Visualizing Transportation Happiness in the Minneapolis-St. Paul Region

Methodology Report

Prepared by:
Yingling Fan
Humphrey School of Public Affairs
University of Minnesota

Travis Ormsby & Peter Wiringa
U-Spatial
University of Minnesota

Chen-Fu Liao
Department of Mechanical Engineering
University of Minnesota

Julian Wolfson
School of Public Health
University of Minnesota

January 8, 2020

CENTER FOR
TRANSPORTATION STUDIES
UNIVERSITY OF MINNESOTA

Overview

This report describes the data and methods used to generate the interactive Minneapolis-St. Paul Transportation Happiness Map at http://maps.umn.edu/transportation-happiness. The map illustrates spatiotemporal differences in travelers' happiness ratings on the streets and roads in the Minneapolis-St. Paul metropolitan region. For policy makers who are interested in improving people's transportation happiness, the map provides important insights on road and street segments that are in need of closer investigations for future improvements.

Figure 1 is the main user interface of the interactive transportation happiness map. Users can select among 10 travel mode(s) options (including the all modes option), 7 time(s) options (including the all times option), 8 emotion options, and two statistic options (mean and median).

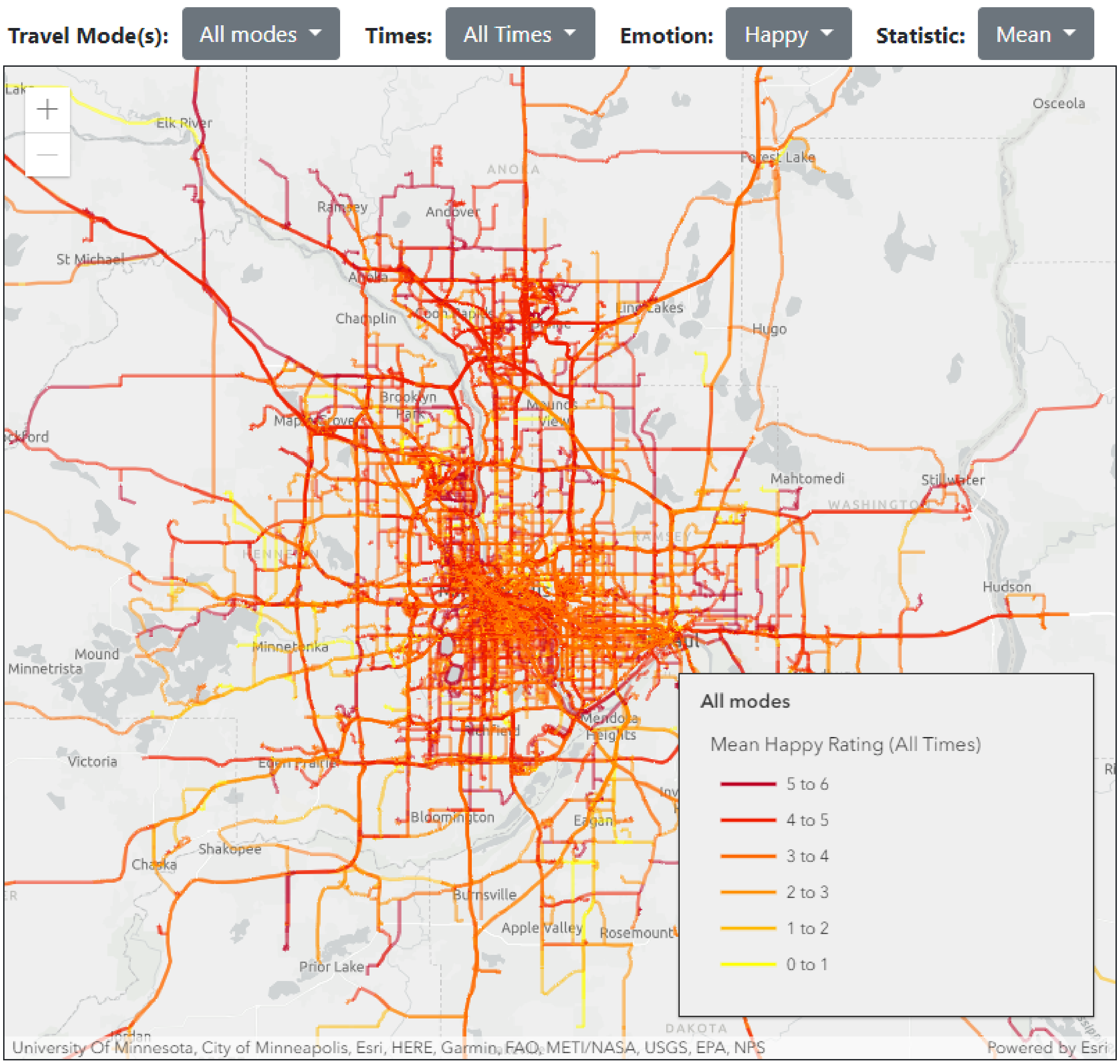
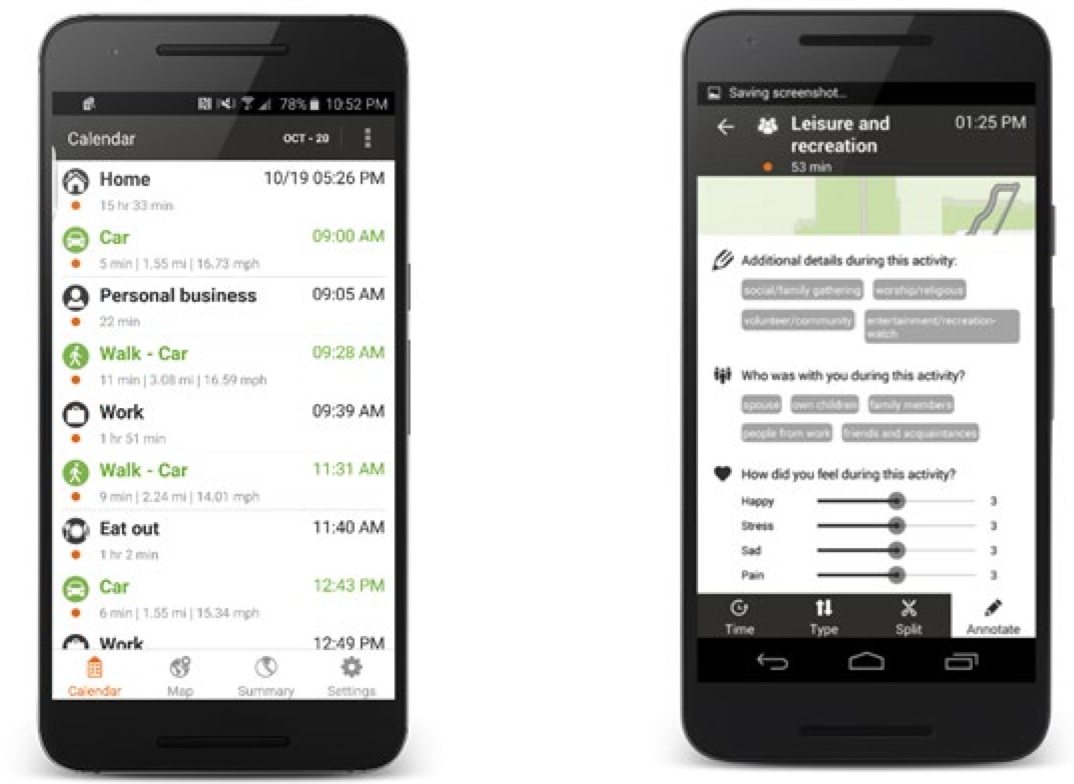


Figure 1. Screenshot showing main user interface of the transportation happiness map

Data

The data used for mapping transportation happiness come from the Minneapolis-St. Paul Metro Area Subjective Well-Being Data: 2016 – 2017 (Fan et al., 2020). The data was collected using an Android-based smartphone application-Daynamica, previously SmarTrAC (Fan et al., 2015). Daynamica is capable of detecting activities and trips in real time to construct sequenced activity/trip episodes throughout the day (Figure 2a). It also allows the user to annotate the detected activities/trips with additional information such as companionship and emotional experiences during each activity/trip at their convenience (Figure 2b).



