

**THERE AND BACK AGAIN:
DEMOGRAPHIC SURVEY SAMPLING
IN THE 21ST CENTURY**

Colm O'Muircheartaigh
NORC, University of Chicago

OVERVIEW

- **History of Demographic Survey Sampling**
- **20th Century Sample Design**
- **New Directions**
- **Evaluation of Lists, GIS, and Maps**
- **Implications**
- **New National Sample Designs**
- **Swiss Cheese**
- **Tailored Samples vs Master Samples**
- **Conclusion**

HISTORY 1

- **A N Kiaer (1895)**
 - **ISI Berne *Representative Enumerations***
 - **Miniature of the population**
 - **Multi-stage design – *places, towns, streets, HUs***
 - **Stratified**
- **US implementation**
 - **Cressy L Wilbur (1896-7) – [vital statistics]**
 - ***small representative areas***
 - **Carroll D Wright (1875 et seq) – [labor statistics]**
 - ***representative statistics***
 - **Non-probability samples**

HISTORY 2 – DEVELOPMENT

- **Bowley (1906)**
 - Theory for simple random sampling
- **Neyman (1934)**
 - Superiority of probability sampling
 - Theory for unequal cluster sampling
- **Hansen Hurwitz Madow 1940s**
 - PPS at higher stages
 - Adequate “representation” of important units
 - Leads to identification of *certainty PSUs*
 - Equal workloads at final stage (HUs)
 - Efficiency of field allocation and estimators
- **1950s: national master samples ISR, NORC, et al.**

THE BASIC NATIONAL DEMOGRAPHIC DESIGN

- **Multi-stage**
 - **Costs**
 - **Feasibility**
- **Some self-representing PSUs**
- **Stratified**
 - **Incorporating knowledge of population and structure**

20th CENTURY DEMOGRAPHIC SURVEY

SAMPLE DESIGN ELSEWHERE

- **Scandinavia**
 - Register-based
- **China**
 - Register-based
 - Late 1980s, registers deteriorated
- **UK**
 - Electoral registers, updated annually
 - 1980s, registers deteriorated
 - Postcode address file (PAF), centrally available
 - Periodic redesign

20th CENTURY DEMOGRAPHIC SURVEY

SAMPLE DESIGN IN USA

- **Decennial update of frame, and**
- **Absence of a current list of population elements**
 - **Selection of a MASTER SAMPLE of PSUs and SSUs**
 - **Listing of the frame for the master sample**
 - **Use as reservoir for the decade**
 - **Updating in the field for the sample only**

NEW DIRECTIONS IN THE USA

- **Availability of current administrative lists**
- **Matching and pre-classification of geographies**
- **GIS and GPS**

- **Tailored samples vs master samples**

WHY LISTS WOULD MAKE A DIFFERENCE

- **Cost parameters would change**
- **Nature of PSU might change**
- **Subsampling fraction might change**
- **Timing of revisions could change**

THE (NON-CENSUS) ADMINISTRATIVE ALTERNATIVE

- **USPS delivery sequence file**
 - Ordered within ZIP by carrier route
 - Within carrier route by walk sequence
- **Available through licensees**
 - Primarily purchased by direct-mail organizations
- **Usability**
 - Basis for MAF in urban areas
 - Addresses in standard format
 - Operational incentives for updating
 - Can be geocoded and mapped
 - Contains PO boxes and rural route boxes (not mappable)

USING/EVALUATING THE LIST

- **1 Direct/non-evaluative use, single city survey, 2001 RTI**
- **2 Evaluation against traditional listing, 2001-2 NORC**
- **3 Inner-city evaluation and use, 2002-3 NORC**
- **4 Direct/non-evaluative use as national frame, 2003 RTI**
- **5 “Rural” evaluation, 2003 NORC**
- **6 Basis for national design template, 2003-4 NORC**
- **7 National comparison with traditional listing, 2004 NORC/ISR**

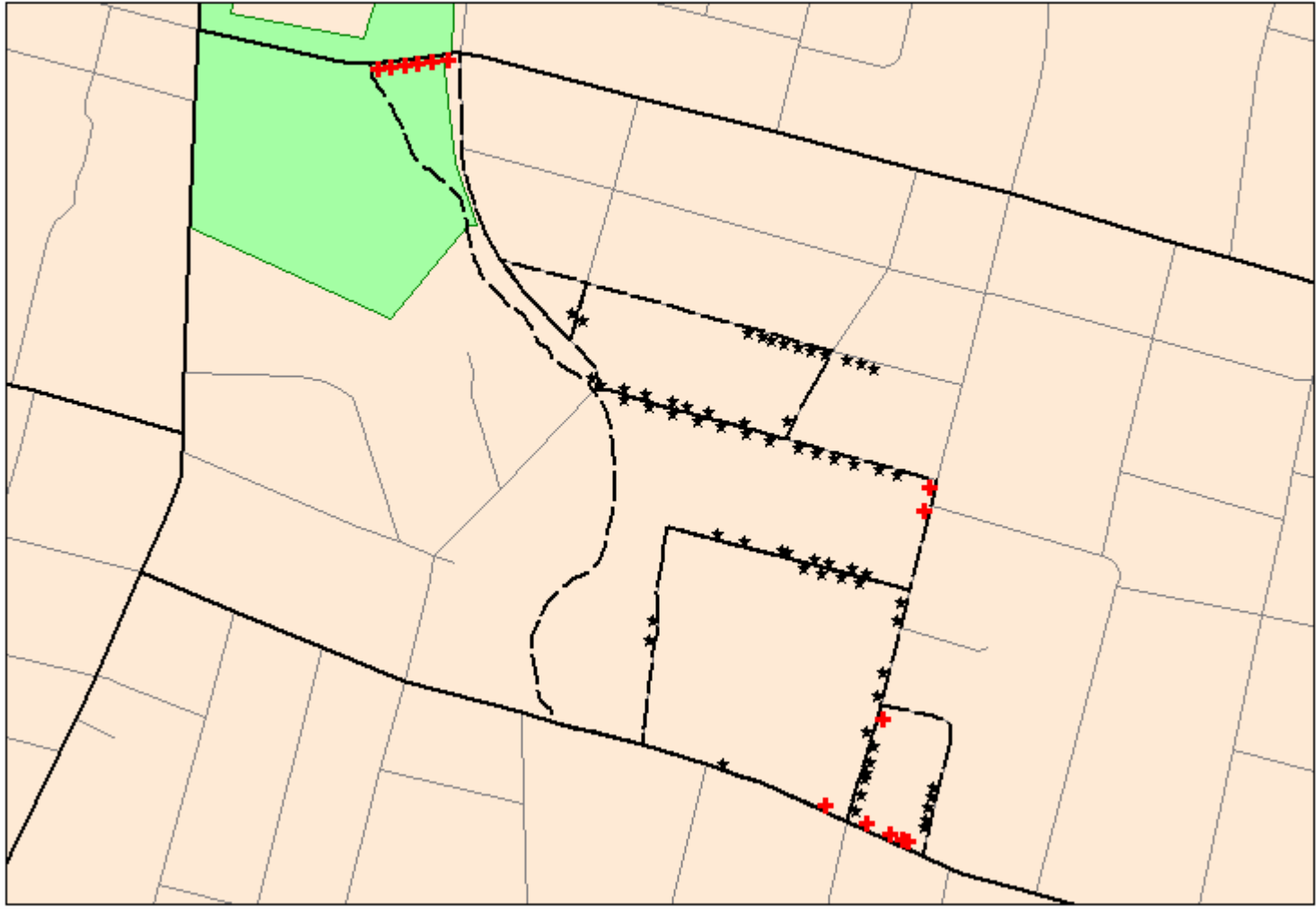
DIRECT USE (RTI-2001)

- **Iannachione, Staab, Redden**
 - **Houston, TX**
 - **Geocoded > 99% of addresses**
 - **Selected sample from list**
 - **97% of selected addresses yielded HUs**
 - **Order of magnitude of list and census count same**
 - **No direct coverage check**

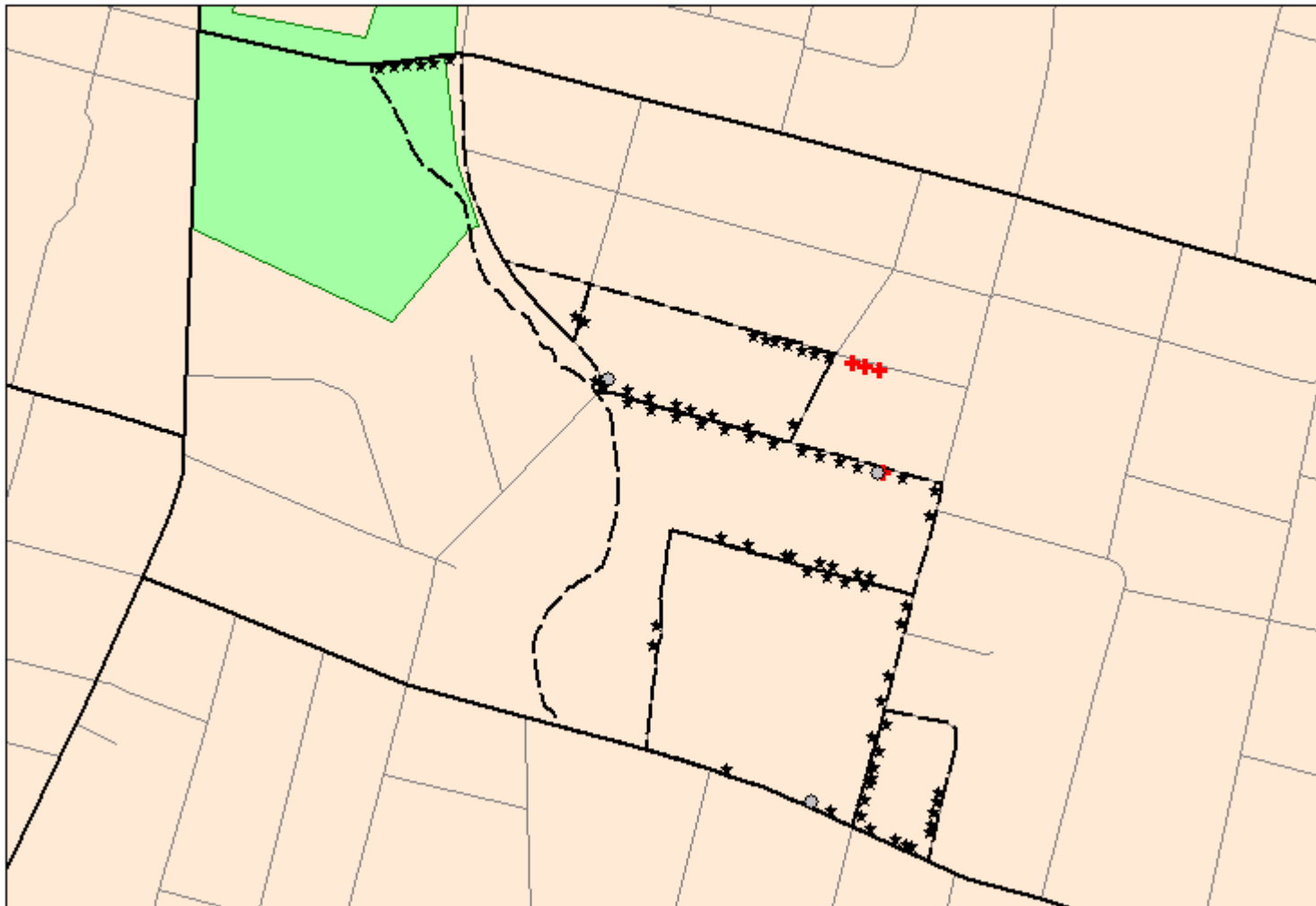
VALIDATING THE LIST (NORC 2001-2)

