Fundamental Statistical Concepts in Presenting Data

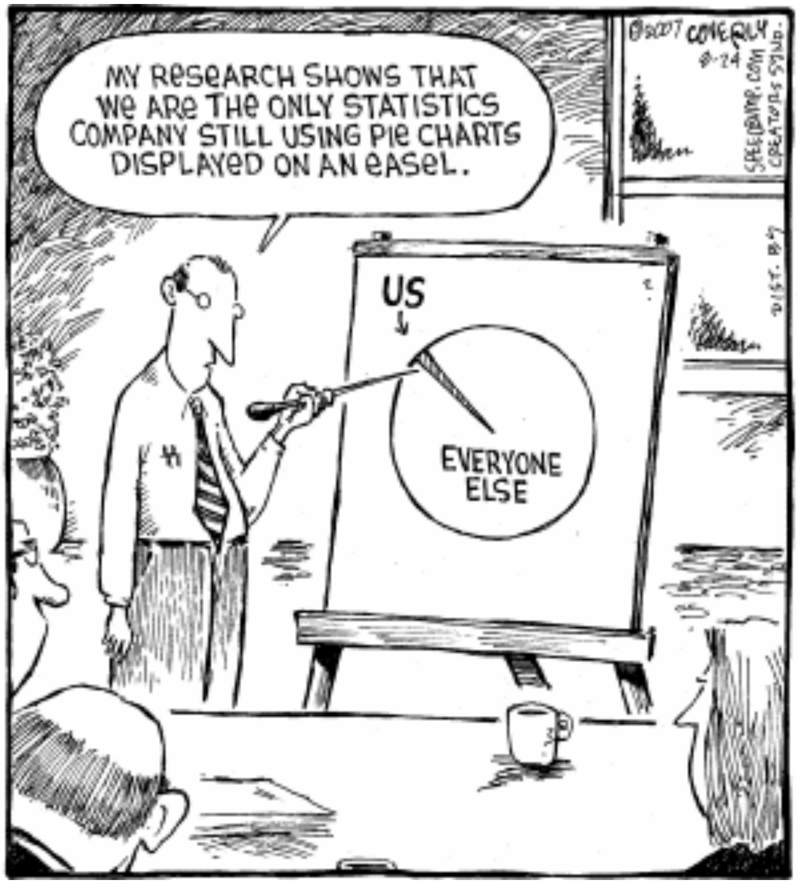
Principles for Constructing Better Graphics

Rafe M. J. Donahue, Ph.D.

Director of Statistics
Biomimetic Therapeutics, Inc.
Franklin, TN

Adjunct Associate Professor
Vanderbilt University Medical Center
Department of Biostatistics
Nashville, TN

Version 2.11
July 2011


MY RESEARCH SHOWS THAT WE ARE THE ONLY STATISTICS COMPANY STILL USING PIE CHARTS DISPLAYED ON AN EASEL.
US
EVERYONE ELSE
©2007 COVERLY 8-24
SPEEDBUMP.COM CREATORS SYND.
DIST. BY

speedbump.com, 2007-08-24

Calls to the call center are databased; that is, every call that comes into the call center has its relevant information stored: where did the call originate, who answered, in what category was the question, when did the call start, when did it stop, and so on. Some of these calls generate "cases"; there is action that needs to be done after the call. Call centers are interested in measuring their capabilities and oftentimes, as in this instance, the time until a case can be called "closed" is a metric that these centers use to grade themselves.

As the consulting statistician, I was told that the leaders of the call center were interested in reducing the time to closure for the incoming calls. Of course, I was told the mean time to closure was some number of minutes, either 2 or 20 or 200 or something, I forget; it really doesn't matter for this discussion. They told me the mean, so naturally I asked for the raw, atomic-level data.

They gave me the data: a printout from an SQL routine that told me, accurate to twenty decimal places (I am not making this up!), the mean time to closure.

No, I need the data that you used to get these means; do you have that data?

After several weeks, I was given a data set with hundreds of call durations.

