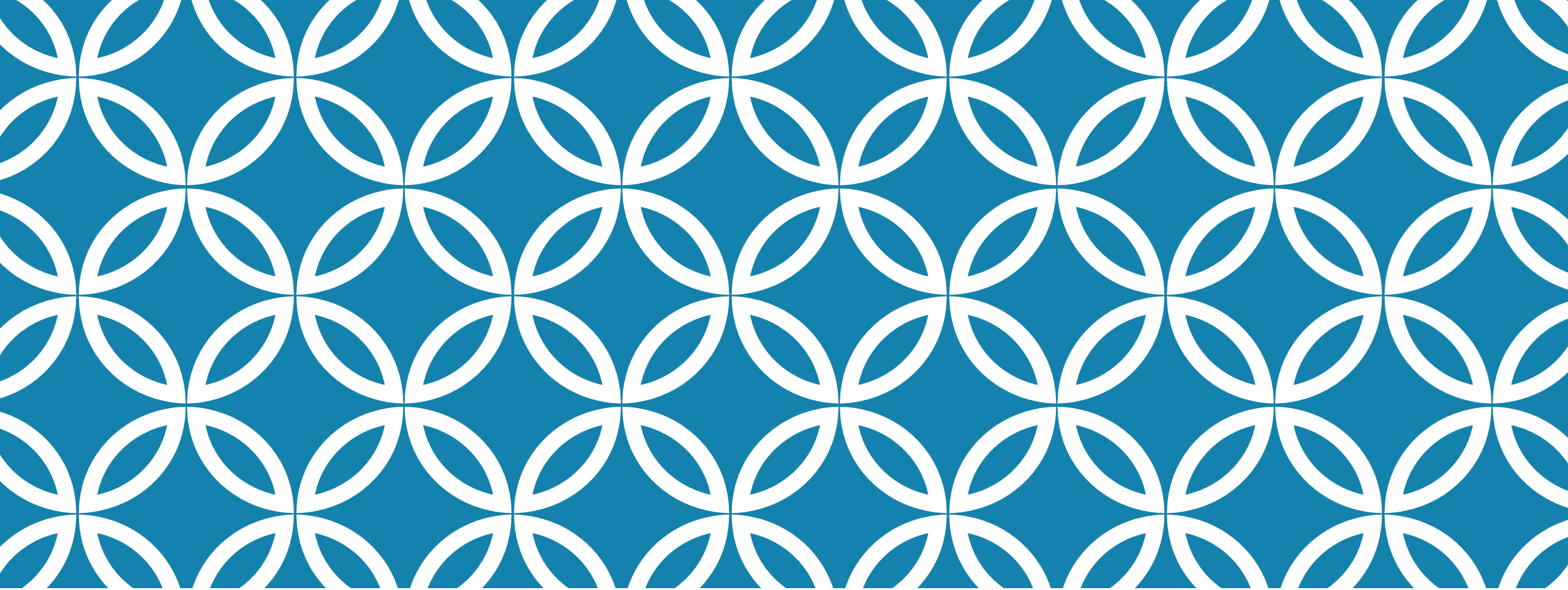


PROGRAMMING FOR SOCIAL SCIENTISTS WITH INTERPOLL

Ben Livshits
Todd Mytkowicz
Microsoft Research

InterPoll 101

We want to give access to polls and survey to the regular developer.

```
1  var people = new MTurkQueryable<Person>(true, 5, 100, 2);
2  var liberalArtsPairs = from person in people
3          where person.Employment == Employment.STUDENT
4          select new {
5                  Person = person,
6                  Value = person.PoseQuestion<bool>(
7                          "Are you a liberal arts major?")
8                  };
```

This is a LINQ query for asking college students whether they are liberal arts majors

We want to make access to human-generated data as easy as access to databases.

SIMPLE LINQ QUERIES

```
var femaleHeight = from person in people where person.Gender ==
Gender.FEMALE select person.PoseQuestion<int>("What is your height?");

var maleHeight = from person in people where person.Gender ==
Gender.MALE select person.PoseQuestion<int>("What is your height?");
```

```
if (maleHeight.ToRandomVariable() > femaleHeight.ToRandomVariable()) {

      Console.WriteLine(

          "Males are taller that females, according to a t-test.");

}
```

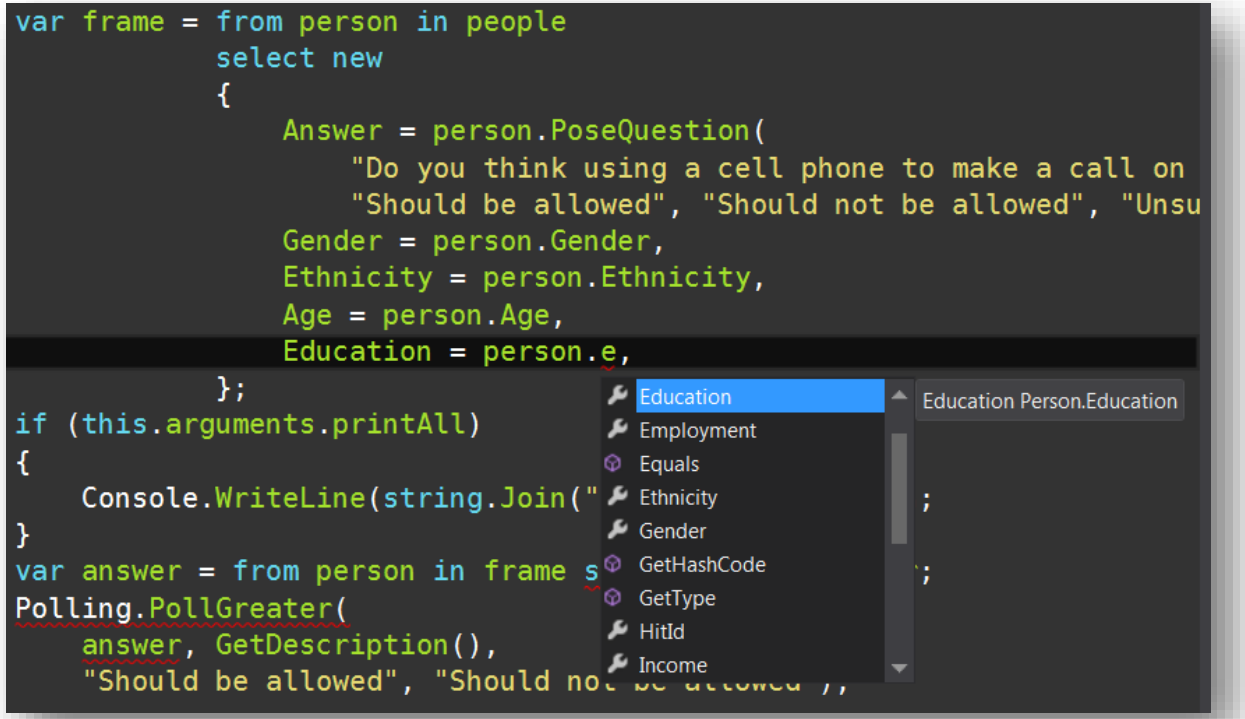
C# + LINQ

LINQ (Language-integrated queries) is the only way to create surveys, but in has several important advantages.

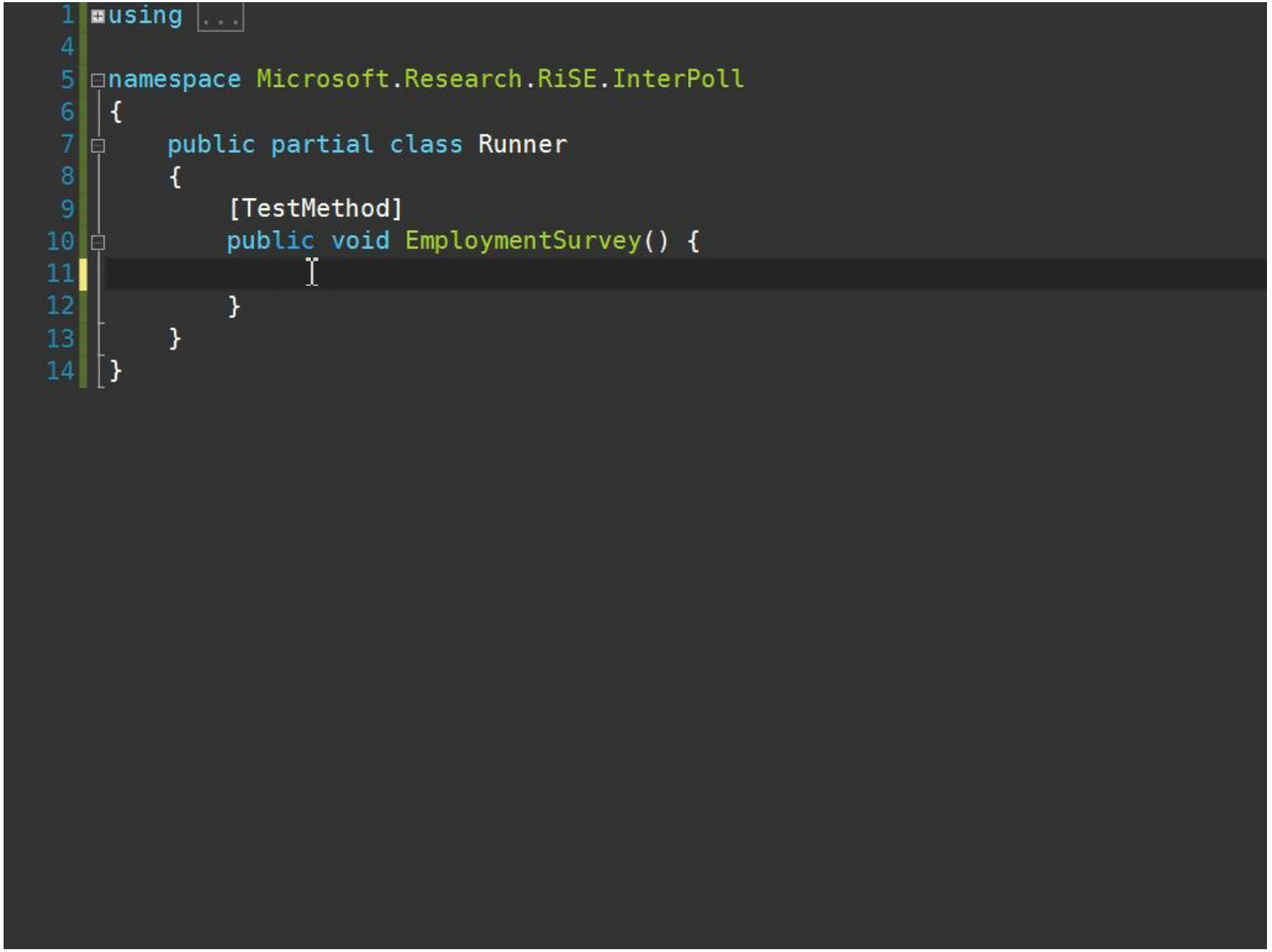
In InterPoll, We are thinking a lot about the developer experience. LINQ is familiar to developers and does not require learning a new domain-specific language.

LINQ queries integrate well with the rest of the application.

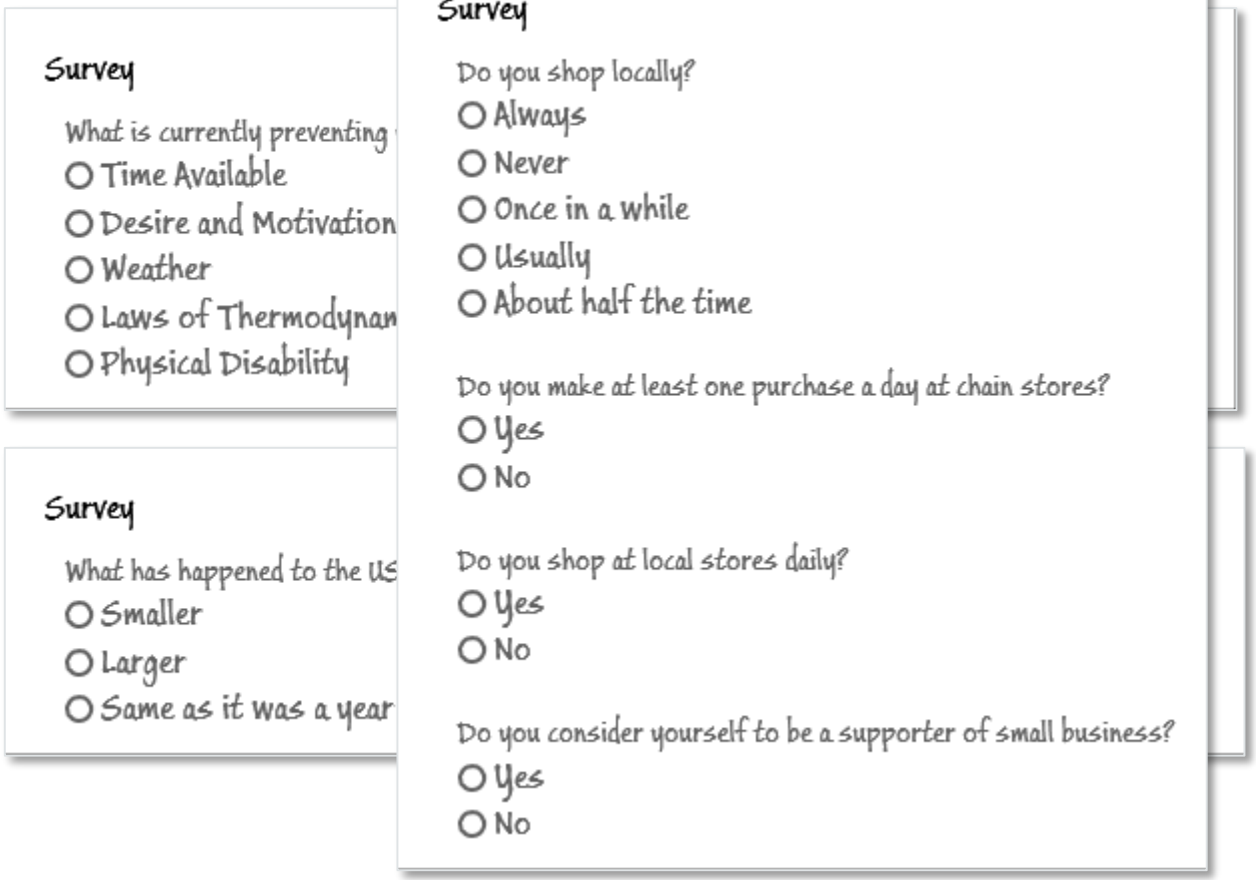
Developers can continue using familiar tools such as Visual Studio, with powerful IntelliSence support and type checking, to avoid making mistakes.



DEMO


using ...
namespace Microsoft.Research.RiSE.InterPoll
{
public partial class Runner
{
[TestMethod]
public void EmploymentSurvey() {
}
}
}

SAMPLE SURVEYS



Survey

What is currently preventing
- Time Available
- Desire and Motivation
- Weather
- Laws of Thermodynan
- Physical Disability

Survey

What has happened to the US
- Smaller
- Larger
- Same as it was a year

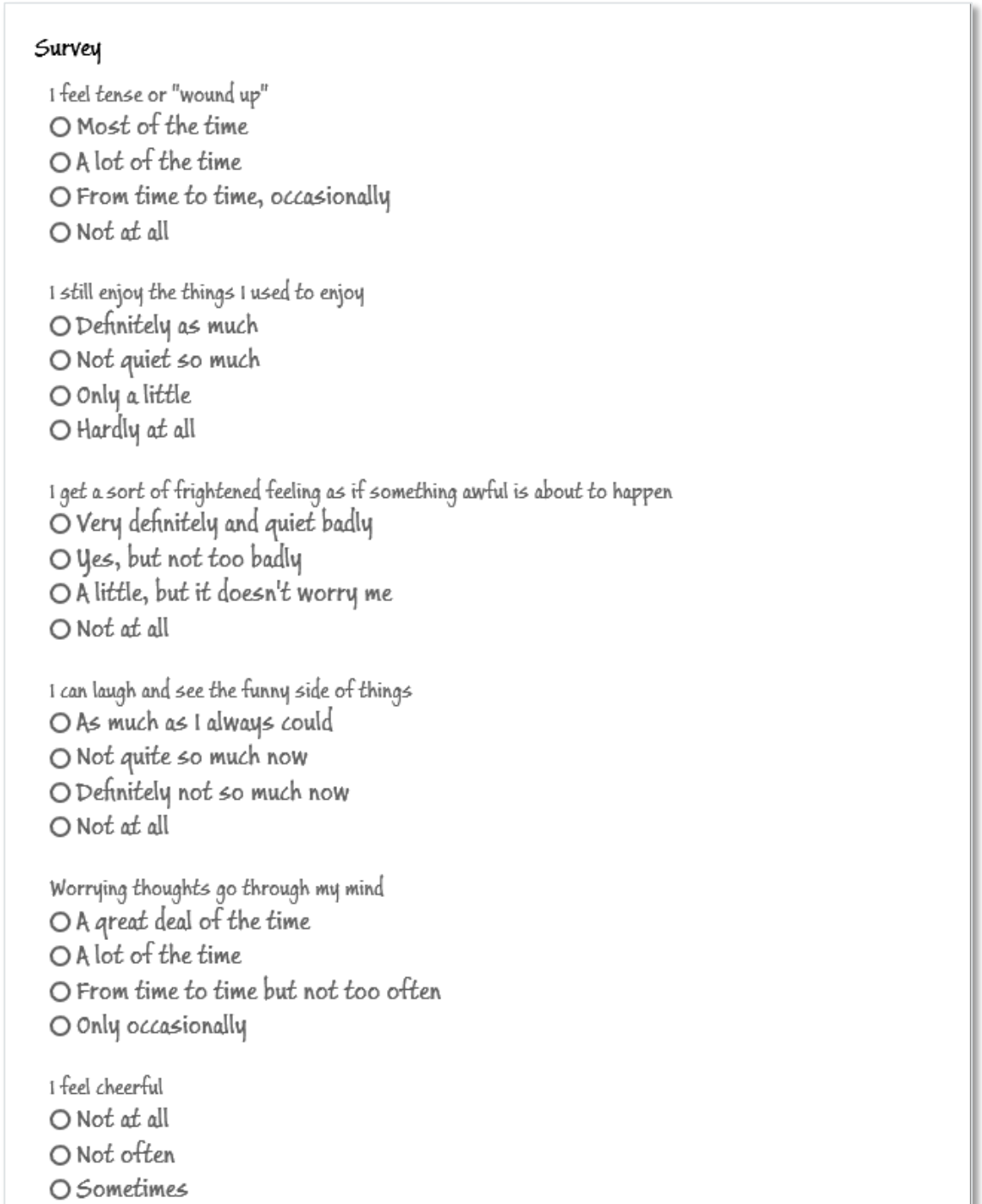
Survey

Do you shop locally?
- Always
- Never
- Once in a while
- Usually
- About half the time

Do you make at least one purchase a day at chain stores?
- Yes
- No

Do you shop at local stores daily?
- Yes
- No

Do you consider yourself to be a supporter of small business?
- Yes
- No



Survey

I feel tense or "wound up"
- Most of the time
- A lot of the time
- From time to time, occasionally
- Not at all

I still enjoy the things I used to enjoy
- Definitely as much
- Not quiet so much
- Only a little
- Hardly at all

