model-building

Sarah McDonald

February 1, 2020

data <- read.csv("C:/Users/Sarah McDonald/Documents/honorsthesis/manipulated\_data/cath\_inst.csv")  
  
#remove 1990 and 1991  
data <- data[data$YEAR!= 1990 & data$YEAR != 1991 & data$YEAR != 1992, ]

# Bishop Accountability

### TOTAL ENROLLMENT

enroll\_total <- lm(EFTOTLT ~ bishop\_pub\_sum\_3 + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Total", ])   
  
summary(enroll\_total)

##   
## Call:  
## lm(formula = EFTOTLT ~ bishop\_pub\_sum\_3 + Percent.Catholic +   
## his\_prop + pop\_dens + income\_per\_capita, data = data[data$LINE ==   
## "Total", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10137.7 -1880.7 -718.2 595.5 20954.5   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.620e+03 1.545e+02 10.488 < 2e-16 \*\*\*  
## bishop\_pub\_sum\_3 7.174e+00 3.436e+00 2.088 0.0369 \*   
## Percent.Catholic 4.076e+00 3.727e+00 1.094 0.2742   
## his\_prop 2.175e+03 3.803e+02 5.719 1.13e-08 \*\*\*  
## pop\_dens 6.178e+05 2.831e+04 21.822 < 2e-16 \*\*\*  
## income\_per\_capita 1.663e-02 3.242e-03 5.129 3.01e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3328 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.1385, Adjusted R-squared: 0.1377   
## F-statistic: 167 on 5 and 5194 DF, p-value: < 2.2e-16

### GRADUATE ENROLLMENT

enroll\_grad <- lm(EFTOTLT ~ bishop\_pub\_sum\_3 + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Graduate", ])   
  
summary(enroll\_grad)

##   
## Call:  
## lm(formula = EFTOTLT ~ bishop\_pub\_sum\_3 + Percent.Catholic +   
## his\_prop + pop\_dens + income\_per\_capita, data = data[data$LINE ==   
## "Graduate", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4472.0 -638.8 -379.7 160.7 9288.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.696e+02 6.121e+01 4.404 1.08e-05 \*\*\*  
## bishop\_pub\_sum\_3 2.135e+00 1.362e+00 1.568 0.1169   
## Percent.Catholic -2.798e+00 1.477e+00 -1.895 0.0581 .   
## his\_prop 1.193e+03 1.507e+02 7.912 3.07e-15 \*\*\*  
## pop\_dens 2.593e+05 1.122e+04 23.112 < 2e-16 \*\*\*  
## income\_per\_capita 9.720e-03 1.285e-03 7.566 4.52e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1319 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.1594, Adjusted R-squared: 0.1586   
## F-statistic: 197 on 5 and 5194 DF, p-value: < 2.2e-16

### UNDERGRADUATE ENROLLMENT

enroll\_undergrad <- lm(EFTOTLT ~ bishop\_pub\_sum\_3 + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Undergraduates", ])   
  
summary(enroll\_undergrad)

##   
## Call:  
## lm(formula = EFTOTLT ~ bishop\_pub\_sum\_3 + Percent.Catholic +   
## his\_prop + pop\_dens + income\_per\_capita, data = data[data$LINE ==   
## "Undergraduates", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5717.5 -1339.9 -373.2 522.2 13491.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.350e+03 1.014e+02 13.322 < 2e-16 \*\*\*  
## bishop\_pub\_sum\_3 5.038e+00 2.255e+00 2.234 0.02550 \*   
## Percent.Catholic 6.874e+00 2.446e+00 2.811 0.00496 \*\*   
## his\_prop 9.825e+02 2.496e+02 3.936 8.38e-05 \*\*\*  
## pop\_dens 3.585e+05 1.858e+04 19.297 < 2e-16 \*\*\*  
## income\_per\_capita 6.909e-03 2.128e-03 3.248 0.00117 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2184 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.1101, Adjusted R-squared: 0.1092   
## F-statistic: 128.5 on 5 and 5194 DF, p-value: < 2.2e-16

