

for

National Institute for Transportation and Communities (NITC)
P.O. Box 751
Portland, OR 97207



March 2020

Technical Report Documentation Page			
1. Report No. NITC-SS-1264	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Visual Exploration of Utah Trajectory Data and their Applications in Transportation		5. Report Date March 2020	
		6. Performing Organization Code	
7. Author(s) Nikola Markovic		8. Performing Organization Report No.	
9. Performing Organization Name and Address University of Utah		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No. 18-1264	
12. Sponsoring Agency Name and Address National Institute for Transportation and Communities (NITC) P.O. Box 751 Portland, OR 97207		13. Type of Report and Period Covered Final (12/2018-11/2019)	
		14. Sponsoring Agency Code	
15. Supplementary Notes			
16. Abstract Transportation agencies have an opportunity to leverage massive GPS trajectory data to support their analyses and enhance their decision making. However, these data are typically purchased from vendors, which means that transportation agencies must understand their benefits before they decide to invest in data acquisition. To help transportation agencies better understand the tremendous value of trajectory data, the PI of the project acquired Utah GPS trajectory data from one of the leading GPS data providers and conducted visual analyses demonstrating applications of trajectory data in various areas of transportation. The analyses focused on (a) evaluating the data itself by exploring its spatial and temporal coverage, and on (b) using visual analytics to demonstrate potential use of trajectory data in planning as well as measuring mobility and performance. The results of the analyses were shared with transportation agencies to help them better understand the value of this type of mobility data in transportation system analyses.			
17. Key Words Big data, data visualization, machine learning		18. Distribution Statement No restrictions. Copies available from NITC: www.nitc-utc.net	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 11	22. Price

ACKNOWLEDGEMENTS

Nikola Markovic would like to acknowledge support from the National Institute for Transportation and Communities (NITC; grant number 1264), a U.S. DOT University Transportation Center.

DISCLAIMER

The contents of this report reflect the views of the authors, who are solely responsible for the facts and the accuracy of the material and information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation University Transportation Centers Program in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof. The contents do not necessarily reflect the official views of the U.S. Government. This report does not constitute a standard, specification, or regulation.

RECOMMENDED CITATION

Markovic, Nikola. *Visual Exploration of Utah Trajectory Data and their Applications in Transportation*. NITC-RR-264. Portland, OR: Transportation Research and Education Center (TREC), 2020.

TABLE OF CONTENTS

Executive Summary.....	1
1.0 Introduction.....	2
2.0 Literature Review	2
3.0 Data	3
3.1 Processing.....	4
3.2 Final Dataset And Penetration Rates.....	5
4.0 Visual Data Exploration.....	6
5.0 Conclusions.....	9
6.0 References	10

LIST OF FIGURES

Figure 1: A trip with artificial waypoints.	1
Figure 2: Interactive animations and visualizations of trajectory data.	1
Figure 3: A sample GPS trajectory.....	4
Figure 4: An example of GPS signal loss.....	4
Figure 5: An example of trajectory map matching.	5
Figure 6: Density of the trajectories and relative distribution of trips.	6
Figure 7: Trips recorded during a 30-minute interval in Salt Lake City	6
Figure 8: Densities of vehicle GPS locations in Salt Lake City.....	7
Figure 9: Chord diagrams of O-D patterns extracted from trajectory data.....	8
Figure 10: Visualization of trip counts.....	9

EXECUTIVE SUMMARY

The objective of this project was to demonstrate the value that massive GPS trajectory data could bring to the analysis and decision-making efforts of transportation agencies. With this goal in mind, the PI purchased one month of Utah trajectory data, which included detailed information about 2.5 million trips that took place in September 2018 (i.e., 130 million recorded locations taking 12 GB of space). First, the data were processed to correct for errors inherent to GPS technology and to match with the state road network, which increased its size by the factor of seven (i.e., over 900 million points taking over 90 GB of space). Second, the data were evaluated by exploring their spatial/temporal coverage and via comparison with ground-truth information from 400 point detectors across Utah. This revealed very good spatial coverage and a capture rate of about 2%, which implies that the acquired trajectory data capture one in every 50 trips in Utah.

