

WHY WE'RE NOT IN LOVE WITH RFM

by Peter Wylie and John Sammis

Back in February and March, Kevin MacDonell published a couple of posts about RFM for this blog ([Automate RFM scoring of your donors with this Python script](#) and [An all-SQL way to automate RFM scoring](#)). If you've read these, you know Kevin was talking about a quick way to amass the data you need to compute measures of RECENCY, FREQUENCY, and MONETARY AMOUNT for a particular set of donors over the last five fiscal years.

In Kevin-like fashion, he was invoking an old time management adage: "Work smarter, not harder." Since I'm a guy who loves fun and hates drudgery, I was intrigued. That's despite the fact that I'm an ignoramus when it comes to the ever-growing digital lexicon. I still don't know what HTML means.

Kevin's posting stirred up a good number of submissions to the data mining listserv that Rob Scott runs out of MIT. (Rob's probably another guy who embraces the smarter, not harder thing.) In almost all those submissions there was a lot of code that I hadn't the foggiest idea of how to decipher. So what, I thought, this is great. This is what professional sharing in the world of data-driven decision making is all about.

Then I started thinking about usefulness. As an analyst you take the time to tee your data up for computing measures of RFM, you compute the five point scales for each of the three components and then ... and then what? What do you do with that stuff?

For example, take a look at Table 1. Although the table doesn't show it directly, there are many possible combinations of the three scores. 1,1,1 would be the lowest combination; 5,5,5 would be the highest; and 3,3,3 would be right smack dab in the middle.

Table 1: Basic Layout of RFM Combinations

R	F	M
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5

Just to be clear:

- A 1,1,1 would mean that the donor had last given five fiscal years ago; had given in only 1 of those last five years; and fell into the lowest 20% of donors in terms of total dollars given during that period.
- A 5,5,5 would mean that the donor had last given in the most recent fiscal year; had given in each of those last five years; and fell into the top 20% of donors in terms of total dollars given during that period.

- A 3,3,3 would mean that the donor had last given in the third fiscal year of the 5-year series; had given in 3 of those last five years; and fell into the middle 20% of donors in terms of total dollars given during that period.

The more I thought about RFM, the more I questioned its usefulness. I had always been bothered by the fact that RFM completely ignores non-donors (or way back in the day donors). For example, John Sammis and I work constantly with databases that have far more non-donors than donors, so RFM isn't even a blip on our radar screen.

And then I got to thinking about M, monetary value. Five bins of 20% each to capture where donors fall on the total dollars dimension? Really, especially given what we know about the extremely top heavy nature of giving? So then I created Tables 2 and 3. They give us a lot of information about the dollar amounts for each of those five categories at the two schools I used to create this piece. Take a look, and then you can compare what you see with what I see.

Table 2: Sum, Mean, Median, Minimum, and Maximum Dollars Given For Each of Five Monetary Categories at School A

MONETARY SCALE	Count	SUM OF TOTAL DOLLARS GIVEN	MEAN DOLLARS GIVEN	MEDIAN DOLLARS GIVEN	MINIMUM DOLLARS GIVEN	MAXIMUM DOLLARS GIVEN
1	1394	\$29,109	\$21	\$20	\$1	\$35
2	1380	\$78,730	\$57	\$50	\$35	\$95
3	1483	\$169,389	\$114	\$100	\$95	\$150
4	1294	\$299,823	\$232	\$225	\$152	\$347
5	1410	\$5,420,710	\$3,844	\$615	\$350	\$1,051,530

Table 3: Sum, Mean, Median, Minimum, and Maximum Dollars Given For Each of Five Monetary Categories at School B