### TOTAL ENROLLMENT - Square Root Transformation

enroll\_total <- lm(sqrt(EFTOTLT) ~ sqrt(bishop\_pub\_sum\_3) + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Undergraduates", ])   
  
summary(enroll\_total)

##   
## Call:  
## lm(formula = sqrt(EFTOTLT) ~ sqrt(bishop\_pub\_sum\_3) + Percent.Catholic +   
## his\_prop + pop\_dens + income\_per\_capita, data = data[data$LINE ==   
## "Undergraduates", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -48.131 -12.389 0.071 9.930 78.354   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.285e+01 9.319e-01 35.248 < 2e-16 \*\*\*  
## sqrt(bishop\_pub\_sum\_3) 7.444e-01 1.758e-01 4.235 2.33e-05 \*\*\*  
## Percent.Catholic 8.579e-02 2.283e-02 3.759 0.000173 \*\*\*  
## his\_prop 9.518e+00 2.302e+00 4.134 3.62e-05 \*\*\*  
## pop\_dens 2.990e+03 1.724e+02 17.349 < 2e-16 \*\*\*  
## income\_per\_capita 6.056e-05 1.982e-05 3.055 0.002264 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20.15 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.1075, Adjusted R-squared: 0.1066   
## F-statistic: 125.1 on 5 and 5194 DF, p-value: < 2.2e-16

### GRADUATE ENROLLMENT - Square Root Transformation

enroll\_grad <- lm(sqrt(EFTOTLT) ~ sqrt(bishop\_pub\_sum\_3) + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Graduate", ])   
  
summary(enroll\_grad)

##   
## Call:  
## lm(formula = sqrt(EFTOTLT) ~ sqrt(bishop\_pub\_sum\_3) + Percent.Catholic +   
## his\_prop + pop\_dens + income\_per\_capita, data = data[data$LINE ==   
## "Graduate", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -57.629 -13.884 -2.206 9.361 66.872   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.219e+01 8.546e-01 14.262 < 2e-16 \*\*\*  
## sqrt(bishop\_pub\_sum\_3) 8.110e-01 1.612e-01 5.032 5.02e-07 \*\*\*  
## Percent.Catholic -7.868e-02 2.093e-02 -3.759 0.000172 \*\*\*  
## his\_prop 2.067e+01 2.111e+00 9.789 < 2e-16 \*\*\*  
## pop\_dens 3.189e+03 1.581e+02 20.173 < 2e-16 \*\*\*  
## income\_per\_capita 1.732e-04 1.818e-05 9.527 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 18.48 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.1612, Adjusted R-squared: 0.1604   
## F-statistic: 199.7 on 5 and 5194 DF, p-value: < 2.2e-16

### UNDERGRADUATE ENROLLMENT - Square Root Transformation

enroll\_undergrad <- lm(sqrt(EFTOTLT) ~ sqrt(bishop\_pub\_sum\_3) + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Undergraduates", ])   
  
summary(enroll\_undergrad)

##   
## Call:  
## lm(formula = sqrt(EFTOTLT) ~ sqrt(bishop\_pub\_sum\_3) + Percent.Catholic +   
## his\_prop + pop\_dens + income\_per\_capita, data = data[data$LINE ==   
## "Undergraduates", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -48.131 -12.389 0.071 9.930 78.354   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.285e+01 9.319e-01 35.248 < 2e-16 \*\*\*  
## sqrt(bishop\_pub\_sum\_3) 7.444e-01 1.758e-01 4.235 2.33e-05 \*\*\*  
## Percent.Catholic 8.579e-02 2.283e-02 3.759 0.000173 \*\*\*  
## his\_prop 9.518e+00 2.302e+00 4.134 3.62e-05 \*\*\*  
## pop\_dens 2.990e+03 1.724e+02 17.349 < 2e-16 \*\*\*  
## income\_per\_capita 6.056e-05 1.982e-05 3.055 0.002264 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20.15 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.1075, Adjusted R-squared: 0.1066   
## F-statistic: 125.1 on 5 and 5194 DF, p-value: < 2.2e-16

### TOTAL ENROLLMENT - Log+1 Transformation

enroll\_total <- lm(log(EFTOTLT+1) ~ log(bishop\_pub\_sum\_3+1) + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Undergraduates", ])   
  
summary(enroll\_total)

