Sampling Distributions

STAT 240 - Fall 2025

Robert Sholl You remind me of 2010s Adam Levine

github.io/stat240_f25

Motivation

You’re a head brewing chemist at a large scale brewery. You’ve been tasked to develop a new product that minimizes costs, maximizes quality, and appeals to a broad audience.

Brewing

- Low cost, high quality, and unoffensive to the masses?
  - That's a German beer
- Barley, hops, and water
- Barley and hop flavor notes are a lie, focus on yield
- Water is the only thing that matters for quality

Cost Reduction

To develop this beer we have to run an experiment.

- Samples can get *very* expensive

In as few samples as possible we need to determine:

- The highest yield crop varieties
- The location with the highest quality water source

In-class Activity

Take 5 minutes and pair up into groups.

- Think about what skills your group members will need
- Each group *at least* needs someone ready to speak up at the end

Challenge 1

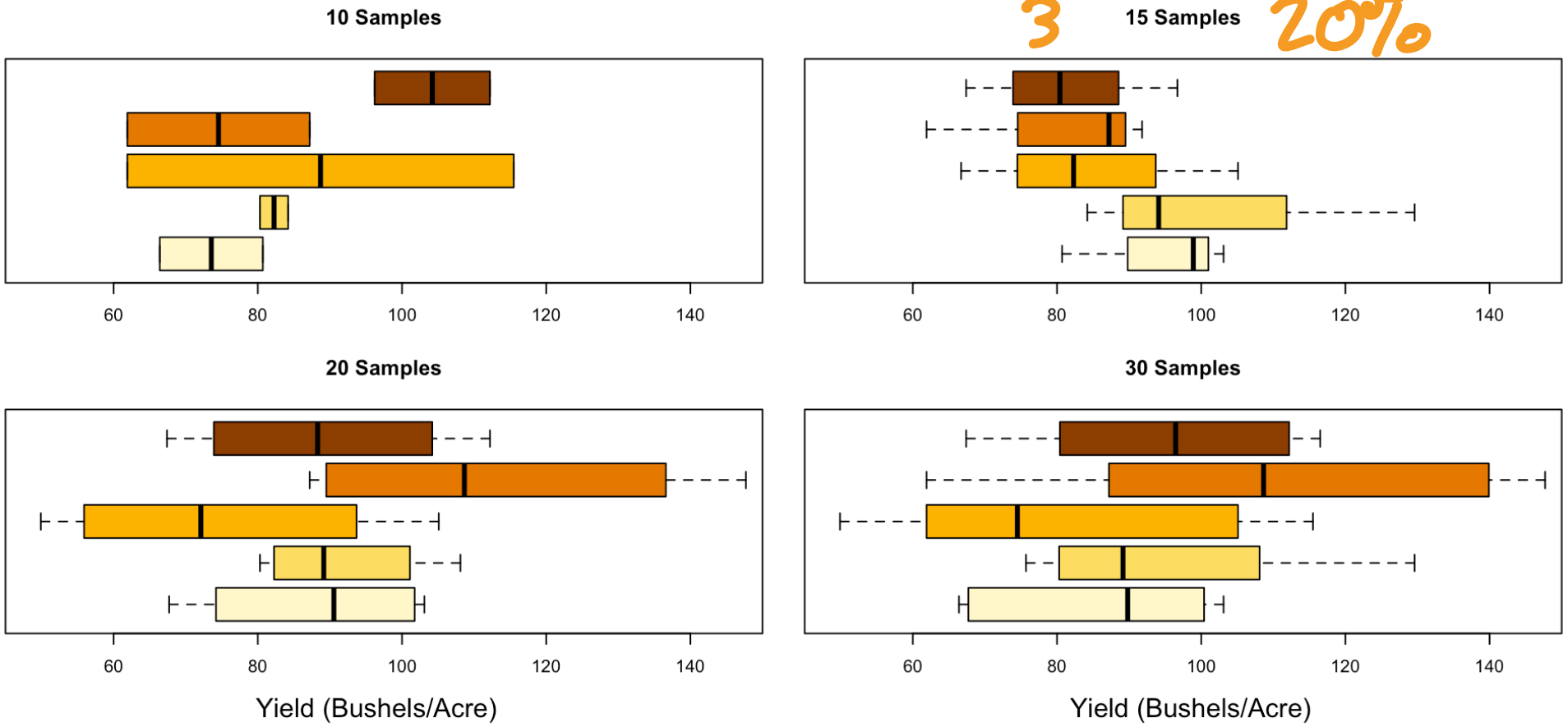
- You have $85000 of company funds at your disposal
- Every sample has a cost to it:
  - $500 per sample for Barley experiments
  - $800 per sample for Hop experiments
  - $2000 per sample for water testing
- The head brewer will only use the recipe you suggest if the data **AND** your explanation justify its use

