

Plan:

1. Discuss data intuition
2. Introduce Fermi Estimation
3. Introduce the concept of the wisdom of the crowds

# Data Intuition

Shannon E. Ellis, Ph.D  
UC San Diego



Department of Cognitive Science  
[sellis@ucsd.edu](mailto:sellis@ucsd.edu)



1011



1375



In today's pattern recognition class my professor talked about PCA, eigenvectors and eigenvalues.

I understood the mathematics of it. If I'm asked to find eigenvalues etc. I'll do it correctly like a machine. But I didn't **understand** it. I didn't get the purpose of it. I didn't get the feel of it.

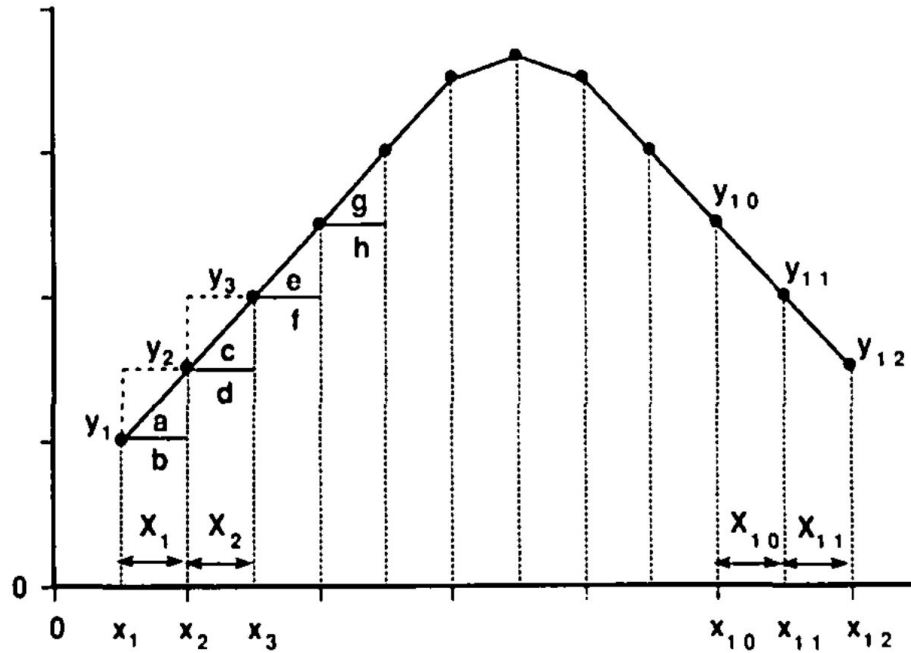
I strongly believe in the following quote:

You do not really understand something unless you can explain it to ~~your grandmother~~ **a layperson**. -- Albert Einstein

Well, I can't explain these concepts to a layman or grandma.

1. Why PCA, eigenvectors & eigenvalues? What was the *need* for these concepts?
2. How would you explain these to a layman?

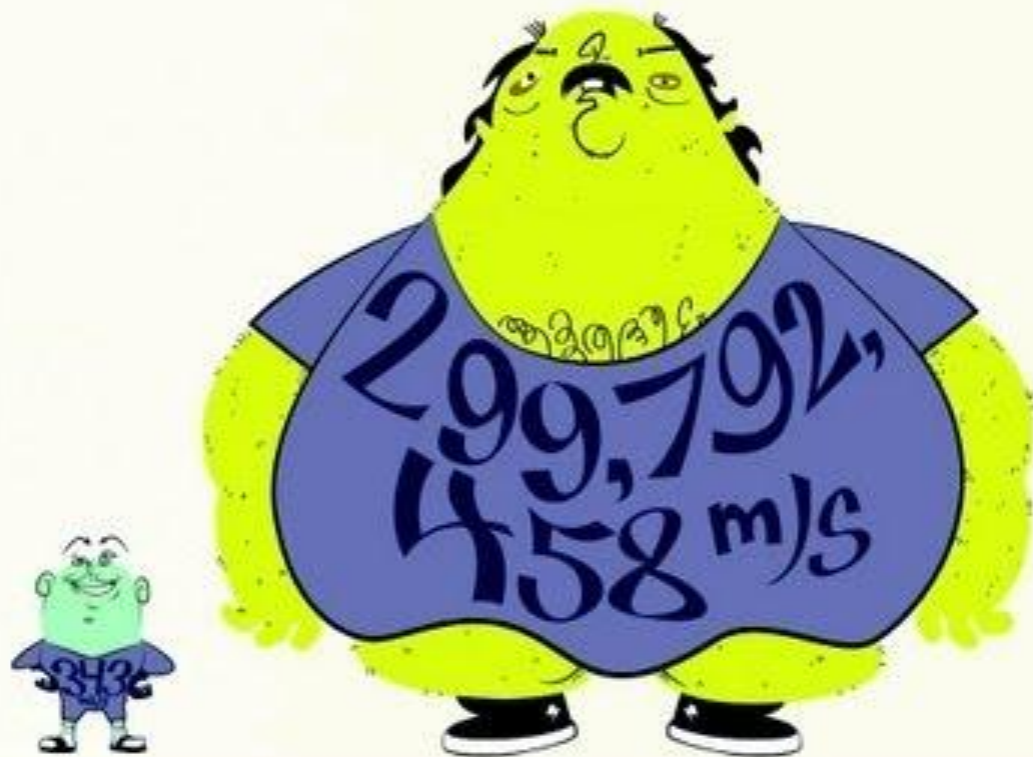
# Theory vs. Practice: “Tai’s model”



