Adaptive Kernel Density in Demographic Analysis

Richard Lycan
Institute on Aging
Portland State University

Context

- Study for Oregon Department of Transportation on the needs for rural transportation for the elderly and disabled by the Center for Transportation Studies at Portland State University
- Need to calculate disability rates for census ***urban areas*** and ***clusters*** and the residual ***rural areas*** of Oregon counties.
- Disability rate data are published in the tables from the American Community Survey. The survey asks about several types of disability such as hearing or problems getting around in the residence.

Data for some urban clusters and rural parts of counties are compromised due to sampling variability.

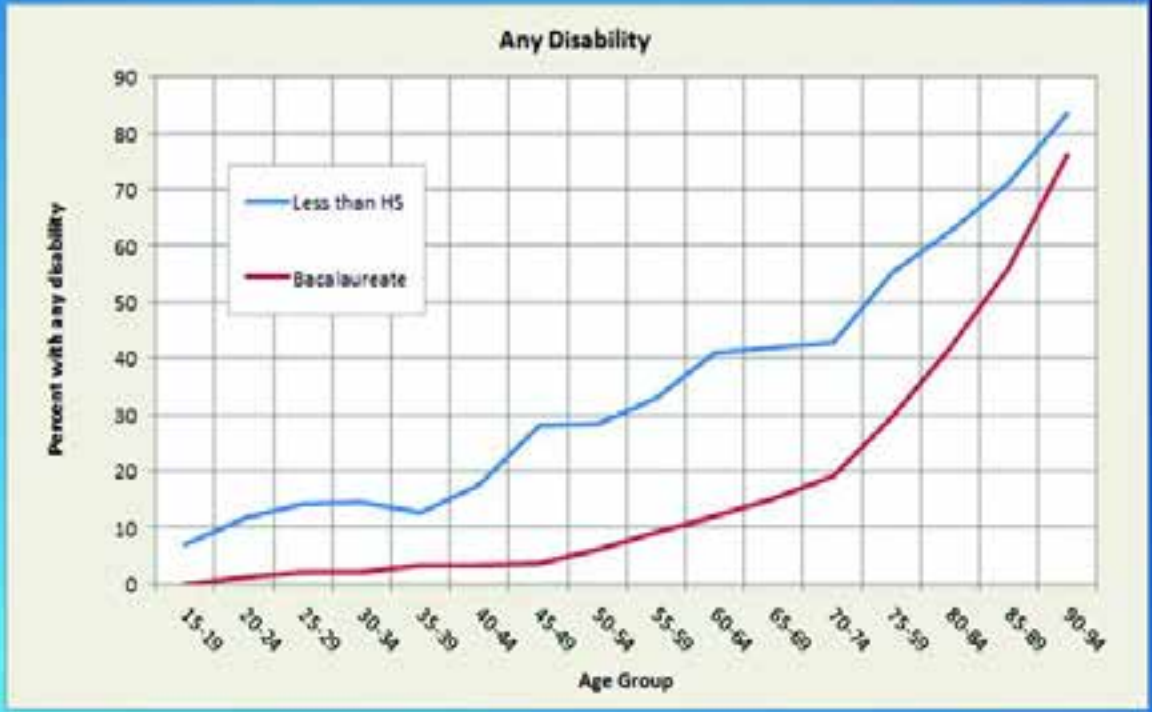
Some characteristics of disability rates

- Partly to escape the problems of small sample size we use the data for ***any disability***. Here are the individual level correlations between the six measures.

	Hearing	Vision	Dressing	Cognitive	Ambulatory	Ind Living
Hearing	1.000	0.230	0.161	0.204	0.222	0.206
Vision	0.230	1.000	0.211	0.244	0.235	0.295
Dressing	0.161	0.211	1.000	0.472	0.504	0.625
Cognitive	0.204	0.244	0.472	1.000	0.386	0.542
Ind Living	0.222	0.235	0.504	0.386	1.000	0.576
Ind Living	0.206	0.295	0.625	0.542	0.576	1.000

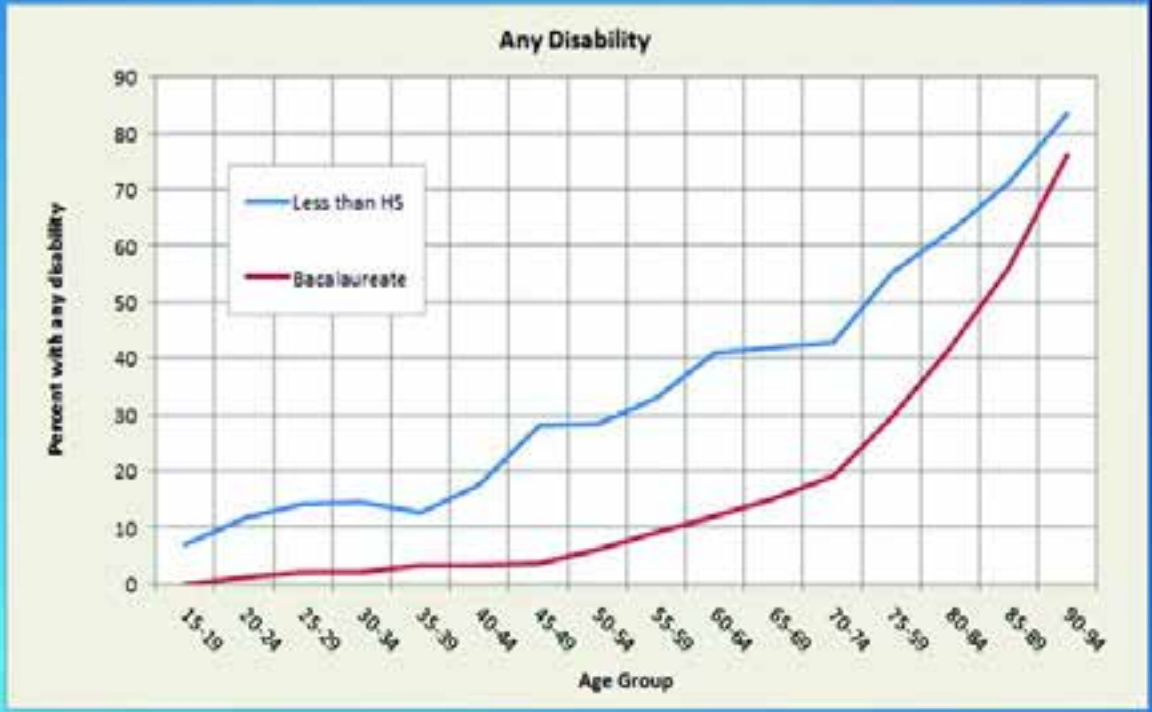
Some characteristics of disability rates

- Partly to escape the problems of small sample size we use the data for ***any disability***. Here are the individual level correlations between the six measures.
- Disability rates rise with age. We use rates for the population age 65 plus but the rates for the very old are high.



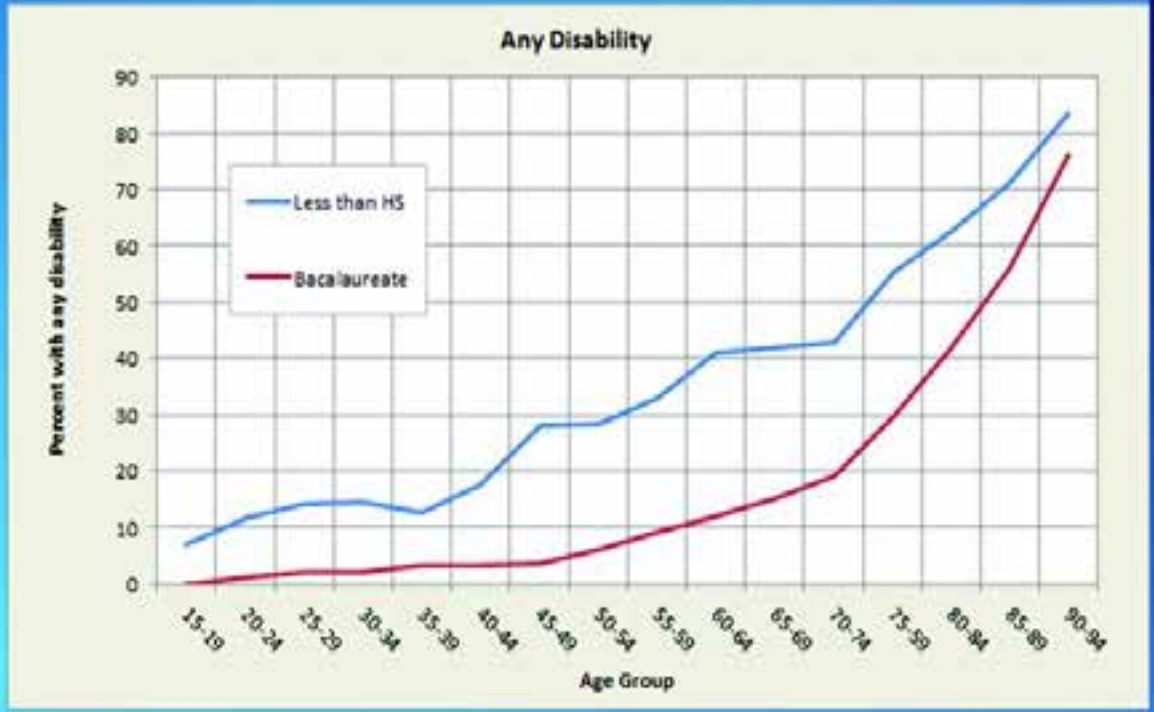
Some characteristics of disability rates

- Partly to escape the problems of small sample size we use the data for ***any disability***. Here are the individual level correlations between the six measures.
- Disability rates rise with age. We use rates for the population age 65 plus but the rates for the very old are high.



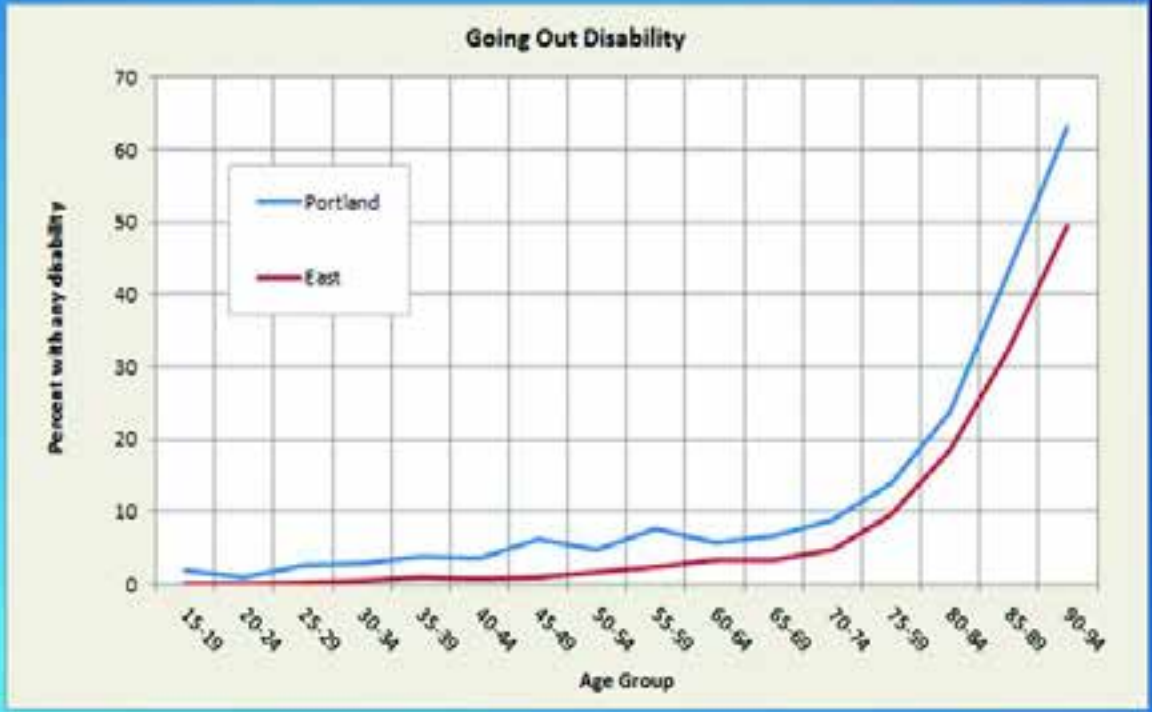
Some characteristics of disability rates

- Partly to escape the problems of small sample size we use the data for ***any disability***. Here are the individual level correlations between the six measures.
- Disability rates rise with age. We use rates for the population age 65 plus but the rates for the very old are high.
- Persons with higher educational levels respond less frequently to the disability questions and incur disabilities later in life.



Some characteristics of disability rates

- Partly to escape the problems of small sample size we use the data for ***any disability***. Here are the individual level correlations between the six measures.
- Disability rates rise with age. We use rates for the population age 65 plus but the rates for the very old are high.
- Persons with higher educational levels respond less frequently to the disability questions and incur disabilities later in life.
- Disability rates are slightly higher for persons living in urban than rural settings.

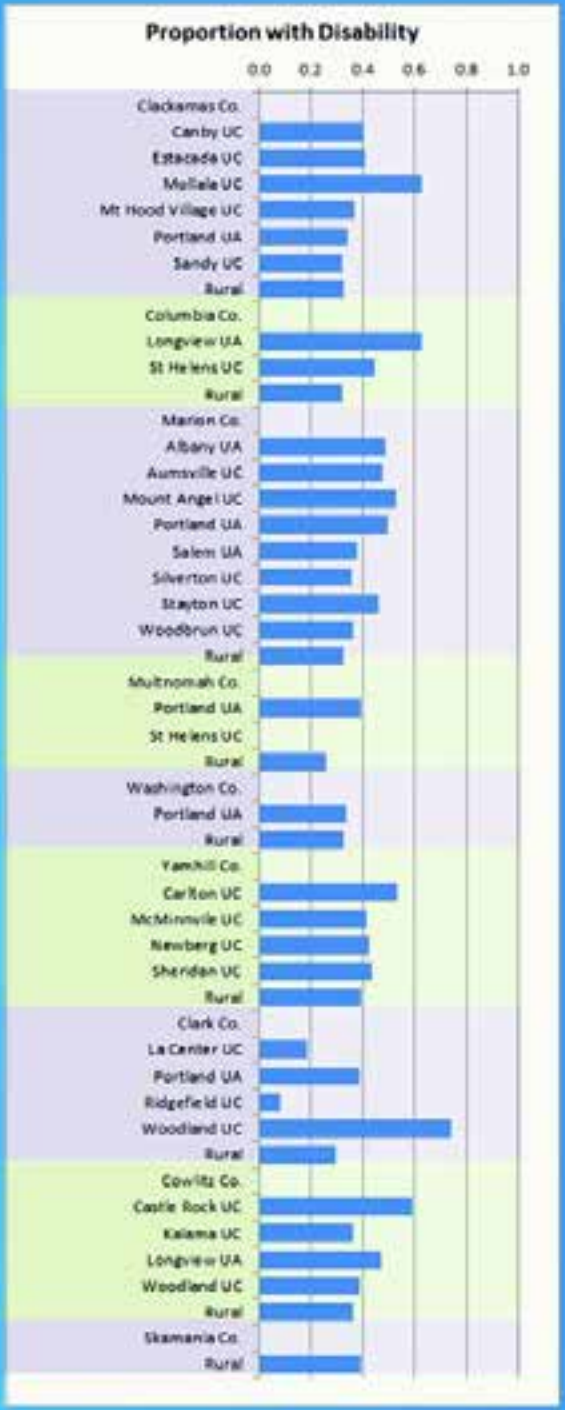


Disability rates for the UA/UC/Rural geography

- The chart to the right shows the proportion with any disability for the urban areas and clusters and the remaining county rural areas for a seven county area surrounding Portland, Oregon.
- The average value is around 0.4 but ranges from 0.2 for the Ridgefield urban cluster to 0.8 for the nearby Woodland urban cluster.

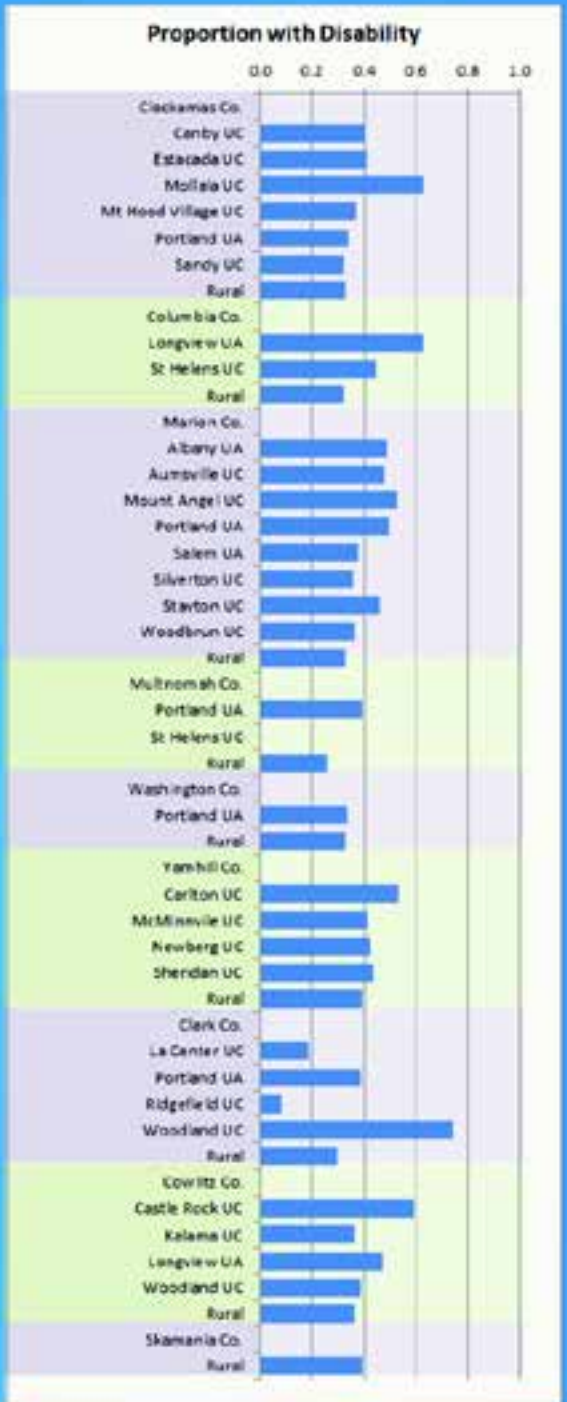
Disability rates for the UA/UC/Rural geography

- The chart to the right shows the proportion with any disability for the urban areas and clusters and the remaining county rural areas for a seven county area surrounding Portland, Oregon.



