Residential Ethnic Clusters in Denver

MATH 6384: SPATIAL DATA ANALYSIS PROJECT

Siu Yin Lee

DEC 7, 2021

1. Introduction/Background

1.1 Ethnic Cluster and its Complex Dynamic

"An ethnic enclave/cluster is a geographical area where a particular ethnic group is spatially clustered and socially and economically distinct from the majority group." (Lim, Yi, De La Cruz, & Chau, 2019) In many cities in the United States, it is not hard to find neighborhoods concentrated with residents of a certain ethnic group. For example, many big cities have Chinatowns, Greek towns and Little Italys etc. While minority neighborhoods can be perceived as a sign of cultural richness and prosperity of a city and serve as great tourist attractions, studies show mixed results regarding the pros and cons of ethnic clusters. (Klaesson & Oner, 2020) On the upside, ethnic clusters are like shelters for minorities. The distinct geographical space offers its residents cultural connections, job opportunities in intra-ethnic businesses and protection from hostile social atmosphere. (Liu & Geron, 2008) On the downside, historically-speaking, many formations of ethnic clusters were 'involuntary' and were an outcome of social struggles. Ethnic groups, very often, were marginalized geographically due to financial, social and political reasons. (Koga, 2019)

1.2 Ethnic Clusters and the Government's Role

Even though ethnic cluster is a complex phenomenon and must be examined and interpreted within a certain region's historical, economic and social context, at the very least, regional governments who recognize the importance of cultural diversity and racial harmony should be aware of the presence and formation of racial clusters in their districts. Governors should understand the underlying reasons for racial clustering and potentially spend resources to acknowledge the needs of minority residents to help them integrate into the society. As an example, in Singapore, in order to promote racial harmony and prevent severe racial segregation over time, the Ethnic Integration Policy (EIP) has been implemented since 1989 to ensure that there is a balanced mix of various ethnic communities in specific towns. (Housing and Development Board of Singapore, n.d.) Other than housing programs, governments can also consider enhancing supports such as translation services, employment supports and education programs for minority neighborhoods. These are all measures that can help prevent minority groups from being marginalized in the society.

1.3 Research Question: Ethnic Clustering/Clusters in Denver

According to the Census in 2020, Denver scored a Diversity Index[1] (DI) of 61.7%, which is higher than Colorado (52.3%) and the US (61.1%) in general (United States Census Bureau, 2021). While the Census' DI seems to suggest that Denver is a relatively diverse city, in a report published in June 2021 by researchers at University of California Berkeley titled *"A study on 21st century racial residential segregation in the US"*, Denver is categorized as one of the "highly segregated cities"(Menendian, Gambhir, & Gailes, 2021).

To understand Denver's racial demographic and its geospatial distribution, this project aims at answering the following research questions:

How are Denver's ethnic groups geospatially distributed? Do they live in clusters? If so, what are those clusters' characteristics?

By examining the geospatial pattern of racial clustering in Denver, the project hopes to offer insights to public and the governments of Denver and Colorado, so that informed policy decisions can be made to promote racial and social harmony for this fast-changing city.

[1] The Diversity Index is one of the ways the Census uses to measure racial diversity of a region. The Diversity Index is the probability of two people chosen at random in a region coming from different ethnic groups.

1.4 Data

The main Data set used in this project is the Denver neighborhood-level data derived from the American Community Survey of 2015-2019. The data was aggregated at the census tract level and then summarized into neighborhoods by the City and County of Denver. The dataset was downloaded from the City and County of Denver website (Denver GIS, 2021) and is observational data published annually by the US Census Bureau.

2. Data Exploration

Five major Ethnic groups were analyzed in this project: White, Hispanic, Black, Asian and Other race. These ethnic groups' population counts and percentages are present in the ACS dataset as variables, except that the Other race analyzed in this project is an aggregation of 4 race groups of the ACS data: Native Americans, Hawaiian and Pacific Islanders, Other race and Multiple races. The aggregation was performed due to these ethnic groups' small population sizes. For data cleaning, two neighborhoods' Median Home Value and Median Gross Rent were replaced with NA as they were listed as 0 in the data.

2.1. Ethnic Groups Overview in Denver

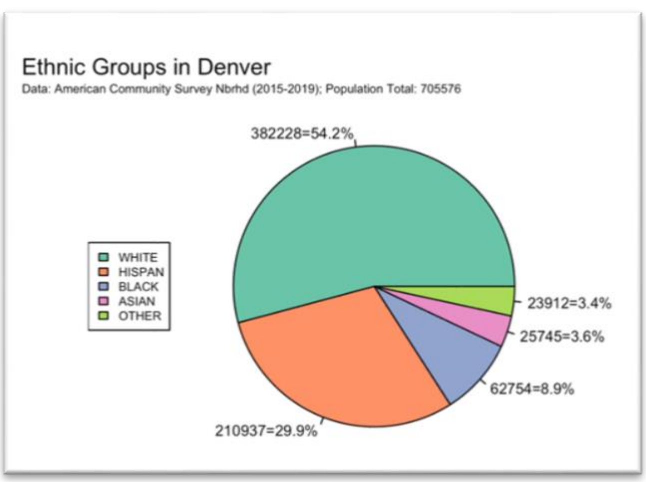

Fig 1: Denver Ethnic Group Percentages (ACS 2015-2019)

	2010-2014	2015-2019	Change
Population	633777	705576	+ 71799
White	52.9 %	54.2 %	+ 1.3 %
Hispanic	31.2 %	29.9 %	- 1.3 %
Black	9.5 %	8.9 %	- 0.6 %
Asian	3.4 %	3.6 %	+ 0.2 %
Other	3.1 %	3.4 %	+ 0.3 %

Fig 2: Change of Denver Ethnic Group Percentages (ACS 2010-2014, 2015-2019)

Denver has a population of over 700k people. As show in Figure 1, more than half (54.2%) of the populations are White. The second biggest ethnic group is Hispanic, which is about 30%, followed up Black (8.9%), Asian (3.6%) and Other race (3.4%). As show in Figure 2, we can see that, for the last decade, there is a slight decrease in the proportion of Hispanic (-1.3%) and Black (-0.6%) residents.

2.2. __Denver Neighborhoods Overview__

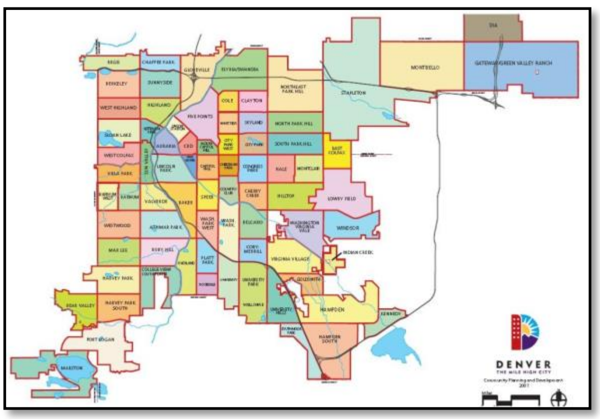

Fig 3: 78 Denver neighborhood map

	78 NBHD Min	78 NBHD Max	78 NBHD Median
MED_HH Income	$ 13,125	$ 197,813	$ 70,936
MED_HOME Value	$ 203.5k	1.1 million	$ 430.3k
White	9.5 %	92.3 %	67 %
Hispanic	3.2 %	82.1 %	19 %
Black	0%	43.6 %	4.1 %
Asian	0%	11.8%	2.5%
Other	0.4%	7.7%	3 %

Fig 4: Quick Statistics Summary for 78 Denver neighborhoods

Denver has 78 neighborhoods of various sizes (Figure 3). These neighborhoods have huge ranges in terms of median household income, median home value and ethnic group proportions. For example, we can see in Figure 4 that the neighborhood (Sun Valley) with the lowest median household income is at 13k while the highest earning neighborhood (Country Club) has a median household income of $198k. Country Club is also the whitest (92.3%) neighborhood with the highest median home value of more than 1 million dollars, while houses in DIA have the lowest median home value of only $203.5k.

2.3. *Quick Glance at Ethnicity Distribution in Denver Neighborhood*

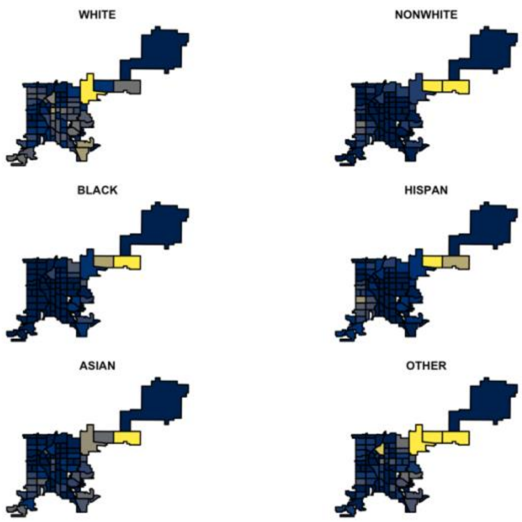

Fig 5: Choropleth Map: Population Count by Ethnicity

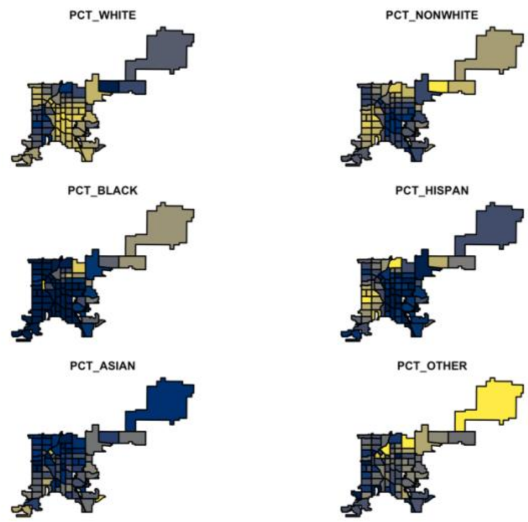

Fig 6: Choropleth Map: Population Percentage by Ethnicity

If we simply use population count and percentage to show the concentration of ethnic groups, we can see in both Figure 5 (Population Count) and Figure 6 (Population Percentage) that there are yellow high concentration spots for each ethnic group. However, depending on how we analyze the data, the high concentration spots' locations can be very different. Therefore, in the next section, we will use more rigorous geospatial statistical methods to detect ethnic clustering.

