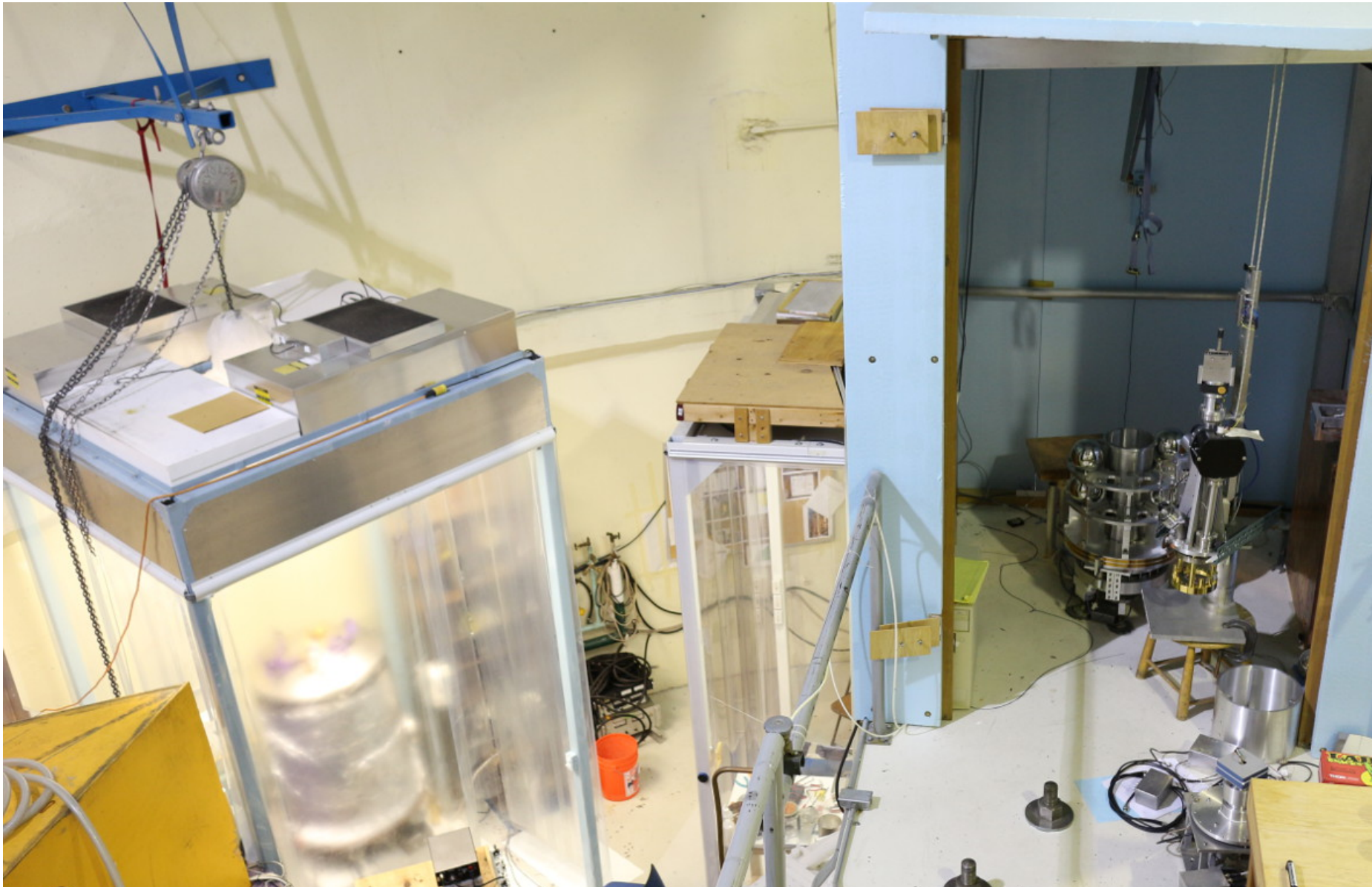
Replicable Analysis and Blind Review

Charlie Hagedorn

October 10, 2014

Who am I?...

- Senior grad student finishing up a short-range inverse square law test



Two topics

- Replicable Analysis
  - It's possible to go from raw data to final publication with a single computer command.
  - That process can be shared along with the publication.
- Blind Review
  - It is possible for a measurement to remain blind until all concerned parties are convinced of correct execution.
  - Irreversible unblinding can be distributed.
  - Opinion: Correct execution of a measurement does not depend upon its final central value.

