

# 04School District Spending in Texas Schools

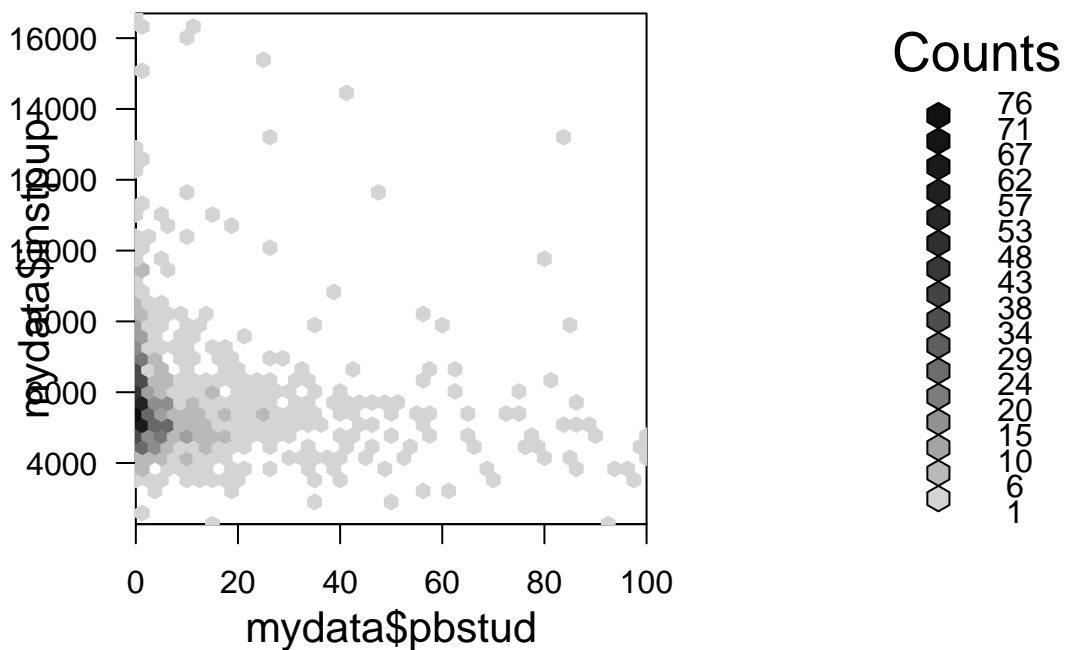
Adam Billen

1/11/2021

```
library(ggforce)
library(haven)
library(ggplot2)
library(dplyr)
library(hexbin)
library(RColorBrewer)
library(tidyverse)
```

```
mydata <- read_dta("\\Users\\adamb\\Documents\\SpendingSD\\finalproject\\code\\school_district (1).dta")

# hexbin to look into dense cluster
a=hexbin(mydata$pbstud,mydata$instpup,xbins=40)
plot(a)
```

