Stat 534 Exam 3: Take Home

This exam will work through the analysis of a dataset collected in Ballina, NSW, Australia with the intent of studying flying fox behavior. The dataset contains counts of flying foxes observed to pass above a property boundary. In addition to the counts, the dataset also contains a set of variables related to property size and land use that researchers believe may help explain flying fox behavior.

```
# modified May 5 2pm
# load(url('https://github.com/Stat534/data/raw/refs/heads/main/ballina.RData'))
# head(ballina_out)

load(url('https://github.com/Stat534/data/raw/refs/heads/main/ballina_small.RData'))

head(ballina_small)
```

```
Simple feature collection with 6 features and 7 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: 153.5569 ymin: -28.78919 xmax: 153.5836 ymax: -28.773
Geodetic CRS:  GDA94
# A tibble: 6 x 8
    CADID Lot_Area_k                   geometry path_count nature developed water
    <dbl>      <dbl>        <MULTIPOLYGON [°]>      <int>  <dbl>     <dbl> <dbl>
1  1.04e8     43814. (((153.575 -28.78523, 15~          1   5.26      31.6    0
2  1.08e8    202338. (((153.5831 -28.78817, 1~          1   0          2.5    0
3  1.69e8    354906. (((153.5717 -28.78667, 1~          2   1.44      0.719   0
4  1.79e8    788696. (((153.5678 -28.78623, 1~          6   0         1.58    6.65
5  1.80e8    522934. (((153.5668 -28.77484, 1~          1   0         0       1.44
6  1.79e8    575363. (((153.5721 -28.77414, 1~          1   0.873     0       5.24
# i 1 more variable: agriculture <dbl>
```

The following variables are included in the dataset:

- CADID: property ID
- Lot_Area_k: square meters of property
- geometry: the polygon defining property boundaries
- path_count: number of flying foxes observed to pass over boundary
- nature: % of area in property boundaries defined as nature
- developed: % of area in property boundaries defined as developed
- water: % of area in property boundaries defined as water
- agriculture: % of area in property boundaries defined as agriculture

Note this dataset contains several disjoint sets of spatial units (cliques).

1. (4 points)

Create a choropleth of the `path_count` variable for the properties in this dataset. Make sure your figure has appropriate titles and labels. Also, add an informative caption to your figure.

```
ggplot() + theme_minimal()
```

Figure 1: Caption here

2. (2 points)

Assess whether path_count exhibits spatial structure.

3. (2 points)

CADID = 183033429 is the Ballina Byron Gateway Airport. How many flying foxes are observed crossing over this property?

4. (2 points)

How many neighbor properties border the Ballina Byron Gateway Airport (CADID = 183033429)?

5. (4 points)

Create a series of figures to explore how `Lot_Area_k`, `nature`, `developed`, `water`, and `agriculture` impact the `path_count`. Write a paragraph describing your what you see in these figures. You are welcome to consider any data transformations that seem appropriate.

6. (4 points)

Write out a statistical model to address the mean structure of the response (`path_counts`). You are welcome to use any transformations you explored in Q5. Note, we will return to add spatial correlation to this model in upcoming questions, so you do not need to include that for this question. Clearly define all notation in your model specification.

7. (4 points)

Fit the model specified in Q6. Provide a written interpretation of all parameters in the model.

8. (4 points)

Add to the statistical model in Q6 to include spatial structure in your proposed model. Clearly define all notation in your model specification.

9. (4 points)

Fit the model specified in Q8. Provide a written interpretation of all parameters in the model.

10. (4 points)

Compare the models from Q7 and Q9. Which do you prefer? Why?