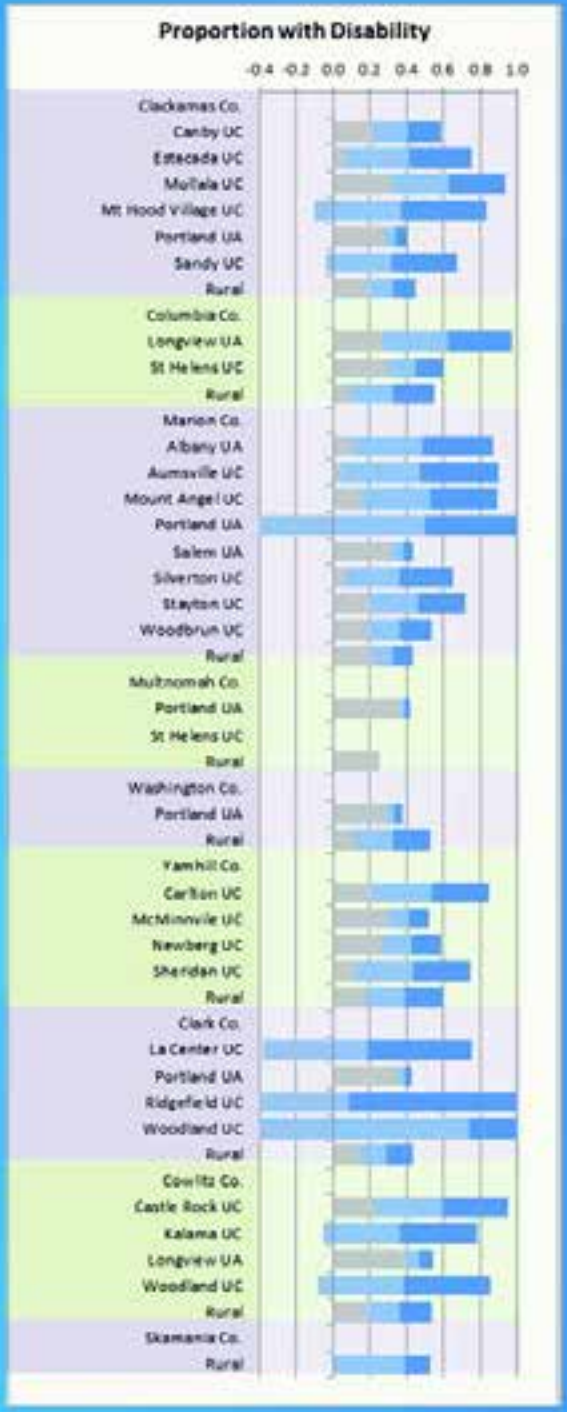
Disability rates for the UA/UC/Rural geography

- The chart to the right shows the proportion with any disability for the urban areas and clusters and the remaining county rural areas for a seven county area surrounding Portland, Oregon.
- The average value is around 0.4 but ranges from 0.2 for the Ridgefield urban cluster to 0.8 for the nearby Woodland urban cluster.
- In general disability rates are higher for the urban areas and clusters than for the surrounding rural areas.
- But when we look at the range of error above and blow the estimate see that the values are not to be trusted.



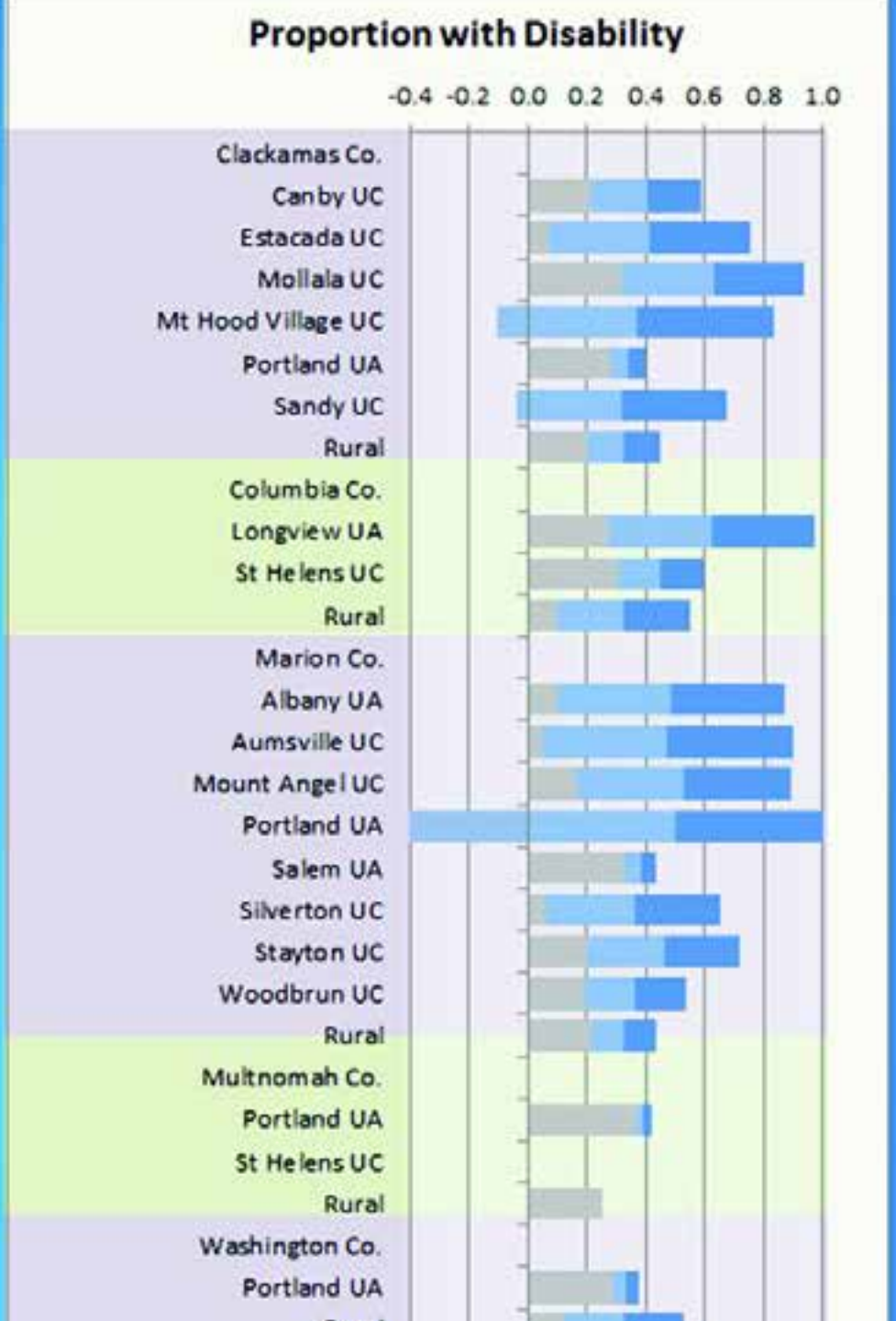
Disability rates for the UA/UC/Rural geography

- The chart to the right shows the proportion with any disability for the urban areas and clusters and the remaining county rural areas for a seven county area surrounding Portland, Oregon.
- The average value is around 0.4 but ranges from 0.2 for the Ridgefield urban cluster to 0.8 for the nearby Woodland urban cluster.
- In general disability rates are higher for the urban areas and clusters than for the surrounding rural areas.
- But when we look at the range of error above and blow the estimate see that the values are not to be trusted.



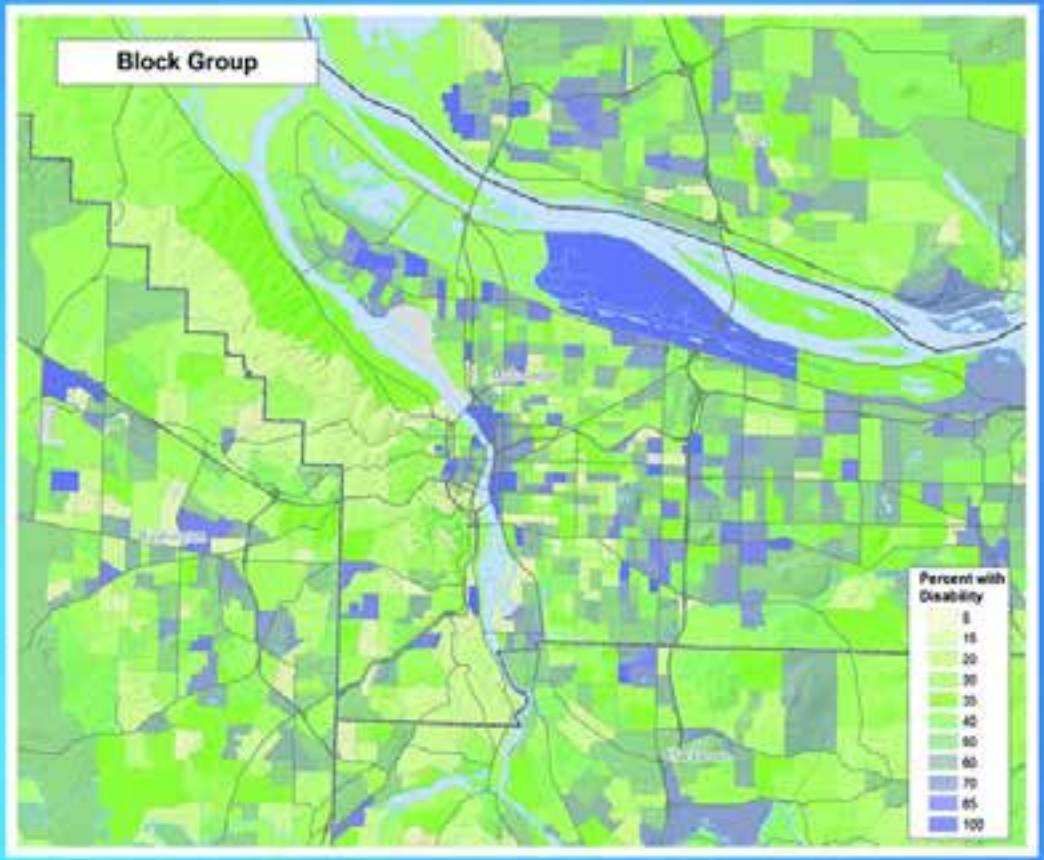
Disability rates for the UA/UC/Rural geography

- The chart to the right shows the proportion with any disability for the urban areas and clusters and the remaining county rural areas for a seven county area surrounding Portland, Oregon.
- The average value is around 0.4 but ranges from 0.2 for the Ridgefield urban cluster to 0.8 for the nearby Woodland urban cluster.
- In general disability rates are higher for the urban areas and clusters than for the surrounding rural areas.
- But when we look at the range of error above and blow the estimate see that the values are not to be trusted.



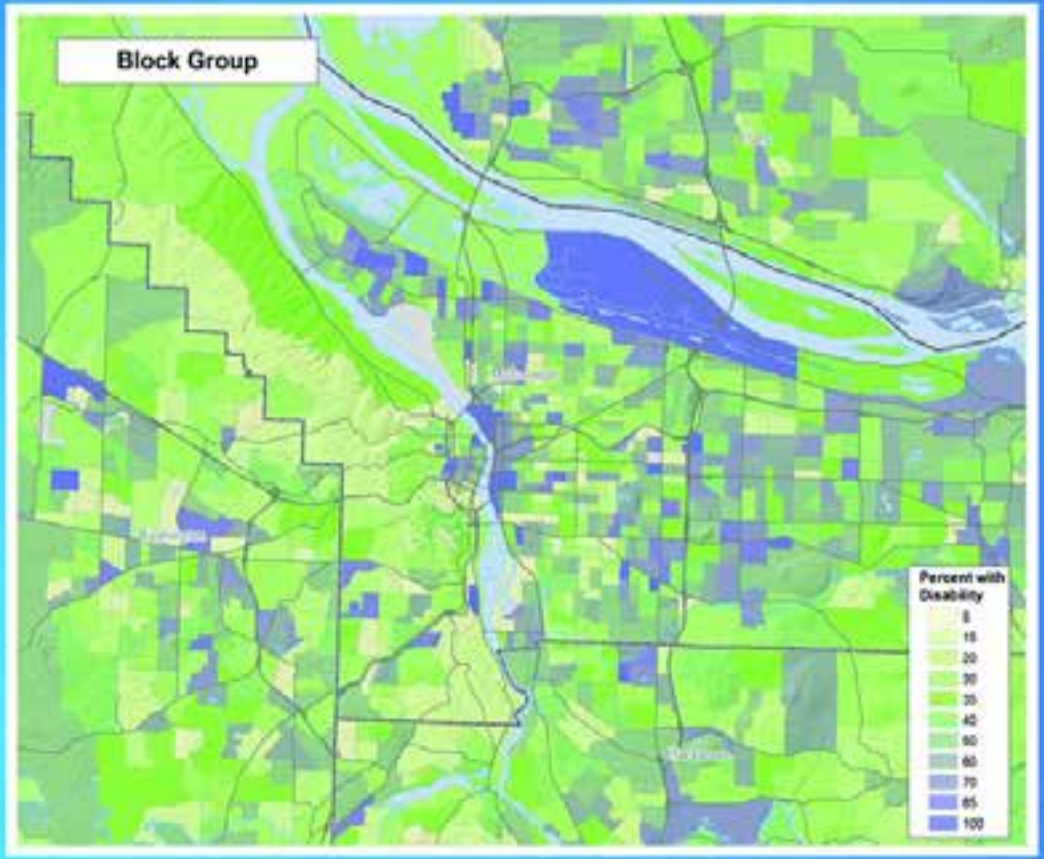
Disability rates for block groups, tracts, and PUMAs

- Block groups, census tracts, and PUMAs show disability rates at varying levels of detail.



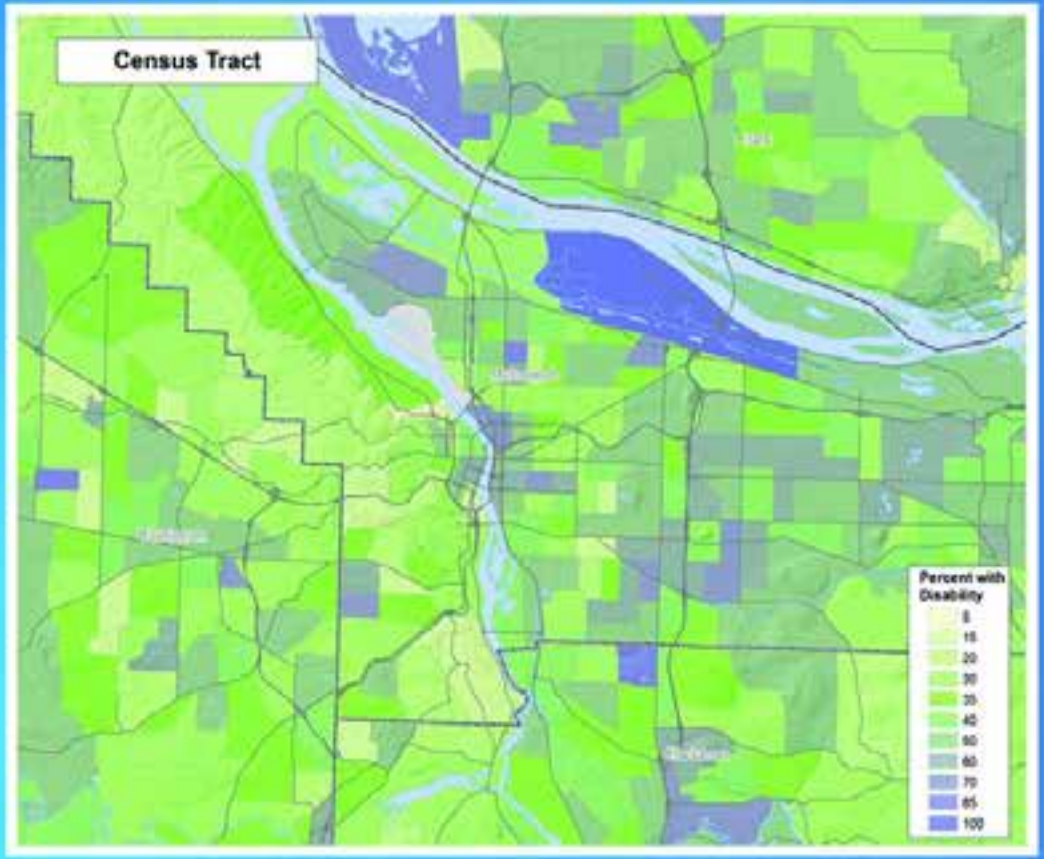
Disability rates for block groups, tracts, and PUMAs

- Block groups, census tracts, and PUMAs show disability rates at varying levels of detail.
- At the block group level we see interesting detail, but know much of the variation is due to sampling error.



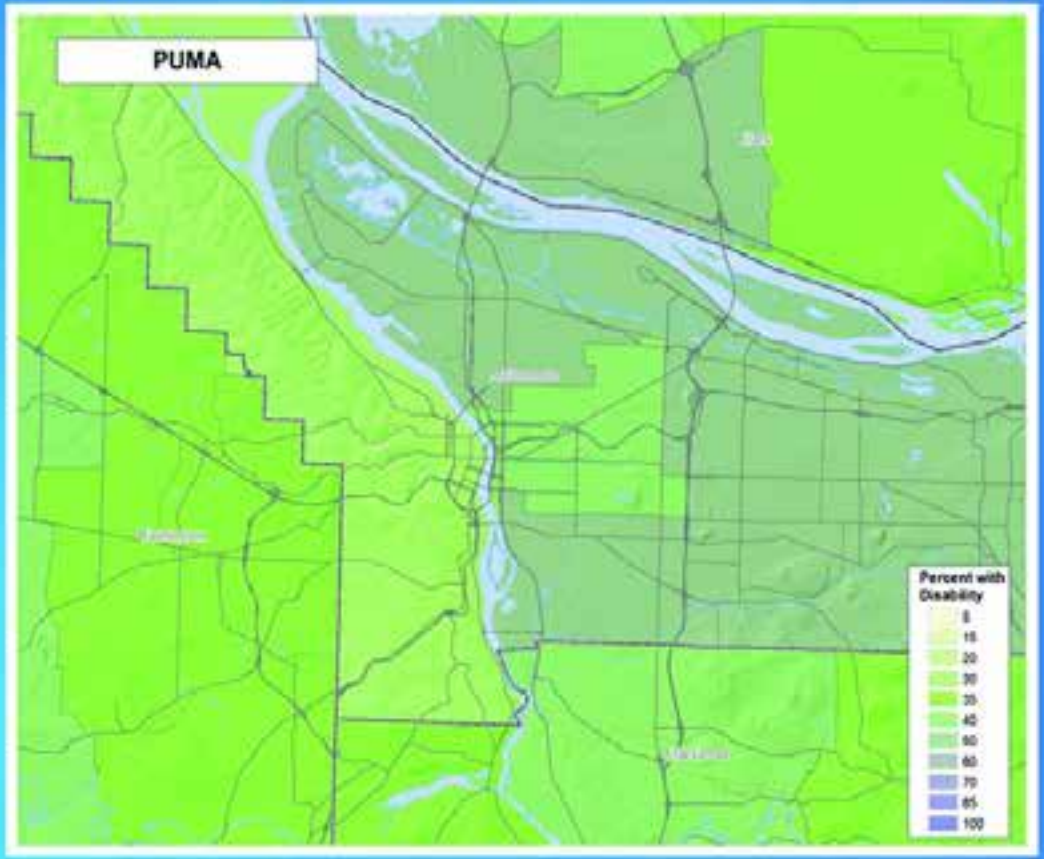
Disability rates for block groups, tracts, and PUMAs

- Block groups, census tracts, and PUMAs show disability rates at varying levels of detail.
- At the block group level we see interesting detail, but know much of the variation is due to sampling error.
- Census tract data averages out some of the sampling error but also reduces the amount of geographical detail.



