Geo-coding Transit Passenger Activity

By: Mike Kowalski

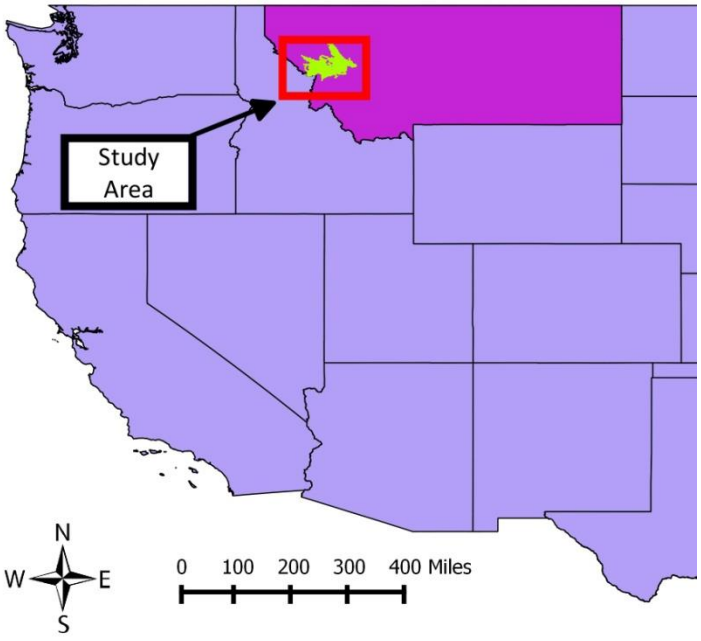
12th of December, 2014

Abstract

Many transit agencies offer on-demand, non-fixed route transit service. This often presents a challenge to analytic tasks that require an easily recognizable location description for bus GPS data once it deviates from a fixed route. Generally speaking, geo-coding, or more specifically, reverse geo-coding of bus GPS data presents a solution to this problem. Reverse geo-coding can assign an intersection, street address and/or other common place name to a set of spatial coordinates.

Introduction

Mountain Line, the public transit agency in Missoula, MT (Figure 1), offers flexible route service where transit vehicles will deviate from fixed routes to make on-demand pickups and drop-offs of passengers. When a bus leaves a fixed route pathway, it becomes difficult to identify where passenger activity takes place. Typically only official bus stops along the fixed route pathway are geo-coded. The geo-coding of bus stops makes it possible for lat/long coordinates, generated by the bus, to be assigned a recognizable location like a stop name.



When passenger activity takes place outside of a fixed route pathway, some sort of physical marker must be used as a location description for the lat/long at which this activity took place (Table 1). One method of generating such physical markers would be identifying each street intersection in the region where the on-demand transit service might operate. Once the intersections are defined, the off route

passenger activity can be associated with the closest intersection to the GPS coordinates where that passenger activity took place.

Another type of physical marker that could be used is a street address. With proper street data and GIS software capability, a street address can be generated for each lat/long point generated by the bus.

Table 1. Raw Bus Data with No Identification by Intersection or Address

LAT	LONG	Bus #	Date	Time	On	Off	???Intersection???	???Address???
46.8494574	-114.1116644	426	11/10/14	10:22:40				
46.84744725	**-114.1087924**	426	11/10/14	10:22:10	1			
46.84756339	-114.1131623	426	11/10/14	10:23:40				
46.85005765	-114.1083824	426	11/10/14	10:23:10				
46.85151256	-114.1020663	426	11/10/14	10:24:40				
46.85596777	-114.0883866	426	11/10/14	10:24:10				
46.85597228	-114.0814195	426	11/10/14	10:25:40				
46.8523438	-114.0814288	426	11/10/14	10:25:10				
46.84874429	-114.0815125	426	11/10/14	10:26:40				
46.8487282	-114.07563	426	11/10/14	10:26:10				
46.84874465	-114.0684443	426	11/10/14	10:27:40				
46.8487291	-114.0555267	426	11/10/14	10:27:10				
46.84874345	-114.0460639	426	11/10/14	10:28:40				
46.84880169	-114.0320998	426	11/10/14	10:28:10				
46.85515013	-114.0281986	426	11/10/14	10:29:40				
46.86248456	-114.0281587	426	11/10/14	10:29:10				
46.86872982	**-114.0223725**	426	11/10/14	10:30:40		1		
46.87289984	-114.0196605	426	11/10/14	10:30:10				
46.87802151	-114.0143324	426	11/10/14	10:31:40				
46.88032221	-114.0138765	426	11/10/14	10:31:10				
46.88099077	-114.0113802	426	11/10/14	10:31:40				

Methods

To identify the intersection of each street at a point, a layer of street data can be intersected onto itself (i.e. the same shape file is loaded twice as the two intersect layers), where the output type is points. The attribute table of the new point layer contains all of the attributes of the two adjacent street line layers. For reasons not quite clear, several points at each intersection are generated where the names of the two adjacent streets are present in different combinations (Figure 2).

