Homework 1C - DATA-312

Jeffrey Williams

Abstract

This writeup explores the dataset, FirstYearGPA, from the R library, Stat2Data. Categorical variables in this dataset utilize boolean values for identification. These variables are as follows: Male, FirstGen, White, CollegeBound. Numerical variables are as follows: GPA, HSGPA, SATV, SATM, HU, SS.

```
head("Stat2Data")
```

```
## [1] "Stat2Data"
```

```
summary("Stat2Data")
```

```
##    Length     Class      Mode
##         1 character character
```

```
sapply("Stat2Data", class)
```

```
##   Stat2Data
## "character"
```

Accessible as a library in R, this dataset comprises information from "a sample of 2019 first year students at a midwestern college", with the original intention of constructing an informed prediction of their first year GPA using various categorical and numerical variables.

```
##          used (Mb) gc trigger (Mb) max used (Mb)
## Ncells 422423 22.6     885067 47.3   643711 34.4
## Vcells 771724  5.9    8388608 64.0  1650011 12.6
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
```

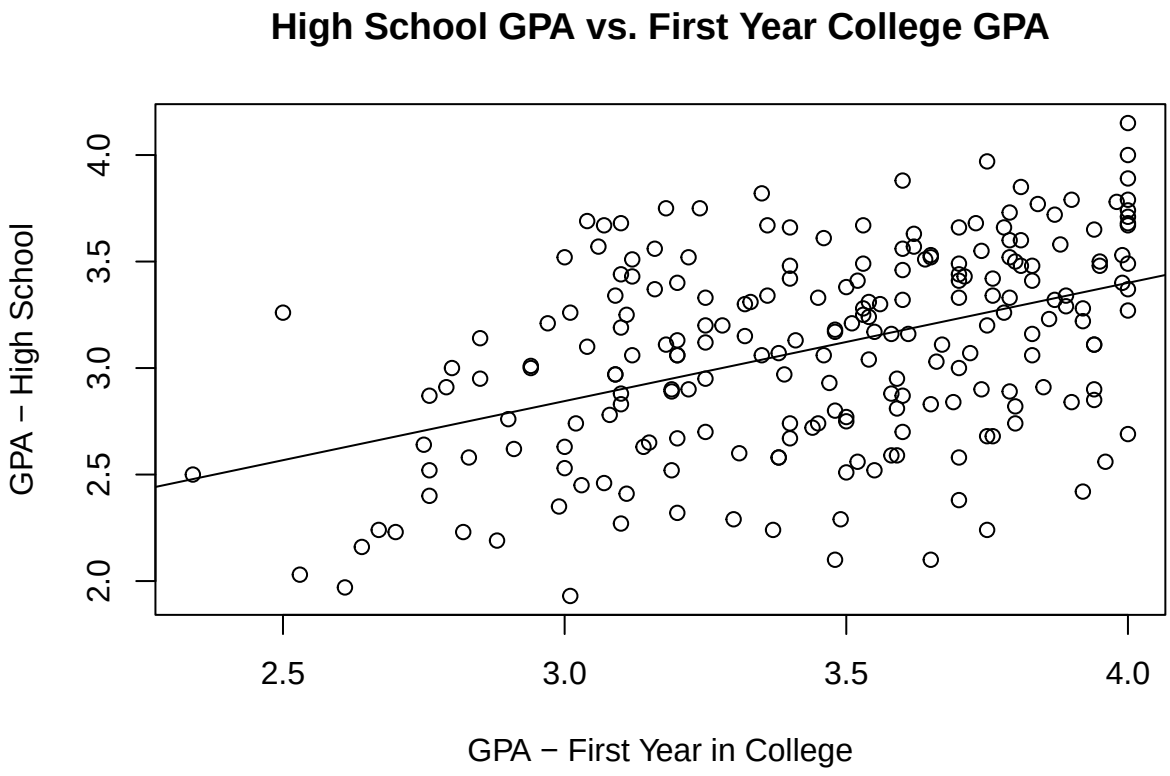
```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
## [1] 3.096164
```

```
## [1] 0.2166678
```