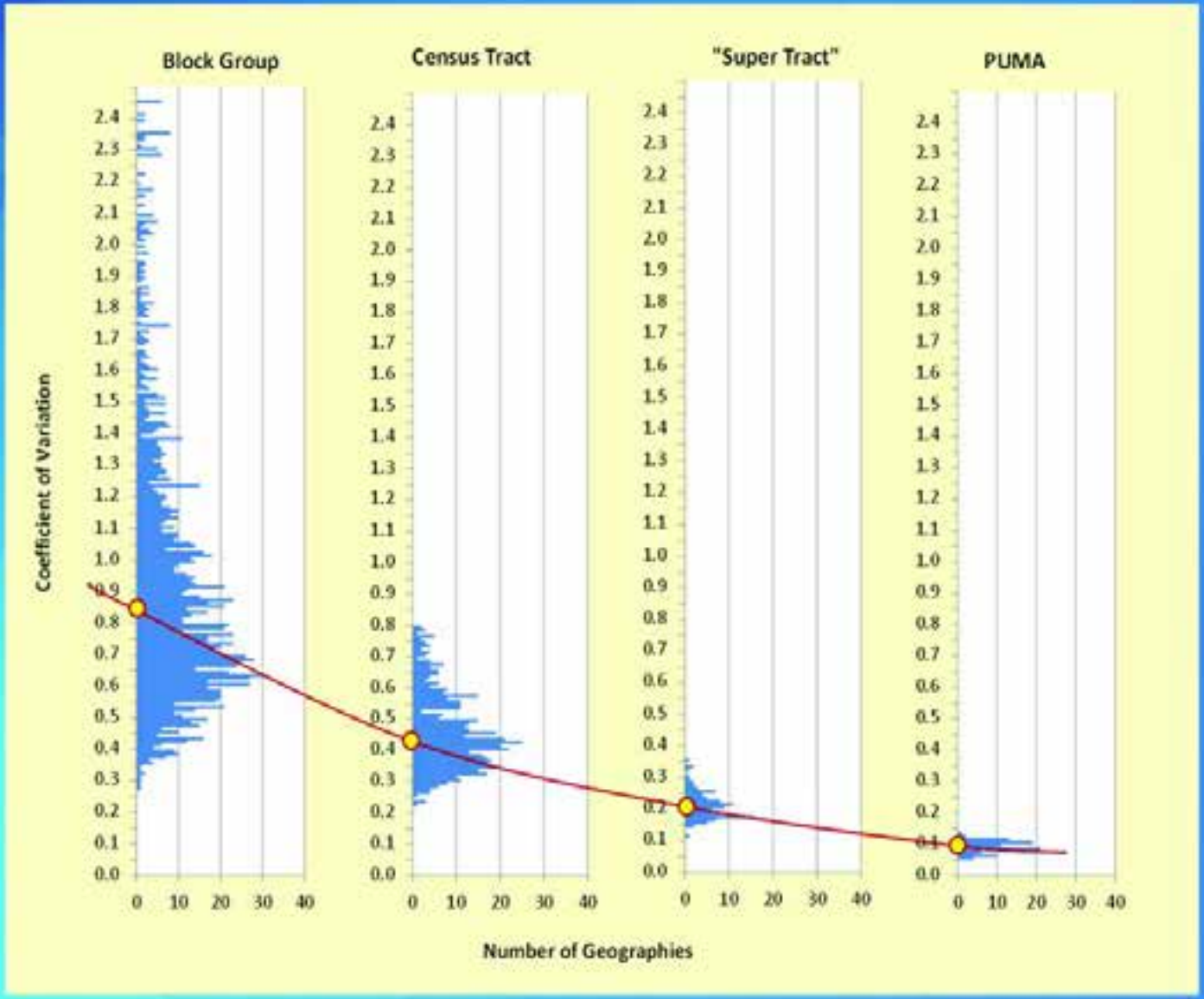
Disability rates for block groups, tracts, and PUMAs

- Block groups, census tracts, and PUMAs show disability rates at varying levels of detail.
- At the block group level we see interesting detail, but know much of the variation is due to sampling error.
- Census tract data averages out some of the sampling error but also reduces the amount of geographical detail.
- Going all the way to PUMA geography, areas with at least 60,000 population, provides stable results, but little geographical detail.



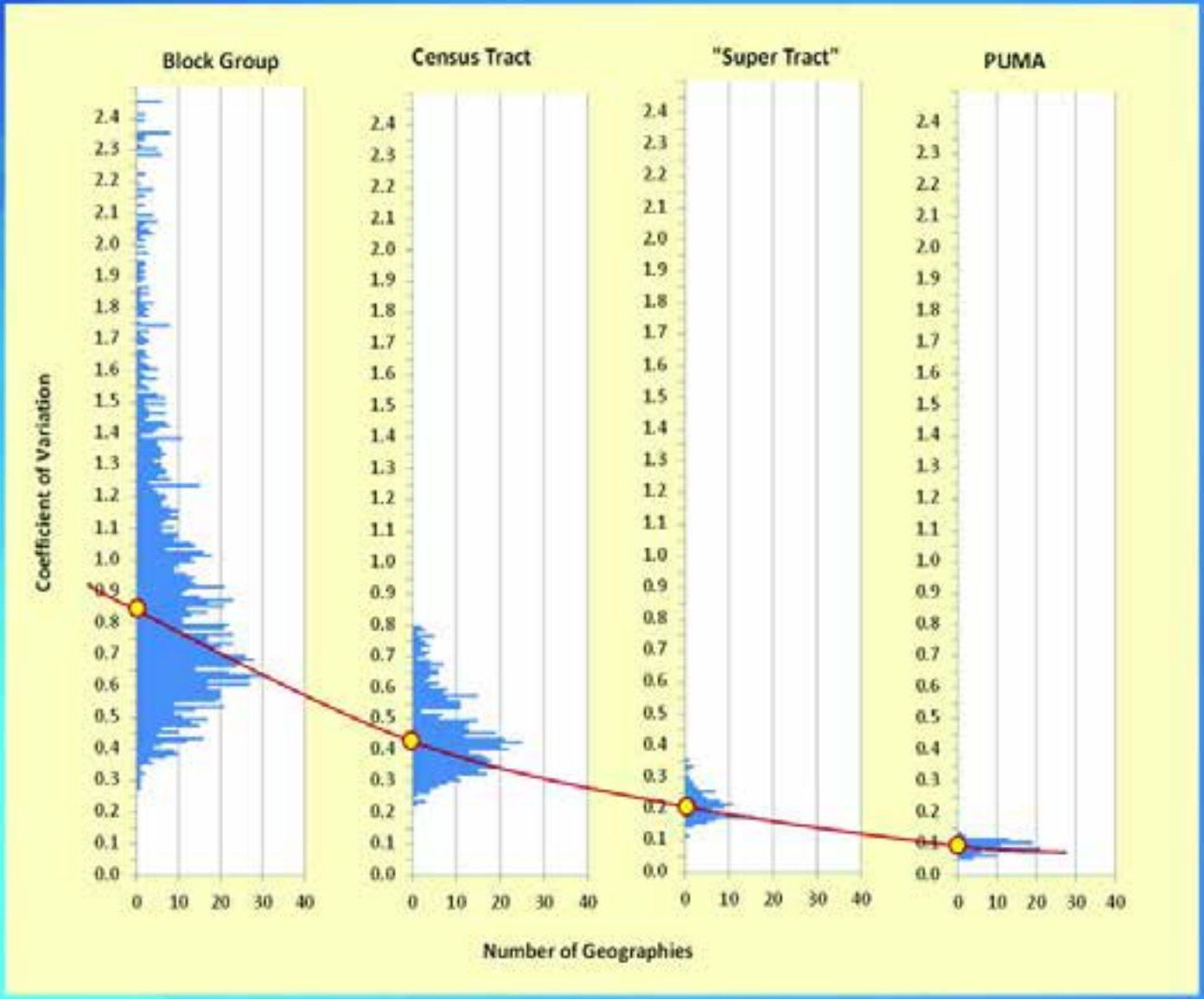
Quantifying the error

- The graphs to the right show coefficient of variation (CV) – the ratio of the MOE to the estimate.



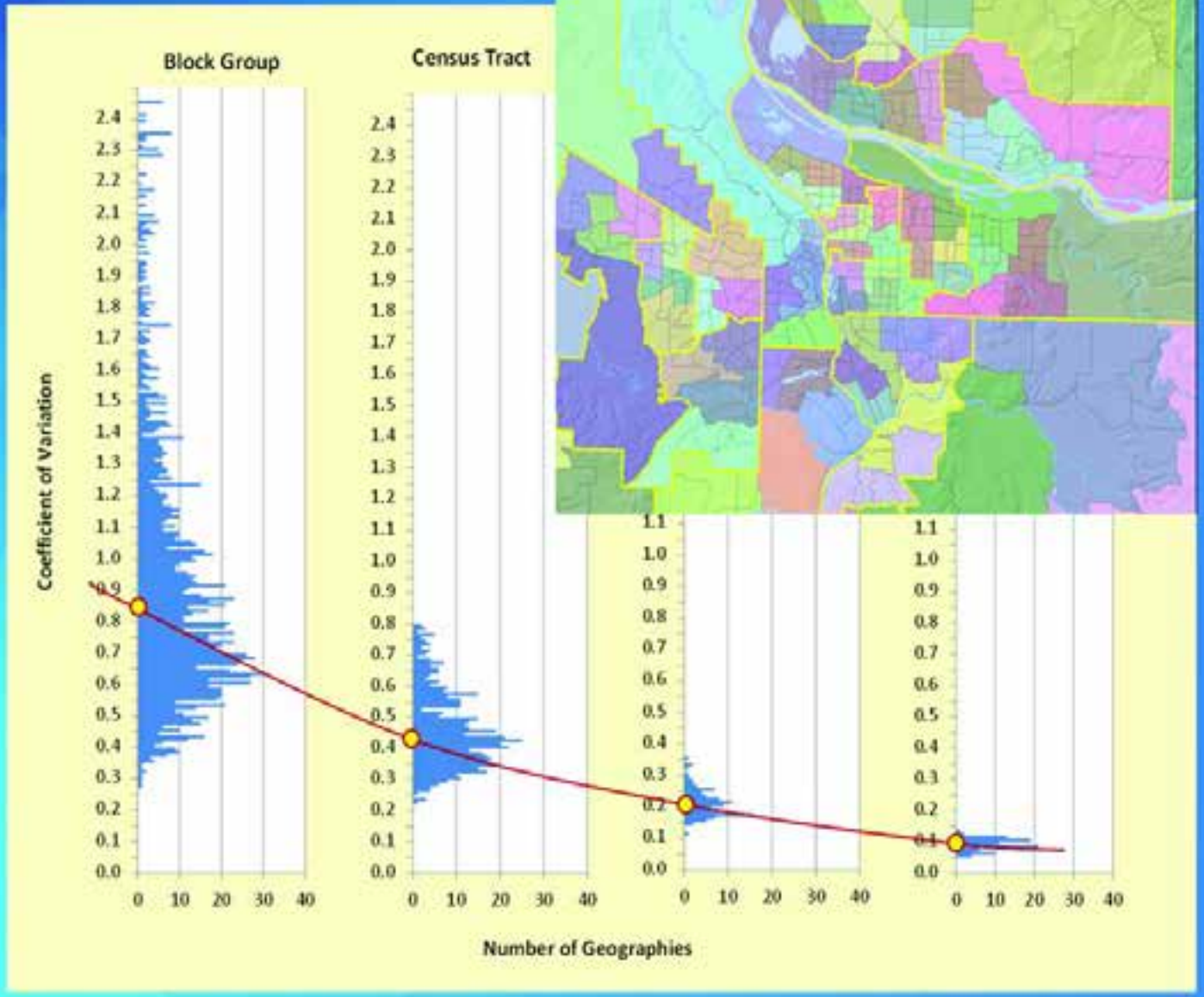
Quantifying the error

- The graphs to the right show coefficient of variation (CV) – the ratio of the MOE to the estimate.
- The CV declines from huge to acceptable as the size of the geography increases.



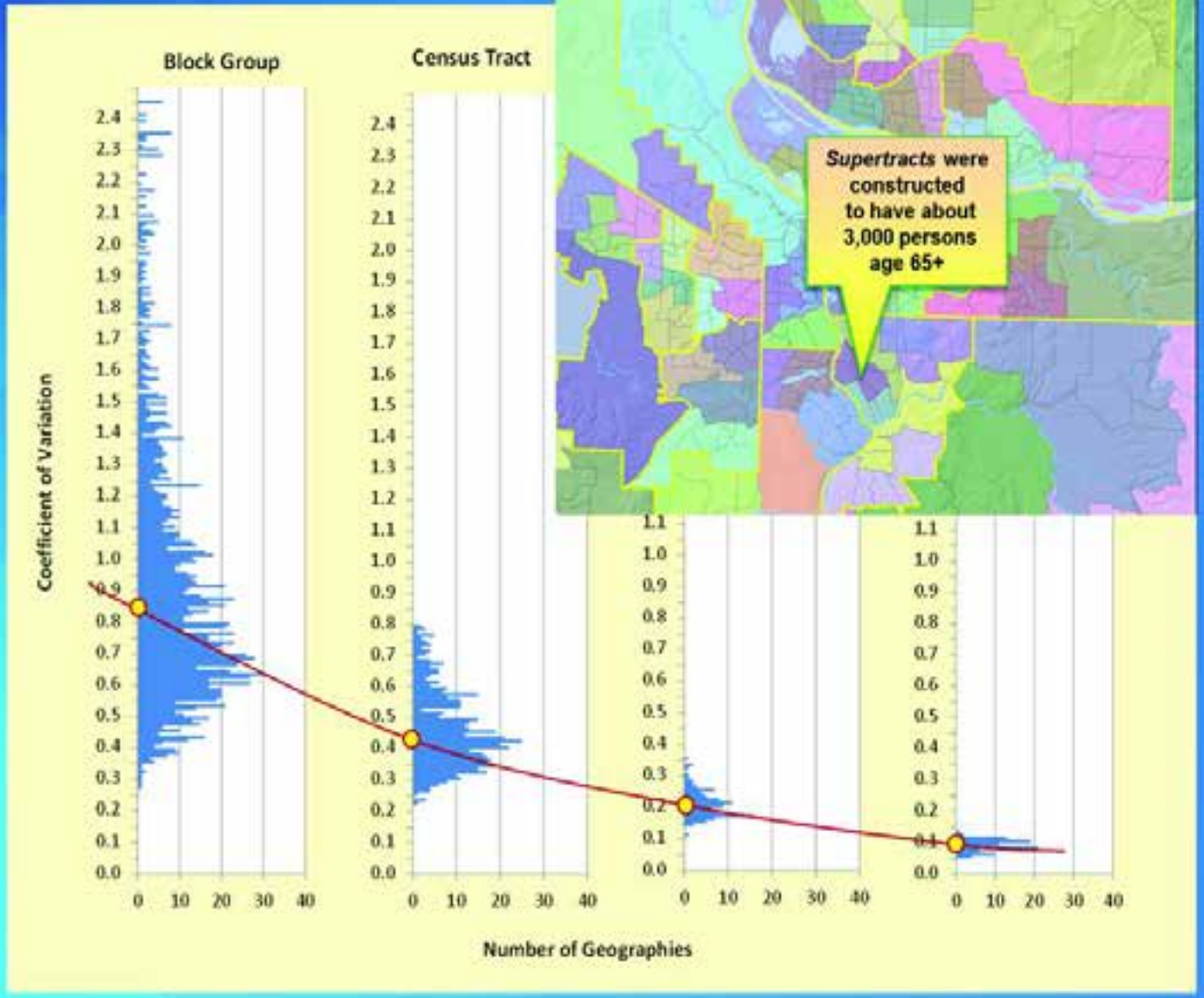
Quantifying the error

- The graphs to the right show coefficient of variation (CV) – the ratio of the MOE to the estimate.
- The CV declines from huge to acceptable as the size of the geography increases.
- We include an intermediate geography referred to as a "super tract" in between the tract and PUMA.