3. Main Analyses and Results

3.1. *Statistical Methods for Spatial Correlation*

As a first step, two methods (Moran's I and Tango's index) were employed to detect ethnic group clustering in Denver. Testing was conducted under the Constant Risk Hypothesis using 9999 Monte Carlos simulations. Alpha level was set at 0.05.

A binary Spatial Proximity Matrix was used: neighborhoods which share region boundary are considered as neighbors. Figure 7 below shows the neighbor relationships among the neighborhoods.

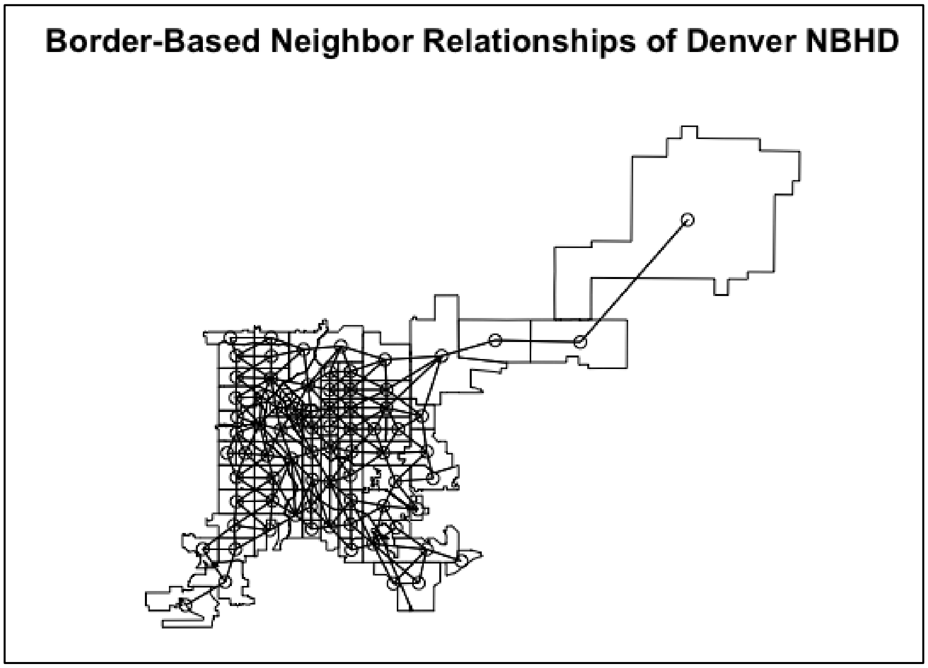

Fig 7: Border-based Neighbor Relationships of Denver Neighborhoods

3.1.1. *Moran's I*

To account for heterogeneity in population and variance, the test statistic used was the Walter (1992) version.

Results of Moran's I:

P-values for all the ethnic groups are 0.0001. Under the CRH assumption, the Moran's I results indicate that there is strong evidence of spatial correlation for all the ethnic groups' cases for the neighborhoods in Denver.

3.1.2. *Tango's Index*

For Tango's index, the Rogerson version was used to detect both clustering within region (the goodness-of-fit component) and clustering between regions (spatial similarity component). According to the distance matrix using the Euclidean calculation method, the intercentroid distances among Denver neighborhoods are between 0.01 and 0.45 units, with mean distance of 0.09 units. For Tango's Index testing, kappa values of 0.01, 0.05, 0.1, 0.2, 0.4 were tested to observe how results may vary according to weaker/stronger spatial correlation.

Results of Tango's Index:

P-values for all the ethnic groups are all nearly 0 (way smaller than alpha level of 0.05) for all the kappa ranges tested. For all the ethnic groups, both the observed goodness-of-fit component and the spatial autocorrelation component are very extreme across all tested kappa values. Just as examples, Fig 8 and 9 show the Tango's statistics for the observed data and simulated data for Other race group using kappa values of 0.01 and 0.4 respectively. The observed data strongly deviates from the simulated data. As all other ethnic groups' Tango plots have very similar patterns, only Fig 8 and 9 are shown as illustration.

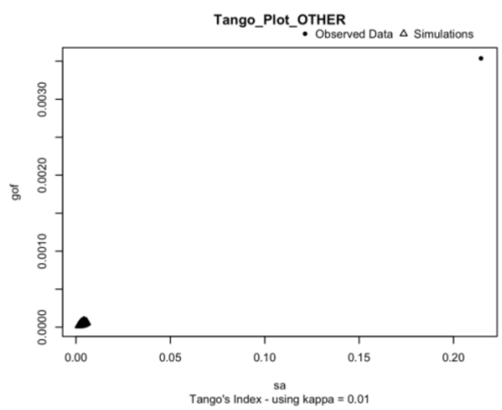

Fig 8: Tango's Stat for Other Ethnic Group using kappa = 0.01

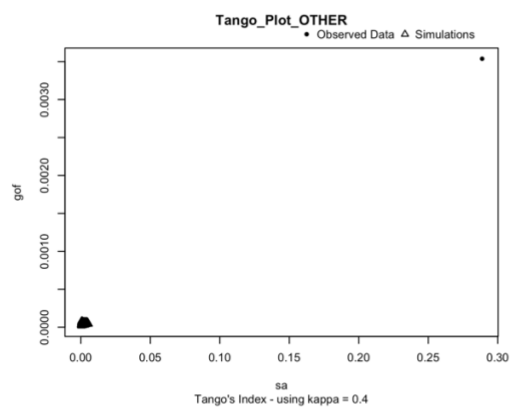

Fig 9: Tango's Stat for Other Ethnic Group using kappa = 0.4

The Tango's Index results indicate that there is strong evidence that the observed proportion of ethnic groups in the neighborhoods in Denver are inconsistent with the CRH.

To conclude, the statistical methods for spatial correlation suggest that there is strong evidence for all ethnic groups in Denver under the CRH assumption. In the next section, we will try to identify the locations of the clusters.

3.2. ***Statistical Methods for Scanning Local Rates***

To detect and locate the specific ethnic group clusters, CEPP (Cluster Evaluation Permutation Procedure), B&N (Besag and Newell) Approach and Spatial Scan Statistics were used. For all these methods, testing was conducted under the Constant Risk Hypothesis. Alpha level was set at 0.05. Various window sizes were tested.

For the CEPP method, N* used include 0.1 and 0.15 of the total Denver population. For the B&N method, C* used include 0.1 and 0.2 of the specific ethnic cases. For the Spatial Scan method, 0.1 and 0.15 were used for ubpop.

Results:

For all the ethnic groups, all methods used identified statistically significant clusters. Some methods identified multiple statistically significant clusters for certain ethnic groups. However, for simplicity and consistency of comparison, only the most-likely-cluster for each ethnic group was examined and plotted in figures below. Even though different methods and window sizes include slightly different neighborhoods, each ethnic group is found to cluster at very specific and distinct area of the city. Results are summarized into the table and graphs below.

	White	**Hispanic**	**Black**	**Asian**	**Other**
MLC location in Denver	Central area	Middle West	Northeast	Southwest	Middle North
Figure # below	11	12	13	14	15

Fig 10: Summary table for Most-likely-cluster location of each ethnic group

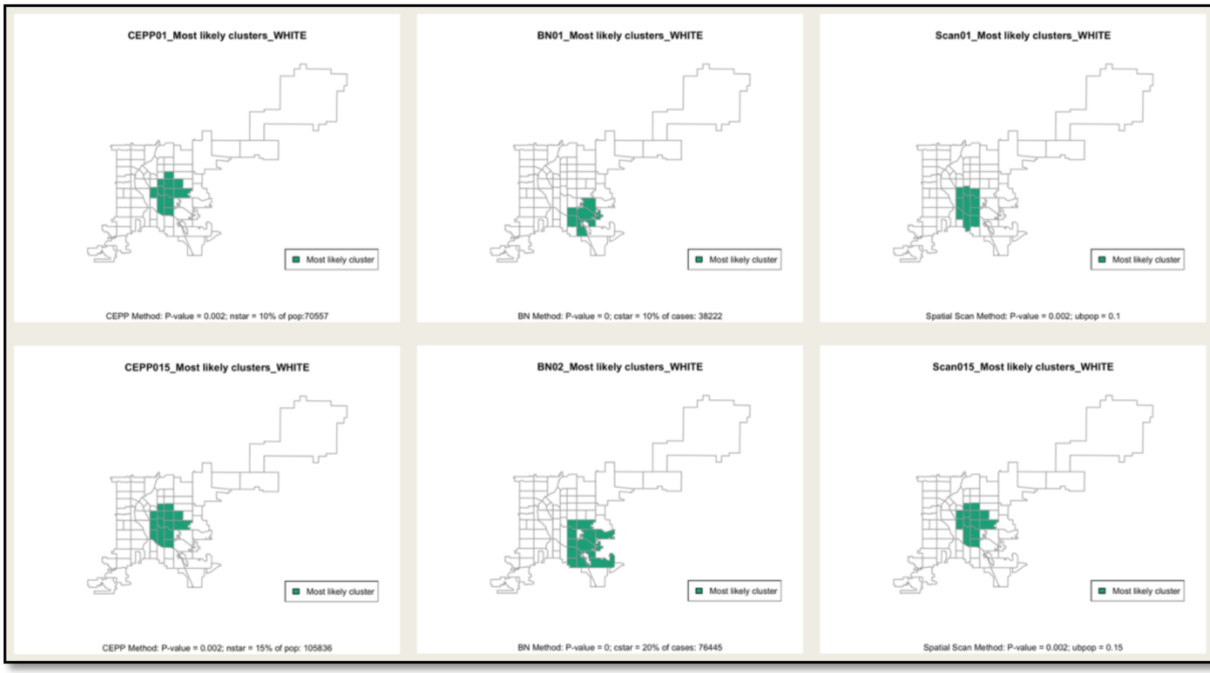

Fig 11: Most-likely-clusters for White ethnic group identified by CEPP, B&N and Spatial Scan methods. Cluster locations for White ethnic group: Central Area of Denver