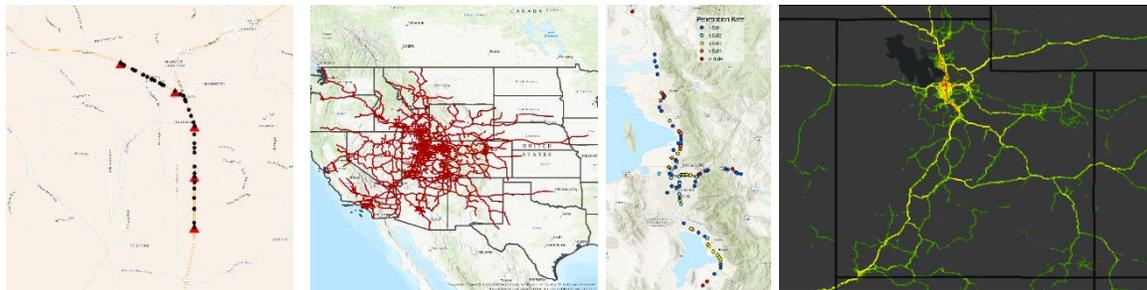


Figure 1: A trip with artificial waypoints imputed using Hidden Markov Model to reconstruct the road-based trajectory; spatial coverage of trajectories spanning across western U.S.; penetration rates of vehicle probes computed at locations with radar data; heat map of waypoints.

Several interactive online animations and visualizations were created to allow transportation agencies to visually explore the data and understand the potential use of GPS trajectories in planning as well as measuring performance and mobility. The interactive tools were shared with various agencies, developers of visual analytics, and data providers. They can be accessed via the following link: <https://markoviclab.civil.utah.edu/>

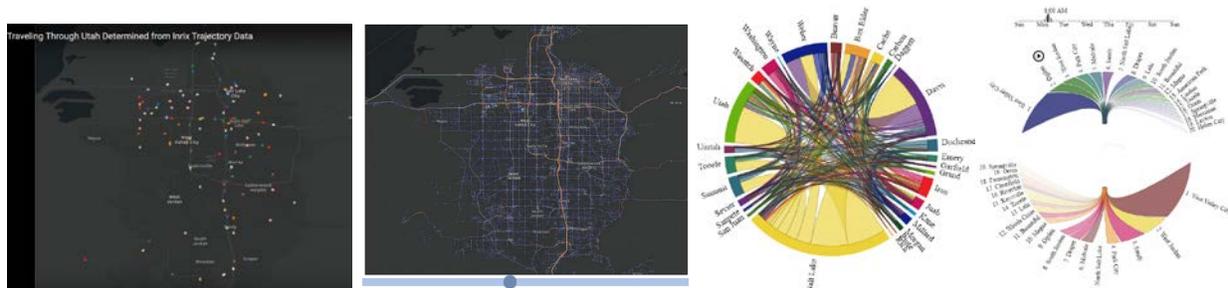


Figure 2: Interactive animations and visualizations facilitate understanding of trajectory data as well as exploration of trip patterns at different levels of spatial and temporal aggregation.

1.0 INTRODUCTION

Transportation agencies have an opportunity to leverage massive GPS trajectory data to support their analyses and enhance their decision making. However, these data are typically purchased from vendors, which means that transportation agencies must understand their benefits before they decide to spend tens of thousands of dollars on data acquisition. To help transportation agencies better understand the tremendous value of trajectory data, we acquired Utah GPS trajectory data from one of the leading GPS data providers and conducted visual analyses demonstrating applications of trajectory data in various areas of transportation. This report will describe the processing that was conducted in order to correct the errors inherent to GPS technology and match trajectory data with the road network. The subsequent analyses will focus on evaluating the data via its comparison with traffic counts from 400 point detectors. Lastly, we will present the developed visual analytics tools that were shared with transportation agencies to help them explore the data and understand its value in planning as well as measuring performance and mobility.

The material summarized in this report stems from a couple of poster presentations as well as a working journal article titled:

- *“Scaling statewide GPS trajectory to match point detector data”* by Seth Miller, Yinhu Wang, Zachary Vander Laan and Nikola Markovic. Presented at the 2019 INFORMS Annual Meeting, Seattle, WA;
- *“Enriching and scaling statewide GPS trajectories”* by Seth Miller, Yinhu Wang, Nikola Markovic, Mark Franz, Abbas Rashidi. Presented at the 2019 UDOT Annual Conference, Salt Lake City, UT;
- *“Enriching and scaling statewide GPS trajectories to match point detector data: Map segmentation for the least-square model”* by Seth Miller, Yinhu Wang, Zachary Vander Laan and Nikola Markovic. In preparation.

Before we proceed with the data description and development of the visual analytics tools, we will review various applications of trajectory data to further emphasize its potential to enhance transportation system analyses.