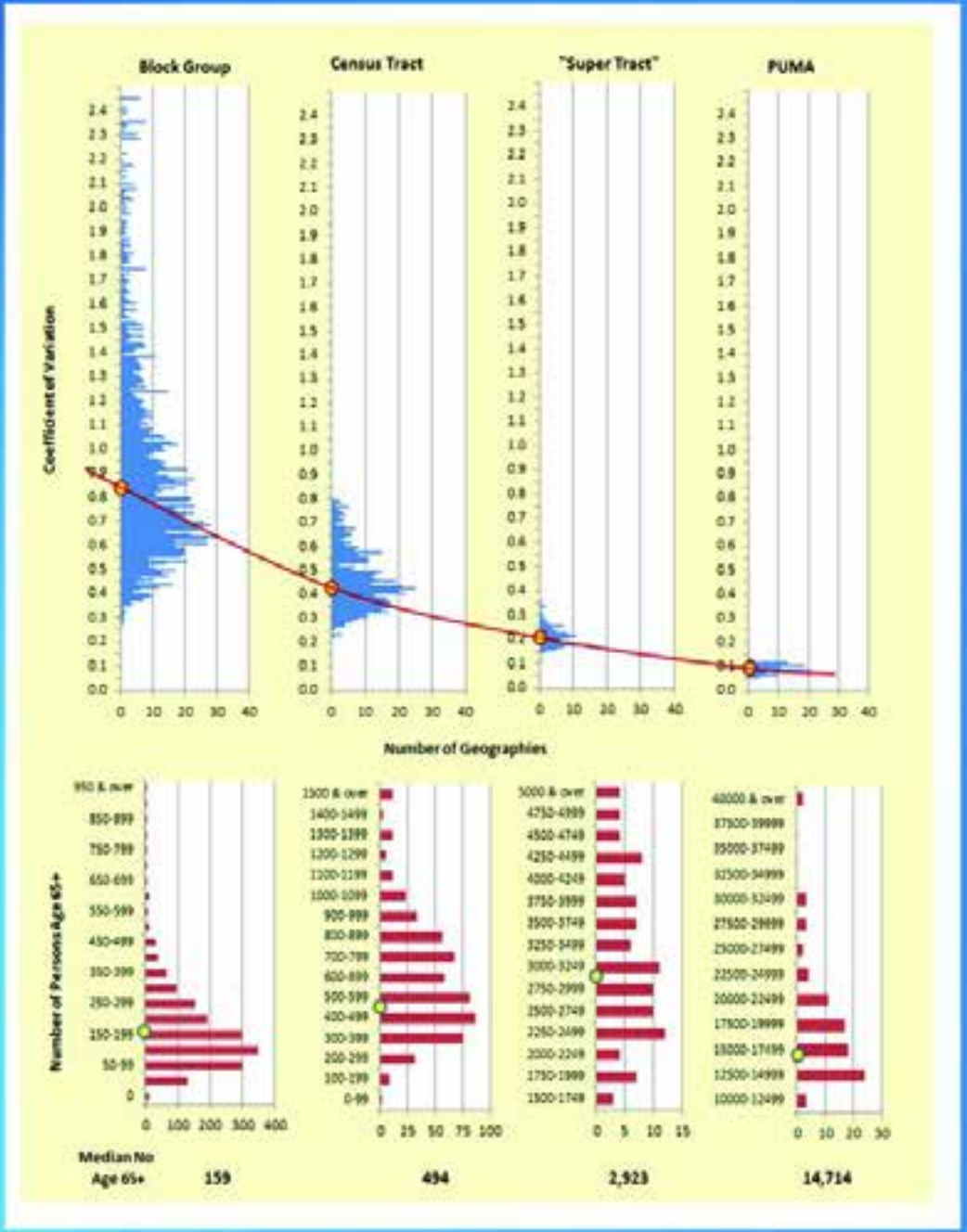
Quantifying the error

- The graphs to the right show coefficient of variation (CV) – the ratio of the MOE to the estimate.
- The CV declines from huge to acceptable as the size of the geography increases.
- We include an intermediate geography referred to as a "super tract" in between the tract and PUMA.



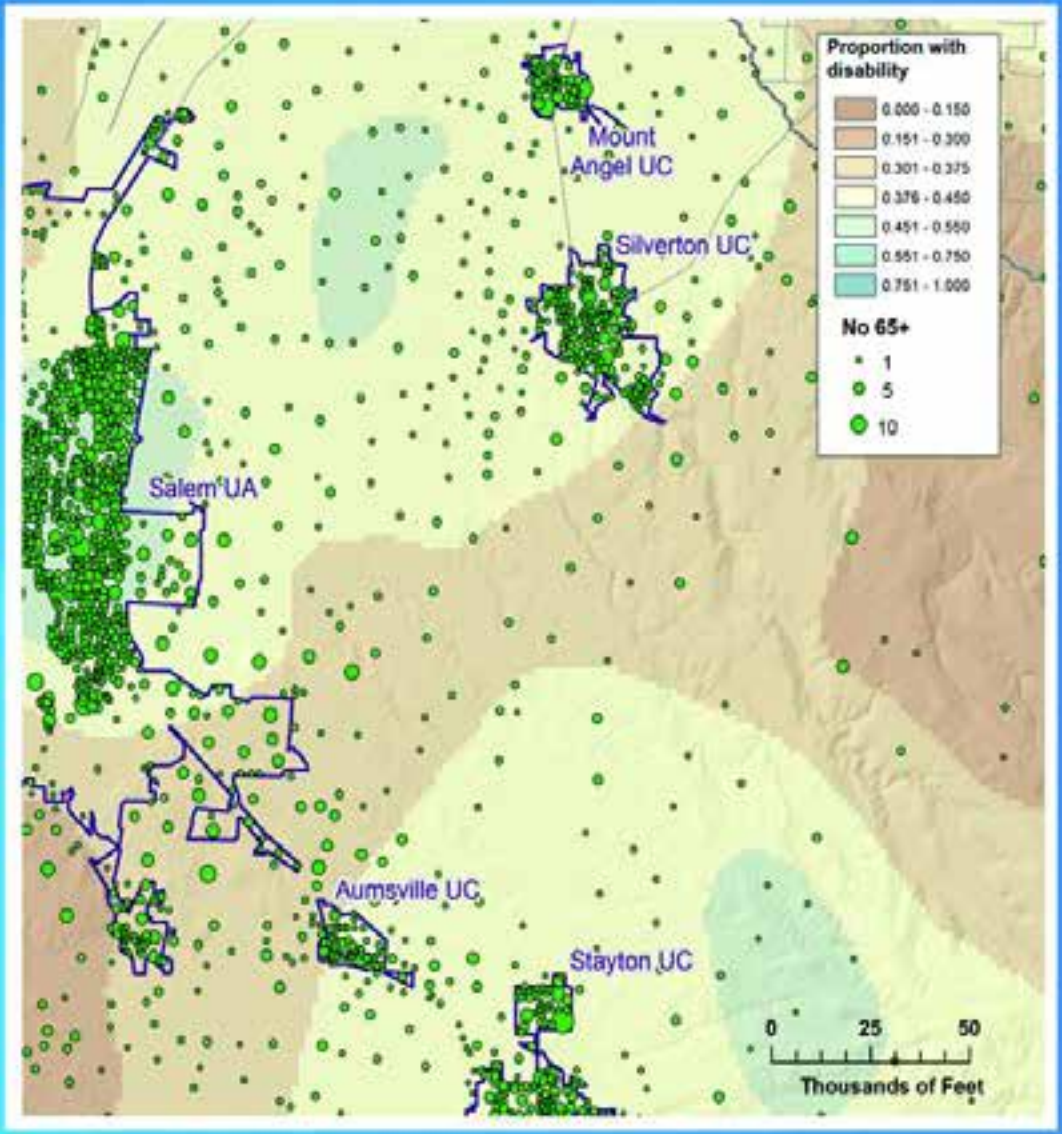
Sample size

- In the following grid based analyses we use data for geographies about the size of the *supertract*.
- The median number of age 65 plus for the *supertract* is 2,923 persons.
- This is about size times the size of the median tract and 1/5th the size of the median PUMA.
- One can argue that it is about midway between the census tract and the PUMA with respect to sample size.



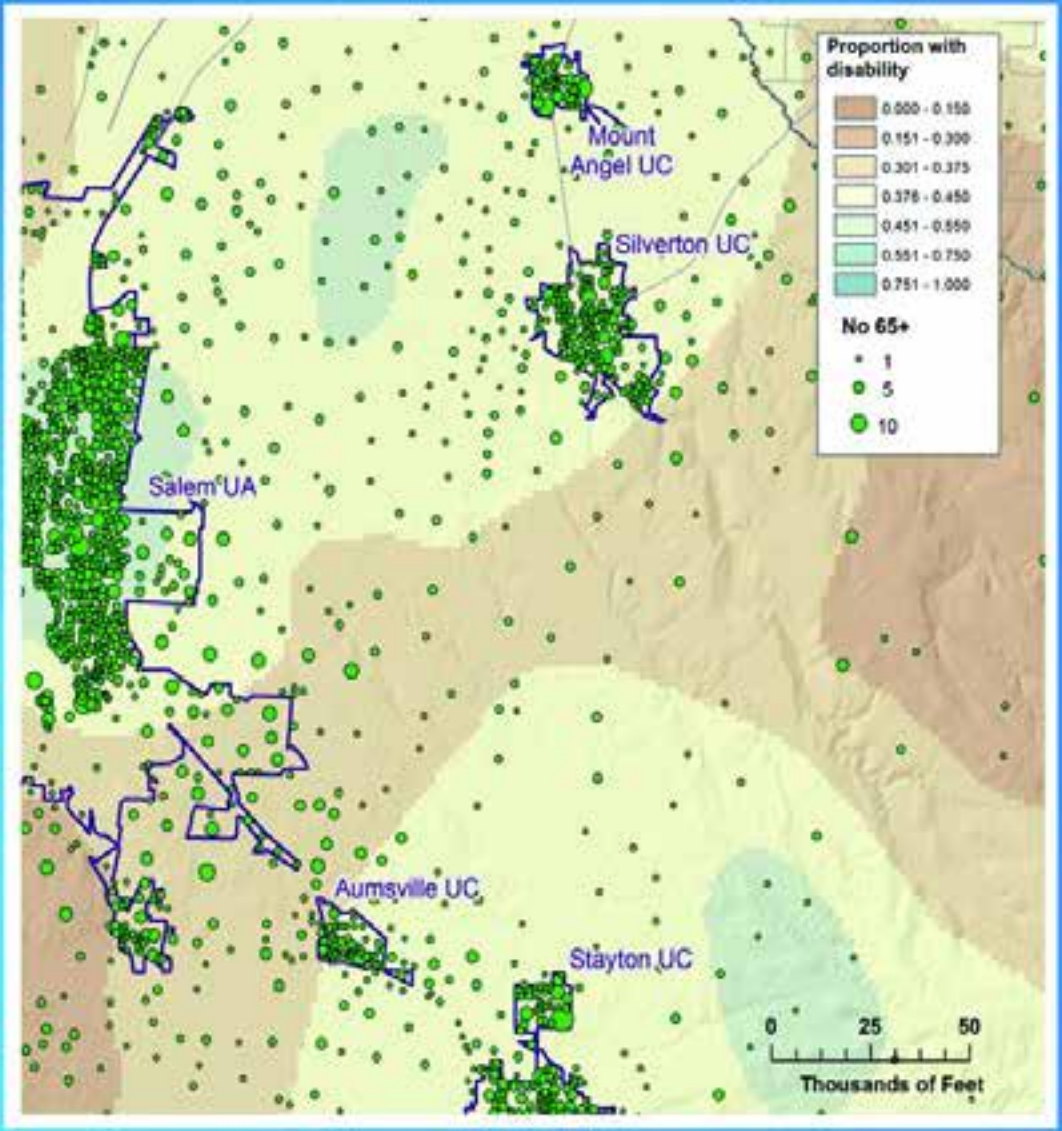
How density grids are used to calculate disability rates

- The map to the right overlays the number of persons age 65+ over proportion with a disability calculated using grid mapping. *Spatial Analyst* extract values to points is used to add the rate to the point file.
- The UA/UC/Rural class also is added to the point file
- Disability rate for UA/UC/Rural is summarized using ArcMap or Excel.



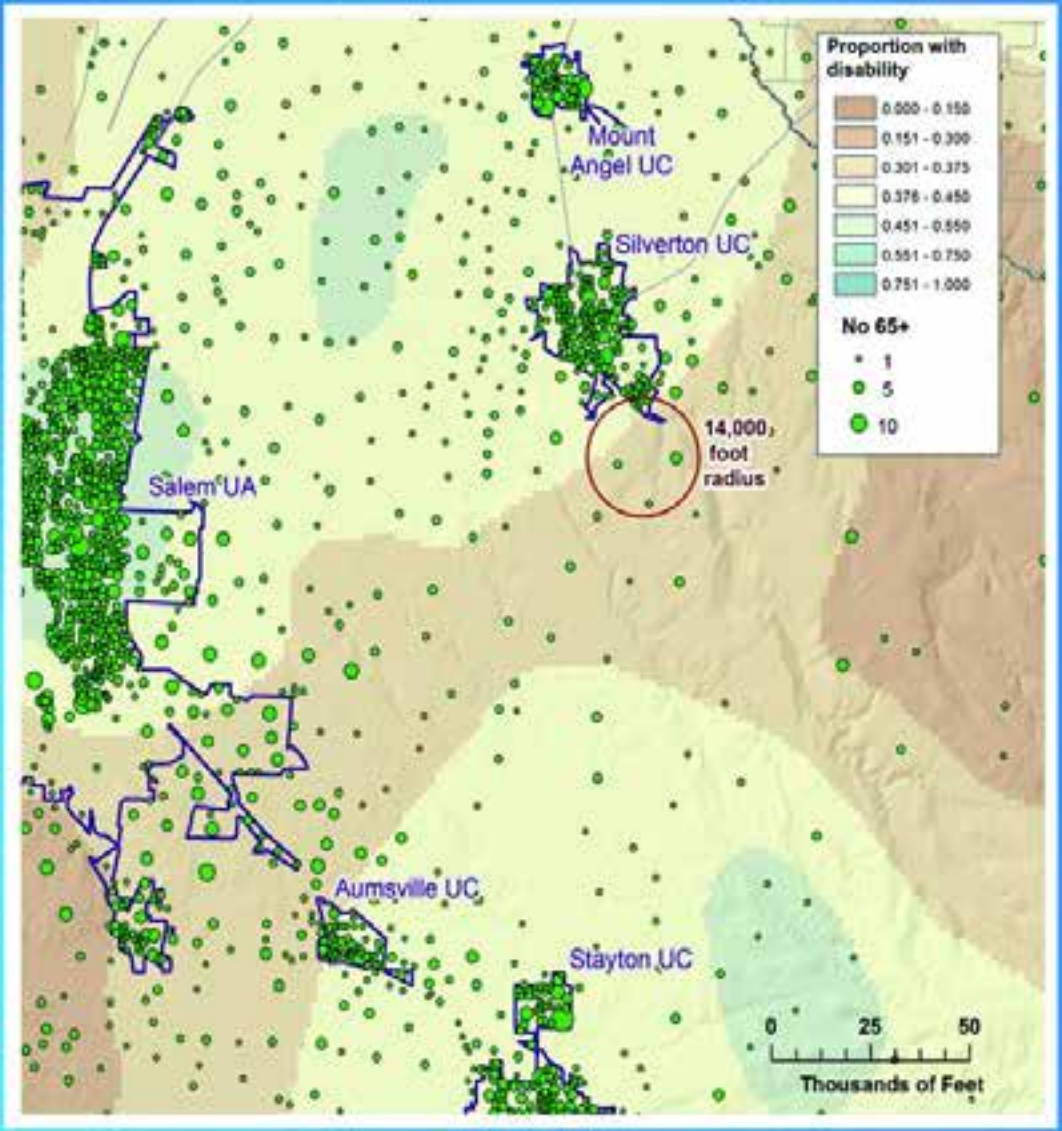
The type of grid mapping affects the results

- A fixed distance grid and an adaptive grid are used to calculate disability rates.



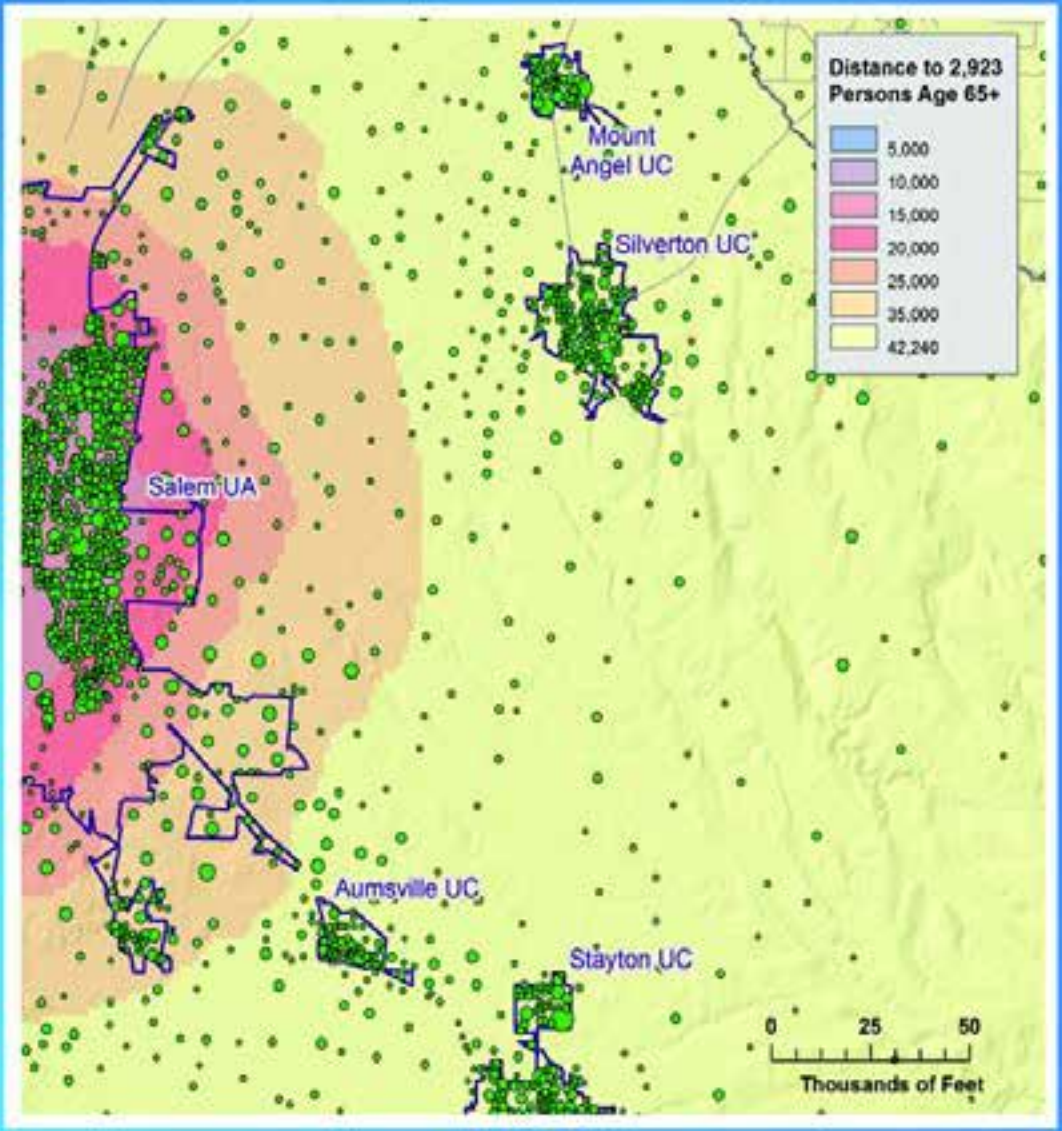
The type of grid mapping affects the results

- A fixed distance grid and an adaptive grid are used to calculate disability rates.
- With a ***fixed density*** grid all the points within a fixed radius from each grid cell are used to calculate density at that grid location. Here the radius is 14,000 feet.