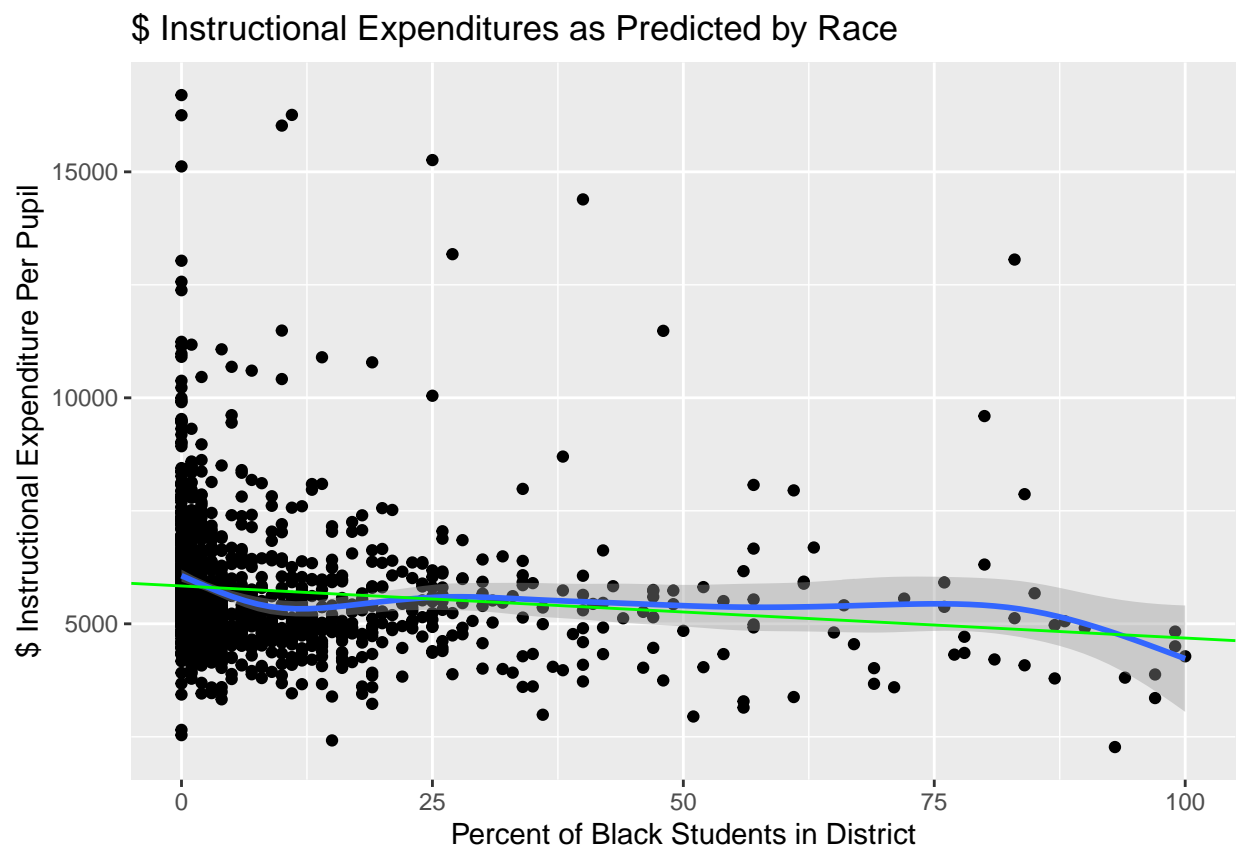


```
ggplot(data = mydata) +
  geom_point(mapping = aes(x = pbstud, y = instpup),
             show.legend = FALSE) +
  geom_smooth(mapping = aes(x=pbstud, y = instpup)) +
  geom_abline(intercept = 5837.26, slope = -11.52, color = "green") +
  geom_abline(intercept = , slope = ) +
  ggtitle("$ Instructional Expenditures as Predicted by Race") +
  xlab("Percent of Black Students in District") +
  ylab("$ Instructional Expenditure Per Pupil")
```

```
## 'geom_smooth()' using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 89 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 89 rows containing missing values (geom_point).
```



```
# Instpup predicted by percent of black students (zoomed in)
ggplot(data = mydata) +
  geom_point(mapping = aes(x = pbstud, y = instpup),
             show.legend = FALSE) +
  geom_smooth(mapping = aes(x=pbstud, y = instpup)) +
  geom_abline(intercept = 5837.26, slope = -11.52, color = "green") +
  geom_abline(intercept = , slope = ) +
  ggtitle("Zoomed In - $ Instructional Expenditures as Predicted by Race") +
```

```

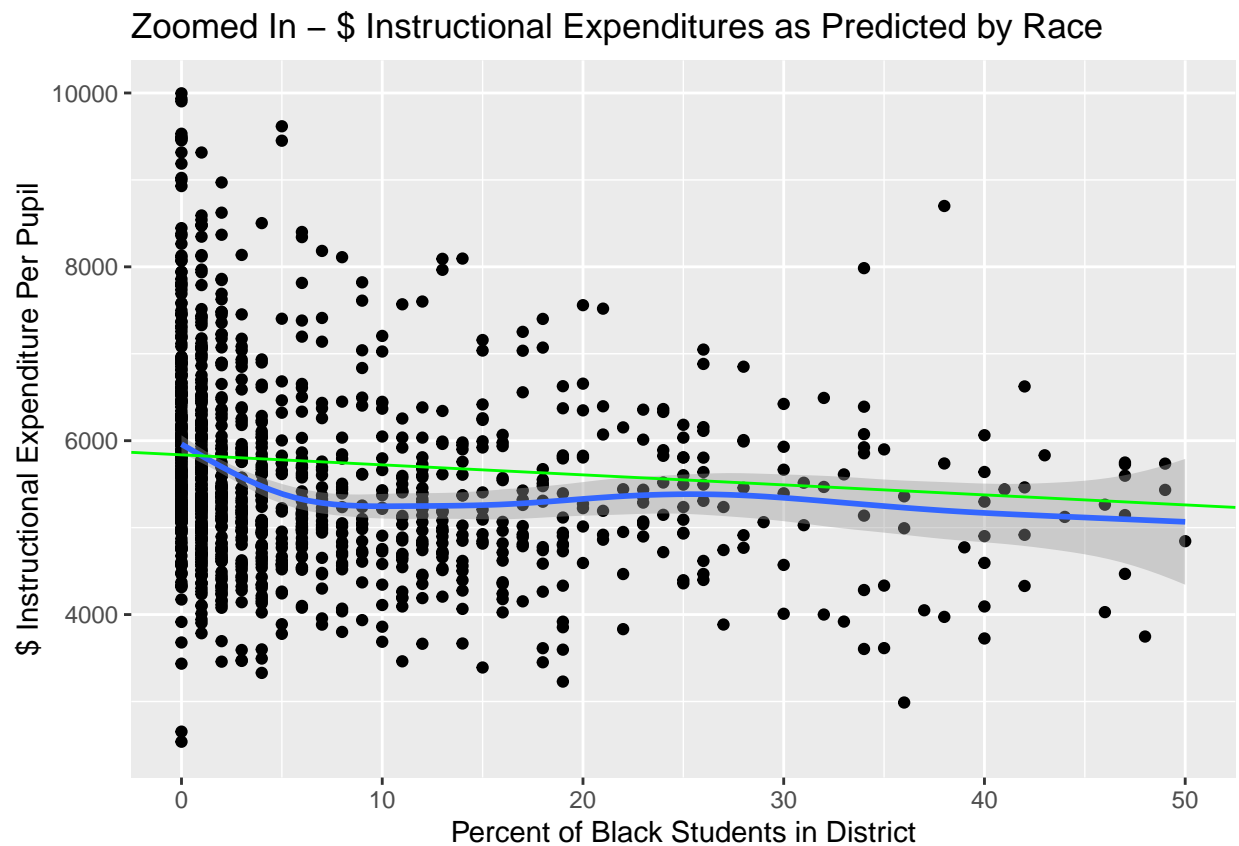
xlab("Percent of Black Students in District") +
ylab("$ Instructional Expenditure Per Pupil") +
xlim(0, 50) +
ylim(2500,10000)