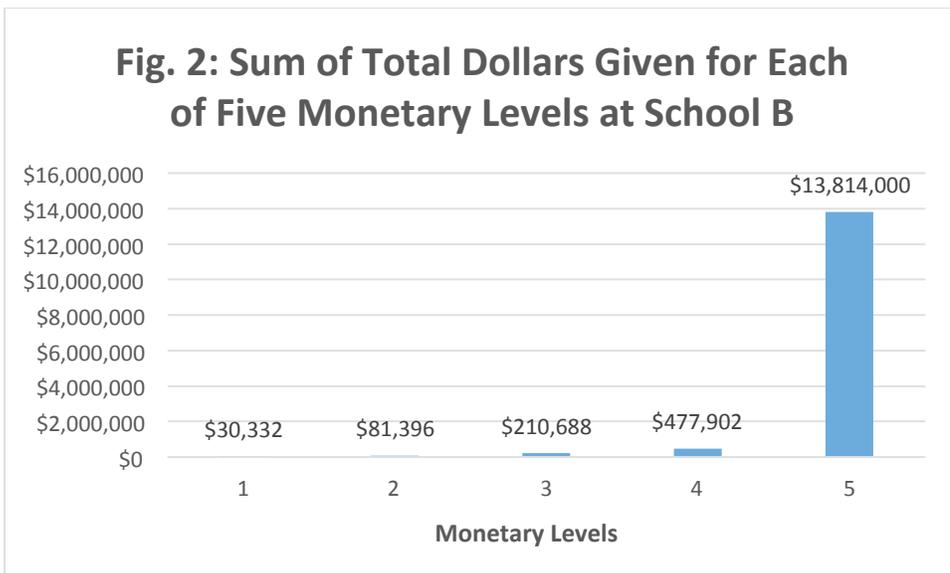
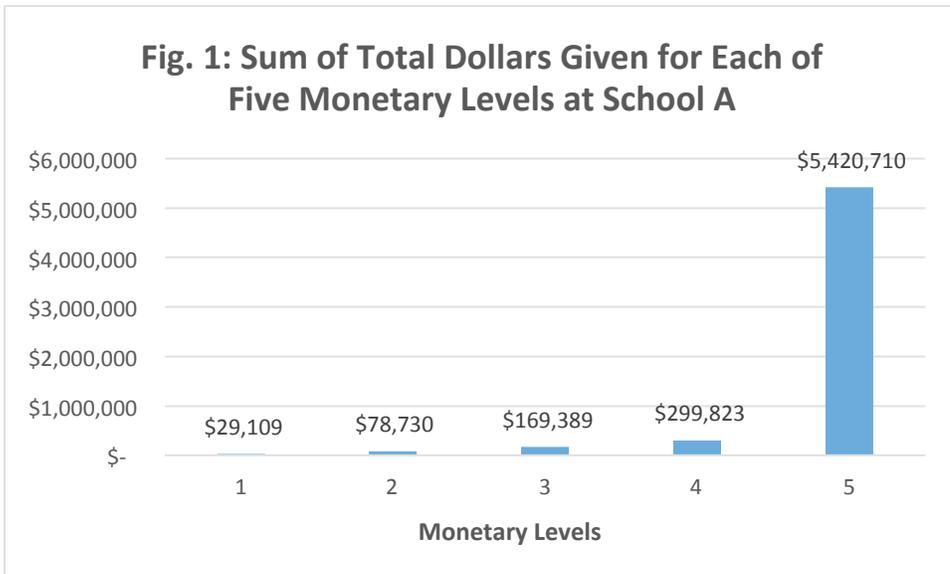
Group	Count	Sum	Mean	Median	Min	Max
1	906	\$30,332	\$33	\$25	\$1	\$50
2	841	\$81,396	\$97	\$100	\$51	\$145
3	968	\$210,688	\$218	\$200	\$150	\$320
4	906	\$477,902	\$527	\$500	\$321	\$850
5	901	\$13,814,000	\$15,332	\$2,100	\$855	\$1,935,490

Mind you, the data you see are from only two schools. But if these schools are at all representative of other higher education institutions in North America, the notion of using a five point scale to capture monetary value is pretty silly. Why? Because (except for measurement and stats experts) a five point scale implies an equal distance between the values of the scale. That is *definitely* not the case here. Consider these three points:

- Start with the most conservative way to compare the five groups: look at the medians. In School A, the median for group 5 (\$615) is 30 times

higher than the median for group 1 (\$20). In School B, the median for group 5 (\$2,100) is almost 100 times higher than the median for group 1 (\$25).

- How about the sum of giving for each of the five groups? Forget about the astronomical differences between group 1 and group 5 in both schools. Look at the differences between group 4 and group 5: in School A, \$299,823 versus \$5,420,710; in School B, \$477,902 versus \$13,814,600. (Figures 1 and 2 below shows these differences graphically.)



- How about the maximum and minimum values for each of the five categories, especially 1 and 5. In School A, for 1 the values go from a

buck to 35 bucks; for 5 the values go from 350 bucks to well over a million. In School B, for 1 the values go from a buck to 50 bucks; for 5 the values go from 855 bucks to well almost two million.

Just when I was about to throw RFM out the window, I had another thought.

Frequency. We talk to people who work in the annual fund all the time.

Whether I agree with their concern or not, these folks worry about years of consecutive giving. Say an alum gives three years running and then misses the next year. Some annual funders want to send an EMS team to the guy's house to make sure he's okay. Well, not really. But they do want to see a lot of 5's on that frequency scale. And they do not want to see a lot of 1's and 2's.

So the thought we had was this: how about using frequency as a **dependent** variable? How about building a model for frequency of giving? If we did that, we could score an entire database of alums with respect to their likelihood of frequent giving. And that would allow annual funders to focus their appeals on alums with high scores and (hopefully) end up with more consistent, consecutive donors.

To be honest, we were pretty sure somebody had already done this. (Maybe you've already done it.) But since we don't know about that, we'll tell you what we did in another blog post we're working on right now.

(TO BE CONTINUED)