Do you have the start and stop times from which you calculated these durations, the actual times the calls came in and when the cases were opened and closed?

After several more weeks, I finally got the data: among other things, start and stop times for each of the calls. A plot of these data is at right.

The horizontal axis shows the day and time during the week when the call "came in" and a case was created. Note that no calls came in on Saturday, as the call center was closed. The vertical axis shows the time until the case that was generated from that call was closed. Note that the time until closure ranges from "negative" to "> 90 days". From this distribution, the call center had been calculating means as a summary measure for that distribution. [Remember: mean if, and only if, total.]

The horizontal axis has been marked with two relevant time points, namely 8am and 8pm, the times when the permanent and contract call staff answered calls. From 8pm until midnight, only contract staff handled all the incoming calls.

The vertical axis is not your typical linear or logarithmic axis. It shows continuous time but the mapping becomes more compressed as time increases in an effort to avoid bunching up the atomic data ink dots in this highly, highly skewed distribution. Within each band, time is uniformly spaced; however, these even-length bands do not contain the same amount of total time.

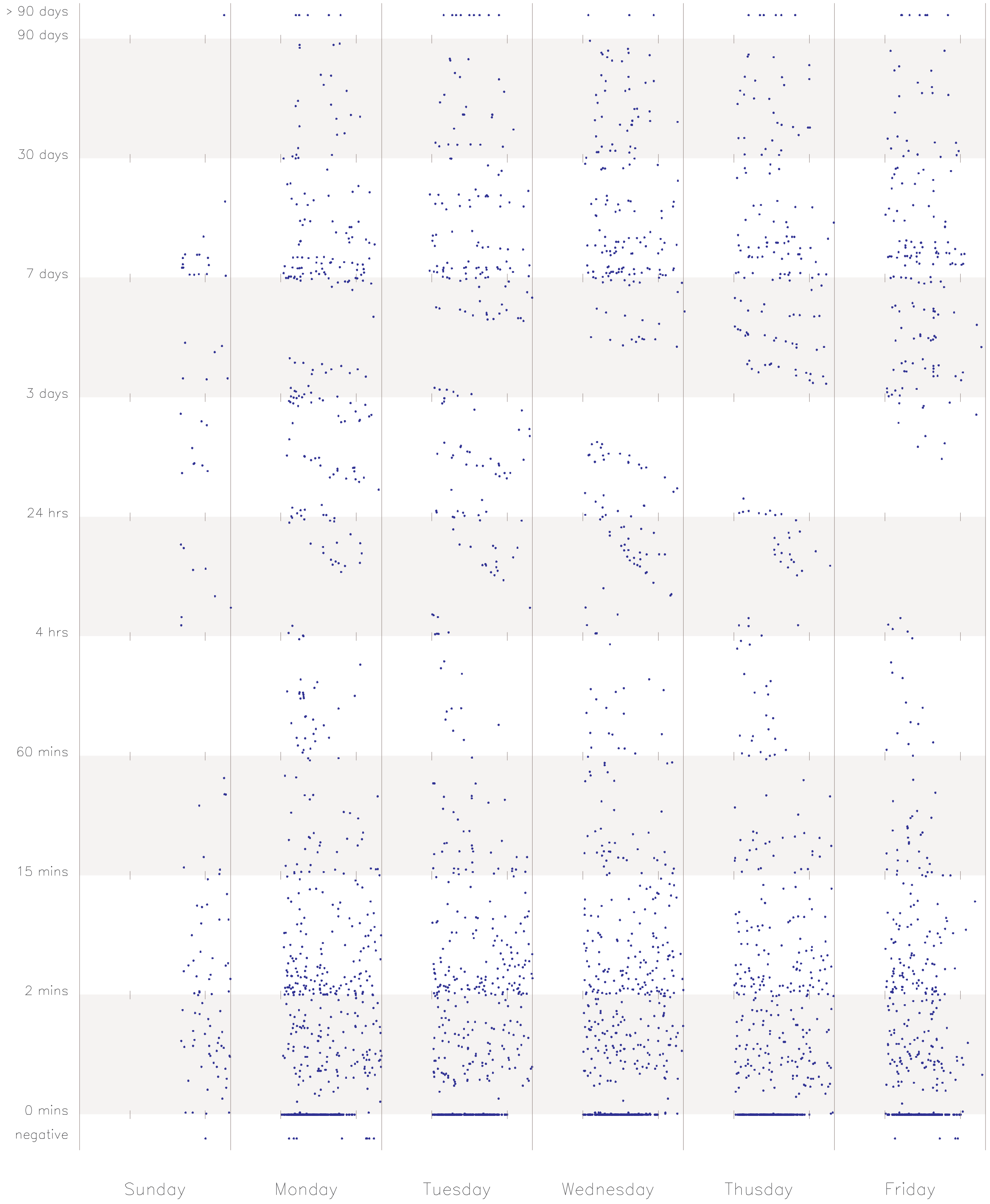
Some observations are in order, once we can see the atomic-level data:

