

# The Economic Value of Census 2000

## MSCDC Economic Report Series No. 2000-1

April 2000

By Douglas H. Bowles

Center for Economic Information, University of Missouri-Kansas City

### *Introduction*

Economic analysis frequently proceeds in terms of categories such as production, distribution, consumption, and investment. All of these categories are relevant to a discussion of the economic value of the decennial census of the United States.

More frequently, however, the economic value of the census is discussed in strictly distributional terms. This is understandable, since the rich informational resource of the decennial census is collected and tabulated under the mandate of Congressional legislation for the specific purpose of providing authoritative distributional guidelines for an entire range of domestic governmental spending programs. Consequently, a measure of the “distributional” economic value of accurate census data for any individual geographic entity--state, county, place, tract, block, etc.--could be measured in terms of per capita spending for any and all governmental programs whose distributional criteria utilize census data.

But while the production of decennial census data may be mandated by Congress to facilitate distributional governmental spending, that is certainly not the limit of its uses. Of course its principal use--constitutionally, rather than legislatively, mandated--is to provide the information necessary for allocation among the states of Congressional delegates to the U.S. House of Representatives. States, in turn, utilize the census to draw the boundaries for Congressional and state legislative districts. For this use, the minimal demographic data obtained from the census short form would be sufficient. But again, that is not the limit of the potential use of even short form demographic data.

Funded by the federal budget, the decennial census can be regarded as part of the aggregate output, or gross domestic product (GDP), of the U.S. economy. All output, or product, is eventually *consumed*, or “used up.” Some output--food, clothing, household furniture and appliances, automobiles, etc.-- is directly consumed by individual households in the more or less routine course of living. Other output--industrial machinery, for example--is indirectly consumed over time in the production process. Economists call that kind of output *capital*. The production of capital goods is called *investment*, so called because the use of capital goods in the production process will typically lead to an increase in output. It is this increase in real output associated with capital investment which underpins the financial concept of *a return on investment*.

Some governmental spending can be characterized as primarily *distributional* (or redistributive). That’s the kind of spending which yields a measure of the *distributional* value of census data. Other governmental spending provides *direct services*--such as the FBI, the Coast Guard, the U.S. Patent Office, or the National Park Service--to individual households, or to business firms, or to both. Most offices of federal, state and local government can be thought of as providing general *governmental services*. Finally, some governmental spending-- the national network of interstate and federal highways, the air traffic control system, and the internet, for example--is best characterized as *capital investment*. The extraordinarily rich

informational resource which the decennial census represents can be plausibly classified among these, as a form of capital investment in the U.S. economy.

This report will address the economic value of the census in terms of both its distributional and its investment value.

### ***The Distributional Value of the Census***

#### *A quantified measure of per capita census-based government spending*

A recent study conducted by the federal Office of Management and Budget found 133 instances where federal congressional legislation required the use of census data in funding formulae.<sup>1</sup> In an article written for the *Missouri County Record* (March/April 1999), Ryan Burson and Debbie Pitts of the Missouri State Census Data Center (MSCDC) identified 122 instances as of FY96, with a total funding obligation of \$182 billion, or roughly \$685 per person.<sup>2</sup> Deployed in conjunction with the problem of *undercount* (or conversely, *overcount*), this quantification of per capita distributional governmental spending provides a rough measure of the distributional value of the decennial census.

#### *The problem of differential undercount*

No decennial census of the U.S. population, has ever been 100 percent accurate. The first census, supervised in 1790 by Thomas Jefferson, was disputed by President Washington, who believed that the count of 3.9 million was too low.<sup>3</sup> The inevitable inaccuracy of the census was acceptable, however, so long as it was assumed to apply proportionately across the population. In the period immediately following the 1940 census, as the nation was undergoing massive mobilization for World War II, it came to the attention of governmental administrators and statisticians that the numbers of African-American men reporting for military service were in significant excess of what had been anticipated based on the 1940 census counts.<sup>4</sup> This anomaly was the first indication of the problem now referred to (and extensively studied) as the *differential undercount*, which simply means that different demographic sectors of the general population are missed by the census count *disproportionately to their real proportion* of the general population. Additionally, where the demographic sectors of the population most susceptible to the problem of undercount are concentrated, the differential undercount problem manifests itself in geographic terms. This means that states, counties, incorporated cities and towns, urban neighborhoods, and urban or rural ethnic enclaves, are all variously affected by the problem of differential undercount. In this context, the inevitable inaccuracy of the census can no longer be dismissed as a problem, since the simple assumption that it applies proportionately across the population can no longer be plausibly asserted.

---

<sup>1</sup> This document can be found on the Census Bureau's website at [www.census.gov/dmd/www/subjects.html](http://www.census.gov/dmd/www/subjects.html). Look for Appendix A.

<sup>2</sup> A reprint of this article can be found on the MSCDC website, at [www.oseda.missouri.edu/mscdc/benefits.html](http://www.oseda.missouri.edu/mscdc/benefits.html).

<sup>3</sup> See "Decennial Census: Overview of Historical Census Issues," a GAO report linked from the C-SPAN website at [www.cspan.org/census2000](http://www.cspan.org/census2000). Follow the text link, "GAO Report to Congress." Also, see the interview with Bureau Director (at the time) Dr. Martha Farnsworth Ritchie, printed in Vol 32, No.10 (October 1997) of the Bureau's monthly newsletter, *Census and You*. A copy is posted on the Bureau's website at [www.census.gov/prod/www/abs/gen-ref.html](http://www.census.gov/prod/www/abs/gen-ref.html).

<sup>4</sup> *Census and You*, October 1997. See earlier Census Bureau website address citation.

When we have quantified a measure of per capita distributional government spending based on census data, as was done above, then we can use it to measure the *cost of the undercount* for any demographic sector of the population, or geographic entity within the U.S.

### *Evaluating the cost of differential undercount in Missouri in the 1990 census*

Since discovery of the problem of differential undercount following the 1940 census, in an effort to understand and measure its scope, the Census Bureau has quantified the amount by which each census undercounts the population. In 1990, the Bureau used the information obtained from a *post-enumeration survey* to produce a set of adjusted tabulations corrected for differential undercount. This decision, which resulted in considerable political controversy that has been carried forward to the design and implementation of Census 2000, gives us the ability to quantitatively estimate the cost of the undercount for demographic population sectors, and for geographic regions. Both the official and the adjusted data for the 1990 decennial census have been posted on the Bureau's website. (The format in which they have been posted, however, is not very conducive to wholesale download in formats that are easily amenable to analysis. The datasets which form the basis for this report were obtained from that site, and are posted here in Microsoft Excel spreadsheet format.)

Our geographic analysis is focused on the State of Missouri. We look at data for 115 Missouri counties, and for 961 incorporated places (cities and towns). Additionally, we have included data for 5 Illinois and 4 Kansas counties which are part of the St. Louis and Kansas City metropolitan statistical areas (MSAs).

The undercount is defined as the adjusted (or estimated) count minus the official (or enumerated) count. Undercount rates are defined as the proportion of the estimated undercount to the adjusted count. In some cases, the estimated undercount is negative, which also yields a negative undercount rate. Negative undercounts and undercount rates are referred to as overcount. We identify both undercounted and overcounted areas.

It's important to keep in mind that high undercount rates for geographic areas or for demographic sectors, while primarily indicative of a disproportionate burden of the undercount problem, can be derived from quite small absolute errors if the population base under consideration is relatively small. Thus, it is important to look at both absolute, or raw undercounts, as well as undercount rates in order to adequately understand the scope of the problem.

The data is presented geographically in six Map Panels. Map Panel 1 presents the raw undercount for Missouri counties, MSAs, and incorporated places for the Total (Map 1a), White (Map 1b), Black (Map 1c), and Hispanic (Map 1d) populations. Map Panel 2 presents undercount rates for the same geography and demographic sectors. Map Panels 3 and 4 present the same data for the St. Louis MSA, and Map Panels 5 and 6 do the same for the Kansas City MSA.

Official and Adjusted 1990 data posted at the Bureau's website is also provided for the 18-and-over age demographic sector, since it is also widely known that children (defined for this purpose as persons under age 18) are also highly susceptible to differential undercount, even controlling for racial/ethnic group differentials. County-level age-18-and-over data is included in the downloaded dataset. This study, however, did not attempt to include any analysis of the under-age-18 demographic sector.

A cursory glance at Map Panel 1 provides an overview of the raw differential undercount for Missouri counties, MSAs, and incorporated places. Individual Maps 1a, 1b, 1c, and 1d provide a

more detailed perspective for each demographic sector. Looking at Map Panel 1, note that the areas of high color intensity reflect high raw undercounts. The points of highest intensity, not surprisingly, are in the two most densely populated areas of the state, the St. Louis and Kansas City MSAs. Note also that the overall distribution of undercount intensity is very similar (excluding overcounts) for Total, White, and Black populations; while there is some visually detectable (and incidentally, necessary) decrease of intensity from Total to White, and from Total to Black, any dissimilarity of intensity distribution between White and Black is not readily apparent to even a studied visual inspection. This tells us at a glance that the raw undercounts between White and Black demographic sectors are of similar orders of magnitude. A quick look at Map 1d also tells us that this similarity does not hold for Hispanics. Map 1d presents a dramatic decrease in the intensity distribution, reflecting a much smaller overall magnitude of raw undercount for Hispanics, compared to Whites and Blacks. This is consistent with our understanding that the Hispanic population in Missouri in 1990 was a much smaller proportion of the Total than that of Whites, and of Blacks as well.

Map Panel 2 presents the undercount rates for Missouri counties, MSAs, and incorporated places. Keeping our previous insights obtained from Map Panel 1 in mind, the principal point of interest to note in Map Panel 2 is the very low intensity of undercount rates for both Total and White populations. Both urban and rural counties, for the most part, are either overcounted, or display undercount rates well under 1 percent, and in many rural county cases, well under 0.1 percent. The incorporated places which form the urban core of the St. Louis and Kansas City MSAs display undercounts rates largely falling on the low side of the 1 to 5 percent range. In sharp contrast to the very low intensities displayed for Total and White populations are the dramatically high intensities reflected in Maps 2c and 2d for the Black and Hispanic populations. We can also note from simple visual inspection that there is a clear increase in the intensity distribution moving from the Black to the Hispanic population. Evaluated alone, these maps show us very clearly how the Black and Hispanic populations in Missouri bear a significantly disproportionate cost of the differential undercount, with the share of the cost borne by Hispanics exceeding even that borne by Blacks. However, remembering the observations obtained from examination of Map Panel 1, we can also note that the raw undercounts for Blacks were overall much higher (approaching those for the White population) than those for Hispanics, whose total population in Missouri was relatively small as of 1990. Thus, it is entirely appropriate to observe that, considering the matter in both absolute and proportionate terms, the cost of the differential undercount borne by the Black population in Missouri far exceeds that of any other racial/ethnic sector.