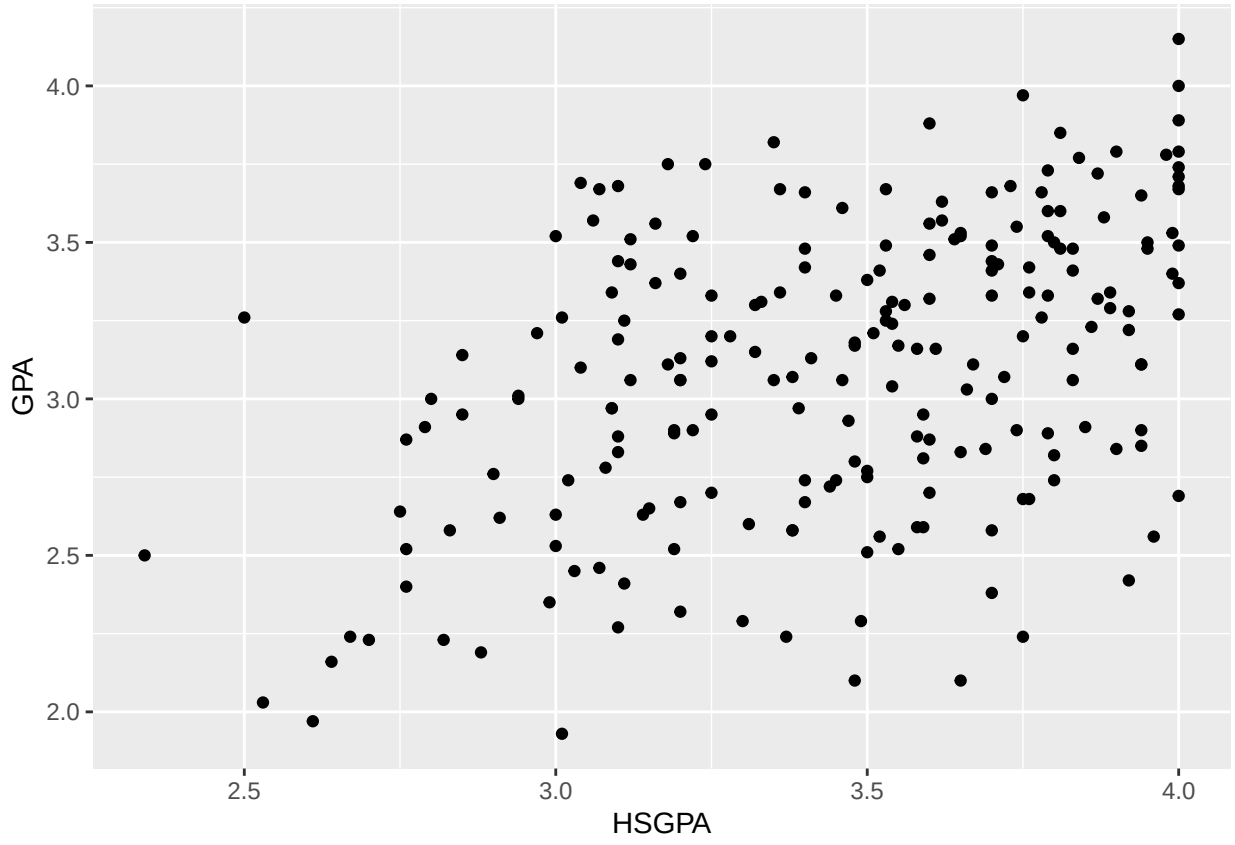
```
## [1] 0.4654759
```

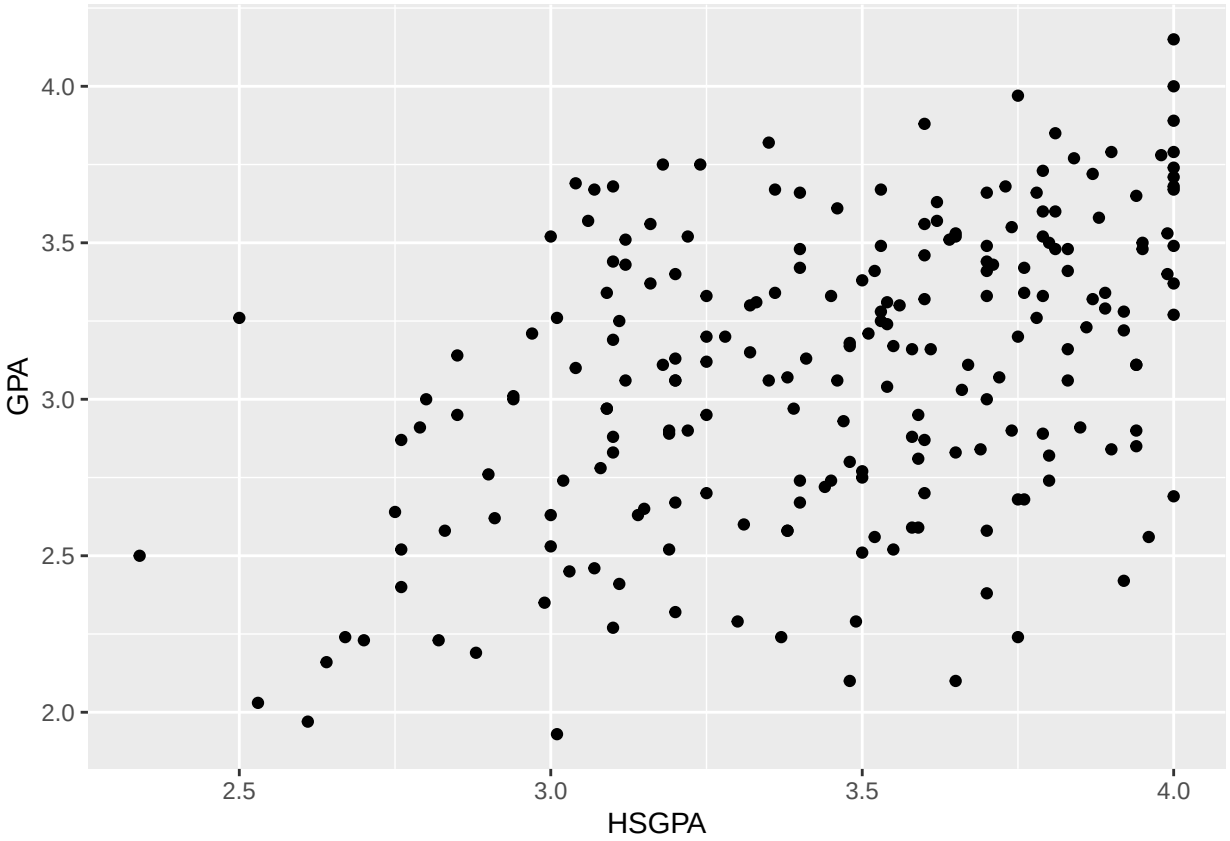
```
## [1] 0.4468873
```

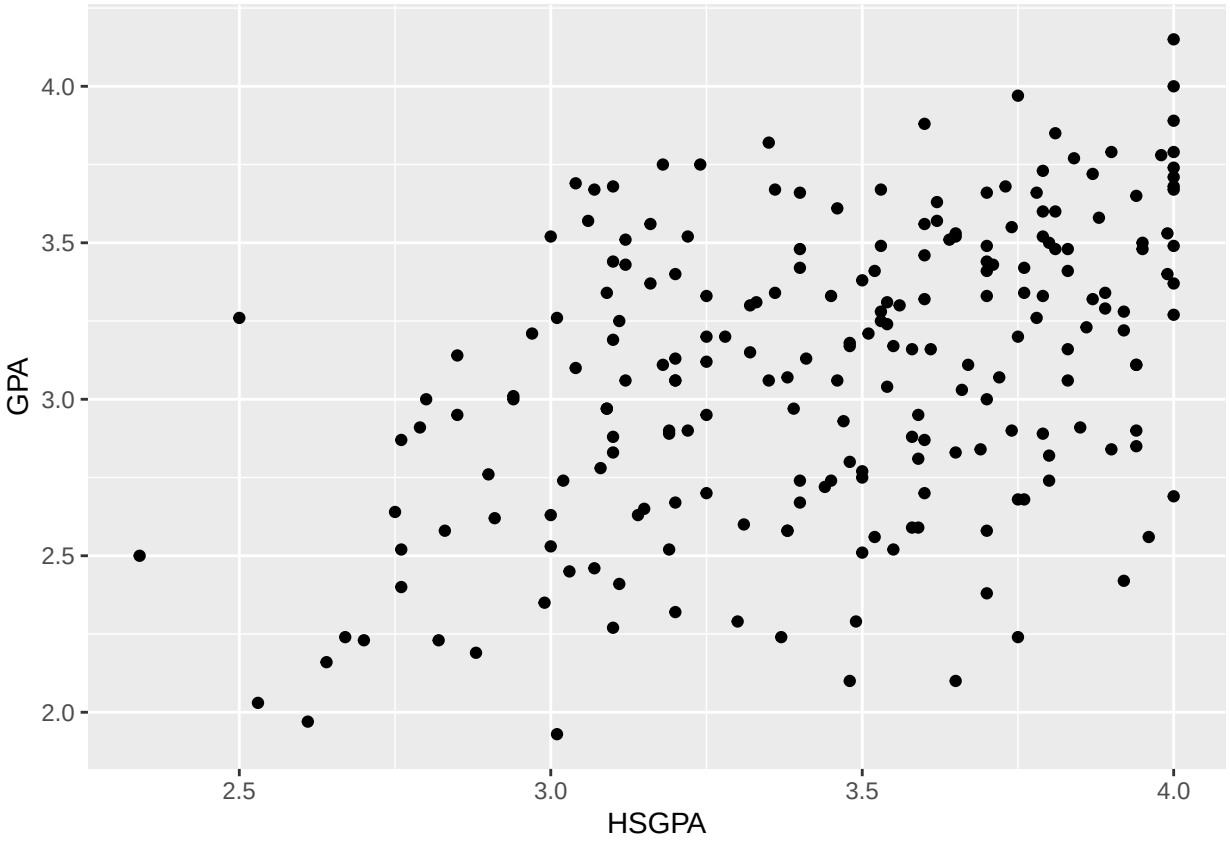


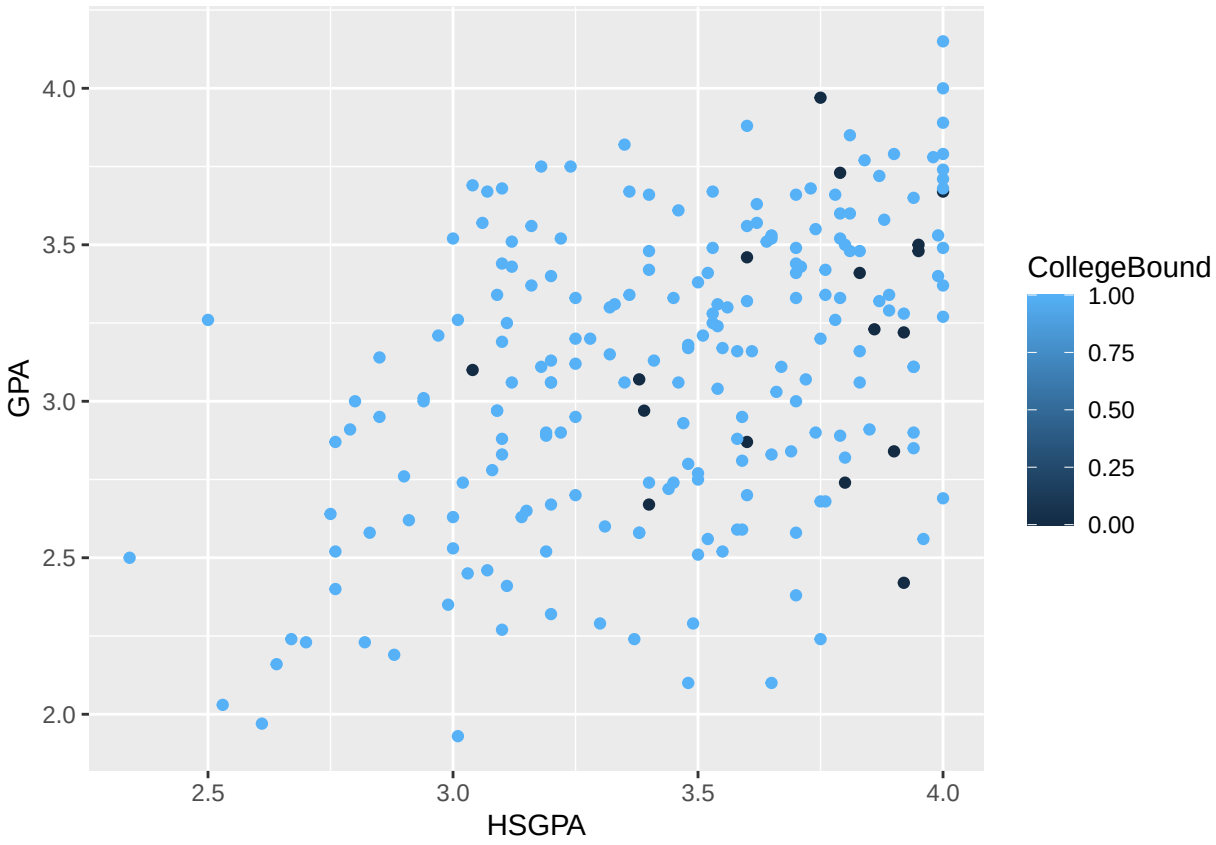

```
##
## Call:
## lm(formula = FirstYearGPA$GPA ~ FirstYearGPA$HSGPA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.10565 -0.31329  0.05871  0.29485  0.82291
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1.17985    0.26194   4.504 1.09e-05 ***
