Visualizing Attribute Uncertainty in the ACS: An Empirical Study of Decision-Making with Urban Planners

Amy L. Griffin, UNSW Canberra
Seth Spielman, Colorado University at Boulder
Jason Jurjevich, Portland State University
Meg Merrick, Portland State University
Nicholas N. Nagle, University of Tennessee-Knoxville
David Folch, Florida State University

NCRN webinar series, 4 February 2015

Goals of the overall project

Census Bureau quantifies the uncertainty in population estimates but….

Does anyone look at these tables?

What are other ways of presenting this information (maps)?

Do users consider the effects of this uncertainty on the conclusions they draw from their use of the data?

Project goals:

1) Understand the extent to which our user group (planners) engages with uncertainty in ACS data.

2) Develop and test methods of representing attribute uncertainty that fit within the workflow & conceptualizations of uncertainty in this user population.

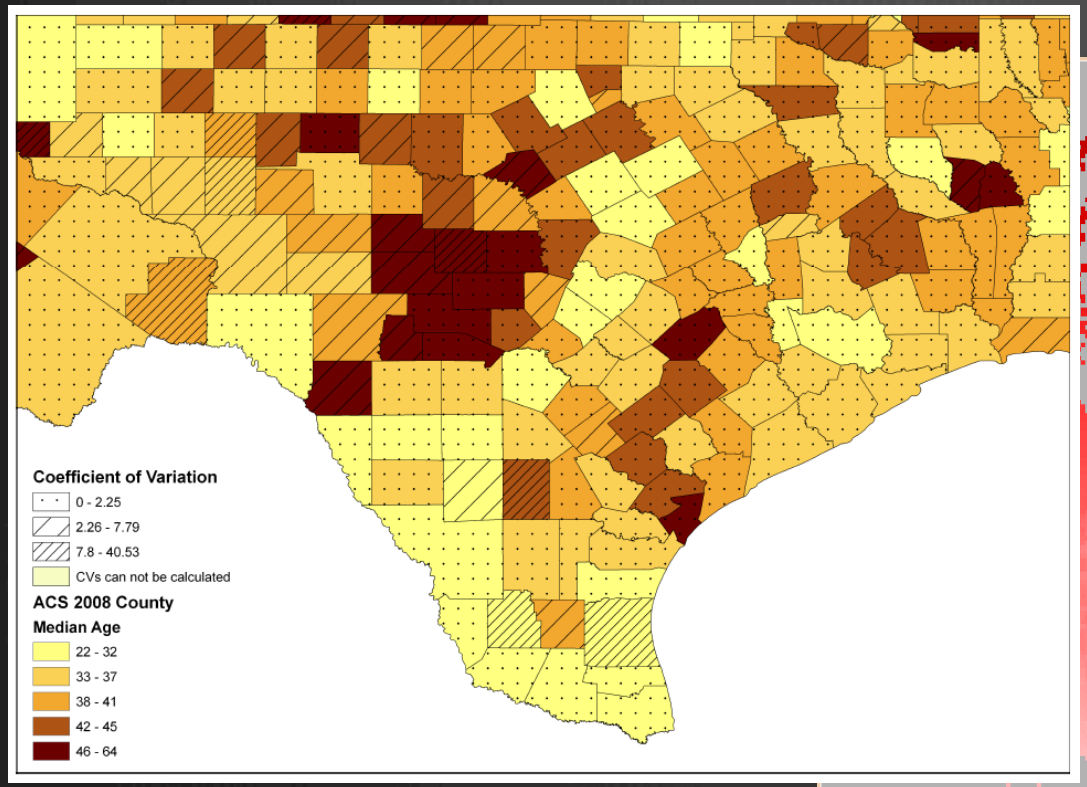
1 - 59 of 59

	United States	
	Estimate	Margin of Error
Total:	308,196,783	+/-21,745
Income in the past 12 months below poverty level:	48,810,868	+/-256,176
Male:	21,938,525	+/-128,136
Under 5 years	2,464,190	+/-30,490
5 years	511,177	+/-13,482
6 to 11 years	2,752,353	+/-35,079
12 to 14 years	1,265,199	+/-19,533
15 years	420,517	+/-10,214
16 and 17 years	784,739	+/-14,155
18 to 24 years	3,254,824	+/-38,523
25 to 34 years	2,682,953	+/-26,570
35 to 44 years	2,204,359	+/-25,235
45 to 54 years	2,239,711	+/-25,932
55 to 64 years	1,932,458	+/-22,850
65 to 74 years	837,917	+/-15,288
75 years and over	588,128	+/-12,841

ACS Data Example, with MOEs

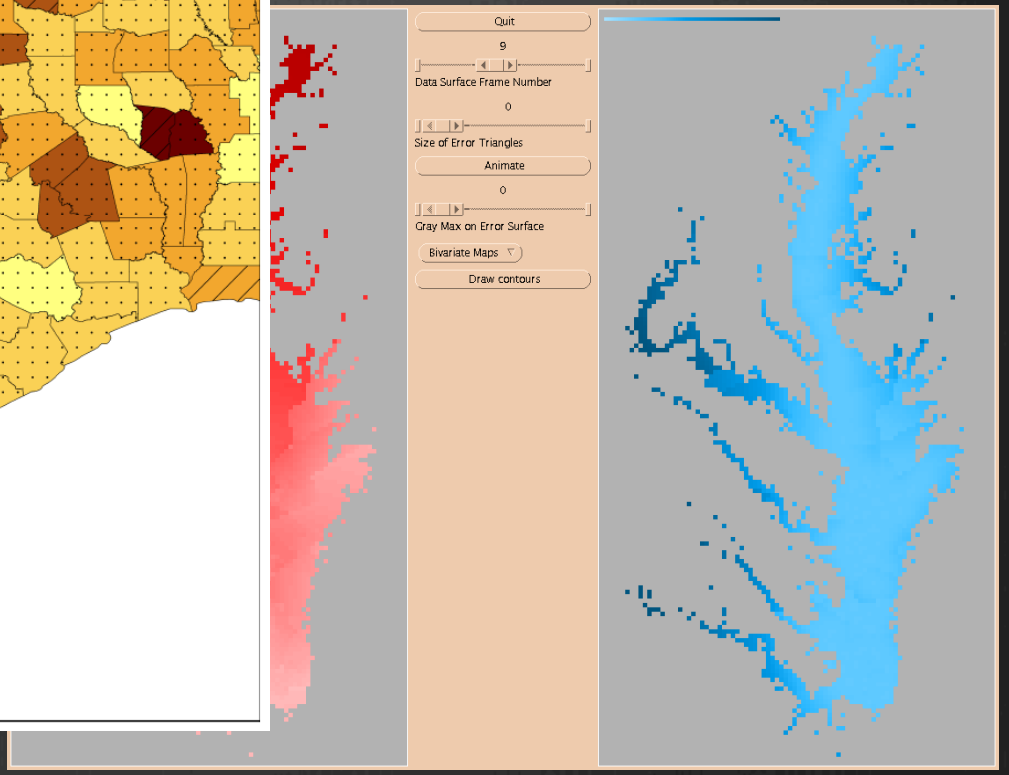
Methods of representing attribute uncertainty in maps