Replicable Analysis

- Elementary School: "Showing your work"
  - Provenance
  - Partial Credit
- Allows you to answer the question, "So, how exactly did you come up with that number?"
- Only become practical in last 10-20 years
  - Fast computers
  - Fast internet
  - Great software
- Many have written on reproducible research software:
  - *The case for open computer programs,* Ince, Hatton, and Graham-Cumming, etc...
  - http://researchcompendia.org/ , etc....

Replicable Analysis

- From data to publication with a single command.
- I need to start an example program while I tell you about a small experiment.

A small experiment

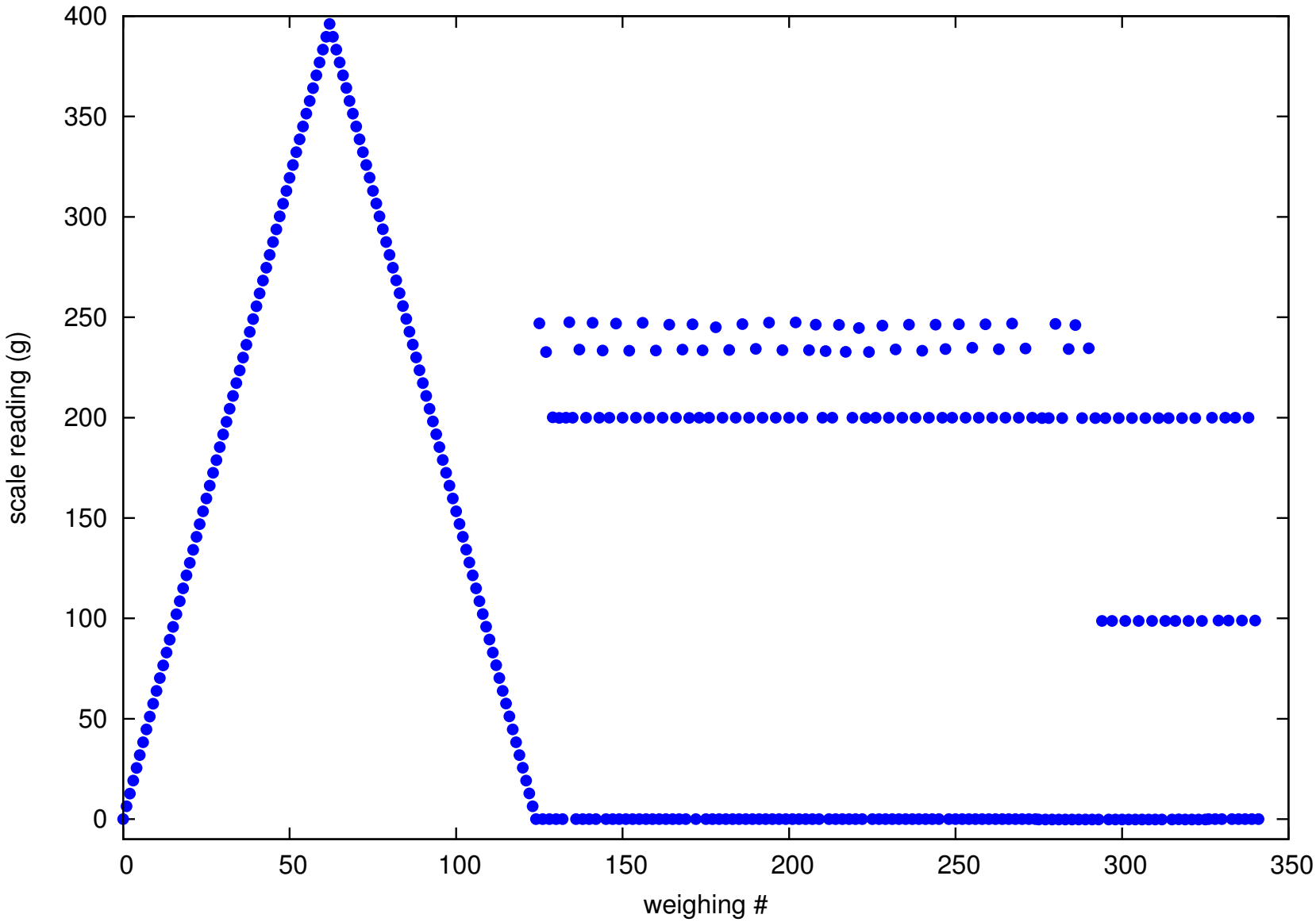
- Measure the density of Seattle City water.