At this point, however, it is important to note that all indications from the intercensal period of the 1990s are that the Hispanic population in Missouri has increased significantly. It will be interesting to see what this kind of analysis will reveal with undercount data from the 2000 Census in respect of that increase.

Map Panels 3 and 4, which zoom in to the St. Louis MSA, and Map Panels 5 and 6, for the Kansas City MSA, allow us to perform similar analysis for the individual counties and incorporated places which comprise them. Links to individual maps are provided below, for ease of navigation. Readers are invited to explore the implications of this mapping for themselves.

Map Panel 3: St. Louis, raw undercount  
Map 3a: Total Population

Map Panel 4: St. Louis, undercount rates  
Map 4a: Total Population

Map 3b: White Population  
Map 3c: Black Population  
Map 3d: Hispanic Population

Map 4b: White Population  
Map 4c: Black Population  
Map 4d: Hispanic Population

Map Panel 5: Kansas City, raw undercount

Map 5a: Total Population  
Map 5b: White Population  
Map 5c: Black Population  
Map 5d: Hispanic Population

Map Panel 6: Kansas City, undercount rates

Map 6a: Total Population  
Map 6b: White Population  
Map 6c: Black Population  
Map 6d: Hispanic Population

### *The Investment Value of the Decennial Census*

To provide a comprehensive discussion of the economic value of the census, its distributional use must be viewed in the larger context of its innumerable other actual and potential uses. It is this multitude of uses that justifies its treatment as a capital investment in the U.S. economy. Even its use in distributional government spending has significant human capital investment implications when we consider the extent to which it helps allocate available government resources efficiently and effectively to otherwise formidably disadvantaged segments of the population.

Uses (not necessarily mutually exclusive) for data collected in the decennial census include:

- ❑ constitutionally mandated apportionment among states of seats in the U.S. Congress
- ❑ determination of state congressional district boundaries
- ❑ distributional, especially including allocation of government and non-profit sector resources to relatively disadvantaged segments of the population
- ❑ baseline data for numerous other governmental and private sector data collection programs
- ❑ business expansion, location, and marketing decisions
- ❑ public, private, and non-profit sector urban and rural planning
- ❑ federal, state, and local governmental administration
- ❑ community and neighborhood development
- ❑ workforce development
- ❑ academic research
- ❑ disaster preparedness and response
- ❑ Geographic Information Systems (GIS) analysis, utilizing TIGER<sup>5</sup> street files and TIGER-based area boundary files developed by the Census Bureau to facilitate the census process
- ❑ public policy debates & decisions

---

<sup>5</sup> TIGER is an acronym for the Topographically Integrated Geographic Encoding and Reference system, which has been developed by the Census Bureau to support its mapping needs for the decennial census and other Bureau programs.

Real capital investment (the production and employment--i.e., *use--*of capital goods) increases real productivity (the level of output of all goods and services). It is the increase in real productivity brought about by capital investment that underlies the financial concept of a return on investment, which is typically employed in the private sector to facilitate investment decisions.

Return on investment is simply the discounted present value of a future revenue stream which is expected to be earned by the real productivity of a proposed investment. In order to judge the desirability of the proposed investment, its expected return must be compared to its opportunity cost--what it would earn in its best alternative use. Any investment's opportunity cost is typically taken to be the going rate of interest. So if the proposed investment's rate of return (technically referred to as the *internal rate of return*, or *IRR*) exceeds the going rate of interest, then the investment will "pay off." Of course, since real and financial resources available for investment are presumably limited, it may be necessary to choose among alternative investment proposals, all of which may pay off to some degree. Such choices can be made (in part on in whole) based on a comparison of their respective rates of return. Or in other words, if choices (as a result of limited resources) must be made among a variety of investment proposals all with positive pay offs, the opportunity cost for any particular proposal will exceed the going interest rate.

Can we calculate such a rate of return for past censuses? Or, after the fashion of business investment practice, can we estimate the expected rate of return for Census 2000? This could be done if we know (or can estimate) two pieces of information: 1) its cost; and 2) the future revenue stream realized (or expected) from it.

Historical and cost calculations for the decennial census are readily available. A GAO report cited earlier in this article estimated the cost of the 1990 census at \$2.6 billion, and the expected cost of Census 2000 to reach 4.0 billion.<sup>6</sup> The real difficulty lies with estimation of an expected future revenue stream against which to evaluate the cost. Total revenue is a function of the price paid for total quantity purchased. Since data from the decennial census is a public good, provided to the public at no charge, the decennial census does not generate a revenue stream that can be estimated for the necessary calculation.<sup>7</sup>

However, it is possible to finesse this difficulty if a plausible method can be proposed by which to impute a revenue stream to the use of census data. We propose to do this by using the prices charged by private sector vendors of intercensal estimates of population, demographics, and socio-economic characteristics. A reasonable current approximation of that price would be around \$.50 per datapoint, or item.<sup>8</sup> At \$.50 per datapoint, private sector vendors are presumably able to cover costs and earn a competitive return on investment. This is probably too low a price, we could argue, to use as the basis for imputation of a revenue stream from the decennial census, since private sector costs for such estimates are already subsidized by the existence of baseline decennial census data. Of course, where real markets exist,<sup>9</sup> prices and quantity demanded (and therefore total revenue) are simultaneously determined. However, so long as we can plausibly estimate the level of use, or quantity demanded, for decennial census data (which can be done

---

<sup>6</sup> "Decennial Census: Overview of Historical Census Issues," linked from the C-SPAN website at [www.cspan.org/census2000](http://www.cspan.org/census2000). Follow the text link, "GAO Report to Congress."

<sup>7</sup> This remains true despite the Census Bureau's current effort to implement a mandate from the Clinton/Gore administration to attempt cost recovery through charges for various Bureau data products.

<sup>8</sup> Based on prices charged by Claritas under recent contract with MSCDC.

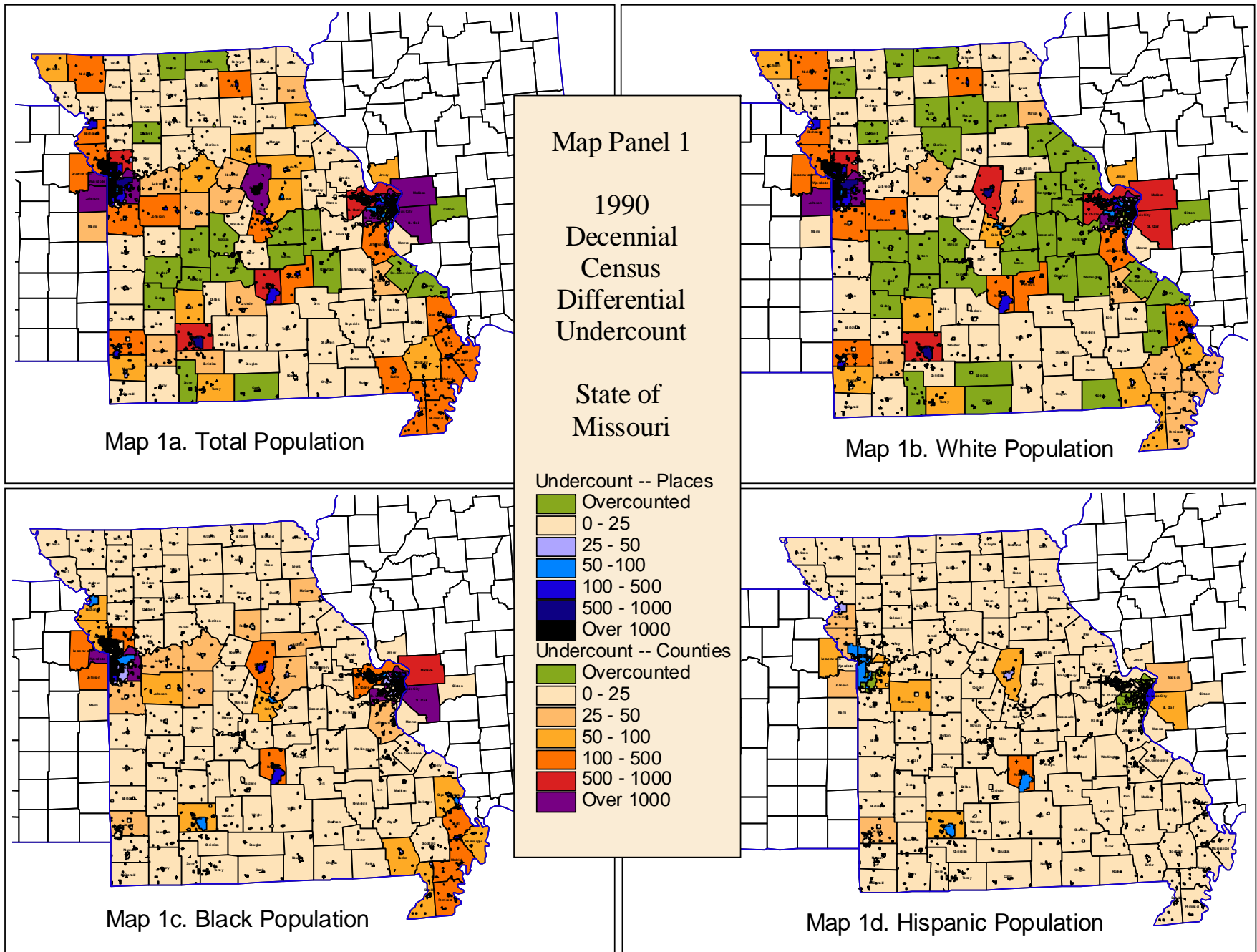
<sup>9</sup> The extent to which "real" markets can be said to exist is, also of course, a matter of hot debate among orthodox and heterodox economists.