```

```
## 'geom_smooth()' using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 167 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 167 rows containing missing values (geom_point).
```



```

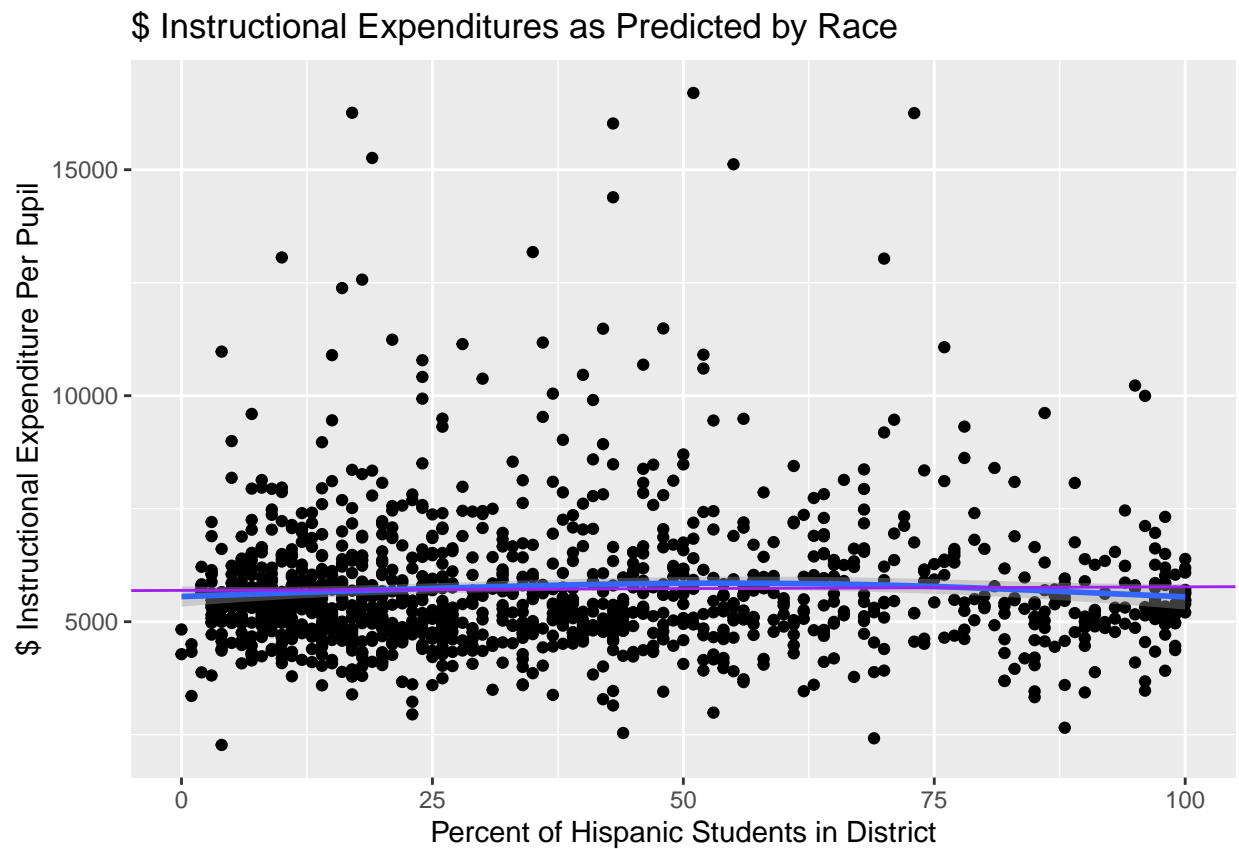
ggplot(data = mydata) +
  geom_point(mapping = aes(x = phstud, y = instpup)) +
  geom_smooth(mapping = aes(x=phstud, y = instpup)) +
  geom_abline(intercept = 5692.4232, slope = .7794, color = "purple") +
  ggtitle("$ Instructional Expenditures as Predicted by Race") +
  xlab("Percent of Hispanic Students in District") +
  ylab("$ Instructional Expenditure Per Pupil")

