study design vs. inference

A SOURCE OF CONFUSION

Mike Baiocchi @Stanford University

why do statistics?

- Make a compelling argument about the world.
- Connect what's going on in the world (data) with a deductive framework (inference).
- If we control the process of data collection then we can have a better sense of how our deductive framework should behave.
- This leads to clearer reasoning about what the underlying structure of the world might be.

- This lecture: What if we don't control the way the data were actually generated? What if we didn't intervene?

study design vs. inference

- 90% of statistics classes are about inference
- Why?
  - It’s useful, getting you those confidence intervals and p-values.
  - The Math is pretty cool.
  - It feels hard.
  - Because many of us don’t really know much about the real world…

design

RANDOMIZATION AND SAMPLING

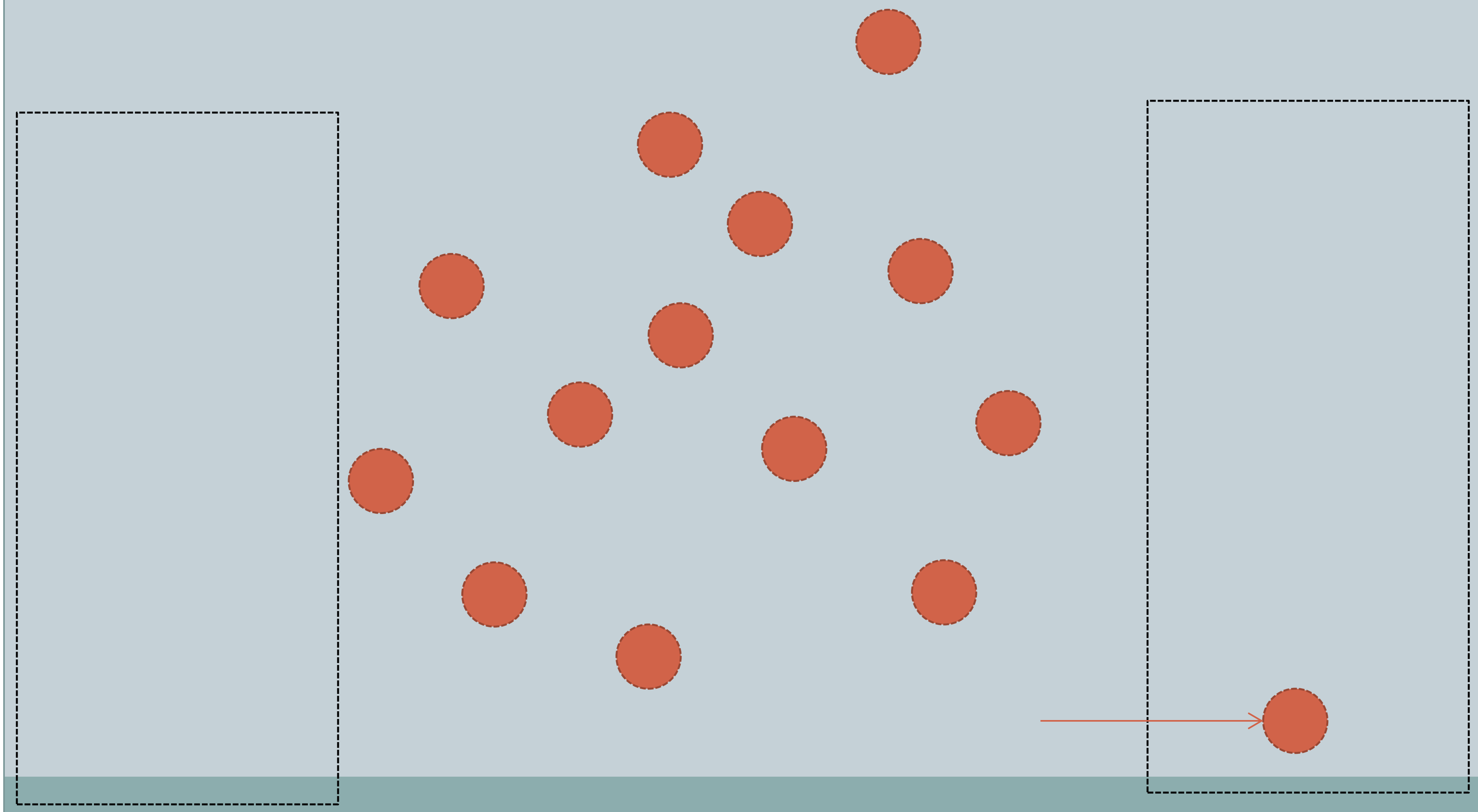
where does the data come from?

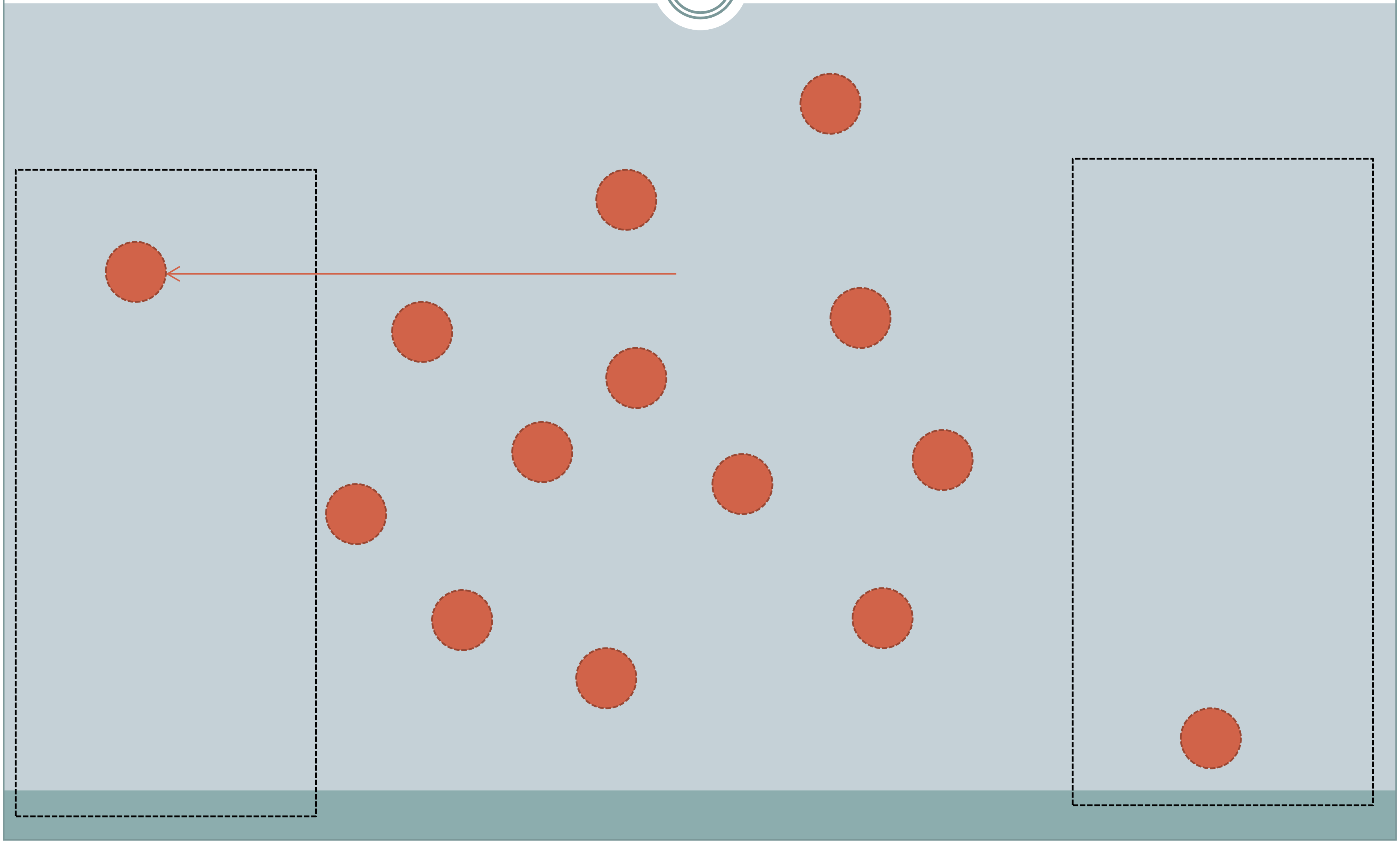
- We design trials.
  - Assign groups that are similar at baseline
  - Examine counterfactuals
- We also design surveys.
  - Representative groups
  - Understand population from subsets of those populations

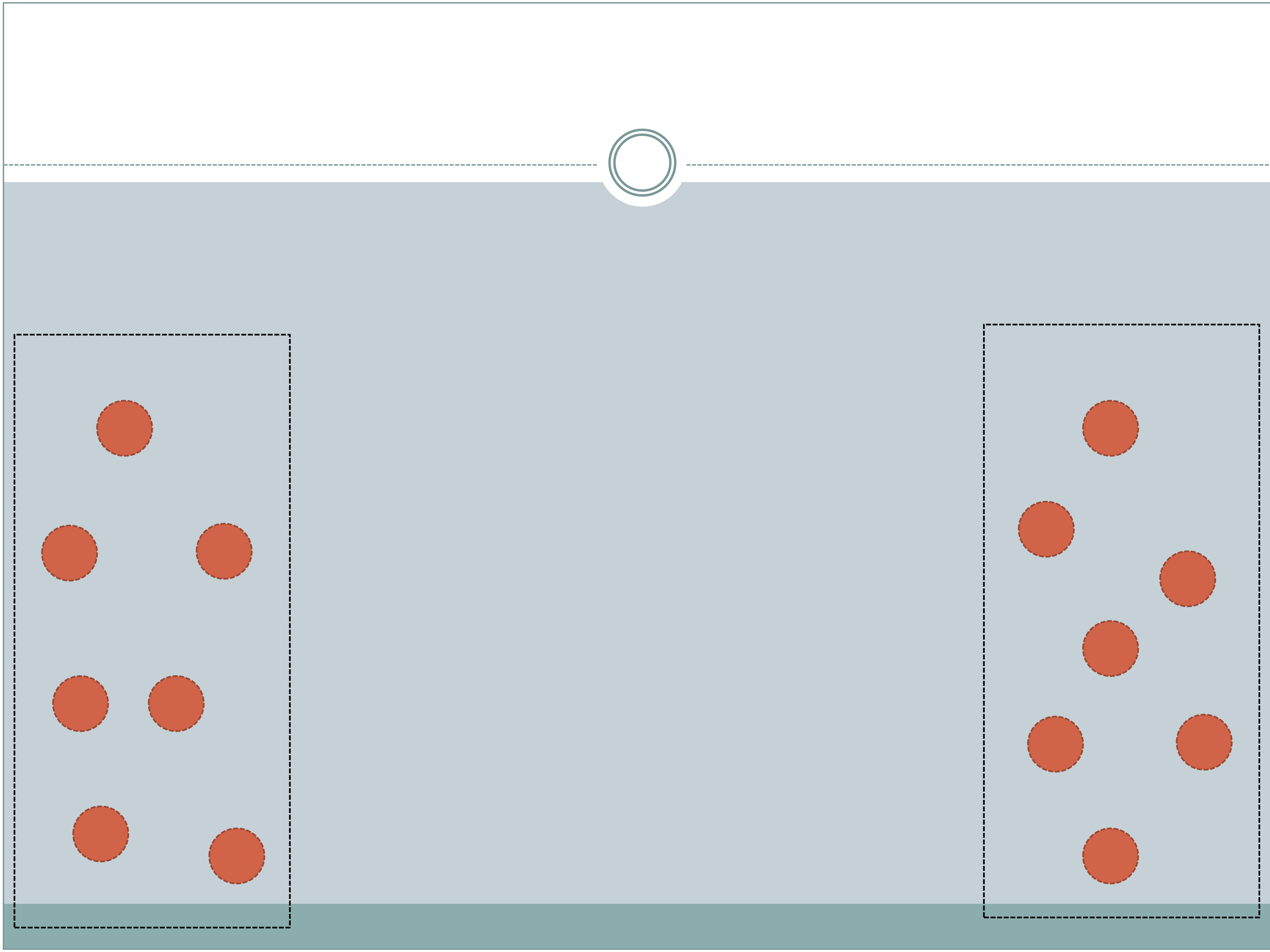
- Both use elements of control and randomness