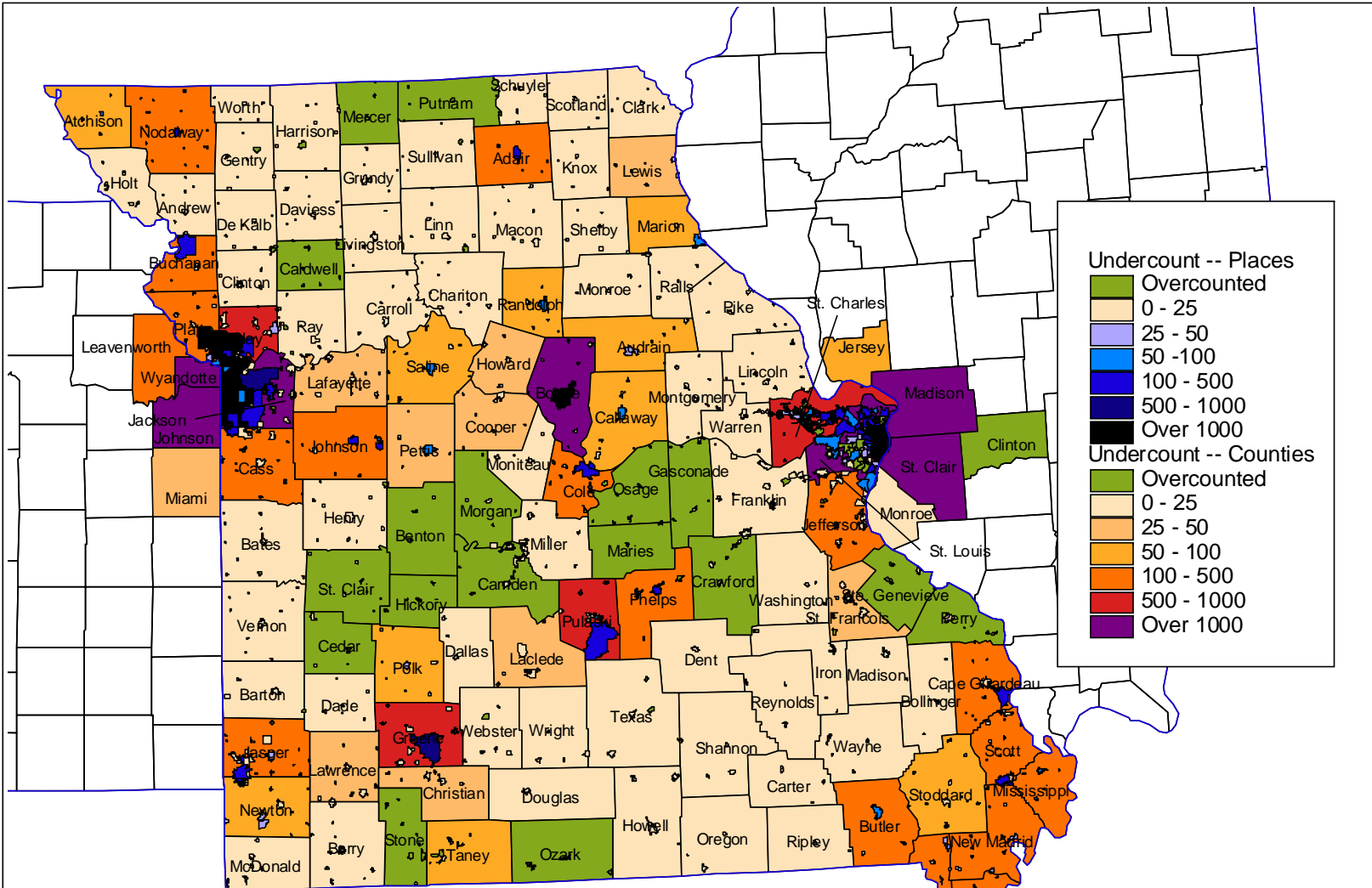
independently of price, since we are not constrained to the use of a market model) \$.50 per datapoint seems a plausible starting point. In other words, we are asserting that the use of a census datapoint is *minimally* equivalent to \$.50 worth of capital investment in the U.S. economy. But what about the level of use (or *quantity demanded*) at that price?

We estimate (roughly) that the tabulation of both short and long form data for the entire heirarchy of census geographies (U.S., state, county, place, tract, block group, etc.) contains approximately 1,173,000,000 (1.173 *billion*) datapoints. Generally speaking, we are talking about what has in the past been called Summary Tape File 3 (STF3), creation and distribution of which originated in the era of reel-to-reel mainframe tape drives. This is often shortened now to Summary File 3 (SF3). It is the most comprehensive tabulation of census data available, but numerous other tabulations are also produced and in use. In addition, any reasonably comprehensive estimation of the use of census data would also have to explicitly consider use of the *public use microdata* files (PUMS), and TIGER geography. But again, if we base our initial estimate on the use of SF3, we can argue that it will give us a *minimal* estimation.

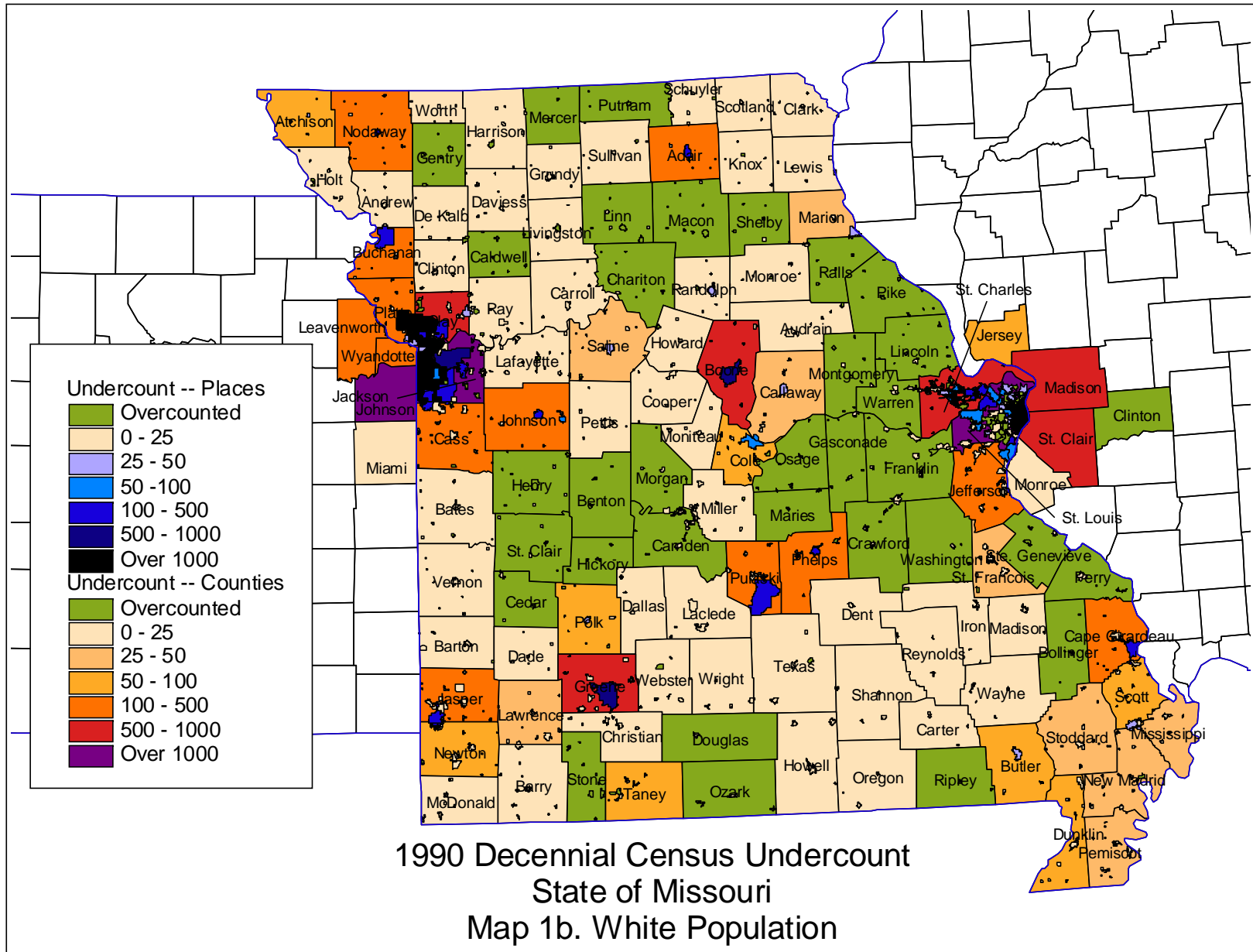
Suppose we were to assume, again *minimally* (reflect on the list of uses for census data provided at the beginning of this section), that the datapoints comprising the SF3 tabulation of the 1990 decennial census were to be used just *one time* every year for the ten years comprising the intercensal period. (This is also a minimal estimate in the sense that decennial census data continues to have use beyond its ten-year intercensal life, in the context of trend analysis and historical studies.) That is the final piece of information we would need to compute a minimal return on investment for the cost of the 1990 decennial census. Given the set of assumptions and estimates we have just laid out, the minimal internal rate of return (IRR) for the \$2.6 billion investment in the 1990 census would be *18.39 percent*. If use of the SF3 datapoints increased to *two times* every year, the IRR would increase to *43.94 percent*. Needless to say, 18.39 percent is a very respectable rate of return, while 43.94 percent could probably be called a “no-brainer.”