2.0 LITERATURE REVIEW

In the past several decades, most of the research about transportation planning and traffic operation relied on travel surveys or traffic counts collected from traffic sensor devices (Cascetta, 1984; Yang et al., 1992; Zhou, Qin and Mahmassani, 2003; Yuan et al., 2014). Such approaches could collect valid data for solving traffic problems. Meanwhile, they also require human labor and frequent maintenance of facilities, which cost time and

money. Nowadays, our mobile phones and most cars are equipped with location-based devices, which create large amounts of trajectory data. The availability of a great amount of trajectory data makes itself widely used in traffic applications.

In the process of transportation planning, it used to be a challenge to obtain the spatial-temporal distribution of people's trips. However, the emergence of big data makes this work rather simple. Research related to uncovering human mobility, estimating origin-destination (OD) matrices and identifying traffic behaviors have been done based on GPS trajectory data (Tang et al., 2015; Toole et al., 2015; Siła-Nowicka et al., 2016; Ge and Fukuda, 2016; Tang et al., 2016). At the level of traffic operations, GPS trajectory data can be used for travel time prediction, traffic state estimation and route choice decisions (Casello and Usyukov, 2014; Wang, Goodchild and McCormack, 2017; Deng et al., 2015; Hess et al., 2015; Hiribarren and Herrera, 2014; Zimmermann, Mai and Frejinger, 2017). Besides, GPS trajectory data has been exploited to estimate vehicle miles travelled and traffic emissions (Li et al., 2017; Sun et al., 2015; Liu et al., 2013; Kan et al., 2018), which are very relevant to transportation agencies.

The literature reviewed above indicates that GPS trajectory data has numerous applications in the field of transportation. However, trajectory data represents only a sample of the overall traffic since not all vehicles are equipped with GPS devices. To address this issue, a few methods were proposed to scale up the number of trips extracted from GPS data. Iqbal et al. (2014) applied a least-square regression method to scale up a transient OD matrix obtained from mobile phone call records for the best match with the observed traffic counts. Based on GPS location data, Yang, Lu and Hao (2017) proposed two OD flow estimation models. The first model used a scaled probe vehicle OD matrix as a prior OD matrix and adopted a conventional generalized least squares framework to conduct an OD correction using link counts.

Our ongoing work is concerned with the development of an enhanced optimization framework to scale GPS trajectories to match point detector data. The ongoing work adopts the least square model and proposes a custom map segmentation for enhancing its performance. The expected contributions of this ongoing work include a) development of a novel method for scaling GPS trajectories to match the population and b) the first work combining the least square model with map segmentation, which effectively advances its performance.

3.0 DATA

The acquired raw data include 2.5 million GPS trajectories recorded in Utah during September 2018. A trajectory is defined with an origin, a number of intermediate vehicle locations called waypoints, and a destination (Figure 3 left). Each location has the corresponding time stamp with a one-second precision. Moreover, the data provider specifies the type of vehicle that produced each trajectory: a vehicle under 14,000 lbs, between 14,000 and 26,000 lbs, and above 26,000 lbs. Figure 3 (right) indicates greater

representation of trucks than passenger cars. In total, the raw GPS data include 130 million waypoints that take 12 GB of storage space. This implies about 50 waypoints per trip, with a median sampling rate of about 50 seconds (Figure 3 middle).

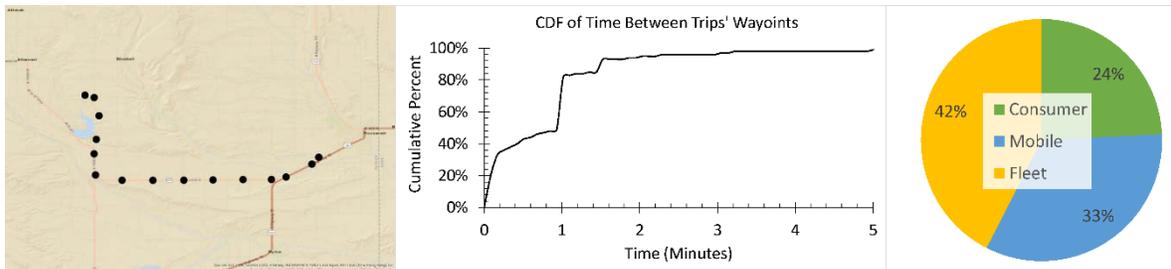


Figure 3: A sample GPS trajectory (left), sampling rate of waypoints (middle), and types of vehicles producing data (right)

3.1 PROCESSING

Raw GPS data contain issues that need to be addressed before further analysis. For example, a common issue is the loss of signal (e.g., due to a tunnel or canyon), which results in resetting the trip counter/ID and generating a new “trip” with an origin at the location where the signal is recovered. This yields artificial trips that may obscure further analysis, such as derivation of OD tables. This issue is resolved by concatenating consecutive trips of the same device ID whenever the average speed between the destination of the first trip and origin of the subsequent trip is greater than 10 km/hr. About 250,000 trips are concatenated through this processing, resulting in about 2.3 million “actual” trips (see Figure 4 right) for sensitivity analysis explaining why 10 km/hr threshold is selected.



