Identifying Employment Subcenters in the LA Metro Area
A Comparison of Three Procedures

Richard Arnott and Daniel McMillen

July 29, 2009

Table of Contents

1. Introduction

The original interest in the identification of subcenters came about in the context of describing the polycentric nature of metropolitan areas in the United States. The seminal paper is Giuliano and Small (1991 – GS hereafter), which defined a subcenter on the basis of an absolute employment density and an absolute employment size, and applied the definition to identify subcenters for the Los Angeles Metropolitan Area (defined as the five counties) for 1981.

Since then the literature has proceeded in three directions. The first has been the identification of subcenters in other metropolitan areas using the GS procedure. The second has been relating various aspects of metropolitan spatial structure (housing rent, land rent, employment density, residential density, etc.) to the location of its various subcenters. The third has been investigating alternative definitions of subcenters that consider relative as well as absolute employment density.

SCAG's interest in identifying subcenters is different yet related. As the designated Metropolitan Planning Organization for Los Angeles, Orange, San Bernardino, Riverside, Ventura and Imperial Counties, SCAG is mandated to draw up plans for transportation, growth management, hazardous waste management, and air quality. Such planning requires forecasting future land use, transportation, and employment. While some of area's increasing needs for floor space will be met through densification, much will also be met through growth at the periphery of the currently developed area. Some of this growth will be dispersed, but much will occur in emerging subcenters. Good planning, especially good transportation and land use planning, requires forecasting these emerging subcenters. For example, for new fixed-route mass transit in peripheral areas, it makes good sense to locate the lines and the stations/stops so as to link up emerging subcenters. The effective development of these corridors will in turn require targeted land use planning.

GS defines an employment subcenter to be a set of contiguous "zones" (perhaps census tracts, zip code areas, or traffic analysis zones), each of which has an employment density exceeding a threshold density d, and which together have a combined employment exceeding a threshold D. The GS definition has been widely used because it is easy to apply and understand. Since it defines an employment center in terms of *absolute* employment density and *absolute* total employment, it is not however well suited to identifying subcenters at the periphery of a metropolitan area, which are characterized by high *relative* rather than absolute employment densities. Since most growth occurs at the periphery of a metropolitan area, the GS definition is also poorly suited to identifying emerging subcenters. For identifying peripheral and emerging subcenters, a subcenter definition that considers relative as well as absolute employment densities is needed.

This final report presents two alternative subcenter definitions that focus on relative rather than absolute employment densities, and compares the subcenters identified in the LA Metro Area using these definitions to the subcenters identified using the GS definition.

2. The Three Procedures

The three procedures are the Giuliano and Small procedure (GS), a non-parametric procedure (NP) developed by Daniel McMillen, and a spatial autocorrelation procedure developed in Beaumont, Ertur, and LeGallo (LISA).

- The Giuliano and Small (GS) procedure

A GS-subcenter is defined to be a set of contiguous zones having the properties that each zone has an employment density exceeding d, and the zones together have a total employment exceeding D. In this report, we investigate two pairs of cutoff levels that are commonly used. In the first, GS20, d = 20 and D = 20,000; the employment density cutoff is 20 employees per acre, and the overall employment cutoff is 20,000. In the second, GS10, d = 10 and D = 10,000; the employment density cutoff is 10 employees per acre, and the overall employment cutoff is 10,000.

- A non-parametric procedure (NP)

The term "non-parametric" derives from statistics and econometrics, and refers to the technique employed in constructing the index. In the current context, the term "spatial smoothing" would perhaps be more appropriate. The employment subcenter identification procedure involves three steps. In the first step, a smoothed employment density surface is derived from the actual employment densities of the TAZ's. In the second step, a TAZ is defined as a *candidate* TAZ, i.e. as a candidate for inclusion in a subcenter, if the ratio of its actual employment density to its smoothed employment density exceeds some cutoff. In the third step, a subcenter is defined as a set of contiguous candidate TAZ's with overall employment exceeding some cutoff. The maps to be presented using the NP procedure identify the candidate TAZ's, and do not go the last step and identify subcenters. We shall however informally talk of NP subcenters when there is a contiguous group of candidate TAZ's. The details of the procedure are given in the Technical Appendix. In this report, we identify a TAZ to be a candidate TAZ if its actual employment density is significantly different from the corresponding smoothed employment density at a 10% level of significance.

According to this procedure, a candidate TAZ is one with a high *relative* employment density, relative to the smoothed employment density surface. Thus, the NP procedure identifies candidate TAZ's based on relative rather than absolute densities.

- A spatial autocorrelation procedure (LISA)

LISA – "Local Indicators of Spatial Association" – is a set of statistical procedures developed by geographers to measure spatial autocorrelation. Following Beaumont, Ertur, and LeGallo (2004), we use the local Moran's I statistic to measure the degree of local spatial autocorrelation for employment density in sets of contiguous tracts. Significantly positive spatial autocorrelation in a TAZ indicates that the tract has higher than average employment density that is positively correlated with density in contiguous tracts. As is the case with the non-parametric approach, LISA approach identifies clusters of tracts with high employment density, but there is no requirement that the total level of employment in the tracts exceeds a critical value. Thus, the both the NP and LISA approaches focus on relative rather than absolute densities.

3. Maps Using 2003 SCAG Employment Data

The set of maps in the following pages are organized according alphabetically by county name: Imperial County, Los Angeles County, Orange County, Riverside County, San Bernardino County, and Ventura County. For each county, as many as four maps are presented, one for each of the three procedures employed.[1] When a particular procedure identifies no subcenters for a county (or for the NP procedure no candidate TAZ's), no map is presented. Thus, for example, only one map is presented for Imperial County since both the GS and LISA procedures identify no subcenters for Imperial County.

We comment briefly on each map in turn, and then provide some summary comments.

- Imperial County

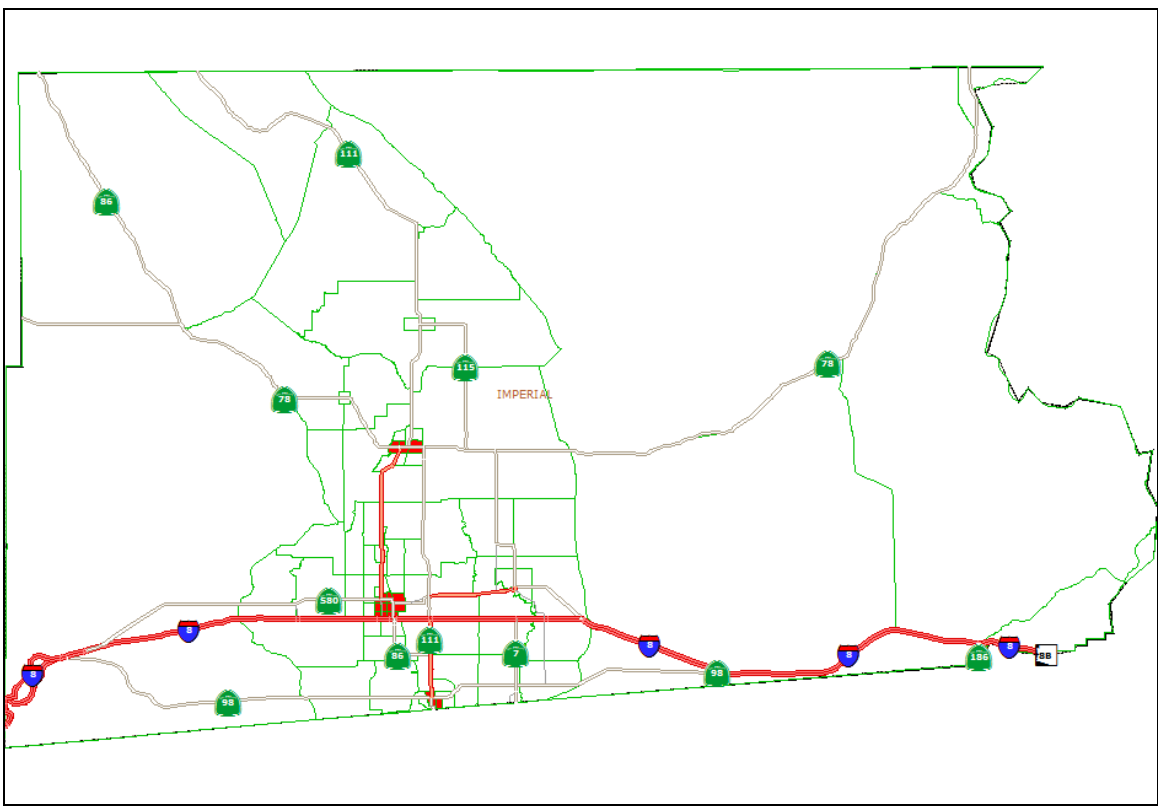


Figure 1: Imperial County, Non-parametric

The GS and LISA procedures identify no subcenters. The NP procedure identifies candidate TAZ's in Brawley, El Centro, and Calexico.

[1] Furthermore, to facilitate visual presentation, more than one map may be presented for a particular county. Los Angeles County may be divided up into a southern and a northern portion, and Riverside and San Bernardino into a western and an eastern portion.

- Los Angeles County

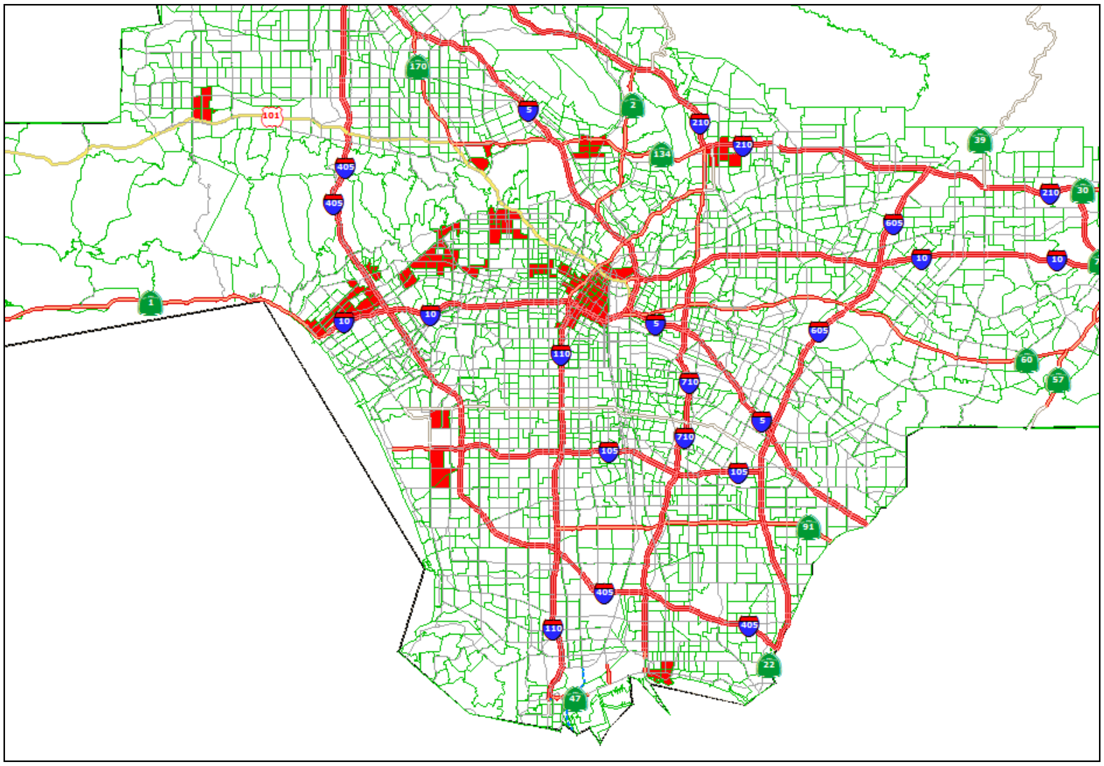


Figure 2: Los Angeles County, GS, d=20, D=20,000

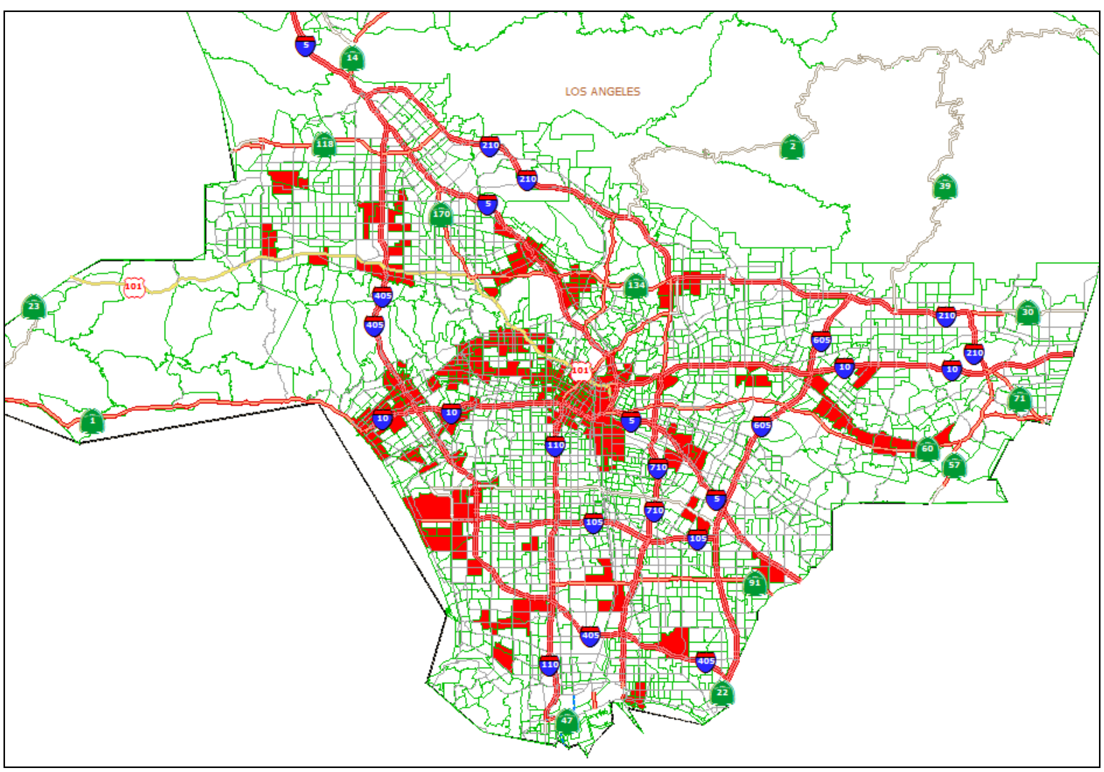


Figure 3: Los Angeles County, GS, d=10, D=10,000

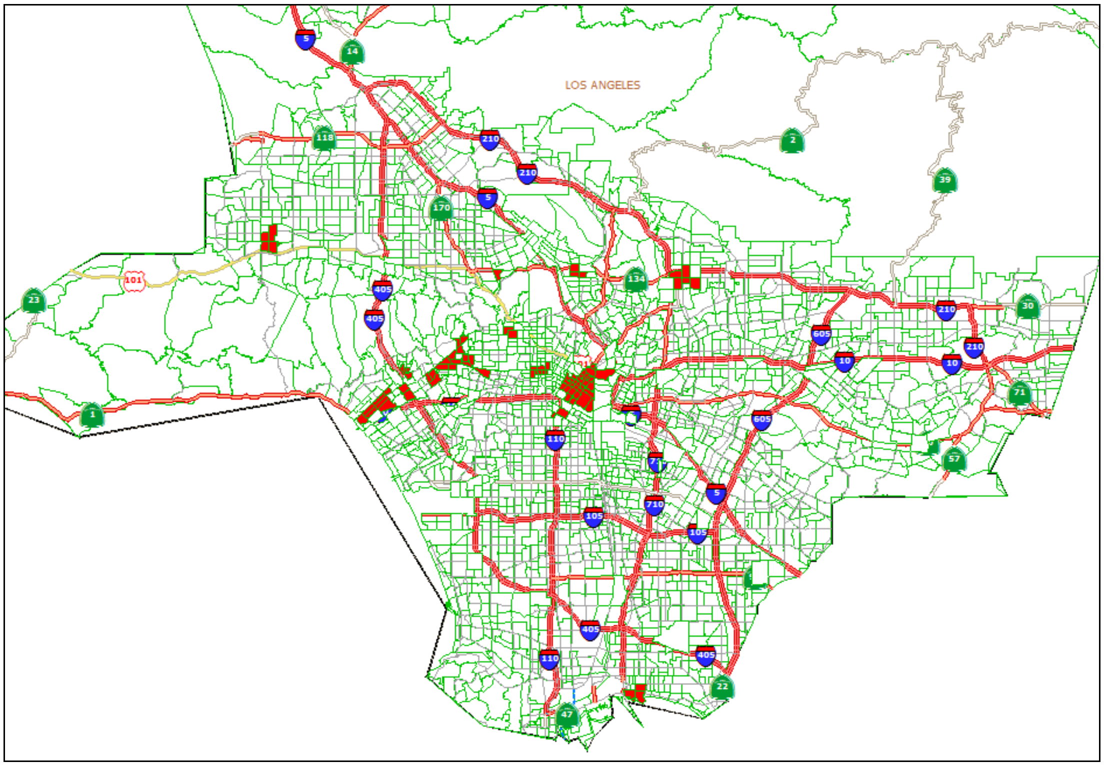


Figure 4: Los Angeles, LISA

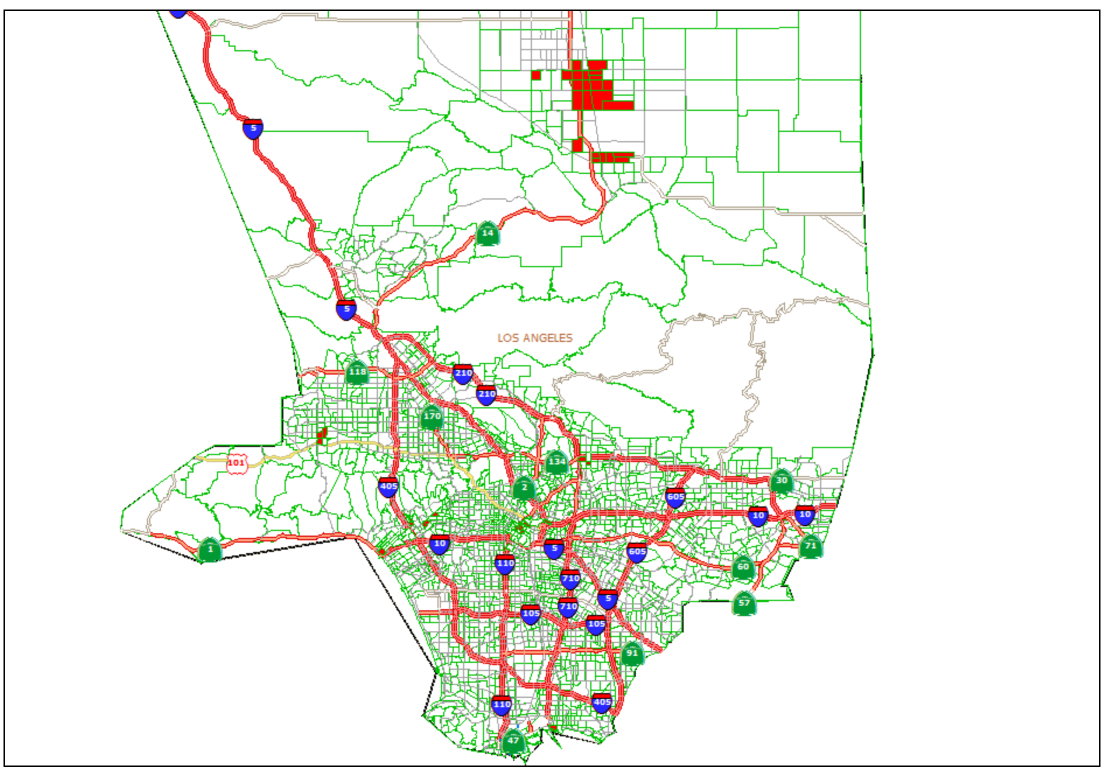


Figure 5: Los Angeles County, Non-parametric

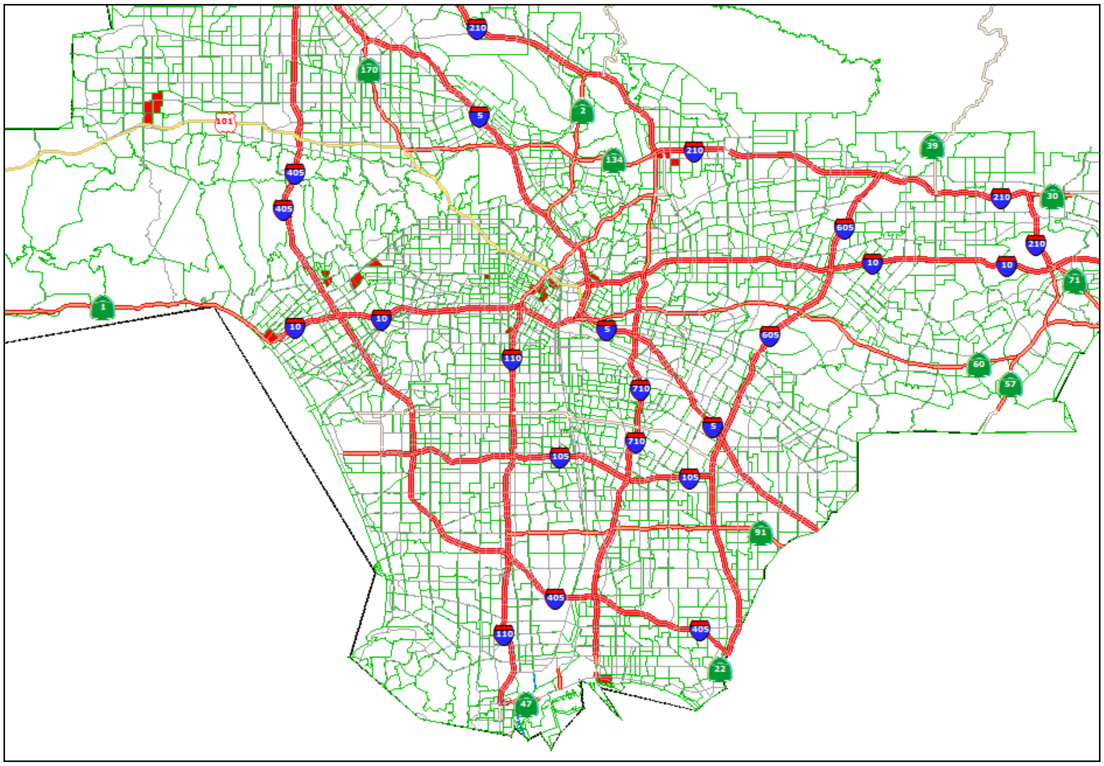


Figure 6: Los Angeles County, Southern Portion, Non-parametric

The GS procedure using the cutoffs of 10 employees/acre and total employment of 10,000 (GS10) identifies some sixteen employment subcenters: Northridge, Calabasas, Sepulveda, Burbank, Glendale, Pasadena, Santa Monica, Los Angeles, Commerce, La Puente, LAX, Compton, Bellflower, Torrance, Long Beach Airport, and Long Beach. It identifies no employment centers in north LA County. Employing the higher cutoffs of 20 employees/acre and total employment of 20,000 (GS20) eliminates Northridge, Sepulveda, Commerce, La Puente, Compton, Bellflower, Torrance, and Long Beach Airport as subcenters. It also reduces the size of the remaining subcenters. And it breaks up the vast Los Angeles "subcenter" into downtown Los Angeles, West Hollywood, and Beverly Hills.

The LISA procedure identifies almost exactly the same set of subcenters as does the GS20; it does not identify LAX as a subcenter while GS20 does. Also, several of the LISA subcenters contain fewer TAZ's than do the corresponding GS20 subcenters.

The NP procedure identifies only a very small proportion of TAZ's as candidate TAZ's. They are in Calabasas, Santa Monica, Beverly Hills, north Hollywood, downtown Los Angeles, Pasadena, and Long Beach. If a total employment criterion were applied, few of these would be identified as subcenters. The NP procedure identifies two subcenters in north LA County, Lancaster and Palmdale. In short, the NP procedure identifies few subcenters in LA County since there are few areas where *relative* employment density (relative to the corresponding smoothed density) is sufficiently high.

- Orange County

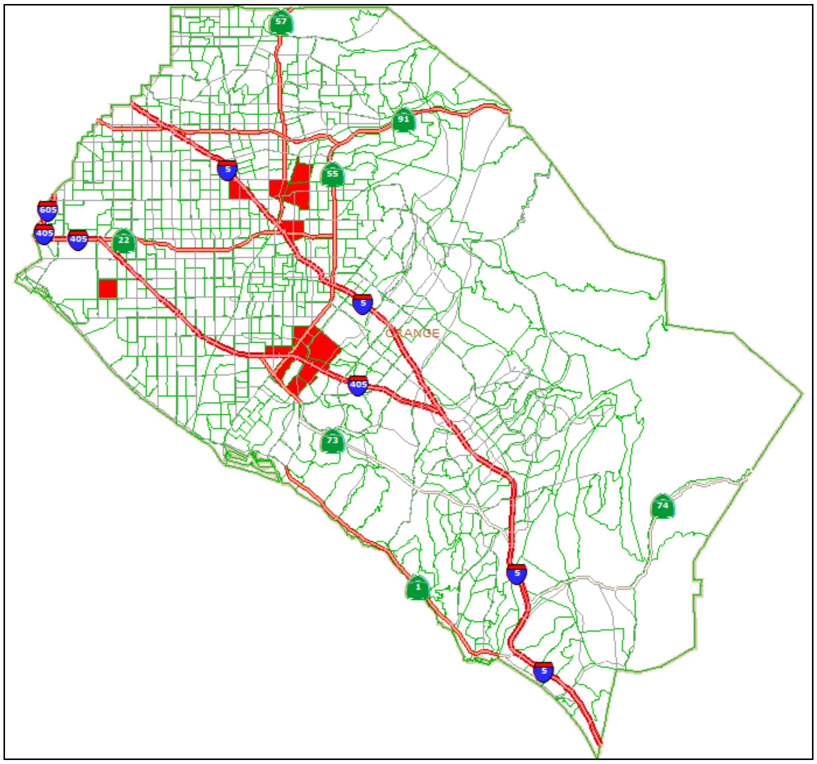


Figure 7: Orange County, GS, d=20, D=20,000

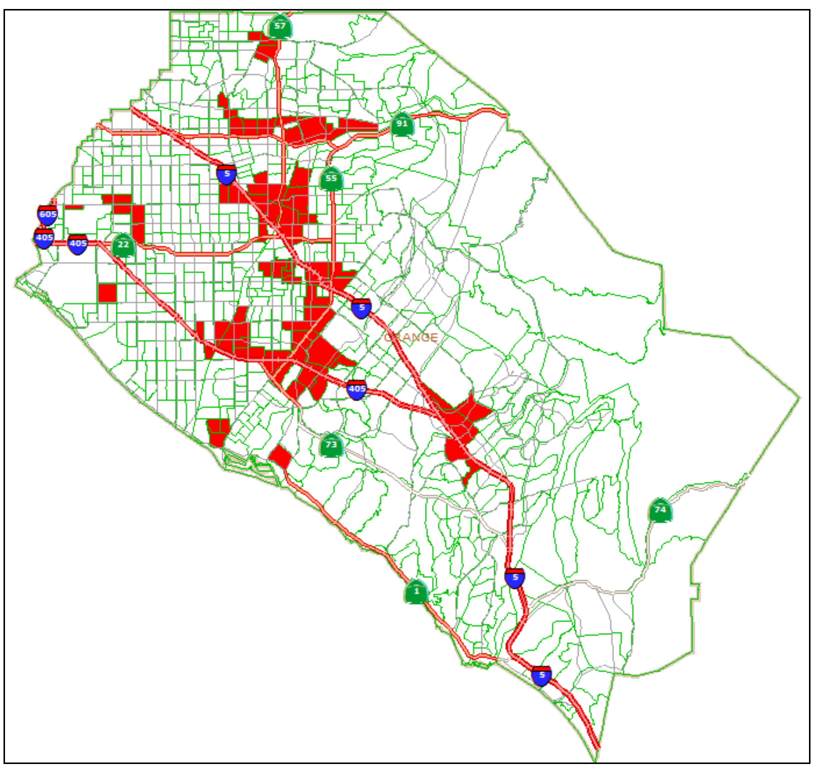


Figure 8: Orange County, GS, d=10, D=10,000

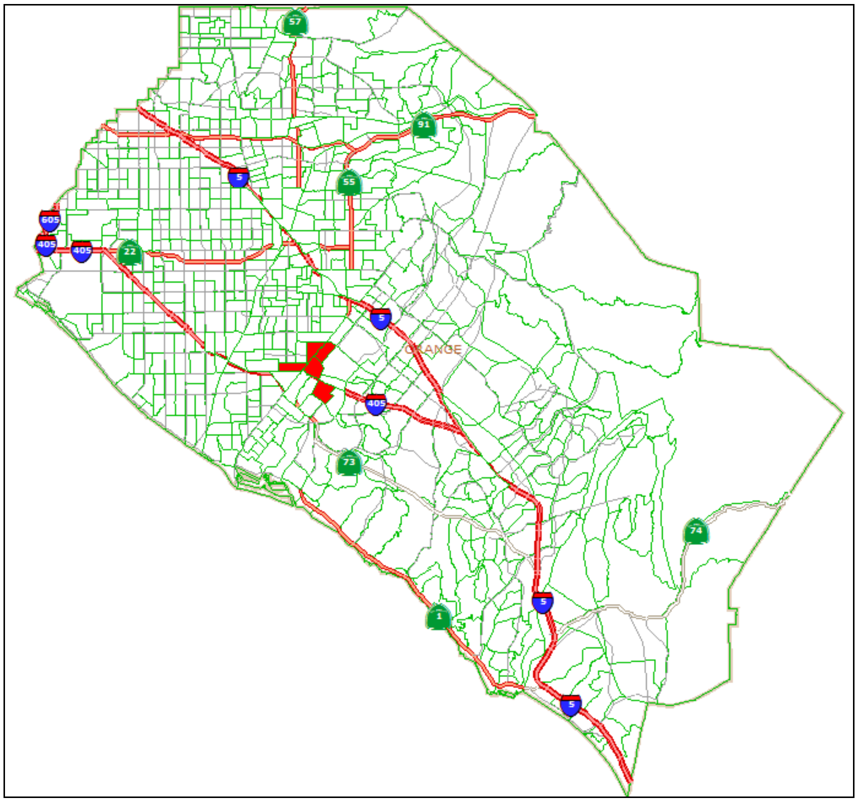


Figure 9: Orange County, LISA

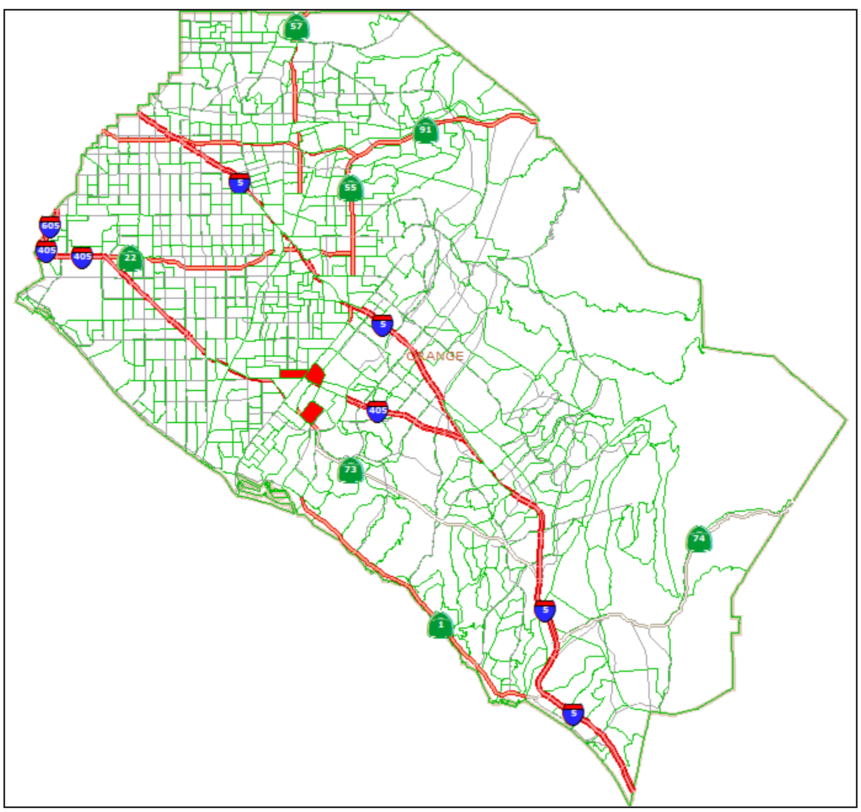


Figure 10: Orange County, Non-parametric

GS10 identifies some half dozen employment subcenters in Orange County: Brea Mall, Placentia, Anaheim, Santa Ana/South Coast Plaza/ Irvine Business Complex/John Wayne, and Irvine Industrial Complex – East. With GS 20, the Anaheim subcenter splits into three smaller subcenters, the Santa Ana/South Coast Plaza/Irvine Business Complex/John Wayne subcenter shrinks to the Irvine Business Complex, and the other three subcenters disappear.

