Portfolio Assignment 2 – Introduction to Cultural Data Science

Part 1.

Load the *'divorce_margarine'* dataset from the *'dslabs'* package. Investigate the correlation between margarine consumption and divorce rates in Maine. Would an increase in the preference for margarine lead to skyrocketing divorce rates?

Part 2.

Load the *'GSSvocab'* dataset from the *'car'* package. This dataset contains people's scores on an English vocabulary test and includes demographic information.

Filter for the year 1978 and remove rows with missing values (the function *na.exclude()* is one way to do this – check out the documentation!).

Is a person's score on the vocabulary test (*'vocab'*) significantly impacted by their level of education (*'educ'*)? Visualize the relationship in a plot and build a model. Briefly explain the results.

Whether a person is the native of an English-speaking country (*'nativeBorn'*) could potentially have an impact on the size of their vocabulary. Visualize the relationship and add the predictor to the model. Briefly explain the results.

Does a person's level of education depend on whether they are a native of the country? Visualize the relationship. Do you think it makes sense to add the relationship as an interaction term? Try creating the model and briefly explain the results.

Which model performs best?

When creating plots, try to play around with the options of *ggplot2* – add titles and axis labels, and consider using colours.