



Bus Arrivals and Bunching

R. Adam Molnar
Bellarmine University

Journal of Statistics Education Volume 16, Number 2 (2008), www.amstat.org/publications/jse/v16n2/molnar.html

Copyright © 2008 by R. Adam Molnar all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Course Project, Data Collection, Exploratory Analysis, Hypothesis Testing, Evaluating Assumptions

Abstract

Finding suitable projects for introductory courses that blend real-world data with relevant questions and feasible instructor effort is often difficult. This paper describes one such project – tabulating the intervals between bus arrivals. By including data gathering, descriptive statistics, hypothesis tests, and regression, it covers most of the topics in a first course. This paper describes the genesis of the project, classroom implementation, analysis results for the student-generated dataset, and adaptations available for other classes and course sizes.

1. Introduction

Students often have trouble seeing the relevance of introductory statistics, particularly when exercises are unrealistic or trivial. For many years, statistics education literature has emphasized the advantages of real data ([Cobb 1992](#)). This sounds simple, but finding relevant tasks is a constant challenge. In particular, finding suitable course projects for introductory courses is difficult. Focused courses like experimental design lend themselves relatively easily to assignments that tie together topics, but most introductory courses cover a wide variety of topics. Often, instructors get around the difficulty of composing a suitable problem by asking students to propose and design an individual or small group experiment. [Albert \(2000\)](#) and [Fillebrown \(1994\)](#) have written about this approach. While student groups tend to create interesting papers, there are complications. Each student or group will have a different question, with an accompanying different dataset. Checking a dozen or more datasets takes a lot of effort. Generally, projects should be due at the end of the course, after all topics are taught; when combined with an in-class final, the grading workload can become overwhelming. Differences in workload also complicate evaluation; which is better, an ambitious project with flaws or a simple project done well? Often, these drawbacks cause instructors to abandon the project idea.

This paper is the story of one project that works around these issues; it uses real-world data where each student collects data and contributes to a single combined analysis dataset. It was used in an introductory course in Chicago during summer 2005 and summer 2006. The subject – bus arrival times and bus bunching – may not be applicable at all schools or all situations, but should be adaptable to many situations. The paper begins with background on Chicago and the bus system. The next section describes the implementation of the project through the quarter. Analysis results then appear. The final section suggests adaptations for different schools and larger classes.

2. Problem Background

According to the [Chicago Transit Authority \(2005\)](#), an average weekday has over 900,000 bus rides on several thousand buses. Theoretically, buses run on a posted and published schedule. Reality differs, as buses are hastened or slowed based on road and weather conditions. One common situation is bus "bunching", when two buses arrive at a stop close together, because a trailing bus has caught the bus in front. Bunching frequently occurs during rush hours. Rush hour traffic congestion delays a bus, causing crowds to build at each stop. Picking up these crowds further detains the bus. The bus behind the slowed bus sees very few passengers, which lets it move more quickly. For example, say the schedule claims an interval of 10 minutes between buses. When bunched, the slowed bus can fall back to the bus behind it. This makes a delay of about 18 minutes followed by two buses within 1 or 2 minutes. Not surprisingly, passengers do not appreciate this pattern.

The neighborhood around the University of Chicago has limited rail service; buses provide most public transit. One bus, the #55, runs along 55th Street from near the school to rail service, then onward to Midway Airport. Students perceive unacceptably poor service, with long waits and lots of bunching. Some have asked for a dedicated shuttle from the campus to the rail service, which has led to much campus debate. The evidence of poor service is anecdotal, as pointed out in a campus newspaper op-ed by [Katz \(2004\)](#). He wrote "The issues drawn out by the prospect of a University shuttle to the Red Line are important. But the final decision should be based at least in part on how reliable the existing service is, and this cannot be determined by casual student observations."

The transit authority claims that bunching occurs less than 10% of the time. There is a published prior work evaluating this claim, released in February 2004 by a Chicago-based nonprofit, the [Campaign for Better Transit](#). They defined "bunching" as a space of 2 minutes or less between buses. Over the 14 routes surveyed (not including the #55), bunching ranged from 1% to 30%, though most routes were between 10% and 15%.