The LISA and NP procedures identify only the Irvine Business Complex as a subcenter.

- Riverside County

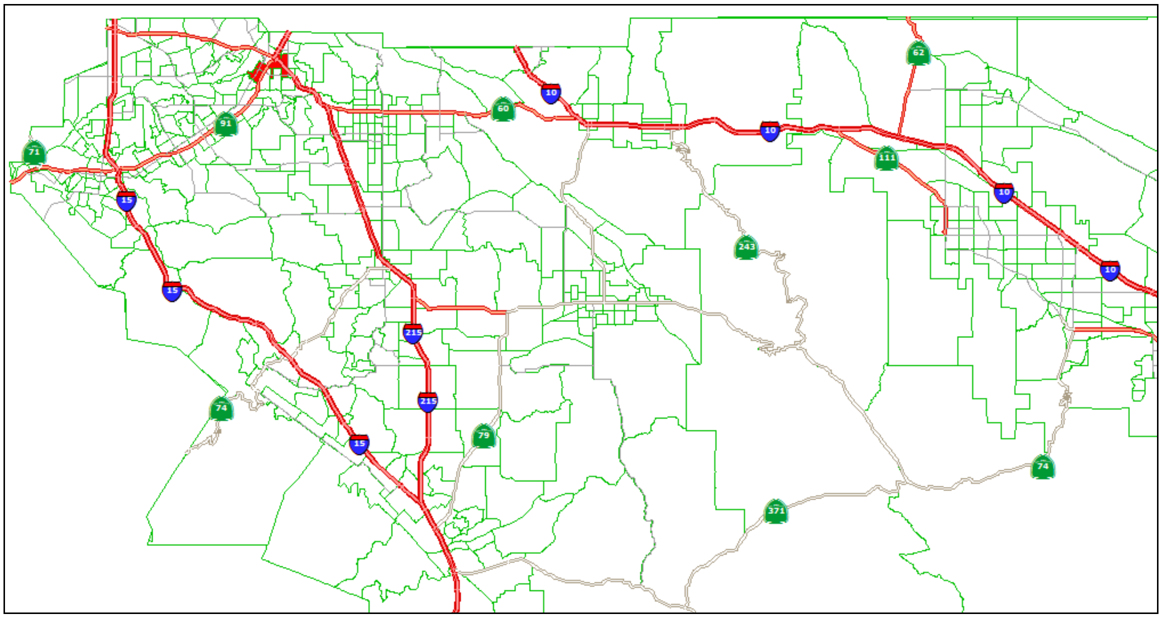


Figure 11: Riverside County, GS, d=10, D=10,000

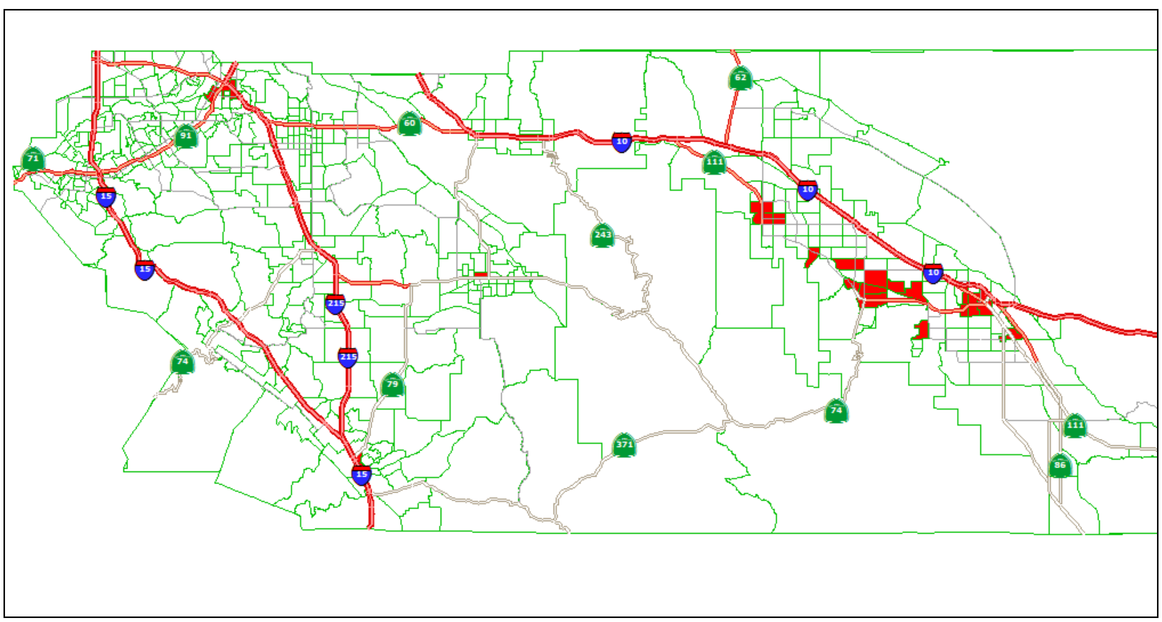


Figure 12: Riverside County, Non-parametric

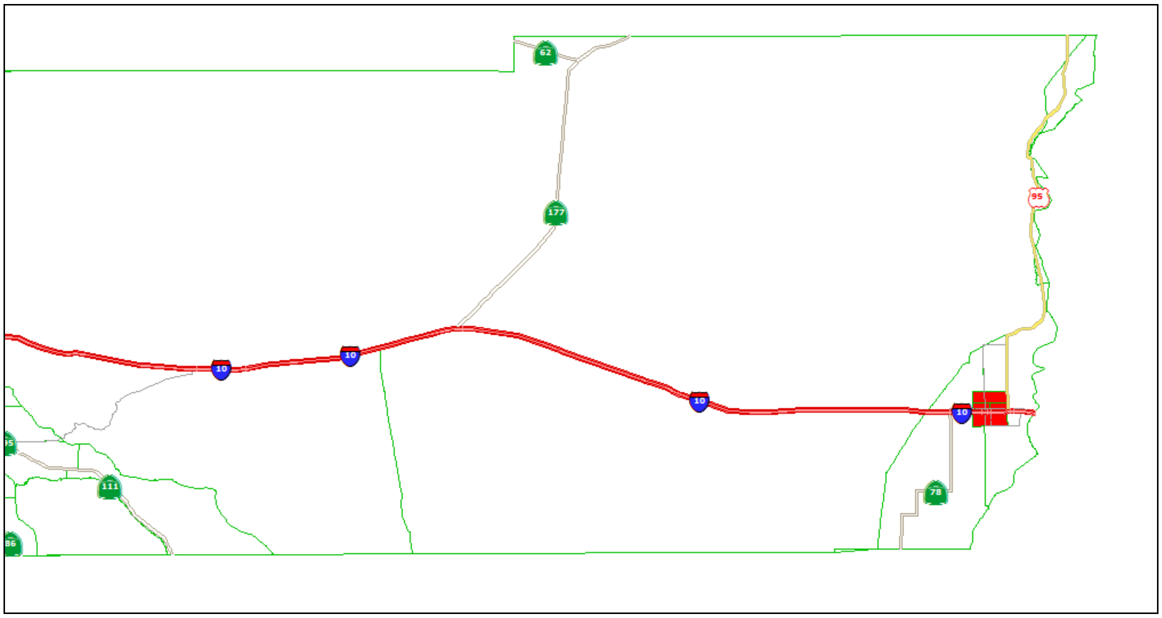


Figure 13: Riverside County, Eastern Portion, Non-parametric

The GS20 and LISA procedures identify no subcenters in Riverside County. GS10 identifies only one, Riverside. The NP procedure generates an embarrassment of riches. Candidate subcenters, in addition to Riverside, include Hemet, Temecula, Blythe, and several of the cities along the Coachella Valley.

These results suggest that none of the procedures alone is satisfactory in identifying subcenters at the metropolitan periphery. On one hand, the city of Riverside is clearly a regional subcenter, but it fails to be identified as such by the GS20 and LISA procedures because its absolute employment density is not sufficiently high. On the other hand, most experts would consider the group of cities along the Coachella Valley from Palm Springs to Coachella to constitute a single subcenter, but the NP procedure identifies six different candidate subcenters, most likely because the relative employment density of almost any desert city is likely to be high.

- San Bernardino County

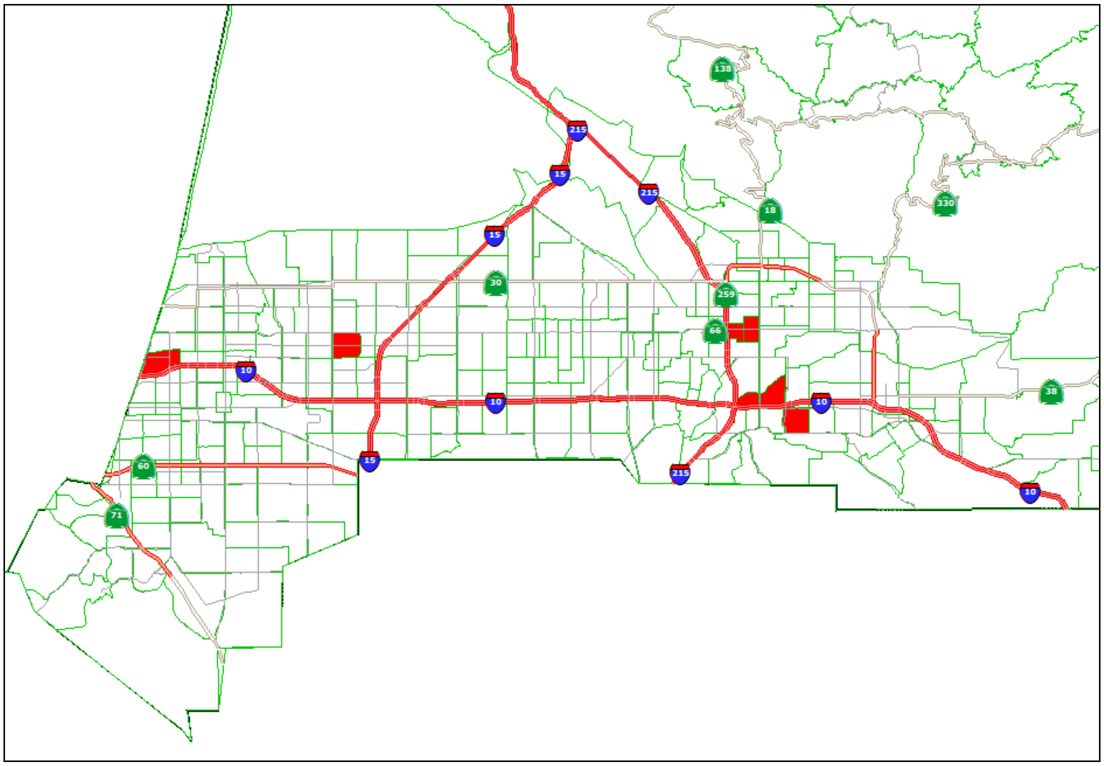


Figure 14: San Bernardino County, GS, d=10, D=10,000

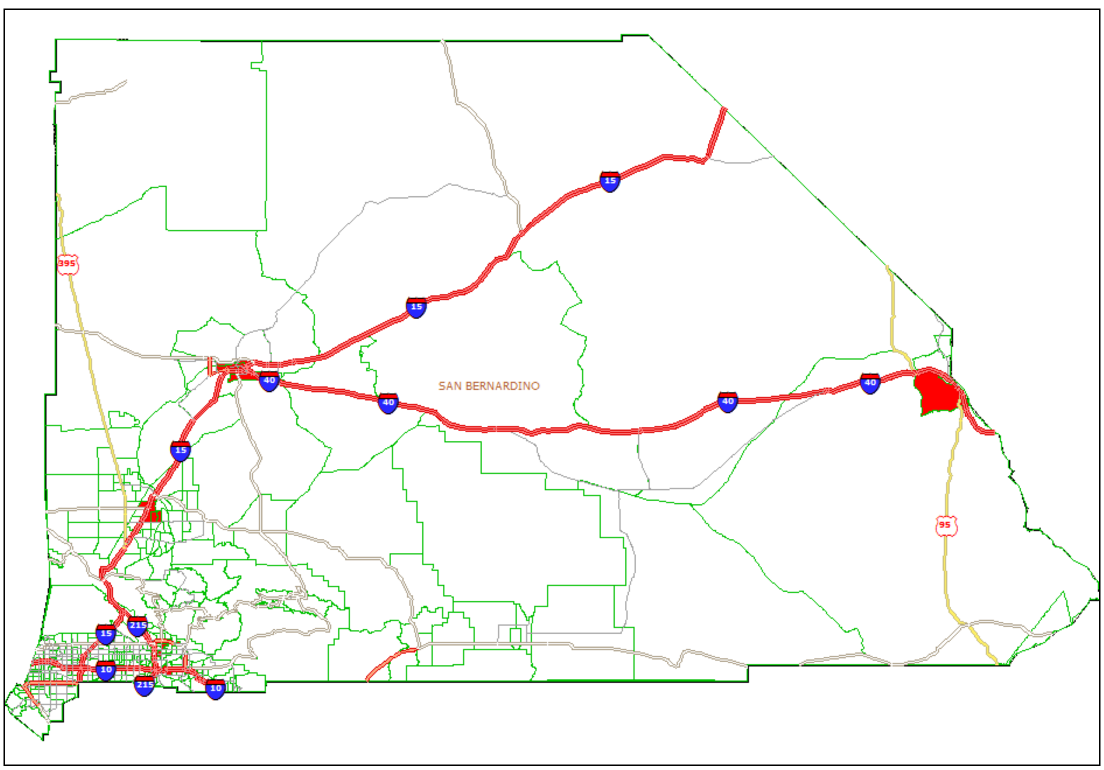


Figure 15: San Bernardino County, Non-parametric

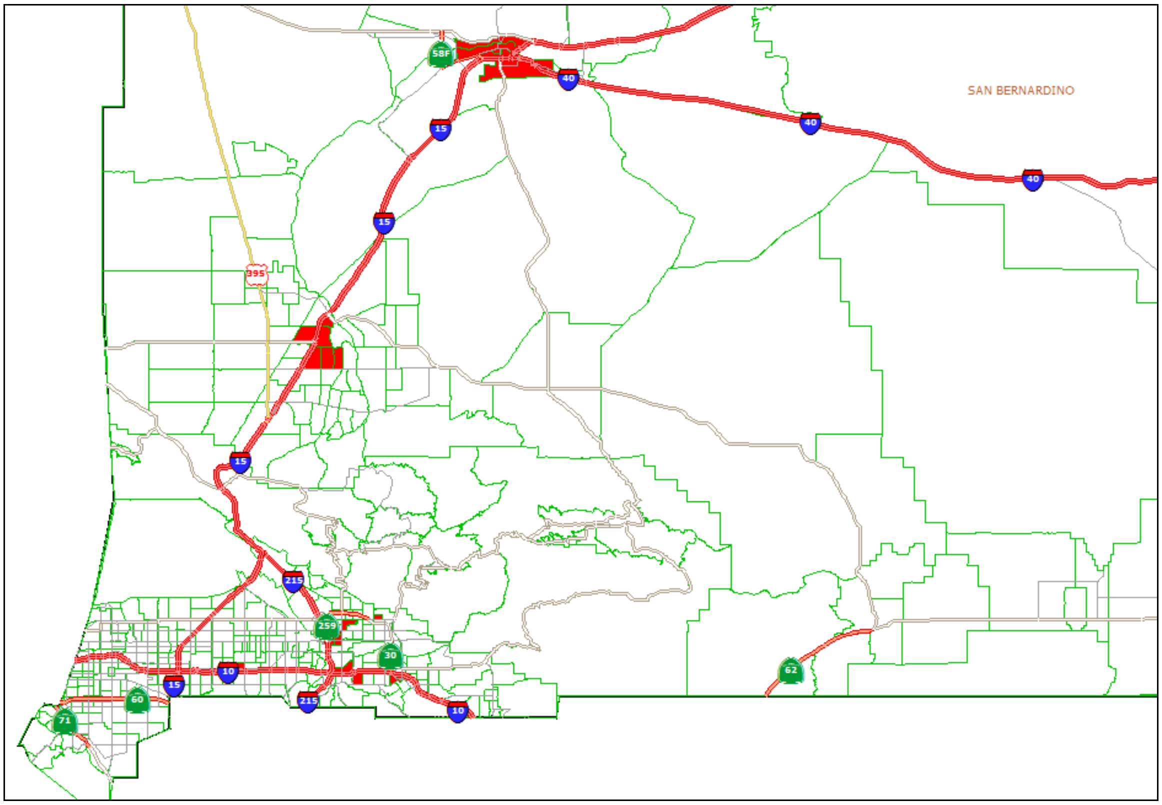


Figure 16: San Bernardino County, Western Portion, Non-parametric

San Bernardino County exhibits qualitatively the same subcenter features as Riverside County. The GS20 and LISA procedures identify no subcenters in San Bernardino County. The GS10 procedure identifies three subcenters, Ontario, Ontario Mills, and San Bernardino. The NP procedure identifies candidate TAZ's in Fontana, San Bernardino, Victorville, Barstow, and Needles, but not in Ontario. Being based on relative employment densities, it identifies "too few" subcenters in the densely populated, southwestern corner of the county, as well as all the desert cities. Unlike in Riverside County, however, each of the desert cities is arguably a "proper" subcenter.

- Ventura County

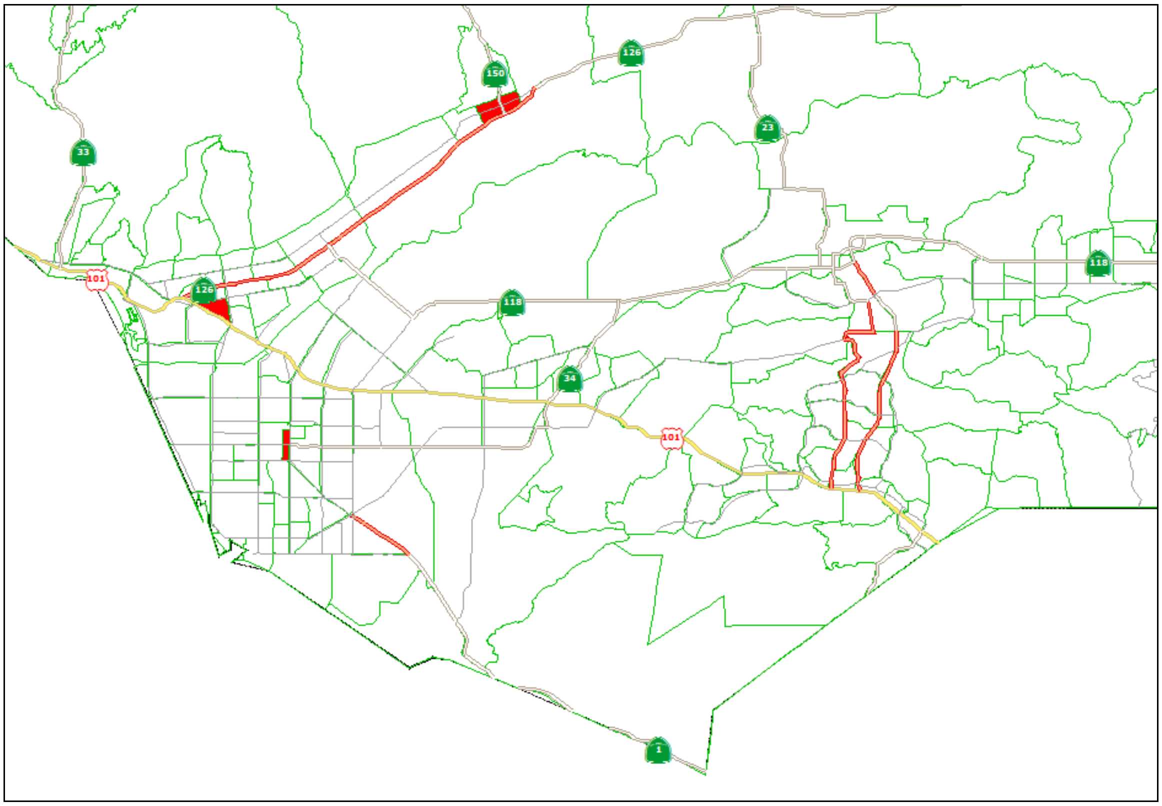


Figure 17: Ventura County, Non-parametric

The GS20, GS10, and LISA procedures identify no subcenters for Ventura County. The NP procedure identifies three candidate subcenters, Oxnard, Ventura, and Santa Paula.

- Discussion

The overall picture is clear.

The GS procedure does well in identifying employment subcenters on the basis of absolute employment density, is intuitive and easy to work with, and through adjustment of the cutoffs allows the researcher to vary the number of subcenters identified. Its major disadvantage is that it fails to identify subcenters in peripheral areas. This is not a flaw of the procedure. The procedure does what it aims to do -- identify subcenters on the basis of absolute employment density -- but in many planning contexts when identifying subcenters one wants to take relative employment density into account as well.

The NP procedure identifies zones with high relative employment density. It is not as straightforward and intuitive as the GS procedure. This is not a flaw of the procedure but reflects the ambiguity of the notion “relative employment density”. The procedure also presents the researcher with the choice of how to aggregate zones with high relative employment density into subcenters. There is no “correct” aggregation procedure. How to choose among the alternatives? The procedure succeeds in doing what it aims to do but this may give rise to counter-intuitive results. It may fail to identify well-known downtown areas as subcenters because relative employment density there is not sufficiently high. And it identifies virtually every desert community as a candidate subcenter because relative employment density there is high.

The LISA procedure performed unsatisfactorily in several respects. Not only did it fail to identify employment subcenters in peripheral areas but it also identified “too few” subcenters in areas of medium employment density. As well, it is not clear intuitively how the LISA procedure identifies a subcenter. It isn’t on the basis of absolute employment density or relative employment density or even some average of the two, but on the basis of the spatial autocorrelation of regression residuals.

We believe that procedures for identifying subcenters are more likely to gain acceptance within the planning community if they are intuitive and easy to understand. A procedure that has desirable statistical properties but is difficult to explain to a non-expert is unlikely to be adopted. For this reason, for the rest of the report, we drop the LISA procedure. The GS procedure is based on absolute densities; that’s intuitive. The NP procedure is based on relative densities; that’s intuitive too. It should be possible to come up with some hybrid or intermediate procedure that, by appropriately weighting absolute and relative densities and employment levels, comes up with a list of subcenters that accord with popular perception.

The maps illustrate another difficulty associated with the concept of an employment subcenter. Those of us brought up in traditional cities think of an employment center as indeed having a center. Most of the employment subcenters in Boston, Washington DC, New York, Philadelphia, and Chicago grew around a sub-regional center with a mini-downtown (such as Newton Center or Waltham Center in the Boston area). But in California many of the employment subcenters are really sections of a freeway corridor.

We also need to keep in mind that it is not “subcenters” that we aim to identify but “employment subcenters”. The traditional CBD was both the employment center and the “consumer” center for a metropolitan area. But today employment centers and consumer centers may be spatially distinct. A regional shopping mall is a consumer center, and often but not always an employment center.

4. Maps Using TAZ Employment Data, Actual for 2003 and Forecast for 2020 and 2035

In this section we discuss what the subcenter maps would look like in 2020 and 2035, if the SCAG TAZ forecasts for those years were realized. We shall not attempt a comprehensive discussion, but rather highlight the most important results.

Two strong empirical regularities are evident. We illustrate them with reference to Riverside County. The maps for the other counties are presented in section 7.

- Riverside County, 2003, 2020, 2035. NP Procedure.

