USDOT Region V Regional University Transportation Center Final Report

NEXTRANS Project No.

"Methods for Improving Bicycle Sharing System Balance" (170OSUY2.2)

By

Morton E. O'Kelly
Professor & Chair
Department of Geography
The Ohio State University

okelly.1@osu.edu

DISCLAIMER

Funding for this research was provided by the NEXTRANS Center, Purdue University under Grant No. DTRT12-G-UTC05 of the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (OST-R), University Transportation Centers Program. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the 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.

USDOT Region V Regional University Transportation Center Final Report

TECHNICAL SUMMARY

NEXTRANS Project No. 170OSUY2.2 Final Report, 12/2/2016

Methods for Improving Bicycle Sharing System Balance

Introduction

Many cities have developed bicycle sharing systems. This project points out that a key element in their success and sustainability is an understanding of the spatial patterns of supply and demand. This project tackles the re-balancing issue where bikes must be removed from full stations and replenished at empty ones. Taking inspiration from the lot size problem, and capacitated vehicle routing, the project devises a rolling window approach to this problem, in essence representing the amount of foresight that the system planner can use to determine solutions. The submitted paper shows that a problem which might be unsolvable with a more myopic approach can actually be solved satisfactorily when a longer window range is used. The penalty for this added window (look-ahead) is of course added computational effort but the paper shows that this effort is manageable for the relatively small sample system analyzed here (Columbus, Ohio). The paper reporting these efforts is in review. The paper is illustrated with results from 54 problem instances derived from variants of the base data. The results are derived on a standard computer with widely available tools such as Python and CPLEX.

Methods (brief summary)

By stringing together all the departures (-1) from station *i* and arrivals (+1) at station *j* in time order, the temporal sequence of number of bikes by station can be found for each time band. Since the data are sorted in time order and each event represents the departure or arrival of one specific bike, each row in the data set is different from the preceding one by exactly one (+ or -) in the position of the station in the column of the table. Define an indicator function I e {+1, -1}. Let I = -1 if the event is a departure from i at time t and I = +1 if the event is an arrival at stop j time t + D. Note that two records are made from every OD pair: create one row labelled [i, -1] at time t, and one row labelled [j, +1] at time t+D

Sort all the resulting records by t, and at any step Station j (t) = Station j (t-1) + I

The data structure for this particular data view lends itself to a dynamic picture of the movements. We can either reset the number of bike each 24 hours, or we can let it roll, or we can introduce optimal rebalancing at particular stages. A fixed number of events (say 500) takes less real time when there is a lot of churn in the system. Our method makes adjustments after a fixed number of events (i.e. more frequently when needed based on system usage).

A paper based on this work is entitled "A Rolling window approach for the multi-period bike-share balancing and inventory problem." Manuscript, EPB-2016-0110. This paper has been submitted to *Environment and Planning B: Planning and Design* and is currently being revised following referees comments.

Findings

As proposed, meaningful results indicate the stations where there are excess arrivals, excess departures, distinctive time of day usage, or day of week usage. The potential uses of this finding will be in adjusting or expanding the system. It may also be possible through operations to refine the way that the equipment is repositioned. The research has identified these stations in the Columbus study and also in the Boston HUBWAY data. [See Report]

The stations in Columbus with the greatest accumulation are: (1) High St & Lincoln St; (2) Schiller Park - Stewart Ave; (3) Bicentennial Park; (4) High St & 2nd Ave; and (5) Columbus Commons - Rich St. The stations with the greatest deficiency (needing replenishment) are: (1) 3rd St & Gay St; (2) Sensenbrenner Park; (3) Nationwide Arena - Front St; (4) Topiary Park - Town St; and (5) North Market.

The stations with very rapid turnover and build up (or down) in the Boston Area are notably adjusted to have larger numbers of bike slots.

Recommendations

Regularity in the usage of a station can be used to model the likelihood of consistently adding or removing bikes from a station. Some stations have highly regular demand. Others are more time varying. The frequency of the need for rebalancing can be found as a simple function of capacity and the rate of addition / removal. The most active stations (going out of kilter most quickly) will determine the need for balancing actions. The stations that are infra-marginal at these times can be ignored but they may also provide needed supplies to the rest of the system. At any time, the sum of bikes in all stations should be less than the total number of bikes, reflecting the bikes in use.

The special case of the Columbus data protocol allows us to see that a bike that checks in to a station and is next seen checking out of another station, *must have been moved by the operator* (for whatever reason -- balancing or maintenance). This allows us to approximately quantify the stations with the largest amounts of repositioning. Note well that a station that often provides bikes back (from its accumulation) to the system may in fact on other occasions also need bikes itself. One discovery from this work is that there are no simple exploitable rules that work all the time.

It is also possible to quantify the amount of time a station is left with zero bikes, or zero slots. In the interests of avoiding extremely frequent small changes, the operator may be able to measure the impact and implications of these out of balance situations and (possibly) tolerate them for a short time. This is less likely to be true for a station with a clear addition or removal trend (i.e. it only gets worse).

Several larger systems show that these issues are scaled up as the systems grow. This may mean that operational practices developed during the growth phase can be continued and expanded and adapted to the larger system. The Boston system data (available on line) shows reports for stations that are empty or full in addition to the length of time that they are in this situation.

In the particular case of Columbus it is noted that despite the study and understanding of the CoGo system, the case was not able to be made to expand the CoGo system include the University district. Instead, the university developed a separate system (Zagster). Although the data is still in its formative phases, it is believe that this failure to exploit what are called network economic effects through integration and expansion has resulted in two less strong systems. It is recommended that future systems developer employ more collaborative negotiating tactics.

Please see associated/attached report that covers some of the details. An additional paper will be available once it is accepted by referees.

NEXTRANS Project No.

“Methods for Improving Bicycle Sharing System Balance” (170OSUY2.2)

Morton E. O’Kelly

The Ohio State University

Funding from NEXTRANS is gratefully acknowledged

Bike Share Research Project

Final Report: Supplementary Material

Preamble

This report summarizes several analyses that were performed for CoGo (Columbus) and HUBWAY (Boston) bike share data. The report emphasizes studies that are completed and on-going as part of the work funded under the Nextran project. The project’s aim of discerning useful general rules about rebalancing has been successfully completed. A peer reviewed article on this topic is under review.