```

```
## 'geom_smooth()' using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 89 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 89 rows containing missing values (geom_point).
```



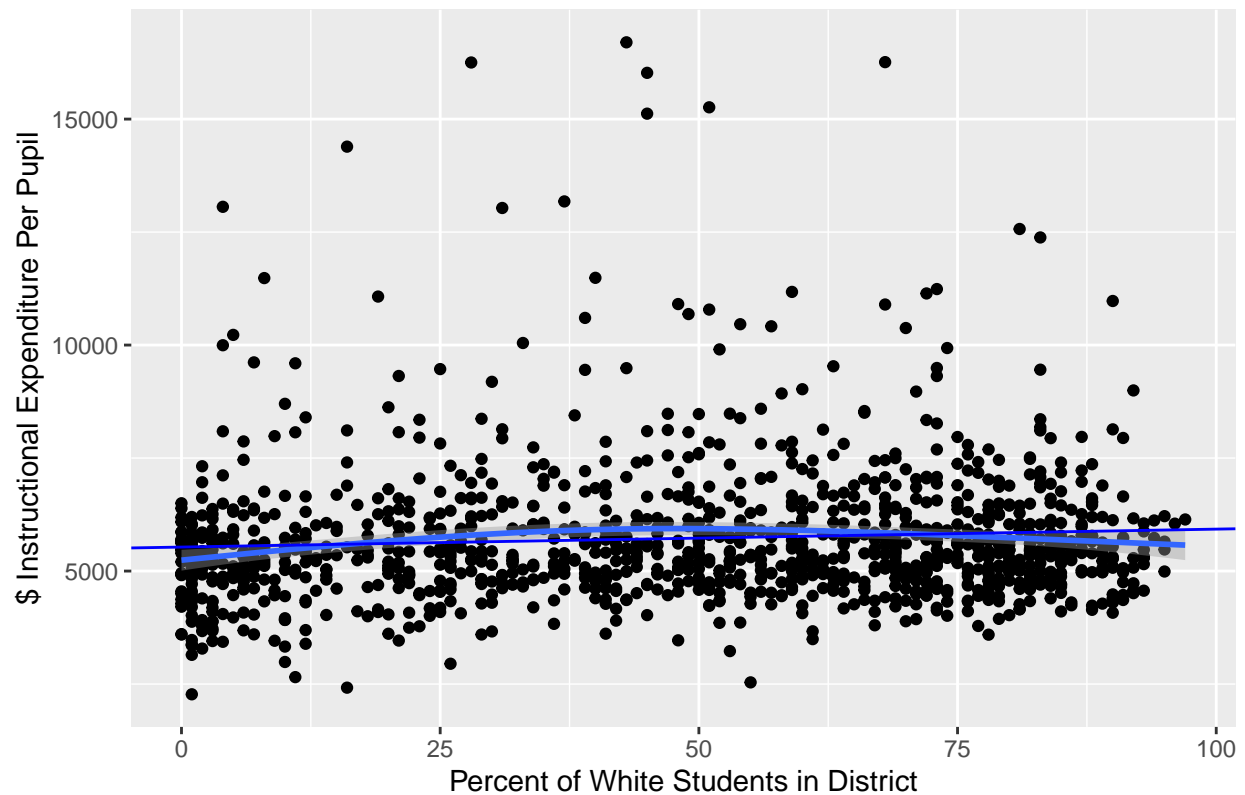
```
ggplot(data = mydata) +  
  geom_point(mapping = aes(x = pwstud, y = instpup)) +  
  geom_smooth(mapping = aes(x=pwstud, y = instpup)) +  
  geom_abline(intercept = 5530.489, slope = 3.969, color = "blue") +  
  ggtitle("$ Instructional Expenditures as Predicted by Race") +  
  xlab("Percent of White Students in District") +  
  ylab("$ Instructional Expenditure Per Pupil")
```

```
## 'geom_smooth()' using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 89 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 89 rows containing missing values (geom_point).
```