Figure 4: An example of GPS signal loss inflating the number of trips (left) and the trip after concatenation (middle). The number of concatenated trips vs. speed threshold, indicates an inflection point at about 3 km/hr (right). The threshold was set to a conservative value of 10 km/hr.

Due to measurement errors inherent to GPS technology, the recorded waypoints are not necessarily located on the physical road network. This issue is exacerbated with common rounding of latitudes and longitudes to the fourth decimal number, which introduces additional error of about 11 minutes. Moreover, the median sampling frequency of 50 seconds is not high enough to include a waypoint along every road link that a vehicle traversed, which produces gaps in road-based trajectories and makes counting of trips

along links a challenge. To overcome these issues, a computationally intense map matching is carried out with the OpenStreetMap routing tool which applies a Hidden Markov model to reconstruct the most likely road-based route from a time-stamped sequence of latitude/longitude pairs. The described processing results in an increased number of waypoints (Figure 5) whose number grows from 130 to 970 million, taking 90 GB of storage.

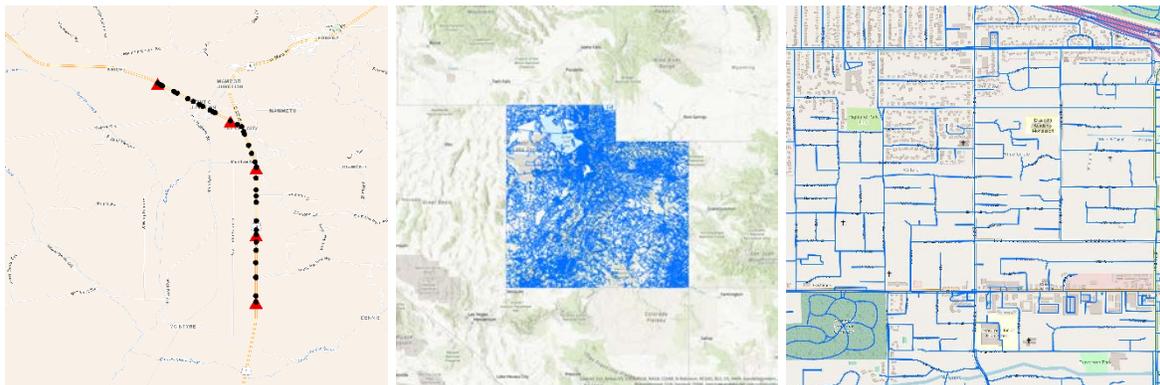


Figure 5: Example of trajectory map matching (left), where triangles indicate recorded locations. The considered road network at the state (middle) at micro level (right).

The above procedure matches waypoints with the OSM road network, which includes valuable information about the road infrastructure such as number of lanes, speed limits, whether opposite directions are physically separated, and others. However, state DOTs typically use a different road network segmentation to measure performance and report to the federal government. In order to make the analysis compatible with local performance measuring efforts, the OSM road segments were conflated both to the NPMRDS and HPMS road networks. This also enables us to leverage additional information contained in the corresponding GIS shapefiles, such as functional road classes, average annual daily traffic (AADT), and others.

3.2 FINAL DATASET AND PENETRATION RATES

To efficiently query the large data we created a SQLite database, optimized for different spatio-temporal queries via a k-d tree and other indexing techniques. The final dataset is presented in Figure 6 (left and middle), which shows spatial density of the trajectories as well as relative distribution of trips between different states. Lastly, since probe vehicles producing trajectory data represent a subset of the overall traffic, it is instructive to quantify their penetration rates via comparison with traffic counts from 400 radar stations in Utah. The average penetration rates vary from less than 1% to over 10% across the 400 locations, with a median of 2% (Figure 6 right).

