Final Report

Russell and Frances

5th October 2022

Contents

1 Health of New Zealand Lakes **1**

1.1 Introduction 1

1.2 Introduction to our data 2

1.3 Methodology 3

1.4 Results 3

1.5 Discussion 3

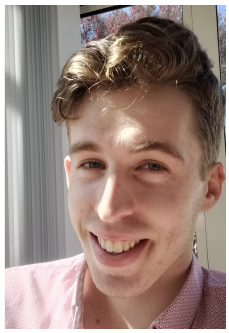
1.6 Bibliography 3

1 Health of New Zealand Lakes

1.1 Introduction

- intro dataset
- leading question
  - secondary questions
- outline of investigation

We are Russell and Frances and we form Group 8. Below are our pictures as well as our contact details and ORCID ID numbers.



Frances Smith
email: frances.j.smith.nz@gmail.com
ORCHID ID: 0000-0002-5168-3134

Russell Syder
email: russellsyder@gmail.com
ORCHID ID: 0000-0002-4582-5909

1.2 Introduction to our data

Our original dataset was extracted from Stats NZ. https://www.stats.govt.nz/indicators/modelled-lake-water-quality/

We made some manipulations to the original dataset such as removing extraneous variables (like Date which was the same for every lake) or variables that we were not interested in. We also reshaped some aspects of the dataset so that the data would be easier to analyse. We also checked for missing values; there were none so imputation was not necessary.

Our final dataset contains information from 3801 lakes in New Zealand measured across 22 variables. Of these 22 variables we selected 10 that we would use for further analysis. They are as follows:

Ammoniacal Nitrogen is a form of nitrogen that supports algae and plant growth, but in large concentrations can be toxic to aquatic life. This is measured in milligrams per litre. The national bottom line for this measure is 1.3mg/L, which none of the observations exceed. It acts as as measure of toxicity.

Chlorophyll-a is an organic molecule found in plant cells that allows plants to photosynthesize. The variable Chlorophyll-a is a measure of the concentration of phytoplankton biomass in milligrams per cubic metre. High concentrations of chlorophyll is a symptom of degraded water quality. The national bottom line for this measure is 12.

Total Phosphorus is the sum of all phosphorus forms in the water, including phosphorus bound to sediment. Large amounts of phosphorus in lakes can reduce dissolved oxygen in the water. This can cause low oxygen areas in the lake, where some aquatic life cannot survive. Total Phosphorus is measured in milligrams per cubic metre and has a national bottom line of 50mg/m3.

Total Nitrogen is the sum of all nitrogens found in the water, including organic nitrogen from plant tissue. An excess of nitrogen in lakes can cause an increase in algae and plant growth, possibly depriving the lake of oxygen. Total Nitrogen is measured in milligrams per cubic metre and the national bottom line for stratified lakes is 750mg/m3, and for polymictic lakes is 800mg/m3. There are observations exceeding the national bottom line.

Clarity is measured in Secchi depth. This is the maximum depth (in metres) a black and white Secchi disk is visible from the surface of the lake.

Area is the surface area of the lake measured in metres squared.

Perimeter is the overall perimeter of the lake, in metres.

Lake Depth is the maximum depth of the lake measured in metres.

Dominant Landcover is split into four types; Exotic Forest, Native, Pastoral and Urban area. There are 12 lakes with no entry for dominant land cover, however in the description of the dataset by Stats NZ, it states all lakes have been categorised, and indicated these empty entries should be another category called 'other' that includes 'Gorse and/or Broom', 'Surface mines and dumps', 'Mixed exotic shrubland', and 'Transport infrastructure' so we have assigned these to the Other category. The category Urban area is applied if urban cover exceeds 15 percent of catchment area. Pastoral is applied if pastoral exceeds 25 percent of catchment area and not already assigned urban. The other three categories; Exotic forest, Native, or Other were assigned according to the largest land cover type by area, if not already assigned urban or pastoral.

Regions in this dataset are; Auckland, Bay of Plenty, Canterbury, Gisborne, Hawke's Bay, Whanganui, Marlborough, Northland, Otago, Southland, Taranaki, Tasman, Waikato, Wellington and West Coast. Each lake corresponds to the region it is located in.

Upon first examining our data we thought that it would be prudent to group certain similar variables together for analysis. Specifically, the 4 variables that gave a measure of the levels of a given substance in a lake, and additionally clarity, we grouped as the "Lake Health variables" as for all of them, high levels of any of these variables can indicate poor lake health, with the exception of clarity, where, in general, higher values indicate better lake health.

We also grouped together lake Area, Perimeter, and Depth and classified this group as the Lake Dimension variables.

1.3 Methodology

- EDA
  - univariate health, dimension
    - distributions
    - cullen and frey
    - small sample stats
  - bivariate h/h, h/d, h/r, h/l, d/d, d/r
    - corr/cov matrices
    - pairs (from presentation cause pretty)
    - sbs box plots
    - small sample stats
- tests for normality

1.4 Results

- test for difference in means health/land + pairwise tests
- PCA
- FA
- LDA

1.5 Discussion

- WHAT DOES IT MEEEAAAAAANNNNNNN
- problems
  - clarity (not always possible as you may just have a shallow lake).

1.6 Bibliography

- I have some references from research into health variables and where data from