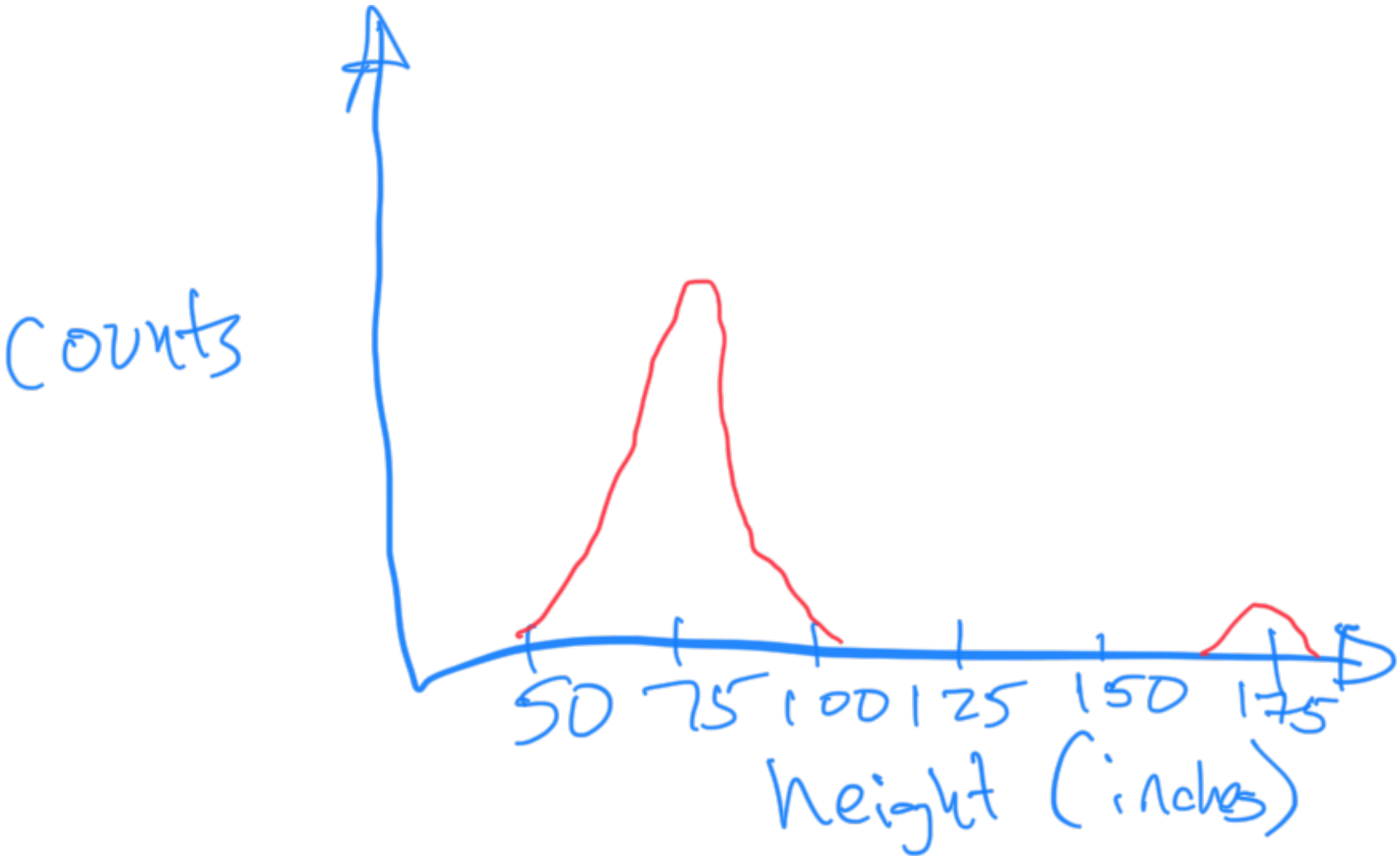