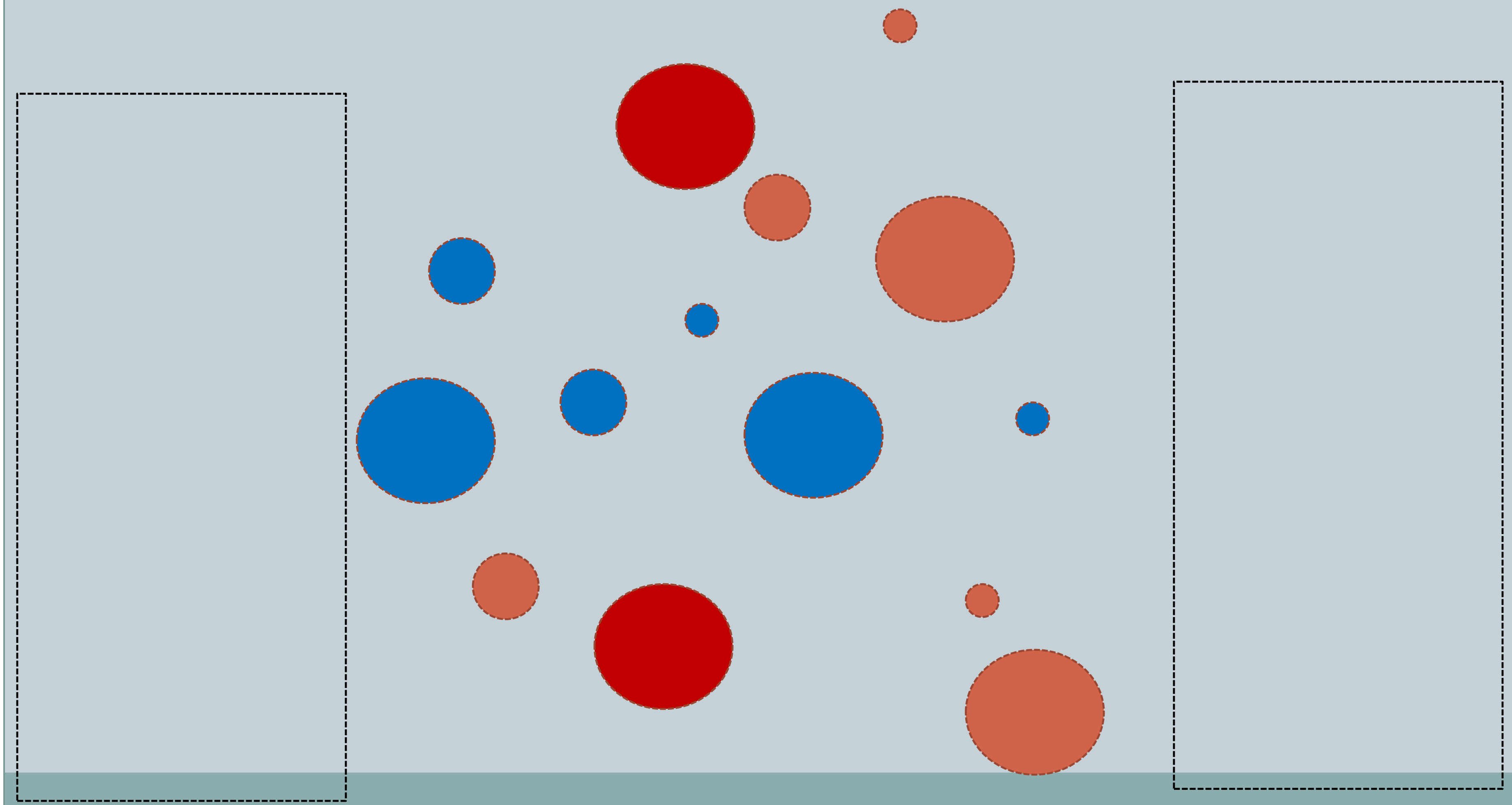
an example

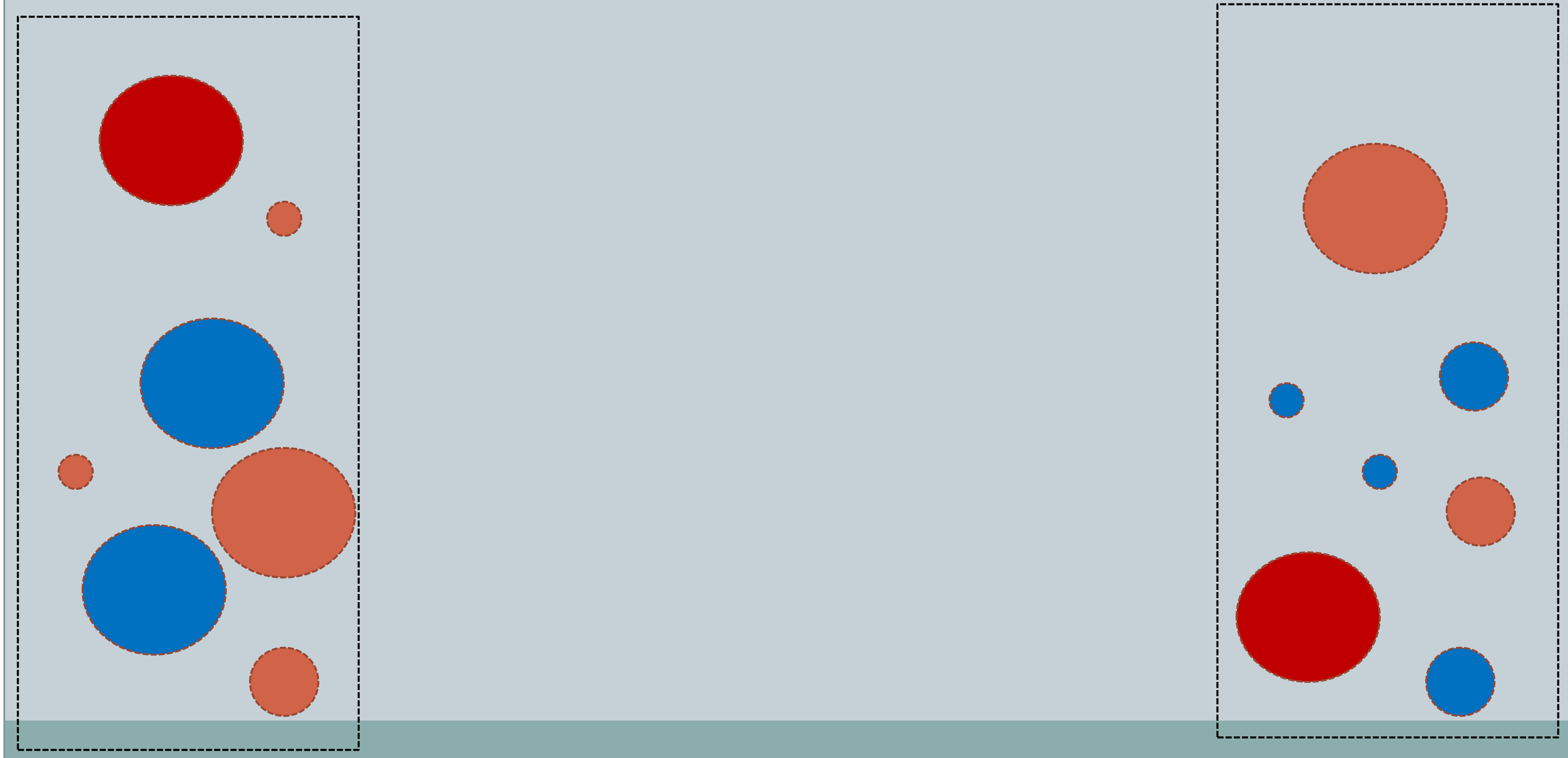
- Want to study a pill.
- Design the study
  - Uniform randomization
  - Matched pairs randomization
  - Crossover design
  - Cluster-randomized
- Inference
  - t-test
  - Matched-pairs t-test
  - Repeated measures model
  - Generalized linear mixed model
  - But… maybe all of those could be GLMM.

