

Stat405

Development

Hadley Wickham

1. Floating point math
2. Optimisation
3. Continuing education
4. Feedback

Your turn

Perform the following calculations in R.
Are the answers what you expect?

```
seq(0.1, 0.9, by = 0.1) - 1:9 / 10
```

```
sqrt(2)^2 - 2
```

What is the property of these numbers
that might cause the problem?

Each number must be stored in a finite amount of space
=> each number can only have a finite number of digits
=> floating point math does not work like normal math

$$(1e-16 + 1) == 1$$

$$(1e-16 + 1) * 10 == 1e-16 * 10 + 1 * 10$$

$$1e9 + 2 - 0.1 - 1e9$$

$$1e10 + 2 - 0.1 - 1e10$$

$$1e11 + 2 - 0.1 - 1e11$$

$$1e12 + 2 - 0.1 - 1e12$$

$$1e13 + 2 - 0.1 - 1e13$$

$$1e14 + 2 - 0.1 - 1e14$$

```
# By default R only shows 7 significant digits
# If the trailing digits are zero, the number will be rounded
(1 / 237)
(1 / 237) * 237
(1 / 237) * 237 - 1

seq(0.1, 0.9, by = 0.1)
seq(0.1, 0.9, by = 0.1) - 1:9 / 10

# Tricky to get to print exactly:
formatC((1 / 237) * 237, digits = 20)
formatC(seq(0.1, 0.9, by = 0.1), digits = 20)
```

```
# When working with floating point numbers (numeric)
# (but not integers, this is the one place where the
# difference is important) never test for equality with ==
```

```
a <- seq(0.1, 0.9, by = 0.1)
```

```
b <- 1:9 / 10
```

```
all(a == b)
```

```
all.equal(a, b)
```

```
all(abs(a - b) < 1e-6)
```

```
# Similarly, need to be careful with < and > etc
```

```
# Places where this matters:  
#  
# * sums  
# * calculating the standard deviation  
# * inverting a matrix (condition)  
# * linear models!  
# * maximum likelihood estimation
```