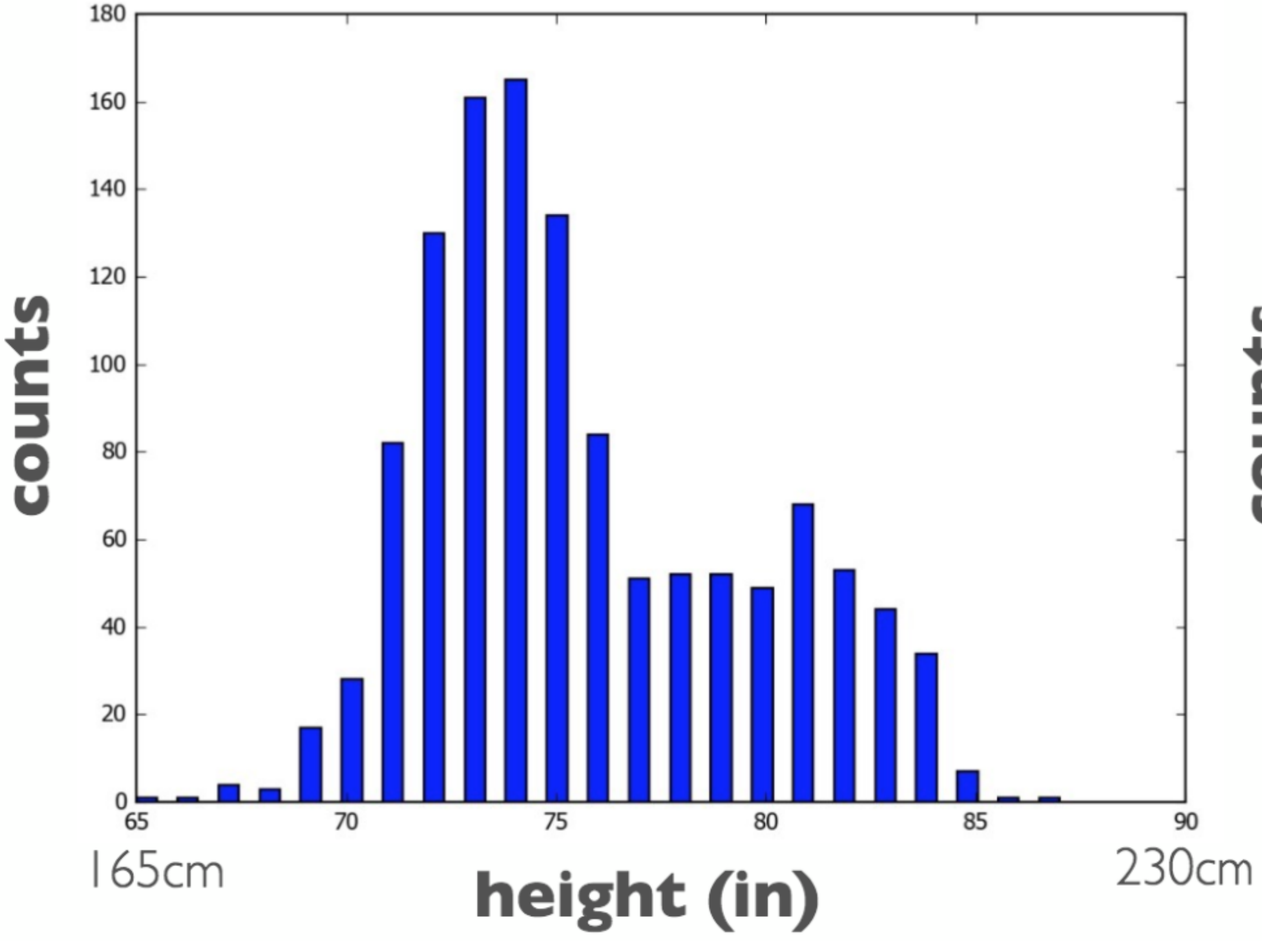
Data Intuition