Problem

The City of Columbus is home to a relatively new bike sharing system, called CoGo. Because the system is still quite small, it has not yet encountered the growth issues that have hampered other larger systems. For this reason an initial study of the re-balancing of the bikes at the stations, the station size and location represent a suite of interdependent problems that are amenable to basic GIS and optimization approaches. Recent news and publicity has made it abundantly clear that large systems have enormous challenges in the area of bike rebalancing. The project carries out research that is needed to combine data from the system with optimization techniques (including linear programming) to design low cost strategies that will ensure operational efficiency within the parameters that are established by the local bike share system.

There are at least three ways to view the problem: individual users and subscribers to the system use the bikes to connect to desired activity spaces in the city. At present, the Columbus system is primarily being used as a connector for weekend and evening recreational activity. From the fixed infrastructure point of view there is the question of where and how many stations to place. From the equipment point of view, there is a need to balance and maintain the system. There are numerous operational and planning questions found in the design and operation of these systems. Once these decisions are taken, there are operational levels of service to be set for prices/time windows, free use and so on. Since the system typically operates under an agreement with the city, there is a need to ensure that the terms of these agreements are both satisfactory for the users and implementable by the operator.

The majority of bike sharing systems have evolved and matured towards 3rd or 4th generation technology --- see Shaheen, Davis Guzman, and Zhang (2010) for a characterization of these fourth generation systems. Among the distinctive features of these systems are smart phone enabled maps, web data about locations and availability, and management techniques that motivate certain advantageous customer usage patterns. It is clear that customers would prefer a system with greater flexibility, and there is a tension in running a sustainable modal alternative. Clearly, alternative modes require efficient adjustments if they are to remain sustainable. On the other hand, there is merit in simply having an additional transport option, one that helps to close the gap between walking and automobile use.

The detailed OD pairs for the CoGo Bike System in Columbus through August 2014 have been obtained by the investigator, and are used in this research. Follow up work with a larger system (Boston) is also carried out.

Approach

The usage pattern here (in Columbus) is <u>not</u> primarily for commuting. The use of the system for recreation and access to amenities in the evenings seems quite robust. The peak hours of use are Saturday and Sunday between noon and 8pm. The system is (or ought to be) designed to comply with those movements. On top of this, there is a need to have a 'trend' and a 'seasonality' that we should try to capture from the data. For example if a station almost assuredly has a declining number of bikes (removal) then the rebalancing operation should probably fill that station back to capacity.

Underlying (proprietary) techniques used by system operators help to avoid major cost overruns and to provide the level of service promised in their contractual obligations with the city. Nevertheless, these systems can present serious issues of balance and cost -- some for example emerging from unanticipated repair and theft (unique factors) but also some that emerge from issues of peak demand, location imbalance, and costs of system re-balance. Bikes that are in the wrong place at the wrong time have to be moved, and the optimal design of these movements, and the elimination of the need for many of them, is a worthy goal for an operational model. There is a need to slice the data in a new way to quantify these effects (see below).

A data analysis approach is used to track the movement of bikes in and out of stations over time. Preliminary efforts have shown that a system of bike share stations can rapidly go out of balance -- with too few bikes available at certain times in key locations, and perhaps less intuitively, too few open slots for returns available at other demand destinations. This can occur during the formative years of a new system (as we have in Columbus) where the primary users tend to be moving towards leisure and other entertainment districts (and parks) from residential areas, and not as yet using the system for commuting or modal integration types of trips. Observations show that certain peak locations are heavy attractors, while some areas are basically stagnant. This could change dramatically with the introduction of more stations, and with the expansion of the system.

Systems such as Dublin (Ireland) are larger and have a usage pattern that includes commuting to work. In this regard it is not too surprising to see that stations at the edge of the city are empty of bikes (in the morning) as these are taken towards the city stations during the commuting rush hour. Apparently from a recent check, their system is allowed to remain out of balance in this respect -- several stations having zero bikes available.

Methodology

The scope of the work is to attempt to anticipate the demand balance that might occur in the system and to provide operational tools (typically optimization or simulation programs) to benchmark the extent of the problem, and to assist the operators to make these adjustments effectively. The method is quite simple: because the movements of bikes in our stations are tagged by bike number and by time of day, it is possible to use a discrete event counter to monitor several key / interesting features: (a) the movement of bikes between geographic stations; (b) the extent to which a bike station would go out of kilter (deficit or surplus); (c) the movement of bikes that are not part of a customer action (i.e. when bikes are moved for repair or for reposition). Sample practice with all these event occurrences have been accomplished, and some preliminary results are reported here.

Bike Movement Timing Analysis

Consider a station where bikes arrive and depart. Use the plus symbol (+) to denote an arrival and a minus symbol (-) to denote a departure. The question then is how long on average do the bikes stay at the station? Bikes 1 and 2 arrive at a station and subsequently depart:

----------(1+)-----------(2+)------------(2-)---------------(1-)-----------

Interestingly, we can ignore the bike name/number and count the time between events, in a variety of ways and get the same answer for the mean (although not the variance). Note that 2 pairs of departures and arrivals create 4 events.

The sum (and average) of stay times resulting from arbitrary pairs between any departure and arrival points, regardless of bike ID, are stable while variance is variable. Here is an example:

Track bike number --- 1 to 1 and 2 to 2: notice the total length of the dotted lines is: a + 2b + c

----------(1+)-----a-----(2+)-----b------(2-)------c--------(1-)-----------

----

Track events first arrival matched to first departure, etc.: notice the total length of the dotted lines is = a + 2b + c

----------(1+)-----a-----(2+)-----b------(2-)------c--------(1-)-----------

----

Our approach then is to compute the time at a station by any bike that is checked in and out, with the understanding that bikes taken <u>off line</u>, say for repair, and are not checked out in the conventional sense. For these cases, after a check in at station A, the bike will reappear at a later time as a check out from station B (we do not know the time of either the removal for system use from A or the arrival at B from the system).

--------- CHECK IN AT A ---------- at some time taken from A ----

--- at some time returned to B ----------- CHECK OUT AT B ---

The times of check in at A and check out at B (for the same bike number) are known, i.e. we know the time of last legitimate check in at A and the next significant departure from some other station. These mismatched events can be used to measure the extent to which the empirical system is actually removing / rebalancing bikes.

Data and results for this pattern can be used to determine the amount of rebalancing going on.

Frequency of Balancing

Consider all the bike movements – each departure and arrival creates a pair (-) and (+) and these are sorted by time. There is a time series of events (essentially every departure is a "-1" and every arrival is a "+1"). We arrange these in a very long list sorted by time. Each successive line clearly differs by 1 from the previous one (an arrival or a departure). If we look at say 500 total + and – actions this corresponds to approximatively 250 trips (each creating two marks). It makes sense to intervene in the system more when there is more movement.