##   
## Call:  
## lm(formula = log(EFTOTLT + 1) ~ log(bishop\_pub\_sum\_3 + 1) + Percent.Catholic +   
## his\_prop + pop\_dens + income\_per\_capita, data = data[data$LINE ==   
## "Undergraduates", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.7546 -0.3939 0.3001 0.7215 2.5422   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.638e+00 5.559e-02 119.421 < 2e-16 \*\*\*  
## log(bishop\_pub\_sum\_3 + 1) 7.429e-02 1.733e-02 4.285 1.86e-05 \*\*\*  
## Percent.Catholic 6.936e-03 1.363e-03 5.090 3.71e-07 \*\*\*  
## his\_prop 3.809e-01 1.372e-01 2.777 0.00551 \*\*   
## pop\_dens 1.342e+02 1.030e+01 13.029 < 2e-16 \*\*\*  
## income\_per\_capita 2.712e-06 1.185e-06 2.289 0.02213 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.201 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.07512, Adjusted R-squared: 0.07423   
## F-statistic: 84.37 on 5 and 5194 DF, p-value: < 2.2e-16

### GRADUATE ENROLLMENT - Log+1 Transformation

enroll\_grad <- lm(log(EFTOTLT+1) ~ log(bishop\_pub\_sum\_3+1) + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Graduate", ])   
  
summary(enroll\_grad)

##   
## Call:  
## lm(formula = log(EFTOTLT + 1) ~ log(bishop\_pub\_sum\_3 + 1) + Percent.Catholic +   
## his\_prop + pop\_dens + income\_per\_capita, data = data[data$LINE ==   
## "Graduate", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.6592 -0.8996 0.8979 1.8668 4.1249   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.847e+00 1.234e-01 31.177 < 2e-16 \*\*\*  
## log(bishop\_pub\_sum\_3 + 1) 2.501e-01 3.848e-02 6.498 8.90e-11 \*\*\*  
## Percent.Catholic -1.776e-02 3.025e-03 -5.870 4.62e-09 \*\*\*  
## his\_prop 2.136e+00 3.045e-01 7.014 2.61e-12 \*\*\*  
## pop\_dens 3.390e+02 2.287e+01 14.825 < 2e-16 \*\*\*  
## income\_per\_capita 2.200e-05 2.630e-06 8.364 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.666 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.1052, Adjusted R-squared: 0.1044   
## F-statistic: 122.1 on 5 and 5194 DF, p-value: < 2.2e-16

### UNDERGRADUATE ENROLLMENT - Log+1 Transformation

enroll\_undergrad <- lm(log(EFTOTLT+1) ~ log(bishop\_pub\_sum\_3+1) + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Undergraduates", ])   
  
summary(enroll\_undergrad)

##   
## Call:  
## lm(formula = log(EFTOTLT + 1) ~ log(bishop\_pub\_sum\_3 + 1) + Percent.Catholic +   
## his\_prop + pop\_dens + income\_per\_capita, data = data[data$LINE ==   
## "Undergraduates", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.7546 -0.3939 0.3001 0.7215 2.5422   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.638e+00 5.559e-02 119.421 < 2e-16 \*\*\*  
## log(bishop\_pub\_sum\_3 + 1) 7.429e-02 1.733e-02 4.285 1.86e-05 \*\*\*  
## Percent.Catholic 6.936e-03 1.363e-03 5.090 3.71e-07 \*\*\*  
## his\_prop 3.809e-01 1.372e-01 2.777 0.00551 \*\*   
## pop\_dens 1.342e+02 1.030e+01 13.029 < 2e-16 \*\*\*  
## income\_per\_capita 2.712e-06 1.185e-06 2.289 0.02213 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.201 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.07512, Adjusted R-squared: 0.07423   
## F-statistic: 84.37 on 5 and 5194 DF, p-value: < 2.2e-16

# Nexis-Lexus

### TOTAL ENROLLMENT

enroll\_total <- lm(EFTOTLT ~ nexis\_sum\_3 + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Total", ])   
  
summary(enroll\_total)

