# Easy Weeder

William Denton

March 29, 2018

## About

Easy Weeder is an algorithm that identifies little-used multiple copies of books and recommends how many copies we really need (minimum 1).

## Busy factor

The *busy factor* will determine which books with multiple copies can be easily weeded.

First note that our calculations will be based on circs in a "circ window" of five years: the last five full academic years, plus the current year to date (so near the end of an academic year this will be almost six years). In April 2018 the circ window includes all of academic 2012, 2013, 2014, 2015, 2106 and also academic 2017 from September 2017 to March 2018.

This means we know three things for each title:

- the number of copies (this is actually "number of items with the same call number," so multiple copies of the same title are counted together, but four separate volumes in a four-volume set are not);

- the number of circs of all those copies in the circ window; and

- the length of circ window (defined as 5 years).

We can combine these into a measure of "circs per copy per year," or how many times a year (on average, over the last five years) each copy of a title circulates.

Define the *busy factor* to be number of circs divided by number of copies divided by length in years of the circ window.

$$busy = \frac{\frac{circs}{copies}}{years} \tag{1}$$

Some examples:

| circs | copies | window | busy |
|------:|-------:|-------:|-----:|
| 10 | 1 | 5 | 2 |
| 5 | 1 | 5 | 1 |
| 7 | 2 | 5 | 0.7 |
| 2 | 1 | 5 | 0.4 |
| 1 | 2 | 5 | 0.4 |
| 2 | 19 | 5 | 0.02 |
| 2 | 698 | 5 | 69.8 |

The last two are real examples: Francis Bacon's Advancement of learning, and New Atlantis has a busy factor of 0.02, and Economics: Canada in the Global Environment is at 69.8, probably because it used to be on reserve but is now in the stacks).

Further, let us set a *target busy factor* of 1. That means in an ideal situation, if we have more than one copy of a book, each is used once a year. (If we only have one copy of a title then Easy Weeder is not used, so its busy factor is not used.)

## Assumptions

1. We don't keep two copies of a book in the circulating collection just in case one gets lost.

2. To minimize human oversight, we will always keep at least 1 copy of a title. Weeding all copies of a title goes beyond Easy Weeder's scope.

3. The inventory is accurate. (If it is not, it can be corrected as weeding is done, and anyway that would rarely affect the number of copies that Easy Weeder will recommend: if we think we have four copies of something but only need one, it doesn't matter that really we only have three.)

## How the logic and calculations work

To revisit the examples above, two are ignored because their busy factor is >= 1. For the remaining four we can calculate the recommended number of
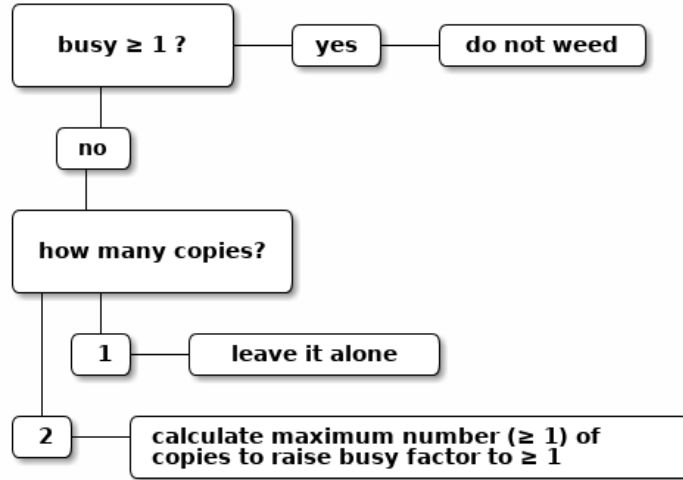
Figure 1: Easy Weeder logic

copies we should keep. Figure 1 shows the decisions made.

For example, for the one with 19 copies, and 2 circs, the calculation goes:

- 19 copies is too many. What if we had 18? Busy factor: 0.022.

- 18 copies is too many. What if we had 17? Busy factor: 0.024.

- ...

- 5 copies is too many. What if we had 4? Busy factor: 0.1.

- 4 copies is too many. What if we had 3? Busy factor: 0.133.

- 3 copies is too many. What if we had 2? Busy factor: 0.2.

- 2 copies is too many. Keep 1 copy.

As it happens, for all four we end up keeping just one copy.

| circs | copies | window | busy | rec_copies |
|-------|--------|--------|------|------------|
| 7 | 2 | 5 | 0.7 | 1 |
| 2 | 1 | 5 | 0.4 | 1 |
| 1 | 2 | 5 | 0.4 | 1 |
| 2 | 19 | 5 | 0.02 | 1 |

3

But, for example, let's say we had a book that had six copies and 12 circs over the last five years, for a busy factor of 0.4. We would end up keeping 4 copies.

- 6 copies is too many. What if we had 5? Busy factor: 0.48.

- 5 copies is too many. What if we had 4? Busy factor: 0.6.

- 4 copies is too many. What if we had 3? Busy factor: 0.8.

- 3 copies is too many. What if we had 2? Busy factor: 1.2.

- Keep 2 copies. (2 is the largest number giving a busy factor $>= 1$.)

## Estimated easy weedables

| home_location | n |
|---|---:|
| BRONFMAN | 901 |
| FROST | 2074 |
| LAW | 556 |
| SCOTT | 120428 |
| STEACIE | 984 |

## To note

- Easy Weeder does not use acquisition year. Low-circing multiple copies acquired within the last five years may need a human decision.

- Easy Weeder does not remove duplicates across branches, but the same data can be used to do that. For example, Captains of Consciousness: Advertising and the Social Roots of the Consumer Culture has ten copies at three branches: Bronfman (7), Frost (1) and Scott (2). Circulation in the circ window is: Bronfman (13), Frost (0), Scott (15). Therefore Easy Weeder recommends this number of copies: Bronfman (2), Frost (1) and Scott (2). Even if it makes sense to have the book at both Bronfman and Scott (which would need 3 copies to get down to a busy factor of 1), Frost does not need a copy, because no one has used it there since 2010.

## Commands

`make` ???

`make push_app` use locally to push the Shiny app up to production

## Installation