

Plan:

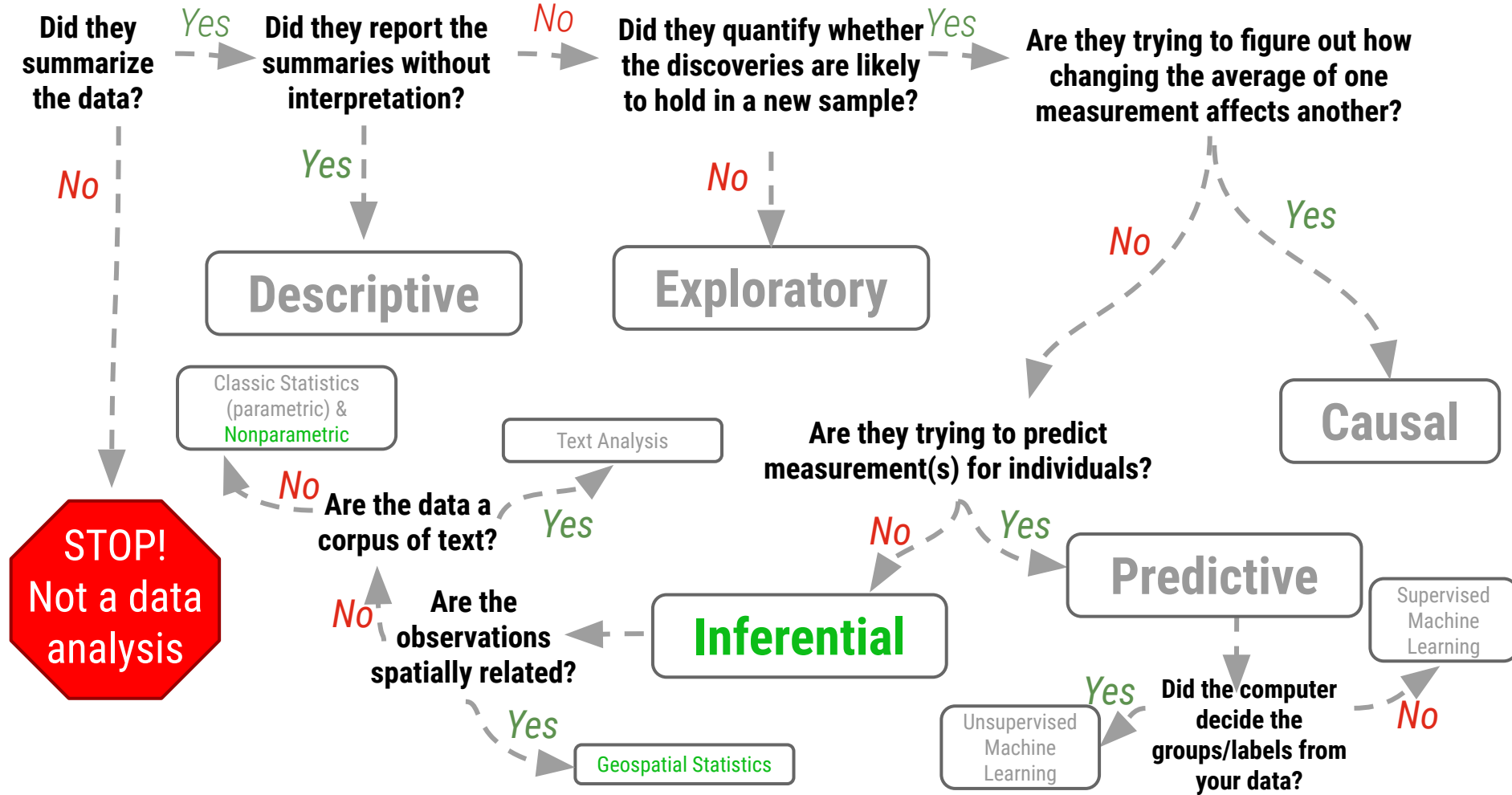
1. Describe why nonparametric statistics are needed & when they are helpful
2. Discuss bootstrapping

