Distribution Agreement

Signature:

______________________________ _______________

Elizabeth Ervin Date

Spatial distributions and patterns of Hantavirus Pulmonary Syndrome in the Western United States

By

Elizabeth Ervin
Master of Public Health

Global Environmental Health

Yang Liu, PhD
Committee Chair

Barbara Knust, DVM, MPH, DACVPM
Committee Member

Paige Tolbert, PhD
Committee Member

Spatial distributions and patterns of Hantavirus Pulmonary Syndrome in the Western United States

By

Elizabeth Ervin

B.S. Environmental Sciences
University of North Carolina at Chapel Hill
2008

Thesis Committee Chair: Yang Liu

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Environmental Health
2013

Abstract

Spatial distributions and patterns of Hantavirus Pulmonary Syndrome
in the Western United States
By Elizabeth Ervin

Hantavirus Pulmonary Syndrome (HPS) is a rare disease in the United States with a high mortality rate of around ~40%. Spread most commonly through deer mice, exposure to virus particles occurs with inhalation of aerosolized rodent feces or contact with infected mice. HPS has been nationally notifiable since 1995 and some 600 cases have since been confirmed. Though deer mice are prevalent across the North American continent, human cases of HPS are not consistently spread; most cases occur in the western United States.

Previous studies have analyzed environmental characteristics and host behaviors, but neither monitoring of rodent density nor identification of common local environmental features appear to be enough in assessing HPS risk for humans. This study analyzes the distribution of human cases across four western States (Washington, Oregon, California, and Nevada) assessing the resulting patterns through spatial data analysis. Hot-spots of increased HPS occurrence were identified and a focused environmental model was built from remote sensing data to examine the local variables that may be influencing HPS cases around the Sierra Nevada Mountains.

Conclusions from this study highlight the importance of spatial relationships between human cases and cases to the environment. The importance in analyzing Hantavirus ecological studies with the inclusion of spatial descriptive statistics in human disease is in developing of more sensitive and accurate models to predict areas of high infection risk for humans.

Spatial distributions and patterns of Hantavirus Pulmonary Syndrome in the Western United States

By

Elizabeth Ervin

B.S. Environmental Sciences
University of North Carolina at Chapel Hill
2008

Thesis Committee Chair: Yang Liu

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Environmental Health
2013

Acknowledgements

There are many people I want to thank. First and foremost are the folks in Viral Special Pathogens Branch, specifically Barbara Knust and Tim Filestra. Barbara's advice and mentorship were invaluable in this process and I am immensely grateful for her advice and support. Tim also was an incredible resource, always willing to discuss the finer details of statistics, keeping me grounded and helping me develop better methods. Special thanks also to Curtis Fritz (CA Dept. of Public Health) and Nicola Marsden Huag (WA Dept. of Health) for their help in completing our records

I'd also like to thank the many professors across Emory University for their support. First of all, my thesis advisor, Dr. Yang Liu, was wonderful in assuring me throughout this semester and keeping me calm. I am especially thankful to the team of instructors in my Spatial Analysis class for their patience in letting me monopolize their time week after week, asking too many questions.

To all my colleagues also completing their MPH, I sincerely thank you for your support and encouragements, and am constantly amazed and inspired by your awesomeness.

Finally, I want to thank my family. Their unwavering support in all that I pursue is beyond words.

Table of Contents

Introduction 1

Literature review 2

Refugia 2

Temporal Influences 4

Landscape and Spatial Ecology 6

Purpose of Study 7

Methods 8

Study Design 8

Data Collection Methods 9

Analysis 13

Descriptive Analysis 13

Spatial Data Statistics 13

Assessment of Environmental Differences 18

Sierra Nevada Environmental Model 19

Results 20

Descriptive Analysis 20

Spatial Data Statistics 23

Environmental Differences 25

Sierra Nevada Modeling Results 27

Discussion 28

Limitations 32

Recommendations 36

Conclusions 38

Tables and Figures 40

Figure 1: Deer mouse habitat in North America 40

Figure 2: Diagnosed HPS cases in the United States mapped by states of exposure 40

Figure 3: Percentage of HPS occurrence by State over Study Area 41

Table 1: Frequency and Percentage Statistics at the COUNTY level 41

Table 2: Frequency and Percentage Statistics at ZCTA level 41

Figure 4: HPS cases within Counties 42

Figure 5: HPS Crude Incidence by County 43

Figure 6: HPS Case Count by ZCTA 44

Figure 7: HPS Crude Incidence by ZCTA 49

Figure 8: Voronoi Map of ZCTA Centroids and Case Counts 54

Figure 9: Voronoi Map of ZCTA Centroids and Incidence Rate 55

Figure 10a: HPS Cases in ZCTA across USGS Ecological Zones 56

Figure 11a: HPS Cases in ZCTAs within Watershed Areas 58

Figure 12a: HPS Cases in ZCTAs across Elevation 60

Table 3: Global Moran's *I* for Case Counts 62

Table 4: Global Moran's *I* for Incidence Rate 63

Figure 13: Global Autocorrelation Spatial Correlogram for case counts 64

Figure 14: Global Autocorrelation Spatial Correlogram for Incidence Rate 64

Figure 15: Local Clustering of Case Counts 65

Figure 16: Significant Clustering of Case Counts 66

Figure 17: Local Clustering of Incidence Rate 67

Figure 18: Significant local clustering of Case Incidence Rate 68
Figure 19: Hot Spot Analysis of Case Counts 69
Figure 20: Hot Spot Analysis of Case Incidence Rates 70
Figure 21: Significant z-scores in Hot Spot Analysis of Case Counts 71
Figure 22: Significant z-scores in Hot Spot Analysis of Case Incidence Rates 72
Figure 23: Elevation.. 73
Figure 24: Ecological Zones 74
Figure 25: Annual Precipitation... 75
Figure 26: Maximum Temperature Averages 1981-2010 76
Figure 27: Minimum Temperature Averages 1981-2010 77
Figure 28: Elevation/Flow Length/Cases in Hot Spots and ZCTA 78
Figures 29 a-d: Hydrology Model Building... 79
Figure 30: Sierra Nevada Study Area 81
Table 5: Sierra Nevada Regression Analysis.. 81
Figure 31: California Aqueducts.. 82
References.. 83

Introduction

Hantavirus Pulmonary Syndrome (HPS) causes severe, sometimes fatal, respiratory distress in humans. This disease was first described in 1993 following an outbreak in the Four Corners region of the United States. The causative agent was quickly identified as a new Hantavirus (family *Bunyavirdae*) and named *Sin Nombre Virus* (SNV). This virus exists chronically in its natural host (the deer mouse, *Peromyscus maniculatus*), which is characteristic of other Hantavirus serotypes [1]. Infection in humans occurs when rodent feces containing infectious virus are aerosolized and inhaled [2-5]. Currently, rodent-to-human transmission is the only route of infection seen in North American; no person-to-person transmission has been reported [1, 4].

Deer mouse populations are extensive and widely distributed across North America including a wide range of environmental conditions and varying human densities (Figure 1). Hantavirus species are endemic in many different rodent populations found all over the world [4, 6]. At this time, over 40 Hantavirus genotypes have been named in the Americas, 17 of which are found in North America. However, only 6 are associated with rodents and currently known to cause HPS in humans [7]. Interestingly, despite documented presence of host and virus across the North American continent, HPS incidence in humans is not evenly distributed. Most HPS cases are reported in the western half of the United States and concentrated in the Four Corners region (Figure 2); the same area where SNV was first identified in 1993.

It is important to note that HPS has been nationally notifiable disease since 1995 [8]. As such, any confirmed diagnosis is expected to be reported to the Centers for Disease Control and Prevention (CDC) through electronic submission within the next reporting cycle [9]. Though still a rare disease in humans, HPS has a high mortality rate: of the 30 cases reported to CDC in 2012, 40% were fatal [10]. Understanding this host-virus relationship is important for predicting risk in humans.

Literature review

Since the 1993 outbreak, a large number of studies have analyzed environmental, host, and human characteristics that may have led to the 1993 outbreak and subsequent Hantavirus Pulmonary Syndrome (HPS) cases in humans. Based on these past surveys and studies, land cover, elevation, and distance to water are basic commonalities associated with increased likelihood of SNV infection in rodent populations. However, the link between confirmed virus presence in mice does not appear to correlate to risk in humans. Despite a growing consensus on similar environmental variables present in previous cases and outbreaks, there is a lack of explanation for the clustering of HPS cases in humans in the western United States. Because the relationship between host, virus, and humans has not been fully explained by environmental characteristics, alternative theories have proposed examining potential mechanisms driving the distribution of and meaning behind disease patterns in humans.

Refugia

Early studies focused on particular environmental conditions potentially affecting rodent densities and how changes in host populations might subsequently change risk of Hantavirus infection in humans. One popular hypothesis focuses on the idea of refugia, or places of ideal environmental conditions where deer mice populations *and* SNV are able to persist over time [4, 7, 11, 12].

For refugia to exist stability must be maintained among desired environmental characteristics or habitat area for the given population; these patches are sustained from year to year despite seasonal or cyclic variations across a number of factors (vegetation, rainfall, human altered landscapes, etc.). The environmental elements consistently linked to higher prevalence of SNV in rodent populations include elevation between 2000m and 3000m [3, 7, 11, 13], diverse vegetation with an extended growing season [5, 13, 14], and measures of precipitation [7, 13, 15-17]. Slope, presence within drainage systems, distance to water, deer mouse population density, and distance from humans are previous identified components [3, 7, 11, 13, 14, 16].

Importantly, the attributes mentioned above may indicate not only consistent presence of deer mice populations, but also the level and frequency of SNV infection [17]. It has been noted with regularity that presence of deer mice alone is not indicative of active, acute SNV infection [2, 7, 13, 16]; some deer mouse populations are sustained without presence of Hantavirus. Of course, in the interest of human population health, identifying the refugia that contain consistent deer mice populations *and* active SNV infection is critical.

Nonetheless, identification of these common environmental characteristics and establishment of refugia have failed to explain the spatial pattern of disease occurrence in humans. Deer mouse populations extend across North America (Figure 1) covering a wide range of environmental conditions with many areas containing all or most of the indicated important characteristics listed above; yet, HPS incidence in humans is not evenly distributed (Figure 2). Despite a growing consensus on similar environmental variables present in previous cases and outbreaks, the basis for clustering of human HPS cases in western States is not known.

Temporal Influences

In addition to understanding key environmental parameters indicative of sustained deer mouse populations, time is an important consideration. Time, on a scale of months, years, and or decades, affects a number of aspects in this complex relationship. The SNV dynamics within deer mice populations varies between breeding seasons, across short-scale and long-scale weather patterns, and as a result of changes in land use practices. Time changes what is seen at the landscape level and therefore, may be an accompanying factor important in understanding the Hantavirus – host – human ecological model.

Mouse life spans are brief in comparison to humans, and represent time on a smaller time scale. Thus, changes in Hantavirus prevalence vary widely over breeding cycles within a single year. One study tested antibody prevalence in mouse samples reporting prevalence differences from less than 5% to over 60% in the same areas over the course of a year [13]. Time, with regard to breeding seasons, may be additionally complicating through a delayed response in Hantavirus prevalence measured through population. It is hypothesized that the breeding season one season prior is the most influential factor in Hantavirus prevalence [4, 5, 7].

Another significant aspect of time may be large scale weather patterns. Seasonal and cyclical weather patterns like the El Niño/La Niña–Southern Oscillation (ENSO) have received much attention, theorized to have important influences on infection rates in mice regardless of presence/absence of refugia [4, 5, 7, 11, 18]. The Trophic Cascade Hypothesis suggests an increase in precipitation increases moisture in the soil triggering more vegetation growth and amplified productivity. Thus, large scale climatic patterns may affect the size and density of deer mouse populations: if more food resources are available, rodent populations can increase [3, 4, 18].

Intuitively, as deer mouse populations increase, susceptible hosts for Hantavirus infection increase and increase the probability of contact with other infected individuals shedding virus. Consequently, as the number of infected mice in a given area grows, the risk to humans intensifies [5, 7, 11].

Despite the logic surrounding the Trophic Cascade Hypothesis and density dependence, the prevalence of Hantavirus infection in deer mice populations frequently do not show a synchronous increase as density increases or vice versa [2, 11]. Further, there is no correlation between high hantavirus infection prevalence in large deer mice populations and cases of human infection [19].

One further aspect of time affecting deer mice populations may be changes in land use practices. Deer mice are described as generalist species in that they are capable of surviving in a variety of environments ([4]. However, changes in landscape features can disrupt food availability and refugia sites. Available habitat directly effects population size by having a finite set of resources capable of sustaining growth [20]. As variations in SNV infection risk in humans may depend on the spatial configuration of rodent populations within the human landscape [17], it is important to consider the larger ecological effects different land use functions can have. The resulting landscape structures, both natural and human influenced, may affect virus distribution within deer mouse populations to a greater extent than ecological variables alone [21].

Landscape and Spatial Ecology

An alternative hypothesis explores the influence of host density and acute Hantavirus infections by examining deer mice populations across large regions of space. Using the theories of landscape and spatial ecology, local rodent populations are not seen as distinct entities, but as part of a larger, more widely distributed population. As such, rates of infection within local communities are partially dependent on the degree of fragmentation and connectivity between smaller populations. Island Biogeography Theory and Metapopulation Theory are two theories offering possible explanations of mice population interactions at the landscape level. These both consider the effect of refugia within a wider distribution of suitable habitat areas or patches.

Island Biogeography Theory suggests there is a mainland that serves as a source of Hantavirus infections in new "islands" and or reinfection for mice populations in once established "islands", even after local extinction of virus or deer mouse populations occur in a given "island" [2, 11, 22]. "Islands" can be in reference to refugia and local habitat but also individual deer mouse or humans as hosts for "colonization" of Hantavirus [23]. Reinfection is expected after periods of isolation when overall population densities are able to grow again to "recolonize" new mouse populations.

Metapopulation Theory lessens the restriction of a single mainland and says there may by multiple refugia or source sites and all individuals within these refugia patches have the ability to populate any of the other available patches [24, 25]. Like Island

Biogeography Theory, the ability to repopulate is determined by the level of fragmentation and connectivity between “islands” or patches. Both theories offer explanations as to why cases seemed to cluster in areas across time periods: they hypothesize that within an area of repeated infection, some sort of refugia or viral reservoir exists.

These concepts may aid in the explanations of why cases seem to cluster in the same areas across time periods and add to the overall picture of Hantavirus-host dynamics. However, to further complicate the relationship, the degree of fragmentation or connectivity changes in response to environmental factors (such as those described in the Trophic Cascade Hypothesis), changes in land use (often from human activities), and changes in population densities in both humans and deer mice populations.

Purpose of Study

Combining the aforementioned theories and studies, the Cascade Mountain Range and Sierra Nevada Mountains is an intriguing area for application. In preliminary mapping of exposure in reported human cases, a distinct pattern emerges: cases appear on the eastern side of these mountains with little to no cases appearing on the western slopes. This study aims to describe the overall spatial pattern and temporal characteristics of HPS cases in California, Nevada, Oregon, and Washington including the identification of clusters of cases considering both the time and space of case location.

Indications to the presence and location of clustering among cases will be assessed using a variety of spatial and temporal analyses and software. In areas with evidence of clustering, statistically significant differences in variables such as landscape features, environmental conditions, and human populations will be assessed comparing areas within a cluster to those outside. The importance in analyzing and understanding hantavirus ecological studies with the inclusion of spatial and temporal descriptive statistics is to develop more sensitive and accurate models to predict areas of high infection risk for humans [4, 7, 17].

Methods

Study Design

This retrospective consecutive case series analysis uses information of confirmed HPS cases from the national HPS surveillance system surveyed through the Viral Special Pathogens Branch at the Centers for Disease Control and Prevention. Because of the study design, characteristics surrounding cases, space, time, and environment, were the sole targets of interest with no assigned control group [26].

The overall objective is a descriptive spatial data analysis of the spatial and temporal characteristics of HPS disease in California, Nevada, Oregon, and Washington to associate spatial clustering with concepts of landscape ecology. Assessments were made using several software packages including ArcGIS 10.1 by ESRI, Cluster Seer 2.3 by BioMedware, and GeoDA from the GeoDa Center for Geospatial Analysis and Computation at Arizona State University. Their use will further investigate two hypotheses:

1. Clustering of cases exists in and amongst the four study states with increased incidence occurring on the eastern slopes of the Cascade and Sierra Nevada Mountain ranges.
2. Clustering can be explained, in part, by particular environmental characteristics surrounding the Sierra Nevada Mountains.

Data Collection Methods

Patients meeting the case definition for HPS (acute febrile illness with respiratory involvement) and having laboratory confirmation of hantavirus infection (positive hanta serum ELISA, RT-PCR, or immunohistochemistry result) were registered by reporting States completing a standard Case Report Form including the ZIP Code, county, State of residence, and exposure if different from residence. Case data were supplied by the Hantavirus Registry managed by the Viral Special Pathogens Branch at the CDC. Microsoft Access and Excel were utilized to gather cases listed as having an exposure in California, Nevada, Washington, and Oregon. ZIP Code of exposure was the targeted variable of interest and every case entry was scrutinized for missing and or incorrect information. Whenever possible, questions were answered by referring to original Case Report Forms submitted by States when reporting a positive case of Hantavirus infection. When no further information was available from these forms, the corresponding state health departments were contacted to identify missing information.

Cases were included in the analysis based on the level of completeness and confidence surrounding exposure information. Most people infected with Hantavirus are exposed in their place of residence or some peri-domestic property nearby. Because of this pattern, often a home address was listed with no explicit exposure address. In these cases, especially when accompanied by a statement aligning to the fact, home address was used as exposure information. In other cases, the address included – either home or exposed – was limited to the town and or county name for privacy regions. Fortunately, many of these areas are rural with sparse populations such that only one ZIP Code is assigned for the corresponding towns and or counties. The United States Postal Service's (USPS) "Look Up a ZIP Code" tool and a Google search for "zip code maps" were used to fill in missing ZIP Code information when only a town or county name was supplied.

Cases were excluded when both exposure address and place of residence address were missing and no supplemental statement explaining possible exposure was present, or when individuals had multiple possibilities for exposures listed, such as frequent or repeated travel within in several endemic regions.

ZIP Code data were obtained through proxies supplied by the U.S Department of Commerce, United States Census Bureau. Though ZIP Codes are a trademark of the USPS, they are not designed as polygons which can be easily mapped, but are instead, a functional network system of lines and points aiding mail delivery [27]. Bridging the gap between USPS function and mapping utility are Zip Code Tabulation Area (ZCTA) files produced by the U.S. Census Bureau. Each ZCTA is an aerial representation of the USPS ZIP Code areas with spatially referenced information for the calculated zip code borders.

The 2010 ZCTA Relationship file was downloaded which included ZCTA-ZIP Code equivalents and comprehensive demographic information collected in the 2010 National Census. Details for every ZCTA include overall area, the area of only land, the area of only water, the estimated 2010 population, and the greater county FIPS (Federal

Information Processing Standard) code to which the ZCTA belonged. Information was provided for the entire United States and all were deleted except entries from California, Nevada, Oregon, and Washington.

Coding HPS cases by ZCTA represents case data at the finest level of spatial resolution possible. This is important because statistical power decreases as false detection of clusters increases with any aggregation of data and ZCTA, though not as spatially resolute as point data, as a good alternative, in theory [28]. However, because of their irregular shape and patchy distribution, analysis of ZCTA is dysfunctional given the input requirements for spatial data analysis. For this reason, HPS cases also were coded at the county level.

Information on counties in the ZCTA file was limited to the county FIPS code, so an additional database was obtained from the Census Bureau. This file included further county information, most importantly, county name. A Geographic Comparability File was chosen for its inclusion of both the county FIPS code and county name. Additionally, this file contained a comparison of every county name and area in 2010 to its name and area in 2000 with notes discussing any differences. County names were added to the ZCTA Relationship File in Excel adding to the overall data available for each ZCTA, and a separate file containing county information and number of HPS cases was created.

Shapefiles for use in ArcMap 10.1 were gathered from a variety of sources. Administrative files including ZCTA, counties, state lines, and urban areas were obtained from the 113th Congressional District TIGER/Line® Shapefiles created by the US Census Bureau. These files were created using the Geographic Coordinate System: North American Datum 1983. Both a Projected Coordinate System preserving distance (the

USA Contiguous Equidistant Conic) and area (the USA Contiguous Albers Equal Area Conic USGS version) were used in different analysis and applied to the TIGER shapefiles at different times.

Elevation data from the United States Geological Survey (USGS) was obtained with 1000 meter resolution for the entire study area. Simple hydrological models were built using these elevation data to create maps of water flow direction, water flow accumulation (without consideration to human-altered landscapes), and water flow length. Basic watershed and ecological zone shapefiles also were obtained through USGS.

Finally, data on additional environmental parameters including precipitation and temperatures were obtained using the United States Department of Agriculture's (USDA) Geospatial Data Gateway. PRISM data, or Parameter-elevation Regressions on Independent Slopes Model [29], is the official climatological data for USDA and is created from an analytical tool incorporating point data, digital elevation models, and other spatial information inputs to generate estimates of climatic parameters including precipitation and temperature. Output from this model extrapolates information from measured inputs (weather stations, for instance) across a gridded layer taking into consideration different time scales (month, year, etc.)" [30]. The PRISM data utilized in this analysis included annual rainfall in California and Nevada averaged from 1981 through 2010, along with maximum and minimum temperature averages calculated across the same time period.

Analysis

Descriptive Analysis

Initially, a series of descriptive analyses were performed for the study area including the cumulative number of counts captured through surveillance data and the incidence rate of cases occurring in each ZCTA, County, and State. A crude incidence rate per 100,000 people was calculated in Excel for each county and ZCTA using the 2010 population estimates for that area as the estimated population at risk.

All shapefiles including State lines, county lines, temperature data, rainfall data, watershed delineations, and elevation data were added to ArcGIS 10.1. A join was made between the ZCTA shapefile in ArcMap containing polygons with assigned ZCTA numbers and the ZCTA Excel file. This created an extended attribute table with added information on Case count, ZCTA area, population (in both ZCTA and county), county name, and county area.

Chloropleth maps of cumulative case counts and crude incidence rates at both the county level and ZCTA level were made including the entire study area as well as for each individual State. Incidences, rates and counts, were mapped over several environmental features including elevation, ecological zones, and watershed areas, visually noting any areas of congruence or pattern.

Spatial Data Statistics

Further analysis involved a variety of spatial data statistical methods. Tests were performed using the software packages ArcGIS 10.1 by ESRI, Cluster Seer 2.3 by

BioMedware, and the GeoDa Center for Geospatial Analysis and Computation at Arizona State University. Functions within these statistical packages characterize the distribution and spatial relationships between areas with and without cases. Relationships were calculated using Global indexes of spatial autocorrelation, comparing the event (incidence rate or case count) in every individual spatial region to the overall study area average. Local indicators of spatial autocorrelation compared every spatial region to every one of its neighbors looking for outliers with significantly high or low event numbers. Finally, the Getis G*i(d) function was used to describe Hot Spot areas most likely to experience increased case occurrence.

Spatial regions, as initially defined in this dataset, included ZCTAs and counties; however, due to complications concerning these regions and the spatial statistical functions, additional polygons were created in two Voronoi diagrams containing Thiessen Polygons. To create these maps, centroids for all ZCTAs were calculated using ArcMap 10.1. Then, using the Geostatistical Analysis function, polygons were created around every centroid such that each line in a given polygon was exactly halfway between that centroid and its nearest neighbor on that side. These diagrams displaying case incidence rate and case count were specifically produced for use in global and local autocorrelation testing.

All four study states were considered in global and local autocorrelation and hot spot analysis measurements: California, Nevada, Washington and Oregon.

Global Spatial Autocorrelation

Global indices of spatial autocorrelation compared the distribution of case incidence rates and case counts within regions across the entire data set. Several statistical tests are available to calculate global autocorrelation and one of the most common and well recognized is the Global Moran's I test [31-34]. The goal of these global spatial autocorrelation summaries was to describe the degree to which similar observations, HPS cases in this study, occur in neighboring observations or regions [32]. This test assumes HPS occurrence is spatially independent of its neighbors, such that disease occurs randomly across regions and the population at risk is evenly distributed.

Moran's I is defined as:

$$I = \frac{\left(\frac{1}{s^2}\right)\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{N}\sum_{j=1}^{N}w_{ij}},$$

[32]

$\bar{y} = \sum_i y_i / N$

$$s^2 = 1/N\sum_{i=1}^{N}(y_i - \bar{y})^2$$

[31]

where *i* and *j* represent the geographic units (or counties for this test), ω_{ij} is a weight determined by the distance between *i* and *j*, y_i is the number of cases in that county, and *N* is the population of the same county.

Using the software ClusterSeer, the global spatial analysis test, Moran's *I*, was used to determine the existence of spatial correlation in both raw incidence and person-time incidence per 100,000 people. Significance was assessed at the 0.05 level with 999 repeated Monte Carlo Simulations. Values of the statistic range between -1 and 1 and can be compared as the spatial version to the Pearson correlation coefficient [33]. If the null is true, no spatial autocorrelation exists and cases occur randomly, the value of *I* will be 0. If the value is positive, there is evidence of positive spatial autocorrelation or clustering between similar values and neighboring regions [32]. Negative *I* values can be interpreted as evidence of negative spatial autocorrelation or regular, repeating patterns [31, 32].

Moran’s *I* tests were also calculated using the software PPA from South Dakota State University. Results of this test differ from the single value output from ClusterSeer: instead of a single Moran’s *I* value calculated for the entire study area, this formulation calculates an *I* value at multiple distances which can be displayed visually in a Spatial Correlogram using Microsoft’s Excel.

Local Spatial Autocorrelation

Moran’s *I* and other tests of global indexes of spatial autocorrelation assess tendencies in the study area for similar values or disease cases to cluster in certain subsets of the overall study area [31, 32, 35]. Because individual regions are compared to the global mean of the overall study area, smaller, more local clusters can be averaged away and missed completely in the results. Additionally, global indexes provide only suggestions of clustering rather than identification of actual clusters. Local indicators of spatial association (LISAs) were developed specifically to address these issues and are routinely used to define the areas or regions where increased disease incidence rates are most likely to occur [32, 35].

Similarly to global indexes, there are several tests of local association; one of the most popular is the Local Moran’s *I.* This test is analogous to the global Moran’s *I* except an *I* statistic is calculated for every geographic region as compared to one *I* for the entire study area [31]. Further, the sum of all calculated Local *I* values is proportional to the overall Moran's *I* statistic from the global autocorrelation test [36]. The goal is identifying spatial outliers or specific locations where similar and dissimilar rates of HPS occur in comparison to neighboring spatial regions. Through this procedure, every region is compared to its neighbors with the expectation or null hypothesis that there is no difference between HPS crude incidence rate per 100,000 people in one region and the HPS incidence rate in its neighboring region.

The Local Moran's I is calculated with the following formula:

$$I_i = Y_i - \bar{Y} \sum_{j=1}^{N} \omega_{ij} \, Y_j - \bar{Y}$$

where *i* and *j* represent the geographic units (counties), ω_{ij} is a weight determined by the distance between *i* and *j*, Y_{ij} are the number of cases in the two regions being compared, and *N* is the population within the selected area. [32]

Using the software GeoDa developed by Arizona State University, a Local Moran's I analysis was conducted. Spatial weights were created using Queen contiguity such that every neighbors in every direction were compared and case incidence in nearer neighbors were given a higher weight. Significance was assessed at the 0.05 level with 999 repeated Monte Carlo Simulations. Output for this test, as stated above, yields an *I* value for every county. Higher values of *I* indicate stronger local correlations through similar counts or incidence between nearby regions, though not necessarily higher count of incidence [32]. A LISA Cluster Map was produced displaying results of the Local test and the significant differences between counties. Relationships are described as High

HPS rate next to High HPS rate (High-High), Low HPS rate next to Low HPS rate (Low-Low), and so on for High-Low, and Low-High.

Hot Spot Analysis

An additional local test was performed at the ZCTA level using the Getis Gi*(d) function in ArcMap 10.1 This test is useful in naming hot-spot areas or spatial regions with significantly higher (or lower) HPS rates as compared to their neighbors in adjacent regions. This is similar to the Local Moran's *I* by identifying local areas different from its neighbors but with one important distinction: it is not related to a Global indicator of spatial association [36] and is determined through vector algebra. Identification of hot-spot regions allowed a focused assessment of differences, specifically across environmental characteristics within a hot-spot area to surrounding spatial regions.

Assessment of Environmental Differences

Drawing from published literature, environmental features of particular interest included elevation, land cover through ecological regions, temperature, and precipitation. Data was collected from USGS and USDA and processed in ArcMap 10.1. For several reasons, investigation of environmental features was restricted to California and Nevada. One was because of the large study area and the need to reduce computing requirements and increase feasibility. The other major reason was because of stability in hot-spot results from the Getis G*i(d) tests between these two states.

All data were imported into ArcMap in Raster format, expect for ecological regions which was recorded by polygons. Initially, all data were reimaged with color:

elevation values were unchanged for this visual appraisal, though maximum and minimum temperatures and annual precipitation were reclassified by standard deviation estimates. Case data of counts and incidence rate by ZCTA and Theissen polygons were overlain on each layer file with slight transparency to retain environmental data underneath. Maps of cases by elevation, maximum temperature, minimum temperature, precipitation, and ecological regions were created.

