IST 687 Group Project

Allison Deming

2/25/2022

```
A <- read.csv("Placement_Data_Full_Class_FINAL.csv", header = TRUE, sep = ",") #Pulling in the d
ata file and naming it A.
getwd() #Getting the working directory
```

```
## [1] "C:/Users/16512/Desktop"
```

```
library(ggplot2)
library(tidyverse)
```

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

```
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.0     v forcats 0.5.1
## v purrr   0.3.4
```

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

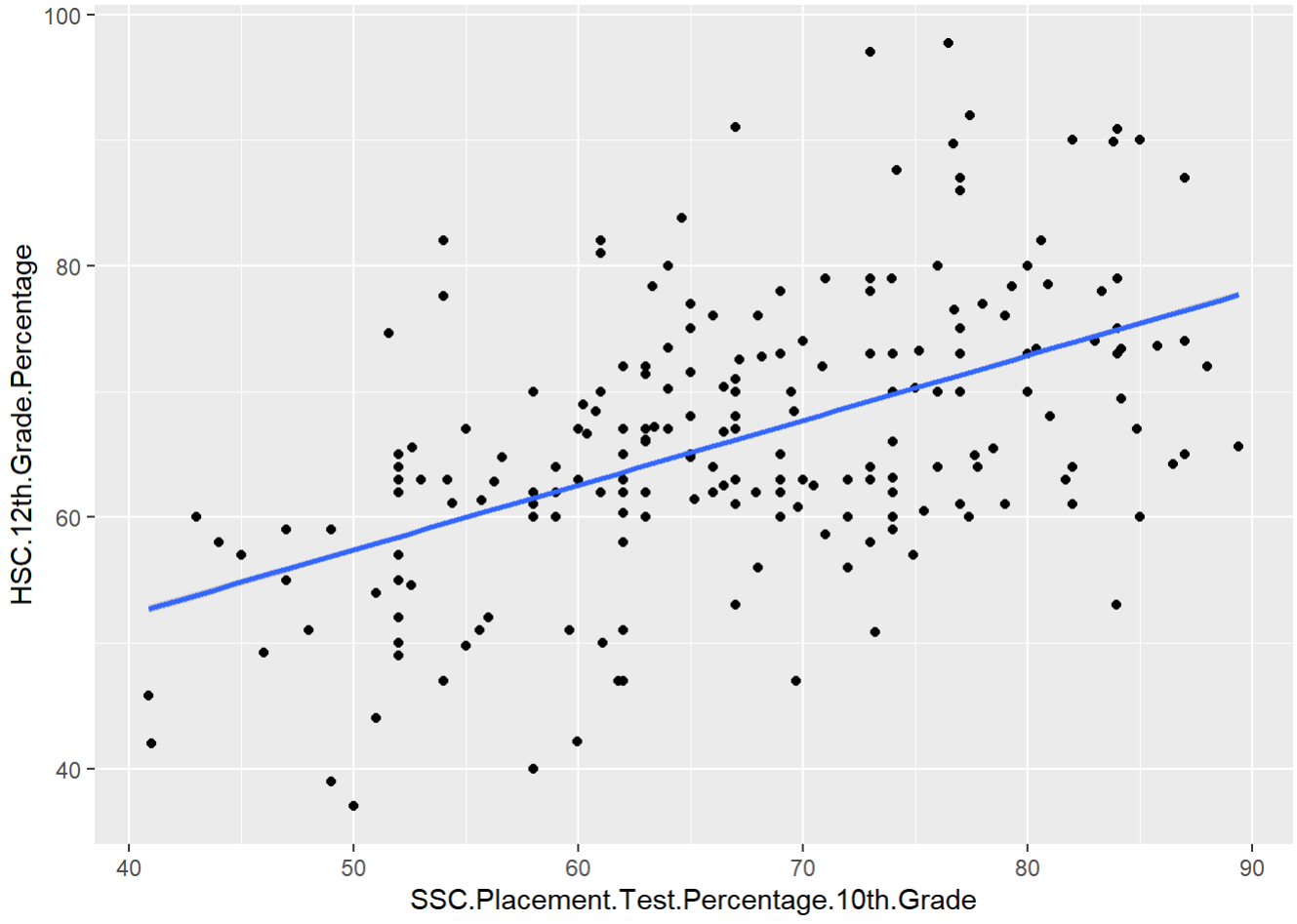
```
#Correlations
cor.test(A$SSC.Placement.Test.Percentage.10th.Grade,A$HSC.12th.Grade.Percentage)
```

```
##
##  Pearson's product-moment correlation
##
## data:  A$SSC.Placement.Test.Percentage.10th.Grade and A$HSC.12th.Grade.Percentage
## t = 86.835, df = 21283, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5014836 0.5213243
## sample estimates:
##       cor
## 0.5114721
```

```
LM0<-lm(formula=A$SSC.Placement.Test.Percentage.10th.Grade ~ A$HSC.12th.Grade.Percentage, data=
A)
summary(LM0)
```

```
##
## Call:
## lm(formula = A$SSC.Placement.Test.Percentage.10th.Grade ~ A$HSC.12th.Grade.Percentage,
##     data = A)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.265  -6.593  -1.041   6.472  23.432
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  33.594712   0.393373   85.40   <2e-16 ***
## A$HSC.12th.Grade.Percentage  0.508172   0.005852   86.83   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.283 on 21283 degrees of freedom
## Multiple R-squared:  0.2616, Adjusted R-squared:  0.2616
## F-statistic:  7540 on 1 and 21283 DF,  p-value: < 2.2e-16
```

```
plotA <- ggplot(data = A, aes(x = SSC.Placement.Test.Percentage.10th.Grade, y =HSC.12th.Grade.Pe
rcentage)) + geom_point()
plotA + geom_smooth(formula = y ~ x, method = "lm")
```



```
#Moderate positive correlation between tests with r=.512 and R squared = 051. This is statistica
lly significant. Due to the large number of values for n.
```

