A20: Single Linear Regression

Your results should be delivered as an HTML webpage generated using R Markdown or Quarto. Make sure to include all code and results along with the answers to the questions. Provide text to describe your methods and results. This should read like the Methods and Results sections of a paper.

Grading Criteria

- Correctness and completeness of code (16 Points)
- Description of process and results (12 Points)
- Answer to questions (4 Points)
- Webpage formatting (4 Points)
- Map output (4 Points)

Note that these data were provided by Dr. Tim Warner and were used in the following publication:

Timothy A. Warner, Nicholas S. Skowronski & Michael R. Gallagher (2017) High spatial resolution burn severity mapping of the New Jersey Pine Barrens with WorldView-3 near-infrared and shortwave infrared imagery, *International Journal of Remote Sensing*, 38:2, 598-616, DOI: 10.1080/01431161.2016.1268739.

Your Task

In this exercise, you will explore single linear regression in R. Your goal will be to predict fire burn severity using different spectral ratios from the WorldView-3 sensor. The normalized burn ratio (NBR) is generally calculated as follows:

$$\frac{Near\ Infrared - Shortwave\ Infrared}{Near\ Infrared + Shortwave\ Infrared}$$

Because WorldView-3 has multiple near infrared and shortwave bands, the question arises as to what combination is best. To address this question, you will create regression models and compare them using root mean square error (RMSE) and R^2.

Data Preparation

You have been provided with a dataset (**burn_wv3.csv**) that contains the burn severity prediction ("Total" column) and the image band data for before (December, labeled in the table as "Dc") and after (April, labeled in the table as "Ap") a fire. From these data, you will need to create all possible differenced normalized burn ratio (dNBR) metrics by differencing the ratios before and after the fire. Each burn ratio change will be as follows:

$$\text{Pre-fire}\left(\frac{Near\ Infrared - Shortwave\ Infrared}{Near\ Infrared + Shortwave\ Infrared}\right) - \text{Post-fire}\left(\frac{Near\ Infrared - Shortwave\ Infrared}{Near\ Infrared + Shortwave\ Infrared}\right)$$

Bands 7 and 8 are near infrared bands while bands 12-16 represent shortwave bands. Generate all possible indices using the image band data and difference the pre- and post-fire data to obtain dNBR.

Regression Analysis

Create separate regression models to predict burn severity ("Total") with each dNBR. From the models and sample data, calculate the RMSE and R-squared metrics for each prediction. You do not need to separate the data into separate training and test data sets.

Graphs

Create two dot or bar plots, whichever you prefer, to compare the resulting RMSE and R-squared metrics for each model. Make sure that one axis represents the metric (RMSE or R-squared) and the other represents the different models. Also, make sure that each model has a descriptive name (maybe the band combination used). So, each model will have one data point associated with it representing the RMSE or R-squared value obtained.

Create a scatter plot using **ggplot2** from the result that provides the best performance based on R^2 and RMSE. dNBR should be mapped to the x-axis while the burn severity should be mapped to the y-axis.

Question

Q1: Which dNBR provided the highest R^2 and lowest RMSE values?