Additionally, a basic hydrology model was created for California and Nevada using elevation data. This was done in addition to precipitation averages as another measure of water availability through investigation of water flow and drainage patterns. Flow direction was assessed for every pixel in the raster data layer in ArcMap 10.1. The direction in which water would flow from this point was calculated producing a new image. This layer was assessed for any gaps or holes in the data, a common error in satellite imagery, and these spots were filled in. A new flow direction map was then created allowing further manipulations to evaluate areas where water should accumulate based solely on elevation data and the length in space water might flow from a given point (drainage). Cases by ZCTA and Voronoi Hot Spots were overlain to consider any patterns between the variables.

Sierra Nevada Environmental Model

Further assessment of the relationship between environmental features and HPS cases centered on one anticipated Hot Spot area: the Sierra Nevada region. Overlaying ZCTA centroid point data on semi-transparent ecological regions and elevation, an oval shape was drawn over the Sierra Nevada Regions with peak elevation serving as the center and guiding line. Centroids within this region were selected for statistical analysis to compare environmental and case differences from one side of the mountains to the other. Raster data from the four layers were extracted to the centroid and added to the attribute table in ArcMap, which was then exported to SAS as well as Excel for analysis in Matlab.

Special consideration was given to the centroid representing Yosemite, CA: ZCTA 95389. In 2012, an outbreak occurred in this area ending with ten confirmed HPS cases. Nine of these ten people stayed in the same location (Curry Village Signature cabins, now closed) over the summer [37]. Though most cases recorded in the Sierra Nevada Region and across the greater study area report an exposure in place of residence or some area, Yosemite is occupied predominantly by visitors, especially during summer months. Additionally, nine of these ten cases represent only one exposure location. For these atypical characteristics, statistical models were built first including the outbreak and then again excluding nine of the ten cases. The tenth case was exposed in a different region of the park and was included in both models. Matlab and SAS were used in regression modeling.

Results

Descriptive Analysis

Of the over 600 cases of HPS that have been reported across the United States and recorded in this database since 1995, 144 or nearly 24% of cases have a likely exposure in California, Nevada, Washington, and Oregon. After investigating original Case Report

Forms and communications with State Public Health Departments, a final count of 124 cases listed among 91 different ZCTAs remained in the dataset based on completeness of exposure information. This represents approximately 86% of all reported cases occurring in nearly 3% of all ZIP Codes from within the four targeted States.

Human cases of HPS occurred at varying frequencies and percentages across States (Figure 3). Between the four states, 59 out of 150 counties, nearly 40%, reported at least one case of HPS since 1995 (Table 1). 39% of counties with HPS occurred in California which was also the state with the most overall reported HPS cases (44% or 55/124 cases). Washington and Oregon contributed similarly in percentage to the overall number of counties with cases, ~25%. Though Nevada’s 10 counties with reported HPS cases made up only 11% of the overall total, this figure represents nearly 60% of all of Nevada’s counties – the highest percentage of counties affected when looking on a State-by-State basis.

Overall HPS case percentages at the ZCTA level of resolution were a degree of magnitude less when compared to the county level. A total of 2962 ZCTAs exist between the four study states and only 3% of all ZCTAs have reported HPS cases (Table 2). Similarly to counties’ statistics, the vast majority of ZCTAs, nearly 60%, came from California and least overall contrition from Nevada, nearly 6%.

Chloropleth maps of case count and crude incidence were mapped in ArcGIS 10.1 at both the county level and the more spatially resolute ZCTA level. Mapping by counties of both case counts and incidence rates resulted in large areas of color, indicative of HPS case occurrence (Figures 4, 5). Contrastingly, ZCTAs returned a more detailed, specific location of previous HPS cases but also highlighted an unanticipated issue – large gaps of space (represented in gray) not included in any ZCTA or ZIP Code area (Figures 6a-d, 7a-d). These areas include national parks, water bodies, other federal lands, and areas without human populations (and without any need of mail delivery services). Over concern of statistical power with so much missing information on the one hand, and over-estimating HPS influence on the other hand [38], the alternative method using Voronoi Diagrams of Thiessen polygons was also coded with cases of cumulative counts and incidence rates (Figures 8, 9).

HPS cases were also mapped by ZCTA across ecological zones, watershed areas, and elevation (Figures 10, 11, 12). When mapping HPS cases with ecological zones, a notable trend can be seen running north to south between edges of different zones denoted by a red line drawn in Figures 10a and 10b. With the exception of the grouped cases in eastern Washington, many case ZCTAs including those in western Washington, eastern Oregon, and possibly central Nevada follow a similar pattern with cases occurring along edges of ecological zones. Central Nevada is difficult to assess accurately given the large size of ZCTAs.

Likewise, cases across watershed areas tend to occur along boundary areas (Figures 11a, 11b). Elevation may be the better explanation as watersheds depend in part on elevation (and gravity). Comparing watershed boundary lines to the elevation map (Figures 12a and 12b), the association between peak elevation regions and lines delineating watershed becomes apparent. This is particularly true in the Sierra Nevada Regions where cases by and large appear on the eastern side of watershed boundaries moving along the mountains’ peak elevation ridge.

Spatial Data Statistics

As stated previously, in an effort to preserve statistical power and limit overestimation of HPS risk, Voronoi Diagrams were used in autocorrelation and cluster analyses. With ZCTAs being small, irregularly shaped, and sometimes separated by large distances of open space, the statistical tests needed for spatial data analysis at both the global and local level were simply not able to perform well under these conditions. Because the global and local statistics function best with similarly shaped neighboring polygons, incidence rates and case counts were initially tested at the county level. However, the large range of size and population densities within counties was too extreme and may overgeneralize Hantavirus presence. The variance between counties reduces reliability in autocorrelation test results and analysis was therefore conducted through Voronoi Maps of case count and incidence rate (Figures 11, 12).

Global Spatial Autocorrelation

Moran's I test results using case count and incidence rate from Voronoi Maps through ClusterSeer yielded a Moran's $I = 0.112608$ for case counts and Moran's $I = 0.121650$ for incidence rate. Though the values are small, only slightly above 0, both are significant at the 0.05 level and both are positive evidence for spatial autocorrelation or clustering at the Voronoi polygon level of resolution (Table 3, 4).

The additional Global Moran's *I* test run with PPA software from South Dakota State University yielded a output with columns of incremental distances and an accompanying Moran's *I* value estimate for every increasing distance level. This information was exported into Microsoft's Excel creating two Spatial Correlograms visually representing the statistic and displaying distances where autocorrelation is maximized (Figure 13, 14). Though the single value resulting from ClusterSeer’s analysis showed only a hundredth of a decimal difference between case counts and incidence rates, the maximized distance displayed in the Correlograms was approximately double for incidence rates as compared to cumulative counts. The slope of the line is smoother as well. This is not surprising given that incidence rates per 100,000 people are a technique used to make differing units or events more comparable.

Local Spatial Autocorrelation

The LISA test in GeoDa yielded two maps for each test of local autocorrelation: a LISA Cluster Map and an accompanying Significance Map (Figures 15-18). Comparing Cluster Maps against each, similar numbers and locations of relationships deemed High-High were generated. The largest discrepancies between case counts and incidence rates were in the distributions of polygons assessed to have a Low-High relationship in comparison to their neighbors. Overall assessment of cluster locations was comparable with some differences occurring across Washington and Oregon.

The Significance Maps displays statistical significance of each polygon. Considering significant clustering of case counts, 208 polygons had significantly different relationships at the 0.05 level compared to their neighbors and 112 polygons were significant at the 0.01 level. Likewise, the incidence rate Significant Map returned 122 polygons significant at the 95% level and 90 significant at the 99% confidence level. Importantly, these Significance Maps are useful in showing statistical significance but not in determining clusters of locally high or low values.

Hot Spot Analysis

Application of the Getis Gi*(d) Local Hot Spot assessment in ArcMap 10.1 yielded some surprising results. While hot spot regions in California and Nevada were similar between application of case counts and incidence rates, results in Washington and Oregon varied significantly. Figures 19 and 20 display the initial output from ArcMap10.1. After correcting for the multiple comparisons and restricting significance to only those polygons with a z-score > 3.71 Figures 21 and 22 were created. The contrast between case count and case incidence rate is more apparent in these images.

When only considering counts, a significant hot spot area is identified in Eastern Washington; however, in the map considering case incidence rate, the hot spot disappears completely and instead, an area in Oregon is identified. Smaller differences between these two images fall in eastern California and central Nevada. With only bearing in mind case counts, the Getis statistic generated two distinct hot spot regions – one mostly in east-central California, slightly spilling over into Nevada and another oblong-shaped cluster in central Nevada. When looking at incidence rates per 100,000 people, however, the two areas were combined into one large region.

Environmental Differences

Displaying cases, either through ZCTA or significant Getis' hot spots, across the different environmental landscapes generated from raster data produced several interesting figures. In many cases, the environmental features encompassed by cases within the two classifications (ZCTAs vs. Hot Spots) were different; and other times, the characteristics were similar. Elevation and cases, for instance, either within ZCTAs or

Voronoi polygons, displayed previous HPS cases falling predominately on areas of higher elevation (Figures 23).

Mapping of precipitation averages and cases, on the other hand, presented a very different situation: Instead of most cases falling in similar parameter ranges (like elevation), the Voronoi hot spot spread across the Sierra Nevada Mountains with half of the hot-spot region in an area receiving high precipitation and the other half in an area with very little annual precipitation (Figure 24). Patterns across Ecological Zones are also less distinct with displaying the hot-spot of Voronoi polygons in comparison to ZCTAs (Figure 25).

Maximum temperature averages, similarly to precipitation, varied across the Sierra Nevada Mountains, such that the hot-spot identified through the Getis G*i(d) contains both extremes captured in the maximum temperature averages. Minimum temperature averages in comparison to the case classifications, alternatively were more similar to patterns seen with elevation. Cases, either in ZCTAs or the hot-spots appear mostly in the lower extremes of minimum temperature averages. Comparing the two temperature images together, cooler temperatures in both seem to contain most of the HPS cases for these states (Figures 26, 27).

Finally, displaying flow length created within the hydrology model together with elevation, created a striking 3-D image presenting water pathways amongst mountain passages down through coastal drainage sites. Cases were inserted on to the image and an interesting pattern between ZCTA cases appeared in the mid-Sierra Nevada range moving south and curving towards the coast. These cases seem to be following both the eastern edge of the mountains and an extensive drainage channel (Figure 28). Additional images resulting from the hydrology model are included (Figures 29a-d).

Sierra Nevada Modeling Results

The study area in which modeling was attempted contained 213 centroid points calculated from ZCTA polygons. Dividing this region in half yielded 112 points coded as being on the “west” side of the Sierra Nevada Range and 102 coded as East (Figure 30). In trying to determine statistical differences between mountain sides and cases counts, results were insignificant until addressing the issue with the 2012 outbreak. Removing nine cases can be justified as these cases were both uncharacteristic and not representative of usual HPS occurrence. Additionally, they signify a single exposure site – it was human behaviors and not environmental characteristics that stimulated this outbreak.

A stepwise linear regression model was selected with total case counts as the dependent variable and environmental characteristics as the independent variables: minimum temperature, maximum temperature, precipitation, elevation, flow accumulation, and orientation in the mountains (East or West). The final model chosen:

$$\text{Total case count} = -7.1508 + 0.112(\text{Max. Temp.}) + 0.001(\text{Elevation}) \\ + 2.904(\text{Orientation}) - 0.085(\text{Min. Temp} * \text{Orientation})$$

Precipitation and flow accumulation fell out of the model. Full results including p-values and test statistics are displayed in Table 5.

Discussion

The significant results of the autocorrelation tests testify against the assumption that populations at risk are evenly distributed. From spatial data analysis including all four study States, Washington, Oregon, California, and Nevada, statistical tests for autocorrelation and clusters were significant at both global and local levels. This indicates that cases or incidence at similar values occur next to other in space. In other words, the distribution of HPS is not random, and some factor or series of factors may explain distributions. Further, physical locations of likely clusters and hot-spots in actual space were identified. This is useful for promoting focused research and directed public health activities.

Additionally, case distributions through either ZCTA or cases within significant hot-spot regions, occurred with repeated patterns across environmental variables. Cases by ZCTA arise most frequently at the intersection or edges of ecological zones and along watershed boundaries. Higher elevation, cooler temperatures, and drier climates were also seen as consistent factors for cases in California and Nevada.

Statistical modeling across the Sierra Nevada Mountain range explains case variation in the area as a function mostly dependent on a ZCTA's orientation East or West of mountain peaks. Precipitation was expected to be a strong contributor in explaining HPS cases and location, but was not significant in this model. The most likely explanation is that the effects of precipitation are expressed through orientation and are therefore, not significant as a separate independent variable. This model is not predictive of HPS cases, but is instead comparing environmental differences between ZCTA centroids and the environmental data collected at that point. Overall, the model explains most of the variance seen in cases given the variables included and that significant differences exist between the Eastern and Western Sierras.

Assessment of spatial patterns in this study was accomplished through use of Voronoi diagrams. Thiessen polygons encompass ZCTA centroids with the intent of compromise between broad data aggregation at the county level and the more spatially resolute information contained by ZCTA but with large gaps of missing information. Because the global and local statistics preform best with similarly shaped neighboring polygons, incidence rates and case counts were not able to be tested through ZCTAs despite being the most precise point of case data.

Counties, likewise, were unreliable because of the wide range of size and population densities within counties, which may overgeneralize Hantavirus presence. Some counties encompass a vast amount of land: Nye County in Nevada, for example, is immense with an estimated square mile area near the area of Connecticut, Delaware, New Jersey, and Rhode Island combined. Yet, population is estimated at only 44,000 people. This can be compared to Los Angeles County which is nearly a fourth in size of Nye County but has a population over 220 times larger. For this extreme variance, use of Thiessen polygons within the Voronoi Diagram was deemed reasonable.

Unanticipated differences between measures of case counts and case incidence rates in the autocorrelation and hot-spot results must be considered. Despite that incidence rate is popular method of analysis in epidemiological studies, HPS is a rare disease most often occurring in areas with low population. Otherwise known as a problem of “small numbers”, adding one case to a zip code dramatically changes disease rate.

To give an example of how small numbers can affect output, consider first, that the most common number of cases per ZIP Code is 0. Then looking only at ZIP Codes with a case, the most common number of cases is 1. In the instance of Hot Spot analysis results, regions in Washington and Oregon deemed significant varied dramatically between analyses of cumulative counts vs. incidence rates. The hot-spot area identified in Oregon with incidence rates was clustered around a single case in twenty years of surveillance. This is compared to Washington where most of the ZIPs identified contained one or more cases. Though these ZIP Codes may in fact be in an area endemic for SNV in deer mouse populations, acknowledging it as true hot spot area might be not be reflective of human risk.

Another possibility affecting interpretations of case incidence rate is the use of Voronoi diagrams given how they were created. Though every Voronoi polygon was centered on an actual ZIP Code or ZCTA, the population within these administrative areas may or may not be included in the Voronoi polygons and may not be reflective of actual population distributions. However, this is unlikely to be a great issue as ZIP Codes are created by the USPS in response to populations, with greater densities containing more ZIP Codes. Therefore, the new polygons are likely to include the population from which incidence rates were calculated.

Since the 1993 outbreak in the four corners region of the United States, when SNV was first isolated, a large number of studies have analyzed environmental and host characteristics. However, information on human behaviors and responses is limited. The Trophic Cascade Hypothesis theorizes increased precipitation in climatic events such as

ENSO will lead to increased productivity, increasing food availability, and a subsequent increase in deer mouse populations. Thus far, the theory seems to connect increased productivity to increased deer mouse densities but does not explain well the spatial distribution of reported human cases. This might be attributable to lack of access to human case data or assessment on too short of a time scale, and warrants further investigation.

Likewise, referencing theories of Island Biogeography and Metapopulations may be of particular use given the similarities of environmental characteristics seen in the comparisons of cases by ZCTA and various environmental data layers. Connectivity between suitable habitat patches, understanding the existence and use of corridors for deer mice would aid prediction of SNV presence.

It’s been shown in previous studies the effect of land use management strategies and how different practices affect either the distribution of deer mice populations, or risk to humans, or some combination of the two. The actual density of deer mice specific to a local area may be less important than the overall environmental regions. This is because the probability of acute infection presence in deer mice is heavily influenced by deer mouse population density patterns within a region considering connective pathways through which SNV could travel [2].

Take for example, that cases occur most frequently on the eastern side of the Sierra Nevada mountain range, as seen in Figures 24-27, and demonstrated through the environmental model. Yes, environmental characteristics are more uniformly similar on one side of the mountain compared to the other, but perhaps SNV activity is greater on the east because connectivity to infected refugia is greater and not that it is drier, or cooler, etc. Perhaps there is less fragmentation as there are less people on the Eastern slopes and SNV is able to persist with more consistently given greater connectivity between patches or “islands”.

Given the significance that populations at risk are not evenly distributed, further research is needed comparing differences between identified hot-spot regions and nearby ZIP Codes not included in the hot-spots. Assessing differences, but then creating policies to mitigate risk are the key next steps.

Limitations

A number of limitations presented throughout the research process. A consistent issue was in dealing with HPS as a rare disease. Because so few cases have been recorded across the entire United States in nearly 20 years of surveillance, analysis must be considered carefully. Reports of confirmed HPS cases in a given State to CDC are most commonly an intermittent event with one or two cases every few years. Additionally, the unbalance of reported cases across the country is another consideration when, for example, extrapolating significant results from one area in the West to a State in the East. Several States in the Eastern and Southeastern United States have never reported a case of HPS, as compared to the Four Corners regions with 90 confirmed cases in New Mexico alone.

Another statistical concern given the rarity of this disease is in the issue of “small numbers”. Because HPS occurs infrequently in most populations, adding a single case incidence to a ZIP Code, for instance, dramatically changes disease rate. In this sample of cases, most ZCTAs with a recorded HPS cases had a cumulative case count of 1… in 20 years of surveillance. Further, populations reported in the 2010 Census recorded at every ZCTA vary greatly. As incidence rates are significantly changed with any additional case, the resulting maps may reflect some combination of true risk for these populations and an inflated risk that is difficult to differentiate. Additionally, it is possible that some cases have been missed completely and never diagnosed or included in the HPS Registry.

Surveillance on positive HPS cases is active across the country; however, certain case information details (name, home address, age, occupation, etc.) are not required on Case Report Forms and are therefore, not consistently reported between States out of privacy concerns. In this study, ZIP codes where chosen to represent data at the most resolute level of information, but full addresses were rarely included in reported data, again, for privacy reasons. Often, only the town or county is included, but as stated earlier, this was less of a problem than initially imagined as cases occurred in rural, less densely populated settings and were often assigned only one ZCTA code. Despite being able to assign a ZCTA to all included cases, detail variations in what is reported from States makes comparisons between States more difficult given the extra time, efforts, and energy needed to adjust case information across States.

Further, use of ZCTA codes as the primary region of interest adds another potential concern. ZIP codes as defined by the United States Postal Service are created as a tool for the Postal Service to aid in the delivery of mail and other post. Just as populations and city centers change over time, so do ZIP codes. ZCTAs, on the other hand, are created by the Census Bureau in 10 year increments: first in 2000 and again in 2010, and are therefore relatively static. To create a ZCTA, the most frequently occurring ZIP Code in an area is chosen to represent that area. This excludes some very small ZIP

Codes completely or renames them such that some addresses end up with a different ZCTA code from their ZIP Code. Additionally, if a ZIP Code was never considered most frequently occurring in a given area when creating ZCTAs, the ZIP Code is simply removed [27].

ZCTA codes widely vary in size across the study area with the smallest zip code measuring 9,145 square feet (located in King County, WA) and the largest: 13,462,551,550 square feet (in Malheur County, OR). Population density, likewise, varies dramatically; per 10,000 square feet: the smallest density near 0 as compared to the highest located in San Francisco County with a ZCTA population density near 197 people/10,000 ft^2. Because explicit addresses could not be obtained for the vast majority of cases, analysis was restricted to the ZCTA code – exploration could not be done at any finer resolution. This leads to the potential for ecological fallacies as associations seen at the zip code level are extrapolated to the individual (human) level which may or may not be accurate [32].

Similarly, using ZIP Codes of exposure, modeled through ZCTAs, instead of individual cases as points, makes analysis results subject to the Modifiable Areal Unit Problem (MUAP). This is a problem when associations between variables or events change when the scope or scale in consideration changes. In this study, ZIP Codes were the variable of interest. For MUAP not to be an issue, results of the analysis would have to yield the same results and patterns regardless of whether point data of cases was used in analysis or ZIP Code regions. Given that ZIP Codes and population density within ZIP Codes vary widely in size and patterns appear different between ZCTAs and Counties, this is a legitimate concern.

Regression modeling of the Sierra Nevada Mountain regions was a useful first attempt at modeling environmental differences between nearby, similar regions. However, data was severely limited and may not reflect variance in features accurately. Environmental data was extrapolated to the ZCTA centroid for analysis and was true for that point, but all values surrounding this centroid were ignored. Ideally, information could be extracted at the ZCTA polygon level to calculate basic descriptive statistics: maximum, minimum, mean, median. Because HPS is rare and is dependent somewhat on human behavior (cleaning, dusting, etc.), even with better environmental data, predicting risk is questionable.

There are several broad issues outside this study further complicating results and interpretations. The first of which, is recognizing that this study focused on human cases alone. Without information on mice populations in the study regions over time, stronger correlations between SNV, humans, and the environment are impossible to make. Though clusters of disease in humans may exist as hot-spot areas and there may be significant environmental differences between areas included within a hot spot as compared to neighboring regions not in a hot spot, the mechanism or series of necessary inputs cannot be fully described with inclusion of deer mouse characteristics.

All information on mice populations – increases or decreases in an area’s population, degree of connectivity (Island and Metapopulation Theory) between mice populations, increases or decreases in Hantavirus prevalence – is missing and seriously limits understanding the complex relationship between mice, virus, and humans. Trapping mice populations and testing Hantavirus prevalence should ideally be done multiple times over many years. Additionally, testing should occur at the same time each year because mice life spans are short and antibody prevalence in mouse samples vary considerably in small time spans, from less than 5% to over 60% in the same areas over the course of a year [13]. A lack of funding and time are severe restrictions in the collection of these data as well as space.

Recommendations

HPS causes severe, sometimes fatal, respiratory distress in humans. This study, describing the spatial patterns of human disease, together with literature describing behaviors and influencers on deer mouse populations, is important in predicting future disease risk for human populations. While examining environmental features is important in realizing large-scale ecological patterns affecting deer mouse populations, it is not enough to establish risk for humans. The human behavior component of this relationship is rarely defined, at least rarely recorded with other case information in the HPS registry. Gathering information on the activities that may have led to exposure might yield further insights into human risk prediction, important to Public Health organizations (though admittedly, more difficult to obtain)

Additionally, strengthening communication and research efforts between public health officials, scientists, and land owners is important. As established, deer mouse populations span most the North American continent. Land included in this area is owned by multiple sources including Federal governments, U.S. State governments, and private citizens. Citation of diagnosed cases may be mandated as notifiable by the Federal government and CDC, but management of land practices is left up to this myriad of supervisors. Given the connectivity of landscapes despite administrative boundaries, it is important for neighboring land managers to engage in communication increasing awareness of any trends or observations of changes to deer mouse populations or HPS occurrence in humans.

Of course, actually creating techniques to implement this sort of management is easier in theory then in practice and more data would be needed on all rodent populations in a chosen landscape. Scale is important as discussed through Island Biogeography and Metapopulations, and choosing the appropriate scale would be an additional hardship. Finally, limitations relate back to people and the complications imposed by different laws and practices at the local, regional, state, and federal level.

Time is likely to have an important role in the Hantavirus-Deer Mouse-Human model but was not able to be incorporated into this analysis. Because HPS is such a rare disease, comparing incidence or counts by year is difficult, but analyzing the documented first day of symptoms recorded in the HPS registry and extrapolating back to an estimated time of exposure could yield important information on seasonality of HPS disease in humans. It should be noted, however, that this might reflect human behavior rather than changes in Hantavirus prevalence or infectivity in mouse populations. This also would need to be considered in future research.

Another interesting application of this study for further investigation comes from Figure 28. It is quite apparent that cases appear to follow water flow length (drainage) starting in the in upper eastern Sierra Nevada Mountains coming south and then curving to the coast. Though data on water cover in California was not included in this study, it appears this drainage system created from elevation data follows a real river system: The Owens River which drains to Owens Valley. Further, this system is the source for the

Los Angeles Aqueduct that starts near the source of Owens River and flows the same path down the Eastern Sierra’s edge curving west towards the coast. Perhaps this area offers increased connectivity to deer mice populations enabling transmission of SNV between populations. Or perhaps, given the intensity of human activities on the landscape, deer mice have been able to flourish due to their ability to adapt as generalist species. It would be worth investigating deer mice populations and the prevalence of SNV along this drainage systems and aqueduct.

All results, except for the focused analysis of the Sierra Nevada Mountains included the 2012 Yosemite outbreak. One potential next step would be conducting the entire spatial data analysis again, but without inclusion of the 2012 Yosemite outbreak. Likely, significant outliers through the LISA test and hot-spot regions from the Getis G*i(d) would be smaller and may reflect distribution patterns of cases by ZCTAs.

Despite the concerns listed about, this study was able to describe spatial patterns in human disease that can be added to the body of knowledge surrounding deer mouse populations and key environmental variables. It is at the intersection between human case occurrence and deer mouse ecology where better risk prediction can occur.

Conclusions

The relationships among and between Hantavirus infection and their host, *P.maniculatus* together with environmental conditions and human interactions are complex and difficult to predict. Various approaches were used to illustrate spatial relationships between human HPS cases to ZIP Codes and atop environmental differences. Significant results suggest clustering of cases occurs in and around specific environmental features, with the Sierra Nevada Mountains as an example. While understanding the necessary environmental characteristics present in previous cases and outbreaks from Washington, Oregon, California, and Nevada, examining the role of ecological islands, patches, fragmentation, together with human interactions are equally important in understanding Hantavirus infection dynamics. Only then can the ultimate goal be accomplished: development of more sensitive and accurate models predicting areas of high HPS risk for humans.

Tables and Figures

Figure 1: Deer mouse habitat in North America

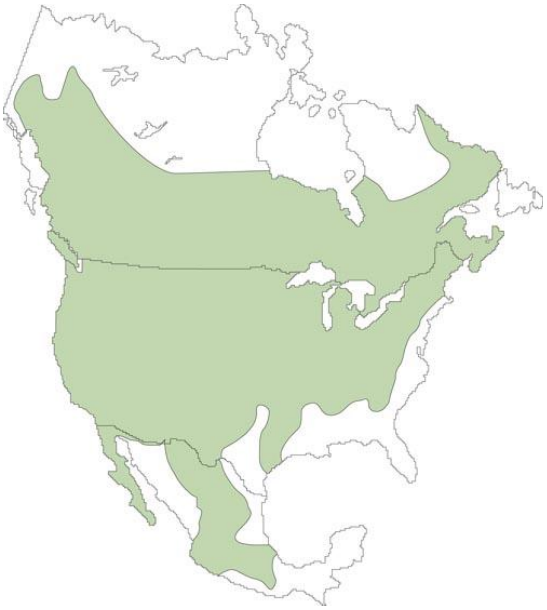


Figure 2: Diagnosed HPS cases in the United States mapped by states of exposure

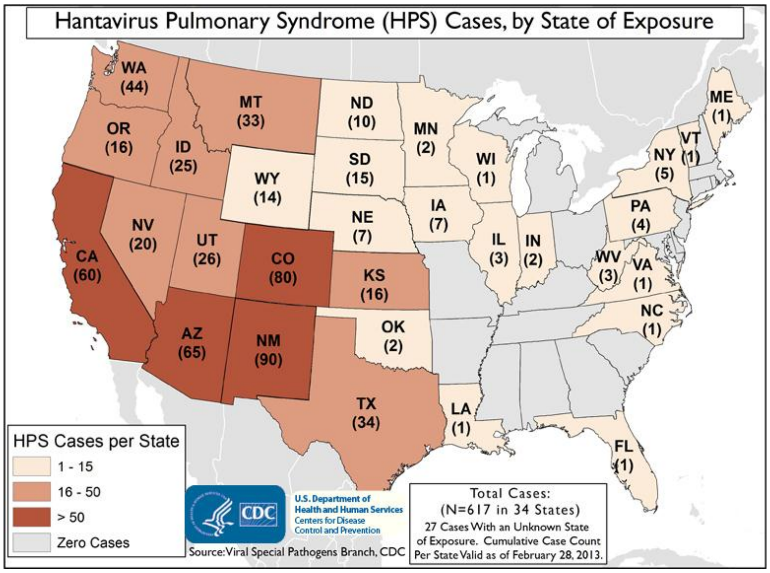

Hantavirus Pulmonary Syndrome (HPS) Cases, by State of Exposure
WA (44)
OR (16)
ID (25)
MT (33)
ND (10)
MN (2)
ME (1)
VT (1)
NY (5)
WI (1)
SD (15)
WY (14)
CA (60)
NV (20)
UT (26)
CO (80)
NE (7)
IA (7)
IL (3)
IN (2)
PA (4)
WV (3)
VA (1)
KS (16)
NC (1)
AZ (65)
NM (90)
OK (2)
TX (34)
LA (1)
FL (1)
HPS Cases per State
1 - 15
16 - 50
> 50
Zero Cases
CDC
U.S. Department of Health and Human Services
Centers for Disease Control and Prevention
Source: Viral Special Pathogens Branch, CDC
Total Cases: (N=617 in 34 States)
27 Cases With an Unknown State of Exposure. Cumulative Case Count Per State Valid as of February 28, 2013.