(a) Daynamica constructs sequenced activity and travel episodes in real time throughout the day.

(b) Daynamica enables user input on each activity or trip episode at their convenience.

Figure 2. Daynamica main interface (Fan, Brown, Das, & Wolfson, 2019)

Using the Daynamica app, the survey successfully recruited 398 residents from six neighborhoods in the Minneapolis-St. Paul region, including four urban and two suburban areas (Figure 3). The recruitment was based upon geographic cluster sampling: random blocks were selected within each neighborhood, and efforts were made to recruit as many households as possible from each block. As shown in Table 1, the distribution of participants across neighborhoods is relatively even, ranging from 55 to 79 participants per neighborhood. Two third of the participants were female, indicating overrepresentation of female participants in the sample.

Table 1. Participants by Gender and Neighborhood in the Minneapolis-St. Paul Region

Neighborhoods	Location	Income Level	Rail Access	Male	Female	Total
Phillips	Urban	Low	Yes	27	48	75
Near North	Urban	Low	No	16	46	62
Prospect Park	Urban	High	Yes	25	43	70
St. Anthony Park	Urban	High	No	24	55	79
Brooklyn Center	Suburban	Low	No	18	39	57
Blaine	Suburban	High	No	18	37	55
			Total	128	268	398

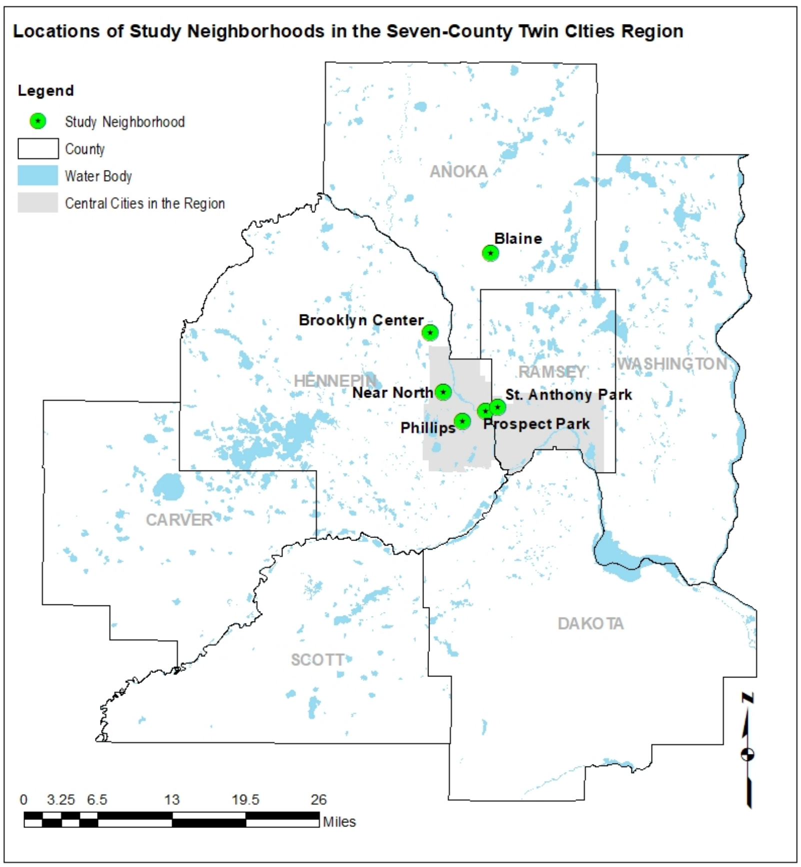