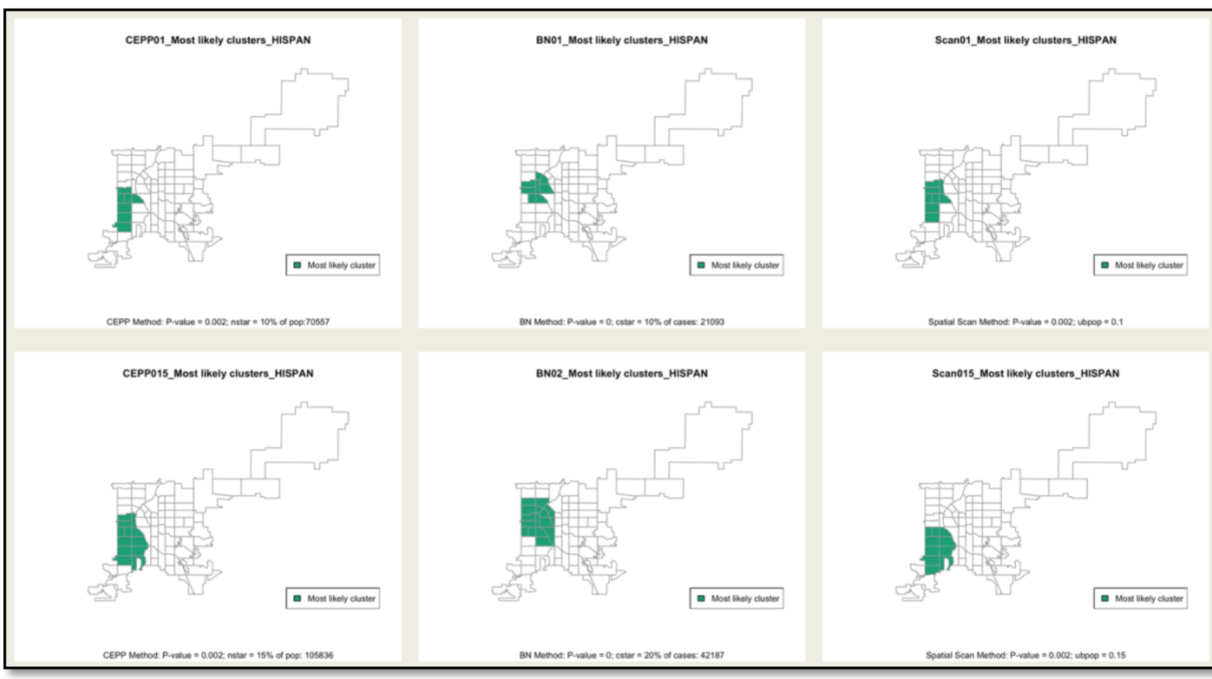

Fig 12: Most-likely-clusters for Hispanic ethnic group identified by CEPP, B&N and Spatial Scan methods. Cluster locations for Hispanic ethnic group: Middle West area of Denver

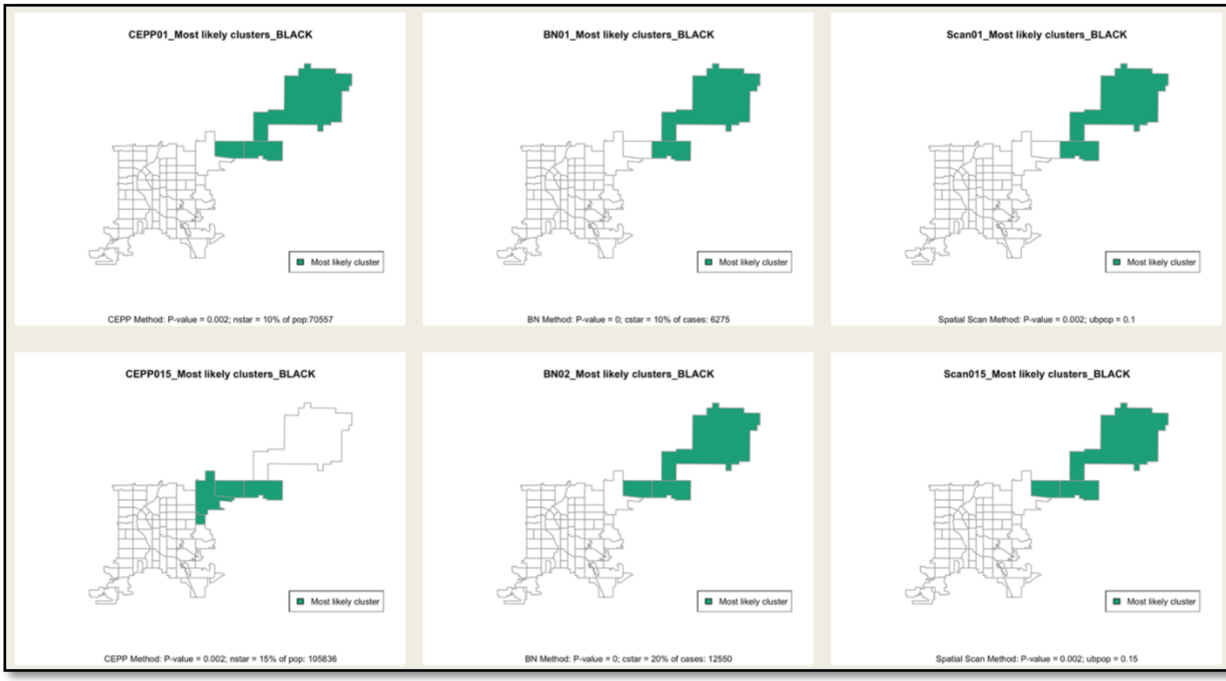

Fig 13: Most-likely-clusters for Black ethnic group identified by CEPP, B&N and Spatial Scan methods. Cluster locations for Black ethnic group: Northeast area of Denver

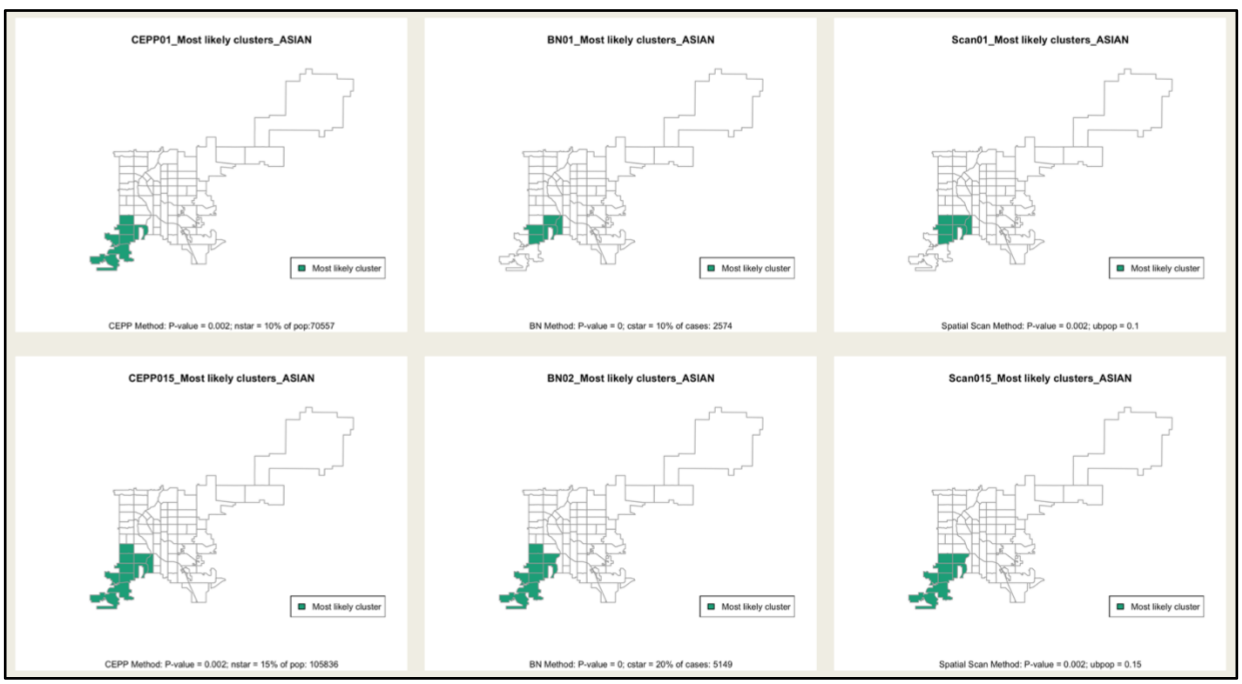

Fig 14: Most-likely-clusters for Asian ethnic group identified by CEPP, B&N and Spatial Scan methods. Cluster locations for Asian ethnic group: Southwest area of Denver

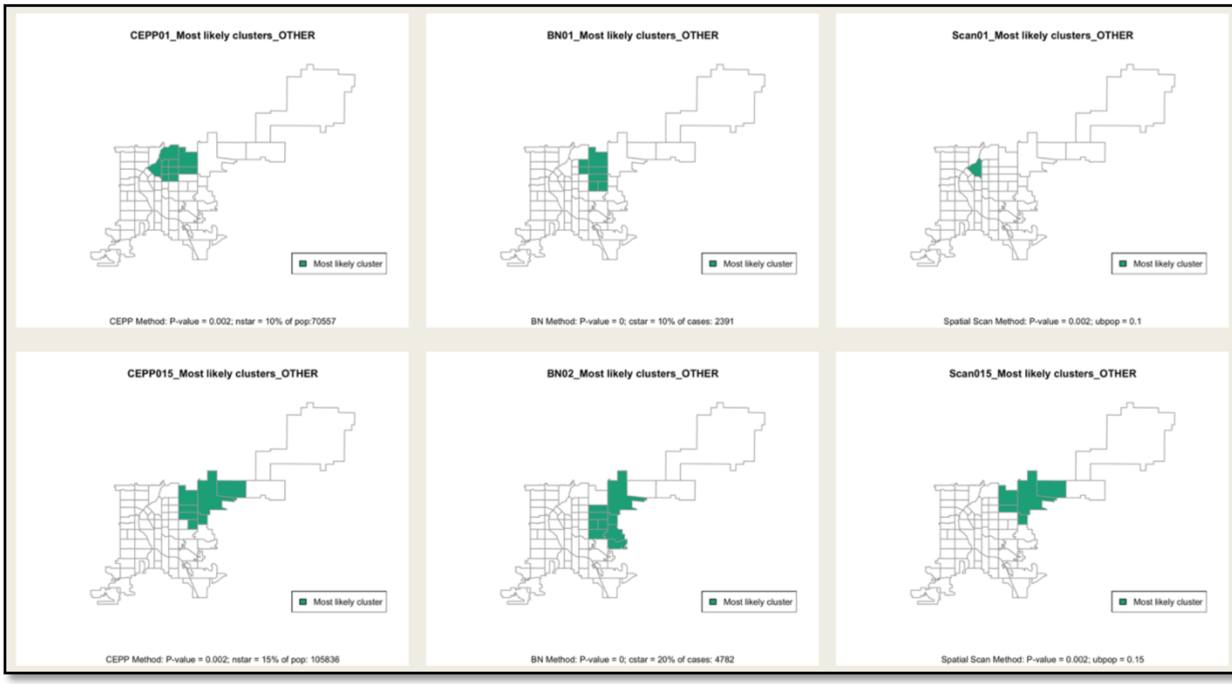

Fig 15: Most-likely-clusters for Asian ethnic group identified by CEPP, B&N and Spatial Scan methods. Cluster locations for Other ethnic group: Middle North area of Denver