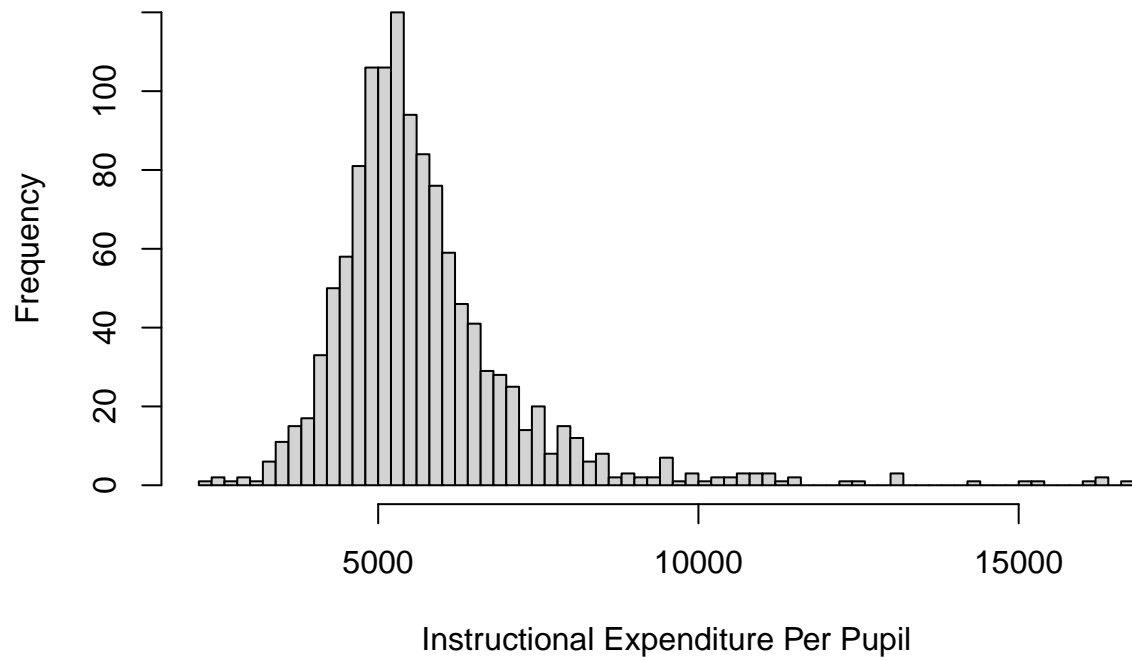
### \$ Instructional Expenditures as Predicted by Race



```
# pbstud: (a=5837.26, b=-11.52) numbers for pbstud as a result of lm(instpup ~ variable, data = mydata)
# phstud: (a=5692.4232, b=.7794)
# pwstud: (a=5530.489, b=3.969)

#histograms, log vs breaks = 100
hist(mydata$instpup,breaks = 100,
     main = "Histogram for Instructional Expenditure Per Pupil",
     xlab = "Instructional Expenditure Per Pupil")
```