I get a sort of frightened feeling as if something awful is about to happen
- Very definitely and quiet badly
- Yes, but not too badly
- A little, but it doesn't worry me
- Not at all

I can laugh and see the funny side of things
- As much as I always could
- Not quite so much now
- Definitely not so much now
- Not at all

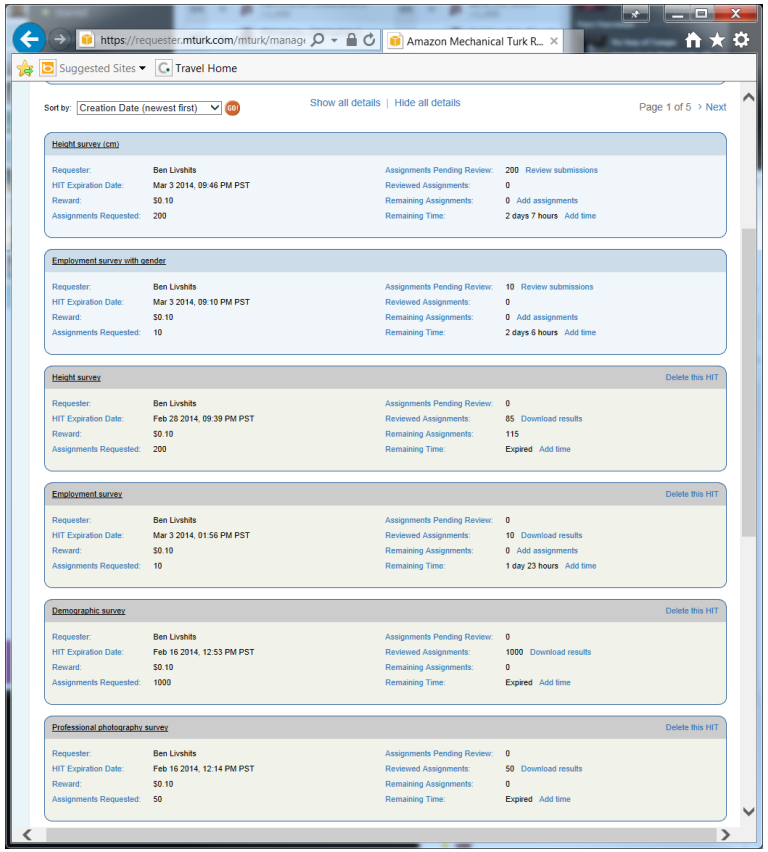
Worrying thoughts go through my mind
- A great deal of the time
- A lot of the time
- From time to time but not too often
- Only occasionally

I feel cheerful
- Not at all
- Not often
- Sometimes

MECHANICAL TURK BACK-END

InterPoll uses Mechanical Turk as a back-end

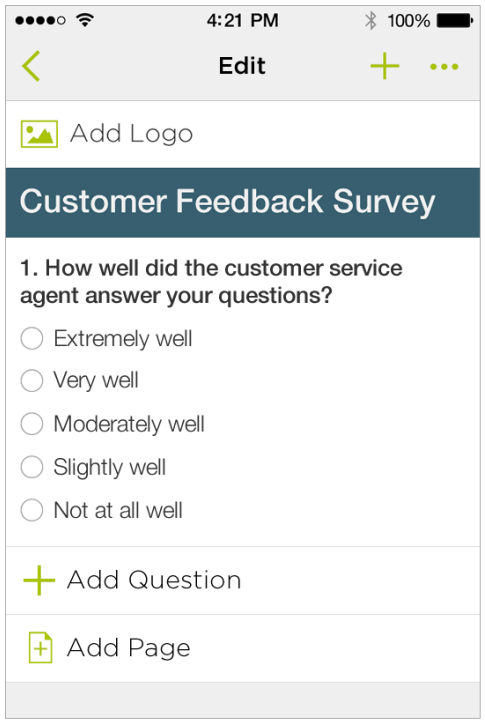
Others like UHRS are possible



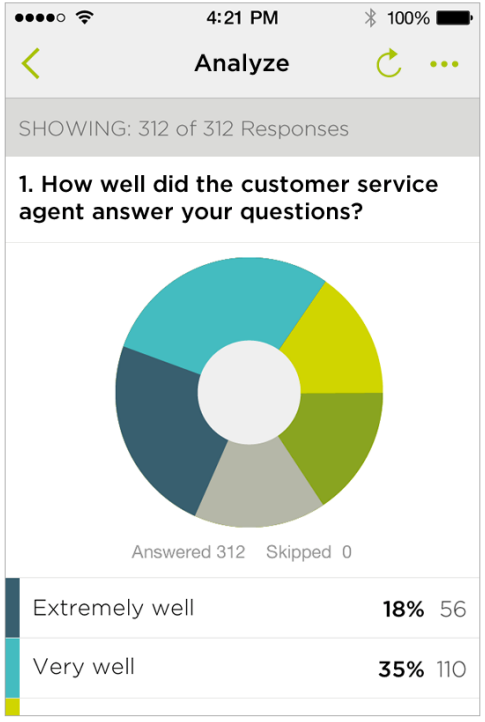
IS THIS NOT A SOLVED PROBLEM?

SurveyMonkey

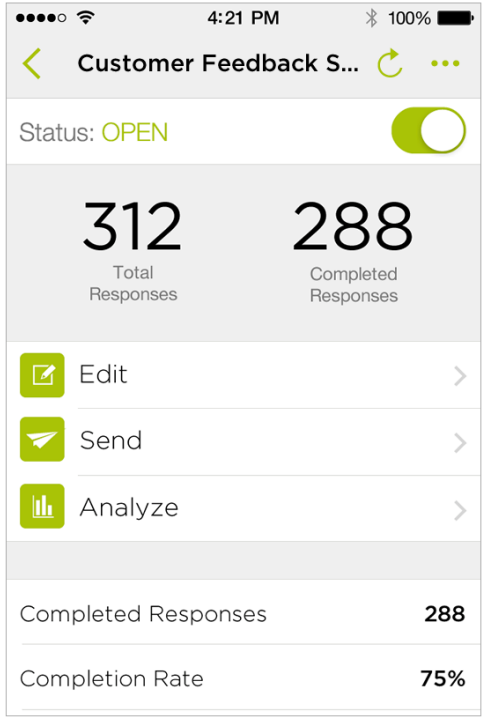
Create surveys anywhere, anytime.



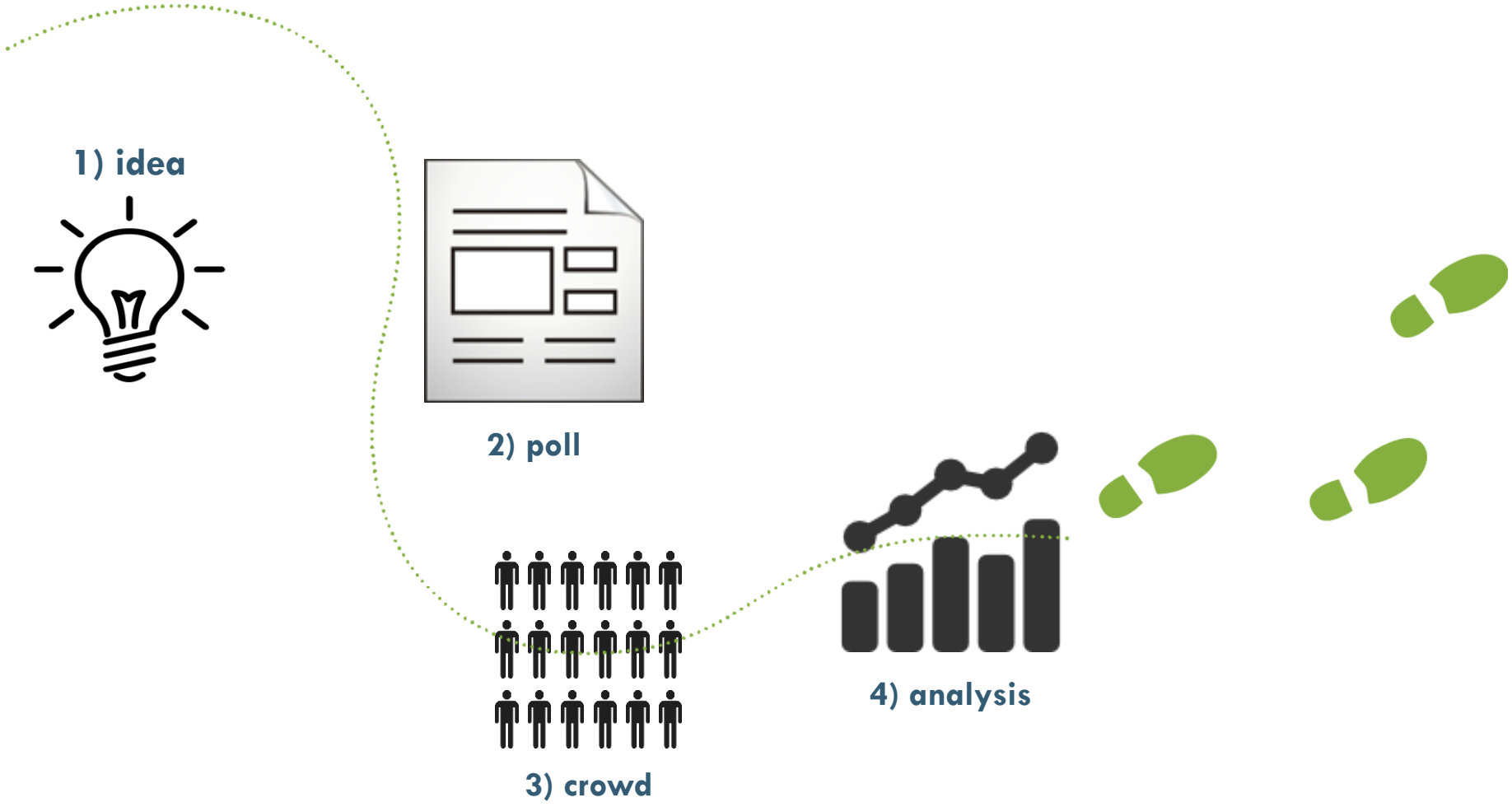
Analyze your survey results on the go.



Track your results in real-time.



OUR FOCUS IS ON AN END-TO-END PROCESS



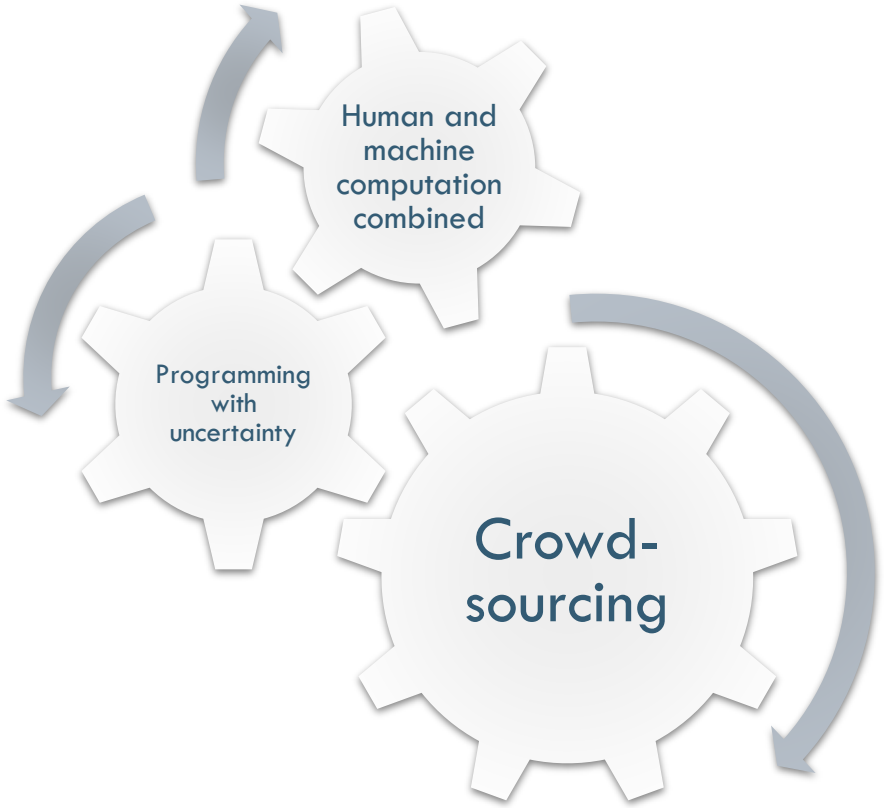
KEY DIFFERENCES BETWEEN THEM AND INTERPOLL

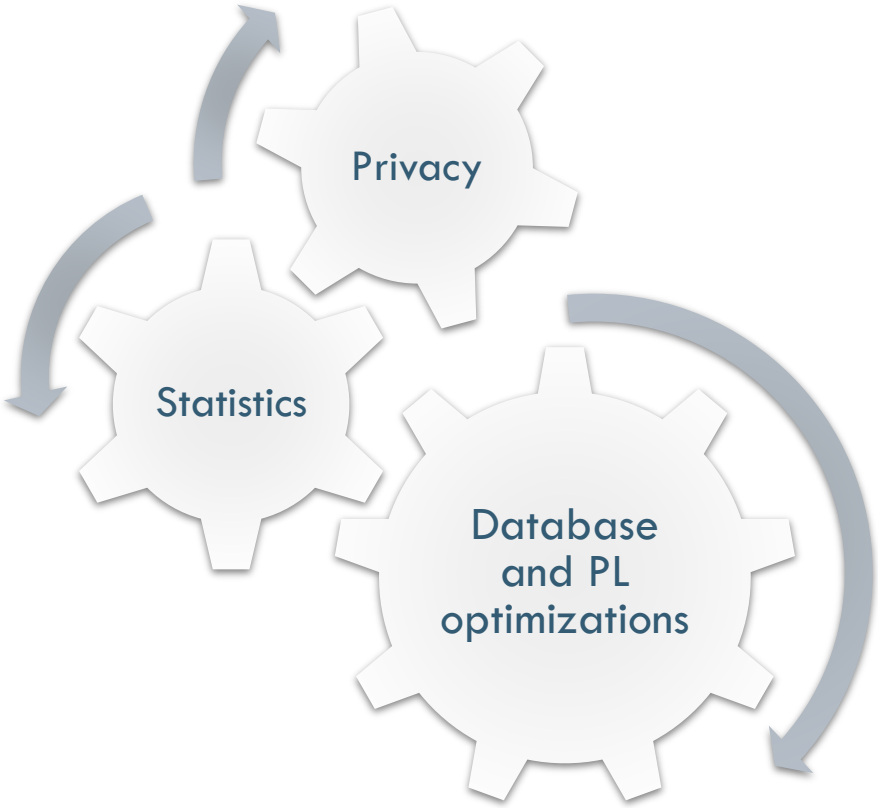
Lack of programmability

InterPoll is considerably cheaper

We give statistical significance to the results through unbiasinging and power analysis

A SMORGASBORD OF TOPICS





One of the wonderful things about InterPoll is that it brings together research fields that are not frequently combined.

This requires researchers who are not afraid to step outside their comfort zone.

"Experimental psychology is
the study of the college sophomore"

Quinn McNemar, 1946

Traditional polling is a painstaking and time-consuming process.

Not everything can be assessed through computer-based polling, but many things can be.

Recent studies show that online polling is frequently better than RDD polls and are good at avoiding the interviewer bias.

Recent results suggest that online polling is actually cheaper, faster, and is devoid of some of the biases such as the interviewer bias present in traditional face-to-face or telephone surveys.

Moreover, in recent years, traditional random-dialing surveys suffer from the fact that young people often do not have a land line.

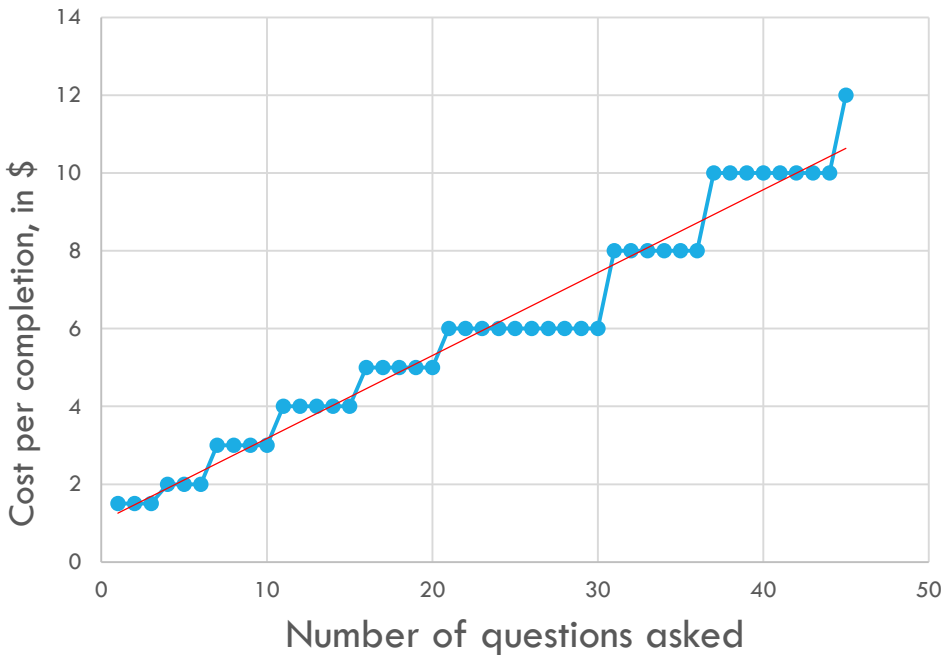
COSTLY… ESPECIALLY AT SCALE

SurveyMonkey Audience pricing

A technology accessory company wants feedback on their latest iPhone case design. They have a 14-question survey and would like 200 responses from an audience of only female iPhone users, delivered on the 2-businsess day schedule.