Use the event counter and take MOD(COUNTER, 250) and MOD(COUNTER, 500) to determine the timing of events at intervals of 250 events and 500 events respectively. When there is a lot of movement in this system the 250 events could be occurring in a short time (actually, average of 0.59 days separation) and when the system is slow or unchanging the 250 events would take a longer time (the maximum at this ratio was 2.52 days)

See below – table

Note that 250 events correspond to 125 move pairs, and 500 events correspond to 250 pairs.

It is felt that the 250 interval (226 moves in the study period of just a few months) is too frequent and inconsistent with the operator willing to let an out-of-balance situation persist for short while

M250		*M500*	
125 pairs of moves	DAYS	*250 pairs of moves*	DAYS
Mean	0.59	Mean	1.17
Standard Error	0.03	Standard Error	0.06
Median	0.57	Median	1.05
Range	2.44	Range	3.03
Minimum	0.08	Minimum	0.21
Maximum	2.52	Maximum	3.24
Sum	133.73	Sum	130.65
Count	226.00	Count	112.00

Table 1: Descriptive statistics for the days between events (M250 and M500).

Obviously when we wait for 250 events to take place (or 500) there can potentially be a lot of build up or depletion of bikes without correction. At M250 interval there were 19 occurrences with 10 or greater accumulated bikes and 15 with -10 or less. These are not necessarily out of balance conditions: a station with capacity of 15 with two current bikes could gain 10 without any issues.

These conditions are however not consistent at any station – in other words there is an approximately random (my view) occurrence of these build ups. Often a station that is out of kilter at a particular time, is no longer out of balance the next (250) steps. In other words some self-correction seems to happen.

When we wait to 500 actions, (the M500 interval)   there are 26 occurrences with 10 or greater when M500 and 23 with -10 or less. Again these are not consistent at any station.

It is clear that the 500 interval means we are rebalancing on average about every day, which makes sense given the real system.

Temporal Patterns

As a means to visualize the trends, the following saw tooth pattern diagram has accumulating (top 2) and decreasing (bottom 2) stations. Notice the dotted line which corresponds to the peak of the first accumulating station. If the rebalance is done at this time, the first two stations would have surplus bikes and the third and fourth stations would have deficits. If the rebalance is delayed (see line to the right of the dotted line) we would have, as shown in the second panel, a situation where stations (1 and 2) would have temporarily (for the length of the red line) had to hold at their maximum values, with bikes perhaps being redistributed to other nearby stations. Similarly the third and fourth stations would have held at their minimal value (the short red line on panel 4).

(Incidentally, the Boston HUBWAY data on line tool has a report that shows the total number of stations that are either empty or full and also the length of time that the stations have remained in that condition. It is apparent from these data that the system tolerates some stations staying out of balance. Also, note that they have several popular stations that are frequently empty.)

The purpose of these diagrams is to suggest that there may be correlation between the overflow situations from one full station to nearby less full ones if we delay the balancing operation. Arguably the situation of a full station with a nearby available station is a tolerable problem (as the user helps the balancing) but an empty station may be a more seriously penalized outcome.

When we examine the ebb and flow of real stations they are rarely if ever monotonic and there are periods of accumulation followed by a period of de-accumulation. As a result, a station appearing to be heading for a full situation could actually be rebalanced if a number of users come and take away from the accumulated bikes.

Regarding the accumulation pattern of each station, several external factors need to be considered.

Case 1) both types of station, accumulators and decliners, indicate that those stations are quite stable with respect to their users' rental/return pattern, which in turn implies the stations' roles are temporally predictable.

Case 2) other stations showing different trends (possibly convex or concave curves) are not the same as case 1 above. They are possibly affected by temporal fluctuations of bike demands with respect to rental/return, for example by tourists who show different movement behavior from residents or workers.

Focusing on urban geography, those observations indicate that the case 1 stations (accumulators and decliners) are likely to be located around transit transfer points, workplace clusters or residential areas. Case 2 can be around popular tourism locations or hotels.

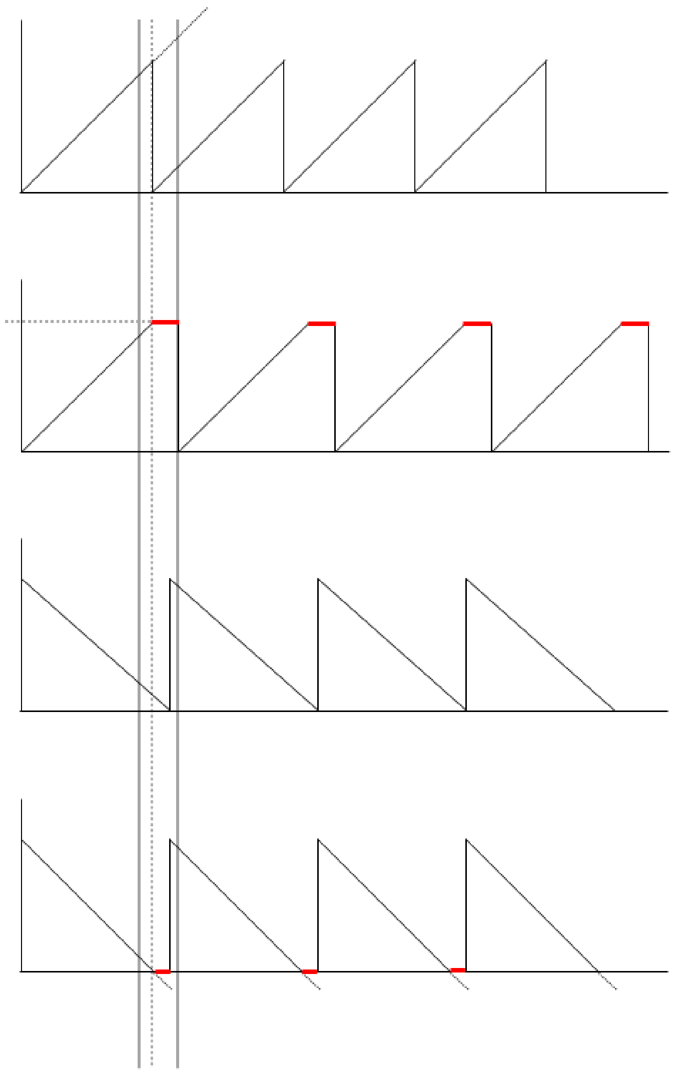


Figure 1: Saw tooth pattern of accumulating and de-accumulating stations

Bike Rebalance (Paper)

An approach to tackling this problem from an optimization perspective is quite complex. It has been attempted in a manuscript, EPB-2016-0110, entitled "A Rolling window approach for the multi-period bike-share balancing and inventory problem." This paper has been submitted to *Environment and Planning B: Planning and Design* and is currently being revised following referees comments.