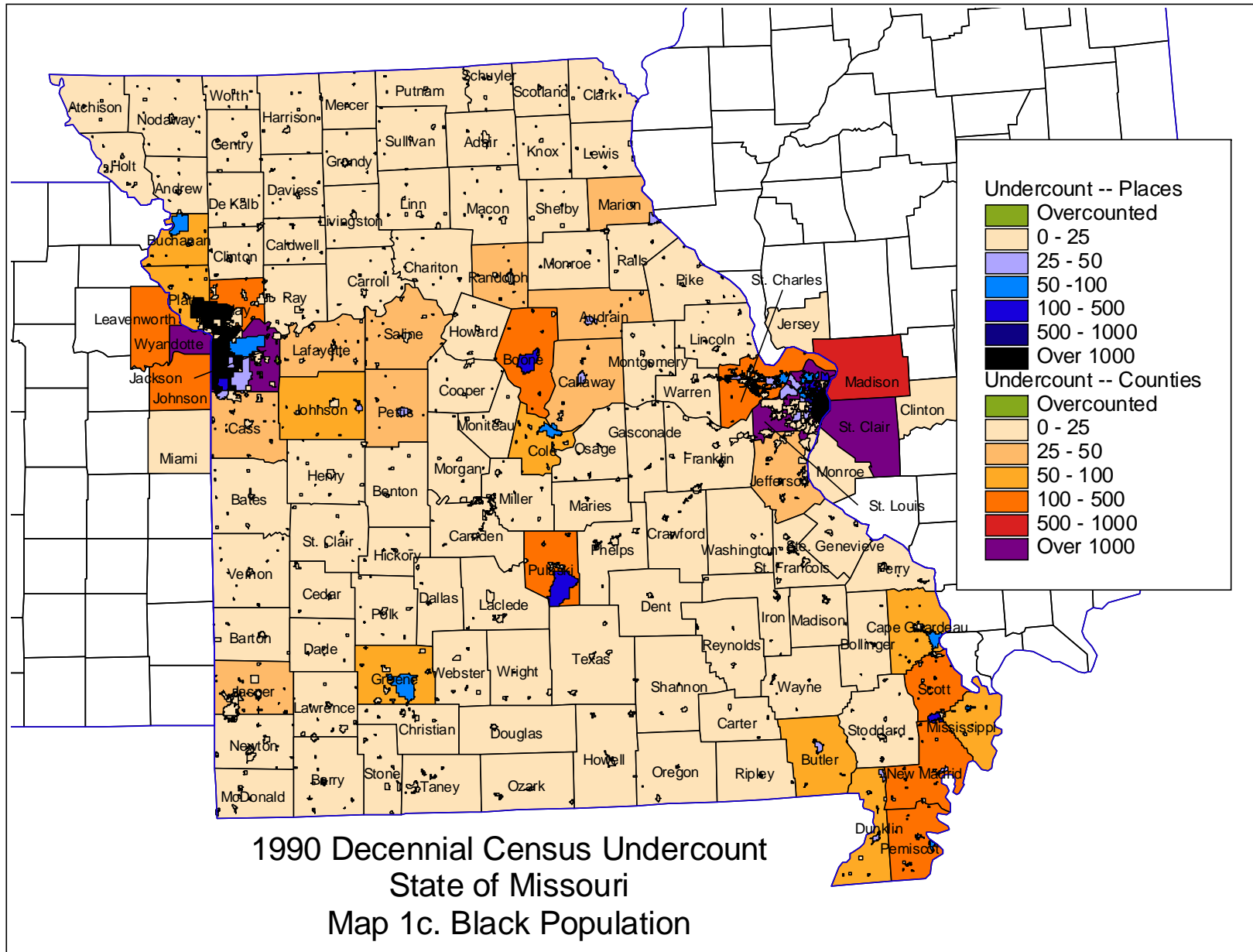
Since we have a cost estimate for Census 2000 (which is approximately 35 percent higher than that for 1990), we could calculate the IRR for Census 2000 as well, and expect to find it to be significantly lower. However, let us keep in mind the minimalist nature of these estimates for 1990. Legitimate comparison of the IRR estimates between 1990 and 2000, furthermore, would require some manipulation of nominal vs. real dollar terms, assumptions concerning expected inflation rates, etc., and the precision of the methods being used here simply do not warrant the effort involved. However, consideration of the clear implications for this analysis of the huge cost increase between 1990 and 2000 could certainly add some weight to the ongoing argument for the use of sampling to 1) reduce the cost, and 2) increase the accuracy of the census.

