

# Journal of Intelligence Studies in Business



## Journal of Intelligence Studies in Business

Publication details, including instructions for authors and subscription information: <https://ojs.hh.se/index.php/JISIB/index>

### Geospatial analysis of census data for targeting new businesses using geoeconomics

Sushant K. Singh<sup>a</sup>

<sup>a</sup>VirtusaPolaris Corporation, Santa Clara, California, USA;  
sushantorama@gmail.com

**To cite this article:** Singh, S.K. (2016) Geospatial analysis of census data for targeting new businesses using geoeconomics. *Journal of Intelligence Studies in Business*. 6 (3) 5-12.

**Article URL:** <https://ojs.hh.se/index.php/JISIB/article/view/175>

## PLEASE SCROLL DOWN FOR ARTICLE

This article is Open Access, in compliance with Strategy 2 of the 2002 Budapest Open Access Initiative, which states:

Scholars need the means to launch a new generation of journals committed to open access, and to help existing journals that elect to make the transition to open access. Because journal articles should be disseminated as widely as possible, these new journals will no longer invoke copyright to restrict access to and use of the material they publish. Instead they will use copyright and other tools to ensure permanent open access to all the articles they publish. Because price is a barrier to access, these new journals will not charge subscription or access fees, and will turn to other methods for covering their expenses. There are many alternative sources of funds for this purpose, including the foundations and governments that fund research, the universities and laboratories that employ researchers, endowments set up by discipline or institution, friends of the cause of open access, profits from the sale of add-ons to the basic texts, funds freed up by the demise or cancellation of journals charging traditional subscription or access fees, or even contributions from the researchers themselves. There is no need to favor one of these solutions over the others for all disciplines or nations, and no need to stop looking for other, creative alternatives.



## Geospatial analysis of census data for targeting new businesses using geoeconomics

Sushant K. Singh<sup>a</sup>

<sup>a</sup>*Virtusa Polaris Corporation, Santa Clara, California, USA* \*Corresponding author: [sushantorama@gmail.com](mailto:sushantorama@gmail.com)

Received 10 October 2016; accepted 25 November 2016

**ABSTRACT** Geoeconomics plays a vital role in encouraging goods and services on new marketplaces. Selecting a “sweet-spot” for new businesses is one of the biggest challenges for new entrepreneurs, enterprises, and investors, especially in the restaurant industry. This paper aims to present a novel geospatial methodological approach for new businesses using census data to answer an important business question: Where I should start my new Asian cuisine restaurant? State and zip code tabulation area (ZCTA) level data on race and income, downloaded from the US census website, were applied for the analysis. ArcGIS software was used as a geospatial analytics tool for hotspot analysis and for producing maps. Based on the state level standard deviation map, California was found to have the second-highest relative Asian population as gauged by the standard deviation (Std. Dev.) from the mean (1.5-2.5 Std. Dev.), after Hawaii (>2.5 Std. Dev.), and followed by New Jersey, New York, Nevada, and Washington. The state of California was selected for further investigation. Seventeen of 58 counties were found to be Asian community hotspots in California. A majority (48%, 854 of 1763) of the ZCTA were found to be Asian community hotspots in these zip codes in this state, and this was statistically significant. Only 9% (163 of 1763) of the ZCTA were not statistically significant Asian community hotspots, while 43% of the ZCTA were found to be statistically significant coldspots of Asian communities in California. Among the 17 hotspot counties of Asian communities, 14 were also derived as hotspots of mean income. The road layer map revealed that these ZCTAs are well connected to major roads in the state. New entrepreneurs, enterprises, and investors, those who are willing to open and or invest in new restaurants, but are not sure about the location, could target hotspot ZCTAs in these counties for Asian cuisine. Integrating ArcGIS with census data for producing maps of statistically significant potential business locations could be used as an important decision-making tool for opening new businesses.