Optimisation

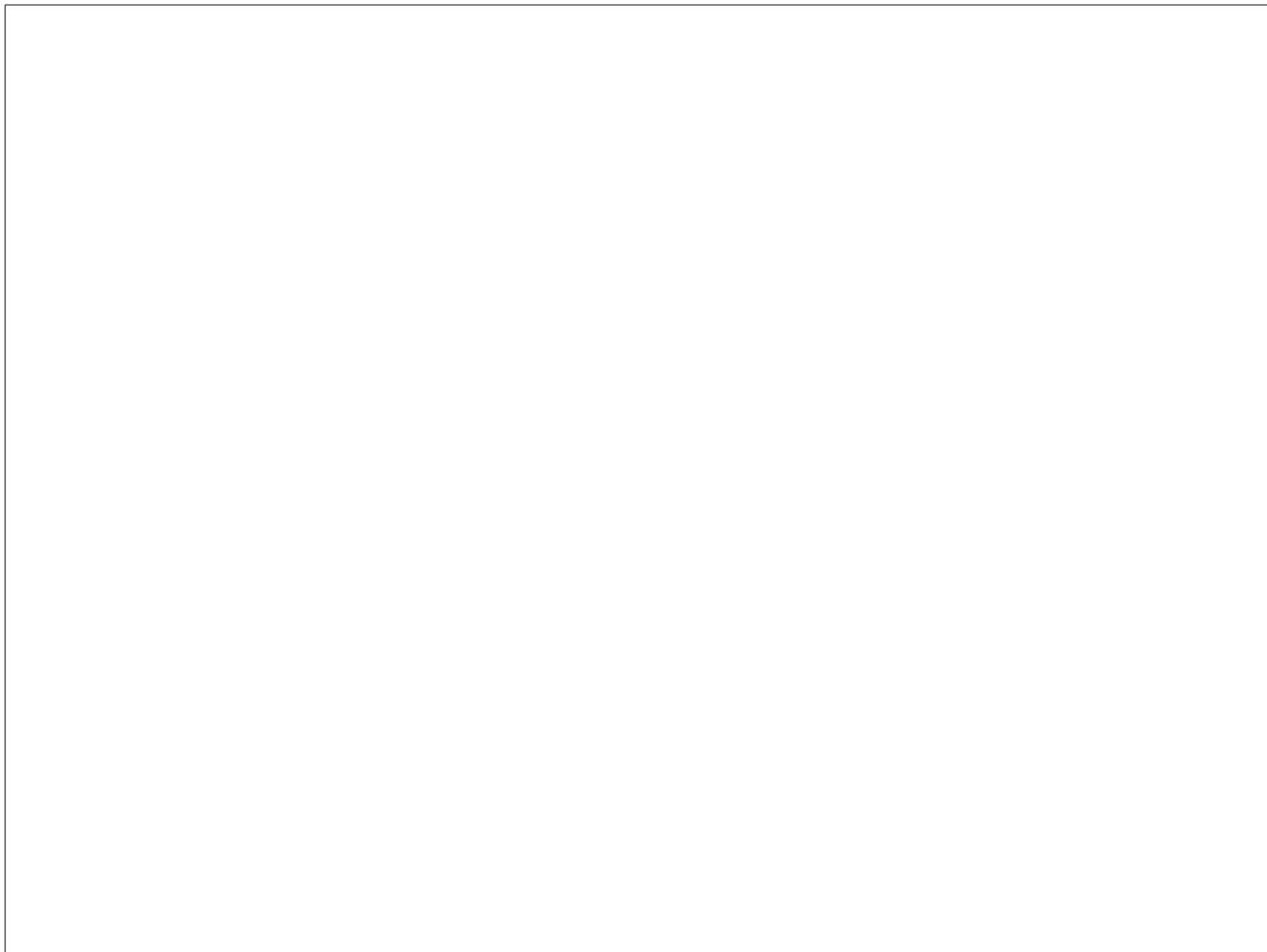
If, and only if, your code is too slow

First use `system.time()` to figure out exactly how long things are taking: you need this so you can check your changes actually speed things up

Then see what is taking the longest amount of time with the `profR` package

General advice

- Start with the slowest part of your code
- Use built-in R functions, where possible
- Use vectorised functions, where possible
- Think through your basic algorithm
 - Knowledge of basic CS algorithms and data structures v. helpful



Continuing education

Learn more about R.

Learn more about your other tools.

Professional development

Mailing list

Sign up to R-help: <https://stat.ethz.ch/mailman/listinfo/r-help>

Make sure to set up filters

Skim interesting subjects and read them

Don't be afraid to post
(use a pseudonym if necessary)

Read books

Phil Spector. Data Manipulation with R.

William N. Venables and Brian D. Ripley.
Modern Applied Statistics with S.

Frank E. Harrell. Regression Modelling
Strategies.

Jose C. Pinheiro and Douglas M. Bates.
Mixed-Effects Models in S and S-Plus.

Read papers

The R Journal: <http://journal.r-project.org/>

The Journal of Statistical Software: <http://www.jstatsoft.org/>

Learn your tools

- Touch typing
- Text editor
- Command line
- Caffeine
- Email

Professional development

The aspects of being a statistician, apart from knowing statistics.

Principally communication: written, spoken, visual and electronic.

Take every opportunity you can to practice these skills.

	Visual	Electronic
Written	Posters Graphics	Email Website Blog Code
Spoken	Oral exam	Video Slidecast

Written

Papers

Vita/Resume

Bibliography

Reviews

Visual

Posters

Graphics

Electronic

Email

Website

Blog

Code

Spoken

Teaching

Short talk

Long talk

Oral exam

Video

Slidecast

Written

Particularly important if you want to be an academic, or if you're PhD student, or want to become one.

“Style: Toward Clarity and Grace”

Sign up for the thesis writing workshops when they come around.

Develop a regular habit.