integrated representations

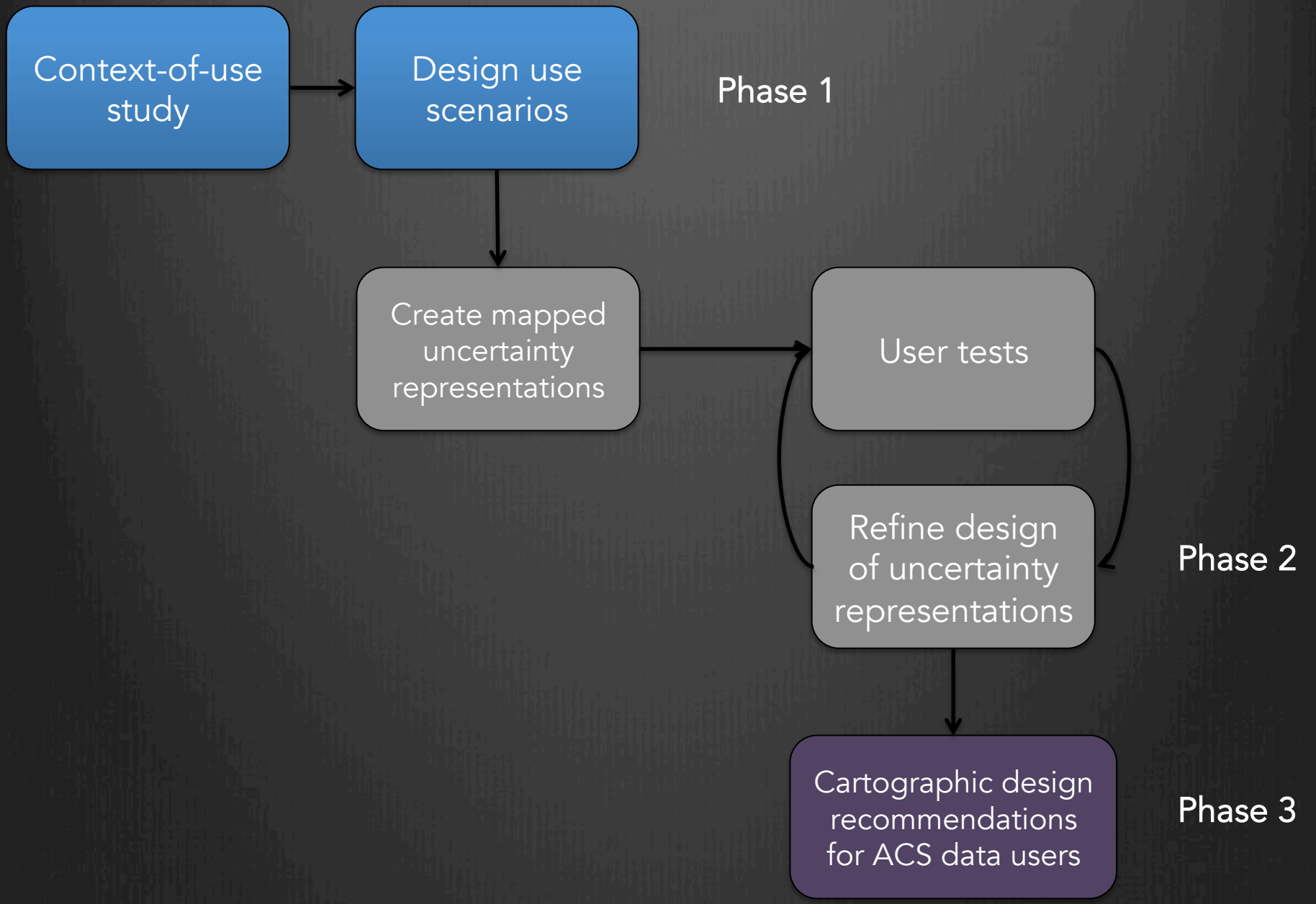


Wong & Sun (2013)

Howard & MacEachren (1996)



separated representations


Context-of-use study
Design use scenarios
Phase 1
Create mapped uncertainty representations
User tests
Refine design of uncertainty representations
Phase 2
Cartographic design recommendations for ACS data users
Phase 3

Phase 1:
Context of Use Study Among Planners

Phase 1 components:

1) In-depth interviews (n = 7)
2) Artifact analysis (n = 30)
3) Survey questionnaire (n = 235)

What types of tasks do our mapped uncertainty representations need to support?

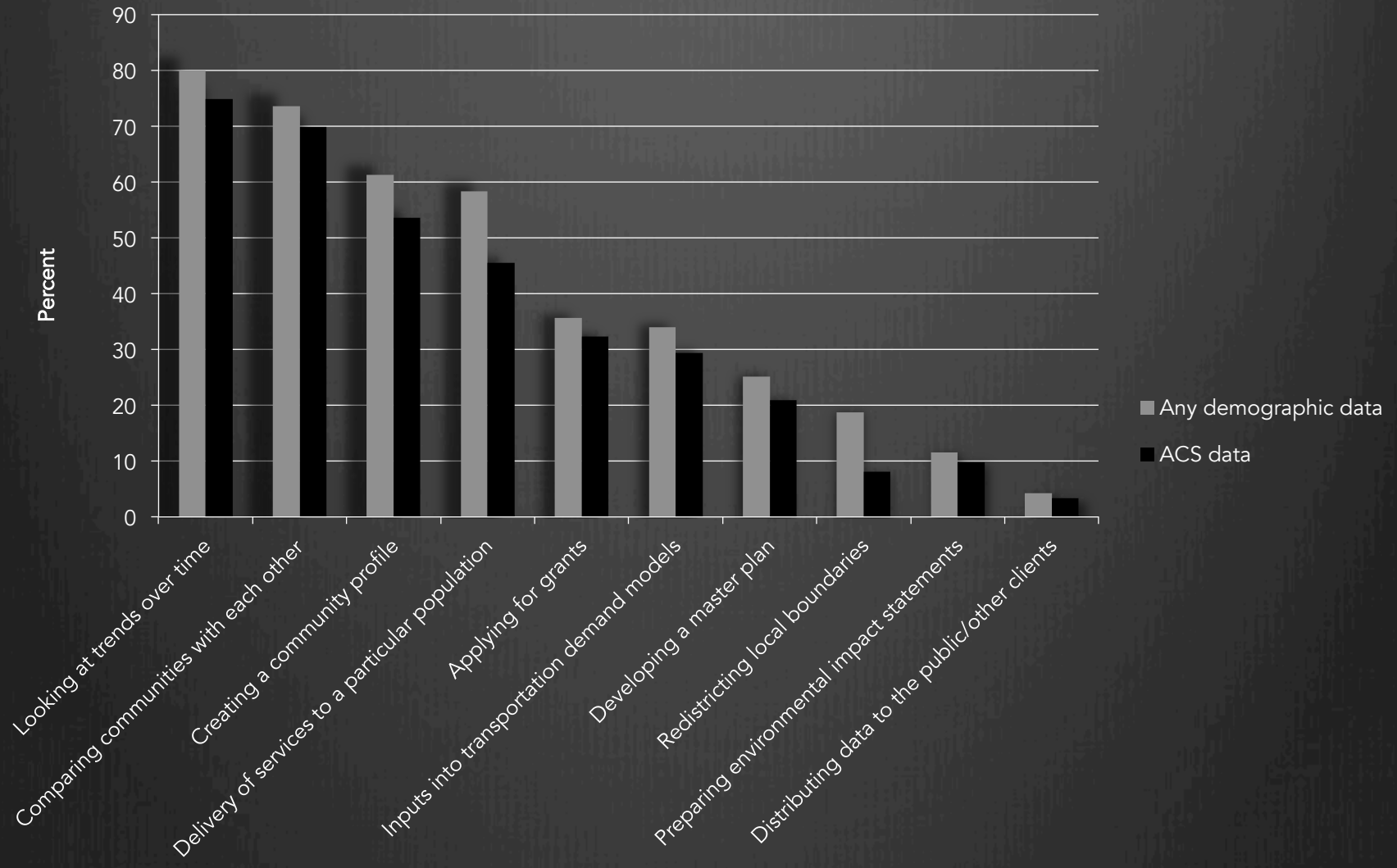
What are reasonable user testing scenarios for this user group?

Survey results

Who were the respondents?

They typically:

- Are at least semi-regular GIS users
  - (51.5% use it at least weekly; 68.9% at least monthly)
- But have little or no formal training in GIS
  - (44.7% have no formal training; 11.9% have taken only one course)
- Have some formal training in statistics
  - (43.4% have taken 3+ statistics courses; only 2.6% have taken no stats courses)
- Have been working as a planner for a range of lengths of time
  - (mean years = 9.6, standard deviation = 9.3)



What were the common tasks respondents used demographic data for?

How are respondents using ACS data? (survey)

Communication method	% of respondents
Table	81.7
Maps	79.6
Bar chart	72.8
Textual description	69.4
Line graph	68.1
Pie chart	58.3
Other statistical graph	34.9

General conceptions of uncertainty, interviews

I would define uncertainty as, I think in the most simple terms, it's a question mark. And when you're working with data is it a big question mark? Or is it a little question mark? (I1)

So there's a margin of error, so…. I have a funny acronym for the margin of error. I call it best available data. It's bad data. (I2)

Yeah, I mean, it's garbage in, garbage out, right? If you put in data that you don't…you have no faith is accurate, that you can't have any faith in the outcome of your analysis using that data. (I4)

What does the word uncertainty mean to you? Reality. (I5)

Uses of ACS uncertainty information, interviews

Uncertainty is typically used to determine when to find other data sources:

And so I just more use it as a caution and….and figures and try and discourage the use of figures where it's an astronomical margin of error…. I don't know if I'd say 'out', but use extreme caution and maybe, maybe the caution is great enough to say 'out', as far as using this… (I3)

You know, like if I got this and I'm looking at it, and I see some of these larger margins of error where the margin of error is bigger than the number, I guess our interaction is probably very basic still of 'wow, that's big, maybe I shouldn't use it' and then figure something else out, you know, which it's a very basic interaction I think with data in that sense. (I1)

We're just going to trust the numbers here are pretty good for evaluation purposes, and frankly, you know, if the margin of error is too large, you know, maybe…you shouldn't use it [the data] at all. I mean that might be the one reason that we use it, but what we would hardly ever present it in any type of document. (I6)

Uses of ACS uncertainty information, interviews

And what 'something else' options do they employ?

Because of the size of [small city] and then sometimes, you know, you're working at a tract level, you do get more uncertainty in that, so we'll go look for physical indicators in the community…or we'll talk to human services providers and just say 'hey who lives out there?' (I1)

I think the answer, like I said earlier, is just finding other data sources to augment it, so you're not relying solely on Census as your only source of information…to the extent that you can. (I4)

Uses of uncertainty information, surveys

19.6% (n = 46) admit to not paying much attention to MOEs when they use ACS data.

When asked: 'What would you consider to be a high margin of error?'