Figure 3: Percentage of HPS occurrence by State over Study Area

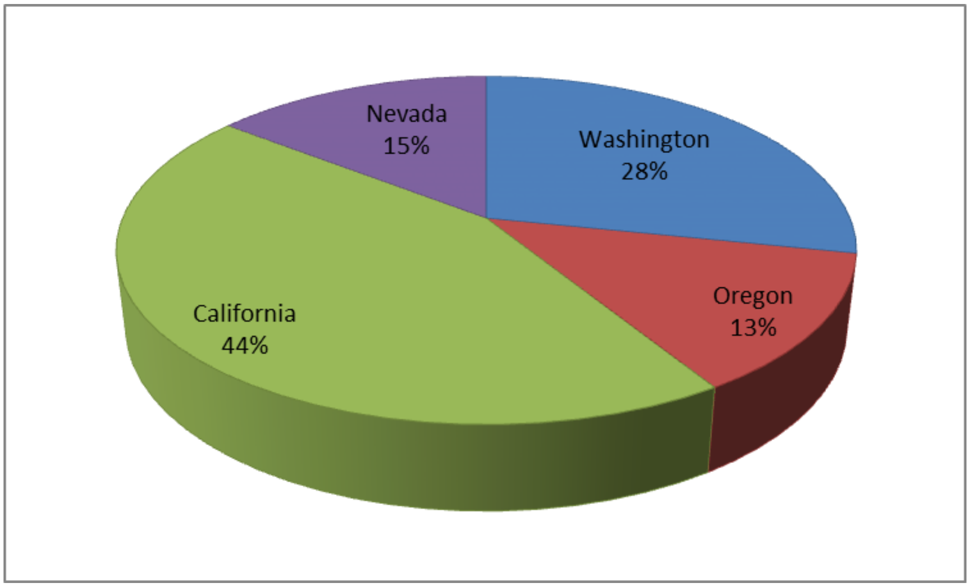


Table 1: Frequency and Percentage Statistics at the COUNTY level

State	HPS Count	Number of HPS Counties	Total Counties	HPS Counties / State Counties (Percent)	HPS Counties/ Total (Percent)
Washington	35	17	39	43.58974359	26
Oregon	16	12	36	33.33333333	24
California	55	20	58	34.48275862	38.6666667
Nevada	18	10	17	58.82352941	11.3333333
TOTAL	124	59	150	39.33333333	1

Table 2: Frequency and Percentage Statistics at ZCTA level

State	HPS Count	Number of ZCTAs reporting HPS	Total ZCTAs	Percent of positive ZCTA within States	Percent of positive ZCTA to Total
Washington	35	32	598	5.351170569	20.1890614
Oregon	16	14	419	3.341288783	14.1458474
California	55	30	1769	1.695873375	59.72316
Nevada	18	15	176	8.522727273	5.9419311
TOTAL	124	91	2962	3.072248481	1