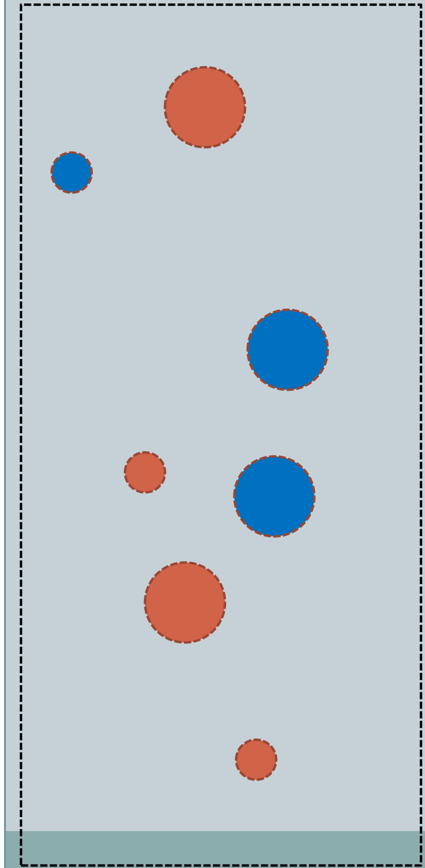




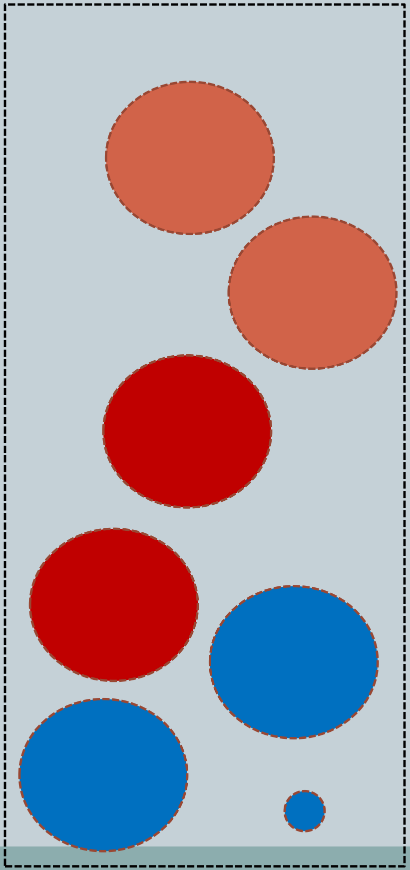


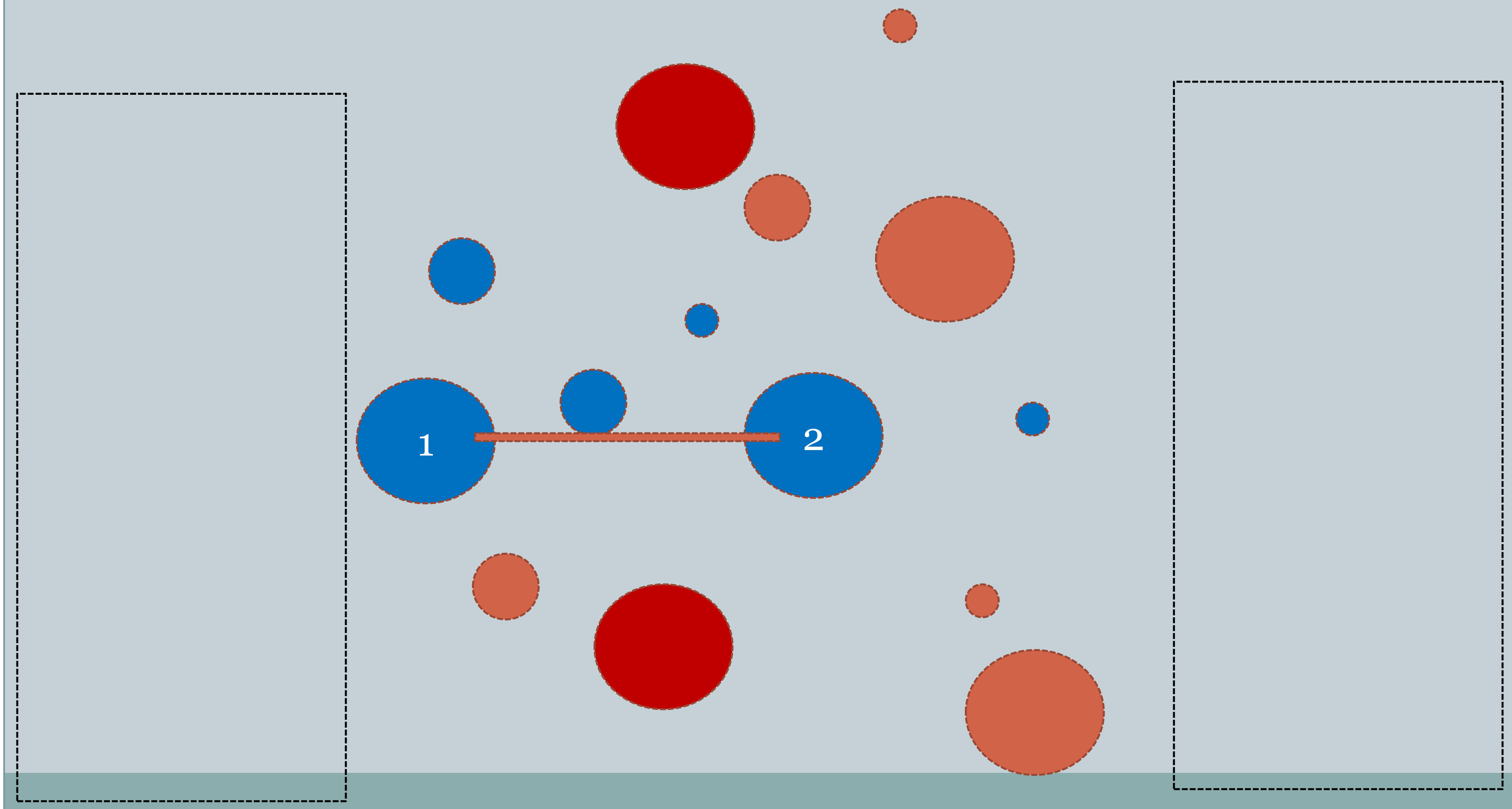


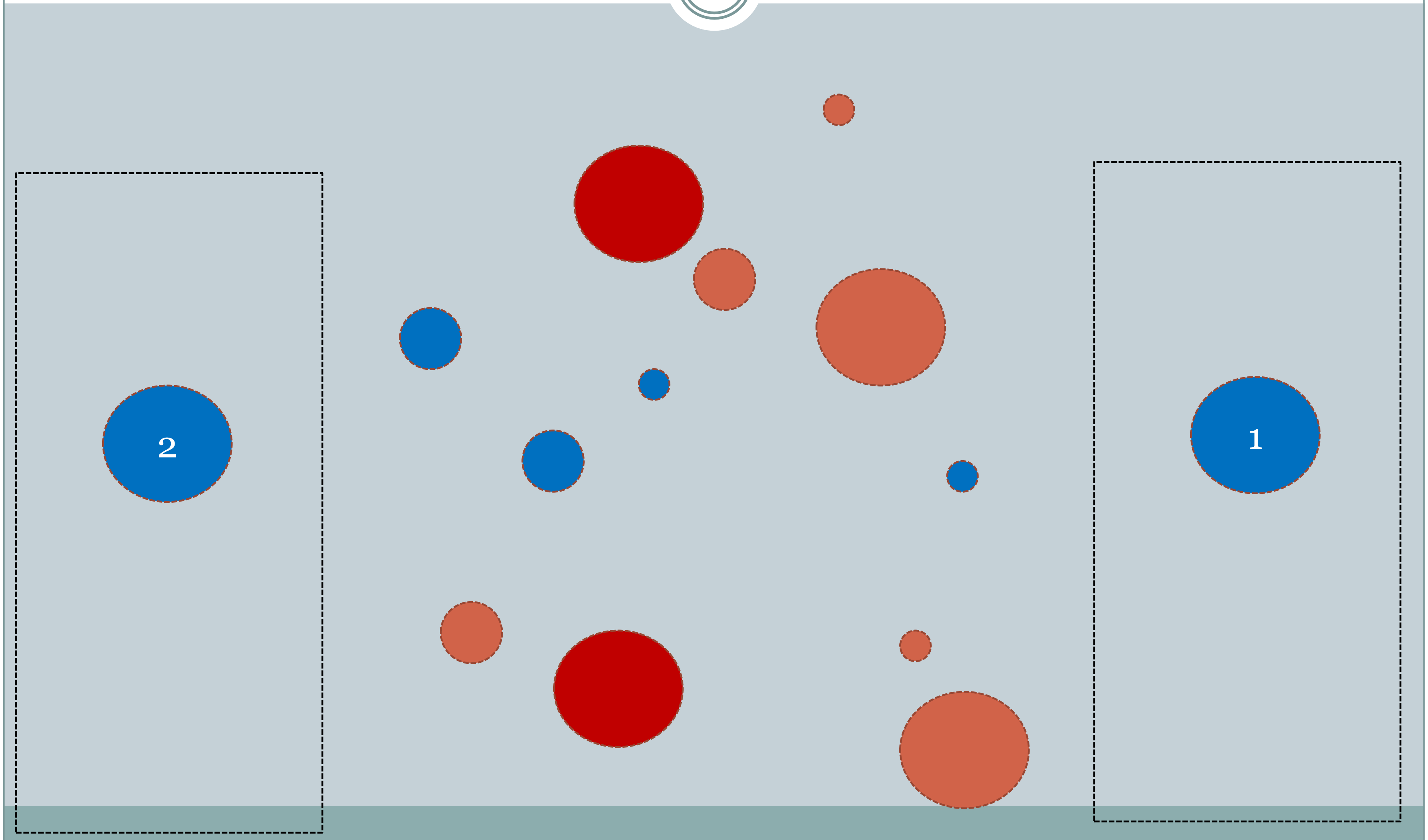


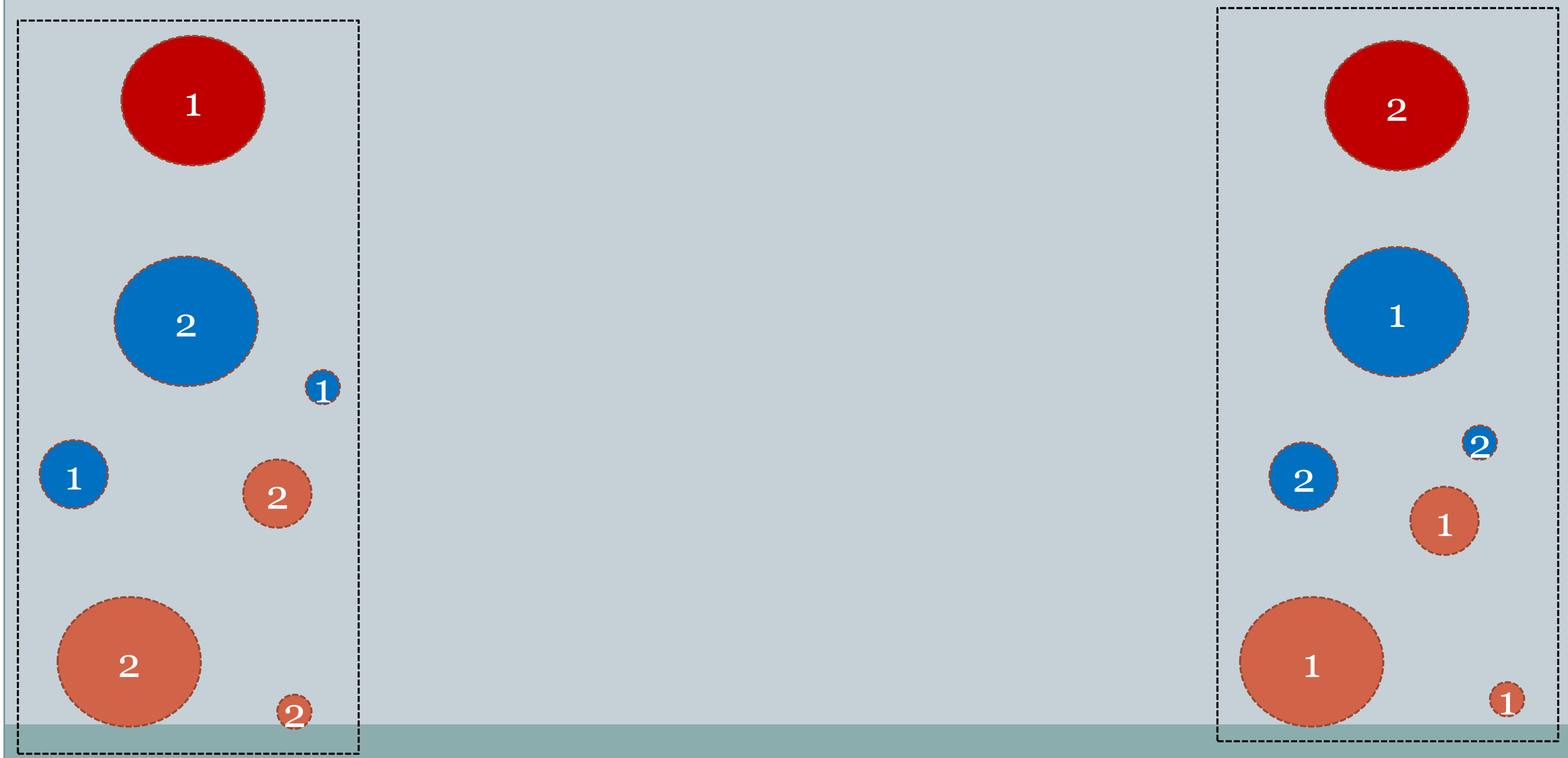


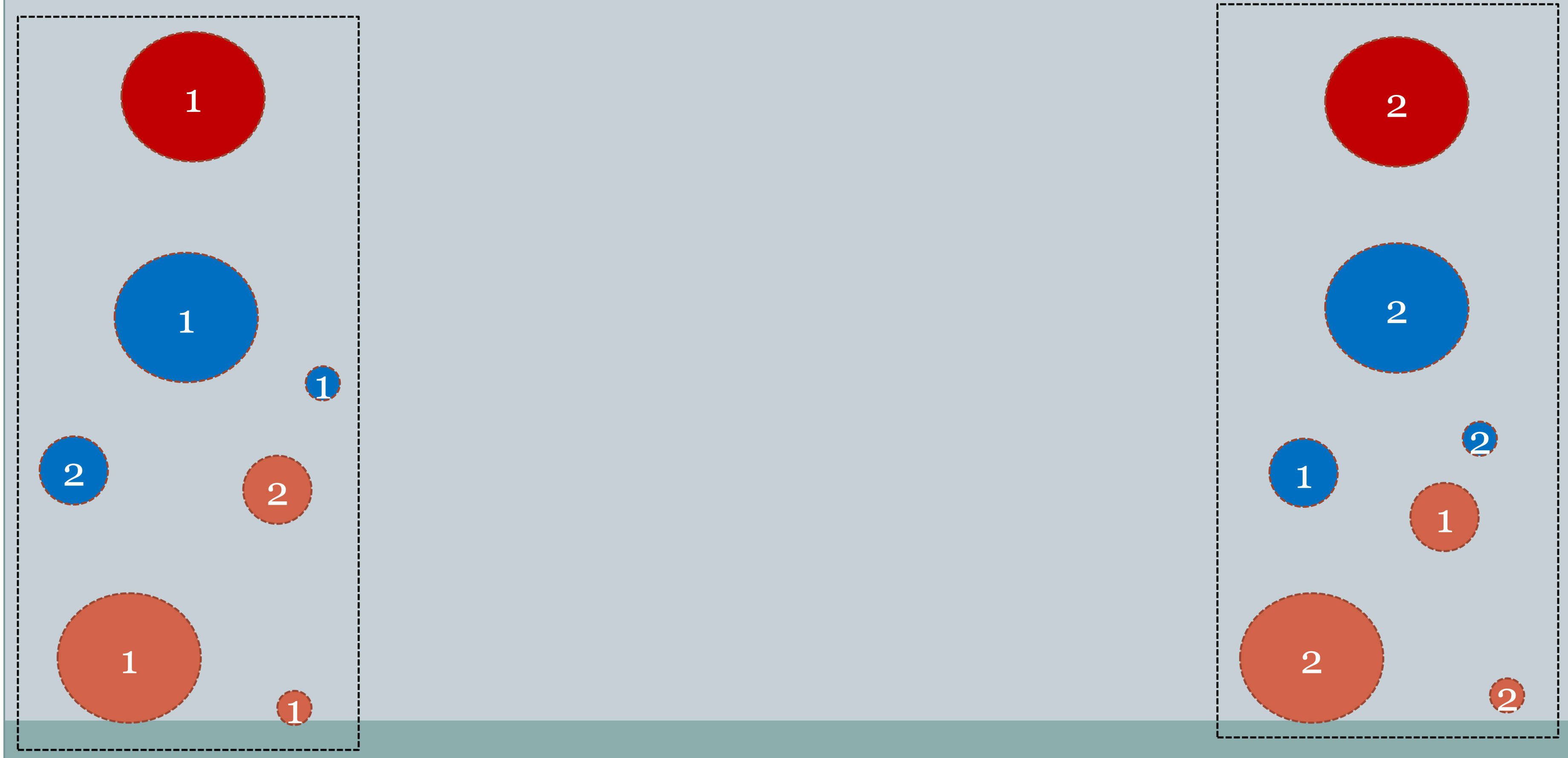
We’re off to a **bad** start.




1
2


2
1


1
2
1
1
2
2
2
2
1
2
2
1
1
1


1
1
1
2
2
1
1
2
2
2
1
1
2
2

the design limits what could have happened

- Uniform randomization allows for quite different possibilities. $({}_nC_r)$
- Matched-pairs randomization limits the size, and range, of possible assignments. $(2^{(\frac{n}{2})})$
- In some sense, we're losing something when we go to matched-pairs…
- … but what are we losing? The "crazy" options that we know are going to lead us astray.

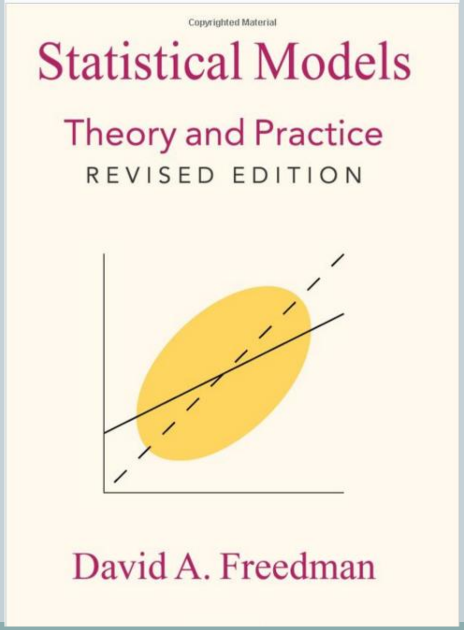
an example

- Want to study an election.
- Design the study
  - Simple random sample
  - Proportional sampling
  - Snowball sampling

- Inference
  - t-test
  - Inverse probability weighting
  - Generalized linear mixed model
  - But… maybe all of those could be GLMM.

different beliefs about where data come from

- RCT and survey
- Structural equation modeling
  - $y_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_p x_{p,i} + \varepsilon_i$
- If you want to be disabused of SEM spend some time reading



inference

picking inference

- Inference requires assumptions
- Linear regression:
  - Linearity and additivity
  - Independent errors
  - Homoskedastiticity
  - Normality of errors
- Permutation test:
  - No interference
- "Fancier" methods tend to have more assumptions… and thus leave you open to more lines of attack.
- These attacks can be obviated by careful preparation during the design phase.

picking inference

- Use the simplest method that gets the job done.
- If you want to accomplish more, collect more data or do additional analyses. ("If have to use something more complicated than a t-test then someone messed up…")

prospective study design

- A lot of the foundations have been worked out:
  - Experimental design
  - Survey sampling

- But, obviously, there are a lot of cool developments still going on:
  - Experimental design: adaptive trials, point-of-care randomization,
  - Sampling: active learning, explore exploit learning

observational (and retrospective) design

- This seems weird