## Histogram for Instructional Expenditure Per Pupil

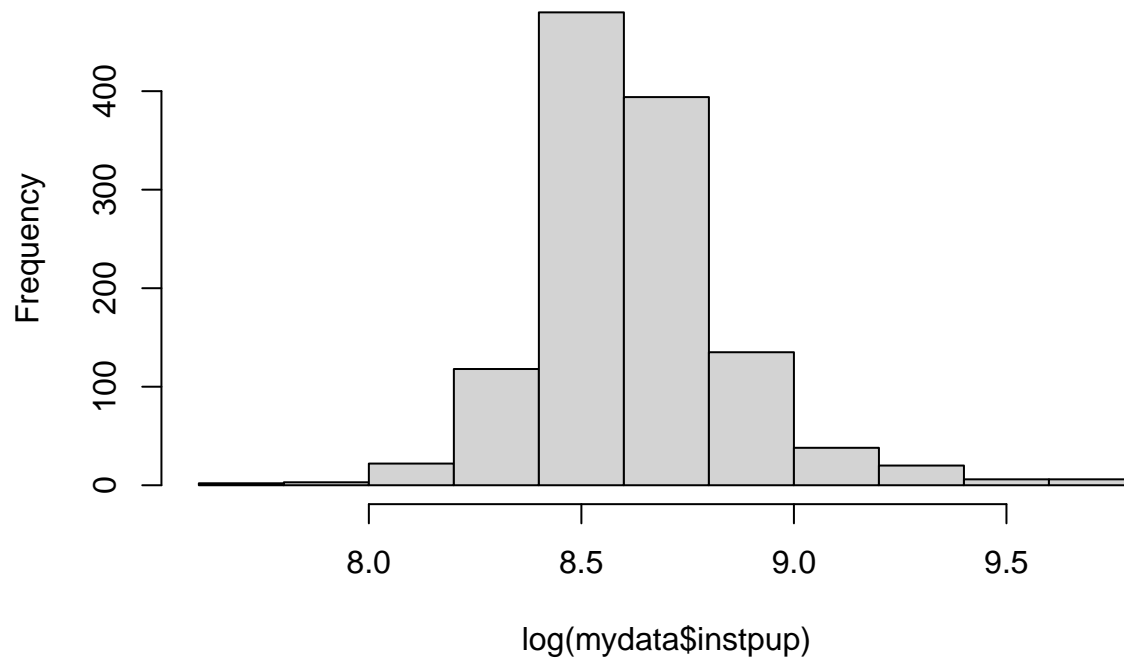


```
"Instructional Expenditure Per Pupil"
```

```
## [1] "Instructional Expenditure Per Pupil"
```

```
hist(log(mydata$instpup))
```

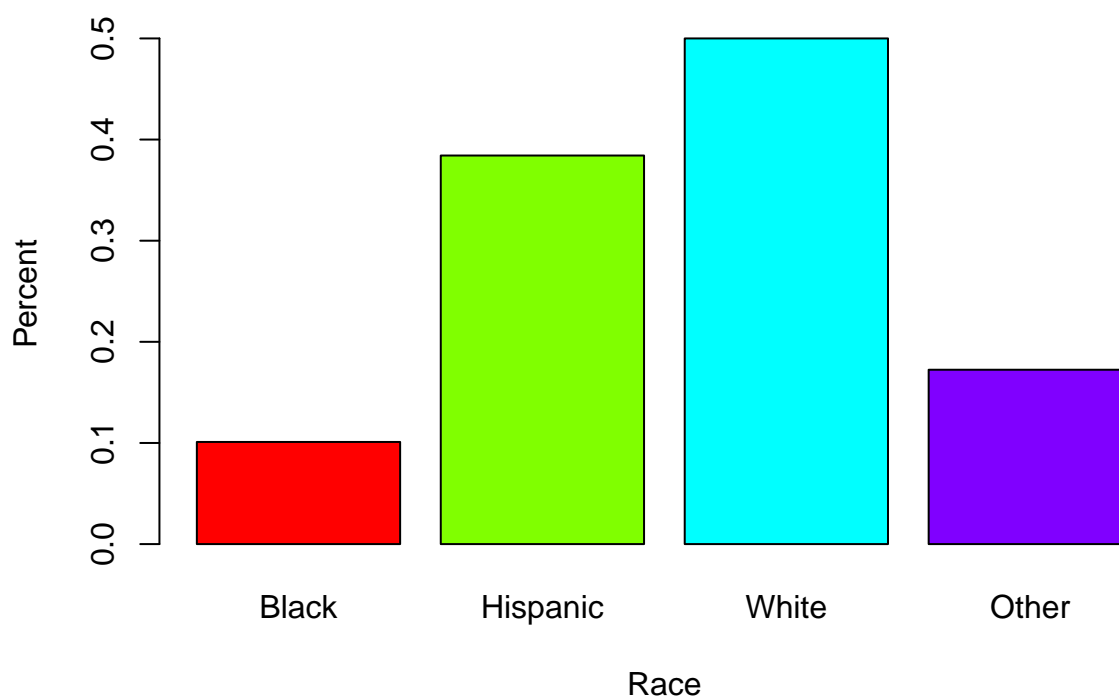
**Histogram of log(mydata\$instpup)**



```
# Mean Percent of Studets by Race
H <- c(.101, .3842, .5, .1724)
M <- c("Black", "Hispanic", "White", "Other")

barplot(H,names.arg=M,
        xlab="Race",
        ylab="Percent",
        main="Mean Proportion of Students by Race",
        col = rainbow(4))
```