Figure 4: HPS cases within Counties

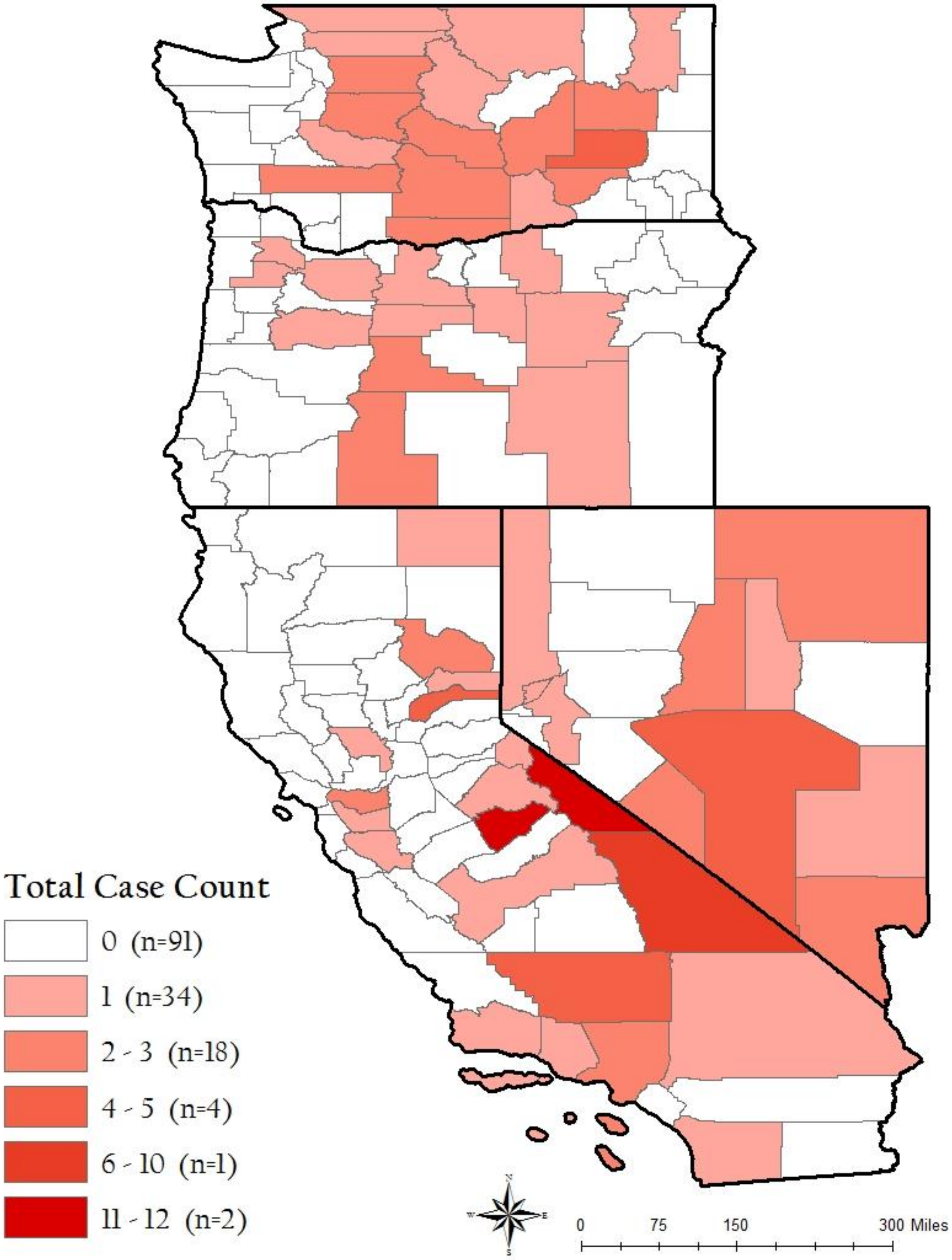


Figure 5: HPS Crude Incidence by County

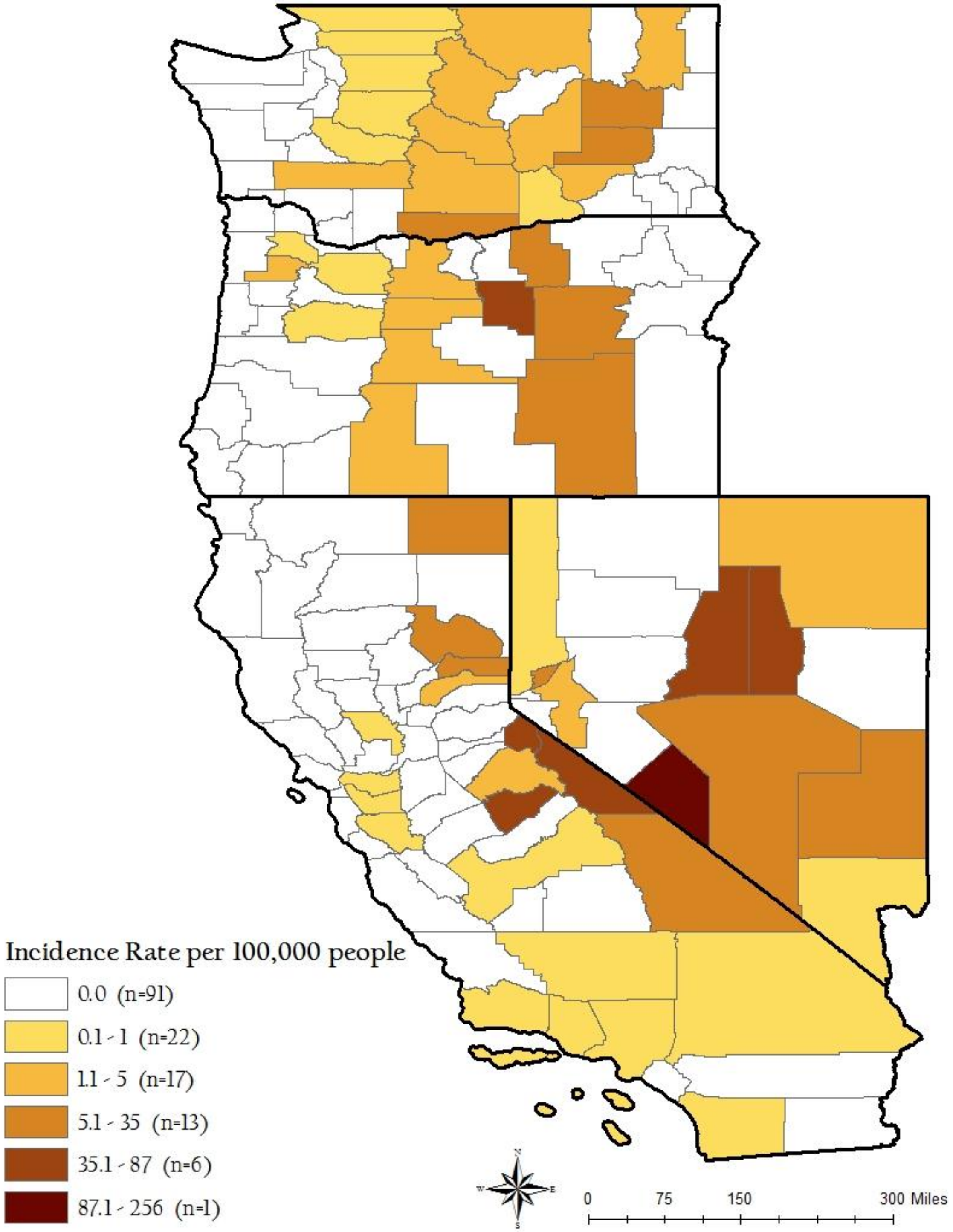


Figure 6: HPS Case Count by ZCTA

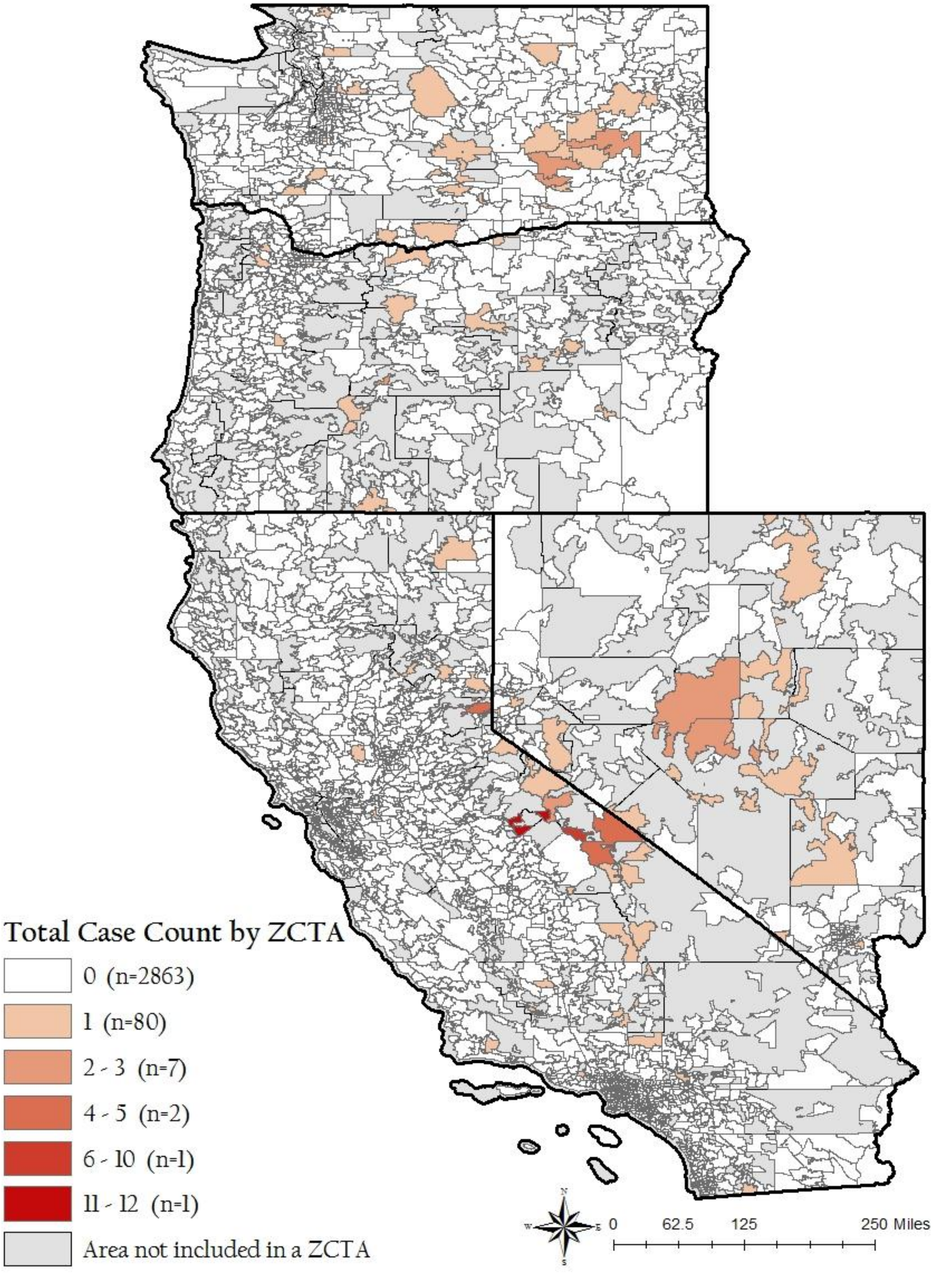


Figure 6a – Washington Cases by ZCTA

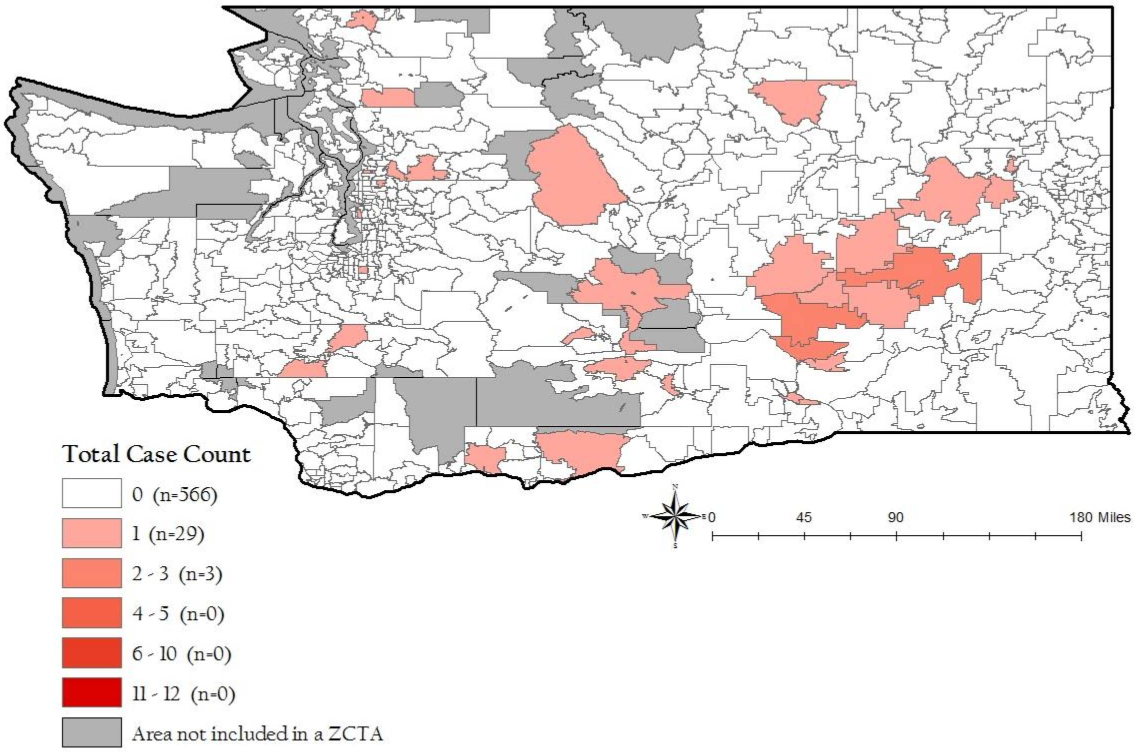


Figure 6b. – Oregon Cases by ZCTA

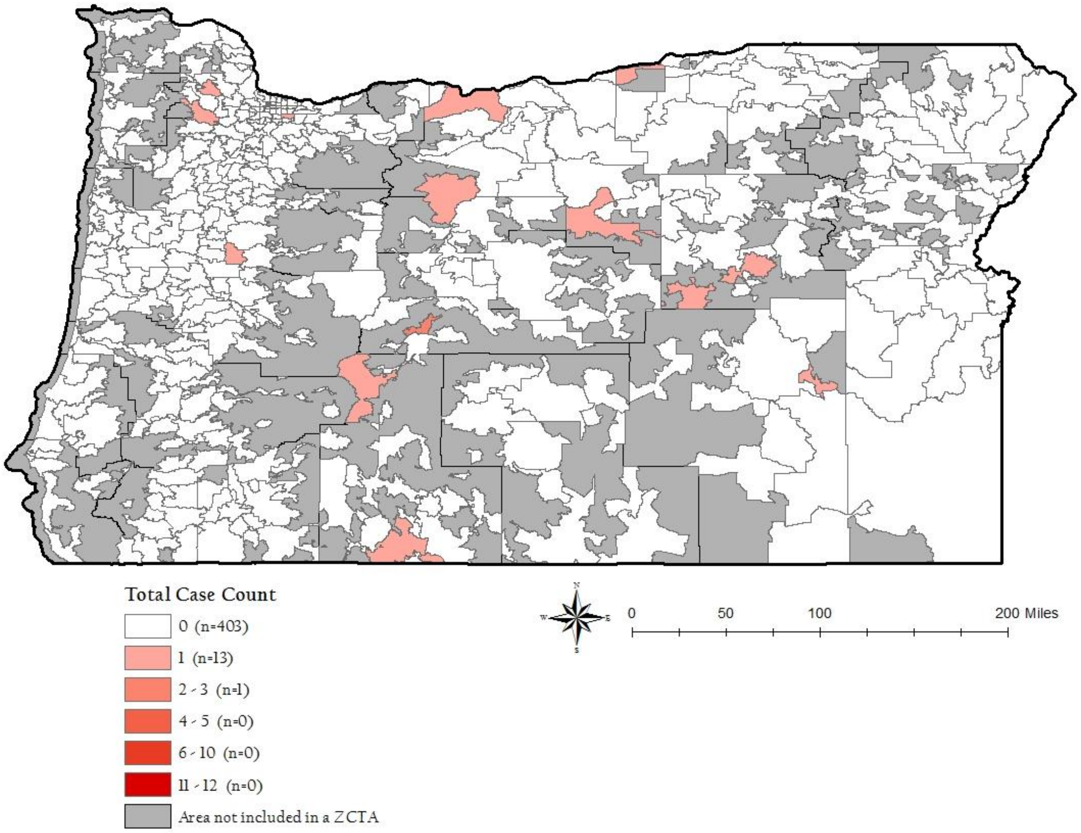


Figure 6c. - California Cases by ZCTA

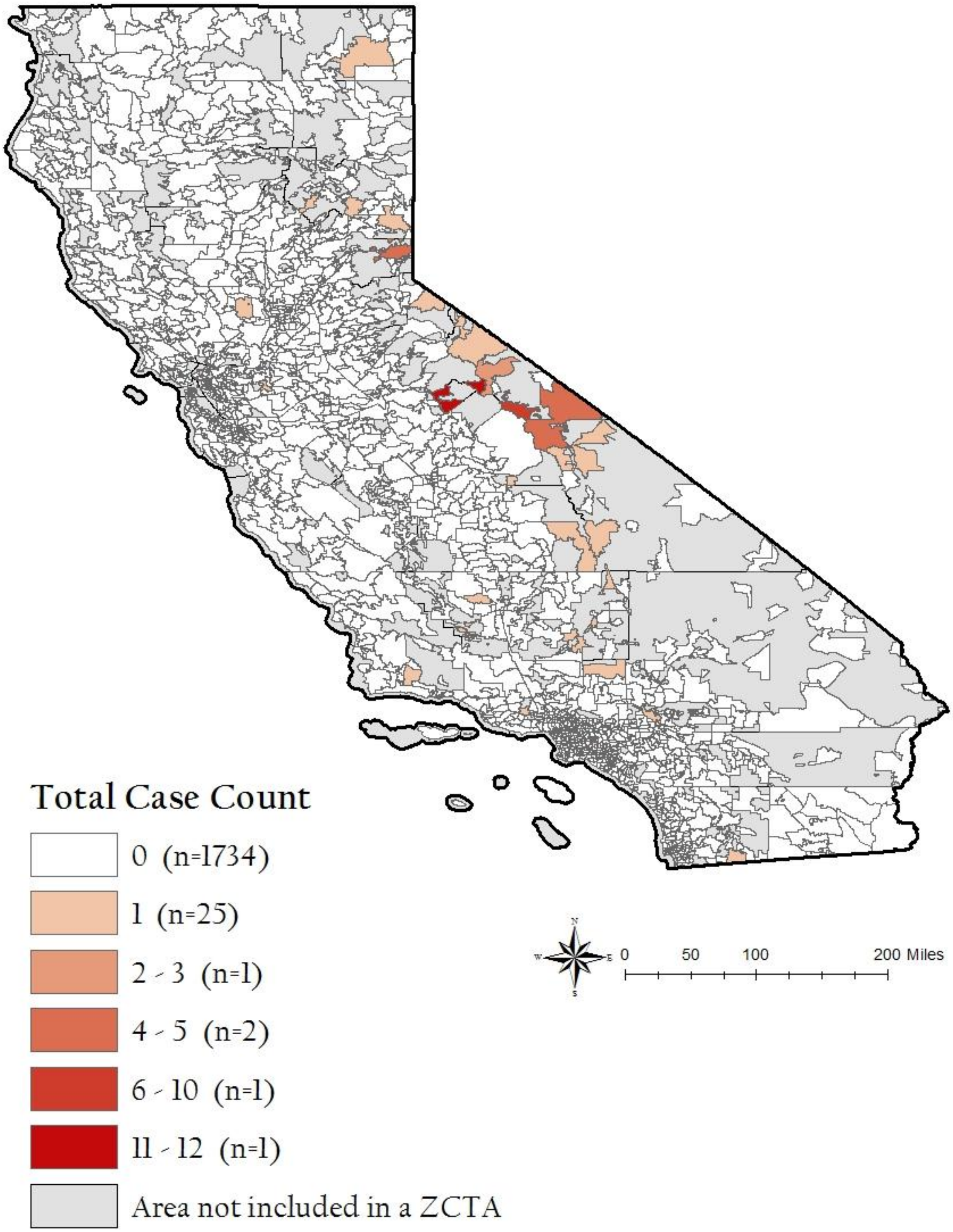


Figure 6d. – Nevada Cases by ZCTA

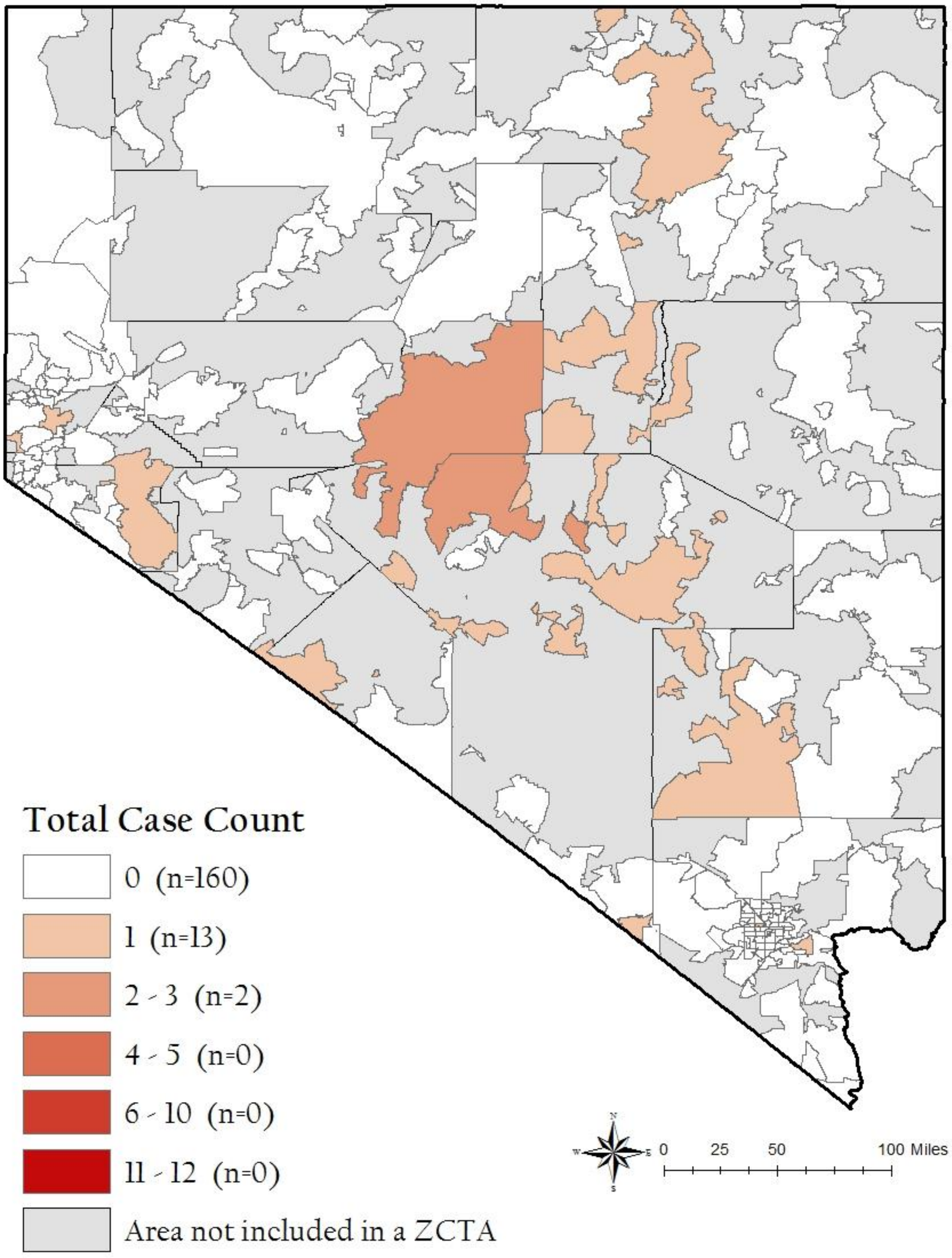


Figure 7: HPS Crude Incidence by ZCTA

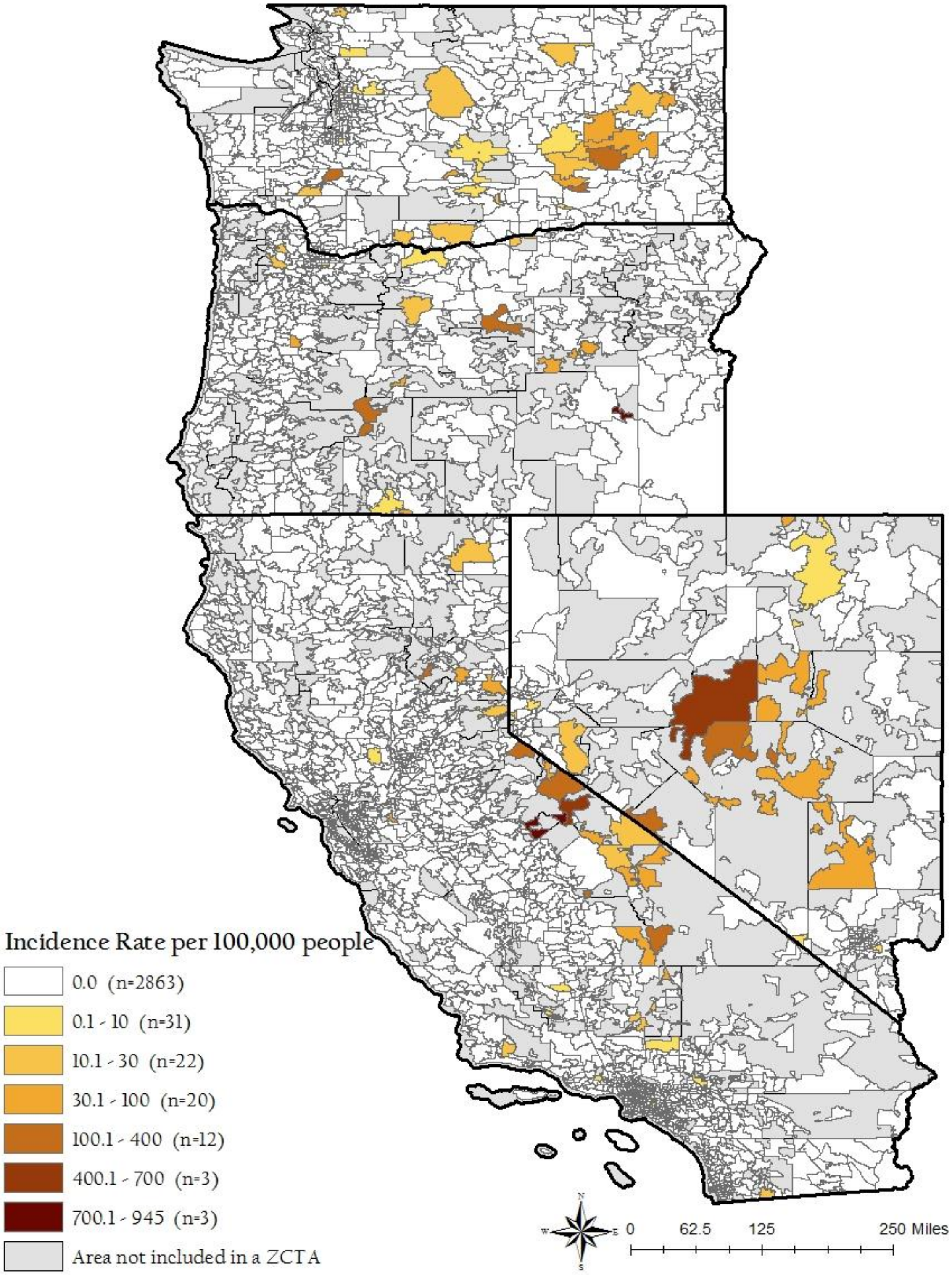


Figure 7a – Washington Crude Incidence by County

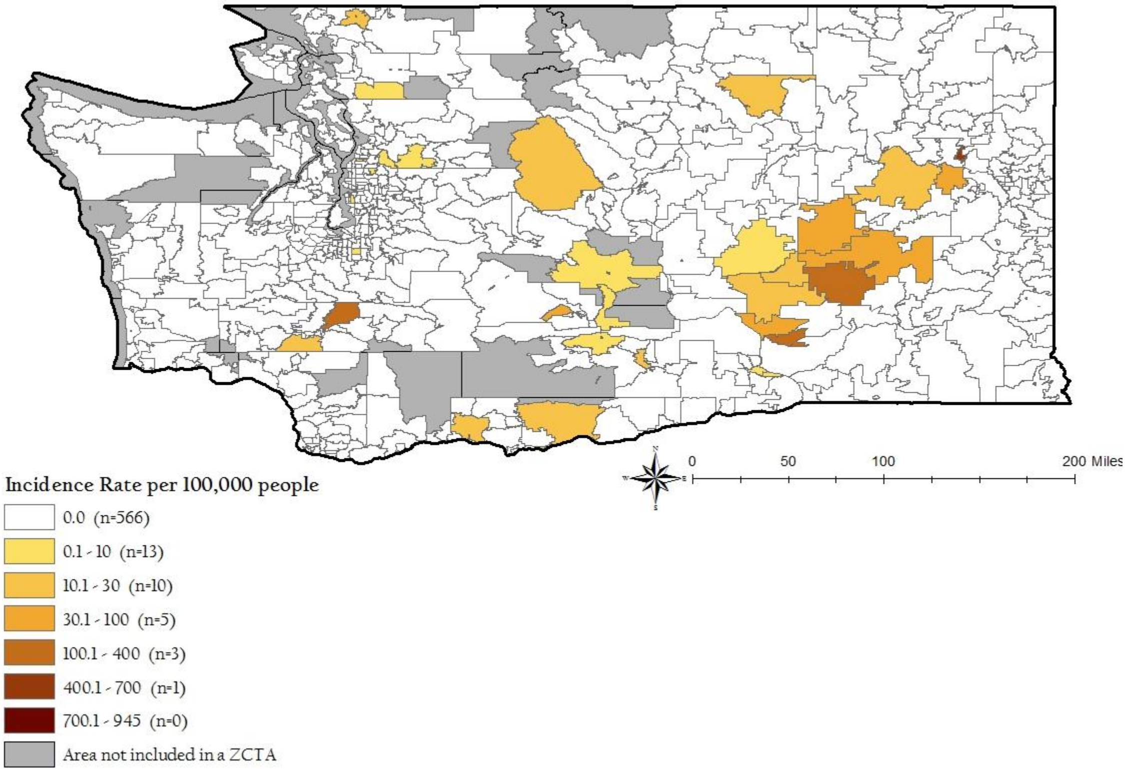


Figure 7b – Oregon Crude Incidence by County

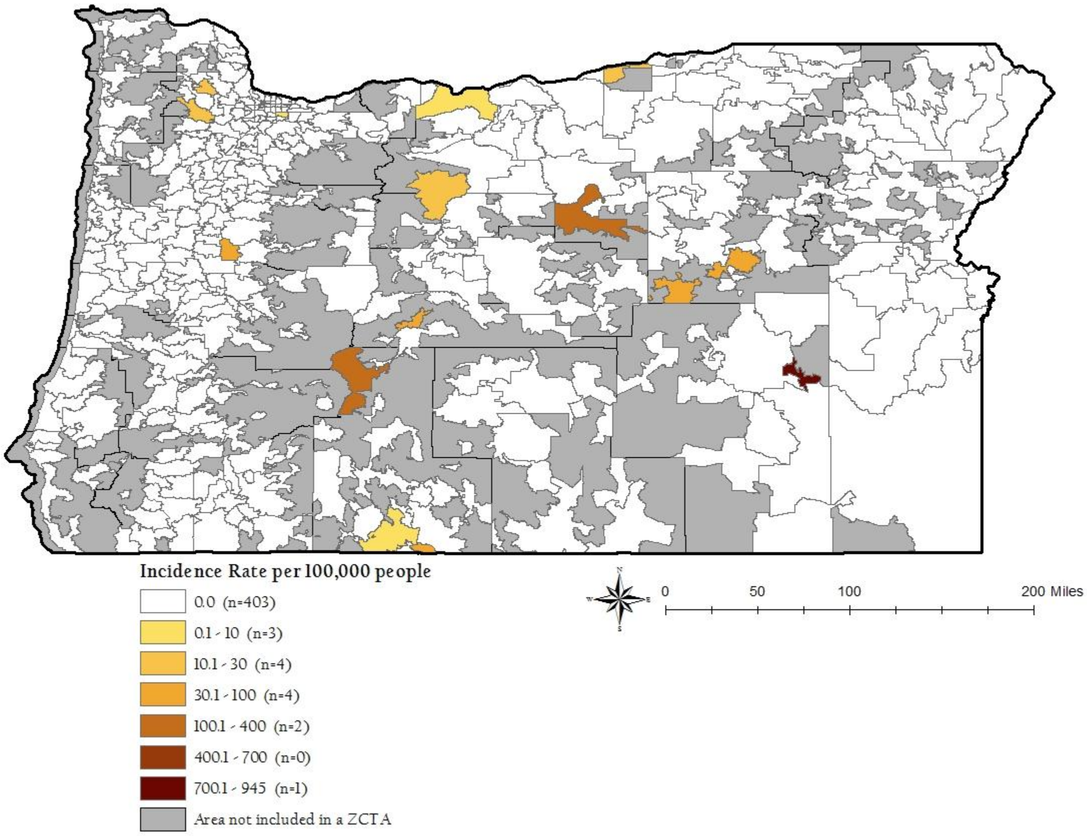


Figure 7c – California Crude Incidence by County

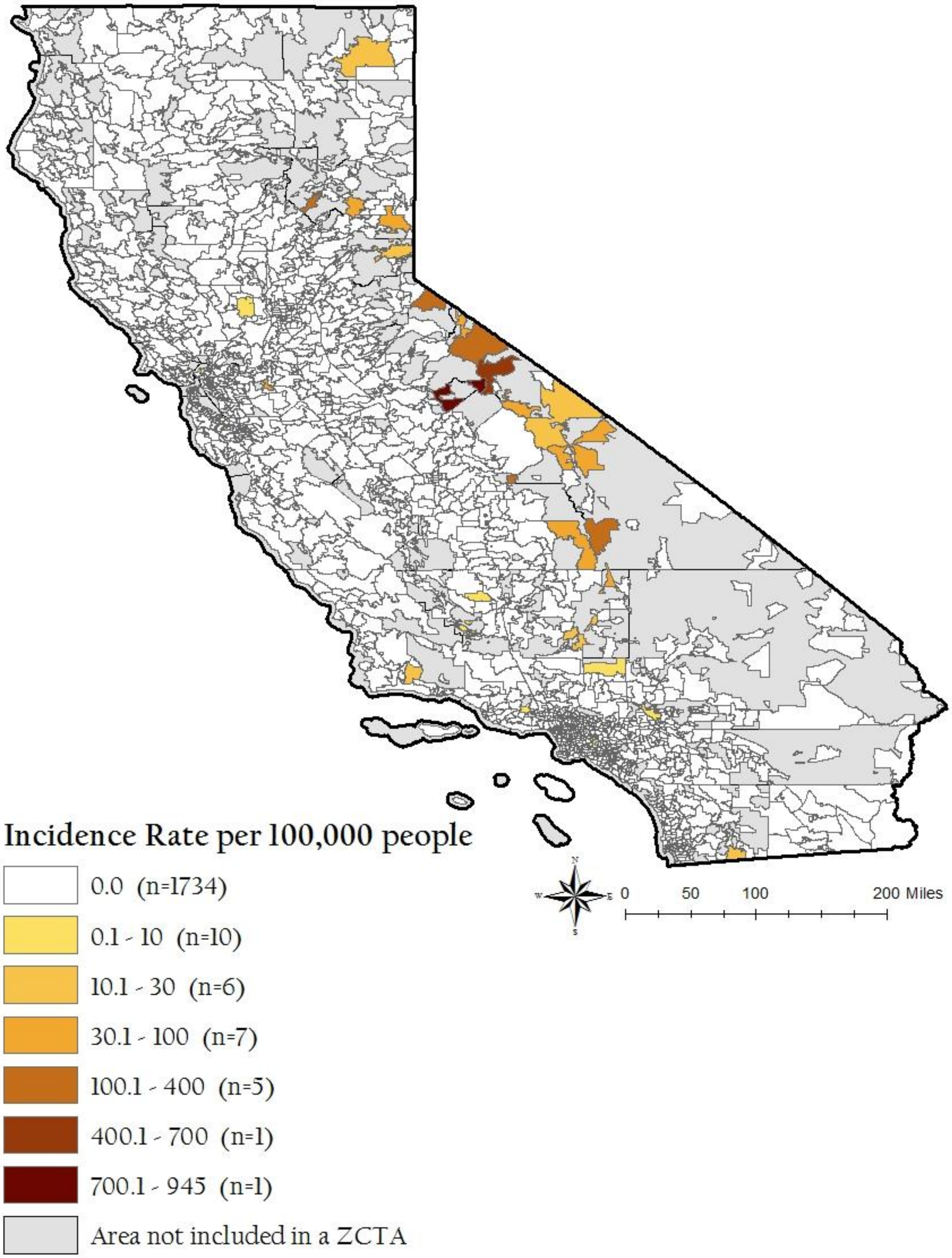


Figure 7d – Nevada Crude Incidence Rate by County

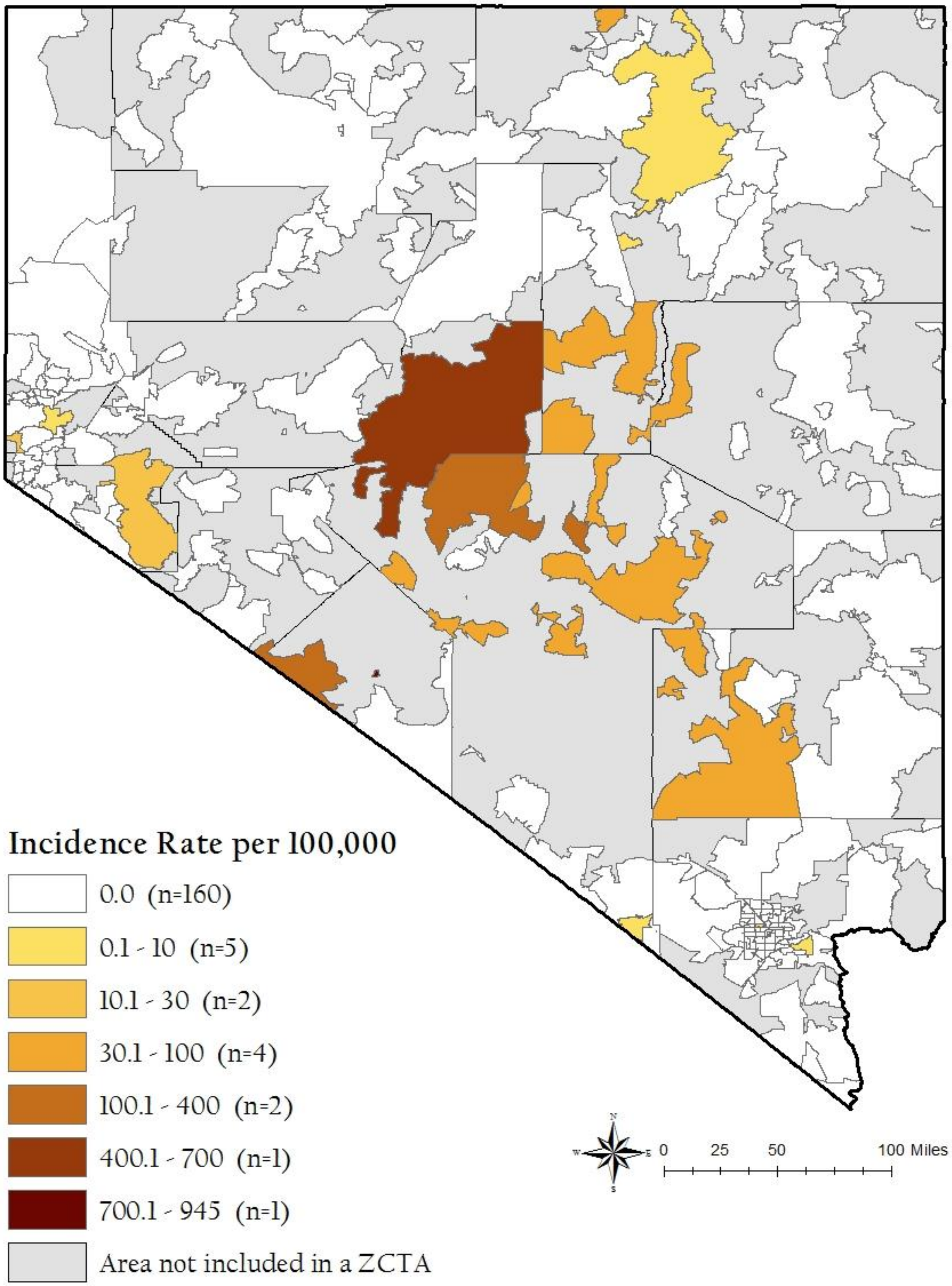


Figure 8: Voronoi Map of ZCTA Centroids and Case Counts

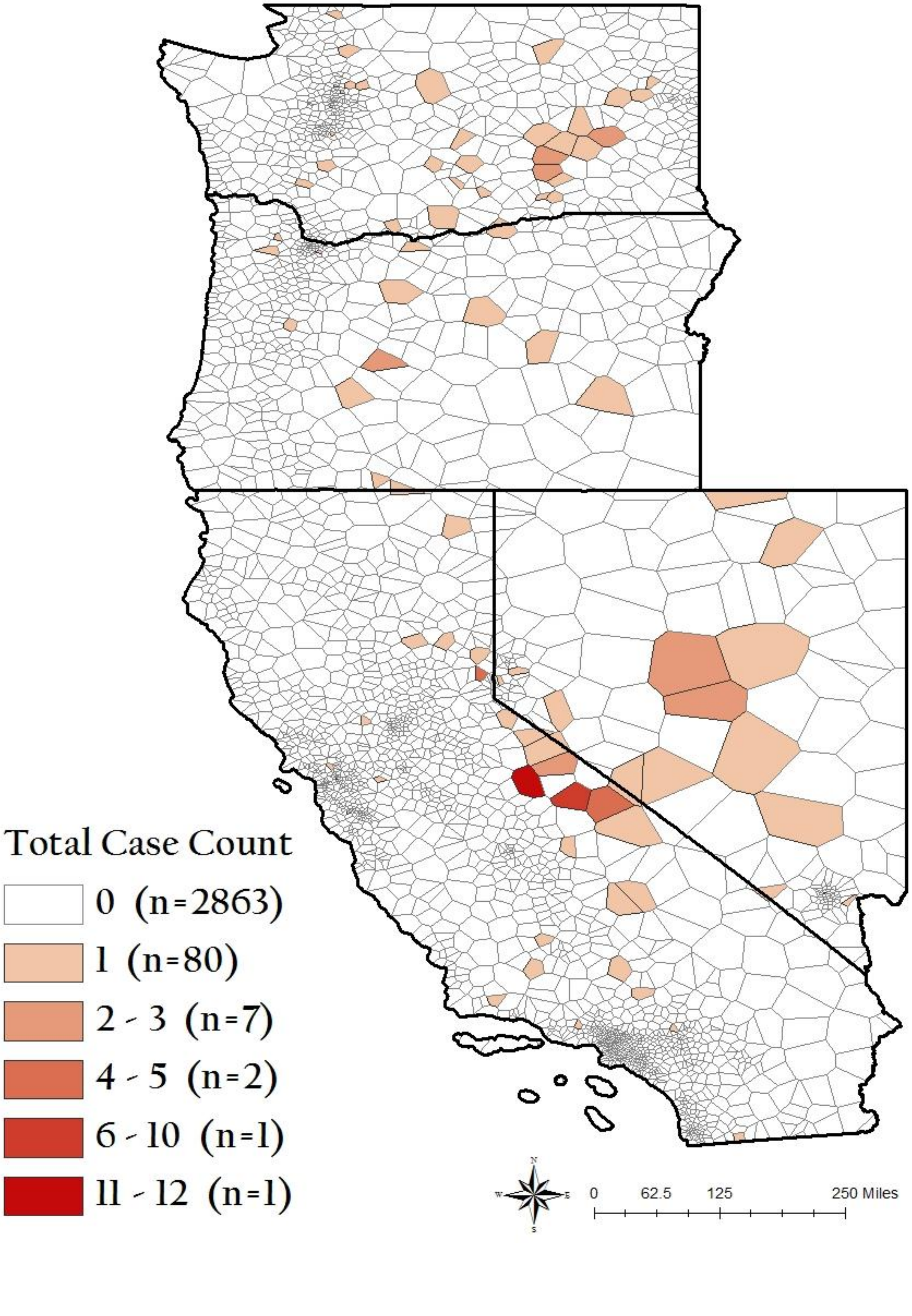


Figure 9: Voronoi Map of ZCTA Centroids and Incidence Rate

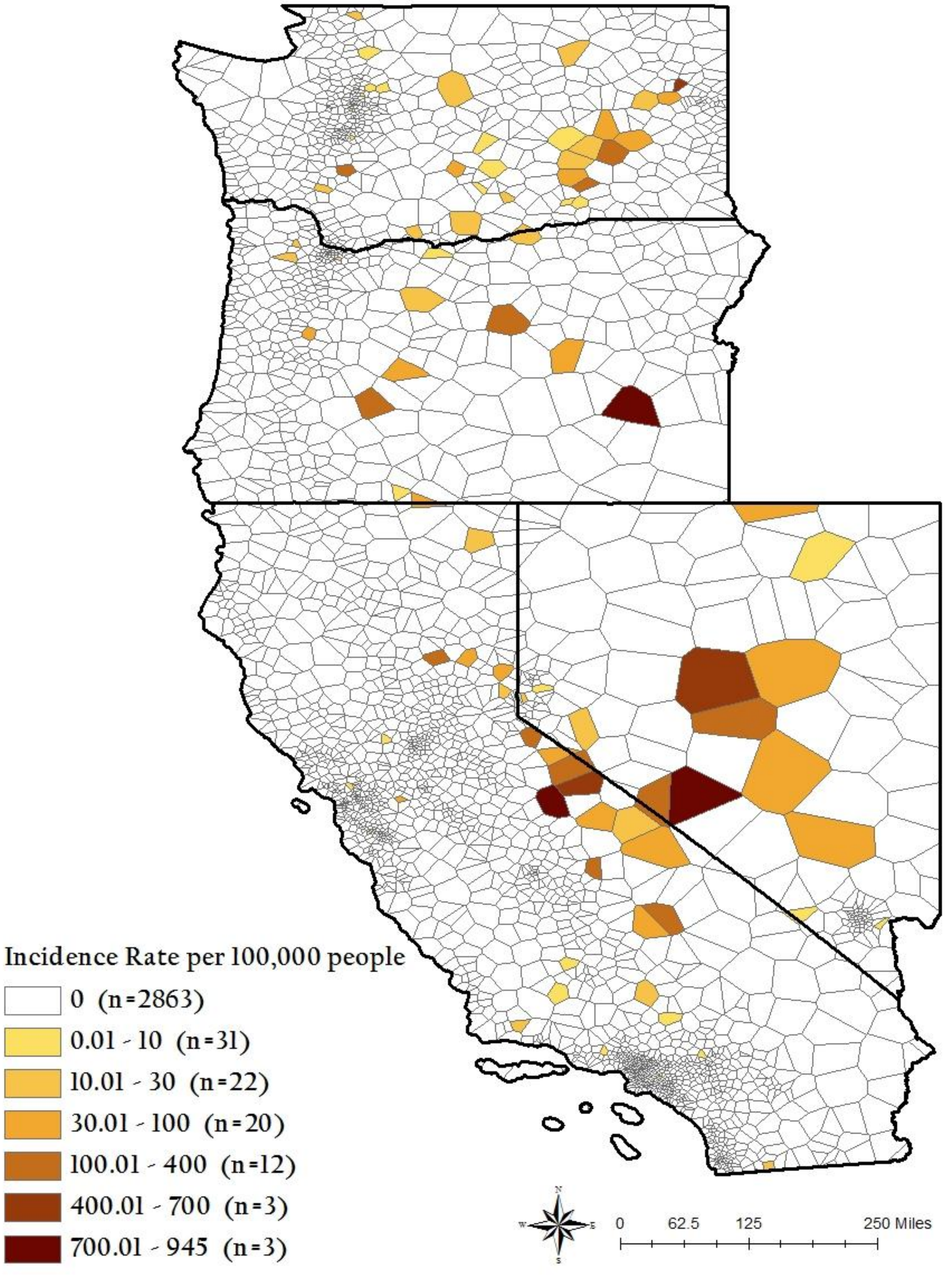


Figure 10a: HPS Cases in ZCTA across USGS Ecological Zones

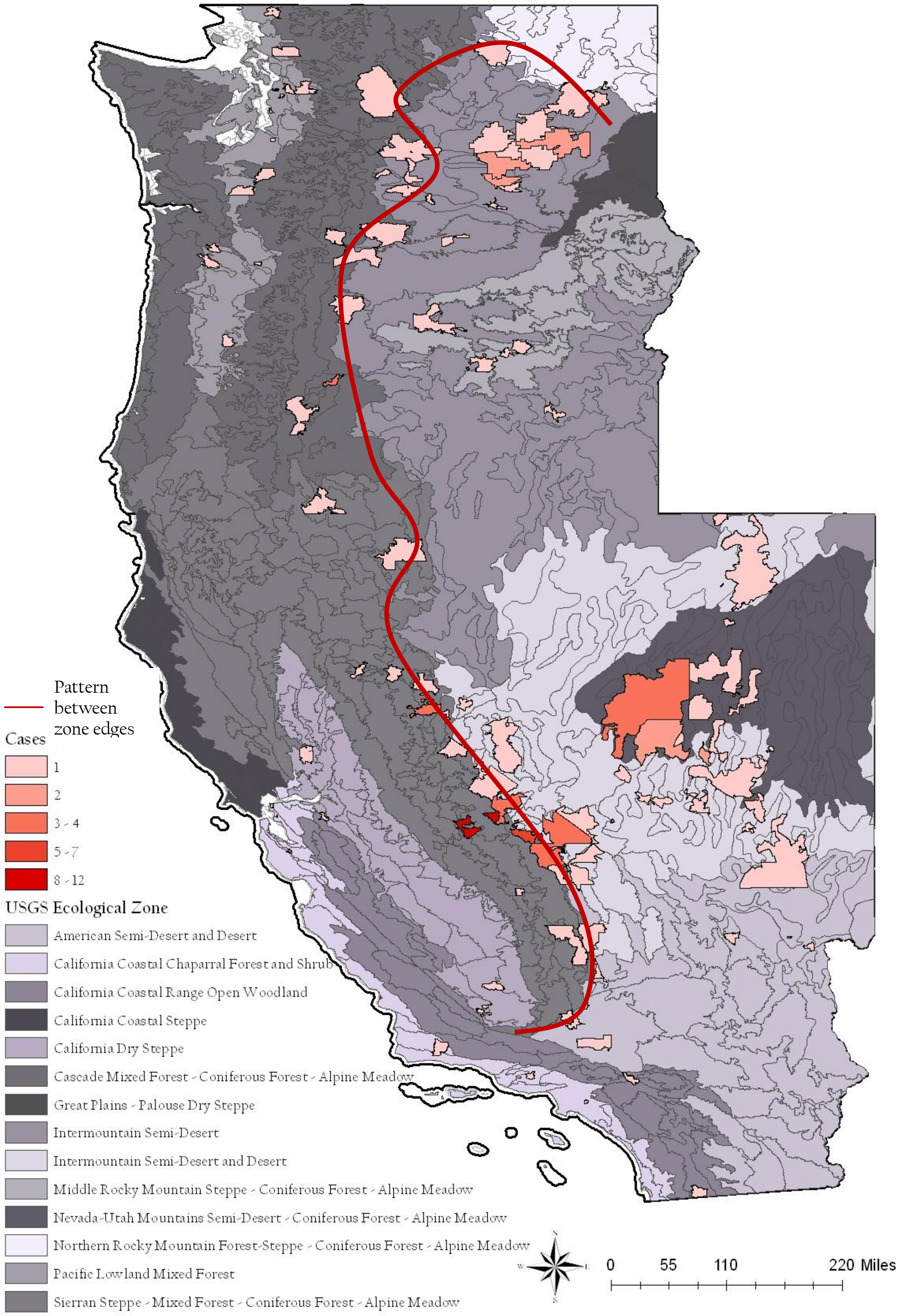

Pattern between zone edges
Cases
1
2
3 - 4
5 - 7
8 - 12
USGS Ecological Zone
American Semi-Desert and Desert
California Coastal Chaparral Forest and Shrub
California Coastal Range Open Woodland
California Coastal Steppe
California Dry Steppe
Cascade Mixed Forest - Coniferous Forest - Alpine Meadow
Great Plains - Palouse Dry Steppe
Intermountain Semi-Desert
Intermountain Semi-Desert and Desert
Middle Rocky Mountain Steppe - Coniferous Forest - Alpine Meadow
Nevada-Utah Mountains Semi-Desert - Coniferous Forest - Alpine Meadow
Northern Rocky Mountain Forest-Steppe - Coniferous Forest - Alpine Meadow
Pacific Lowland Mixed Forest
Sierran Steppe - Mixed Forest - Coniferous Forest - Alpine Meadow
N
W
E
S
0
55
110
220 Miles