- **O’Muircheartaigh, Eckman, and Weiss**
- **NORC GSS Field Test 2001: 14 segments**
 - **First, traditional listing (T)**
 - **Then, geocoded USPS list for the areas (U)**
 - **Finally, independent enhanced list (E) built from U**
- **Comparison of coverage**
 - **T – Traditional**
 - **U – USPS addresses geocoded inside segment**
 - **E – U enhanced in the field**
 - **USPS – full USPS list whether geocoded inside or not**

Segment 100008-1000107
E,T HUs



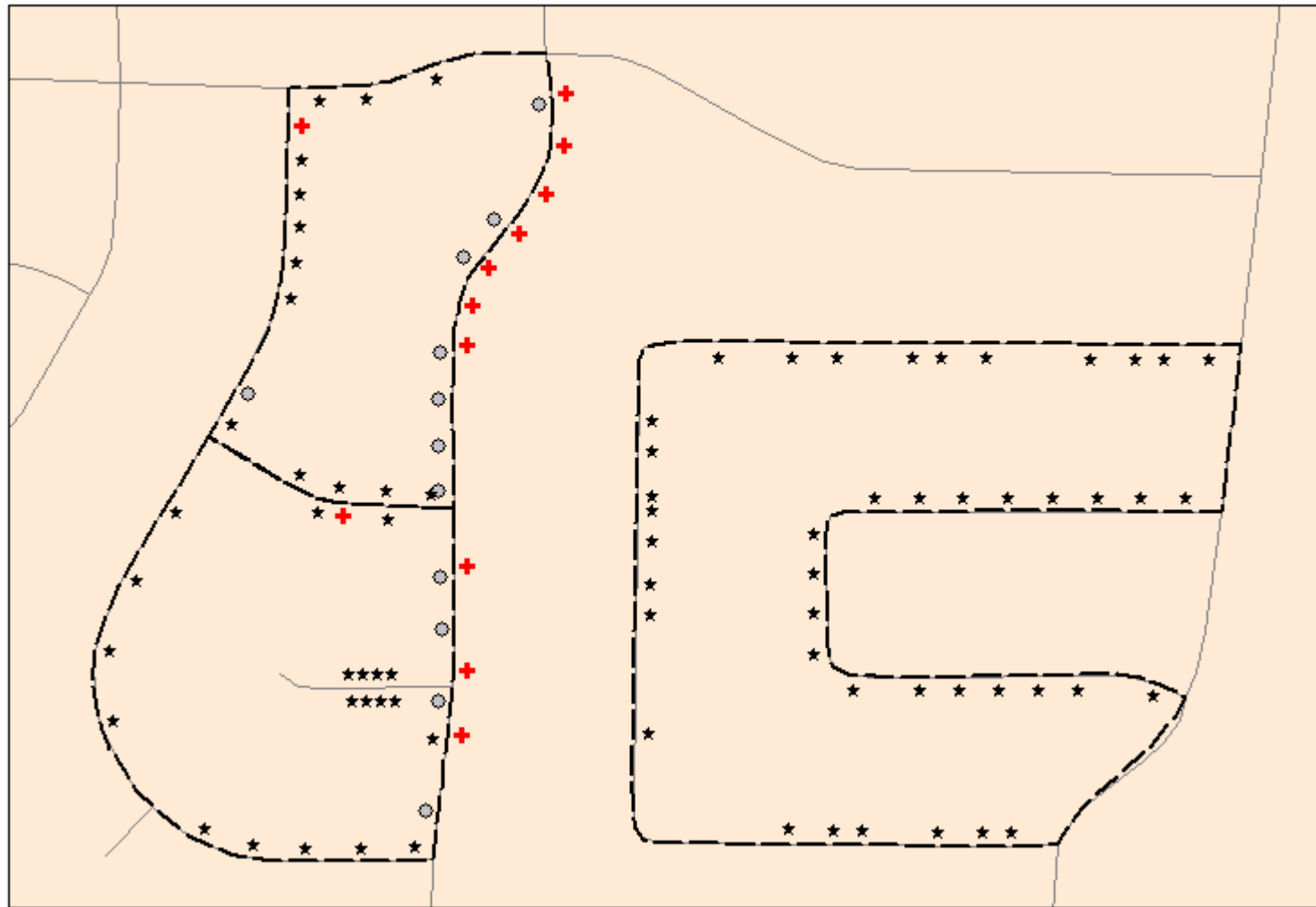
Segment 100008-1000107
E,U HUs



ISSUES ENCOUNTERED IN ENHANCED LISTING

- **Issues with USPS list**
 - missing apartment numbers
 - addresses removed at request of resident
 - PO boxes, rural route boxes unusable
 - includes hard to find HUs missed by field listers
- **Geocoding issues**
 - block boundaries
 - side-of-street errors
- **Matching geographies**
 - ZIPs vs blocks, block groups, tracts

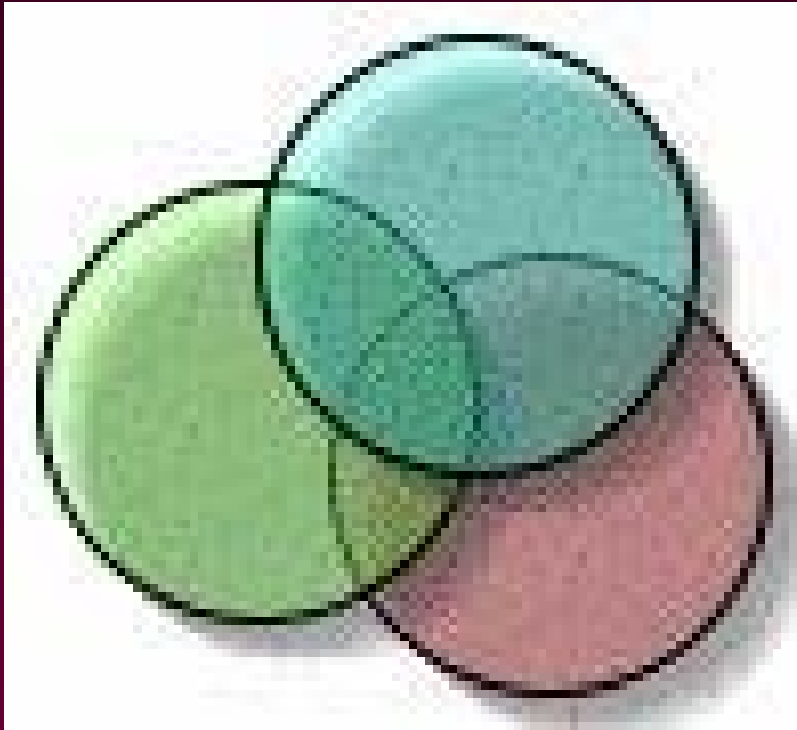
Segment 100279-1000017
E,U HUs



COST COMPARISON T vs E

- **Travel costs, etc.**
 - Equal
- **Listing costs**
 - T approximately twice as expensive as E

COMPARISON OF T, U, AND E



- **U in E** **95%**
- **E in U** **93%**
- **T in U** **87%**
- **E in T** **81%**

- **E in USPS** **96%**

INNER CITY EVALUATIONS (NORC 2002-3)