**KEYWORDS** Analytics, ArcGIS, Asian, business, California, census, geospatial, restaurants, ZCTA

### 1. INTRODUCTION

Geoeconomics is described as a theoretical and an applied science, and a methodical trend in socioeconomic geography, and can be applied in temporal, spatial, and political economic systems encouraging goods and services in new marketplaces (Alayev 1983, Anokhin and Lachininskii 2015). It is also considered to be a multidisciplinary science investigating

economic activities and is defined as “the study of spatial, cultural, and strategic aspects of resources, with the aim of gaining a sustainable competitive advantage” (Søilen 2012). Since geoeconomics lies on a trifold of scientific domains, including sociology, geography, and economics, each component plays a vital role in promoting new economic activities at local and global levels (Renner

1942, Lachininskii 2013, Anokhin and Lachininskii 2015). Therefore, the geoeconomic space, which is a complex network transborder system, is vital in promoting new business activities (Gay 2012, Søylen 2012, Anokhin and Lachininskii 2015). Furthermore, the local economic groups may strongly influence the regional economic performance (Porter 2003). Besides, the economic, legal, political, infrastructure, ecological, technical, cultural, and social factors are vital macro-environments that help in business decision-making processes (Søylen 2012).

Selecting a “sweet-spot” for new businesses is one of the biggest challenges for new entrepreneurs, enterprises, and investors. In this study, a “sweet-spot” is defined as a geographic location where the likelihood of maximizing the benefits is the highest for a new business, considering the local environments. The location information plays a vital role in the establishment of new industries as it impacts the economic growth of the firms as well as the socioeconomic, environmental, and political status of the establishment area (Bhat et al. 2014, Demiriz and Ekizoğlu 2015, Mishra et al. 2015). The locational information could help in retail site selection (Karadeniz 2009), preventing retail banking fraud (Demiriz and Ekizoğlu 2015), financing commercial real estate acquisitions by real estate investment trusts (Conklin et al. 2016), fast-food industry (Austin et al. 2005) and much more. Moreover, the location could also impact the business-level innovation (Jordan 2015). The impact of taxes, subsidies and incentives, environmental regulations, quality of life and amenities, labor costs and availability, technical infrastructure, transportation, and accessibility have been reported as the most important factors in the assessment of finding potential locations for new businesses (Kimelberg and Williams 2013, Bhat et al. 2014). However, the socioeconomics and demographics of the potential customers are largely ignored in most of the studies. This could adversely impact a newly established business or firm. For example, in the restaurant industry, the theme, food quality, ambiance, aesthetics, the service, and economic shifts play a vital role in the success and the failure of the business (Murillo 2010, Kimelberg and Williams 2013). Nevertheless, selecting the wrong customer neighborhood, poor accessibility, and a less dense population in the surroundings may unfavorably impact the new establishments (Murillo 2010).

According to location theory, firms or enterprises tend to assess where and why economic activities happen so that they can maximize benefits (North 1955, Kimelberg and Williams 2013, Dubé et al. 2016). In this process, in most cases, non-spatial data, which are outcomes of small or large surveys, have been applied assessing the potential locations for establishing a new business (Kimelberg and Williams 2013). These surveys are very expensive and have inherent reliability challenges. Therefore, targeting a location for a new business based on the analysis of the survey data with a small sample size could be a big concern for entrepreneurs, investors, and enterprises.

The United States (US) census captures socioeconomics, demographics, and business information of the US population that could be used in business decision-making. However, this non-spatial data still lacks a geographical/spatial context. Spatial data have the advantage of showing patterns on maps and letting the users connect the dots, taking neighborhood geographies into consideration. Historically, presenting facts and figures on a map have played a vital role in both the political and the economic contexts (Søylen 2012). Geospatial analytics have emerged as an important method for the spatial and temporal analysis of data in various domains for informed decision-making (Prato et al. 1995, Boulos et al. 2011, Rey et al. 2015, Singh 2015, Singh and Vedwan 2015, Supak et al. 2015, Singh et al. 2016). However, in business, it is still in the rudimentary stages. The reason could be the lack of vision for integrating geospatial tools, such as Arc Geographical Information System (ArcGIS), in business decision-making. Similarly, the census data has been always available for public and private use, though it has not been integrated and/or used in business decision-making.

The main goal of this paper is to attempt to bridge the above-mentioned gaps and integrate US census data with ArcGIS to help new businesses answer an important business question: Where I should start my new Asian cuisine restaurant? There could be several business questions similar to this.