Figure 10b: HPS Incidence Rate in ZCTA across USGS Ecological Zones

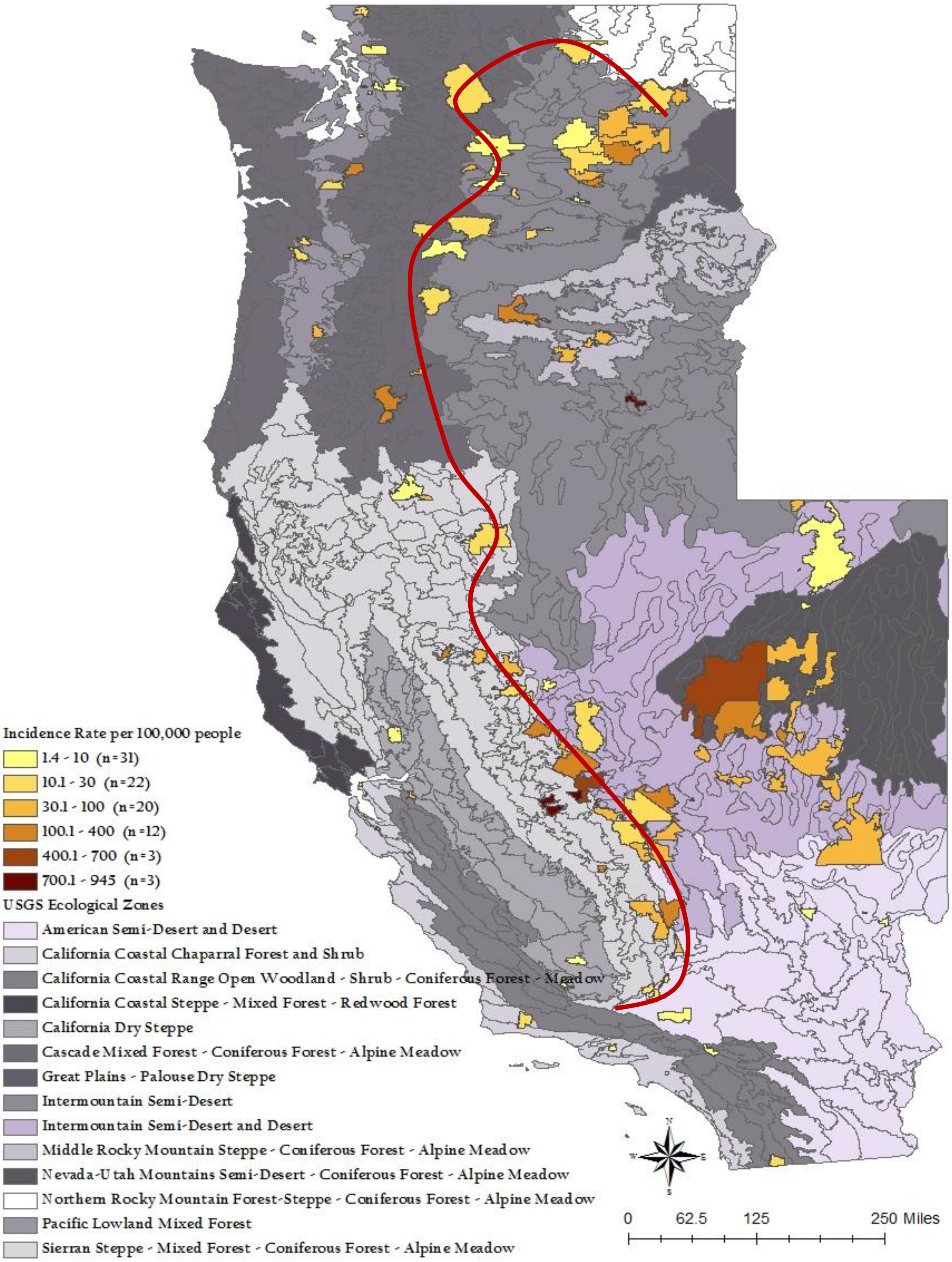

Incidence Rate per 100,000 people
1.4 - 10 (n=31)
10.1 - 30 (n=22)
30.1 - 100 (n=20)
100.1 - 400 (n=12)
400.1 - 700 (n=3)
700.1 - 945 (n=3)
USGS Ecological Zones
American Semi-Desert and Desert
California Coastal Chaparral Forest and Shrub
California Coastal Range Open Woodland - Shrub - Coniferous Forest - Meadow
California Coastal Steppe - Mixed Forest - Redwood Forest
California Dry Steppe
Cascade Mixed Forest - Coniferous Forest - Alpine Meadow
Great Plains - Palouse Dry Steppe
Intermountain Semi-Desert
Intermountain Semi-Desert and Desert
Middle Rocky Mountain Steppe - Coniferous Forest - Alpine Meadow
Nevada-Utah Mountains Semi-Desert - Coniferous Forest - Alpine Meadow
Northern Rocky Mountain Forest-Steppe - Coniferous Forest - Alpine Meadow
Pacific Lowland Mixed Forest
Sierran Steppe - Mixed Forest - Coniferous Forest - Alpine Meadow
N
W
E
S
0
62.5
125
250 Miles

Figure 11a: HPS Cases in ZCTAs within Watershed Areas

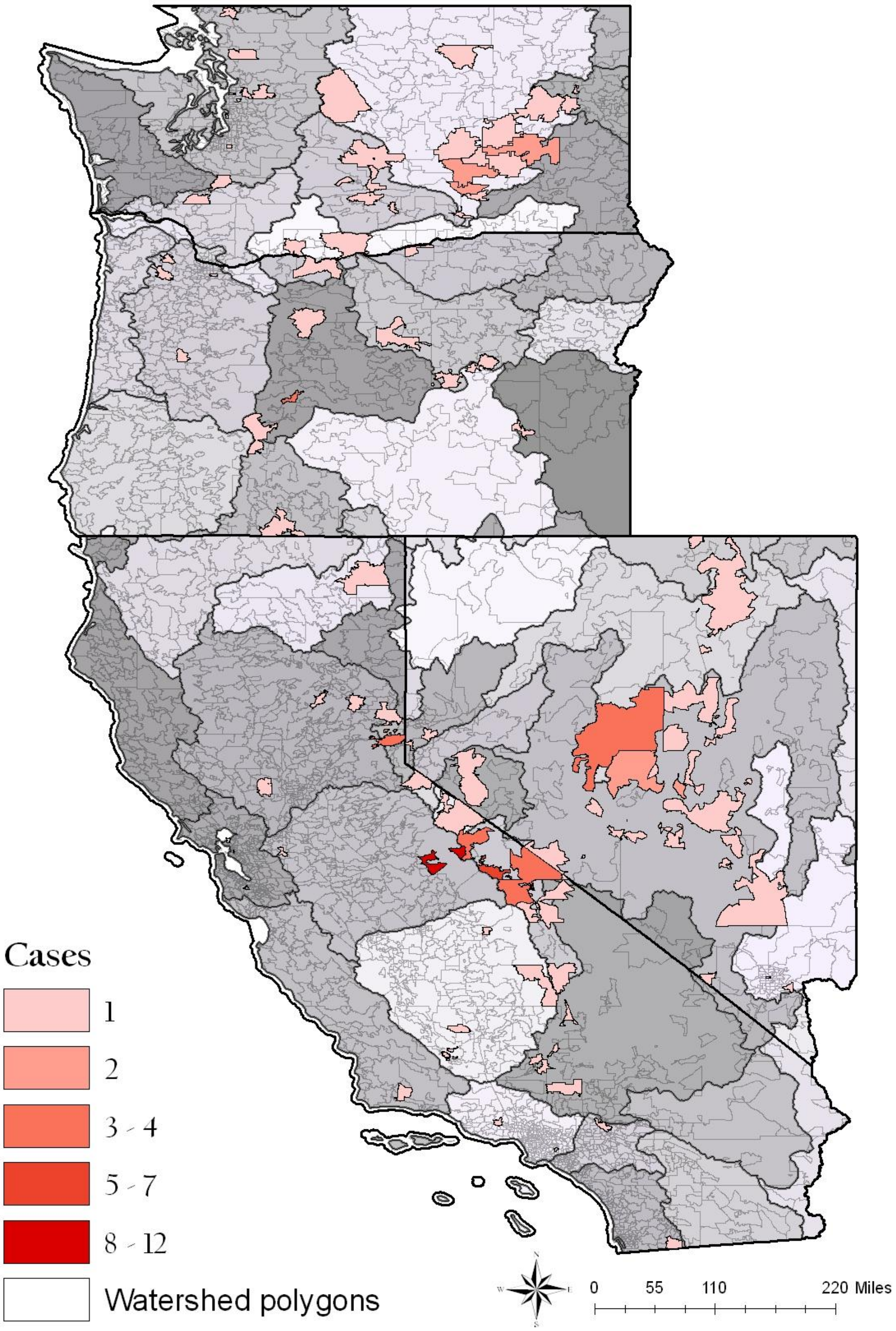


Figure 11b: HPS Incidence Rate in ZCTA within Watershed Areas

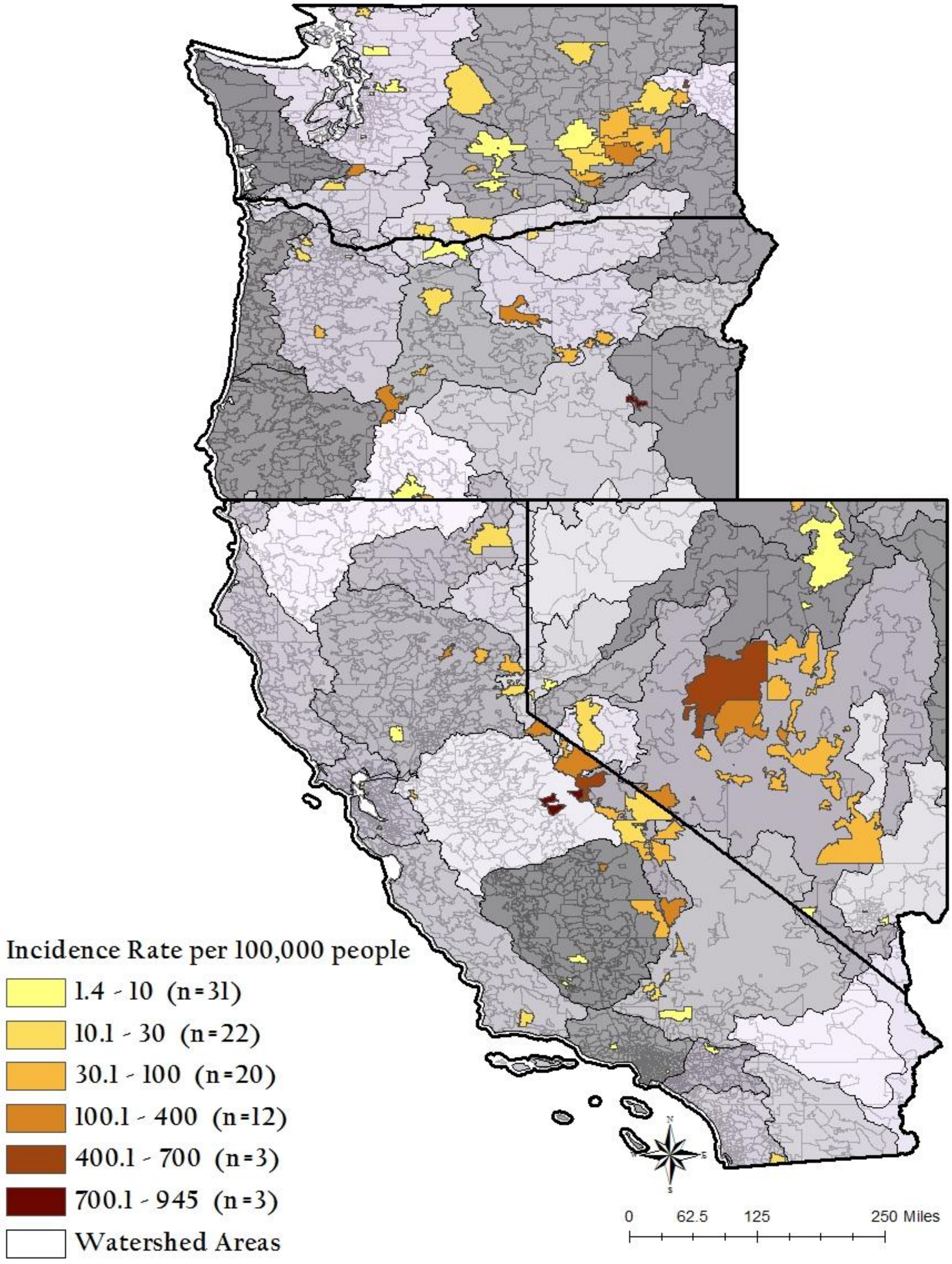


Figure 12a: HPS Cases in ZCTAs across Elevation

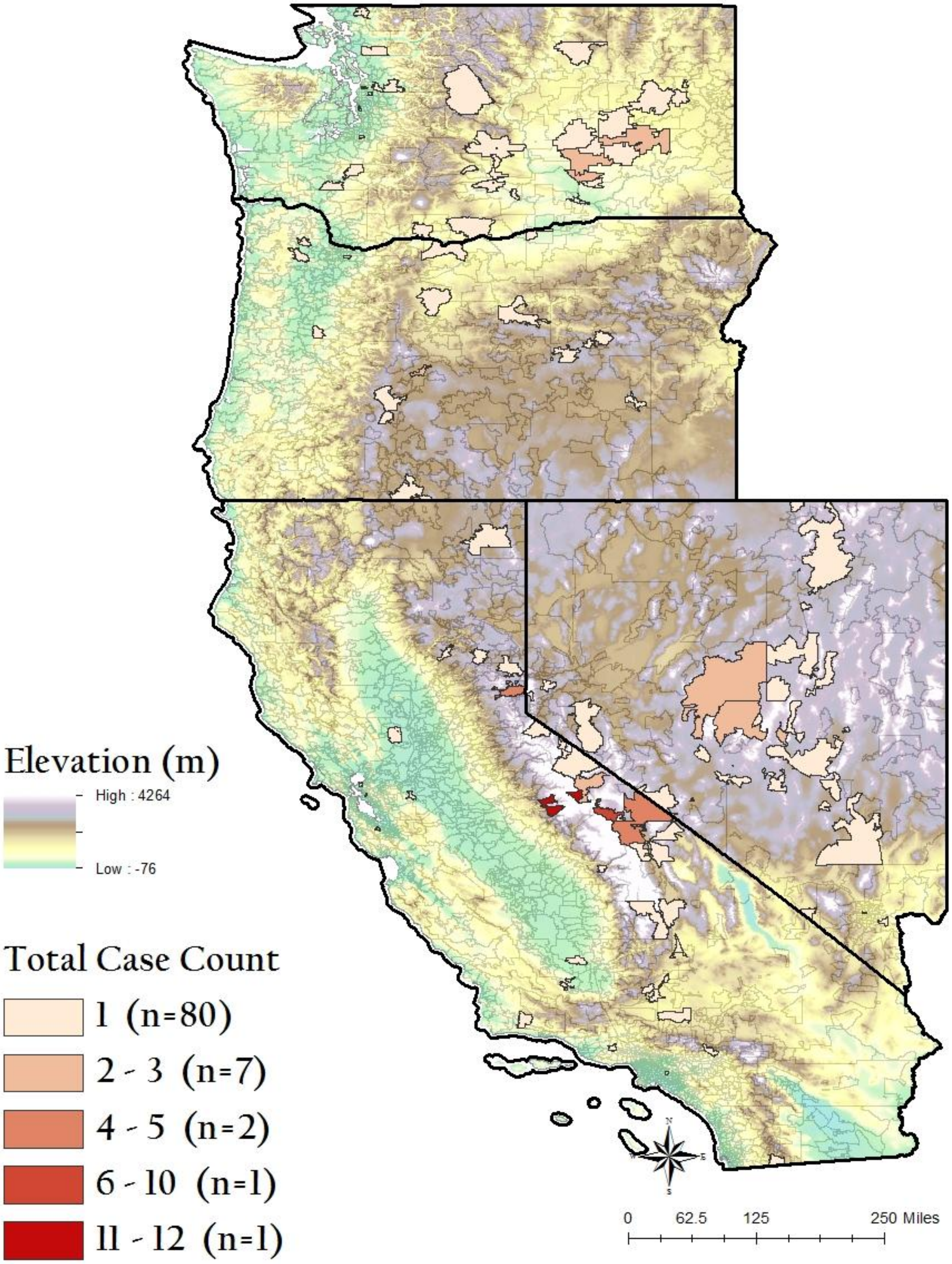


Figure 12b: HPS Incidence Rate in ZCTAs across Elevation

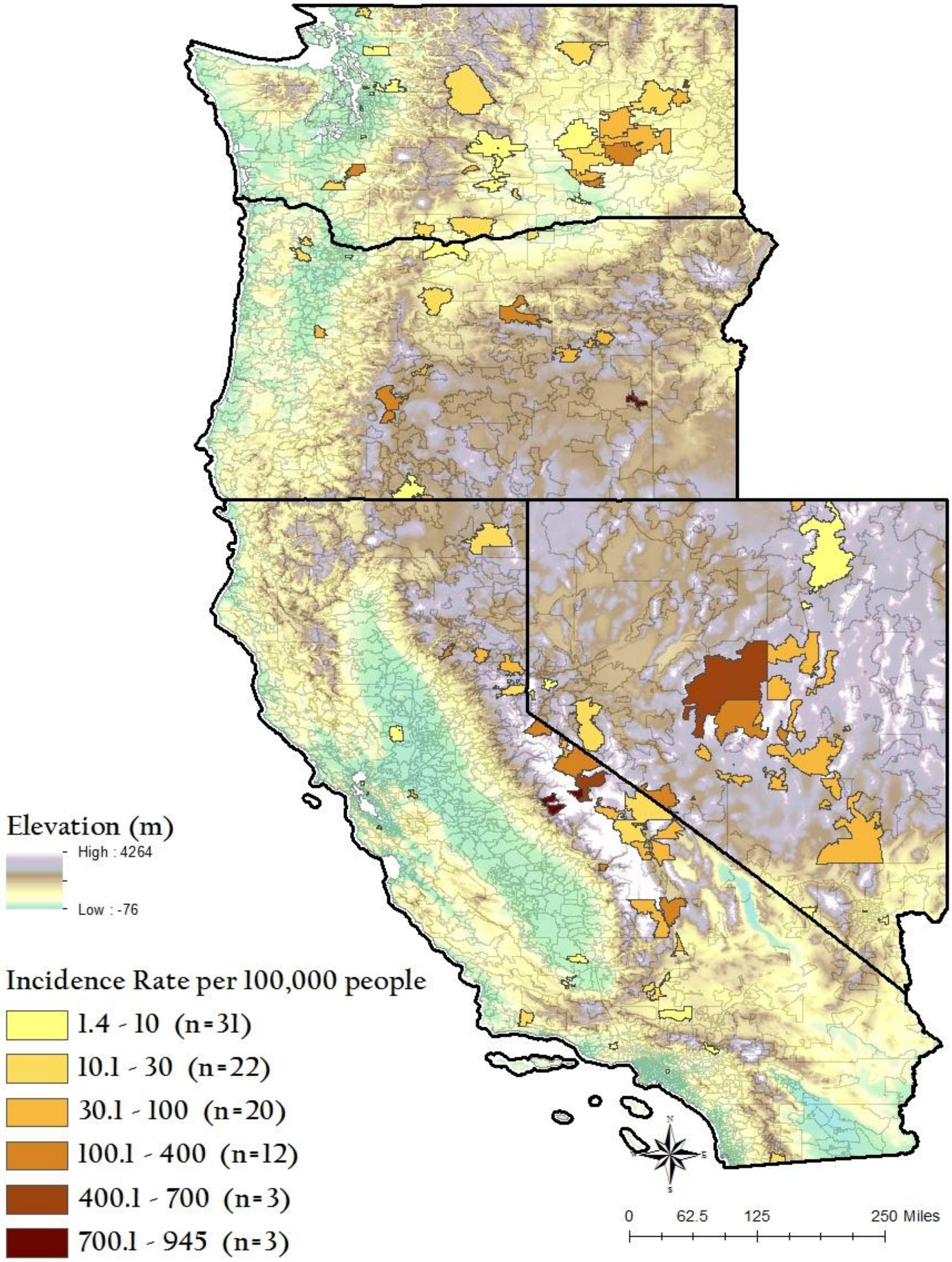
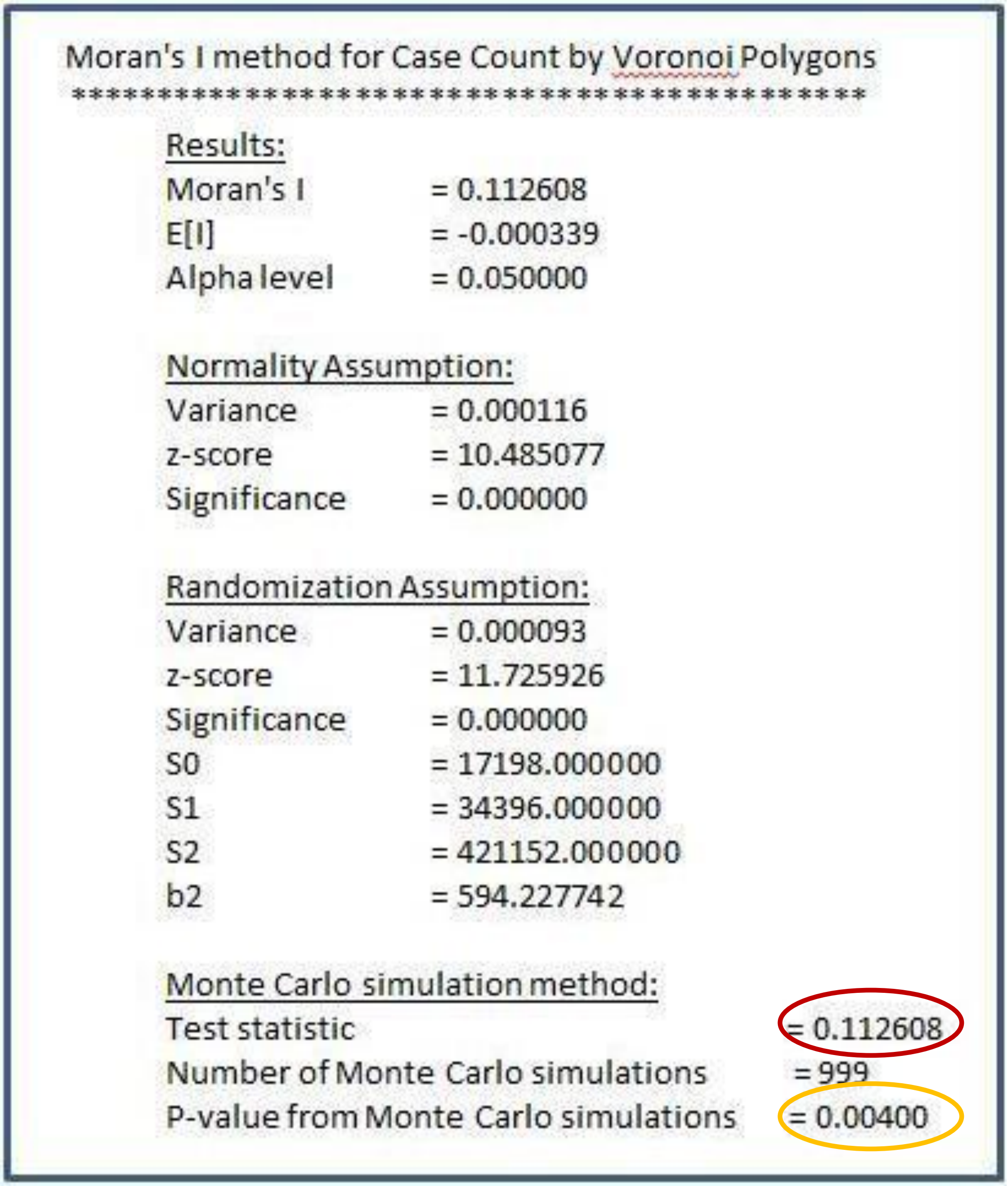
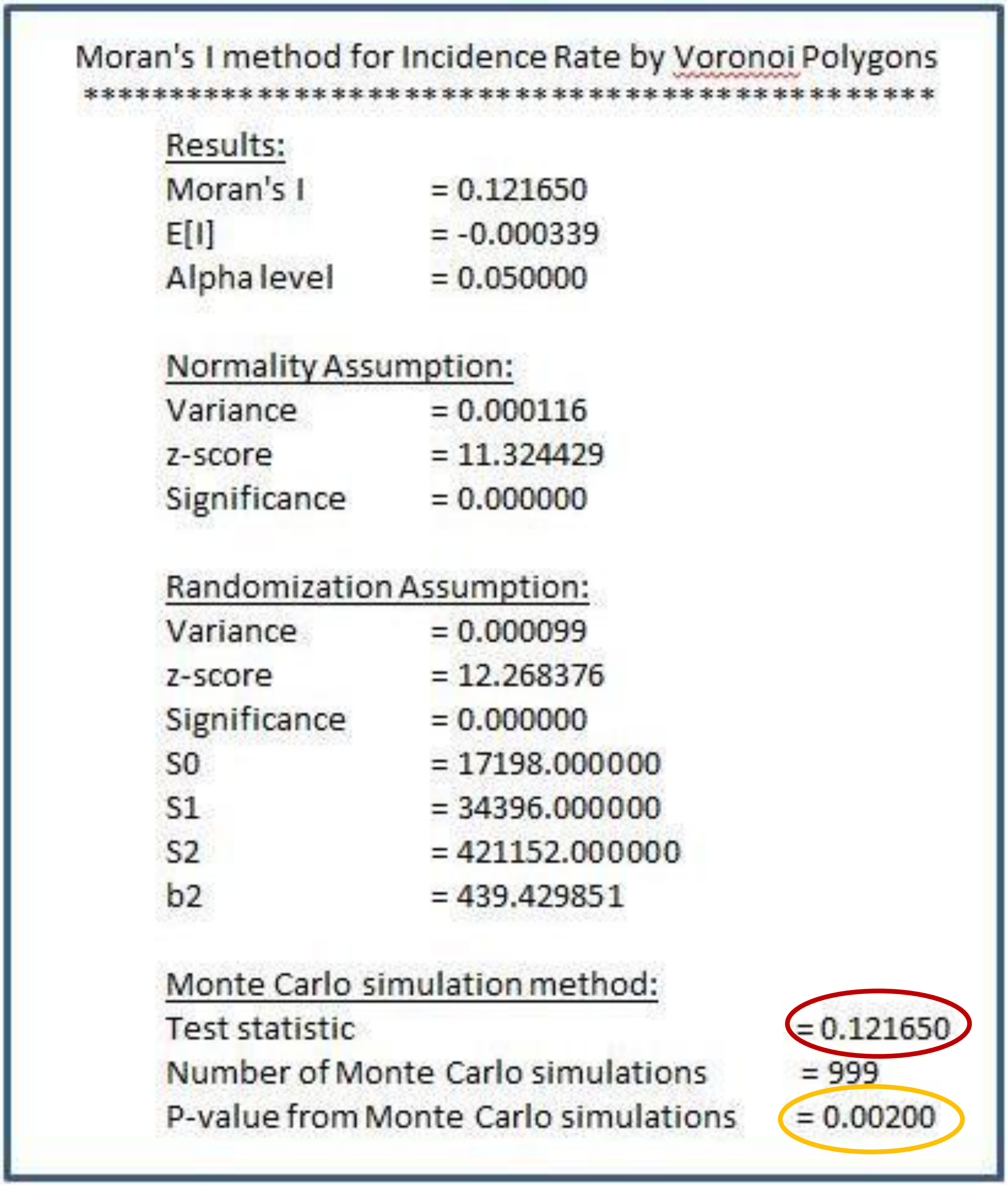


Table 3: Global Moran's *I* for Case Counts



```
Moran's I method for Case Count by Voronoi Polygons
*******************************************

    Results:
    Moran's I            = 0.112608
    E[I]                 = -0.000339
    Alpha level          = 0.050000

    Normality Assumption:
    Variance             = 0.000116
    z-score              = 10.485077
    Significance         = 0.000000

    Randomization Assumption:
    Variance             = 0.000093
    z-score              = 11.725926
    Significance         = 0.000000
    S0                   = 17198.000000
    S1                   = 34396.000000
    S2                   = 421152.000000
    b2                   = 594.227742

    Monte Carlo simulation method:
    Test statistic                             = 0.112608
    Number of Monte Carlo simulations           = 999
    P-value from Monte Carlo simulations       = 0.00400
```

Table 4: Global Moran's *I* for Incidence Rate



```
Moran's I method for Incidence Rate by Voronoi Polygons
*************************************************

     Results:
     Moran's I            = 0.121650
     E[I]                 = -0.000339
     Alpha level          = 0.050000

     Normality Assumption:
     Variance             = 0.000116
     z-score              = 11.324429
     Significance         = 0.000000

     Randomization Assumption:
     Variance             = 0.000099
     z-score              = 12.268376
     Significance         = 0.000000
     S0                   = 17198.000000
     S1                   = 34396.000000
     S2                   = 421152.000000
     b2                   = 439.429851

     Monte Carlo simulation method:
     Test statistic                              = 0.121650
     Number of Monte Carlo simulations            = 999
     P-value from Monte Carlo simulations        = 0.00200
```