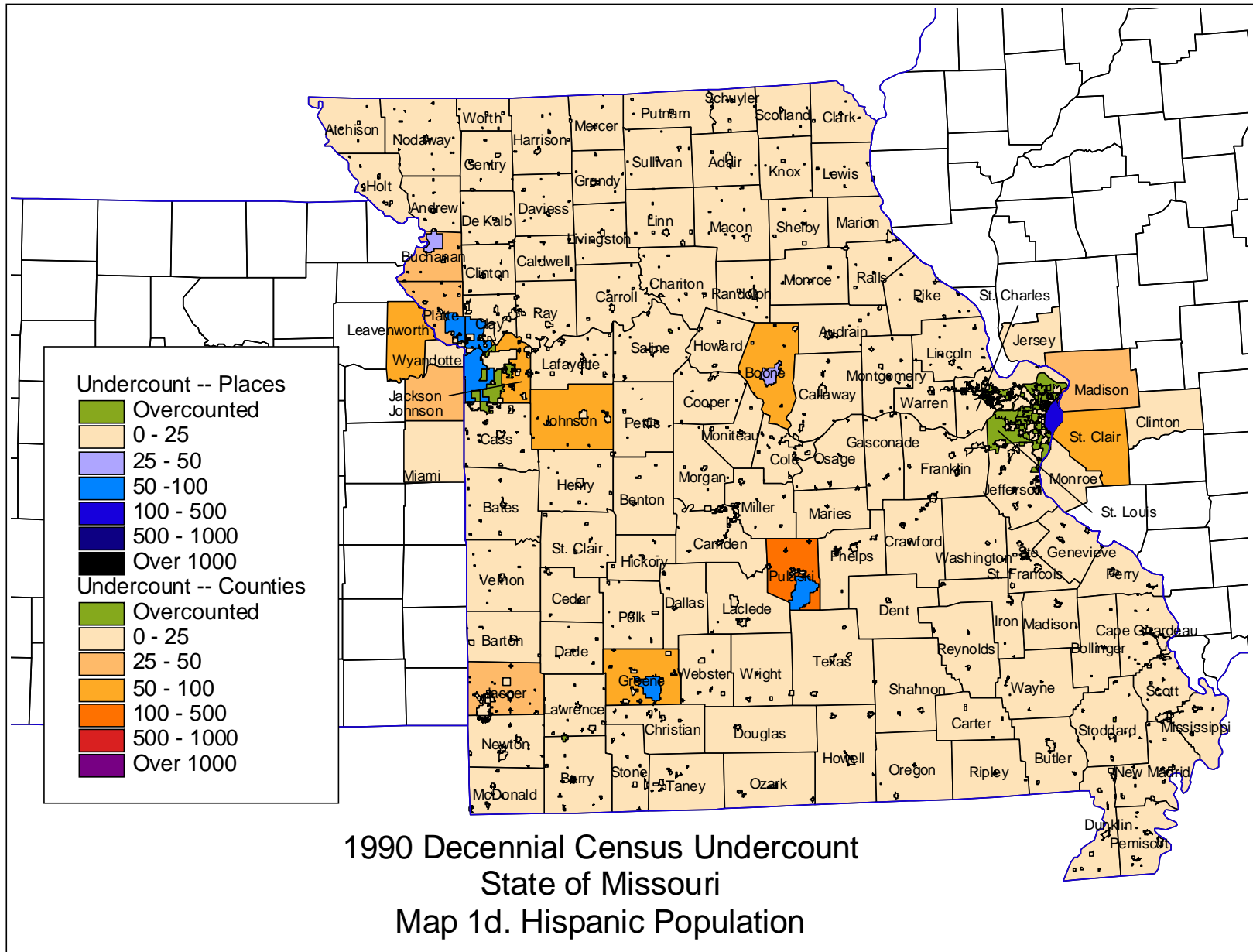


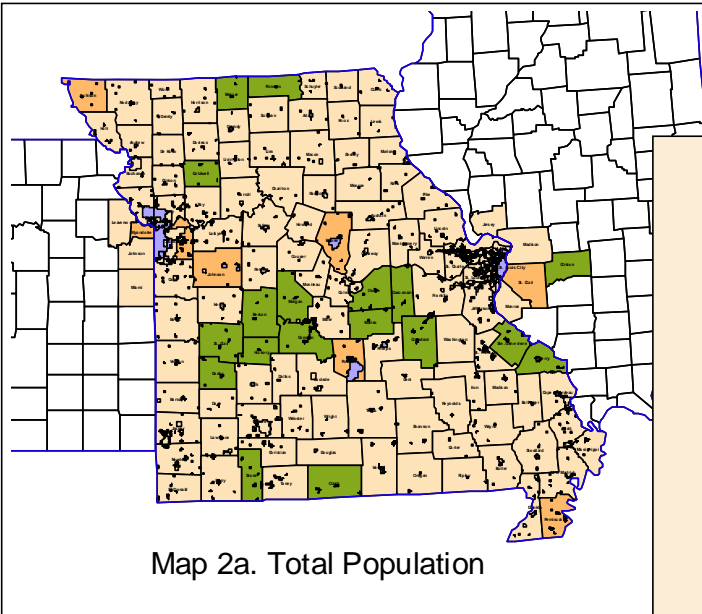


1990 Decennial Census Undercount  
 State of Missouri  
 Map 1a. Total Population







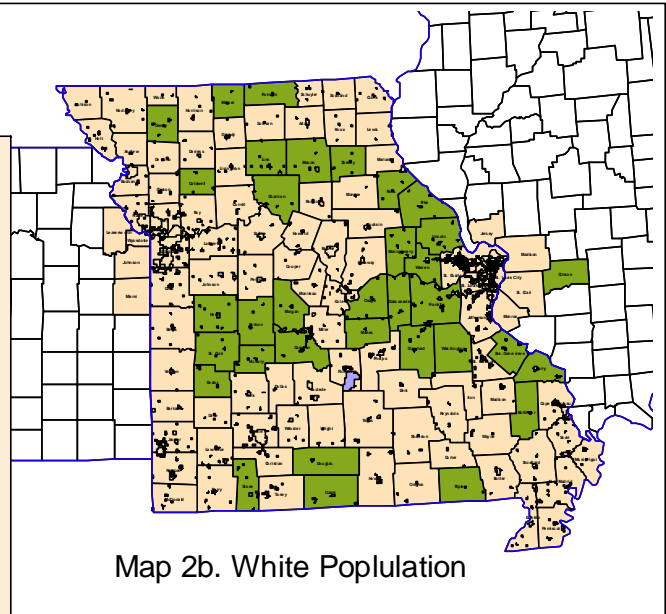


Map 2a. Total Population

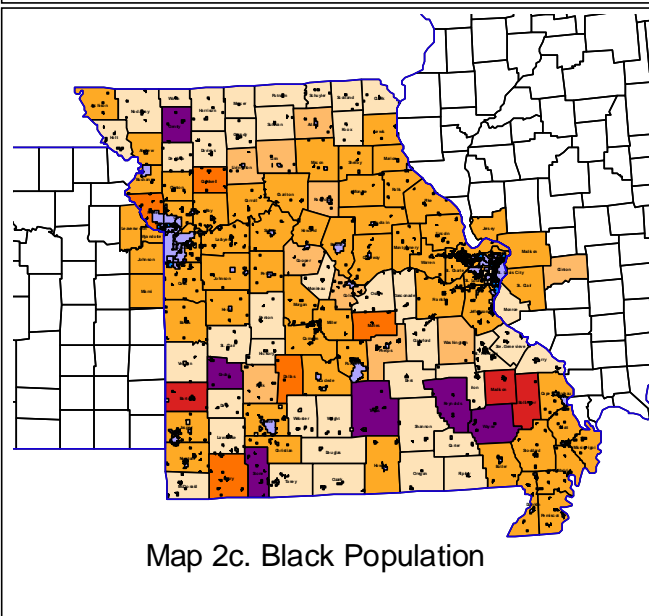
Map Panel 2

1990  
Decennial  
Census  
Differential  
Undercount  
Rates

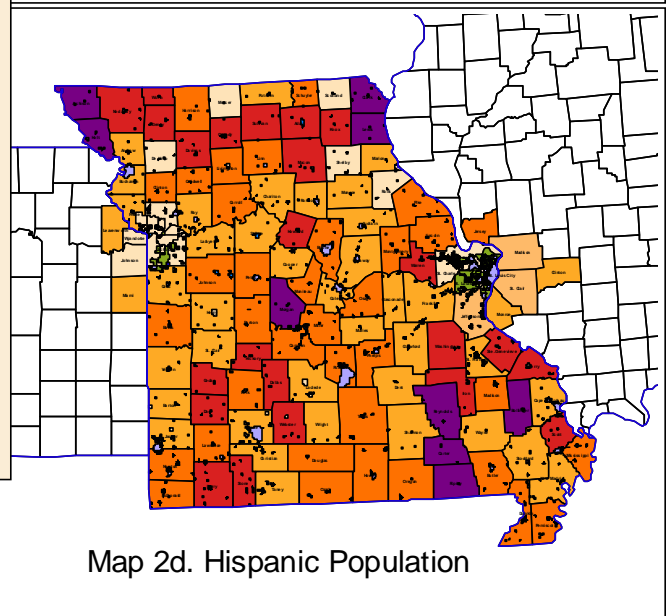
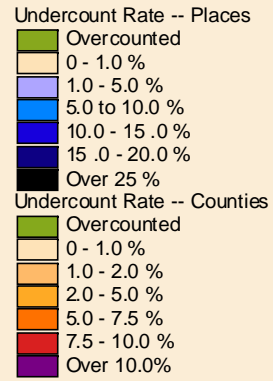
State of  
Missouri