The state of California in the USA has been the center of economic growth with maximum wages and opportunities (Porter 2003). Because of the high rate of diverse immigration, the state became a hotspot for ethnic cuisine specially restaurants (Porter 2003, Capps 2007). The restaurant industry is

one of the fastest growing industries, and small to large business entities could be impacted if a poor site is chosen to start a business. The small business entities may not have enough resources to evaluate site selection. Therefore, the current approach, described in this paper, could be a cost-effective and more efficient way of selecting a site for new businesses within the restaurant industry. However, the approach could be adapted for any other industries.

## 2. MATERIALS AND METHODS

The state and the zip code level population data for Asian immigrants and household mean income in the zip code areas were used. State and zip code tabulation area (ZCTA) shapefiles were applied to perform geospatial analyses and to create maps.

### 2.1 Data collection

#### 2.1.1 State and ZCTA level Asian population data

The state level race data was downloaded from the American Community Survey (ACS). A detailed description of the ACS and the data is found in the 2016 US Census (US-Census 2016). The ACS offers a total of nine different combinations of races at ZCTA level. For the current study, only HD01\_VD05 (i.e. Asian

alone) was selected. A detailed description on the ZCTA can be found at the US census site (US-Census 2016). In brief, the “ZCTAs are generalized areal representations of the United States Postal Service (USPS) ZIP Code service areas,” however, the “USPS ZIP Codes are not areal features but a collection of mail delivery routes” (US-Census 2016).

#### 2.1.2 ZCTA level income data

The ZCTA level mean income data was downloaded from the ACS. The ACS offers a total of 27 different combinations of mean income at ZCTA level. For the current study, only HC02\_EST\_VC02 (i.e. estimated mean income in dollars by all households) was selected (US-Census 2016).

#### 2.1.3 State and ZCTA shapefiles

State and ZCTA shapefiles were downloaded from the US census website (US-Census 2014, 2015, 2016). These shapefiles were used to produce maps for geospatial analysis.

### 2.2 Data integration to ArcGIS, analysis, and mapping

The state and the ZCTA data were joined to the shapefiles within the ArcGIS environment using GEOID as a join key (ESRI 2012). ArcGIS is a mapping software, developed by