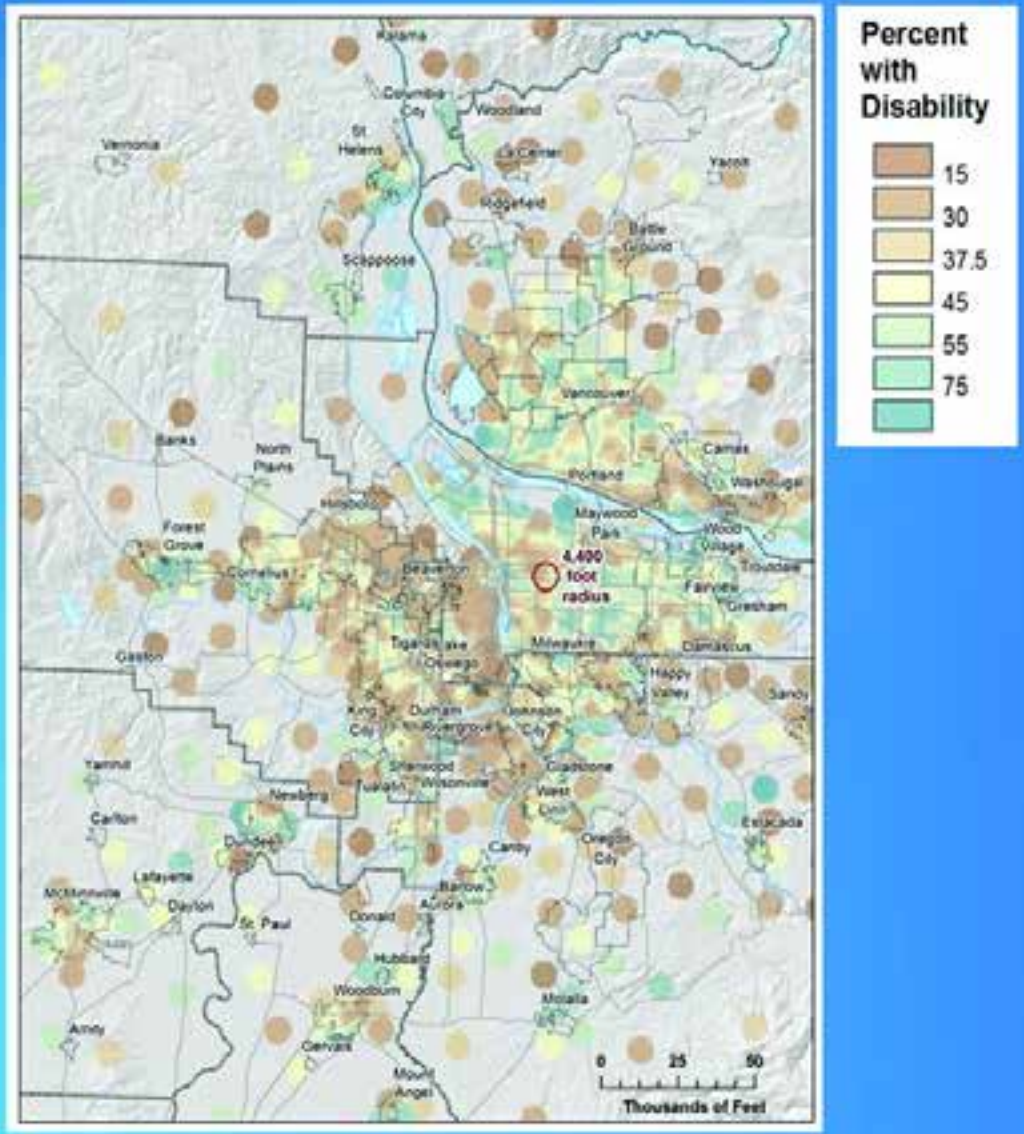
The type of grid mapping affects the results

- A fixed distance grid and an adaptive grid are used to calculate disability rates.
- With a ***fixed density*** grid all the points within a fixed radius from each grid cell are used to calculate density at that grid location. Here the radius is 14,000 feet.
- With an ***adaptive density*** grid the search regions is extended from each grid cell until the required population is included. Here the criteria is 2,923 persons age 65+.



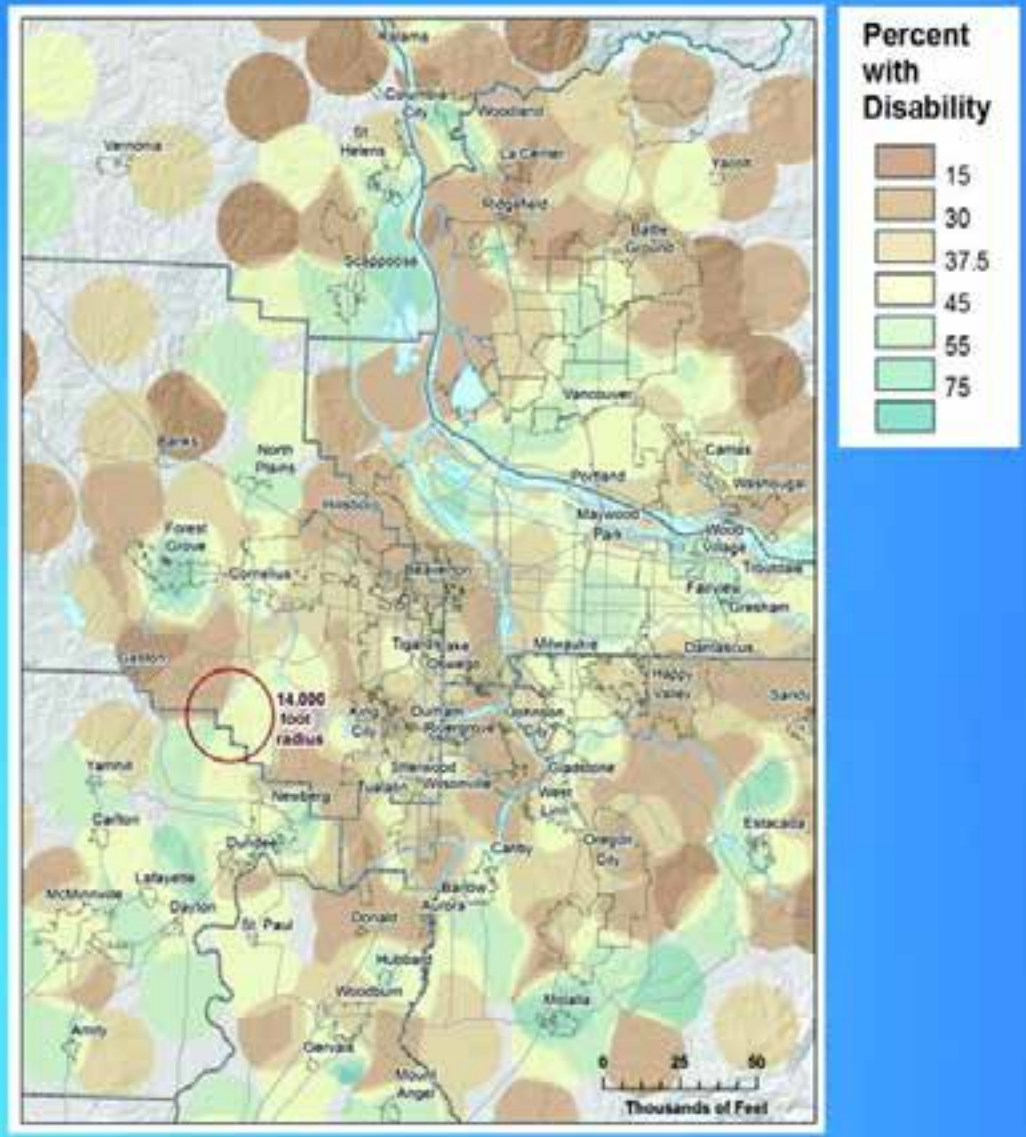
Regional scale comparison

- The fixed radius grid with a 4,400 foot distance provides considerable local detail but only spotty coverage overall.



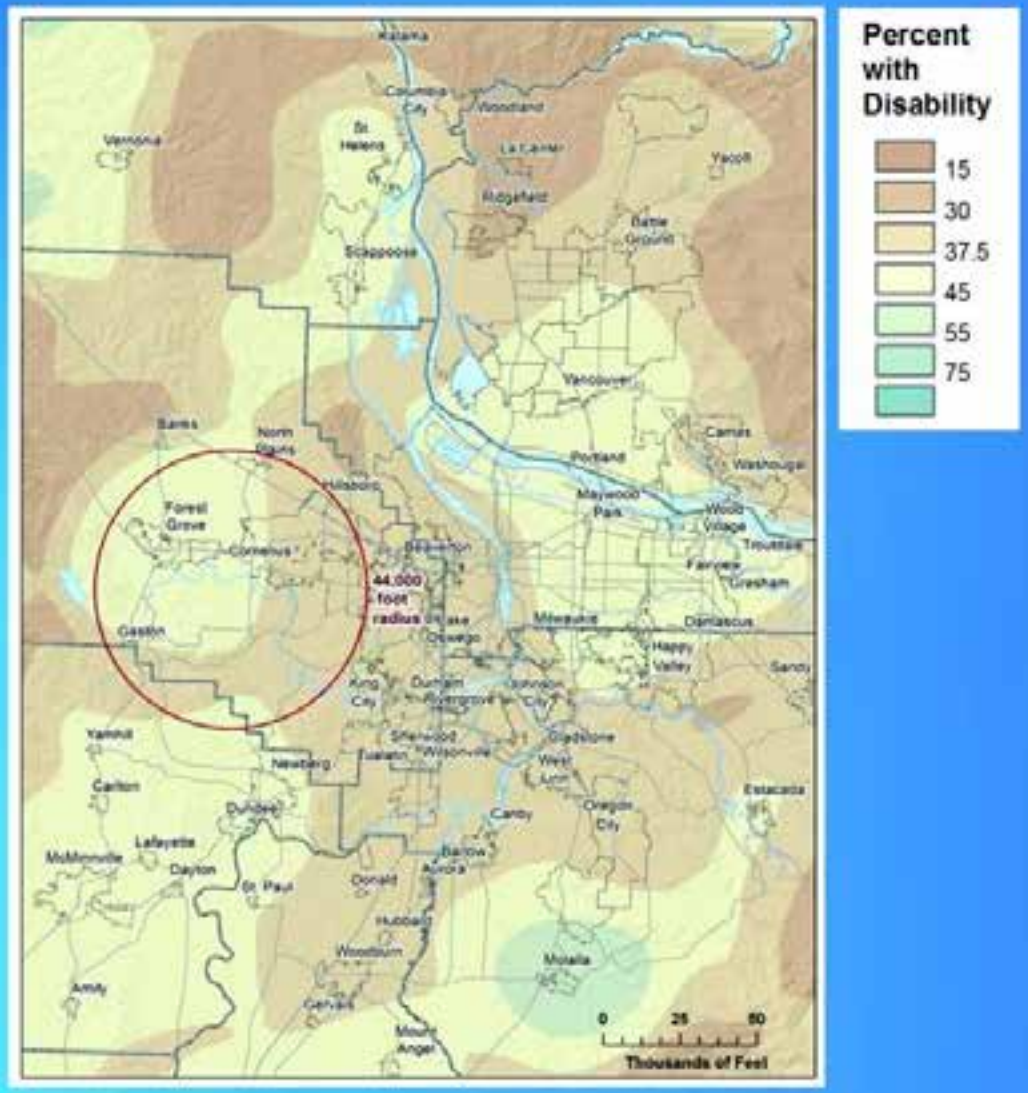
Regional scale comparison

- The fixed radius grid with a 4,400 foot distance provides considerable local detail but only spotty coverage overall.
- Expanding to a 14,000 foot distance provides some local detail but still does not cover the entire area.



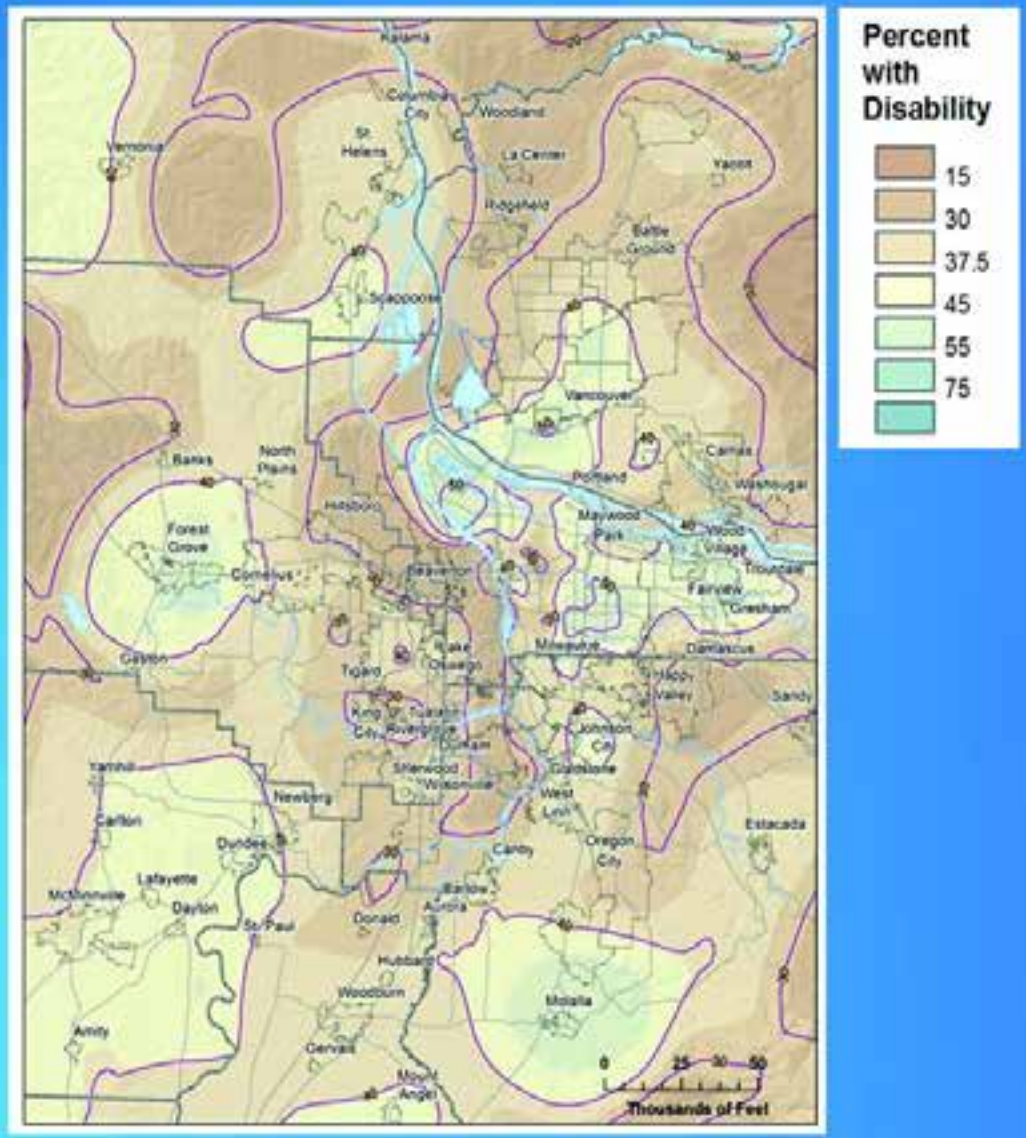
Regional scale comparison

- The fixed radius grid with a 4,400 foot distance provides considerable local detail but only spotty coverage overall.
- Expanding to a 14,000 foot distance provides some local detail but still does not cover the entire area.
- Bumping all the way up to a 44,000 distance provides grid detail for the entire area but highly generalized.



Regional scale comparison

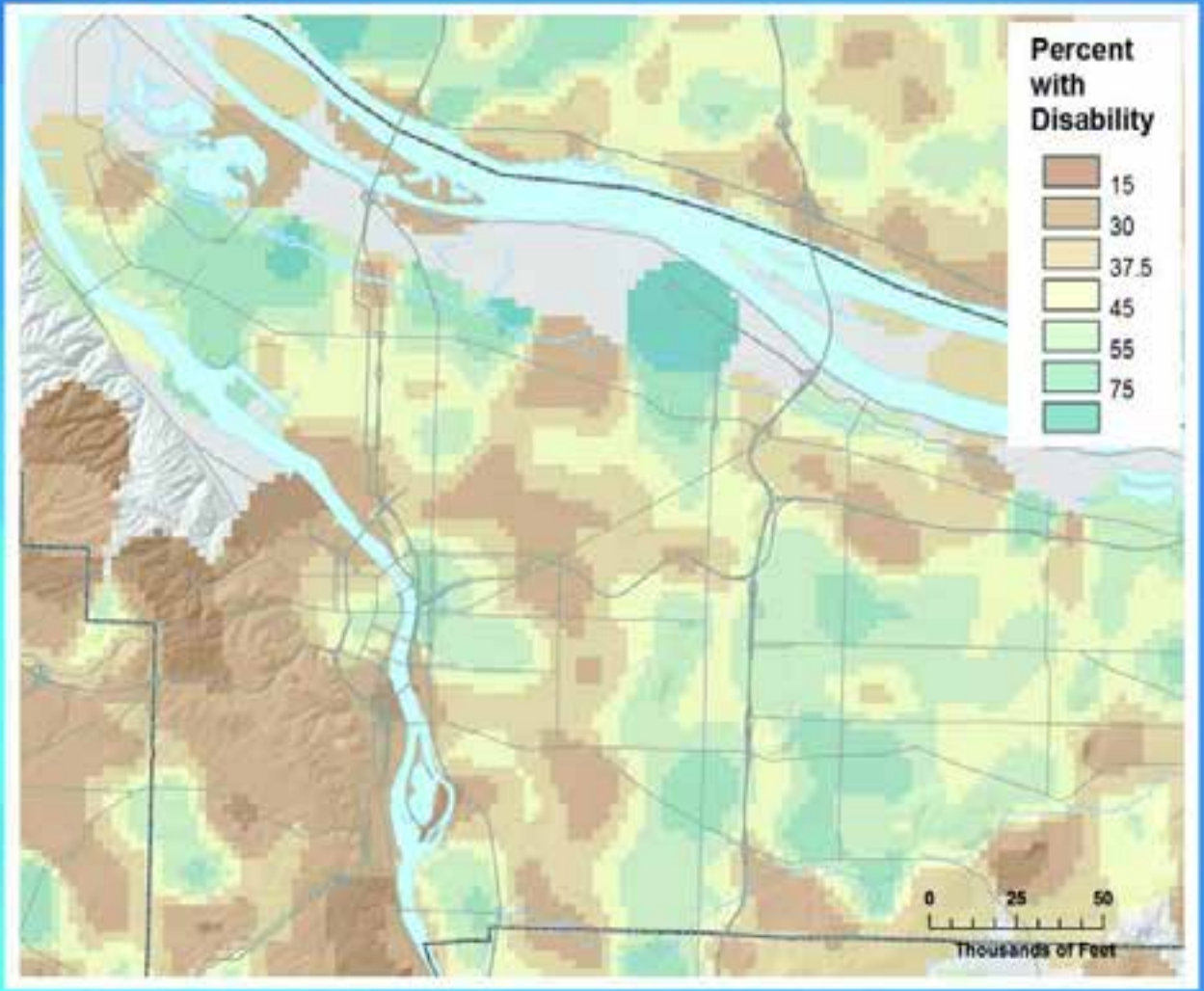
- The fixed radius grid with a 4,400 foot distance provides considerable local detail but only spotty coverage overall.
- Expanding to a 14,000 foot distance provides some local detail but still does not cover the entire area.
- Bumping all the way up to a 44,000 distance provides grid detail for the entire area but highly generalized.
- The adaptive grid method using a quota of 2,923 persons age 65+ is similar to the fixed grid at 44,000 feet in the rural areas but also provides reasonably good detail in densely populated areas.