The essential idea in that paper is to use a time window approach to trigger the rebalancing operations. The tradeoff (as can be appreciated) is that further foresight requires more computational power. Also all these approaches require data from an existing operational system to gather some ingredients for the time trend of the stations.

Among the decision variables that the analyst must consider are:

- When to rebalance
- Which stations to rebalance
- How many bikes to take away from full (or approaching full) stations
- How many bikes to return to empty (or approaching empty) stations
- Safety stock levels (so that stations are neither entirely empty or full)
- Decision about the deferral of rebalancing (tolerate constraint violation)

These issues are considered in the paper that is cited above under review.

CoGo Data: Trends

Begin with data from a system that is currently operational where bikes are removed and added to stations over time. These data are available for Columbus. The data can be used to track individual bikes, or to aggregate the service at each station, or by time. The research sorts these observed actions in time order. Then, observing these movements, we can quickly understand that some stations accumulate bikes, while depletion occurs at others. The system is the result of the operator's rebalancing activity, as it is clear that the physical system cannot permit negative bikes or bikes above the threshold.

Based on accumulating (red, i.e. where bikes stop) and depleting (green, i.e. where bikes start) stations, it is apparent that there is a spatial rebalancing needed. Think of red stations where demand sinks / stops and green stations where the demand starts or is created. We need to bring bikes to the green stations, and to take them from the red ones. The challenge is to get a good way of doing this.

This research has experimented with multiple strategies, and a brief summary is here:

The basic system is small (30 stations). However, there are comparable data available from Boston, and the initial small scale efforts will be amplified to deal with the Boston case in later work. At least initially, the problem is a bounded transportation problem (stock of bikes at current location and a required redeployment at the other stations). Of course this is not the complete idea of how to route the flows, but it does give a quick way to gauge the amount of movement needed and the apparent frequency of adjustment. For example, if a station is currently not actually causing an unbalanced condition (it is neither too full nor too empty), that station could still be used as a component of the rebalance operation if the available bikes could be used to satisfy a greater need at some other locations. It is expected that the dual variables from the associated linear program might give some insight on stations that have valuable / costly conditions. The research also explores more complete sub-problems once the logical balancing operation is understood. In other words, the implied optimization step right now is a simple bounded transportation problem (redistribute 220 bikes from existing points to some other positons …). However, once the logic is working, that sub step will be examined and possibly replaced by a capacitated vehicle routing module.

The research simulates the impact of various heuristics for rebalancing. For example, after looking at a system of this type, note that at any moment, the total number of bikes in the racks is something less than the maximum number of bikes overall (because some are checked out). The rebalance operator should treat this available stock of bikes as an equality constraint. On the other hand the number bikes to be placed back in the racks may fall inside a target range for each station. The bikes checked out at the start of the rebalance step will reappear downstream and the next time a rebalance is needed there could very well be a different stock of bikes to move. These insights are important to avoid framing the problem with infeasible conditions. It is also notable that the system could be adjusted to make rebalancing less frequently needed.

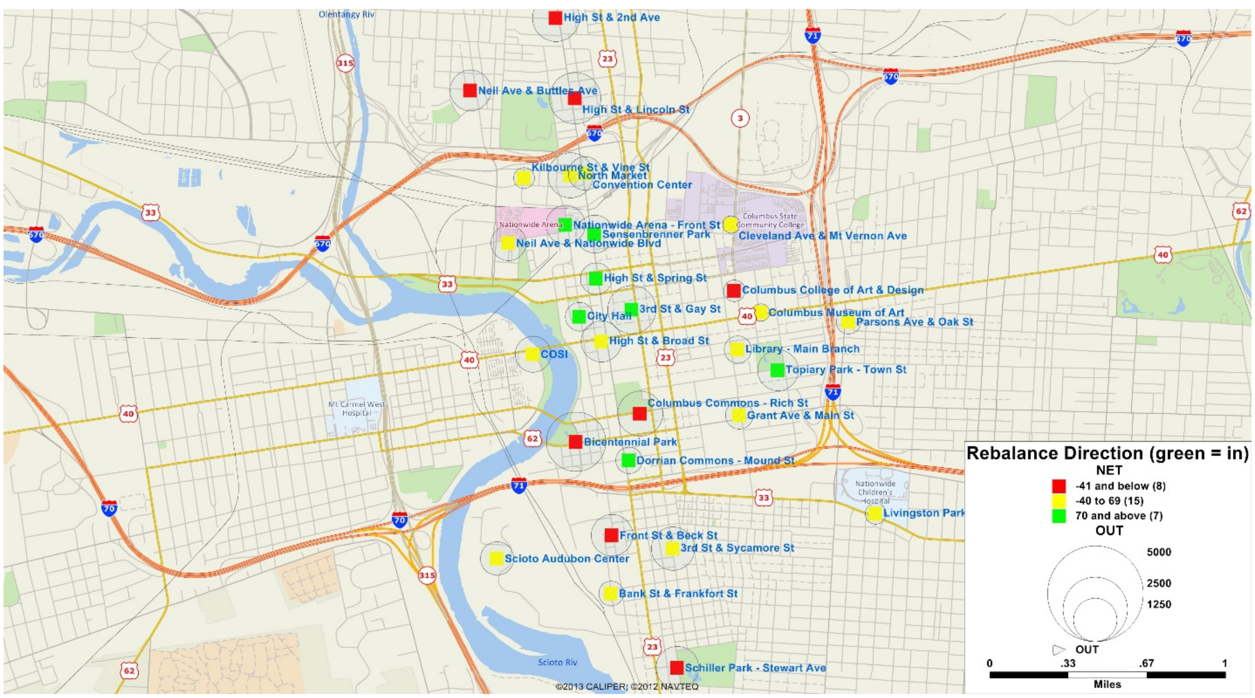


Figure 2: Accumulating (red) and de-accumulating (green) stations. Columbus OH, Source: Author's map from CoGo Data Analysis.