39.6% (n = 93) specified a particular percentage of the estimate, ranging from 7% of the estimate to 'more than 100%' of the estimate

13.6% (n = 32) provided an answer that indicated that what was 'high' would depend on the geographic scale or the context of the data.

> …The data should be considered in the context of the data universe, the geography (the law of large numbers is meaningful here), and the purpose to which the data are being applied.

A few respondents cited guidelines developed by ESRI or the Census Bureau:

> Any estimate with a coefficient of variation greater than 15%, as recommended in the Census Bureau's handbook for State and Local Governments

Uses of uncertainty information, surveys

11.9% (n = 28) say that they would try to warn the end user, most typically through caveats published with a table or text.

But there are a few people who discussed graphical communication:

> Depending on the use of the data -- that is, if no capital or human life issues are involved -- I might present the value as a range rather than just the number. Graphing can be helpful in this regard, as it allows the viewer to understand that the value is "somewhere" in the range but we can't be precise enough to name it.

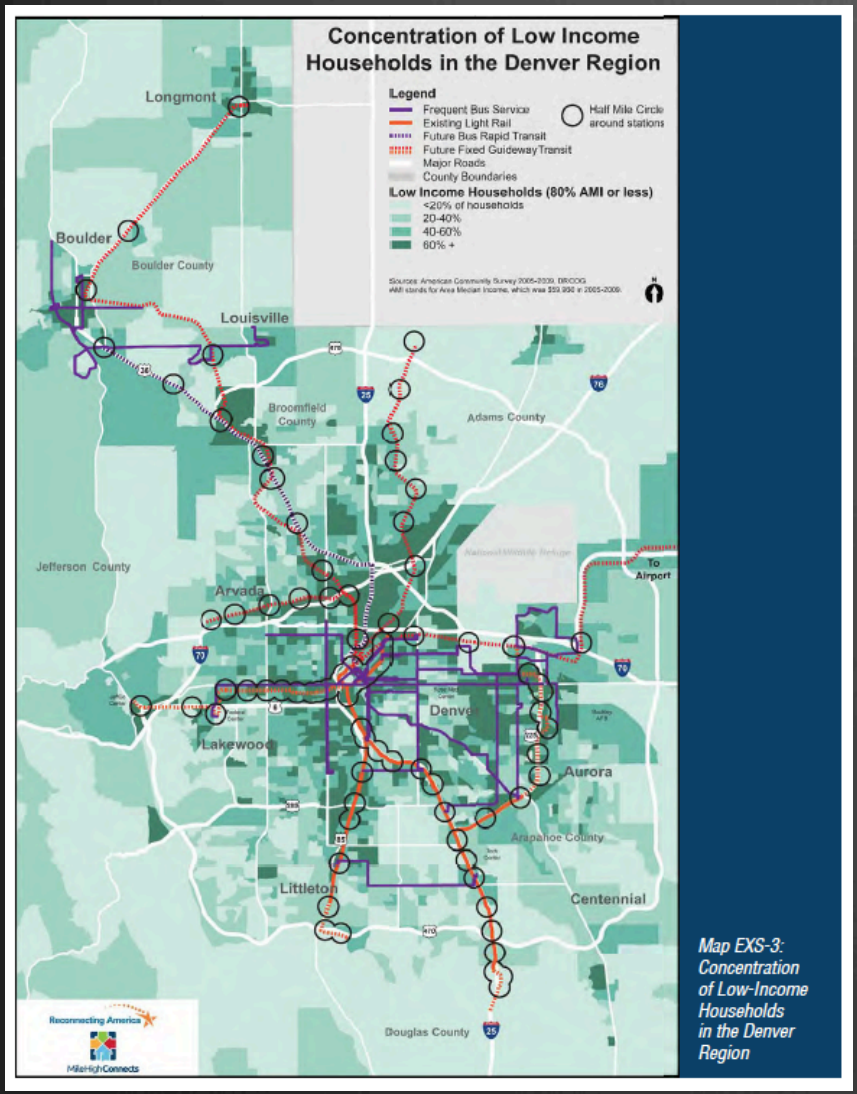
> If we were just displaying data in a lookup table, we would list the data point and the margin of error. In addition we would provide resources for people to understand what it means and what to do with it. If we were displaying it in a graphic, we would be clear about it's comparability to other data points.

Artifact Analysis Results, brief summary

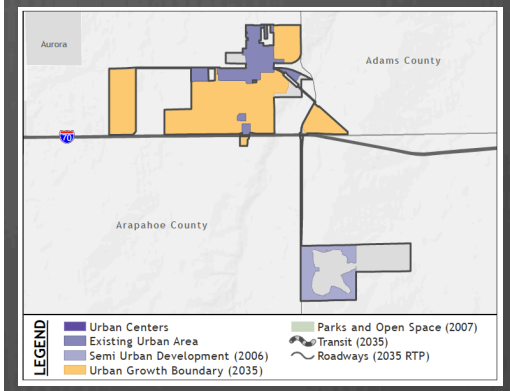
Uncertainty is almost never communicated to end users of information.

Maps are commonly used for contextual reference rather than for communicating thematic information.

Thematic information is more commonly conveyed with graphs.



Map EXS-3: Concentration of Low-Income Households in the Denver Region

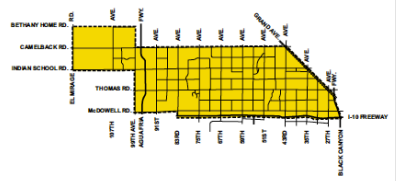


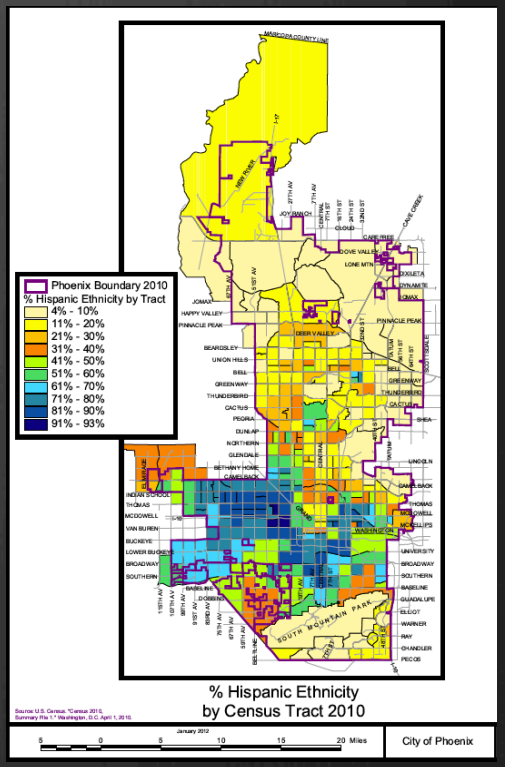
PRESENT & PROJECTED* SOCIO-ECONOMIC PROFILE

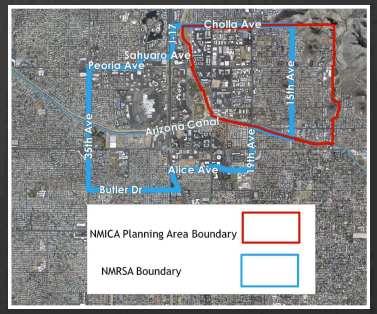
Population		Households		Employment	
1995	2020	1995	2020	1995	2020
172,100	238,900	52,800	76,100	52,700	71,200

*Source: Village Size - Village GIS Coverage, Phoenix Planning Department, October 1997
Projections - Whole Traffic Analysis Zones, Maricopa Association of Governments, July 1997

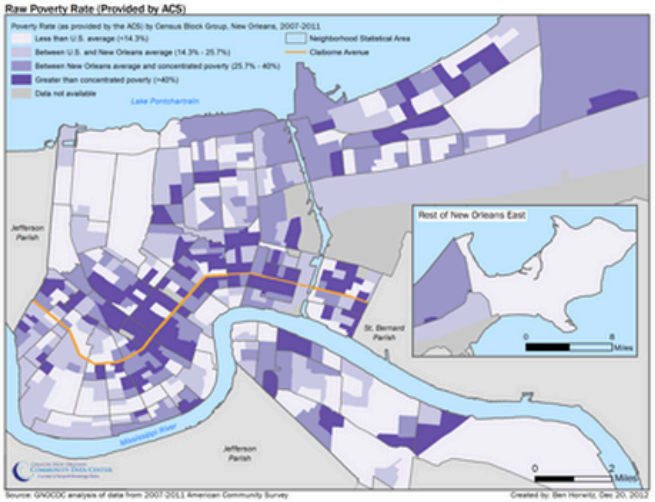
MARYVALE VILLAGE



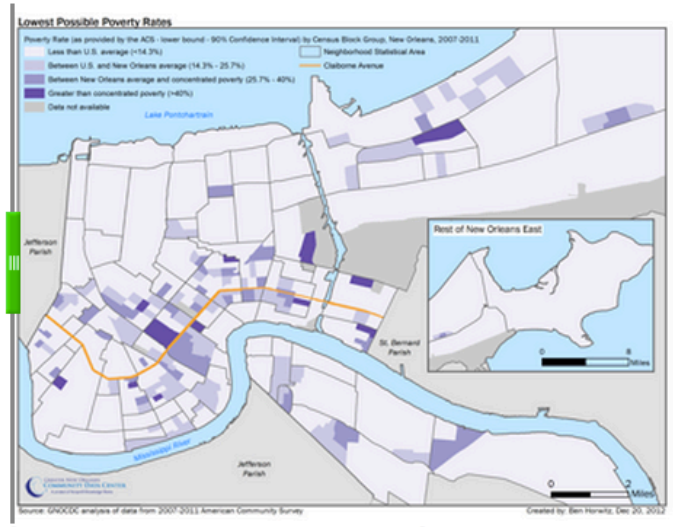




Raw Poverty Rate (provided by ACS)