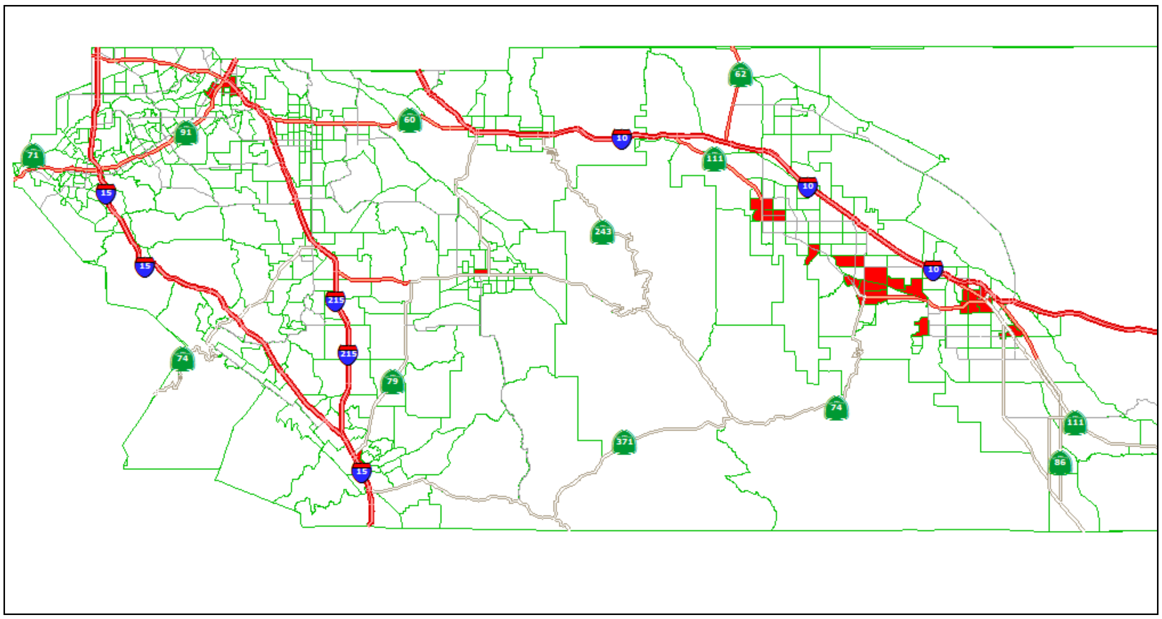


Figure 18: Riverside County 2003, Non-parametric

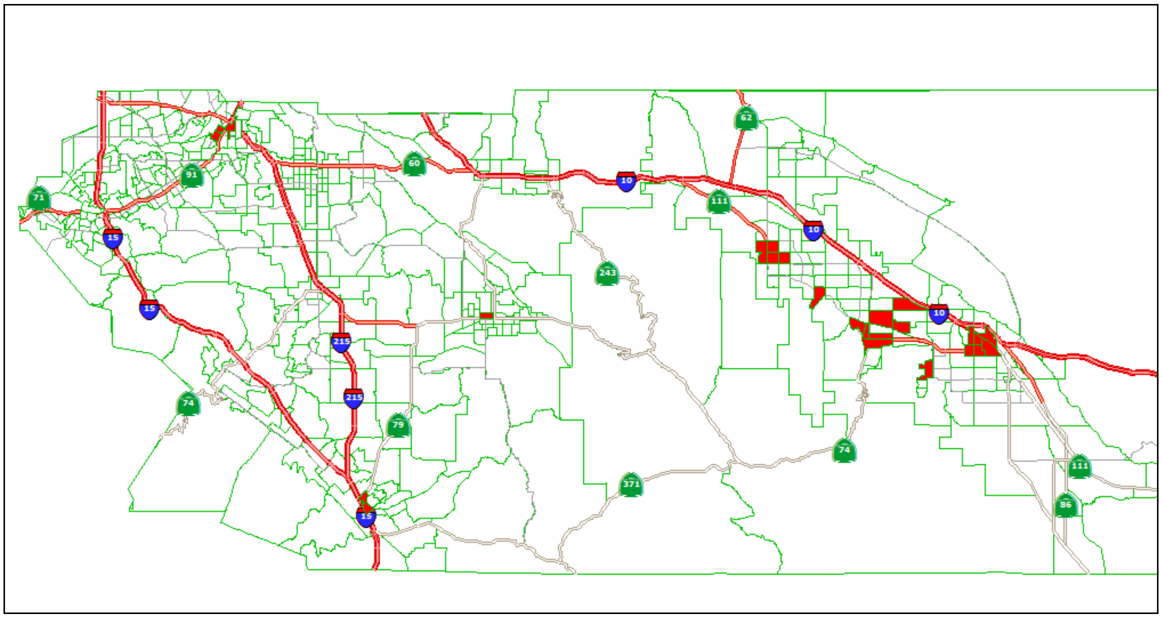


Figure 19: Riverside County 2020, Non-parametric

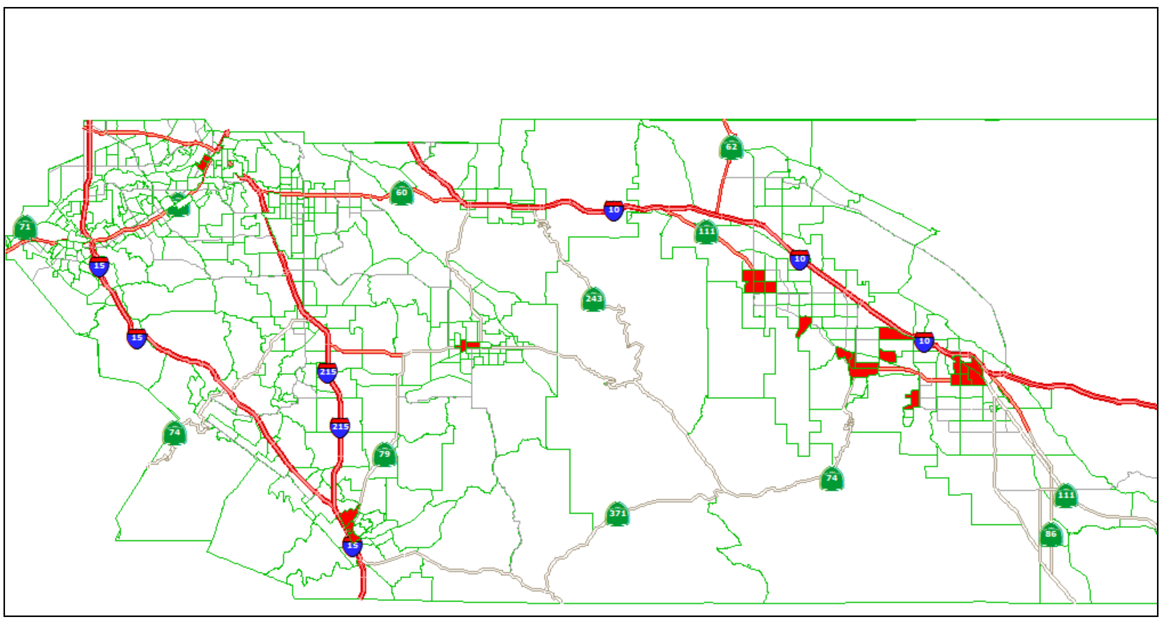


Figure 20: Riverside County 2035, Non-parametric

We start with the NP procedure. The maps for 2003, 2020, and 2035 look almost identical. Some differences can however be observed. As time proceeds, more TAZ's in Hemet and Temecula become candidate TAZ's and fewer TAZ's in Riverside and in the Palm Springs area. A likely explanation of the former result is that the employment forecasts take into account that some agricultural land around Hemet and Temecula will transition to urban use. A likely explanation of the latter results is that the employment forecasts take into account that "employment infill" will occur in the Riverside and Palm Springs areas, so that in the future fewer TAZ's there will have sufficiently high relative employment density to qualify as candidate TAZ's. Nevertheless, these changes are "second order". The striking feature of the maps is their sameness, reflecting that over the thirty-year period relative employment densities are not forecast to change much.

- Riverside County, 2003, 2020, 2035. GS Procedure.

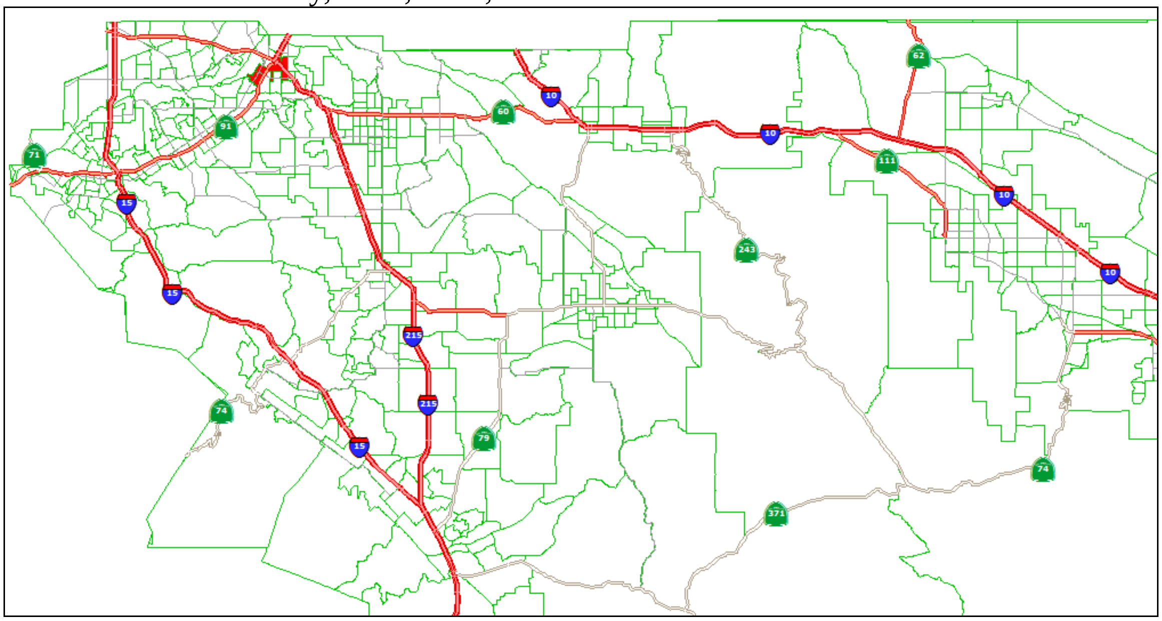


Figure 21: Riverside County 2003, GS, d=10, D=10,000

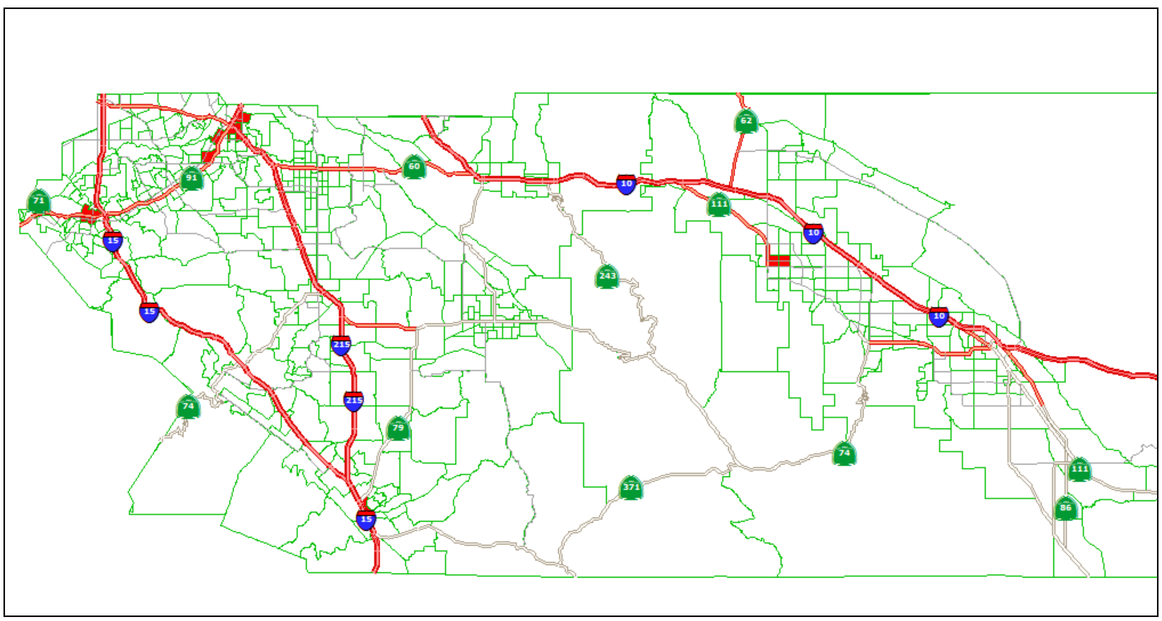


Figure 22: Riverside County 2020, GS, d=10, D=10,000

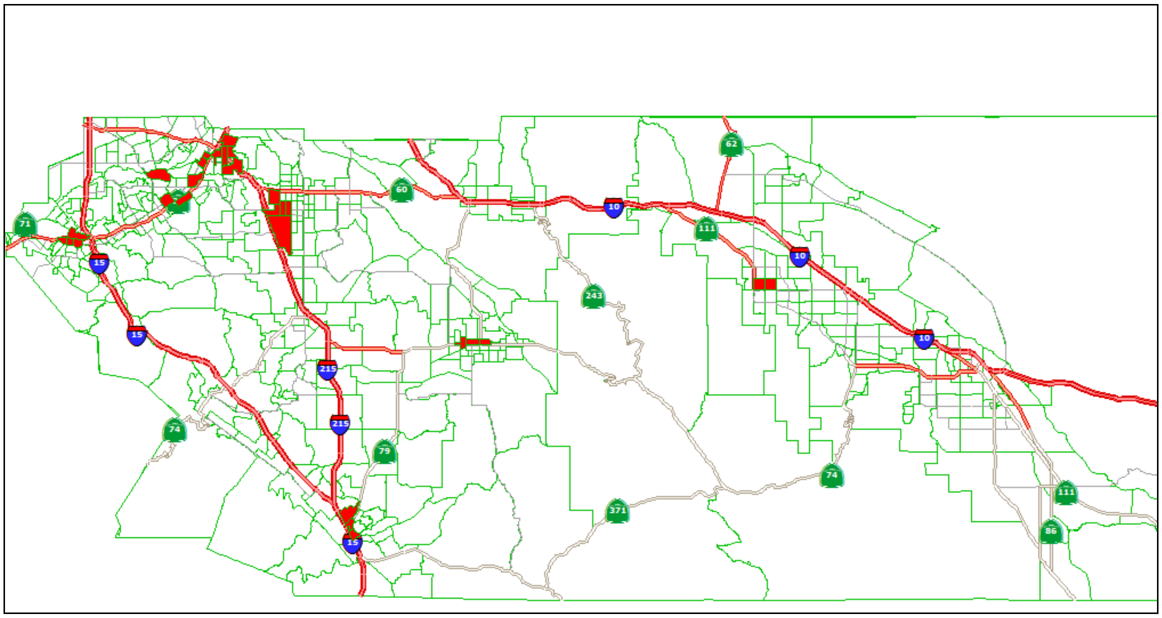


Figure 23: Riverside County 2035, GS, d=10, D=10,000

As for 2003, the GS20 procedure identifies no subcenters in Riverside County in either 2020 or 2035. It does identify some TAZ's in the area of Riverside that meet the minimum density criterion but there is insufficient total employment, even in 2035, for the set of contiguous TAZ's to meet the subcenter criterion.

The GS10 procedure generates more interesting results. In 2003, Riverside is the only subcenter. By 2020, the number of TAZ's in the Riverside subcenter has increased, new subcenters have formed at Palm Springs and Corona, and one TAZ at Temecula meets the minimum employment density but not the total employment criterion. By 2035, the Riverside, Palm Springs, and Corona subcenters have expanded considerably, Moreno Valley and Temecula are added to the list of subcenters, and Hemet has several TAZ's that meet the density criterion but together do not the total employment criterion. The results for the GS10 procedure with forecast employment are especially interesting since they suggest a procedure for identifying emerging subcenters. If one were to ask anyone familiar with Riverside County where the next employment subcenters will form, they would likely give answers very similar to the GS10 results. Thus, the GS10 results with forecast employment are consistent with intuition concerning the location of emerging employment subcenters.

- Comparison of Riverside County NP, 2003, and Riverside County GS10, 2035

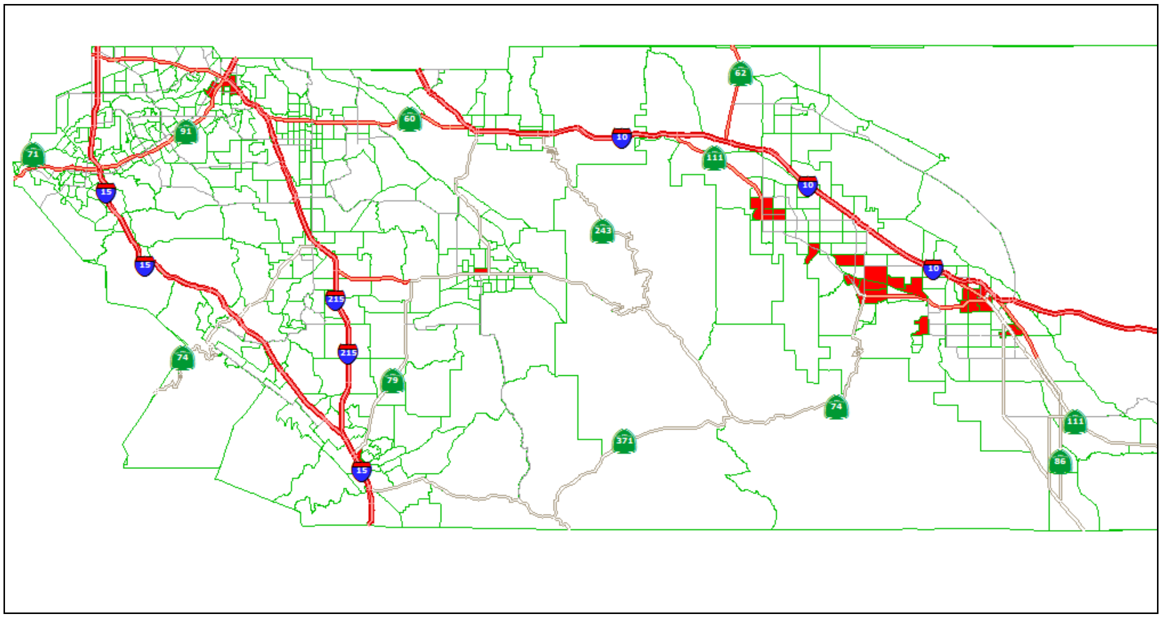


Figure 24: Riverside County 2003, Non-parametric

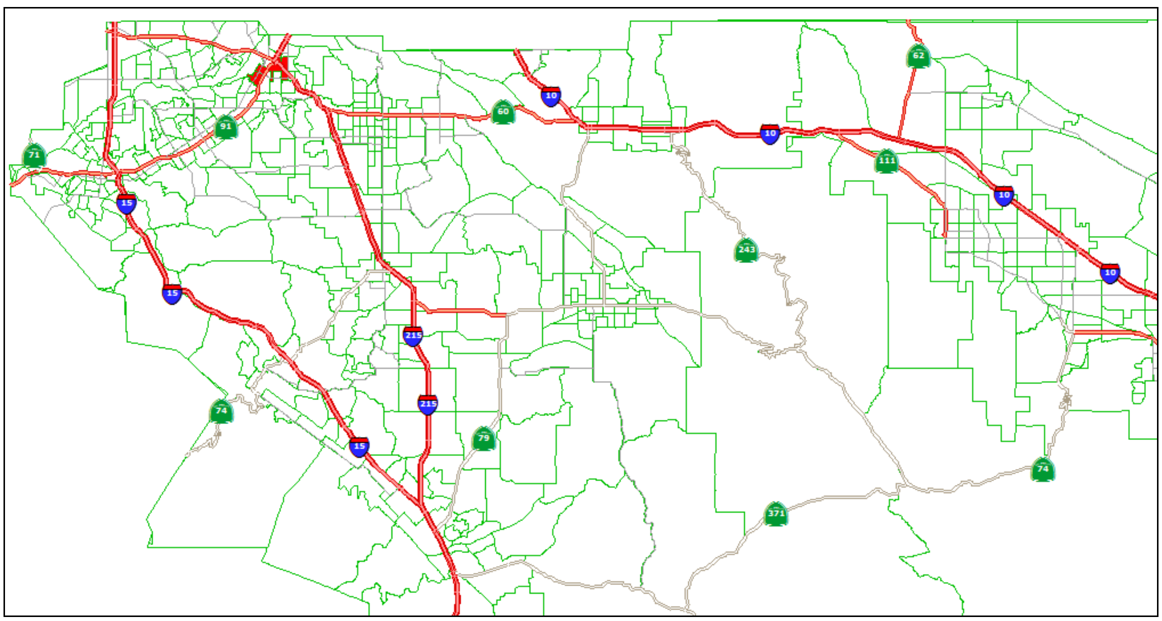


Figure 25: Riverside County 2003, GS, d=10, D=10,000

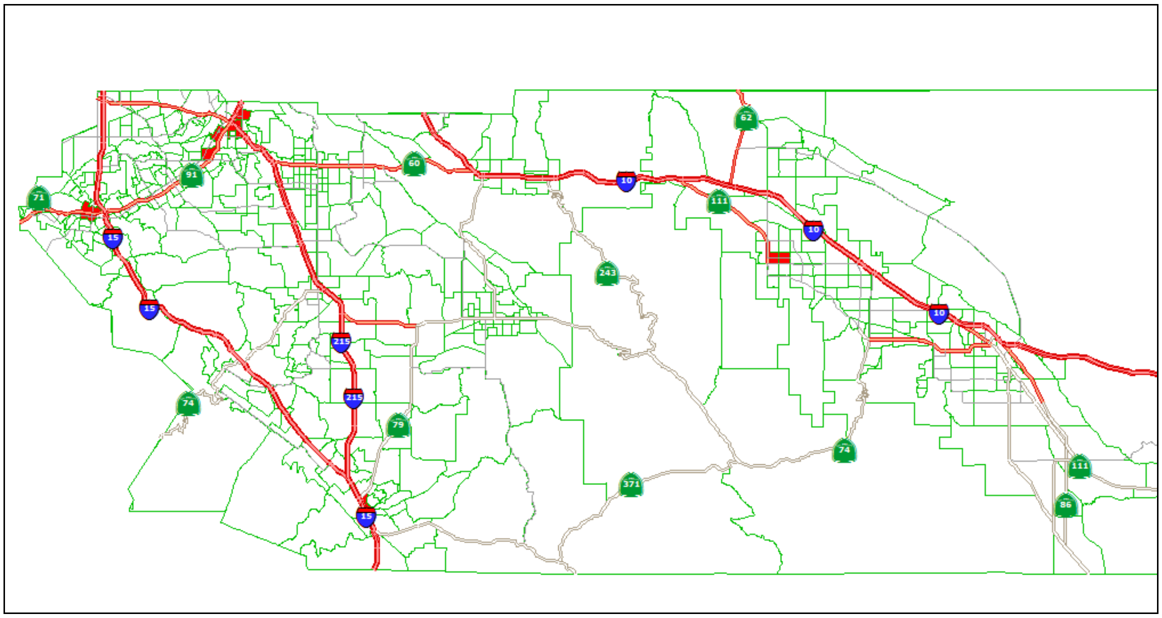


Figure 26: Riverside County 2020, GS, d=10, D=10,000

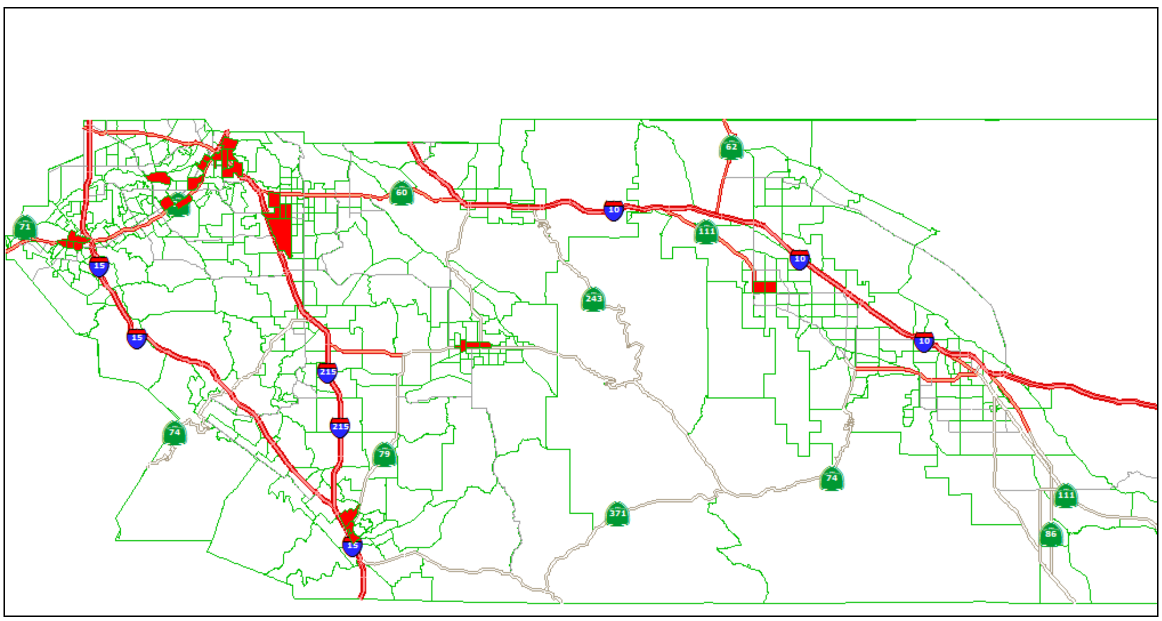


Figure 27: Riverside County 2035, GS, d=10, D=10,000

A possible alternative procedure for identifying emerging subcenters is to identify areas with high relative employment density and some minimal level of total employment. At least for Riverside County, this procedure does not seem to be very successful. The procedure identifies several cities in the Palm Springs Area as candidate TAZ's, as well as one TAZ in Temecula and another in Hemet, but fails to identify Moreno Valley and Corona.

When employment forecasts are available, we conjecture that applying GS10 with forecasted employment will turn out to be a more generally satisfactory procedure than NP in identifying emerging subcenters. This should not be surprising since GS10 with forecasted employment uses more information about future growth prospects – that contained in the employment forecasts. When employment forecasts are not available, NP should be somewhat successful in identifying emerging subcenters.

5. Directions for Future Work: Alternative Hybrid Procedures

There is no right way to define a subcenter. One might want to define subcenters on the basis of the use to which the definition is to be put. Thus, one might wish to define subcenters differently in the context of transportation planning than in the context of planning for retail development. But if the concept of a subcenter is to gain popular acceptance, it needs to accord reasonably well with our intuitive, albeit inchoate, concept of what a subcenter is.

The GS definition of an employment subcenter, based on absolute densities, does a good job of identifying subcenters in the metropolitan core, but identifies too few subcenters at the metropolitan periphery. The NP definition, based on relative densities, succeeds in identifying subcenters at the metropolitan periphery but not in the metropolitan core. What is needed is some hybrid procedure that does a good job of identifying subcenters throughout the metropolitan area.

We have informally started investigating several hybrid procedures.

1. Using the GS definition of a subcenter but allowing the cutoffs to vary by county (suggested by Huasha Liu).

This procedure has the appeal of simplicity, and also accommodates the political reality that much planning is done on a county basis. One disadvantage it has is that the definition it employs is discontinuous. Consider, for example, an employment "area" that straddles the boundary between Los Angeles and San Bernardino Counties. If a lower cutoff is applied to San Bernardino County than to Los Angeles County, that part of the employment area in San Bernardino County might be classified as an employment center but not that part in Los Angeles County. Another disadvantage is that, while it takes into account differences in "background" employment density between counties, it does not do so within counties.

2. Using the GS definition of a subcenter with a high cutoff and the NP definition of a subcenter with a high cutoff, and defining a hybrid subcenter to be one that meets *either* of the two definitions (also suggested by Huasha Liu).

At central locations, this procedure would tend to identify subcenters on the basis of absolute densities, and at peripheral locations on the basis of relative densities. One disadvantage of this procedure is that it would likely identify too few subcenters at intermediate locations, where neither the absolute nor the relative employment density criteria are met.

3. Using some scoring procedure that attaches weight to both absolute and relative density criteria in defining a subcenter. The GS and NP procedures would therefore lie at the two extremes of this procedure. This procedures holds promise, but how to make it operational?

4. There are different kinds of subcenters. Employment subcenters are particularly easy to work with since employment data are readily available. But one might wish to include non-employment-related criteria in deciding what is and is not a subcenter. One could, for example, define a subcenter on the basis of trip originations and destinations. Intuitively, such a definition would tend to identify shopping centers, to which there are a large number of trips of relatively short duration. One might term this an “activity” subcenter. One could also define a general subcenter on the basis of some weighted average of employment and trip criteria, or on the basis of the total number of person-hours spent in different TAZ’s.

For both procedures, thought needs to be given as to how to treat “corridor subcenters”. Some of the difficulties are illustrated by the Palm Springs area, where the NP procedure identified multiple subcenters, whereas intuition strongly suggests that the whole area from Palm Springs to Coachella is one subcenter. Perhaps the definition of contiguity should be relaxed somewhat.

One wants an intuitive definition of a subcenter. One also wants a standardized definition to ensure comparability between different studies. Unfortunately, reasonable people may differ on an intuitive definition of a subcenter, so that there is a tradeoff between standardization in definition and flexibility.

6. Conclusions

The focus of this final report has been on the comparison of three alternative methods of identifying employment subcenters, applied to the Los Angeles Metropolitan Area. We have also touched on a number of related issues.

The Giuliano-Small (GS) procedure defines a subcenter to be a set of “zones”, each of which has an employment density exceeding a cutoff level d, and all of which together have a total employment exceeding a cutoff level D. It bases the definition on *absolute* employment densities and total levels of employment. The GS procedure can be defined as a two-step procedure, the first step identifying candidate zones for inclusion in subcenters on the basis of employment density, the second identifying sets of contiguous candidate zones that together meet the overall employment criterion. Conditional on being based on absolute employment densities and levels, the GS procedure is a good one. It is intuitive and easy to apply. Its major weakness is that, since it is based on absolute employment densities and levels, it fails to identify subcenters at the periphery of a metropolitan area that will likely become GS subcenters in the future.

There are many possible non-parametric (NP) procedures. The one employed here is a two-step procedure. The first step identifies a zone as a candidate for inclusion in an employment subcenter on the basis of the ratio of its employment density to the average employment density around it. Thus, the condition for candidacy is *relative* employment density. One way in which NP procedures differ is in how they calculate the average employment density around each zone (or, alternatively, the smoothed employment density of that zone). The second step is identical to the second step of the GS procedure. It identifies sets of contiguous candidate zones that together meet an overall employment criterion. Conditional on being based on relative employment densities, the NP procedure is a good one. Though the statistics involved in formalizing the notion of "relative employment density" are quite complex, the notion itself is intuitive. Its major weakness is that it identifies "too few" employment subcenters in the metropolitan core, and too many in the peripheral desert areas of Metropolitan Los Angeles since every city in the desert has a high relative employment density.

We also considered a third procedure, LISA, which is based on spatial autocorrelation, but rejected it, since its method is more technical/statistical and less intuitive than the other two procedures and since its results are no more satisfactory.

There is no right definition of an employment subcenter. The best definition depends on context. SCAG will be employing the definition in the context of transportation, land use, and environmental planning. For this purpose identifying emerging subcenters and traffic corridors is particularly important. As well, since the transportation planning process involves close consultation with the county planning authorities, it is important to identify the subcenters of each county, including Imperial, Riverside, San Bernardino, and Venture Counties, whose overall employment densities are considerably lower than those of Los Angeles and Orange Counties. In this context, neither the GS definition by itself or the NP definition by itself is satisfactory. The GS procedure identifies too few subcenters at the metropolitan periphery, while the NP procedure identifies too many subcenters at the periphery and too few at the core. What is needed is an intermediate or hybrid procedure that takes into account both absolute and relative employment densities in its definition of a subcenter, and that succeeds in identifying what are intuitively the subcenters in both the core and the periphery.

Developing such an intermediate or hybrid procedure was not an element of the current project. However, we plan to submit a follow-up proposal with this aim. The work in the follow-up proposal would present some half dozen alternative intermediate/hybrid procedures for identifying subcenters at the metropolitan periphery and/or emerging subcenters, apply them to the SCAG region, and comment on the strengths and weaknesses of each. These procedures would differ in the weights they apply to absolute and relative densities, and to employment and traffic, in determining candidate zones for inclusion in subcenters, and in whether their aim is to identify current peripheral subcenters (based on current data) or emerging subcenters (using forecasted data). Hopefully SCAG would find that one or more of these procedures is well suited to a subset of its planning tasks.