A linear program has been designed and experimented with to perform this analysis. It is essentially a “transportation problem.” An interesting aspect of the small system in Columbus is that we can observe daily variations in the usage of the system, and incidentally hour by hour variations even within these days. The system is small enough that we can get a good handle on these variations. There is also some evidence from hackathon efforts with HUBWAY data that there are some similar stable / repetitive patterns in the data for Boston. Clearly, the system operator can exploit recurring patterns.

By and large, Columbus stations have peak use over the weekend period (late Thursday to Sunday).

Markov model for the transitions

A useful operational tool includes measuring the interactions between stations, and the times and usage patterns of the bike system. The analytical design is to use the data as observed to extract parameters that can be used as drivers for the simulations and optimization. There is a very detailed analytical design including stochastic flow representation. Implications include:

1. The patterns of movement (of the bikes) between stations.

2. The apparent spatial patterns of the accumulation and depletion of bikes.

Example: Imagine the bike system starting say at 4pm, just before evening recreation time. A lot of bikes are taken out for use. The number of bikes is now low (everywhere) because bikes are checked out. So the dynamic control of this system has a variable number of bikes and a somewhat fuzzy target – it cannot for example enforce that all the bikes be set back to some base condition because there are simply not that number of bikes available at any instant. (We are assuming here that the rebalance operation takes place before the next move. This is not a loss of generality because the rebalance resultant will leave plenty of capacity for these returning bikes.)

The CoGo system consists of N stations. CoGo signs out bike at these locations, and the data show that a bike that is borrowed at location i has probability P_{ij} of being checked back in at station j.

Suppose T_{ij} is the raw count of the flows between stations i and j (transfer or transition probabilities).

Compute transition probabilities according to the well-known method of Anderson and Goodman (1968):

$$pij = \frac{Tij}{\sum_{j'} Tij'}$$

The limit after a large number of transitions is a vector such that:

$$\pi = \pi P$$

The elements of π are the equilibrium shares in each of the states.

These values are computed using a MATLAB script.

Semi-Markov model: Holding time distributions

(a) In a semi-markov model there is a probability of transition and a time that the mover holds in a state before completing the move (or you can think of it as deciding to move to j from i, and then holding a random amount of time h_{ij}, and then completing the move whereupon the process starts over again at j, (going to k for example).

(b) The reason this is not a perfect match to the CoGo system is that the bike rider has a probability of transition from i to j, and there is a distribution of duration that someone making an i to j move might hold, but, upon arrival at j, the process does not immediately start because there is a time during which some of the bikes sit idle --- it would be as if we could get the busy fraction of the bike from the DURATIONS they are signed out --- this is clearly just a small fraction of the total day

Suppose hj is the average time spent at node j. The limiting behavior of the process is then:

$$\phi j = \frac{\pi j \, hj}{\sum_{j'} \pi j' \, hj'}$$

Where πj is the limiting probability from P and hj is the mean holding time in state j

$\pi j \geq \phi j$ implies that the process makes many transitions into station j, but makes short stays there

$\pi j \leq \phi j$ implies that the process makes few transitions in to station j, but makes long stays there

For example we find the equilibrium for station 1 (3rd and Gay) as 0.0495, but it has 0.5610 (quick turn over), with an equilibrium weighted hold percentage of 3.75%. In other words, bikes deplete from this station. (In other words, it would be expect that bikes have to be replenished there.)

A complete analysis of station equilibrium is in Table 2.

STATION	Departures from	Arrivals at	Time weighted equilibrium	Markov equilibrium
3rd St & Gay St	1008	996	5.16%	4.95%
3rd St & Sycamore St	1020	1045	5.44%	4.63%
Bank St & Frankfort St	367	325	1.67%	1.63%
Bicentennial Park	1646	1711	8.93%	8.23%
City Hall	480	438	2.27%	1.74%
Cleveland Ave & Mt Vernon Ave	202	184	0.94%	0.79%
Columbus College of Art & Design	541	518	2.68%	1.67%
Columbus Commons - Rich St	940	952	4.91%	5.20%
Columbus Museum of Art	216	184	0.95%	0.95%
Convention Center	639	565	2.93%	2.90%
COSI	822	855	4.41%	3.37%
Dorrian Commons - Mound St	210	198	1.03%	1.64%
Front St & Beck St	765	792	4.10%	4.46%
Grant St & Main St	311	263	1.35%	1.89%
High St & 2nd Ave	1139	1270	6.59%	5.66%
High St & Broad St	857	863	4.48%	3.99%

High St & Lincoln St	1261	1356	7.10%	6.84%
High St & Spring St	586	473	2.47%	2.25%
Kilbourne St & Vine St	254	278	1.45%	1.29%
Library - Main Branch	333	296	1.52%	2.06%
Livingston Park	140	131	0.67%	1.38%
Nationwide Arena - Front St	319	301	1.54%	3.46%
Neil Ave & Buttles Ave	781	816	4.23%	4.42%
Neil Ave & Nationwide Blvd	81	82	0.39%	3.33%
North Market	910	927	4.85%	4.91%
Parsons Ave & Oak St	467	462	2.40%	1.85%
Schiller Park - Stewart Ave	1044	1063	5.54%	4.95%
Scioto Audubon Center	594	595	3.09%	3.00%
Sensenbrenner Park	662	673	3.51%	2.93%
Topiary Park - Town St	669	652	3.40%	3.63%

Table 2: Markov Chain and Semi-Markov Results (Source: author’s calculations)

Bike Balancing

Bike balancing is a fundamental task in bike share systems. The main idea is to know more about how many moves to make, where, and when the system needs to be rebalanced. The system may be allowed to stay out of balance for a time but the system operator would have to compensate the inconvenienced user, perhaps with added free time. So, in essence, there is a cost to being out of balance and the goal would be to minimize the impact of these displacements.

There are also some very simple and logical connections in the system which our research has examined and proposes to exploit. For example, stations tend to either accumulate or deplete. In the case of Columbus for example, it is well known that Bicentennial Park and its attractive restaurant area can be a magnet for trips, and that this station tends to have bikes accumulate. On the other hand 3rd and Gay, a residential and multiuse area tends to deplete over time. One might think that these are special cases and that there is not a generalizable pattern, but our investigation has highlighted this prevailing pattern of accumulation and depletion in both Columbus, and the much larger Boston HUBWAY system. The problem is that the rates of these ebbs and flows are not uniform so that there are some fast accumulators, some slow accumulators, and some similar depletion sites. To place the question in clearer context, would we wish to trigger a rebalancing operation as soon as any station hits the high or low point (safety stock – a concept discussed further below). Or, would it be better to wait for several stations to turn critical and to make the moves at that stage?