- Usually a data set is in front of you, so you just analyze it
- It takes some thought to see this

Let’s do an example.

observational study design

NEONATAL INTENSIVE CARE UNITS

Application: Regionalization

- Hospitals vary in their ability to care for premature infants.
- The American Academy of Pediatrics recognizes levels: 1, 2, 3A, 3B, 3C, 3D and Regional Centers.
- ***Regionalization of care*** refers to a policy that suggests or requires that high-risk mothers deliver at hospitals with greater levels of capabilities.

Outcome

Outcome

The data

- Every baby delivered in a 10+ year period
  - California
  - Pennsylvania
  - Missouri
- Mothers' information
  - ICD9 codes
    - Delivery
    - Post-delivery complications
    - Some pre-delivery
  - Some SES information
  - Zip code of residence
- Birth/death certificates
- Census information
  - PA and MO have zip code level
  - CA will have block group

Pre-delivery Severity?

	Variable Type	High NICU	Low NICU	sd	Δ/sd
Mortality	Outcome	2.26%	1.25%	13.33%	0.08
Difference in Travel Time	Instrument	4.57	19.00	17.18	-0.84
% attending high level NICU	Treatment	100.0%	0.0%	49.7%	2.01
Birth weight	Preemie covariates	2,454.07	2,693.24	739.27	-0.32
Gestational age		34.61	35.69	2.80	-0.39
GI	% of preemies with type of congenital disorders	0.9%	0.6%	8.7%	0.04
GU		0.9%	0.8%	9.0%	0.01
CNS		0.9%	0.4%	8.3%	0.05
Pulmonary		0.8%	0.7%	8.8%	0.01
Cardio		1.4%	0.7%	10.5%	0.06
Skeletal		0.7%	0.9%	9.0%	-0.02
Skin		0.0%	0.0%	0.0%	0.00
Chromosomes		0.4%	0.3%	6.3%	0.02
Other_Anomaly		0.8%	0.1%	7.0%	0.09
Gestational_DiabetesM	Mother covariates	4.9%	4.3%	21.0%	0.03
Mother's education		3.76	3.58	1.19	0.16
Insurance - Fee for service		24.0%	24.5%	42.8%	-0.01
Insurance - HMO		32.3%	27.8%	46.0%	0.10
Insurance - Government		23.5%	24.2%	42.6%	-0.02
Insurance - Other		16.8%	21.4%	39.1%	-0.12
Uninsured		2.2%	1.6%	13.7%	0.04
Prenatal care		2.51	2.37	1.30	0.11
Single birth (y/n)		79.0%	86.1%	38.3%	-0.18
Parity		2.08	2.09	1.31	-0.01
Mother's age		28.41	27.71	6.25	0.11
Median income	Census level covariates	41,484.25	40,258.92	14,587.24	0.08
Median home value		97,663.00	95,083.15	48,762.43	0.05
% completed high school		79.9%	80.0%	9.7%	-0.01
% completed college		22.2%	19.4%	13.1%	0.21
% renting		31.4%	27.9%	12.8%	0.28
% below poverty line		13.4%	11.8%	9.9%	0.16