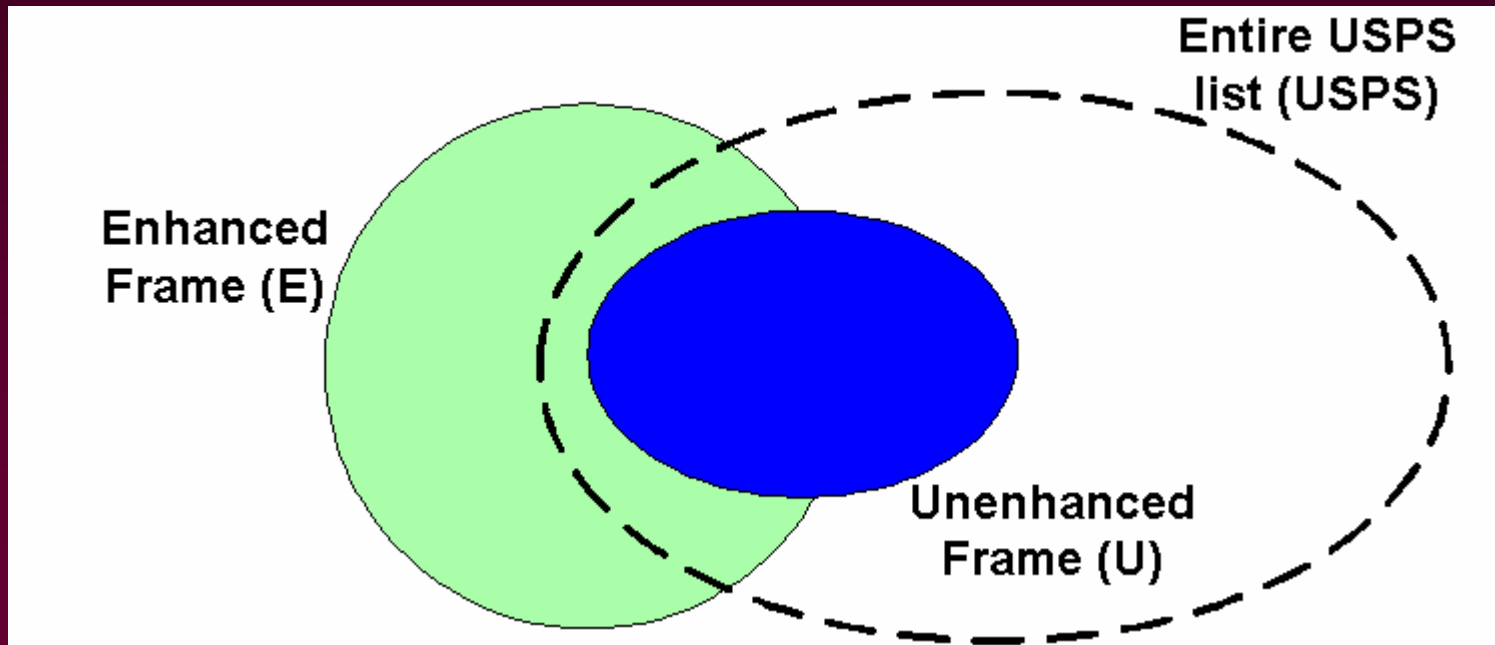
- **O’Muircheartaigh, Eckman, English, and Haggerty**
- ***The Making Connections Project***
 - **Funded by Annie E. Casey Foundation**
- **10 Deprived Inner-City Communities**
 - **Denver, Des Moines, Indianapolis, San Antonio, Seattle**
 - **Milwaukee, Hartford, Providence, Oakland, Louisville**

INNER CITY EVALUATIONS

- **Purchased USPS lists for ZIPs surrounding whole community**
 - **Geocoded all**
 - **With U as base:**
 - **Produced E with in-person listing**
 - **Compared U and E for coverage**
 - **Compared U and E coverage during fieldwork**

INNER CITY EVALUATIONS

- **Two key measures:**
 - **How much of E is in U (the geocoded part of USPS)?**
 - **How much of E is in USPS as a whole**



INNER CITY EVALUATIONS

- **Overall results**
 - **90% of E in U**
 - **94% of E in USPS**
 - **Difference due to geocoding/map inaccuracies**
- **Range across cities:**
 - **82% - 95% of E in U**
 - **83% - 99% of E in USPS**
- **Characteristics of missed HUs**
 - **In most severe cases, many vacant HUs**
- **MHU**
 - **Only moderately successful**

DIRECT USE NATIONAL FRAME (RTI 2003)

- **Staab, Iannachione**
- **Used postal frame exclusively for EuroQol study**
- **Used postal geographies**
- **Ignored ZIPs with no residential addresses**
- **Ignored residents without street addresses**

NATIONAL LIST EVALUATION (NORC/ISR 2004)

- **O'Muircheartaigh, Lepkowski, Heeringa**
 - **HRS and NSHAP**
 - **National listing of 549 segments by ISR**
 - **Purchase of USPS lists for 100 segments**
 - **Comparison of T and U**
 - **Nationally representative**

USE FOR NORC NATIONAL SAMPLE DESIGN 2003

- **Geographic units**
- **Preclassification of list quality**
- **Stratification**
- **Optimal design**

THE POPULATION

- **8.2 million census blocks**
- **66,275 tracts**
- **3219 counties**
- **281 (C)MSAs in Census 2000**
 - **Now 362 MSAs and 565 Micropolitan SAs**
- **Variable population density**
- **Variable list quality**

PRECLASSIFICATION OF GEOGRAPHIES

- **Census classification of blocks [TEA – type of enumeration area]**
 - Available for all blocks
 - Indicator of feasibility of using USPS list as frame
 - Whether suitable for mail-out
 - Address type
- **Type A tracts**
 - 95% of HUs in tract are in blocks classified as TEA=1
- **Type B tracts**
 - All other tracts

THE DESIGN – 1

Categorize MSAs/counties according to population density and list quality

Large MSAs (likely certainty areas) with high-density population dominated by Type A tracts [category 1]

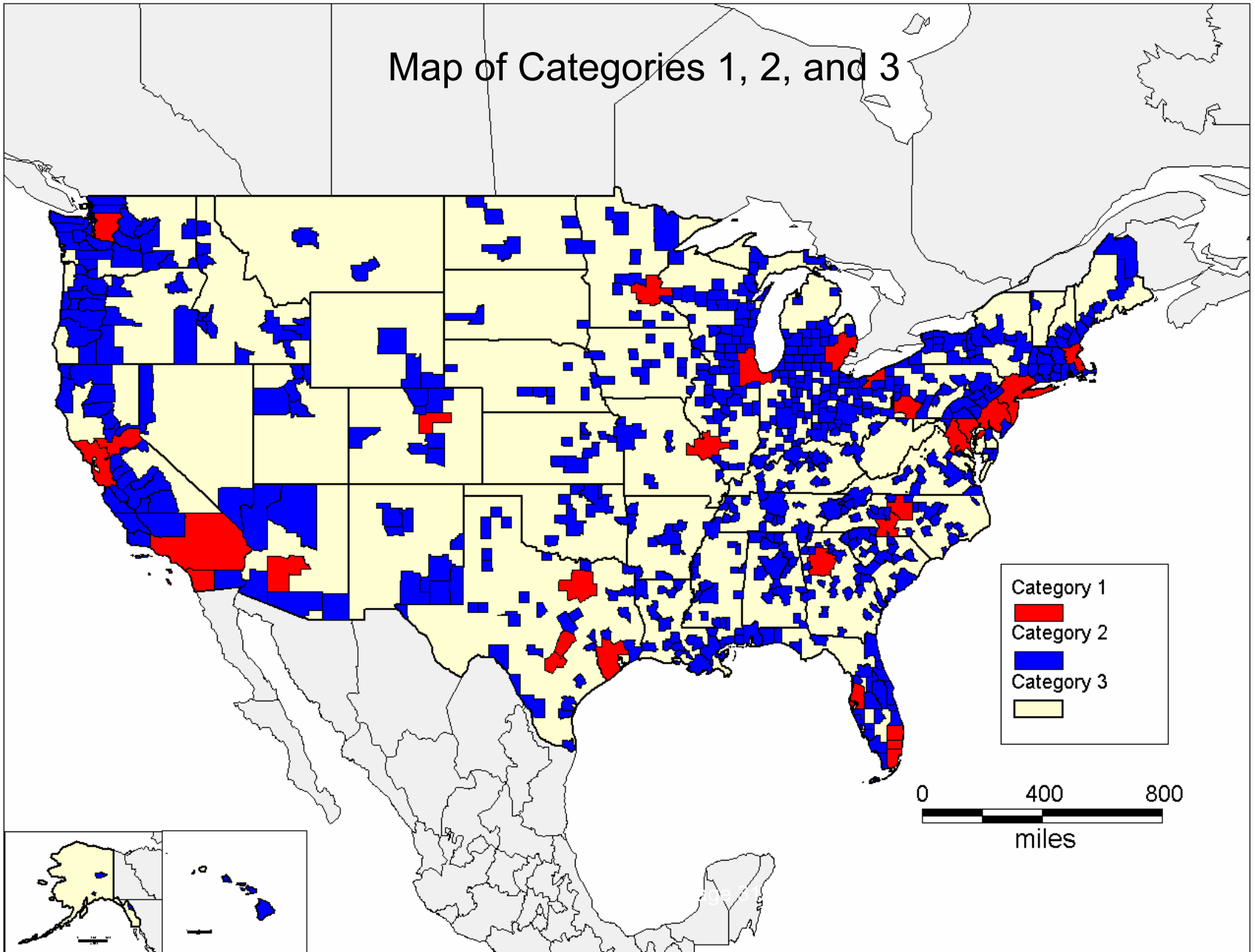
Small counties with less than 30% of population in type A tracts or less than 15,000 population [category 3]

All other counties/MSAs [category 2]

THE DESIGN – 2

- **Category 1**
 - 45% of population in 4.5% of the area
- **Category 2**
 - 40% of population in 25% of the area
- **Category 3**
 - 15% of population in 70% of the area