My habit

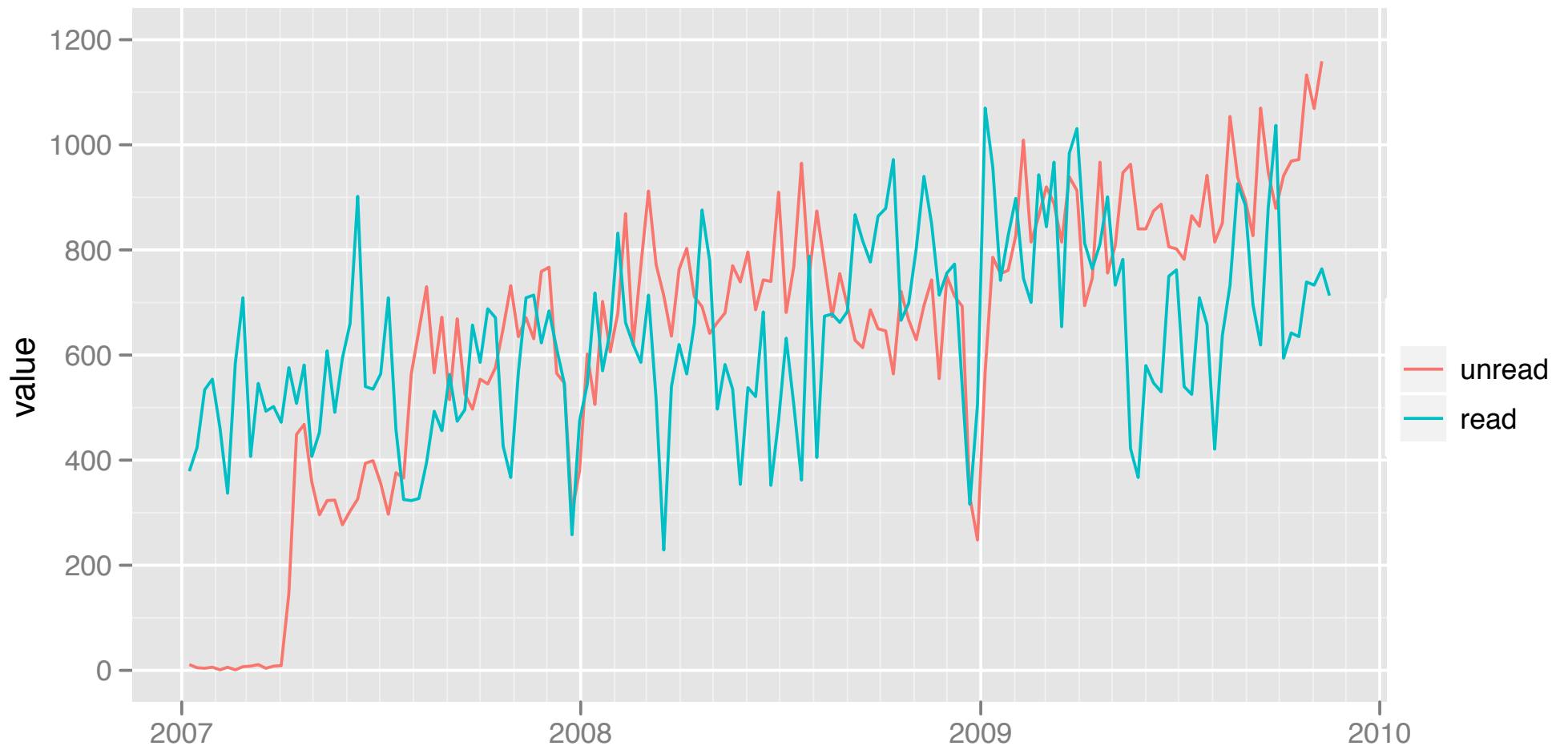
- Roll out of bed at 7am
- Boil water
- Make tea
- Drink tea
- Write for an hour