## FirstYearGPA$HSGPA  0.55501    0.07542   7.359 3.78e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4174 on 217 degrees of freedom
## Multiple R-squared:  0.1997, Adjusted R-squared:  0.196
## F-statistic: 54.15 on 1 and 217 DF,  p-value: 3.783e-12
```

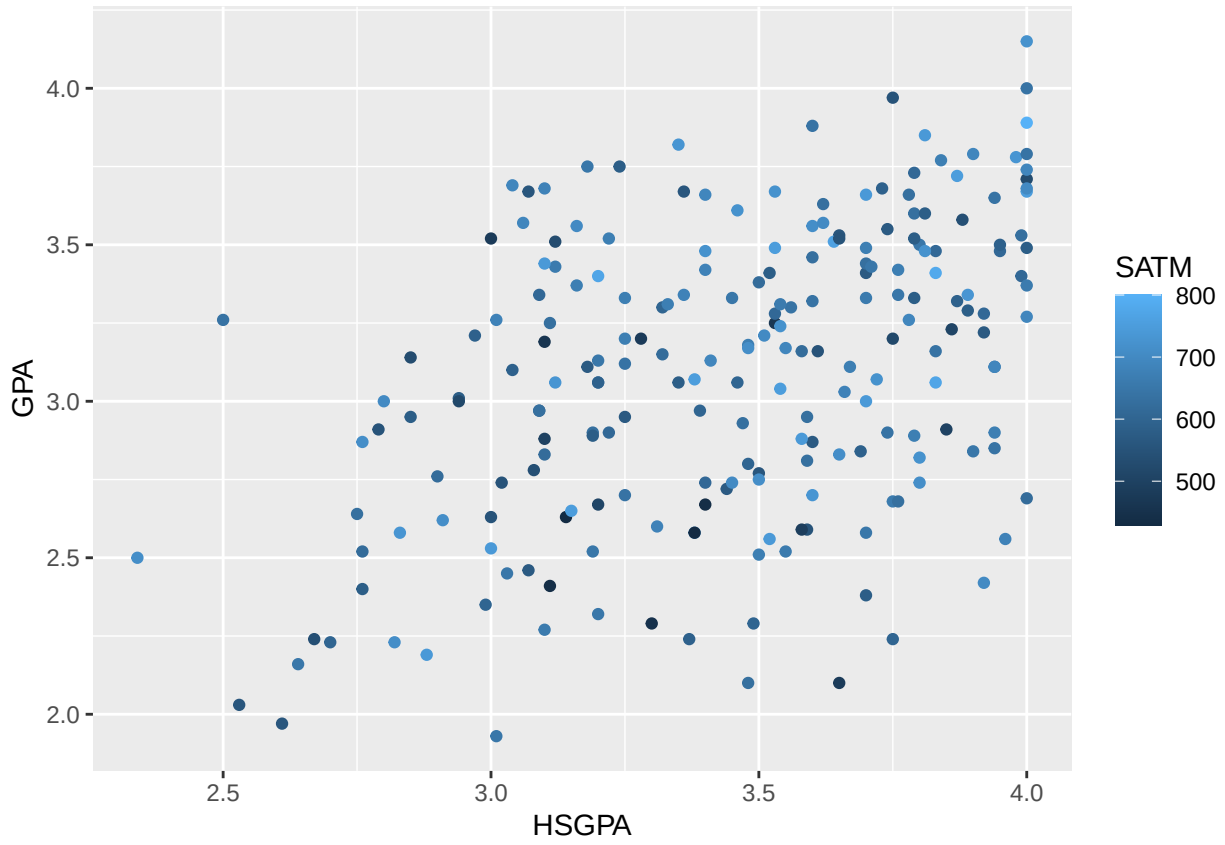
```
##        (Intercept) FirstYearGPA$HSGPA
##          1.1798507          0.5550125
```