Based on the university debate, the summer 2004 class incorporated some homework problems on the concept of bus bunching. Students in the class responded positively, calling the idea interesting and relevant. When looking for a large assignment for the 2005 course, bus bunching came to mind.

3. Project Implementation

The introductory quarter course in statistics is open to both undergraduate and graduate students; through the summer school office, visiting students and high school students may also enroll. The textbook is *Moore and McCabe's Introduction to the Practice of Statistics*. Being outside the normal academic year, summer classes are small. In summer 2005, there were 12 students, and summer 2006 had 7 pupils. The course lasts nine weeks. After completing the sections on describing data, correlation, regression, and collecting data, the project was introduced early in Week 4. The introductory slide asked a few questions. How can the question of "bunching" be precisely defined? Is there prior work on the subject? (The instructor mentioned the study from the Campaign for Better Transit.) What data would be useful? How could this data be collected, considering budget, ethics, methods, and so forth? How might this data be stored and analyzed?

Since almost all Chicago students ride buses, they were interested and enthusiastic. In 2005, the discussion carried on for about 30 minutes; in 2006 the dialogue took 25. Proposals from the 2005 class formed the list of potential predictors, shown below in [Figure 1](#), the final project handout for 2006. The 2006 class made proposals before viewing the 2005 list, but came up with no new observable variables and less ideas overall; the instructor chose to use the same predictor set to allow students to estimate with both years' worth of data.

In addition to the list of enumerable variables, both classes had ideas that were not workable. For instance, several students wanted to count the number of people on each bus. More crowded buses likely had to make more stops to pick up people, so passenger count would approximate the time waiting for people to board. Unfortunately, students couldn't quickly board a bus to count people without disrupting the schedule, nor could they try to look through the windows. Counting the number of people entering and leaving at the one stop became the substitute, though an inferior one. Another example is traffic congestion, which many students wanted to measure, but proved a challenge. There are automated congestion reports in Chicago, but none about roads in Hyde Park. Instead, a student suggested counting cars that passed through the intersection. That led to several sub-questions. Could cars be counted while the light was green? That is extremely difficult. Instead, since the chosen stop is at an intersection with a traffic light, cars could be counted while stopped. This led to more questions; when counting cars, the class had to choose a direction (the same as the bus) and a red light (the light immediately after the bus leaves).

After working through the design, the next phase was gathering the data. Given the course budget, the students would need to collect data themselves. Instead of stretching the limited number of students across days and hours, it was decided to focus on one key interval, the weekday afternoon rush. During the sixth or seventh week of the quarter, each student took a different two-hour block during afternoon rush hours, 3-7 PM, Monday-Friday. Scheduling to avoid overlap took just a few minutes. Each student sat at the same designated point, noting the time of every arrival and other information related to the bus. It was an experience taken good-naturedly. Several students commented that it helped them understand the practical difficulties of data collection.

Before the fieldwork, about 20 class minutes were spent more precisely defining the variables. To give two examples, how does every collector keep the same time? (Synchronize timepieces with a particular TV clock.) When and where should temperature be measured? (It was measured at the beginning of each shift, using a nearby bank thermometer.) After collecting the data, students were allowed to submit their results on paper. The instructor entered the results and built a dataset; this took about 15 minutes per student. [Allowing paper copies was a mistake. Having counts submitted electronically, in a standardized format, would be much more efficient.]

Each student attempted to follow the agreed upon instructions, and the instructor walked by the collection point a few times to spot check results. In general, the students followed directions and submitted quality data. One quality issue arose through a fortunate accident in 2005. One student consistently overestimated a variable, the number of cars at the next red light. Instead of the typical range of 0 to 8, the student reported values from 10 to 17. In class, the student volunteered what happened; cars were counted through the light cycle, not just stopped cars. There was no way to go back and find the exact value; some sort of ad hoc correction was necessary. The class decided to roughly equalize the range by subtracting 8 from that set of values. [In a class project, data mistakes like this are not negative. They provide the opportunity to discuss error handling. The proposed solution was pretty good; it assumes a constant amount of travel through the green light. Perhaps a division would be better, assuming proportional travel, but that's not clearly better, and without clear superiority, the student idea is preferable.]

