Project proposal

Passionate Puffins

#install.packages("tidyverse")  
library(tidyverse)  
library(broom)  
library(readxl)

## 1. Introduction

Our data comes from all High Schools in the Androscoggin Area (this will be updated for specificity once we have specific school names). They surveyed their students about drug use, alcohol consumption, smoking habits, eating habits, home lives, and how much support they felt they were recieving from their parents. We will be using this data to evaluate the health of students at Androscoggin County High school compared to how much support they feel they are recieving from their parents. We would also like to compare the differences in support and unhealthy behavior between gender (i.e.: are female students more likely to smoke/drink/do drugs if they have low familial support than male students?).

In a study by the NIH (<https://pmc.ncbi.nlm.nih.gov/articles/PMC8656923/>), the association between parental support and the rate of smoking in youth was found to be positive. It will be interesting to compare these results to our findings about the relationship between support and unhealthy behavior.

Research Question: How big a role do home support systems play in the health of Androscoggin youth?

## 2. Data

Drinking\_data <- read\_xlsx("../data/How many times in the past year (12 months) have you been drunk or high at school? (hn126) Percentage of students who answered at least 1 time.xlsx")  
#I don't have access to this data file to rename it. (Katie)  
  
family\_support\_data <- read.csv ("/cloud/project/data/love\_support\_family.csv")  
  
Drinking\_vs\_family <- full\_join(Drinking\_data,   
 family\_support\_data,   
 by = "Grade")

#Drinking\_data <- Drinking\_data |>  
 # rename("Percent Total Ague" = "Percent Total Age")

view (Drinking\_vs\_family)  
view (Drinking\_data)

glimpse(Drinking\_data)

## Rows: 9  
## Columns: 10  
## $ Age <chr> NA, NA, "<=14", "15.0", "16.0", "17.0", ">=18"…  
## $ `Percent Total Ague` <dbl> 0.125, NA, 0.108, 0.132, 0.116, 0.130, 0.140, …  
## $ `Female Age` <dbl> 0.115, NA, 0.086, 0.134, 0.112, 0.104, 0.152, …  
## $ `Male Age` <dbl> 0.136, NA, 0.142, 0.132, 0.123, 0.158, 0.130, …  
## $ Grade <chr> NA, NA, "Grade 9", "Grade 10", "Grade 11", "Gr…  
## $ `Percent Total Grade` <dbl> NA, NA, 0.109, 0.115, 0.142, 0.133, NA, NA, NA  
## $ `Female by Grade` <dbl> NA, NA, 0.130, 0.091, 0.119, 0.122, NA, NA, NA  
## $ `Male by Grade` <dbl> NA, NA, 0.083, 0.147, 0.165, 0.144, NA, NA, NA  
## $ Race <chr> NA, NA, "American Indian/Alaskan Native", "Asi…  
## $ `Percent Total By Race` <chr> NA, NA, "0.218", "\*", "0.048", "0.305", "\*", "…

glimpse(Drinking\_vs\_family)

## Rows: 17  
## Columns: 39  
## $ Age.x <chr> NA, NA, "<=14", "15.0", "16.0", "17.0", ">=18"…  
## $ `Percent Total Ague` <dbl> 0.125, NA, 0.108, 0.132, 0.116, 0.130, 0.140, …  
## $ `Female Age` <dbl> 0.115, NA, 0.086, 0.134, 0.112, 0.104, 0.152, …  
## $ `Male Age` <dbl> 0.136, NA, 0.142, 0.132, 0.123, 0.158, 0.130, …  
## $ Grade <chr> NA, NA, "Grade 9", "Grade 10", "Grade 11", "Gr…  
## $ `Percent Total Grade` <dbl> NA, NA, 0.109, 0.115, 0.142, 0.133, NA, NA, NA…  
## $ `Female by Grade` <dbl> NA, NA, 0.130, 0.091, 0.119, 0.122, NA, NA, NA…  
## $ `Male by Grade` <dbl> NA, NA, 0.083, 0.147, 0.165, 0.144, NA, NA, NA…  
## $ Race <chr> NA, NA, "American Indian/Alaskan Native", "Asi…  
## $ `Percent Total By Race` <chr> NA, NA, "0.218", "\*", "0.048", "0.305", "\*", "…  
## $ Age.y <chr> NA, NA, "14 or younger", "15", "16", "17", NA,…  
## $ Age.. <chr> NA, NA, "79.80%", "77.20%", "81.80%", "79.70%"…  
## $ Age.CI <chr> NA, NA, "(76.0%-83.5%)", "(75.3%-79.2%)", "(76…  
## $ Age.N <chr> NA, NA, "489", "845", "871", "718", NA, NA, NA…  
## $ Female.Age.. <chr> NA, NA, "81.60%", "76.00%", "82.50%", "83.40%"…  
## $ Female.Age.CI <chr> NA, NA, "(70.7%-92.5%)", "(69.2%-82.8%)", "(79…  
## $ Female.Age.N <chr> NA, NA, "306", "444", "467", "388", NA, NA, NA…  
## $ Male.Age.. <chr> NA, NA, "77.70%", "78.70%", "81.90%", "75.80%"…  
## $ Male.Age.CI <chr> NA, NA, "(65.2%-90.2%)", "(74.3%-83.2%)", "(72…  
## $ Male.Age.N <chr> NA, NA, "181", "397", "400", "330", NA, NA, NA…  
## $ Grade.. <chr> NA, NA, "79.20%", "78.50%", "80.60%", "79.00%"…  
## $ Grade.CI <chr> NA, NA, "(76.6%-81.9%)", "(73.9%-83.2%)", "(78…  
## $ Grade.N <int> NA, NA, 872, 859, 795, 692, NA, NA, NA, NA, NA…  
## $ Female.grade.. <chr> NA, NA, "81.00%", "77.50%", "83.60%", "80.00%"…  
## $ Female.grade.CI <chr> NA, NA, "(73.5%-88.5%)", "(73.4%-81.5%)", "(79…  
## $ Female.grade.N <int> NA, NA, 497, 465, 405, 386, NA, NA, NA, NA, NA…  
## $ Male.grade.. <chr> NA, NA, "77.10%", "80.10%", "78.30%", "77.70%"…  
## $ Male.grade.CI <chr> NA, NA, "(66.0%-88.1%)", "(71.9%-88.3%)", "(73…  
## $ Male.grade.N <int> NA, NA, 371, 387, 390, 304, NA, NA, NA, NA, NA…  
## $ Race.Ethnicity <chr> NA, NA, "American Indian or Alaskan Native\*\*",…  
## $ Race.. <chr> NA, NA, "61.60%", "73.50%", "74.60%", "77.40%"…  
## $ Race.CI <chr> NA, NA, "(33.7%-89.5%)", "(67.9%-79.2%)", "(71…  
## $ Race.N <chr> NA, NA, "28", "36", "159", "114", NA, NA, NA, …  
## $ Race.Female.. <chr> NA, NA, "^", "^", "^", "^", NA, NA, NA, "^", "…  
## $ Race.Female.CI <chr> NA, NA, "^", "^", "^", "^", NA, NA, NA, "^", "…  
## $ Race.Female.N <chr> NA, NA, "^", "^", "^", "^", NA, NA, NA, "^", "…  
## $ Race.Male.. <chr> NA, NA, "^", "^", "^", "^", NA, NA, NA, "^", "…  
## $ Race.Male.CI <chr> NA, NA, "^", "^", "^", "^", NA, NA, NA, "^", "…  
## $ Race.Male.N <chr> NA, NA, "^", "^", "^", "^", NA, NA, NA, "^", "…