Figure 3. Locations of Study Neighborhoods

Each participant carried a smartphone equipped with the Daynamica app for seven consecutive days. In addition, home-based entry/exit surveys were conducted for each participant to collect individual socio-demographic data as well additional individual-level data including personality attributes, life evaluation, and satisfaction with various aspects of life including work, health, leisure, home, and the neighborhood. Figure 4 below illustrates a participant's detailed activity and trip data along with self-reported happiness ratings over seven consecutive days that were collected by the Daynamica app.

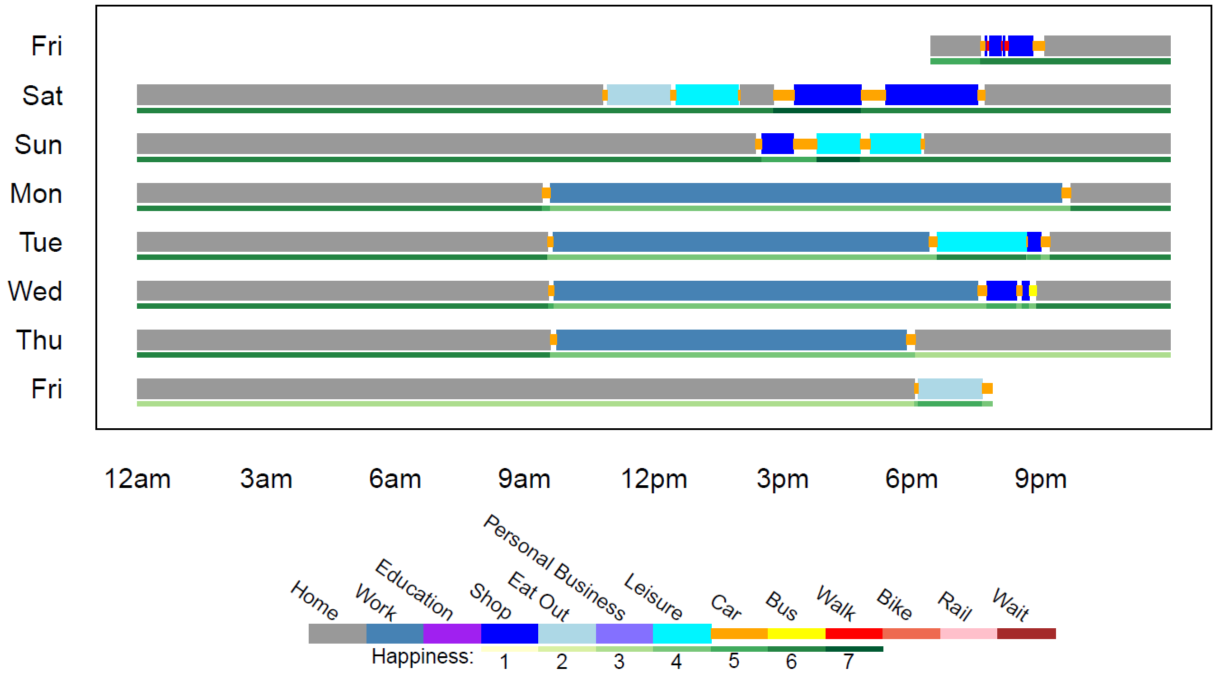


Figure 4. Illustration of 7-day activity-trip data with happiness information

To summarize, the following data were generated in 2016-2017 Minneapolis-St. Paul Well-Being survey efforts:

- Episode-level activity-travel data over seven consecutive days for each participant, including spatiotemporal details, type, companionship, and happiness ratings for each activity/trip episode. There are a total of 11,182 activity episodes and 10,968 trip episodes in the dataset.
- Individual-level data on socio-demographics, personality, life evaluation, and perceive satisfaction with various aspects of life including work, health, leisure, home, and the neighborhood.