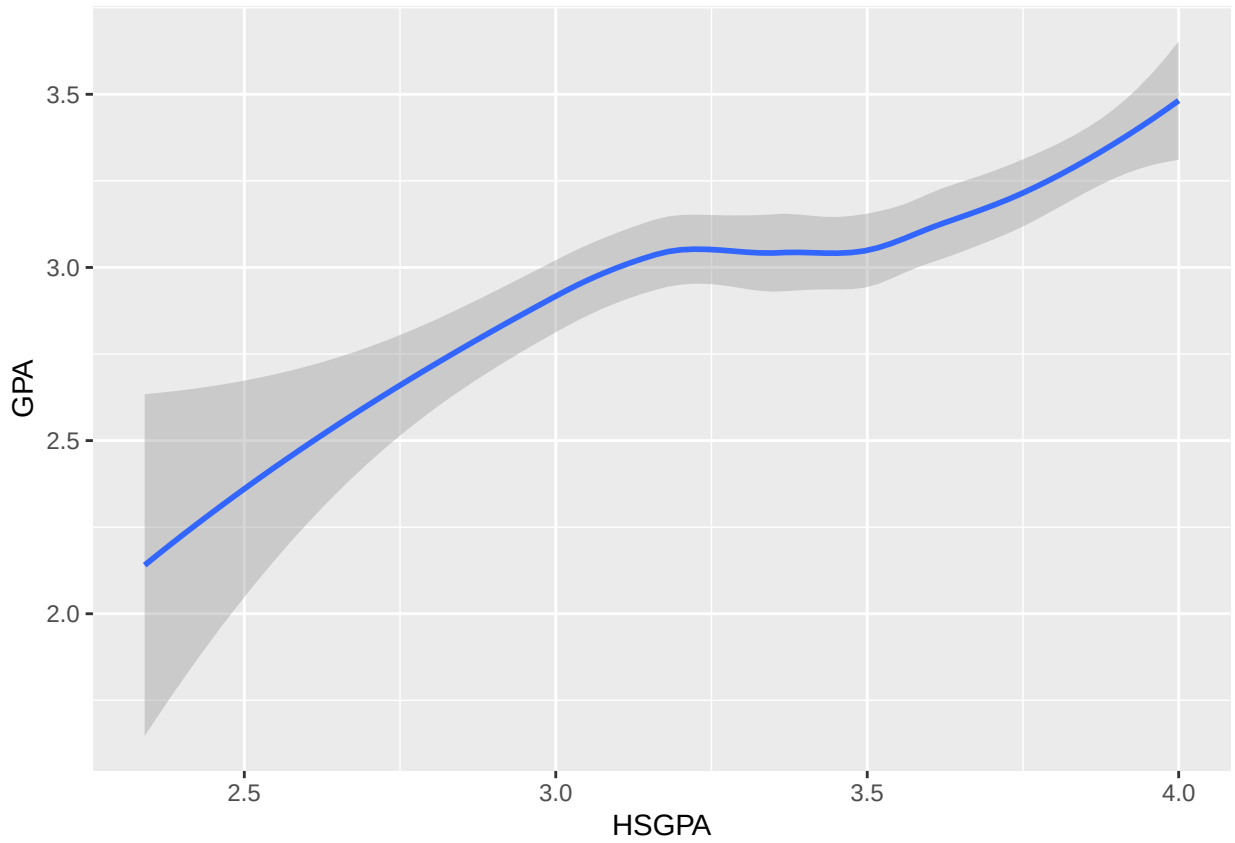

GPA
4.0
3.5
3.0
2.5
2.0
2.5
3.0
3.5
4.0
HSGPA


GPA
4.0
3.5
3.0
2.5
2.0
2.5
3.0
3.5
4.0
HSGPA

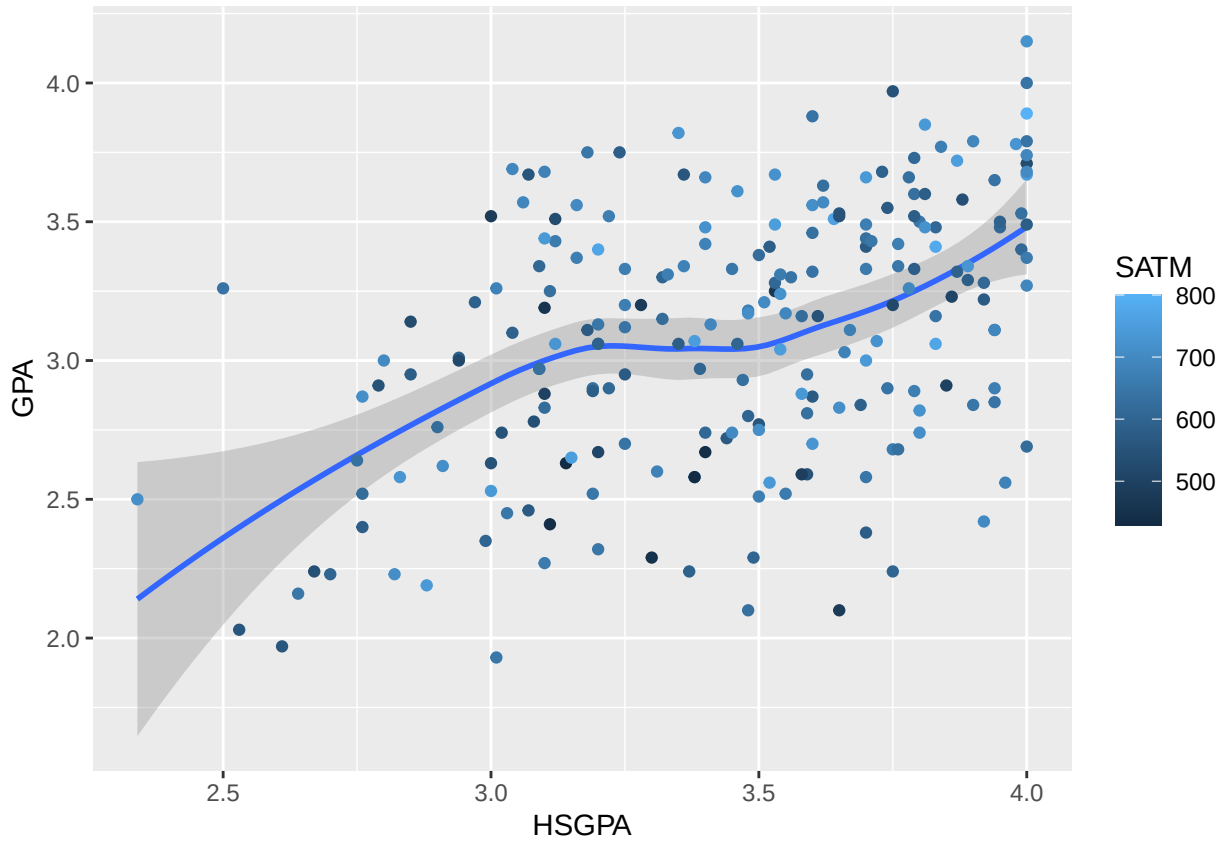

GPA
4.0
3.5
3.0
2.5
2.0
2.5
3.0
3.5
4.0
HSGPA
CollegeBound
1.00
0.75
0.50
0.25
0.00