**Figure 1**—Total area under the curve is the sum of individual areas of triangles a, c, e, and g and rectangles b, d, f, and h.

# Fermi Estimation

Approximately how many piano tuners do you think there are in the city of Chicago?



<https://www.youtube.com/watch?v=0Yzvup0X8ls>

***Has humanity produced enough paint to  
cover the entire land area of the Earth?***  
—Josh (Bolton, MA)



# Fermi Estimation

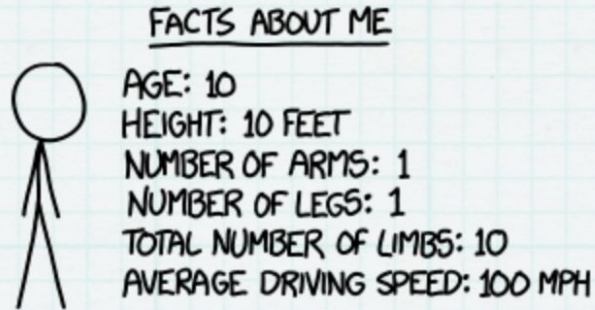
Has humanity produced enough paint to cover the entire land area of the Earth?

This answer is pretty straightforward. We can look up the size of the world's paint industry, extrapolate backward to figure out the total amount of paint produced. We'd also need to make some assumptions about how we're painting the ground. Note: When we get to the Sahara desert, I recommend not using a brush.





But first, let's think about different ways we might come up with a guess for what the answer will be. In this kind of thinking—often called **Fermi estimation**—all that matters is getting in the right ballpark; that is, the answer should have about the right number of digits. In Fermi estimation, you can round <sup>[1]</sup> all your answers to the nearest order of magnitude:



# Paint two room estimate

Let's suppose that, on average, everyone in the world is responsible for the existence of two rooms, and they're both painted. My living room has about 50 square meters of paintable area, and two of those would be 100 square meters. 7.15 billion people times 100 square meters per person is a little under a trillion square meters—an area smaller than Egypt.

NOT ENOUGH	EXACTLY ENOUGH	MORE THAN ENOUGH
/		

# Lifetime spent painting estimate

Let's make a wild guess that, on average, one person out of every thousand spends their working life painting things. If I assume it would take me three hours to paint the room I'm in, [2] and 100 billion people have ever lived, and each of them spent 30 years painting things for 8 hours a day, we come up with 150 trillion square meters ... just about exactly the land area of the Earth.

NOT ENOUGH	EXACTLY ENOUGH	MORE THAN ENOUGH
/	/	

# Light bulb comparison estimate

How much paint does it take to paint a house? I'm not enough of an adult to have any idea, so let's take another Fermi guess.

Based on my impressions from walking down the aisles, home improvement stores stock about as many light bulbs as cans of paint. A normal house might have about 20 light bulbs, so let's assume a house needs about 20 gallons of paint. [\[3\]](#) Sure, that sounds about right.



# Light bulb comparison estimate

The average US home costs about \$200,000. Assuming each gallon of paint covers about 300 square feet, that's a square meter of paint per \$300 of real estate. I vaguely remember that the world's real estate has a combined value of something like \$100 trillion, <sup>[4]</sup> which suggests there's about 300 billion square meters of paint on the world's real estate. That's about one New Mexico.

NOT ENOUGH	EXACTLY ENOUGH	MORE THAN ENOUGH
//		

Of course, both of the building-related guesses could be overestimates (lots of buildings are not painted) or underestimates (lots of things that are not buildings [\[5\]](#) are painted) But from these wild Fermi estimates, my guess would be that there probably isn't enough paint to cover all the land.