Local scale comparison

- This sequence shows the same maps but zoomed in to the Portland City Area.

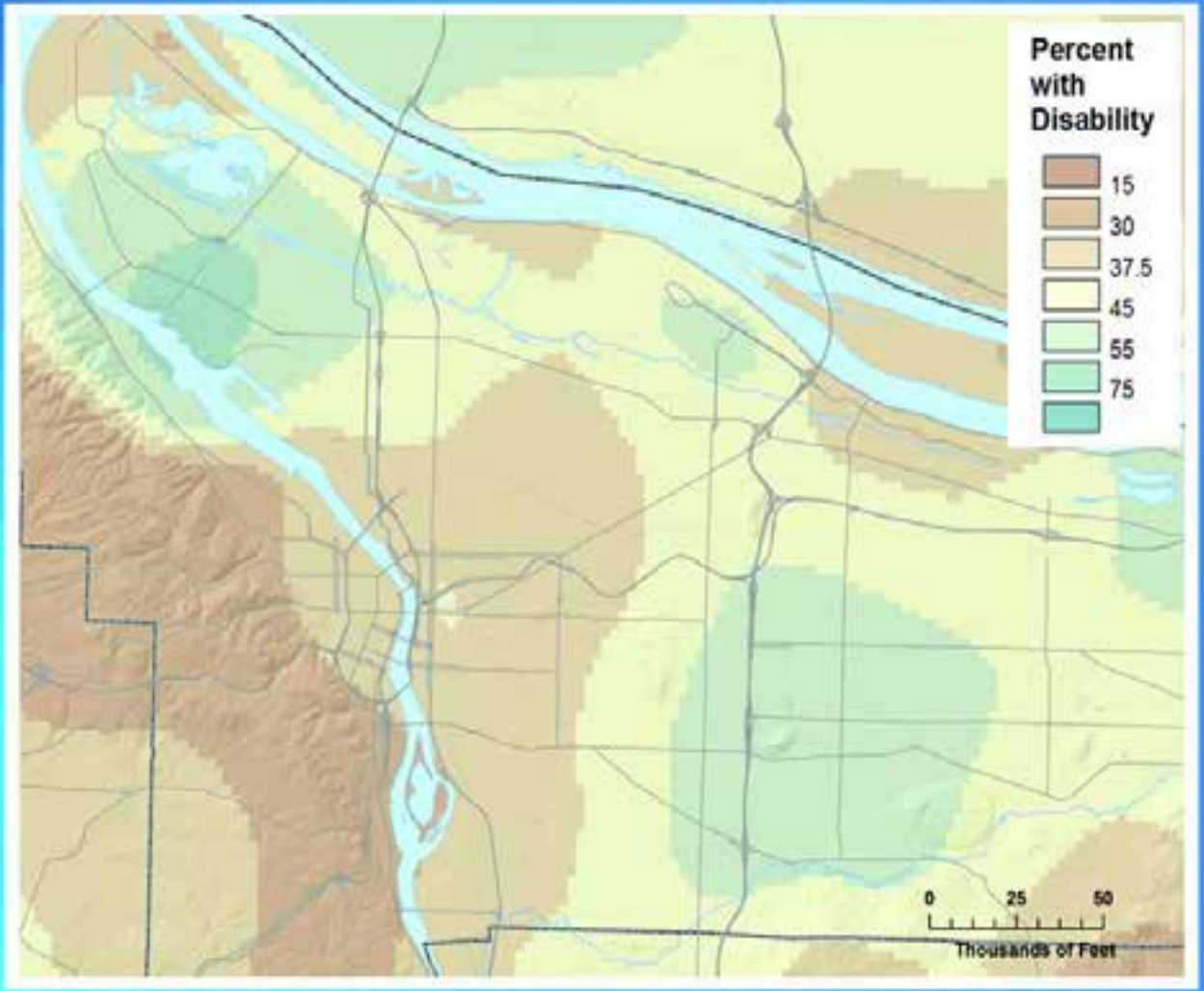
Fixed 4,400



Local scale comparison

- This sequence shows the same maps but zoomed in to the Portland City Area.

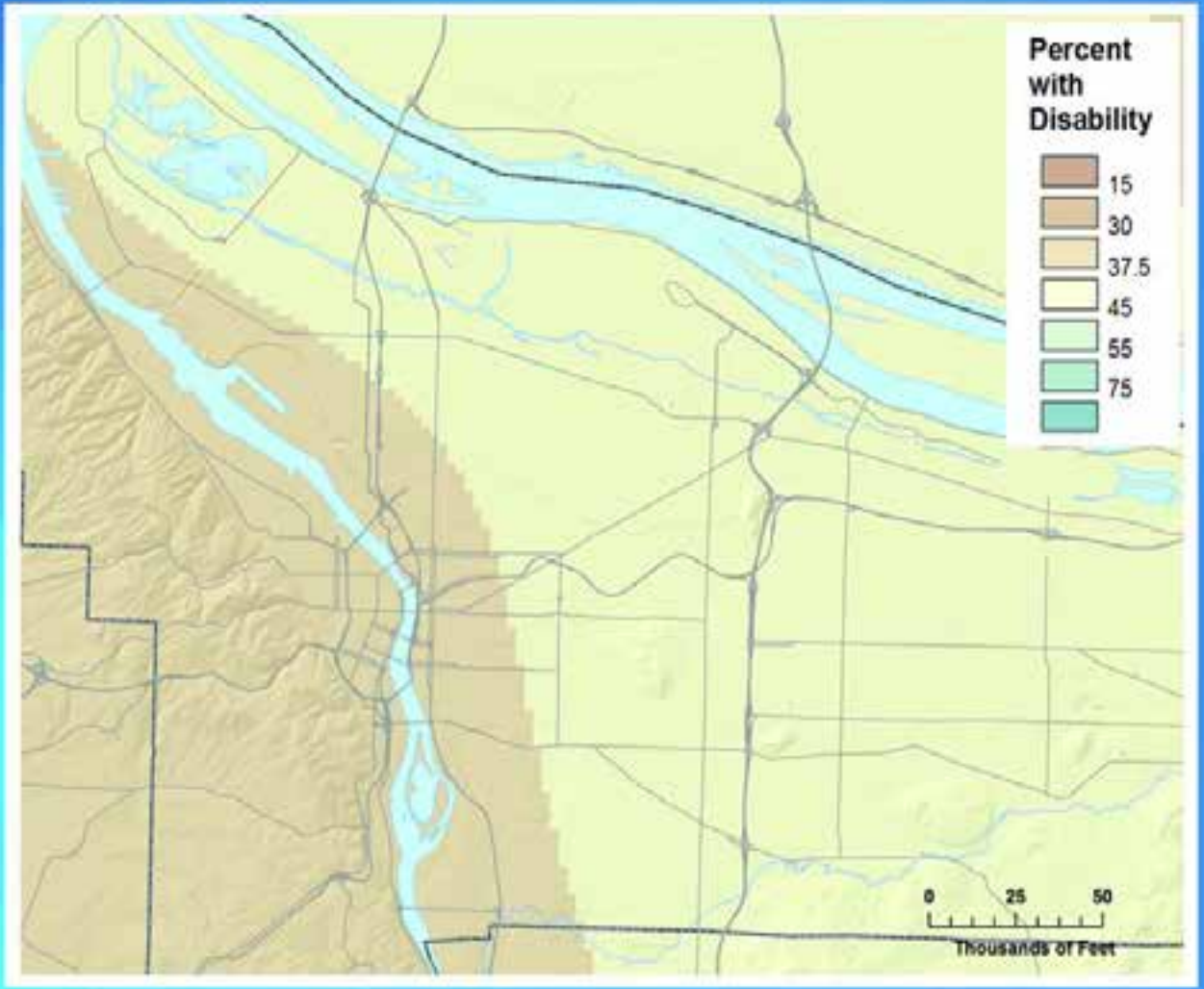
Fixed 14,0000



Local scale comparison

- This sequence shows the same maps but zoomed in to the Portland City Area.

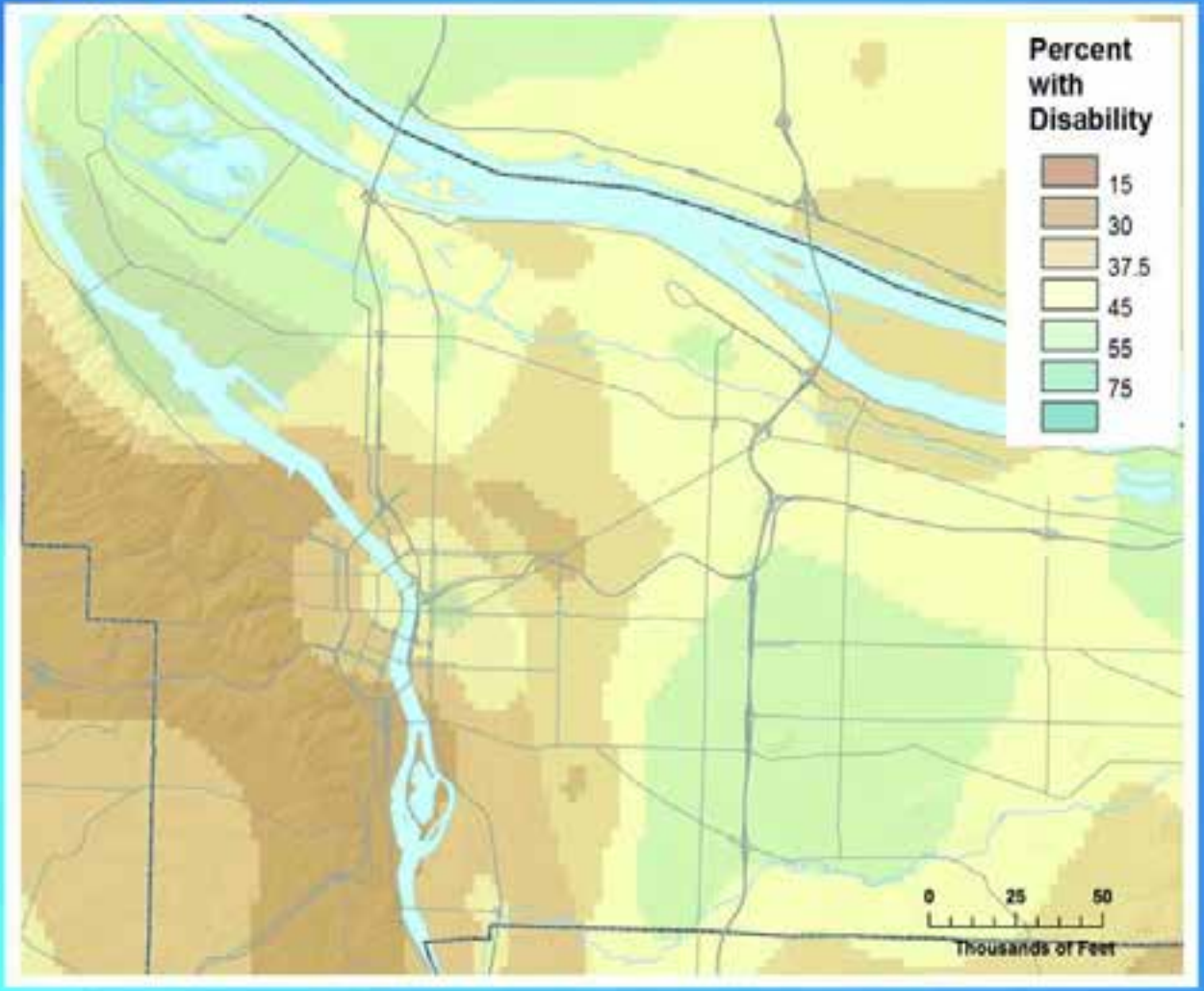
Fixed 44,000



Local scale comparison

- This sequence shows the same maps but zoomed in to the Portland City Area.

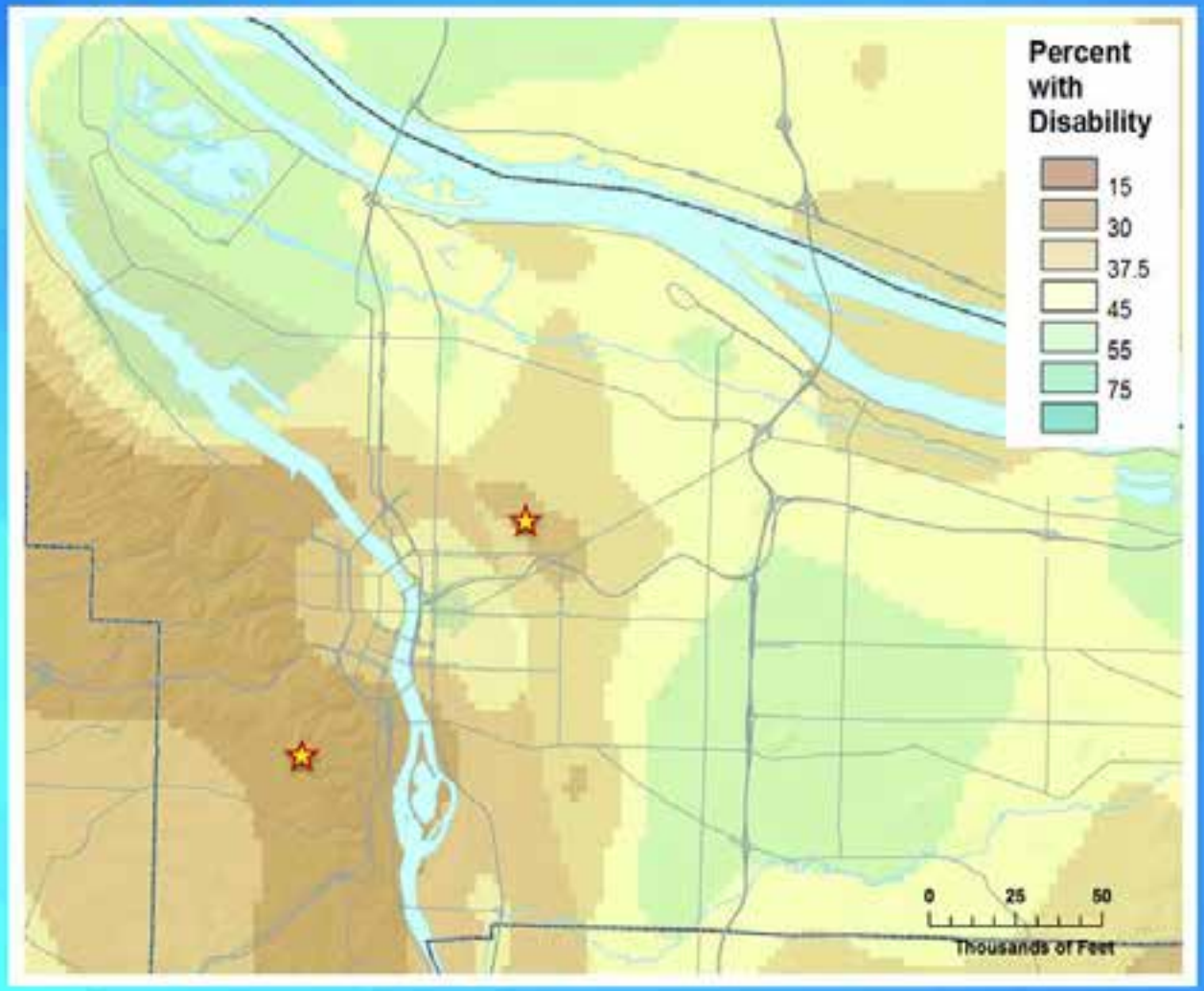
Adaptive 2,923



Local scale comparison

- This sequence shows the same maps but zoomed in to the Portland City Area.
- The zoomed in view of the adaptive density map shows the lower disability rates in the higher SES neighborhoods