Figure 13: Global Autocorrelation Spatial Correlogram for case counts

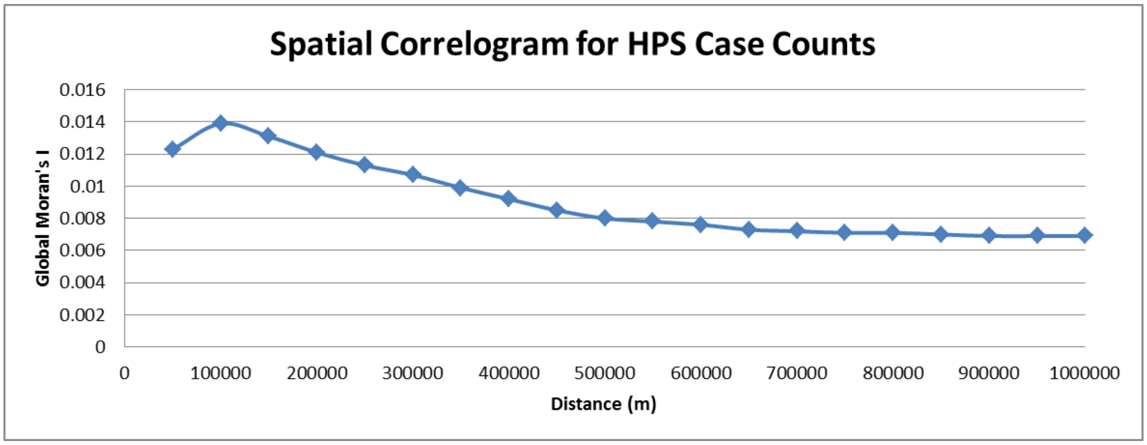


Figure 14: Global Autocorrelation Spatial Correlogram for Incidence Rate

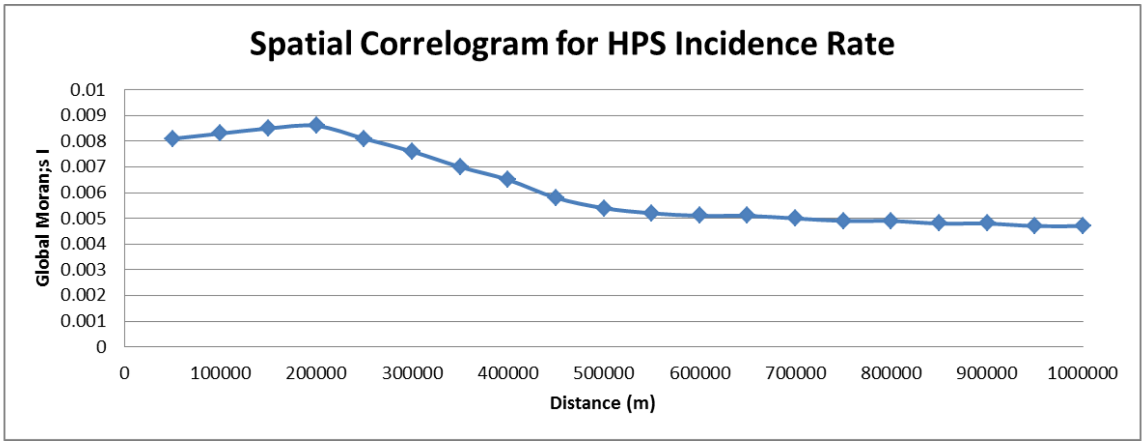


Figure 15: Local Clustering of Case Counts

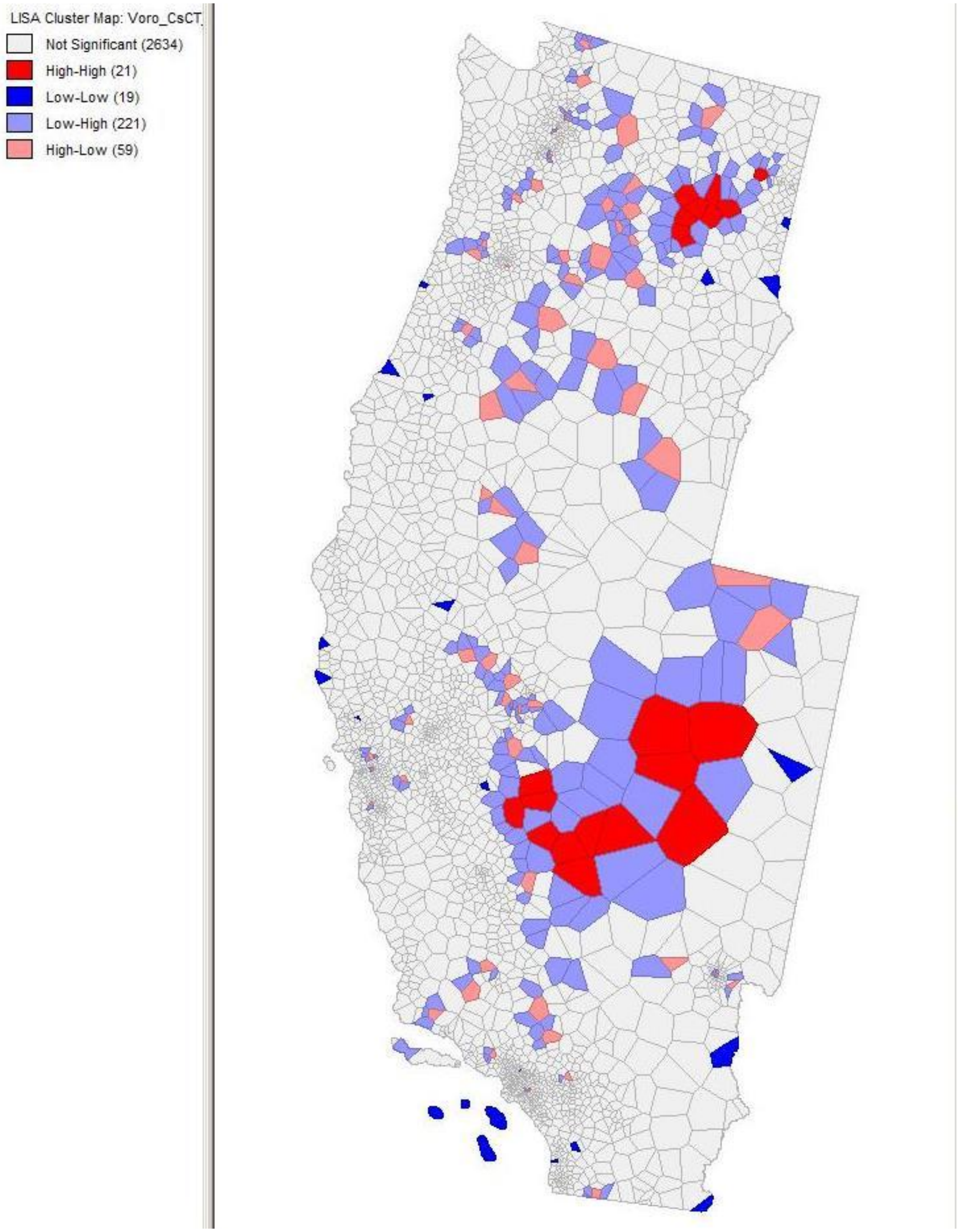


Figure 16: Significant Clustering of Case Counts

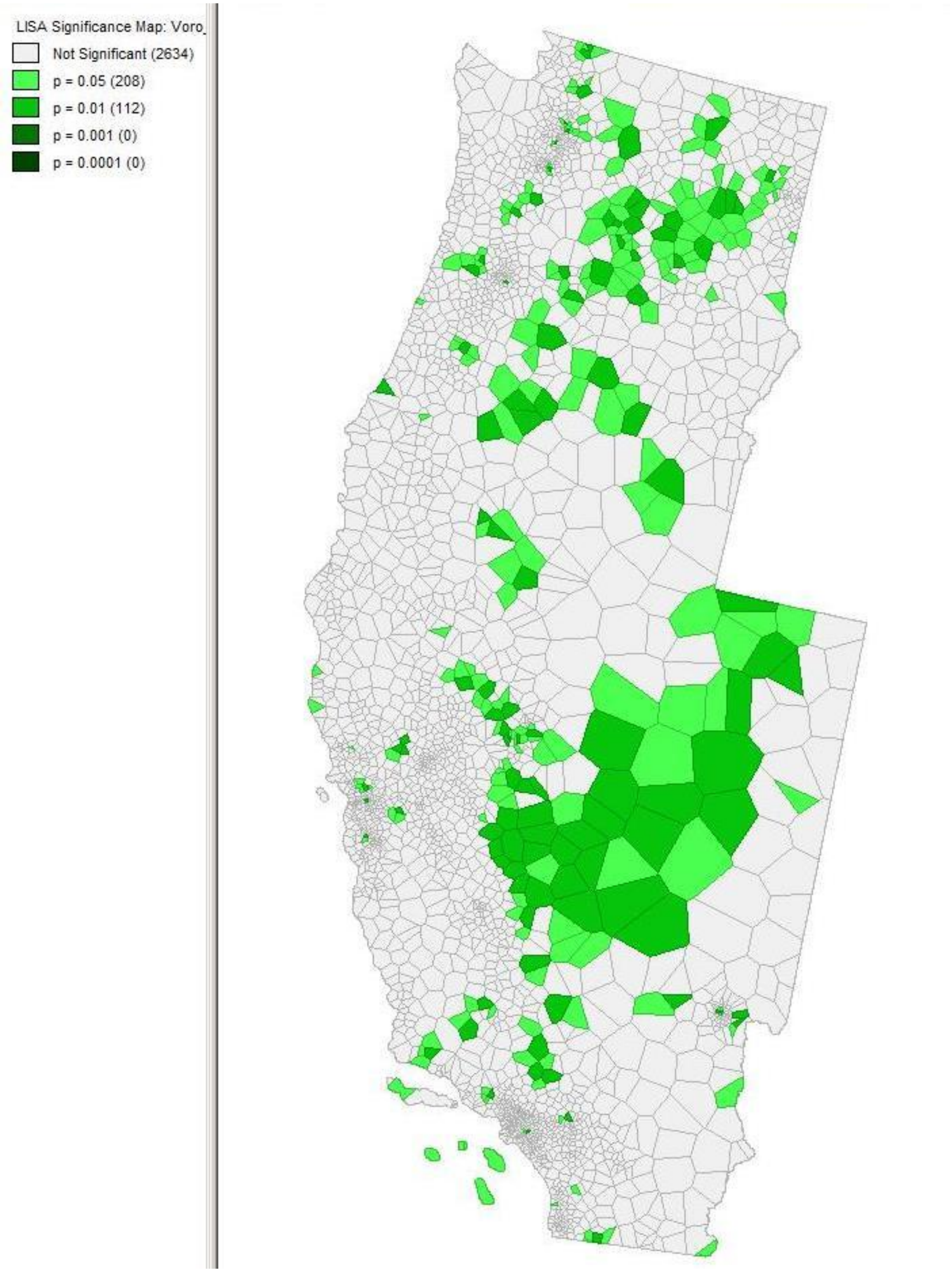


Figure 17: Local Clustering of Incidence Rate

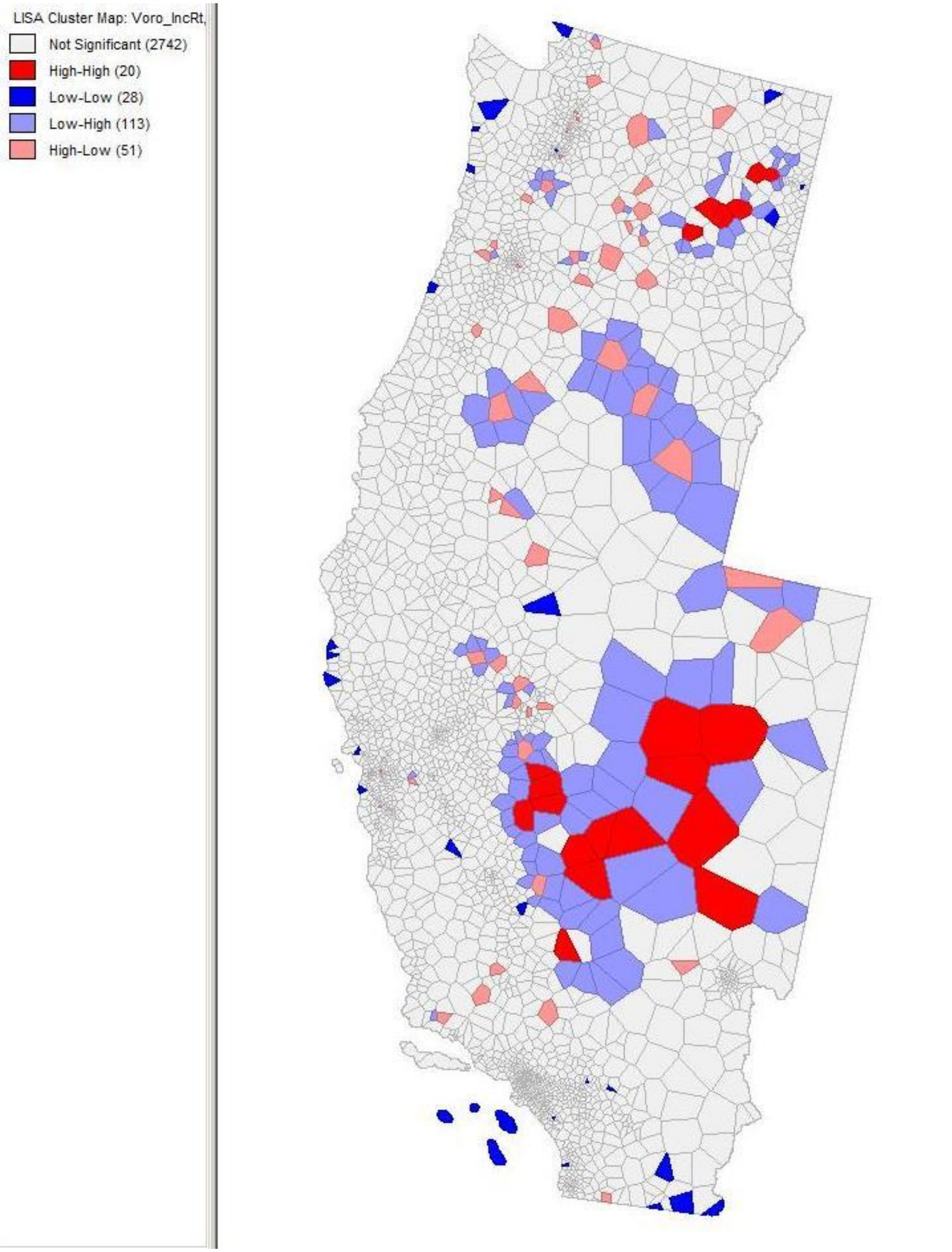


Figure 18: Significant local clustering of Case Incidence Rate

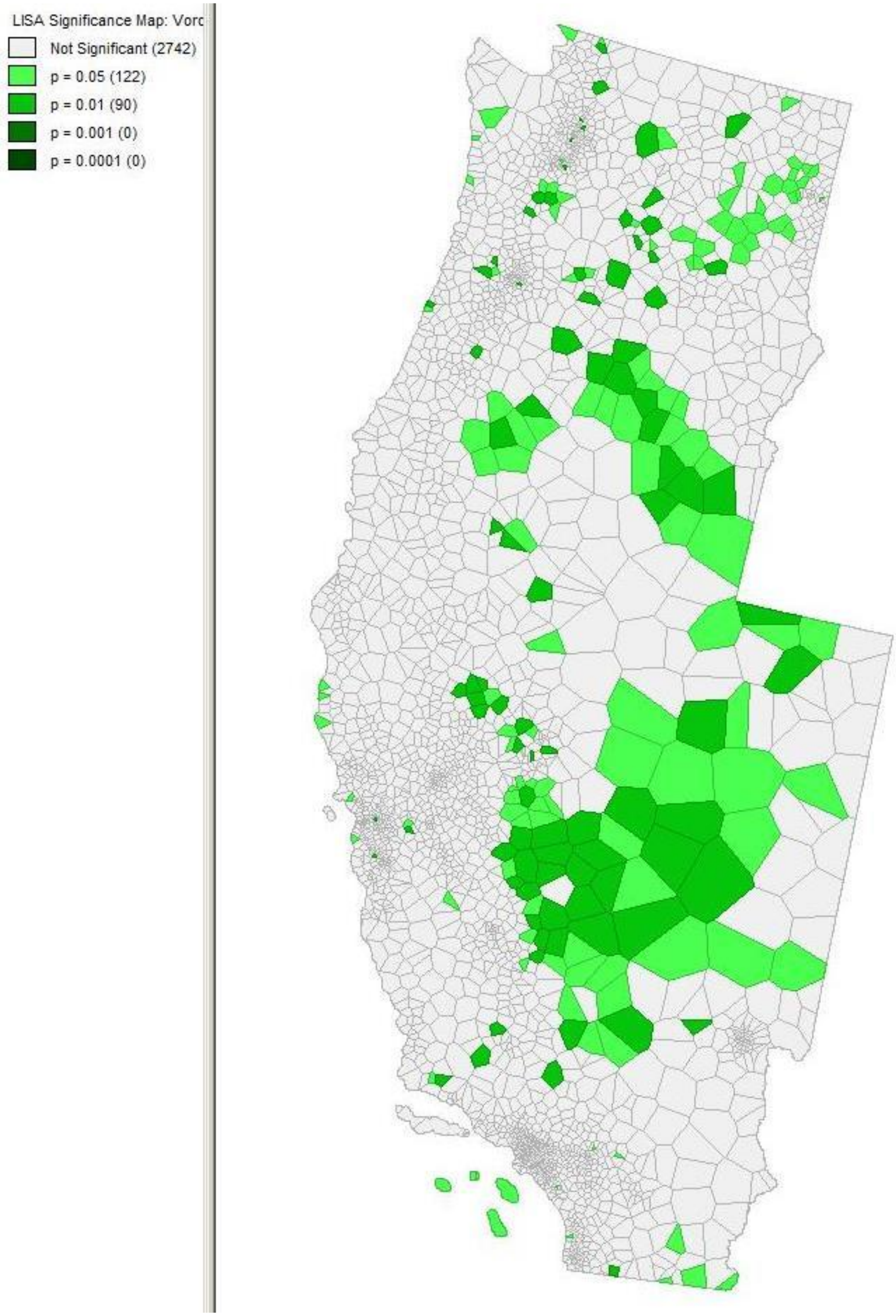


Figure 19: Hot Spot Analysis of Case Counts

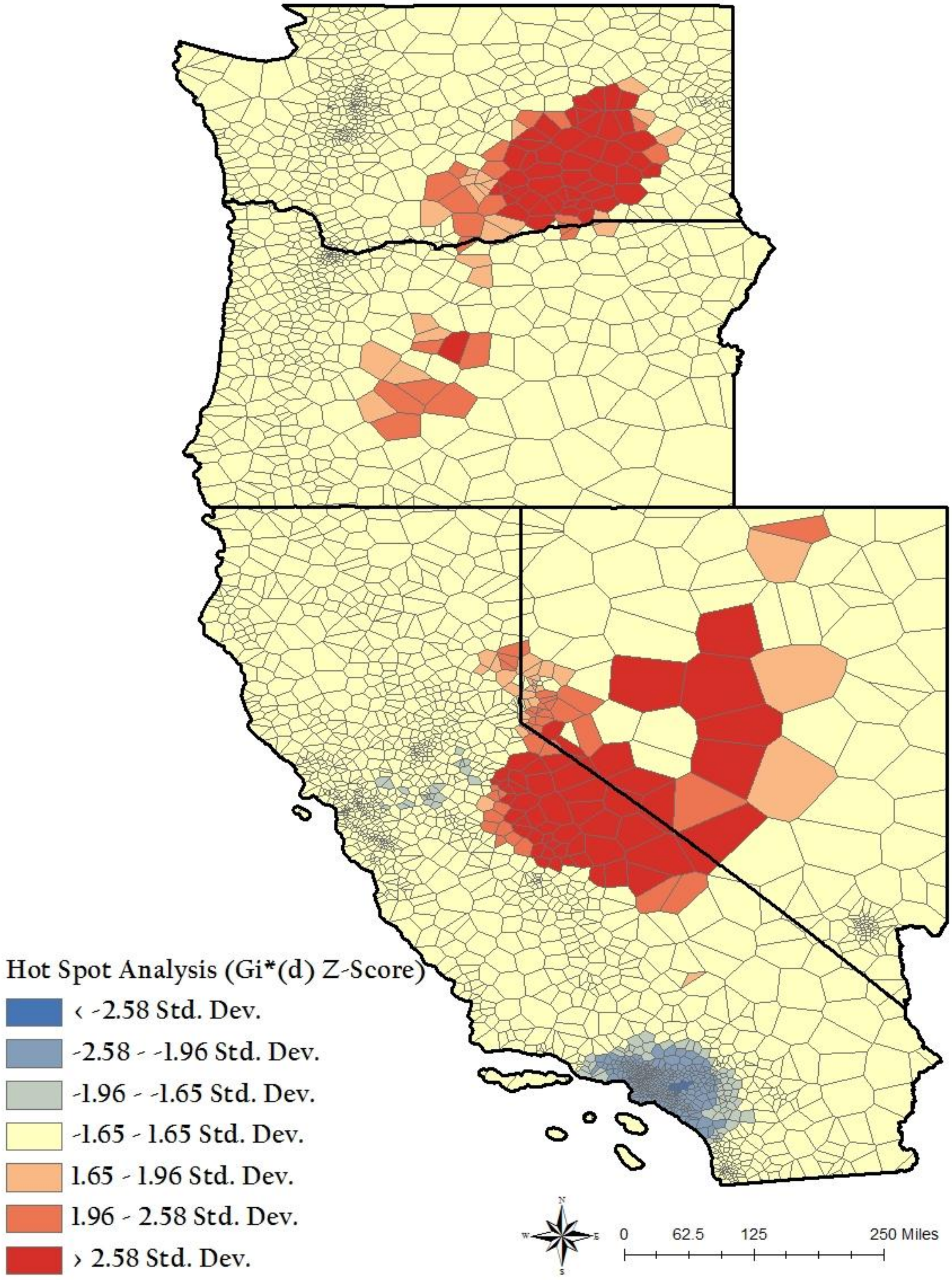


Figure 20: Hot Spot Analysis of Case Incidence Rates

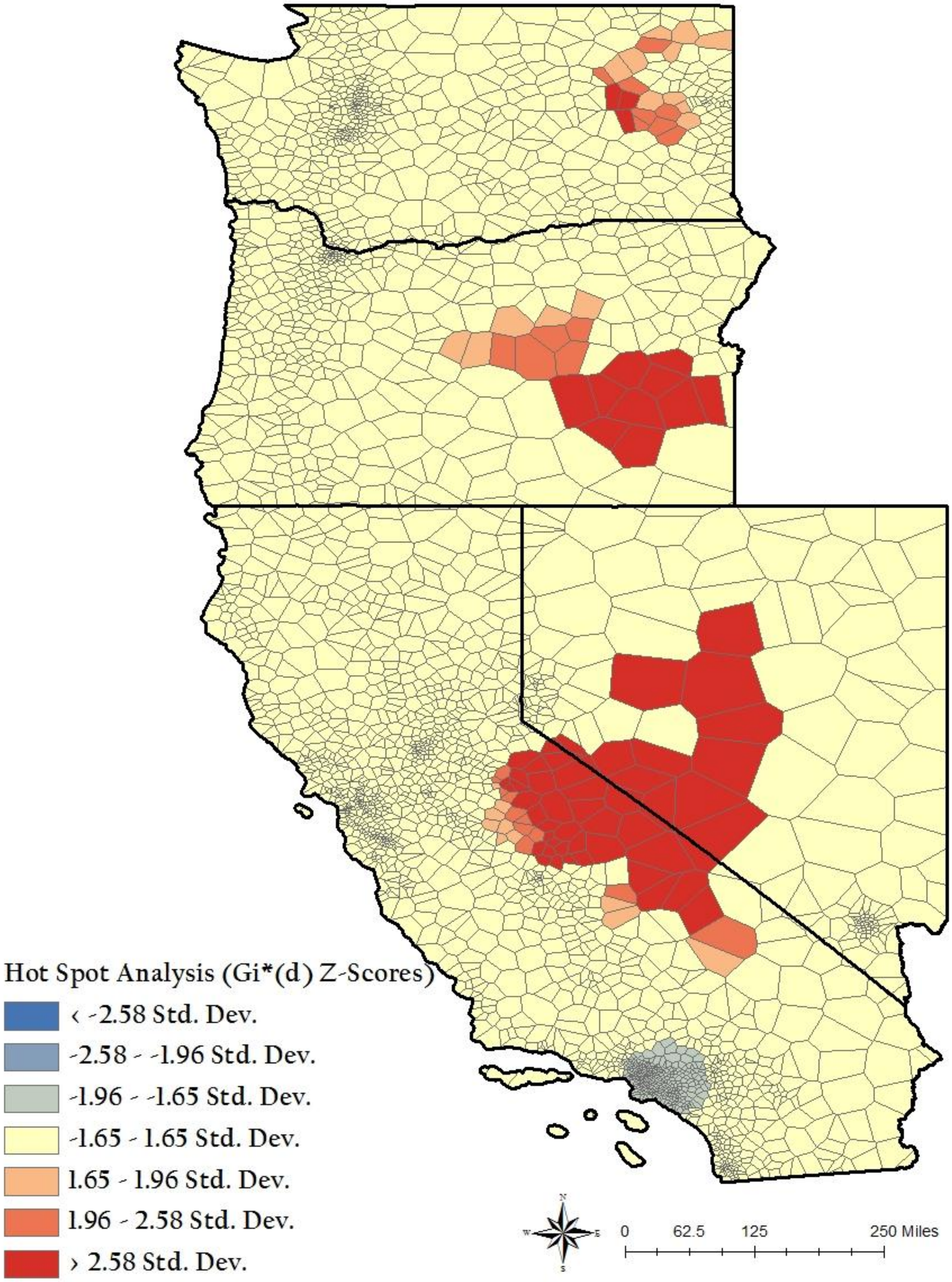


Figure 21: Significant z-scores in Hot Spot Analysis of Case Counts

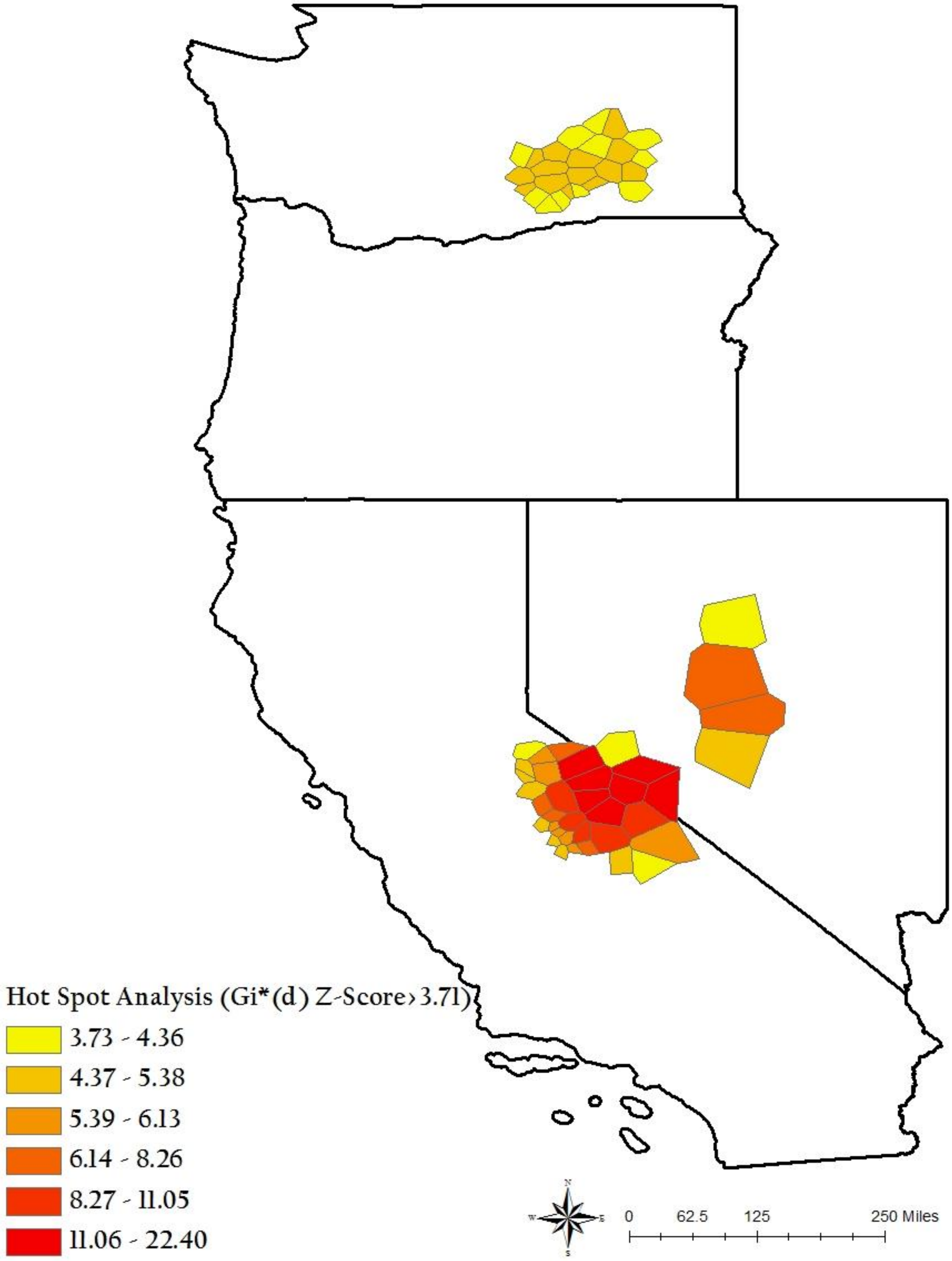


Figure 22: Significant z-scores in Hot Spot Analysis of Case Incidence Rates

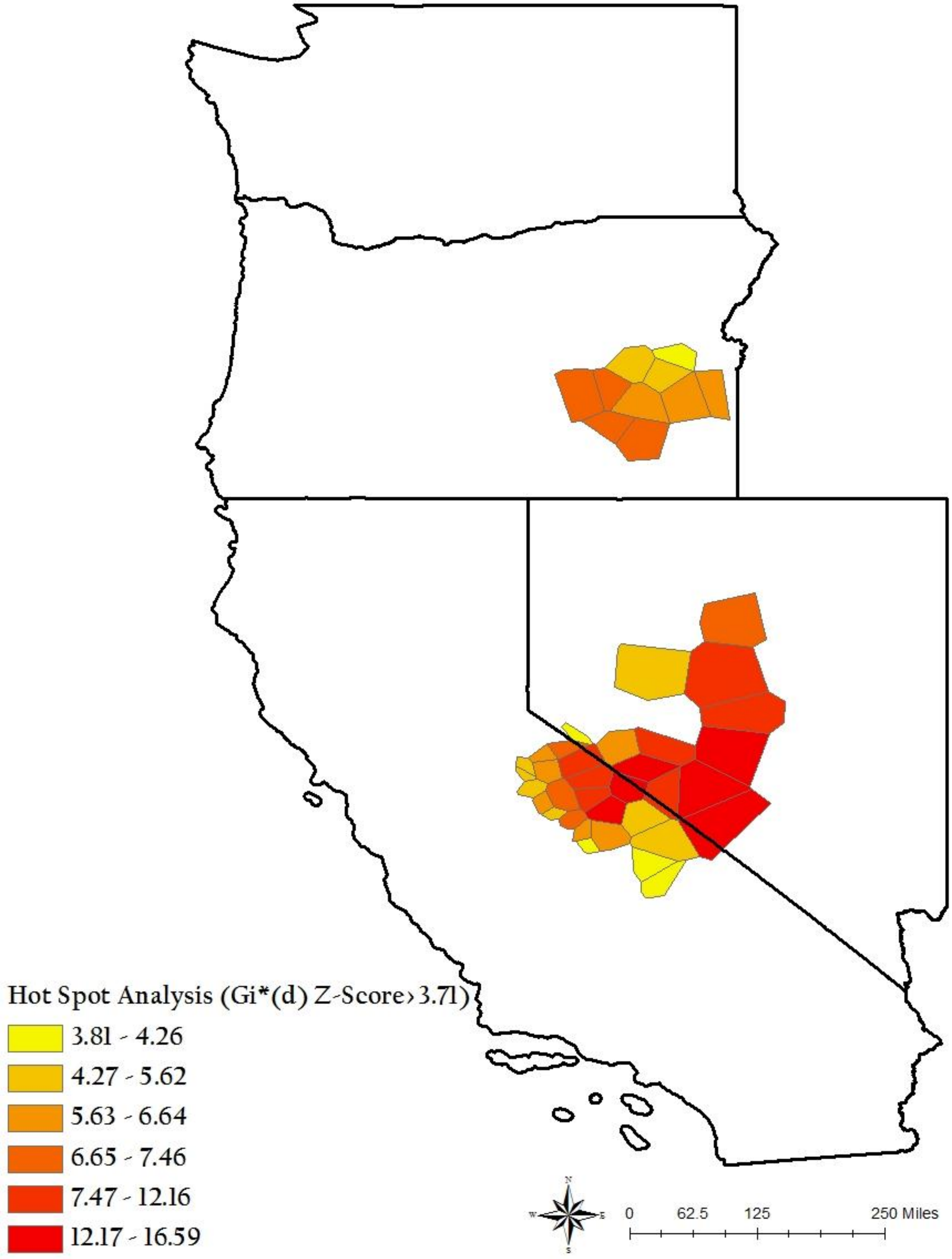
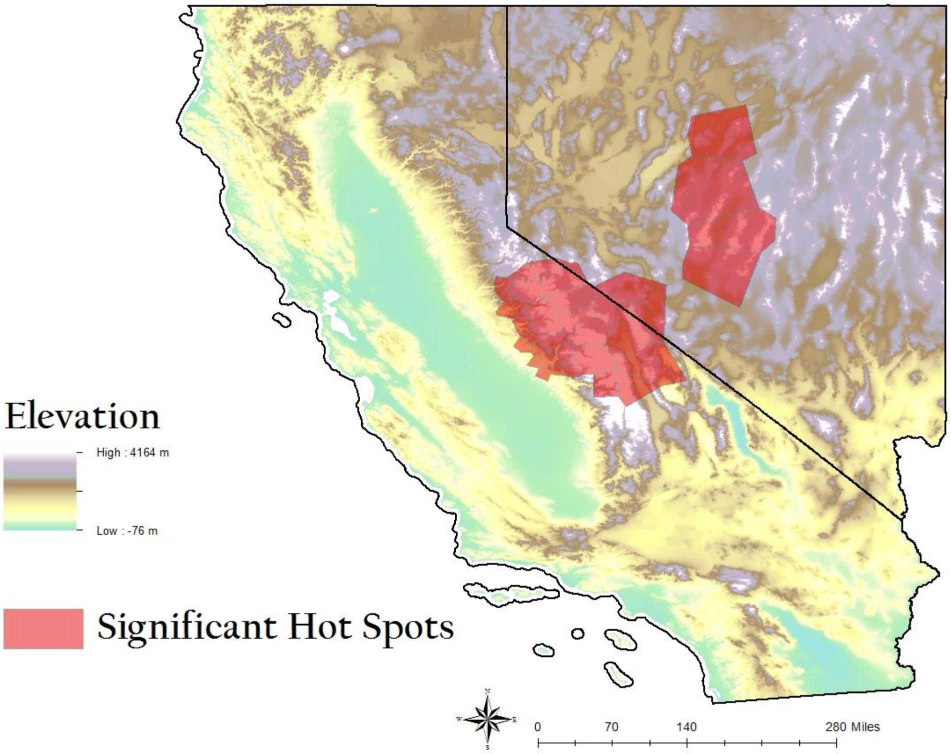


Figure 23: Elevation

Elevation with Cases by Getis' Hot Spots (z-scores>3.71) in Voronoi Polygons



Elevation with Cases by ZCTA

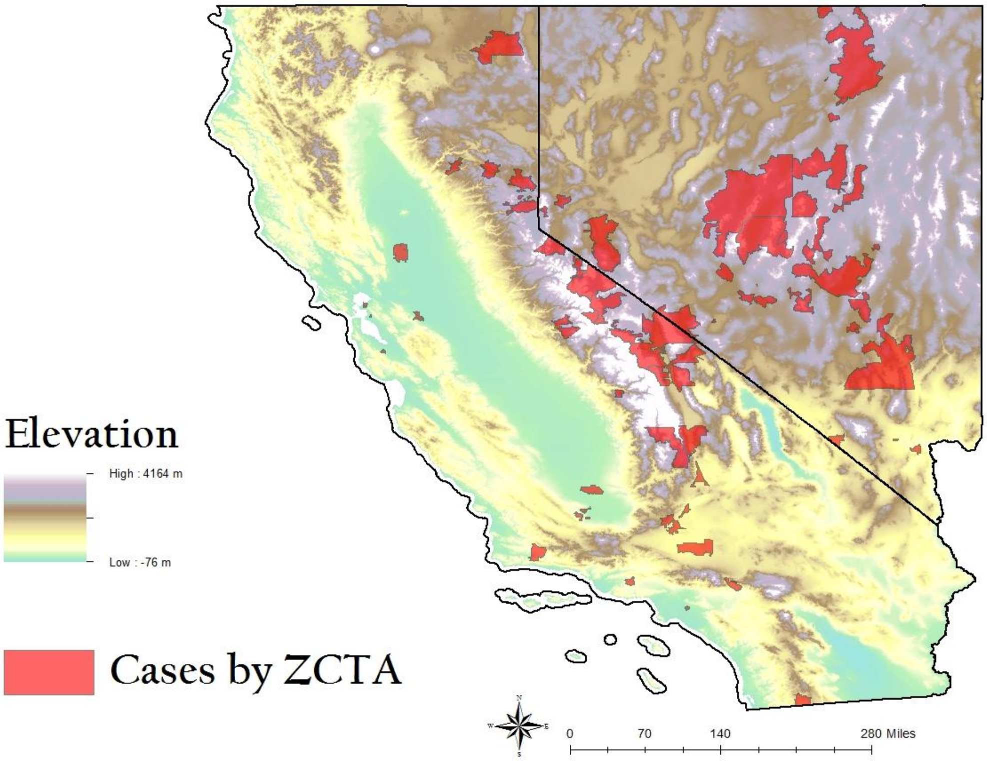
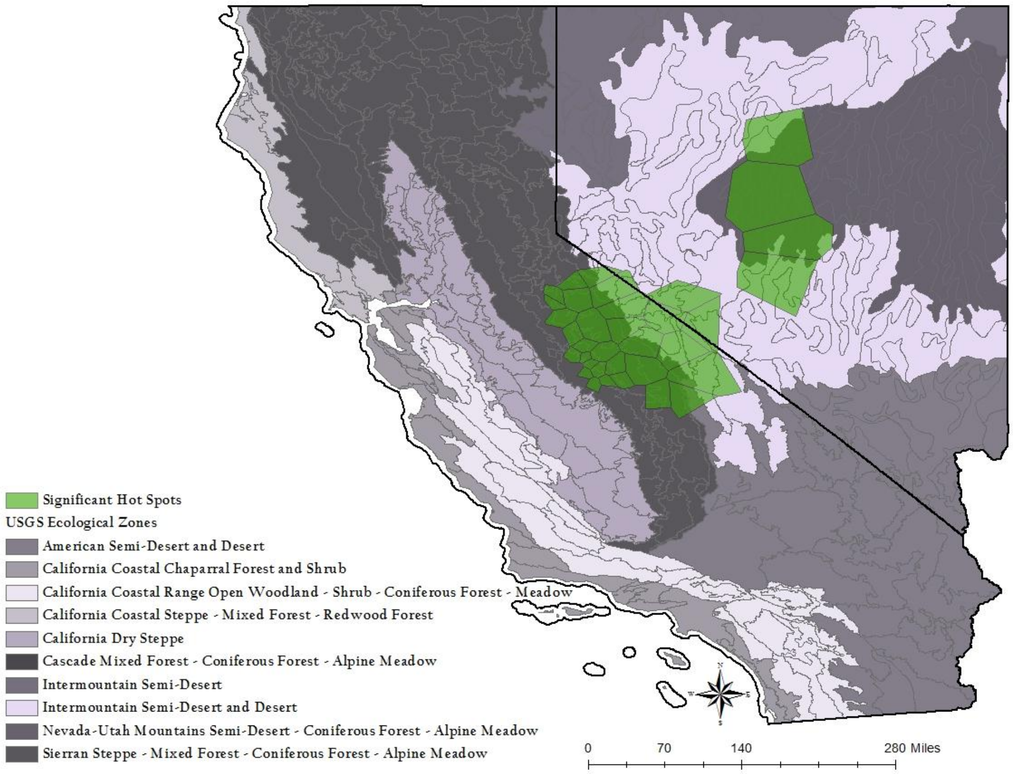


Figure 24: Ecological Zones

Ecological Zones with Cases by Getis' Hot Spot using Voronoi polygons


Significant Hot Spots
USGS Ecological Zones
American Semi-Desert and Desert
California Coastal Chaparral Forest and Shrub
California Coastal Range Open Woodland - Shrub - Coniferous Forest - Meadow