14 question survey	$1.50 per response
2 specific targeting options added	$1.25 per response (Gender targeting & iPhone ownership targeting)
2-business day turnaround	$1.00 per response
Project cost for 200 responses	**$750.00 ($3.75 per response)**

Instant.ly cost per completed survey



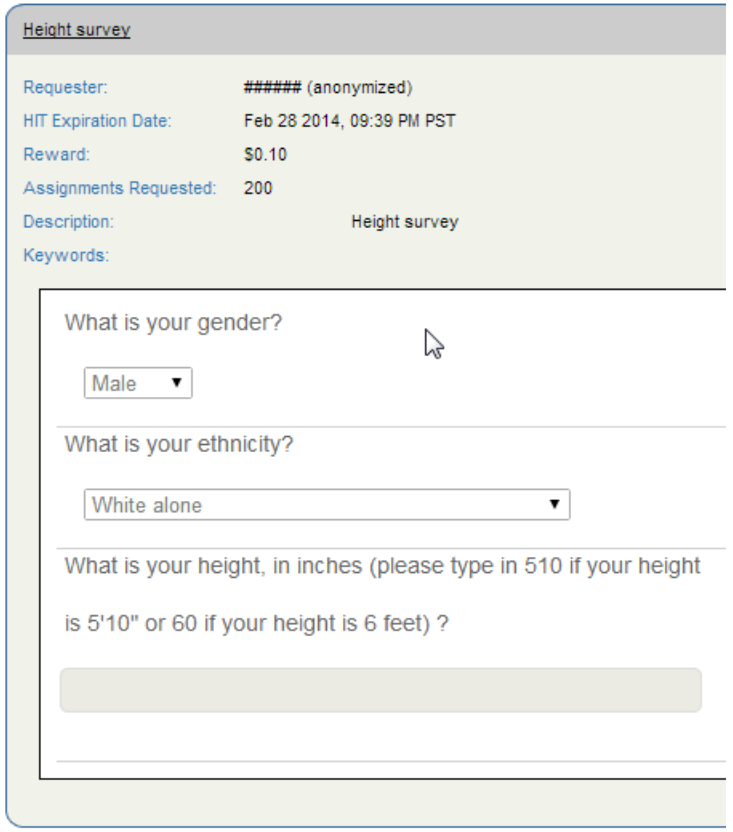
MECHANICAL TURK (MTURK)

Our main backend is Mechanical Turk

Generally can get answers to most polls for $.10/survey completion

Latency drops as a result of higher rewards, but the quality generally doesn't seem to increase

InterPoll takes care of creating appropriate HITs, generating XML forms for the UI that appears on Mechanical Turk and collecting the results

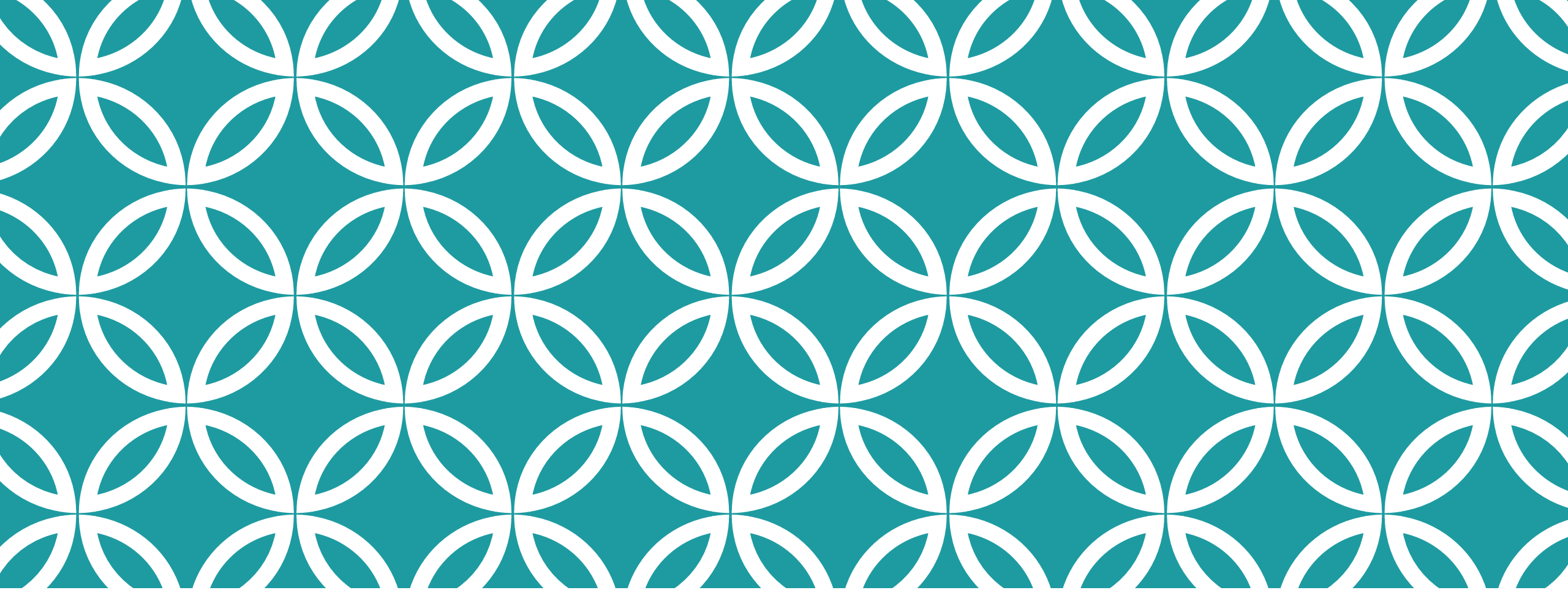


How to integrate human and computer computation is a wide open problem in programming languages.

Our approach is to use LINQ to provide seamless language extensions.

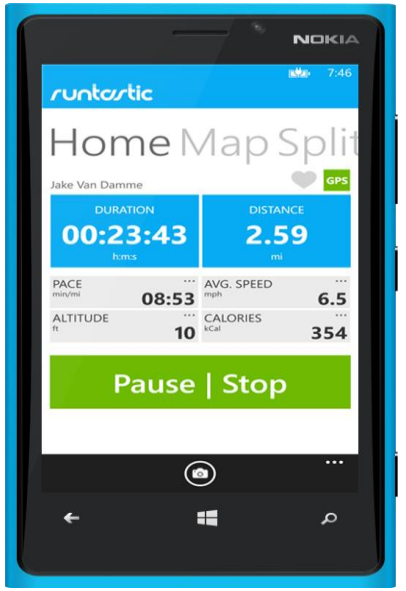
But many topics remain unaddressed. How does one program with long-running crowd operations?

How does one combine lazy and eager computation?



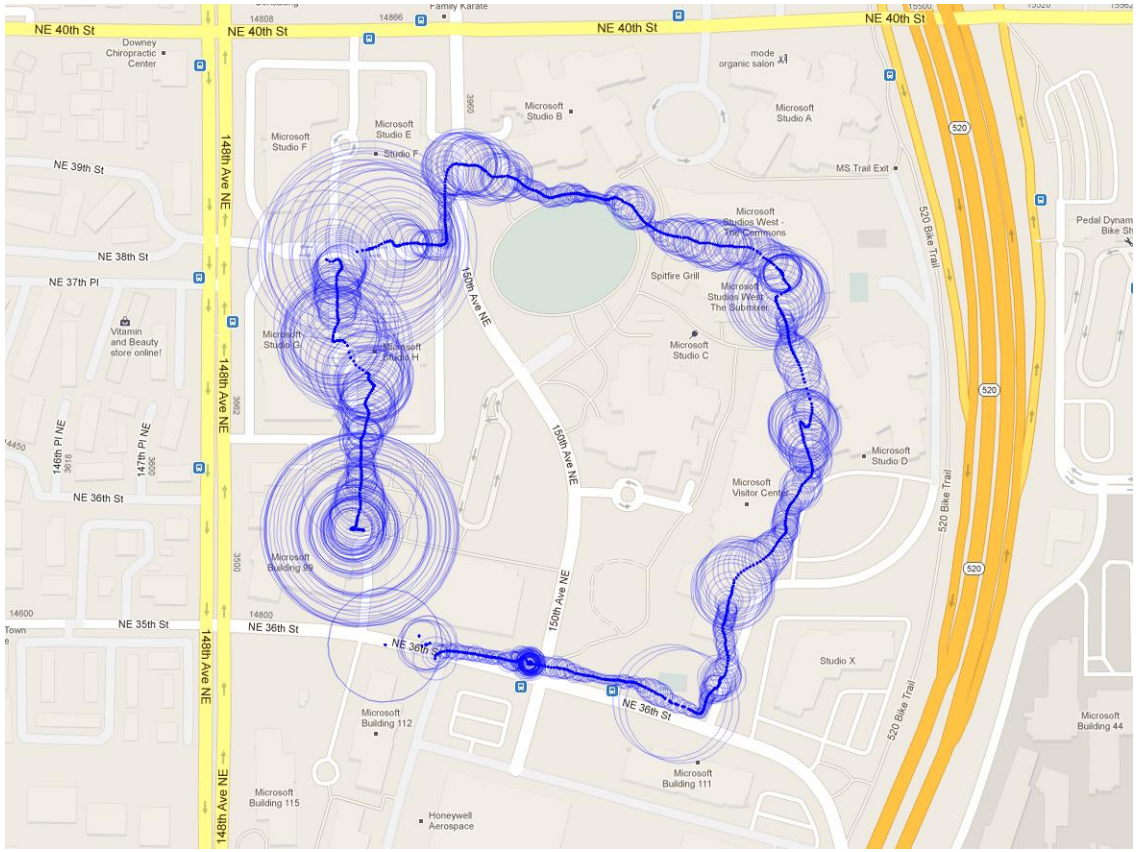
Uncertain<T>

24 mph


NOKIA
runtastic
7:46
Home Map Split
Jake Van Damme
GPS
DURATION
00:23:43
h:m:s
DISTANCE
2.59
mi
PACE
min/mi
08:53
AVG. SPEED
mph
6.5
ALTITUDE
ft
10
CALORIES
kCal
354
Pause | Stop

59 mph

A WALK WITH A GPS



HOW DO YOU WRITE THIS KIND OF CODE?

```
GeoCoordinate Location;
```

HOW DO YOU WRITE THIS KIND OF CODE?

~~GeoCoordinate Location;~~

```
Uncertain<GeoCoordinate> Location;
```

Uncertain<T> is an uncertain type abstraction.

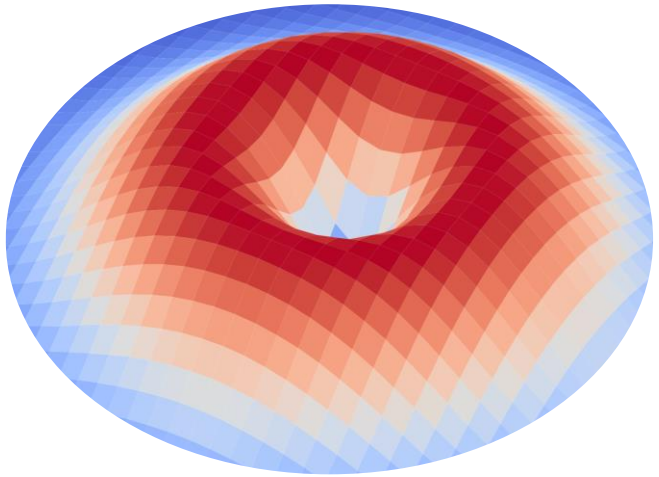
It encourages developers to
explicitly reason about uncertainty.

SYNTAX AND SEMANTICS

```
Uncertain<GeoCoordinate> LastLoc = GPS.GetLocation();
```

A variable of type Uncertain<T> is a random variable, represented by a distribution.

Gaussian distribution

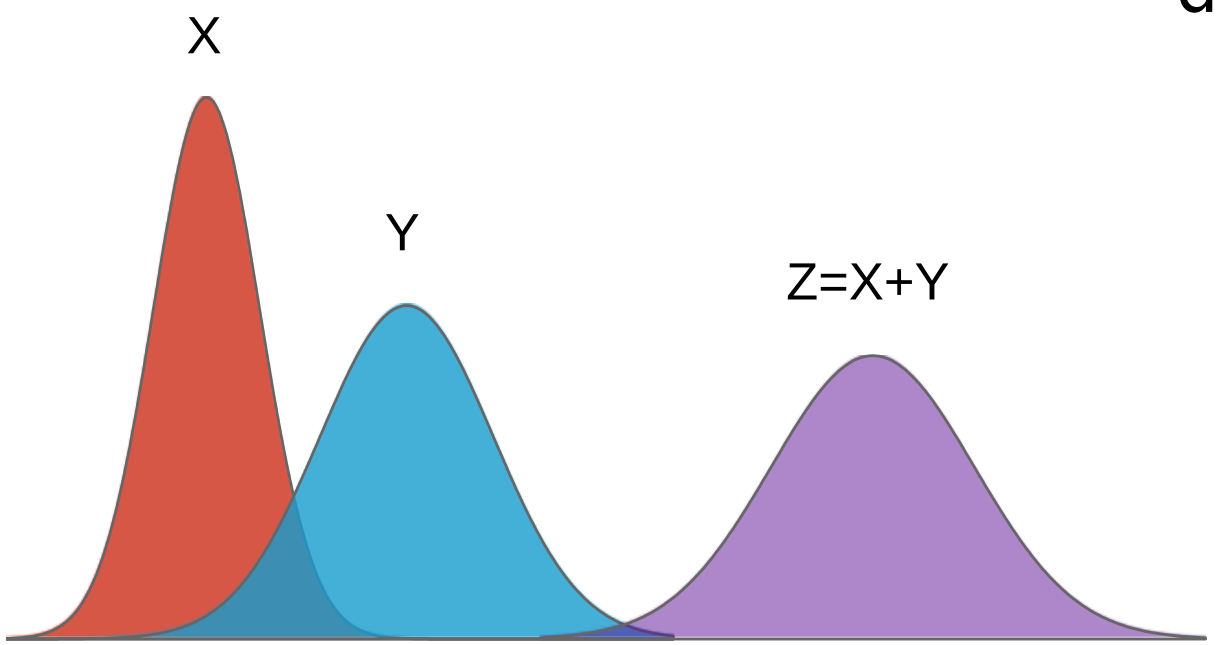


3 D Raleigh Distribution

Arbitrary non-continuous functions

PROGRAMMING WITH Uncertain<T>

```
Uncertain<double> Speed = Dist / 5;
```

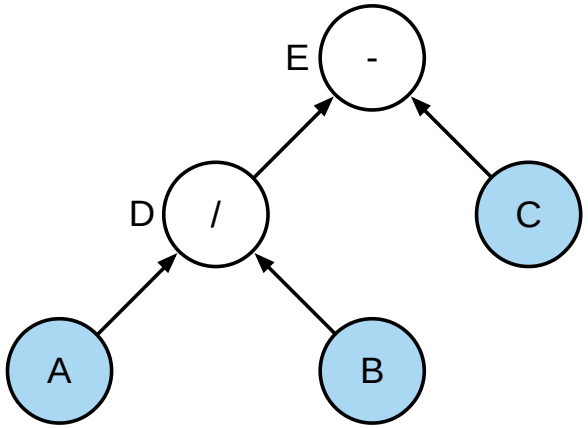


Or more generally, Z = X + Y, if X and Y are distributions.

If	*x*	is a sample of *X*
and	*y*	is a sample of *Y*
then	*x+y*	is a sample of *X+Y* *`

* if X and Y are independent

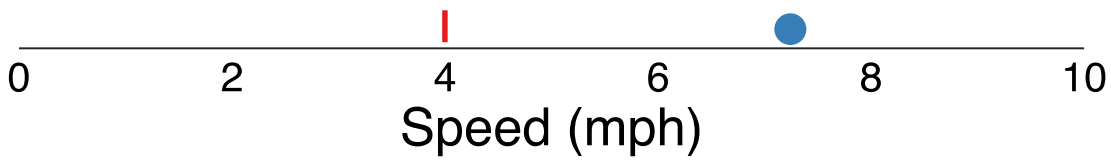
BAYESIAN NETWORK REPRESENTATION:



Sampling function for E recursively samples children.

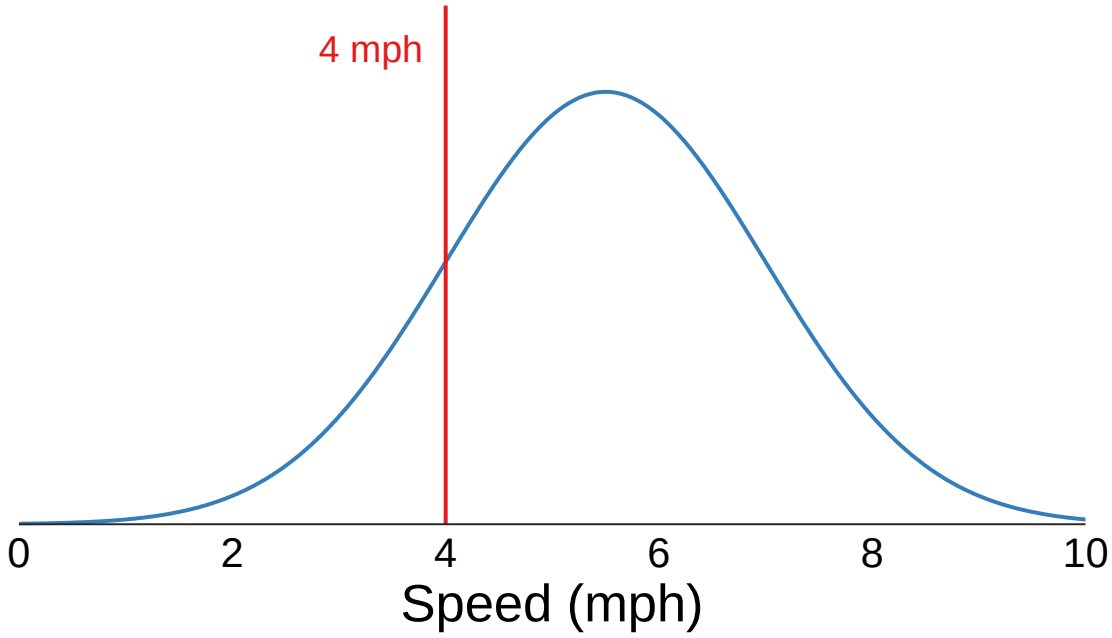
TESTING DISTRIBUTIONS

```
if (Speed > 4) print("Great job!");
```



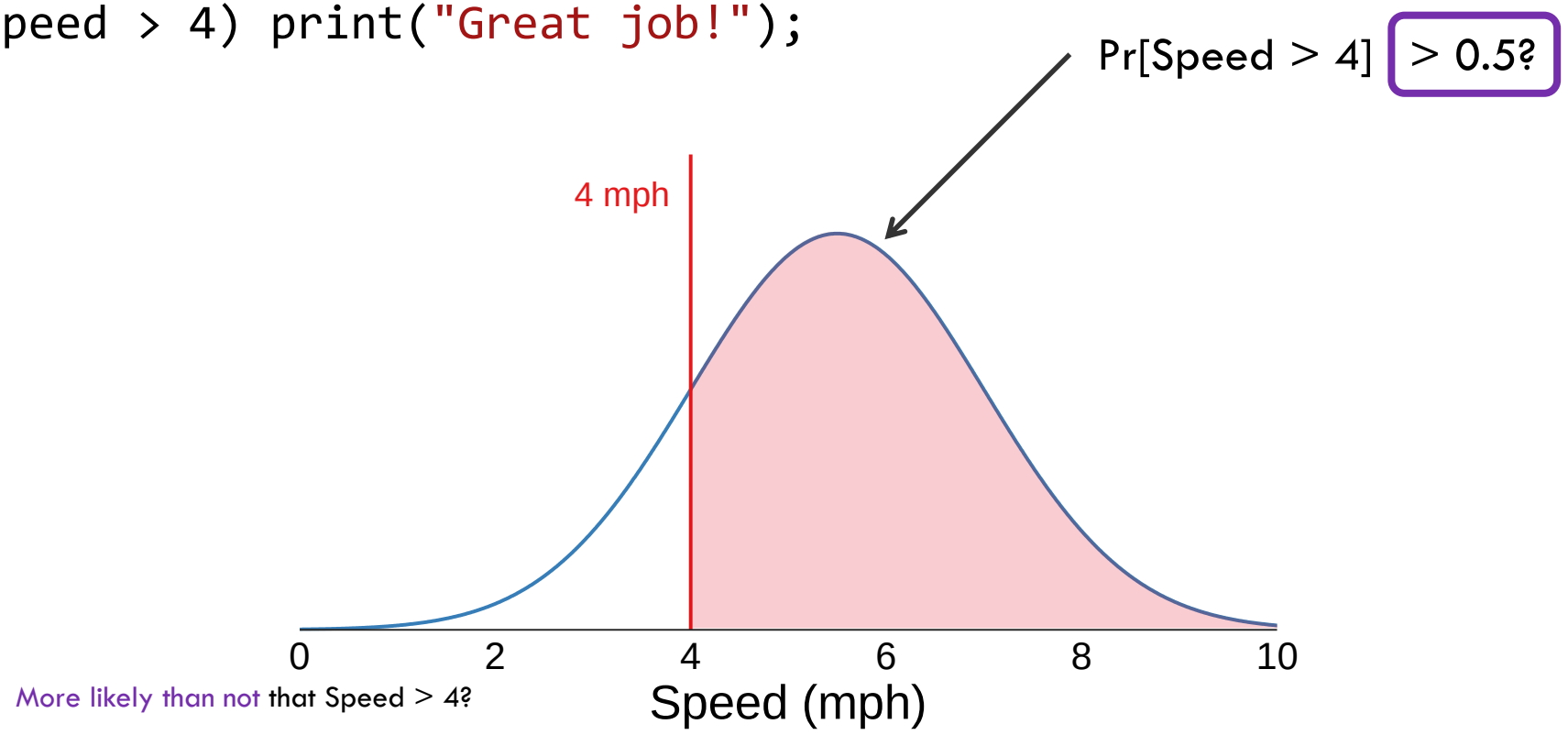
TESTING DISTRIBUTIONS

```
if (Speed > 4) print("Great job!");
```



