Education 240 Final Project

Professor Kisha Daniels

Matthew Cui

3/31/2020

In this project, I will be using statistical testing method, data visualization techniques, and content covered in lectures and readings to carry out an analysis on a dataset collected made available by the University of California, Irvine Department of Machine Learning.

First, let's read and load the data. In this chunk, I will also load the libraries that I will need to perform the necessary analysis used in the rest of this report.

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.0     v purrr   0.3.3
## v tibble  3.0.0     v dplyr   0.8.5
## v tidyr   1.0.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

This dataset has 649 rows and 33 columns, meaning that the researchers took a sample of 649 students and collected information about them on 33 different variables. The original data was titled `Student Alcohol Consumption` as it categorizes students into different consumption levels from 1-5 both on workdays and weekends, as shown by the variabels `dalc` and `walc` respectively.

However, this dataset also contains valuable insights into the students' academic performance as it contains their grades from year 1 to 3, shown in the variables `g1`, `g2`, and `g3` respectively. I will be creating a series of visualizations using this information and attempting to explain the trends I see using information learned in class and in my readings.