

Pretty pictures we need for the paper, *publication quality*

RBLandau 20180808

Notes on pictures:

- Aspect ratio 16:10, raw size 1600x1000 or greater for publication, 800x500 for easier viewing while editing. [Do journals have standards for sizes and aspects? *Science* does, but I've never looked at others.]
- All log-log.
- Almost all need decade lines for loss rates: 1%, 0.1%, 0.01%, I think. Horizontal dotted lines with small labels. Would be good if the lines were easily distinguished, e.g., by weight, but not clear that's possible. What would Tufte say?
- X half-life scales 2-1000 or 2-100.
- Legends upper right (if I can force legends at all; may have to be manual annotations, yuck).
- Attempt consistent color labeling for number of copies, but not sure this is reasonable in R.
- Describe data plots: #copies, losses, error injection, log scales x and y, percent losses, trimmed means (25%=midmean, maybe only 10%) of sample size=21, colors=black or blue usually best case recommendation.
- Describe why low half-lives and why 3,4 copies: otherwise numbers too small to make clear pictures, best to look at where the differences are exaggerated but parallel.
- Note that one would never actually use disks with such low half-lives, but they are needed to show the phenomena.

(Note: the numbers are just for reference during discussions and editing.)

idea / assertion		basics	details	wayzit
1	for copies=1 losses are unacceptable and auditing can't help	cop=1, lifem=2-1000	line at 0.1% or so maybe decades below 1%, many annotations	

	idea / assertion	basics	details	wayzit
2	without auditing, need many copies to minimize losses	cop=1,2,3,5,10 lifem=2-1000	1% line and 0.1%, clear legend for n-copies	
3	with annual auditing, need only a few copies	cop=3,4,5 lifem=2-1000	1% line and 0.1% line, shorten lifem=2-100	
4	is 5 sufficient for longer periods, 30-50 years?	lifem=2-1000	lines 1% and decades below	
5	is 4 sufficient in calm periods?	cop=4,5 lifem=2-100, various audit methods	decade lines for all graphs, I think	
6	are 6 necessary in shock periods?	cop=5,6 lifem=2-100 shocks f=2yr dur=1yr 50, 67, 100%	see which is most striking visually	
7a	even generous random auditing still worse than total auditing	WITH vs WITHOUT replacement, cop=3 maybe also 5	one segment total, 4 or 12 segments random	
7b	include very sparse random auditing vs total	WITH replacement, cop=3,4,5	random 20% per year, yikes	
8	frequent segments improve survival slightly in no-shock			
9	frequent segments help a lot in high-shock			
10	increase total audit frequency vs segmentation	cop=3,4,5 lifem=2-100		

idea / assertion		basics	details	wayzit
11	larger docs are larger targets for random errors	docsize=5,50,500,5000MB usual lifem, losses are 100% for large docs at low half-life	legend must be clear	
12	TABLE: tradeoff of docsize vs error rate	docsize5-5000 lifem=2-10000	nicely shaded png or pdf of spreadsheet excerpt; graph overlays too much to make the point	
13	TABLE: threats to content, from google doc	make it look like the other tables, bad word wraps	redo in md so it matches if possible to change text color emphasis	
14	glitch just like higher error rate	cop=5 show line shifted to lower half-life with significant glitch	probably one month per quarter 50% or 67% increase in error rate, cherry-pick for clearest appearance	
15a	shocks: moderate not tolerated with 3 or 4 copies, the current pic is not too bad		redo with larger samples to smooth out the numbers	
15b	severe shocks may require 6 copies	copies=4,5,6	freq2yr dur1yr span3 imp50,80	
15c	severe shocks may require 6 copies	copies=4,5,6	freq2yr dur1yr span2 imp50,80	
15d	severe shocks may require 6 copies	copies=4,5,6	freq1yr dur0.5yr span2 imp50,80	

idea / assertion		basics	details	wayzit
16	how many copies for < 1% long loss? no audit	long=30yr		
17	how many copies for < 1% long loss? annual audit	long=30yr		
18	how long before lose 1% of collection	lifem-losses plot	separate lines for 10,30,50 years, copies=3,4,5 with annual audit	
19	TABLE: calibration test of simulation results	theoretical vs simulated results	copies=1, 100 samples, vs straight Poisson, in ppm; have in spreadsheet, get nice png or pdf	tables/calibration
20				
21				