So, how did Fermi do?

According to the report [\*\*The State of the Global Coatings Industry\*\*](#), the world produced 34 billion liters of paints and coatings in 2012.

There's a neat trick that can help us here. If some quantity—say, the world economy—has been growing for a while at an annual rate of  $n$ —say, 3% (0.03)—then the most recent year's share of the whole total so far is  $1 - \frac{1}{1+n}$ , and the whole total so far is the most recent year's amount times  $1 + \frac{1}{n}$ .

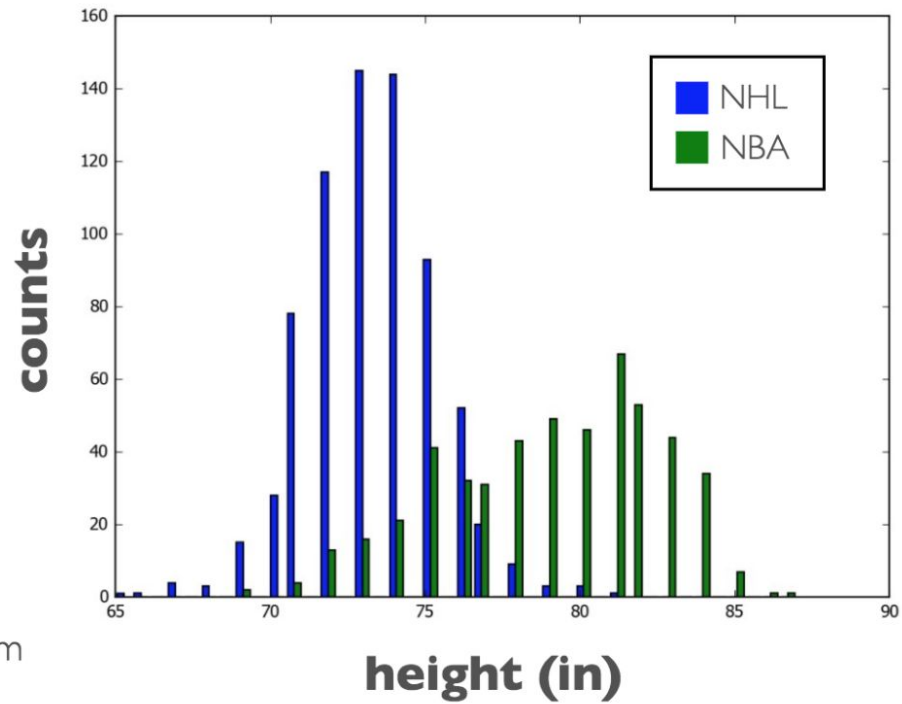
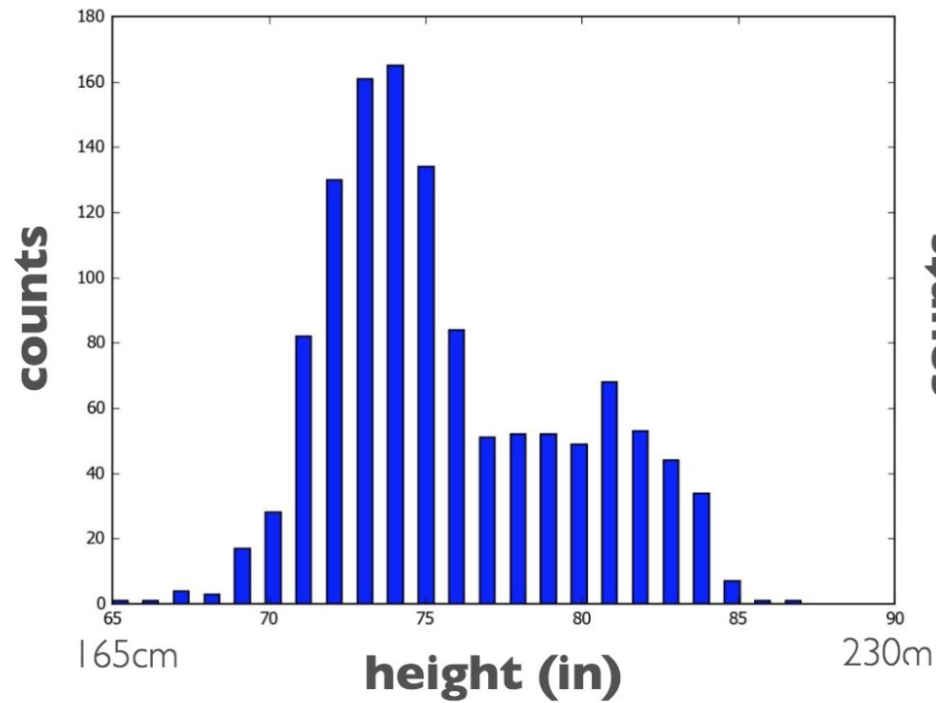
If we assume paint production has, in recent decades, followed the economy and grown at about 3% per year, that means the total amount of paint produced equals the current yearly production times 34.<sup>[6]</sup> That comes out to a little over a trillion liters of paint. At 30 square meters per gallon,<sup>[7]</sup> that's enough to cover 9 trillion square meters—about the area of the United States.