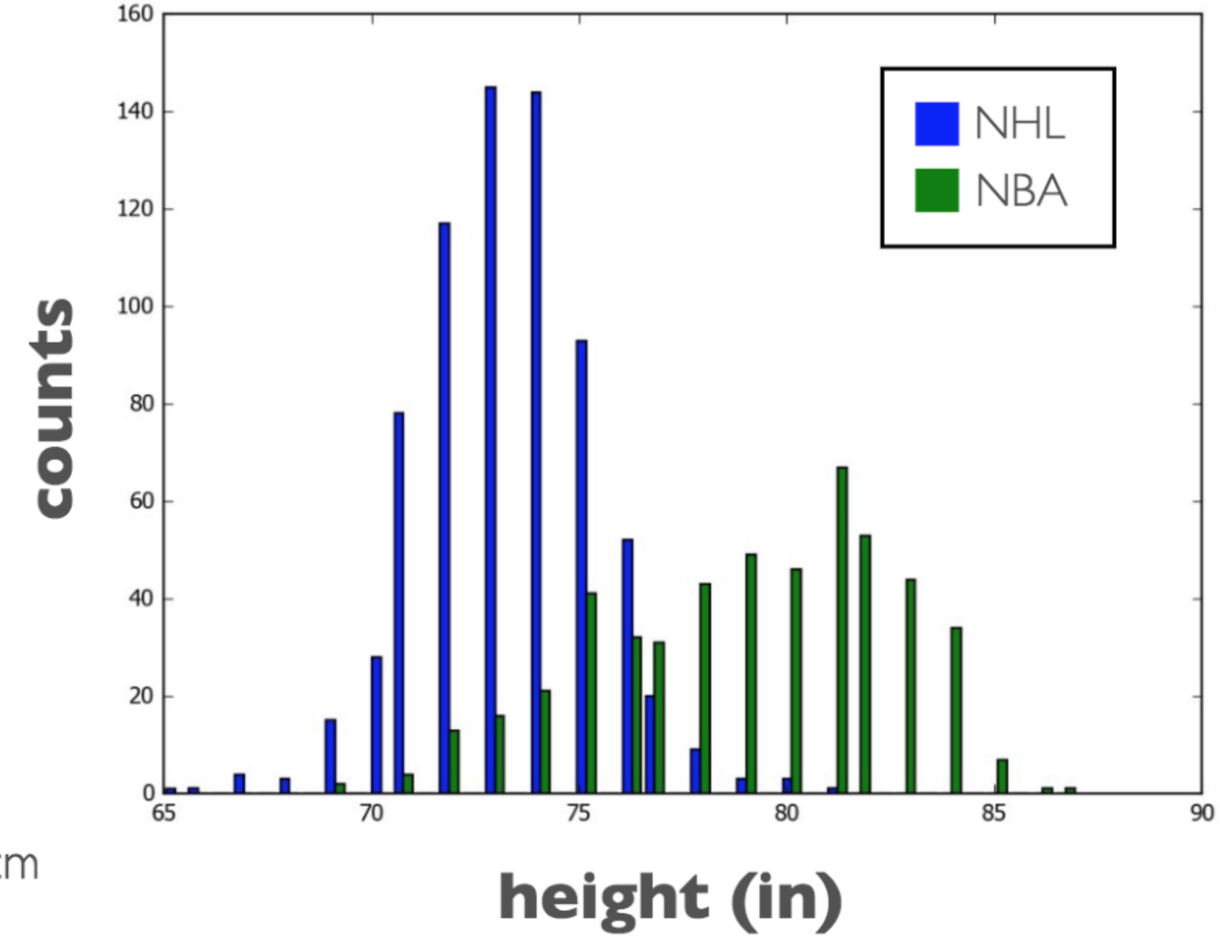
Data intuition

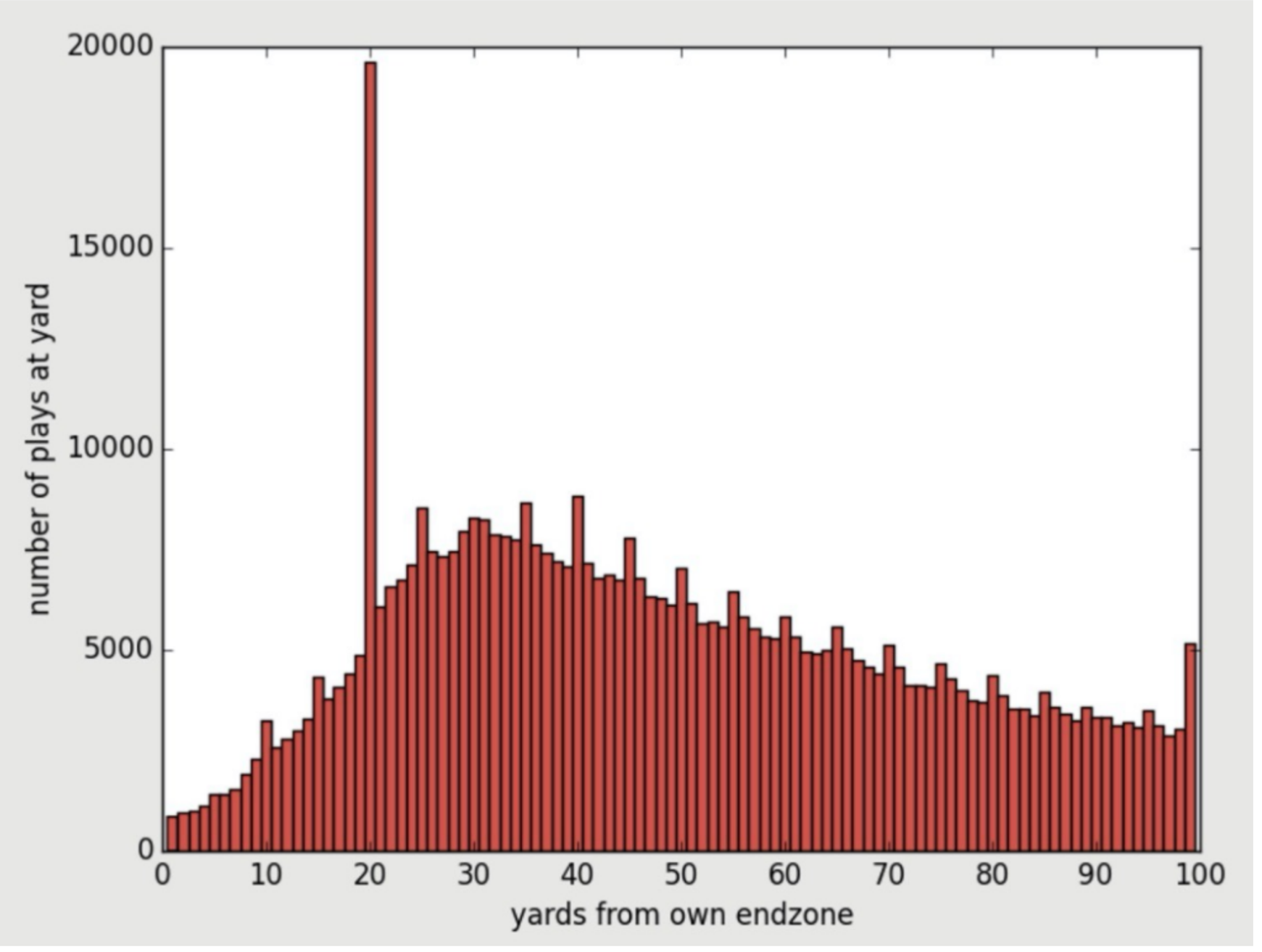
- Knowledge and judgement: a generic toolbox vs domain specific
  - Always looking for outliers in a column and examining them vs. knowing that 5774 mg/dl blood glucose is impossible for a living human
- Ability to ask the right questions of the data
  - After calculating a mean / variance, asking if the data should be broken up into different subgroups (stratified) by some category we have in the data
- Ability to pick a reasonable hypothesis
  - Did the subgroup variable cause the measured differences or is there perhaps something else going on here