The Minneapolis-St. Paul Transportation Happiness Map utilizes the spatiotemporal details, travel mode information, and happiness ratings associated with the 10,968 trip episodes in the dataset. Note that each trip may contain multiple transportation modes. The data used to visualize the transportation happiness map include a total of 13,924 mode segments with happiness rating information.

Tables 2 and 3 respectively illustrate the point-level and trip segment-level data records that we used to generate the Minneapolis-St. Paul Transportation Happiness Map. Using the smartphone-based location sensing technology, the Daynamica app records location time series of the phone. The app further segments location time series into activity vs. trip segments based upon self-designed algorithms. As a result, each activity/trip segment captured by the app contain information on all point locations associated with the segment. The Location Timestamp column in Table 2 presents information on the exact time in milliseconds since the Unix epoch when a specific activity/trip segment passes through the specific location. The Segment ID column presents information on the ID of the activity/trip segment with which this point location is associated. This Trip Segment ID allows us to match segment-level happiness ratings (as shown in Table 3) to each point location in the trip segment. As shown in Table 2, each location in the dataset come with speed and accuracy information that was directly generated by the smartphone's location sensors.

Table 2. Examples of raw data records at the point level

Index	Segment ID	Location Timestamp	Latitude	Longitude	Speed	Accuracy	Geometry
411706	15086	1481050134234	44.9495631	-93.2627	1.016949	7	POINT (-93.2627403 44.9495631)
411707	15086	1481050129233	44.9496088	-93.2627	5.058428	5	POINT (-93.2627369 44.9496088)
411708	15086	1481050124227	44.9498365	-93.2627	12.0558	7	POINT (-93.2627490 44.9498365)
411709	15086	1481050119236	44.9503779	-93.2627	15.61236	6	POINT (-93.2627399 44.9503779)
411710	15086	1481050114228	44.9510812	-93.2627	12.31548	7	POINT (-93.2627132 44.9510812)

Table 3 illustrate the segment-level data record. As mentioned earlier, the Daynamica app breaks location time series into activity vs. trip segments. These segments construct the activity and trip episodes shown in Figure 2(a). In general, an activity episode contains a single activity segment record, while a trip episode may contain more than one trip segment records. As shown in Table 3, the segments #9198 and #9199 are trip segments with different transportation modes: car and walk. As each trip departs from the previous activity location to the next activity location, consecutive trip segments in the data represent that multiple transportation modes were used to travel from the previous activity location to the next activity location. The Start Timestamp and End Timestamp columns in Table 3 are the start and end time of each segment in milliseconds since the Unix epoch.

Table 3. Examples of raw data records at the activity/trip segment level

Segment ID	Person ID	Start Timestamp	End Timestamp	Segment Type	Subtype	Happy	Tired	Stress	Sad	Pain	Meaningful
9197	1007	1478192370002	1478194440018	ACTIVITY	HOME	5	6	2	1	1	4
9198	1007	1478194440018	1478194740006	TRIP	CAR	5	3	3	1	1	3
9199	1007	1478194740006	1478195610000	TRIP	WALK	5	3	3	1	1	3
9200	1007	1478195610000	1478205870001	ACTIVITY	EDUCATION	4	3	4	1	1	4
9201	1007	1478205870001	1478206560005	TRIP	WALK	3	2	3	1	1	3
9202	1007	1478206560005	1478208210002	TRIP	CAR	3	2	3	1	1	3
9203	1007	1478208210002	1478218320010	ACTIVITY	PERSONAL	6	3	3	1	1	6
9204	1007	1478218320010	1478220090000	TRIP	CAR	6	4	2	1	1	6
9205	1007	1478220090000	1478261250001	ACTIVITY	HOME	7	5	2	1	1	4

The happiness ratings were collected via the annotation function in the Daynamica app. Each app user was asked to answer questions about their emotional experiences associated with each activity/trip. The questions include a total six different types of emotions: happy, meaningful, sad, painful, tired, and stressful. All emotion variables are scaled 0–6. Take Happy as an example, the original question was, "From 0 to 6, where a 0 means you were not happy at all and a 6 means you were very happy, how happy did you feel during this time?" Note that information on emotions was collected at the episode level for each activity/trip. For a trip with multiple mode segments, these segments have the same data value for each emotion variable. For example, segments # 9198 and #9199 in Table 3 have the same values for emotion variables because they belong to the same trip. Note that only trip-related data were used in this project.