##   
## Call:  
## lm(formula = EFTOTLT ~ nexis\_sum\_3 + Percent.Catholic + his\_prop +   
## pop\_dens + income\_per\_capita, data = data[data$LINE == "Total",   
## ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10161 -1863 -713 588 20940   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.603e+03 1.551e+02 10.335 < 2e-16 \*\*\*  
## nexis\_sum\_3 8.607e-01 1.479e+00 0.582 0.561   
## Percent.Catholic 5.096e+00 3.696e+00 1.379 0.168   
## his\_prop 2.201e+03 3.816e+02 5.768 8.51e-09 \*\*\*  
## pop\_dens 6.192e+05 2.836e+04 21.839 < 2e-16 \*\*\*  
## income\_per\_capita 1.714e-02 3.258e-03 5.261 1.49e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3329 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.1378, Adjusted R-squared: 0.137   
## F-statistic: 166 on 5 and 5194 DF, p-value: < 2.2e-16

### GRADUATE ENROLLMENT

enroll\_grad <- lm(EFTOTLT ~ nexis\_sum\_3 + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Graduate", ])   
  
summary(enroll\_grad)

##   
## Call:  
## lm(formula = EFTOTLT ~ nexis\_sum\_3 + Percent.Catholic + his\_prop +   
## pop\_dens + income\_per\_capita, data = data[data$LINE == "Graduate",   
## ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4477.1 -637.5 -382.3 163.2 9269.7   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.688e+02 6.145e+01 4.375 1.24e-05 \*\*\*  
## nexis\_sum\_3 5.940e-01 5.859e-01 1.014 0.3107   
## Percent.Catholic -2.562e+00 1.464e+00 -1.750 0.0803 .   
## his\_prop 1.192e+03 1.512e+02 7.883 3.88e-15 \*\*\*  
## pop\_dens 2.593e+05 1.123e+04 23.082 < 2e-16 \*\*\*  
## income\_per\_capita 9.761e-03 1.291e-03 7.562 4.65e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1319 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.1592, Adjusted R-squared: 0.1584   
## F-statistic: 196.7 on 5 and 5194 DF, p-value: < 2.2e-16

### UNDERGRADUATE ENROLLMENT

enroll\_undergrad <- lm(EFTOTLT ~ nexis\_sum\_3 + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Undergraduates", ])   
  
summary(enroll\_undergrad)

##   
## Call:  
## lm(formula = EFTOTLT ~ nexis\_sum\_3 + Percent.Catholic + his\_prop +   
## pop\_dens + income\_per\_capita, data = data[data$LINE == "Undergraduates",   
## ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5726.4 -1347.4 -373.4 523.0 13485.7   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.334e+03 1.018e+02 13.107 < 2e-16 \*\*\*  
## nexis\_sum\_3 2.667e-01 9.705e-01 0.275 0.783512   
## Percent.Catholic 7.658e+00 2.426e+00 3.157 0.001602 \*\*   
## his\_prop 1.009e+03 2.505e+02 4.030 5.66e-05 \*\*\*  
## pop\_dens 3.599e+05 1.861e+04 19.342 < 2e-16 \*\*\*  
## income\_per\_capita 7.381e-03 2.138e-03 3.452 0.000561 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2185 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.1093, Adjusted R-squared: 0.1084   
## F-statistic: 127.4 on 5 and 5194 DF, p-value: < 2.2e-16

### TOTAL ENROLLMENT - Square Root Transformation

enroll\_total <- lm(sqrt(EFTOTLT) ~ sqrt(nexis\_sum\_3) + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Undergraduates", ])   
  
summary(enroll\_total)

##   
## Call:  
## lm(formula = sqrt(EFTOTLT) ~ sqrt(nexis\_sum\_3) + Percent.Catholic +   
## his\_prop + pop\_dens + income\_per\_capita, data = data[data$LINE ==   
## "Undergraduates", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -47.489 -12.433 0.078 9.911 78.336   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.336e+01 9.490e-01 35.154 < 2e-16 \*\*\*  
## sqrt(nexis\_sum\_3) 3.311e-01 1.205e-01 2.747 0.00604 \*\*   
## Percent.Catholic 9.947e-02 2.247e-02 4.428 9.71e-06 \*\*\*  
## his\_prop 9.479e+00 2.310e+00 4.103 4.14e-05 \*\*\*  
## pop\_dens 2.994e+03 1.738e+02 17.221 < 2e-16 \*\*\*  
## income\_per\_capita 5.732e-05 2.064e-05 2.777 0.00550 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20.17 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.1057, Adjusted R-squared: 0.1048   
## F-statistic: 122.8 on 5 and 5194 DF, p-value: < 2.2e-16