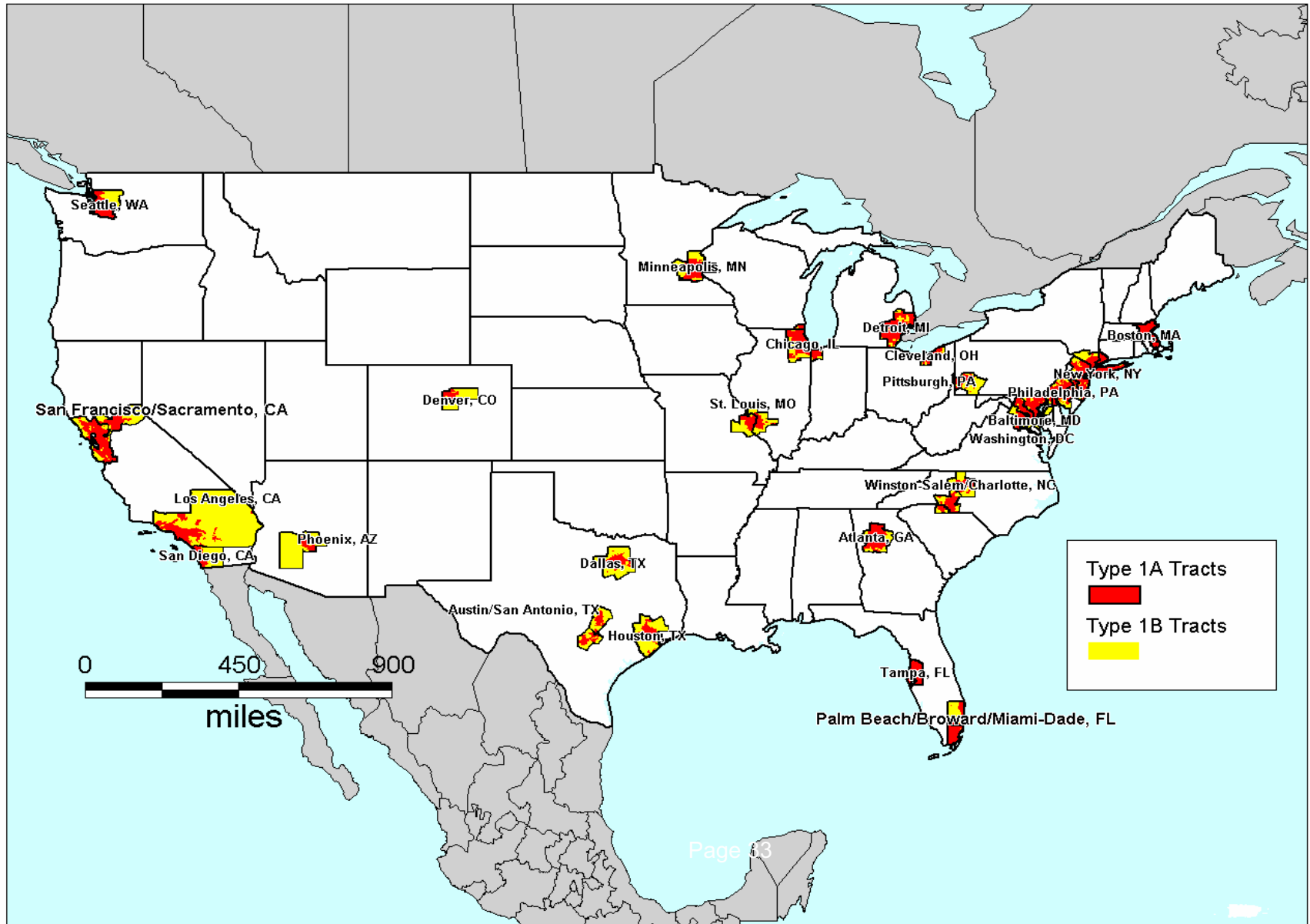
Map of Categories 1, 2, and 3



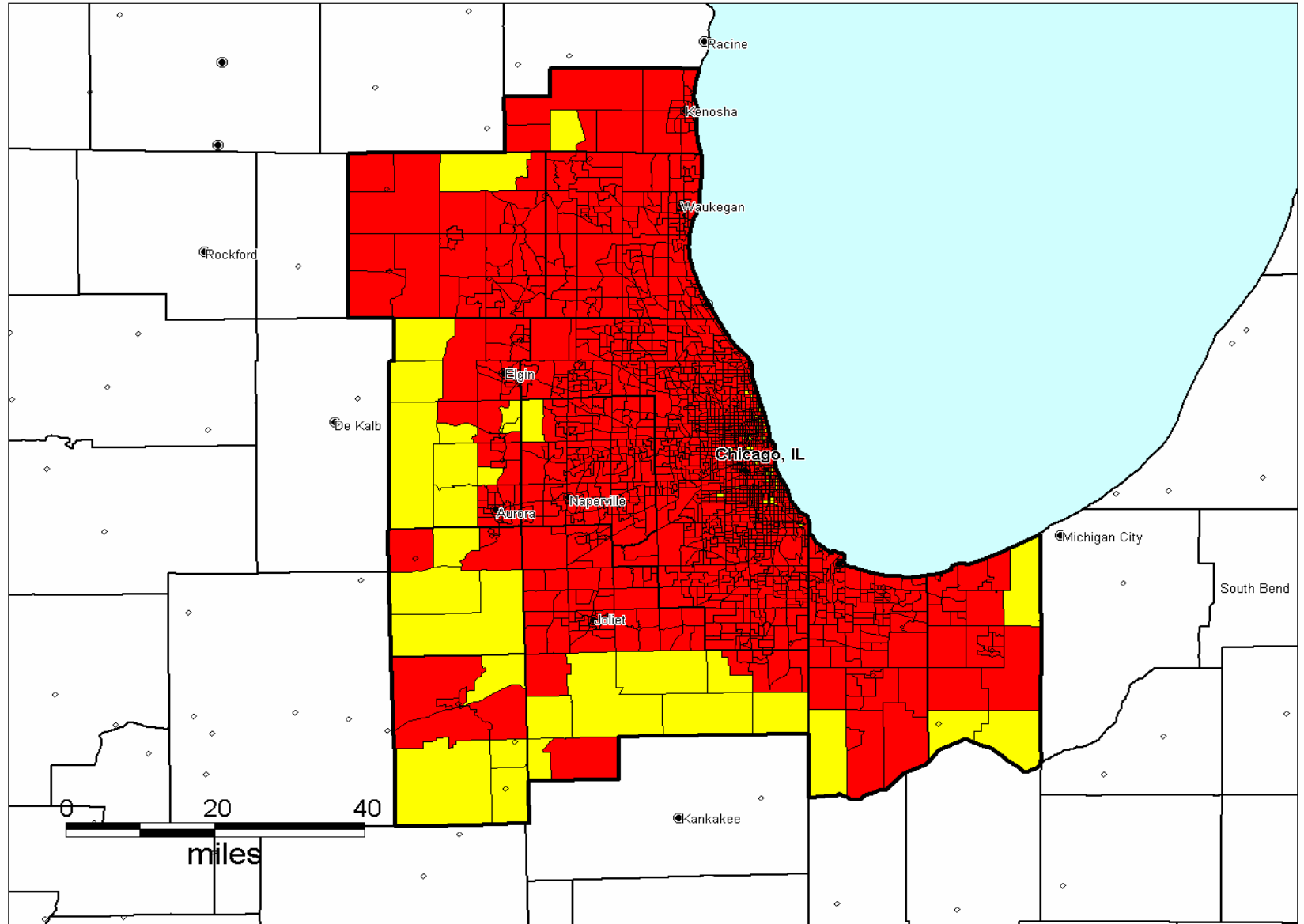
THE DESIGN – 3

- **Different designs are appropriate for the different categories**
- **A major problem:**
 - **Even in the high density urban MSAs rural (non-street-style address) areas are interspersed with urban (street-style address) areas**

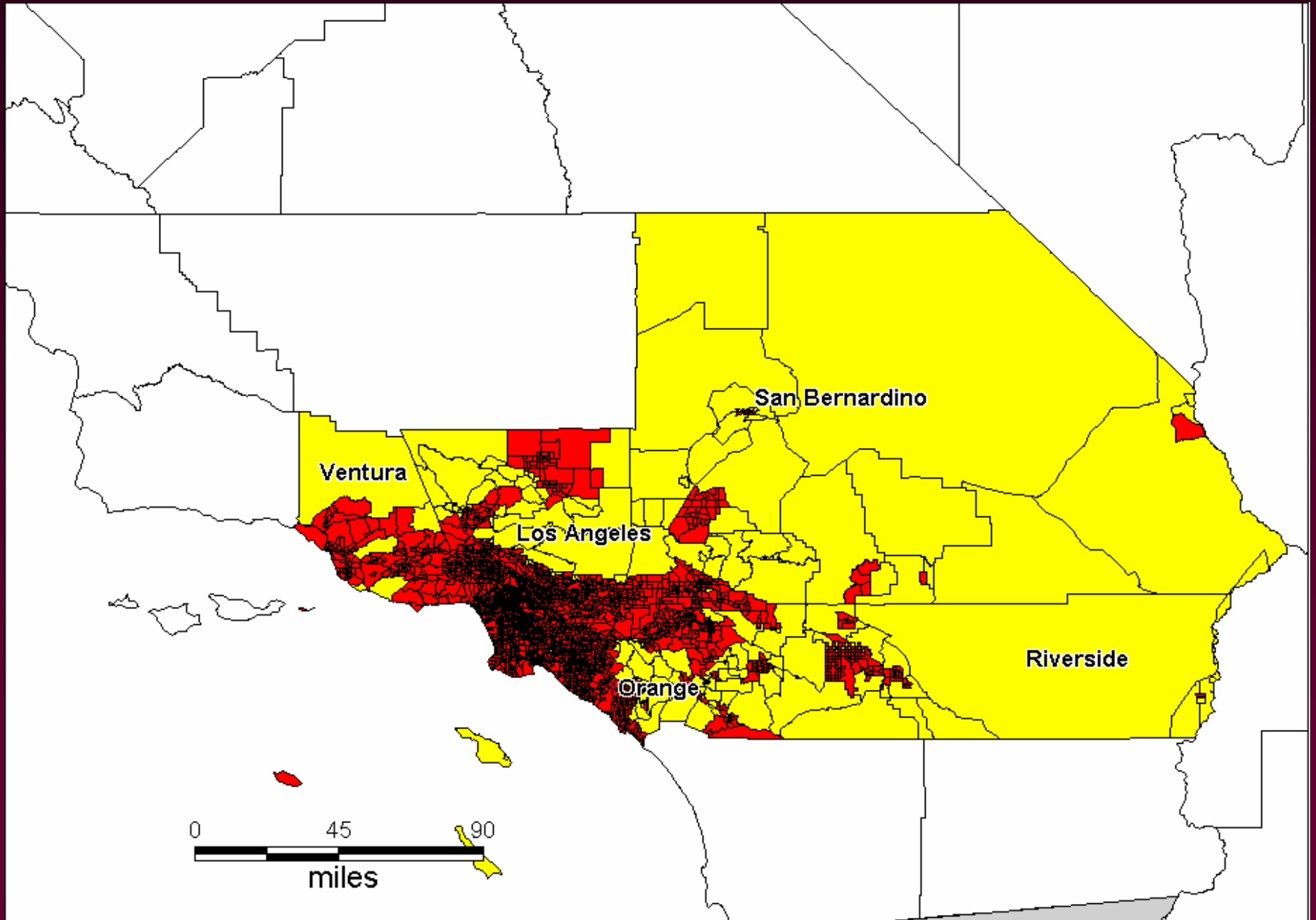
24 Category 1 Areas Showing Type A and Type B Tracts