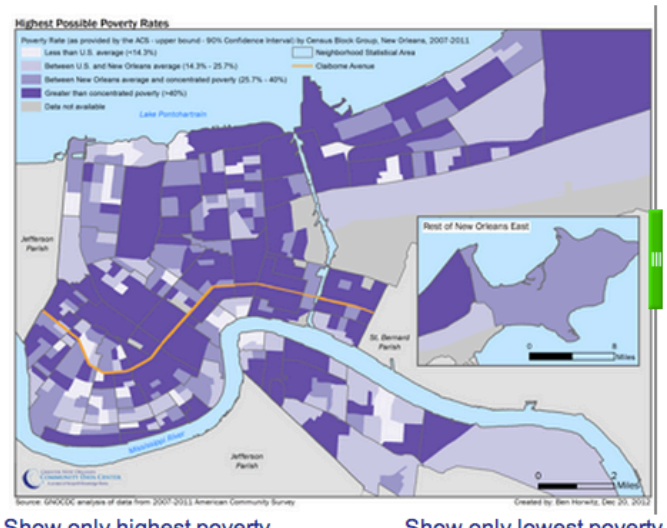


Explore Highest & Lowest Possible Poverty Rate



Show only highest poverty Show only lowest poverty

Explore Highest & Lowest Possible Poverty Rate



Show only highest poverty Show only lowest poverty

Greater New Orleans Data Center (2012)

(Mis)conceptions about uncertainty in the ACS

Well, it's not really a margin of error problem in the sense of…it's, it's. Like I said, it all depends on whether people actually care. And I mean it is apples to apples. Everyone, what [I6] said, was using the same…as long as we're all using ACS data, it doesn't really matter because they all have the same margin of error, you know, in terms of figuring out, ok….And I also found when I started bringing it in this time around, I eliminated the MOE. I took it out. Because it's just all these extra columns that I don't need. (I2)

Any good statistics class, software, person who just does statistics will show… you have to include a margin of error when you do the type of sampling. However, we just don't use it. Nobody….unless you're a statistics type person presenting to statistics professors where you have to have your footnotes in there…for the actual real world studies, what I said is the case. If you're comparing ACS to ACS, it really doesn't matter. They're going to have the same margin of error, more or less. (I6)

(Mis)conceptions about uncertainty in ACS data, surveys

Only 53% of respondents (n = 125) indicated that they agreed that the reliability of ACS data was not the same in all places.

74% of respondents (n = 174) knew that they needed to be more careful (ie that uncertainty was higher) when using estimates for smaller than for larger places.

→ They generally understand geographic scale issues with ACS, but not that uncertainty varies across space.

Slightly less than half of respondents (n = 101, 42.9%) agreed that it is necessary to consider MOEs when making comparisons between places.

Other barriers to appropriate consideration of uncertainty information, surveys & interviews

Time

I SHOULD not use the data or provide a range from 0-200 but often I don't have the time to look in detail at the MOEs for as many geographies and years of data that we have to provide data for. It gets overlooked much too often but it's hard to have a good solution when there isn't better data available. (survey response)

Difficulty communicating uncertainty

If you get too deep into using margin of error, in a powerpoint presentation to a councilperson, or whomever, it will just…you've lost everyone. (I6)

So it's just kind of breaking it down for them…the average citizen that doesn't deal with stats or planning and those things. (I7)

Summary of the Most Important Phase 1 Results

1) Some planners have some important misconceptions about uncertainty in ACS data, and planners make widely varying judgments about how much uncertainty is 'too much'.

2) Planners seem to use ACS uncertainty information in somewhat limited ways – prescreening data and finding alternatives if the uncertainty is unacceptably high rather than communicating it directly to decision makers.

3) There are specific tasks for which planners could benefit from effective methods for conveying uncertainty information, especially comparisons between places (tract to tract, city to city, city to state, city to nation) or time -- benchmarking.

How did we use this information?

Some use scenarios we considered:

1) Establishing whether a trend over time is real, given the uncertainty in the data.
2) Establishing whether a location's estimate for a particular variable meets some threshold (eg a grant cutoff), given the uncertainty in the data.
3) Delineating where a significant population resides within a community → to target delivery of services.

Which uncertainty measure should we use?

How should we class uncertainty data?

Experimental task, repeated 4 times

with and without uncertainty information, two locations

A new federal program provides tax incentives for investment in the "poorest" part of the metropolitan area. The city administration has asked you to identify the "poorest part" of the city, which is defined as a contiguous group of five census tracts. The poorest part of the metro area will be identified as a "special opportunity zone" and all investment in that area will be tax-deductible.

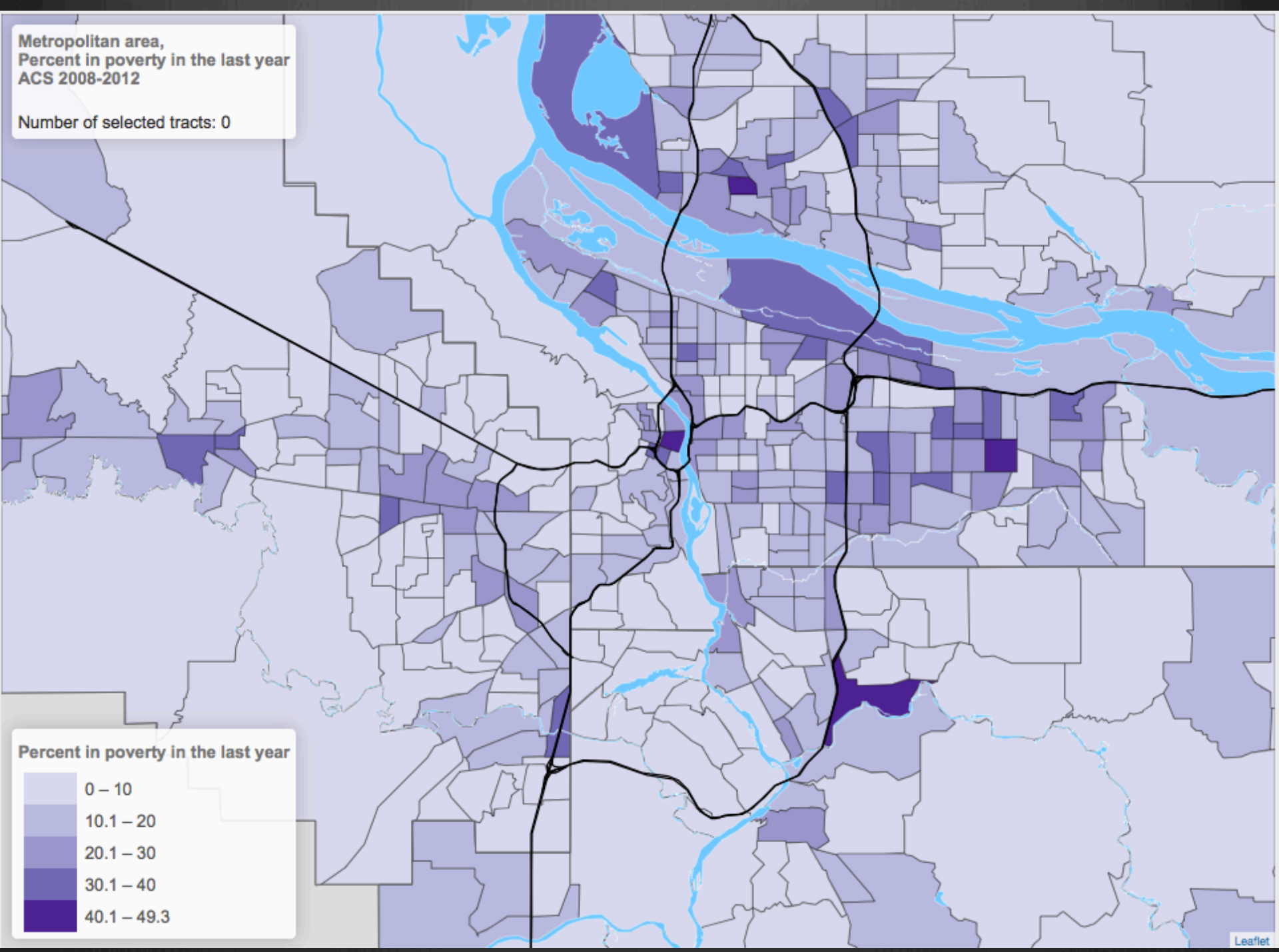
Your first task is to identify a group of five census tracts with the metropolitan area that meets the eligibility criteria so that your department can target advertising about the program most effectively.