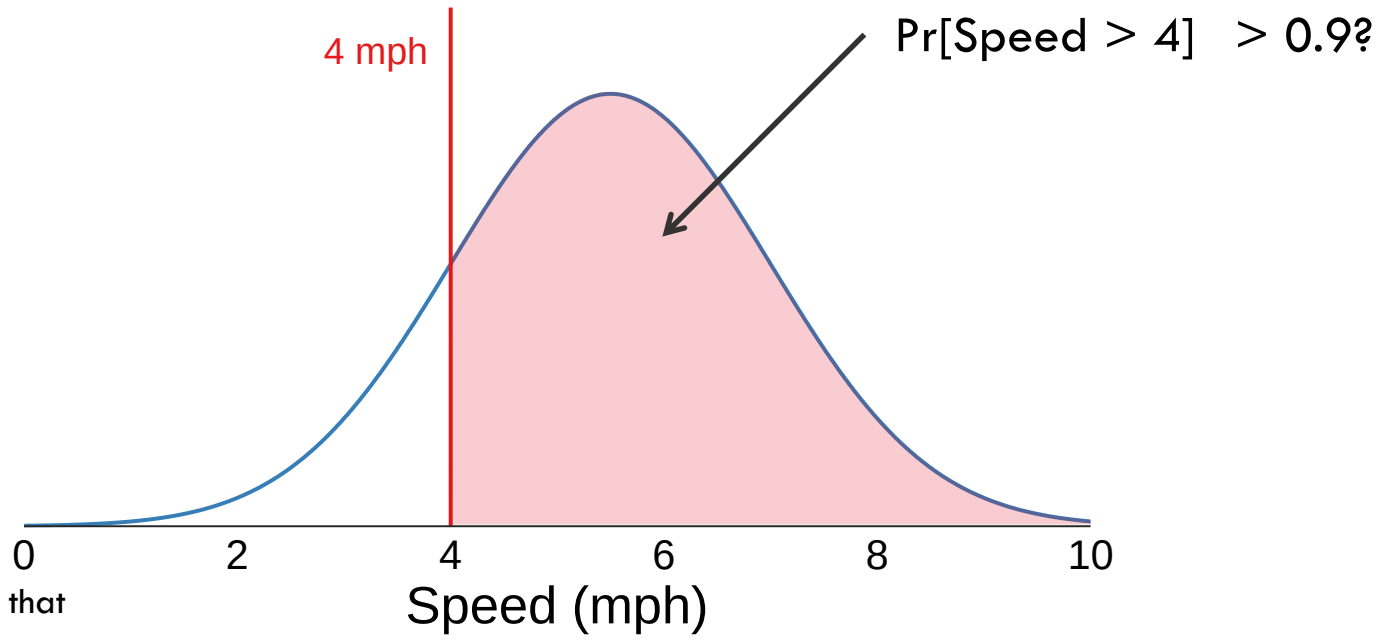
TESTING DISTRIBUTIONS

```
if (Speed > 4) print("Great job!");
```



TESTING DISTRIBUTIONS

```
if (Speed > 4) print("Great job!");
```



At least 90% likely that Speed > 4?

TESTING DISTRIBUTIONS

```
if (Speed > 4).Pr(0.9) print("Great job!");
```

null hypothesis H_0: $\Pr[\text{Speed} > 4] \leq 0.9$

alternate hypothesis H_A: $\Pr[\text{Speed} > 4] > 0.9$

approximate!

How many samples?

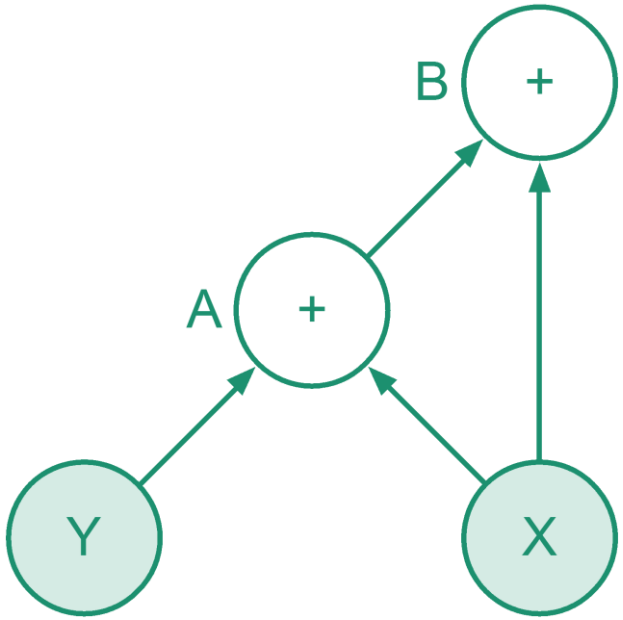
Too many = too slow

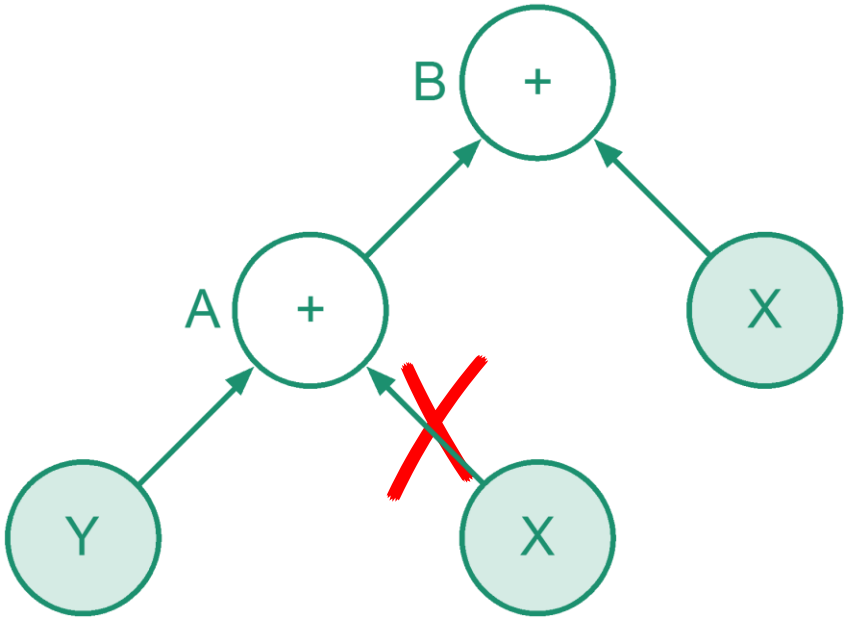
Too few = too noisy

Sequential sampling: sample size depends on progress

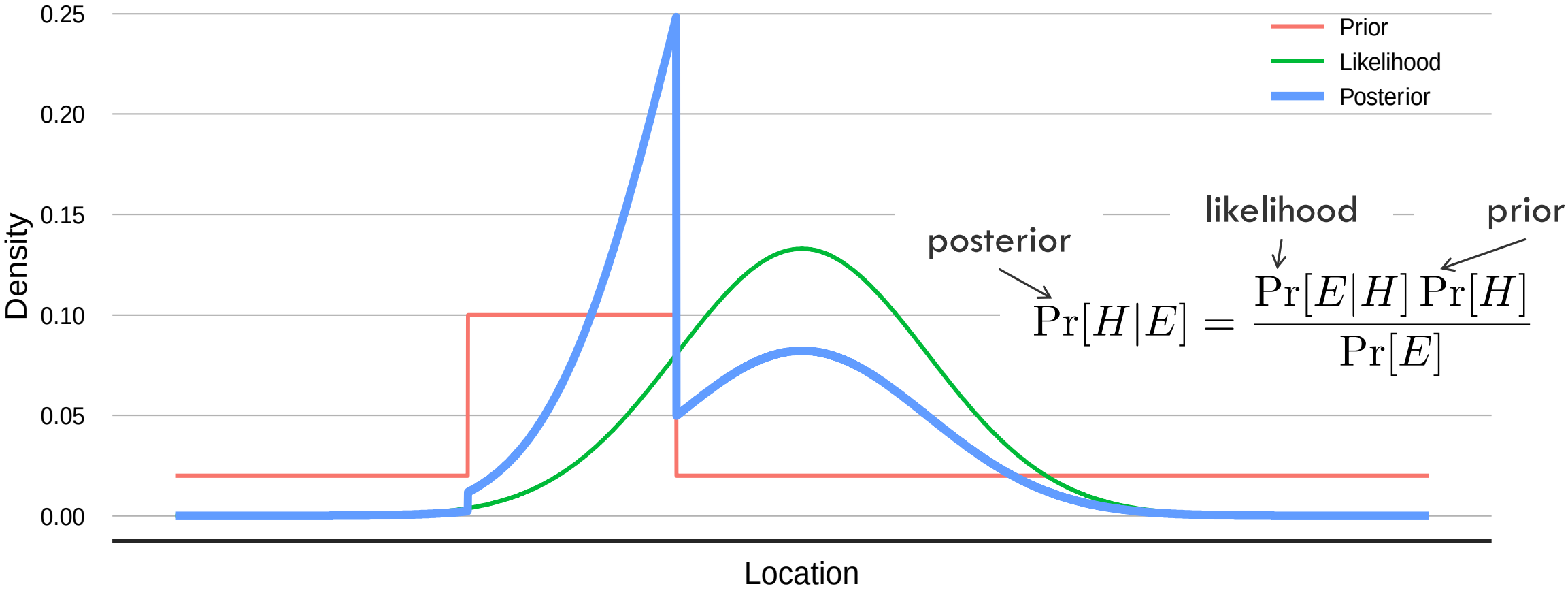
BAYESIAN NETWORK ESTIMATION

A and B depend on X – not independent – take one X sample not two!