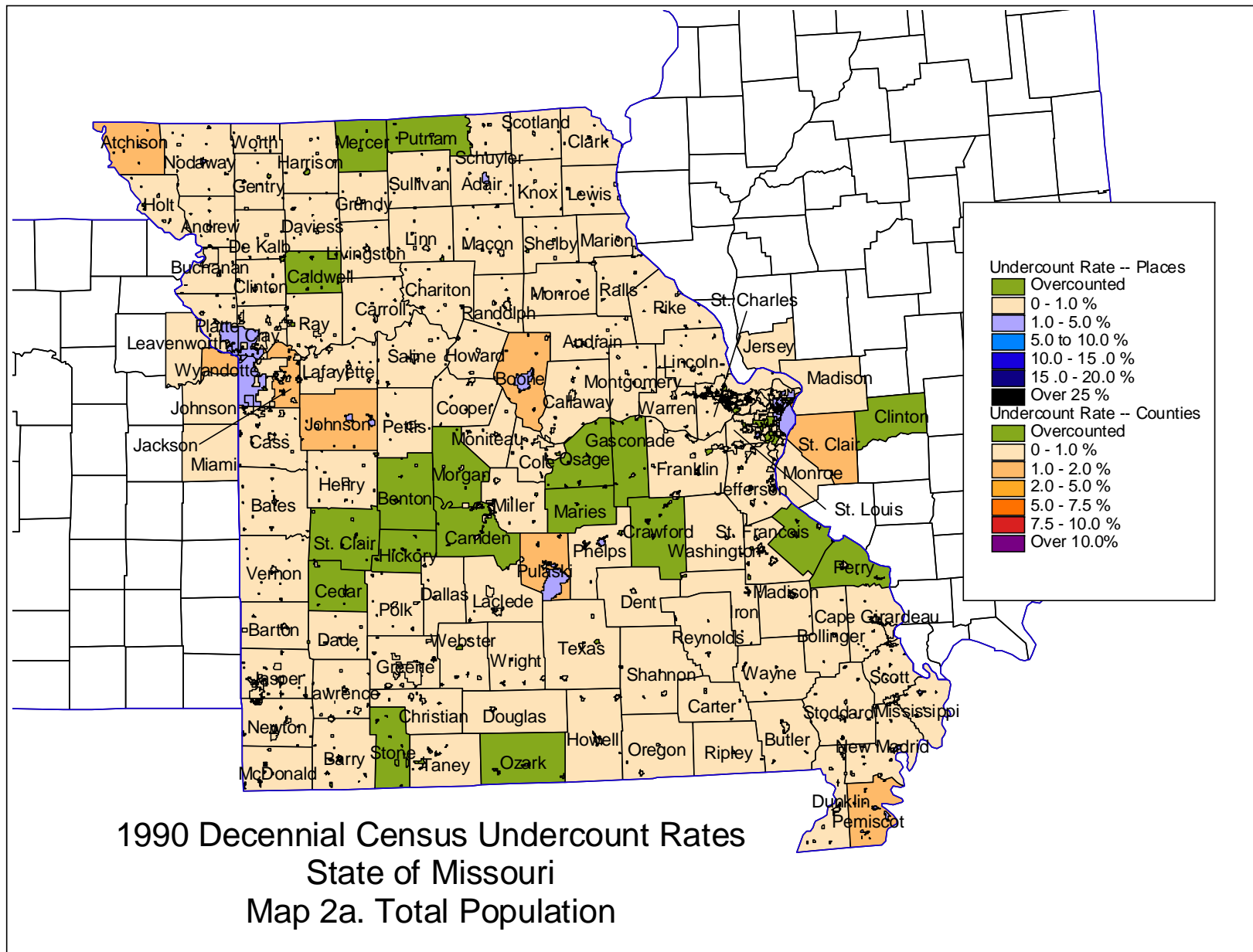
Map 2b. White Population

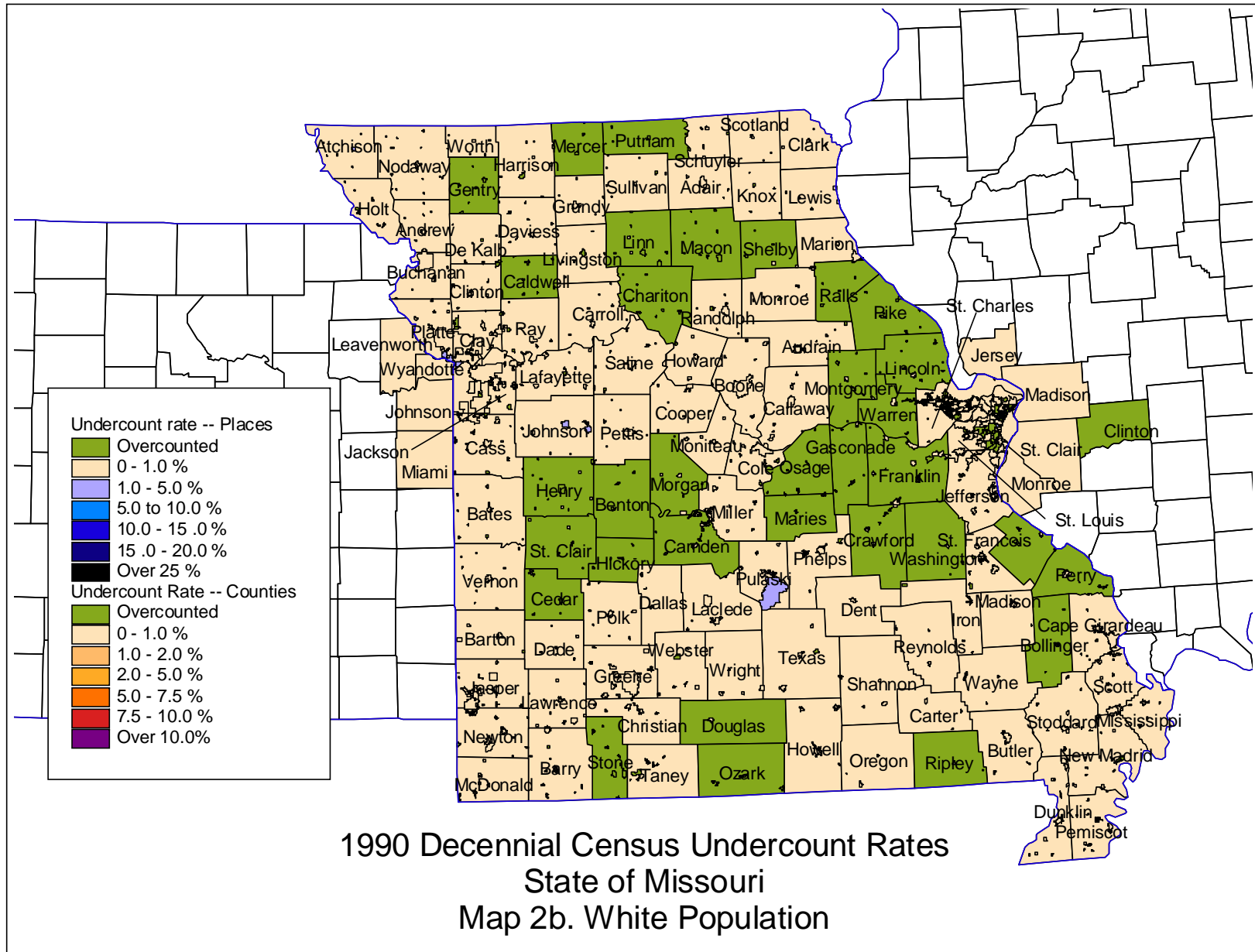


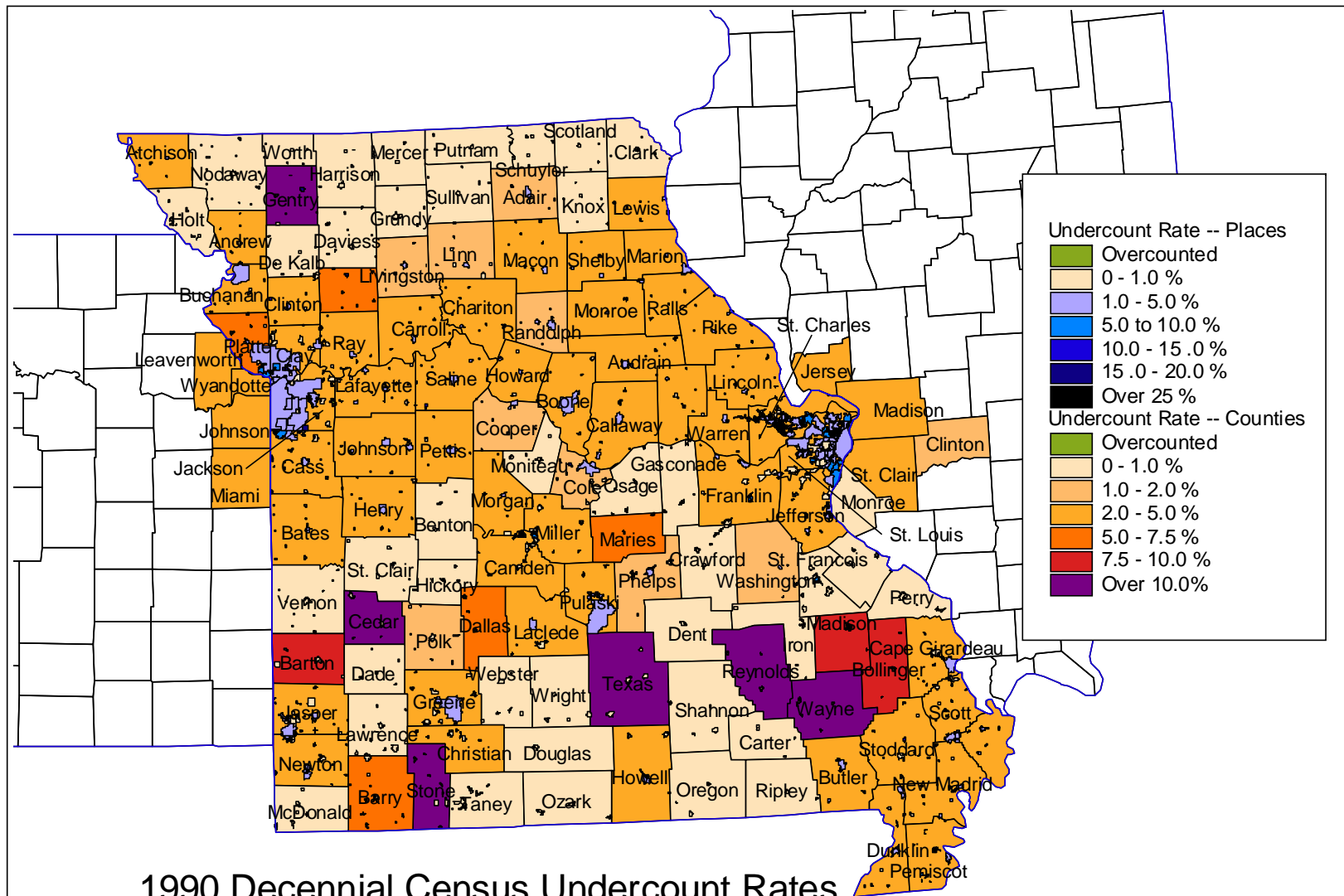
Map 2c. Black Population



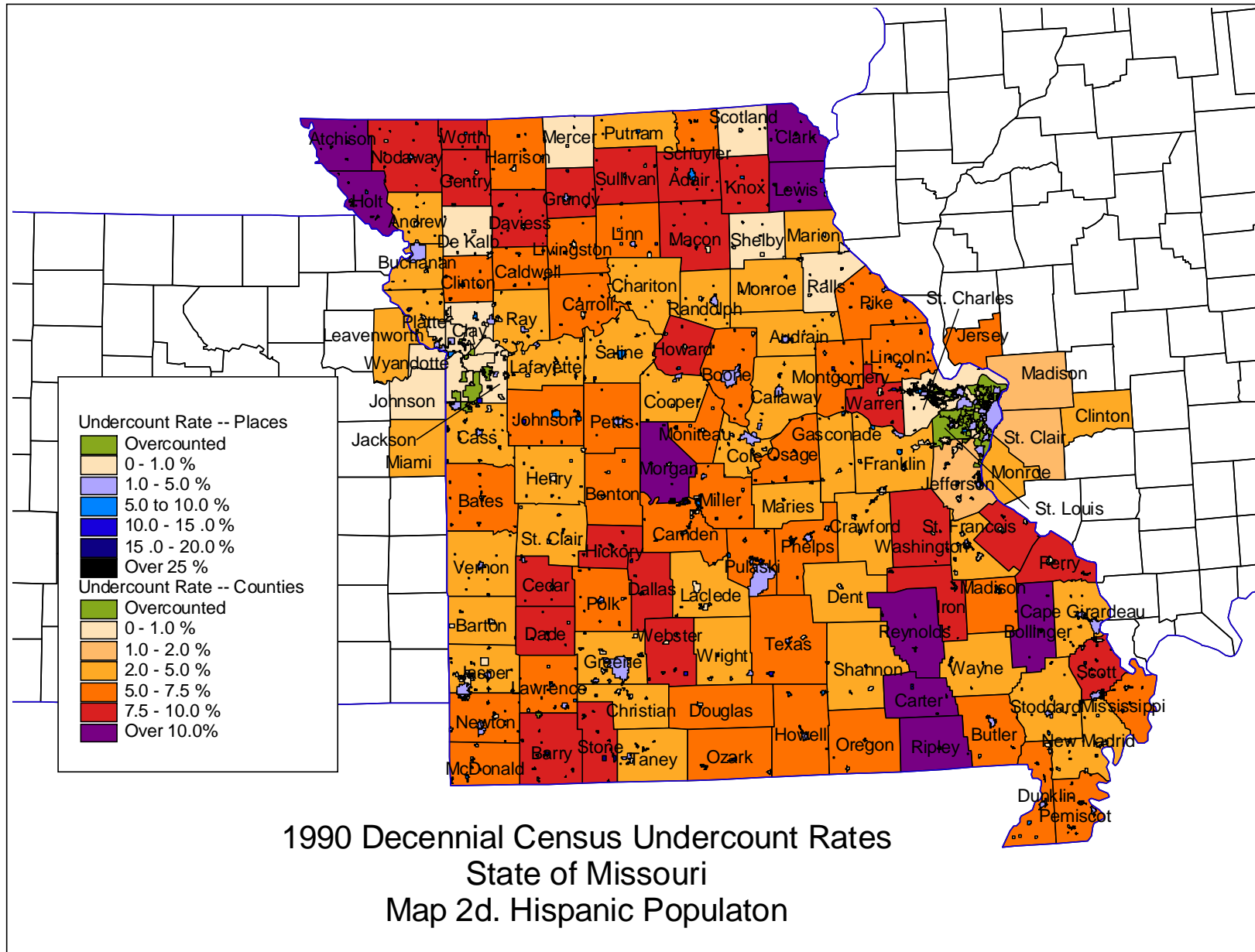
Map 2d. Hispanic Population

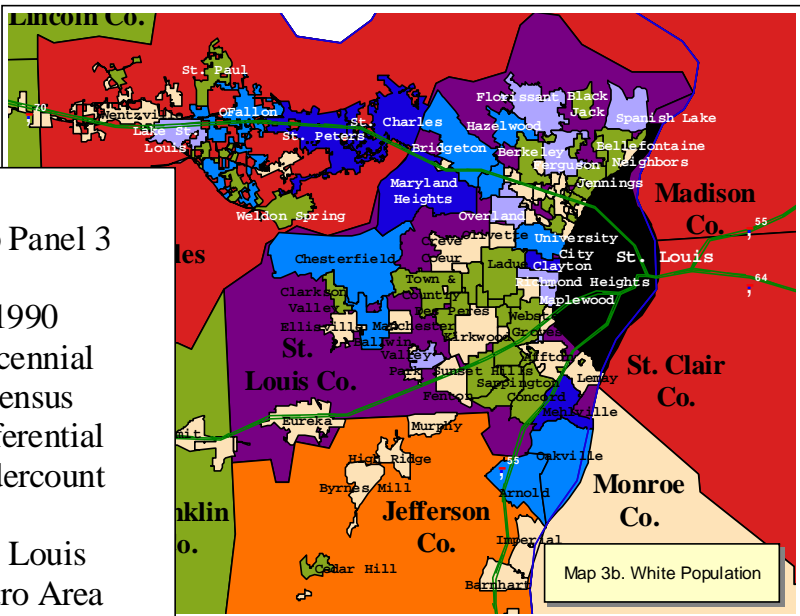
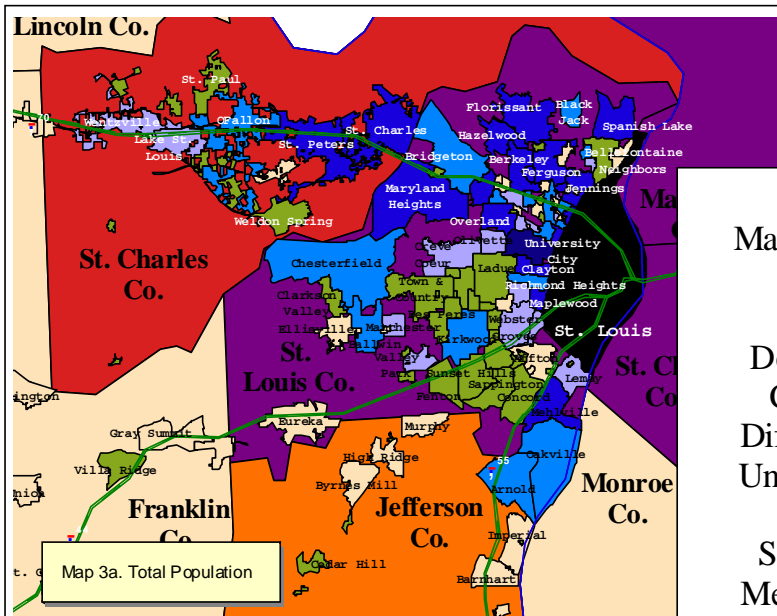






