Project 1 Graphs

```r
library(readr)
MSHS_DropoutRates <- read_csv("STAT 3220_ Project Part 1 - Sheet1.csv")
```

```
## Rows: 122 Columns: 10
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): County, Region of Virginia - Snigdha, Student Behavior
## dbl (6): Annual Drop out Rate, Percentage of People Over 25 with less than a...
## num (1): Average Individual Income - Karen
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
names(MSHS_DropoutRates) <- gsub(" ", "_", names(MSHS_DropoutRates))
#View(MSHS_DropoutRates)
```

Selection of Data Rows

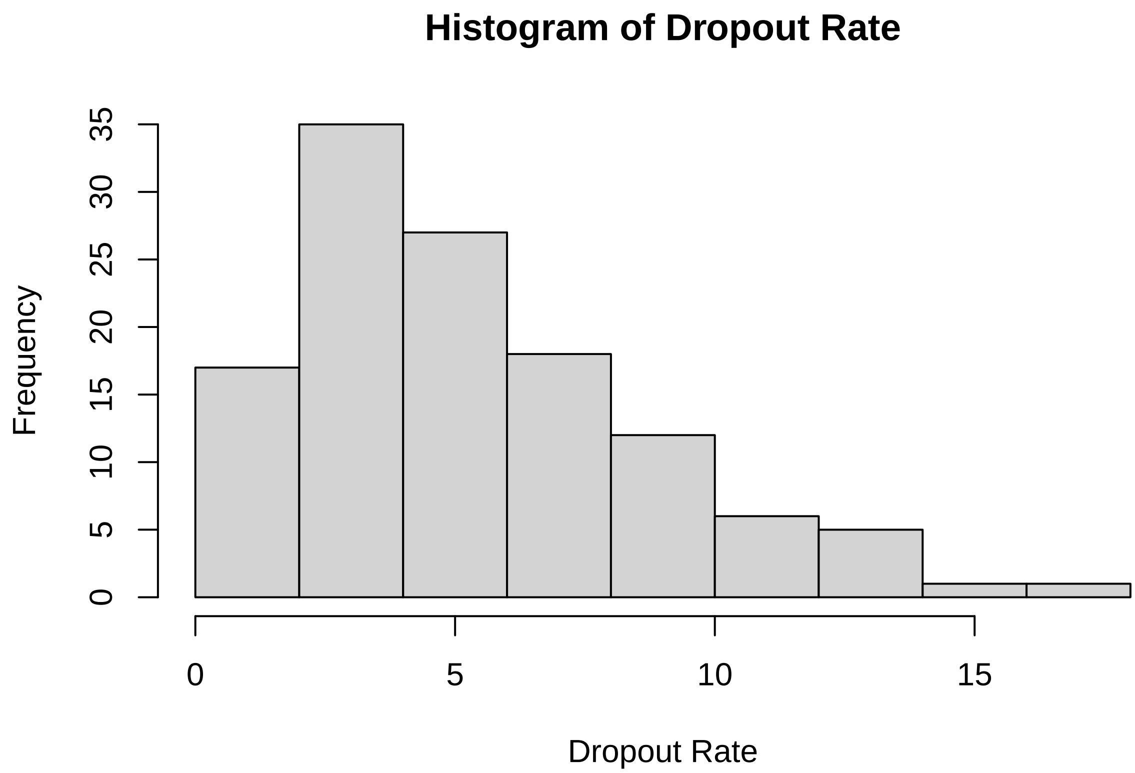
```r
#for appendix B
head(MSHS_DropoutRates)
```