7. Other Maps

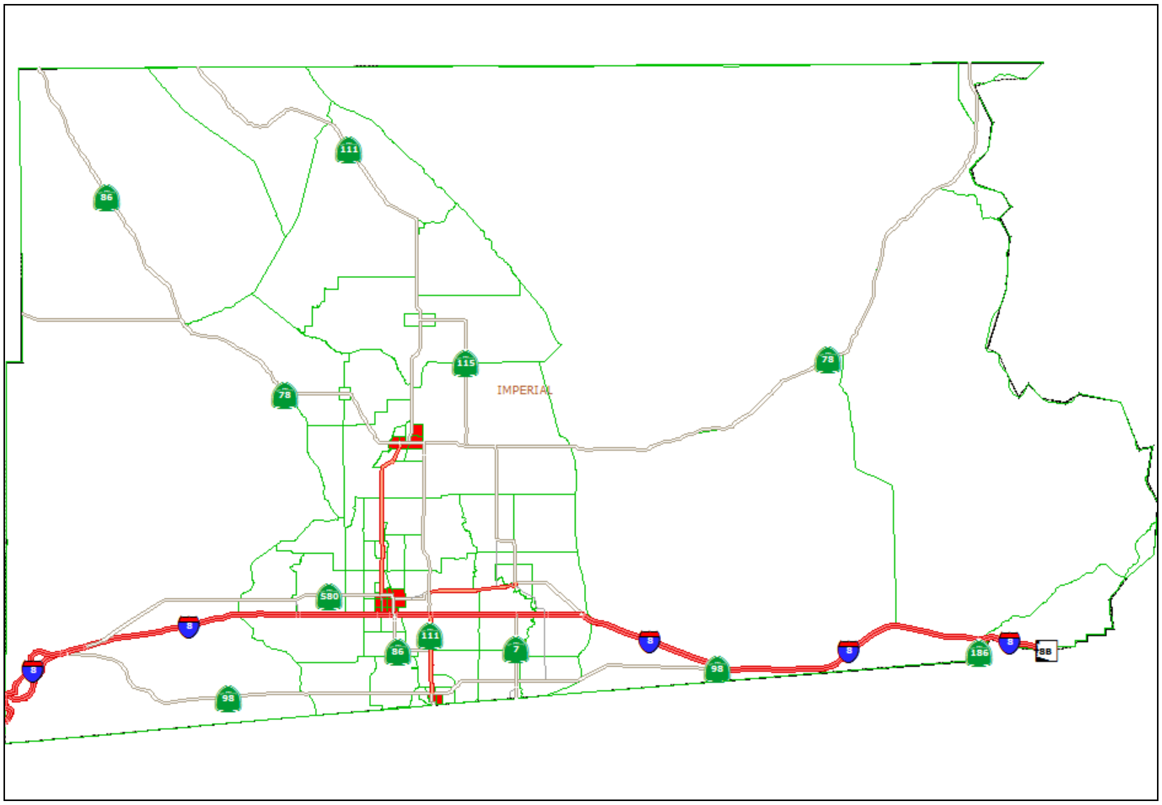


Figure 28: Imperial County 2020, Non-parametric

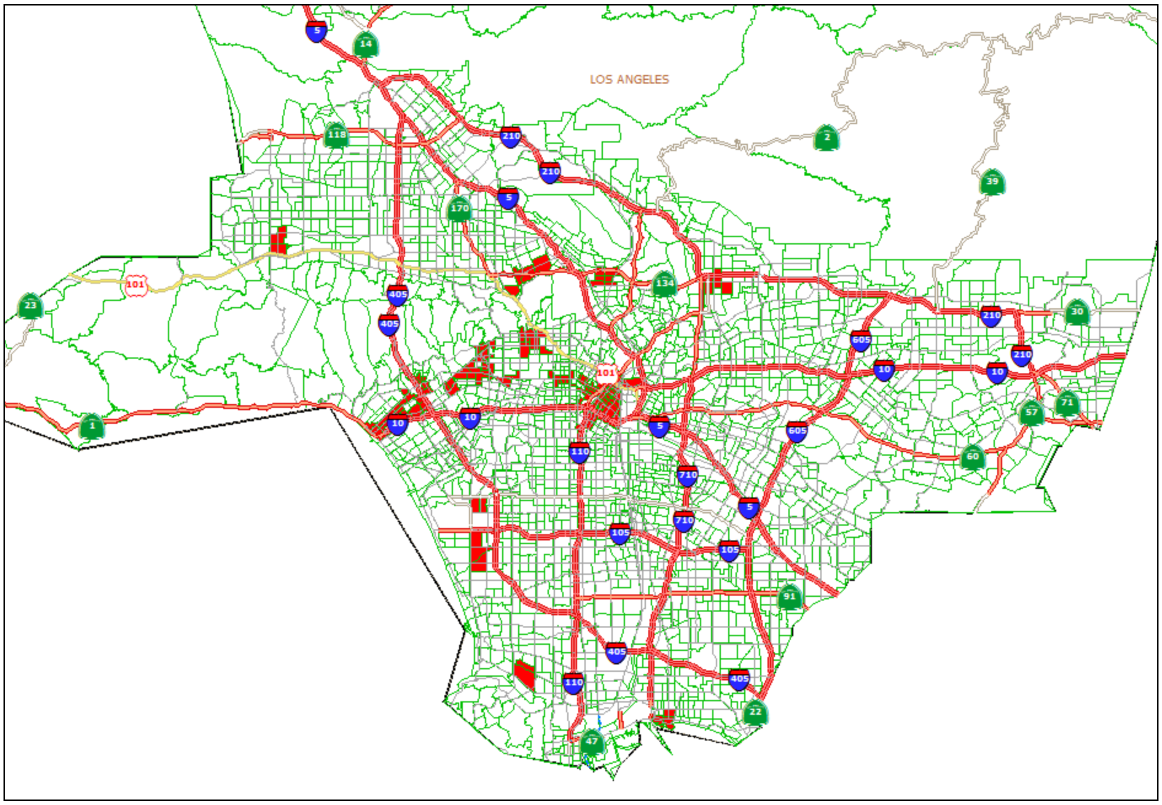


Figure 29: Los Angeles County 2020, GS, d=20, D=20,000

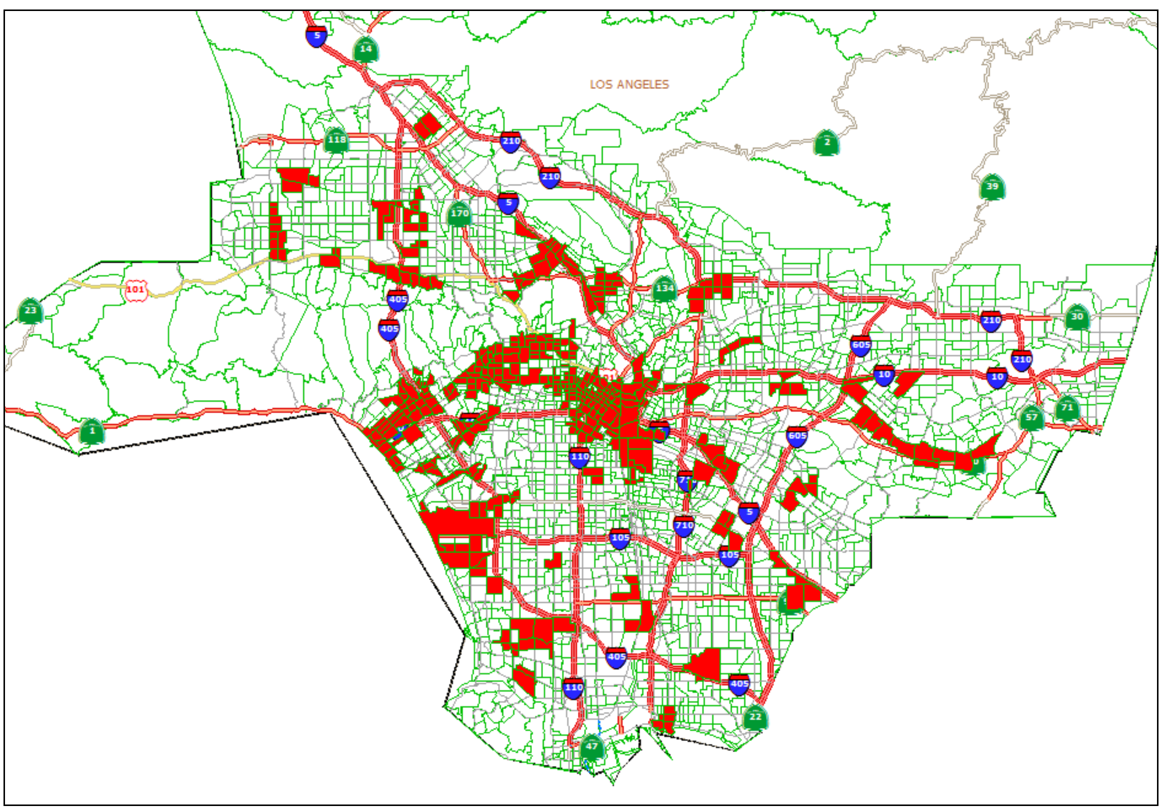


Figure 30: Los Angeles County 2020, GS, d=10, D=10,000

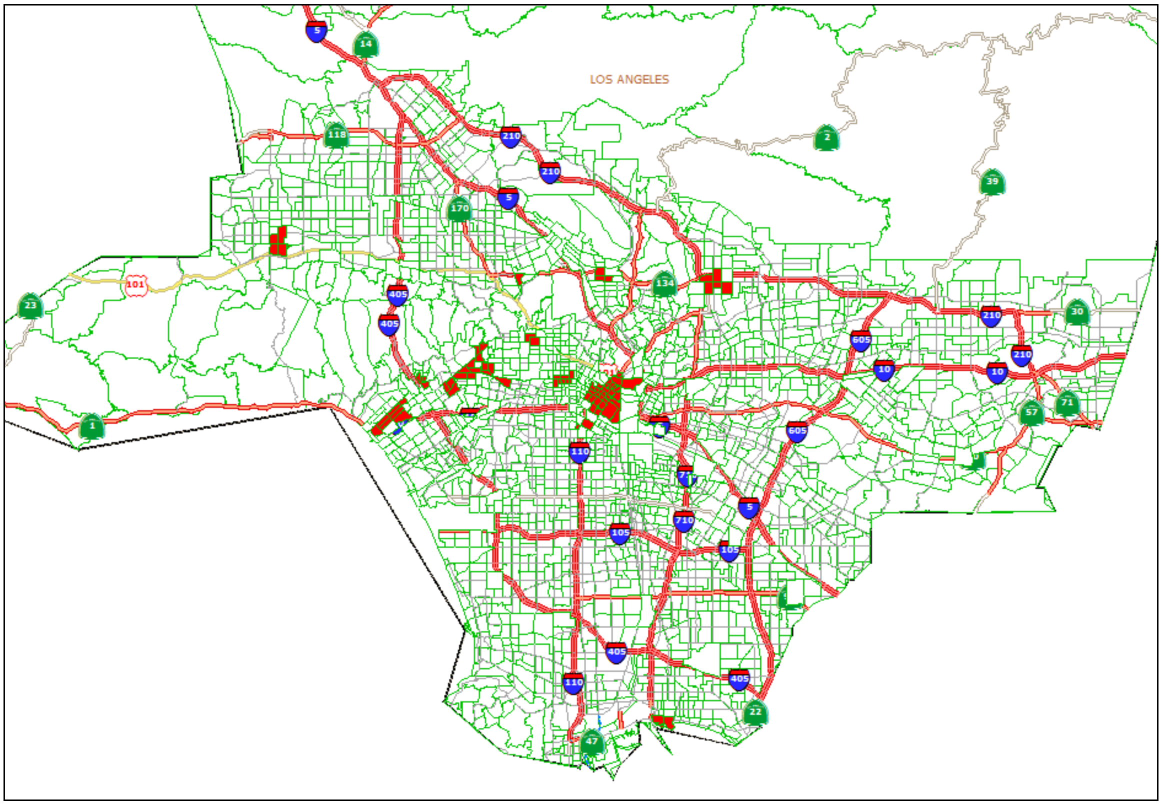


Figure 31: Los Angeles County 2020, LISA

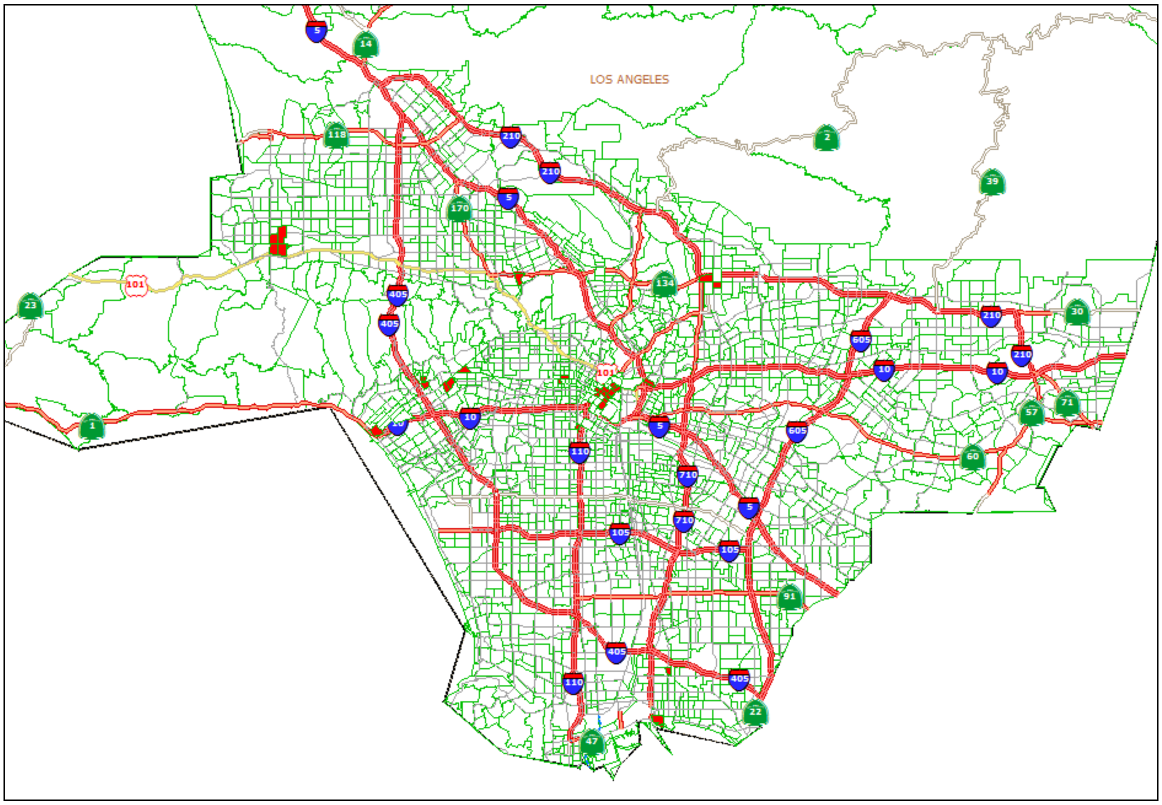


Figure 32: Los Angeles County 2020, Non-parametric

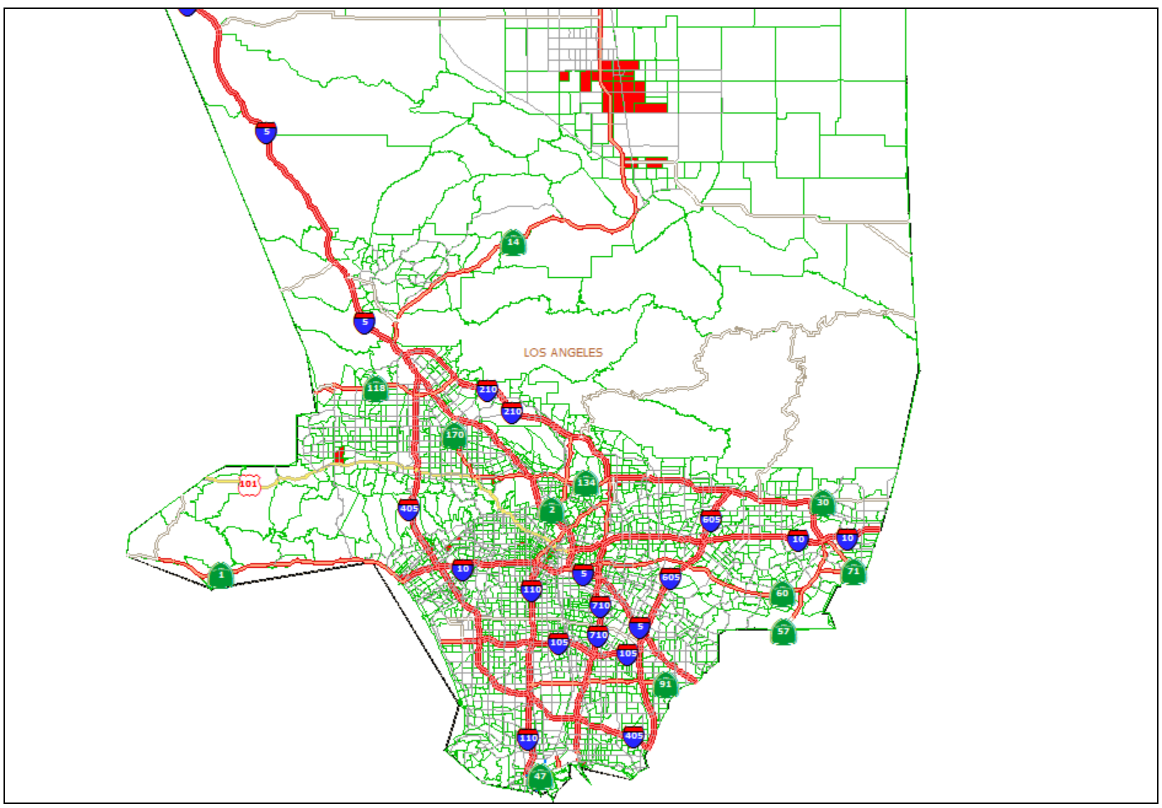


Figure 33: Los Angeles County 2020, Full County, Non-parametric

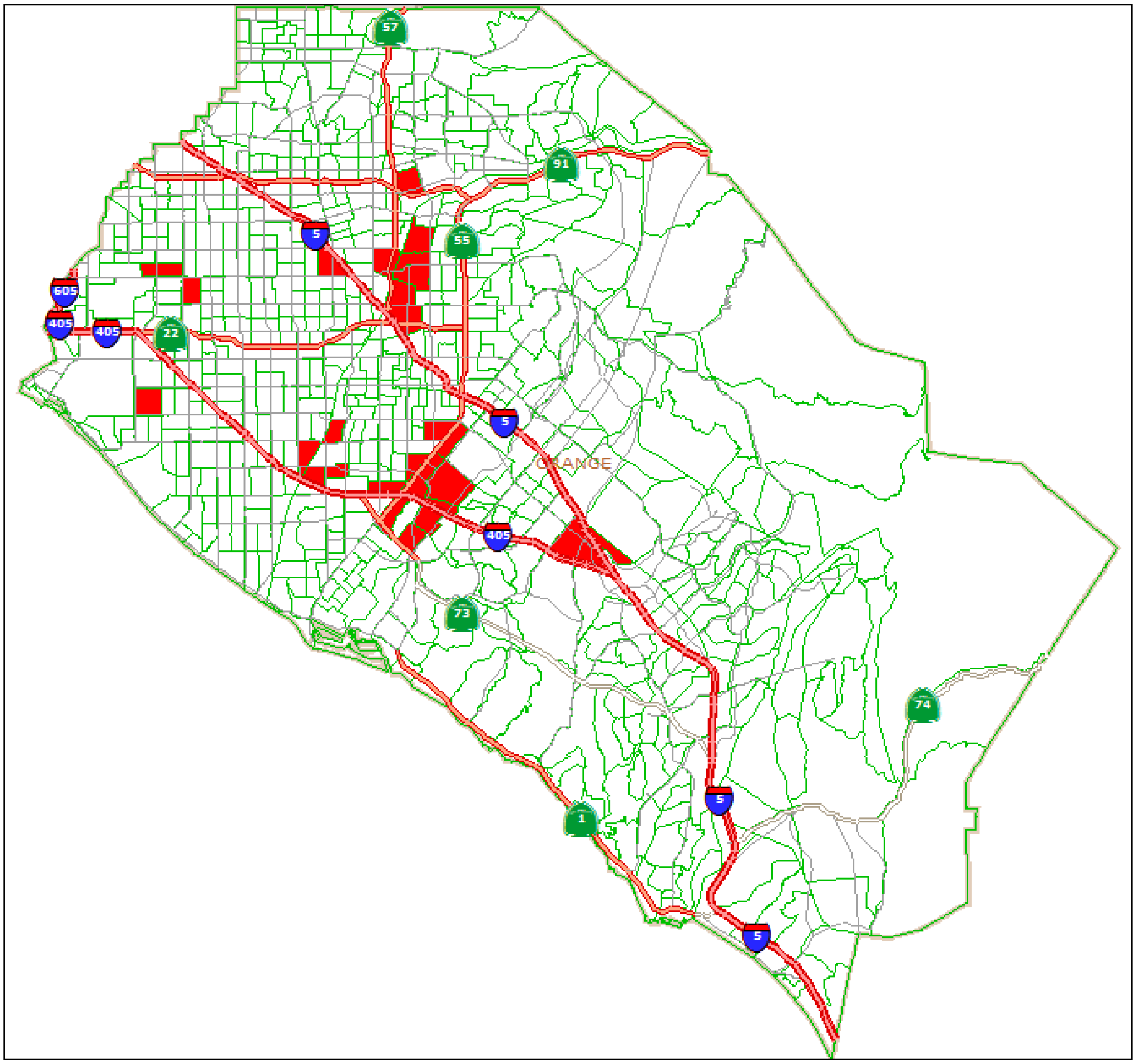


Figure 34: Orange County 2020, GS, d=20, D=20,000

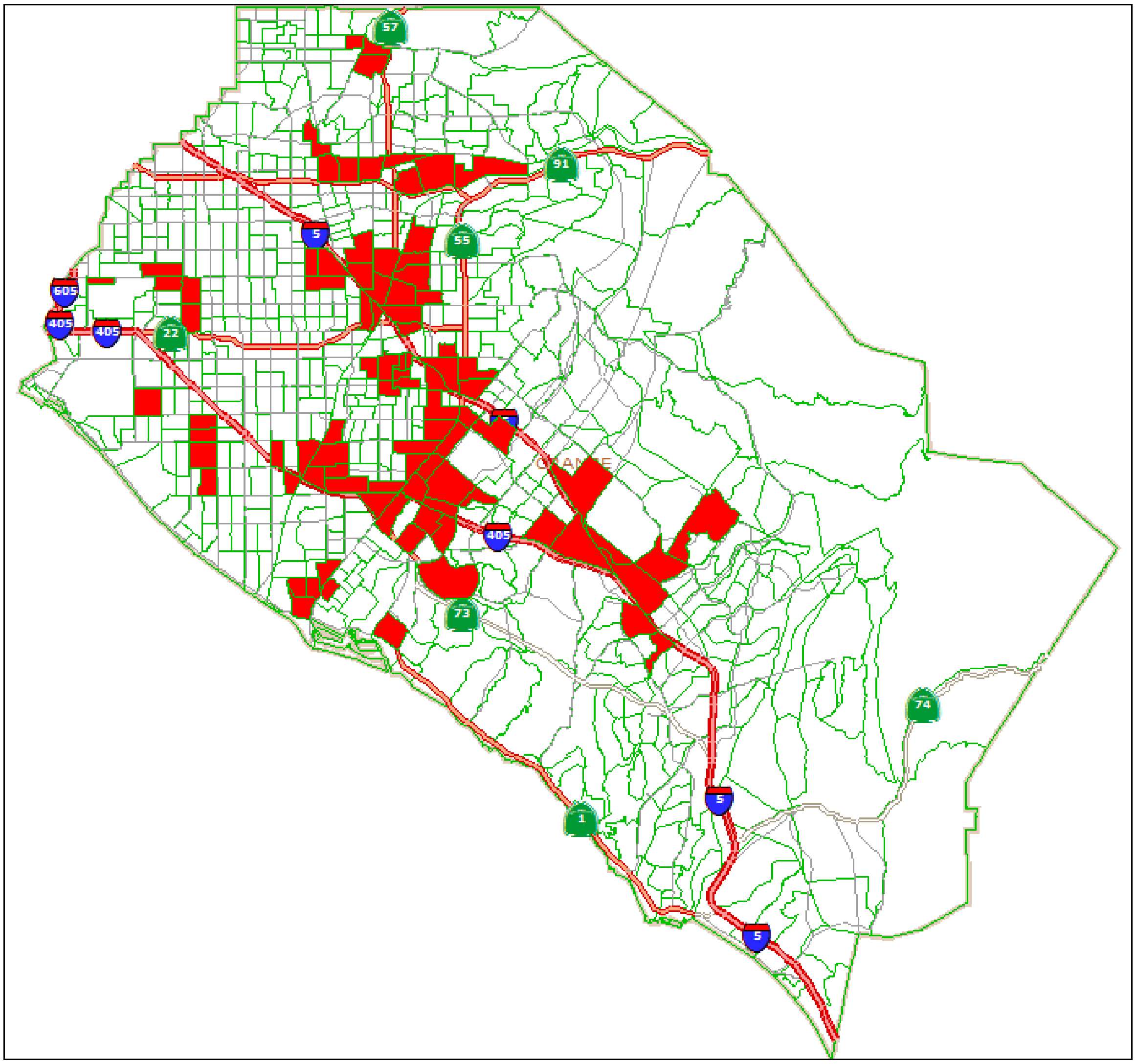


Figure 35: Orange County 2020, GS, d=10, D=10,000

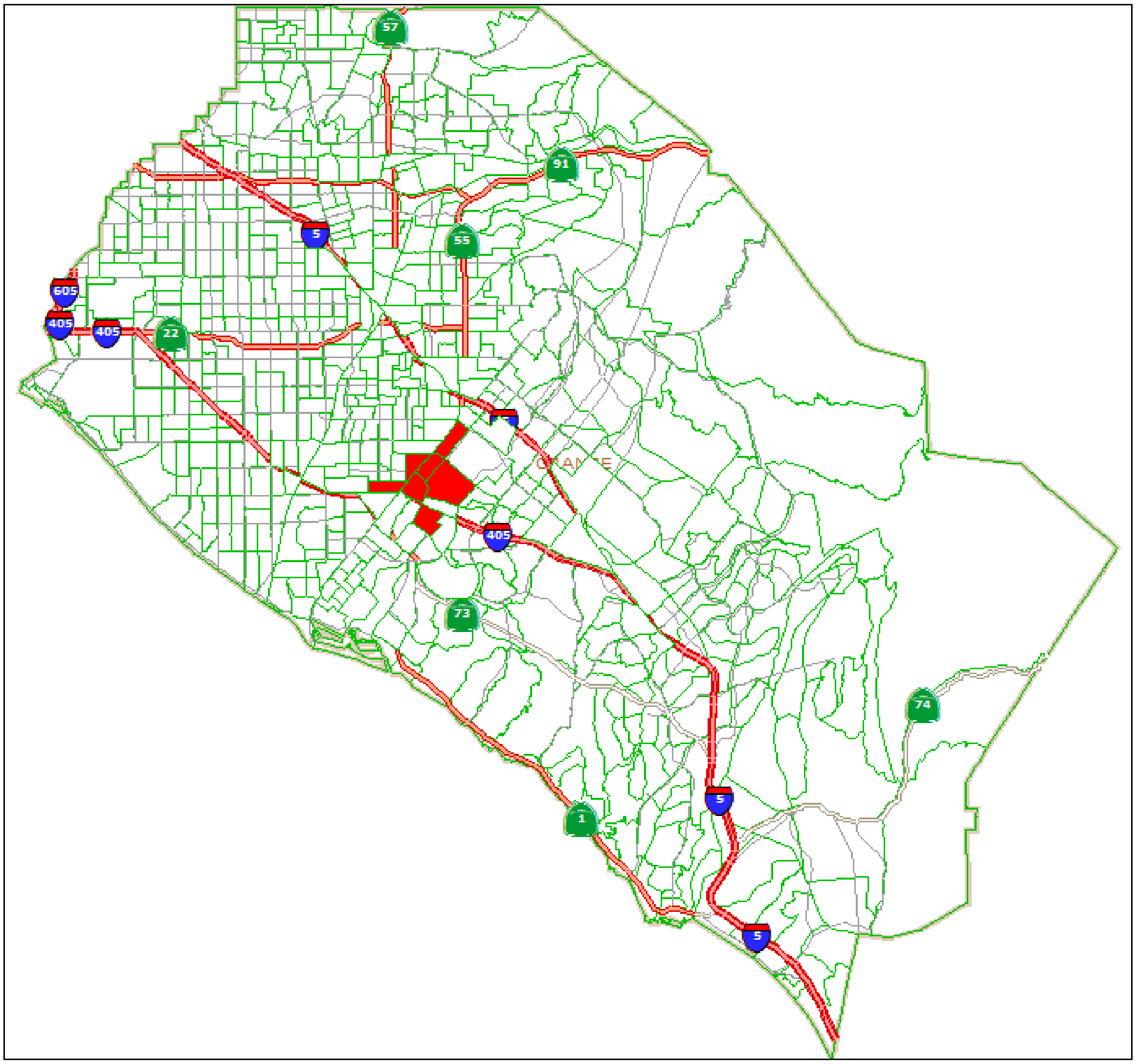


Figure 36: Orange County 2020, LISA

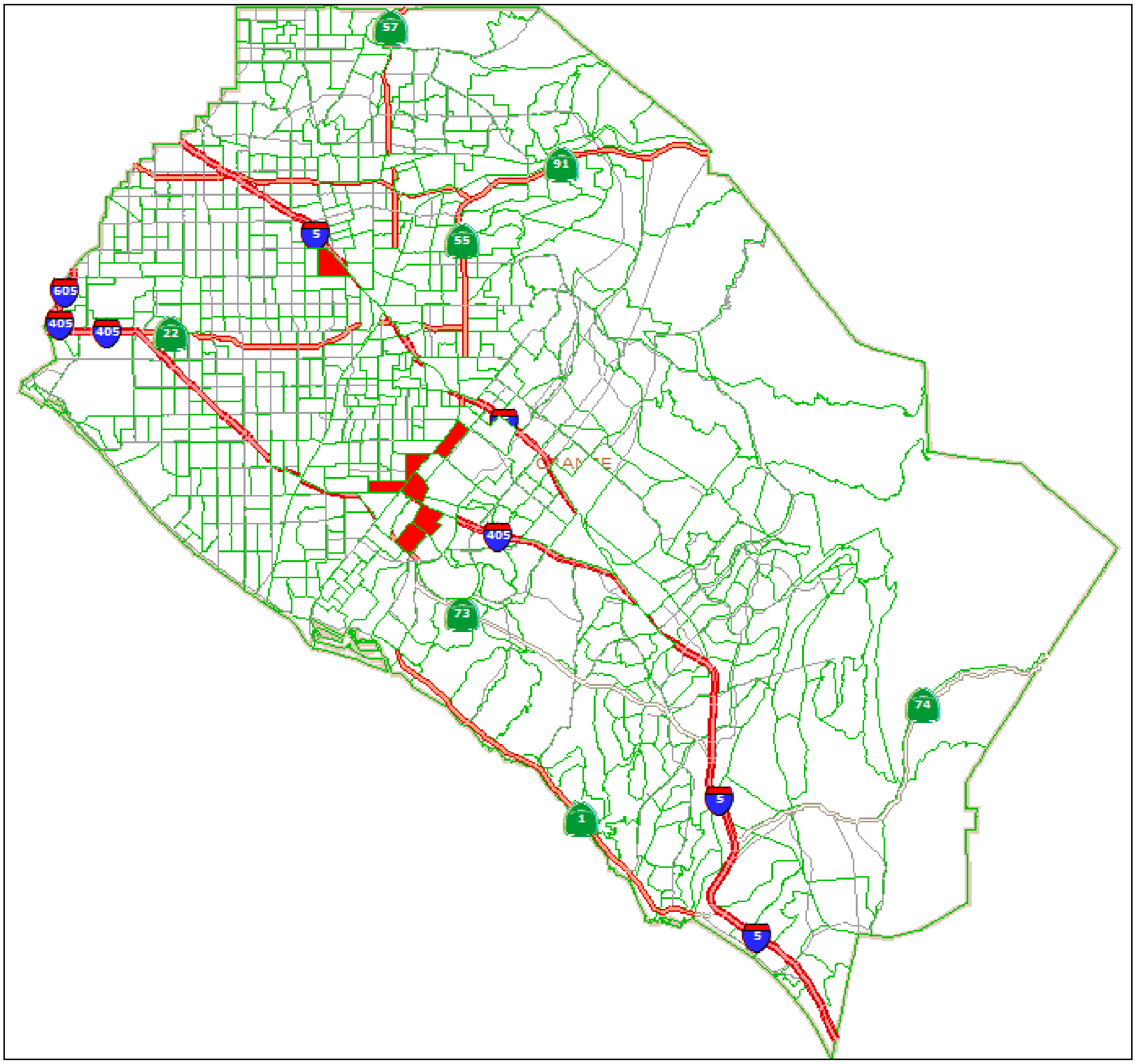


Figure 37: Orange County 2020, Non-parametric

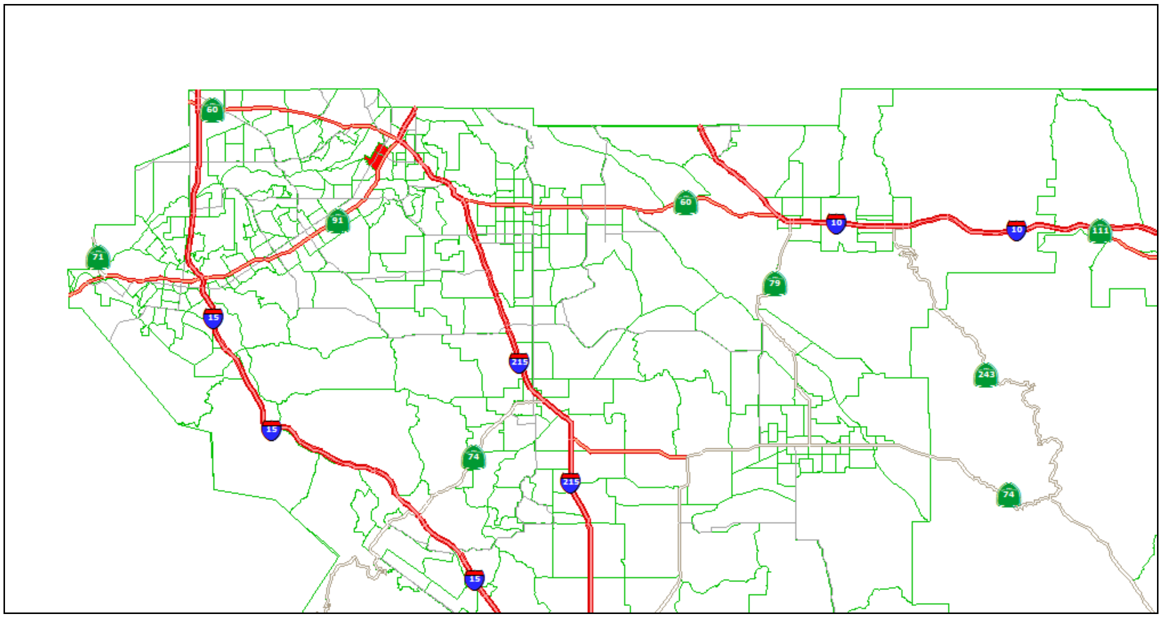


Figure 38: Riverside County 2020, GS, d=20, D=20,000