Challenge 1

- For each “correct” ingredient, you’ll see a 150% return on investment
  - “Second best” ingredients see 80% return
  - “Third best” see 20%
  - Anything else results in a 30% loss
- You’ll have 2 minutes with each data source to make your decisions
- You’ll have 5 minutes to develop your “defense” of your recipe

Barley

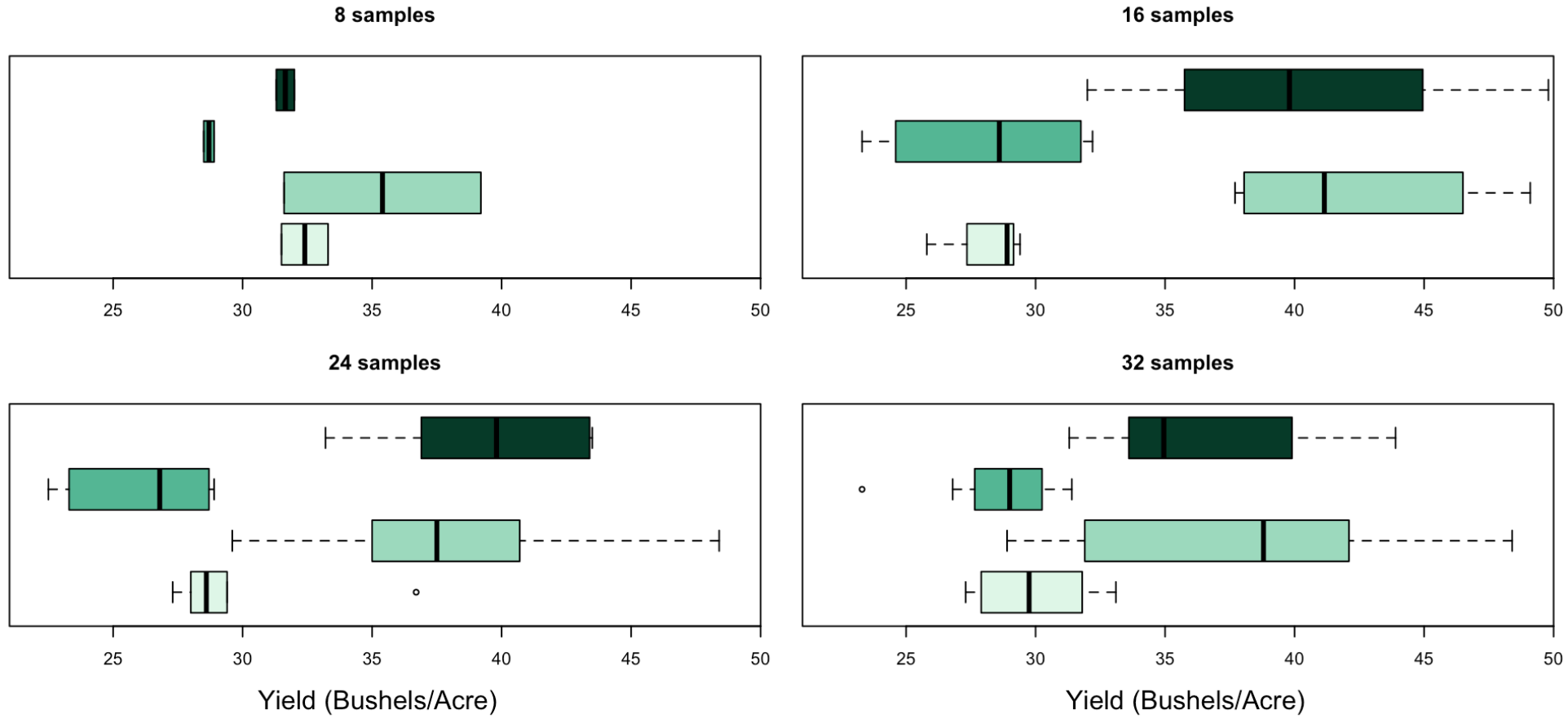
1 150% -130%
2 80%
3 20%



2 1 3

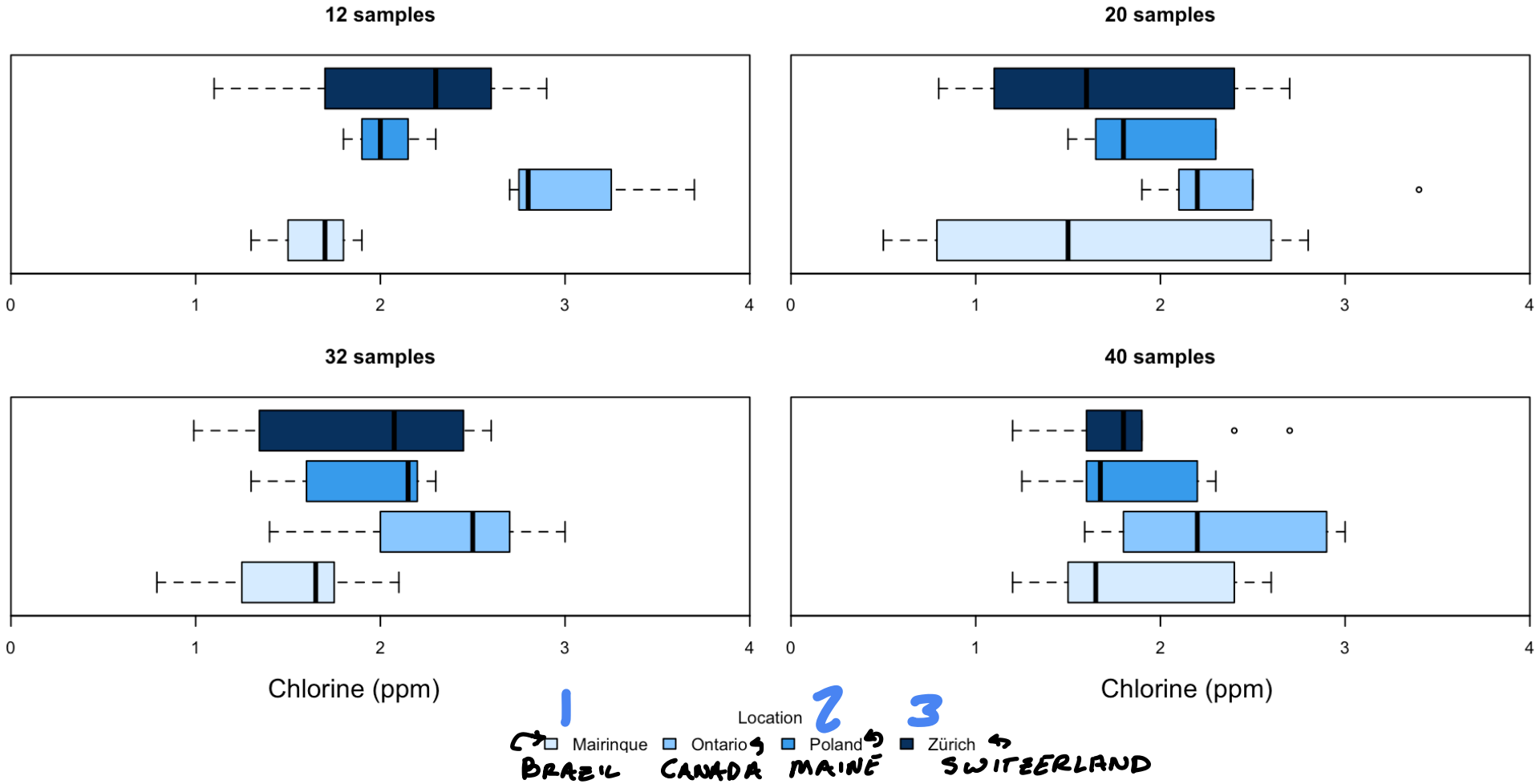
github.io/stat240_f25

Hops



github.io/stat240_f25

Water


12 samples
20 samples
32 samples
40 samples
0
1
2
3
4
Chlorine (ppm)
Chlorine (ppm)
Location
Mairinque
Ontario
Poland
Zürich
1
2
3
BRAZIL
CANADA
MAINE
SWITZERLAND