A tibble: 6 x 10

County Annual_Drop_out_Rate Average_Individual_I~1 Percentage_of_People~2 1 Accomack 5.5 54153 18.4 2 Albemarle 3.77 91201 6.2 3 Alexandria 12.8 105239 6.3 4 Alleghany ~ 2.48 46250 11.8 5 Amelia Cou~ 4.76 52977 13.6 6 Amherst 2.34 44723 12.3 # i abbreviated names: 1: `Average_Individual_Income_-_Karen`, # 2: `Percentage_of_People_Over_25_with_less_than_a_High_School_Degree_` # i 6 more variables: `Num.Truancies_2023-2024_(Counts)_-_Stephanie` , # `Percentage_White_-_Snigdha` , `Region_of_Virginia_-_Snigdha` , # `Fare-free_public_transportation_Project_-_Karen` , # `Rural_(0)_vs._Urban_(1)_-_Antigone` , Student_Behavior

Histogram

```r
hist(MSHS_DropoutRates$Annual_Drop_out_Rate, xlab="Dropout Rate", main="Histogram of Dropout Rate")
```




```
summary(MSHS_DropoutRates$Annual_Drop_out_Rate)
```

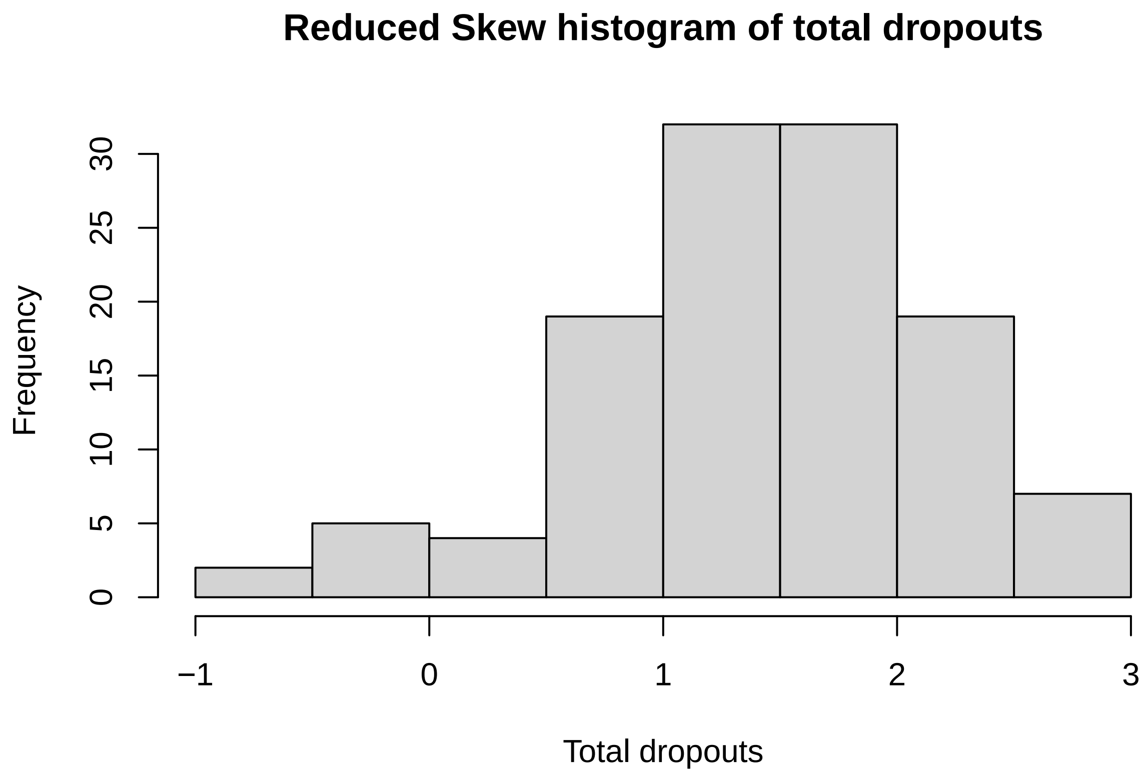
Min. 1st Qu. Median Mean 3rd Qu. Max. 0.000 2.623 4.370 5.260 7.138 17.690

```
sd_value <- sd(MSHS_DropoutRates$Annual_Drop_out_Rate, na.rm = TRUE)

cat ("Standard Deviation: ", round(sd_value,3))
```