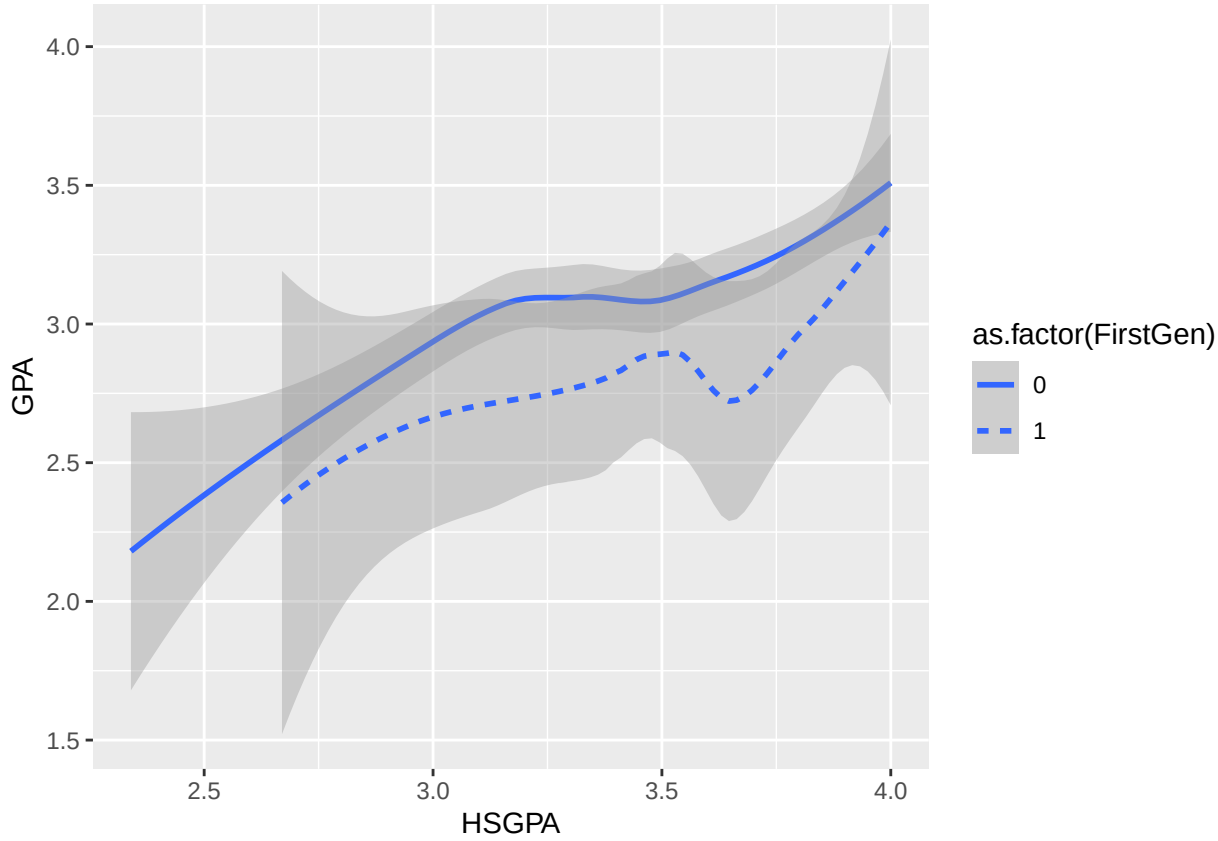

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



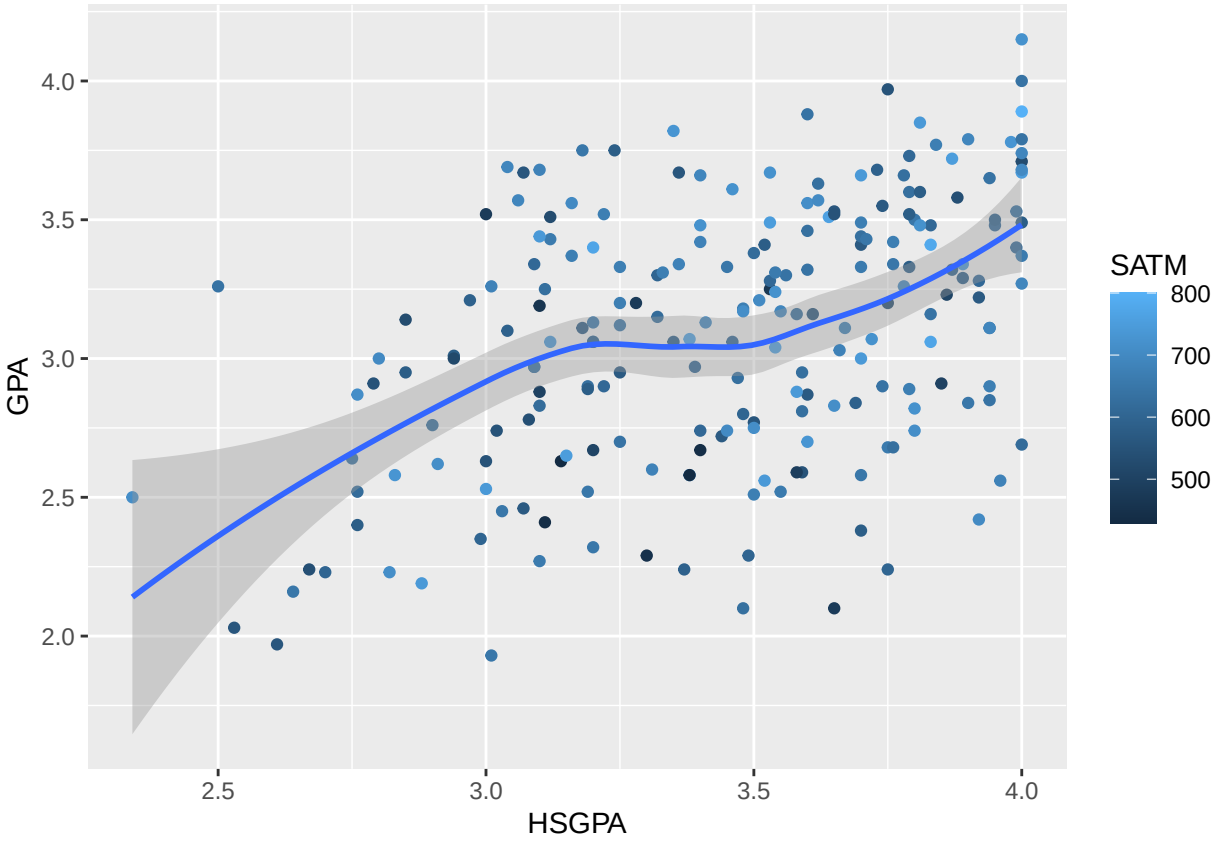

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

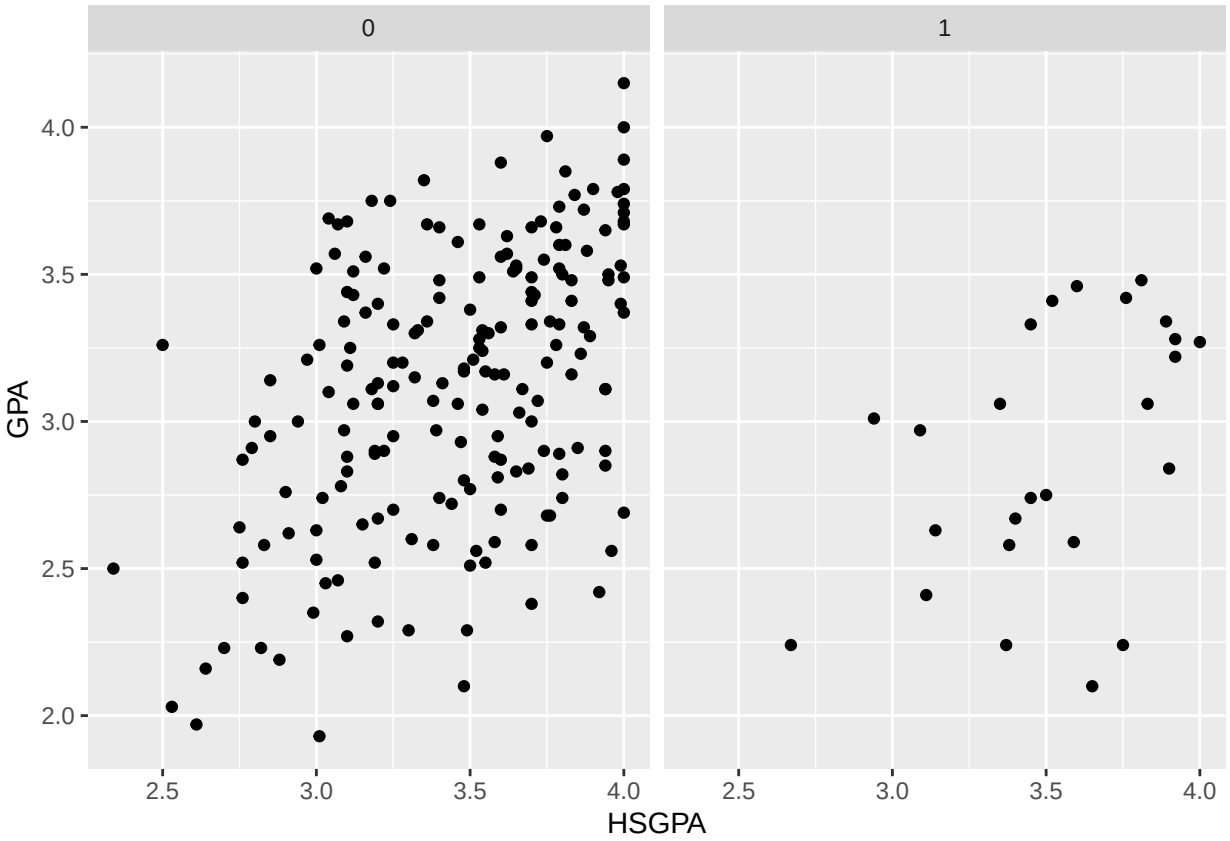


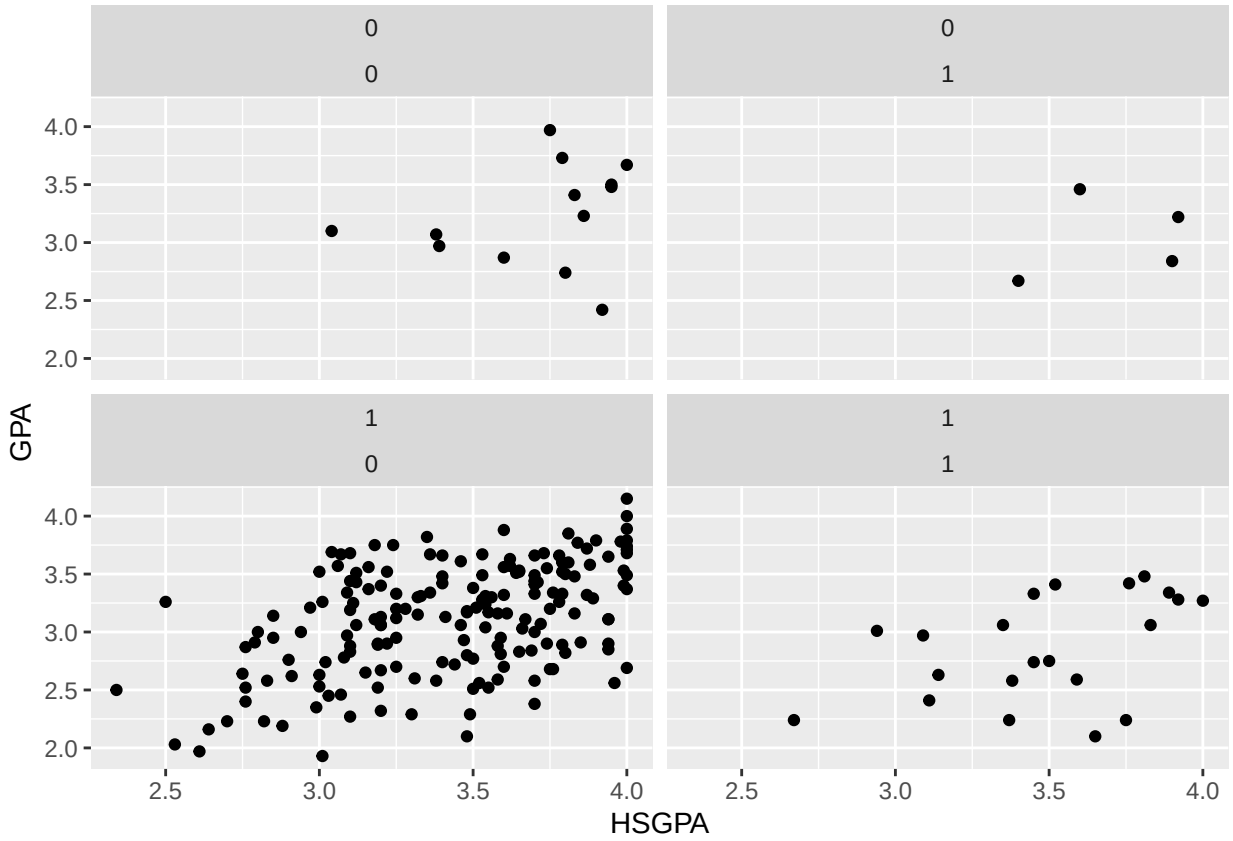

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