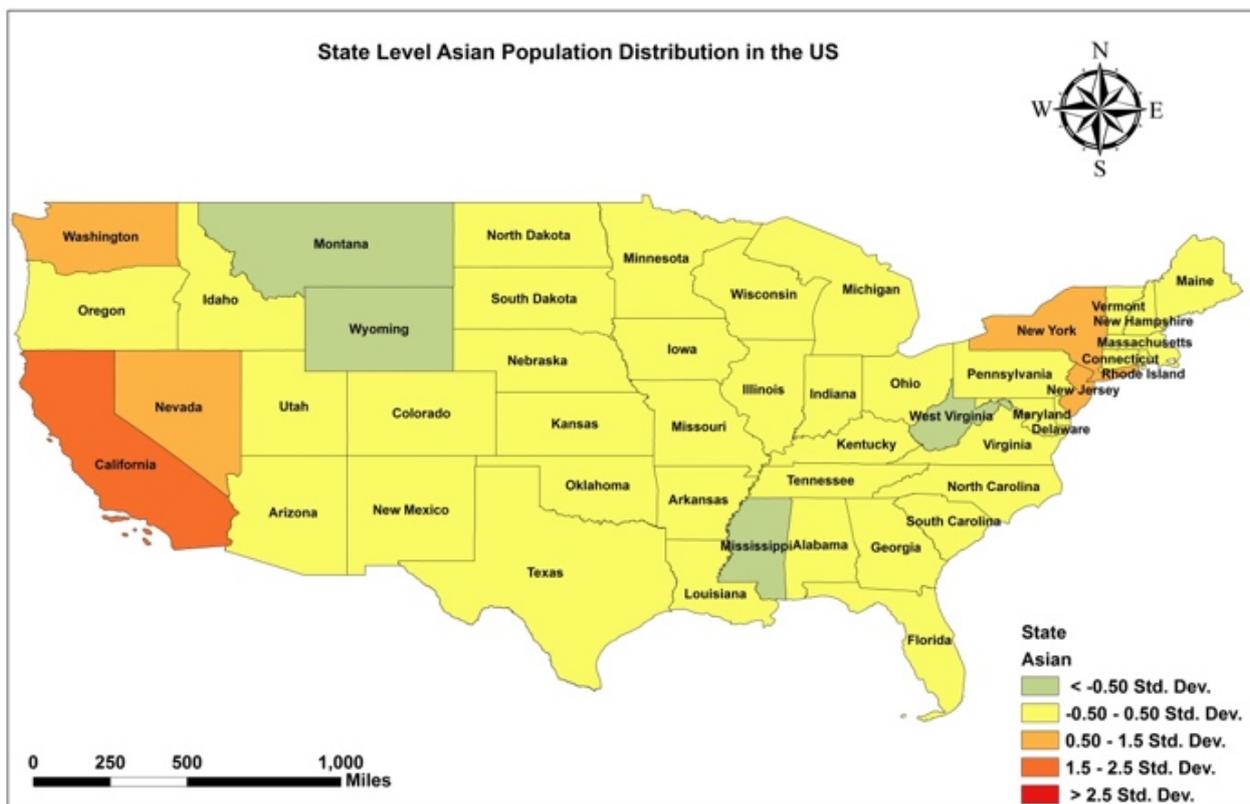


Figure 1 A standard deviation map of the Asian population in the United States.