After the data was combined, making the one change in 2005 and none in 2006, directions for the report were distributed at the beginning of the next-to-last week of the course, replacing the final week's homework. Students reported working around 10 hours – more than a typical homework, but not enough to lead to negative complaints. It took the course staff between 20 and 30 minutes to grade each project. The project had an open-ended format, so each project looked a little different; identical phrasing would indicate likely cheating. The handout summarizing the 2006 assignment appears in [Figure 1](#) below. The file `buses06.csv` in the archive contains the data in comma delimited format. A more complete data description appears in the archive as `buses06.txt`.

The 207 observations were collected over 11 days. Because the 2006 class was very small, their dataset included the 2005 observations as well as their own. These students could also search for CTA improvement between years. The first observation each day has no prior bus, so there are 196 buses that could be bunched. The arrival time of each bus was measured in hours and minutes. By keeping these variables separate in the dataset, it makes it easy to compute results by hour. The first bus of each day has a special value of -1 for variables that rely on a prior bus. Because an actual value of -1 is not possible, the instructor coded this special value to make it easy to remove these buses from calculations when needed. For instance, the Stata command `drop if bunch == -1` will remove these observations.

Final Project
Statistics 220, Summer 2006

The bus dataset has 207 observations of #55 buses (not including any X55), in order from earliest to latest. 147 are from 2005, and 60 are from 2006. The dataset contains a header row with the variable names. Each bus record contains the following variables:

1. Year: Either 5 (for 2005) or 6 (for 2006).
2. Day: One letter signifying the day of the week, M, T, W, H, or F.
3. Date: Day of the month. Some 2005 data was collected from August 1-3.
4. Temp: The recorded temperature for that session, in degrees Fahrenheit.
5. Rain: Equal to 1 if raining, 0 if not. There's not enough rain to make this useful, just one day each year, so it is unlikely to work in a regression model.
6. Hour: Hour the bus arrived.
7. Minute: Minute the bus arrived.
8. Time: Time in minutes since last bus. The first bus each day is coded as -1.
9. Bunch: A created variable. It equals -1 for the first bus each day, 0 if the bus arrived 3 or more minutes after the last 55 bus, and 1 if the bus was "bunched", arriving 0 to 2 minutes after the last bus.
10. On: Number of people who got on the bus.
11. Off: Number of people who got off the bus.
12. Cars: Number of cars at the next red light after the bus left.
13. Onlast: Number of people who got on the last bus, coded as -1 for the first bus each day.
14. Offlast: Number of people who got off the last bus, with -1 for the first bus of the day.
15. Carslast: Number of cars at the next red light after the previous bus left, with -1 for the first bus of the day.

Write a report answering the following three questions. You may use any computer package you want. Up to 10 points per part will be awarded, based 80% on content (use of statistics, proper conclusions) and 20% on style (graphs, labels, grammar, descriptions). There are no length minimums or maximums, but 10-15 pages is the "expected" amount. Projects are due at the time of the final exam. You may not collaborate or work in groups. You may ask questions of the instructor and course assistant.

1. Waiting Time Distribution: Say you arrive at the bus stop to see that a bus has just departed. (Here we assume that buses leave immediately after arrival.) How would you describe the distribution of the time spent waiting for the next bus? Does this distribution differ by year? Does it differ between early hours (3-5 PM) and late hours (5-7 PM)? Present the distributions graphically and/or numerically.
2. Bunching: The CTA claims that bunching (defined by our bunch variable) occurs less than 10 percent of the time. How often does bunching occur? Test the CTA claim. Does the proportion differ significantly between 2005 and 2006?
3. Predicting Waiting Time: Take Time as the response, and the other variables (except Bunch and Minute) as potential predictors. What is the best regression model to predict waiting time? For On, Off, and Cars, you'll want to use the immediately prior numbers, Onlast, Offlast, and Carslast, to predict the time.