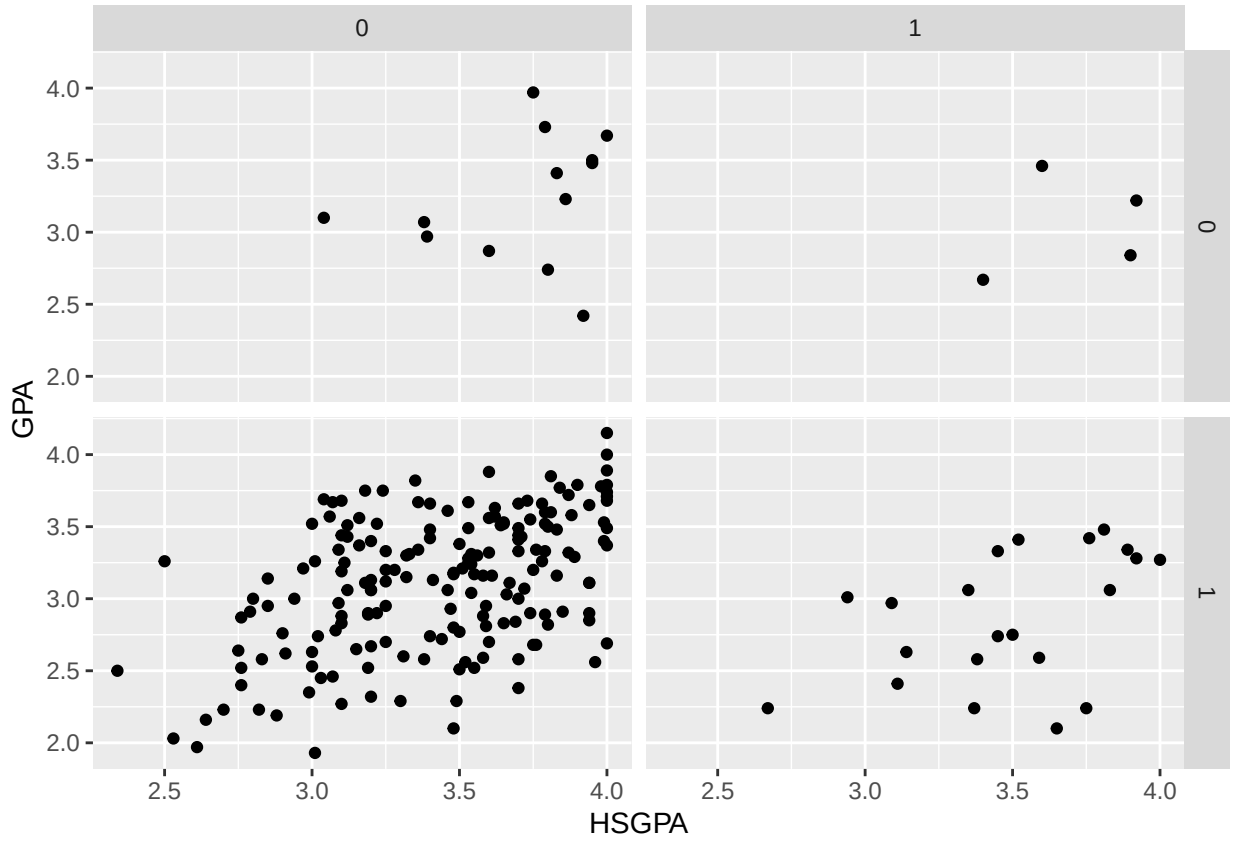



```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

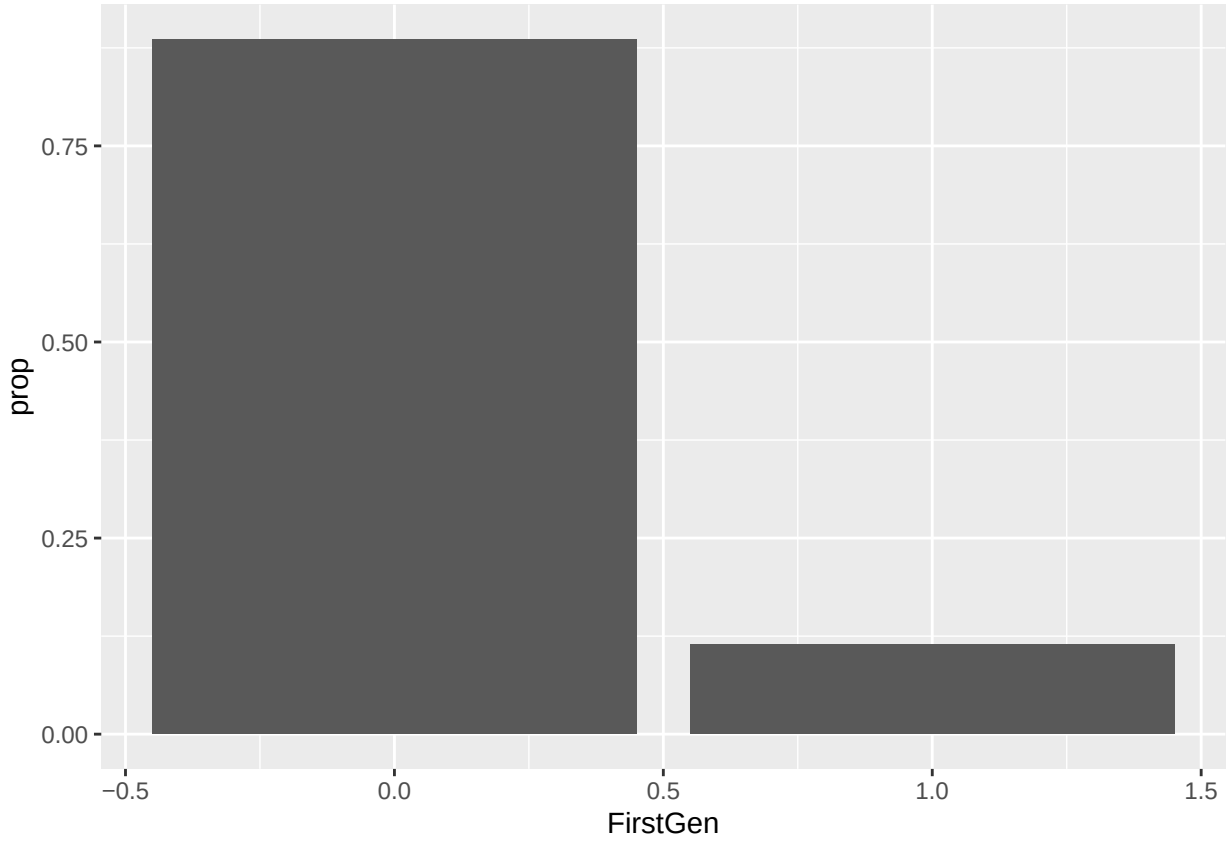

GPA
4.0
3.5
3.0
2.5
2.0
2.5
3.0
3.5
4.0
HSGPA
SATM
800
700
600
500


0
1
GPA
4.0
3.5
3.0
2.5
2.0
2.5
3.0
3.5
4.0
HSGPA


0
0
0
1
1
0
1
1
GPA
HSGPA
4.0
3.5
3.0
2.5
2.0
2.5
3.0
3.5
4.0




```
##
##  Welch Two Sample t-test
##
## data:  nonwhite$GPA and iswhite$GPA
## t = -3.8836, df = 62.441, p-value = 0.0002511
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4865808 -0.1559198
## sample estimates:
## mean of x mean of y
##  2.842391  3.163642
```



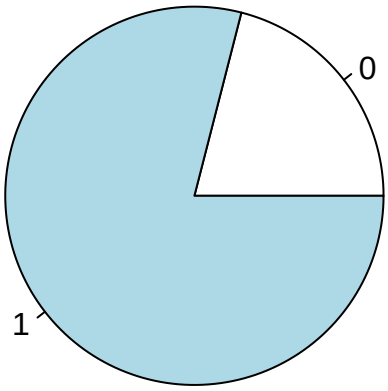