### GRADUATE ENROLLMENT - Square Root Transformation

enroll\_grad <- lm(sqrt(EFTOTLT) ~ sqrt(nexis\_sum\_3) + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Graduate", ])   
  
summary(enroll\_grad)

##   
## Call:  
## lm(formula = sqrt(EFTOTLT) ~ sqrt(nexis\_sum\_3) + Percent.Catholic +   
## his\_prop + pop\_dens + income\_per\_capita, data = data[data$LINE ==   
## "Graduate", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -58.064 -13.977 -2.173 9.443 66.309   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.259e+01 8.710e-01 14.456 < 2e-16 \*\*\*  
## sqrt(nexis\_sum\_3) 2.541e-01 1.106e-01 2.296 0.02170 \*   
## Percent.Catholic -6.155e-02 2.062e-02 -2.985 0.00285 \*\*   
## his\_prop 2.080e+01 2.120e+00 9.811 < 2e-16 \*\*\*  
## pop\_dens 3.217e+03 1.595e+02 20.166 < 2e-16 \*\*\*  
## income\_per\_capita 1.756e-04 1.894e-05 9.270 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 18.52 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.158, Adjusted R-squared: 0.1572   
## F-statistic: 194.9 on 5 and 5194 DF, p-value: < 2.2e-16

### UNDERGRADUATE ENROLLMENT - Square Root Transformation

enroll\_undergrad <- lm(sqrt(EFTOTLT) ~ sqrt(nexis\_sum\_3) + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Undergraduates", ])   
  
summary(enroll\_undergrad)

##   
## Call:  
## lm(formula = sqrt(EFTOTLT) ~ sqrt(nexis\_sum\_3) + Percent.Catholic +   
## his\_prop + pop\_dens + income\_per\_capita, data = data[data$LINE ==   
## "Undergraduates", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -47.489 -12.433 0.078 9.911 78.336   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.336e+01 9.490e-01 35.154 < 2e-16 \*\*\*  
## sqrt(nexis\_sum\_3) 3.311e-01 1.205e-01 2.747 0.00604 \*\*   
## Percent.Catholic 9.947e-02 2.247e-02 4.428 9.71e-06 \*\*\*  
## his\_prop 9.479e+00 2.310e+00 4.103 4.14e-05 \*\*\*  
## pop\_dens 2.994e+03 1.738e+02 17.221 < 2e-16 \*\*\*  
## income\_per\_capita 5.732e-05 2.064e-05 2.777 0.00550 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20.17 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.1057, Adjusted R-squared: 0.1048   
## F-statistic: 122.8 on 5 and 5194 DF, p-value: < 2.2e-16

### TOTAL ENROLLMENT - Log + 1 Transformation

enroll\_total <- lm(log(EFTOTLT+1) ~ log(nexis\_sum\_3+1) + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Undergraduates", ])   
  
summary(enroll\_total)

##   
## Call:  
## lm(formula = log(EFTOTLT + 1) ~ log(nexis\_sum\_3 + 1) + Percent.Catholic +   
## his\_prop + pop\_dens + income\_per\_capita, data = data[data$LINE ==   
## "Undergraduates", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.7496 -0.4004 0.2978 0.7127 2.5193   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.693e+00 5.655e-02 118.364 < 2e-16 \*\*\*  
## log(nexis\_sum\_3 + 1) 5.748e-02 1.402e-02 4.101 4.18e-05 \*\*\*  
## Percent.Catholic 7.648e-03 1.337e-03 5.723 1.11e-08 \*\*\*  
## his\_prop 3.697e-01 1.374e-01 2.691 0.00715 \*\*   
## pop\_dens 1.318e+02 1.041e+01 12.664 < 2e-16 \*\*\*  
## income\_per\_capita 1.706e-06 1.261e-06 1.353 0.17602   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.201 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.07485, Adjusted R-squared: 0.07396   
## F-statistic: 84.04 on 5 and 5194 DF, p-value: < 2.2e-16