There is point mass at zero. Note that this is not approximately zero, but actually zero. A look at the raw data revealed that the time of closure was identical to the time of opening these cases. They are not *rounded* to zero; they are *exactly* zero, down to the precision of the data collection device, the timeclock on the computers where these cases were logged. Note that this point mass only occurs, however, between 8am and 8pm.

Time to closure of cases by day of week and time

Subset of cases: all cases, April through September

Day and time of case creation is shown on the horizontal axis. Time to closure is shown on the vertical axis. Small tick marks show 8am and 8pm. Time on the vertical axis is uniformly spaced between demarcations.



Rafe Donahue, PhD

There are values in excess of 90 days. All these values, except the one that was generated late on a Sunday evening, occur between 8am and 8pm.

The non-uniform scaling of the vertical axis produces a non-uniform window that demonstrates the absence of any closures taking place on weekends. Note that for cases that were opened late on Sunday, there are no closures in the time window that spans 5 to 7 days into the future. This is because such closures would necessarily occur on Saturday or early on Sunday, at times when the call center is not in operation.

This sliding weekend effect appears to get wider as the week goes on, a consequence of the unequal spacing on the vertical axis. Cases being opened on Friday get closed within slightly more than 4 hours, or they don't get closed until after the weekend. As the time scale becomes more compressed, we can see day/ night differences; look, for example, at Thursday and the nearly parallel curves showing the case closures 4, 5, 6, and 7 days after the cases were opened.

Note that those cases that were opened late in the day on Friday (after approximately 5pm) were either closed quickly, within about 15 minutes, or were delayed until Sunday. Very few cases were opened after 8pm on Friday.

Once again, our data display is the model so our model here states that case-creation time is a source of variation in time to closure. A lesson one can learn from these data is that the call center might be able to reduce its average time to closure simply by tinkering with the days and hours that the center is staffed.

But from a data point of view, one must be concerned with the presence of the negative times to closure and the point masses at zero. I investigated these anomalies by tracing the data back through its sources until I found the database programmers huddling in the basement of the building. After showing them the plot and explaining the problem, they told me why such an instance should not be viewed as a problem at all: the server that collected the case-open times and the server that collected the case-close times weren't necessarily the same server. As such, because the servers weren't necessarily sychronized, it was possible to have negative values sometimes. *That is ok*, they said. *We usually just delete the negative values from the database, since we know they cannot be right.*

How can one know, then, I asked, *that values that are at, say, 74 seconds shouldn't really be at 98 seconds because the clocks are out of sync by 24 seconds?*

There is no way to tell, I was informed, because it is not possible to know which server is making the time stamp. But I was assured that this wasn't a problem because values of 74 seconds are possible but values of -20 seconds are not!

And what of the point mass at zero? Surely the server synchronization problem didn't explain that too?