California Coastal Steppe - Mixed Forest - Redwood Forest
California Dry Steppe
Cascade Mixed Forest - Coniferous Forest - Alpine Meadow
Intermountain Semi-Desert
Intermountain Semi-Desert and Desert
Nevada-Utah Mountains Semi-Desert - Coniferous Forest - Alpine Meadow
Sierran Steppe - Mixed Forest - Coniferous Forest - Alpine Meadow
0
70
140
280 Miles

Ecological Zones with Cases by ZCTA

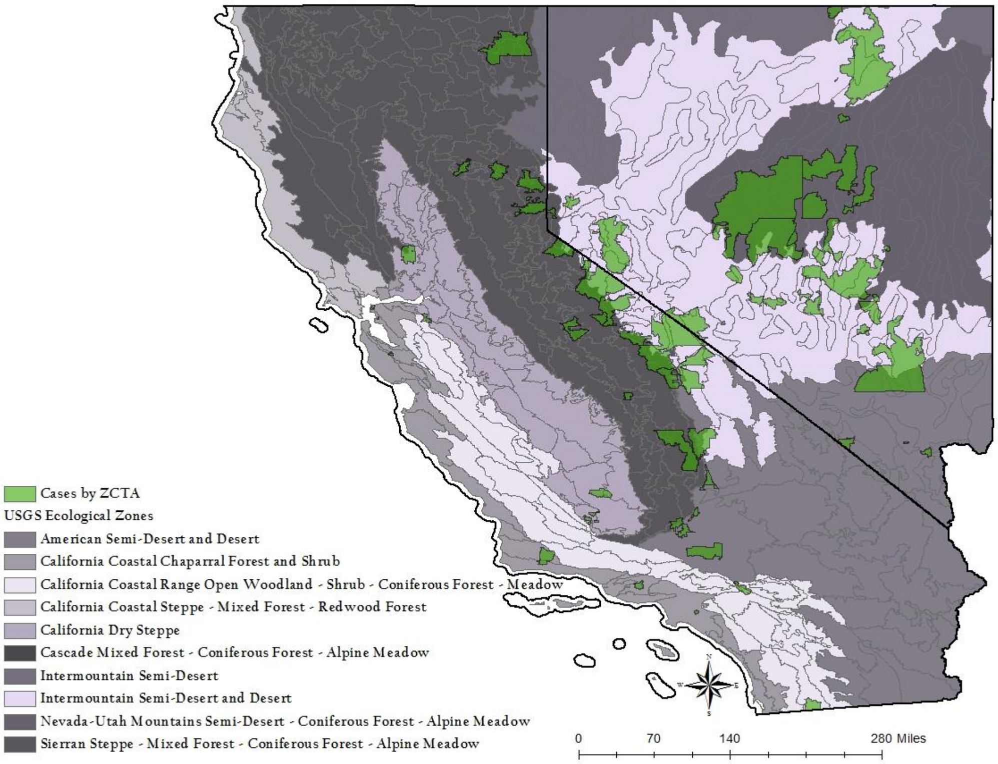
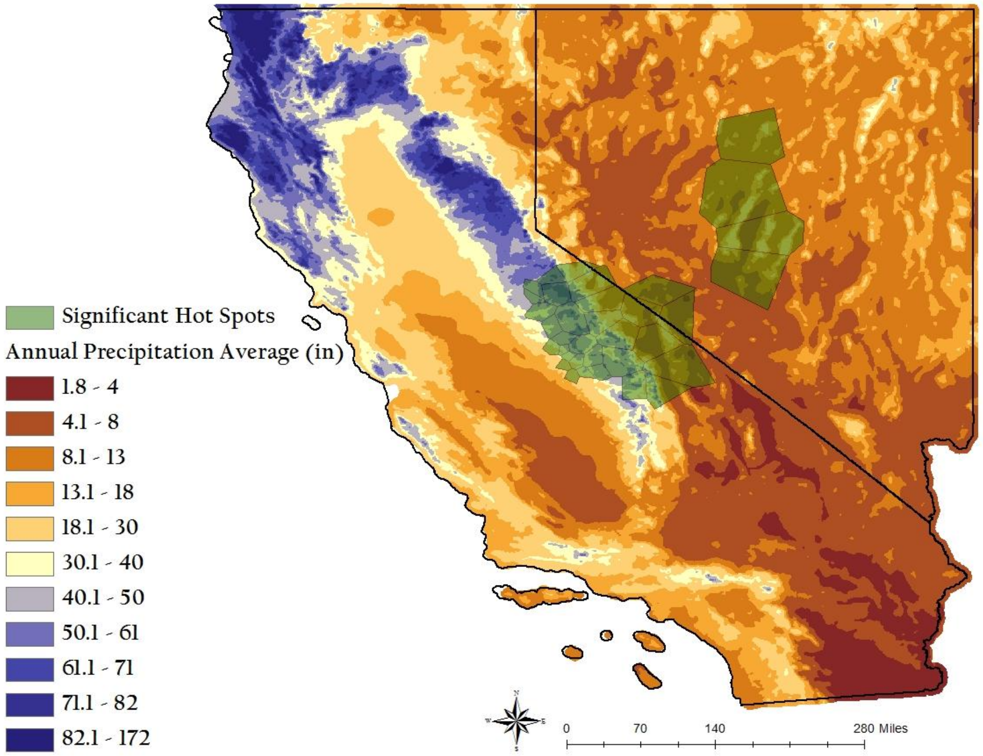

Cases by ZCTA
USGS Ecological Zones
American Semi-Desert and Desert
California Coastal Chaparral Forest and Shrub
California Coastal Range Open Woodland - Shrub - Coniferous Forest - Meadow
California Coastal Steppe - Mixed Forest - Redwood Forest
California Dry Steppe
Cascade Mixed Forest - Coniferous Forest - Alpine Meadow
Intermountain Semi-Desert
Intermountain Semi-Desert and Desert
Nevada-Utah Mountains Semi-Desert - Coniferous Forest - Alpine Meadow
Sierran Steppe - Mixed Forest - Coniferous Forest - Alpine Meadow
0
70
140
280 Miles

Figure 25: Annual Precipitation

Annual Precipitation with Voronoi Hot Spots (z-score>3.71)



Annual Precipitation with Cases

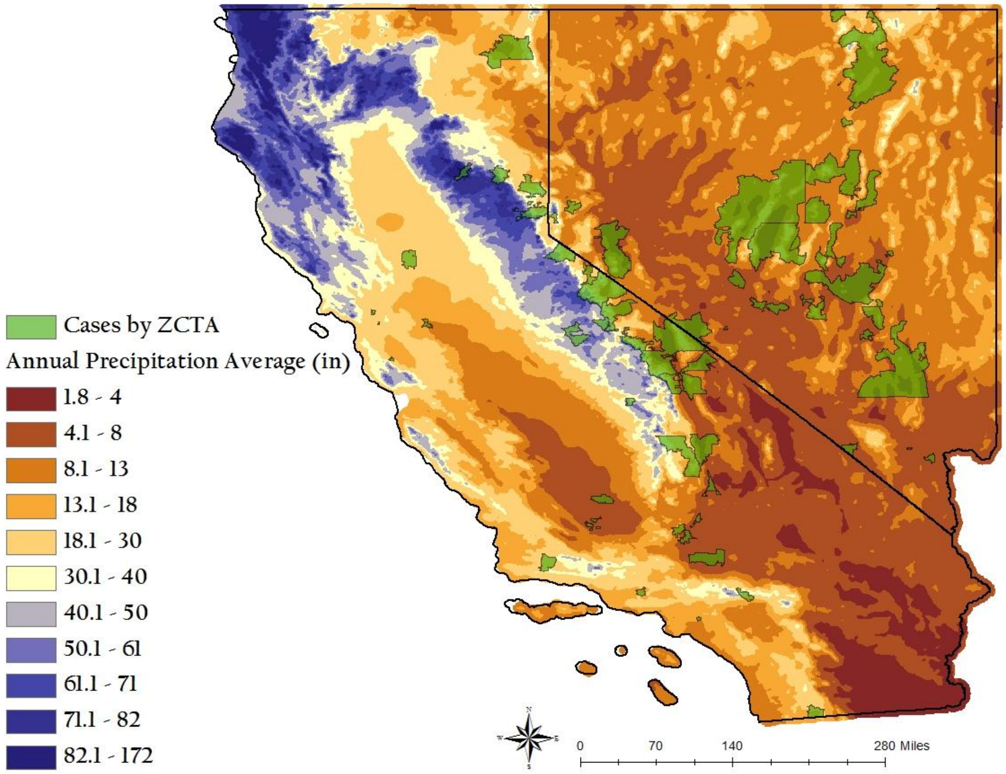
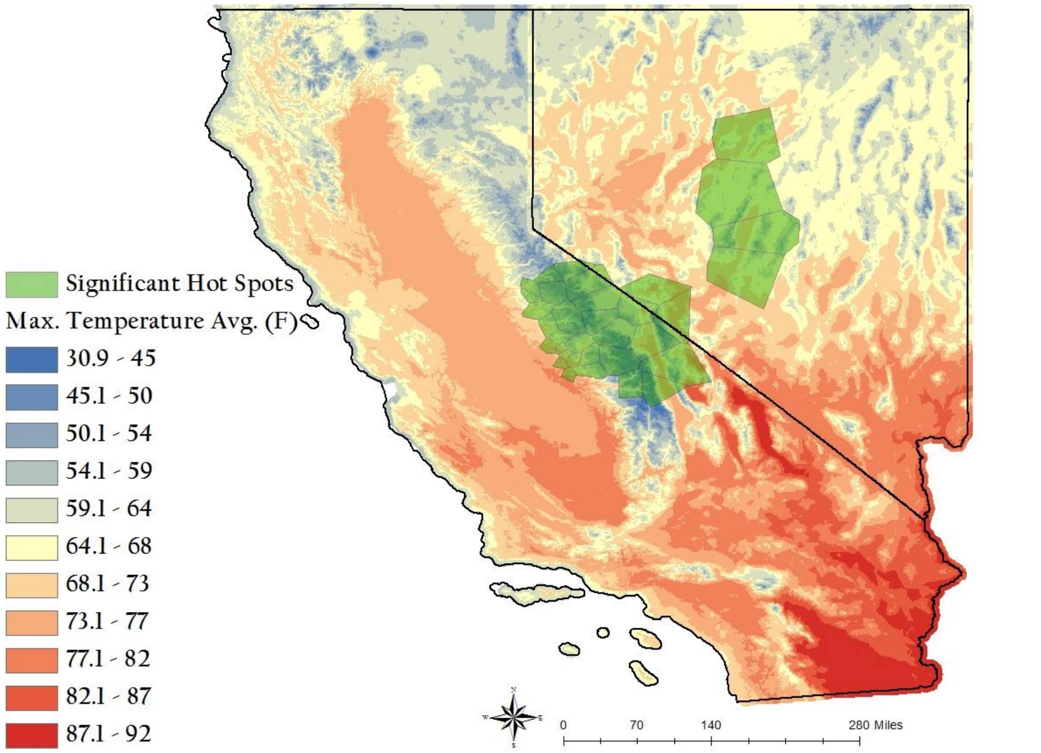


Figure 26: Maximum Temperature Averages 1981-2010

Maximum Temperature Averages with Voronoi Hot Spots (z-scores>3.71)



Maximum Temperature Averages with Cases

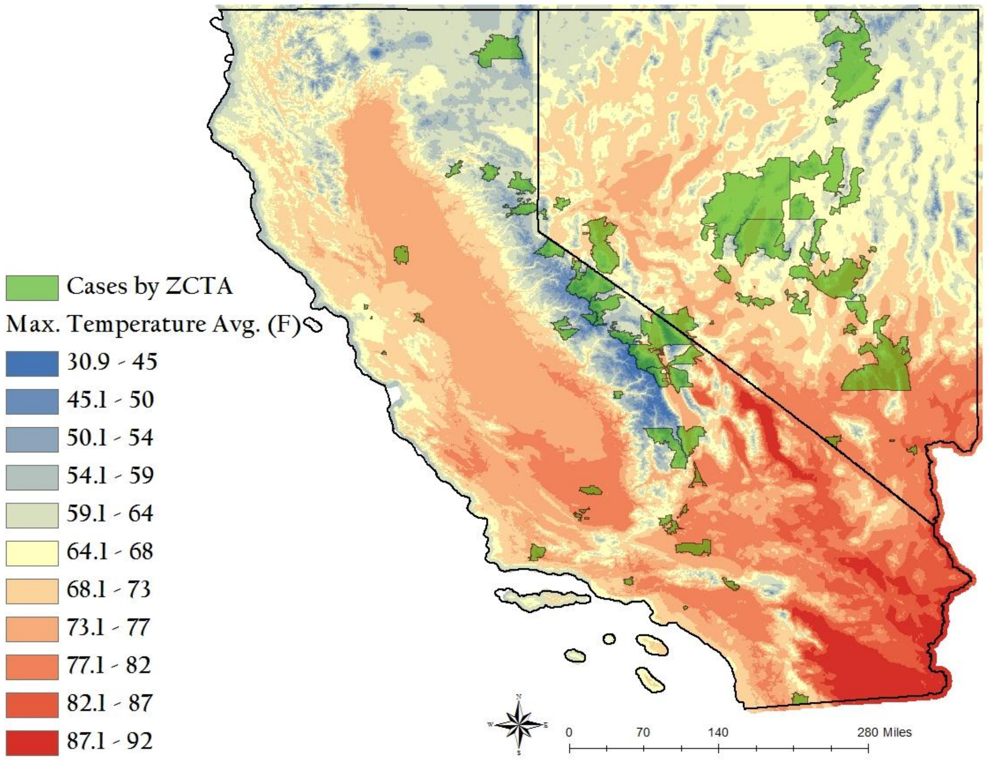
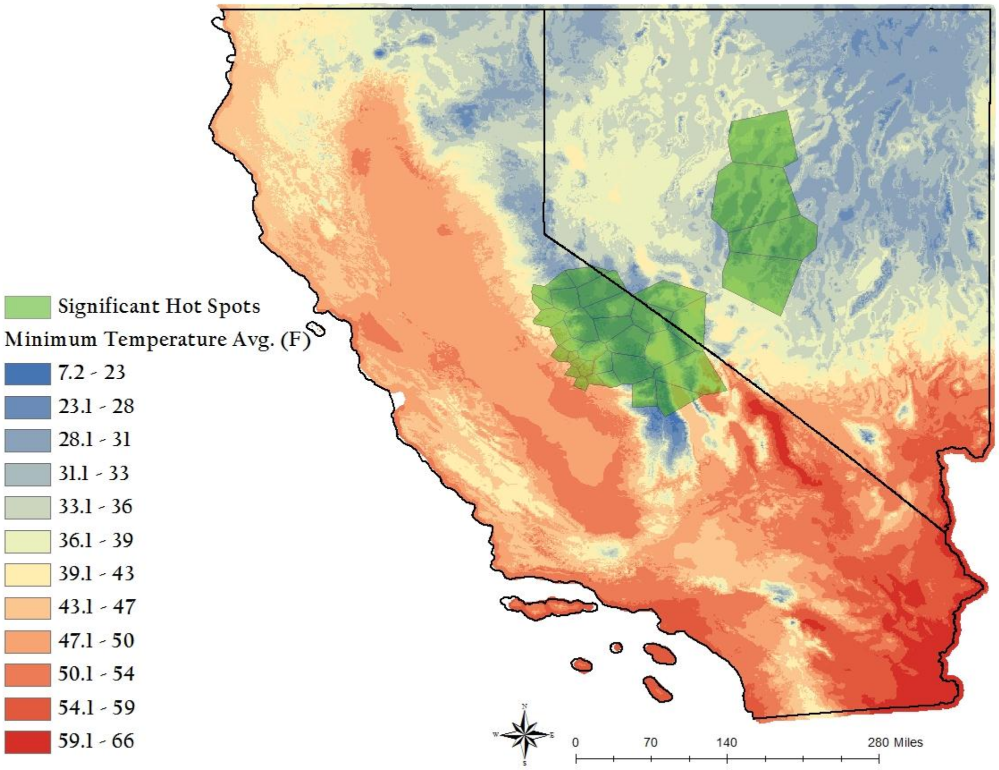


Figure 27: Minimum Temperature Averages 1981-2010

Minimum Temperature Averages with Voronoi Hot Spots (z-scores>3.71)