3.3. ***Summarizing Ethnic Clusters Characteristics***

Method for averaging

As different methods of scanning local rates include slightly different neighborhoods for each ethnic group's most-likely-cluster, to summarize these clusters' characteristics, an average approach is taken for all neighborhoods identified. For example, if method A has included neighborhoods #1, #2, #3 in the cluster, while method B has included neighborhoods #2, #3 only, to calculate the median home value for the most-likely-cluster for a specific ethnic group, we summed the median home values for neighborhood #1, #2, #3, #2, #3 and then divide it by 5. That was the averaging approach for variables of median values. To illustrate the averaging method for calculating percentages, let's use 'the percentage of renter-occupied household' as an example: First, we summed the total counts of renter-occupied households for neighborhoods #1, #2, #3, #2, #3. Then, we summed the total counts of households for those neighborhoods. At the end, we divided the sum of renter-occupied households by the sum of total households in the selected regions.

Cluster Characteristics Summary in 4 categories

To present the characteristics more clearly, the selected summary variables were grouped into 4 categories: basic demographics, household, economics and housing.

Demographic

Ethnic Clusters	# NBHD in cluster	Foreign Born	Under 18	Above 65
White	11	9%	14%	20%
Hisp	10	23%	26%	12%
Black	3	26%	30%	8%
Asian	8	21%	25%	15%
Other	6	13%	23%	13%

Fig 16: Ethnic Clusters Demographic Summary

Household

Ethnic Clusters	Male HH No Wife	Fem HH No Husb	Family HH	Only English
White	3%	5%	39%	84%
Hisp	6%	15%	59%	51%
Black	7%	15%	74%	49%
Asian	7%	14%	62%	57%
Other	5%	12%	56%	71%

Fig 17: Ethnic Clusters Household Summary

Economic

Ethnic Clusters	Commute 30min +	College Grad	Med HH Income	Family Poverty
White	18%	54%	$103.5k	3.5%
Hisp	22%	17%	$55.9k	20.3%
Black	23%	15%	$62k	8.7%
Asian	24%	17%	$57.6k	10.4%
Other	18%	36%	$91.3k	9%

Fig 18: Ethnic Clusters Economic Summary

Housing

Ethnic Clusters	House Built 2000 +	Med Home Value	Med Rent	Renter Occupied
White	12%	$633.9k	$1604	52%
Hisp	11%	$335.8k	$1016	53%
Black	52%	$249.8k	$1442	32%
Asian	7%	$279.1k	$1115	44%
Other	29%	$479.1k	$1520	41%

Fig 19: Ethnic Clusters Housing Summary

One of the most interesting distinctions among the ethnic clusters is the difference in housing prices. For most people, the main factor that determines where they live is the housing prices. Therefore, housing prices can be a strong contribution to ethnic clustering if there is are distinct differences in ethnic groups' economic classes. As shown in Figure 19, for median home value, the white cluster's neighborhoods are the highest in average, which are at the $600k range. The second highest is Other ethnic cluster at $480k, followed by Hispanic cluster at 330k. The lowest are the Black and Asian clusters' neighborhoods, which are at the 200k range. Also, we can see that Black cluster is located in very new neighborhoods. About 52% of the houses in the cluster were built after 2000.

More descriptive summaries are provided below:

- **Black Cluster NBHD (Northeast) has**

the highest % of

  - *Foreign Born population*
  - *Population below 18*
  - *Householders without spouse present*
  - *Family households*
  - *Other-than-English speakers*
  - *Houses built after 2000*

the lowest

  - *Median home value*
  - *% of population with Bachelor's degree*

- **Asian Cluster NBHD (Southwest) has**
  - *The highest % of 30+ minutes commuters*
  - *The lowest % of houses built after 2000.*

- **Other- race Cluster NBHD (Middle North)**
  - *Generally falls in the middle rank among all categories*

- **Hispanic Cluster NBHD (Middle West) has**

the highest % of

  - *Family Poverty*
  - *Renter-Occupied residence*

the lowest

  - *Median household income*
  - *Median rent*

- **White Cluster NBHD (Central) has**

the highest

  - *% of population above 65*
  - *% of English-only speakers*
  - *% of population with Bachelor's degree*
  - *Median household income*
  - *Median home value*
  - *Median rent*

the lowest

  - *% of foreign Born population*
  - *% of population below 18*
  - *% of householders without spouse present*
  - *% of family Poverty*

4. *Conclusions and Policy Recommendations*

In this project, strong evidence of ethnic clustering and clusters were found in Denver.

Each ethnic group is found to cluster in distinct and specific areas (see Fig 10 and descriptive summary above) of the city which have very different demographic and social-economic characteristics. White cluster is found to be in the most well-off neighborhoods.

As policy recommendations, the City of Denver should be aware of the distinct ethnic clusters present and be cautious of the social dynamic among clusters to prevent ethnic clustering growing into negative residential ethnic segregation. Specifically, as the ethnic clusters have very distinct ranges of housing and rent prices, governments should try to further investigate if financial reason is the main driving force for the formation of these ethnic clusters. If so, government can consider taking Singapore as an example and implement housing programs that can encourage racial integration.

Supports for young people, children, single-parents and education should be emphasized in the newly-developed neighborhoods in the Northeast such as Central Park, Montbello, Green Valley Ranch and DIA. As these neighborhoods are where the Black ethnic group cluster, support programs can be designed to tailor their social and cultural needs. Government should also consider enhancing translation services in hospitals, courts and voting stations of these neighborhoods, as well as the middle West part of the city such as West Colfax to Westwood areas because these are where many other-than-English speakers reside.

Good transportation network should be maintained for the Southwest area of the city such as Harvey Park downwards to assist long commuters. Further study can be performed to understand if there are specific areas Asian workers commute to in order to understand their employment situations. Also, employment and career training programs should be increased for the West part of the city to lower family poverty. And as that is where Hispanic community clusters, these programs can be designed to be more Spanish-friendly.

Finally, more elderly-care services can be considered in the central part of Denver to assist the relatively older white community there.

5. ***Lessons learnt and future study suggestion***

Racial diversity is a complex topic that needs to be studied from multiple angles. The use of spatial analysis helps us discover patterns and insights that could not be revealed by simple indices and percentages. As Denver develops rapidly in recent years with the influx of population into Colorado, the process of gentrification is also intensified. (Porter, 2020) To monitor and study the changes of geospatial patterns of ethnic demographic during the gentrification process, the cluster analyses performed in this project can be replicated over a certain study period. Also, the scale of the study area can be expanded to the entire Colorado to study the movement of ethnic clusters. Through doing so, we can further understand if ethnic groups relocate to different locations in Colorado as the outcome of gentrification in Denver, or if the ethnic groups are gradually moving out of Colorado. County and State governments should acknowledge these changes and work together to implement policies that promote long-term social equality and racial harmony.

As cluster analyses are very sensitive to the methodologies used and their specific window sizes, researchers should explore and compare how various methods and numeric settings affect the concluding results.

Lastly, in this project, due to relatively smaller population sizes, Native American, Hawaiian and Pacific Islanders, Other race and Multiple race groups are aggregated together as “Other race”, while these ethnic groups can be culturally and socially very different. For future study, researchers can consider studying them individually when data size and resources allow.

Works Cited

Denver GIS. (2021, 1 25). *American Community Survey Nbrhd (2015-2019)*. Retrieved from Denver Government Open Data Catalog: https://www.denvergov.org/opendata/dataset/american-community-survey-nbrhd-2015-2019

Housing and Development Board of Singapore. (n.d.). *Ethnic Integration Policy and SPR Quota*. Retrieved from Housing and Development Board: https://www.hdb.gov.sg/residential/buying-a-flat/resale/eligibility/ethnic-integration-policy-and-spr-quota

Klaesson, J., & Oner, O. (2020, January 20). Ethnic enclaves and segregation - self-employment and employment patterns among forced migrants. *Small Business Economics*, pp. 985-1006.

Koga, J. (2019, December 12). *Unpacking Ethnic Enclaves*. Retrieved from The Kenan Institute for Ethics at Duke University: https://kenan.ethics.duke.edu/unpacking-ethnic-enclaves/

Lim, S., Yi, S. S., De La Cruz, N. L., & Chau, T.-S. (2019). Defining ethnic enclave and its associations with self-reported health outcomes among Asian American adults in New York City. *Journal of Immigrant and minority health*, 138-146.

Liu, M., & Geron, K. (2008). Changing Neighborhood: Ethnic Enclaves and the Struggle for Social Justice. *Social Justice*, 18-35.