Spoken

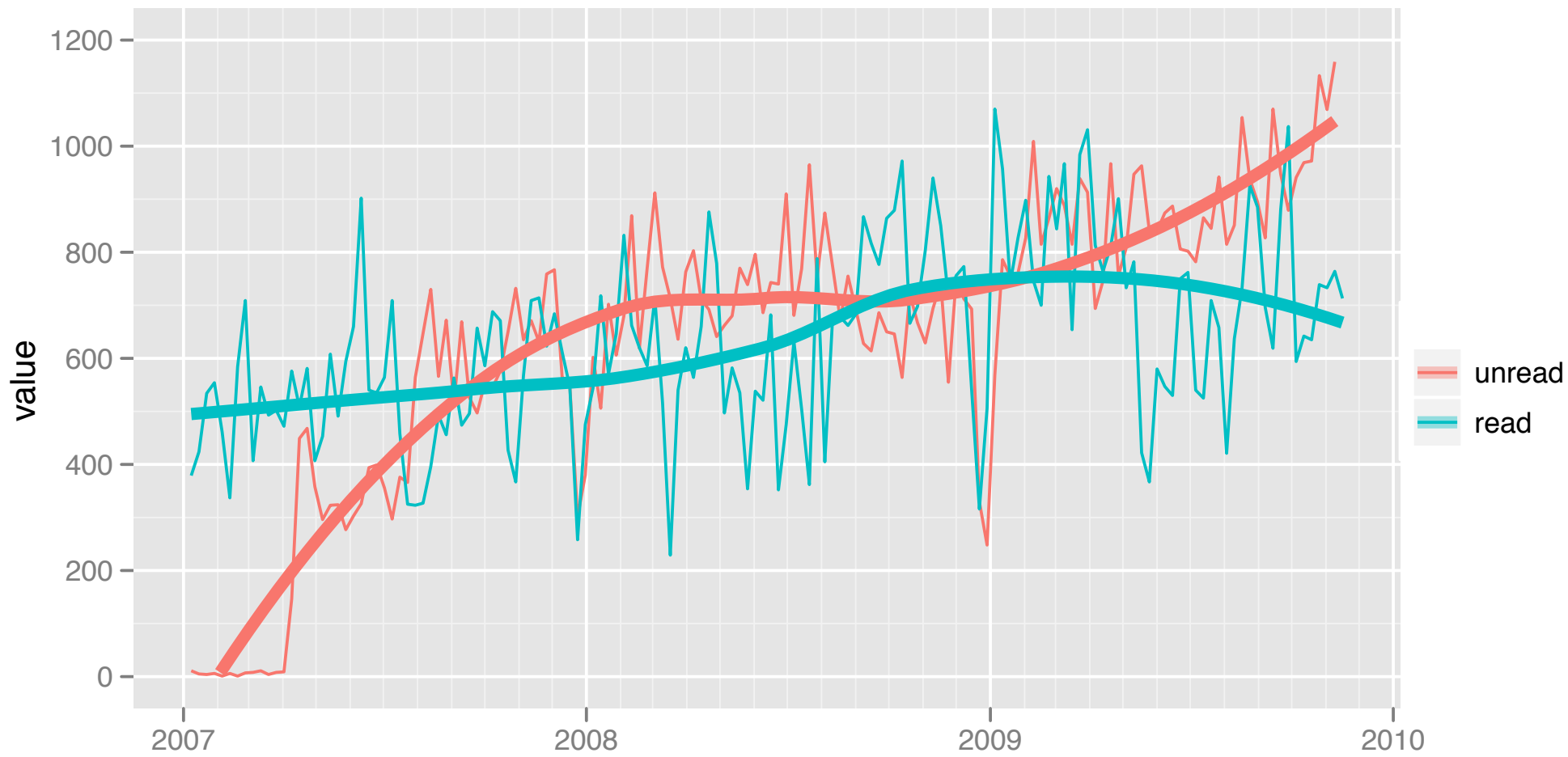
Seize every opportunity to practice.

