

Computer Science 50  
Fall 2010  
Scribe Notes

Week 10 Wednesday: November 10, 2010  
Andrew Sellergren

## Contents

<b>1</b>	<b>Announcements and Demos (0:00–7:00)</b>	<b>2</b>
<b>2</b>	<b>A Short Course in Numeric Self Defense (7:00–58:00)</b>	<b>3</b>

## 1 Announcements and Demos (0:00–7:00)

- This is CS50.
- 0 new handouts.
- Check out [The Wilderness Downtown](#), a very cool customizable film built using HTML5 and the Google Earth API.
- The next installment of CS50 Lunch will take place this Friday. If you'd like to attend, please [RSVP](#).
- Now's the time to [apply](#) to be a TF or CA for next year's CS50! Don't let your CS50 fun end after just one year!
- [HarvardLunch](#), created by a former CS50 student, has just been released! If you'd like to meet new people, just plug in your name and e-mail address and you'll be randomly paired with another student with whom you can have lunch!
- If you're inclined to register your own domain for your final project, we recommend GoDaddy as a registrar. Although their marketing is a bit over the top, their interface for administering your domain is actually very powerful. We can take care of hosting the website for you on the CS50 Cloud. You'll only need to tell GoDaddy or whichever registrar you choose to use different nameservers.
- If you're interested in meeting others who are interested in a specific topic in technology, take a look at these [mailing lists](#) to see if there's one or more you'd like to join.
- Have an amazing idea that you need some funding for? Condense it to 90 seconds and enter the [Elevator Pitch Competition](#) for a chance to win \$1000.
- If your final project aims to improve student life at Harvard and you'd like it to live on beyond the end of CS50, consider applying to [HackHarvard](#), a self-described web app incubator!
- [TECH](#), or Technology and Entrepreneurship Center at Harvard, is an excellent resource for all things entrepreneurial.
- The [Harvard College Innovation Challenge](#) is coming up soon. Enter for a chance to win some of the \$50,000 worth of prizes.
- If you're interested in any of the things I just typed out, you might consider joining the [Harvard College Entrepreneurship Forum](#).

## 2 A Short Course in Numeric Self Defense (7:00–58:00)

- Let's give a warm CS50 welcome to Brian Kernighan, the very same professor who taught CS50 to David in 1996! Oh, and besides that, he also literally wrote [the book](#) on C.
- In his time at Princeton, Professor Kernighan has been teaching a course called Computers in Our World, roughly the equivalent of the Bits course here at Harvard. The course is intended to empower you to make sense of all the numbers that are floating around these days.
- Let's begin with a simple question that arises from a suggestion that we use the Strategic Petroleum Reserve's oil to stay the rising costs of gasoline:
  - “Perhaps the Bush administration could use the 660-billion-barrel Strategic Petroleum Reserve to push prices down.” (Newsweek, 5/24/04 (when gas was under \$2/gallon)
  - “The average vehicle uses 550 gallons/year.”

Given these data, how long would this oil reserve last?

- To answer this question, we need to first answer two others:
  - How many vehicles are there?
  - How big is a barrel?

Here, we'll make some assumptions. First, let's assume there is one vehicle per person in the United States, which means there are about 300 million vehicles total. Second, let's assume that a barrel is about 50 gallons, roughly the same volume as a 55-gallon drum. Given these assumptions, the math works out like so:

$$\begin{aligned} 500 \text{ gallons/year} &= 10 \text{ barrels/year} \\ 300 \text{ million cars} \times 10 \text{ barrels/car} \\ &= 3 \text{ billion barrels/year} \\ &= 200 \text{ years before SPR runs out} \end{aligned}$$

So what are we worried about then? Well, it turns out that Newsweek was off by a factor of 1000. The SPR actually has 660 **million** barrels of oil, not 660 billion. This would only last us a few months.

- It turns out that mistakes like these are not uncommon. Here are a few prominent ones from the last two years:
  - The bailout will cost \$700 billion, not million. (New York Times, 9/25/08)
  - The U.S. budget deficit is projected to total \$1.2 trillion this year, not \$1.2 billion. (Wall Street Journal, 1/13/09)

- Total credit limits are \$3.4 trillion, not billion. (New York Times, 11/11/09)
- The rescue package for ailing European economies is 750 billion euros, not 750 million. (New York Times, 5/25/10)
- The yearly budget deficit stands at \$1.3 billion. (New York Times, 10/24/10)

Considering just the last example, we can calculate that if there are 300 million people in the U.S., each person could contribute just \$4 and we'd be out of debt. Unfortunately, however, the New York Times was wrong. The deficit is in the trillions, not the billions.

- If people struggle with millions and billions in the context of dollars, a unit they know well, how are they to comprehend millions and billions in the context of bytes, a unit they don't know so well? Take this example from the WSJ and the NYT related to the Nook:
  - The Nook e-book reader has 2 GB of memory, “enough to hold about 1,500 digital books.” (Wall Street Journal, 12/9/09)
  - A zettabyte (10<sup>21</sup> bytes) “is equivalent to 100 billion copies of all the books in the Library of Congress.” (New York Times, 12/10/09)

With these two pieces of information, we can calculate that the Library of Congress only has about 10,000 books. Hmm. . .