Experimental method

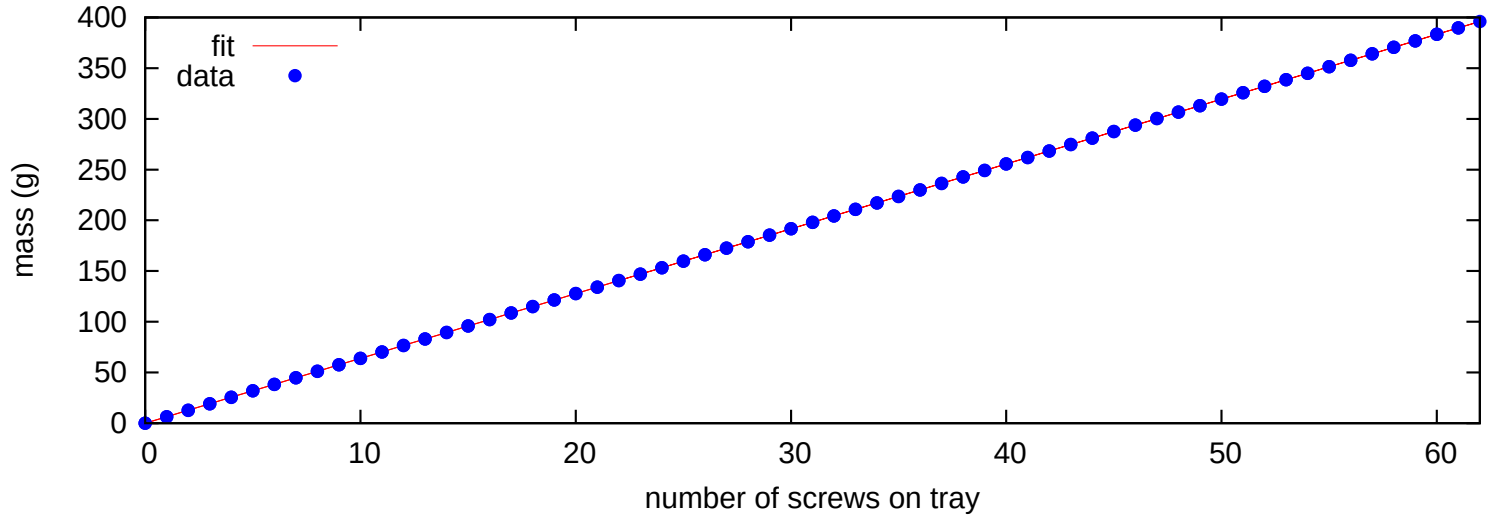
- Weigh plastic mass M
- Alternate weighing to extract mass of displaced water W
  - calibration mass
  - cup with water $(W_0 + W)$
  - calibration mass
  - cup with water+plastic $(W_0 + M)$
  - calibration mass
  - ...
- Using measured volume V of plastic mass M,

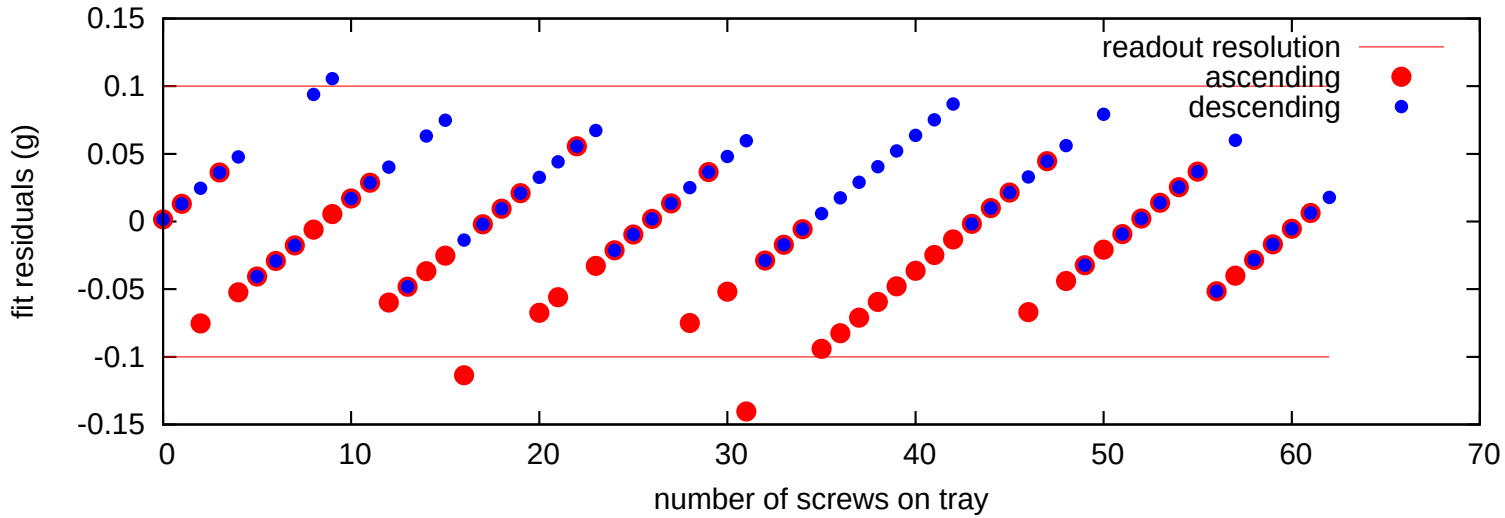
$$\rho = \frac{(W_0 + W) - ((W_0 + M) - M)}{V} = \frac{W}{V}$$