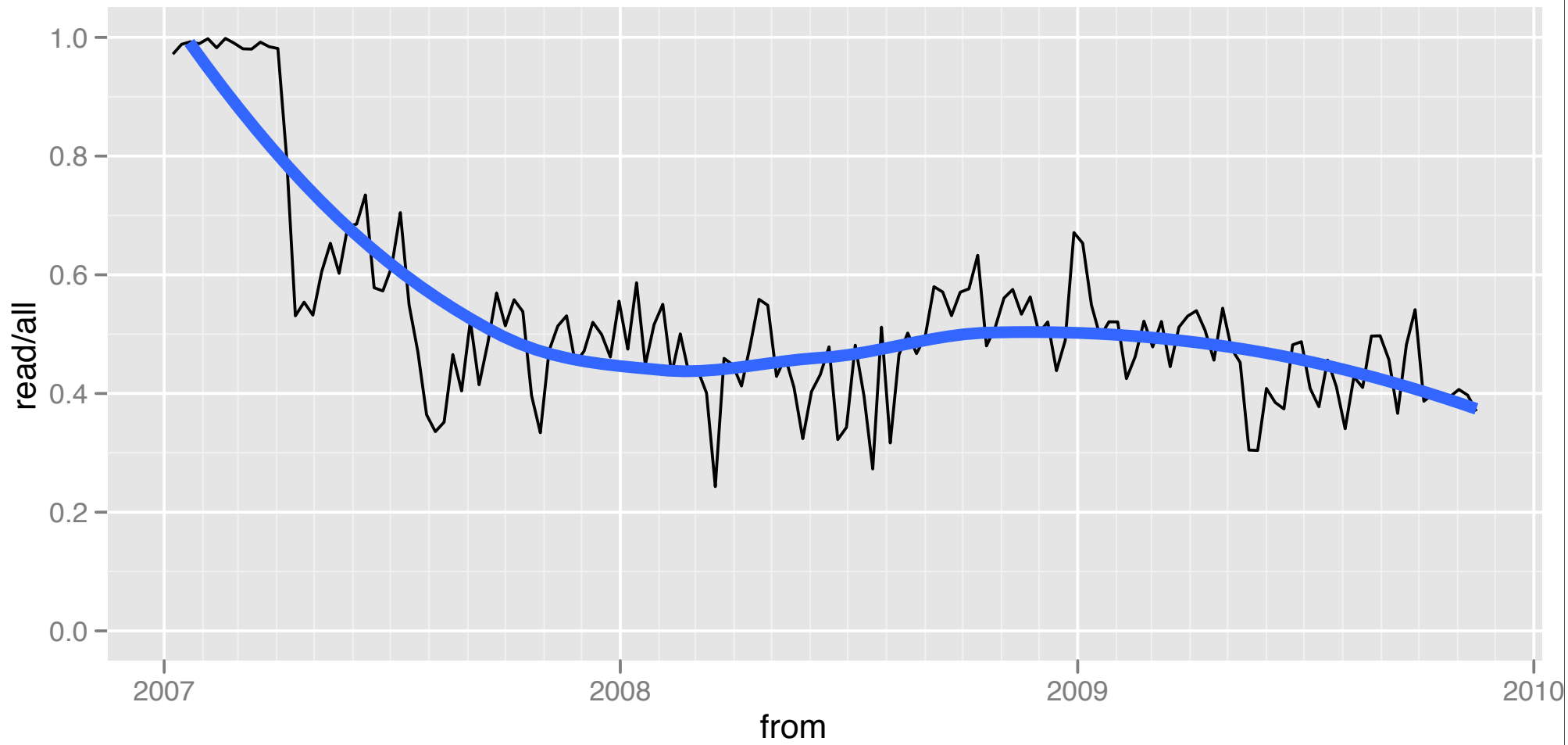
Make use of Tracy Volz - tmvolz@rice.edu.
She is a fantastic resource - if you had to pay for her, you wouldn't be able to afford it.