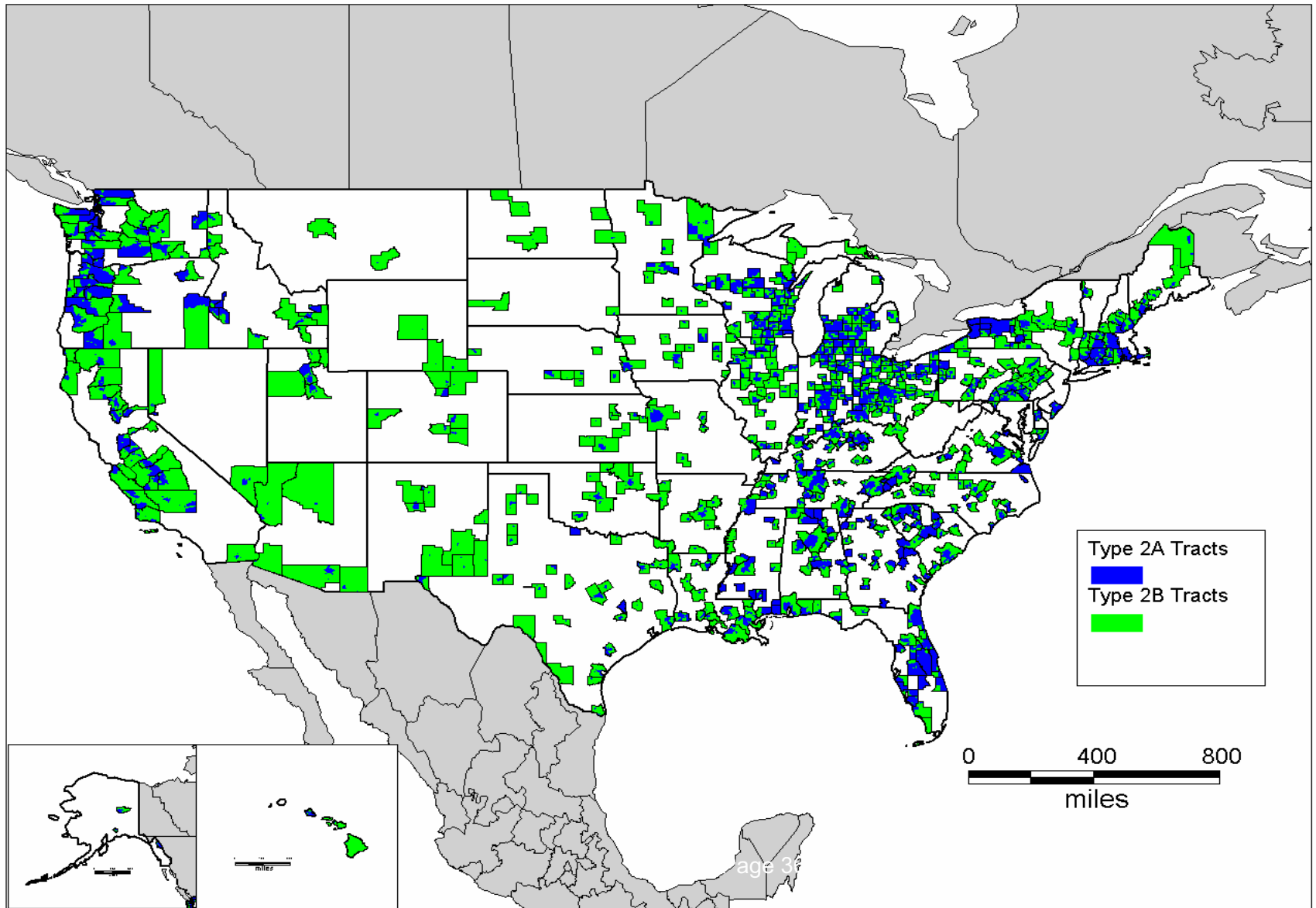
Chicago Category 1 MSA Showing Type A and B Tracts



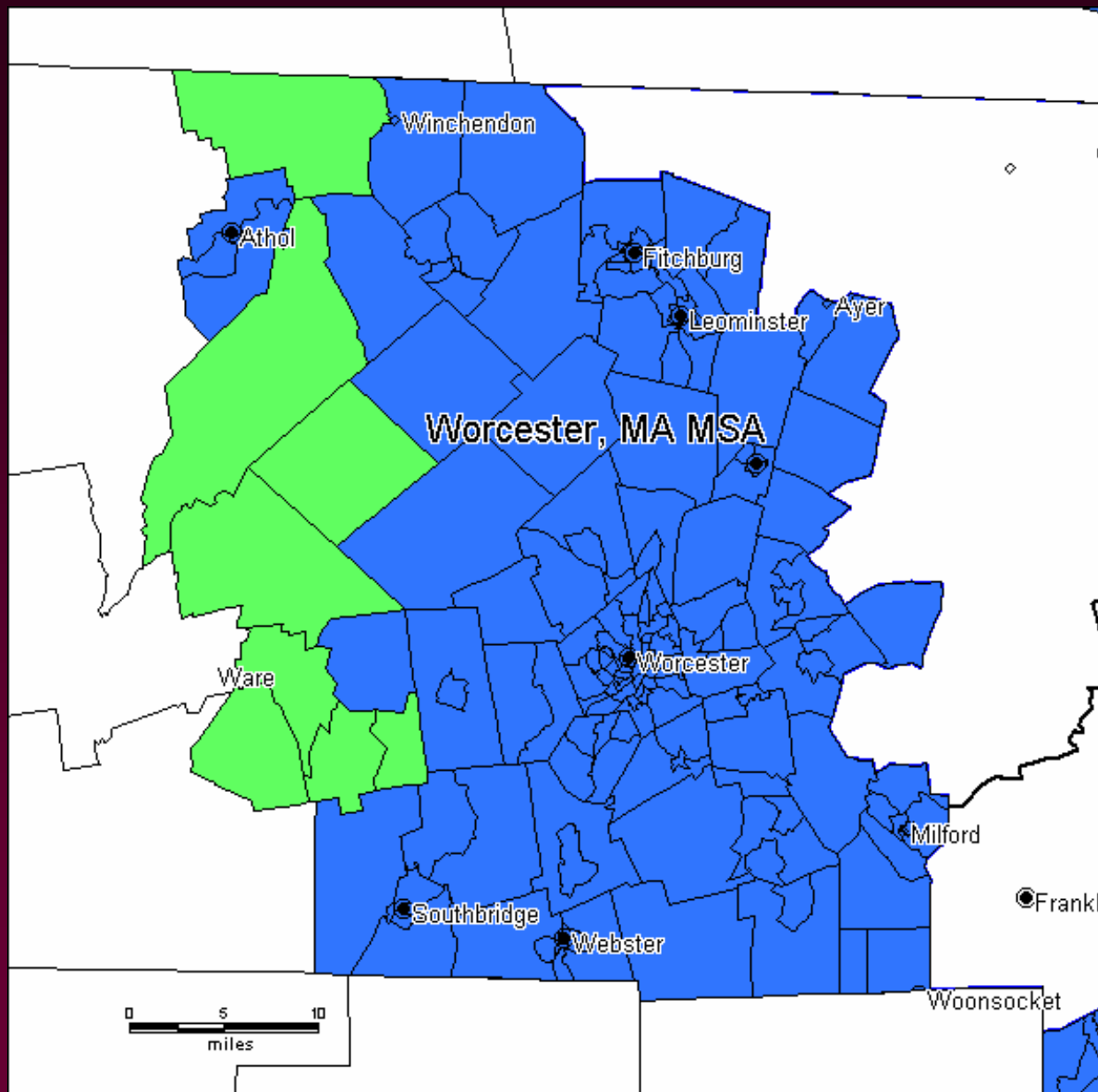
Los Angeles Category 1 MSA Showing Type A and B Tracts