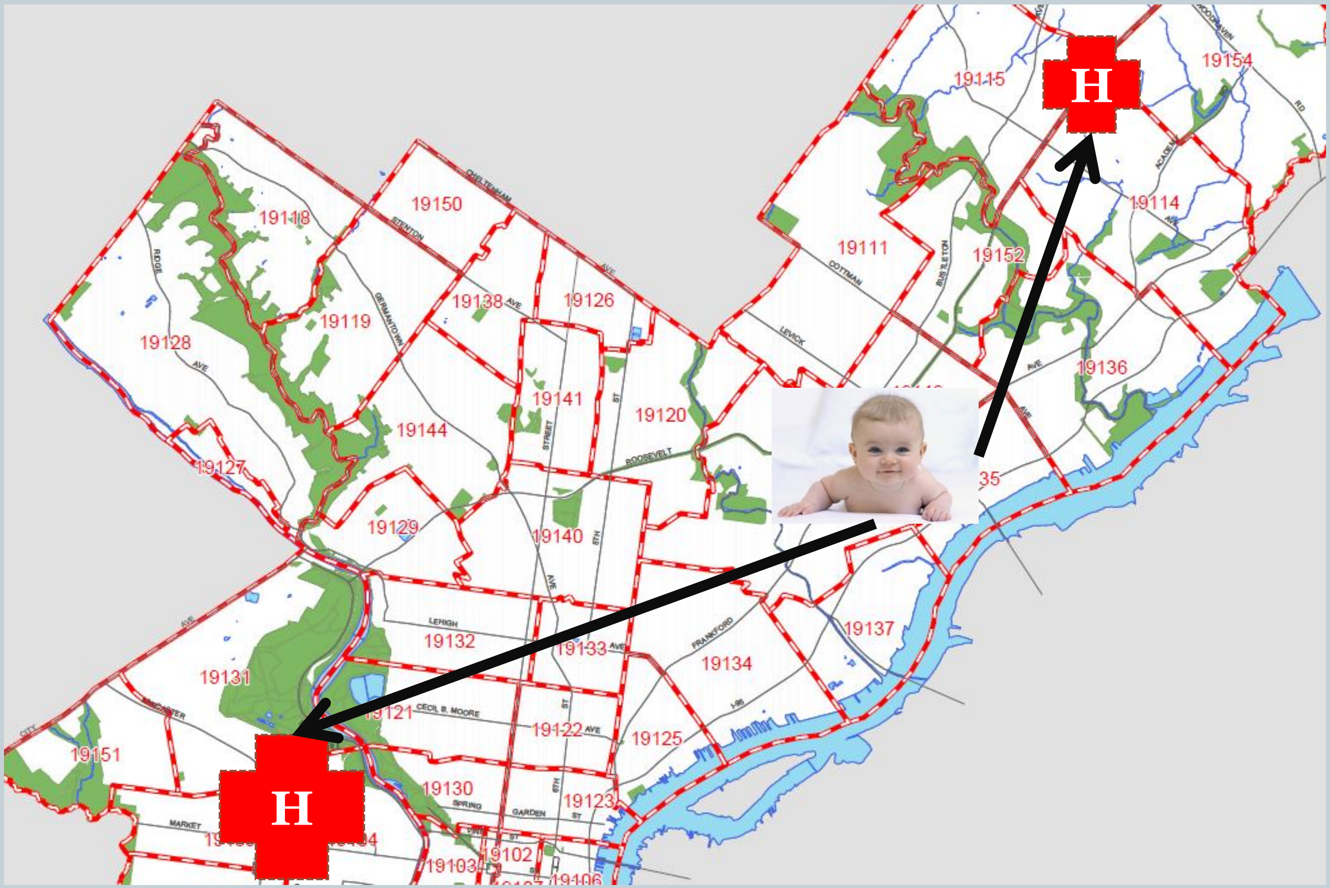
Summary of Problem

- Want to quantify effect of level of NICU on rate of death
- Observational data
- Sorting bias
- Some sorting variables are **unobserved**

real world randomness

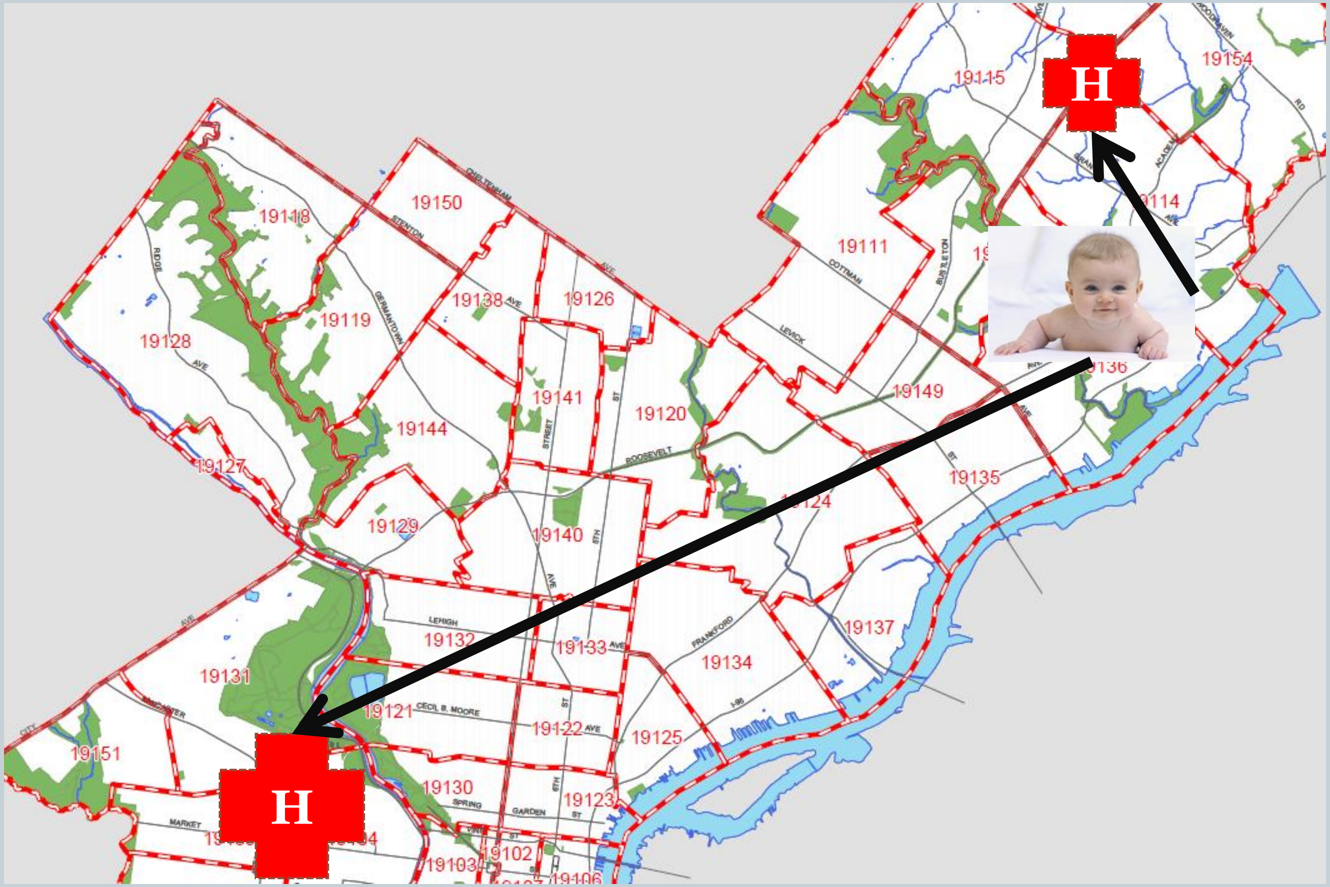
(MAYBE)