Method

We use the Open Source Routing Machine (OSRM) version 5.22 to match the trip-related points to streets and roads in the Minneapolis-St. Paul region. We created the necessary Open Source Routing Matching (OSRM) files from the March 20, 2017 planet extract of OpenStreetMap (OSM) data, for a bounding box covering the Minneapolis-St. Paul Metro area. A Python script was developed to feed consecutive trip-related GPS points with an accuracy better than 25 meters through the OSRM API (Project-osrm.org, 2019), which has been made available for download at https://z.umn.edu/hrmm.

Specifically, the Python script was run separately for each of the three OSRM profiles:

- The driving profile was used for points associated with three trips modes: 1 (car), 2 (bus), and 7 (in vehicle).
- The walking profile was used for pointed associated with trip mode 3 (walk).
- The biking profile was used for points associated with trip mode 4 (bike).

Points associated with trip mode 5 (rail) were not matched because there is no OSRM profile for rail travel. Points associated with trip mode 6 (other) were not matched because there is insufficient information available to determine the appropriate OSRM profile to use. For modes 1, 2, 3, 4, and 7, this Python script was able to output each individually matched trip segment as a feature in a shapefile.

Not all points associated with a trip segment ID have high levels of accuracy. We decided to only match points with an accuracy value lower than 25 meters to a street or road. As shown in Table 2, each point has an accuracy value measured in meters. The lower the accuracy value, the higher the accuracy level. According to Developer.android.com (2020), if the accuracy value is 25, then there is a 68% chance the true location of the device is within 25 meters of the reported coordinates. For trip segments that contain points with accuracy values higher than 25, the matching process returned multiple features with the same trip segment ID for the portions that could be matched. Using ArcGIS Pro version 2.4, the features in each of the output shapefiles were dissolved by trip segment ID, so features associated with the same trip segment ID are consisted of a single polyline. These feature classes were then merged into a single feature class with all matched trip segments for every mode. The features were then projected into the WGS 1984 UTM Zone 15 N (EPSG:32615) coordinate reference system, so that further distance-based analysis could use Cartesian, instead of geodetic, distance calculations.

To create the rail segments, the Merge Divided Roads tool in ArcGIS Pro was used to generalize the parallel tracks of the METRO Blue and Green lines into a single track. The rail lines were segmented with breaks at station locations, and midpoints were calculated. This was done because the GPS data are not accurate enough to match to a particular direction without using the

OSRM algorithm. Using a one-to-many intersection spatial join, every rail trip within 50 meters of a rail line was matched to the midpoint of the rail network segment it passed near.

Since not all GPS traces could be matched to a route, there was data loss with the map matching process. Table 4 summarizes the final route matched data by trip mode in comparison to the raw GPS data. A total of 12,544 trip mode segments were finally matched to 77,063 street/road segments in the seven-county Minneapolis-St. Paul region.

Table 4. Data loss during the route matching process

	Raw GPS data	**Route Matched data**
Total Single-Mode Trip Segments	13,924	12,544
Driving	7,324	6,896
Bus	626	597
Walking	3,608	3,138
Biking	520	508
Rail	220	190
Other	316	0
Vehicle	1,310	1,215

Note: Data loss was due to either discarded because the GPS accuracy score was higher than 25 meters, or lost because the segment could not be matched by OSRM (or spatial join to the rail network).