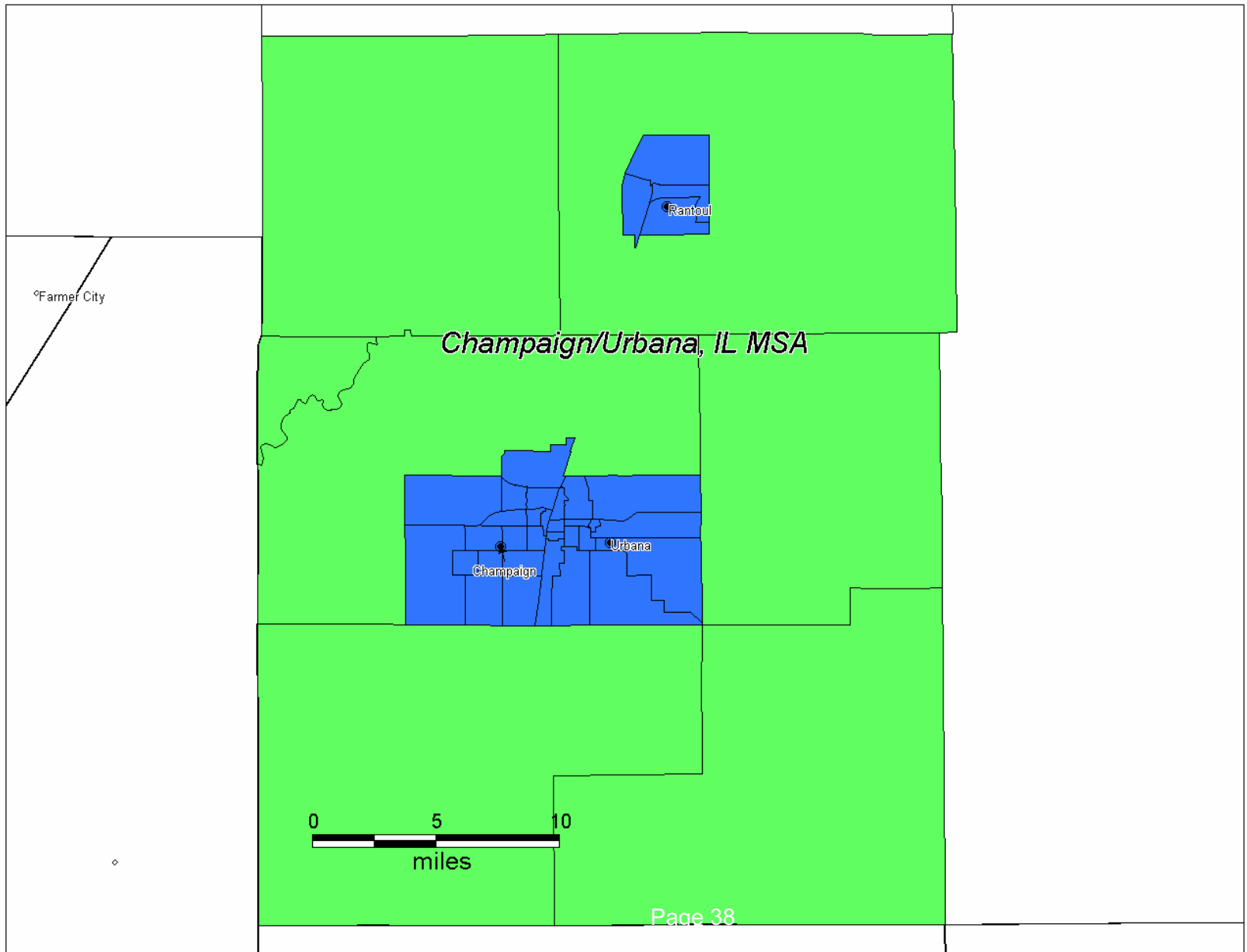
Category 2 Areas Showing Type A and B Tracts



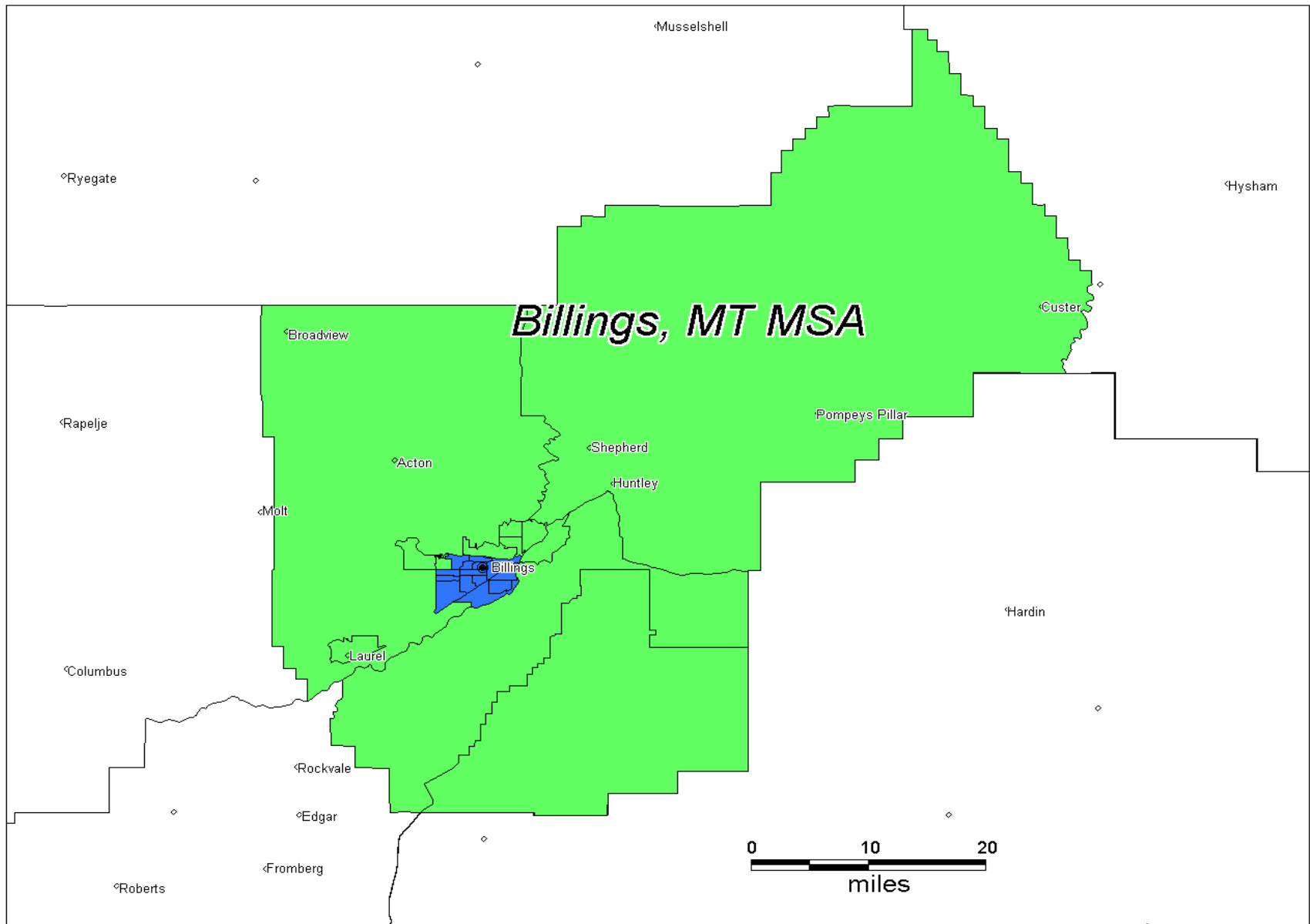
Type A and B Tracts In Worcester, MA [a category 2 MSA]



Type A and B Tracts In Champaign/Urbana, IL [a category 2 MSA]



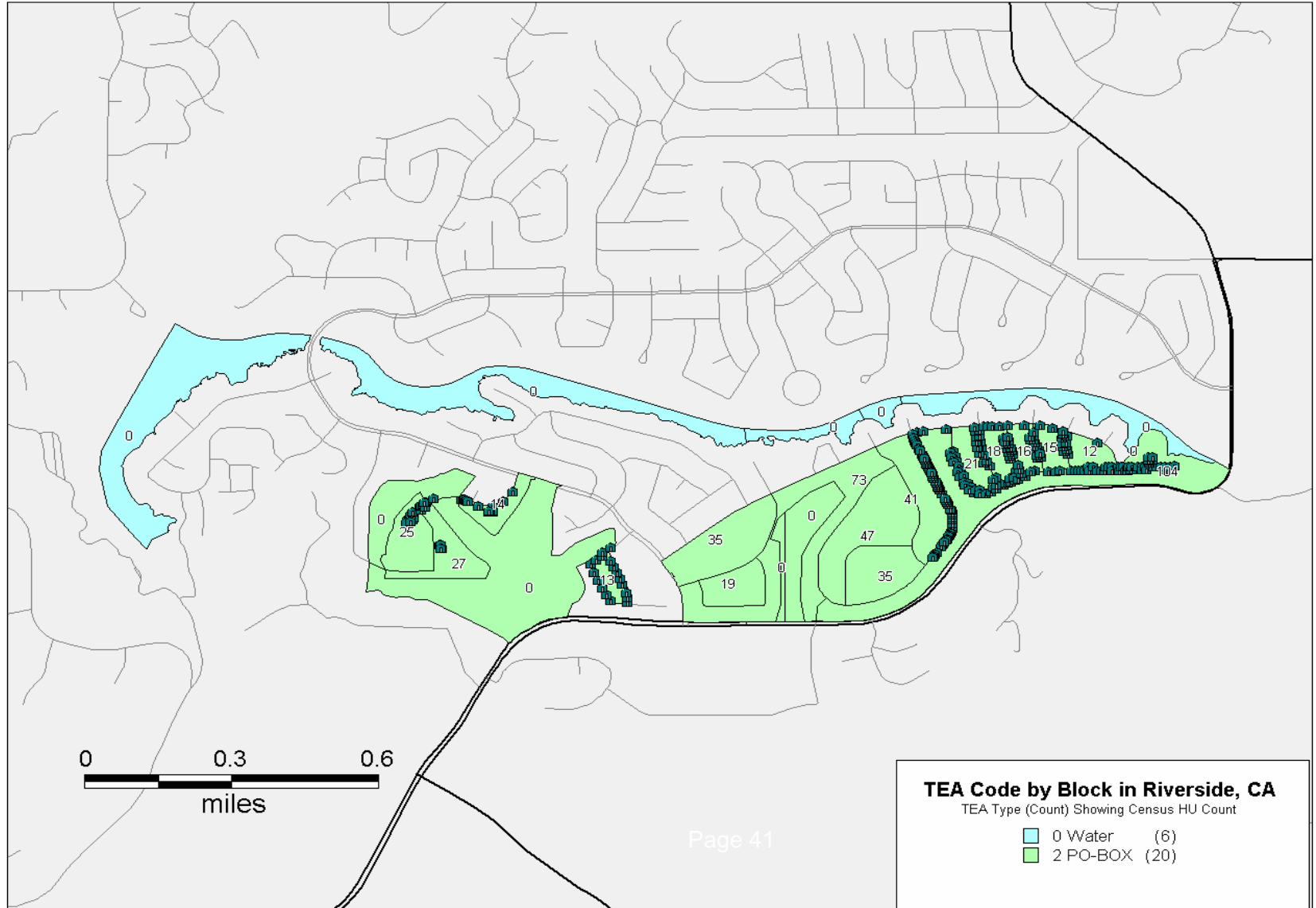
Type A and B Tracts In Billings, MT [a category 2 MSA]