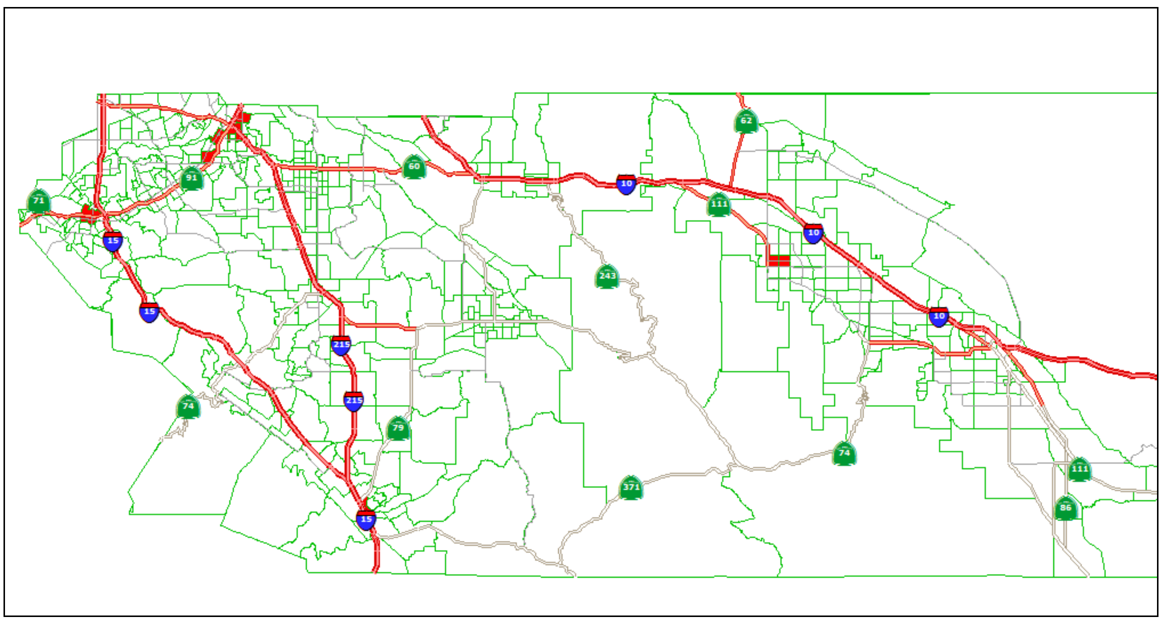


Figure 39: Riverside County 2020, GS, d=10, D=10,000

LISA: none

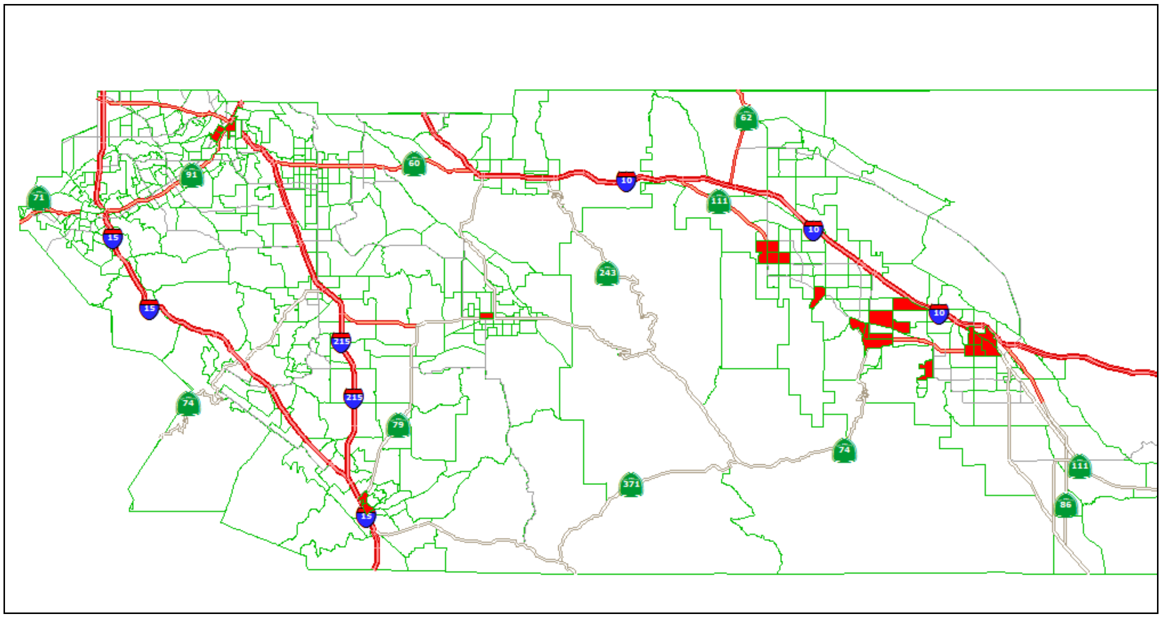


Figure 40: Riverside County 2020, Non-parametric

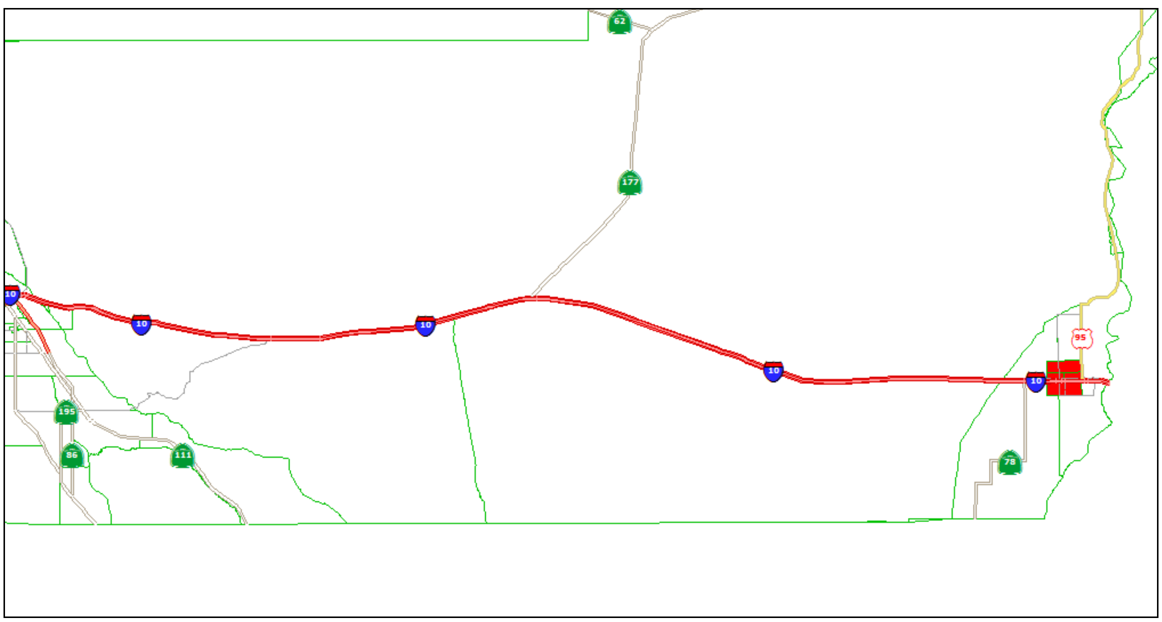


Figure 41: Riverside County 2020, Eastern Portion, Non-parametric

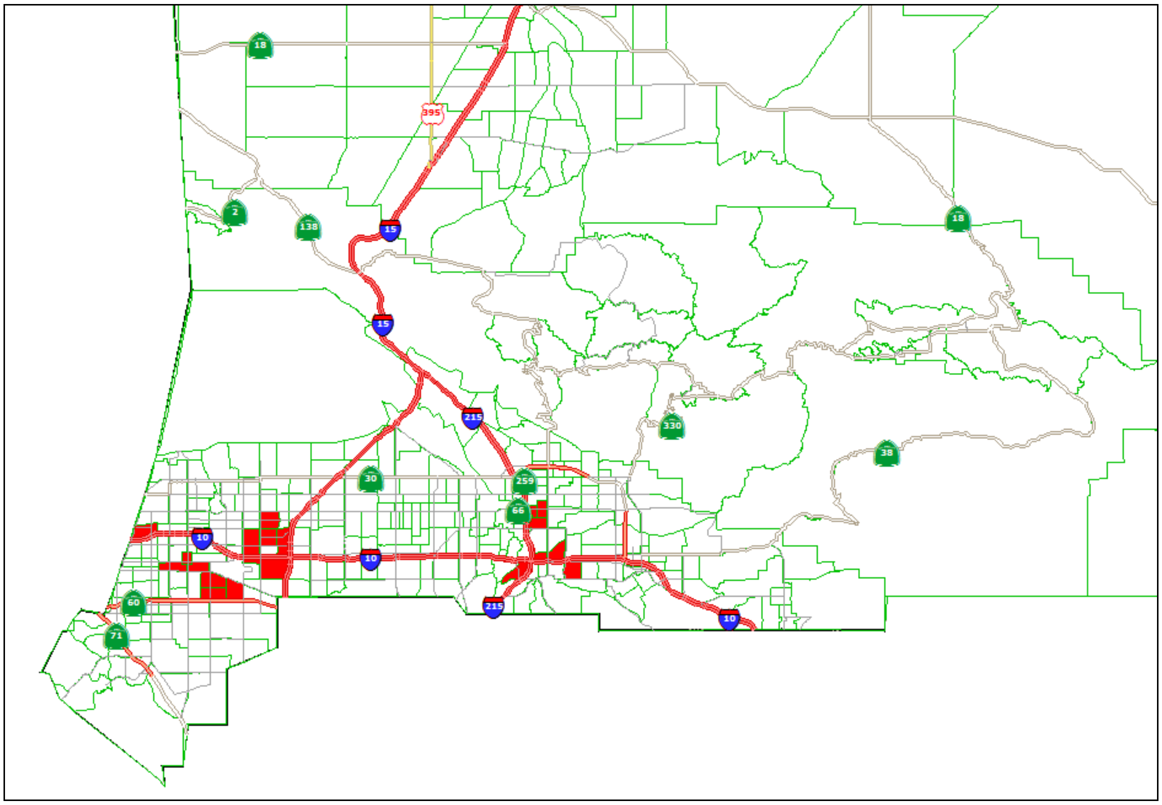


Figure 42: San Bernardino County 2020, GS, d=10, D=10,000
(none with d=20, D=20,000)
(none by LISA)

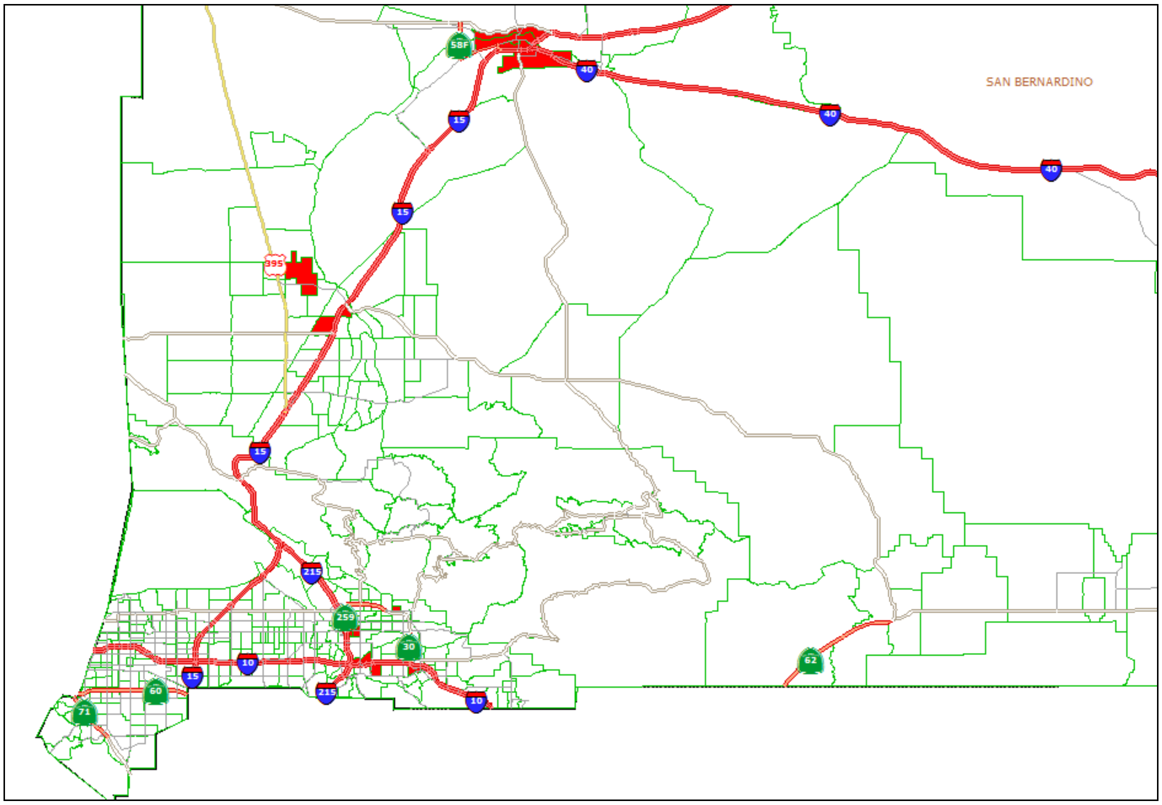


Figure 43: San Bernardino County 2020, Non-parametric

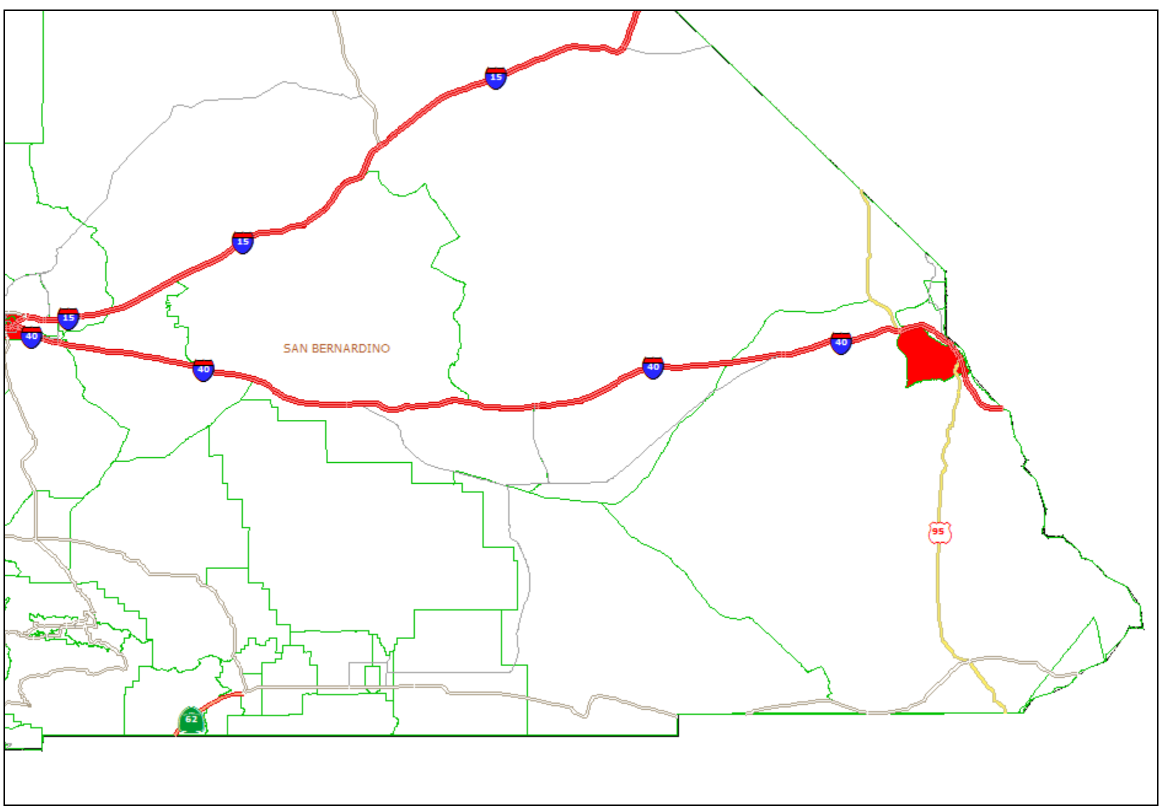


Figure 44: San Bernardino County 2020, Eastern Portion, Non-parametric

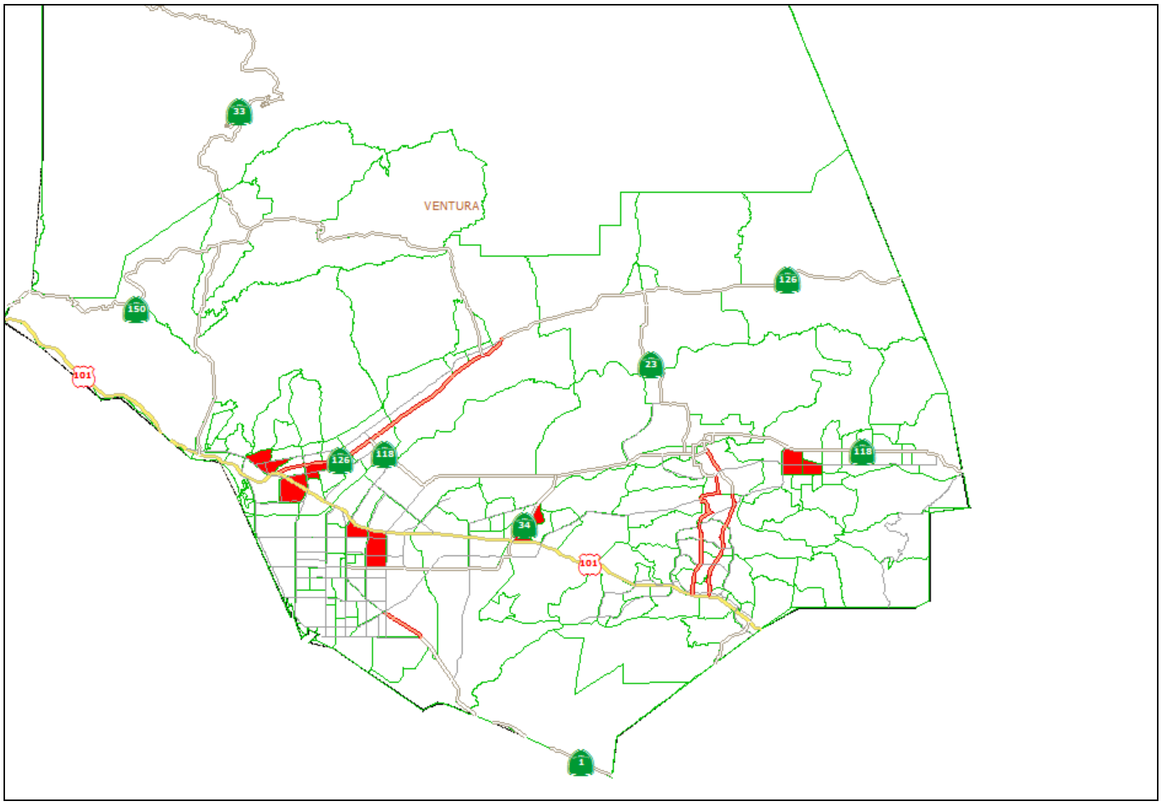


Figure 45: Ventura County 2020, GS, d=10, D=10,000

(none with d=20, D=20,000)
(none by LISA)

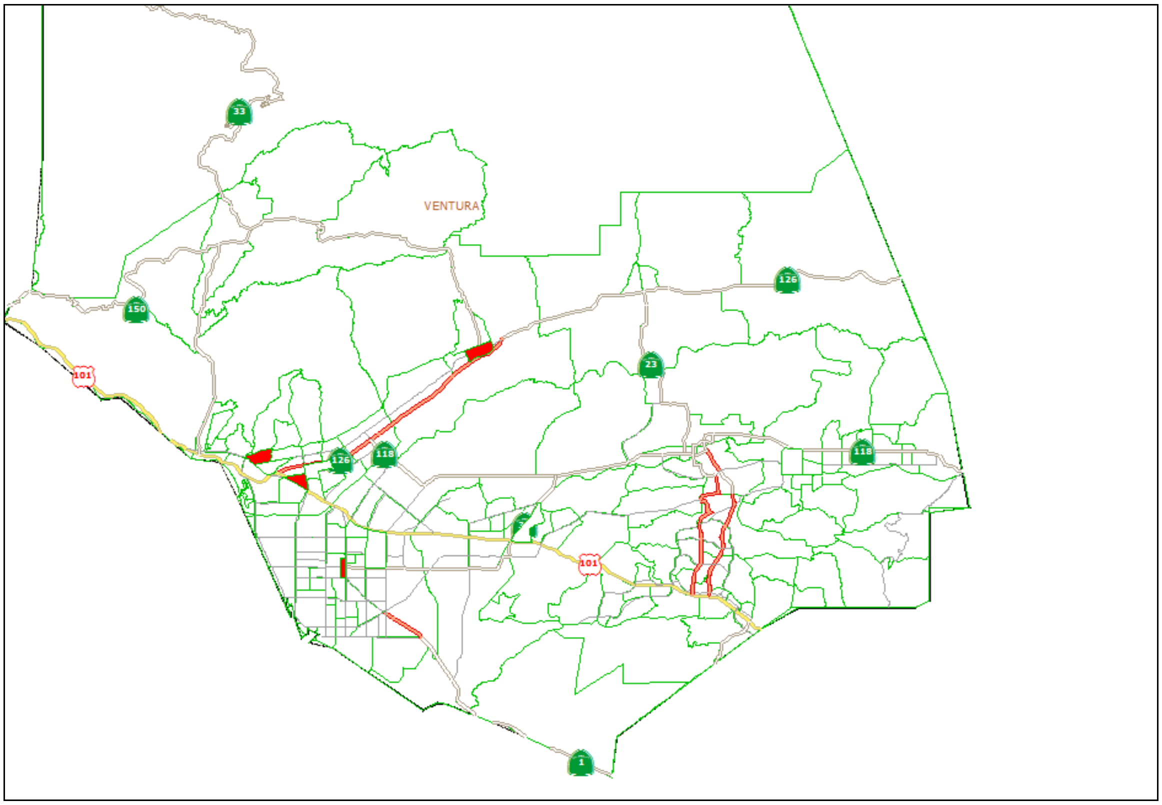


Figure 46: Ventura County 2020, Non-parametric

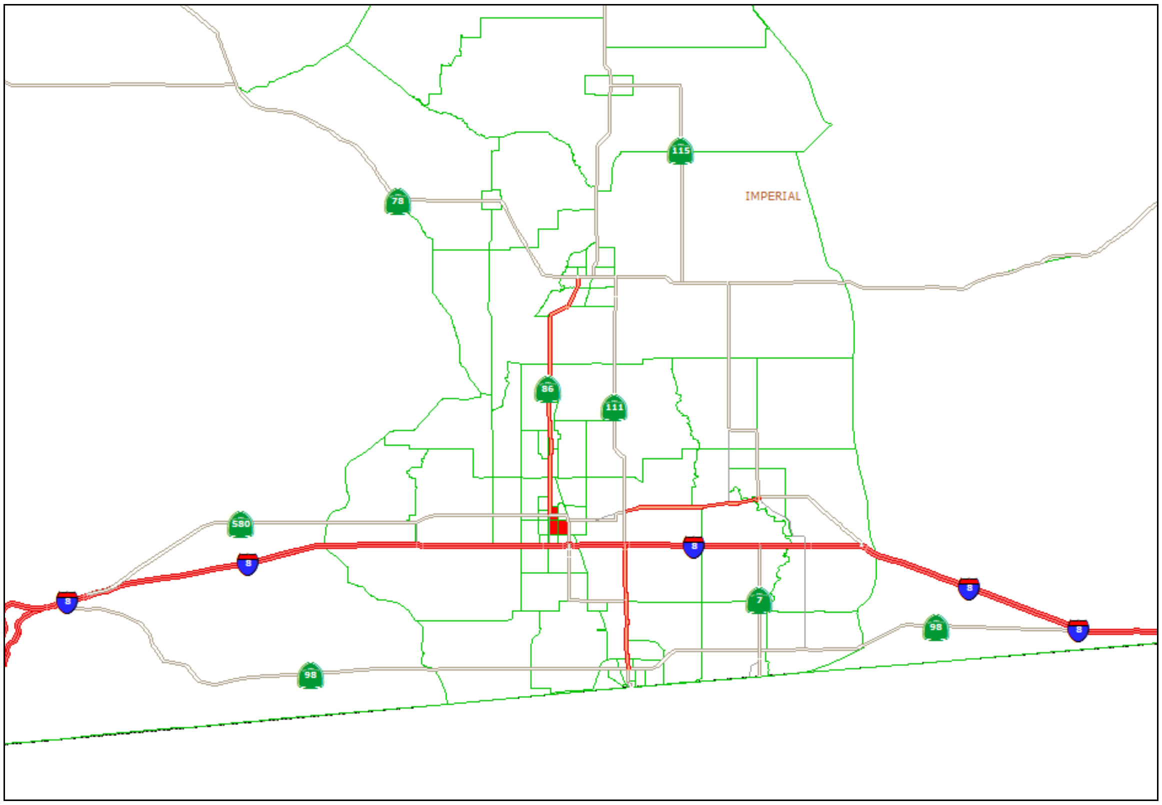


Figure 47: Imperial County 2035, GS, d=10, D=10,000

(none with d=20, D=20,000)
(none by LISA)