```
##
##       0   1
##   0  34 160
##   1  12  13
```

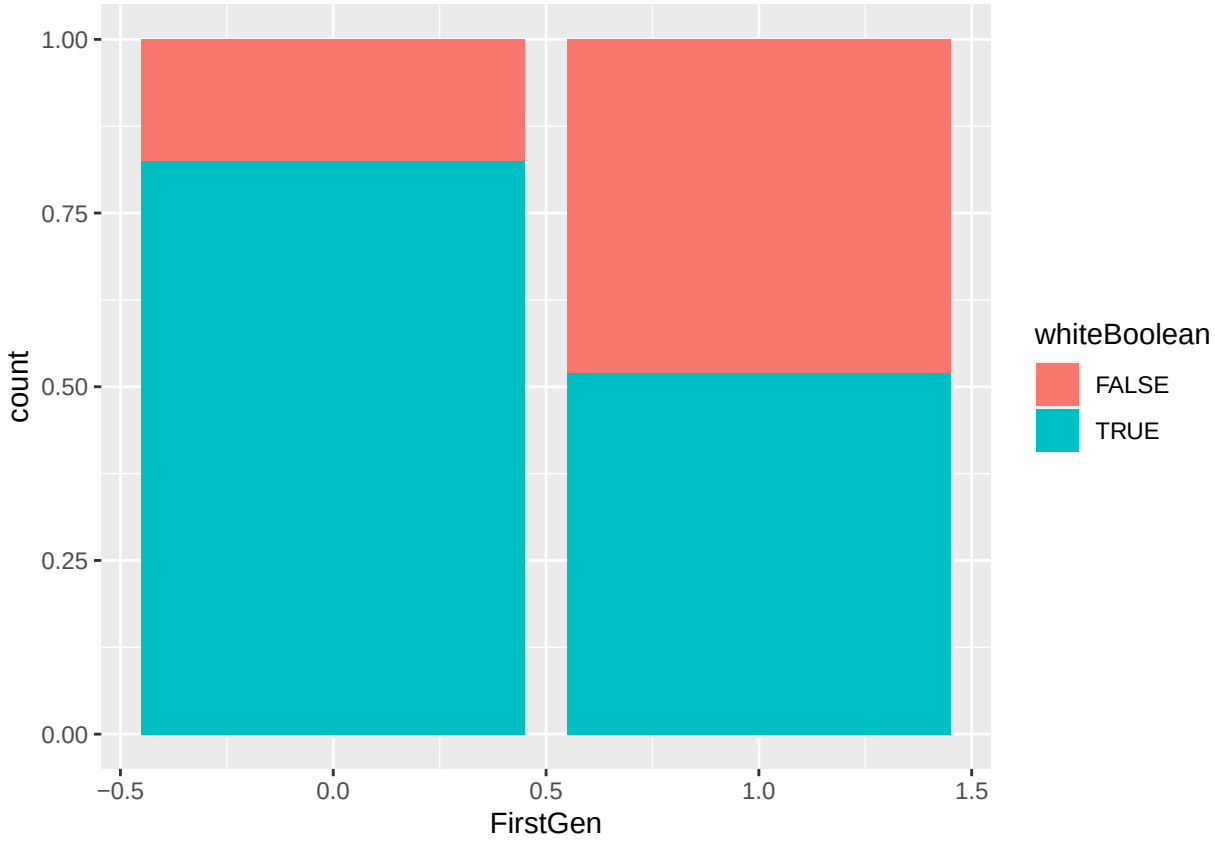
```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  newtable
## X-squared = 10.626, df = 1, p-value = 0.001115
```

```
##
##       0   1
##   0  34 160
##   1  12  13
```

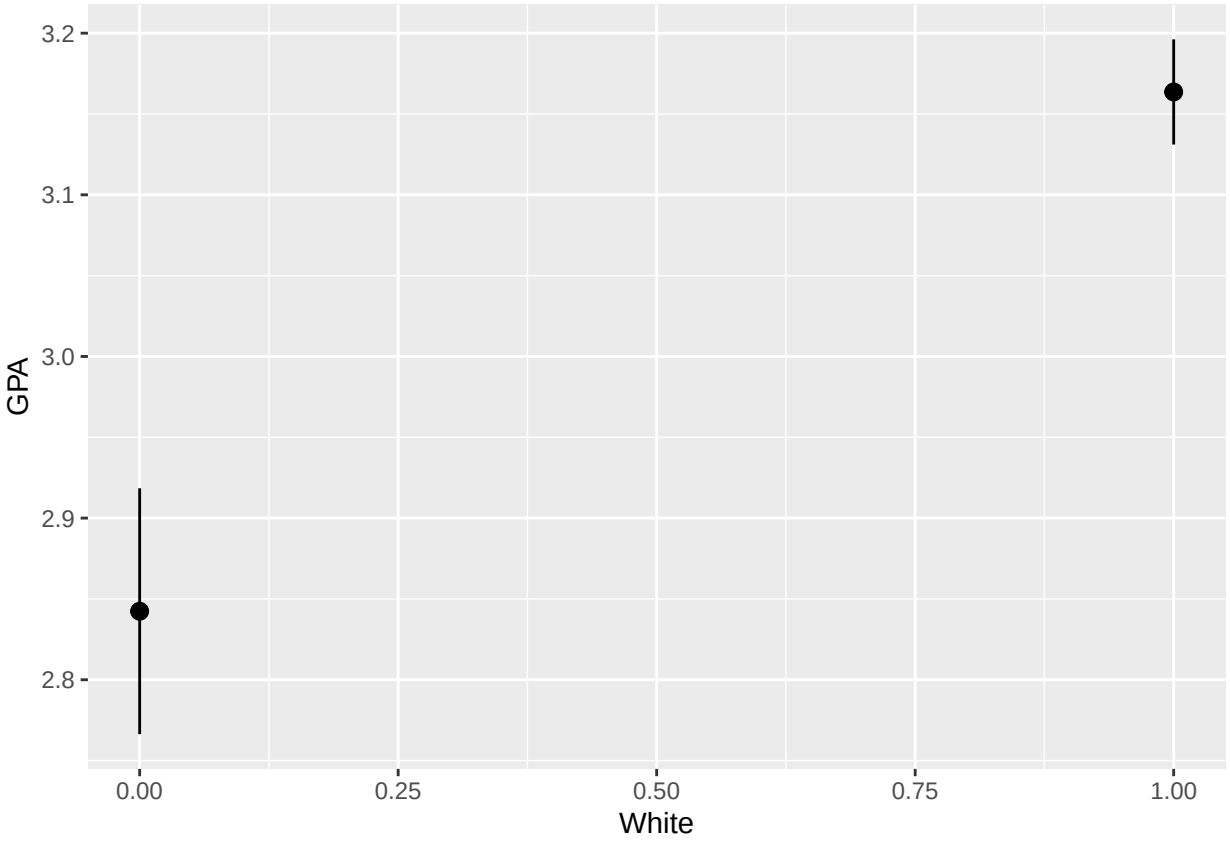
```
##
##             0         1
##   0 40.748858 153.25114
##   1  5.251142  19.74886
```

```
##
##         0      1
##   0 40.75 153.25
##   1  5.25  19.75
```


% of White Students
0
1




```
## No summary function supplied, defaulting to `mean_se()`
```


3.2
3.1
3.0
2.9
2.8
GPA
0.00
0.25
0.50
0.75
1.00
White