Nonparametric Statistics I

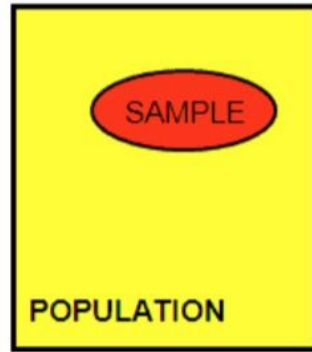
Shannon E. Ellis, Ph.D
UC San Diego



Department of Cognitive Science
sellis@ucsd.edu



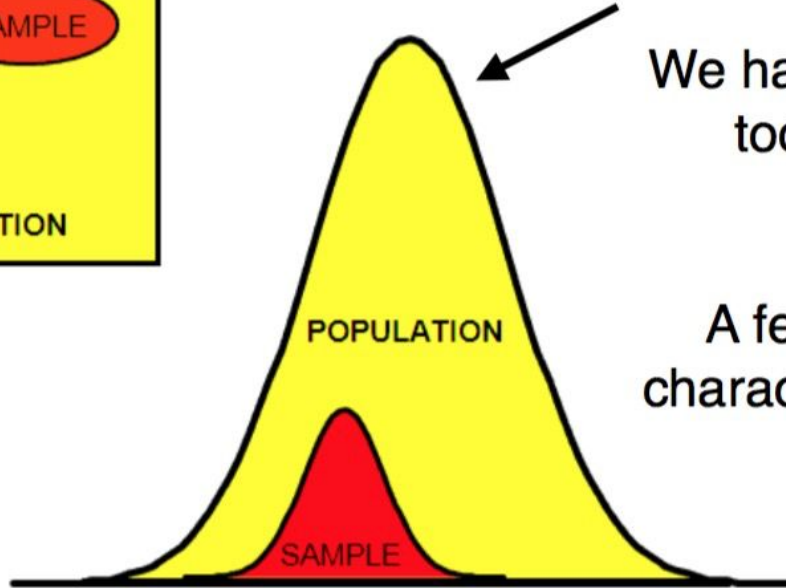
Non-parametric Statistics: The Why



Normal distribution
(nice and friendly)

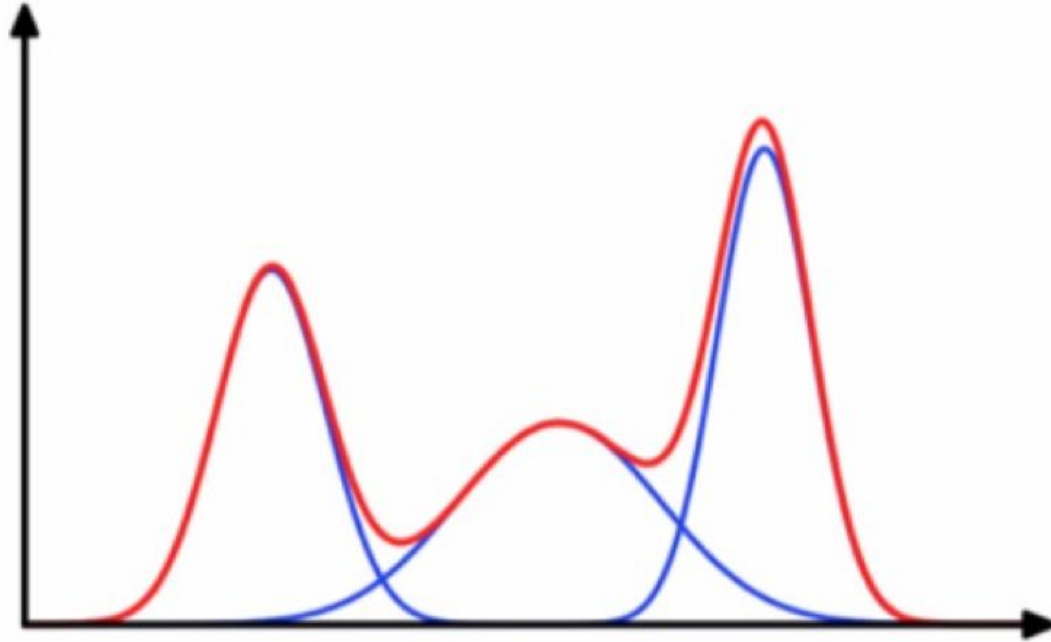
We have good math
tools for this.

A few parameters **fully**
characterize the distribution.

