

## **IS415 Geospatial Analytics and Applications**

### **Take-home Exercise 3 (15%)**

### **Delineating Social Area Analysis using Geographic Segmentation Approach**

#### **Setting the Scene**

Social area analysis (SAA), which is a generic name for a number of methods employing census and other data to classify small areas into similar socioeconomic groups, is an approach which quantifies data in a useful fashion and has important applications in regional development, urban planning, medical and health services research. The potentialities of the approach for Liveable City planning and Smart City preparedness, however, have yet to be explored.

#### **The Task**

In this take-home exercise, you are tasked to segment Singapore at the planning subzone level into homogeneous socioeconomic areas by combining geodemographic data extracted from Singapore Department of Statistics and urban functions extracted from the geospatial data provided.

From the Singapore Residents by Planning Area Subzone, Age Group, Sex and Type of Dwelling, June 2011-2019 provide by Singapore Department of Statistics, you are required to extract the following indicators for 2019:

- Economy active population (i.e. age 25-64)
- Young group (i.e. age 0-24)
- Aged group (i.e. 65 and above)
- Population density
- HDB1-2RM dwellers
- HDB3-4RM dwellers
- HDB5RM-Ec dweller
- Condominium and apartment dwellers
- Landed property dwellers

From the geospatial data provided, you are required to extract the following urban functions:

- Government including embassy
- Business

- Industry
- Shopping
- Financial
- Upmarket residential area

## The Data

For the purpose of this study, five geospatial data sets are provided, they are: Govt\_Embassy, Private residential, Shopping, Business and Financial. These data sets are in ESRI shapefile format.

You are also required to download the following data from the relevant government portals:

- Singapore Residents by Planning AreaSubzone, Age Group, Sex and Type of Dwelling, June 2011-2019, and
- URA Master Plan 2014 Planning Subzone boundary

## Grading Criteria

This exercise will be graded by using the following criteria:

1. **Geospatial Data Wrangling:** This is an important aspect of geospatial analytics. You will be assessed on your ability to employ appropriate R functions from various R packages specifically designed for modern data science such as readr, tidyr, dplyr, sf just to mention a few of them, to perform the entire geospatial data wrangling processes, including. This is not limited to data import, data extraction, data cleaning and data transformation. Besides assessing your ability to use the R functions, this criterion also includes your ability to clean and derive appropriate variables to meet the analysis need. **(Warning:** All data are like vast grassland full of land mines, your job is to clear those mines and not to step on them). (20 marks)
2. **Geospatial Analysis:** In this exercise, you are expected to use hierarchical clustering and spatially constrained clustering techniques and other related R functions introduced in class to analysis the geospatial data prepared. You will be assessed on your ability:
  - a. to delineate social areas using hierarchical clustering technique and describe the analysis result; and
  - b. to delineate social areas using spatially constrained clustering technique and describe the analysis result. (45 marks)
3. **Geovisualisation:** In this section, you will be assessed on your ability to communicate the complex spatial statistics results in business friendly visual representations. This course is geospatial centric, hence, it is important for you to demonstrate your

competency in using appropriate geovisualisation techniques to reveal and communicate the findings of your analysis. (20 marks)

4. Bonus: Demonstrate your ability to employ methods beyond what you had learned in class to gain insights from the data. The methods used must be geospatial in nature. (15 marks)

## Deliverables

- The project folder in a single zip file format. The project folder should consist of the followings:
  - The project sandbox in data sub-folder (all raw, intermediate and final data files)
  - An R Markdown file contains all code chunks used and the written statements.
  - A html report knitted from the R Markdown document.
- A published version of the report on [RPubs](#).

## Submission Instructions

- The final deliverable (e.g. R Markdown file, project sandbox and report) must be zipped in a single zip file format.
- Name the zip file according to the course code and assignment, for example: **IS415\_Take-home\_Ex03**.
- The deliverable is to be submitted in softcopy. You are required to upload the zipped into the Dropbox of LMS before the stated assignment due date. Late work will be severely penalised. Students must check and confirm on LMS the assignment due date.

## Due Date

24<sup>th</sup> May 2020 (Sunday), 11.59pm (midnight).

## References

Social Area Analysis (<https://www.encyclopedia.com/social-sciences/dictionaries-thesauruses-pictures-and-press-releases/social-area-analysis>)

T. S. Brindley ; J. W. Raine (1979) "Social Area Analysis and Planning Research" *Urban Studies*, Vol.16 (3), p.273-289. (<https://journals-sagepub.com.libproxy.smu.edu.sg/doi/pdf/10.1080/713702552>)

Chavent, Marie et. al. (2018) "ClustGeo: an R package for hierarchical clustering with spatial constraints", *Computational Statistics*, Vol.33 (4), p.1799-1822.  
(<https://arxiv.org/abs/1707.03897>)