I went to talk to the people who actually took the calls and opened and closed the cases. The call reps took calls over headsets and were talking while looking up various pieces of information. If a case was to be opened, that is, if the call reps couldn't derive the answer in real time, then they filled out an on-screen form and sent it into "the system".

But some of the senior permanent staff had figured out a trick. Because calls came fast and furiously during the day and time-on-hold and number-of-rings-before-answering were also gradable metrics, the call reps learned to do what they could to keep their non-talk time to a minimum. And this meant making sure the computer was ready when the calls came in and setting things up so as to not have to wait for the computer to respond when they needed information.

What all this meant is that the experienced permanent staff had learned to open up oodles of the on-screen case-open forms at the start of their shift, *before* their session of call-taking began, filling their large computer screens with empty case screens. By doing so, they were avoiding a certain downtime while fielding calls by having these forms open, instead of waiting for the case form to open when they first answered the call.

They also found a way to close a case immediately upon opening it (or open it as it closed), so that they would get credit for resolving issues, and resolving them quickly, in the event that they were able to answer the question without actually creating a case. So, in essence, the zero-length cases were a different kettle of fish altogether, an artifact not of the way cases were perceived by management but as a way to keep the system running, and not getting bogged down in waiting for the computers.

Learning that there were two different types of operators of the system, I wondered if operator type was a source of variation. We can split the data, then, on the basis of contract or permanent staff and look at the resulting distribution of times to closure for each group.

For simplicity's sake, I simply re-ran the previous plot twice, once subsetting on only those cases taken by the contracted staff and once subsetting on only those cases taken by the permanent staff. The two re-runs are shown on the following spread (reduced to fit and allow for some text).

Cases for the contract staff are on the left and cases for the permanent staff are on the right and the side-by-side differences are astounding. The compelling distribution we saw in the last plot is really a mixture distribution from two highly different processes. The process on the left shows the variance that is naturally in the process when one fills out the on-screen forms according to plan while the process on the right shows the use of the trickery of data entry on the part of the permanent call reps.