Why do you need intuition?




counts
height (in)
165cm
230cm


counts
height (in)
NHL
NBA


20000
15000
10000
5000
0
number of plays at yard
0
10
20
30
40
50
60
70
80
90
100
yards from own endzone

Chicago piano tuners

https://forms.gle/C982naWtU9RvHqAb7

https://www.youtube.com/watch?v=0YzvupOX8Is

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

—Josh (Bolton, MA)

https://what-if.xkcd.com/84/

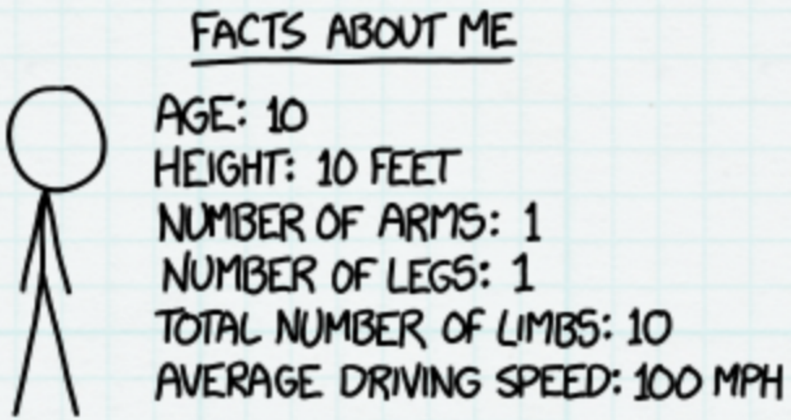
Paint the town red

https://forms.gle/1cjJd91Mo37ATK8v9

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.

