**BST 5600 Project**

**Fall 2022**

**Overview**

The project should allow you the chance to apply spatial methods to a problem you have encountered. Most students in this class apply spatial regression (the BYM model), but other methods are acceptable. You are encouraged to talk to me about your topic, especially if it involves something other than the BYM model. Throughout the term, I have discussed the kinds of data that you will have to acquire. Now that we have covered the BYM model along with how to use INLA to get the Bayesian estimates, you should be able to proceed.

Here is an outline about how such a project would look. Models for areal data involve disease counts (or some other count) for cases that occur in predefined regions such as counties. You as the researcher might assume that the disease count depends on the population of the county and on other demographic variables for the county. For example, we will study cases of SIDS (sudden infant death syndrome) in North Carolina counties. Predictor variables might involve average income in the county, percent minority in the county, percent of college graduates in the county, etc. In applying these models you will be able to answer questions such as “What is the effect of variables X1, X2, and X3 on the rate of cases Y when we take into account the spatial correlation?”

You will likely need three sources of data for such a project:

1. the shape files for the state and its counties (Georgia would work great, because we already have the shape files for Georgia, but it’s not hard to find the shape files for other states and counties.)

2. the outcome variable (a count variable for each county in the state)

3. demographic characteristics for each county in the state.

There will probably be three or more sources for the data.

**Deliverables**

1. You should prepare a paper, written as if you were submitting it to a journal. The cover page should contain a title, your name, and an abstract. In the body of the paper,

1. describe the data (and sources),
2. suggest a research question (e.g., do variables X1, X2, or X3 affect the outcome Y?)
3. your methods (probably the BYM model with parameters estimated with INLA),
4. your results (regression analyses, maps, other figures, etc.),
5. discussion (where you give the answers to the questions you posed, and
6. references.

Quality is more important than length, but think of 5 to 10 pages, including the cover page, maps, and references.

2. A 12-15 minute presentation on your work.

**Competencies**

The following is from the syllabus:

|  |  |  |
| --- | --- | --- |
| **Program Competency** | | **Assessment** |
| MPH-BST 3 | Create statistical reports and presentations using appropriate graphical and numerical summaries and narrative explanation. | Homework assignments 3 – 6 and the term project will assess students’ ability to present and explain graphical and numerical summaries along with narrative explanation. |
| MS-BSTHA 6 | GEOSPATIAL DATA MANAGEMENT: Acquire, manage, analyze, and display geospatial health data. | Homework assignments 1 and 2, and the final project will assess students’ ability to acquire and manage data. The final project will assess their ability to analyze and display geospatial health data. |

**Grading**

Both the written report and the presentation will be graded using similar criteria:

1. Have you posed a relevant research question?
2. Did you follow the guidelines?
3. Did you apply the methodology correctly?
4. Are your maps, figures, tables, etc., helpful in conveying your message?
5. Are you able to answer questions about your work?
6. Do your report and presentation demonstrate the course competencies described above?

**Timing**

May 9, Four presentations

May 11, Remaining eight presentations

May 13, 5:00 pm. Written project due.