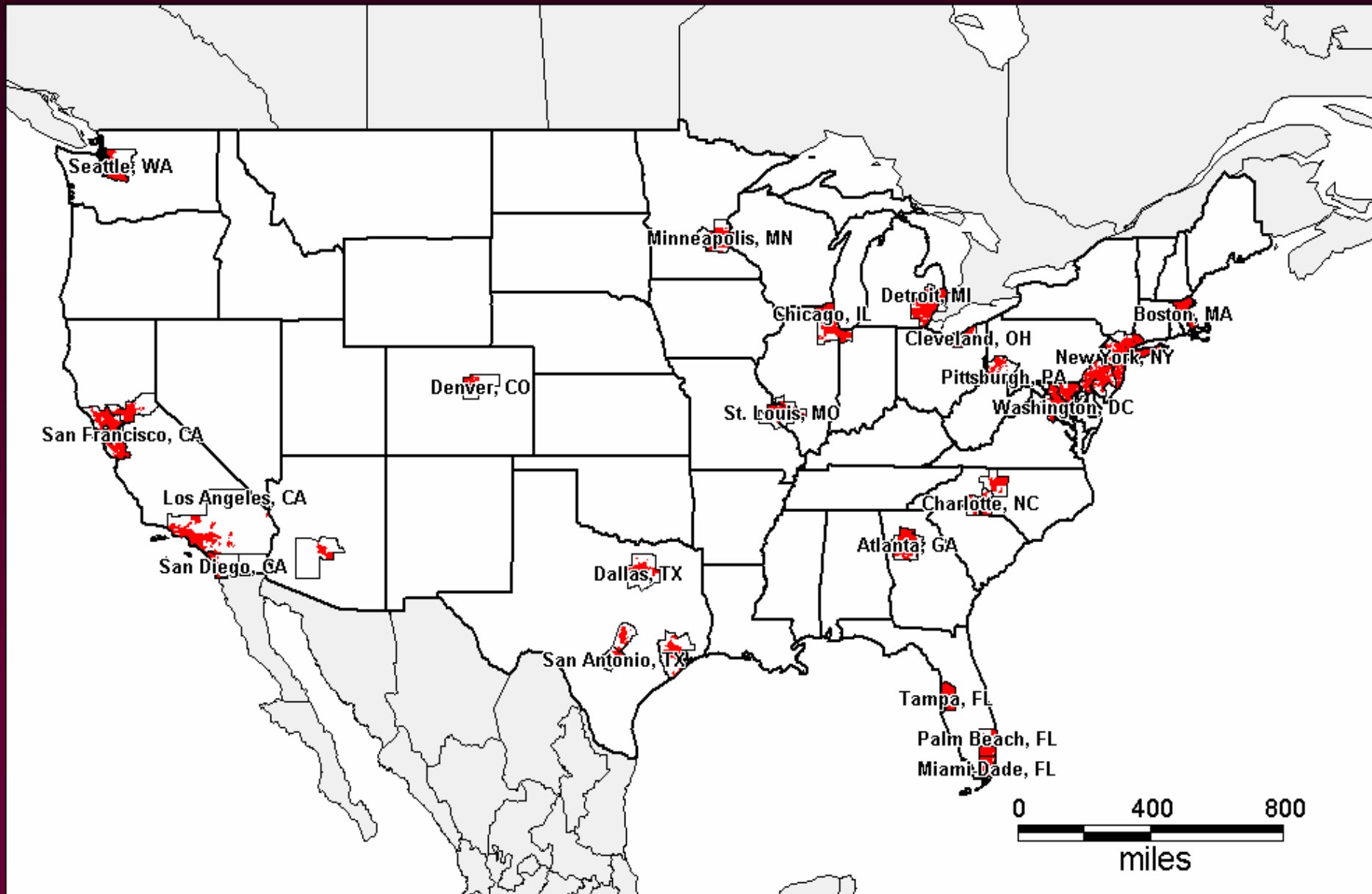
THE DESIGN SOLUTION

- **The Swiss cheese frame**
 - **Stratum 1 contains all type A tracts in category 1**
 - **In this stratum, the tract is the PSU**
 - **Stratum 2 contains all type A tracts in category 2**
 - **In this stratum the MSA/county is the PSU**
 - **All remaining tracts (category 1B, category 2B, and category 3)**
 - **In this stratum, the MSA/county is the PSU**
 - **Supplementary tracts from category 1B**

Type 1B Segment in Riverside CA, showing TEA Type, Census Count, and USPS Address Locations



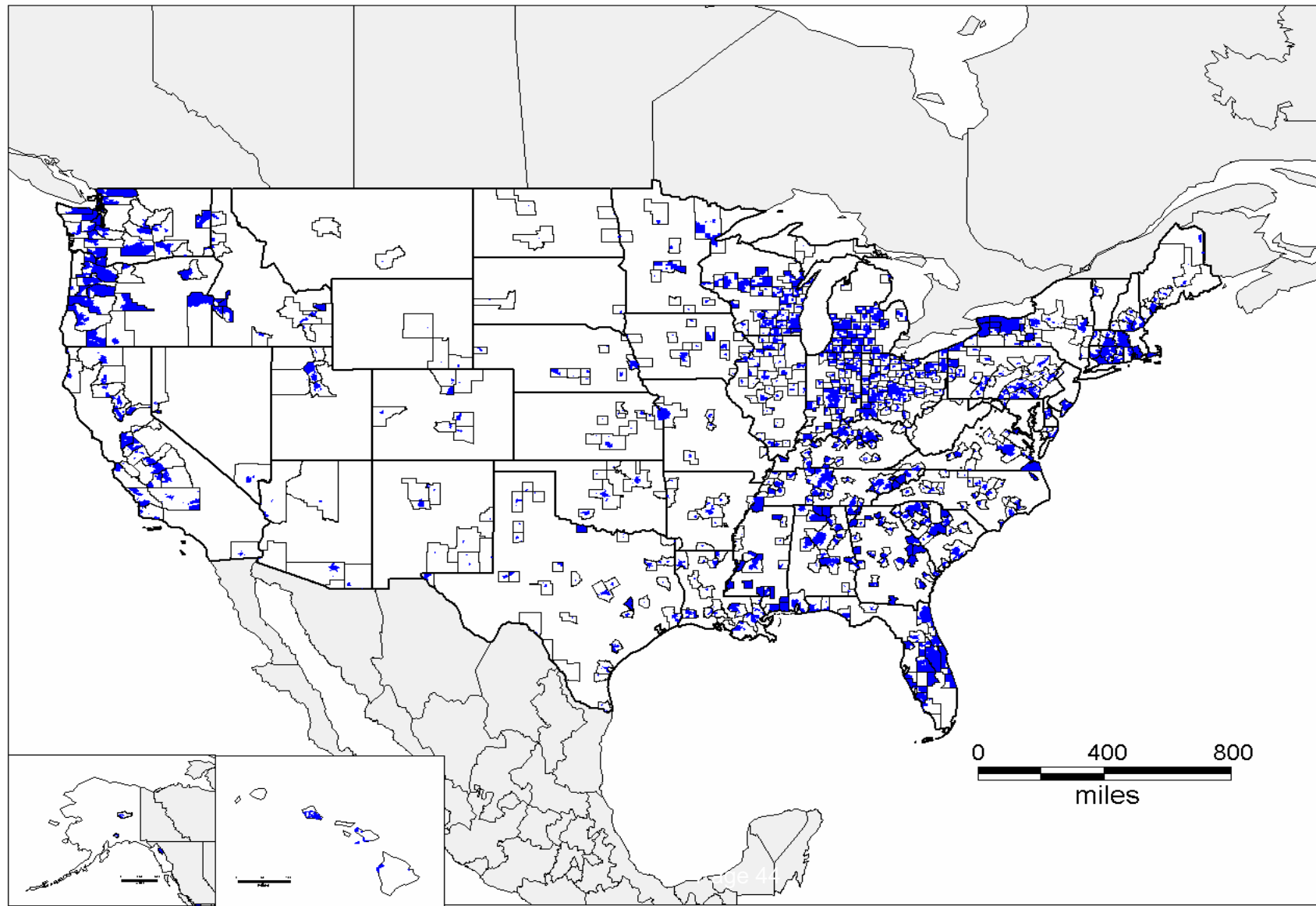
Stratum 1 – All Type A Tracts in Category 1 MSAs



STRATUM 1

- **42% of population, 2% of area, 24 certainty areas**
- **Direct selection of tracts as PSUs**
- **Contemporaneous USPS list with MHU procedures for HU selection**