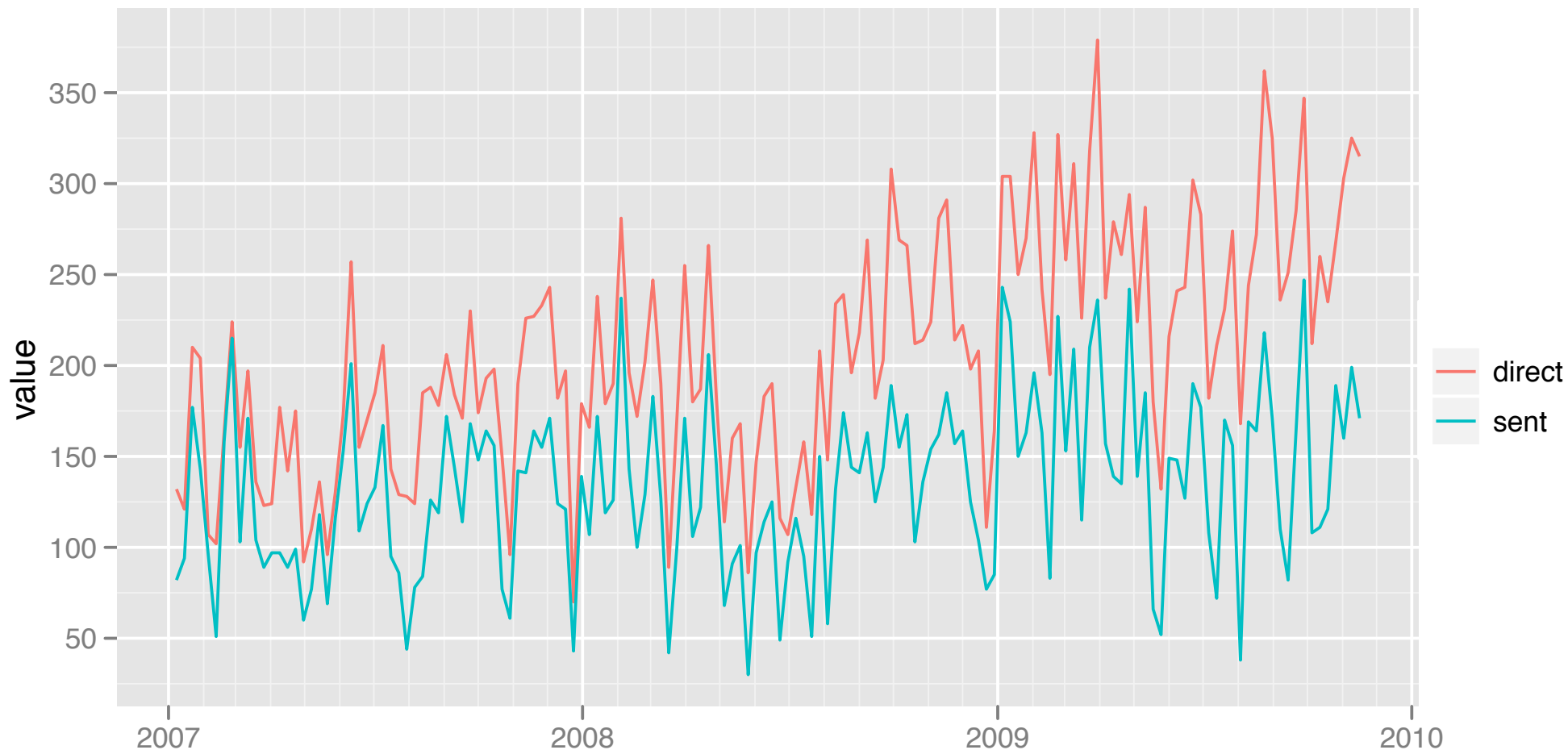
Email

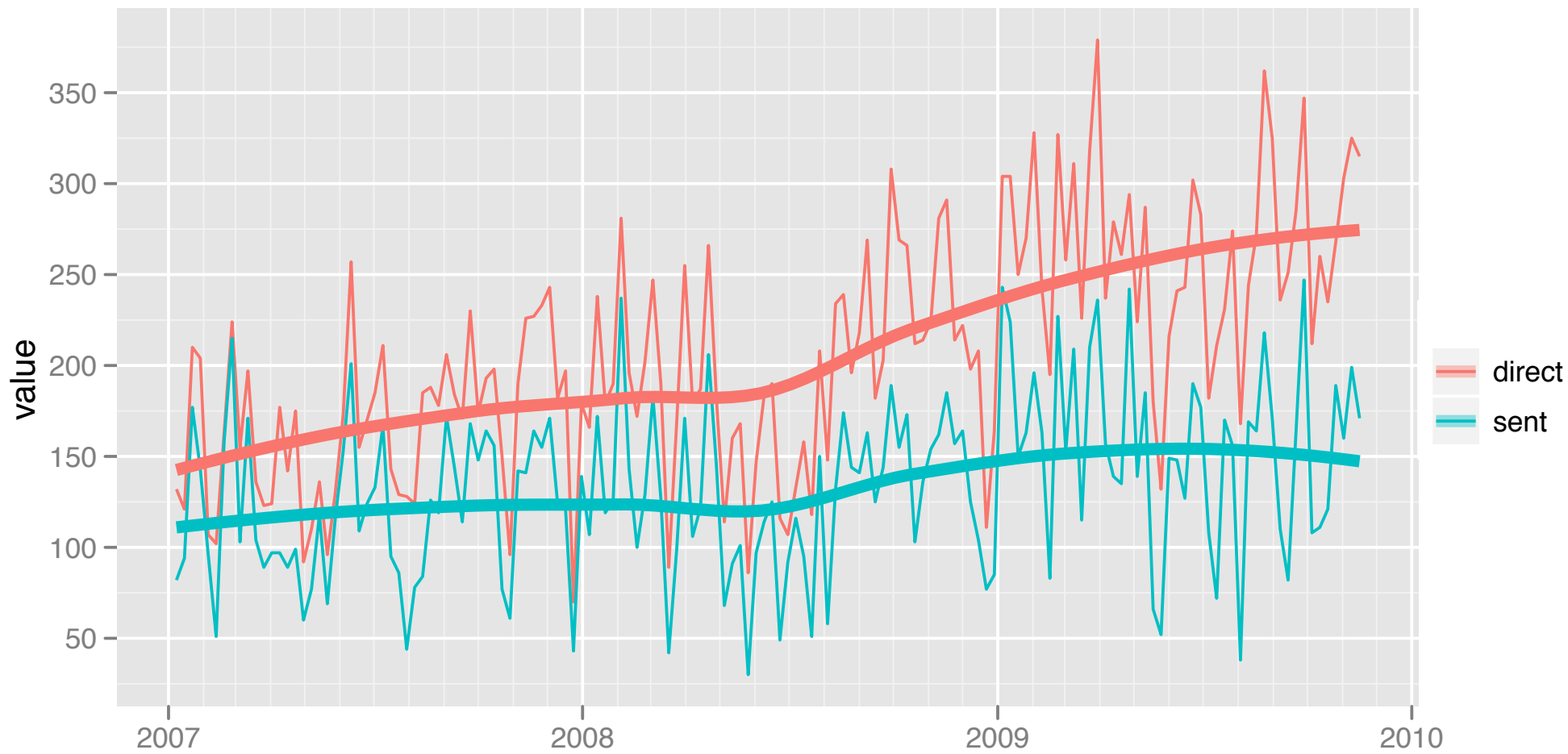


265,000 emails
134,000 unread!









Inbox Zero

<http://www.43folders.com/izero>

Merlin Mann

There is no way you will ever be able to respond to — let alone read in exquisite detail — every email you ever receive for the rest of your life. If you take issue with this, just wait six months, because, believe me, we're all getting a lot more email (and other sundry demands on our attention) every day. What seems like a doddle today is going to get progressively more difficult — even insurmountable — unless you put a realistic system in place now.

Your time is priceless
(and wildly limited)

You need an agnostic system for
dealing with mail that isn't based on
nonces, exceptions, and guilt.

[The] ultimate goal is for you to spend
less time playing with your email and
more time doing stuff.

Key concepts

Regularly empty your inbox

Minimal response

Delete, delete, delete

Filters

Email dashes

Response does **not** need to be
proportional to request

“Do you still need this?”

“I don’t know”

“Good idea. I’ll add it to my to do list.”

“Here’s a link that might be what you’re
looking for...”

[Delete]

Delete!

Most minimal response is none.

“Just remember that every email you read, re-read, and re-re-re-re-re-read as it sits in that big dumb pile is actually incurring mental debt on your behalf.”

Be brutally honest - if you're not going to do anything with the email delete it now.