dim(Drinking\_data)

## [1] 9 10

dim(Drinking\_vs\_family)

## [1] 17 39

Our first data set is called Drinking\_data which is a 10x9 deta set with 10 different variables and 70 observations that we are going to be able to incorporate into visualizations. There is no code book for the data set as I had to manually input all of the data into an excel sheet as the format we were given the data was not suitable to use in r. We have added an additional data set called Grow\_vs\_family that has the dimensions of 17x39 with roughly 250 observations and 15 variable. We are still in the process of finalizing our data. We have categorical variables as well as discrete numerical variables within our set. As I add more data to the set we will be able to include continuous numerical variables which will allow us to see a range of percentages for variables.

## 3. Data analysis plan

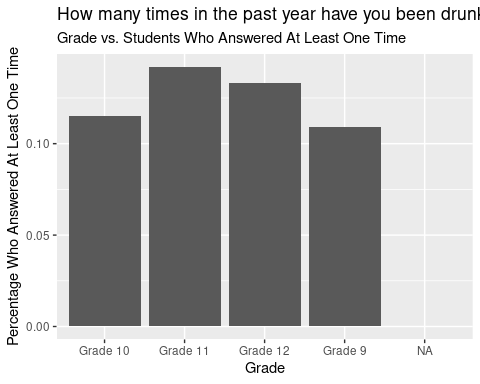
Based on our research question, we plan to compare various indications of good support systems against indications of unhealthy youth behavior. For example, we plan to create a visualization and analysis to contrast how much a child feels supported by their parents against how easy it would be for them to attain drugs. How effective is a good familial support system (how much the student feels supported by their parents) in preventing unhealthy behavior (buying/using drugs)?

To answer the question effectively, we will create multiple visualizations similar to this, including:

* How many times in the past year (12 months) have you been drunk or high in school (percentage of students who answered one time or more)? vs. Students who answered ‘Very or often’ or ‘Extremely or almost always’ to ‘I have parents who are good at talking with me about things.’
* During the past 30 days, on how many days did you smoke cigars, cigarillos, or little cigars (percentage of students who answered ‘at least one day’)? vs. Students who answered ‘Very or often’ or ‘Extremely or almost always’ to ‘I have parents who are good at talking with me about things.’
* During the past 30 days, how many times did you drive a car or other vehicle when you had been drinking alcohol (among students who drove a car or other vehicle during the past 30 days, the percentage who answered at least 1 time)? vs. Percentage of students who answered ‘Very or often’ or ‘Extremely or almost always’ to ‘I have a family that gives me love and support.’

We plan to categorize and consider variables within the datasets such as grade, age, race, and gender. We will likely use a bar graph similar to the one below to visualize the patterns between the data.

Drinking\_data |>  
 ggplot(aes(x = Grade, y = `Percent Total Grade`)) +  
 geom\_col(na.rm = TRUE) +  
 labs(x = 'Grade',   
 y = 'Percentage Who Answered At Least One Time',   
 title = 'How many times in the past year have you been drunk/high at school?',   
 subtitle = "Grade vs. Students Who Answered At Least One Time")



## 4. Data Ethics Review

The data ethics review section will be introduced in a separate class and is not part of the original proposal deadline.