Case Study 3

Introduction

Figure 1: Bigfoot

For your final case study, you are tasked with helping Professor LaCombe with one of his passion projects- finding Bigfoot. He has collected (fake) data on what counties have and have had Bigfoot sightings in the last year, and collected county information he feels might be helpful in determining which county he should visit. These include county population, the percentage of the population that is rural, median income, the percentage of the county that has at least a college degree, the percentage of the population that trusts the federal government at least sometimes, the percentage of a county covered by trees, and a binary indicator for whether the county has a national park.

He is unsure of which variables to include, so your job is to identify the model that does the best job explaining Bigfoot sightings. Professor LaCombe was kind enough to check the linearity condition for you before sharing the data, so you do not need to worry about that condition. Your submission should include a brief description of the question and data, an overview of your methods (how you decided on the final model), and completion of the tasks below. No other visualizations are needed.

Tasks:

1. Determine which logistic regression model has the best explanatory power while not violating other conditions.
2. Interpret the regression model, noting the direction and significance of each coefficients (up to you if you prefer log odds or odds ratio- but only interpret your final model).
3. Based on the regression results, which county should Professor LaCombe visit to maximize his probability of seeing Bigfoot for his summer vacation? Make sure to report the county state and name, as well as the probability of seeing Bigfoot.

The data is accessible via github using the sample code below:

```
library(readr)
urlfile <- "https://raw.githubusercontent.com/lacombe129/SDS291_Fall23/main/inclass_r/bigf
mydata<-read_csv(url(urlfile), show_col_types=F)
```