Standard Deviation: 3.452

Graphical Summary #1

```
#transformed, adding constant to avoid log(0)
transformed_hist <- log(MSHS_DropoutRates$Annual_Drop_out_Rate+1)

hist(log(MSHS_DropoutRates$Annual_Drop_out_Rate), xlab="Total dropouts", main="Reduced Skew histogram o
```




```r
summary(transformed_hist)
```

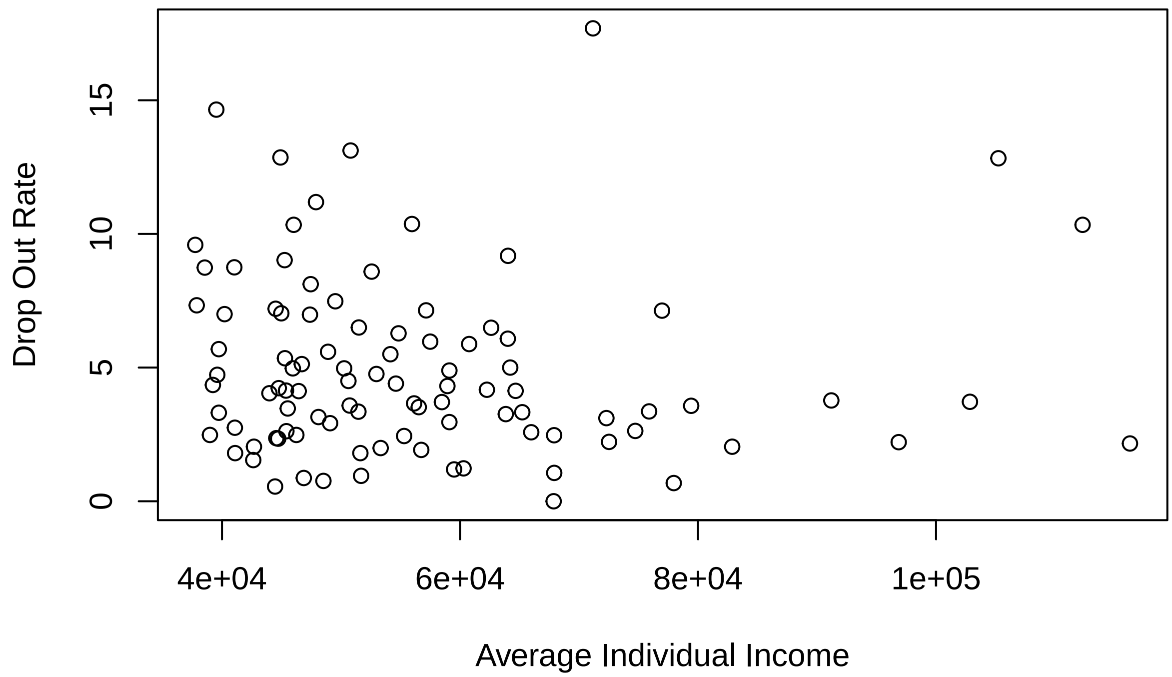
Min. 1st Qu. Median Mean 3rd Qu. Max. 0.000 1.287 1.681 1.677 2.096 2.928

```r
transformed_sd_value <- sd(transformed_hist, na.rm = TRUE)

cat ("Standard Deviation: ", round(transformed_sd_value,3))
```