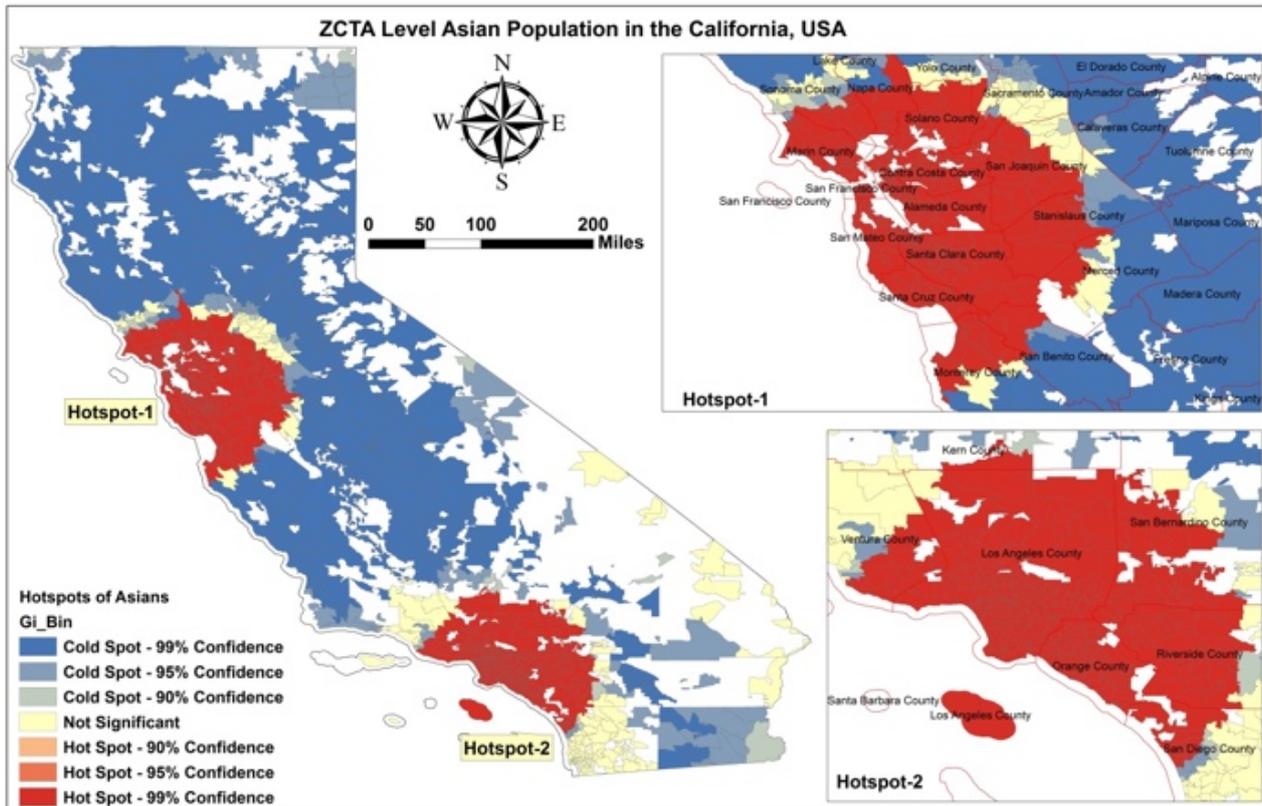


Figure 2 Hotspot and coldspot map of Asian communities in the Zip Code Tabulation Areas of California, United States.

the Environmental Systems Research Institute (ESRI), that offers several geospatial analytical tools (ESRI 2014). For all mapping and geospatial analyses, ArcGIS version 10.3.1 was used (ESRI 2014).

### 2.2.1 Standard deviation mapping

There are seven standard classification methods (manual, defined, equal, geometrical interval, quantile, natural breaks, and standard deviation) available in ArcGIS to spatially display numerical data on a map (ESRI 2012, 2014, 2016). The state level Asian population data, used in this study, is available in absolute numbers and in percentage. The standard deviation classification method was applied to produce a classification map of Asian populations at the state level. In this method, ArcMap derives the mean and standard deviation and produce maps displaying which feature polygons deviate (positively and negatively) from the mean (ESRI 2014, 2015). Based on the positive deviation and the highest standard deviation values, California was selected for further analysis.

### 2.2.2 Hotspot analysis

The hotspot analysis is one of the spatial statistical analysis tools in ArcMap that was applied for mapping spatial statistically significant clusters of high values (hotspots)

and low values (coldspots) (ESRI 2014, 2016). The output feature class is in the form of a shapefile with a Giz-score, Gip-value, and Gi\_Bin. The Giz-score and Gip-value measure the statistical significance and the Gi\_Bin represents the confidence intervals at 90, 95, and 99% (ESRI 2014, 2016).

## 3. RESULTS AND DISCUSSION

### 3.1 State level distribution of Asian population in the US

The analysis revealed that the highest relative percentage (38%) of people identifying as Asian live in Hawaii, followed by 13% in California, 8.6% in New Jersey, 7.6% in New York, 7.4% in Nevada, and 7.3% in Washington. Although Hawaii has the highest relative Asian population, California was used for this case study because of the availability of other relevant data and the interest of new entrepreneurs, enterprises, and investors in California (Figure 1).

Later, the focus for further investigation was California and all the ZCTAs in the state.

### 3.2 ZCTA level distribution of Asian populations in the US

The ZCTA is the smallest census unit and may offer more specific information on the



counties with the hotspots of Asian communities and mean income were derived (Table 1).

*Table 1* Hotspot counties and the coverage of Asian population and mean income in California, United States.  
\*NHS = not a hotspot.

Sl. No	Asian population Hotspots	Coverage of the county	
		Asian	Mean Income
1	Contra Costa County	Entire	Entire
2	Los Angeles County	Entire	Entire
3	Marin County	Entire	Entire
4	Merced County	Partial	NHS*
5	Monterey County	Partial	Partial
6	Orange County	Entire	Entire
7	Riverside County	Partial	Partial
8	San Bernardino County	Partial	NHS*
9	San Diego County	Partial	Partial
10	San Francisco County	Entire	Entire
11	San Joaquin County	Partial	Partial
12	San Mateo County	Entire	Entire
13	Santa Clara County	Entire	Entire
14	Santa Cruz County	Entire	Entire
15	Solano County	Partial	Entire
16	Stanislaus County	Partial	NHS*
17	Ventura County	Partial	Entire

Out of 17 counties with hotspots of Asian communities (Figure 2), 14 counties were either fully or partially identified as the hotspots of mean income (Table 1, Figure 2 and 3) in California. Only three hotspot counties with Asian communities, including Merced, San Bernardino, and Stanislaus were not hotspots of

mean income (Table 1). Therefore, the above 14 counties could be targeted for new business establishments of Asian cuisine restaurants (Table 1).