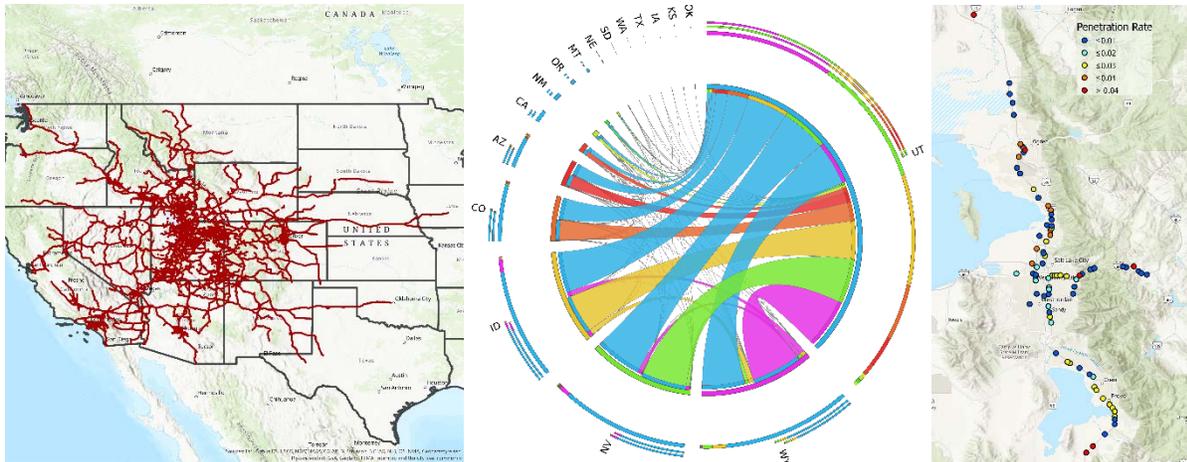


Figure 6: Trajectories spanning across the western U.S. and beyond (left). Distribution of interstate trips (middle) shows that most of these trips take place between Utah and four neighboring states. Comparison with ATR data in Utah indicates probe vehicle penetration rate of about 2% (right).

4.0 VISUAL DATA EXPLORATION

We have prepared a number of interactive online visualizations in order to enable UDOT and other transportation agencies to explore and better understand the data and potential uses. For starters, we developed an animation that plots trips recorded during a 30-minute interval in the Salt Lake City region (Figure 7). This animation enables practitioners to understand that data and sample size.

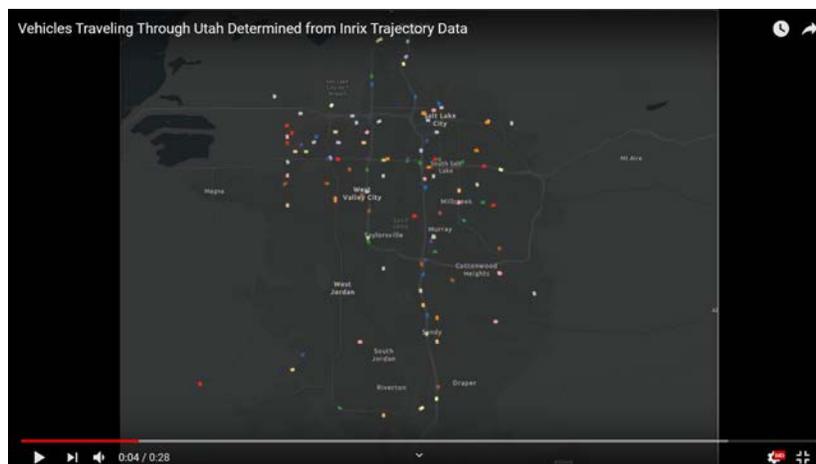
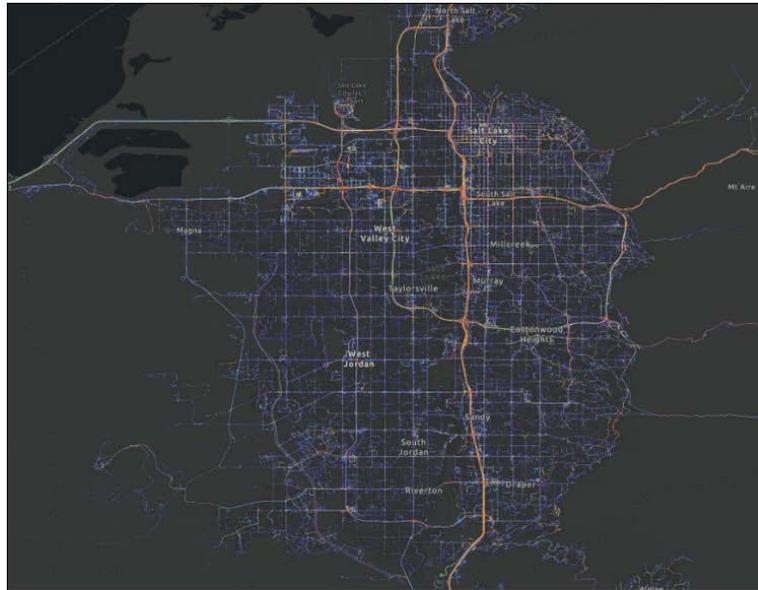