Raw Data

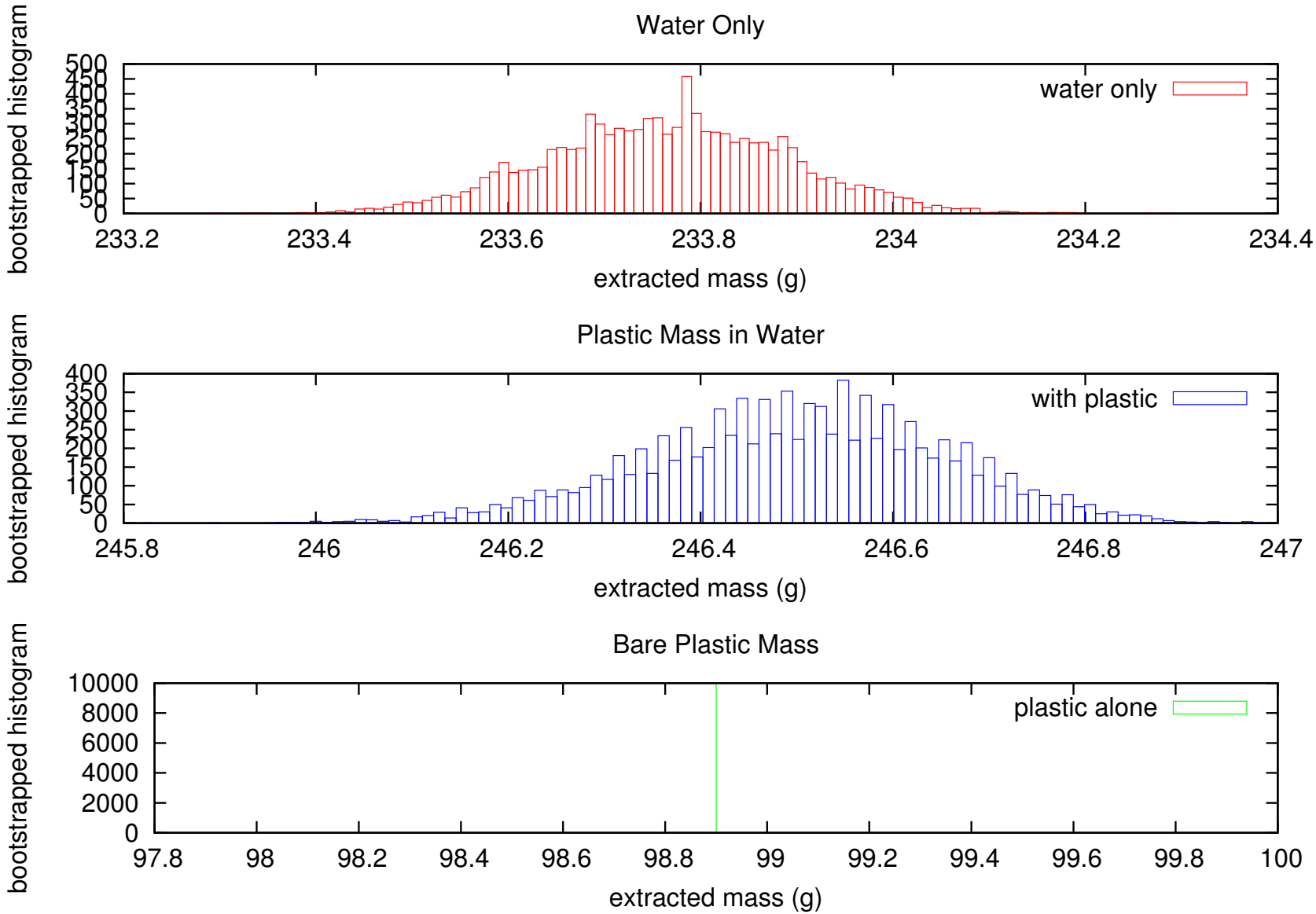


Linearity


fit
data
mass (g)
number of screws on tray
0
50
100
150
200
250
300
350
400
0
10
20
30
40
50
60


readout resolution
ascending
descending
fit residuals (g)
number of screws on tray
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
0
10
20
30
40
50
60
70

Bootstrapped uncertainties....



Result.... before unblinding

Quantity	Magnitude
Auto-zeroing	0.1 g
Scale nonlinearity	0.1 g
Hysteresis	0.1 g
Buoyancy, plastic mass	110 mg
Dissolved solids/impurities	50 mg
Buoyancy shift, calibration mass	2×10^{-4}
Plastic Cylinder Mass	(99.01 ± 0.17) g
Displaced Water Mass	(86.27 ± 0.36) g
Plastic Cylinder Diameter	(44.550 ± 0.056) mm
Plastic Cylinder Length	(55.520 ± 0.074) mm
Thermal volumetric expansion	2×10^{-3}
Plastic Cylinder Volume	(86544 ± 451) mm^3
Water Density	(996.9 ± 6.9) kg/m^3

Expected: 997 ± 3 kg/m^3.

Freely available.

- Github

Software tools

- **All tools are freely available, with auditable source code**
- Git - a source-code management system
- GNU Make - manages the recipe
- GNU Octave - does most of the computation
- Gnuplot - nicer plots
- LyX/LaTeX - typesetting
- Assorted GNU/Linux standard tools (shell scripts, sed, etc.)