### 3.4 Hotspot ZCTAs with roads, availability in California, US

As discussed in the introduction, the accessibility to the facilities plays an important role in the success of a newly established business. Therefore, the hotspot map of the Asian population was overlaid with a road layer to see whether the hotspot ZCTAs are close enough to roads and accessibility is not a constraint (Figure 4).

Both of the hotspots were found to have a good network of roads (Figure 4). Consequently, the 14 listed counties (Table 1) meet three major criteria for opening an Asian cuisine restaurant in California: high Asian population, higher mean income, and greater road network.

## 4. CONCLUSIONS

Geoeconomics can be examined at the lowest, low, middle, upper, and top geographical levels (Anokhin and Lachininskii 2015). In this study, the lowest geographic unit, the ZCTA, was used to show how social, economic, and geographic components could potentially impact new businesses. Additionally, ArcGIS offers several tools for geospatial analysis and could be used as decision-making tool for informed business management within geoeconomics. Moreover, census data offer various significant insights for the establishments of new businesses as well as for the existing businesses. Integrating ArcGIS with census data could help businesses accomplish their goals, starting from locating new sites to predicting their business growth. The socioeconomic and demographic data are freely available at the ZCTA level. These could be applied for informed business decision-making through geospatial analytics using ArcGIS. In this study, only race, mean income, and road network data were used for the analysis. However, the US census offers several other important attributes such as gender, age, occupation, and household, housing, and business data, and much more. These attributes could be applied in the selection of new business sites and/or for other business needs. The current geospatial approach described in this paper is a cost-effective, easy, and efficient way of selecting