Imagine a station where bikes accumulate (if no intervention) – we know this as a result of the accumulation of bikes at that station (i.e. the result of a lot of +1 arrivals, and -1 departures).

Suppose the slope of this "graph" is $N(t) = a + bt$. (Assuming for simplicity a linear accumulation function).

The ith accumulative station has $N(i,t) = a_i + b_i t$

Similarly for decreasing stations: $N(j,t) = a_j - b_j t$

Now if we set a floor of zero (or some safety stock) we can see that the j th decreasing station will hit zero at

$t_j = a_j / b_j$

and an accumulative stations will max out at

$t = (c_i - a_i) / b_i$

where we note that the stations may have variable upper capacity. There may be stations that do not fit either model. For example a station may have a variable demand pattern that places its usage as fluctuating between the lower and upper limit. Take all the accumulating stations and place their sloped line at the origin – the steepest accumulator then reaches its limit faster than the smaller sloped lines. (There may be different maximum capacity sizes at the larger stations and so a fast accumulating station may have a larger upper limit.)

We can decide on a time to activate the rebalance and of course at that time the slowly accumulating stations may not have yet reached a point where anything needs to be done. The stations that are rebalanced are set back to some smaller level (and this then is another important question – what is the best level to reset a typically accumulating station to?)

The results from HUBWAY data suggest that a number of stations reach a critical full or empty situation within 2 days. The sum of all the accumulating stations and the depleting stations should be approximately in balance – we say approximately because there is a possibility that there are bikes in active use that are not yet checked in to their destination station. Let's assume away that issue and that the fleet of bikes is somewhere in the system.

In any case, the replenishment / rebalance decision will occur at some compromise time (not all the decreasing rates and accumulating rates hit the same marks) and we could have a slowly decreasing station that does not need to be touched at all until maybe several iterations.

The decision to activate a rebalancing action is designed to top up the decreasing stations and remove from the increasing stations. The optimal timing of the rebalance step seems to be very difficult, but it needs to be triggered by the first station to reach max or min and to recur thereafter.

Heuristic for gauging the rebalance needs

Generalized rates of accumulation and depletion from stations are used to make a heuristic for gauging the rebalance needs. Work has highlighted the difficulty of optimizing the entire rebalance operation. The problem is quite difficult, but most cities have some very obvious spots that have to be covered, and the rest of the system has to mesh/blend in with those constraints. So fixing A B C ... implies fairly limited choices for D, E, .. etc. It could be for example that the bike docks have to be within one or two city blocks of other stations to ensure overlap in coverage.

The basic finding is that if stations can be identified as reliable accumulators or de-accumulators (adding or shedding bikes) then the system operator can exploit these recurrences by removing and adding bikes systematically. If a station is an extremely reliable accumulator (for example) it might be possible to reduce it to the safety level as opposed to say some half way point. The data processing steps used in this work allow the identification of stations that are reliable in this regard (they have a very high R^2). There is no clear finding about whether these changes (up or down) should be all the way to the capacity or floor of the station – by analogy with stock replenishment models, it seems that the system should use a safety stock (for example 2 more than zero, or 2 less than capacity).

Rebalance Direction for Columbus

The following stations provided the largest amount of transfer bikes to be picked up (because these stations accumulate)

Station
High St & Lincoln St
Schiller Park - Stewart Ave
Bicentennial Park
High St & 2nd Ave
Columbus Commons - Rich St

The following stations needed the largest amount of bikes dropped off (because these stations tend to run down their stock)

Station
3rd St & Gay St
Sensenbrenner Park
Nationwide Arena - Front St
Topiary Park - Town St
North Market

To be clear a station with a lot of drop off or pick up also experiences the opposite type of action, possibly because the operator is not optimally planning these moves.

Boston data

The Boston site provides lots of data (HUBWAY).

STEPS

[1] Download and open the compressed file (it is 190 Mb). The file is Stationstatus data, with information about available bikes and empty docks per station and minute back to August 2011 (30 million records), is now available: download 190MB CSV (tar.gz)

[2] Copy all the START STATIONS AND START TIMES to a list with a new field "-1" (for departure)

AND concatenate i.e. below the first list copy all the END STATIONS AND THE END TIMES with a new field "+1" (for arrival)

[3] Sort it by TIME

[4] Select a station, we can plot the departure and arrival …. this will show the cumulative changes by station

[5] "hubway_trips_time sorted.csv" is the concatenated file, and when sorted by station its recorded datetime contains both of start_date and end date in the raw data but with different codes (-1 for start date and +1 for end date respectively).

[6] If you filter in EXCEL the records for station 3, for example,

1) The first record, which is the earliest station 3 record, is a bike rental at station 3 (freq: -1).

2) Then, the next record is a bike return at the corresponding station (freq: +1).

Those patterns are accounted for by the cum_freq column that indicates cumulative frequency based on the freq column.

[7] From these data, create the cumulative graphs for individual stations (png files in the folder). Overall, stations are biased toward either departure / check out or return / check in.

Results: Station vs Time

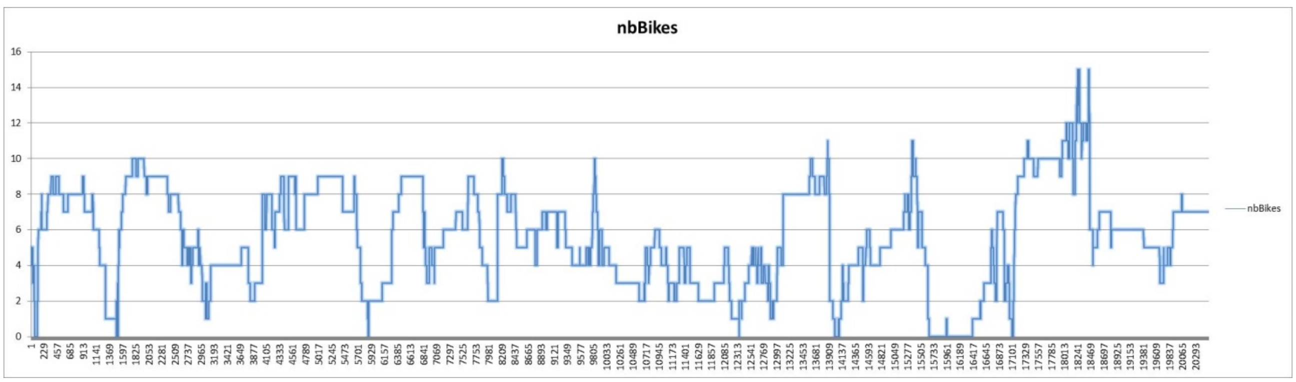
A large data set **Stationstatus** data, with information about available bikes and empty docks per station and minute back to August 2011 (30 million records), is now available: download 190MB CSV (tar.gz). Extracts of this data series have been accessed and analyzed.