Filters grey mail

“noisy, frequent, and non-urgent items which can be dealt with all at a pass and later.”

facebook, comments, university/
department memos, newsletters, mailing
lists

Good catch all: contains **unsubscribe**

<http://www.43folders.com/2006/03/13/filters>

1300/3500
(5/day!)

bannerpcard@rice.edu, carlyn@rice.edu,
cchat@rice.edu, cmtcomment@rice.edu,
giving@rice.edu, payroll@rice.edu,
registrar@rice.edu, sandra@rice.edu,
sallie@rice.edu subject:(weekly message),
alldepts@rice.edu, list:"k2i-members.rice.edu",
list:"mailman.rice.edu"

allfaculty@stat.rice.edu, faculty@stat.rice.edu,
statdept@stat.rice.edu, colloquium@stat.rice.edu,
undergrad@stat.rice.edu

from:(statements@wageworks.com)

from:(TIAA-CREF_eDelivery@tiaa-cref.org)

Patricia Wallace, a techno-psychologist, believes part of the allure of e-mail—for adults as well as teens—is similar to that of a slot machine. “You have intermittent, variable reinforcement,” she explains. You are not sure you are going to get a reward every time or how often you will, so you keep pulling that handle.”

Email dashes

Don't have your email open all day. Schedule times when you respond to emails.

You can tackle emails a lot faster when you batch them up.

Lack self control (like me)? Try an internet blocker: <http://macfreedom.com/>

<http://www.43folders.com/2006/03/15/email-dash>

Feedback

[http://hadley.wufoo.com/
forms/stat405-final-feedback/](http://hadley.wufoo.com/forms/stat405-final-feedback/)