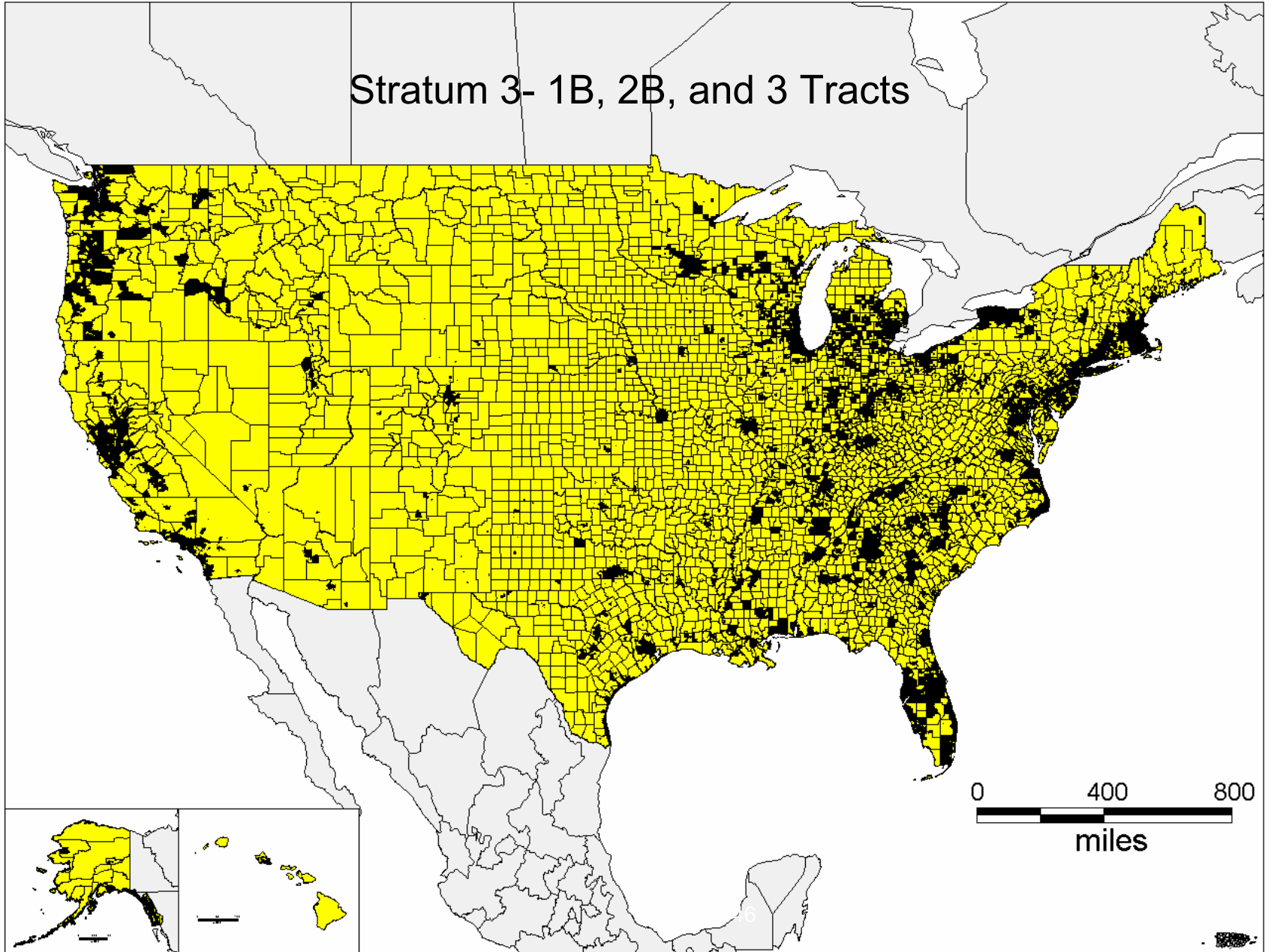
Stratum 2 – All Type A Tracts in Category 2 PSUs



STRATUM 2

- **30% of population, 3% of area, 607 MSAs/counties (or parts thereof)**
- **60 MSAs/counties (or parts thereof) as primary selections**
- **Selection of tracts as SSUs**
- **Contemporaneous USPS list with MHU procedures for HU selection**

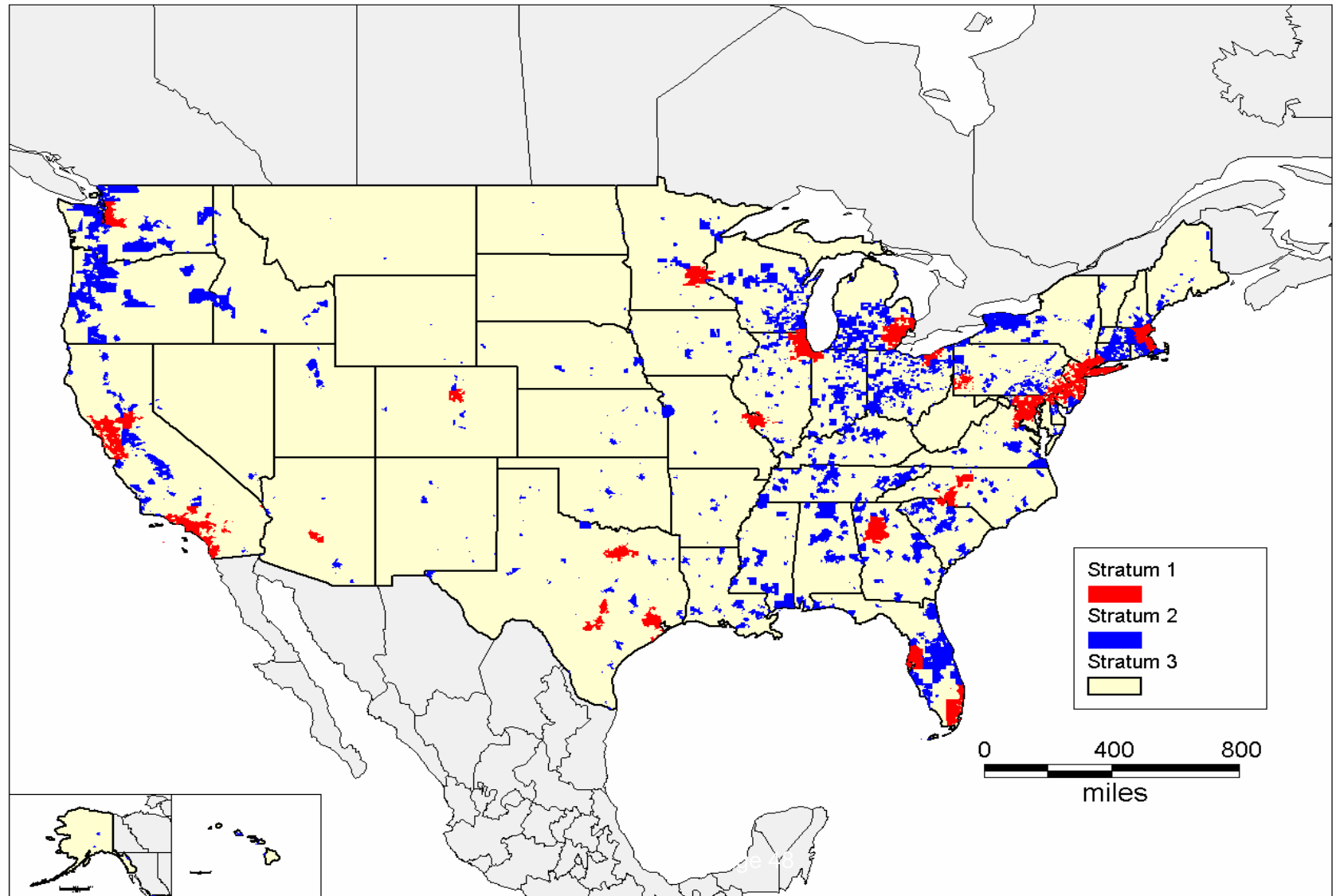
Stratum 3- 1B, 2B, and 3 Tracts



STRATUM 3 [composite of categories 3, 2B, and 1B]

- **28% of population, 93% of area, 3074 MSAs/counties (or parts thereof)**
- **Selected of 28 MSAs/counties (or parts thereof) as PSUs**
- **Constructed segments (blocks or groups of blocks) as SSUs**
- **Listed master sample of HUs within segments**
 - **Collect geocode during listing (GPS devices)**
 - **Reservoir for decade**

Map Showing Strata 1, 2, and 3



IMPLICATIONS OF LISTS FOR SAMPLE DESIGNS

- *Tailored samples vs Master samples*
- **Rural – no change from previous designs**
 - **Definition of rural?**
- **Non-rural**
 - **For timeliness, coverage, and cost, E superior to T**
 - **Is U superior to T?**
 - **Not desirable to construct very much in advance**
- **Non-rural can be extended as quality permits**

FEATURES OF NEW DESIGNS

- **Flexibility for tailored designs**
 - Accommodates modified stratification within strata 1 and 2 using ACS and/or other information during decade
 - Permits updates to HU frame using USPS lists
 - Allows different definition and number of PSUs per stratum depending on size of sample and precision requirements
- **Timeliness**
 - Can take advantage of any list upgrades or updates

THERE ...

- **19th Century**
 - **Multi-stage cluster sample of HUs**
 - **Stratified by urbanicity**
 - **Use of lists where possible**
 - **Selection from street addresses or registers**
 - **Designs tailored to specific projects**

- **Mid-20th Century**
 - **Area sampling as conceptual framework**
 - **Decennial listing/master samples**
 - **Re-design decennially**

... AND BACK AGAIN

- **21st Century**
 - Lists as frames
 - GIS/location as unique identifier
 - Designs differentiated by cost/feasibility
- **The Mechanisms**
 - Available (high) quality lists
 - GIS – identification and tracking
 - Pre-classification of geographies
 - Computer power

- **The Result**
 - **Tailored samples**
 - **Cheaper, better samples**
 - **Unnecessary uniformity minimized**
 - **Subject matter can inform sample design**
 - **Database linkages for analysis**

CHALLENGES

- **For designers:**
 - **Matching list geographies and census geographies**
 - **Better map data bases**
 - **Unique identifiers for addresses**
 - **Confidentiality/anonymity concerns**
- **For users:**
 - **Taking advantage of the potential**
- **Overall, most exciting time for sampling since Neyman in 1934 and the subsequent CPS design**