## Mean Proportion of Students by Race



*#summary statistics / descriptive statistics - add in regression numbers here*

*# PBSTUD summary*

```
summary(mydata$pbstud)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##      0.0      1.0      3.0    10.1    12.0    100.0      85
```

```
#  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
#   0.0      1.0      3.0    10.1    12.0    100.0      85
```

*# instpup predicted by pbstud regression*

```
lm(formula = pbstud ~ instpup, data = mydata)
```

```
##
```

```
## Call:
```

```
## lm(formula = pbstud ~ instpup, data = mydata)
```

```
##
```

```
## Coefficients:
```

```
## (Intercept)      instpup
```

```
##   17.325738   -0.001285
```

```
mod <- lm(pbstud ~ instpup, data = mydata)
```

```
summary(mod)
```



```
##
## Call:
## lm(formula = pbstud ~ instpup, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.066  -9.032  -6.304   1.770  88.175
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.3257384  1.7781305   9.744 < 2e-16 ***
## instpup      -0.0012849  0.0002998  -4.285 1.97e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.32 on 1222 degrees of freedom
## (89 observations deleted due to missingness)
## Multiple R-squared:  0.0148, Adjusted R-squared:  0.014
## F-statistic: 18.36 on 1 and 1222 DF,  p-value: 1.969e-05
```

```
#Residuals:
#      Min       1Q   Median       3Q      Max
# -14.066  -9.032  -6.304   1.770  88.175
#Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept) 17.3257384  1.7781305   9.744 < 2e-16 ***
# instpup      -0.0012849  0.0002998  -4.285 1.97e-05 ***
```

```
# PHSTUD summary
summary(mydata$phstud)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      0.00  15.00   31.50   38.42  56.00  100.00    85
```

```
#      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
#      0.00  15.00   31.50   38.42  56.00  100.00    85
```

```
# instpup predicted by phstud summary
lm(formula = phstud ~ instpup, data = mydata)
```

```
##
## Call:
## lm(formula = phstud ~ instpup, data = mydata)
##
## Coefficients:
## (Intercept)      instpup
##  3.697e+01    2.488e-04
```

```
mod <- lm(phstud ~ instpup, data = mydata)
summary(mod)
```

```
##
## Call:
## lm(formula = phstud ~ instpup, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -38.169 -23.191  -6.769  17.666  61.738
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.697e+01  3.031e+00  12.197  <2e-16 ***
## instpup      2.488e-04  5.111e-04   0.487    0.626
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.83 on 1222 degrees of freedom
## (89 observations deleted due to missingness)
## Multiple R-squared:  0.000194, Adjusted R-squared: -0.0006242
## F-statistic: 0.2371 on 1 and 1222 DF, p-value: 0.6264
```

```
# Residuals:
#      Min       1Q   Median       3Q      Max
# -38.169 -23.191  -6.769  17.666  61.738
# Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept) 3.697e+01  3.031e+00  12.197  <2e-16 ***
# instpup      2.488e-04  5.111e-04   0.487    0.626
```

```
# PWSTUD summary
summary(mydata$pWstud)
```

```
## Warning: Unknown or uninitialised column: 'pWstud'.
```

```
## Length Class Mode
##      0    NULL  NULL
```

```
#  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
#  0.00  23.75   51.00   48.19  73.25   97.00     85
```

```
#instpup predicted by pwstud summary
lm(formula = pwstud ~ instpup, data = mydata)
```

```
##
## Call:
## lm(formula = pwstud ~ instpup, data = mydata)
##
## Coefficients:
## (Intercept)      instpup
##  40.581871      0.001355
```

```
mod <- lm(pwstud ~ instpup, data = mydata)
summary(mod)
```

```
##
## Call:
## lm(formula = pwstud ~ instpup, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -54.273 -24.332   3.242  25.396  48.099
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.058e+01  3.126e+00  12.98  <2e-16 ***
## instpup      1.355e-03  5.271e-04   2.57  0.0103 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 28.69 on 1222 degrees of freedom
## (89 observations deleted due to missingness)
## Multiple R-squared:  0.005378, Adjusted R-squared:  0.004564
## F-statistic: 6.607 on 1 and 1222 DF, p-value: 0.01028
```

```
# Residuals:
#      Min       1Q   Median       3Q      Max
# -54.273 -24.332   3.242  25.396  48.099
# Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept) 4.058e+01  3.126e+00  12.98  <2e-16 ***
# instpup      1.355e-03  5.271e-04   2.57  0.0103 *
```

```
# INSTPUP summary
summary(mydata$instpup)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##      2273   4860   5405    5722   6182   16698        89
```

```
#      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
#      2273   4860   5405    5722   6182   16698        89
```

```
# Controlled Regressions
lm(formula = instpup ~ pbstud + phstud + pwstud + enroll + stear + lowinc, data = mydata)
```

```
##
## Call:
## lm(formula = instpup ~ pbstud + phstud + pwstud + enroll + stear +
##      lowinc, data = mydata)
##
## Coefficients:
```

```
## (Intercept)      pbstud      phstud      pwstud      enroll      stear
##  9.579e+03    4.160e+00    7.242e+00   -8.608e-01    1.832e-03   -3.427e+02
##      lowinc
##  3.180e+00
```

```
mod <- lm(instpup ~ pbstud + phstud + pwstud + enroll + stear + lowinc, data = mydata)
summary(mod)
```

```
##
## Call:
## lm(formula = instpup ~ pbstud + phstud + pwstud + enroll + stear +
##      lowinc, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3781.6  -577.8  -200.6   256.6  8287.0
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.579e+03  8.467e+02  11.314  <2e-16 ***
## pbstud       4.160e+00  9.331e+00   0.446   0.656
## phstud       7.242e+00  8.724e+00   0.830   0.407
## pwstud      -8.608e-01  8.723e+00  -0.099   0.921
## enroll       1.832e-03  2.958e-03   0.619   0.536
## stear       -3.427e+02  1.246e+01 -27.504  <2e-16 ***
## lowinc       3.180e+00  2.427e+00   1.310   0.190
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1190 on 1216 degrees of freedom
## (90 observations deleted due to missingness)
## Multiple R-squared:  0.4191, Adjusted R-squared:  0.4162
## F-statistic: 146.2 on 6 and 1216 DF, p-value: < 2.2e-16
```

```
# Call:
# lm(formula = instpup ~ pbstud + phstud + pwstud + enroll + stear +
#      lowinc, data = mydata)

# Coefficients:
# (Intercept)      pbstud      phstud      pwstud      enroll      stear      lowinc
#  9.579e+03    4.160e+00    7.242e+00   -8.608e-01    1.832e-03   -3.427e+02    3.180e+00

# Call:
# lm(formula = instpup ~ pbstud + phstud + pwstud + enroll + stear +
#      lowinc, data = mydata)

# Residuals:
#      Min       1Q   Median       3Q      Max
# -3781.6  -577.8  -200.6   256.6  8287.0

# Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept)  9.579e+03  8.467e+02  11.314  <2e-16 ***
```

```

# pbstud      4.160e+00  9.331e+00  0.446  0.656
# phstud      7.242e+00  8.724e+00  0.830  0.407
# pwstud     -8.608e-01  8.723e+00 -0.099  0.921
# enroll      1.832e-03  2.958e-03  0.619  0.536
# stear      -3.427e+02  1.246e+01 -27.504  <2e-16 ***
# lowinc      3.180e+00  2.427e+00  1.310  0.190
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Residual standard error: 1190 on 1216 degrees of freedom
# (90 observations deleted due to missingness)
# Multiple R-squared:  0.4191, Adjusted R-squared:  0.4162
# F-statistic: 146.2 on 6 and 1216 DF, p-value: < 2.2e-16