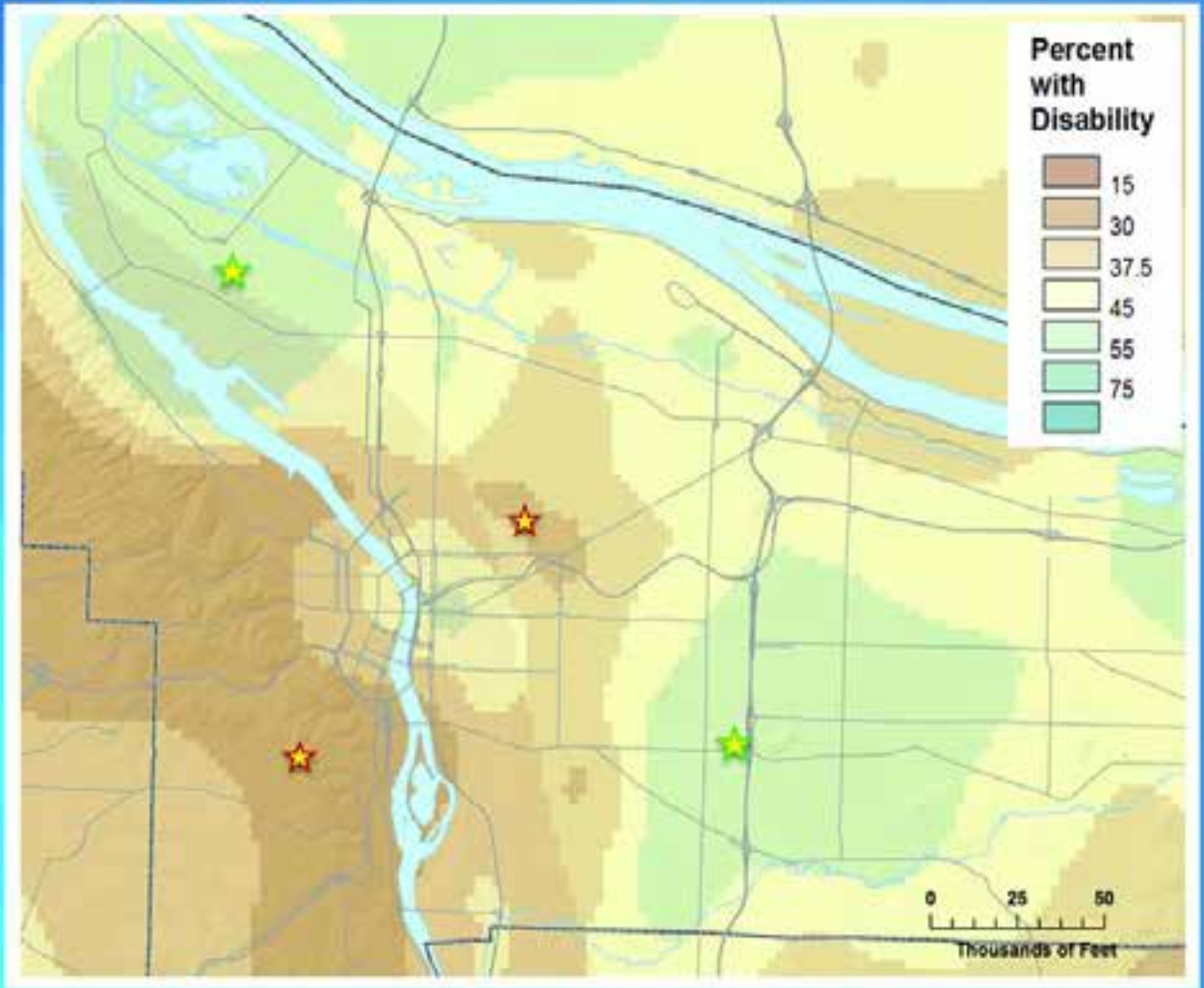
Adaptive 2,923



Local scale comparison

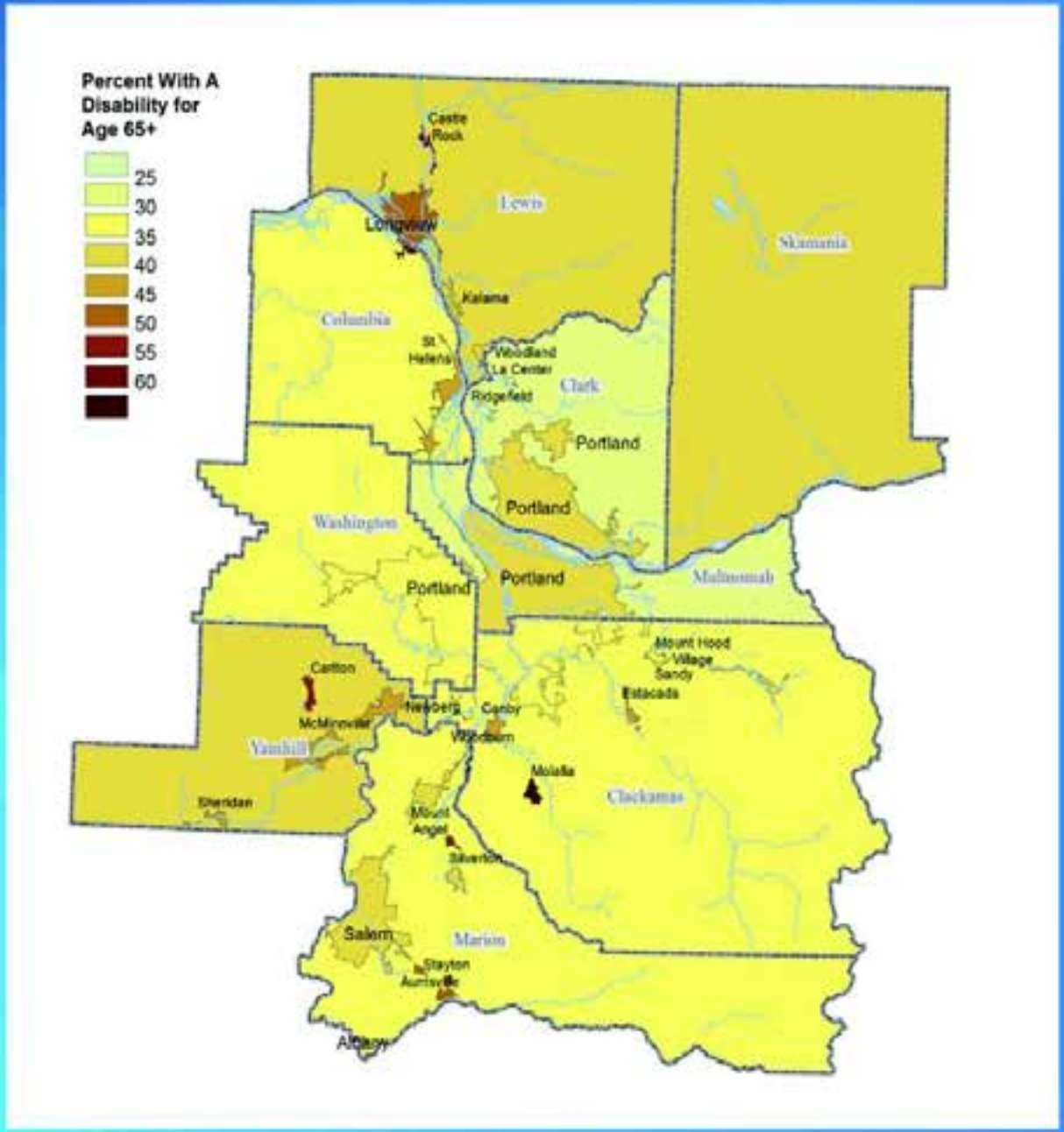
- This sequence shows the same maps but zoomed in to the Portland City Area.
- The zoomed in view of the adaptive density map shows the lower disability rates in the higher SES neighborhoods
- And higher rates in surrounding lower SES areas.

Adaptive 2,923



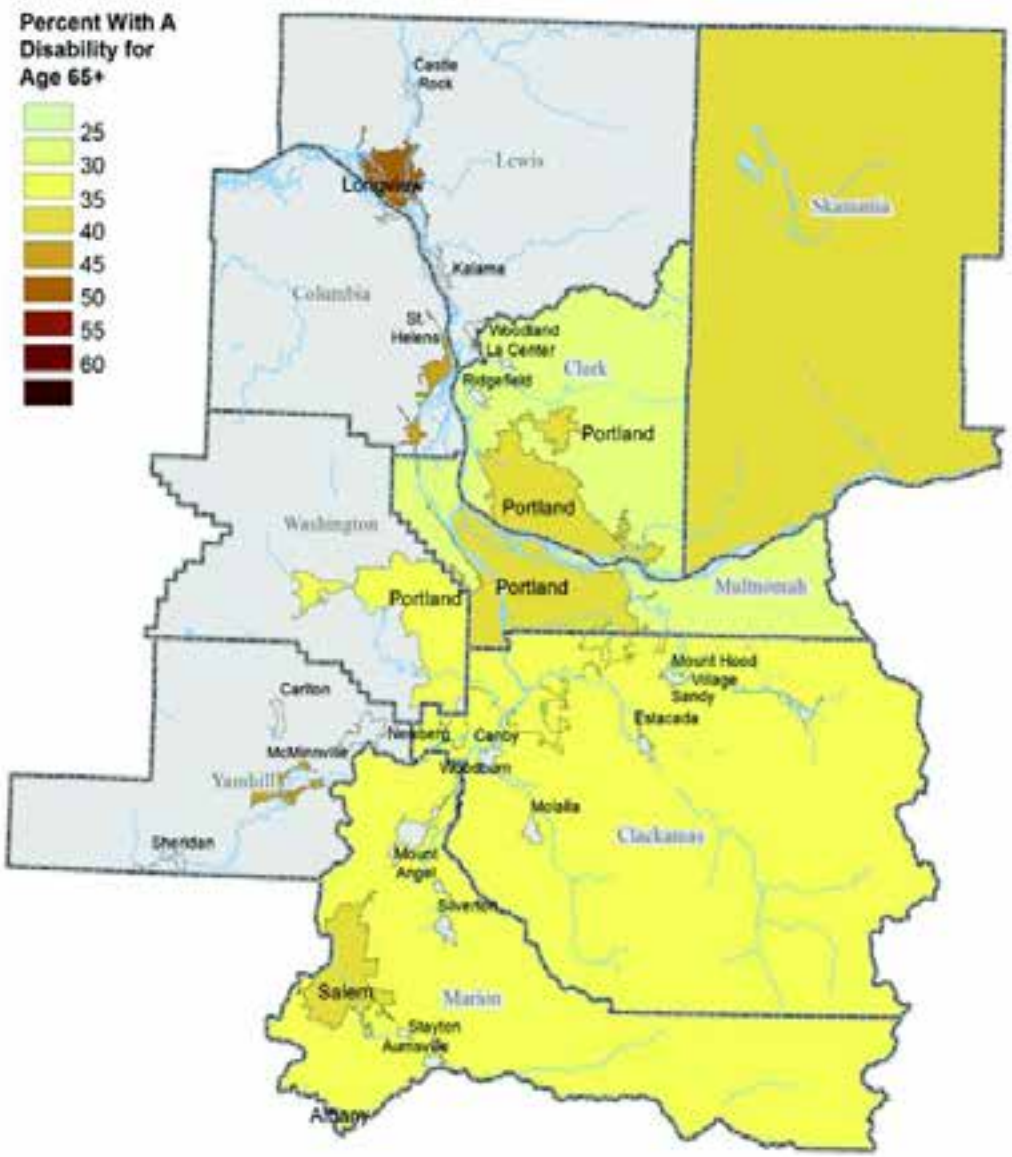
Comparing methods

- This map shows the disability rates as reported in the 2008-13 American Community Survey for *urban areas* and *clusters* and *rural* areas of counties.



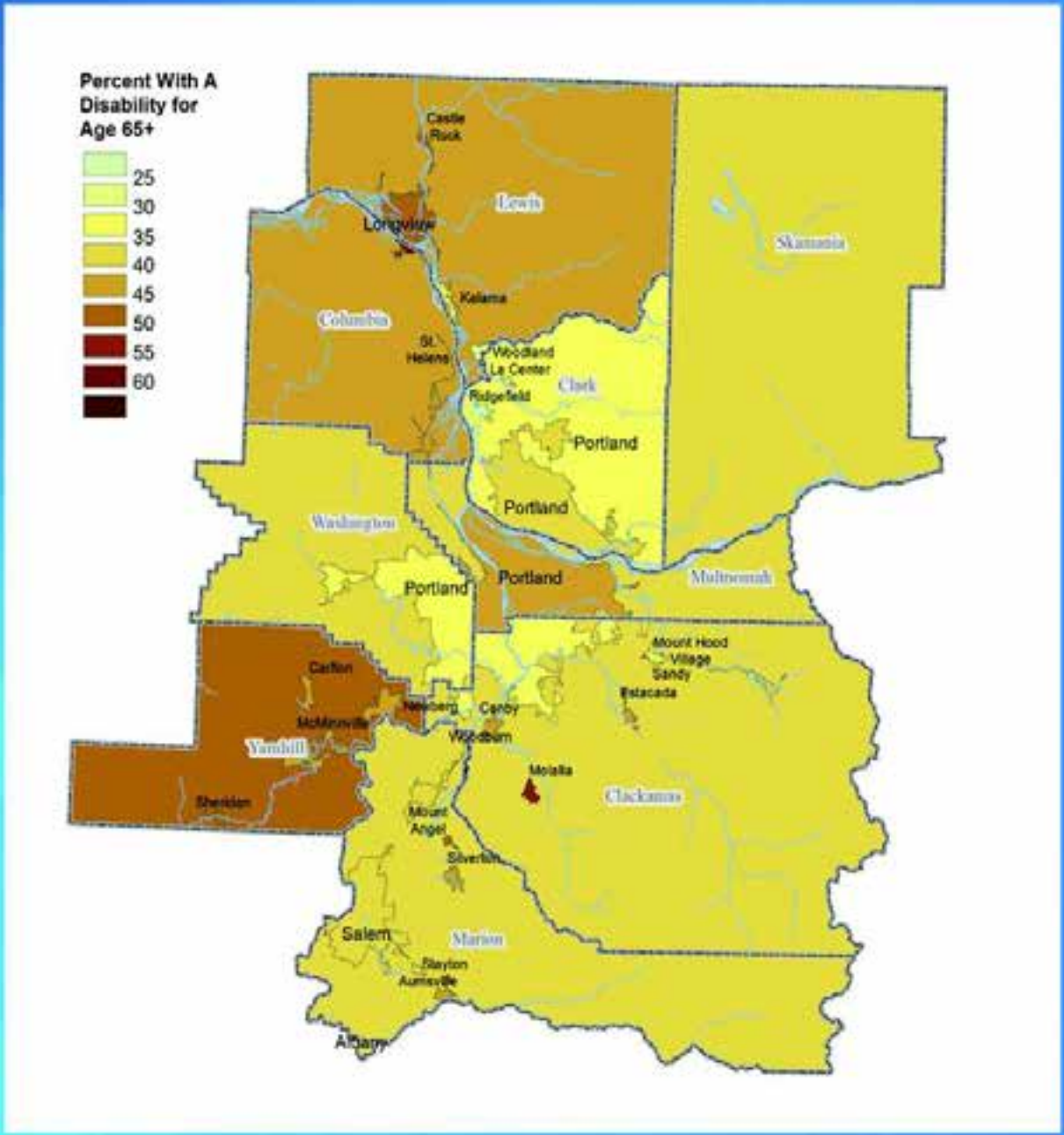
Comparing methods

- This map shows the disability rates as reported in the 2008-13 American Community Survey for *urban areas* and *clusters* and *rural* areas of counties.
- Then we mask out those areas where the MOE is 15% or more of the estimate. We lose the rural portions of several counties and most of the urban clusters.



Comparing methods

- This map shows the disability rates as reported in the 2008-13 American Community Survey for *urban areas* and *clusters* and *rural* areas of counties.
- Then we mask out those areas where the MOE is 15% or more of the estimate. We lose the rural portions of several counties and most of the urban clusters.
- The map of disability rates calculated from the adaptive grid method provides data for all of the geographies, but draws high and low values to the center.

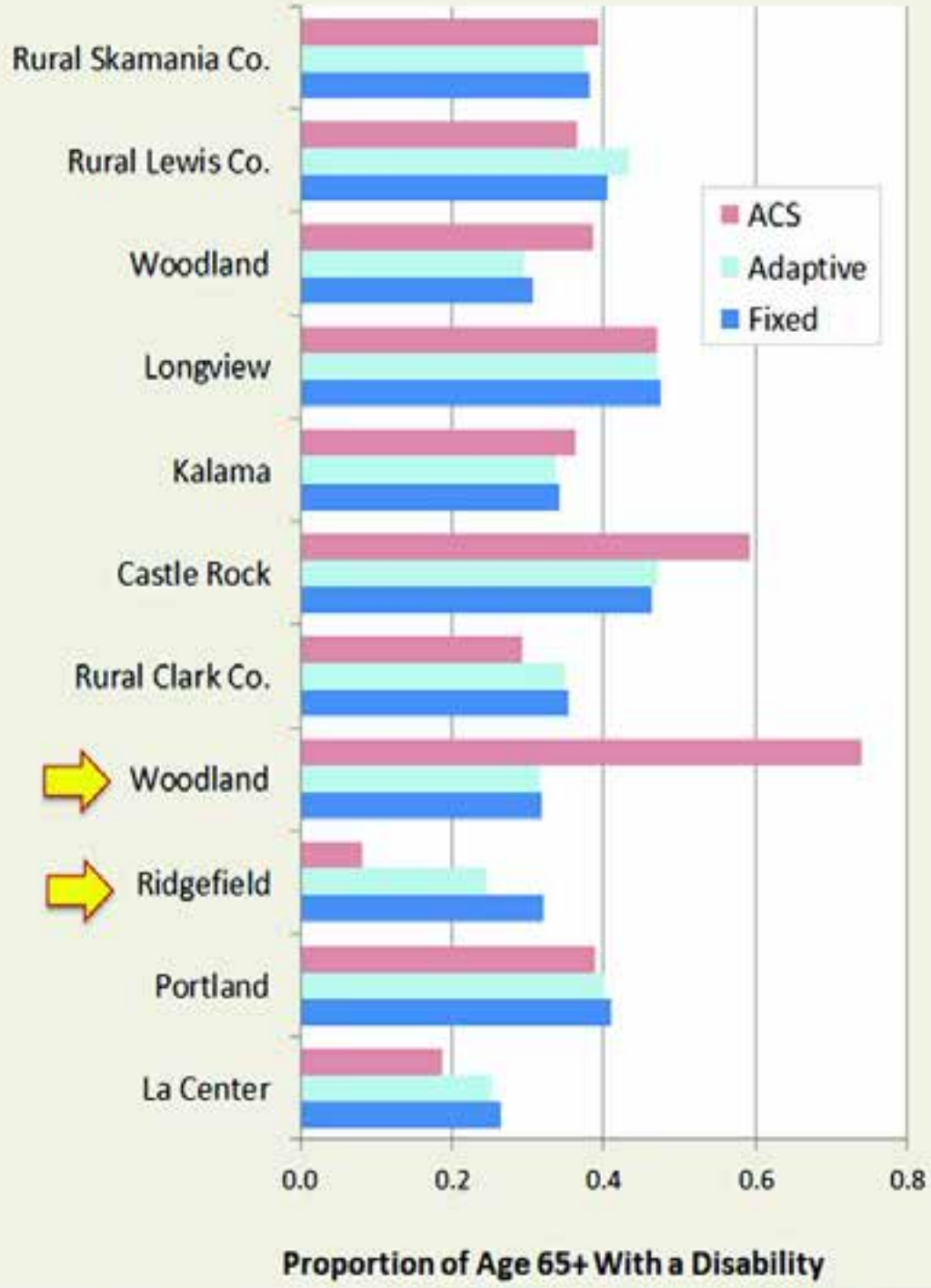


Comparing fixed and adaptive

- The use of either grid based approach limits the very high and very low rates, such as the 78% disabled for the Clark County portion of Woodland and 8% for nearby Ridgefield.