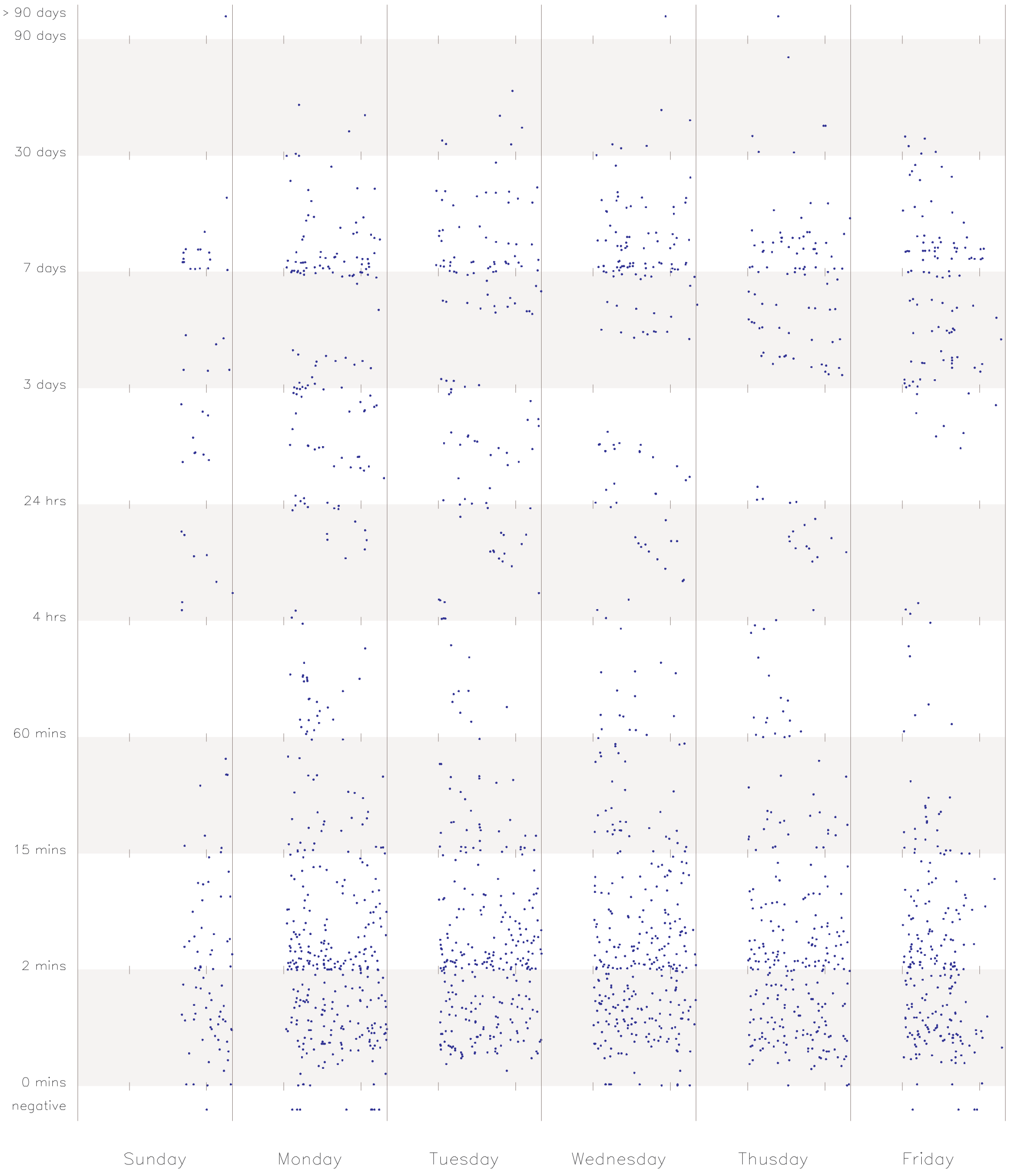
The left distribution shows essentially none of the zeros but all of the negatives. Obviously, then, there is something about the way the permanent staff are tricking the system that allows them to avoid the obvious server synchronization problem; whether or not their non-negative times are biased by synchronization issues cannot be assessed.

The distribution of times to closure for the permanent staff is stunning in its uniformity, once the point mass is taken out of consideration. These permanent staff created all but three of the cases that have dragged on for over 90 days.

Are these two staffs behaving differently or are they getting calls of different natures? Are the "tough" calls being sent more often to the permanent staff? Is there any selection bias in whom gets what type of call?

Are the staffs even doing the same thing? The presence of the point mass at zero for the permanent staff and the negatives for the contract staff and the disparities in the distributions lead one to wonder if these are really the same process. Should both groups really be combined in an effort to examine times to closure?

And what of the mean time for the total mixture distribution? Does the trick employed by the permanent staff carry enough weight to offset the impact caused by the over-90 days cases? Could we lower the mean significantly (whatever that means) by just teaching the contract staff the trick with the on-screen windows? If that were enough to impact the mean time to closure, what does that say about the process as a whole and the use of the mean as a summary statistic for measuring performance?