### GRADUATE ENROLLMENT - Log + 1 Transformation

enroll\_grad <- lm(log(EFTOTLT+1) ~ log(nexis\_sum\_3+1) + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Graduate", ])   
  
summary(enroll\_grad)

##   
## Call:  
## lm(formula = log(EFTOTLT + 1) ~ log(nexis\_sum\_3 + 1) + Percent.Catholic +   
## his\_prop + pop\_dens + income\_per\_capita, data = data[data$LINE ==   
## "Graduate", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.5419 -0.8739 0.8975 1.8297 4.1861   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.941e+00 1.260e-01 31.287 < 2e-16 \*\*\*  
## log(nexis\_sum\_3 + 1) 7.286e-02 3.122e-02 2.334 0.0197 \*   
## Percent.Catholic -1.414e-02 2.977e-03 -4.748 2.11e-06 \*\*\*  
## his\_prop 2.187e+00 3.060e-01 7.146 1.02e-12 \*\*\*  
## pop\_dens 3.486e+02 2.319e+01 15.033 < 2e-16 \*\*\*  
## income\_per\_capita 2.282e-05 2.808e-06 8.129 5.38e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.675 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.09888, Adjusted R-squared: 0.09802   
## F-statistic: 114 on 5 and 5194 DF, p-value: < 2.2e-16

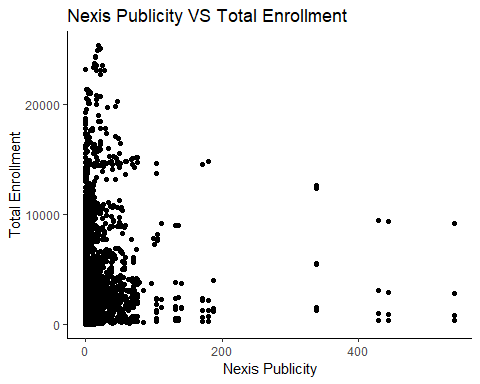
### UNDERGRADUATE ENROLLMENT - Log + 1 Transformation

enroll\_undergrad <- lm(log(EFTOTLT+1) ~ log(nexis\_sum\_3+1) + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Undergraduates", ])   
  
summary(enroll\_undergrad)

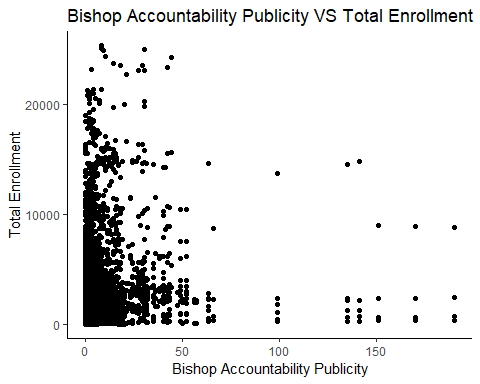
##   
## Call:  
## lm(formula = log(EFTOTLT + 1) ~ log(nexis\_sum\_3 + 1) + Percent.Catholic +   
## his\_prop + pop\_dens + income\_per\_capita, data = data[data$LINE ==   
## "Undergraduates", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.7496 -0.4004 0.2978 0.7127 2.5193   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.693e+00 5.655e-02 118.364 < 2e-16 \*\*\*  
## log(nexis\_sum\_3 + 1) 5.748e-02 1.402e-02 4.101 4.18e-05 \*\*\*  
## Percent.Catholic 7.648e-03 1.337e-03 5.723 1.11e-08 \*\*\*  
## his\_prop 3.697e-01 1.374e-01 2.691 0.00715 \*\*   
## pop\_dens 1.318e+02 1.041e+01 12.664 < 2e-16 \*\*\*  
## income\_per\_capita 1.706e-06 1.261e-06 1.353 0.17602   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.201 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.07485, Adjusted R-squared: 0.07396   
## F-statistic: 84.04 on 5 and 5194 DF, p-value: < 2.2e-16