Menendian, S., Gambhir, S., & Gailes, A. (2021, June 30). *Most to Least Segregated Cities*. Retrieved from Othering & Belonging Institute: https://belonging.berkeley.edu/most-least-segregated-cities

Porter, J. (2020, July 8). *Study: Denver is now 2nd most gentrified city in the nation*. Retrieved from The Denver Channel: https://www.thedenverchannel.com/news/local-news/study-denver-is-now-2nd-most-gentrified-city-in-the-nation

United States Census Bureau. (2021, August 12). *Racial and Ethnic Diversity in the United States: 2010 Census and 2020 Census*. Retrieved from United States Census Bureau: https://www.census.gov/library/visualizations/interactive/racial-and-ethnic-diversity-in-the-united-states-2010-and-2020-census.html

Appendix: All codes for this project

```
knitr::opts_chunk$set(echo = FALSE)
## Load library
library(spdep)
library(sf)
library(smerc)
library(tigris)
library(RColorBrewer)
library(leaflet)
library(plotrix)
library(magicfor)
library(sp)

## Change Work directory
setwd("/Users/siuyinlee/OneDrive/School/Math 6384 Spatial Data/Project")

## read shapefile for the 2015-2019 ACS data
acs19 <- sf::st_read("./Data/american_community_survey_nbrhd_2015_2019/americ
an_community_survey_nbrhd_2015_2019.shp")
#data <- sf::st_read("/Users/siuyinlee/OneDrive/School/Math 6384 Spatial Data
/Project/Data/american_community_survey_nbrhd_2013_2017/american_community_su
rvey_nbrhd_2013_2017.shp")
#data <- sf::st_read("./Data/Census_BG_Data_2020/BG_Data_2020.shp")

## Plot the border of Denver to make sure the data is right
plot(sf::st_geometry(acs19), border="grey60")

## Create a Non_white column and its percentage
NONWHITE <- acs19$TTL_POPULA - acs19$WHITE
PCT_NONWHITE <- 100 - acs19$PCT_WHITE

## Create a Other column combining ethnic groups (Native Am, Hawaiian Pacific
 Islander, Other Race and Multiple race)
## they consists of small numbers of percentage
OTHER <- acs19$NATIVE_AME + acs19$HAWAIIAN_P + acs19$OTHER_RACE + acs19$TWO_O
R_MOR
PCT_OTHER <- acs19$PCT_NATIVE + acs19$PCT_HAWAII + acs19$PCT_OTHERR + acs19$P
CT_TWOORM

## create a dataframe of the created columns and bind them neighborhood
NB<- acs19$NBHD_NAME
df_NEW <- data.frame(NB, NONWHITE, PCT_NONWHITE, OTHER, PCT_OTHER)

## Merge the Non_white column
acs19<-geo_join(acs19, df_NEW, 'NBHD_NAME', 'NB', by = NULL, how = "left")

## Change column names
```

```r
colnames(acs19)[3] <- "HISPAN"

## Create a list of all the ethnic groups
ETH <- c("WHITE", "NONWHITE", "BLACK", "HISPAN", "ASIAN", "OTHER")

## Extract centroids of each neighborhood
coords <- st_coordinates(st_centroid(st_geometry(acs19)))


## Replace two 0 values with NA
acs19[acs19$MEDIAN_HOM == 0, ]$MEDIAN_HOM <- NA
acs19[acs19$MED_GROSS_ == 0, ]$MED_GROSS_ <- NA

##############################################
#Exploratory plots and tables
#############################################

## Plot the map of Denver
plot(st_geometry(acs19))


## In order for par mfrow to work, legends need to be given up
## Make Choropleth maps using counts
png(file="Count_plot.png",
    res = 300,
    width = 5, height = 4, units = 'in',
    pointsize = 8)
par(mfrow = c(3, 2))
for (i in ETH) {
  plot(acs19[i], pal = viridisLite::cividis, key.pos = NULL, reset = FALSE)
}
dev.off()
par(mfrow=c(1,1))

## Get the top region counts for each race
for (i in ETH) {
  print(head(acs19[order(acs19[[i]], decreasing = TRUE), c("NBHD_NAME", i)],
3))
}


## Make Choropleth maps using percentages
png(file="Pct_plot.png",
    res = 300,
    width = 5, height = 4, units = 'in',
    pointsize = 8)
par(mfrow = c(3, 2))
for (i in paste0("PCT_",ETH)) {
  plot(acs19[i], pal = viridisLite::cividis, key.pos = NULL, reset = FALSE)
```

```
}
dev.off()
par(mfrow=c(1,1))

## Get the top region percentages for each race
for (i in ETH) {
  print(head(acs19[order(acs19[[paste0("PCT_",i)]], decreasing = TRUE), c("NB
HD_NAME", paste0("PCT_",i))], 3))
}

## 2019 ETHNIC group pie charts
pie_col = brewer.pal(5, "Set2")

Prop_2019 <- c(sum(acs19$WHITE), sum(acs19$HISPAN), sum(acs19$BLACK), sum(acs
19$ASIAN), sum(acs19$OTHER))
pielabel_19 <- c("WHITE", "HISPAN", "BLACK", "ASIAN", "OTHER")
piepercent_19 <- round(100*Prop_2019/sum(Prop_2019),1)

png(file="Ethnic_Grp_2019.png",
    res = 300,
    width = 5, height = 4, units = 'in',
    pointsize = 8)
pie(Prop_2019, labels = paste0(Prop_2019, "=", piepercent_19, "%"),
    col = pie_col)
mytitle = "Ethnic Groups in Denver"
mysubtitle = paste0("Data: American Community Survey Nbrhd (2015-2019); Popul
ation Total: ", sum(Prop_2019))
mtext(side=3, line=1, at=-2, adj=0, cex=1.5, mytitle)
mtext(side=3, line=0, at=-2, adj=0, cex=0.8, mysubtitle)
#mtext(side = 3, line = 0.25, at = 1, adj = -2, mysubtitle)
legend("left",
       c("WHITE", "HISPAN", "BLACK", "ASIAN", "OTHER"),
       cex = 0.8, fill = pie_col)
dev.off()



##################################
#Now let's do some testing (Local Rates)
##################################

##################################
## 1. CEPP
##################################

##################################
## A. CEPP - NSTAR = 0.1 of population
##################################
```

```r
## Assign population
pop <-acs19[["TTL_POPULA"]]
alpha = 0.05
nstar = floor(0.1*sum(acs19[["TTL_POPULA"]]))
#nstar = floor(0.15*sum(acs19[["TTL_POPULA"]]))


## Set up the CEPP parameters using alpha = 0.05
## Using magicfor to store scan result
magic_for(silent = TRUE, progress = TRUE)
for (i in ETH) {
  case = acs19[[i]]
  CEPP01 = cepp.test(coords = coords,
                     cases = case,
                     pop = pop,
                     nstar = nstar,
                     alpha = alpha)

  ## Get the 3 most-likely clusters
  CEPP01_MLC_3 = head(summary(CEPP01),3)

  #List out the neighborhood_ID in the 5 most likely cluster
  CEPP01_MLC_1_NB_ID = CEPP01$clusters[[1]]$locids
  CEPP01_MLC_2_NB_ID = CEPP01$clusters[[2]]$locids
  CEPP01_MLC_3_NB_ID = CEPP01$clusters[[3]]$locids


  #List out the neighborhood in the 5 most likely cluster
  CEPP01_MLC_1_NB = acs19$NBHD_NAME[CEPP01_MLC_1_NB_ID]
  CEPP01_MLC_2_NB = acs19$NBHD_NAME[CEPP01_MLC_2_NB_ID]
  CEPP01_MLC_3_NB = acs19$NBHD_NAME[CEPP01_MLC_3_NB_ID]

  cat("\n================================================")
  cat("\nThe following are MLC NB for ", i, "\n, using alpha= ", alpha, "\n a
nd nstar=", nstar)
  cat("\nMLC_1_NB\n", CEPP01_MLC_1_NB)
  cat("\nMLC_2_NB\n", CEPP01_MLC_2_NB)
  cat("\nMLC_3_NB\n", CEPP01_MLC_3_NB)
  cat("\n================================================")

  # create vector of colors to show results
  mycol = brewer.pal(3, "Dark2")

  # default is white (no clustering)
  ctcol = rep("white", nrow(acs19))

  # Color the 3 most likely clusters
  ctcol[CEPP01_MLC_1_NB_ID] = mycol[1]
  #ctcol[CEPP01_MLC_2_NB_ID] = mycol[2]
```

```
  #ctcol[CEPP01_MLC_3_NB_ID] = mycol[3]

  #Make the plot
  png(file= paste0("CEPP01_MLC_plot_",i ,"_nstar_", nstar, ".png"),
      res = 300,
      width = 5, height = 4, units = 'in',
      pointsize = 8)
  plot(sf::st_geometry(acs19), border="grey60",
       col = ctcol,
       #main = paste0("CEPP01_3 most likely clusters_",i),
       main = paste0("CEPP01_Most likely clusters_",i),
       sub = paste0("CEPP Method: P-value = ", round(CEPP01$clusters[[1]]$pva
lue, 4), "; nstar = 10% of pop:", nstar))
  legend("bottomright",
         #legend = c("1st most-likely cluster", "2nd most-likely cluster", "3
rd most-likely cluster"),
         legend = "Most likely cluster",
         fill = mycol, border = "black" )
  dev.off()
  put(CEPP01, CEPP01_MLC_3, CEPP01_MLC_1_NB, CEPP01_MLC_2_NB, CEPP01_MLC_3_NB
)
}
CEPP01_test_results <- magic_result()
#CEPP015_test_results <- magic_result()




##################################
## B. CEPP - NSTAR = 0.15 of population
##################################

## Assign population
pop <-acs19[["TTL_POPULA"]]
alpha = 0.05
#nstar = floor(0.1*sum(acs19[["TTL_POPULA"]]))
nstar = floor(0.15*sum(acs19[["TTL_POPULA"]]))


## Set up the CEPP parameters using alpha = 0.05
## Using magicfor to store scan result
magic_for(silent = TRUE, progress = TRUE)
for (i in ETH) {
  case = acs19[[i]]
  CEPP015 = cepp.test(coords = coords,
                     cases = case,
                     pop = pop,
                     nstar = nstar,