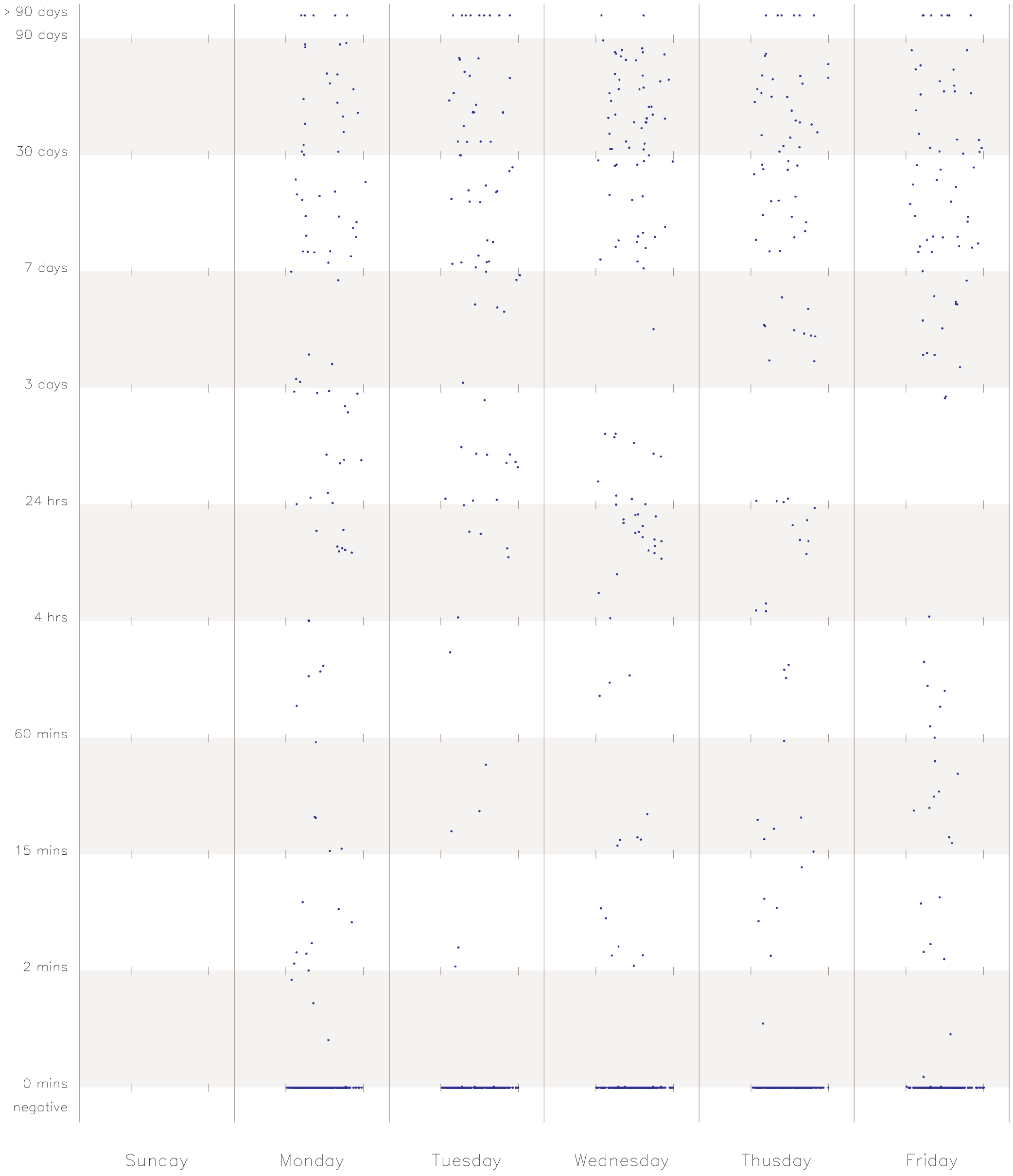


A few notes on features of the plots themselves. Note that we could have employed a you-are-here method to these plots by putting the complementary data points in the background of the plots. Unfortunately, when these plots were made, I had not yet envisioned that idea.

The time demarcations are alternating, light-weight stripes in the background, improving our ability to focus on the data. Again, like a good wait-staff, they are there when you need them, but essentially invisible when you don't.

The original plots, subsets of which are shown here, included titles and credits and explanations like the mixture distribution, so as to provide background information that supports the data and their interpretation.

The vertical axis is neither linear nor logarithmic but a hybrid that allows us to deal with the point mass, the anomalies (negatives and over-90s), and the extreme skewness in the data. The linear scaling between connection points allows us the ability to easily interpolate between the connection points if we need to do so.


> 90 days
90 days
30 days
7 days
3 days
24 hrs
4 hrs
60 mins
15 mins
2 mins
0 mins
negative
Sunday
Monday
Tuesday
Wednesday
Thusday
Friday