Note that in these data that the capacity of each stations seems to be able to change but there is a "modal" class -- the one which occurs most often for the station: for example station 3, has "15" as the most common value (but also 10, 12, 13, 14)...

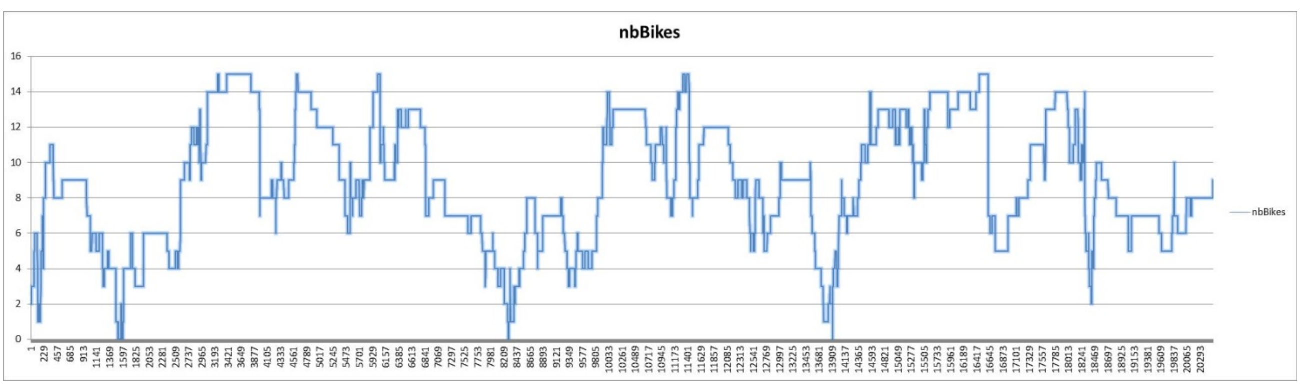
A graphic of bike share station status in Boston over time oscillates back and forth between the capacity and zero. An alternative idea is to get a summary of the contents of each station by time (every one minute interval). We are able to analyze about 20,000 minutes of data for selected stations within an excel model.

This of course produces a lot of data but the results are informative.

Example Station 5: has an average over these observations of 5.57 (say 6) bikes available and an average of just over 8 empty stations. (The capacity of this station varies for some reason -- with an average of 13.6.)

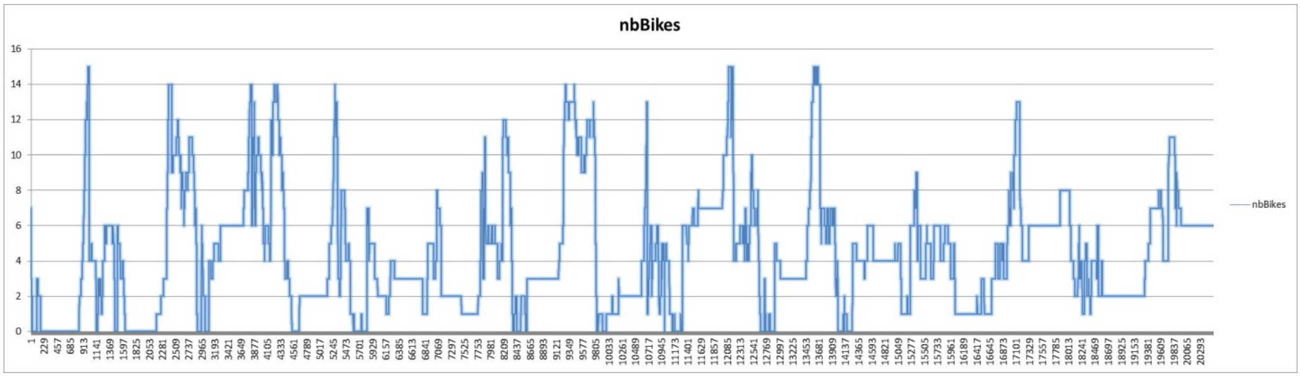


Example Station 8: has an average of 8.9 bikes and 6.1 empty stations. The size of this station is 15 bikes.



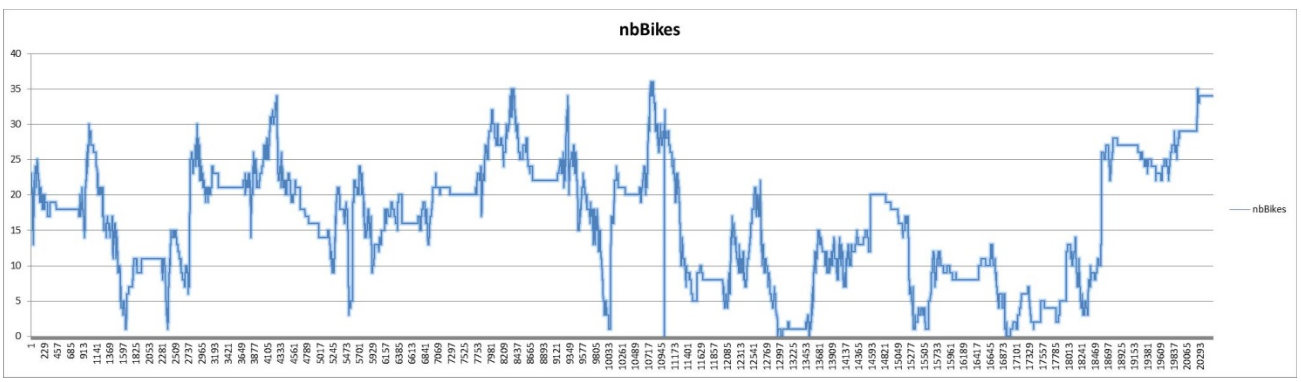
Example Station 48

From the data here, we see 4.3 bikes available and larger number of empty slots (10.5). The station has a capacity on average of between 14 and 15.



These traces show the intense variability of real stations and also the difficulty of representing the stations with a simple linear model for growth or decline.

From this we note that Station 22 is a very large station (much larger than anything in Columbus:



These traces show that every station is essentially a controlled or limited process as it approaches the upper or lower limit, some correction is performed.