Figure 7: Animation showing trips recorded during a 30-minute interval in the Salt Lake City metropolitan area. The animation can be further explored via this link <https://markoviclab.civil.utah.edu/vis-travel/>

Additional insight into data coverage is provided in Figure 8. This map shows the density of over 900 million waypoints (i.e., vehicle GPS locations) that were recorded over the observed one-month period for which trajectory data were purchased. The colors change from dark to light as densities increase. Densities for different time points can be explored by dragging the button below the video.



11:00 AM

Figure 8: Densities of vehicle GPS locations in the Salt Lake City metropolitan area for different times of the day. The animation can be further explored via this link: <https://markoviciab.civil.utah.edu/vis-trafficdensity/>

Exploration of trip origin-destination patterns is crucial to applications related to transportation planning. To demonstrate simple use of trajectory data to this end, we have developed interactive chord diagrams to enable transportation agencies to easily explore trip patterns between counties in Utah. By hovering a cursor over a specific county, the animation will emphasize trips between the selected county and the remaining counties.

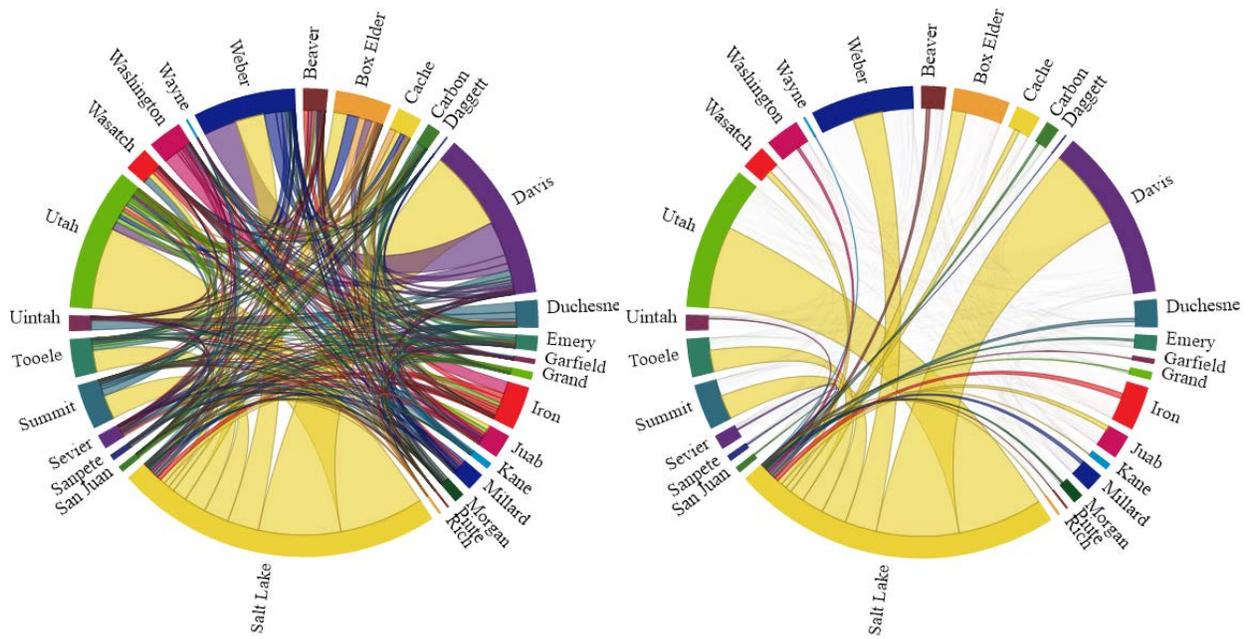


Figure 9: Chord diagrams enable intuitive visual exploration of origin-destination patterns extracted from trajectory data. By hovering the cursor over an individual geography, the animation emphasizes corresponding trips. An example includes trips between counties in Utah, with the second plot emphasizing trips to/from Salt Lake County. The animation can be further explored via this link: <https://markoviclab.civil.utah.edu/vis-circos/>

The presented chord diagrams provide very good insights into travel patterns between multiple selected geographies. However, the main downside is that chords show trips taking place in both directions. In order to facilitate directional exploration of trips, another animation was developed to show trips both starting and ending in the selected geographic region (Figure 9 right).