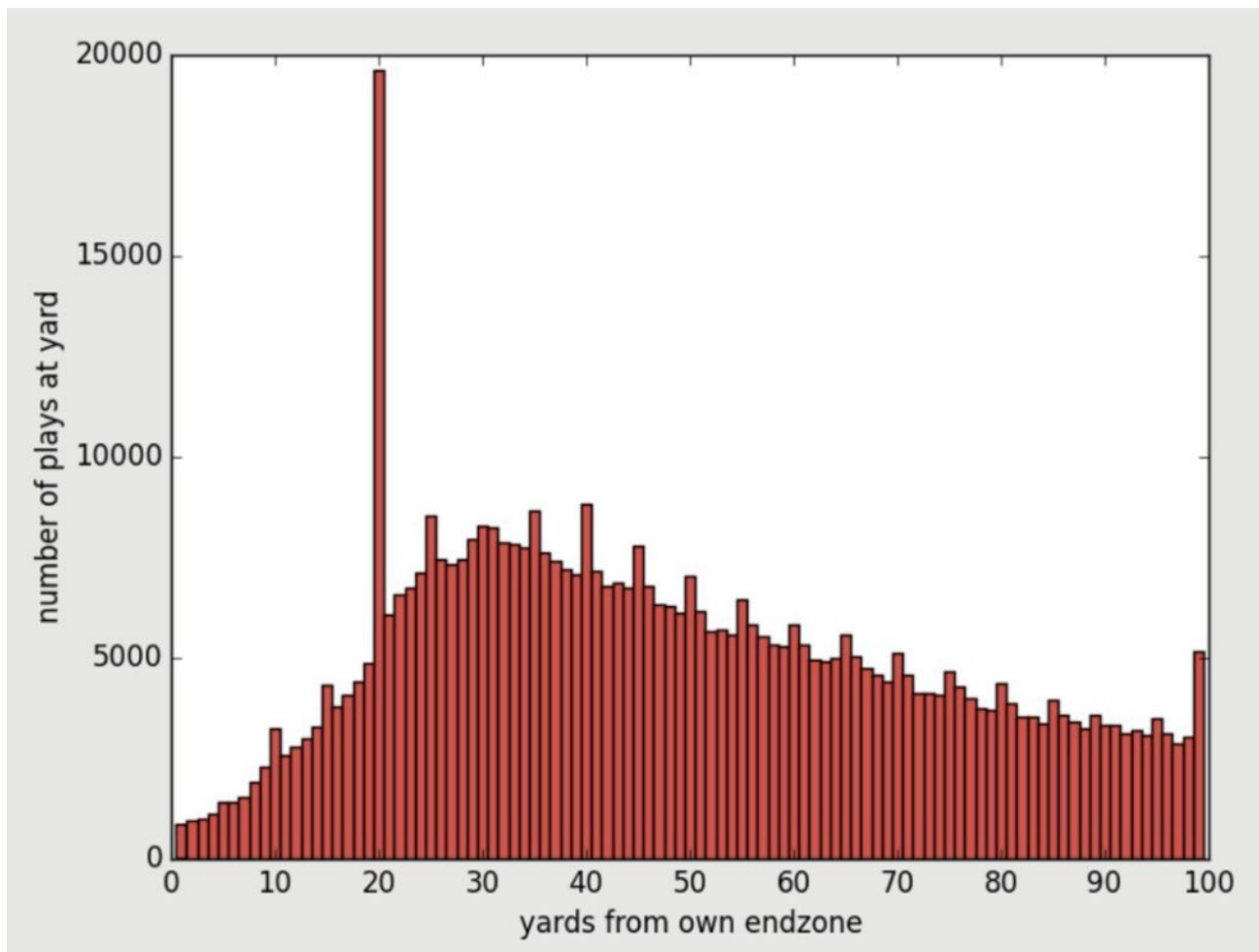
So the answer is no; there's not enough paint to cover the Earth's land, and—at this rate—probably won't be enough until the year 2100.