# Other Things I Have Tried

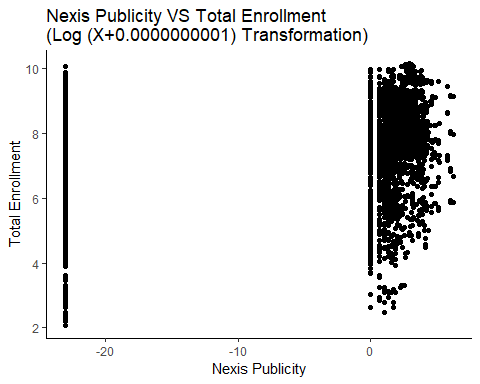
library(ggplot2)  
  
ggplot(data[data$LINE == "Total", ], aes(x= nexis\_sum\_3, y = EFTOTLT))+  
 geom\_point() +  
 theme\_classic() +  
 labs(title = "Nexis Publicity VS Total Enrollment", x = "Nexis Publicity", y = "Total Enrollment")



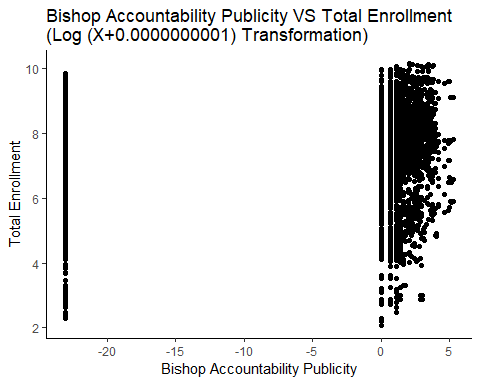
ggplot(data[data$LINE == "Total", ], aes(x= bishop\_pub\_sum\_3, y = EFTOTLT))+  
 geom\_point()+  
 theme\_classic()+  
 labs(title = "Bishop Accountability Publicity VS Total Enrollment", x = "Bishop Accountability Publicity", y = "Total Enrollment")



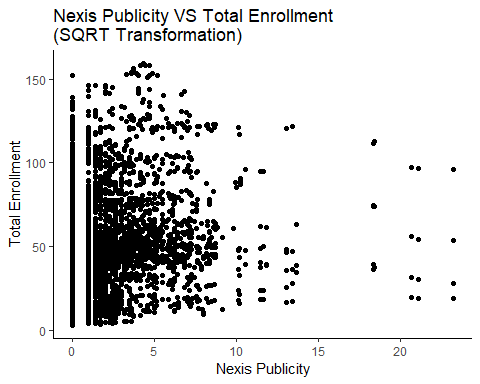
ggplot(data[data$LINE == "Total", ], aes(x= log(nexis\_sum\_3+0.0000000001), y = log(EFTOTLT+0.0000000001)))+  
 geom\_point() +  
 theme\_classic() +  
 labs(title = "Nexis Publicity VS Total Enrollment \n(Log (X+0.0000000001) Transformation)", x = "Nexis Publicity", y = "Total Enrollment")



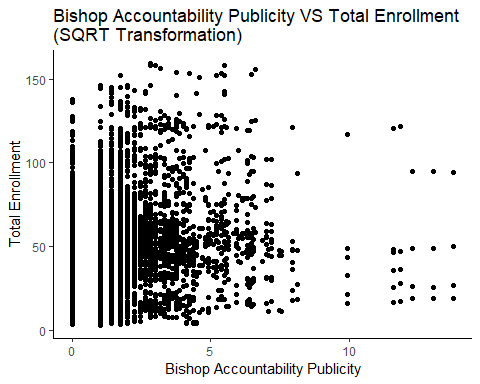
ggplot(data[data$LINE == "Total", ], aes(x= log(bishop\_pub\_sum\_3+0.0000000001), y = log(EFTOTLT+0.0000000001)))+  
 geom\_point()+  
 theme\_classic()+  
 labs(title = "Bishop Accountability Publicity VS Total Enrollment \n(Log (X+0.0000000001) Transformation)", x = "Bishop Accountability Publicity", y = "Total Enrollment")