Standard Deviation: 0.588

Graphical Summary #2

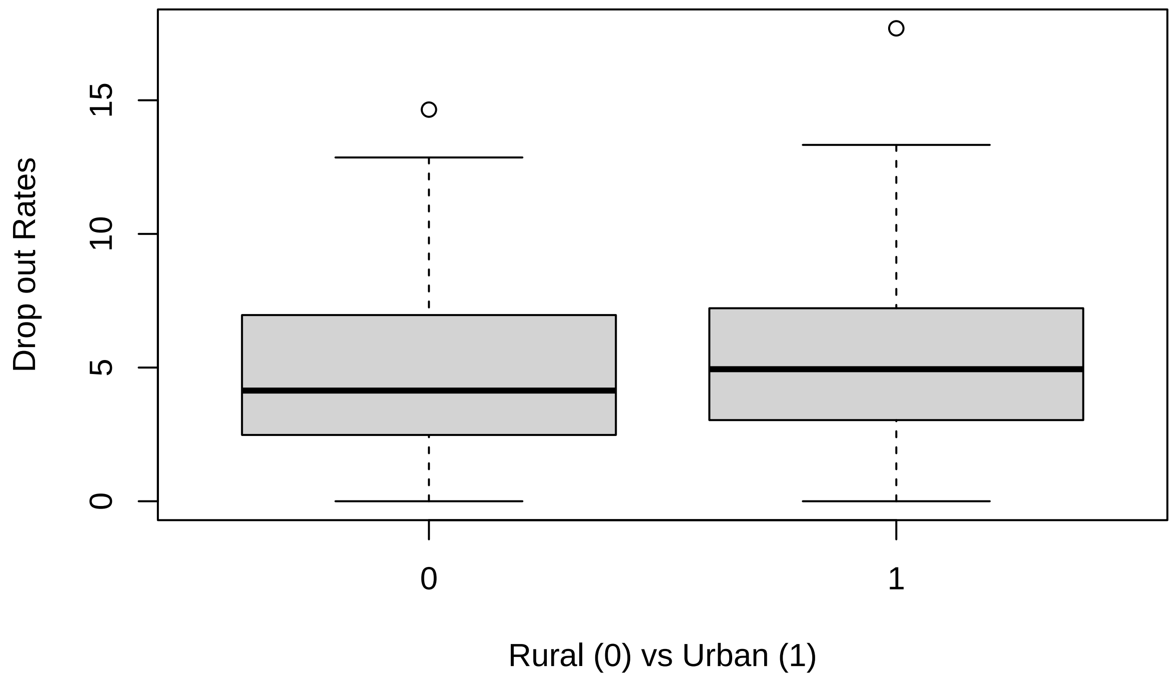
```r
plot(MSHS_DropoutRates$`Average_Individual_Income_-_Karen`, MSHS_DropoutRates$Annual_Drop_out_Rate, xlal
```



This graph allows us to further explore our question "Do high school students in areas with a higher individual income stay in school more frequently than those in areas with a lower personal income?" It ranges from 30,000 to 120,000 and graphs the dropout rate for each county.

Graphical Summary #3

```
boxplot(Annual_Drop_out_Rate~`Rural_(0)_vs._Urban_(1)_-_Antigone`, MSHS_DropoutRates, xlab = "Rural (0)
```




```
tapply(MSHS_DropoutRates$Annual_Drop_out_Rate, MSHS_DropoutRates$`Rural_(0)_vs._Urban_(1)_-_Antigone`, s
```