In addition to visualizing transportation happiness of each of the six specific modes (driving, bus, walking, biking, rail, and vehicle), the interactive Minneapolis-St. Paul Transportation Happiness Map allows visualization for four different mode combinations: Driving & In Vehicle, Biking & Walking, Bus & Rail, and All Modes.

To create street segments with aggregated happiness scores from multiple modes, the street segments with matched single-mode happiness scores were split at every intersection using ArcGIS Pro's Planarize tool regardless of whether the street segments were roads, bike paths, or footpaths. The linear features in OpenStreetMap data ("ways") may span multiple blocks and it is possible for a trip to traverse only a portion of the way; ways were planarized to construct approximate minimum travel units, e.g. a city block. Trip geometries were then matched to the planarized segments.

A new feature class representing the midpoints of each of these street/road segments was then created to facilitate the spatial joins in the next step. The midpoint-based spatial joining is to ensure that trips were not matched to multiple street/road segments at intersections. Using a one-to-many intersection-type spatial join with a search radius of 0.1 meters, every mode-specific segment was matched to the midpoint of every street/road segment it passed through. As part of the join process, the mean and median of each emotion variable, total trips, and count of unique users for trips that matched to the midpoint were calculated.

To allow users to explore the mode-specific and time varying nature of transportation happiness, we not only calculated the mean and median of each emotion variable of all trips matched to a single segment, but also calculated the means and medians among trips of different time bucket categories and different travel mode categories. In the end, given the mode/time/emotion/statistic variations, the Minneapolis-St. Paul Transportation Happiness Map allows a total of 1,120 combinations of mapping options (10 mode options * 7 time options *8 emotion options * 2 statistic options). These options are listed in Table 4 below.

Table 4. Available Mapping Options at the Minneapolis-St. Paul Transportation Happiness Map

Mode options (10)	Time buckets (7)	Emotion ratings (8)	Statistics (2)
All modes	All times	Happy	Mean
Driving	Weekday AM Rush	Meaningful	Median
Bus	Weekday PM Rush	Sad	
Walking	Weekday AM + PM Rush	Pain	
Biking	Weekday non-Rush	Tired	
Rail	Weekday All Times	Stress	
In vehicle	Weekend All Times	Net affect	
Driving + In vehicle		Peak affect	
Walking + Biking			
Bus + Rail			

The net affect and peak affect values were calculated as below.

- Net Affect = mean (happy, meaningful) - mean (sad, pain, tired, stress).
- Peak Affect = max (happy, meaningful) - max (sad, pain, tired, stress).

Limitations

The map matching algorithm represents a baseline estimate. The GPS data collected by smart phones is not always accurate enough to match to a particular route. While the vast majority of these cases are rejected by the algorithm (i.e. not matched to a road segment), there are some spurious matches in the dataset.

Map matching is more accurate than snapping GPS points to the nearest segment, especially at intersections, as the process takes into account direction of travel using the GPS trace data. Snapping to the nearest segment at an intersection without consideration for previous and future points (direction) may result in snapping to the incorrect segment. Despite the improvement observed using the map matching process, we cannot be 100% confident in the results. The algorithm does report the confidence level of each match, and the project could be enhanced by taking that confidence level into account at one or more points in the workflow.

Because the algorithm does not contain a profile for rail travel, it was not possible to use OSRM to match rail trips, and we were forced to fall back on snapping to the nearest rail line. Because there are very few intersections on the rail network relative to its length, this is less of a problem than it would be for the road network, but would require manual review, a new OSRM profile, or other steps to establish the direction of rail trips. Given the small number of rail trips in the dataset, the time necessary to create this profile is likely to have a low return on investment. This same limitation also applies to dedicated transit right-of-ways for buses, as the map matching algorithm will attempt to route these trips of those right-of-ways (since there is no profile to differentiate between a bus and a car). There are very few of these dedicated right of ways in the Minneapolis-St. Paul Metro, but this limitation could be more impactful in other locations.