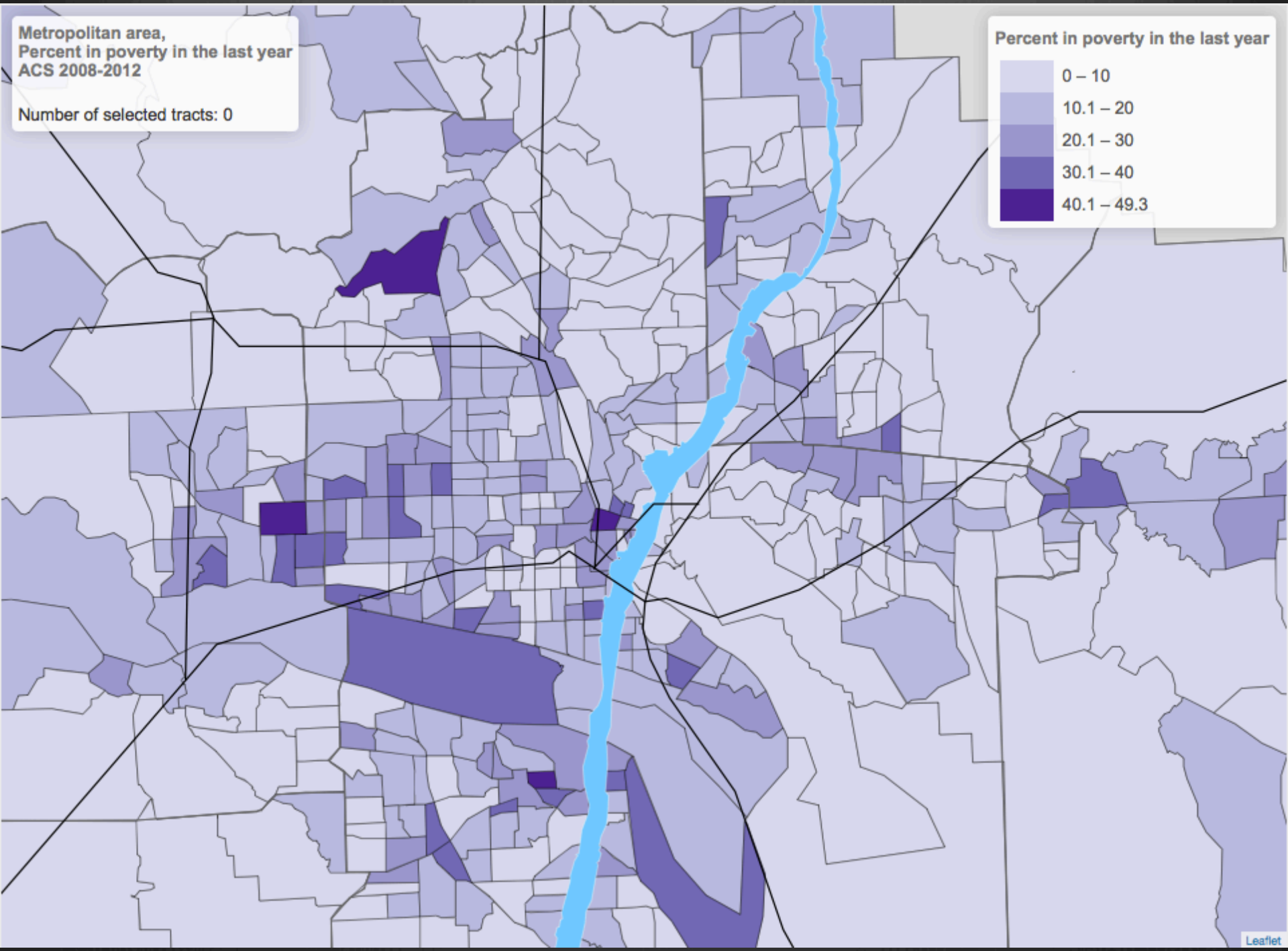
Create a special opportunity zone by selecting a group of five census tracts from the map below that qualify for this program. The tracts must be contiguous (i.e., touching each other) so that it is possible to walk/drive between all parts of the zone without leaving it.

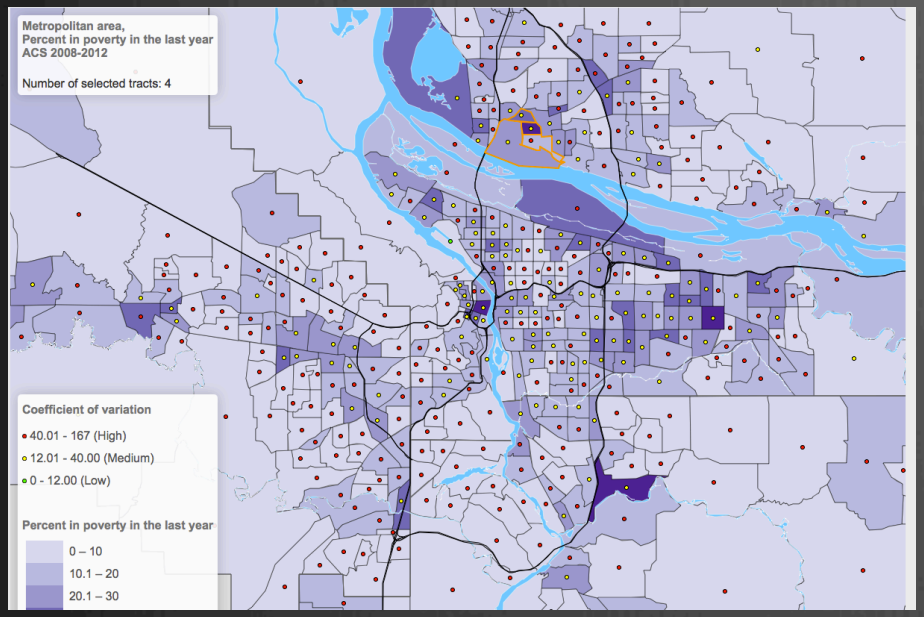
You can select a census tract by clicking on it, and you can deselect a tract by clicking on it a second time.

Experiment Design

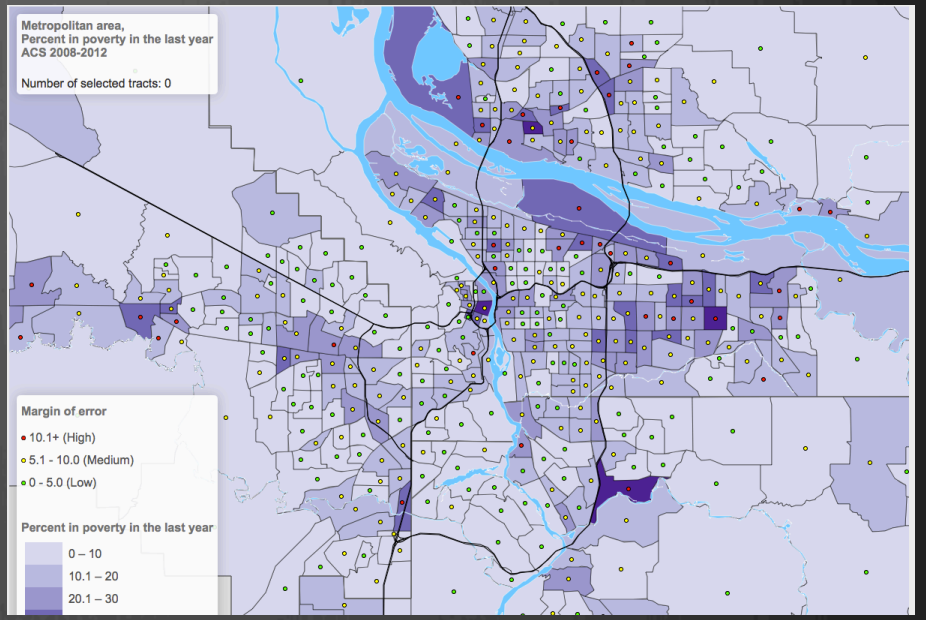
- Run in two populations:
  - planning students (Masters of Urban and Regional Planning)
  - professional planners
- 2 x 2 (between subjects) x 2 (within subjects)
  - Uncertainty representation (between) (stoplight/sketchy)
  - Uncertainty measure (between) (margin of error/coefficient of variation)
  - Mapped location (within) (disguised Portland/Portland)
- Behavioural data (students & planners)
  - Time to complete task
  - Task answer (areas selected)
- Other data (planners only)
  - Eye movements
  - Concurrent verbal protocol
  - Participant demographics & experience data


Metropolitan area,
Percent in poverty in the last year
ACS 2008-2012
Number of selected tracts: 0
Percent in poverty in the last year
0 – 10
10.1 – 20
20.1 – 30
30.1 – 40
40.1 – 49.3
Leaflet