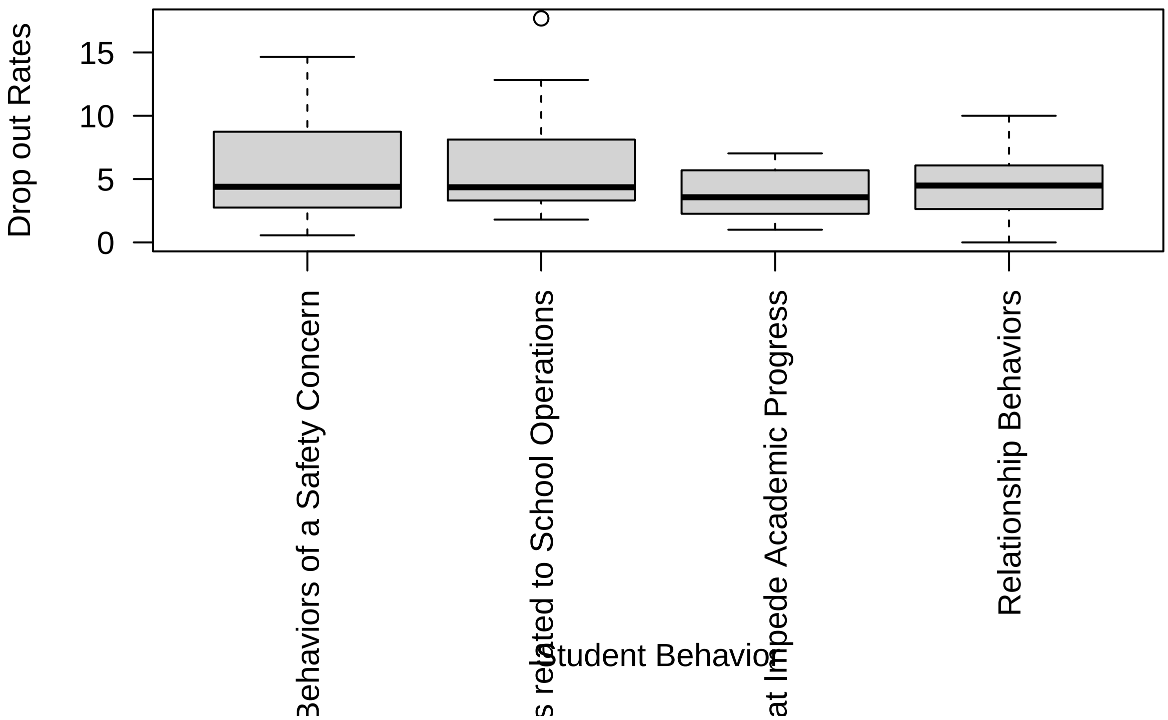
$0 Min. 1st Qu. Median Mean 3rd Qu. Max. 0.000 2.480 4.140 4.881 6.965 14.650

$1 Min. 1st Qu. Median Mean 3rd Qu. Max. 0.000 3.035 4.940 5.665 7.220 17.690

This box plot helps us contextualize our qualitative variable of rural vs urban. This directly relates to our question of the relationship between the category of the student's county and the dropout rate for that county.

Graphical Summary #4

```
par(mar = c(12, 4, 4, 2) + 0.1)
boxplot(Annual_Drop_out_Rate~Student_Behavior, MSHS_DropoutRates, xlab = "", ylab = "Drop out Rates", la
mtext("Student Behavior", side = 1, line = 10)  # adjusting x label
```




```
tapply(MSHS_DropoutRates$Annual_Drop_out_Rate, MSHS_DropoutRates$Student_Behavior, summary)
```

`$Behaviors of a Safety Concern` Min. 1st Qu. Median Mean 3rd Qu. Max. 0.560 2.750 4.390 5.758 8.740 14.650

`$Behaviors related to School Operations` Min. 1st Qu. Median Mean 3rd Qu. Max. 1.800 3.315 4.355 5.911 7.890 17.690

`$Behaviors that Impede Academic Progress` Min. 1st Qu. Median Mean 3rd Qu. Max. 1.000 2.370 3.565 3.820 5.293 7.030