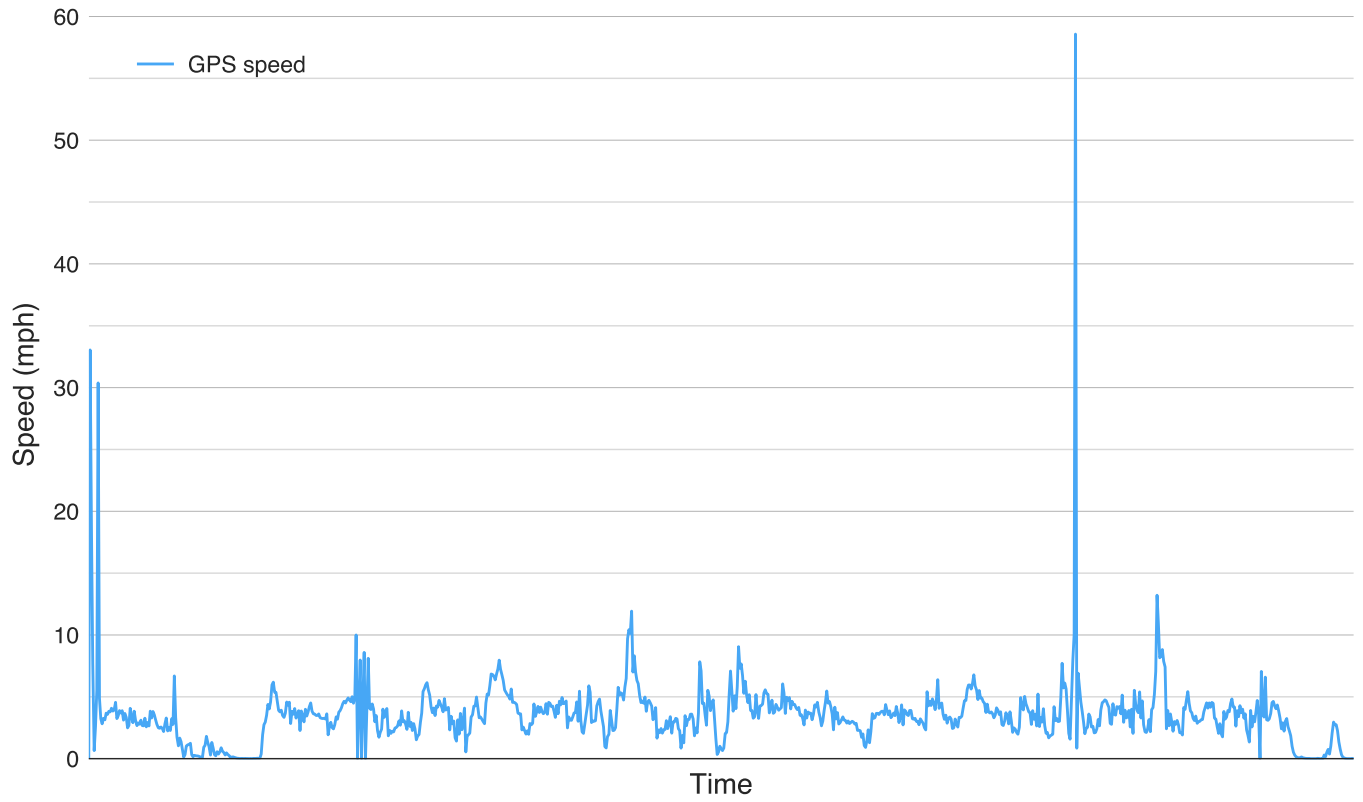


INCORPORATING DOMAIN KNOWLEDGE

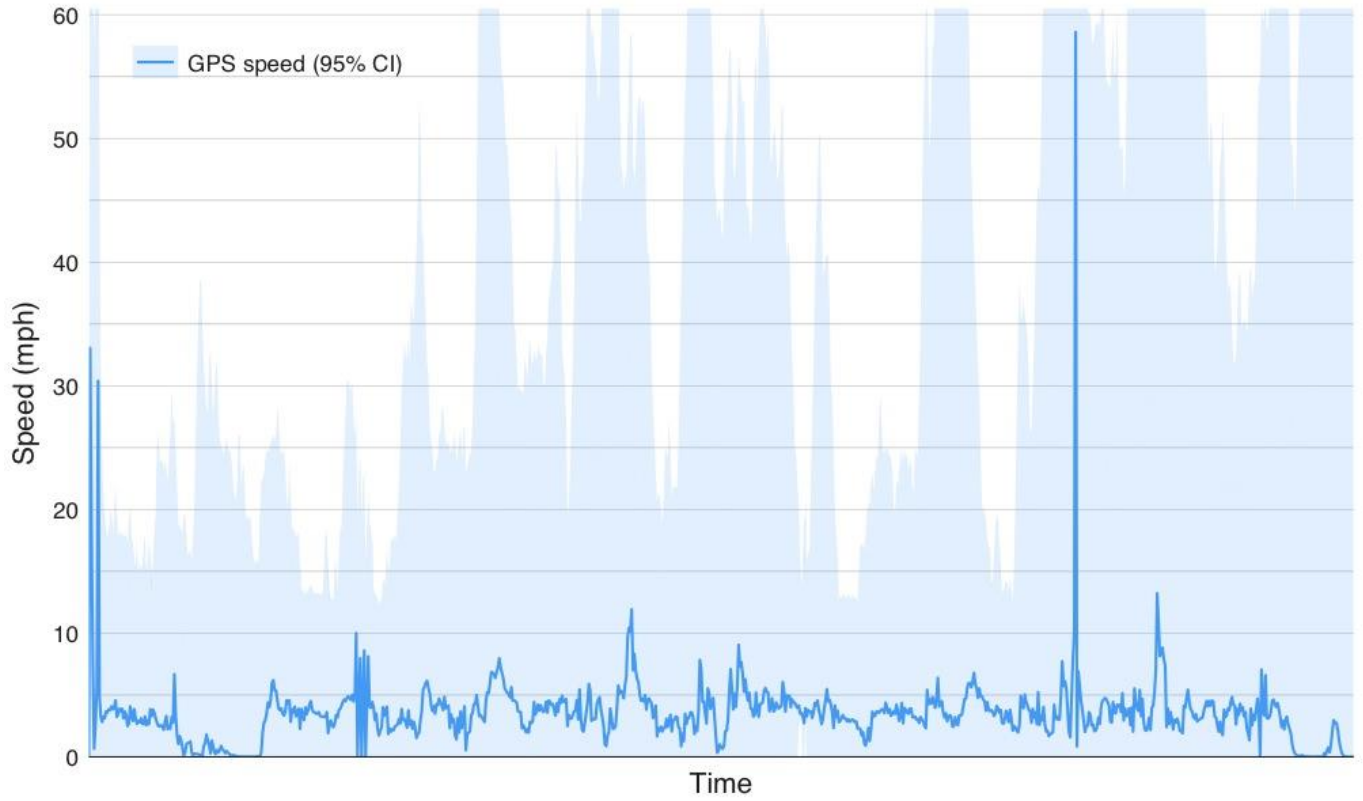


Uncertain<T> language semantics implements Bayes Rule

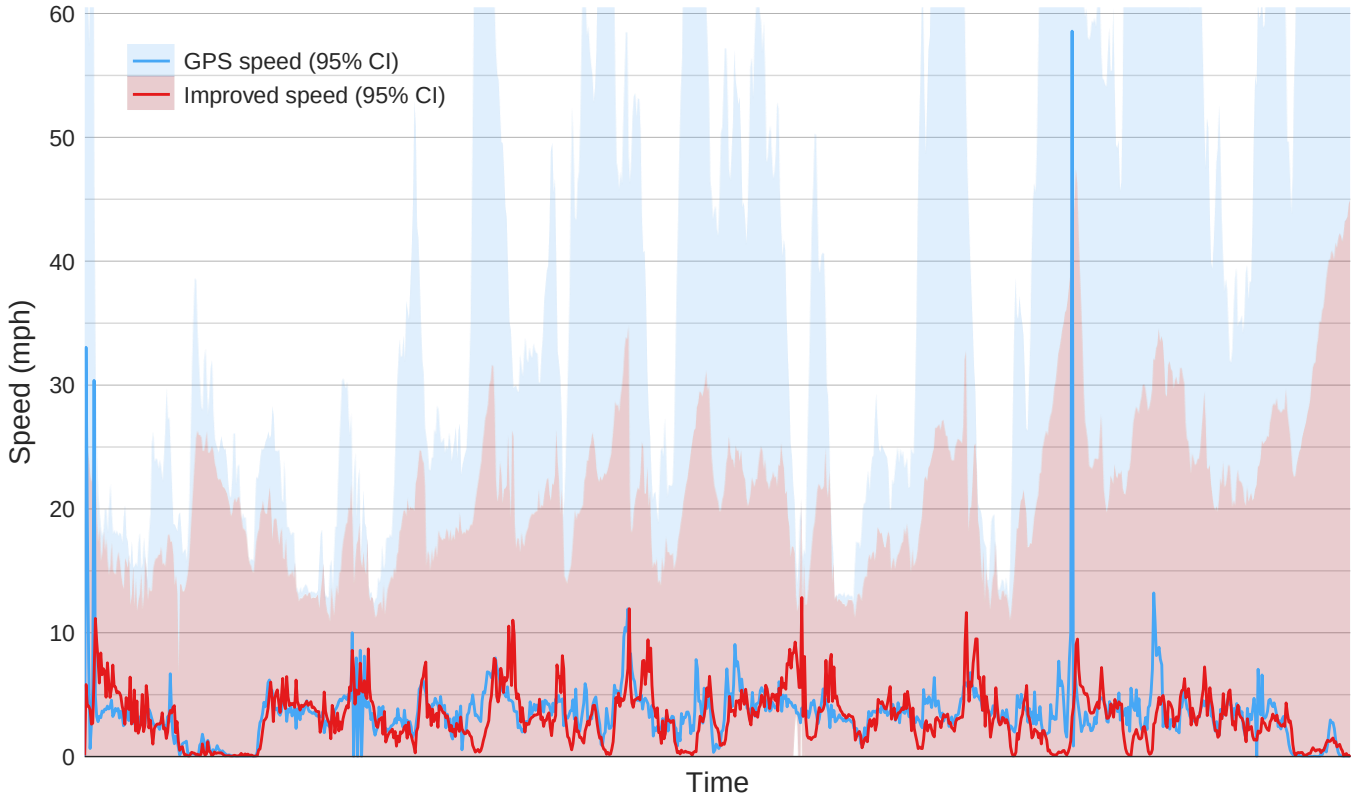
GPS SPEED



GPS SPEED



GPS SPEED

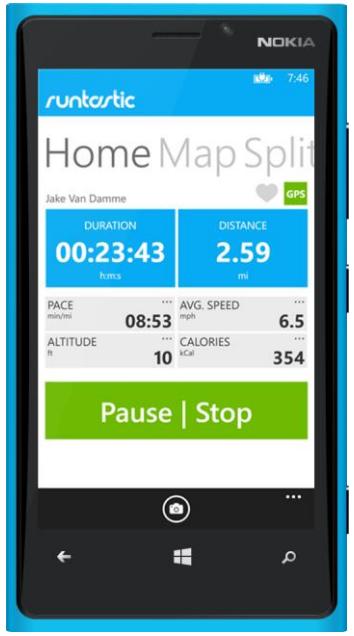

GPS speed (95% CI)
Improved speed (95% CI)
60
50
40
30
20
10
0
Speed (mph)
Time

Uncertain<T> is an uncertain type abstraction.

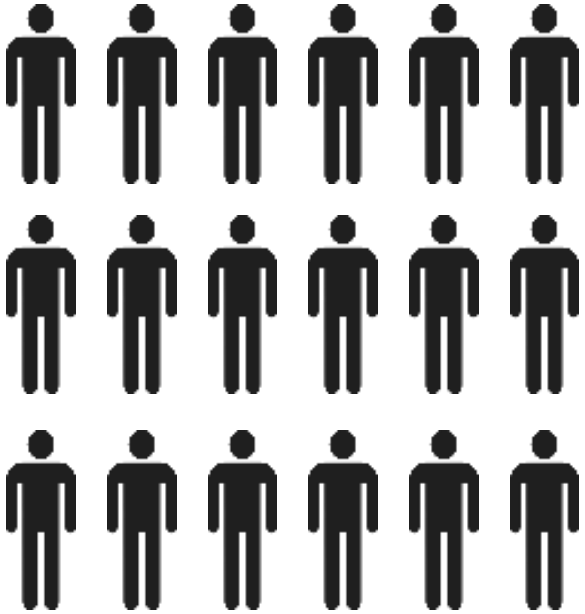
It encourages developers to
explicitly reason about uncertainty.