```

```
hist(log(mydata$instpup))
```

```
#summary of pother summary(mydata$pother) Min. 1st Qu. Median Mean 3rd Qu. Max. NA's 1.00 15.00
18.00 17.24 20.00 34.00 88
```

```
#testing - figuring out logs, not working yet ggplot(data = mydata) + geom_point(mapping = aes(x =
pbstud, y = instpup)) + lm(formula = pbstud ~ instpup, data = mydata) mod <- lm(pbstud ~ instpup,
data = mydata) summary(mod)
```

```
geom_smooth(mapping = aes(x=pbstud, y = instpup)) + geom_abline(intercept = 5837.26, slope = -11.52,
color = "green") + scale_y_log10()
```

```
#size plot ggplot(data = mydata) + geom_point(mapping = aes(x = enroll, y = instpup), show.legend =
FALSE) + geom_smooth(mapping = aes(x=enroll, y = instpup))
```

## histogram

```
hist(mydata$instpup,breaks = 100)hist(log(mydata$instpup))
```

```
geom_histogram(mydata$instpup) + scale_x_log10()
```

```
hist(log(mydata$pbstud))
```

```
hist(log(mydata$phstud))
```

```
hist(log(mydata$pwstud))
```

```
loghist(mydata$instpup)
```

```
hist(mydata$instpup,freq = T)
```

```
hist(mydata$pbstud,breaks = 50) +
```

```
hist(mydata$pbstud,freq = T)
```

```
hist(mydata$phstud,breaks = 50)hist(mydata$phstud,freq = T)
```

```
hist(mydata$pwstud,breaks = 50)hist(mydata$pwstud,freq = T)
```