Instrument: Excess Travel Time



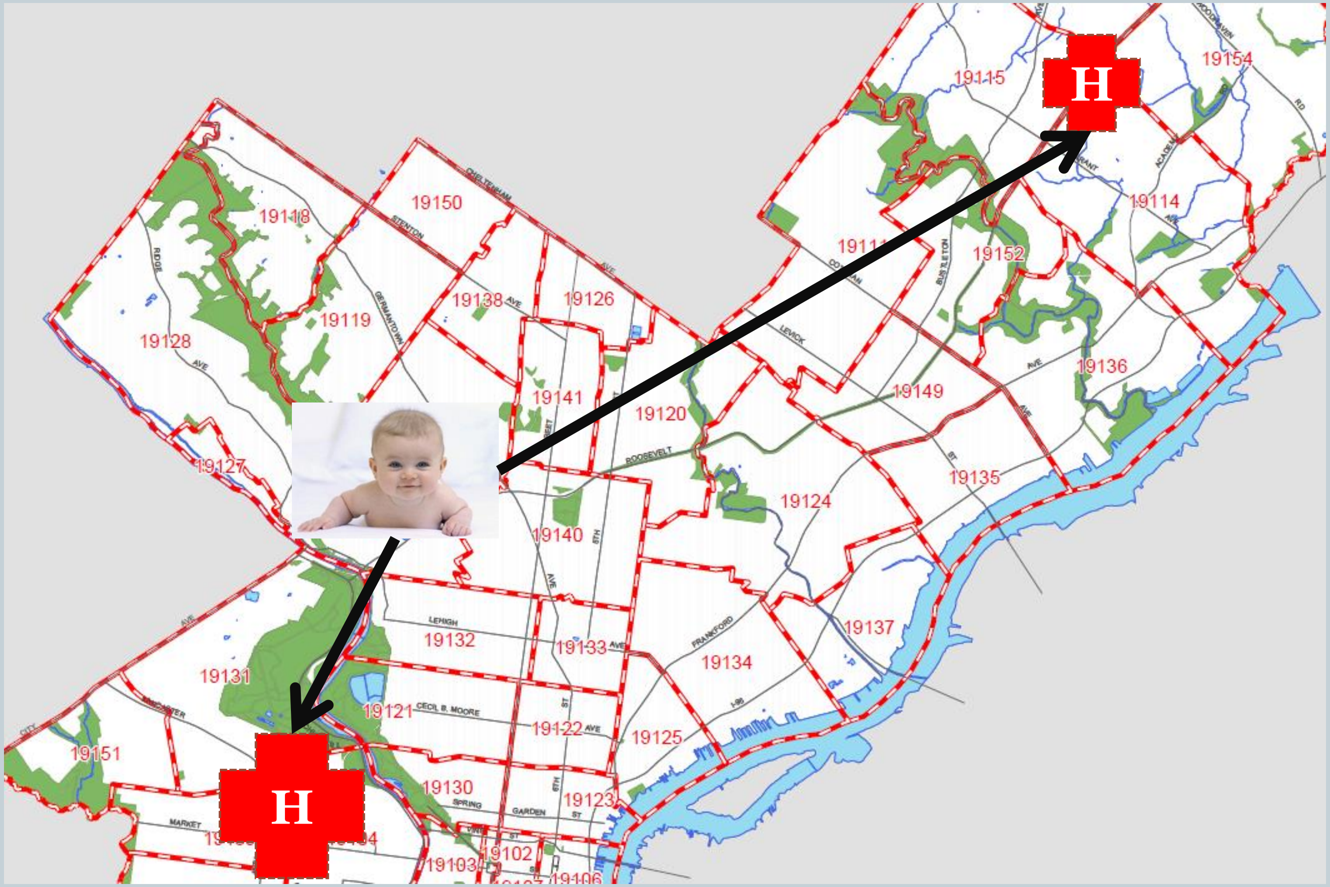
Excess Travel Time

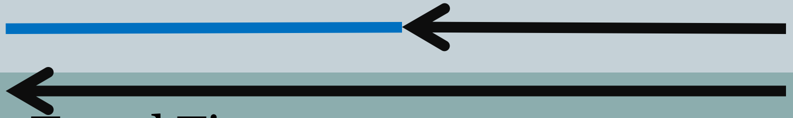
Instrument: Excess Travel Time



Excess Travel Time

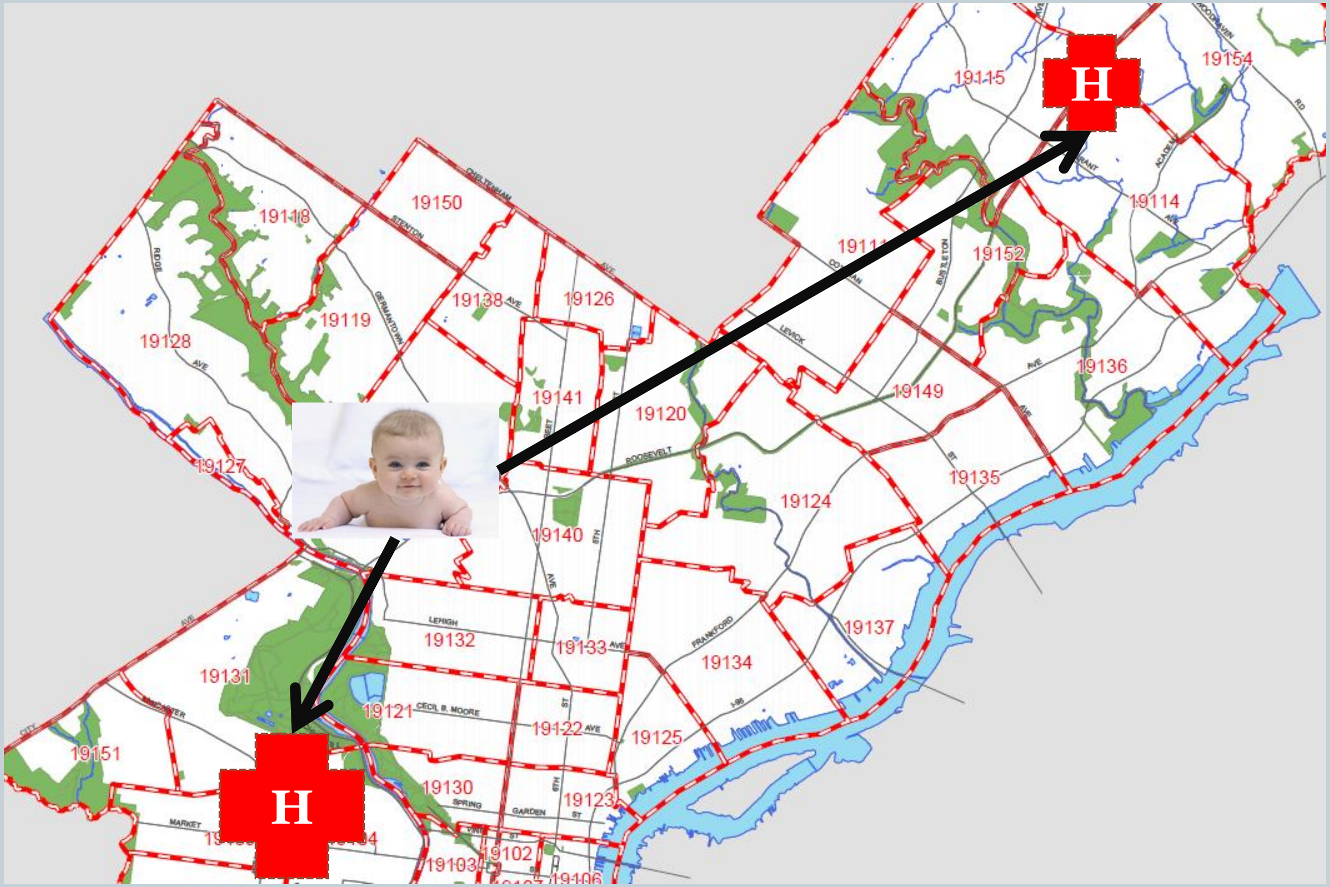
Instrument: Excess Travel Time





Excess Travel Time

Instrument: Excess Travel Time

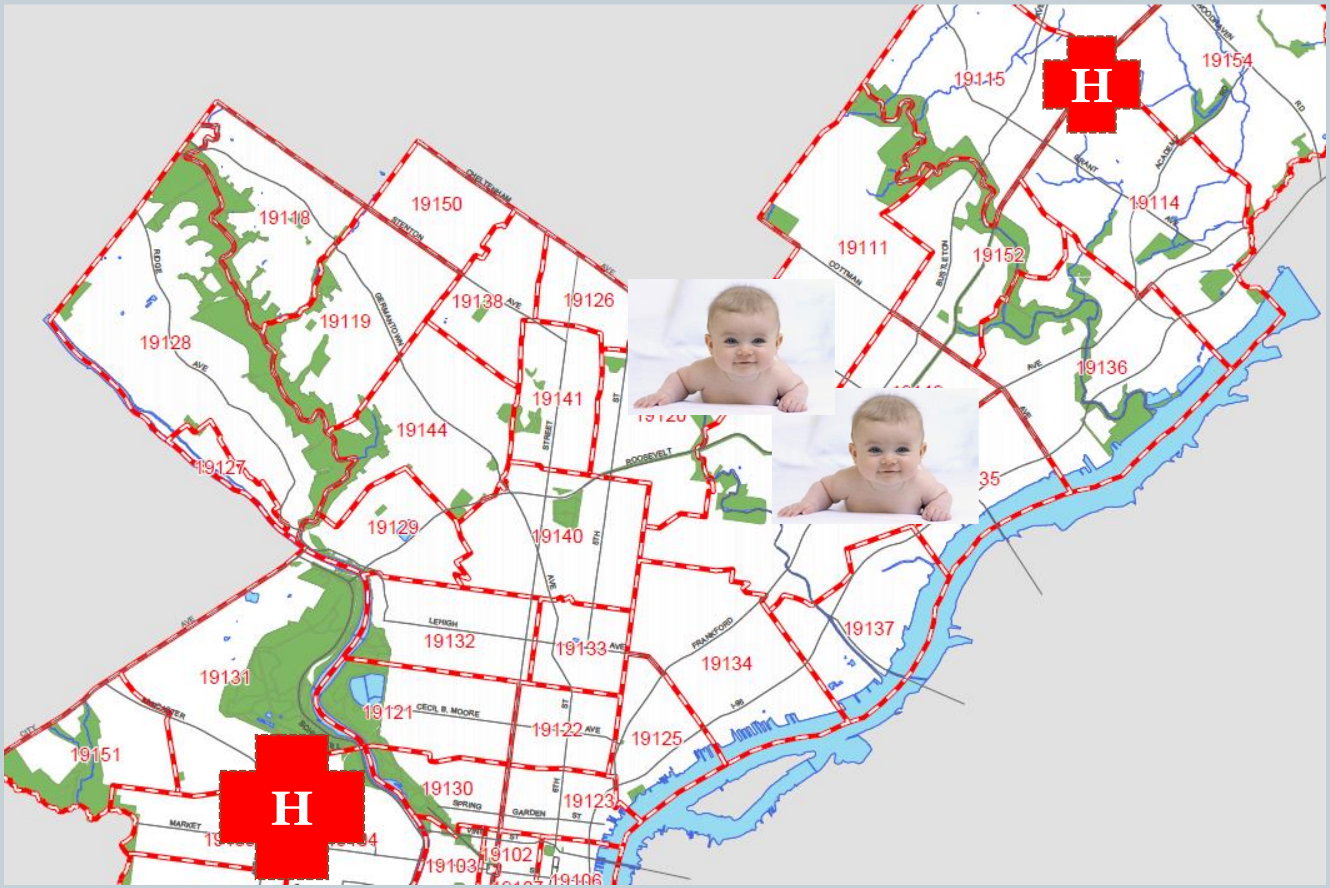


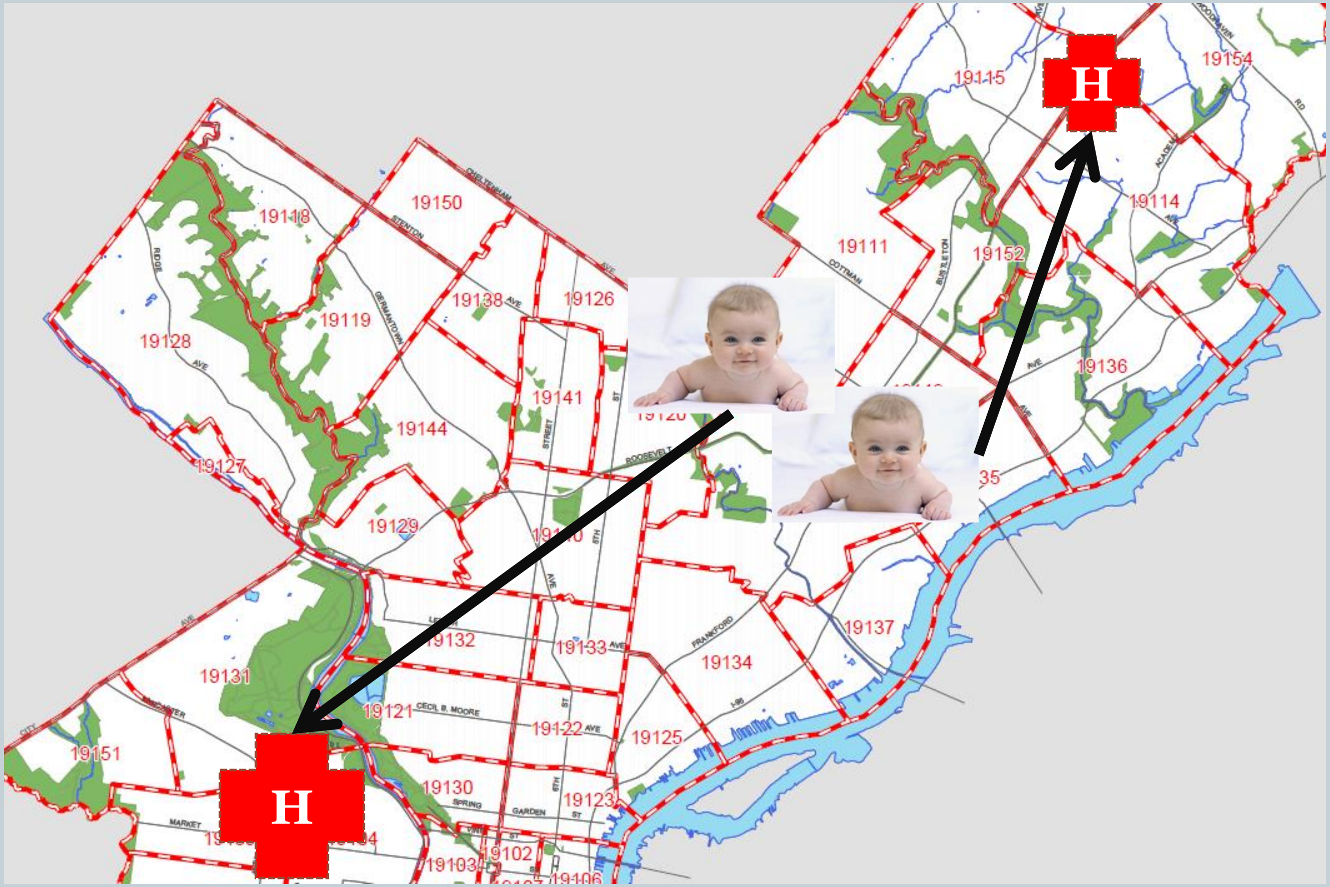
McClellan, McNeil & Newhouse; "**Does more intensive treatment of acute myocardial infarction reduce mortality?**" *JAMA*. 272(11): 859-66, September 1994