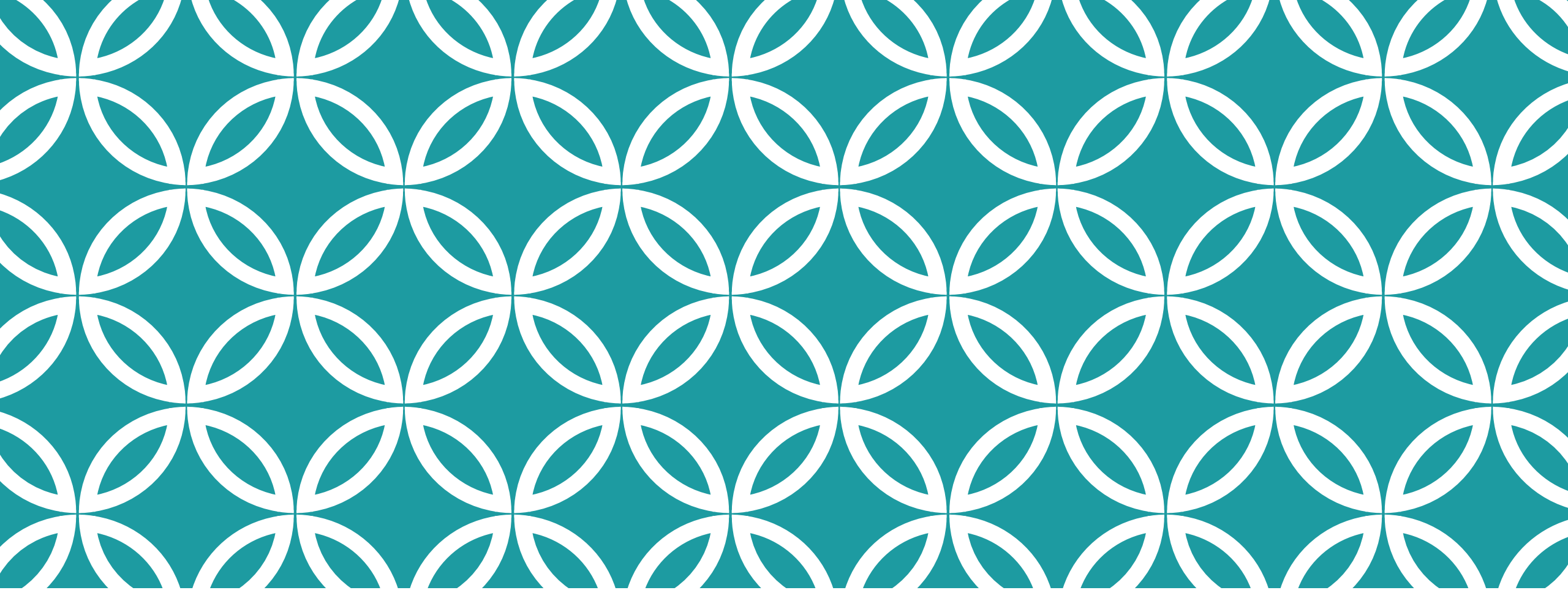
BRINGING IT ALL TOGETHER

sensors



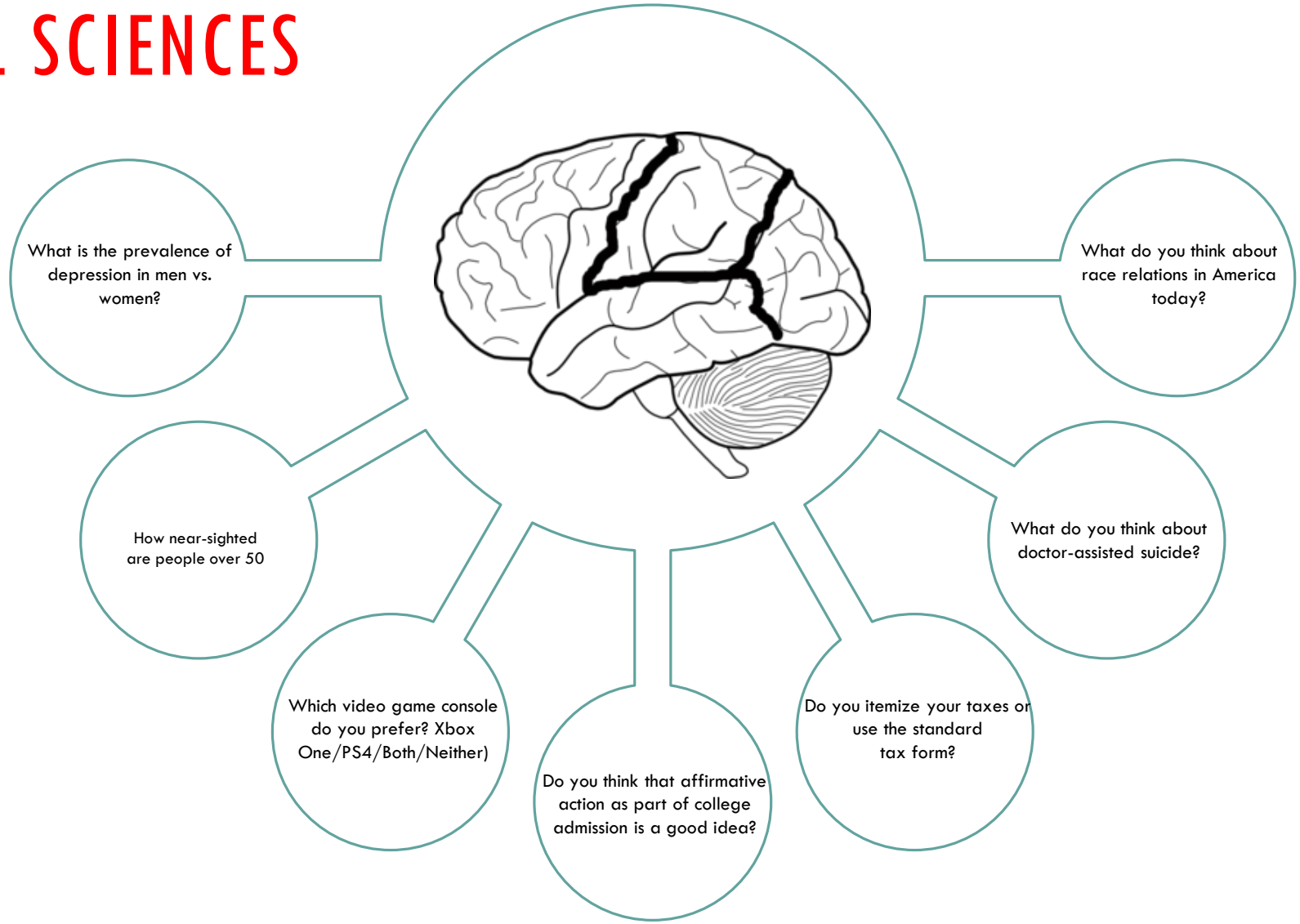
people



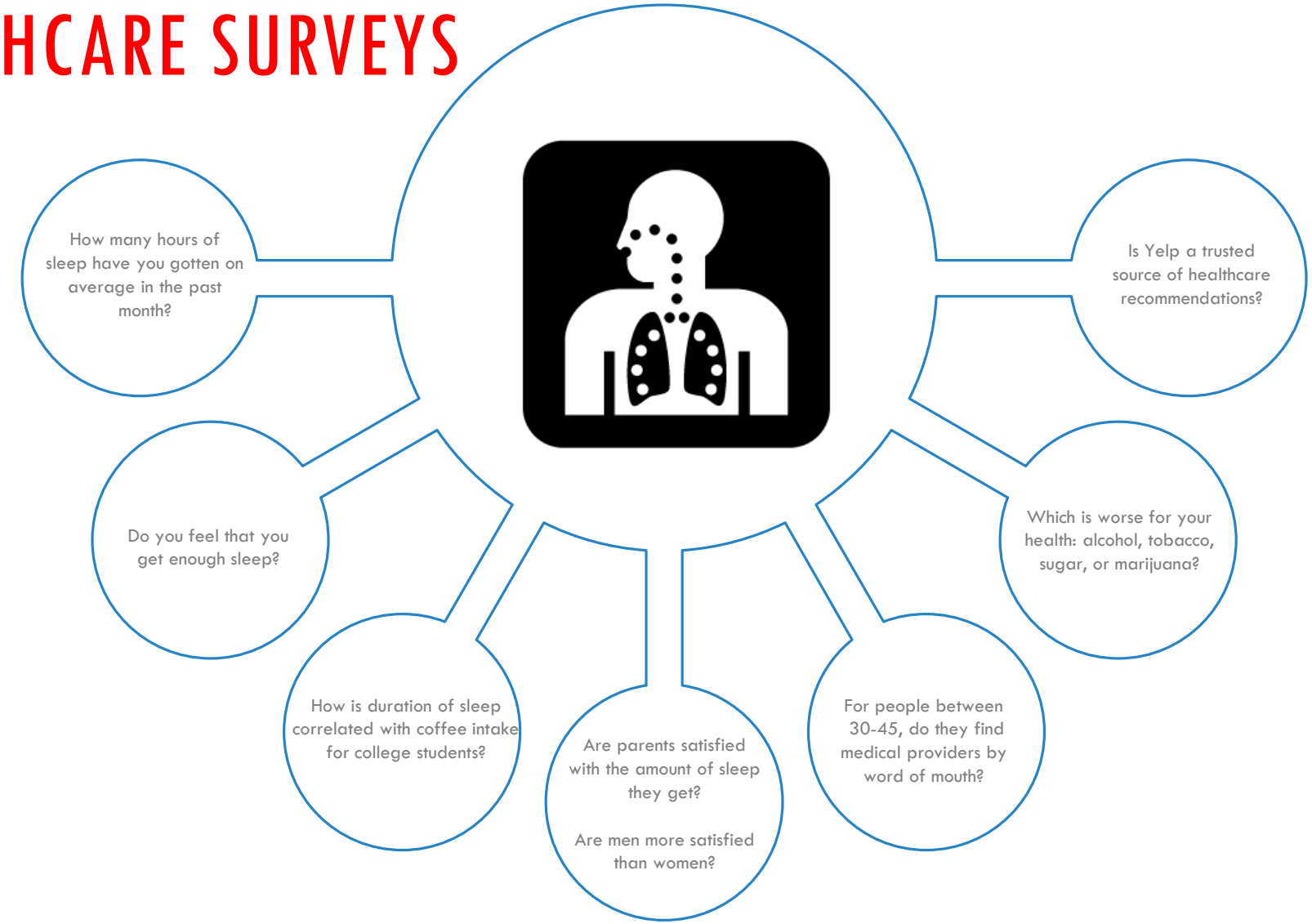


BUILDING INTERPOLL APPLICATIONS

SOCIAL SCIENCES



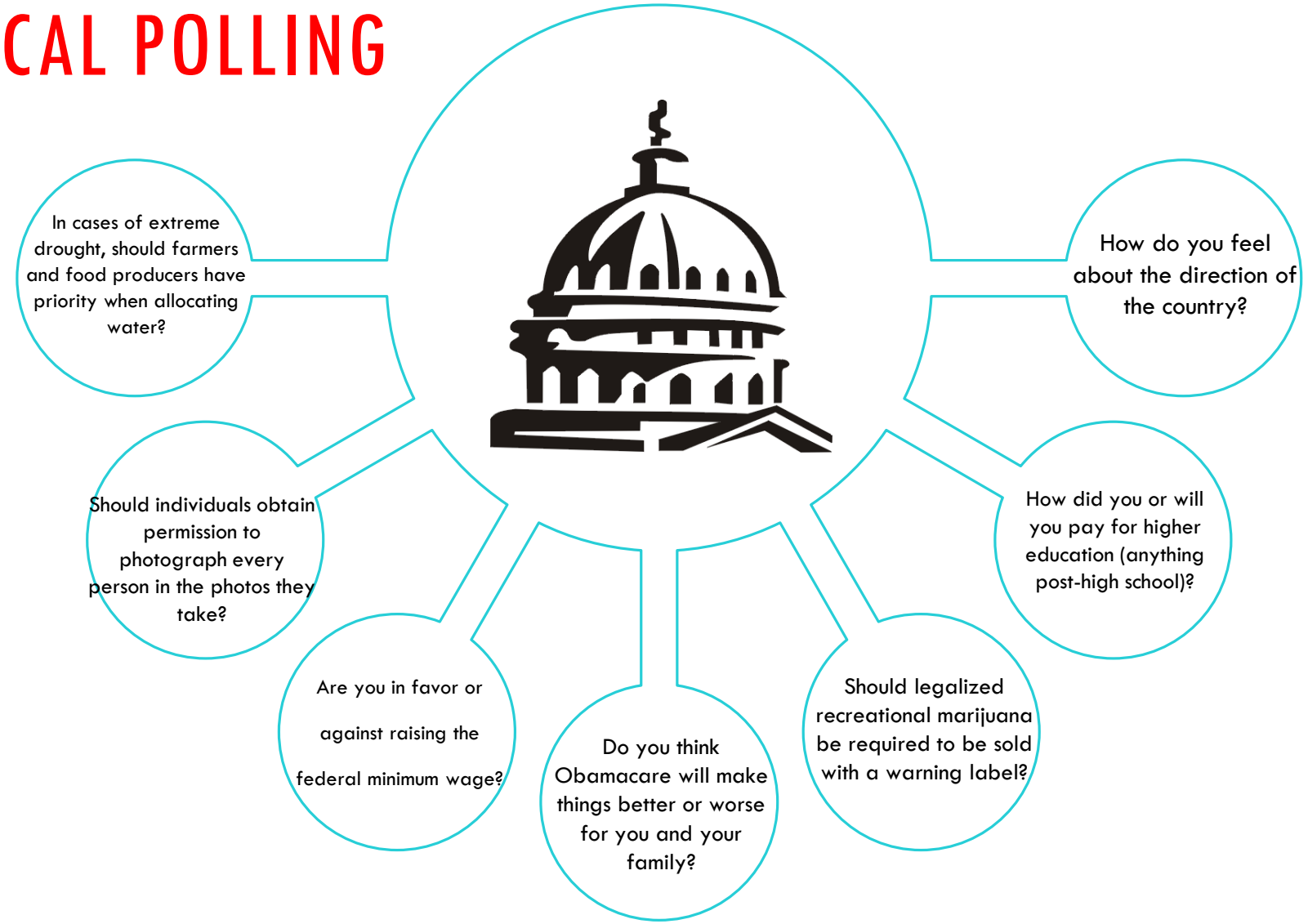
HEALTHCARE SURVEYS



MARKETING

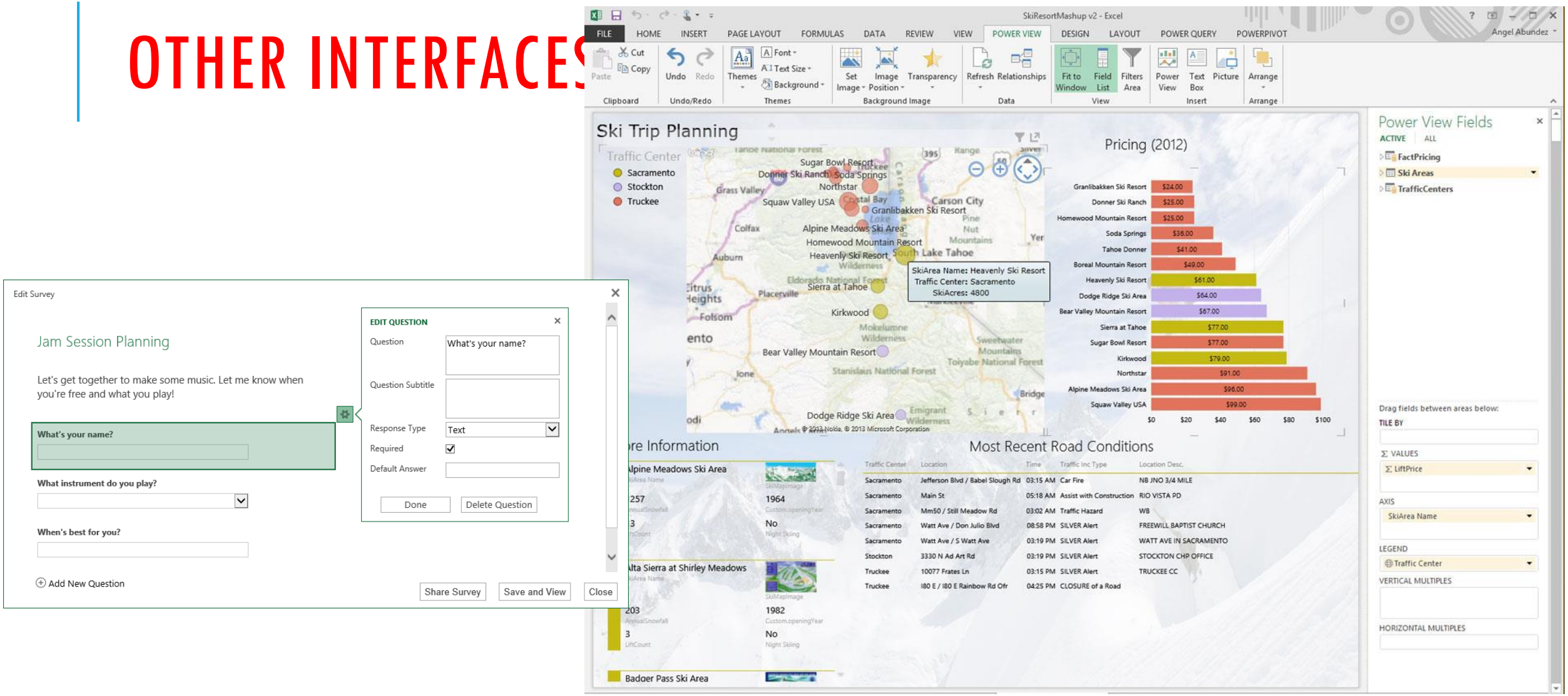

Are men or women responsible for shopping for pet food?
What percentage of the population between 30 and 50 use both cable and streaming services like Netflix?
Do people between 20-30 play more video games than those between 30-45?
Do you like the new Microsoft company logo better than the old Microsoft company logo?
Which video game console do you prefer?
What kind of jeans do you wear? (Skinny/straight/loose, etc.)
Have you paid for a cable TV subscription in the last year and, if so, are you likely to have a Hulu subscription?

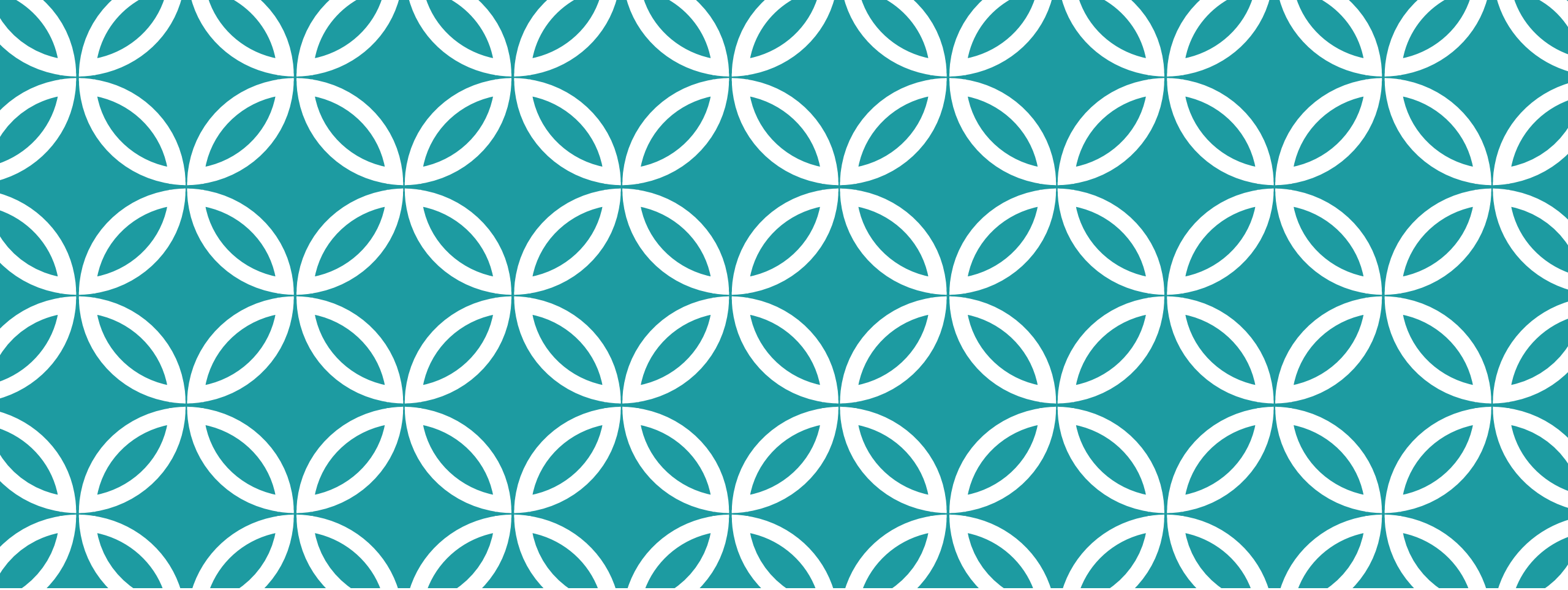
POLITICAL POLLING



Demo: basic surveys

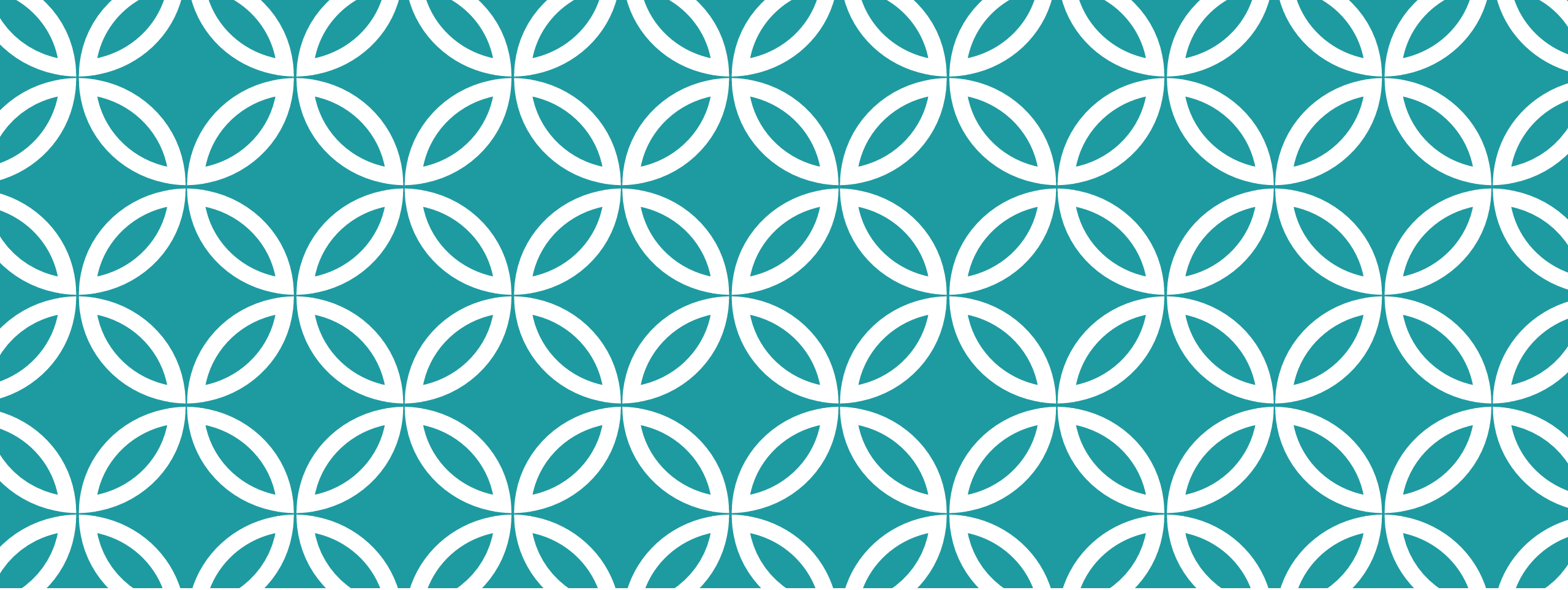
OTHER INTERFACES





FUNDAMENTAL PROBLEMS

MOVING AWAY FROM SMALL AND UNREPRESENTATIVE SAMPLES



POWER ANALYSIS

POWER ANALYSIS

Determine the number of samples for a query

We can sample from the crowd sequentially until we satisfy or disprove our hypothesis.

We will poll the crowd for more until our stopping criterion is reached.

The sopping criterion allows us to conclude that he hypothesis can be proven or disproven with the required level of confidence.

EXAMPLE QUESTION: HEIGHT

```
var people = GetPeople(GetDescription(), 200, false, false);
var height = from person in people
             select new
             {
                 Height = person.PoseQuestion<int>(
                  "What is your height, in centimeters " +
                  "(if you know your height in inches, you can conver
                  "centemeters using a calculator here: http://www.ca
                  "Please be careful when typing in your height. Inva
                  ),
                 Gender = person.Gender,
                 Ethnicity = person.Ethnicity,
             };

var males = from person in height
            where person.Gender == Gender.MALE
            select person.Height;
var females = from person in height
              where person.Gender == Gender.FEMALE
              select person.Height;

if (males.ToRandomVariable(false) > females.ToRandomVariable(false))
{
    Console.WriteLine("Males are taller than females.");
}
```

N=**29**

Once we remove the outliers

height = from person in height where

where

person.Height >=140 &&

person.Height <=220)

N=**27**

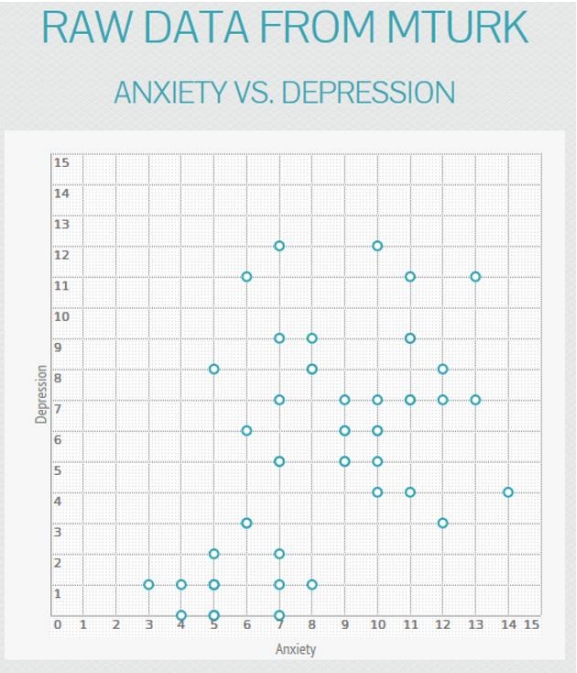
ARE THE TWO CORRELATED?

```
var happinessData = from person in people
                    select new
                    {
                        Consider = person.PoseCodedQuestion(
                        "I feel tense or 'wound up'", new Tuple<string,
                        Enjoy = person.PoseCodedQuestion(
                        "I still enjoy the things I used to enjoy", new
                        Awful = person.PoseCodedQuestion(
                        "I get a sort of frightened feeling as if somet
                        Laugh = person.PoseCodedQuestion(
                        "I can laugh and see the funny side of things",
                        Worry = person.PoseCodedQuestion(
                        "Worrying thoughts go through my mind", new Tup
                        Cheerful = person.PoseCodedQuestion(
                        "I feel cheerful", new Tuple<string, int>("Not
                        Sit = person.PoseCodedQuestion(
                        "I can sit at ease and feel relaxed", new Tuple
                        Slowed = person.PoseCodedQuestion(
                        "I feel as if I am slowed down", new Tuple<stri
                        Frightened = person.PoseCodedQuestion(
                        "I get a sort of frightened feeling like 'butte
```

```
var scores = from data in happinessData
             select new
             {
                 Anxiety =
                     data.Consider + data.Awful + data.Worry +
                     data.Sit + data.Frightened + data.Restless + data.Panic,
                 Depression = data.Enjoy + data.Laugh + data.Cheerful + data.Slowed +
                 data.LostInterest + data.LookForward + data.GoodBook,
                 Gender = data.Gender,
                 Income = data.Income,
                 Education = data.Education,
                 Ethnicity = data.Ethnicity,
                 Employment = data.Employment,
             };
```

Hospital anxiety and depression scale (HADS)

- Normal 0-7
- Mild 8-10
- Moderate (11–14)
- Severe (15–21)



DOES MONEY BUY HAPPINESS (OR AT LEST TRANQUILITY)?

```
var rich = from person in scores
           where
               person.Income == Income.INCOME_35_000_TO_49_999 ||
               person.Income == Income.INCOME_75_000_AND_OVER ||
               person.Income == Income.INCOME_50_000_TO_74_999
           select person.Anxiety;

var poor = from person in scores
           where
               person.Income == Income.INCOME_1_TO_4_900 ||
               person.Income == Income.INCOME_10_000_TO_14_999 ||
               person.Income == Income.INCOME_15_000_TO_24_999
           select person.Anxiety;
```

```
if (rich.ToRandomVariable(false) < poor.ToRandomVariable(false))//(p >
{
    Console.WriteLine("Rich are more anxious than poor: " + "\tYes");
}
else
{
    Console.WriteLine("Rich are more anxious than poor: " + "\tNo");
}
```

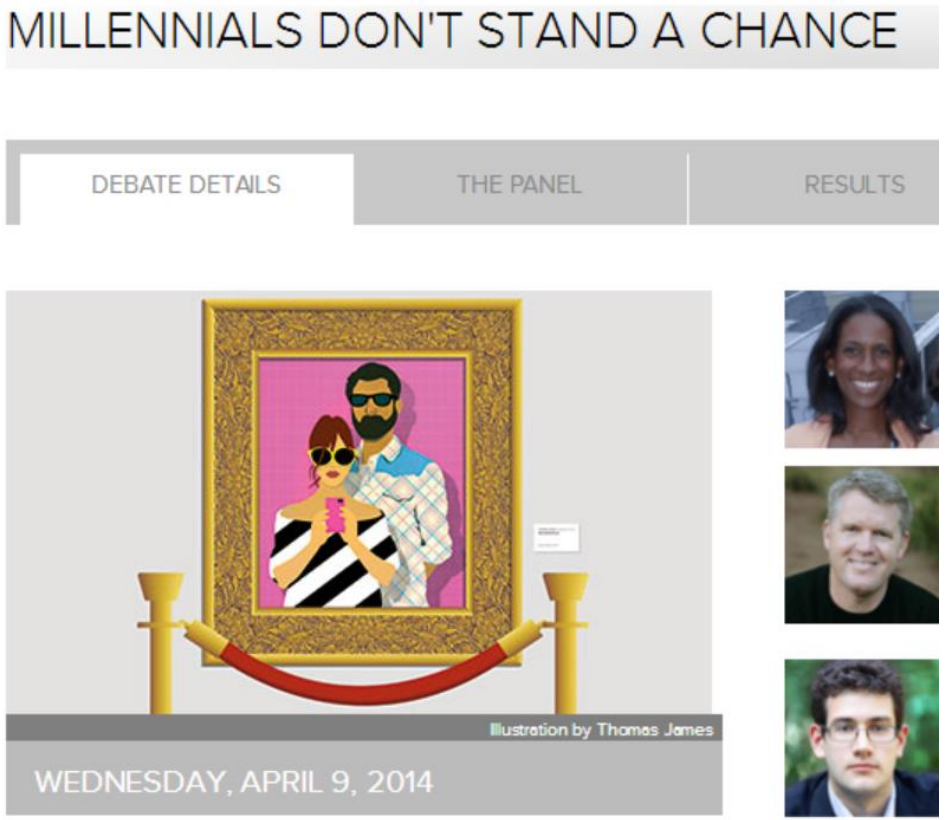
Are rich more anxious than poor?