- **How to include a number? TeX child documents.**
  - $(996.9 \pm 6.9)\ \mathrm{kg/m^3}$
  - `$(996.9 \pm 6.9)$~kg/m$^{3}$`

Unblinded!

Quantity	Magnitude
Auto-zeroing	0.1 g
Scale nonlinearity	0.1 g
Hysteresis	0.1 g
Buoyancy, plastic mass	110 mg
Dissolved solids/impurities	50 mg
Buoyancy shift, calibration mass	2×10^{-4}
Plastic Cylinder Mass	(99.01 ± 0.17) g
Displaced Water Mass	(86.27 ± 0.36) g
Plastic Cylinder Diameter	(44.550 ± 0.056) mm
Plastic Cylinder Length	(55.520 ± 0.074) mm
Thermal volumetric expansion	2×10^{-3}
Plastic Cylinder Volume	(86544 ± 451) mm^3
Water Density	(996.9 ± 6.9) kg/m^3

Expected: 997 ± 3 kg/m^3.

Replicable Analysis

- Strengths:
  - Repeatable, clear analysis that's easy to audit.
  - You catch your own errors.
  - Encourages clarity
  - It's liberating. Your responsibility is quality, above all else.
- Weaknesses:
  - It is work, and it takes time.
  - Hardware execution cannot be documented with comparable precision.
  - Can get trapped in a framework as analysis evolves.
  - It's intimidating, as everyone will see your work.
- Replicable Analysis
  - It's possible to go from raw data to final publication with a single computer command.
  - That process can be shared along with the publication.