**What does this mean  
in COGS 108?  
(and in your future)**

---



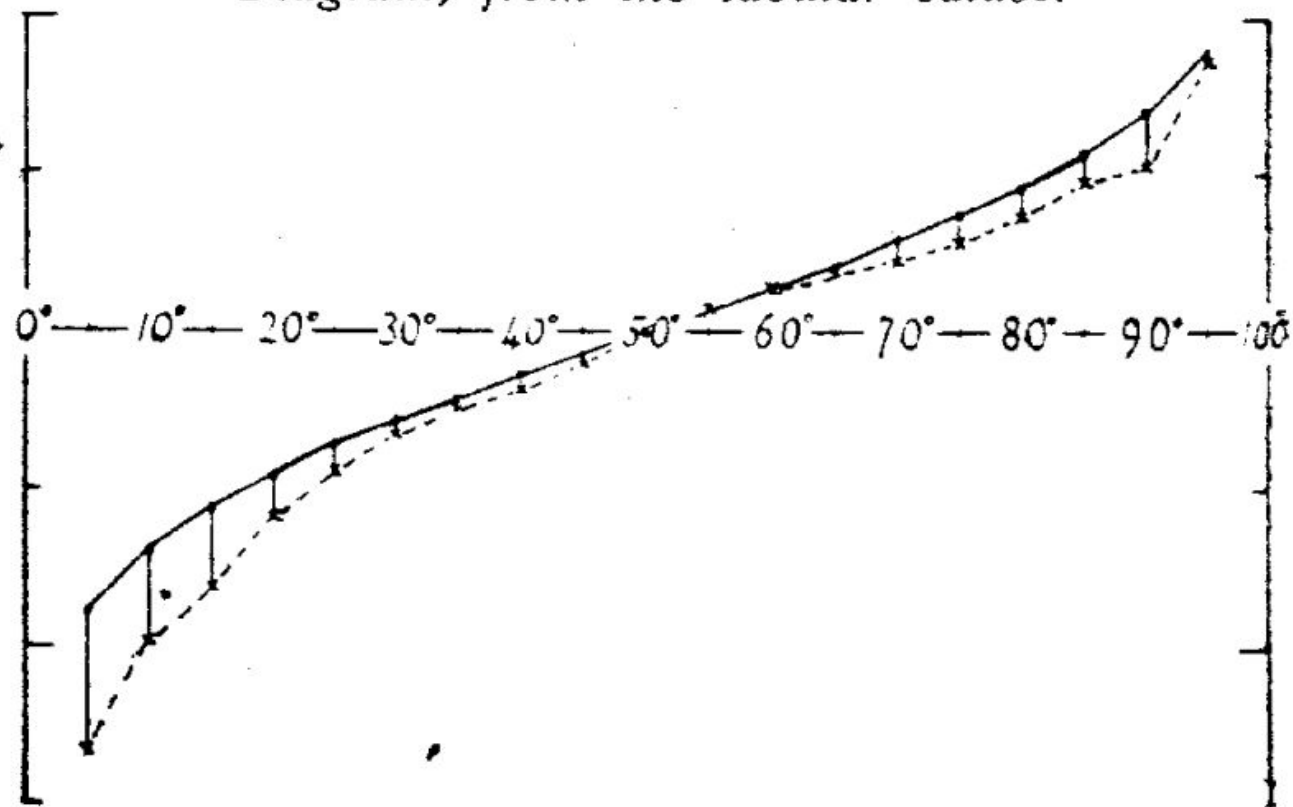


# Data Intuition

1. Think about your question and your expectations
2. Do some Fermi calculations (back of the envelope calculations)
3. Write code & look at outputs <- think about those outputs
4. Use your gut instinct / background knowledge to guide you
5. Review code & fix bugs

Diagram, from the tabular values.

*Vox Populi*



# The Wisdom of the Crowds

- Diversity of opinion: Each person should have private information....even if it's just an eccentric interpretation of the known facts
- Independence: People's opinions aren't determined by the opinions of those around them
- Decentralization: People are able to specialize and draw on local knowledge
- Aggregation: Some mechanism exists for turning private judgements into a collective decision

On your own (meaning w/o Googling), please fill out quickly:

[http://bit.ly/fermi\\_fa20](http://bit.ly/fermi_fa20)

---