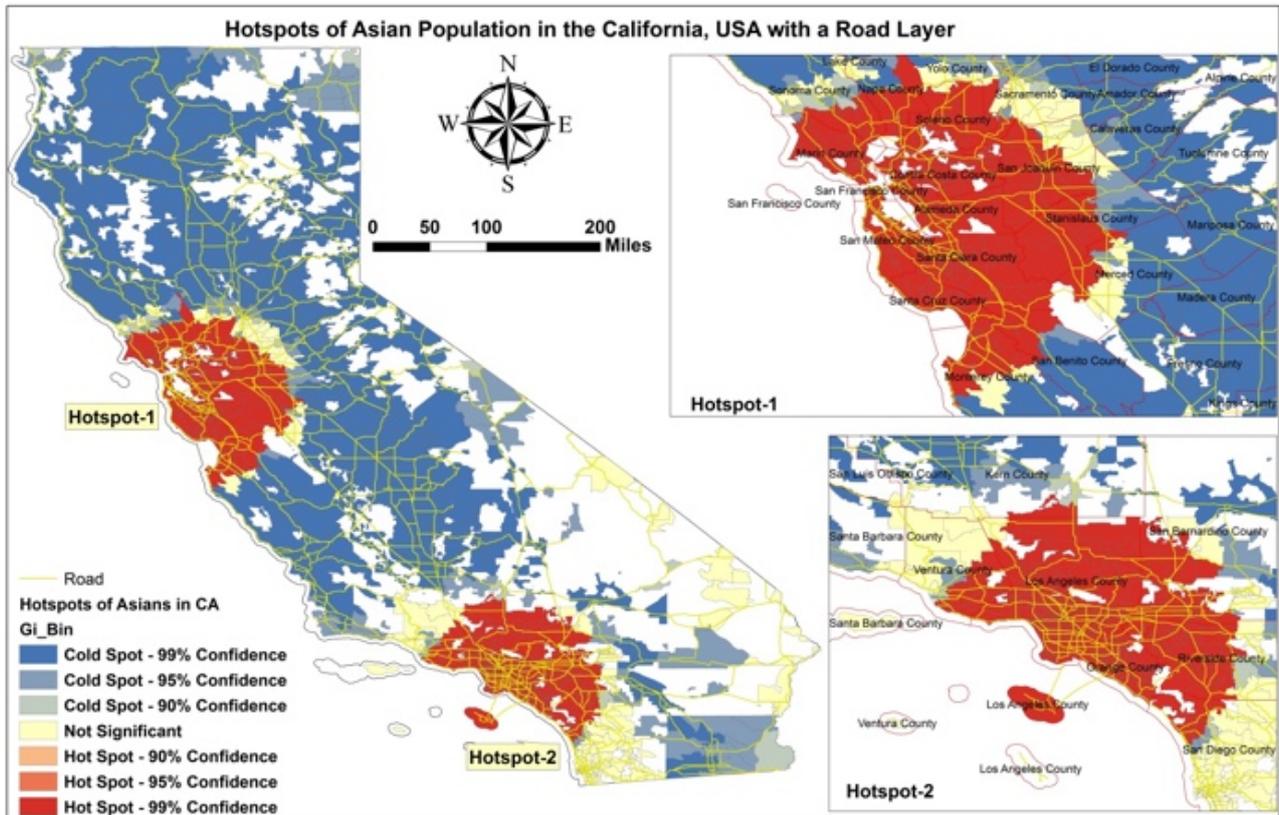


Figure 4 Hotspot and coldspot map of the Asian population overlaid with a road layer in the Zip Code Tabulation Areas of California, United States.

new business sites and could be used by small as well as large business entities within and outside the restaurant industry.

### Acknowledgements

This work is purely for research purposes and does not necessarily represent the official views of the organization author is associated with. The author did not receive any grant for this study and declares no conflict of interest publishing this article.

### 5. REFERENCES

- Alayev, E. (1983). "Sotsial'no-ekonomicheskaya geografiya. Ponyatiyno-terminologicheskii slovar'." M. Mysl.
- Anokhin, A. and S. Lachininskii (2015). "Evolution of the ideas and contents of geoeconomic studies." *Regional Research of Russia* 5(1): 90-95.
- Austin, S. B., S. J. Melly, B. N. Sanchez, A. Patel, S. Buka and S. L. Gortmaker (2005). "Clustering of fast-food restaurants around schools: a novel application of spatial statistics to the study of food environments." *American Journal of Public Health* 95(9): 1575-1581.
- Bhat, C. R., R. Paleti and P. Singh (2014). "A spatial multivariate count model for firm location decisions." *Journal of Regional Science* 54(3): 462-502.
- Boulos, M. N. K., T. Viangteeravat, M. N. Anyanwu, V. R. Nagisetty and E. Kucu (2011). "Web GIS in practice IX: a demonstration of geospatial visual analytics using Microsoft Live Labs Pivot technology and WHO mortality data." *International Journal of Health Geographics* 10(1): 1.
- Capps, R. (2007). "The characteristics of unauthorized immigrants in California, Los Angeles County, and the United States."
- Conklin, J. N., M. Diop and M. Qiu (2016). "The Importance of Location in the Financing of Property Acquisitions by REITs." Available at SSRN 2741222.
- Demiriz, A. and B. Ekizoğlu (2015). Using location aware business rules for preventing retail banking frauds. *Anti-Cybercrime (ICACC), 2015 First International Conference on, IEEE.*
- Dubé, J., C. Brunelle and D. Legros (2016). "Location Theories and Business Location Decision: A Micro-Spatial Investigation in Canada." *The Review of Regional Studies* 46(2): 143.
- ESRI (2012). *ArcGIS for Desktop 10.1*. USA, Environmental Systems Research Institute.
- ESRI (2014). *ArcGIS 10.3 for Desktop*, Environmental Systems Research Institute Redlands, CA USA.