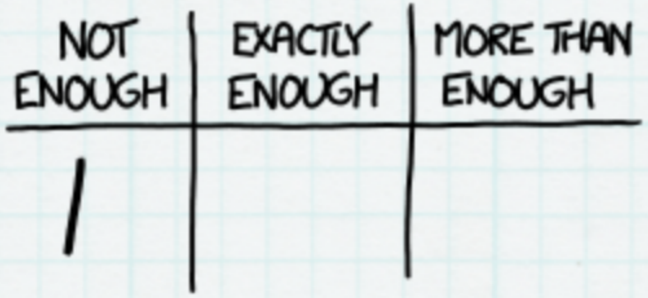
https://what-if.xkcd.com/84/

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 [1] all your answers to the nearest order of magnitude:



https://what-if.xkcd.com/84/

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.



https://what-if.xkcd.com/84/

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.

https://what-if.xkcd.com/84/

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.

https://what-if.xkcd.com/84/

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,[4] which suggests there's about 300 billion square meters of paint on the world's real estate. That's about one New Mexico.

https://what-if.xkcd.com/84/

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?

https://what-if.xkcd.com/84/

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 $1 + \frac{1}{n}$.

https://what-if.xkcd.com/84/

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.[6] That comes out to a little over a trillion liters of paint. At 30 square meters per gallon,[7] 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.

https://what-if.xkcd.com/84/

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

https://forms.gle/CREcpMkYDLYTUp2s6

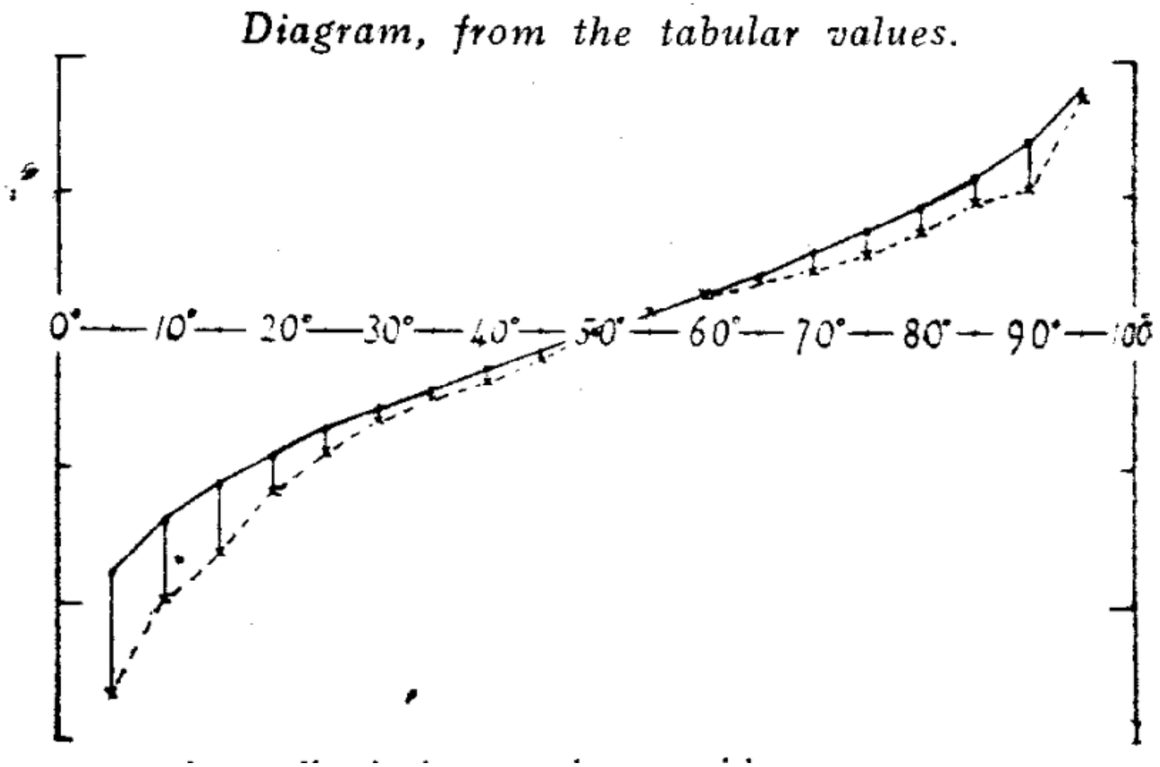
4 min think
3 min pair
3 min share

Other kinds of guessing and intuitions

Vox Populi

The wisdom of the crowd

At a 1906 country fair in Plymouth, 800 people participated in a contest to estimate the weight of a slaughtered and dressed ox. Statistician Francis Galton observed that the median guess, 1207 pounds, was accurate within 1% of the true weight of 1198 pounds.



The Wisdom of the Crowds

Criteria for validity

- 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

Research

- Lookup the topic in the scientific literature…
  - If nobody has asked your exact question, read about the nearest thing to it
- No existing useful knowledge? Then lay what-if with one of the “reasoning” LLMs or with your teammates
  - Don’t just ask an LLM for the answer, ask it to generate a few different possibilities and to explain why those might be true. Converse, criticize.