- According to Dear Abby, Americans receive 2 million tons of junk mail daily. That works out to be 13 pounds of mail per person per day. What went wrong? Probably Abby meant **yearly**, not daily.
- With examples like these and others, we can perform numeric triage. By converting these numbers to ones we can better understand, we can decide if the numbers make sense, if they agree with or run counter to our experience, and if they have reasonable or ridiculous implications.
- How do we prevent mistakes like these? It helps to know some basic facts and conversions, for example:
  - 6.8 billion people (China 1.3B, India 1.2B, USA 300M, NYC 9M)
  - 1 gallon of water weighs about 8 lb
  - 1 cubic foot of water weighs about 60 lb (rock 200, steel 500)
  - 1 liter is just over 1 quart
  - 1 kilogram is about 2.2 lb (28 gms in an ounce)
  - 1 ton is 2000 lb (a metric ton is 1000 kg or about 2200 lb)
  - 1 meter is a little over 3 feet or 1 yard
  - MP3 music is 1 MB/minute

- electricity costs 15-20 cents/KWH
  - 100K seconds in a day, 500K minutes (30M seconds) in a year
  - speed of light: 1 foot/nanosecond, speed of sound: 1000 feet/sec
  - 250 working days/year, 2000 hours/year...
- Here's an example interview question from Google: how many golf balls can fit in a school bus? Instead of looking like a deer in the headlights, you could approach this problem by assuming that a golf ball is actually cubic with 1-inch sides and a school bus is roughly 8 feet by 8 feet by 30 feet. Have at it!
  - One of Professor Kernighan's favorites is: how many petabytes will fit in this room?
  - According to Gambling Magazine, 10,000 Americans turn 50 years old every day. Is this number reasonable? To figure this out, we can use Little's Law, which states that the number in process is equal to the arrival rate multiplied by the service time. For example, in the case of Harvard, there are 6600 undergrads "in process," since the arrival rate is 1650 undergrads per year (new freshmen) and the service time is 4 years. If we consider the case of America, then there are 300 million people "in process," and the service time is roughly 75 years (the average lifespan). This means that  $(300/75)$  million people reach any given milestone in a year. This works out to about 10,000 a day, so the number provided by Gambling Magazine seems reasonable.
  - Here are some things to watch out for as you mount your numeric self-defense:
    - errors of dimensionality
    - oddly precise numbers
    - graphical trickery
    - flaky data
    - advocacy bias
  - According to the Newark Star-Ledger writing about bears, "young males can roam 60 to 100 square miles looking for food and mates, but females stay close to the cave, foraging within a 10-mile radius." Unfortunately, they committed an error of dimensionality, as a 10-mile radius actually corresponds to 314 square miles.
  - When editor of The Yacht Report declared that any yacht over 328 is so big that it loses its intimacy, he was using an oddly precise number. Why 328 and not 329?

- The Privacy Rights Clearinghouse reported in 2006 that the “rough figure” for the number of compromised pieces of personal information was 93,753,333. This number is problematic because the PRC has a definitive interest in making this number appear as high and as authoritative (read: “precise”) as possible.
- If you take nothing else away from this lecture, at least remember the name of a book: [How To Lie with Statistics](#) by Darrell Huff. It’s well worth the \$10. One of Huff’s favorite devices is the gee-whiz graph. With the right scale for the y-axis, you can make a graph of GPA versus first initial of student name appear to be meaningful. Based on the picture, you might be tempted to conclude that students with names starting with C or D are at a significant disadvantage. However, upon closer examination, we can see that the difference in their GPAs is infinitesimal: 3.35 compared to 3.36 for other initials. Another popular method of graphical trickery is the one-dimensional graph. The growth of Starbucks and money available to Princeton grad students can appear to be much greater when the data points are multi-dimensional. A three-dimensional cylinder or a two-dimensional tiger will grow with the cube and the square of their height change, but it’s only the height change that’s meaningful.
- Here are a few more examples of advocacy bias:
  - “Four thousand teens will try their first cigarette today.” (advertisement, New York Times, 11/18/05, endorsed by American Academy of Pediatrics, American Heart Assoc, American Lung Assoc, American Medical Assoc, National PTA.)
  - “Every day 5000 teenagers try pot for the first time.” (advertisement, New York Times, 11/4/05, Coalition for a Drug-Free America)
  - “Each year, according to the [American Anorexia and Bulimia] Association, 150,000 American women die of anorexia.” (Naomi Wolf, *The Beauty Myth*, 1990, p 148)
  - “The number of American children killed by guns has doubled every year since 1950.” (Nancy Day, *Violence in Schools: Learning in Fear*, 1996)

The numbers given for teens trying cigarettes and pot for the first time are at least in line with our estimate of 10,000 people reaching a given milestone every day, but they must be taken with a grain of salt given their sources, who have an interest in making the numbers appear shocking. Similarly, we can analyze the number given for anorexia deaths and determine that it’s somewhat exaggerated: about 2 million women die every year; does it seem reasonable that 10% of them will die of anorexia? The last example of American children killed by guns speaks for itself.

- In summary, here are a few defenses you can avail yourself of as you attempt to maintain reason:

- recognize the enemy
  - \* flaky numbers / excessive precision
  - \* arithmetic errors/ wrong units / wrong dimensions
  - \* bad statistics / deceptive presentation / suspect motives
- beware of the source
  - \* how do they know? how could they know?
  - \* why do they care? what are they trying to sell you?
- learn some useful numbers, facts, shortcuts
  - \* populations, rates, sizes, areas, conversions, etc.
  - \* approximate arithmetic, Little's Law, Rule of 72, powers of 2
- use your common sense and experience
  - \* much too big? much too small?
  - \* does it makes sense? what would it imply?