Figure 1: Handout describing the project assignment.

4. Analysis Summary

As with many real-world datasets, it is not hard getting output or results from the computer. The difficult part of analysis is evaluating assumptions behind the tests. For instance, a histogram of bus waiting times by year appears below. Stata or another statistical program will report that the mean waiting time in 2005 was 9.36 minutes, while the 2006 mean was 9.45 minutes. The overall average is 9.38 minutes with standard deviation 7.00 minutes. For instance, a histogram of bus waiting times by year appears below as [Figure 2](#).

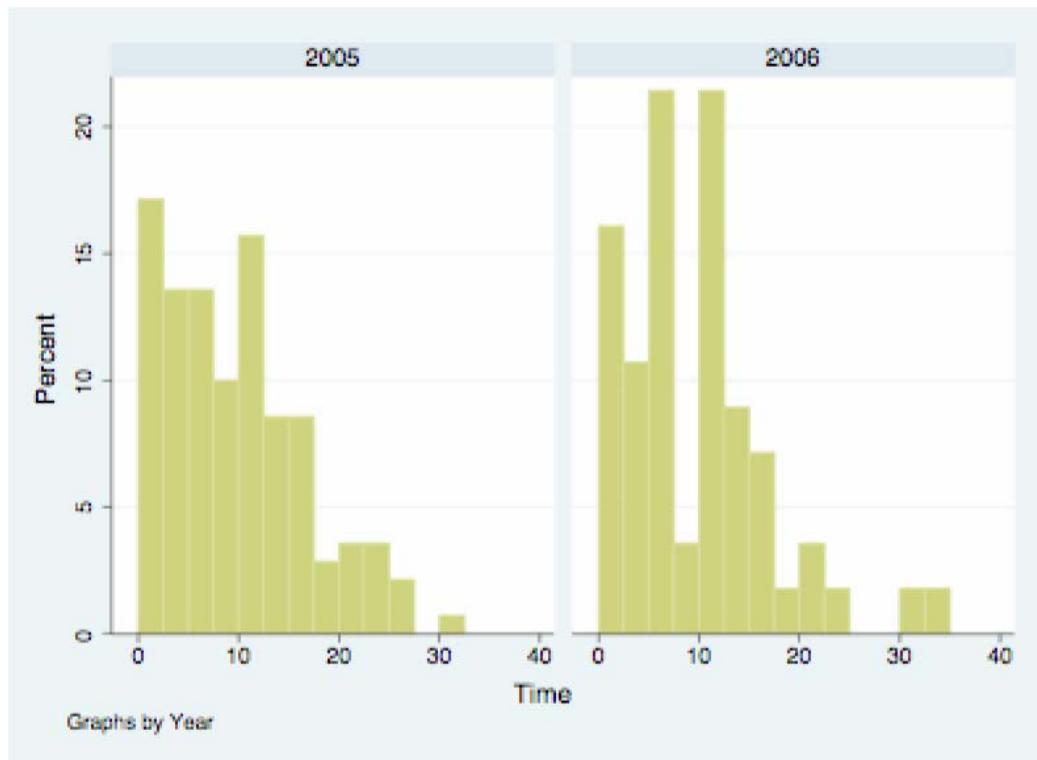


Figure 2: Bus waiting times by year.

Does the distribution differ between the two years? Looking at the means and histograms, there appears to be little difference. Formally, students might attempt a two-sample hypothesis test, using the t statistic and relying on the central limit theorem for approximate normality. That approach is flawed; in bunching, a very long wait time will be more likely followed by a short interval, making successive times dependent. Since the CLT presented in most introductory classes relies on independence, the conditions for approximation do not apply. The question suggests graphical and numerical summaries to nudge students away from formal tests. Most students got the hint, and a couple even noticed the dependence.