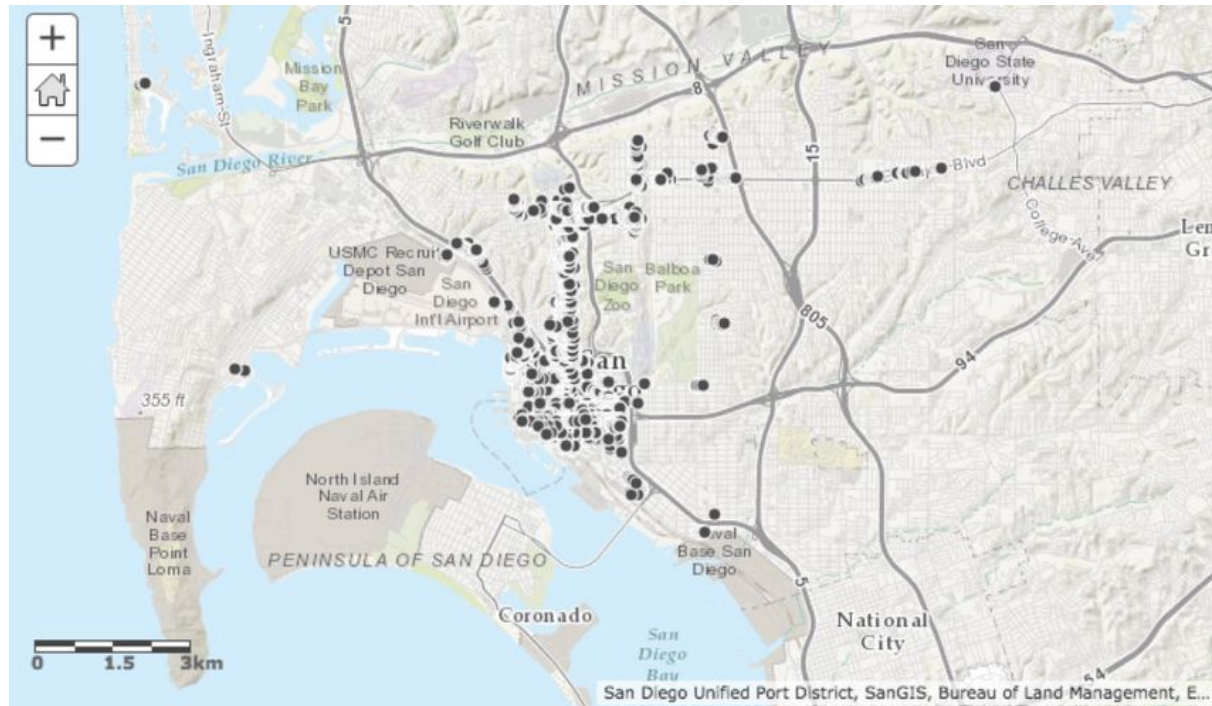


Non-parametric Statistics:

What if your distribution looks like this?



Non-parametric Statistics: ...or like this?

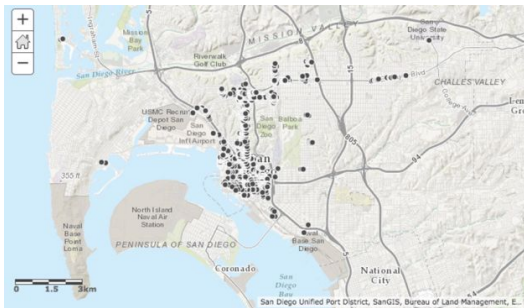


Parameters (like mean and variance) cannot fully and accurately capture this distribution!

Hence, we require **non-parametric statistics**.

When to turn to non-parametric statistics...

- When underlying distributions are non-normal, skewed, or cannot be parameterized simply.



Artwork by @allison_horst

- When you have ranked (ordinal) data, *e.g.*, preferences.

Like	Like Somewhat	Neutral	Dislike Somewhat	Dislike
1	2	3	4	5

- When you need to build an empirical “null” distribution.

NOMINAL

UNORDERED DESCRIPTIONS



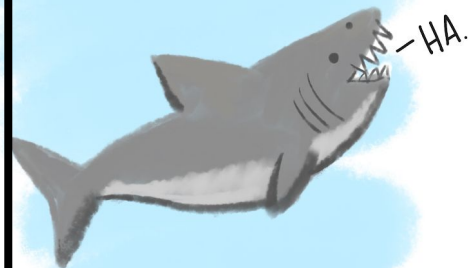
ORDINAL

ORDERED DESCRIPTIONS



BINARY

ONLY 2 MUTUALLY
EXCLUSIVE OUTCOMES



@allison-horst

Non-parametric Statistics: distribution-free

- **Myth:** Non-parametric statistics does not use parameters.
- **Fact:** Non-parametric statistics does not make *assumptions about* / parametrize the underlying distribution generating the data.

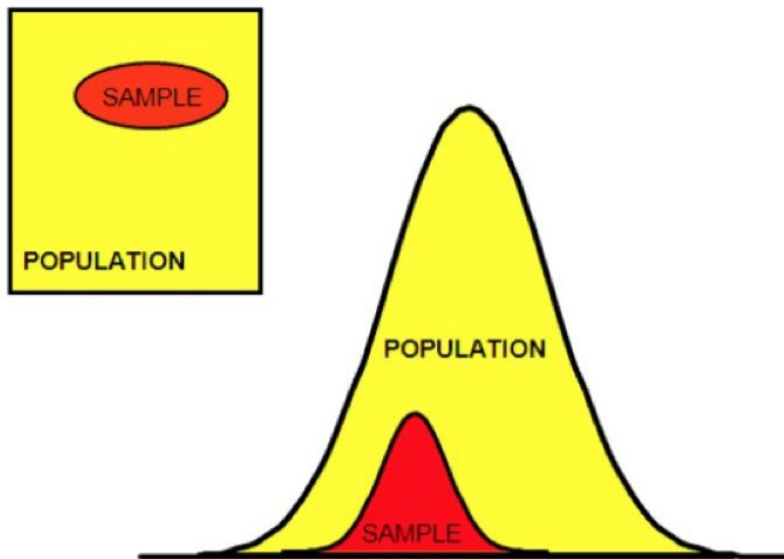
- **“Distribution-Free” statistics**
 - Meaning, it does not assume data-generating process (like heights) result in, *e.g.*, normally-distributed data