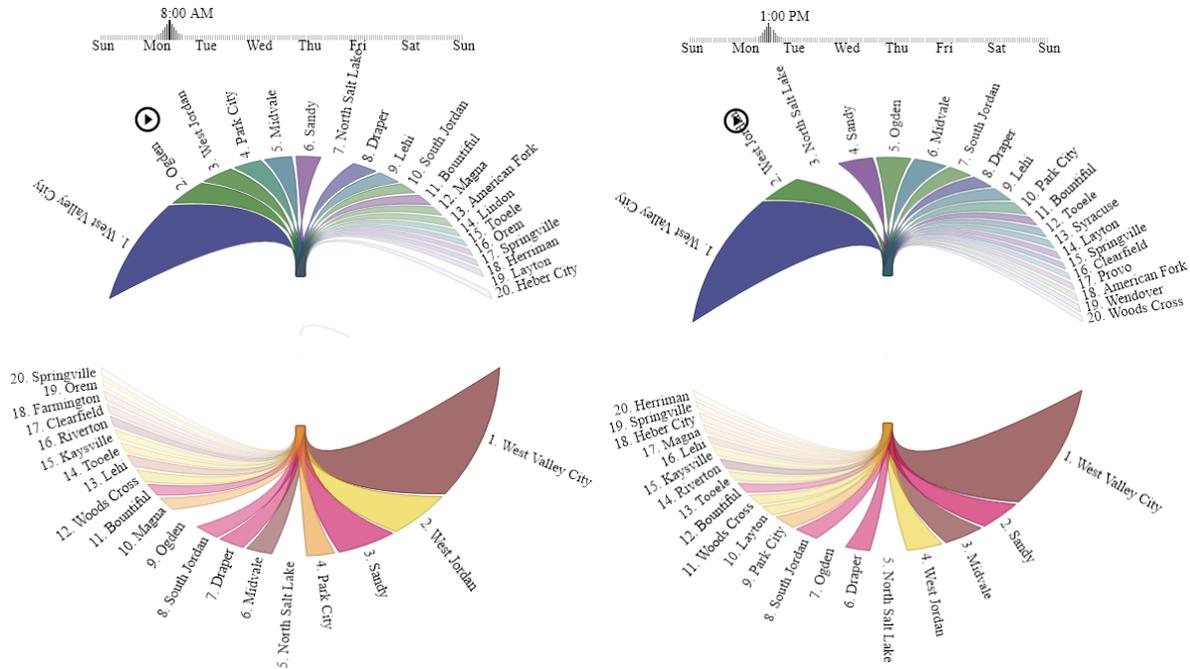


Figure 10: This visualization shows the trip counts that start or end in Salt Lake City in a single hour. The top shows trips originating from Salt Lake City and ending elsewhere and the bottom shows the trips originating from elsewhere and ending in Salt Lake City. To change the day of the week, click the days on the timeline. This animation can be further explored via this link: <https://markoviclab.civil.utah.edu/vis-enterexit/>

5.0 CONCLUSIONS

As part of this project we have acquired one month of Utah GPS trajectory data from one of the leading data providers and conducted analysis to help transportation agencies understand the tremendous value of this data source. First, we have showcased the computationally intense data processing that is needed to correct for any errors that are inherent to GPS data. Second, we have evaluated the data's spatial coverage via comparison with ground-truth traffic information from 400 radar locations, which showed a satisfactory capture rate of about 2%. Lastly, a number of interactive animations and visualizations were prepared and shared with several transportation authorities. These interactive tools facilitate understanding of the trajectory data as well as its application potential in planning as well as measuring mobility and performance. As of today, we are exploring opportunities for further acquisitions of Utah trajectory data and their use to support analysis and decision making of local transportation agencies.