N = 105

expected value for poor=8.5714,

expected value for rich=7.9619

DEBATES: INTELLIGENCE SQUARED



Task	Outcome	Power	Cost
MilennialsDontStandAChance	No	37	$3.70
MinimumWage	No	43	$4.30
RichAreTaxedEnough	No	51	$5.10
EndOfLife	No	53	$5.30
BreakUpTheBigBanks	Yes	73	$7.30
StrongDollar	No	85	$8.50
MarginalPower	No	89	$8.90
GeneticallyEngineeredBabies	Yes	135	$13.50
AffirmativeActionOnCampus	Yes	243	$24.30
ObesityIsGovernmentBusiness	No	265	$26.50

CONVERGENCE CURVES: SEQUENTIAL PROBABILITY RATIO TEST (SPRT) OR WALD, 1945

Sequential probability ratio test: To implement this, we build a sequential acceptance plan. Let $H_0 : p + \epsilon$ and $H_A : p - \epsilon$ where $p = 0.5$ by default and can be overloaded by a programmer. `Uncertain⟨T⟩` calculates the cumulative log-likelihood ratio for each sample:

$$\Delta_L = k \log(H_A/H_0) + (n - k) \log(H_0/H_A)$$

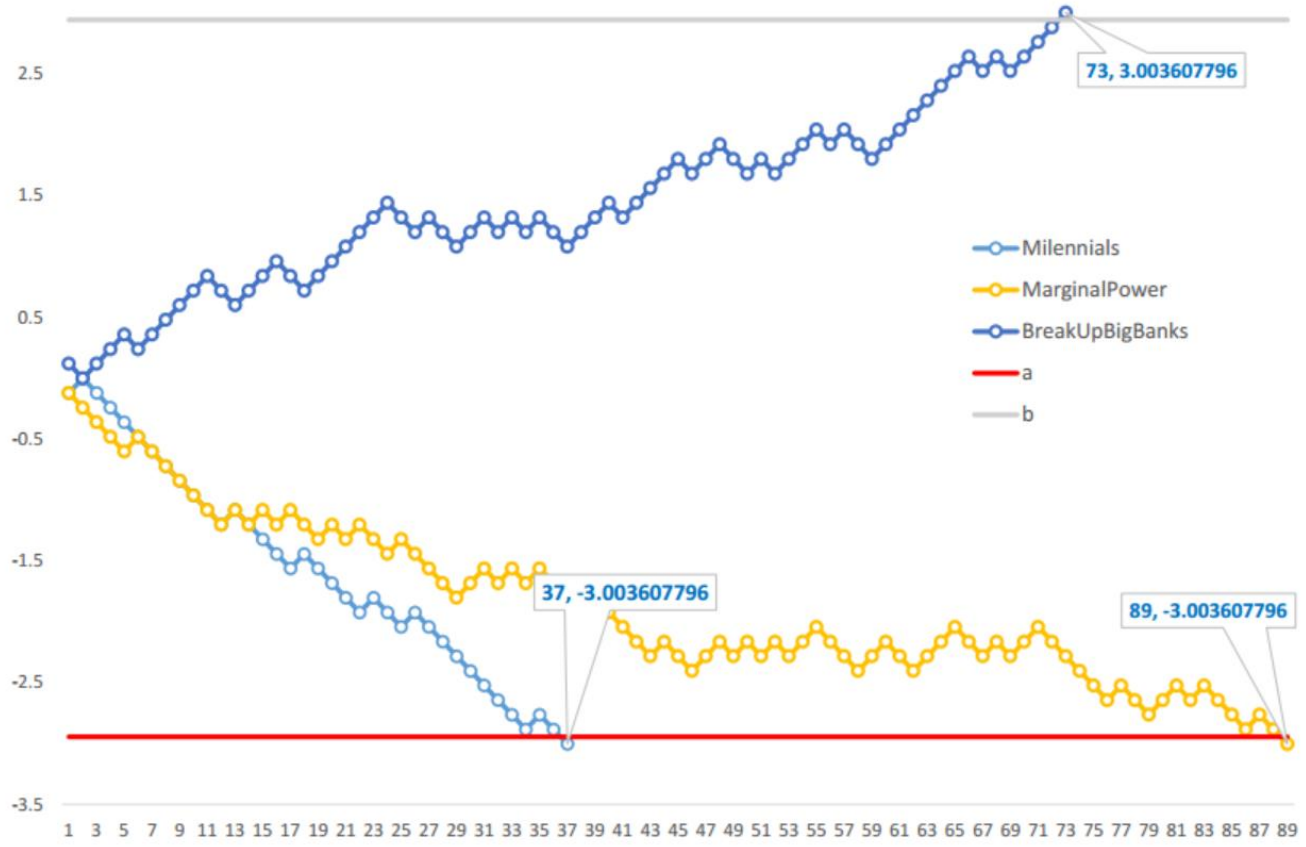
where n is the number of samples taken thus far and k is the number of successes out of those n trials. If

$$\Delta_L \leq \log(alpha/(1 - alpha)) \quad = a$$

then `Uncertain⟨T⟩` evaluates the conditional as false while if

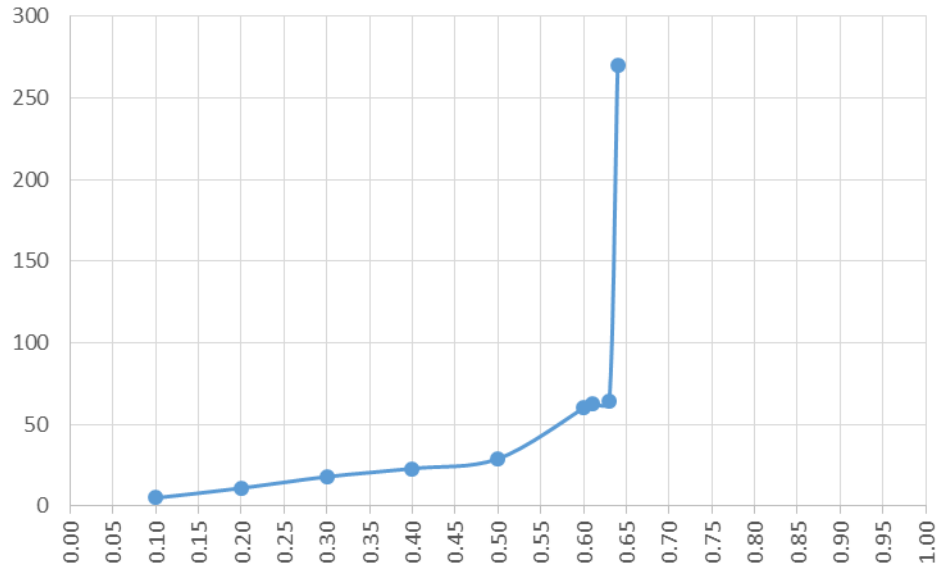
$$\Delta_L \geq \log((1 - alpha)/alpha) \quad = b$$

the conditional is true.

