GEOGRAPHY 5303
Exercise 2 – 100 points
Due: February 15, 2020, by 5 PM
Numerical Summarization of Spatial Data

Purpose of Exercise

This exercise further explores the datasets introduced in Exercise 1. Moving on from mostly graphical depictions (e.g., boxplots, histograms, and line graphs) this exercise will primarily focus on computing summary statistics for point and areal data. The combination of both graphical and numerical methods is a powerful tool for fully understanding, properly analyzing, and effectively summarizing a variety of data sources, so Exercises 1 and 2 have been designed to develop these skills.

Both datasets from Exercise 1 are again used here (except questions A.3 and A.4; see info in that section). The same general information regarding the dataset and guidelines for the preparation, presentation, and submission of exercises described Syllabus Appendix 1 are still in effect. You will want to use SPSS where you can (and is indicated) and Excel for the rest.

A. Cross-sectional Social and Demographic Analysis of Oklahoma

This exercise begins with an exploration of various measures of central tendency and dispersion so we can uncover some geographic traits of a state that is great except for its weight.

Summary Statistics for Oklahoma's Territory: Basic summary statistics are computed to determine the most "central" location of the state and its dispersion.

1. Compute the centroid and standard distance for Oklahoma based on tracts weighted by:

   - tract land areas (areas are given in square miles);
   - tract populations; and
   - unweighted (each tract weight = 1).

   For simplicity, regard the decimal degree coordinates for each tract as X (longitude) and Y (latitude) coordinates of points on a flat plane.

First, be sure to analyze only tracts by going to "Data", "Select Cases" and selecting **if Scale = "T"** (quotes are necessary for SPSS to recognize text as a "value"). You can do both weighted and unweighted centroids in SPSS simply by computing the means of latitude and longitude, using the proper weighting variables where appropriate ("Data", "Weight Cases", "Weight cases by"). Be aware that once you weight cases, SPSS maintains that specific weighting until you go back and either change the weight variable or revert to "Do not weight cases".

You can also derive standard distances by having SPSS compute unweighted or weighted variances for latitude and longitude, adding the latitude and longitude variances, and then taking the square root of the sum. Report all computed final values in Word in a summary table in addition to the required map and narrative.

a. Map the locations of the three computed centroids as well as the Census centroid, compare these four locations for similarity, and discuss the methodological and accuracy issues involved with each (Note: the Census Bureau's "official" centroid for Oklahoma is: 35.572285°N, 97.043688°W).

b. Compare the computed dispersion statistics and discuss possible causes for variations observed in two-dimensional dispersion between unweighted and weighted centroids.

Which "Center" is Best and Why? Having computed several centroids for the entire state in order to compare various weighting schemes above, we switch to comparing/contrasting different measures of central tendency. In order to make the work a little more manageable, however, the focus is on just a portion of the state since one of the measures below becomes more challenging with more observations.

The state would like to find good locations for 5 rural regional health risk assessment centers to serve the population of each region and will use measures of central tendency to provide first approximations. You will examine one of the 15-county regions introduced in Exercise 1 (see region assignment Excel sheet). Ultimately, population distributions at a finer scale than the county combined with network distances should be used for the final site selection, but oh well.

2. Compute the following geographic centers for your assigned 15-county region, using the county-level data provided for your assigned region:

   - Unweighted mean center,
   - Weighted (by population) mean center, and
   - Weighted (by population) Euclidean median. Follow the procedure outlined in the textbook (section 2.6.3) and iterate 5 times. Use the population weighted mean center from above as your algorithm starting point (Iteration 0).

   a. Map the locations of these three centers, and compare their locations by discussing the various influences that resulted in these three locations. Overall, are the three locations relatively similar within the study area?

   b. Which location would you choose for the facility, and why?

Measures of Concentration: Areal patterns are studied in a different way by focusing on just two variables and using a family of techniques that build upon the shares that each sub-unit (here, a county) makes up of the larger region (here, the state) total of a given activity.

3. Location quotients (LQs) allow us to compare all observations in a dataset. You will compute LQs for two variables not previously analyzed, voter registrations (as of January 15, 2019) and voting in the 2018 Oklahoma gubernatorial election (November 6, 2019), by county, for the entire state (n=77). One-third of the class will each compute LQs for Republican, Democratic, and Independent voter registrations, and the class will be split between computing LQs for votes for Kevin Stitt (R) and Drew Edmondson (D). See Student Variable Assignment Excel file.

   To perform this analysis, you will need to access raw voter registration totals and vote totals for each county from the County-level Excel file (OSEB_18_Counties.xlsx) provided on Canvas. Use total registered voters per county (not total votes cast) as the normalizing variable for both sets of LQ computations.

   a. Compute the LQ computations for both variables in a submitted Excel file, showing your work, with the counties listed in their original alphabetical order.

   b. Submit a choroplethic map of your LQs for each variable. Be sure to explain *what* classification method you chose and why, *how many* categories you chose to portray and why, and otherwise adhere to sound cartographic practice.

   c. Describe and discuss the patterns/trends you see on the individual LQ maps. Do the patterns match your expectations? Are there any big surprises? Where and why?

   d. Comment on any spatial associations you might observe between the two maps.

B. Home Sales in Milwaukee, 2012

We will explore spatio-temporal patterns in the Milwaukee home sales data by looking at sales by month. This will begin an arc of several explorations across the next few exercises using relatively equal time chunks (months) to determine if various traits of the home sold fluctuate during the year.

Each student will study three consecutive months of the home sales data (see Student Variable Assignment file). As with the selection of Tracts in part A, if using SPSS you will have to select data with **if SaleDate="2012-##"** where ## is a two-digit month number from 01 (January) to 12 (December).

1. Compute the mean center and standard distance of homes sold for each of your three months; that is, you will compute three separate means and standard distances, not pooled for all three of your months. Report these values in Word and plot the centroids on a separate map/scatterplot for each month (that is, plot all the homes sold each month along with the centroid), labeling the figures appropriately. Also, compute the mean and standard deviation of SalePrice for each month. Summarize all computed statistics requested below in a well-organized and labelled table in Word – do not just say "See Excel."

2. Discuss the general movement of your three centroids and whether the movement appears random or might reflect some underlying trend. Also, discuss whether sales price seems to vary significantly between the three months.

3. Discuss the varying sizes of your three standard distances and any rationale for the changes. Also, discuss whether the variability of sales price seems to vary significantly between the three months.

What you should upload to Canvas:

- Word document answering all numbered questions above (don't forget to label lettered subparts and to answer all subparts, labeled or otherwise) with SPSS charts and several maps inserted directly where discussed.
- One Excel file for Questions 2-3 (Part A), and a second Excel file if you did Part B in Excel (probably easier, given the chart and graphing requirements). If Part B was done in SPSS, then your results will still be reported in a table in Word (see last sentence of question B.1 above) and your graphics will also still be inserted in Word.