1990 Decennial Census Undercount Rates  
 State of Missouri  
 Map 2c. Black Population

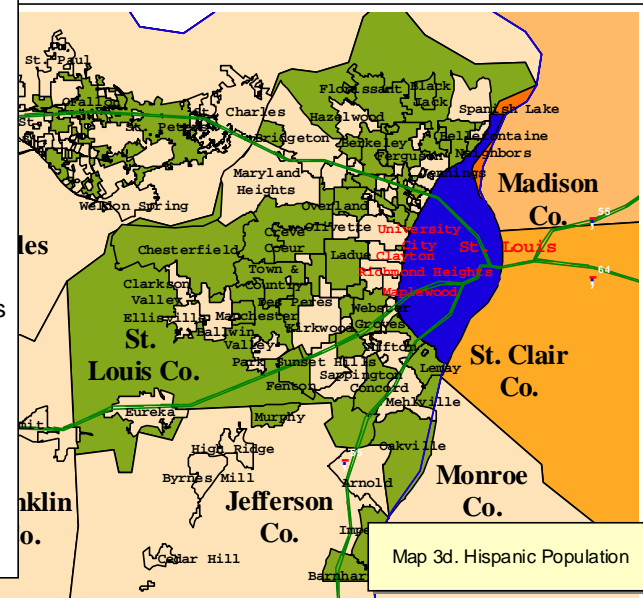
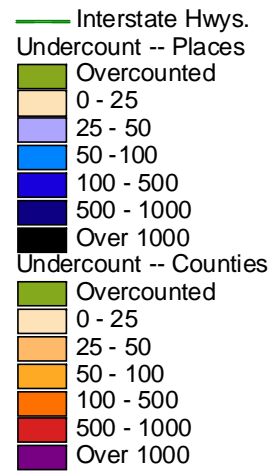
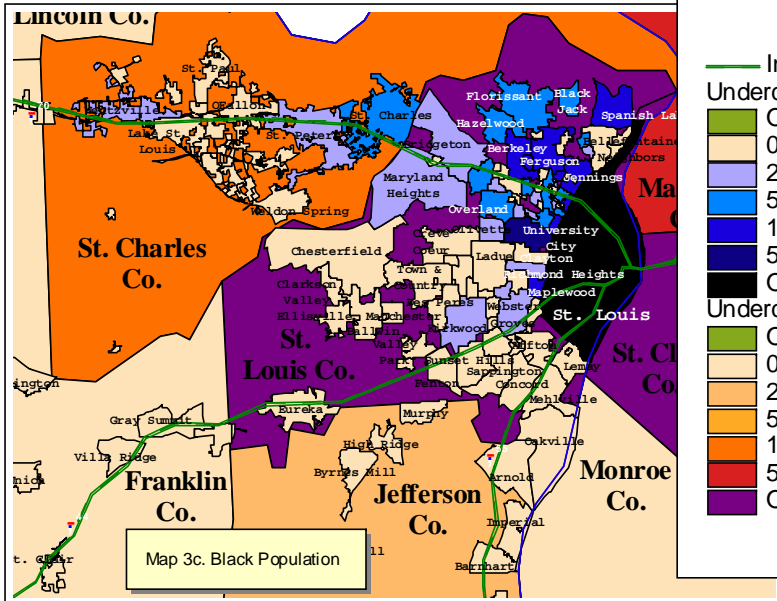




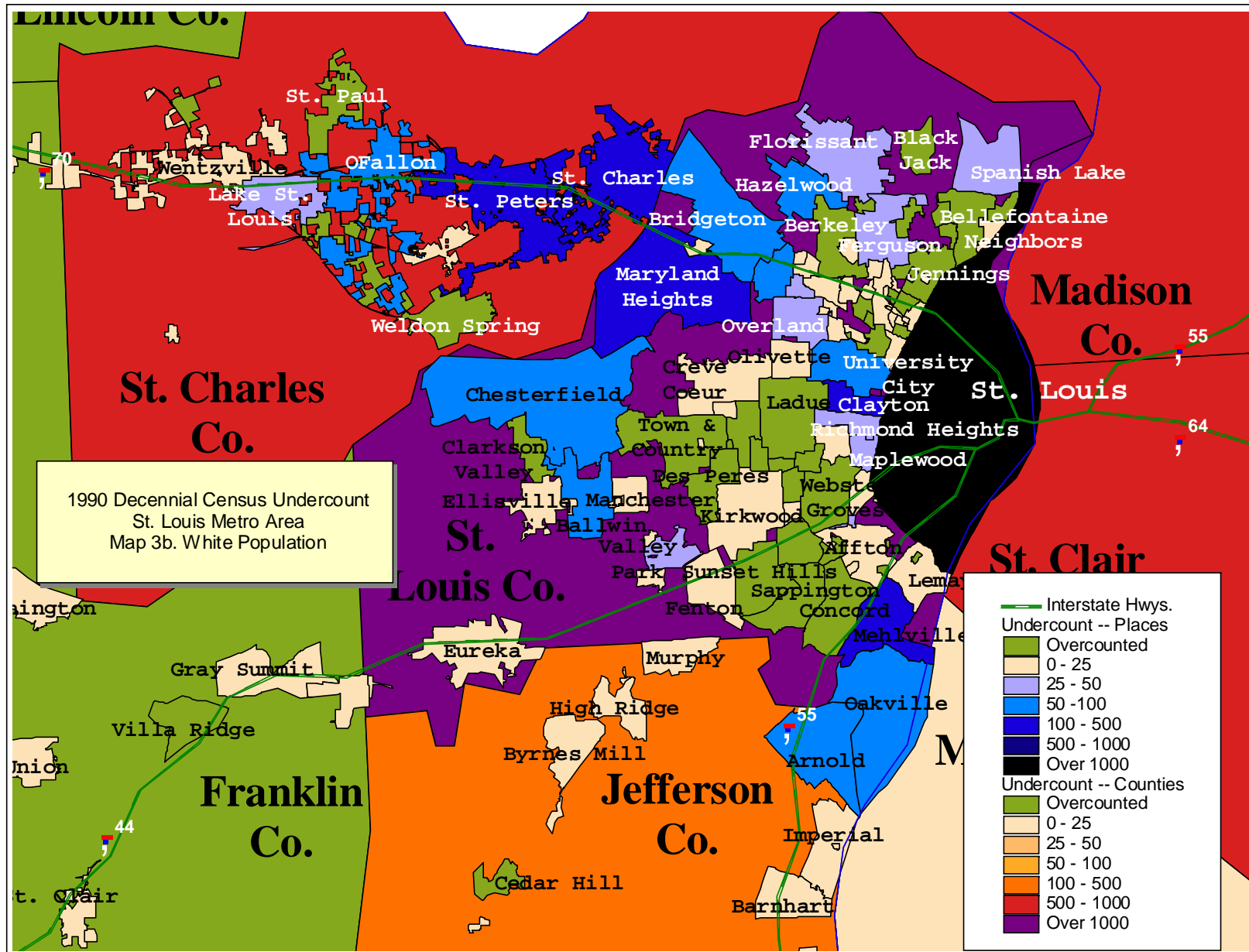
Map Panel 3

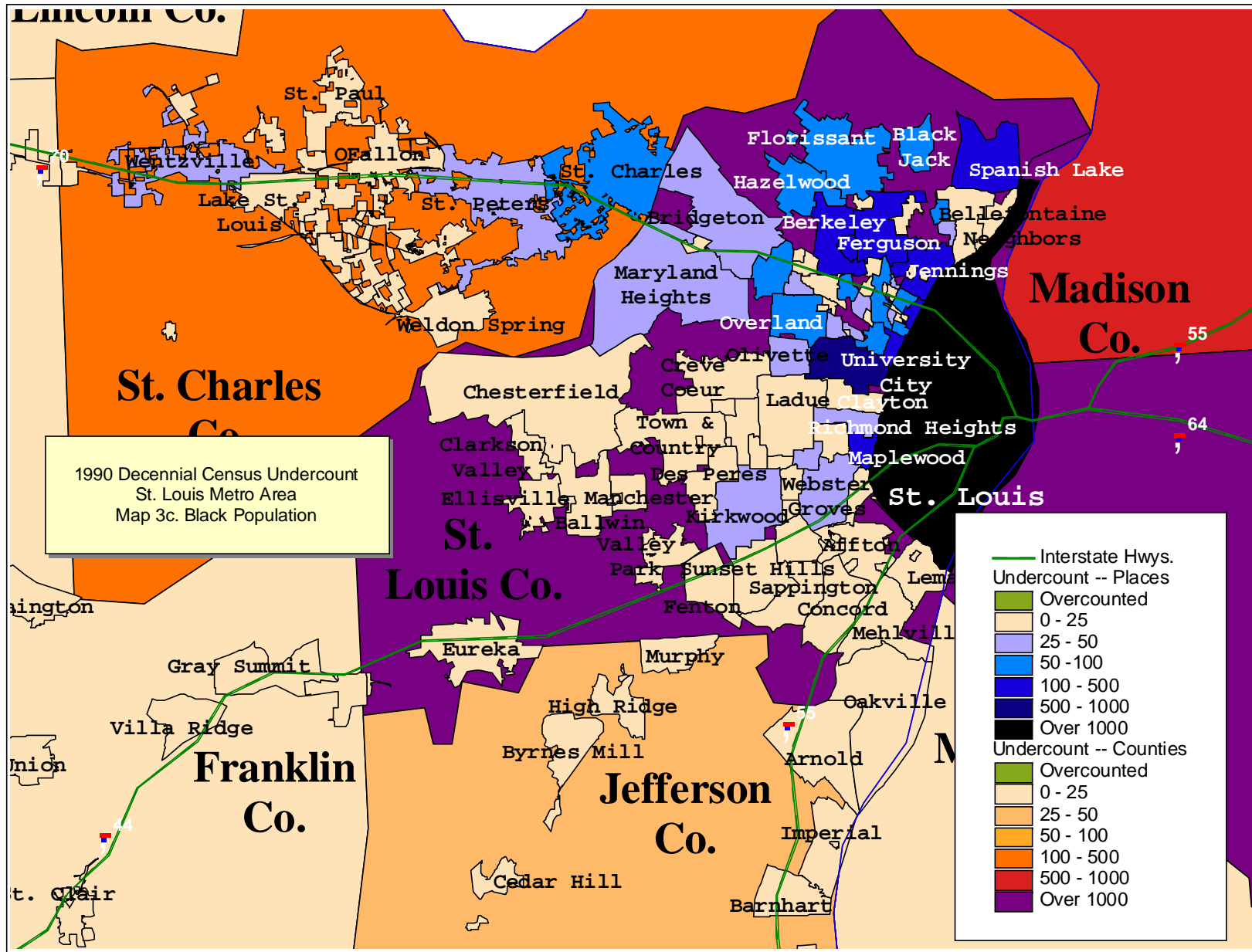
1990  
Decennial  
Census  
Differential  
Undercount