                     alpha = alpha)
```

```
  ## Get the 3 most-likely clusters
  CEPP015_MLC_3 = head(summary(CEPP015),3)

  #List out the neighborhood_ID in the 5 most likely cluster
  CEPP015_MLC_1_NB_ID = CEPP015$clusters[[1]]$locids
  CEPP015_MLC_2_NB_ID = CEPP015$clusters[[2]]$locids
  #CEPP015_MLC_3_NB_ID = CEPP015$clusters[[3]]$locids


  #List out the neighborhood in the 5 most likely cluster
  CEPP015_MLC_1_NB = acs19$NBHD_NAME[CEPP015_MLC_1_NB_ID]
  CEPP015_MLC_2_NB = acs19$NBHD_NAME[CEPP015_MLC_2_NB_ID]
  #CEPP015_MLC_3_NB = acs19$NBHD_NAME[CEPP015_MLC_3_NB_ID]

  cat("\n===============================================")
  cat("\nThe following are MLC NB for ", i, "\n, using alpha= ", alpha, "\n a
nd nstar=", nstar)
  cat("\nMLC_1_NB\n", CEPP015_MLC_1_NB)
  cat("\nMLC_2_NB\n", CEPP015_MLC_2_NB)
  cat("\n===============================================")

  # create vector of colors to show results
  mycol = brewer.pal(3, "Dark2")

  # default is white (no clustering)
  ctcol = rep("white", nrow(acs19))

  # Color the 3 most likely clusters
  ctcol[CEPP015_MLC_1_NB_ID] = mycol[1]
  #ctcol[CEPP015_MLC_2_NB_ID] = mycol[2]
  #ctcol[CEPP015_MLC_3_NB_ID] = mycol[3]

  #Make the plot
  png(file= paste0("CEPP015_MLC_plot_",i ,"_nstar_", nstar, ".png"),
      res = 300,
      width = 5, height = 4, units = 'in',
      pointsize = 8)
  plot(sf::st_geometry(acs19), border="grey60",
       col = ctcol,
       main = paste0("CEPP015_Most likely clusters_",i),
       sub = paste0("CEPP Method: P-value = ", round(CEPP015$clusters[[1]]$pv
alue, 4), "; nstar = 15% of pop: ", nstar))
  legend("bottomright",
        legend = "Most likely cluster",
        fill = mycol, border = "black" )
  dev.off()
  put(CEPP015, CEPP015_MLC_3, CEPP015_MLC_1_NB ,CEPP015_MLC_2_NB
      #, CEPP015_MLC_3_NB
```

```
      )
}
#CEPP01_test_results <- magic_result()
CEPP015_test_results <- magic_result()



##################################
## 2. BN
##################################

##################################
## A. BN - CSTAR - 10% of CASES
##################################

## Assign population
pop <-acs19[["TTL_POPULA"]]
alpha = 0.05
cstar_prop = 0.1
#cstar_prop = 0.2


## Set up the BN parameters using alpha = 0.05
## Using magicfor to store scan result
magic_for(silent = TRUE, progress = TRUE)
for (i in ETH) {
  cstar = floor(cstar_prop*sum(acs19[[i]]))
  case = acs19[[i]]
  BN01 = bn.test(coords = coords,
                    cases = case,
                    pop = pop,
                    cstar = cstar,
                    alpha = alpha)

  ## Get the 3 most-likely clusters
  BN01_MLC_3 = head(summary(BN01),3)

  #List out the neighborhood_ID in the 3 most likely cluster
  BN01_MLC_1_NB_ID = BN01$clusters[[1]]$locids
  BN01_MLC_2_NB_ID = BN01$clusters[[2]]$locids
  BN01_MLC_3_NB_ID = BN01$clusters[[3]]$locids


  #List out the neighborhood in the 3 most likely cluster
  BN01_MLC_1_NB = acs19$NBHD_NAME[BN01_MLC_1_NB_ID]
  BN01_MLC_2_NB = acs19$NBHD_NAME[BN01_MLC_2_NB_ID]
  BN01_MLC_3_NB = acs19$NBHD_NAME[BN01_MLC_3_NB_ID]

  cat("\n===============================================")
```

```r
  cat("\nThe following are MLC NB for ", i, "\n, using alpha= ", alpha, "\n a
nd cstar=", cstar)
  cat("\nMLC_1_NB\n", BN01_MLC_1_NB)
  cat("\nMLC_2_NB\n", BN01_MLC_2_NB)
  cat("\nMLC_3_NB\n", BN01_MLC_3_NB)
  cat("\n===============================================")

  # create vector of colors to show results
  mycol = brewer.pal(3, "Dark2")

  # default is white (no clustering)
  ctcol = rep("white", nrow(acs19))

  # Color the 3 most likely clusters
  ctcol[BN01_MLC_1_NB_ID] = mycol[1]
  #ctcol[BN01_MLC_2_NB_ID] = mycol[2]
  #ctcol[BN01_MLC_3_NB_ID] = mycol[3]

  #Make the plot
  png(file= paste0("BN01_MLC_plot_",i ,"_cstar_", cstar, ".png"),
      res = 300,
      width = 5, height = 4, units = 'in',
      pointsize = 8)
  plot(sf::st_geometry(acs19), border="grey60",
       col = ctcol,
       main = paste0("BN01_Most likely clusters_",i),
       sub = paste0("BN Method: P-value = ", round(BN01$clusters[[1]]$pvalue,
 4), "; cstar = 10% of cases: ", cstar))
  legend("bottomright",
         #legend = c("1st most-likely cluster", "2nd most-likely cluster", "3
rd most-likely cluster"),
         legend = "Most likely cluster",
         fill = mycol, border = "black" )
  dev.off()
  put(BN01, BN01_MLC_3, BN01_MLC_1_NB, BN01_MLC_2_NB, BN01_MLC_3_NB)
}
BN01_test_results <- magic_result()




##################################
## B. BN - CSTAR - 20% of CASES
##################################

## Assign population
pop <-acs19[["TTL_POPULA"]]
alpha = 0.05
```

```
#cstar_prop = 0.1
cstar_prop = 0.2


## Set up the BN parameters using alpha = 0.05
## Using magicfor to store scan result
magic_for(silent = TRUE, progress = TRUE)
for (i in ETH) {
  cstar = floor(cstar_prop*sum(acs19[[i]]))
  case = acs19[[i]]
  BN02 = bn.test(coords = coords,
                 cases = case,
                 pop = pop,
                 cstar = cstar,
                 alpha = alpha)

  ## Get the 3 most-likely clusters
  BN02_MLC_3 = head(summary(BN02),3)

  #List out the neighborhood_ID in the 3 most likely cluster
  BN02_MLC_1_NB_ID = BN02$clusters[[1]]$locids
  BN02_MLC_2_NB_ID = BN02$clusters[[2]]$locids
  #BN02_MLC_3_NB_ID = BN02$clusters[[3]]$locids


  #List out the neighborhood in the 3 most likely cluster
  BN02_MLC_1_NB = acs19$NBHD_NAME[BN02_MLC_1_NB_ID]
  BN02_MLC_2_NB = acs19$NBHD_NAME[BN02_MLC_2_NB_ID]
  #BN02_MLC_3_NB = acs19$NBHD_NAME[BN02_MLC_3_NB_ID]

  cat("\n================================================")
  cat("\nThe following are MLC NB for ", i, "\n, using alpha= ", alpha, "\n a
nd cstar=", cstar)
  cat("\nMLC_1_NB\n", BN02_MLC_1_NB)
  cat("\nMLC_2_NB\n", BN02_MLC_2_NB)
  #cat("\nMLC_3_NB\n", BN02_MLC_3_NB)
  cat("\n================================================")

  # create vector of colors to show results
  mycol = brewer.pal(3, "Dark2")

  # default is white (no clustering)
  ctcol = rep("white", nrow(acs19))

  # Color the 3 most likely clusters
  ctcol[BN02_MLC_1_NB_ID] = mycol[1]
  #ctcol[BN02_MLC_2_NB_ID] = mycol[2]
  #ctcol[BN02_MLC_3_NB_ID] = mycol[3]
```

```
  #Make the plot
  png(file= paste0("BN02_MLC_plot_",i ,"_cstar_", cstar, ".png"),
      res = 300,
      width = 5, height = 4, units = 'in',
      pointsize = 8)
  plot(sf::st_geometry(acs19), border="grey60",
       col = ctcol,
       main = paste0("BN02_Most likely clusters_",i),
       sub = paste0("BN Method: P-value = ", round(BN02$clusters[[1]]$pvalue,
 4), "; cstar = 20% of cases: ", cstar))
  legend("bottomright",
         #legend = c("1st most-likely cluster", "2nd most-likely cluster", "3
rd most-likely cluster"),
         legend = "Most likely cluster",
         fill = mycol, border = "black" )
  dev.off()
  put(BN02, BN02_MLC_3, BN02_MLC_1_NB, BN02_MLC_2_NB
      #, BN02_MLC_3_NB
      )
}
BN02_test_results <- magic_result()




####################################
## C. BN - CSTAR - 15% of CASES
####################################

## Assign population
pop <-acs19[["TTL_POPULA"]]
alpha = 0.05
#cstar_prop = 0.1
#cstar_prop = 0.2
cstar_prop = 0.15

## Set up the BN parameters using alpha = 0.05
## Using magicfor to store scan result
magic_for(silent = TRUE, progress = TRUE)
for (i in ETH) {
  cstar = floor(cstar_prop*sum(acs19[[i]]))
  case = acs19[[i]]
  BN015 = bn.test(coords = coords,
                  cases = case,
                  pop = pop,
                  cstar = cstar,
                  alpha = alpha)

  ## Get the 3 most-likely clusters
```

```
  BN015_MLC_3 = head(summary(BN015),3)

  #List out the neighborhood_ID in the 3 most likely cluster
  BN015_MLC_1_NB_ID = BN015$clusters[[1]]$locids
  BN015_MLC_2_NB_ID = BN015$clusters[[2]]$locids
  #BN015_MLC_3_NB_ID = BN015$clusters[[3]]$locids


  #List out the neighborhood in the 3 most likely cluster
  BN015_MLC_1_NB = acs19$NBHD_NAME[BN015_MLC_1_NB_ID]
  BN015_MLC_2_NB = acs19$NBHD_NAME[BN015_MLC_2_NB_ID]
  #BN015_MLC_3_NB = acs19$NBHD_NAME[BN015_MLC_3_NB_ID]

  cat("\n===============================================")
  cat("\nThe following are MLC NB for ", i, "\n, using alpha= ", alpha, "\n a
nd cstar=", cstar)
  cat("\nMLC_1_NB\n", BN015_MLC_1_NB)
  cat("\nMLC_2_NB\n", BN015_MLC_2_NB)