ggplot(data[data$LINE == "Total", ], aes(x= sqrt(nexis\_sum\_3), y = sqrt(EFTOTLT)))+  
 geom\_point() +  
 theme\_classic() +  
 labs(title = "Nexis Publicity VS Total Enrollment \n(SQRT Transformation)", x = "Nexis Publicity", y = "Total Enrollment")



ggplot(data[data$LINE == "Total", ], aes(x= sqrt(bishop\_pub\_sum\_3), y = sqrt(EFTOTLT)))+  
 geom\_point()+  
 theme\_classic()+  
 labs(title = "Bishop Accountability Publicity VS Total Enrollment \n(SQRT Transformation)", x = "Bishop Accountability Publicity", y = "Total Enrollment")



zip model

M1 <- glm(EFTOTLT ~ nexis\_sum\_3 + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Total", ],  
 family = 'poisson')  
  
summary(M1)

##   
## Call:  
## glm(formula = EFTOTLT ~ nexis\_sum\_3 + Percent.Catholic + his\_prop +   
## pop\_dens + income\_per\_capita, family = "poisson", data = data[data$LINE ==   
## "Total", ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -137.53 -39.78 -14.82 10.18 224.36   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 7.565e+00 8.323e-04 9089.6 <2e-16 \*\*\*  
## nexis\_sum\_3 2.920e-04 6.775e-06 43.1 <2e-16 \*\*\*  
## Percent.Catholic 2.402e-03 2.012e-05 119.4 <2e-16 \*\*\*  
## his\_prop 5.795e-01 1.894e-03 305.9 <2e-16 \*\*\*  
## pop\_dens 1.044e+02 1.035e-01 1008.3 <2e-16 \*\*\*  
## income\_per\_capita 6.203e-06 1.556e-08 398.7 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 15638012 on 5199 degrees of freedom  
## Residual deviance: 13601844 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## AIC: 13650257  
##   
## Number of Fisher Scoring iterations: 5

E2 <- resid(M1, type = "pearson")  
N <- nrow(data)  
p <- length(coef(M1))   
sum(E2^2) / (N - p)

## [1] 1064.226

#overdispersion

library(MASS)  
M2 <- glm.nb(EFTOTLT ~ nexis\_sum\_3 + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita, data=data[data$LINE == "Total",])  
summary(M2)

##   
## Call:  
## glm.nb(formula = EFTOTLT ~ nexis\_sum\_3 + Percent.Catholic + his\_prop +   
## pop\_dens + income\_per\_capita, data = data[data$LINE == "Total",   
## ], init.theta = 1.055981708, link = log)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.1065 -0.8475 -0.2518 0.2072 2.7355   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 7.469e+00 4.534e-02 164.718 < 2e-16 \*\*\*  
## nexis\_sum\_3 4.940e-04 4.323e-04 1.143 0.25311   
## Percent.Catholic 3.604e-03 1.080e-03 3.336 0.00085 \*\*\*  
## his\_prop 8.939e-01 1.116e-01 8.013 1.12e-15 \*\*\*  
## pop\_dens 1.695e+02 8.288e+00 20.447 < 2e-16 \*\*\*  
## income\_per\_capita 4.765e-06 9.524e-07 5.003 5.65e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(1.056) family taken to be 1)  
##   
## Null deviance: 6707.8 on 5199 degrees of freedom  
## Residual deviance: 5964.1 on 5194 degrees of freedom  
## (25 observations deleted due to missingness)  
## AIC: 93747  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 1.0560   
## Std. Err.: 0.0184   
##   
## 2 x log-likelihood: -93733.3000

E2 <- resid(M2, type = "pearson")  
N <- nrow(data)  
p <- length(coef(M2)) + 1 # '+1' is for variance parameter in NB  
sum(E2^2) / (N - p)

## [1] 0.3397552

#underdispersion?

#M3 <- zeroinfl(EFTOTLT ~ nexis\_sum\_3 + Percent.Catholic + his\_prop + pop\_dens + income\_per\_capita,  
 # dist = 'poisson',  
 # data=data[data$LINE == "Total",])  
  
#summary(M3)