revised design

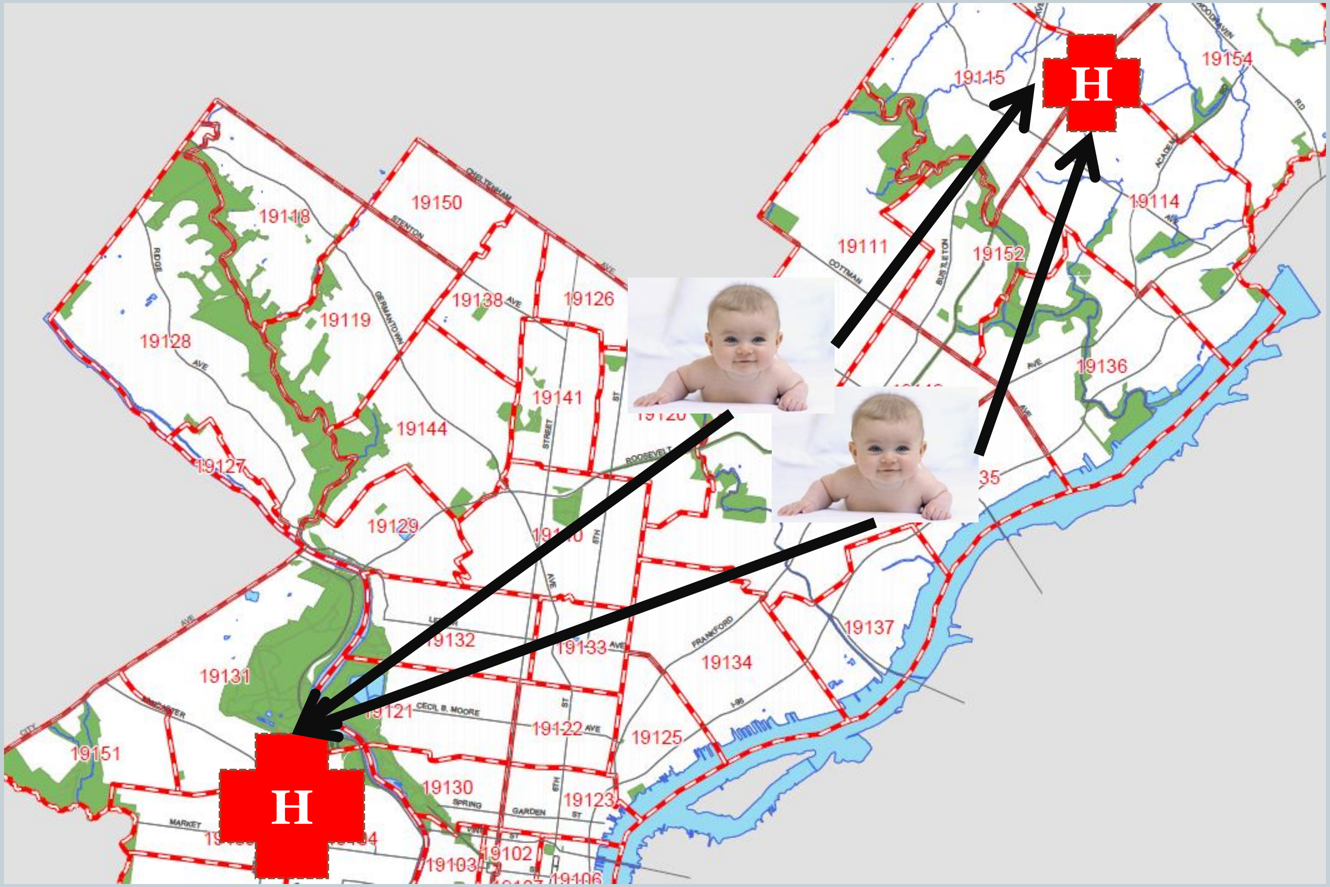
NEAR-FAR MATCHING

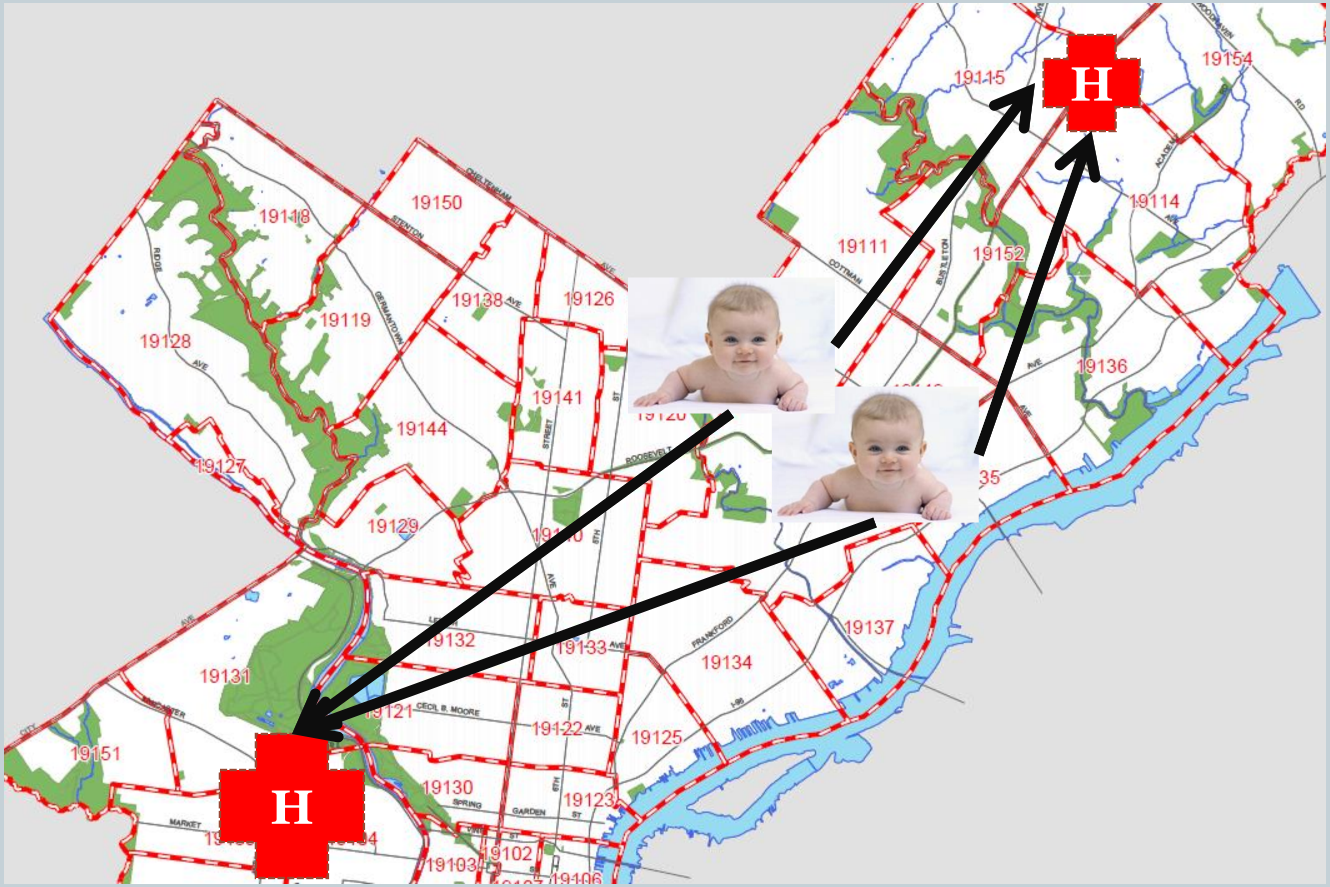
H
19115
19150
CHELTENHAM
STENTON
19118
RIDGE
GERMANTOWN
19128
19138
19126
19141
19127
ROOSEVELT
19140
LEHIGH
19132
19133
FRANKFORD
19134
19137
19131
19121
CECIL B. MOORE
19122
19125
19151
MARKET
H


H
19154
19115
19114
19111
19152
19150
CHELTENHAM
STENTON
19118
RIDGE
19138
19126
19119
19128
COTTMAN
BUSTLETON
19136
19141
19144
19127
ROOSEVELT
19129
19140
19137
LEHIGH
19132
19133
FRANKFORD
19134
19131
19121
CECIL B. MOORE
19122
19125
I-95
19151
H
19130
SPRING
GARDEN
19123
MARKET
19102
19103
19106