- ESRI. (2015). "ArcMap: Classifying Numerical Fields for Graduated Symbolology." Retrieved September 03, 2015, 2016, from <http://desktop.arcgis.com/en/arcmap/10.3/map/working-with-layers/classifying-numerical-fields-for-graduated-symbols.htm>.
- ESRI. (2016). "ArcGISPro: Data Classification Methods." Retrieved September 03, 2016, 2016, from <http://pro.arcgis.com/en/pro-app/help/mapping/symbols-and-styles/data-classification-methods.htm>.
- ESRI. (2016). "ArcMap: Hot Spot Analysis (Getis-Ord Gi\*)." Retrieved September 03, 2016, 2016, from <http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/hot-spot-analysis.htm>.
- Gay, B. (2012). "Competitive Intelligence and Complex Systems." *Journal of Intelligence Studies in Business* 2(2).
- Jordan, D. (2015). "Location and business-level product innovation in Vietnam: Regional differences and drivers." *Australasian Journal of Regional Studies* 21(2): 232.
- Karadeniz, M. (2009). "The importance of retail site selection in marketing management and hypothetical approaches used in site selection." *Deniz Bilimleri ve Mühendisliği Dergisi* 5(3).
- Kimelberg, S. M. and E. Williams (2013). "Evaluating the importance of business location factors: The influence of facility type." *Growth and Change* 44(1): 92-117.
- Lachinskii, S. (2013). "Experience on the typology of geoeconomic risks." *Geography and Natural Resources* 34(2): 111-117.
- Mishra, S., S. P. Mohanty, S. K. Pradhan and R. Hota (2015). "Importance of Location Classification using Rough Set Approach for the Development of Business Establishment." *International Journal of Computer Applications* 119(24).
- Murillo, L. (2010). "Elements food group marketing plan."
- North, D. C. (1955). "Location theory and regional economic growth." *The Journal of Political Economy*: 243-258.
- Porter, M. (2003). "The economic performance of regions." *Regional studies* 37(6-7): 549-578.
- Prato, T., Y. Wang, T. Haithcoat, C. Barnett and C. Fulcher (1995). "Converting hydric cropland to wetland in Missouri: a geoeconomic analysis." *Journal of soil and water conservation* 50(1): 101-106.
- Renner, G. T. (1942). *Human Geography in the Air Age (a Text for High School Students)*, Macmillan.
- Rey, S. J., L. Anselin, X. Li, R. Pahle, J. Laura, W. Li and J. Koschinsky (2015). "Open geospatial analytics with PySAL." *ISPRS International Journal of Geo-Information* 4(2): 815-836.
- Singh, S. K. (2015). Assessing and mapping vulnerability and risk perceptions to groundwater arsenic contamination: Towards developing sustainable arsenic mitigation models (Order No. 3701365). Available from ProQuest Dissertations & Theses Full Text. (1681668682). Ph.D. Dissertation/Thesis, Montclair State University.
- Singh, S. K., S. A. Brachfeld and R. W. Taylor (2016). Evaluating hydrogeological and topographic controls on groundwater arsenic contamination in the mid-Gangetic Plain in India: Towards Developing Sustainable Arsenic Mitigation Models. *Emerging Issues in Groundwater Resources, Advances in Water Security*. A. Fares. Switzerland, Springer International Publishing.
- Singh, S. K. and N. Vedwan (2015). "Mapping Composite Vulnerability to Groundwater Arsenic Contamination: An Analytical Framework and a Case Study in India." *Natural Hazards* 75(2): 1883-1908.
- Søilen, K. S. (2012). *Geoeconomics*, Bookboon.
- Supak, S., G. Brothers, D. Bohnenstiehl and H. Devine (2015). "Geospatial analytics for federally managed tourism destinations and their demand markets." *Journal of Destination Marketing & Management* 4(3): 173-186.
- UC-Census. (2014). "Geography: TIGER/Line® Shapefiles and TIGER/Line® Files: Roads." Retrieved September 03, 2016, 2016, from <http://www2.census.gov/geo/tiger/TIGER2014/ROADS/>.
- US-Census. (2015). "Geography: TIGER/Line® Shapefiles and TIGER/Line® Files." Retrieved September 03, 2015, 2015, from <https://www.census.gov/geo/maps-data/data/tiger-line.html>.
- US-Census. (2016). "American Fact Finder: American Community Survey." Retrieved September 03, 2016, 2016, from <http://factfinder.census.gov/faces/nav/jsf/pages/programs.xhtml?program=acs>.
- US-Census. (2016). "Geography: Cartographic Boundary Shapefiles - ZIP Code Tabulation Areas (ZCTAs)." Retrieved September 03, 2016, 2016, from [https://www.census.gov/geo/maps-data/data/cbf/cbf\\_zcta.html](https://www.census.gov/geo/maps-data/data/cbf/cbf_zcta.html).
- US-Census. (2016). "Geography: ZIP Code Tabulation Areas (ZCTAs)." Retrieved September 03, 2016, 2016, from <https://www.census.gov/geo/reference/zctas.html>.