  #cat("\nMLC_3_NB\n", BN015_MLC_3_NB)
  cat("\n===============================================")

  # create vector of colors to show results
  mycol = brewer.pal(3, "Dark2")

  # default is white (no clustering)
  ctcol = rep("white", nrow(acs19))

  # Color the 3 most likely clusters
  ctcol[BN015_MLC_1_NB_ID] = mycol[1]
  #ctcol[BN015_MLC_2_NB_ID] = mycol[2]
  #ctcol[BN015_MLC_3_NB_ID] = mycol[3]

  #Make the plot
  png(file= paste0("BN015_MLC_plot_",i ,"_cstar_", cstar, ".png"),
      res = 300,
      width = 5, height = 4, units = 'in',
      pointsize = 8)
  plot(sf::st_geometry(acs19), border="grey60",
       col = ctcol,
       main = paste0("BN015_Most likely clusters_",i),
       sub = paste0("BN Method: P-value = ", round(BN015$clusters[[1]]$pvalue
, 4), "; cstar = 20% of cases: ", cstar))
  legend("bottomright",
         #legend = c("1st most-likely cluster", "2nd most-likely cluster", "3
rd most-likely cluster"),
         legend = "Most likely cluster",
         fill = mycol, border = "black" )
  dev.off()
  put(BN015, BN015_MLC_3, BN015_MLC_1_NB, BN015_MLC_2_NB
```

```
      #, BN015_MLC_3_NB
  )
}
BN015_test_results <- magic_result()








################################################
## 3. Spatial Scan method using CRH for each race
################################################

################################################
## A. Spatial Scan ubpop = 0.1
################################################

## Assign population
pop <-acs19[["TTL_POPULA"]]
alpha = 0.05
ubpop = 0.1
#ubpop = 0.15

#Change plot margin
par(mar=c(1,1,1,1))

## Set up the scan parameters using alpha = 0.05
## Using magicfor to store scan result
magic_for(silent = TRUE, progress = TRUE)
for (i in ETH) {
  case = acs19[[i]]
  scan01 = scan.test(coords = coords,
                     cases = case,
                     pop = pop,
                     ex = sum(case)/sum(pop)*pop,
                     alpha = alpha,
                     ubpop = ubpop)

  ## Get the 3 most-likely clusters
  scan01_MLC_3 = head(summary(scan01),3)

  #List out the neighborhood_ID in the 3 most likely cluster
  scan01_MLC_1_NB_ID = scan01$clusters[[1]]$locids
  scan01_MLC_2_NB_ID = scan01$clusters[[2]]$locids
  scan01_MLC_3_NB_ID = scan01$clusters[[3]]$locids
```

```
  #List out the neighborhood in the 3 most likely cluster
  scan01_MLC_1_NB = acs19$NBHD_NAME[scan01_MLC_1_NB_ID]
  scan01_MLC_2_NB = acs19$NBHD_NAME[scan01_MLC_2_NB_ID]
  scan01_MLC_3_NB = acs19$NBHD_NAME[scan01_MLC_3_NB_ID]


  cat("\n===============================================")
  cat("\nThe following are MLC NB for ", i, "\n, using alpha= ", alpha, "\n a
nd ubpop=", ubpop)
  cat("\nMLC_1_NB\n", scan01_MLC_1_NB)
  cat("\nMLC_2_NB\n", scan01_MLC_2_NB)
  cat("\nMLC_3_NB\n", scan01_MLC_3_NB)
  cat("\n===============================================")

  # create vector of colors to show results
  mycol = brewer.pal(3, "Dark2")

  # default is white (no clustering)
  ctcol = rep("white", nrow(acs19))

  # Color the 3 most likely clusters
  ctcol[scan01_MLC_1_NB_ID] = mycol[1]
  #ctcol[scan01_MLC_2_NB_ID] = mycol[2]
  #ctcol[scan01_MLC_3_NB_ID] = mycol[3]

  #Make the plot
  png(file= paste0("Scan01_MLC_plot_",i ,"_up_", ubpop, ".png"),
      res = 300,
      width = 5, height = 4, units = 'in',
      pointsize = 8)
  plot(sf::st_geometry(acs19), border="grey60",
       col = ctcol,
       main = paste0("Scan01_Most likely clusters_",i),
       sub = paste0("Spatial Scan Method: P-value = ", round(scan01$clusters[
[1]]$pvalue, 4), "; ubpop = ", ubpop))
  legend("bottomright",
         legend = "Most likely cluster",
         fill = mycol, border = "black" )
  dev.off()
  put(scan01, scan01_MLC_3, scan01_MLC_1_NB, scan01_MLC_2_NB, scan01_MLC_3_NB
)
}
scan01_test_results <- magic_result()



################################################
## B. Spatial Scan ubpop = 0.15
```

```r
################################################

## Assign population
pop <-acs19[["TTL_POPULA"]]
alpha = 0.05
#ubpop = 0.1
ubpop = 0.15

#Change plot margin
par(mar=c(1,1,1,1))

## Set up the scan parameters using alpha = 0.05
## Using magicfor to store scan result
magic_for(silent = TRUE, progress = TRUE)
for (i in ETH) {
  case = acs19[[i]]
  scan015 = scan.test(coords = coords,
                      cases = case,
                      pop = pop,
                      ex = sum(case)/sum(pop)*pop,
                      alpha = alpha,
                      ubpop = ubpop)

  ## Get the 3 most-likely clusters
  scan015_MLC_3 = head(summary(scan015),3)

  #List out the neighborhood_ID in the 3 most likely cluster
  scan015_MLC_1_NB_ID = scan015$clusters[[1]]$locids
  scan015_MLC_2_NB_ID = scan015$clusters[[2]]$locids
  scan015_MLC_3_NB_ID = scan015$clusters[[3]]$locids


  #List out the neighborhood in the 3 most likely cluster
  scan015_MLC_1_NB = acs19$NBHD_NAME[scan015_MLC_1_NB_ID]
  scan015_MLC_2_NB = acs19$NBHD_NAME[scan015_MLC_2_NB_ID]
  scan015_MLC_3_NB = acs19$NBHD_NAME[scan015_MLC_3_NB_ID]


  cat("\n===============================================")
  cat("\nThe following are MLC NB for ", i, "\n, using alpha= ", alpha, "\n a
nd ubpop=", ubpop)
  cat("\nMLC_1_NB\n", scan015_MLC_1_NB)
  cat("\nMLC_2_NB\n", scan015_MLC_2_NB)
  cat("\nMLC_3_NB\n", scan015_MLC_3_NB)
  cat("\n===============================================")

  # create vector of colors to show results
  mycol = brewer.pal(3, "Dark2")
```

```
  # default is white (no clustering)
  ctcol = rep("white", nrow(acs19))

  # Color the 3 most likely clusters
  ctcol[scan015_MLC_1_NB_ID] = mycol[1]
  #ctcol[scan01_MLC_2_NB_ID] = mycol[2]
  #ctcol[scan01_MLC_3_NB_ID] = mycol[3]

  #Make the plot
  png(file= paste0("Scan015_MLC_plot_",i ,"_up_", ubpop, ".png"),
      res = 300,
      width = 5, height = 4, units = 'in',
      pointsize = 8)
  plot(sf::st_geometry(acs19), border="grey60",
       col = ctcol,
       main = paste0("Scan015_Most likely clusters_",i),
       sub = paste0("Spatial Scan Method: P-value = ", round(scan015$clusters
[[1]]$pvalue, 4), "; ubpop = ", ubpop))
  legend("bottomright",
         legend = "Most likely cluster",
         fill = mycol, border = "black" )
  dev.off()
  put(scan015, scan015_MLC_3, scan015_MLC_1_NB, scan015_MLC_2_NB, scan015_MLC
_3_NB)
}
scan015_test_results <- magic_result()





#####################################################
# Now let's do some Spatial Autocorrelation testing
#####################################################

##################################################################
# Need to determine neighbors relationship for Moran's I
##################################################################

## Extract centroids of each neighborhood
nb_centroids <- st_centroid(st_geometry(acs19))

#Border-based nb
(nb_acs19 <-spdep::poly2nb(acs19))
plot(st_geometry(acs19))
plot(nb_acs19, st_centroid(st_geometry(acs19)), add = TRUE)
title("Border-Based Neighbor Relationships of Denver NBHD")

#Distance-based nb using 10 km as upper dist threshold
(nb_acs19_dnn <-spdep::dnearneigh(nb_centroids, d1 = 0, d2 = 10))
```

```
plot(st_geometry(acs19))
plot(nb_acs19_dnn, st_centroid(st_geometry(acs19)), add = TRUE)
title("Neighbor (within 10 km) relationships of Denver")

# KNN nb using K=5
nb_acs19_knn <- spdep::knn2nb(spdep::knearneigh(nb_centroids, k=5))
plot(st_geometry(acs19))
plot(nb_acs19_knn, st_centroid(st_geometry(acs19)), add = TRUE)
title("Neighbor (5 nearest) relationships of Denver")


###################################################################
# Decide which neighbor relationship to use (border, dnn or knn)
###################################################################

Neigh_RS = nb_acs19 #border
#Neigh_RS = nb_acs19_dnn #distance-based
#Neigh_RS = nb_acs19_knn #5 nearest neighbor

#Set proximity matrix

# assume adjacency weights (w_ij = 1 if regions i and j share a boundary)
# proximity matrix, binary style.  W is row standardized.
w = nb2mat(Neigh_RS , style = "B")
# see ?nb2listw for more options
# proximaty matrix in list format
lw = nb2listw(Neigh_RS , style = "B")


#######################################
## 4. Moran's Icr Monte Carlos under CRH
#######################################
set.seed(108)
# some preliminaries
N = length(acs19$NBHD_NAME) # number of regions