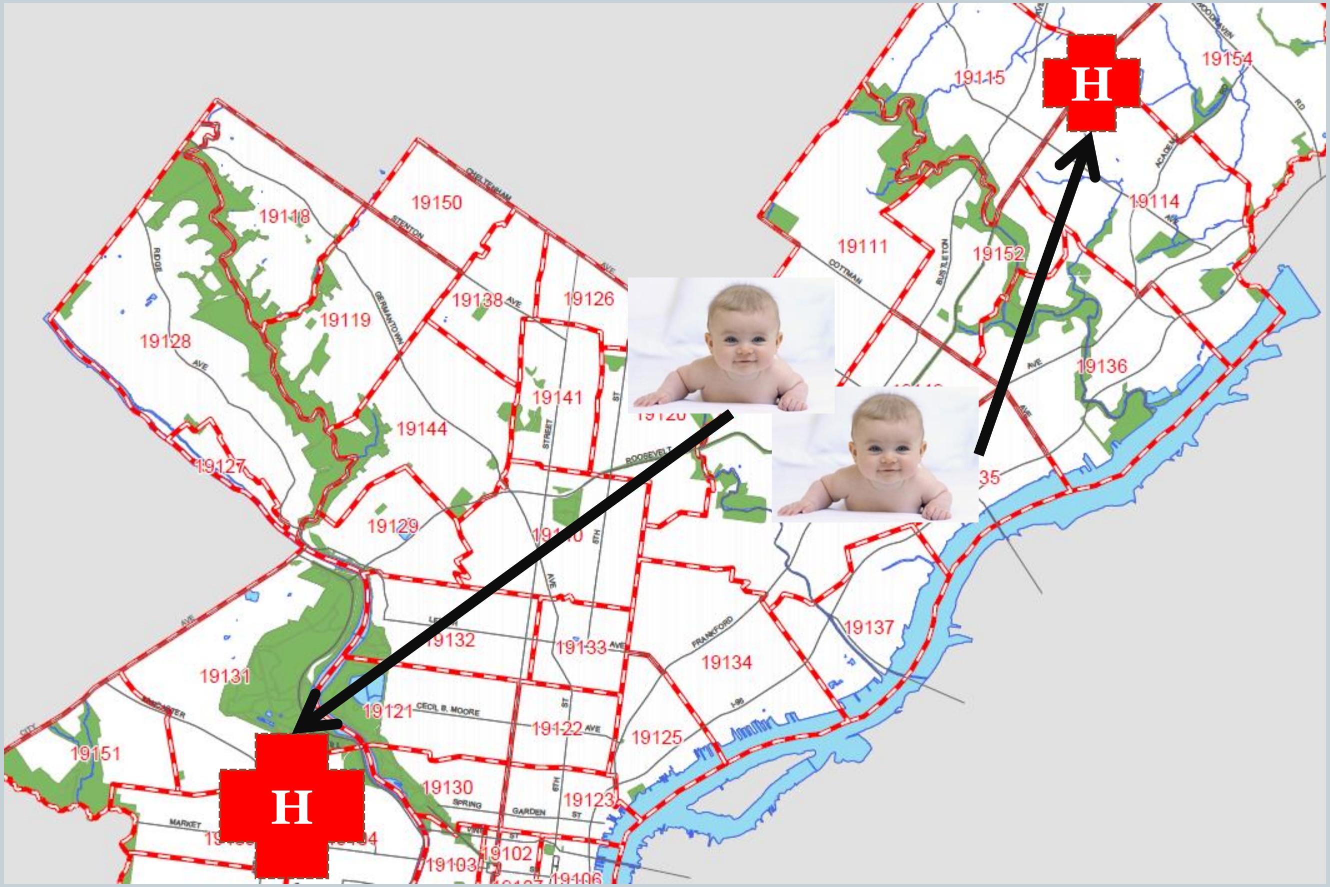


19154
19115
19150
19118
19114
19111
19152
19138
19126
19119
19128
19136
19141
19144
19127
19129
19132
19133
19137
19134
19131
19121
19122
19125
19151
19130
19123
19102
19103
19106
H
H

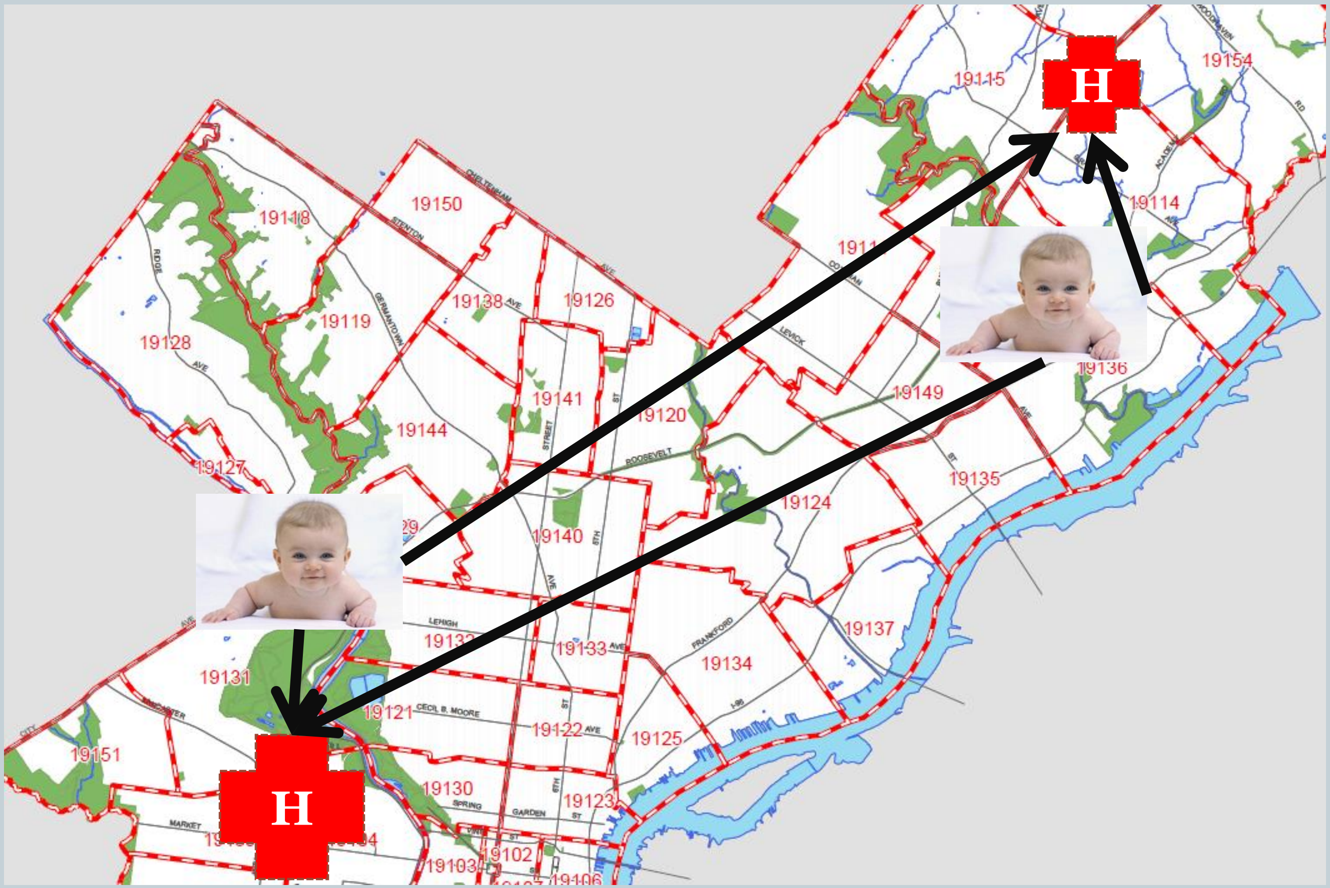

H
19115
19154
19114
19111
19152
19136
19118
19150
19119
19128
19138
19126
19141
19144
19127
19129
19137
19132
19133
19134
19131
19121
19122
19125
19151
19130
19123
19102
19103
19106
H



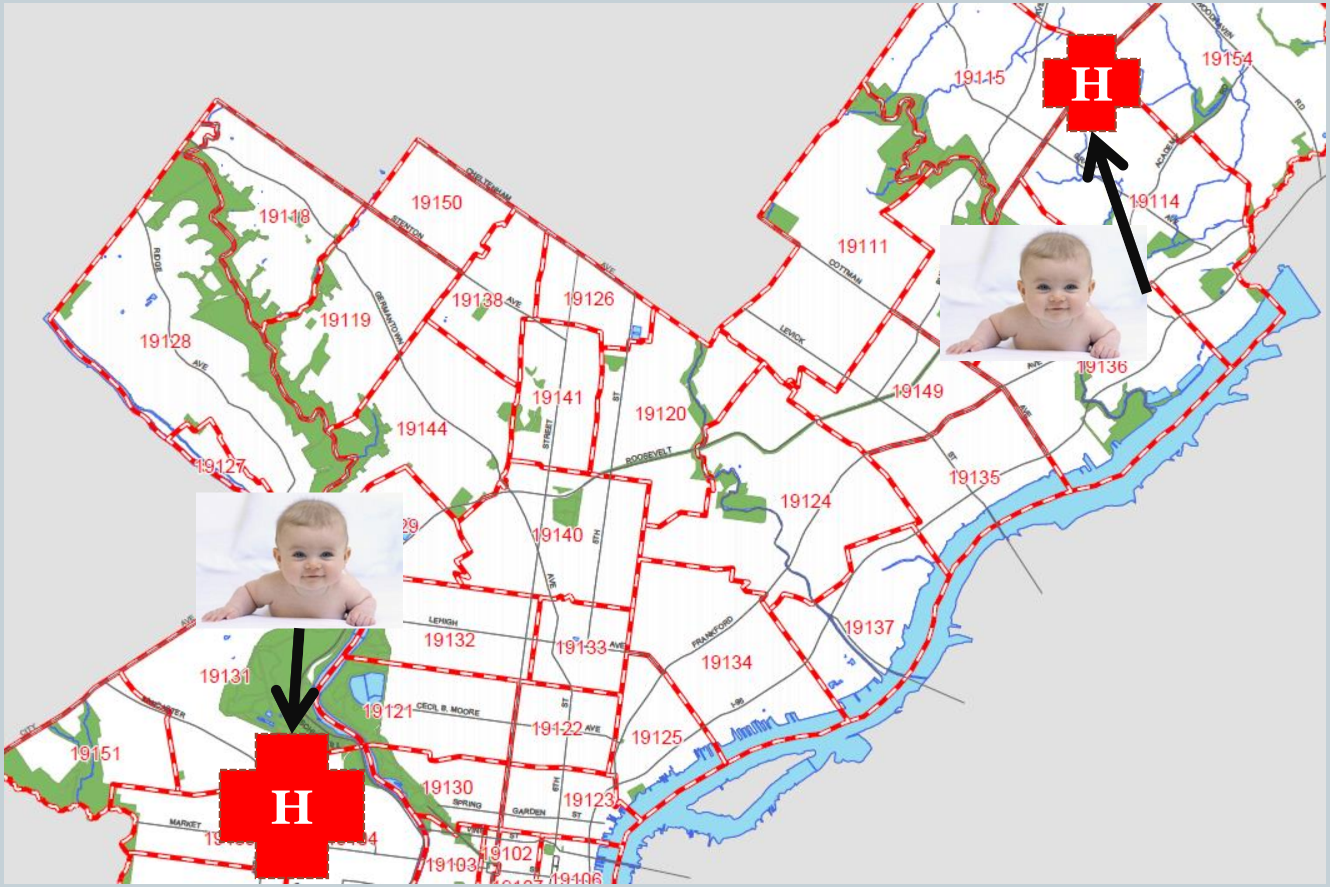
Sorting is potentially biased!



Sorting is potentially biased!



Sorting largely due to the randomness!



Sorting largely due to the randomness!

linking to inference

the fundamental form

- For RCTs, the fundamental form of inference is based on permutation tests (a.k.a. randomization tests)
- For survey, the fundamental form of inference is bootstrap (debatable)
- Everything else are necessary concessions to the particularities of a given problem
- Connect the structure of the data to your form of inference

takeaways

takeaways

- Design comes it two flavors
  (actually, three… but the third one is not very healthy)
- In prospective studies
  - design is an obvious consideration
  - and one that MUST be passed through in order to obtain data
- In retrospective studies,
  - design is a less obvious consideration
  - but one that MUST be passed through… unfortunately without much attention paid