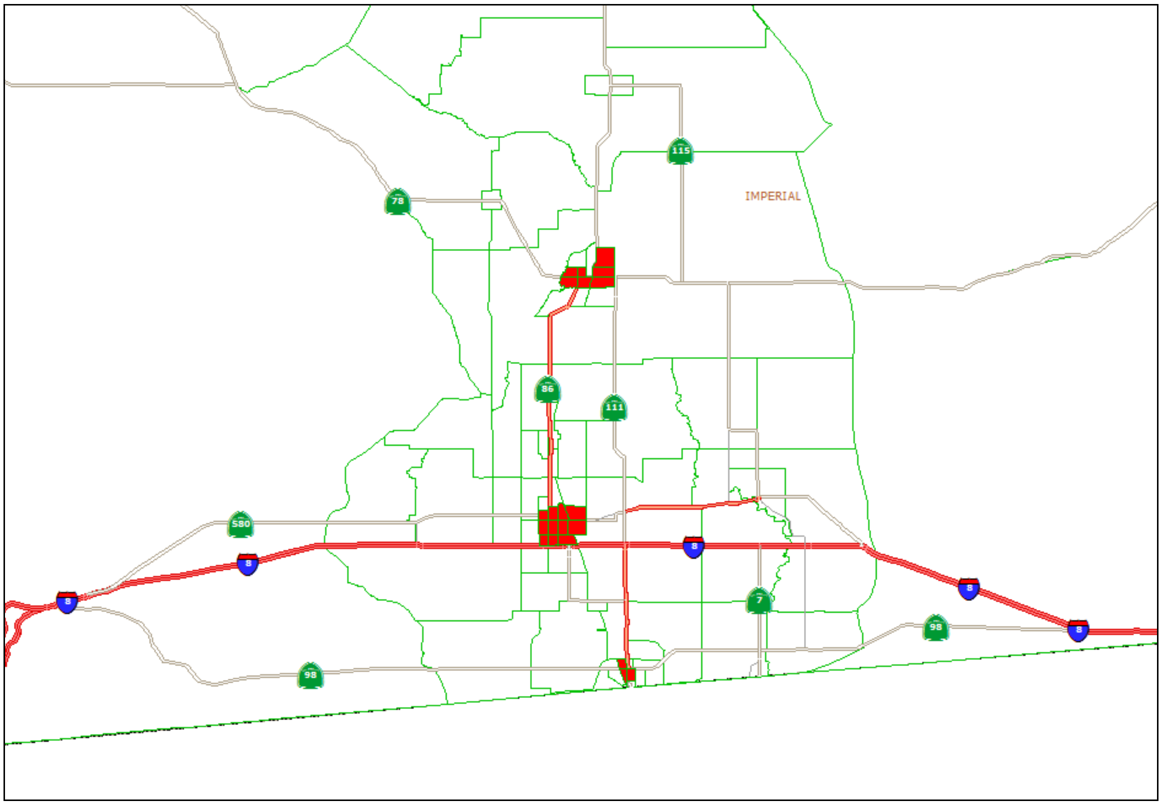


Figure 48: Imperial County 2035, Non-parametric

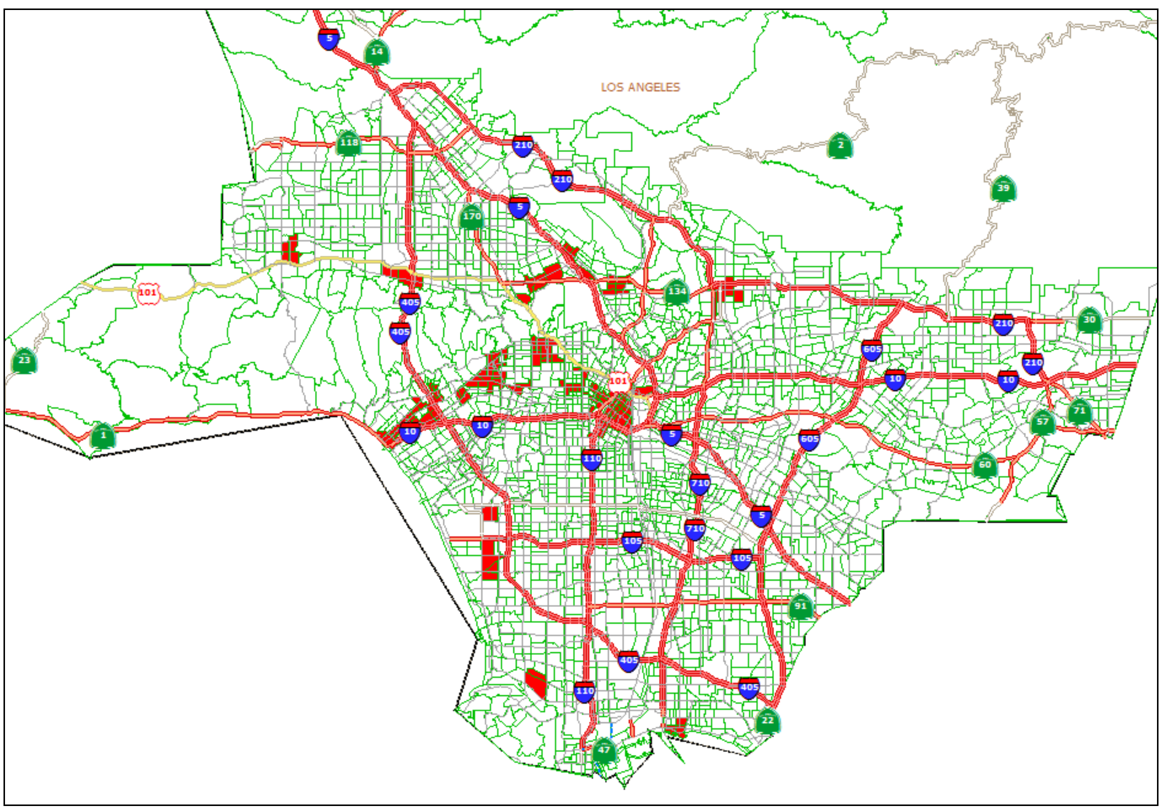


Figure 49: Los Angeles County 2035, GS, d=20, D=20,000

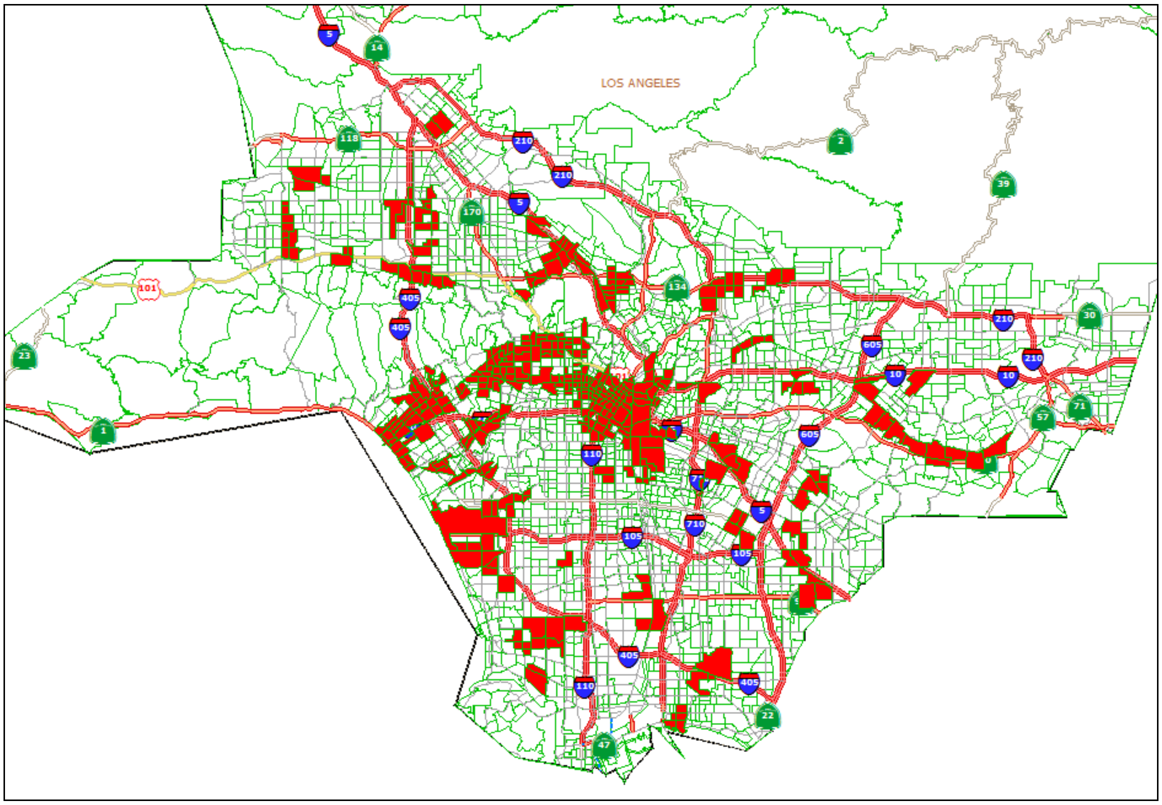


Figure 50: Los Angeles County 2035, d=10, D=10,000

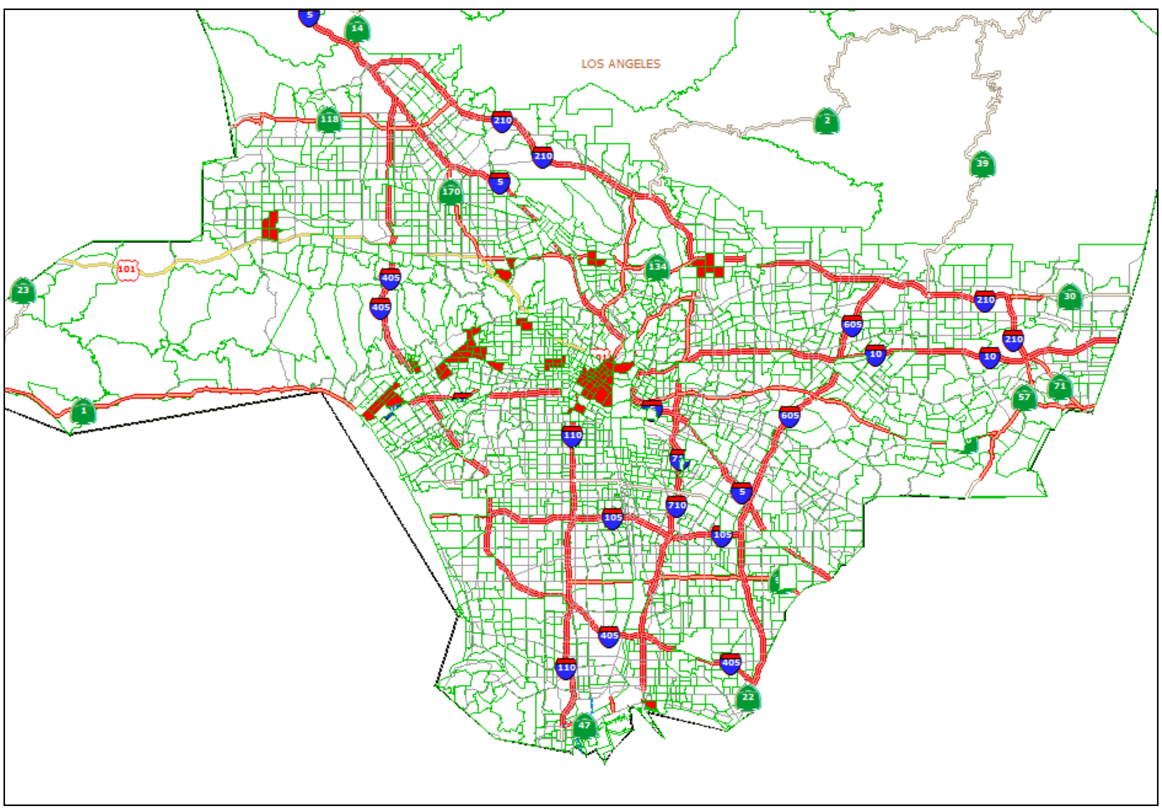


Figure 51: Los Angeles County 2035, LISA

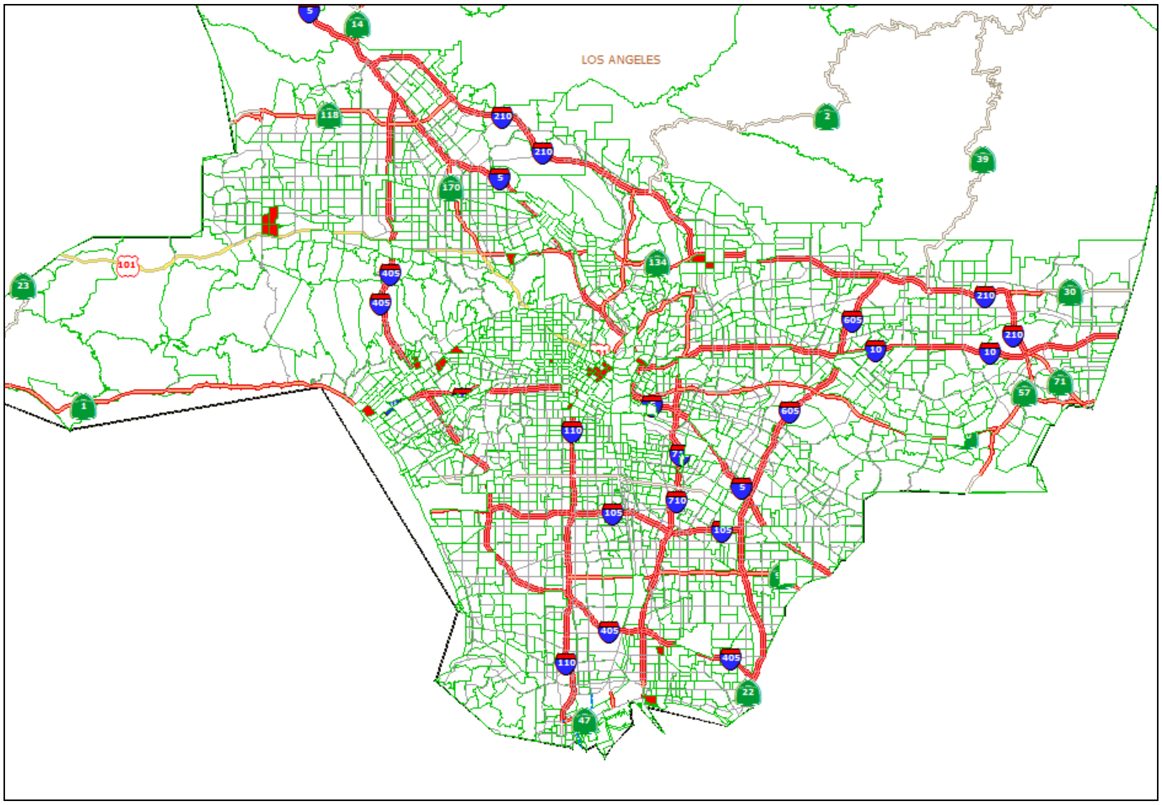


Figure 52: Los Angeles County 2035, Non-parametric

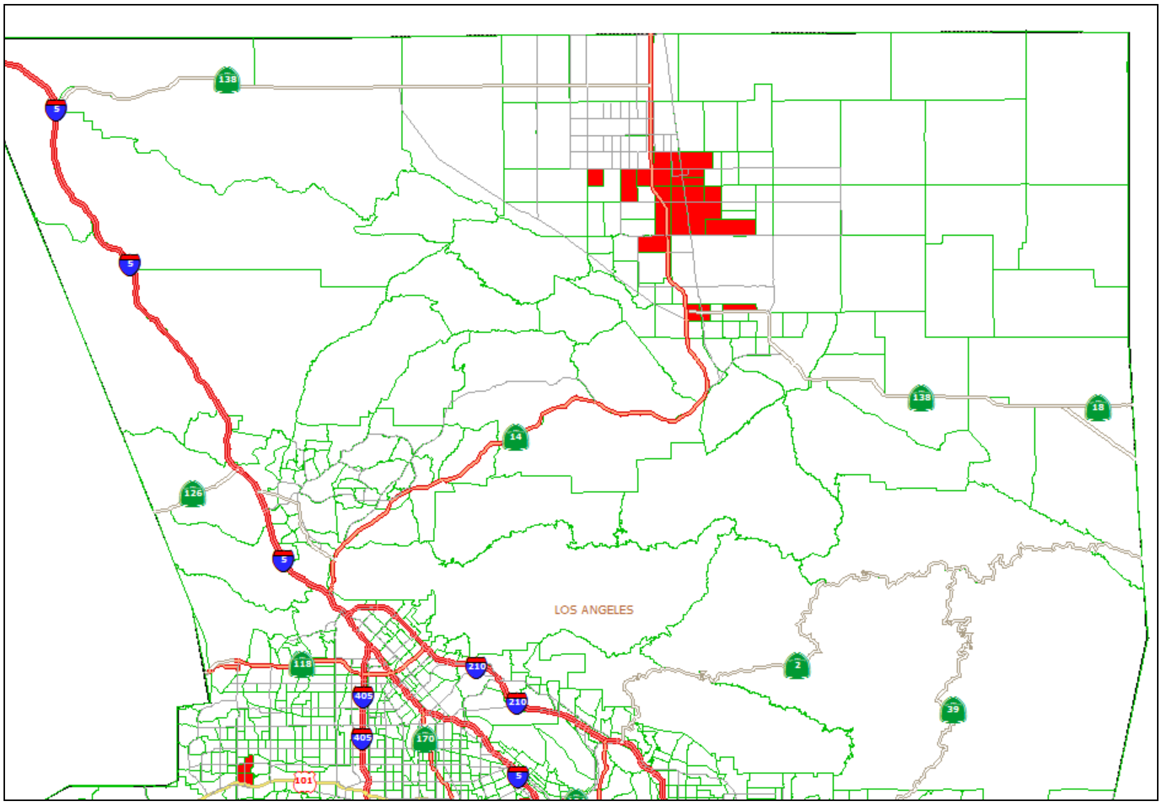


Figure 53: Los Angeles County 2035, Northern Portion, Non-parametric

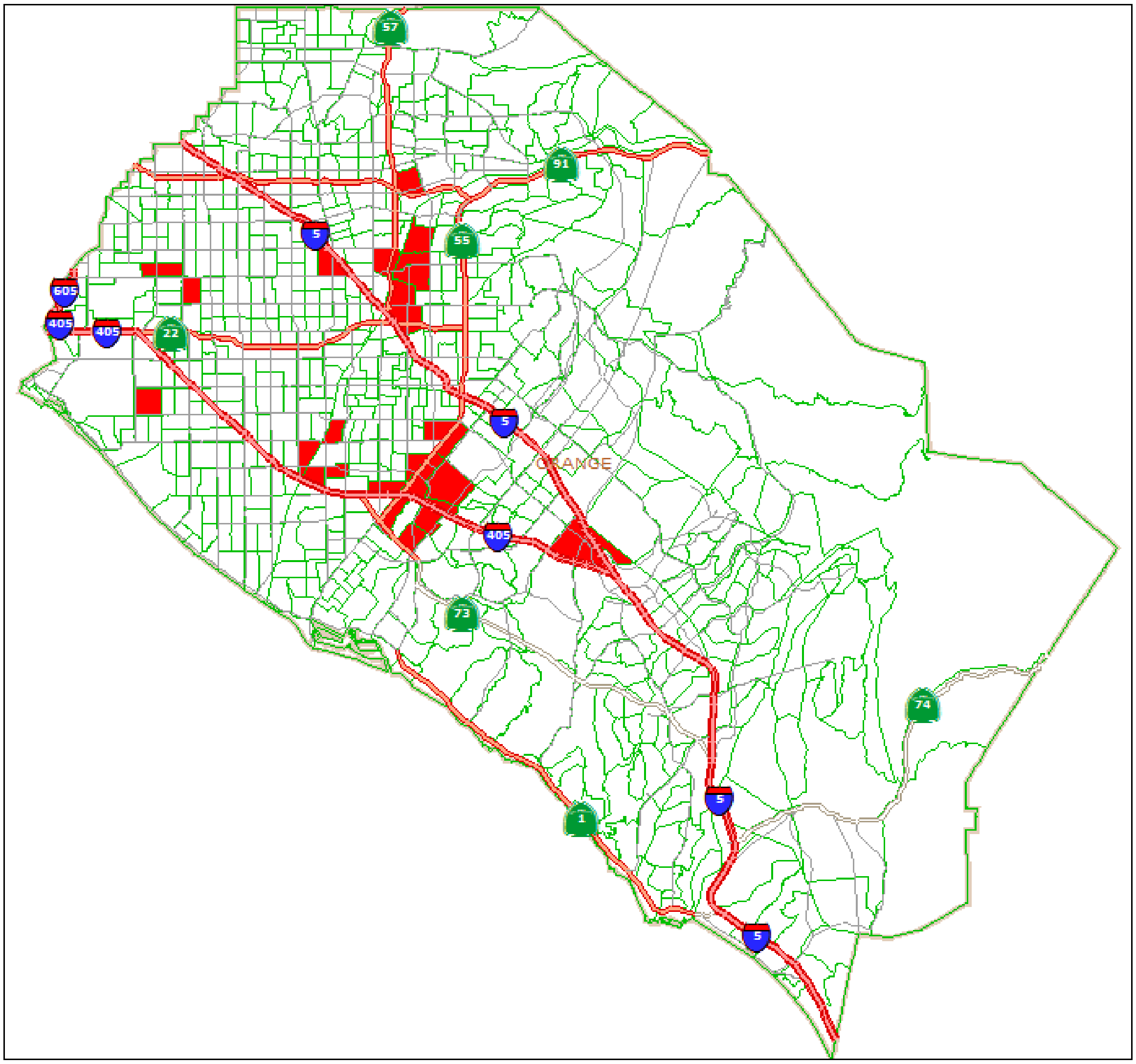


Figure 54: Orange County 2035, GS, d=20, D=20,000

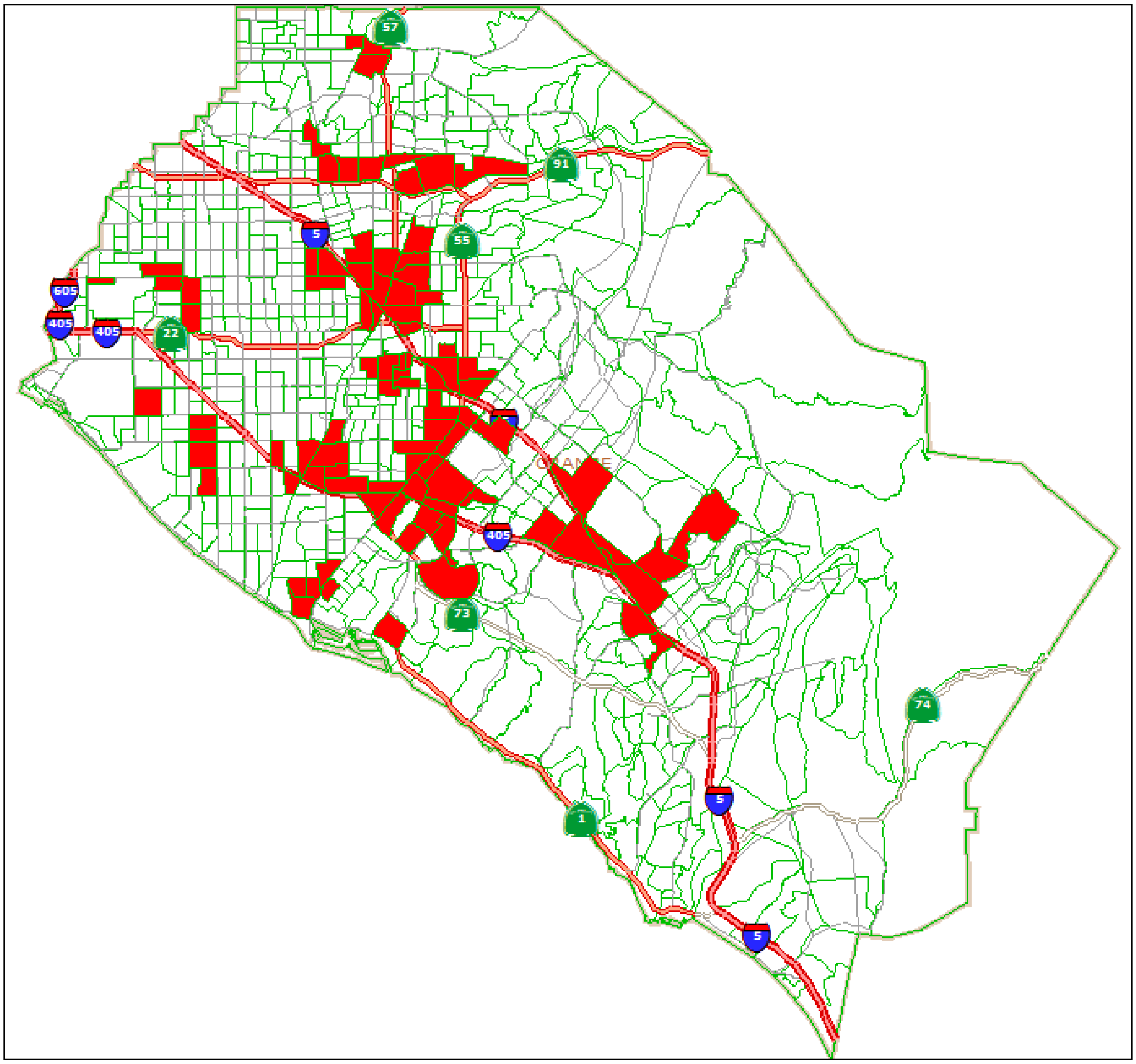


Figure 55: Orange County 2035, GS, d=10, D=10,000

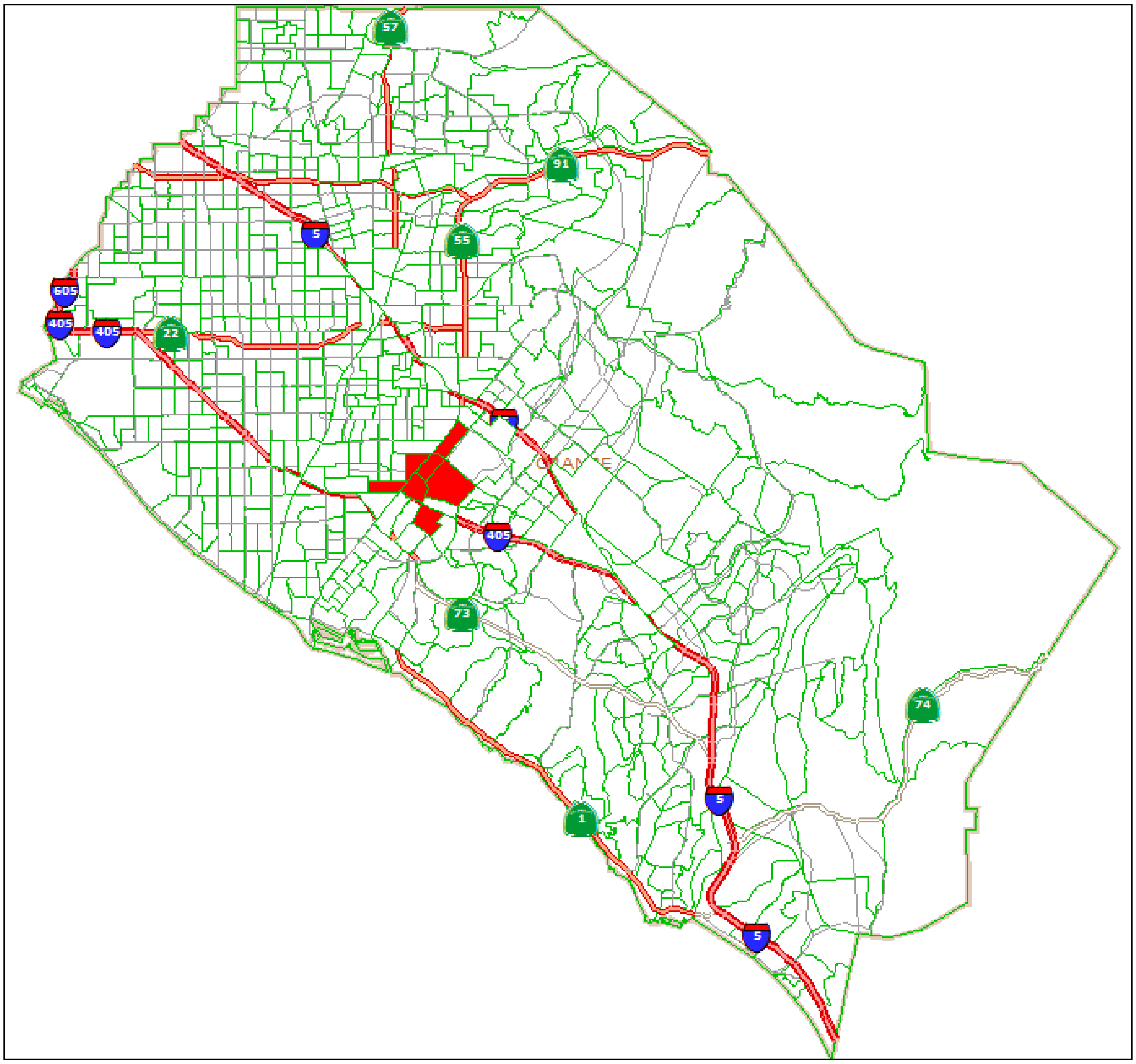


Figure 56: Orange County 2035, LISA

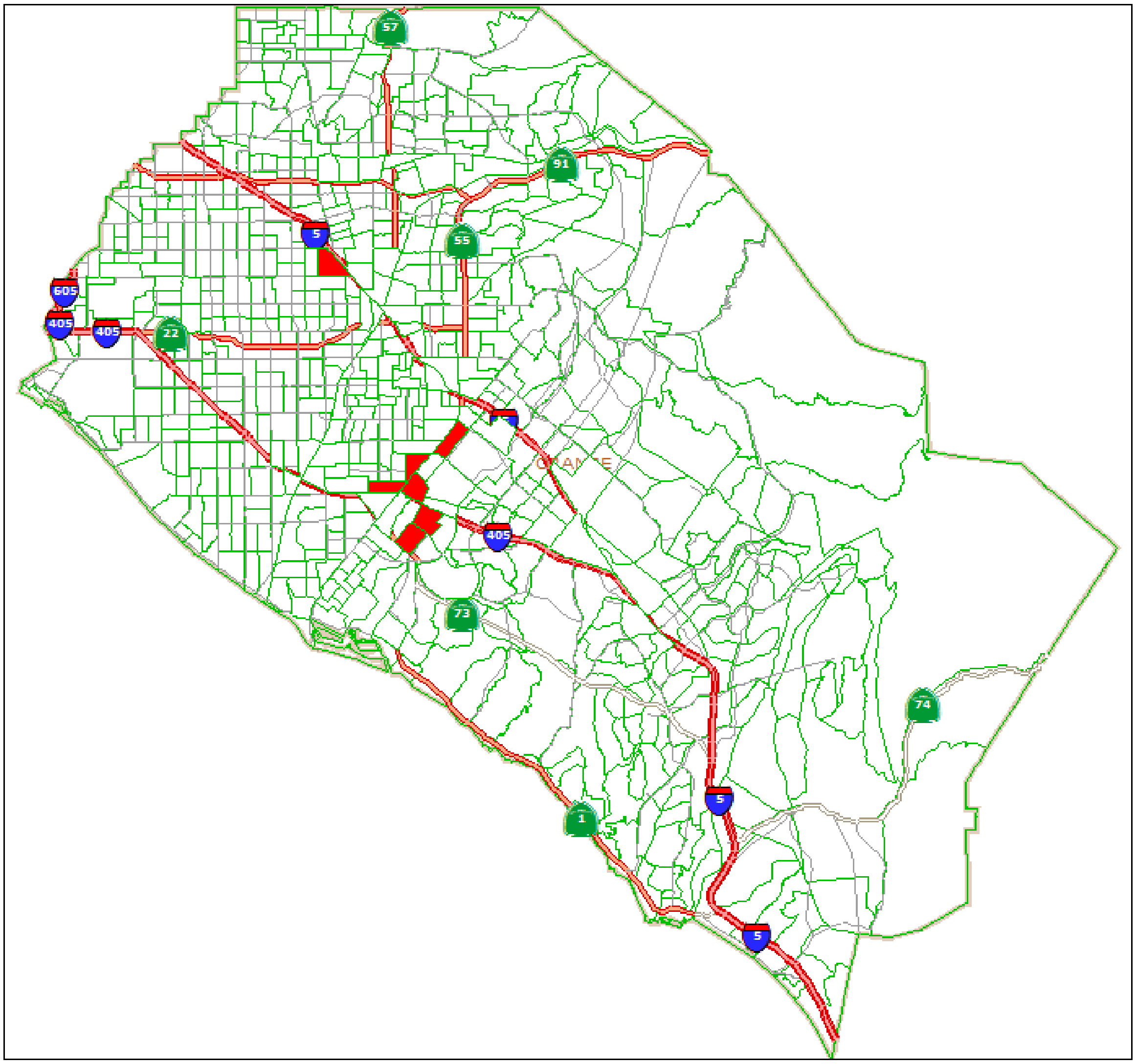


Figure 57: Orange County 2035, Non-parametric

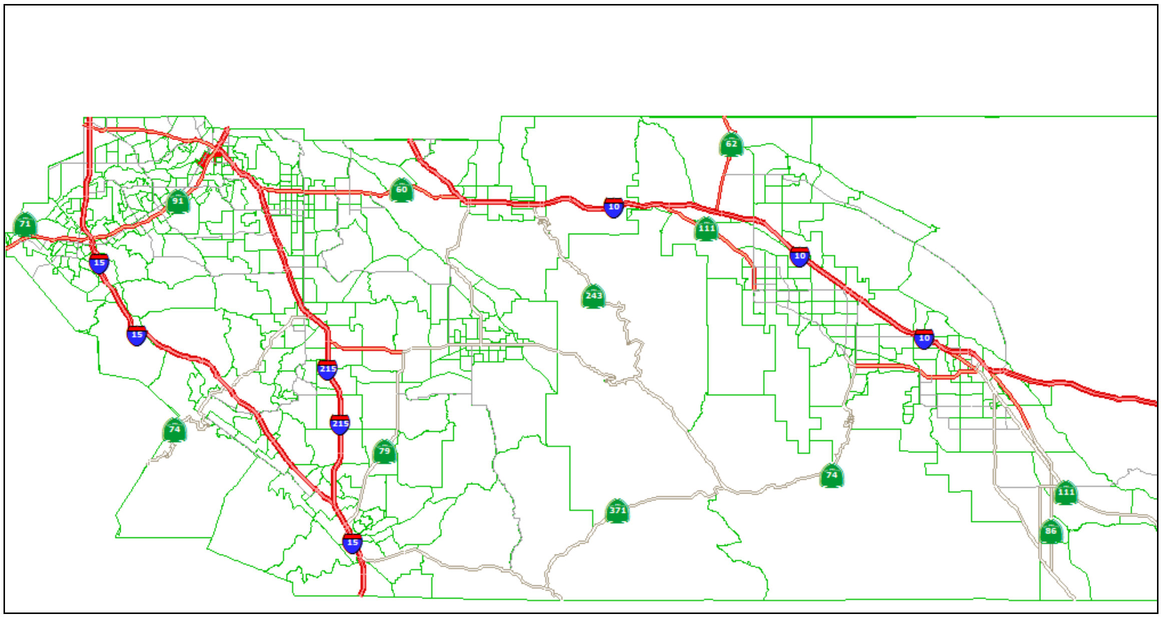


Figure 58: Riverside County 2035, GS, d=20, D=20,000

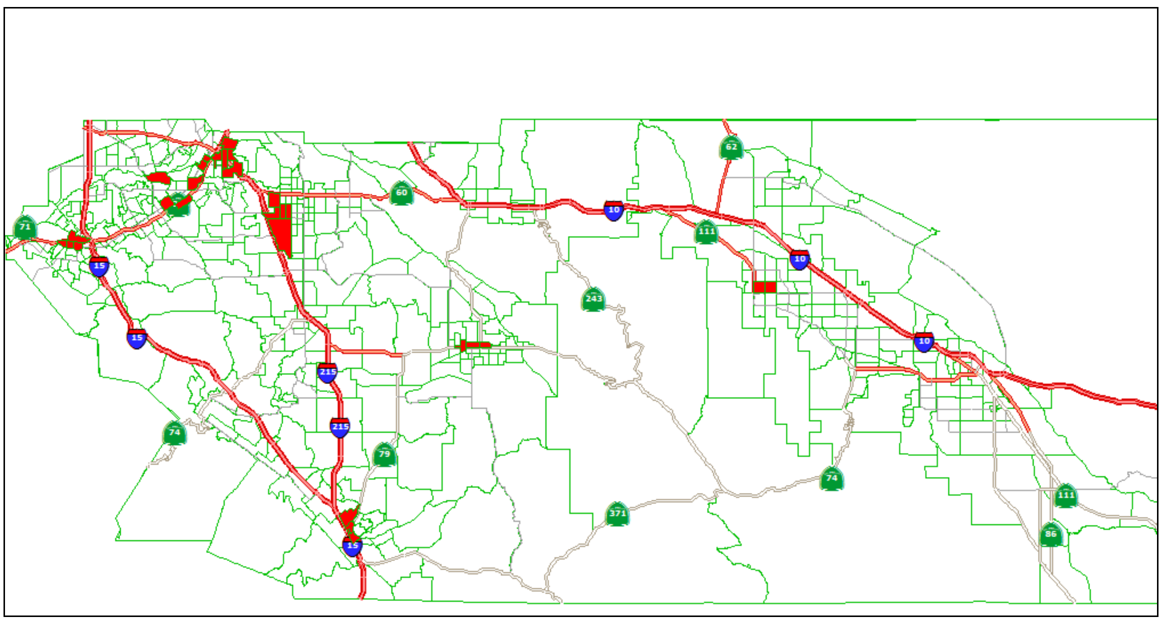


Figure 59: Riverside County 2035, GS, d=10, D=10,000

(none by LISA)

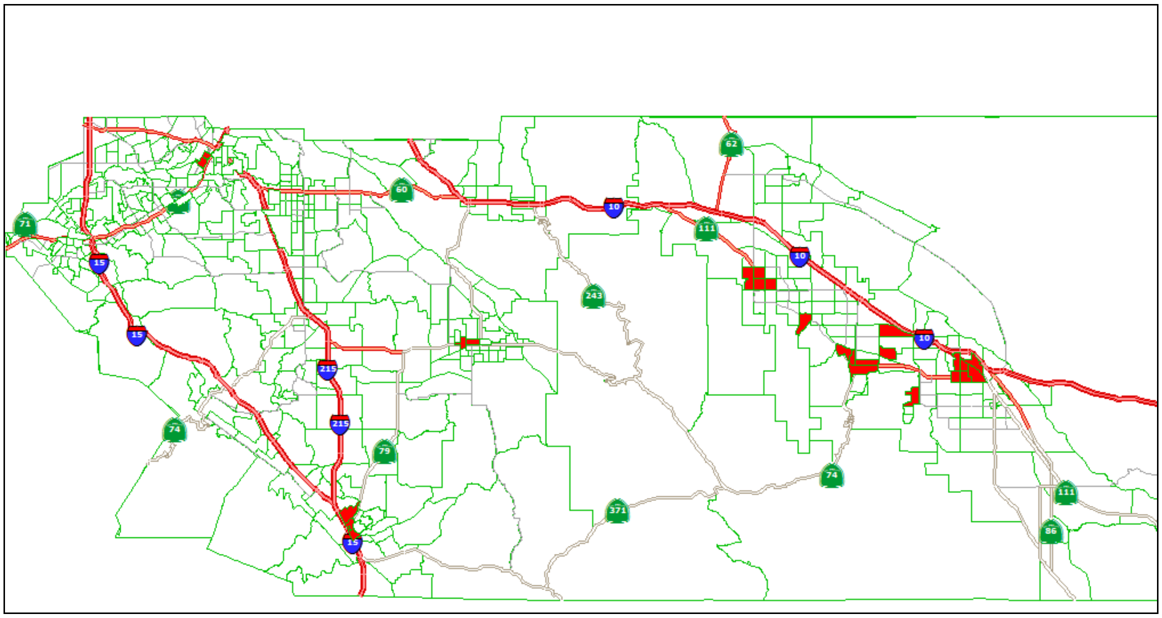


Figure 60: Riverside County 2035, Non-parametric

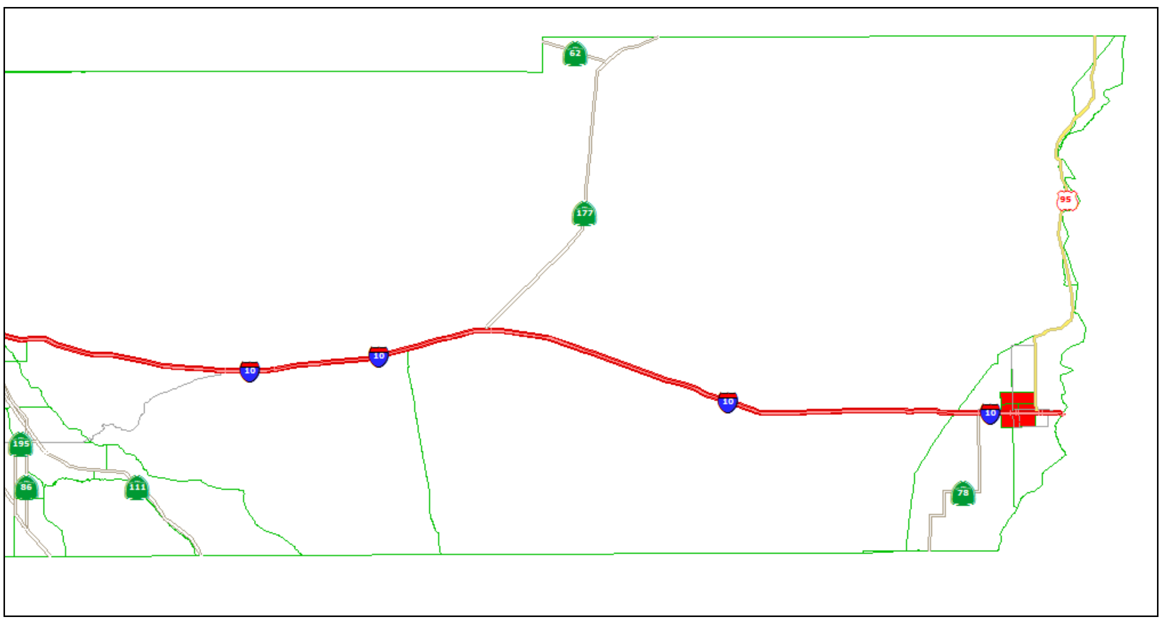


Figure 61: Riverside County 2035, Eastern Portion, Non-parametric

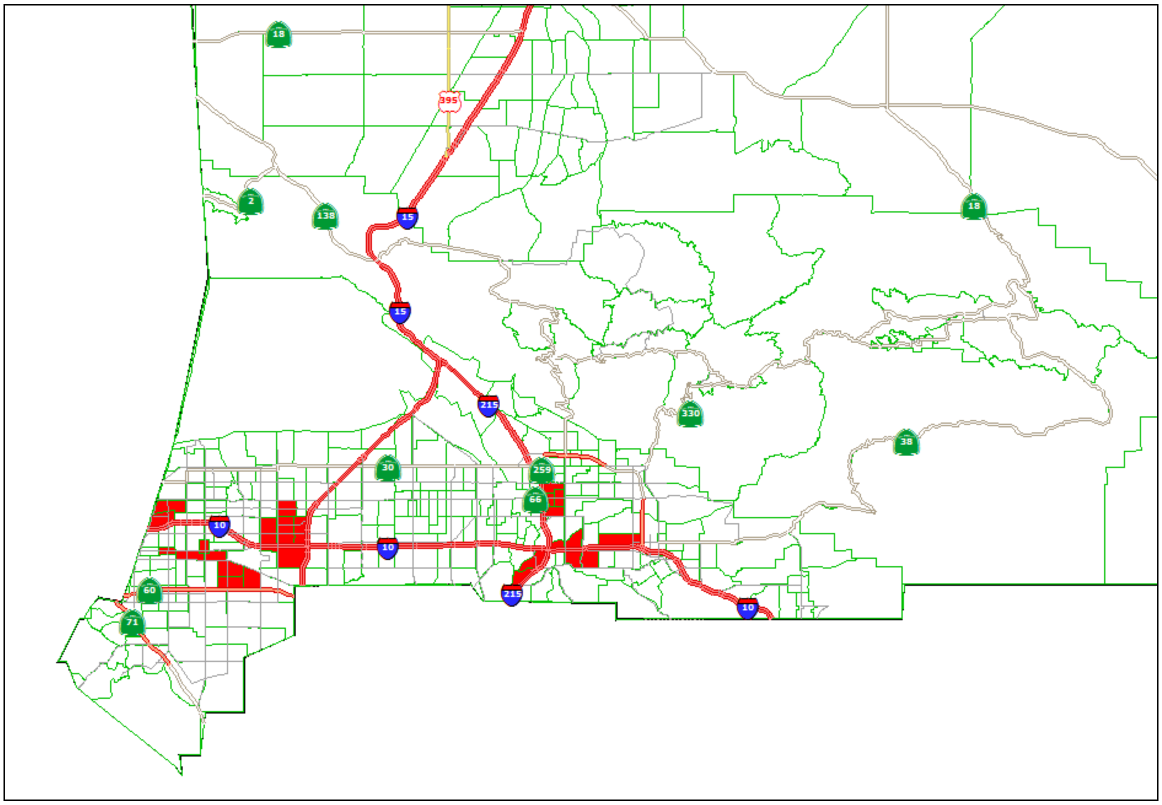