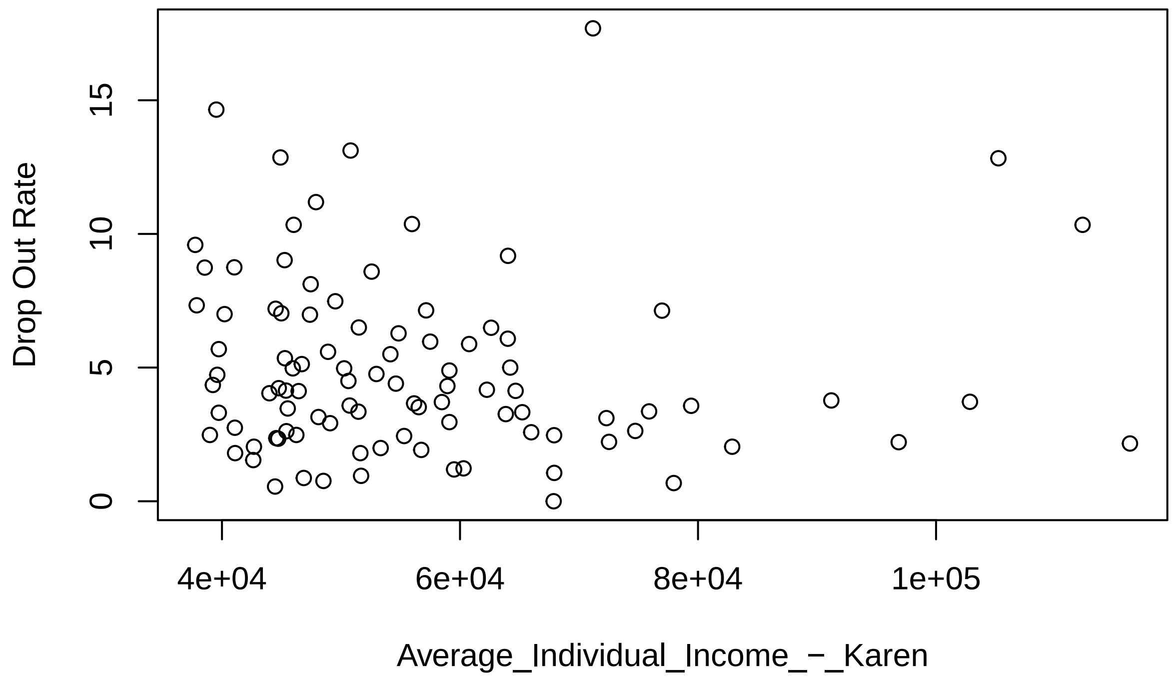
`$Relationship Behaviors` Min. 1st Qu. Median Mean 3rd Qu. Max. 0.000 2.810 4.490 4.390 5.935 10.000

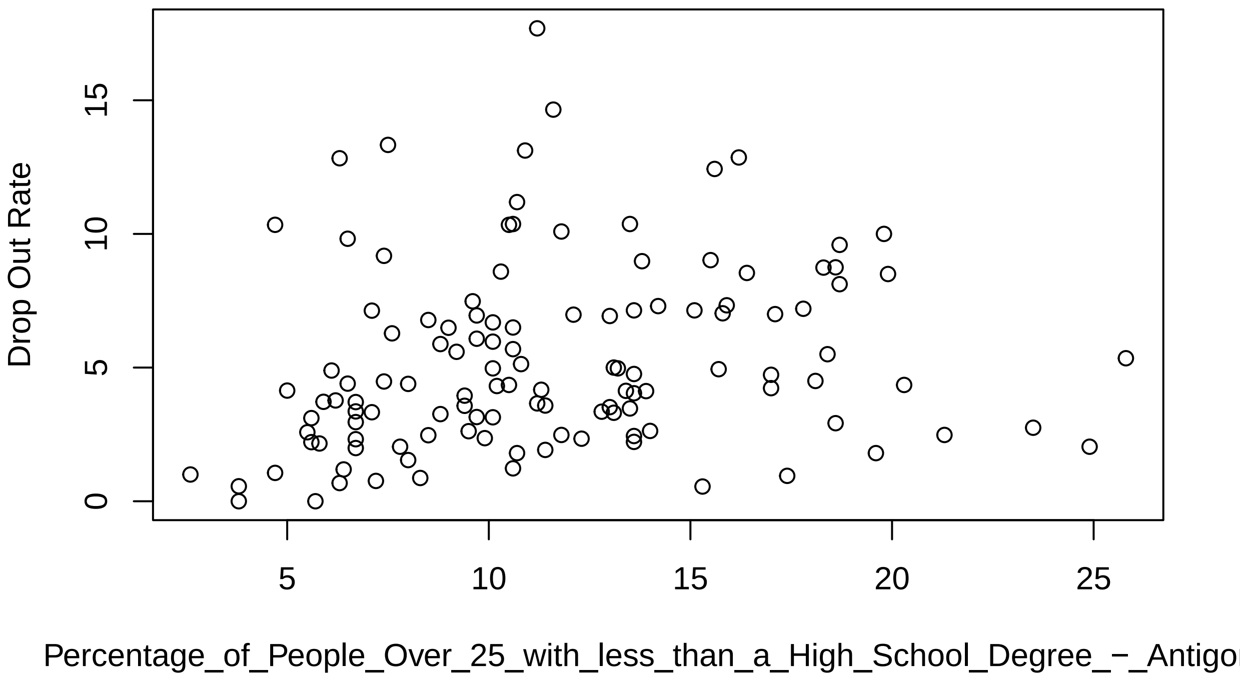
This is a box plot that plots the dropout rate of counties categorized by their highest student behavior issue.