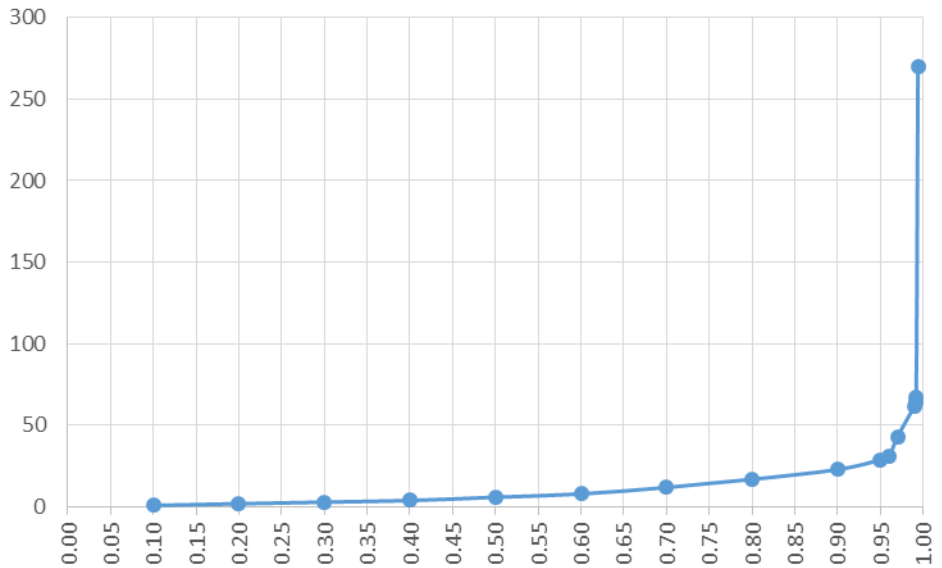


BE CAREFUL WITH YOUR PARAMETERS

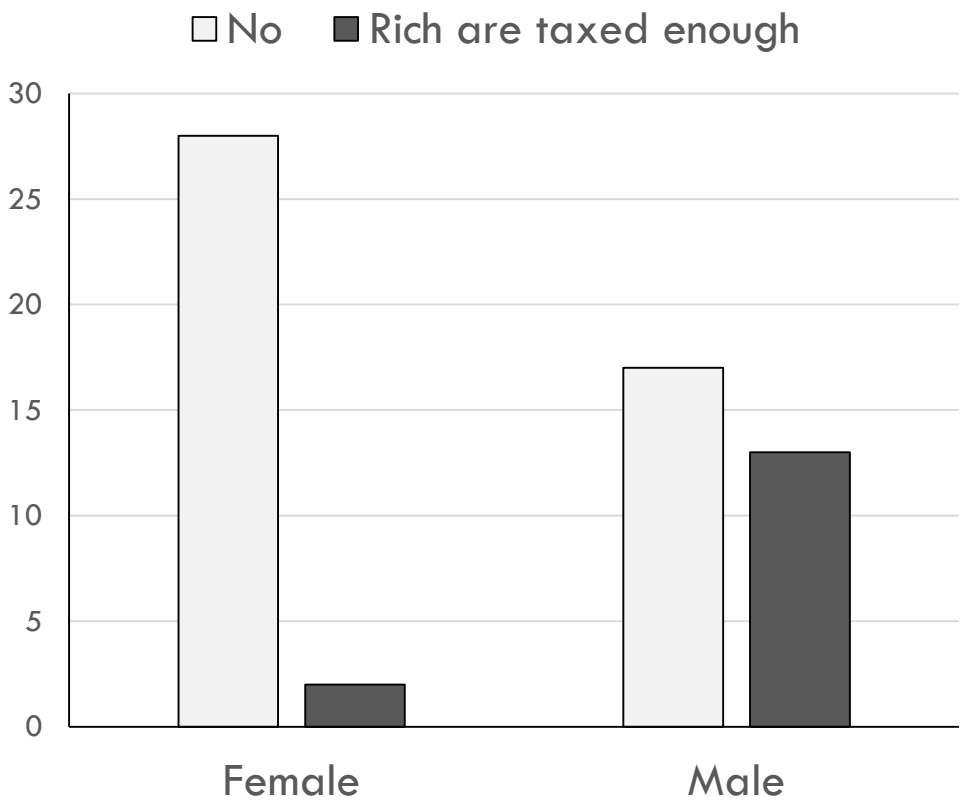
As we vary the probability

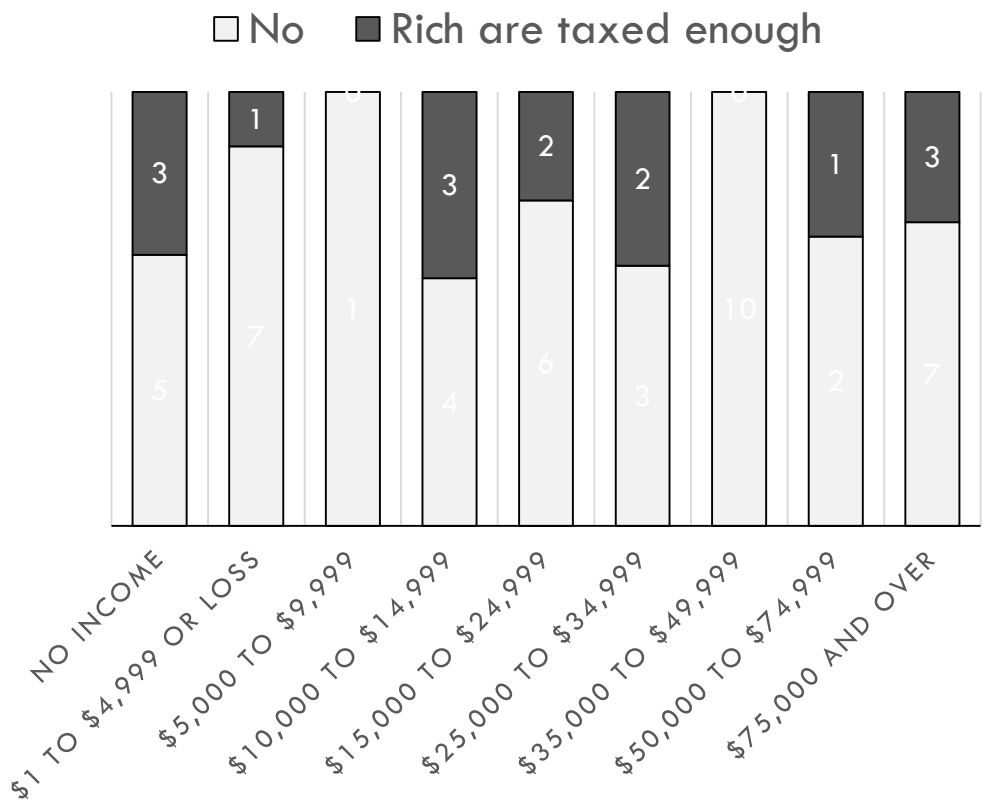


As we vary the confidence



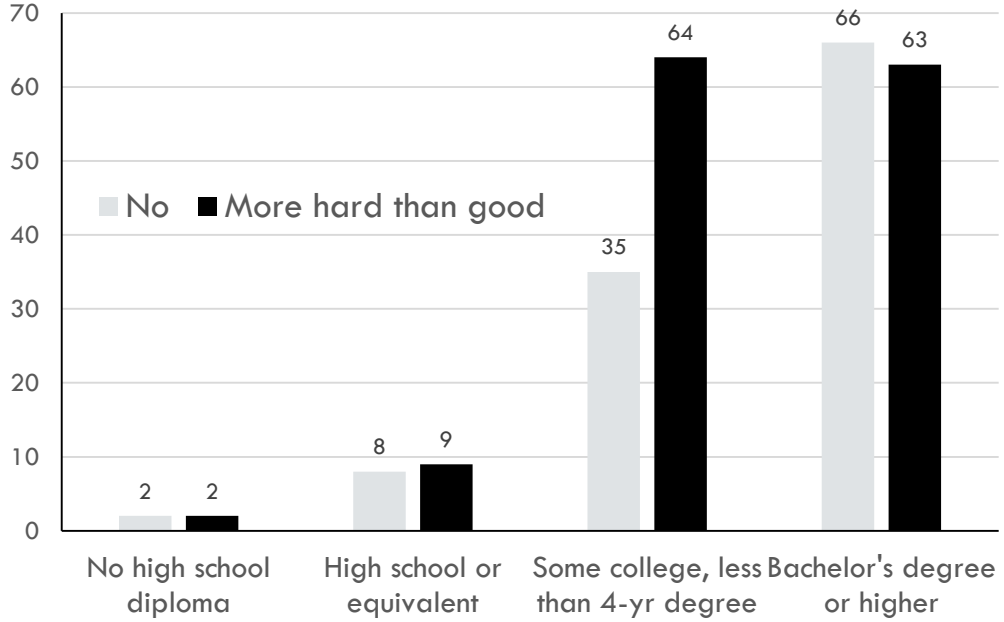
TAXATION, BY GENDER AND INCOME



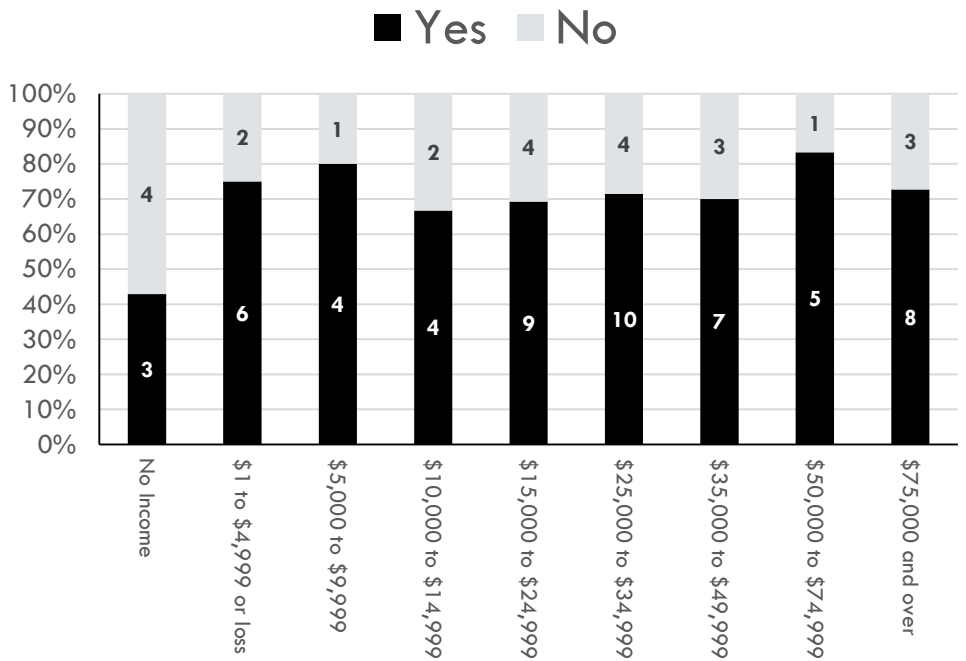


TWO MORE CONUNDRUMS

Affirmative action does more harm than good

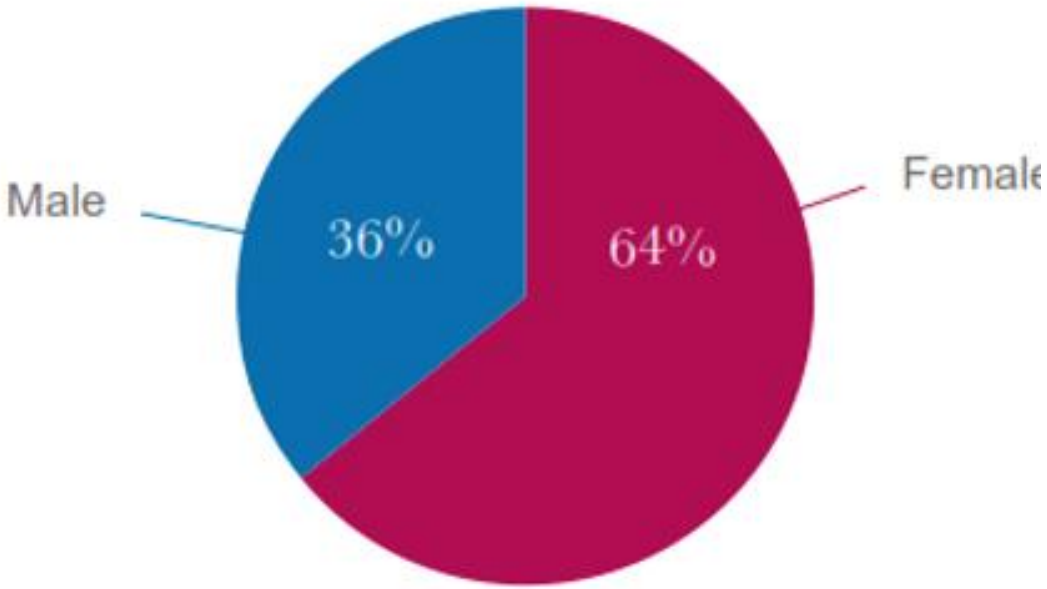


We should break up the big banks

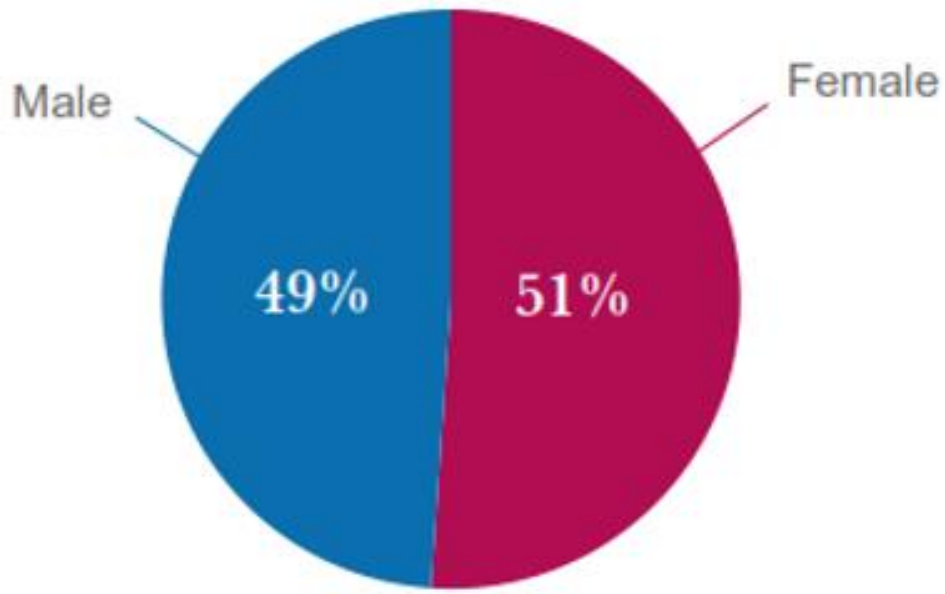


PRIORS FOR THE CROWD

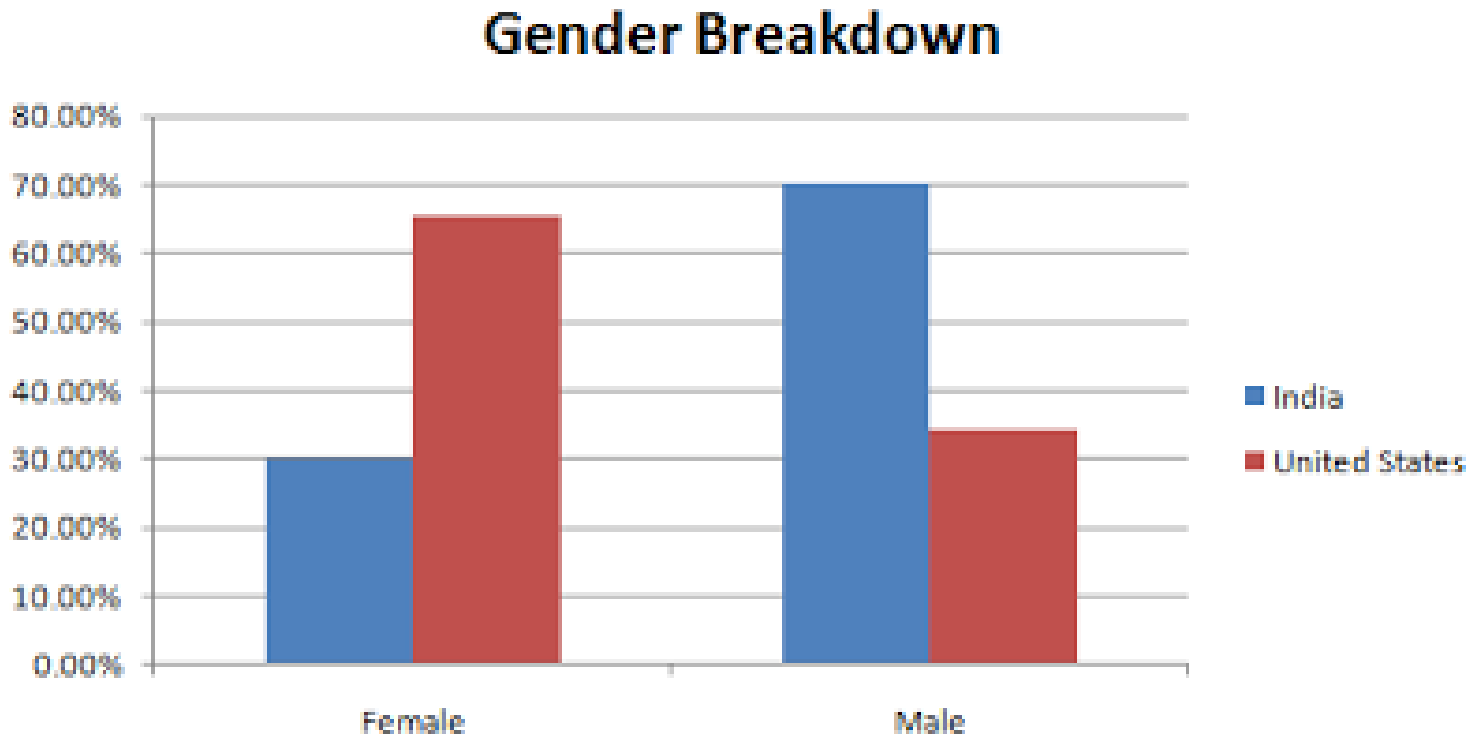
Instant.ly crowd



US census



SEGMENTS OF THE CROWD LOOK DIFFERENT



Panos Ipeirotis, 2010

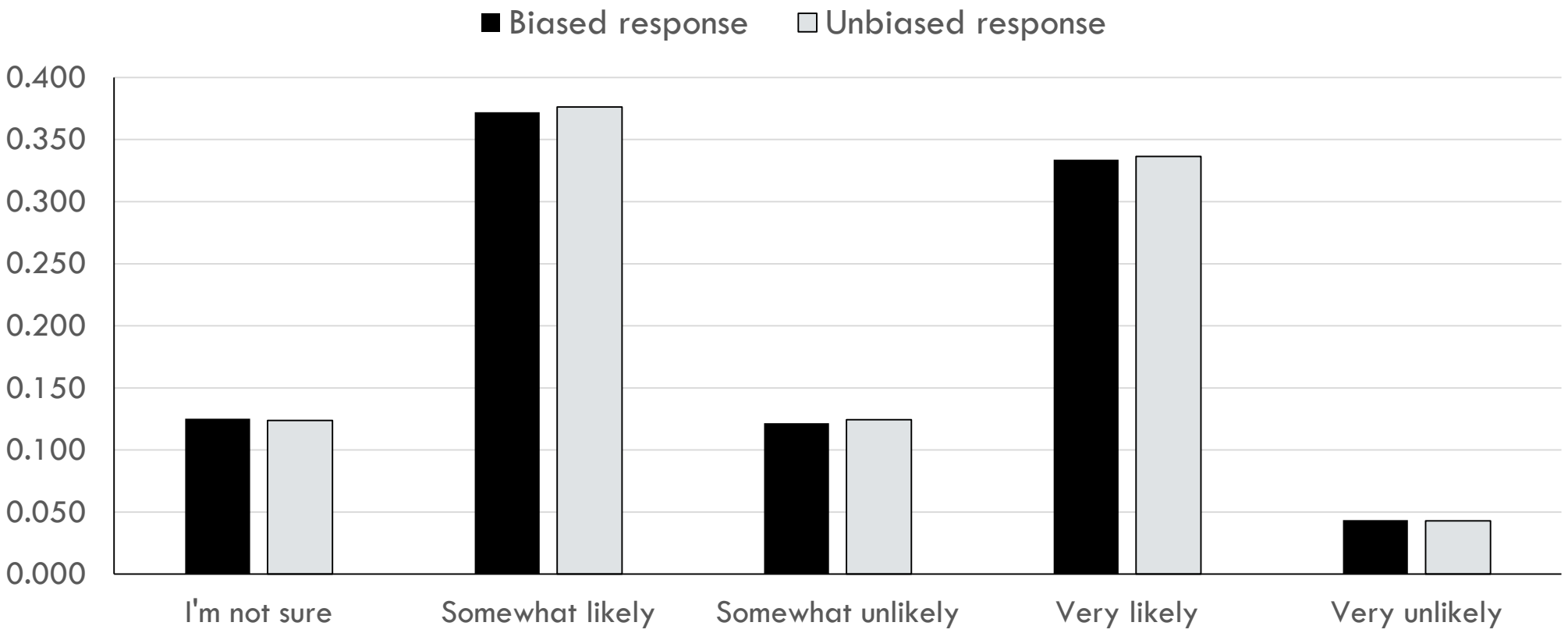
THE UNBIAS OPERATOR

```
var photoAttitudes = (from person in people
                      select new
                      {
                          Used = person.PoseQuestion<bool>(
                           "Have you ever hired a professional photograph
                          WorthIt = person.PoseQuestion<bool>(
                          "Do you feel the money you spent was worth the
                          Quality = person.PoseQuestion(
                          "How would you rate the quality of the pictures
                          "\u2605", "\u2605\u2605", "\u2605\u2605\u2605",
                          HowLikely = person.PoseQuestion(
                          "How likely are you to
                          "Very likely", "Somewh
                          WhatDidYouEnjoy = pers
                          "What did you most enj
```

```
// priors for demographics.
var mturk = MTurkPriors.DefaultPriors;
var census = CSPSlicedPriors.DefaultPriors;

var correctedAttitudes = Unbiasing.Unbias(photoAttitudes,
    p => p.HowLikely,
    p => p.Gender, mturk, census);
```

UNBIASING IN INTERPOLL



LOCAL OPTIMIZATIONS

```
var males = from person in height
            where person.Gender == Gender.MALE
            select person.Height;

var females = from person in height
              where person.Gender == Gender.FEMALE
              select person.Height;
```

Poll twice?

Poll once and reuse the population?

```
var unemployed =
    from person in people
    where person.Employment==Employment.WORKED_FULL_TIME_YEAR_F
    select person;

foreach (var person in unemployed)
{
    if (person.Employment == Employment.DID_NOT_WORK)
    {
        Console.WriteLine("dead code");
    }
}
```

Dead code – reason about collection properties and query/code interactions

CONDITIONAL BLENDING

```
var unemployed =
    from person in people
    where person.Employment==Employment.WORKED_FULL_TIME_
    select person;

foreach (var person in unemployed)
{
    if (person.Gender == Gender.FEMALE)
    {
        Console.WriteLine("unemployed female: {0}", per
    }
}
```

```
var unemployedFemales =
    from person in people
    where
        person.Employment==Employment.WORKED_FULL_TIME_YEAR
        person.Gender == Gender.FEMALE
    select person;

foreach (var person in unemployedFemales)
{
    Console.WriteLine("unemployed female: {0}", person);
}
```

CROSS-CROWD SAMPLING STRATEGIES

Having more than one crowd can help us assess the issue of ignorability: does the fact that we are polling online affect the outcome?

Test respective biases brought about by different crowds

Provide answers for different population mixes: 70% Indian, 30% American

Specific optimizations

- US -> India
- US and India, blended
- US during the day, India at night
- Use India to find missing population segments

MORE OPTIMIZATIONS

Can we cache answers to common questions and reuse them across surveys? Is that always the correct thing to do?

Can we determine questions that are time-sensitive?

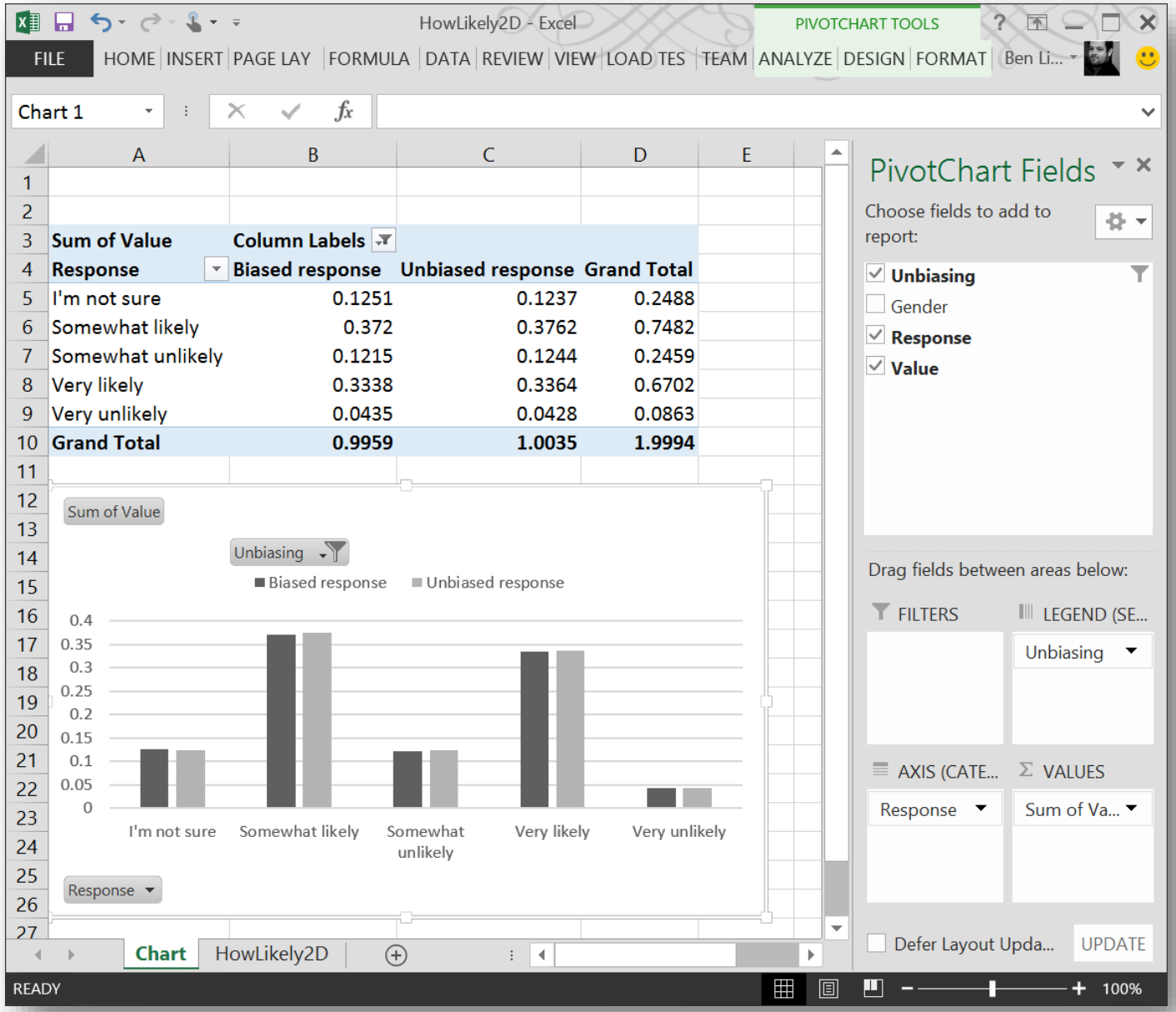
Can we decide if keeping answers to specific questions may lead to privacy violations.

Can we determine questions suitable for different crowds?

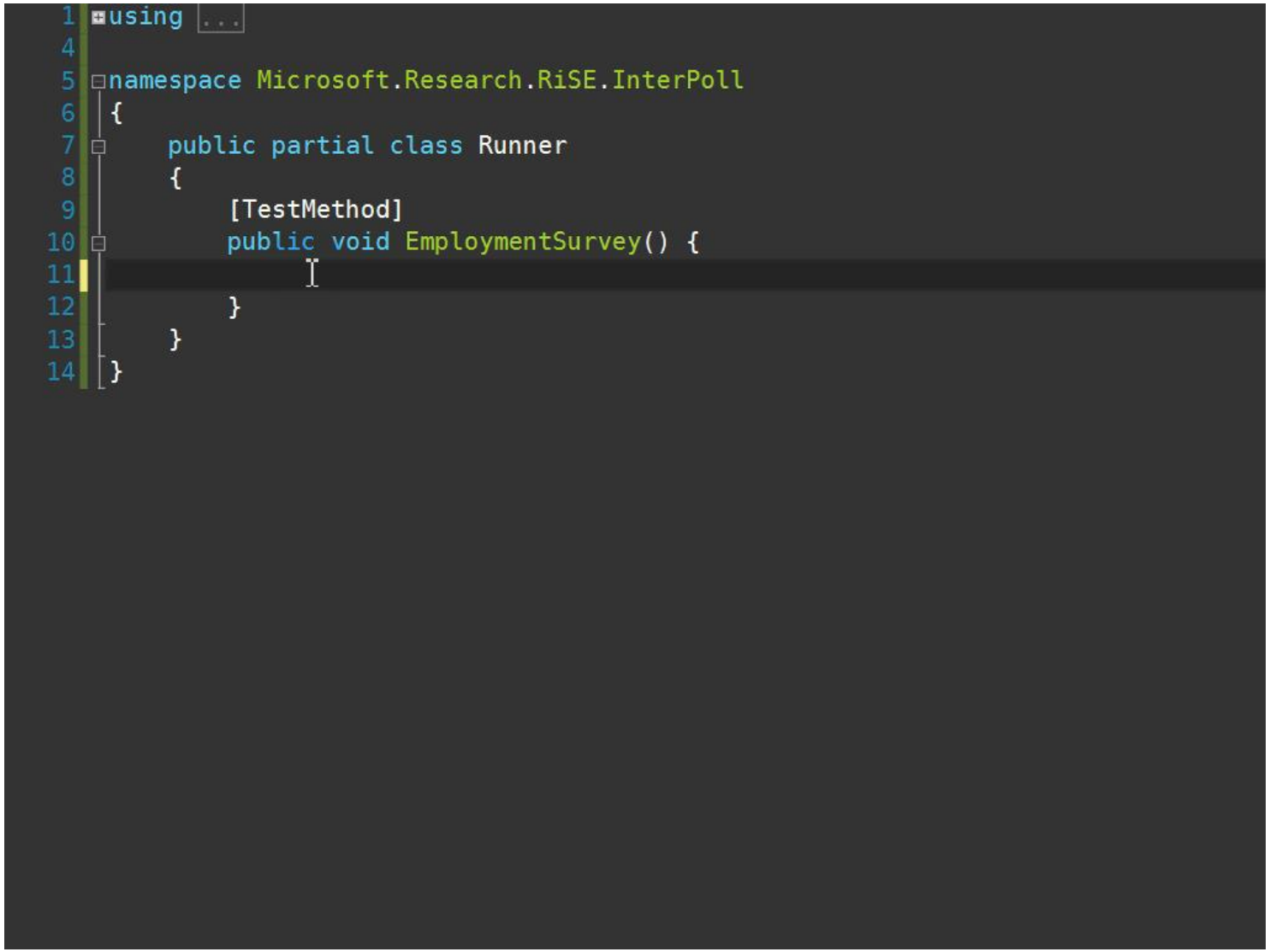
Can we determine if someone is likely to cheat or compensate for the possibility of cheating?

EXPLORING THE DATA

Age	Gender	Race	Education	Income	Employm	Relationsh	Zip	Answer 9	Answer 1	Answ
48	Female	Two or m	Bachelor's	$75,000 ar	Worked fu	Married, S	98275	From time	Definitely	Not a
24	Female	Black or A	Bachelor's	$5,000 to	Worked le	Never Mai	60638	Not at all	Only a littl	Not a
22	Male	White alor	High scho	$15,000 to	Worked le	Married, S	81003	From time	Hardly at a	A litt
31	Male	White alor	Bachelor's	$35,000 to	Worked fu	Married, S	14226	A lot of th	Not quiet	Yes,
18	Male	White alor	High scho	$1 to $4,9	Did not w	Never Mai	77388	From time	Definitely	A litt
53	Male	White alor	Some colle	$75,000 ar	Worked fu	Married, S	90606	A lot of th	Not quiet	A litt
28	Female	Asian alor	Bachelor's	$50,000 to	Worked le	Never Mai	98056	From time	Not quiet	Yes,
50	Male	White alor	High scho	$75,000 ar	Worked fu	Married, S	29730	From time	Definitely	Not a
22	Female	White alor	Bachelor's	$1 to $4,9	Worked le	Never Mai	6612	From time	Not quiet	Yes,
28	Male	White alor	Bachelor's	$10,000 to	Worked fu	Never Mai	95776	A lot of th	Hardly at a	Very
33	Female	White alor	Bachelor's	No Income	Did not w	Married, S	89012	From time	Not quiet	A litt
30	Female	Two or m	Some colle	$1 to $4,9	Did not w	Never Mai	92869	From time	Not quiet	Not a



INTERFACES


using ...
namespace Microsoft.Research.RiSE.InterPoll
{
public partial class Runner
{
[TestMethod]
public void EmploymentSurvey() {
}
}
}

CONCLUSIONS

InterPoll: a system for large-scale crowd-sourced polling

Geared toward

- Developers who want to incorporate human data into their applications
- But also social scientists
- Marketing professionals
- Campaign pollsters

Have explored power analysis

and are doing experiments

on unbiasing

and various optimizations