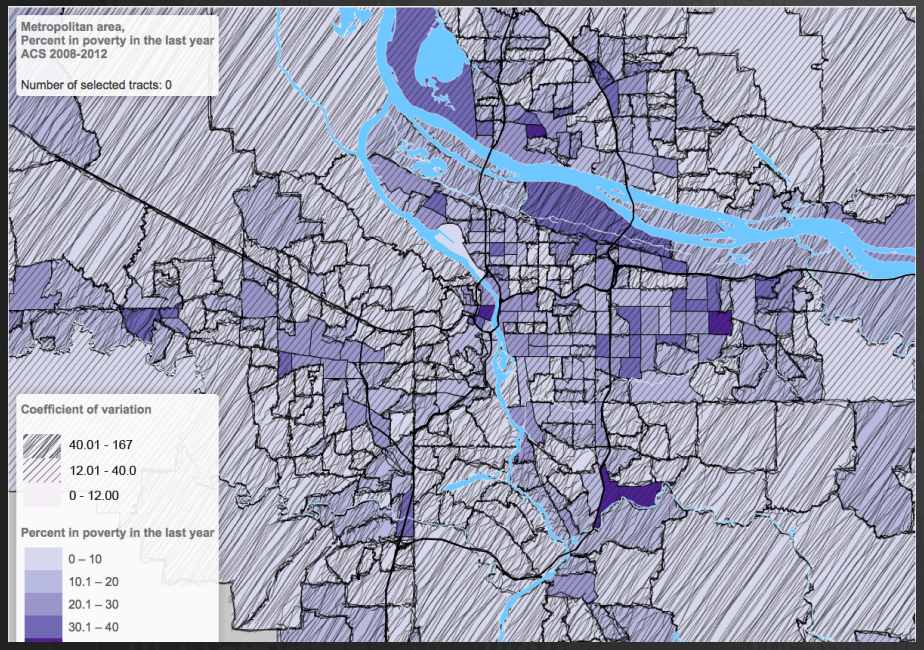

Metropolitan area,
Percent in poverty in the last year
ACS 2008-2012
Number of selected tracts: 0
Percent in poverty in the last year
0 – 10
10.1 – 20
20.1 – 30
30.1 – 40
40.1 – 49.3
Leaflet



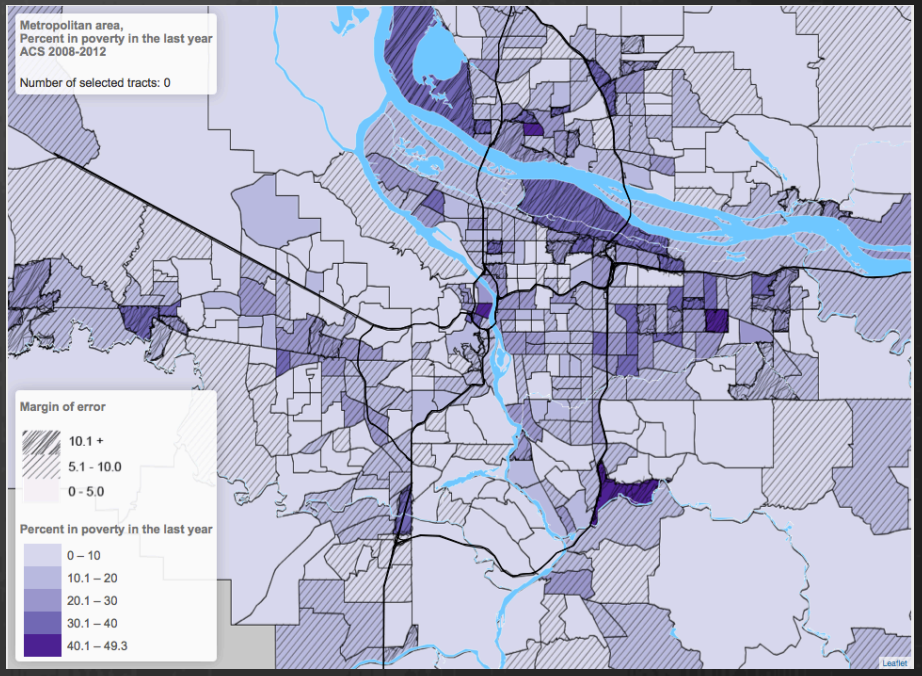
Stoplight CV



Stoplight MOE



Sketchy CV



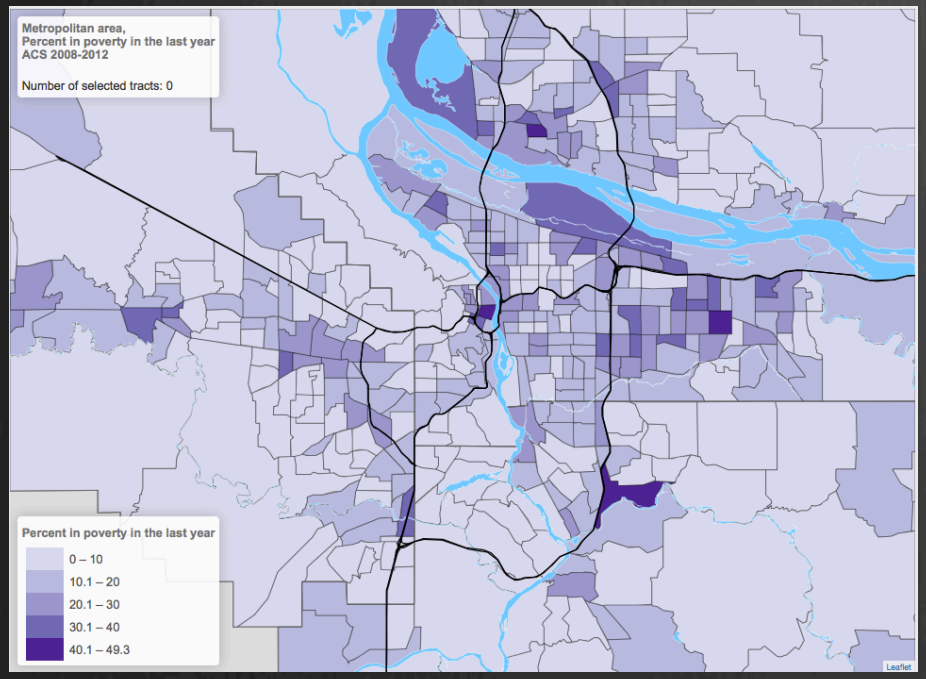
Sketchy MOE

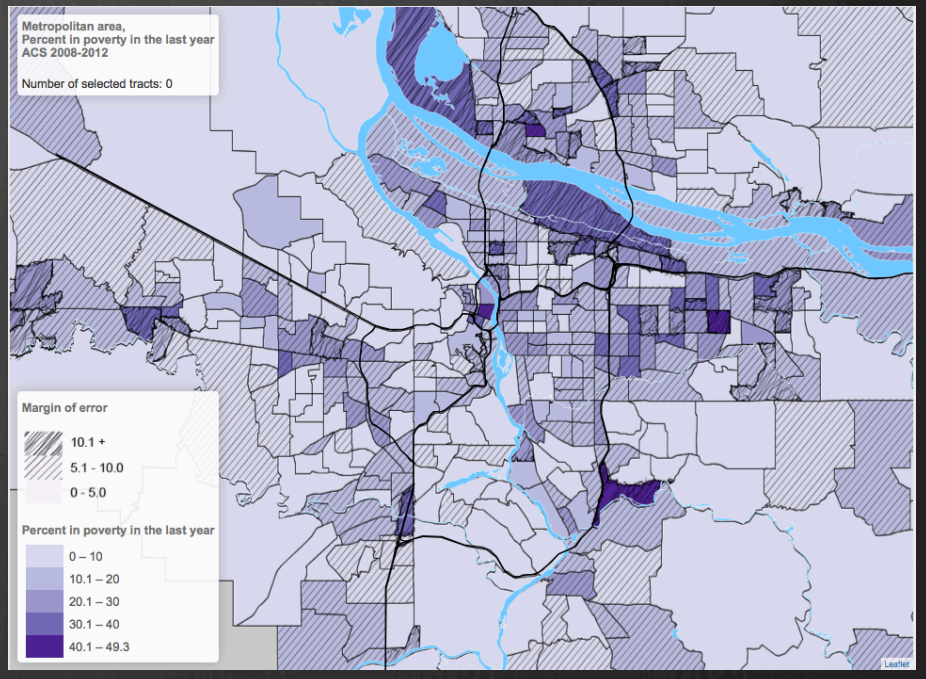
A few preliminary results

- Does representing uncertainty affect decision-making?
  - Compare areas selected in the 'no uncertainty' map with areas areas selected in the uncertainty maps.
- Did the area selected differ between the disguised Portland & Portland, both with and without uncertainty represented?
  - Compare areas selected in 'no uncertainty' maps for the disguised Portland and Portland, and the uncertainty maps for the disguised Portland and Portland.
  - What role does place-based knowledge play?
- Does it matter how we represent uncertainty (visual method)?
  - Are there differences by condition?

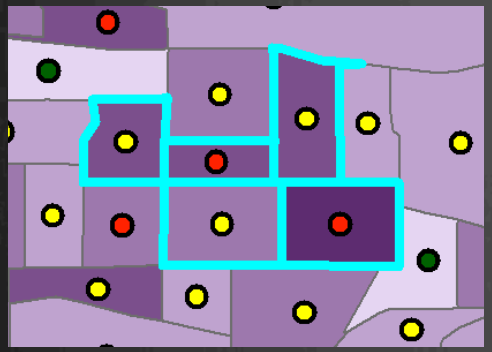
Does representing uncertainty affect decision-making?