Scatterplots

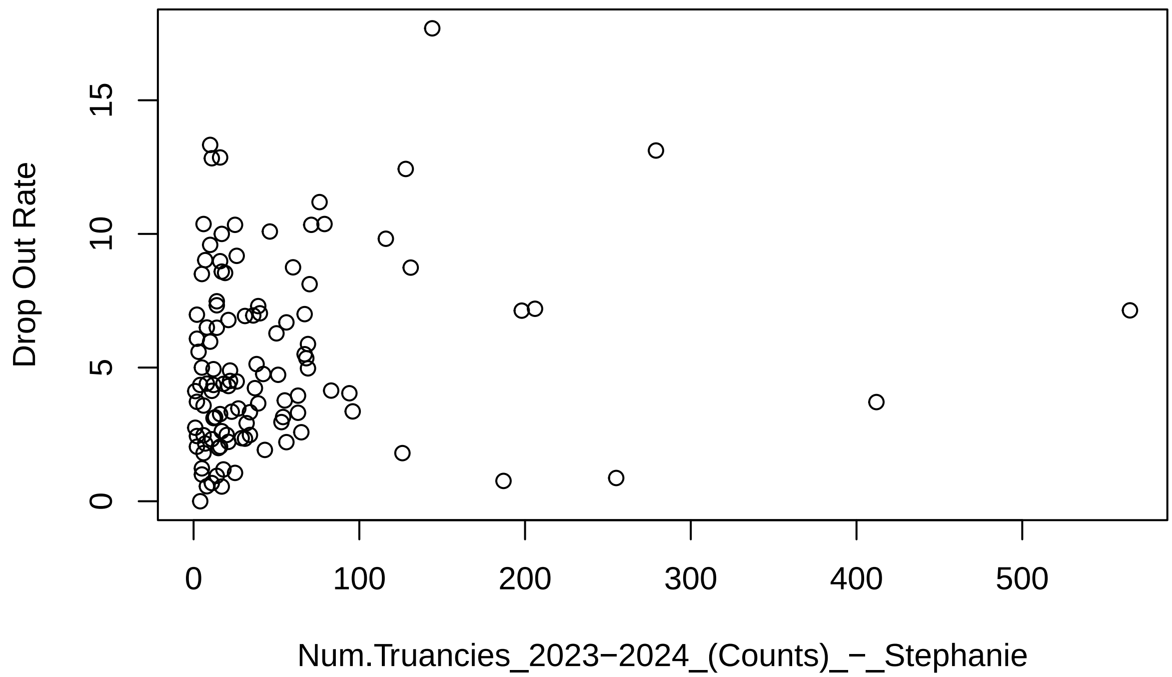
not necessarily directly related to our research questions but can be used if needed (can fix labels later)