Ideally, only a single point would exist at each intersection with the two different names of each adjacent street present. To get to that point, the attribute table for the point layer was exported and loaded into SPSS statistics software. A short script procedure was developed to remove all cases for each point, where duplicate information is present. This data was then loaded back into GIS so that a single point exists for each intersection.

Some hypothetical sample data of off route bus GPS points was used to test out this procedure. It consists of roughly (60) GPS points that would have been taken on board a bus as it left to pick up and drop off a passenger (Figure 3).

FULLROADNA	FULLROAD_1
LINCOLN RD	LINCOLN RD
LINCOLN RD	RATTLESNAKE DR
LINCOLN RD	RATTLESNAKE DR
LINCOLN RD	CREEK CROSSING
RATTLESNAKE DR	LINCOLN RD
RATTLESNAKE DR	RATTLESNAKE DR
RATTLESNAKE DR	RATTLESNAKE DR
RATTLESNAKE DR	CREEK CROSSING
RATTLESNAKE DR	LINCOLN RD
RATTLESNAKE DR	RATTLESNAKE DR
RATTLESNAKE DR	RATTLESNAKE DR
RATTLESNAKE DR	CREEK CROSSING
CREEK CROSSING	LINCOLN RD
CREEK CROSSING	RATTLESNAKE DR
CREEK CROSSING	RATTLESNAKE DR
CREEK CROSSING	CREEK CROSSING

Figure 2. Repetitive combinations of street names in point layer that results from intersection.

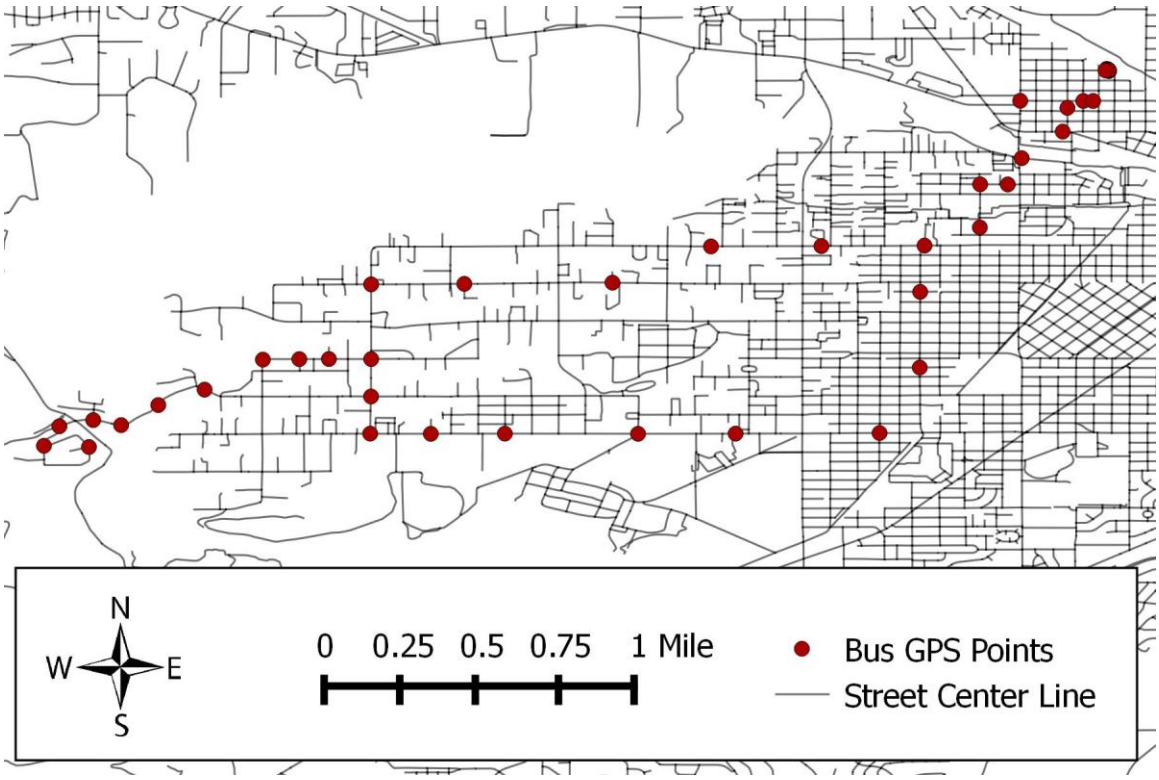


Figure 3. Hypothetical Off-Route Bus Data

The Near function was utilized to determine which intersection was closest to each GPS point generated on board the bus. More specifically, the Near Table function was used to input all (60) points and

generate the closest intersection along with the distance to that intersection. The resulting near table could then be joined back to the original data using SPSS or GIS.