Compare areas selected in the 'no uncertainty' map with areas areas selected in the uncertainty maps.

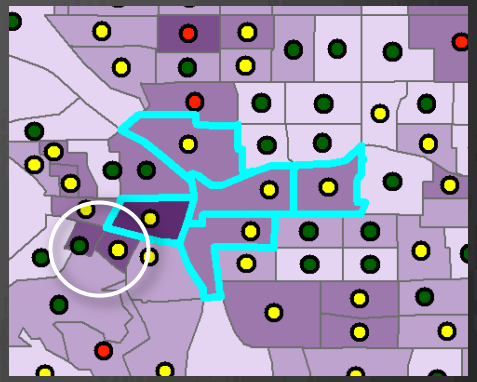




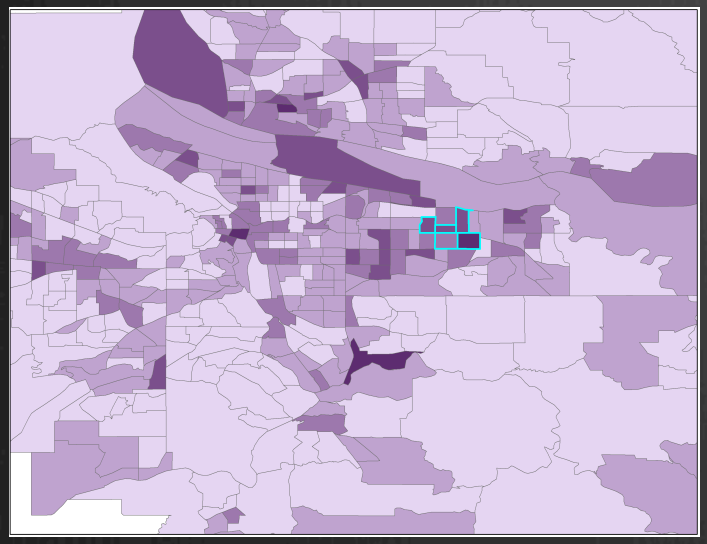
Example of effects on decision-making



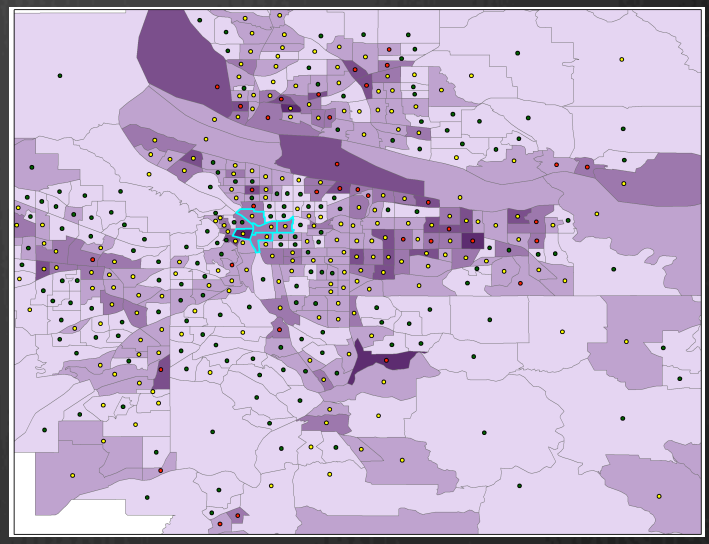
Uncertainty of original selection



Uncertainty of new selection



No uncertainty, Portland



MOE uncertainty, Portland

Decision changes: the case of downtown

Most participants chose either the area in downtown or in east Portland to build their region.

Minor differences in which polygons selected from the core area of higher poverty, but generally 'the same' areas selected by:

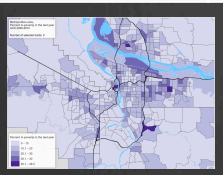
Location	DP no Uncertainty	DP with Uncertainty	Portland no Uncertainty	Portland Uncertainty
Downtown	8	18	3	6
East Portland	45	31	48	43
Vancouver	0	0	2	3
Other	2	4	2	3

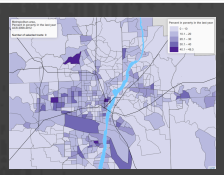
Did the area selected differ between the disguised Portland & Portland, both with and without uncertainty represented?

Condition	Decision did not change between disguised/ real Portland	Decision did change
N/A	26	29

Condition	Decision did not change	Decision did change
Sketchy CV	5	9
Sketchy MOE	3	10
Stoplight CV	5	9
Stoplight MOE	5	9