Blind Review

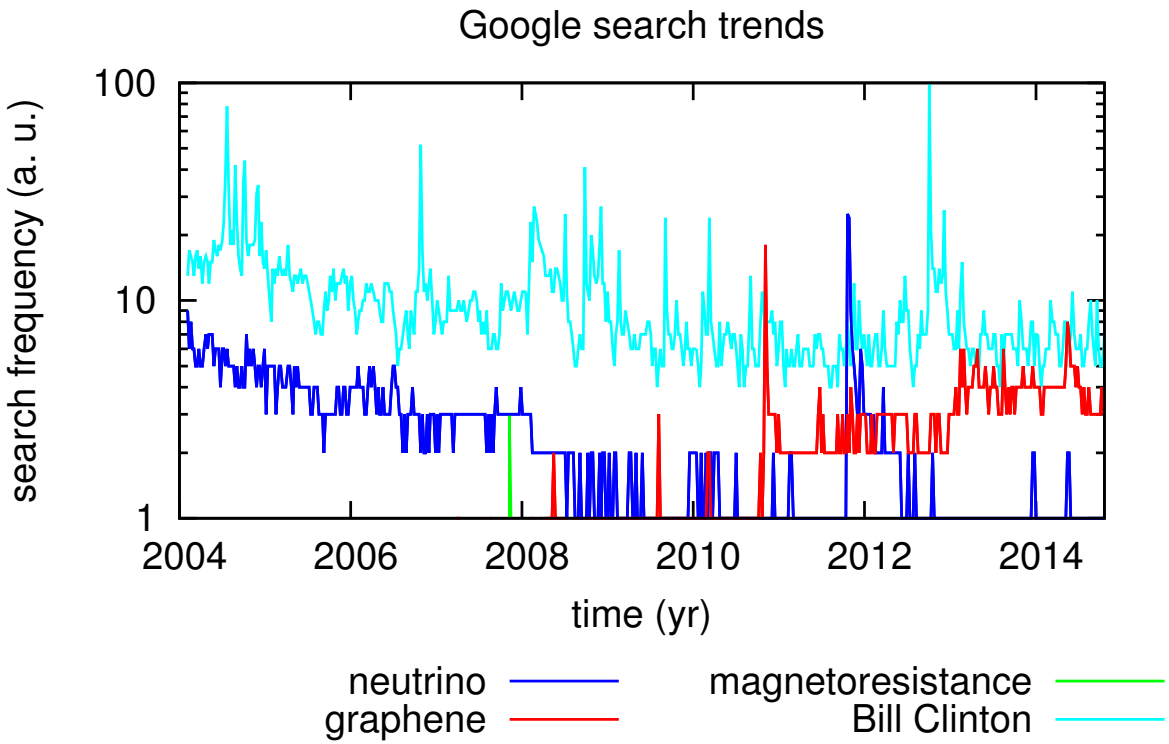
- How do I know when to stop looking for systematics in a null measurement?
- Reproducibility problem
  - G has one, but many others do/did too (Neutron lifetime, etc.)
  - Amgen (Begley and Ellis) found that 47/53 "landmark" cancer papers could not be reproduced
- Too few negative results in the literature.

Important objections to blind analyses

- "No experiment is ever truly blind."
  - If there's a gross error after the opening of a blind, you'll revisit the measurement before publication.
  - Agreed: **Unblinding must be irreversibly public.**
- "Just because a measurement is blind doesn't mean that it's correct."
  - Blindness alone is insufficient to guarantee anything.
  - Agreed: **There is no substitute for quality work. Blinds are tools to temper our own flaws.**
- "If a measurement is blind, it's harder to find problems."
  - Comparison with known results yields efficiency gains.
  - Agreed: **Yep. Blind measurements force consideration of all possible errors.**
  - Upside: **You stop looking for systematics when you've considered everything you can think of.**

Risks and Consequences in precision measurement

- Measurements that agree with the status quo are safe.
  - Higher precision $\Rightarrow$ impress peers, get/keep a great job
- Measurements that disagree with the status quo are scary.
  - Small chance of major revolution and big payoff
  - Long term consequences of failure



Blind Review

Chalkboard time...

- Strengths
  - Allows blind peer review to the highest hierarchy of error-checking
  - Asynchronous irreversible public unblinding
  - Blind is immutable, but analysis may change; even a nefarious actor may be uncovered.
  - Shared responsibility; (acknowledge accepting referee?)
  - Requires no change to existing structures.
  - Results accepted for quality, sensitivity and import, not statistical significance of the result.
- Weaknesses
  - Still depends upon a strong/unmeasureable blind.
  - Requires time and care.

Two topics

- Replicable Analysis
  - It's possible to go from raw data to final publication with a single computer command.
  - That process can be shared along with the publication.
- Blind Review
  - It is possible for a measurement to remain blind until all concerned parties are convinced of correct execution.
  - Irreversible unblinding can be distributed.
  - Opinion: Correct execution of a measurement does not depend upon its final central value.