Minimum Temperature Averages with Cases

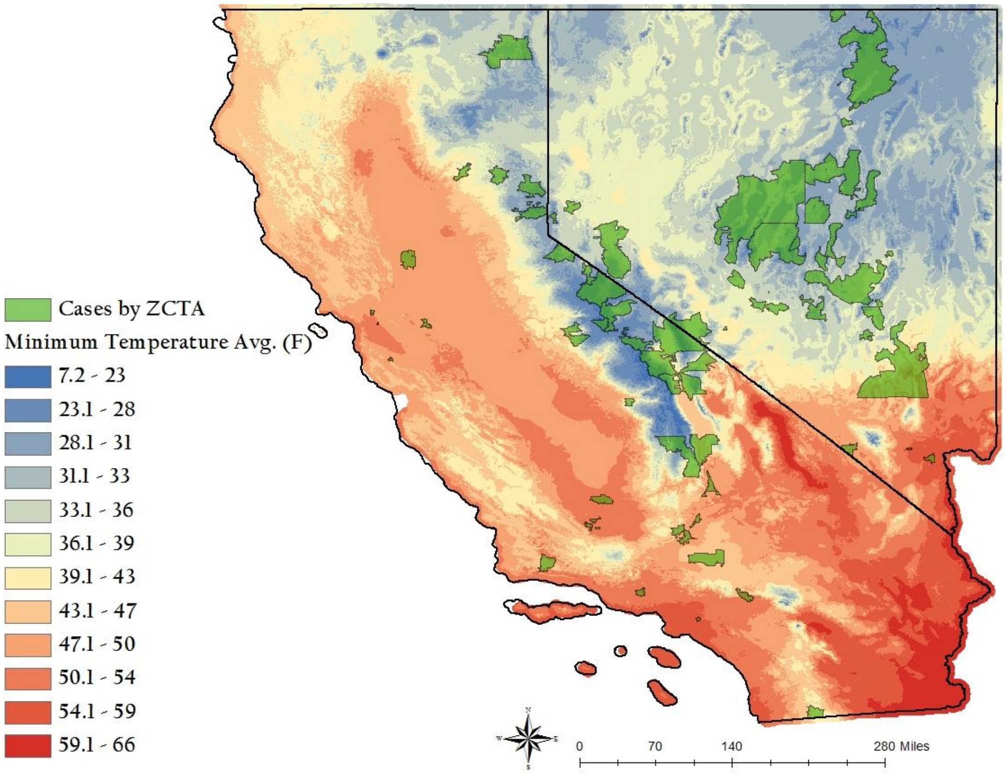
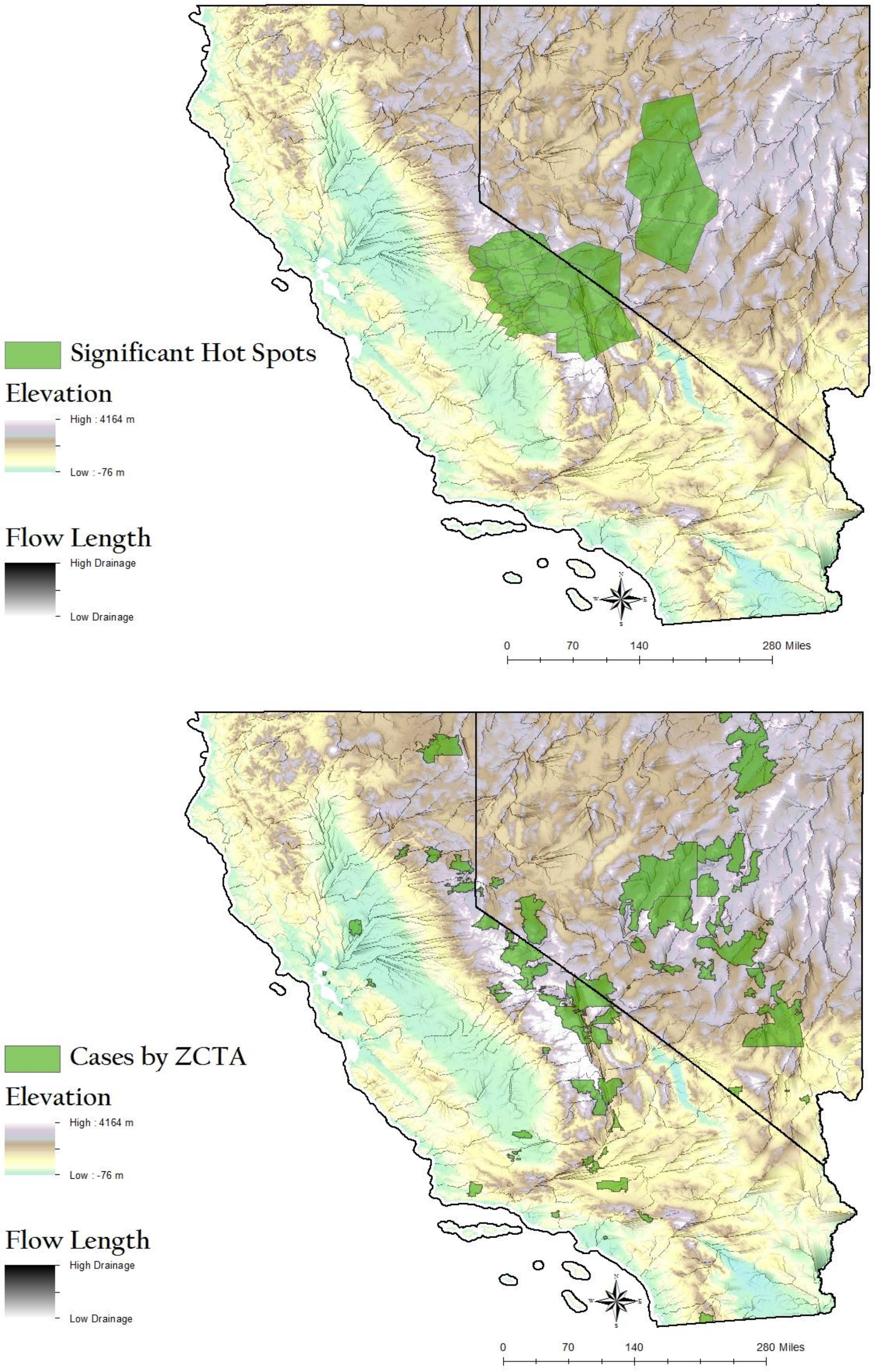
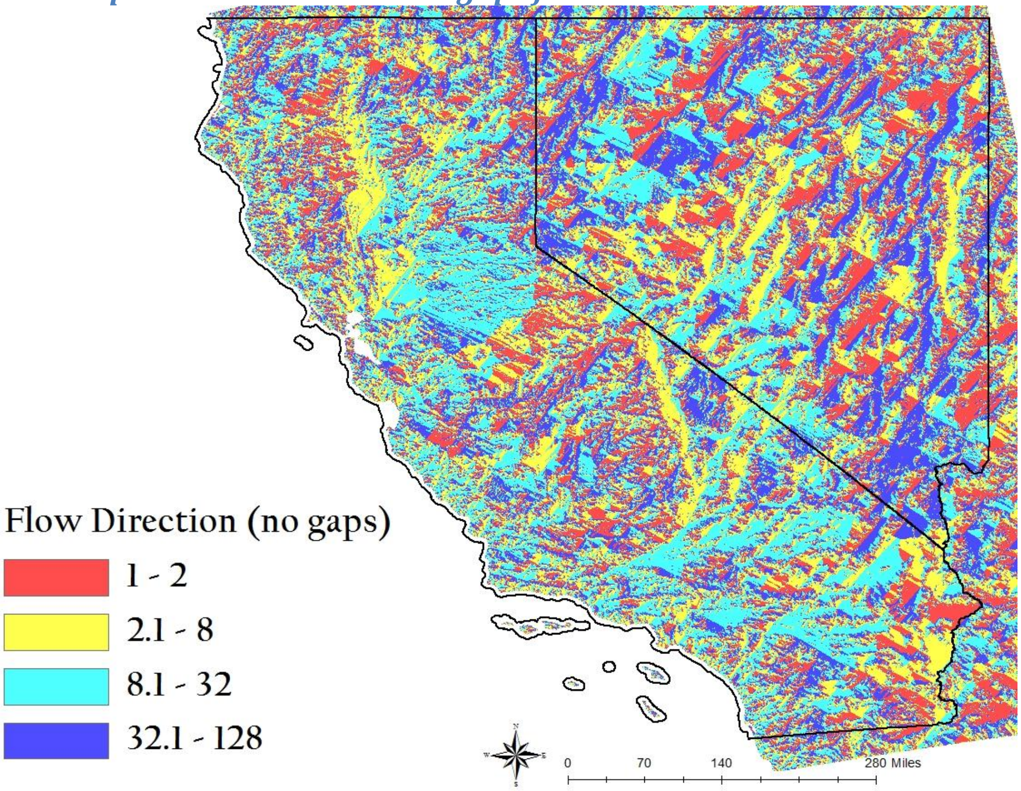


Figure 28: Elevation/Flow Length/Cases in Hot Spots and ZCTA

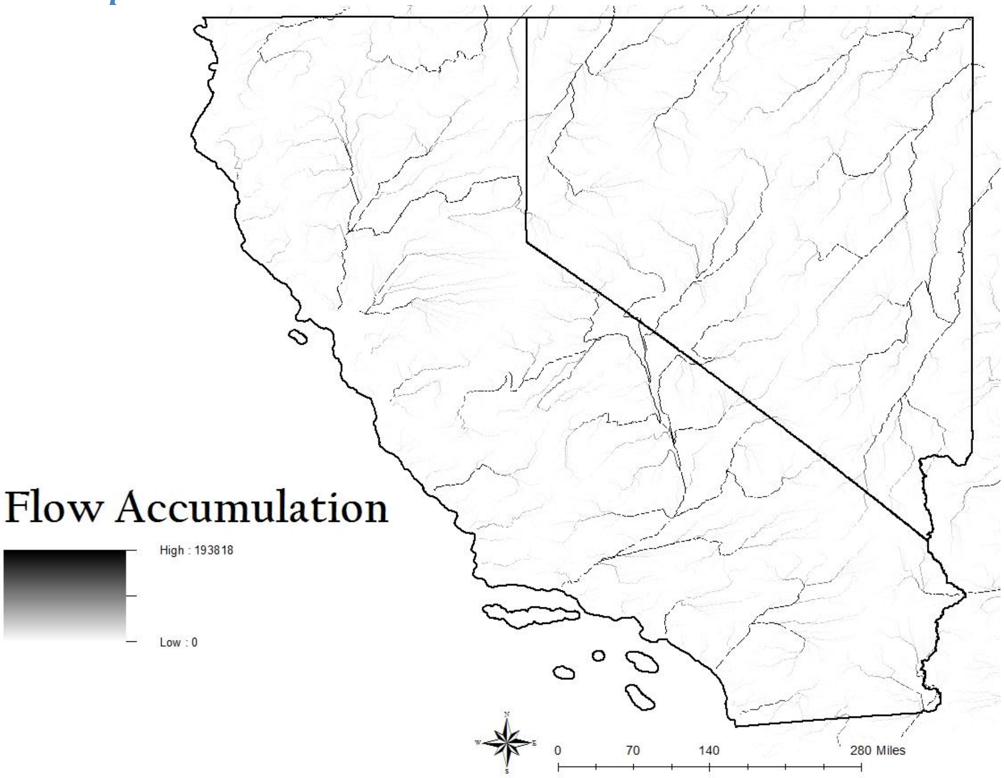


Figures 29 a-d: Hydrology Model Building

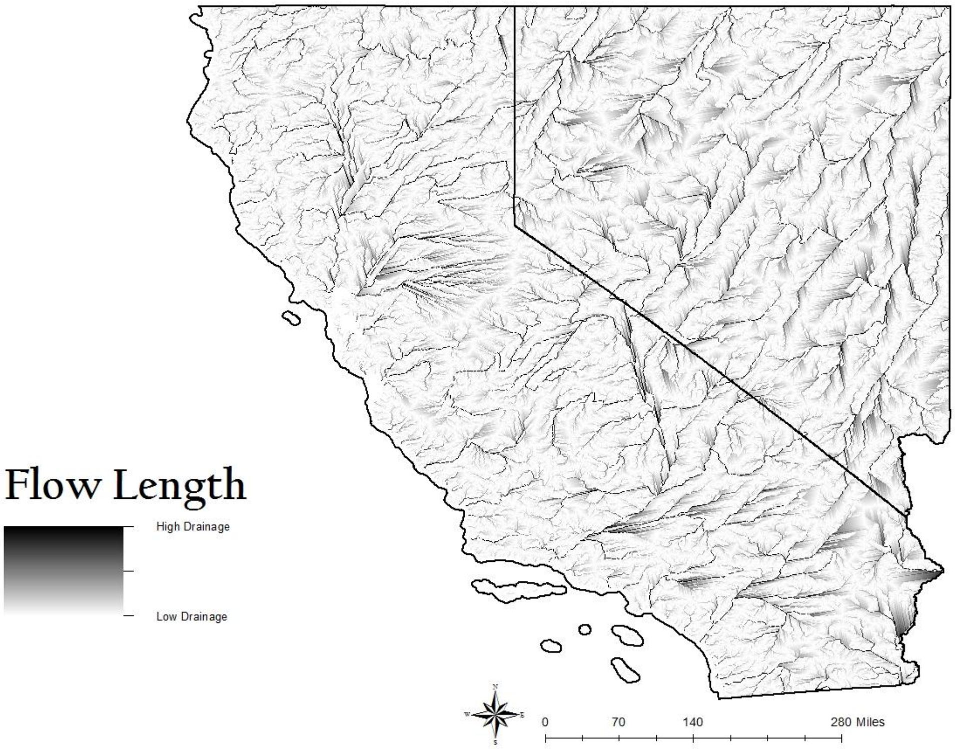
a. Step 1: Flow Direction with gaps filled in



b. Step 2: Flow Accumulation



c. Step 3: Flow Length



d. Step 4: Flow Length with Elevation

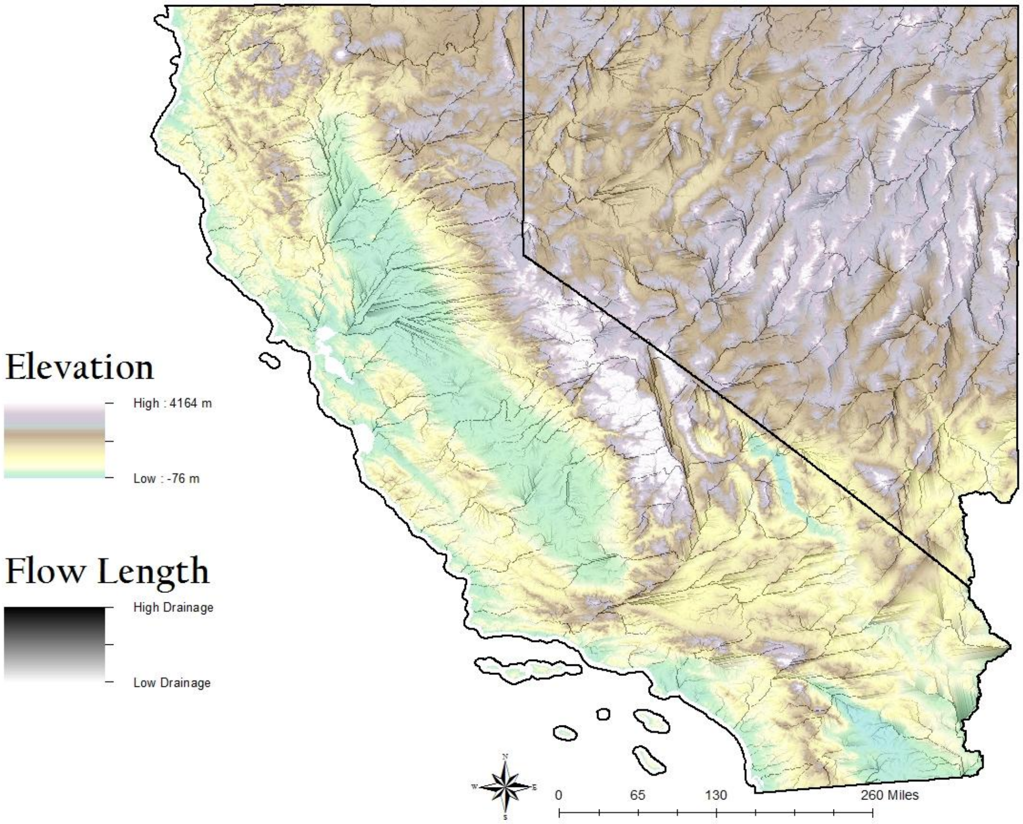


Figure 30: Sierra Nevada Study Area

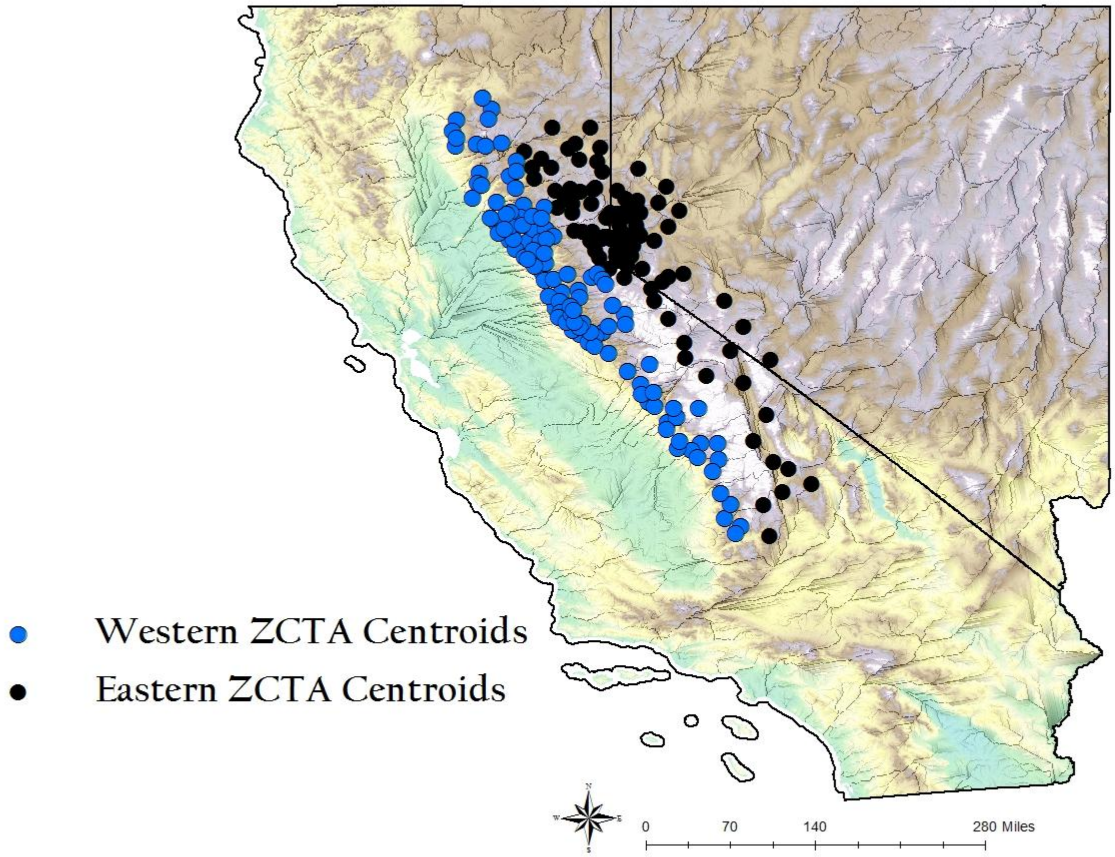
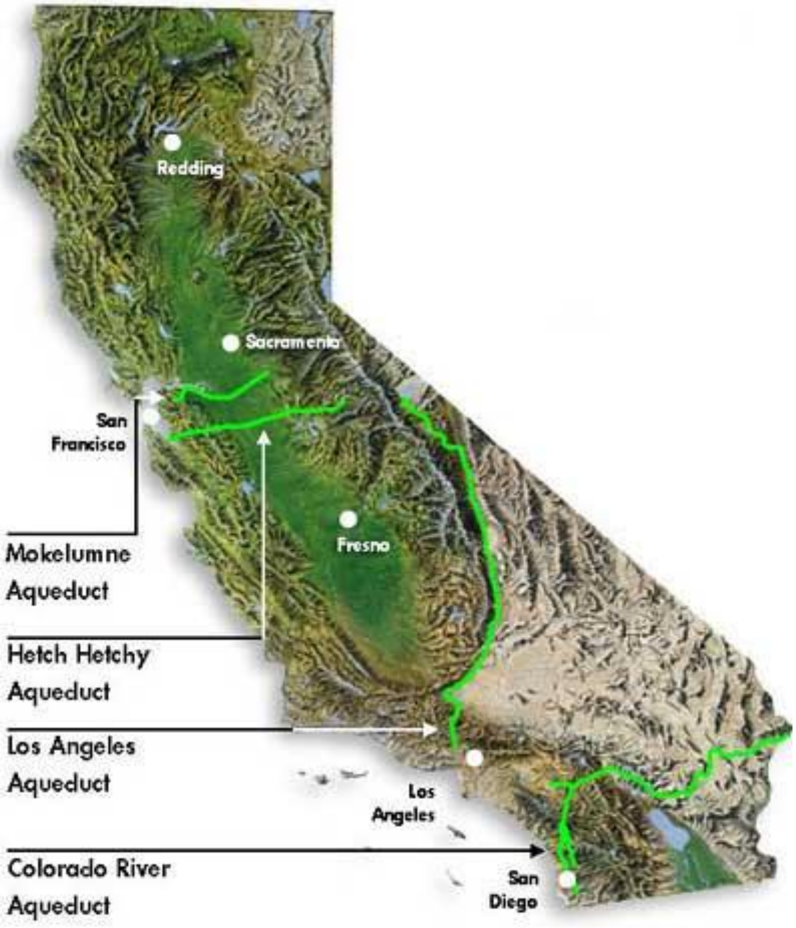


Table 5: Sierra Nevada Regression Analysis

	Estimate	Std. Error	t-Stat	p-Value
(Intercept)	-7.1508	1.9598	-3.6487	0.00033363
Min. Temp	-0.032443	0.022301	-1.4548	0.14724
Max. Temp	0.11198	0.026415	4.2391	3.38E-05
Elevation	0.00094764	0.00026749	3.5427	0.00048942
Orientation	2.9042	0.88401	3.2853	0.0011964
Min. Temp.* Orientation	-0.085116	0.024234	-3.5123	0.00054541

Figure 31: California Aqueducts



California Dept. of Water Resources:
http://www.water.ca.gov/pubs/surfacewater/local_water_projects_informational_maps/050809local2.jpg

References

1. Zeier, M., et al., *New ecological aspects of hantavirus infection: a change of a paradigm and a challenge of prevention--a review.* Virus Genes, 2005. **30**(2): p. 157-80.

2. Boone, J.D., et al., *Infection dynamics of Sin Nombre virus after a widespread decline in host populations.* Am J Trop Med Hyg, 2002. **67**(3): p. 310-8.

3. Eisen, R.J., et al., *A spatial model of shared risk for plague and hantavirus pulmonary syndrome in the southwestern United States.* The American Journal of Tropical Medicine and Hygiene, 2007. **77**(6): p. 999-1004.

4. Dearing, M.D. and L. Dizney, *Ecology of hantavirus in a changing world.* Ann N Y Acad Sci, 2010. **1195**: p. 99-112.

5. Bagamian, K.H., et al., *Population density and seasonality effects on Sin Nombre virus transmission in North American deermice (Peromyscus maniculatus) in outdoor enclosures.* PLoS One, 2012. **7**(6): p. e37254.

6. Dragoo, J.W., et al., *Phylogeography of the deer mouse (Peromyscus maniculatus) provides a predictive framework for research on hantaviruses.* The Journal of General Virology, 2006. **87**(Pt 7): p. 1997-2003.

7. Mills, J.N., B.R. Amman, and G.E. Glass, *Ecology of hantaviruses and their hosts in North America.* Vector Borne Zoonotic Diseases, 2010. **10**(6): p. 563-74.

8. *National Notifiable Diseases Surveillance System: Hantavirus pulmonary syndrome*. [Web] Dec. 7, 2012 [cited 2013 March 30]; Available from: http://wwwn.cdc.gov/NNDSS/script/conditionsummary.aspx?CondID=76.

9. *Protocol for Public Health Agencies to Notify CDC about the Occurrence of Nationally Notifiable Conditions, 2013*, C.f.S.a.T. Epidemiologists, Editor., Centers for Disease Control and Prevention.

10. *Hantavirus: Annual U.S. HPS Cases and Case-Fatality, 1993-2011*. Centers for Disease Control and Prevention [Web] March 22, 2012 [cited 2012 Oct. 17]; Available from: http://www.cdc.gov/hantavirus/surveillance/state-of-exposure.html.

11. Glass, G.E., et al., *Persistently highest risk areas for hantavirus pulmonary syndrome: potential sites for refugia.* Ecological Applications : a publication of the Ecological Society of America, 2007. **17**(1): p. 129-39.

12. Kumar, N., R.R. Parmenter, and V.M. Kenkre, *Extinction of refugia of hantavirus infection in a spatially heterogeneous environment.* Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics, 2010. **82**(1 Pt 1): p. 011920.

13. Boone, J.D., et al., *Remote sensing and geographic information systems: charting Sin Nombre virus infections in deer mice.* Emerg Infect Dis, 2000. **6**(3): p. 248-58.

14. Jonsson, C.B., L.T. Figueiredo, and O. Vapalahti, *A global perspective on hantavirus ecology, epidemiology, and disease.* Clinical Microbiology Reviews, 2010. **23**(2): p. 412-41.

15. Luis, A.D., et al., *The effect of seasonality, density and climate on the population dynamics of Montana deer mice, important reservoir hosts for Sin Nombre hantavirus.* J Anim Ecol, 2010. **79**(2): p. 462-70.

16. Previtali, M.A., et al., *Roles of human disturbance, precipitation, and a pathogen on the survival and reproductive probabilities of deer mice.* Ecology, 2010. **91**(2): p. 582-92.

17. Lambin, E.F., et al., *Pathogenic landscapes: interactions between land, people, disease vectors, and their animal hosts.* International Journal of Health Geographics, 2010. **9**: p. 54.

18. Glass, G.E., et al., *Using remotely sensed data to identify areas at risk for hantavirus pulmonary syndrome.* Emerging Infectious Diseases, 2000. **6**(3): p. 238-47.

19. Saasa, N., et al., *Ecology of hantaviruses in Mexico: genetic identification of rodent host species and spillover infection.* Virus Research, 2012. **168**(1-2): p. 88-96.

20. Bennett, A.F., J.Q. Radford, and A. Haslem, *Properties of land mosaics: Implications for nature conservation in agricultural environments.* Biological Conservation, 2006. **133**(2): p. 250-264.

21. Langlois, J., et al., *Landscape structure influences continental distribution of hantavirus in deer mice.* Landscape Ecology, 2001. **16**(3): p. 255-266.

22. MacArthur, R.H. and E.O. Wilson, *The theory of island biogeography*. Monographs in population biology. 1967, Princeton, N.J.,: Princeton University Press. xi, 203 p.

23. Reperant, L.A., *Applying the theory of island biogeography to emerging pathogens: toward predicting the sources of future emerging zoonotic and vector-borne diseases.* Vector Borne Zoonotic Dis, 2010. **10**(2): p. 105-10.

24. Levins, R., *Evolution in changing environments; some theoretical explorations*. Monographs in population biology,. 1968, Princeton, N.J.,: Princeton University Press. ix, 120 p.

25. Hanski, I. and O. Ovaskainen, *Metapopulation theory for fragmented landscapes.* Theoretical Population Biology, 2003. **64**(1): p. 119-127.

26. National.Cancer.Institute. *NCI Dictionary of Cancer Terms: consecutive case series*. [cited 2013; Available from: http://www.cancer.gov/dictionary?CdrID=285747.

27. United.States.Census.Bureau. *ZIP Code Tabulation Areas (ZCTA)*. [cited 2013; Available from: http://www.census.gov/geo/reference/zctas.html.

28. Ozonoff, A., et al., *Effect of spatial resolution on cluster detection: a simulation study.* Int J Health Geogr, 2007. **6**: p. 52.

29. Oregon.State.University. *PRISM Climate Group*. [cited 2012; Available from: http://prism.oregonstate.edu/.

30. USDA and NRCS. *PRISM*. [cited 2013 April 4, 2013]; Available from: http://www.wcc.nrcs.usda.gov/climate/prism.html.

31. Jackson, M.C., et al., *Comparison of tests for spatial heterogeneity on data with global clustering patterns and outliers.* Int J Health Geogr, 2009. **8**: p. 55.

32. Waller, L.A. and C.A. Gotway, *Applied Spatial Statistics for Public Health Data*. Wiley Series in Probability and Statistics ed. D.J. Balding;, et al. 2004, Hoboken, New Jersey John Wiley & Sons, Inc.

33. Gomez-Rubio, V. and A. Lopez-Quilez, *Statistical Methods for the Geographical Analysis of Rare Diseases.* Rare Diseases Epidemiology, 2010. **686**: p. 151-171.

34. Oden, N., *Adjusting Morans-I for Population-Density.* Statistics in Medicine, 1995. **14**(1): p. 17-26.

35. Waller, L.A., E.G. Hill, and R.A. Rudd, *The geography of power: Statistical performance of tests of clusters and clustering in heterogeneous populations.* Statistics in Medicine, 2006. **25**(5): p. 853-865.

36. Anselin, L., *Local Indicators of Spatial Association - Lisa.* Geographical Analysis, 1995. **27**(2): p. 93-115.

37. Osadebe, L.U., et al., *Notes from the Field: Hantavirus Pulmonary Syndrome in Visitors to a National Park - Yosemite Vally, California, 2012*, in *Morbidity and Mortality Weekly Report (MMWR)*. 2012, Centers for Disease Control and Prevention. p. 952.

38. Jones, S.G. and M. Kulldorff, *Influence of spatial resolution on space-time disease cluster detection.* PLoS One, 2012. **7**(10): p. e48036.