In addition to relating GPS points to the closest street intersection, another potentially useful physical location that could be used is street address. ArcMap has a feature that allows for Reverse Geo-coding. The reference data for reverse geo-coding is street data that contains address information for each street segment. To run the feature, you load in the street data and the GPS points to be reverse geo-coded. As output, an address field is added to the input GPS point data.

Results

The two analyses that identified a street address and the nearest intersection were appended to the raw bus data (Table 2).

Table 2. Raw Bus Data with Intersections and Addresses Identified

LAT	LONG	Bus #	Date	Time	On	Off	Intersection	Address
46.8494574	-114.1116644	426	11/10/14	10:22:40			O'BRIEN CREEK RD & RED HAWK VW	2400 TRIPLE CREEK DR
46.84744725	**-114.1087924**	426	11/10/14	10:22:10	1		DOUBLE TREE LN & MESA CT	8051 MESA CT
46.84756339	-114.1131623	426	11/10/14	10:23:40			DOUBLE TREE LN & TRIPLE CREEK DR	2501 TRIPLE CREEK DR
46.85005765	-114.1083824	426	11/10/14	10:23:10			BLUE HERON LN & RIVER PINES RD	201 BLUE HERON LN
46.85151256	-114.1020663	426	11/10/14	10:24:40			HANSON DR & SOUTH AVE W	144 RIVER PINES RD
46.85596777	-114.0883866	426	11/10/14	10:24:10			MOUNT AVE & PLEASANT AVE	4399 MOUNT AVE
46.85597228	-114.0814195	426	11/10/14	10:25:40			CLEMENTS RD & MOUNT AVE	1598 CLEMENTS RD
46.8523438	-114.0814288	426	11/10/14	10:25:10			CLEMENTS RD & NORTH AVE W	4101 NORTH AVE W
46.84874429	-114.0815125	426	11/10/14	10:26:40			CLEMENTS RD & SOUTH AVE W	4099 SOUTH AVE W
46.8487282	-114.07563	426	11/10/14	10:26:10			SOUTH AVE W & SUNLITE LN	3851 SOUTH AVE W
46.84874465	-114.0684443	426	11/10/14	10:27:40			36TH AVE & SOUTH AVE W	2398 36TH AVE
46.8487291	-114.0555267	426	11/10/14	10:27:10			31ST AVE & GUARDSMAN LN	2399 31ST AVE
46.84874345	-114.0460639	426	11/10/14	10:28:40			SERVICE & SOUTH AVE W	2799 SOUTH AVE W
46.84880169	-114.0320998	426	11/10/14	10:28:10			SCHILLING ST & SOUTH AVE W	2399 SCHILLING ST
46.85515013	-114.0281986	426	11/10/14	10:29:40			BURLINGTON AVE & JOHNSON ST	1701 JOHNSON ST
46.86248456	-114.0281587	426	11/10/14	10:29:10			JOHNSON ST & S 8TH ST W	999 JOHNSON ST
46.86872982	**-114.0223725**	426	11/10/14	10:30:40		1	S 2ND ST W & S CATLIN ST	313 S S CATLIN ST
46.87289984	-114.0196605	426	11/10/14	10:30:10			MONTANA ST & N WASHBURN AVE	1561 MONTANA ST
46.87802151	-114.0143324	426	11/10/14	10:31:40			COWPER ST & W BROADWAY	1313 W W BROADWAY
46.88032221	-114.0138765	426	11/10/14	10:31:10			COWPER ST & SHERWOOD ST	931 COWPER ST
46.88099077	-114.0113802	426	11/10/14	10:31:40			BURTON ST & PHILIPS ST	1435 PHILIPS ST

Conclusion

By identifying points that represent each intersection in a given set of street data, bus GPS points from off route activity can be more easily identified using intersections. Likewise, street data containing necessary address information can be used to identify points by their street address. Further discussion, investigation and feedback will be necessary to determine whether intersection or address would be the best single location to report. It would seems that intersection would be the most useful. However, there may be cases where a GPS point lies far enough away from an intersection that using an intersection as _the location of the point wouldn't be appropriate. In that cases, perhaps street address would serve as a better location identification.

Acknowledgments

County of Missoula GIS Department - Street Centerline Data.