Resampling statistics: The What

- Bootstrap (Monte Carlo)
- Rank Statistics (Mann Whitney U)
- Kolmogorov-Smirnoff Test
- Non-parametric prediction models

1) Bootstrapping (resampling)

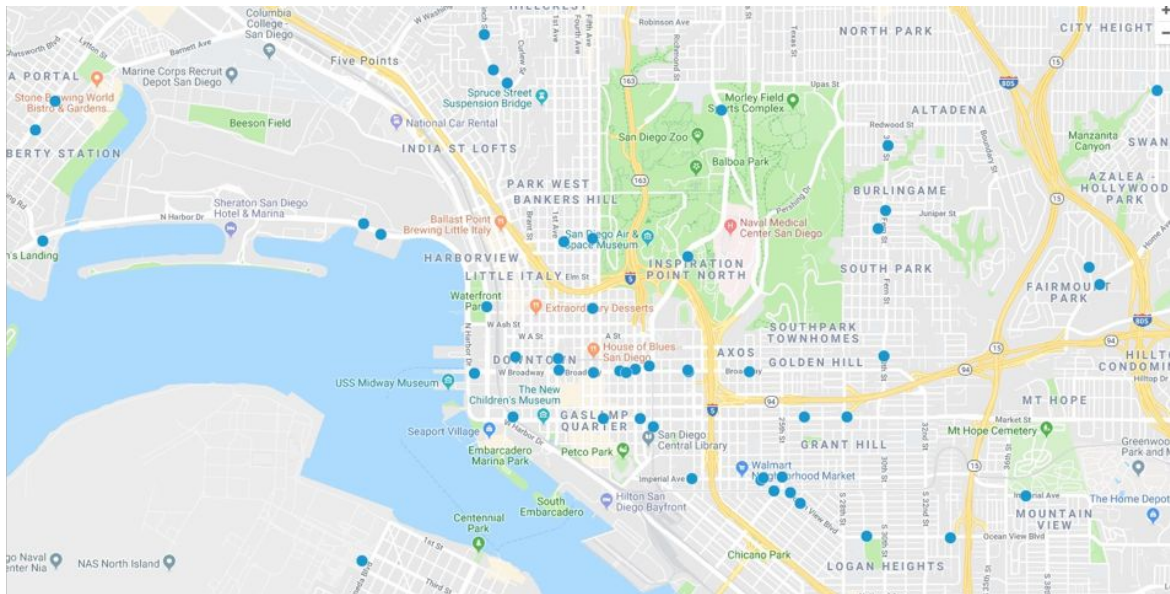
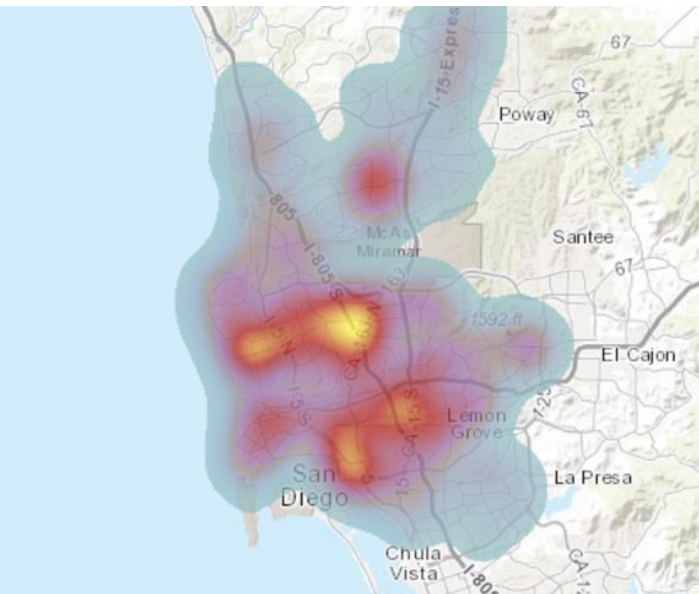
- How can we build a more realistic “null distribution” for the sample estimate without knowing the population it’s drawn from?



Bootstrapping (resampling)

Example Question:

- Are San Diego's pot holes closer to bus stops than not?



Bootstrapping (resampling)

