Litterati crowd-sourced data as a supplement or complement to municipal litter audit data

1. Introduction
   1. Marine debris
      1. History of study
      2. Distribution
      3. Relationship to Plastic pollution
      4. Ecosystem effects
   2. Looking upstream
      1. River pollution
      2. Street litter
   3. Methods for studying
      1. Scientists
         1. Municipal audits
            1. Benefits
            2. Limitations
      2. Citizen science
         1. Litterati
            1. Benefits
            2. Limitations

Presence-only data

Correlaries in animal science observations

Statistical techniques to overcome it

Zero-inflated poisson

Methodological techniques to overcome it

Reporting effort

Examples from other projects

Preferences/understanding of “pollution”

Gum & glass

* 1. Objectives for this chapter

1. Investigation of Raw data for Vancouver (1x for Litterati, 1x for municipal data)
   1. Description of data
      1. Number of users/audits
         1. 75 users, 971 items, average 6.18 items/user-session (or is it 1431 items?)
         2. 3 audits, total 5536 items (avg 1496/year or 17/site/yr), 108 sites
      2. Collection area
         1. Avg user length:
         2. Site size: ~330m2
      3. Dates & frequency of collection
         1. Litterati: March 2017 – December 2019
         2. Audits: same week in September 2017, 2018, 2019
      4. Accuracy of data
         1. Number of null values
            1. Entries lacking “tag”: 179 (or is it 269)
         2. GPS accuracy
   2. Table: Trash items & frequency (and percent of total items made up by top 10 types)
   3. Map: Submissions + Audits
   4. Comparison
2. Spatial distribution
   1. Heat map by census block
      1. Comparison of litterati & municipal data
   2. Truncated/zero-inflated poisson [on census block or 100m grid cells?]
      1. Population, landuse, roadways, active users?
      2. Random effect: census block id
   3. Comparison
3. Best practices, uses for each data-type

DATA FORMATS NEEDED

1. Item types:
   1. Litterati: Year, month, day, user, municipal region, category, material, tag Y/N
      1. One row per item, filtered to only include 2017-2019
      2. Summarize by item material
   2. Municipal: summary in pdf report
2. Item Locations
   1. Litterati
      1. Municipal region with highest litterati count. Municipal region with highest litterati count/population.
         1. municipal region, year, litterati count
      2. Month with widest litterati spatial representation (visually?)
         1. Item type litterati, summarized by month (12 month x 3 years)
   2. Municipal
      1. Regression linking site characteristics and municipal counts
         1. Siteid, small count, large count, site characteristics
3. Item Counts:
   1. Litterati
      1. First, separately calculate mean per day count to report
         1. How quickly are litterati submissions coming in? And how different is the max litterati day from the municipal audit blitz in terms of total items characterized? How about max day and spatial spread?I
            1. “Biginflux.csv” includes all litterati submissions from 06/29-30/2019: estimate area surveyed during these two dates and compare against a municipal survey area

Expected evidence: if you advertise enough (and add features to data collection that ensure full area is surveyed, you can get similar coverage to a municipal audit.

* + 1. Summarize: municipal region, year, month, litterati count
       1. Answer questions about differences in monthly hotspots around the city (e.g. is the municipal September snapshot close to the annual average? Or higher, lower? How about which areas of the city show up as hotspots?)
    2. Summarize to join: municipal region, year, litterati count
  1. Municipal
     1. Group sites by municipal region
     2. Join with litterati count above & census data to get:
        1. Municipal region, year, litterati count, municipal count, census variables