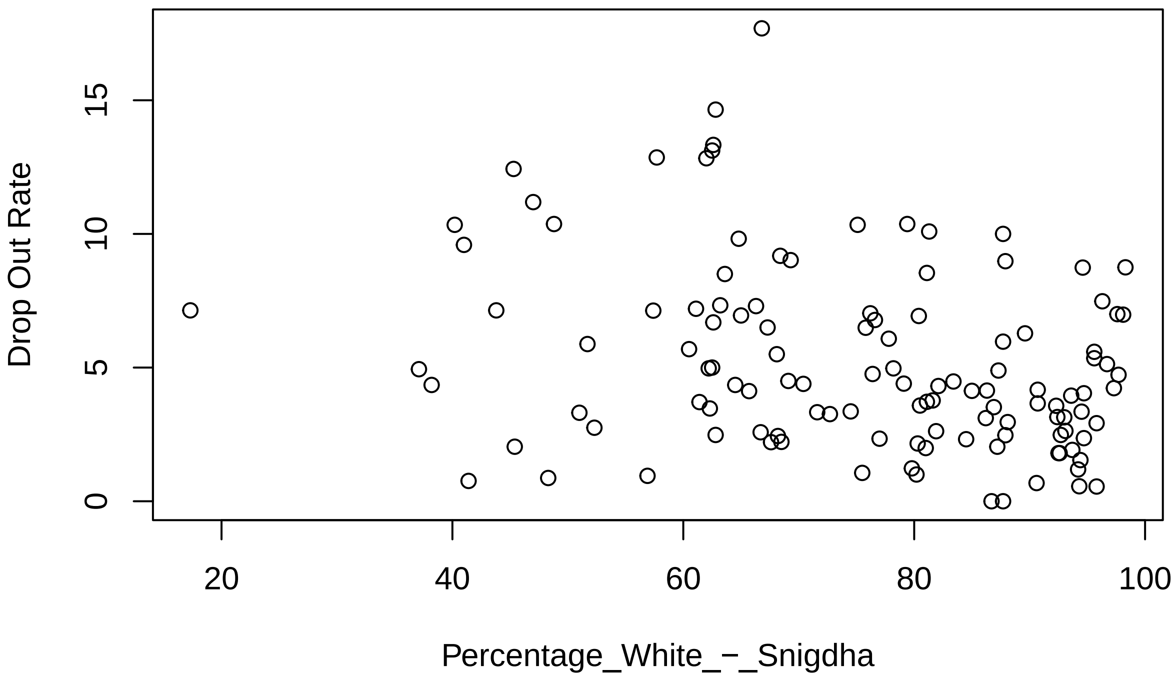
```
for (i in names (MSHS_DropoutRates)[3:6]){
  plot(MSHS_DropoutRates[[i]], MSHS_DropoutRates$Annual_Drop_out_Rate, xlab = i, ylab="Drop Out Rate")

}
```


Drop Out Rate
0
5
10
15
4e+04
6e+04
8e+04
1e+05
Average_Individual_Income_−_Karen



Drop Out Rate
0
5
10
15
5
10
15
20
25
Percentage_of_People_Over_25_with_less_than_a_High_School_Degree_−_Antigo


Drop Out Rate
0
5
10
15
0
100
200
300
400
500
Num.Truancies_2023−2024_(Counts)_−_Stephanie




```
round(cor(MSHS_DropoutRates[3:6], MSHS_DropoutRates$Annual_Drop_out_Rate),3)
```

```
[,1]
```

Average_Individual_Income_-*Karen NA Percentage_of_People_Over_25_with_less_than_a_High_School_Degree-Antigone 0.188 Num.Truancies_2023-2024* (Counts)_-*Stephanie NA Percentage_White*-_Snigdha NA