Master of Data Science - University of British Columbia Okanagan

Minutes for May 13, 2020 - 11:00 AM - 11:30 PM PST via Microsoft Teams Meeting

Present: Bruno St-Aubin (Statistics Canada), Marian Radulescu (Statistics Canada), Sofia Bahmutsky, Ngan Lyle, Kaitlyn Hobbs (Minutes), Shreeram Murali

Agenda

Discussion Points

1 — Discuss methods thus far.

2 — Discuss handling inconsistencies between scraped data and ODHF (facility source type).

3 — Discuss proposal to create a new database as opposed to adding metadata to ODHF for LTCs (ie. adding a column for non-profit/for profit LTCs).

4 — Discuss proximity measure aggregation and DB overlap with HR (Health Regions).

5 — Clarify usability of arcGIS versus QGIS

Meeting Notes

4 —

- **Aggregation of proximity measures:** weighted average for data aggregation alternatives: taking average or median. Sometimes DB may have low population (e.g. outskirts of city). In populated areas take high proximity measures.
- **Dealing with overlaps between DB and HR boundaries: (1)** Look at centroid of DB and proportion that overlaps HR. May occur for larger dissemination blocks. Assume that population dispersion across DB is equal. Marian will forward an e-mail from colleague regarding this methodology. **(2)** Could also run algorithm to classify the DB into an HR based on proportion overlap (HR acts as a decision boundary).

5 —

- **Technically discerning overlap of DB and HR:** QGIS can be used to intersect two layers of data, which will immediately identify overlaps. Processing can be done after overlap through QGIS. StatCan has python scripts relating to this that they can share with us. Similar spatial packages also exist in R.
- QGIS has more bugs but is more than enough for this project. arcGIS is a larger environment with cloud resources and servers.

2 —

- Forward discrepancies between scraped data and ODHF to Bruno and Marian once complete.

Integrating Census Navigator Attributes for Co-morbidities —

Data is from PHAC. We could potentially automate data retrieval. May be useful for visual relationships. Co-morbidities approach is already being looked into with PHAC but integrating with proximity data may be a more novel approach.

Action Items

- Inform Scott about decision to use QGIS.
- Install QGIS.
- Send Marian list of long-term care facility discrepancies with scraped data and ODHF.

Next Meeting: Wednesday, May 20, at 11:00 AM PST with Bruno and Marian via Microsoft Teams Meeting