Figure 62: San Bernardino County 2035, GS, d=10, D=10,000

(none with d=20, D=20,000)
(none by LISA)

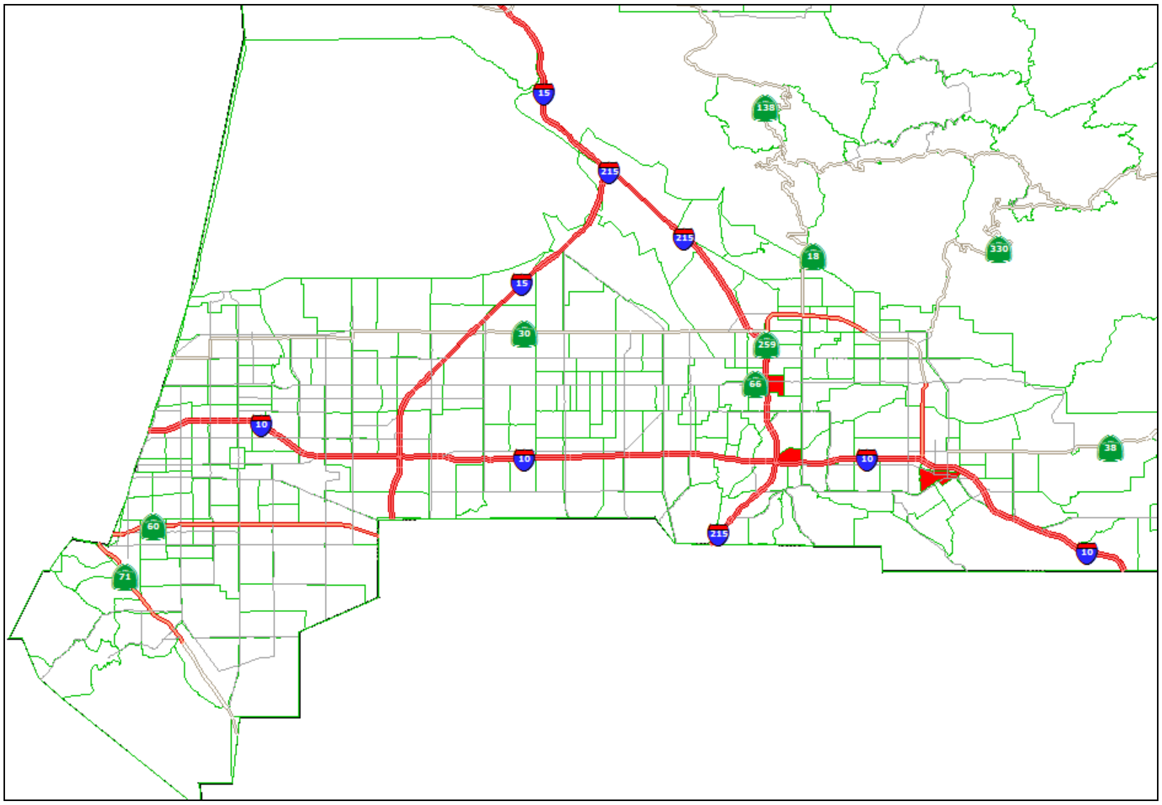


Figure 63: San Bernardino 2035, Southwestern Portion, Non-parametric

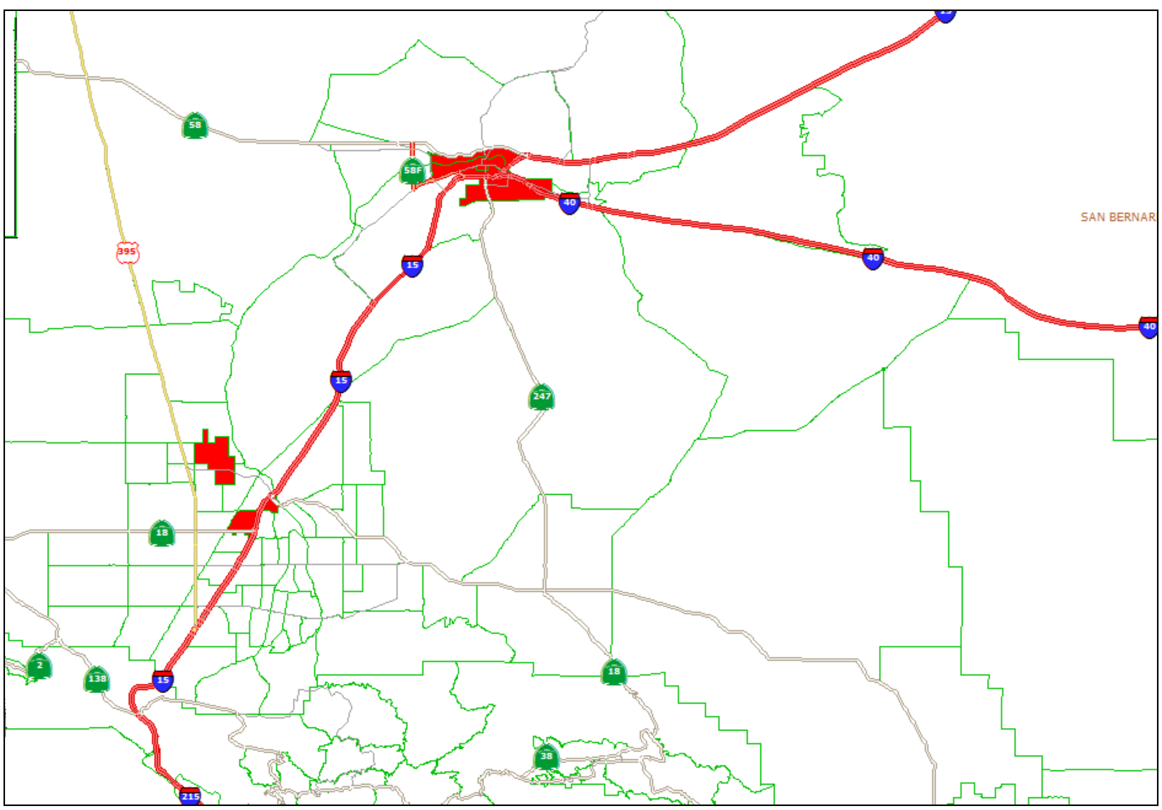


Figure 64: San Bernardino 2035, West-Central Portion, Non-parametric

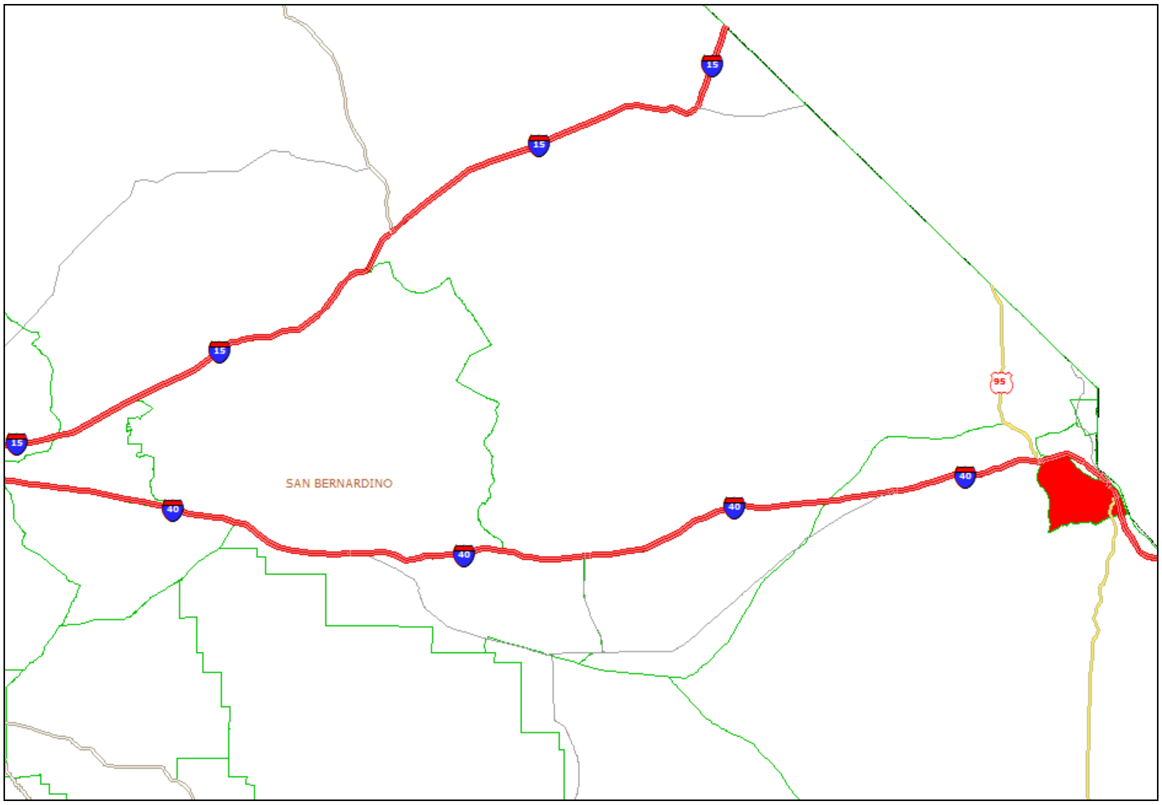


Figure 65: San Bernardino County 2035, Eastern Portion, Non-parametric

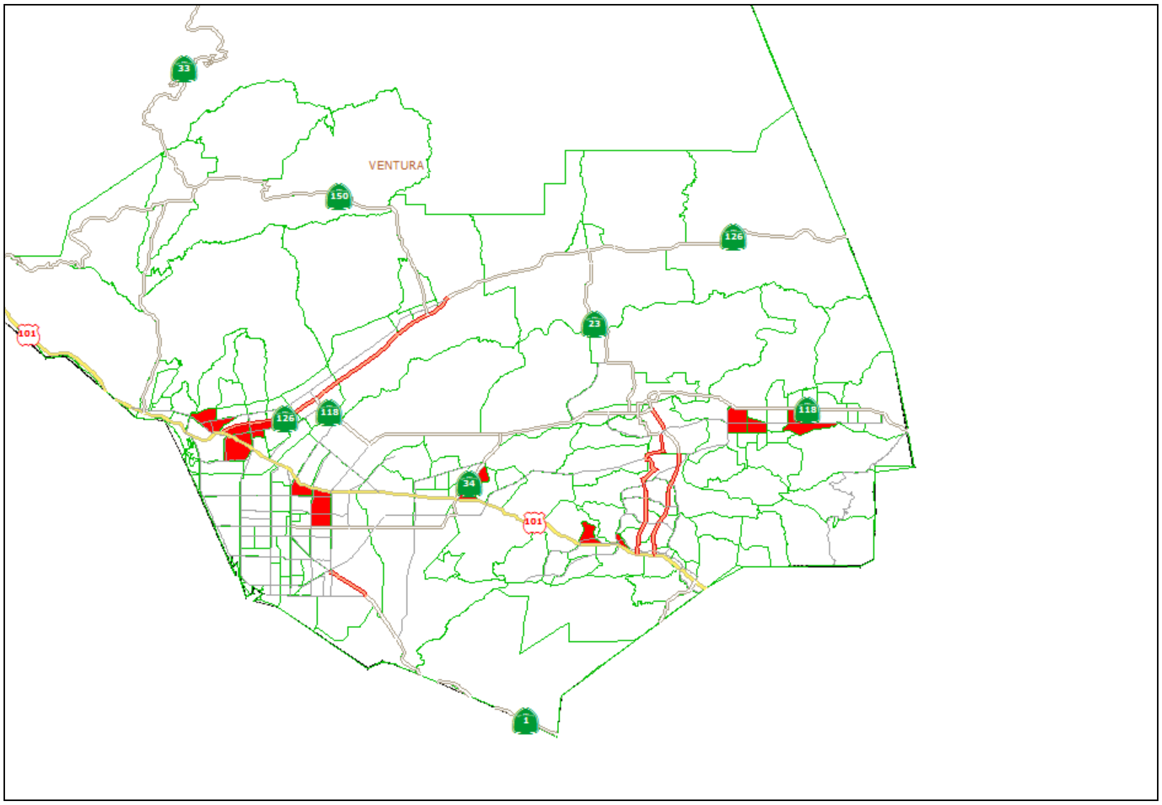


Figure 66: Ventura County 2035, GS, d=10, D=10,000

(none with d=20, D=20,000)
(none by LISA)

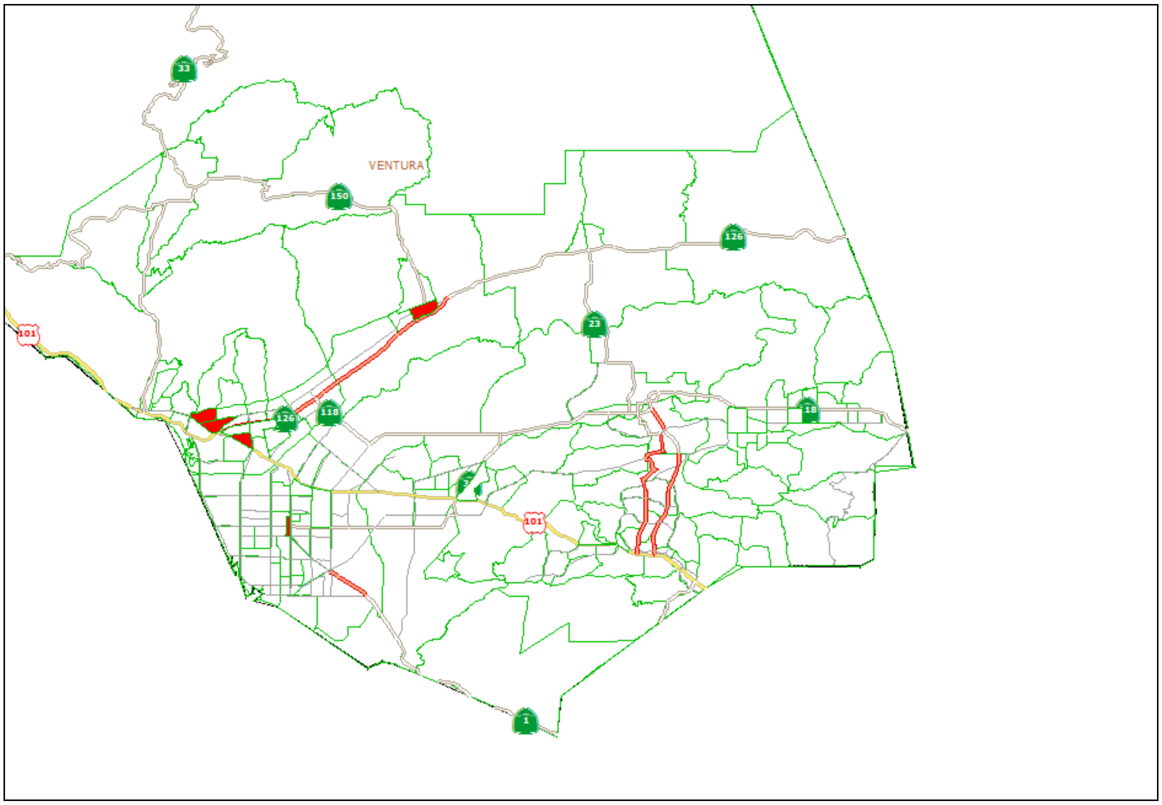


Figure 67: Ventura County 2035, Non-parametric

8. *Technical Appendix*

We have developed functions in the computer program R that allow the subcenter identification procedures to be implemented easily. One of the packages in R's library, "spdep", makes it possible to read shape files directly into R. Another free software program, GeoDa, is used to construct a spatial contiguity matrix for the shape file. The "locfit" package is used for the non-parametric identification procedure. These programs can be downloaded at the following sites;

http://www.r-project.org/ and http://geodacenter.asu.edu/software/downloads

We stored our R code in a computer directory named "\scag" and we stored the shape files in the subdirectory "/scag/maps". All of the data used in the analysis were stored in the TAZ shape file.