Plots of the station dynamics for Boston 2013

The process to develop these charts was to sort each stations time stamps (+1, -1) and to take the cumulative trace of the time stamps. Clearly the rebalance operation would intervene before a station went out of range, but it gives an idea for the rate at which bikes are accumulated or de-cumulated from that center.

A graph of the stations depleting (left) and adding (right) is here: there are clearly needs arising about every second day. (See ovals around stations with a 2-4 day range.)

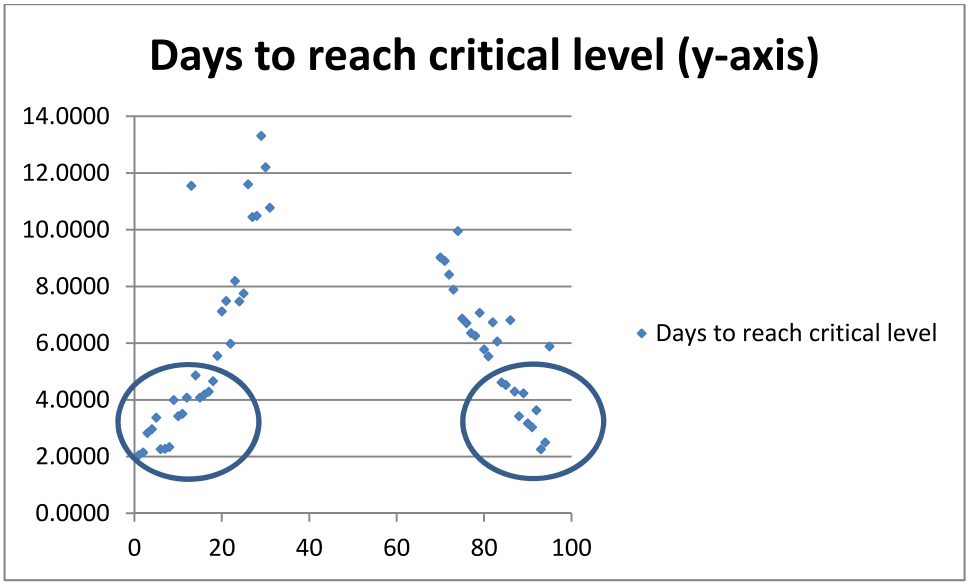


Figure 3: Station and slope/fit: impacts

Highlight the stations with at least a 0.9 goodness of fit. Characterize these good fitting situations as either declining ("dec") or accumulating ("acc"). [See Appendix.]

Selected examples of these plots are shown on the next page.

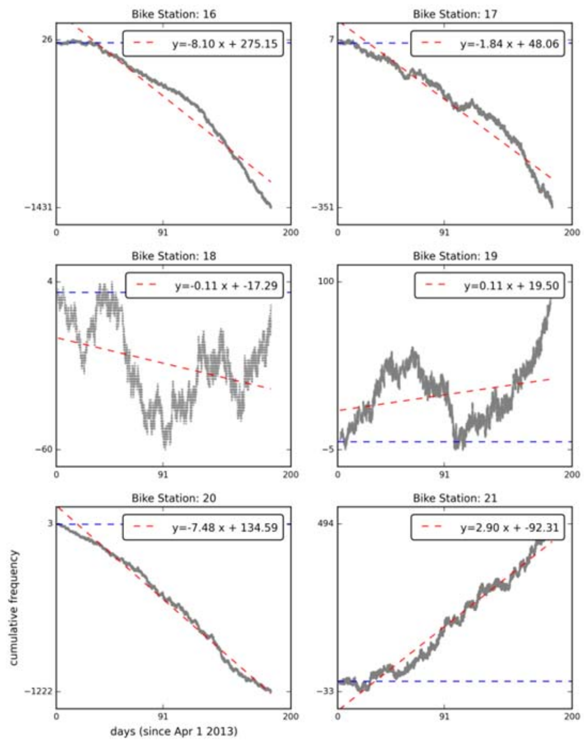


Figure 4: Tends at selected stations, Boston Hubway System. Source: Yongha Park analysis and graphics.

This series shows two good fitting decreasing and increasing stations (20 and 21) as well as two reasonably good decreasing stations (16 and 17) and examples of two stations with score outside the acceptance range (18 and 19).

All further examples in the Appendix (available from okelly.1@osu.edu)

Conclusions

The work completed so far has provided an added empirical facts about bike sharing systems. There are apparent exploitable regularities. There are rapidly changing locations which require frequent maintenance and there are other stations that vary without much discernable pattern. The ability to detect repetitive and reliable trends at some stations should provide a recipe for a recurring pattern of balancing. These locations might also give rise to demand management techniques to incentivize behavior that is beneficial to the shared system.

References

Camargo, R., F. Lima, M.E. O'Kelly, 2016 "A Rolling window approach for the multi-period bike-share balancing and inventory problem." This paper has been submitted to *Environment and Planning B: Planning and Design*

Chemla, Daniel; Frédéric Meuniera, Roberto Wolfler Calvo Bike sharing systems: Solving the static rebalancing problem, Discrete Optimization 10 (2013) 120–146

García-Palomares, Juan Carlos; Javier Gutiérrez, Marta Latorre (2012) Optimizing the location of stations in bike-sharing programs: A GIS approach, Applied Geography 35, 235-246

Lin, Jenn-Rong; Ta-Hui Yang (2011) Strategic design of public bicycle sharing systems with service level constraints Transportation Research Part E 47 (2011) 284–294

Lin, Jenn-Rong; Ta-Hui Yang, Yu-Chung Chang (2013) A hub location inventory model for bicycle sharing system design: Formulation and solution. Computers & Industrial Engineering 65, 77–86

Martinez, Luis M; Luís Caetano, Tomás Eiró, Francisco Cruz (2012) An optimisation algorithm to establish the location of stations of a mixed fleet biking system: an application to the city of Lisbon Procedia - Social and Behavioral Sciences 54, 513 – 524. 15th meeting of the EURO Working Group on Transportation

Shaheen, Susan; Davis Guzman, Stacey, Zhang, Hua, (2010) Bikesharing in Europe, the Americas, and Asia: Past, Present, and Future, Institute of Transportation Studies

Contacts

For more information:

Morton O'Kelly
The Ohio State University
1036 Derby Hall, 154 N Oval Mall
Columbus OH 43210

okelly.1@osu.edu

(614) 292-8744

NEXTRANS Center
Purdue University - Discovery Park
3000 Kent Ave.
West Lafayette, IN 47906

nextrans@purdue.edu
(765) 496-9724

www.purdue.edu/dp/nextrans