6.0 REFERENCES

- Cascetta, Ennio. 1984. "Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator." *Transportation Research Part B: Methodological* 18 (4-5): 289–299.
- Casello, Jeffrey M, and Vladimir Usyukov. 2014. "Modeling cyclists' route choice based on GPS data." *Transportation Research Record* 2430 (1): 155–161.
- Deng, Bin, Steve Denman, Vassilis Zachariadis, and Ying Jin. 2015. "Estimating traffic delays and network speeds from low frequency GPS taxis traces for urban transport modelling." *European Journal of Transport and Infrastructure Research* 15 (4).
- Ge, Qian, and Daisuke Fukuda. 2016. "Updating origin–destination matrices with aggregated data of GPS traces." *Transportation Research Part C: Emerging Technologies* 69:291–312.
- Hess, Stephane, Mohammed Quddus, Nadine Rieser-Schüssler, and Andrew Daly. 2015. "Developing advanced route choice models for heavy goods vehicles using GPS data." *Transportation Research Part E: Logistics and Transportation Review* 77:29–44.
- Hiribarren, Gabriel, and Juan Carlos Herrera. 2014. "Real time traffic states estimation on arterials based on trajectory data." *Transportation Research Part B: Methodological* 69:19–30.
- Iqbal, Md Shahadat, Charisma F Choudhury, Pu Wang, and Marta C González. 2014. "Development of origin–destination matrices using mobile phone call data." *Transportation Research Part C: Emerging Technologies* 40:63–74.
- Kan, Zihan, Luliang Tang, Mei-Po Kwan, and Xia Zhang. 2018. "Estimating vehicle fuel consumption and emissions using GPS big data." *International journal of environmental research and public health* 15 (4): 566.
- Li, Meng, Yinghao Jia, Zuojun Shen, and Fang He. 2017. "Improving the electrification rate of the vehicle miles traveled in Beijing: A data-driven approach." *Transportation Research Part A: Policy and Practice* 97:106–120.
- Liu, Haobing, Xiaohong Chen, Yuqin Wang, and Shu Han. 2013. "Vehicle emission and near-road air quality modeling for shanghai, china: Based on global positioning system data from taxis and revised moves emission inventory." *Transportation Research Record* 2340 (1): 38–48.2

- Sila-Nowicka, Katarzyna, Jan Vandrol, Taylor Oshan, Jed A Long, Urška Demšar, and A Stewart Fotheringham. 2016. "Analysis of human mobility patterns from GPS trajectories and contextual information." *International Journal of Geographical Information Science*30 (5): 881–906.
- Sun, Zhanbo, Peng Hao, Xuegang Jeff Ban, and Diange Yang. 2015. "Trajectory-based vehicle energy/emissions estimation for signalized arterials using mobile sensing data." *Transportation Research Part D: Transport and Environment*34:27–40.
- Tang, Jinjun, Han Jiang, Zhibin Li, Meng Li, Fang Liu, and Yinhai Wang.2016. "A two-layer model for taxi customer searching behaviors using GPS trajectory data." *IEEE Transactions on Intelligent Transportation Systems*17 (11): 3318–3324.
- Tang, Jinjun, Fang Liu, Yinhai Wang, and Hua Wang. 2015. "Uncovering urban human mobility from large scale taxi GPS data." *Physica A: Statistical Mechanics and its Applications*438:140–153.
- Toole, Jameson L, Serdar Colak, Bradley Sturt, Lauren P Alexander, Alexandre Evsukoff, and Marta C González. 2015. "The path most traveled: Travel demand estimation using big data resources." *Transportation Research Part C: Emerging Technologies*58:162–177.
- Wang, Zun, Anne Goodchild, and Edward McCormack. 2017. "A methodology for forecasting freeway travel time reliability using GPS data." *Transportation research procedia*25:842–852.
- Yang, Hai, Tsuna Sasaki, Yasunori Iida, and Yasuo Asakura. 1992. "Estimation of origin-destination matrices from link traffic counts on congested networks." *Transportation Research Part B: Methodological*26 (6): 417–434.
- Yang, Xianfeng, Yang Lu, and Wei Hao. 2017. "Origin-destination estimation using probe vehicle trajectory and link counts." *Journal of Advanced Transportation*2017.
- Yuan, Yufei, Hans Van Lint, Femke Van Wageningen-Kessels, and Serge Hoogen-doorn. 2014. "Network-wide traffic state estimation using loop detector and floating car data." *Journal of Intelligent Transportation Systems*18 (1): 41–50.
- Zhou, Xuesong, Xiao Qin, and Hani S Mahmassani. 2003. "Dynamic origin-destination demand estimation with multiday link traffic counts for planning applications." *Transportation Research Record*1831 (1): 30–38.
- Zimmermann, Maëlle, Tien Mai, and Emma Frejinger. 2017. "Bike route choice modelling using GPS data without choice sets of paths." *Transportation research part C: emerging technologies*75:183–196.