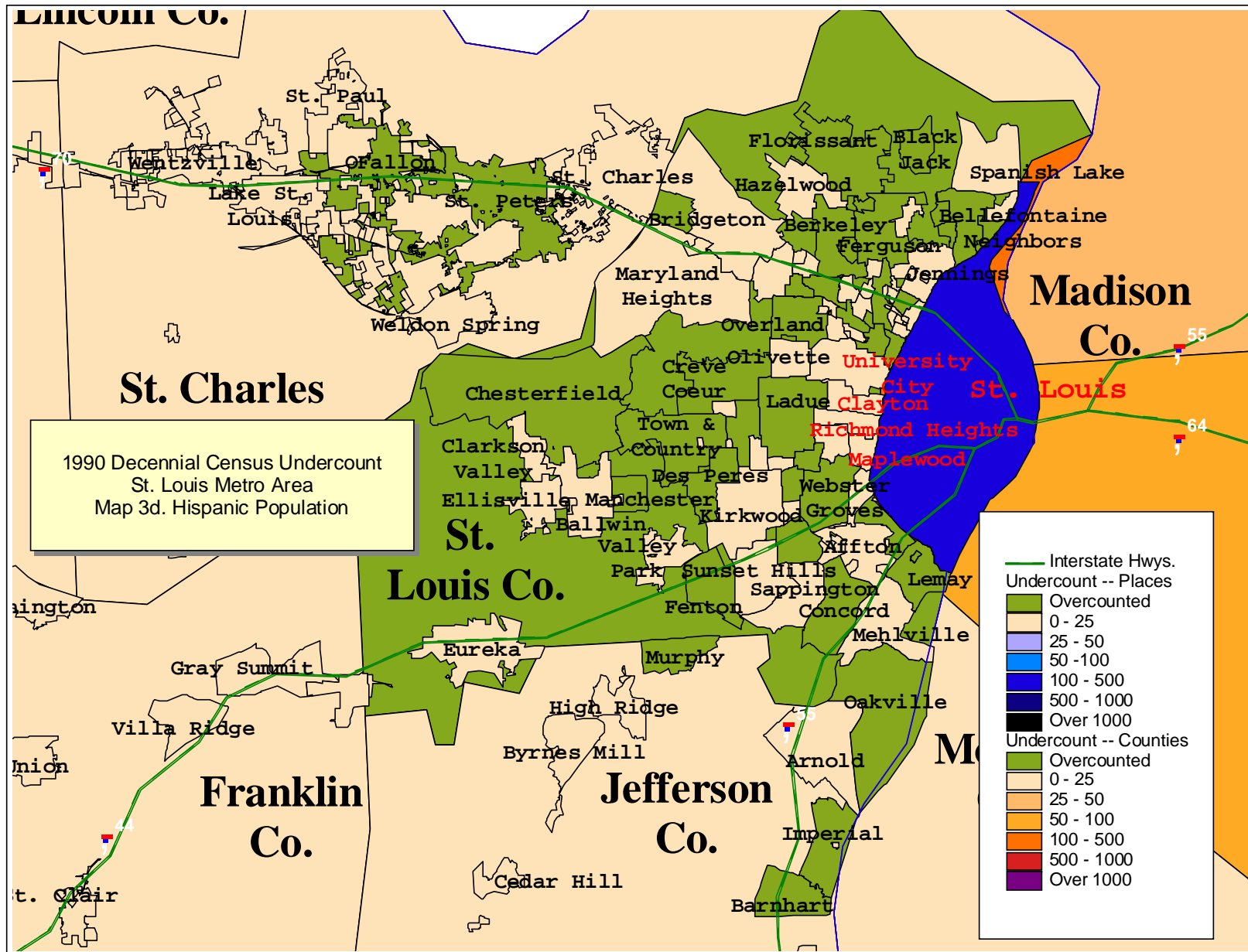
St. Louis  
Metro Area

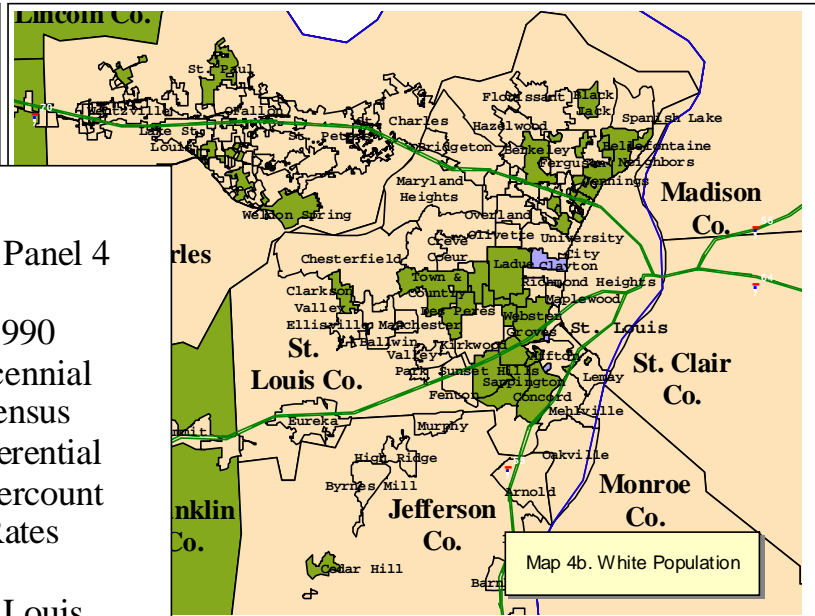
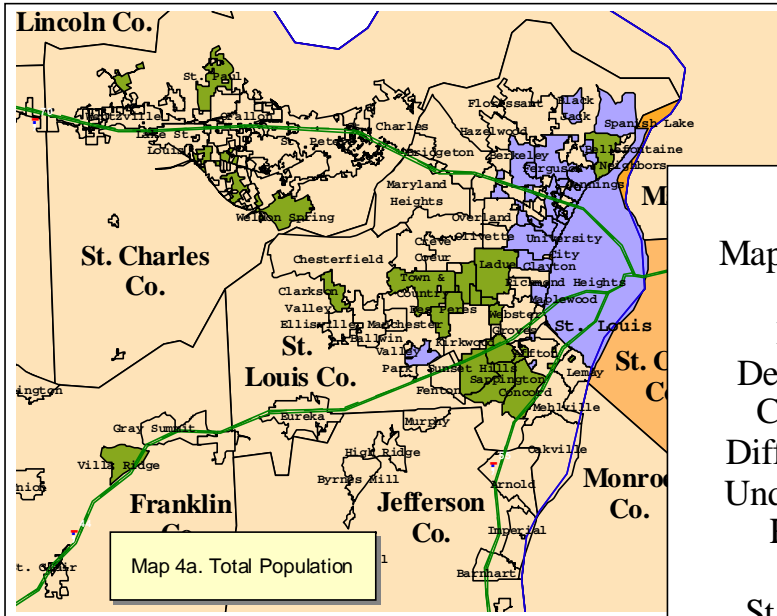




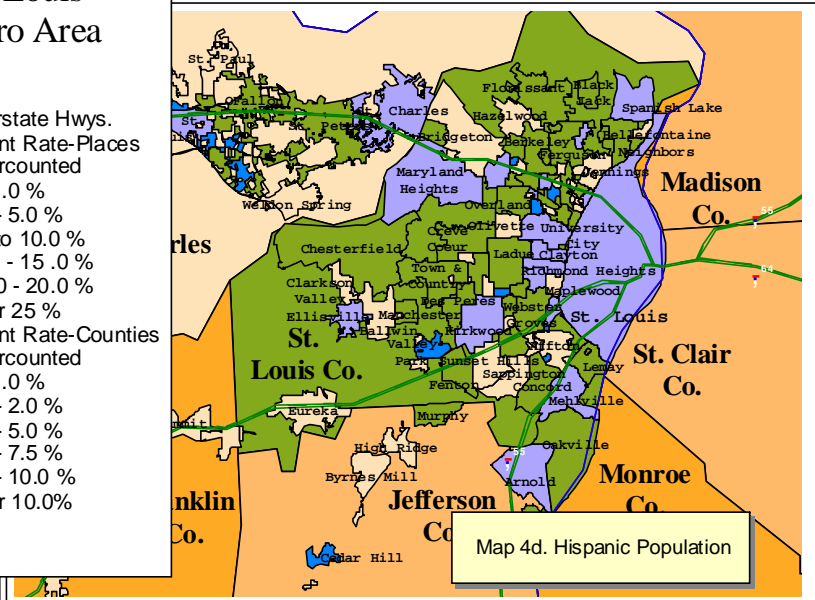
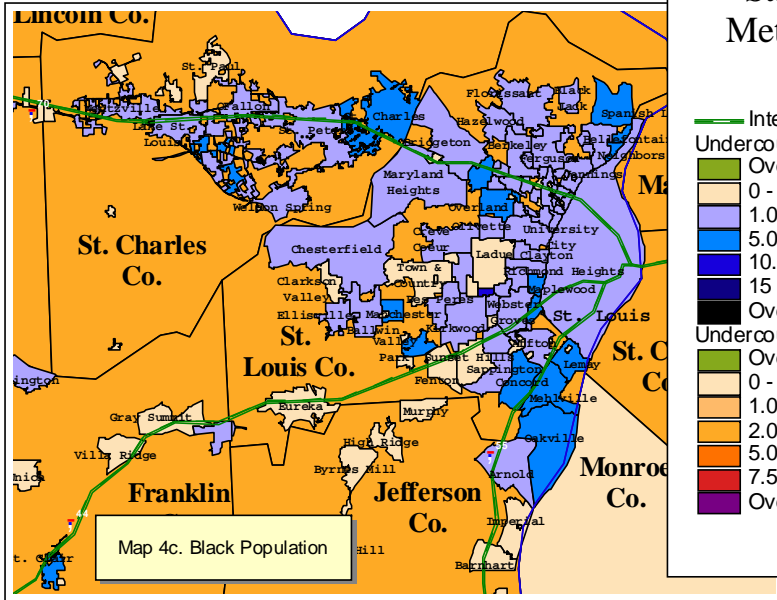








Map Panel 4  
1990  
Decennial  
Census  
Differential  
Undercount  
Rates  
St. Louis  
Metro Area



- Interstate Hwys.
- Undercount Rate-Places
  - Overcounted
  - 0 - 1.0 %
  - 1.0 - 5.0 %
  - 5.0 to 10.0 %
  - 10.0 - 15.0 %
  - 15.0 - 20.0 %
  - Over 25 %
- Undercount Rate-Counties
  - Overcounted
  - 0 - 1.0 %
  - 1.0 - 2.0 %
  - 2.0 - 5.0 %
  - 5.0 - 7.5 %
  - 7.5 - 10.0 %
  - Over 10.0 %