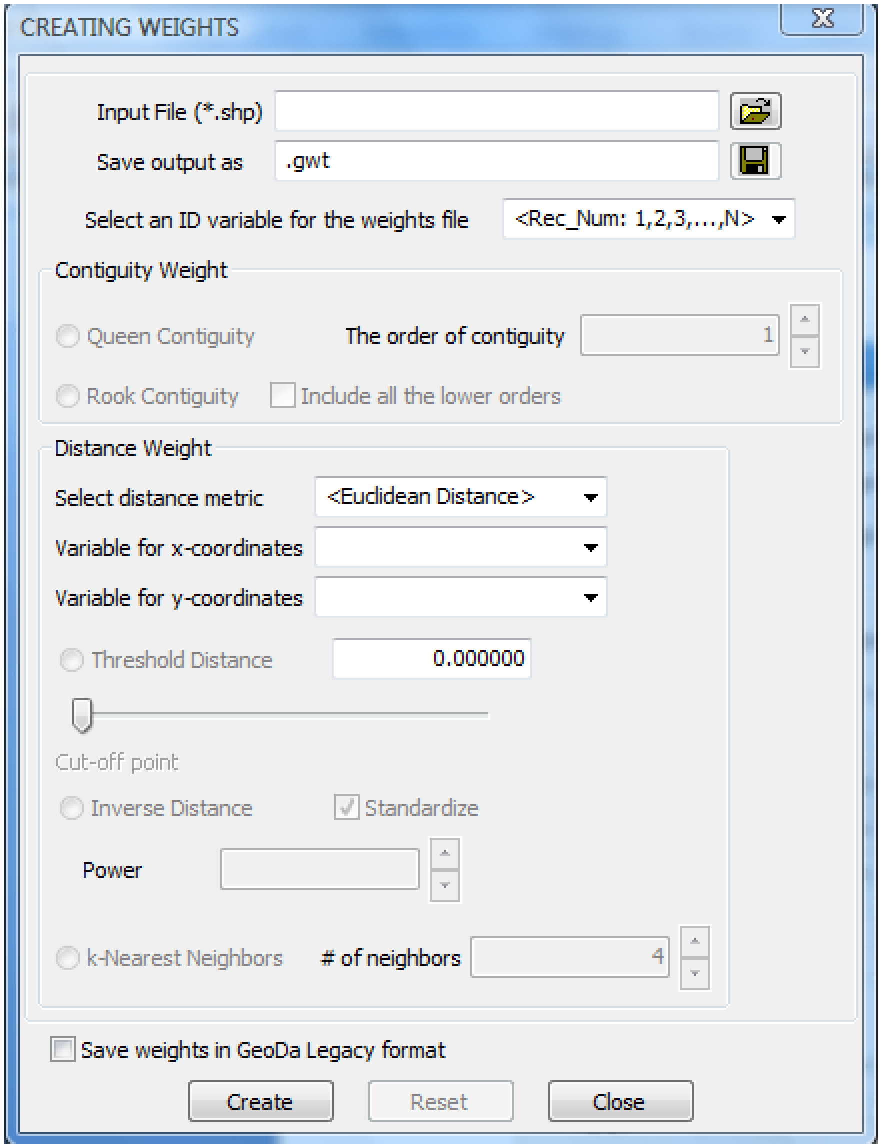
Before running the R programs, GeoDa must be used to set up a file which then will be used by R to generate the contiguity matrix. We used a "Queen" definition of Contiguity. The file used to generate the contiguity matrix in R is set up as follows using GeoDa:

1. Launch GeoDa.

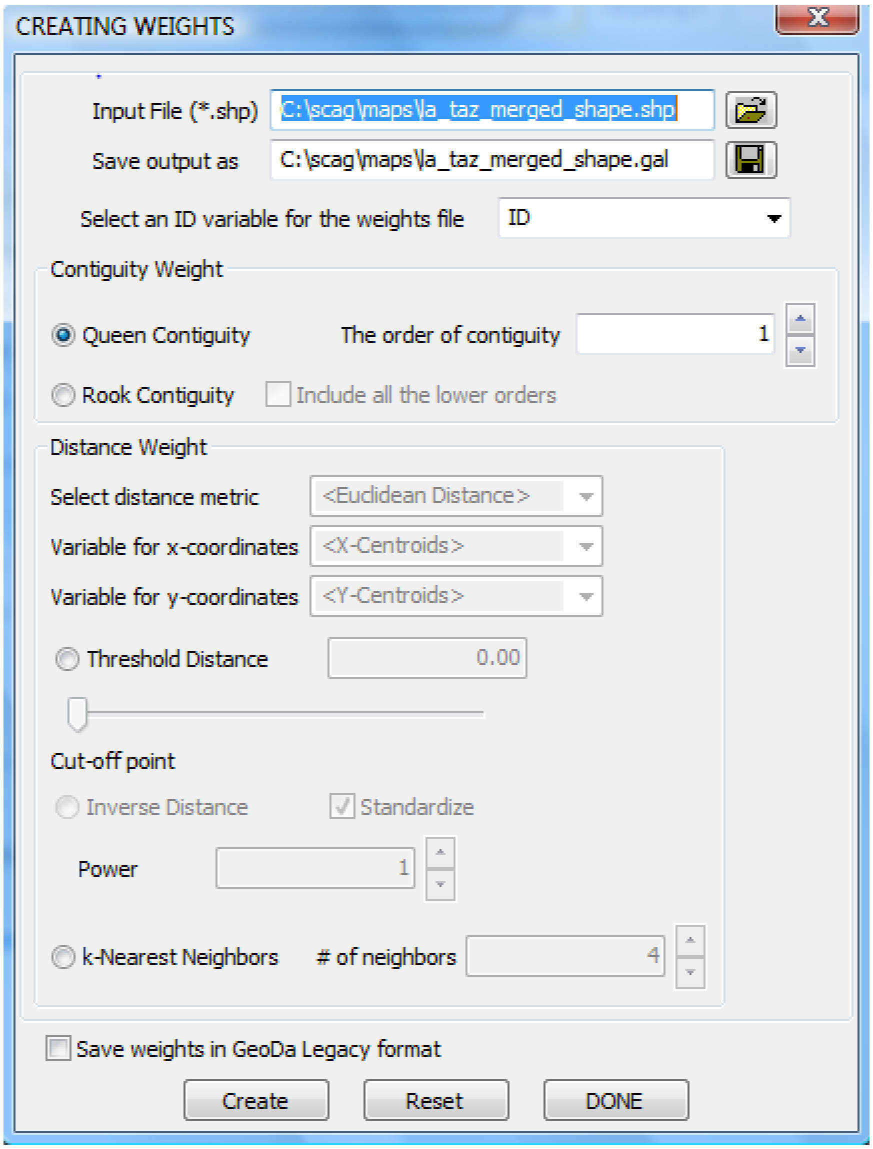
2. Choose "tools" from the main toolbar:

3. Under "tools", choose "weights" and then "create". The menu shown below will appear. The shape file can be entered directly or by browsing. By default, the "gal file" output will be stored in the same directory as the shape file, with the "gal" extension added.

4. Enter the ID variable (we simply used "ID", which is a variable in the TAZ file). Choose the "Queen Contiguity" option under "Contiguity Weight," and then click on "Create". The gal file will then be stored automatically in the shape file directory, where it will be ready for input into R.


CREATING WEIGHTS
Input File (*.shp)
Save output as
.gwt
Select an ID variable for the weights file
<Rec_Num: 1,2,3,...,N>
Contiguity Weight
Queen Contiguity
The order of contiguity
1
Rook Contiguity
Include all the lower orders
Distance Weight
Select distance metric
<Euclidean Distance>
Variable for x-coordinates
Variable for y-coordinates
Threshold Distance
0.000000
Cut-off point
Inverse Distance
Standardize
Power
k-Nearest Neighbors
# of neighbors
4
Save weights in GeoDa Legacy format
Create
Reset
Close

Same screen with options chosen:



In addition to storing the shape files and creating the contiguity matrix, the program calls three functions that can either be included directly in the R program or stored in a computer. These functions were prepared by Daniel McMillen for this project. The functions – "sub_gs", "sub_np", and "sub_lisa" – carry out the subcenter identification procedures. We shall assume that these functions are stored in a computer directory called "Rfunctions." Each function includes a set of parameters (such as the density cutoffs for the Giuliano-Small procedure and the window size for the non-parametric approach) that can be varied by the user. Each function also includes default values that correspond to the parameter values that we found most useful for the Los Angeles region. The functions can be read into the main program using R's "source" command, or they can be included directly in the main program.

After these preliminaries, the following R program is all that is required to identify subcenters using the GS, LISA, and NP procedures. We shall discuss the commands by section.

```
# Section 1 – read data files
library(spdep)
shp_file<- readShapePoly("/scag/maps/la_taz_merged_shape.shp",ID="ID")
gal_file <- read.gal("/scag/maps/la_taz_merged_shape.gal",region.id=shp_file$ID)

# Section 2 – finding geographic coordinates and re-labeling data

lmat <- coordinates(shp_file)
longitude <- lmat[,1]
latitude <- lmat[,2]
employment <- shp_file$EMP03
area <- shp_file$ACRE
id <- shp_file$ID
taz2k <- shp_file$TAZ2K

# Section 3 – Data manipulation and definition of contiguity matrix

empdens <- employment/area
empdens <- ifelse(is.na(empdens),0,empdens)
summary(empdens)
lndens <- ifelse(empdens>0,log(empdens),log(.5/area))
wmat <- nb2mat(gal_file,style="B")
```

```
# Section 4 – Giuliano-Small method
source("/rfunctions/sub_gs.r")
gssub <- sub_gs(file=gal_file,dens=empdens,emp=employment,
        mind=10,totemp=10000,wmat=wmat)
subobs_gs10 <- gssub$subobs
gssub <- sub_gs(file=gal_file,dens=empdens,emp=employment,
        mind=20,totemp=20000,wmat=wmat)
subobs_gs20 <- gssub$subobs

# Section 5 – LISA method
source("/rfunctions/sub_lisa.r")
lisasub <- sub_lisa(file=gal_file,dens=empdens,pval=.01)
subobs_lisa <- lisasub$subobs

# Section 6:  Non-parametric Method
source("/rfunctions/sub_np.r")
npsub <- sub_np(lndens,lat=latitude,long=longitude,window=.5,pval=.10)
subobs_np <- npsub$subobs

# Section 7:  Write results to file
submat <- cbind(id,taz2k,longitude,latitude,
        subobs_gs10,subobs_gs20,subobs_lisa,subobs_np)
write.csv(submat,"//scag/maps/subcenter.csv")
```

Discussion:

Section1. The first line of Section 1 provides access to the spdep package (which first must be downloaded from one of the CRAN mirrors at http://www.r-project.org/). The second line reads the TAZ shape file into R. The third line reads the gal file, which was set up previously using GeoDa. Note that the ID names in lines 2 and 3 must correspond to the variable labeled as ID when constructing the gal file.

Section 2. The first line of section 2 finds the geographic coordinates associated with the shape file. The first column of the matrix "lmat" stores the longitudes and the second column has the latitudes. The variable "employment" is defined to be the shape file variable "emp03"; this variable can be changed to work with data for 2020 or 2035. It is not strictly necessary to define the variables "area", "id", and "taz2k" since they are defined already as shp_file$EMP03, shp_file$ID, shp_file$TAZ2K. Providing the new labels to these variables simply make them easier to refer to in subsequent parts of the program.

Section 3. This section begins by setting up the employment density variable, "empdens." For a small number of TAZ's, data for employment or area may be missing; the second line of Section 3 re-defines these observations as having employment density of zero. The third line defines the natural logarithm of employment density, "lndens".

For observations with a density of zero, the log-density is arbitrarily set to ln(.5/area). The results are not sensitive to this choice. We chose to include the observations rather than omit them because it is easier to work with full data sets in R, and it also is easier subsequently to merge the program's results back into the GIS program. The last line of Section 3 defines the contiguity matrix using the spdep package command "nb2mat."

Section 4. This section identifies subcenters using the Giuliano-Small procedure. The first line reads in our R function "sub_gs.r". Alternatively, the full program could be included directly in the main R program at this point. The first four arguments in the function simply feed the required data into the function. These arguments include the gal file (which was given the name gal_file in the third line of section 1), the employment density variable ("dens=empdens"), the variable representing total employment in a TAZ ("emp=employment), and the contiguity matrix ("wmat=wmat"). The names on the right hand side of the equal signs can be varied if desired; the names on the left hand side are required by the function. The last two arguments are the cutoff values for the GS procedure: "mind" is the minimum employment density and "totemp" is the total employment cutoff. (These values are referred to as "d" and "D" in the main report. The variable gssub_obs equals one if the TAZ observation is indicated to be part of a subcenter; it equals zero otherwise.

Section 5. This section identifies subcenters using the LISA method. The arguments for the function include the gal file, the variable for which the local Moran's I statistics are to be calculated, and the p-value used to define whether the statistics are statistically significant. The p-value can be varied; the default is the 1% value used here. Note also that the subcenter identification could be based on the natural logarithm of employment density rather than the level, although we found that lndens led to less reasonable results in this application. The output from the function, the variable lisasub$subobs, is a variable that equals one if the TAZ observations is indicated to be part of a subcenter; it equals zero otherwise.

Section 6. This section identifies subcenters using the non-parametric approach. The non-parametric regressions are estimated using the package "locfit", which must be installed from one of the CRAN mirrors. The first three arguments for the function are the dependent variable (the natural log of employment density, "lndens") and the two explanatory variables (latitude and longitude). The last two arguments in the functions are the window size for the local regression and the p-value used to determine whether the residuals from the spatial smooth are statistically significant. The defaults are a 50% window size and a p-value of 10%. The output from the function, "npsub$subos", is a variable that equals one if the TAZ observation is indicated to be part of a subcenter; it equals zero otherwise.

Section 7. The final section of the program writes the matrix "submat" to an excel csv file. This file can then be read into a GIS program to map the subcenter locations.

The R functions used to identify subcenters are "sub_gs.r", "sub_lisa.r", and "sub_np.r". The code for these functions is presented next along with technical details of the estimation procedures.

Giuliano-Small Procedure

```
sub_gs <- function(file,dens,emp,wmat=0,mind=10,totemp=10000) {
  library(spdep)
  if (sum(wmat)==0) {wmat <- nb2mat(file,style="B")}
  dens <- ifelse(is.na(dens),0,dens)
  obs <- seq(1:length(dens))
  densobs <- obs[dens>mind]
  wmat <- wmat[dens>mind,dens>mind]
  n = nrow(wmat)
  amat <- matrix(0,nrow=n,ncol=n)
  amat[row(amat)==col(amat)] <- 1
  bmat <- wmat
  wmat1 <- wmat
  newnum = sum(bmat)
  cnt = 1
  while (newnum>0) {
    amat <- amat+bmat
    wmat2 <- wmat1%*%wmat
    bmat <- ifelse(wmat2>0&amat==0,1,0)
    wmat1 <- wmat2
    newnum = sum(bmat)
    cnt = cnt+1
  }
  emat <- emp[dens>mind]
  tmat <- amat%*%emat
  obsmat <- densobs[tmat>totemp]

  subemp <- array(0,dim=length(dens))
  subemp[obsmat] <- tmat[tmat>totemp]
  subobs <- ifelse(subemp>0,1,0)

  tab <- tabulate(factor(subemp))
  numsub = sum(tab>0)-1

  cat("Number of Subcenters = ",numsub,"\n")
  cat("Total Employment and Number of Tracts in each Subcenter","\n")
  print(table(subemp))
  out <- list(subemp,subobs)
  names(out) <- c("subemp","subobs")
  return(out)
}
```

The sub_gs function uses an algorithm proposed by McMillen (2003) to identify subcenters using the Giuliano-Small procedure. McMillen's procedure takes advantage of the properties of contiguity matrices to produce an efficient algorithm that can be applied to large data sets. The base contiguity matrix, "wmat," is an nxn matrix, where n is the number of observations in the data set. A typical entry of the matrix, w_{ij}, equals one if observation i is contiguous to observation j and $w_{ij} = 0$ otherwise. Working with such a large matrix can be demanding of both computer time and memory. Fortunately, it is only necessary to work with the subset of the observations that meet the minimum employment density threshold. The command wmat <- wmat[dens>mind,dens>mind] selects this subset of observations from the full contiguity matrix.

One interpretation of a contiguity matrix is that each non-zero entry shows the number of paths that can be taken to get from one observation to another. For example, if observation 1 is next to observation 2, then there is (trivially) one path that can be taken to get from observation 1 to observation 2 in one step. The entries of a contiguity matrix's square – another nxn matrix – show the number of paths that can be taken to get from observation i to observation j in two steps. Succeeding powers of a contiguity matrix – w2 = wmat*wmat, w3 = w2*wmat, and so on through w(k) = w(k-1)*wmat – show how many paths can be taken to get from observation i to observation j in k steps.

McMillen's (2003) algorithm takes advantage of this property of contiguity matrices to define subcenters. The square of the contiguity matrix for the observations with dens>mind shows which TAZ's are next to other TAZ's that also meet the minimum employment density threshold. Subsequent multiplication indicates whether additional TAZ's meeting the density threshold are contiguous to these TAZ's. The process continues until no new observations are added to the cluster. The final step of the algorithm is to find the total amount of employment in each cluster of contiguous TAZ's. If this total exceeds the total employment threshold ("totemp"), the cluster is defined to be a subcenter.

A simple example of the algorithm is presented in McMillen (2003). An important feature of the approach is that it can detect long strings of contiguous tracts paralleling freeway routes as easily as it can detect circular concentrations of employment. The key requirement is that the TAZ's must be contiguous; two subcenters separated by one TAZ that fails to meet the minimum density threshold will be indicated to be separate subcenters when they may look like a single subcenter on a map. Moreover, the two separate areas may each fail to qualify as a subcenter if neither meets the total employment threshold.

The primary arguments for the procedure are:

- file: the gal file defined by GeoDa. If this file is specified without providing the argument wmat, the procedure will use spdep to generate the contiguity matrix.
- dens: the employment density variable
- emp: the variable showing total employment in a TAZ

- wmat: the contiguity matrix. Although this matrix does not have to be provided to the function if the “file” is provided in the function’s first argument, it will save a great deal of time to specify wmat if the sub_gs function is called several times using different density cutoffs. If “file” is specified but not “wmat”, the program will set up the contiguity matrix every time the function is called. Since setting up the contiguity matrix can be quite time consuming for large data sets, it is preferable to calculate the contiguity matrix before calling sub_gs, providing it to the function using the “wmat” argument.
- mind: the minimum density cutoff. The default is 10 employees per acre.
- totemp: the total employment density cutoff. The default is 10,000.

The LISA Procedure

```
sub_lisa <- function(file,dens,pval=.01,wlist="NO") {
  if (wlist[1]=="NO") {wlist <- nb2listw(file)}
  lisaout <- localmoran(dens,wlist)
  subobs <- ifelse(lisaout[,5]<=pval,1,0)
  cat("Number of tracts identified as part of subcenters:  ",sum(subobs),"\n")
  out <- list(subobs)
  names(out) <- c("subobs")
  return(out)
}
```

The LISA procedure was used by Baumont, Ertur, and LeGallo (2004) to identify subcenters in Dijon, France. The procedure uses a local version of the Moran’s I statistic to find tracts with levels of employment density that are positively spatially autocorrelated with the levels in contiguous tracts. The statistic for observation i is:

$$I_i = \frac{(x_i - \bar{x})}{\sum_{k=1}^{n}(x_k - \bar{x})^2 / (n-1)} \sum_{j=1}^{n} w_{ij}(x_j - \bar{x})$$

where x_i represents employment density in observation i, x_k is employment density for observation j, $\bar{x}$ is the average value of employment density across all observations, and w_{ij} is a variable indicating whether observation i is contiguous to observation j. High values of I_i indicate either that (1) observation i has higher than average employment density while neighbors also have higher than average density, or (2) observation i and its neighbors have lower than average densities. Only those observations for which case (1) holds qualify as subcenters.

The primary argument for this function is “pval”, which indicates the significance level for the I statistics. The default is 1%. The full set of arguments is:

- file: the gal file
- dens: the employment density variable
- pval: the significance level for the local Moran's I
- wlist: If this term is not specified, the program sets up the list of contiguous tracts using the gal file specified with "file." Since calculating the list is somewhat time-consuming, it is also possible to provide the contiguity list directly to the function using the wlist option. If specified, wlist should be a "neighbors list" specified using the spdep package's command nb2listw. The default is simply wlist="NO", meaning that no wlist is provided and that the program should calculate the neighbors list using the gal file specified using "file".

The Non-parametric Approach

```
sub_np <- function(lndens,lat,long,window=.5,pval=.10) {
  library(locfit)
  fit <- locfit(lndens~lp(lat,long,nn=window,deg=1),kern="tcub",ev=dat(cv=F))
  mat <- predict(fit,se.fit=T,band="pred")
  yhat <- mat$fit
  sehat <- mat$se.fit
  upper <- yhat - qnorm(pval/2)*sehat
  subobs <- ifelse(lndens>upper,1,0)

  cat("Number of tracts identified as part of subcenters:  ",sum(subobs),"\n")
  out <- list(subobs)
  names(out) <- c("subobs")
  return(out)
}
```

The arguments for the function include the variable representing the natural logarithm of employment density ("lndens"), the geographic coordinates ("lat" and "long"), a window size ("window", to be discussed below), and the significance level for the log-employment density residuals ("pval"). The window size and significance levels can be varied readily.

The non-parametric approach is based on a procedure proposed in McMillen (2001). The idea is that a subcenter comprises a set of tracts that have higher employment density than would be expected given the broad spatial trend of the data. Unlike the Giuliano and Small approach, which is based on absolute density levels, the non-parametric approach has lower minimum density thresholds in remote areas than in dense areas near the traditional city center.

The first step of the procedure is a locally weighted regression of the natural logarithm of employment density (lndens) on the geographic coordinates (lat and long). A separate regression is estimated for a set of target locations. The regressions place more weight on tracts that are close to the target locations.

The sub_np function takes advantage of an excellent R package, locfit, which computes locally weighted regression extremely quickly. This package can be downloaded from one of the CRAN mirrors. The locfit procedure differs somewhat from McMillen's approach in that it uses simple Euclidean measures of distance rather than geographic distance. Thus, if the geographic coordinates for a target location are given by (lat, long), the distance of tract i from this location is defined as:

$$d_i = \sqrt{(lat_i - lat)^2 / s_{lat}^2 + (long_i - long)^2 / s_{long}^2}$$

The geographic coordinates are normalized by dividing them by their sample standard deviations, s_{lat}^2 and s_{long}^2, to keep the distance measures from being affected by scale differences. Although this distance measure could easily be based on geographic rather than Euclidean distance, the results are not sensitive to the choice.

Having defined this distance measure, the weight applied to observation i when estimating the regression for the target location is defined as follows:

$$w_i = \left(1 - \left(\frac{d_i}{d_{\max}}\right)^3\right)^3 I(d_i > d_{\max})$$

In this equation, I is an indicator function that equals one if the condition is true. This function, which is known as a "tri-cube" function, places a weight of one on an observation located at the target point (so that $d_i = 0$). The weights decline to 0 at a distance given by the parameter d_{max}, and the weights also equal 0 for any distance greater than d_{max}. The value of d_{max} is determined by the choice of "window size". By default, sub_np sets d_{max} at the median value of d_i. This choices means that a "window" of the 50% of the observations that are closest to the target location is used to estimate the regression at the target point. This window can be varied by changing the parameter "window." A value of .75 for window means that 75% of the observations are used to estimate the regressions, while a value of 0.10 means that only the closest 10% of the observations are used. Larger window sizes produce smoother functions; smaller values produce greater local variability in the estimates.

After defining this weighting function, the predicted value of lndens at the target location is simply the prediction from the weighted least squares regression of the natural logarithm of employment density on the geographic coordinates:

$$\hat{y} = x\left(\sum_{i=1}^{n} w_i x_i' x_i\right)^{-1} \sum_{i=1}^{n} w_i x_i y_i$$

Where y represents the natural log of employment density and $x_i = (1 \quad lat_i \quad long_i)$. Thus, the approach simply uses a locally weighted regression to approximate an unknown function at a set of target locations. McMillen (2001) estimated locally weighted regressions using each data point, in turn, as a target location. The locfit package takes advantage of a sophisticated interpolation procedure to estimate the function at only a subset of the full data set. The estimates are then interpolated to each data point.

One might expect that a procedure using n regressions to predict densities at n data points would exhaust all degrees of freedom, or would at least produce highly variable results. However, the locally weighted regression approach actually imposes a sufficient degree of smoothness on the estimates that only a modest number of degrees of freedom are used. The estimator is linear in the sense that, for the full set of n observations, the estimator can be written as $\hat{Y} = LY$, where L is an nxn matrix. The degrees of freedom used in estimation can be calculated as the trace of the L matrix (Loader, 1999). By this measure, only about 12 degrees of freedom are used in calculating the locally weighted regression estimates for the Los Angeles TAZ data set using the default values of sub_np.

After calculating the predicted log-employment density for every point in the data set, the NP approach identifies candidate subcenter sites as the observations with significant residuals. The variance of the predicted value of y at a given target location can be estimated using a standard robust covariance matrix:

$$\sigma^2 \left(\sum_{i=1}^{n} z_i{}' \sqrt{w_i} z_i \right)^{-1} \left(\sum_{i=1}^{n} z_i{}' w_i z_i \right) \left(\sum_{i=1}^{n} z_i{}' \sqrt{w_i} z_i \right)^{-1}$$

where $z_i = (1 \quad lat_i - lat \quad long_i - long)$. The variance of the predicted value at the target location is simply the leading diagonal of this 3x3 matrix. McMillen (2001) used a kernel regression procedure to estimate values for σ^2 that vary across target points. Here we use a global estimate of σ^2 : $\hat{\sigma}^2 = \frac{1}{n - tr(L) - tr(L'L)} \sum_{i=1}^{n} e_i^2$, where e_i is the residual for observation i. This estimate is the default value in locfit (Loader, 1999). Letting v_i denote the leading diagonal element of the covariance matrix, the confidence interval for the prediction at observation i is $\hat{y}_i \pm c_{pval} \sqrt{1 + v_i}$, where c_{pval} is the critical value from the normal distribution function. (For example, $c_{pval} = 1.96$ when pval = 0.05.) The set of *candidate* subcenter locations is the set of TAZ's with value of y_i that fall above this confidence interval.

A final set of subcenter locations can be defined by identifying which candidate tracts provide statistically significant explanatory power for estimated employment density functions (McMillen, 2001), or a variant of Giuliano and Small's procedure could be used (McMillen, 2003). In the former approach, measures of proximity to each subcenter supplement distance to the traditional city center as explanatory variables for employment density functions. In the latter approach, a subcenter is a set of contiguous tracts with significant residuals from the locally weighted regression estimates that together exceed a threshold for total employment.

References

C. Baumont, C. Ertur, and J. LeGallo, Spatial analysis of employment and population density: The case of the agglomeration of Dijon 1999," *Geographical Analysis*, **36**, 146-176 (2004).

G. Giuliano and K. A. Small, Subcenters in the Los Angeles region, *Regional Science and Urban Economics*, **21**, 163–182 (1991).

C. Loader, *Local Regression and Likelihood*, Springer, New York (1999).

D. P. McMillen, Identifying subcenters using contiguity matrices, *Urban Studies*, **40**, 57-69 (2003).

D. P. McMillen, Nonparametric Employment Subcenter Identification, *Journal of Urban Economics*, **50**, 448-473 (2001).