Proposal

Group 3: Person A, Person B, Person C, Person D

due 4/13 [*6 pts*]

Once you have settled on a dataset, the next step is to identify interesting questions you hope to answer through a considered, in-depth analysis. Research questions in statistics can focus on both the motivating application (scientific knowledge) and on the statistical methodologies used (statistical knowledge). For example, an analysis of earthquake data using a Poisson point process model might be able to answer questions about the timing and spatial structure of earthquakes as well as what characteristics of Poisson point processes work well in that setting, and which might not work so well.

Your group will submit a proposal document to Canvas that details three primary research questions you plan to answer and what statistical methods you anticipate will be useful in answering those questions. The document will also include a description of what each member of your research group will do to contribute. The document should address the following questions:

1. What three primary research questions do you plan to answer?
    - Are there relationships between the frequency of road collisions and the frequency potholes, traffic lights, and/or liquor businesses? If so, what is the nature of these relationships?
    - Are the location and frequency of road conditions, potholes, traffic lights, and liquor businesses useful predictors of the frequency of collisions?
    - Do any of these findings change with the day / night cycle?
    - Descriptive Statistics: average collisions, max collisions in day, percent of collisions within XX miles of a liquor location
2. What statistical methods do you plan to use to answer those questions? For each method, explain why it is appropriate for your data and how it will help to answer at least one of your research questions.

    Data type: discrete, areal, equal-area units

    Response variable:

    1. collisions**:** discrete counts

    Predictor variables:

    1. overall condition index: “continuous”, $\in [0,100]$
    2. liquor businesses: discrete counts
    3. traffic lights: discrete counts

4. pothole counts: discrete counts
5. Day/Night (undefined atm)

Eligible Statistical Analyses:

1. multiple linear regression: a GLM with a Poisson or negative binomial response distribution, a log link function, and a CAR/SAR model for the linked residuals.

   collisions_count ~ overall_condition_index + pothole_count + liquor_count + streetlight_count + I(day/night)?

2. (exploratory) geographically weighted regression:
   GWR allows the relationships between the independent and dependent variables to vary by locality. GWR is useful as an exploratory technique – it allows visualization of stimulus-response relationships and if/how that relationship varies in space. It also accounts for spatial autocorrelation of variables.

3. cases and controls: Binary Regression Using Generalised Additive Models.

   We generated a control set of “collisions” that we know is independent of all predictor variables.

   We fit a GLM model with a linear component for distance to nearest XXXX, fit a GAM that accounts for small spatial dependence in the residuals.

   We ask, is one of these models better at uncovering the controls than the other? Why?

4. Related ideas:

   Random spatial index (point process) slides RLab 4 – Well drilling causing earthquakes

   SAR model (discrete, fixed spatial index slides RLab 3 - Bike injuries)

   Binary Regression Using Generalised Linear Models (non-gaussian spatial data slides Asthma example RLab 5 – “Does the expected number of electric vehicles increase or decrease as population increases?” we can change the question to “Does the expected number of collisions increase or decrease as traffic lights increase?”)

3. What will each person in the group do to contribute? Provide a detailed description of the responsibilities for each group member. Note: if these seem out of balance, I reserve the right to ask you to adjust these responsibilities.

	Person A	Person B	Person C	Person D
Report	Summarize findings in final report.	Introduction to data set in final report Ready all plots and graphics for both reporting and video. Construct appendix for final report.	Review class material for possible methods. Construct first draft of all milestones due for turn-in. Describe research goals in final report.	Explain the methods used in final report. Edit final draft of report.
Video	Edit the presentation video. Summarize findings in video.	Create supporting slides for video. Introduce dataset in video.	Video introduction, research goals in video.	Explain methods in video.
Coding	Bounding box, investigate missing values.	Plot tweaks, hex grid selection, standardize projections/CRS.	Organize code, save and store important data objects.	Geocoding, count aggregation, OCI to collision mapping.