1. 51K
2. 52K
3. 64K
4. 71K
5. 96K
6. 81K
7. 31K
8. 41K

Results

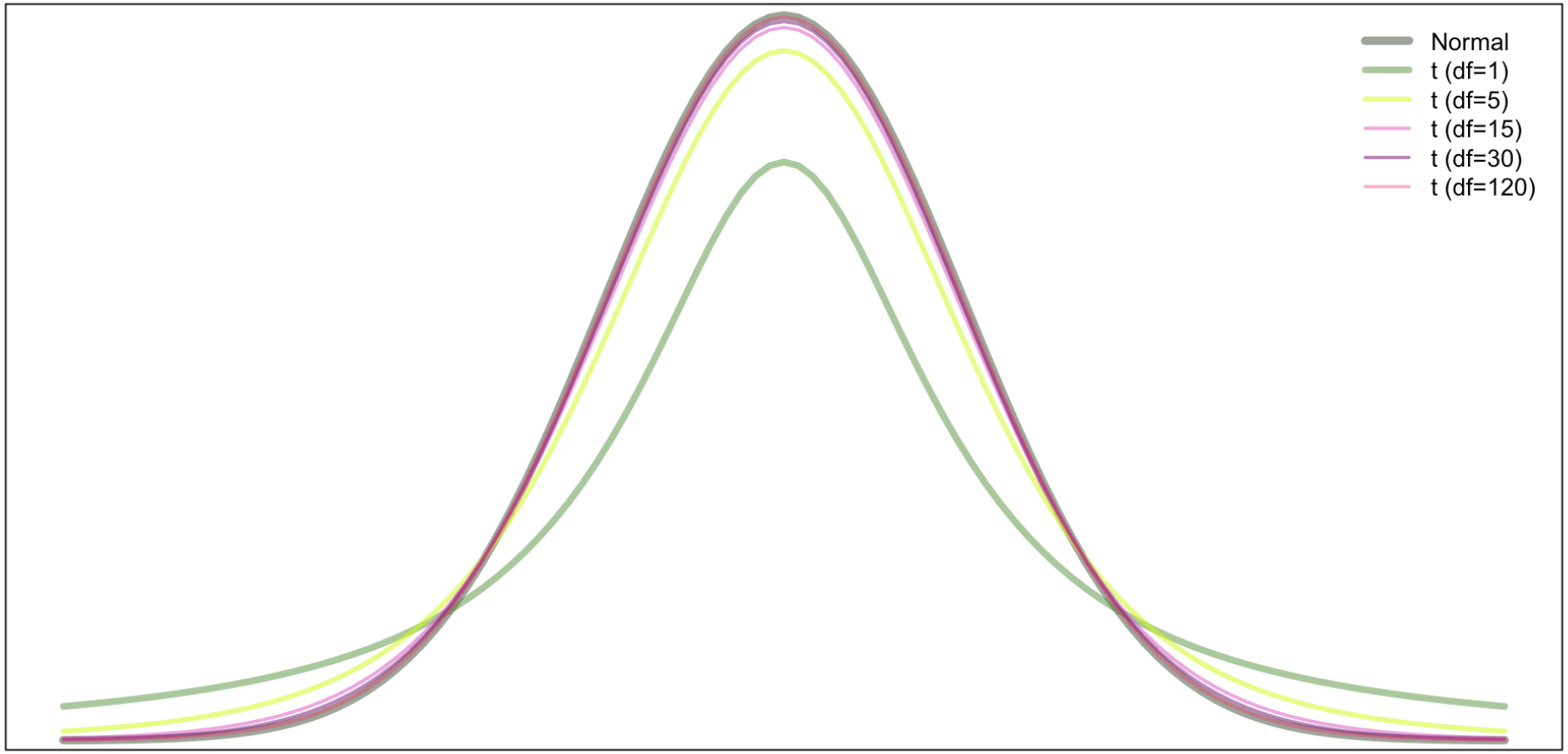
Barley: Trebi

Hop: Nugget

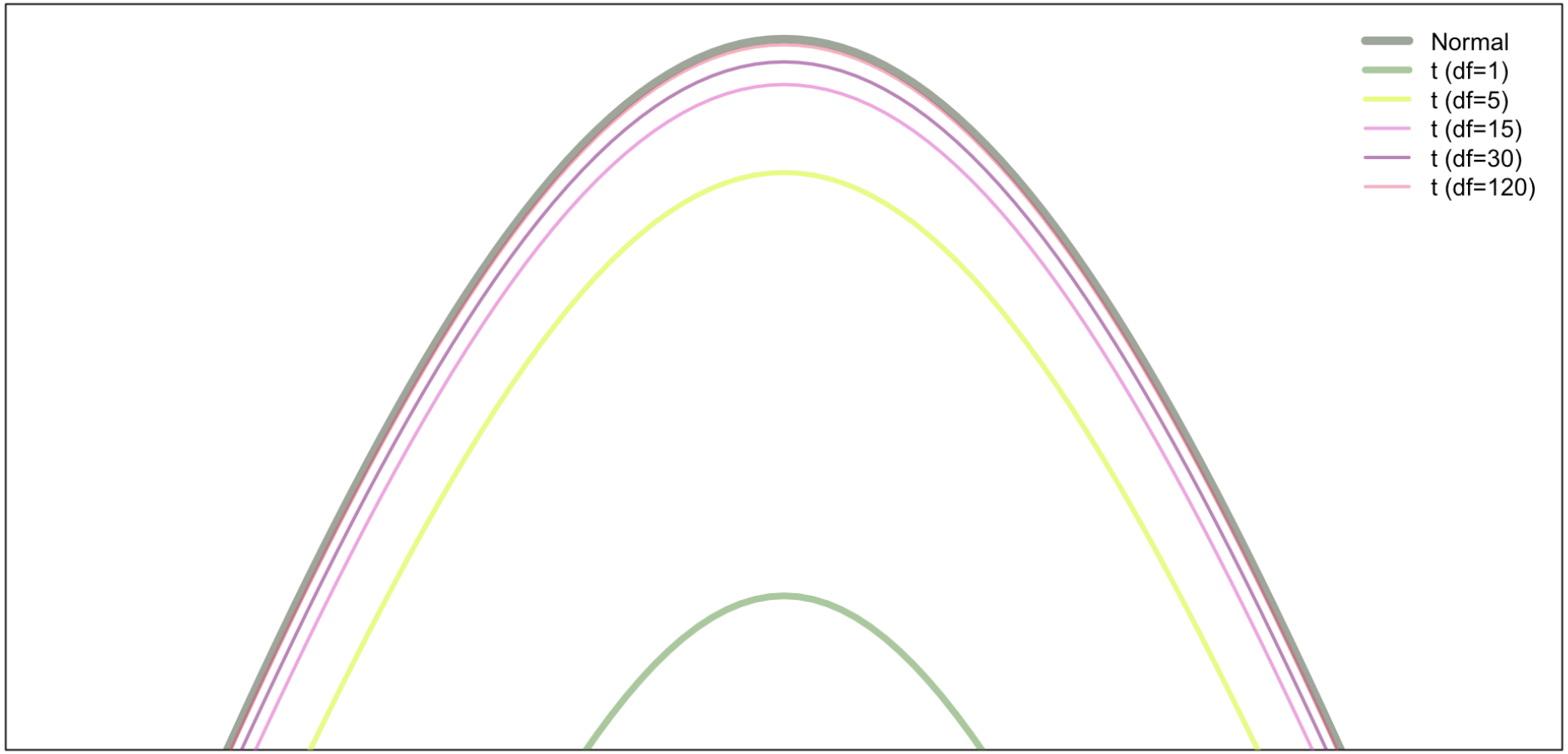
Water: Marinique

Sampling Distributions

Student's t-distribution



Student's t-distribution



Central Limit Theorem

- I'll attempt this explanation
- I am not as good as this guy
  - I highly suggest you watch this, but it's 31 minutes so that's a *you* problem

Central Limit Theorem

- Pull cards from a deck of 52 playing cards
  - Pull cards until you hit the queen of hearts
  - Record how many cards you pulled to get there, shuffle the deck, and repeat
- If you shuffle perfectly:
  - All possible positions are equally likely
  - Thus the chance of hitting a “center” queen is equally fair