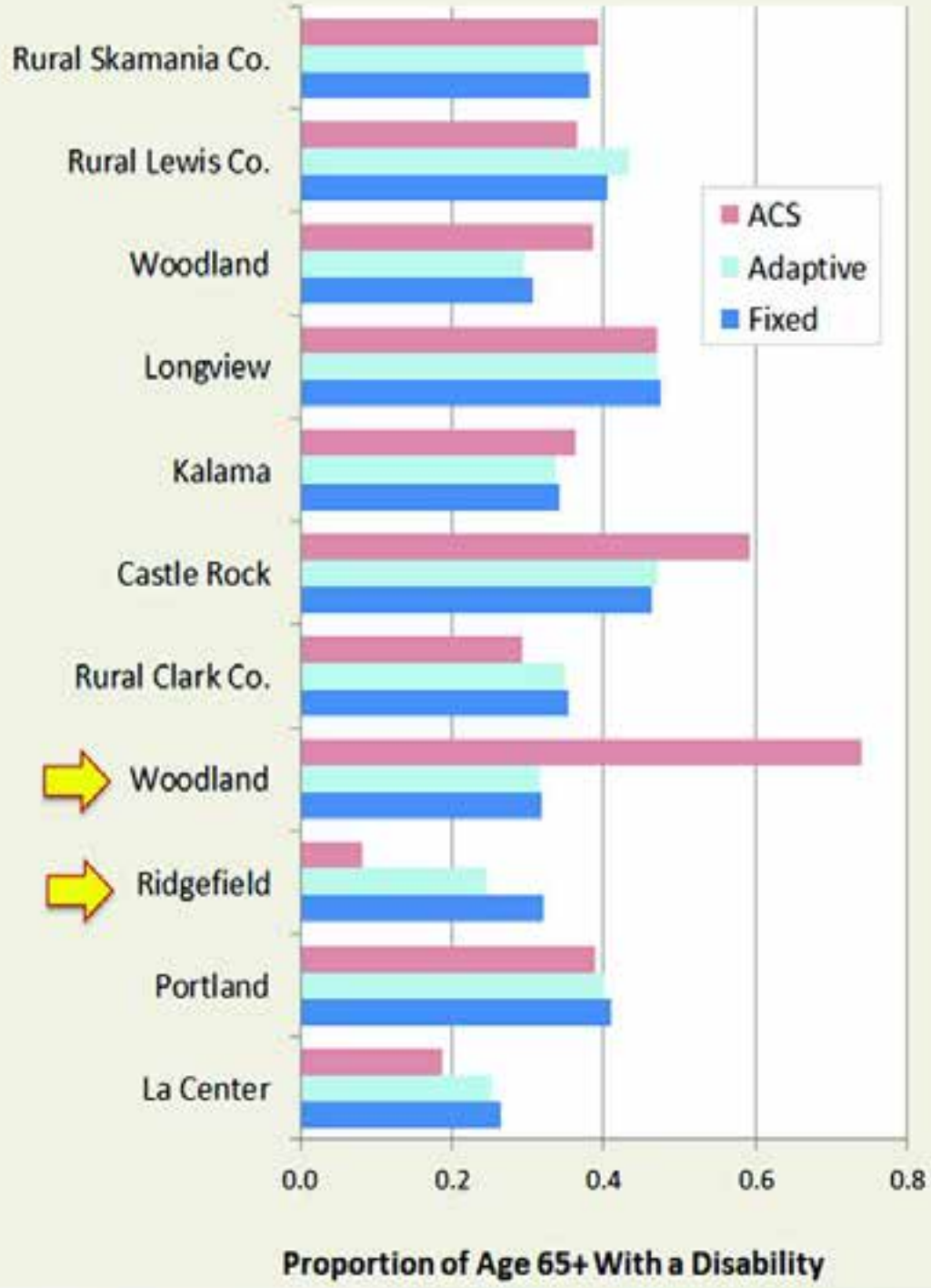
Comparing fixed and adaptive

- The use of either grid based approach limits the very high and very low rates, such as the 78% disabled for the Clark County portion of Woodland and 8% for nearby Ridgefield.



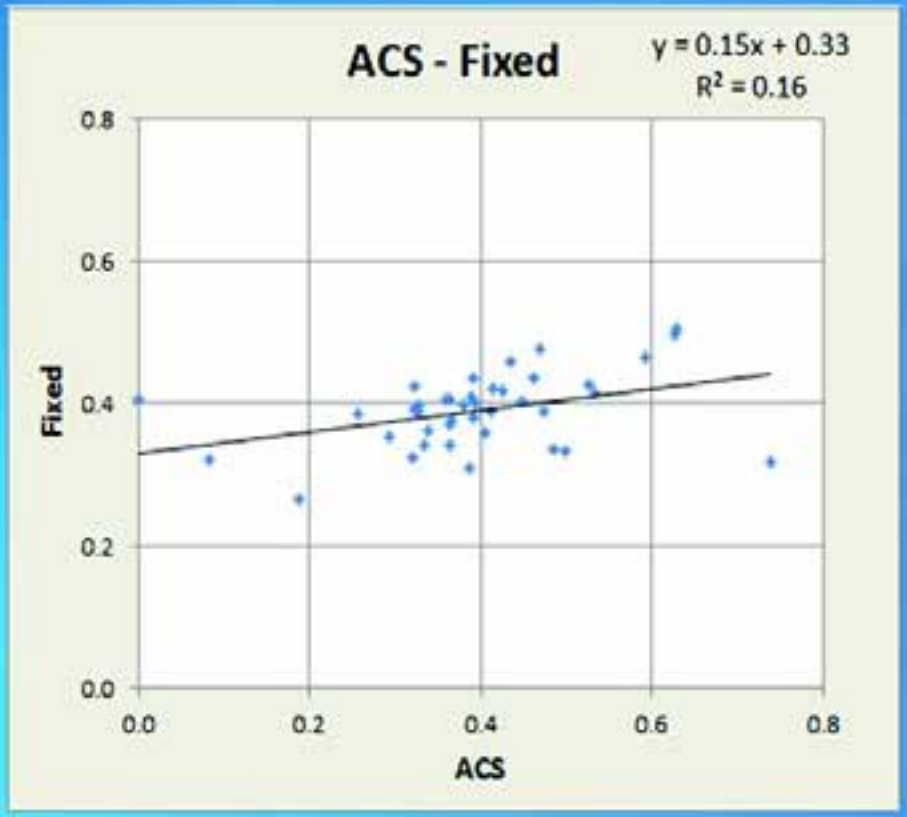
Comparing fixed and adaptive

- The use of either grid based approach limits the very high and very low rates, such as the 78% disabled for the Clark County portion of Woodland and 8% for nearby Ridgefield.
- The grid based values for Woodland and Ridgefield are based on the reported numbers of disabled for age 65+ in block groups in the cities plus a surrounding band of block groups determined by the grid model.



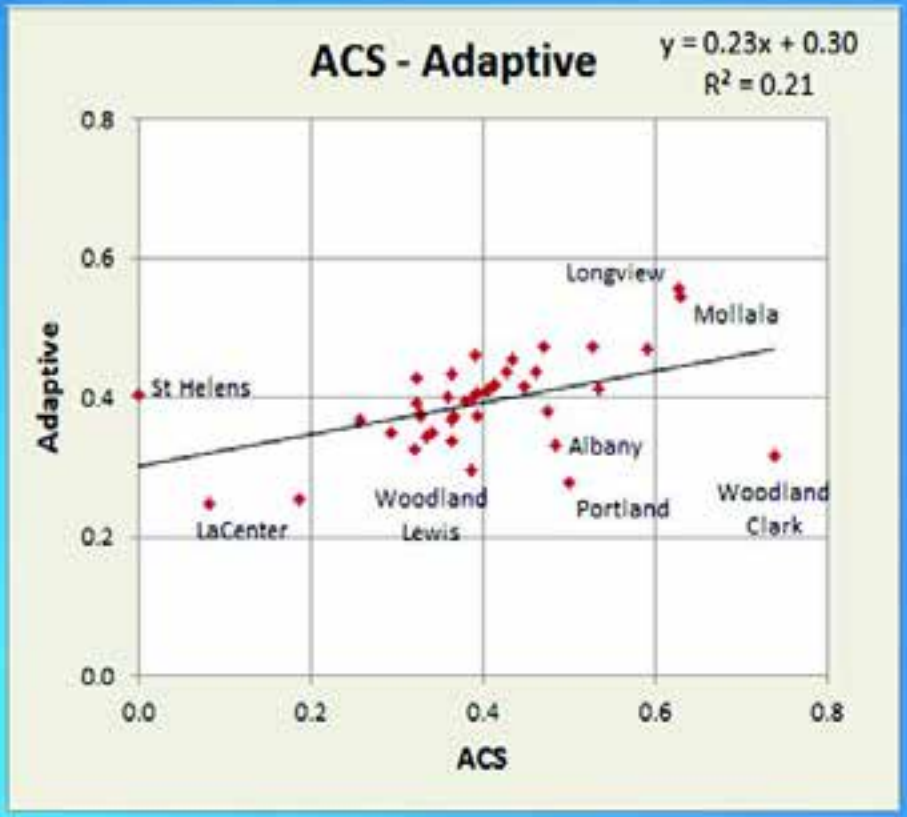
Comparing estimates by fixed and adaptive

- The disability rates computed by use of the 44,000 foot fixed distance grid only have an 0.21 R^2 with the values reported for the UA/UC geographies in the ACS. The also are constrained to a narrower range.



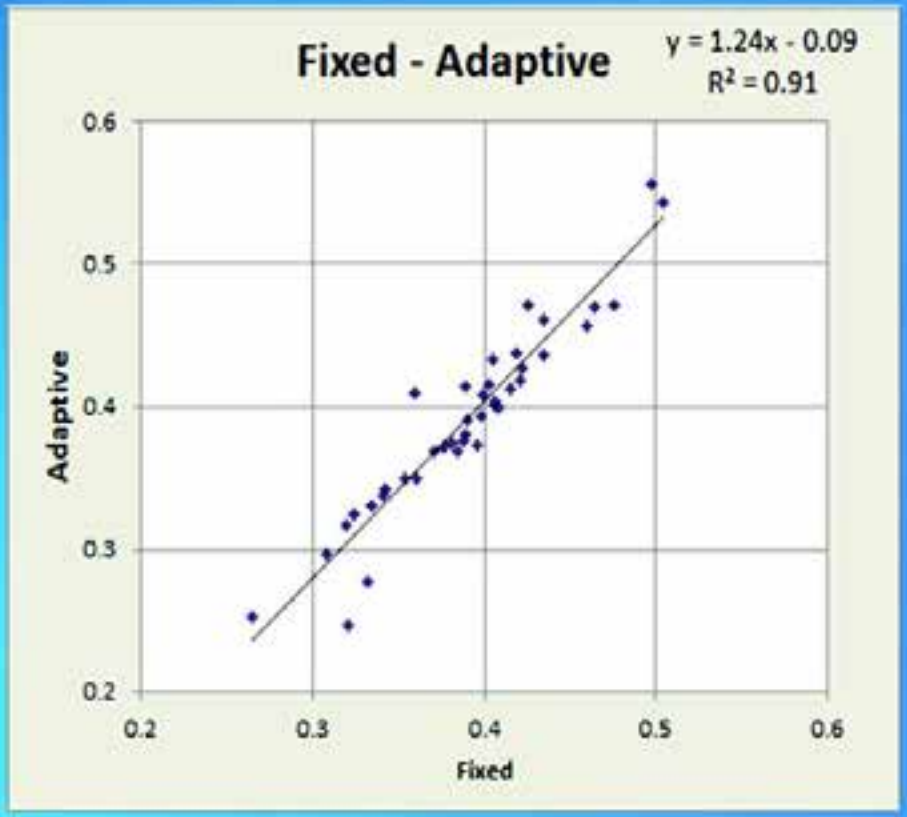
Comparing estimates by fixed and adaptive

- The disability rates computed by use of the 44,000 foot fixed distance grid only have an 0.21 R^2 with the values reported for the UA/UC geographies in the ACS. The also are constrained to a narrower range.
- Rates computed from the adaptive grid method are similar but show a slightly greater range.



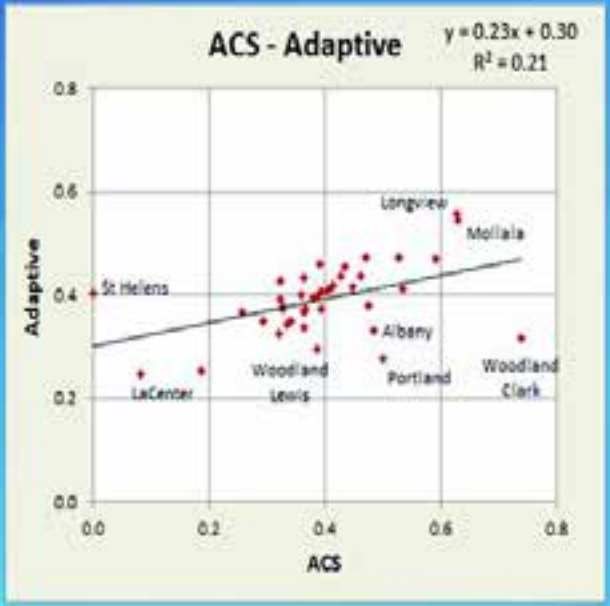
Comparing estimates by fixed and adaptive

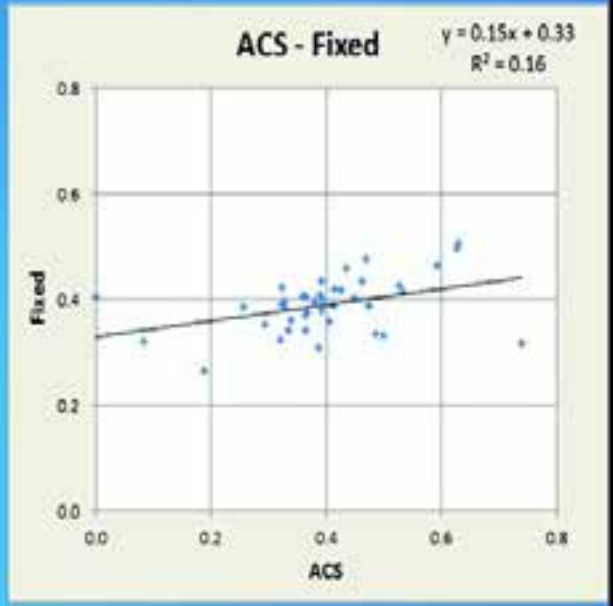
- The disability rates computed by use of the 44,000 foot fixed distance grid only have an 0.21 R^2 with the values reported for the UA/UC geographies in the ACS. The also are constrained to a narrower range.
- Rates computed from the adaptive grid method are similar but show a slightly greater range.
- A comparison of the rates calculated by the two methods shows only minor differences for individual places, but the range of those computed by the adaptive grid is greater by a factor of 1.24.

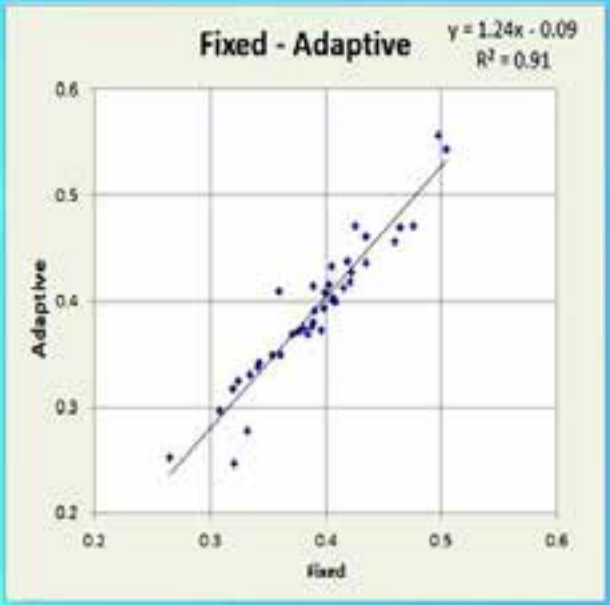


Comparing estimates by fixed and adaptive

- The disability rates computed by use of the 44,000 foot fixed distance grid only have an 0.21 R^2 with the values reported for the UA/UC geographies in the ACS. The also are constrained to a narrower range.
- Rates computed from the adaptive grid method are similar but show a slightly greater range.
- A comparison of the rates calculated by the two methods shows only minor differences for individual places, but the range of those computed by the adaptive grid is greater by a factor of 1.24.







Conclulsions

- Interpolation of block group data using grid based rates may be appropriate for reducing the sampling variability in ACS data.
- Applying grid methods to the Census Urban Area geographies may not have been a good idea. The urban/rural density gradients are very steep for these data, particularly in Oregon. We may test use for block group to neighborhood geographies.
- The recent release of numerous block group data tables from the ACS begs for tools to manage the large sampling variability in these data.
- There is a tradeoff between generalizing maps and data and risking spikey tabular results. The use of adaptive bandwidth grids makes this tradeoff more explicit.
- The choice of bandwidth in the adaptive model is simpler than in the fixed model but understanding how the data are selected by distance in the adaptive model may be more difficult for the viewer to understand.
- It would be helpful if Spatial Analyst included and adaptive bandwidth density tool. A number of articles suggest it is preferred over the fixed bandwidth model in a variety of settings.
- If Spatial Analyst were to include an adaptive bandwidth density tool the type of computation used in SatScan might be more useful that that used in CrimeStat.

Richard Lycan
Institute on Aging
College of Urban and Public Affairs
Portland State University, Oregon
lycand@pdx.edu

Portland State
UNIVERSITY