Another part of the question about distributions asks the students to evaluate the difference between early hours and late hours. The 117 buses that arrive before 5:00 PM have a mean interbus wait of 8.49 minutes, while the 79 buses after 5:00 PM average 10.70 minutes. This is not surprising, given that the Chicago rush hour runs from 3 – 5 PM; the CTA would tend to schedule more frequent service during rush hour. Students might attempt a two-sample t -test to determine if the means differ, getting a p -value around 0.03. This would lead to a conclusion of a significant difference at the standard 0.05 level if the conditions behind the test held. Like the comparison between years, though, the key assumption of independence is not satisfied. Successive bus times are not independent, and the test is not designed for dependent samples. One can only point to a difference between early and late hours, not claim statistical significance.

This trouble with assumptions extends into the second question, evaluating the CTA's claim. The basic count is that 33 of 196 buses are bunched, a rate of 16.8%. The rates by year are similar. 2005 had 24 of 140 bunched, 17.1%, while the 2006 numbers were 9 of 56 or 16.1%; any test or comparison will find little difference between the years. The more interesting problem is overall bunching. One possible test here is an exact binomial test. In the binomial test, the base or null hypothesis believes the CTA's claim that bunching occurs no more than 10% of the time. The test uses the binomial distribution to determine the likelihood under the 10% assumption that a sample with this much bunching or more would occur by chance. The Stata command to perform this test is "bitesti 196 33 .10", which yields a one-sided p of 0.002. If the assumptions behind the test hold, this is strong evidence against the CTA's claim.

Two questions arise about these assumptions. First, is it fair to compare this sample to the CTA population value? The sample is not fully representative, including some of the more difficult hours. Students should consider the effect of limited sampling; outside rush hours the percentage might decrease. On the other hand, the CTA claim has no time qualifications, and serves as a baseline; it's a fair starting point. Second, is the binomial test appropriate? The maximum possible proportion is much less than 1, which would occur only with a multiple hour gap followed by a caravan. Even 50% bunching would be difficult; that would indicate that buses almost always arrived in pairs. The exact distribution of the bunching variable is unknown. That said, the minimum is zero, if the buses ran on schedule. The dependence and caps make it unlikely that the distribution is more weighted towards high proportions than the binomial. A binomial proportion test is conservative and thus adequate.

Are students expected to consider these issues? Ideally, yes. In the papers, some students commented about the less-than-complete sample, but no student worried about independence. They chose the most appropriate test that they knew.

Finally, the regression can lead to many results. None of the available variables are strongly predictive. The R^2 from a model with all available predictors is only 13%. This shouldn't be surprising, since the basic hypothesis is that bunching is caused by delays earlier on the route, which

looking at one stop does not measure. Hour is the most useful, as shown from the prior summaries; earlier hours have more scheduled buses and lower average bus waits. Beyond that, the course covers very basic variable selection, mostly on significance, so there are a variety of reasonable models here. The main point was to show logic in model building.

5. Discussion

Does this project apply to other places? Bus bunching and transit service are important in many places, though not all. Larger cities have similar bus systems, and this project would adapt easily. Other universities don't have municipal transit, instead offering bus shuttles around campus. For a campus shuttle, the question would change. Instead of looking at waits, the students might examine on-time performance. This project did not ask about on-time performance, because the relevant bus schedule says merely "7 to 12 minutes". With a fixed timetable, punctuality and reliability become relevant questions to test instead of bunching. Then, there are some places where bus transit is relatively unimportant to students. For instance, the instructor has moved from Chicago to Bellarmine University in Louisville. At Bellarmine, most students drive and community bus service is infrequent. Therefore, most course sections don't use this project.

Another alternative to consider is online timing. A few systems, including Seattle and Portland, have installed GPS technology to provide online data. In Chicago, the CTA has installed online tracking on one route. In cities with online results, the data collection effort is reduced. Depending on the goals of the class, this may or may not be a good thing. As the students noted, there are benefits to working through a small amount of actual collection; in their case, at the Starbucks!

For an integrated project such as this, there is time commitment throughout the term. In class, introducing the idea and discussing variables requires about 30 minutes. Later, scheduling times for fieldwork took roughly 10 minutes and discussing the mechanics of data collection about 20. Outside class, the instructor spent about 15 minutes per person to examine and enter the collected data. Asking the students to provide results in a common electronic format would reduce this time, though someone would still need to examine the results for potential collection errors. Once projects were submitted, each report took roughly 20 minutes to grade.