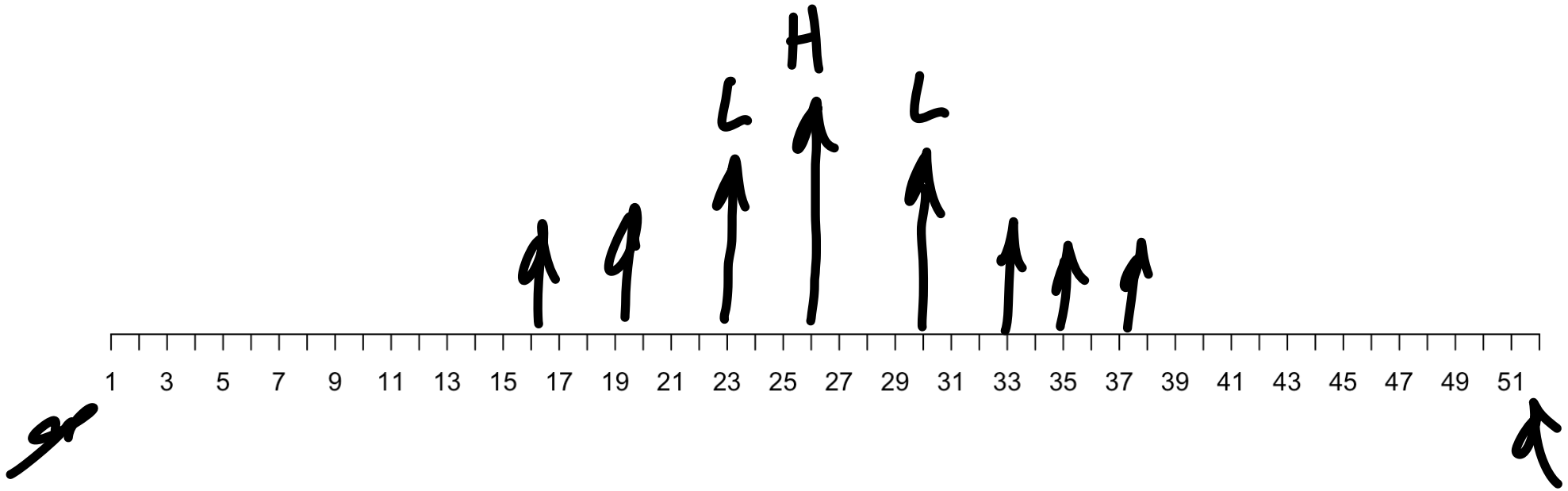
Central Limit Theorem

- The proper representation of reality would be:

> How often do people pull the queen in the center of the deck, on average?

- Zoom out and look at average results, i.e.
  - Each person pulls cards until the queen of hearts 20 times
  - That number is averaged and plotted out

Central Limit Theorem



Why does this matter?

Who actually won our little competition?

- The answer feels obvious, but it isn't
- The CLT holds as $n \longrightarrow \infty$
- At smaller samples we need the t-distribution
  - That's not very convenient though

Distribution of Statistics

In practice, the convergence can happen much quicker

- For any sample mean $\bar{x}$:

$$\text{When } n > 30, \quad \bar{x} \sim N(\mu_{\bar{x}}, \sigma^2_{\bar{x}})$$

Where:

$$\mu_{\bar{x}} = \mu \quad \text{and} \quad \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$$

$$\sqrt{\sigma^2_{\bar{x}}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

Challenge 2

The “correct” ingredient can be described as the one that has the highest probability of the best outcome

- You’ll be given the mean and variance of each ingredient
- With the sample sizes you chose:
  - Find the probability that you achieve the “best” ingredients Q_3 (noted with *)
- The team with the highest probabilities for each wins

Barley

$\mu_{\bar{x}} = \mu = 110$ $\quad \frac{\sigma^2}{n} \neq \frac{\sigma}{\sqrt{n}}$

Ingredient	Mean	Variance	Q_3
Manchuria	103	674	115
Peatland	110	455	120
Svantosa	102	677	138
Trebi*	127	1345	138
Velvet	103	1066	123

$$\frac{138 - 110}{\sqrt{455/n}}$$

Hops

Ingredient	Mean	Variance	Q_3
Cascade	31.7	42.7	33.0
Millennium	37.8	31.4	42.4
Mt. Hood	28.4	10.4	31.2
Nugget*	37.5	29.2	41.1

Water

LOCATION

~~Ingredient~~	Mean	Variance	Q_3
Zürich	2.00	0.469	1.5
Ontario	2.32	0.416	2
Poland	1.93	0.178	1.6
Marinique*	1.70	0.317	1.3

Results

Go away