The data are currently hosted on ArcGIS Online. That limits the size of the dataset that can be posted, as well as slowing down the speed at which data move between the server and the client. The current load speeds are tolerably fast, but that could change if the dataset became much larger. Additionally, because most of the graphics rendering takes place in the user's web browser, users with older/slower computers will see relatively poor performance compared to more powerful machines. Performance could be improved by changing the hosting or delivery methods.

Scalability

Within a single geographic region, we expect the map matching process to scale very well as the number of trips increases. However, the map matching process for the Minneapolis-St. Paul region covers a small geographic area.

The quality of the map matching is dependent on the quality of the OpenStreetMap data. Quality is a measure of accuracy and completeness of coverage for the area of interest. For the Minneapolis-St. Paul area we judge OSM data quality to be high; the application may have difficulty handling regions where the quality is lower.

Because the data are aggregated to road segments before hosting them on ArcGIS Online, the only effect of additional trips will come from rendering road segments that were not traversed by previous trips. Because the dataset already covers the most traveled roads, additional trips are unlikely to dramatically increase the number of trip segments the app needs to render.

The scalability is a little worse as the geographic extent of the trips increases. The Python code that implements the algorithm is tuned for the Minnesota dataset, and there will be an additional time cost associated with generalizing the program. Additionally, because the algorithm requires building a graph of all the nodes in the OpenStreetMap data in the relevant area, it is best to keep this area small. For example, if matching trips in both Minneapolis and Chicago, it would be better to build two separate graphs and run the analysis twice, rather than building the graph for a single region that encompasses both cities.

The larger issue with expanding the geographic extent of trips is increasing the number of road segments that need to be stored on the server and rendered in the user's browser. For the current 77,000 road segments, the rendering performs acceptably. Attempting to substantially increase the number of segments being rendered may require technical changes to make the experience for users more palatable.

Concluding Remarks

The Minneapolis-St. Paul Transportation Happiness Map provides mode- and time-specific information on travelers' emotional experiences on the streets and roads in the Minneapolis-St. Paul metropolitan region. Map users can interactively explore street and road segments that are associated with positive and/or negative emotional experiences based upon their interested travel modes and travel time periods. The map is admittedly exploratory in nature because only 398 residents contributed a week of trip data. The driving-based transportation happiness map has the largest coverage in space as the majority of trips in the dataset are driving trips. Overall, biking-based transportation happiness levels are higher than the transportation happiness levels associated with other modes. To provide map users a sense of confidence in the data, the Map provides information on the number of unique residents who travelled along each specific road/street segment and contributed to the transportation happiness data. A closer look at the streets in downtown Minneapolis show that the South 8th Street and South 4th Street are the least happy streets while the South 2nd street and the 3rd Avenue South are the happiest streets.

Acknowledgements

The authors acknowledge the Center for Transportation Studies at the University of Minnesota for providing funding resources that made this online map possible. Data used in this map come from the SRN: Integrated Urban Infrastructure Solutions for Environmentally Sustainable, Healthy and Livable Cities project funded by the National Science Foundation (Award Number: 1444745).

References

Developer.android.com, 2020. Accessed on February 22, 2020 at https://developer.android.com/reference/android/location/Location.html#getAccuracy()

Fan, Y., Wolfson, J., Adomavicius, G., Das, K. V., Khandelwal, Y., & Kang, J. (2015). *SmarTrAC: A Smartphone Solution for Context-Aware Travel and Activity Capturing*. Retrieved from University of Minnesota website: http://www.cts.umn.edu/Publications/ResearchReports/reportdetail.html?id=2424

Fan, Y., Das, K., Ramaswami, A., Cao, J. 2020. *Minneapolis-St. Paul Metro Area Subjective Well-Being Data: 2016-2017*. Sustainable Healthy City Network: University of Minnesota, Minneapolis, Minnesota, USA.

Project-osrm.org, 2019. Open Source Routing Machine (OSRM) API Documentations. Accessed on November 2, 2019 at http://project-osrm.org/docs/v5.5.1/api/#general-options.