pop = acs19[["TTL_POPULA"]] #population sizes
nsim = 9999

### Use CR Moran's I for inference
# make a function out of this process
i_cr = function(y, rni, w) {
  y_std = matrix((y - rni)/sqrt(rni))
  return(sum(w * y_std %*% t(y_std))/sum(w))
}

## Set up the Moran's Icr parameters
## Using magicfor to store test result
magic_for(silent = TRUE, progress = TRUE)
for (i in ETH) {
```

```
  case = acs19[[i]] # number of cases
  r = sum(case)/sum(pop) # estimated risk (global)
  e = r * pop # expected per region
  print(paste0("For ", i))
  print(paste0("Obs count: ", case, ". Exp count: ", ceiling(e)))

  tsimc = numeric(nsim)
  print(paste0("Observed Moran's Icr statistic for ", i))
  t0c = i_cr(case, e, w) # observed statistic
  print(t0c)

  # statistics for data simulated under CRH
  for (j in 1:nsim) tsimc[j] = i_cr(rpois(N, e), rni = e, w = w)

  # p-value
  print(paste0("P-value for Moran's Icr statistic for ", i))
  p_value = (sum(tsimc >= t0c) + 1)/(nsim + 1)
  print(p_value)

  put(r, e, t0c, p_value)

}

Moran_Icr_results <- magic_result()


##################################
## 5. Tango
##################################

coords <- st_coordinates(st_centroid(st_geometry(acs19)))
d <- as.matrix(dist(coords))
max(d)
mean(d)

# # Code to plot weights and "effective range"
plotdist <- seq(0, max(d), 0.01)
kappa <- 0.005
plot(plotdist,
     exp(-plotdist/kappa),type="l",xlab="Distance",ylab="exp(-distance/kappa)
",
     cex.lab=1.5,cex.axis=1.25,ylim=c(0,0.1))
#rug(dist)
title(paste("kappa = ",kappa),cex.main=2.0)
effrange <- -kappa*log(0.005)
segments(0,0.005,effrange,0.005)
segments(effrange,0,effrange,0.005)
```

```r
## Assign population
pop <-acs19[["TTL_POPULA"]]
# Set different weights using different kappa
# May need to double check the intercentroid distances' range
#k_value = 0.005
#k_value = 0.01
k_value = 0.05
#k_value = 0.1
#k_value = 0.2
#k_value = 0.4
#k_value = 1
#k_value = 10
#k_value = 20
#k_value = 50
wstar <- dweights(coords, pop = pop, kappa = k_value, type = "rogerson")
#wstar <- dweights(coords, kappa = k_value)

set.seed(108)
## Set up the Tango.test parameters
## Using magicfor to store test result
magic_for(silent = TRUE, progress = TRUE)
for (i in ETH) {
  case = acs19[[i]]
  print(paste0("Tango Test Result for ", i))
  tango = tango.test(cases = case,
                         pop = pop,
                         w = wstar,
                         nsim = 9999)
  print(tango)

  #Make the plot of simulated result to check Goodness-of-fit and Spatial Ass
ociation
  png(file= paste0("Tango_plot_",i ,"_kappa_", k_value, ".png"),
      res = 300,
      width = 5, height = 4, units = 'in',
      pointsize = 8)
  plot(tango,
       main = paste0("Tango_Plot_",i),
       #sub = paste0("Tango's Index - using kappa = ", test))
       sub = paste0("Tango's Index - using kappa = ", k_value))
  legend("bottomright",
         legend = c("Observed Data", "Simulations"),
         pch = c(20, 2),
         inset=c(0,1), xpd=TRUE, horiz=TRUE, bty="n")
  dev.off()
  put(tango)
}
```

```
tango_test_results <- magic_result()



#################################################################
# Summarize Ethnic Cluster NBHD Characteristics
#################################################################
magic_free()
## Nested loop make tables
table_name <- c("CEPP01","CEPP015","BN01","BN02","scan01","scan015")
outside_list<- list()
for (t in table_name) {
  temp_list<- list()
  for (i in 1:length(ETH)) {

    ## Get the MLC regions ID for each ethnic group
    temp_name<- paste0(t,"_test_results")
    MLC_ID<-eval(parse(text=temp_name))[[t]][[i]][["clusters"]][[1]][["locids
"]]
    print(MLC_ID)

    ## Subset acs dataframe using regions ID
    subset_df <- acs19[MLC_ID, ]

    ## Create a vector of interested variables
    cal_var <- c("TTL_POPULA",
                 ETH[i],
                 paste0("PCT_",ETH[i]),
                 "FOREIGN_BO",
                 "MEDIAN_AGE",
                 "AGELESS18",
                 "AGE65PLUS",
                 "TOTAL_COMM",
                 "COMMUTE_30",
                 "COMMUTE_45",
                 "COMMUTE_60",
                 "MED_HH_INC",
                 "MED_GROSS_",
                 "MEDIAN_HOM",
                 "PCT_FAM_PO",
                 "BACHELORS_",
                 "TTLPOP_5PL",
                 "ONLY_ENGLI",
                 "TTL_HOUSIN",
                 "BUILT_2014",
                 "BUILT_2010",
                 "BUILT_2000",
                 "RENTER_OCC",
                 "OWNER_OCCU",
```

```r
                 "TTL_HOUSEH",
                 "FAMILY_HOU",
                 "MALE_HHLDR",
                 "FEMALE_HHL"
    )
    subset_df<-subset_df[ ,cal_var, drop=TRUE]

    ## Calculate cluster average for desired variables
    Avg_df<- data.frame(t(sapply(subset_df, mean)))
    Avg_df["Regions_Count"] <- length(MLC_ID)

    ## Calculate cluster sum for desired variables
    Sum_df<- data.frame(t(sapply(subset_df, sum)))
    Sum_df["Regions_Count"] <- length(MLC_ID)

    temp_list[[i]] <- list(subset_df = subset_df, Avg_df = Avg_df, Sum_df = S
um_df)
  }
  names(temp_list) <- ETH
  outside_list[[which(table_name == t)]] <- temp_list
}

names(outside_list) <- paste0(table_name, "_NB_summary")

## Shoot the NB_summary_list to the global environment
list2env(outside_list, envir = .GlobalEnv)


#################################
## Make final summary ethnic df
################################

magic_for(silent = TRUE, progress = TRUE)
for (i in ETH) {
  CEPP01<-CEPP01_NB_summary[[i]][["Sum_df"]]
  CEPP015<-CEPP015_NB_summary[[i]][["Sum_df"]]
  BN01<-BN01_NB_summary[[i]][["Sum_df"]]
  BN02<-BN02_NB_summary[[i]][["Sum_df"]]
  scan01<-scan01_NB_summary[[i]][["Sum_df"]]
  scan015<-scan015_NB_summary[[i]][["Sum_df"]]

  temp_df<-rbind(CEPP01, CEPP015, BN01, BN02, scan01, scan015)
  put(temp_df)
}

final_df_all <- magic_result()

White_df<- final_df_all[["temp_df"]][[1]]
Nonwhite_df<- final_df_all[["temp_df"]][[2]]
```

```
Black_df<- final_df_all[["temp_df"]][[3]]
Hisp_df<- final_df_all[["temp_df"]][[4]]
Asian_df<- final_df_all[["temp_df"]][[5]]
Other_df<- final_df_all[["temp_df"]][[6]]

avg_white<- round(colMeans(White_df[, -c(2:3)], na.rm = TRUE))
avg_nonwhite<- round(colMeans(Nonwhite_df[, -c(2:3)], na.rm = TRUE))
avg_black <- round(colMeans(Black_df[, -c(2:3)], na.rm = TRUE))
avg_hisp<- round(colMeans(Hisp_df[, -c(2:3)], na.rm = TRUE))
avg_asian<- round(colMeans(Asian_df[, -c(2:3)], na.rm = TRUE))
avg_other<- round(colMeans(Other_df[, -c(2:3)], na.rm = TRUE))

End_Result <- rbind(avg_white, avg_hisp, avg_black, avg_asian, avg_other)
End_Result <- as.data.frame(End_Result)

End_Result[["C_PCT_FB"]] <- round(End_Result$FOREIGN_BO/End_Result$TTL_POPULA
, 2)
End_Result[["C_PCT_AGEL18"]] <- round(End_Result$AGELESS18/End_Result$TTL_POP
ULA, 2)
End_Result[["C_PCT_AGE65P"]] <- round(End_Result$AGE65PLUS/End_Result$TTL_POP
ULA, 2)
End_Result[["C_PCT_COMM"]] <- round(End_Result$TOTAL_COMM/End_Result$TTL_POPU
LA, 2)
End_Result[["C_PCT_COMM_L"]] <- round((End_Result$COMMUTE_30+End_Result$COMMU
TE_45+End_Result$COMMUTE_60)/End_Result$TTL_POPULA, 2)
End_Result[["C_PCT_BACH"]] <- round(End_Result$BACHELORS_/End_Result$TTL_POPU
LA, 2)
End_Result[["C_PCT_FL"]] <- round(End_Result$TTLPOP_5PL/End_Result$TTL_POPULA
, 2)
End_Result[["C_PCT_ENG_ONLY"]] <- round(End_Result$ONLY_ENGLI/End_Result$TTL_
POPULA, 2)
End_Result[["C_PCT_NEWH"]] <- round((End_Result$BUILT_2000+End_Result$BUILT_2
010+End_Result$BUILT_2014)/End_Result$TTL_HOUSIN, 2)
End_Result[["C_PCT_MHH"]] <- round(End_Result$MALE_HHLDR/End_Result$TTL_HOUSE
H, 2)
End_Result[["C_PCT_FHH"]] <- round(End_Result$FEMALE_HHL/End_Result$TTL_HOUSE
H, 2)
End_Result[["C_PCT_RenterOcc"]] <- round(End_Result$RENTER_OCC/End_Result$TTL
_HOUSEH, 2)
End_Result[["C_PCT_FamHH"]] <- round(End_Result$FAMILY_HOU/End_Result$TTL_HOU
SEH, 2)
End_Result[["C_PCT_FamPov"]] <- round(End_Result$PCT_FAM_PO/End_Result$Region
s_Count, 2)
End_Result[["C_Med_HH_INC"]] <- round(End_Result$MED_HH_INC/End_Result$Region
s_Count, 2)
End_Result[["C_Med_Home_Val"]] <- round(End_Result$MEDIAN_HOM/End_Result$Regi
ons_Count, 2)
End_Result[["C_Med_Rent"]] <- round(End_Result$MED_GROSS_/End_Result$Regions_
Count, 2)
```