6. Modifications for Larger Classes

With a small number of students, as in summer session, in-class discussion was uncomplicated, students collected data individually to maximize the amount of information, and students analyzed data individually, making decisions in the process. Larger or different classes necessitate some changes to the project, described below.

1. Introducing the project: In a 150 student lecture hall, or with a very quiet class, discussion is difficult. An alternative would assign the design as homework, and then summarize the results afterwards. It's less active, and not as effective, but might be necessary in a huge lecture. Additionally, data collection instructions could be distributed as a handout, with little comment.
2. Group work: Allowing the students to work in small groups both reduces the amount of papers to mark and encourages interaction. In a large class, the workload reduction is likely necessary. In a small class, instructors might prefer that students benefit from working with others. On the other hand, in solo work students develop research and individual idea creation skills. University of Chicago culture tends towards individual efforts and away from group projects, which led to the request for independent projects. Other schools will differ. One way to gather data from more time periods, while maintaining the overall effort level, would be to make each team responsible for more than 2 hours. The two hour window was chosen to keep the project at a reasonable amount of effort.
3. Collecting data: With more people available, a large class could expand from the limited time window (weekday afternoons) and collect data over an entire bus schedule, both non-peak and peak. Scheduling survey times gets more complicated. Automation is strongly recommended. It's very difficult for students to copy data from prior years, since the temperature and rain variables would likely give that away.
4. Overlap in data collection: With a larger number of groups, it would be better to schedule some students to overlap. Having two results from the same time frame would introduce the concept of repeated measures, measurement error, and dealing with differences in data. These lessons are very practical, but are minimized in many textbooks and classes.

7. Acknowledgements

In addition to the students, the author would like to thank Shali Wu for suggestions throughout the project. Additionally, the author wishes to thank the editor and referees for comments that improved the explanation of several issues.

8. References

Albert, J. (2000), "Using a Sample Survey Project to Assess the Teaching of Statistical Inference," *Journal of Statistics Education [Online]*, 8 (1). (<http://www.amstat.org/publications/jse/secure/v8n1/albert.cfm>)

Campaign for Better Transit (2004), "The Late State of the Buses." (<http://www.bettertransit.com/busstudy.pdf>)

Chicago Transit Authority (2005), "Bus Ridership by Route: July 2005." Available through <http://www.transitchicago.com>. Chicago Transit Authority Bus Tracker, at <http://ctabustracker.com/bustime/home.jsp>.

Cobb, G. (1992), "Teaching Statistics," in *Heeding the Call for Change: Suggestions for Curricular Action*, ed. L. A. Steen, pp. 3-43, Washington, DC: Mathematical Association of America.

Fillebrown, S. (1994), "Using Projects in an Elementary Statistics Course for Non-Science Majors," *Journal of Statistics Education [Online]*, 2

(2). (<http://www.amstat.org/publications/jse/v2n2/fillebrown.html>)

Gal, I. and Ginsburg, L. (1994), "The Role of Beliefs and Attitudes in Learning Statistics: Towards an Assessment Framework," *Journal of Statistics Education [Online]*, 2(2). (<http://www.amstat.org/publications/jse/v2n2/gal.html>)

Katz, R. (2004), "Shuttle debate needs Statistics." Chicago Maroon, May 25, 2004, Viewpoints section.
(http://maroon.uchicago.edu/viewpoints/articles/2004/05/25/shuttle_debate_needs.php)

Moore, D., and McCabe, G. (2006), *Introduction to the Practice of Statistics*, fifth edition. New York: W.H. Freeman and Company.

R. Adam Molnar
Bellarmine University
2001 Newburg Road
Louisville, KY 40205
amolnar@bellarmine.edu

[Volume 16 \(2008\)](#) | [Archive](#) | [Index](#) | [Data Archive](#) | [Resources](#) | [Editorial Board](#) | [Guidelines for Authors](#) | [Guidelines for Data Contributors](#) | [Home Page](#) |
[Contact JSE](#) | [ASA Publications](#)