No uncertainty



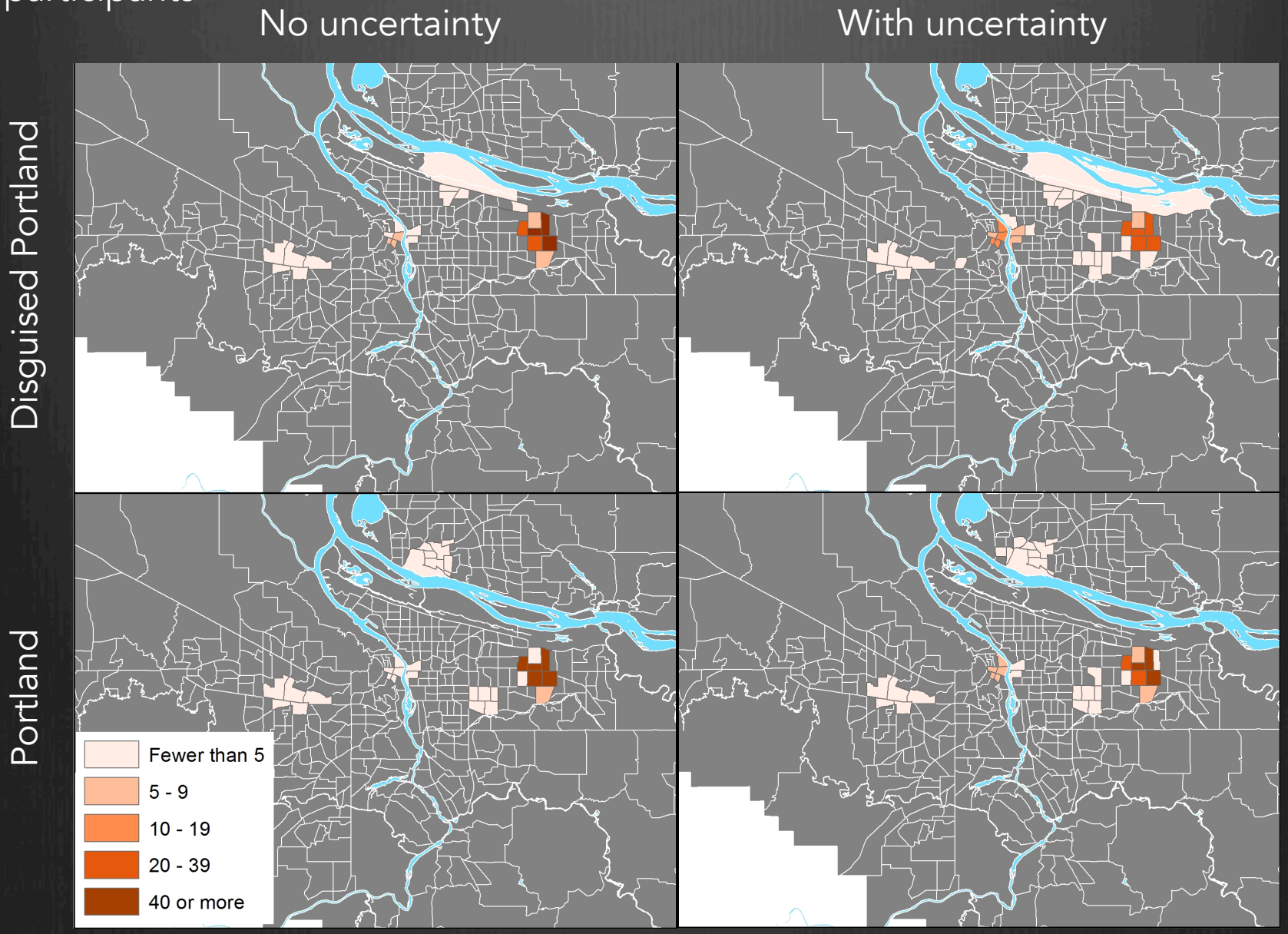


With uncertainty

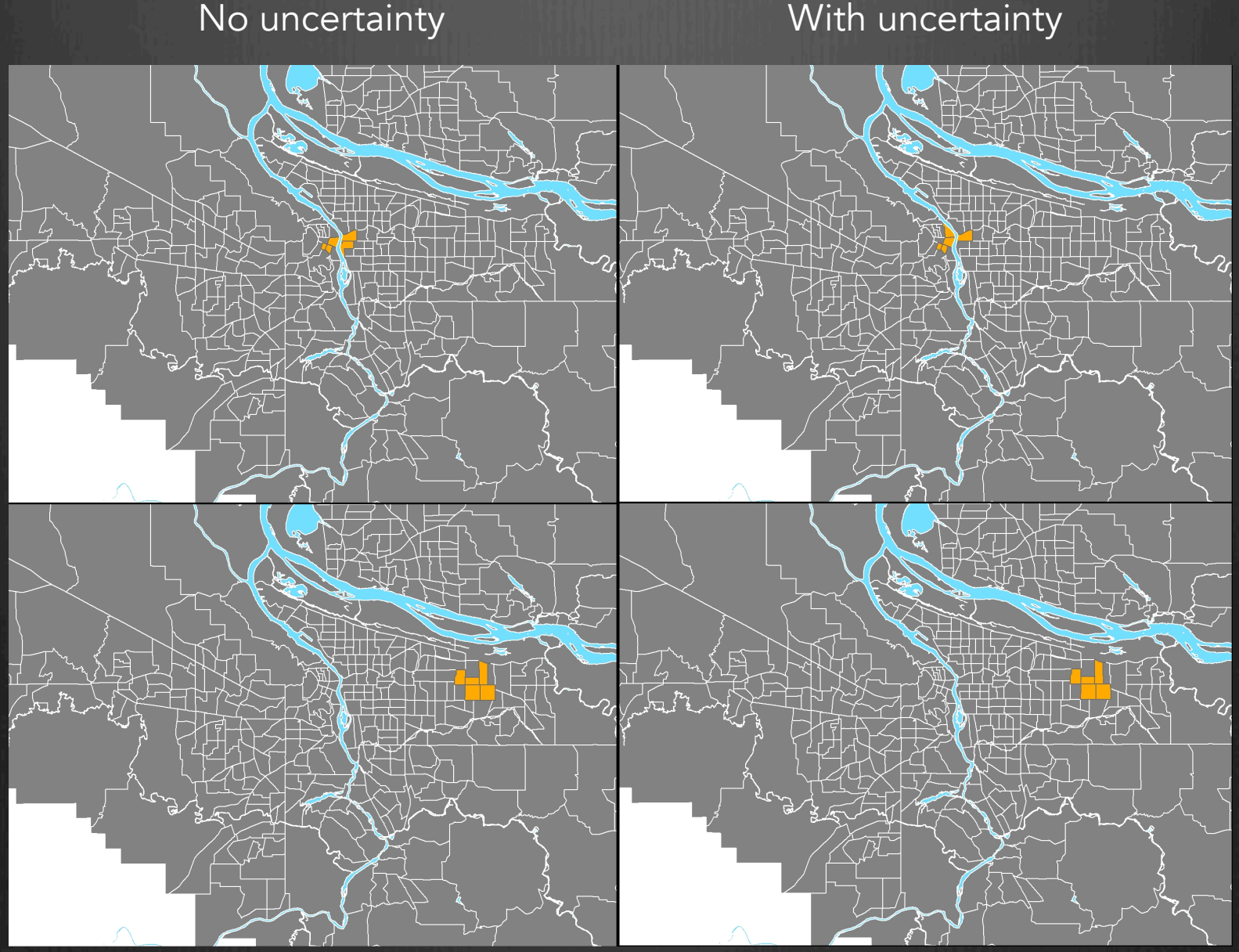
More decisions were different between disguised Portland and Portland when uncertainty is represented (67%) than when it is not (53%).

Map patterns are the same in this comparison
→ also potential evidence for an interaction between place-based knowledge and uncertainty information in the decision-making process.

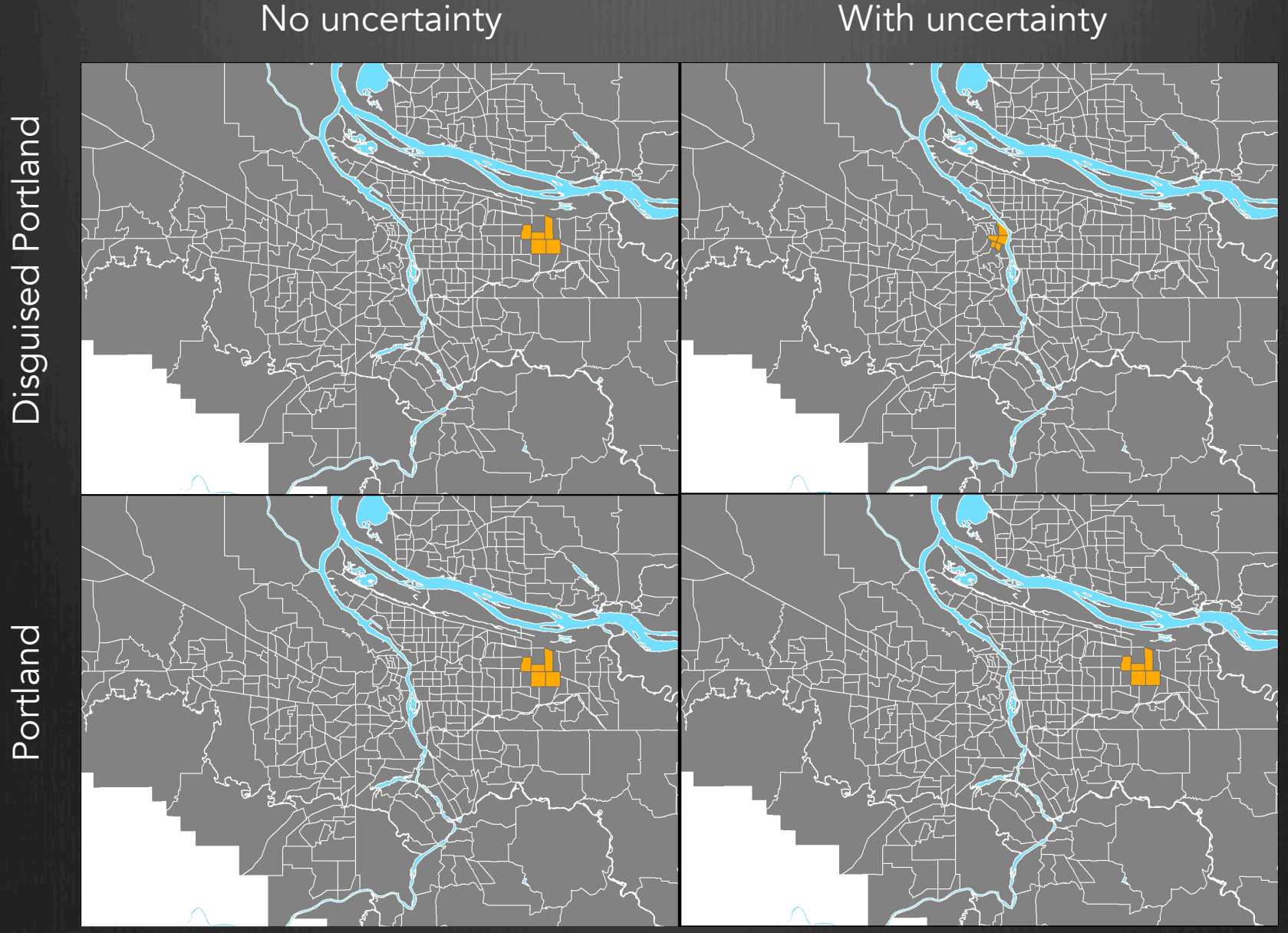
All participants



P55


No uncertainty
With uncertainty
Disguised Portland
Portland

P44


No uncertainty
With uncertainty
Disguised Portland
Portland

Does the particular visual representation used for uncertainty matter?

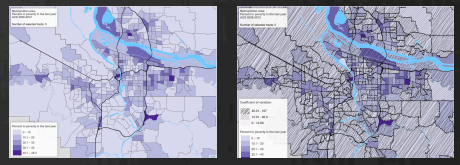
Overall (across both 'cities'), planners changed their decisions equally often in the sketchy and stoplight conditions (40%), but they changed it more often in the MOE condition (46%) than the CV condition (32%).

Condition	Decision did not change	Uncertainty changed decision
Sketchy CV	7	7
Sketchy MOE	3	10
Stoplight CV	12	2
Stoplight MOE	6	8
Total	28	27

Disguised Portland

Condition	Decision did not change	Uncertainty changed decision
Sketchy CV	12	2
Sketchy MOE	10	3
Stoplight CV	7	7
Stoplight MOE	10	4
Total	39	16

Portland



Uncertainty seems to be taken into account more often when planners had less place-based knowledge (49% of participants in the fake city versus 29% of participants in Portland).

Conclusions thus far

Planners may be more likely to rely on what they think they know for a place that they know than considered examination of uncertainties, especially if a quick glance at the mapped attribute confirms what they think they know.

→ Might need a way of highlighting (drawing attention to) important uncertainties.

→ We might need different visual representations of uncertainty depending on how much we know about that place.

Comments / Questions?

Acknowledgements

We would like to thank the planners who took part in both Phase 1 and Phase 2 of our study.

This material is based upon work supported by the National Science Foundation under Grant Number 1132008.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.