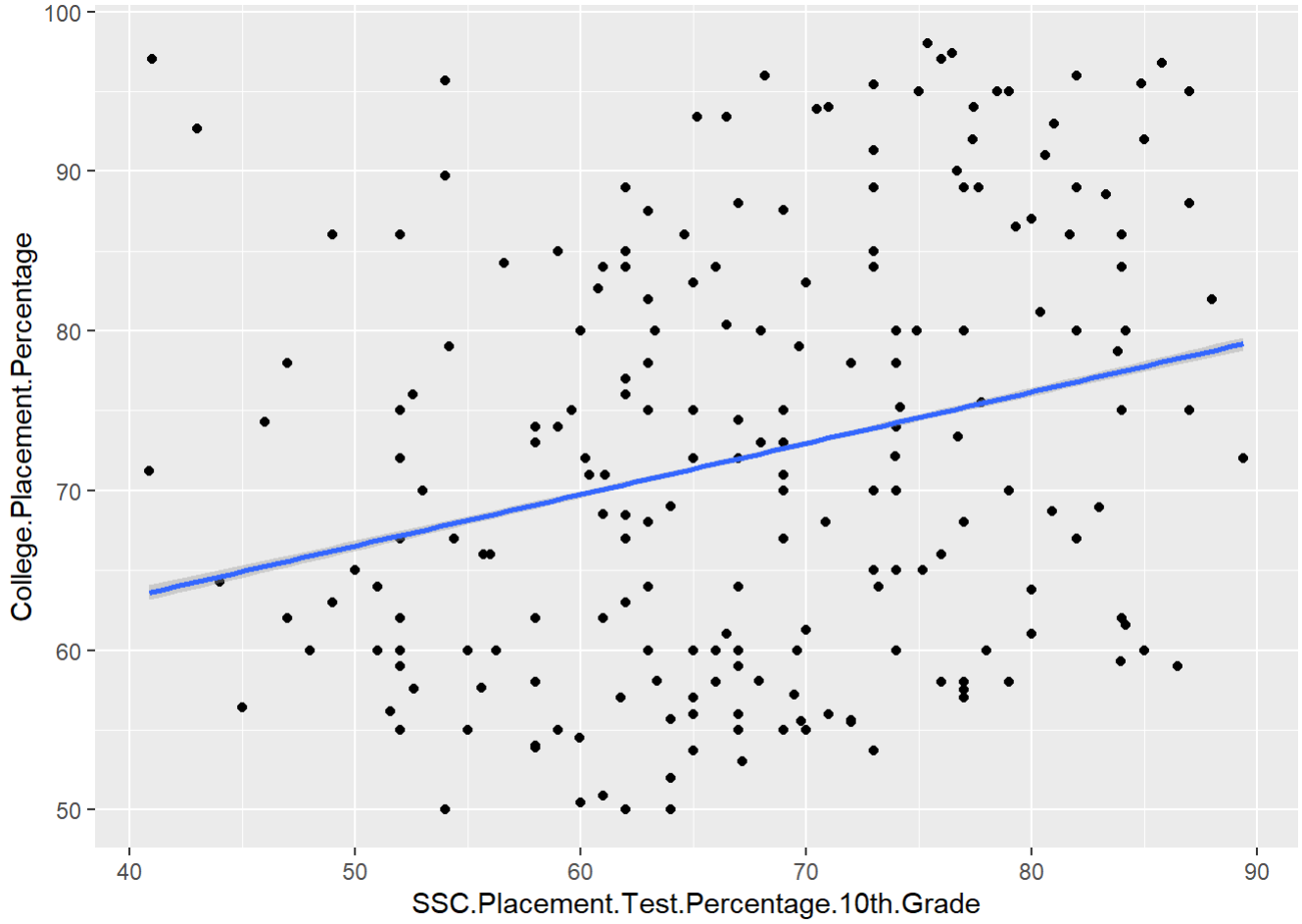
```
cor.test(A$SSC.Placement.Test.Percentage.10th.Grade,A$College.Placement.Percentage)
```

```
##
##  Pearson's product-moment correlation
##
## data:  A$SSC.Placement.Test.Percentage.10th.Grade and A$College.Placement.Percentage
## t = 39.605, df = 21283, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2494363 0.2744610
## sample estimates:
##       cor
## 0.2619927
```

```
LM00<-lm(formula=A$SSC.Placement.Test.Percentage.10th.Grade ~ A$College.Placement.Percentage, da
ta=A)
summary(LM00)
```

```
##
## Call:
## lm(formula = A$SSC.Placement.Test.Percentage.10th.Grade ~ A$College.Placement.Percentage,
##     data = A)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -31.624  -6.668   0.923   7.214  22.118
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   51.897804   0.395493   131.2   <2e-16 ***
## A$College.Placement.Percentage  0.213668   0.005395    39.6   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.43 on 21283 degrees of freedom
## Multiple R-squared:  0.06864,    Adjusted R-squared:  0.0686
## F-statistic:  1569 on 1 and 21283 DF,  p-value: < 2.2e-16
```

```
plotB <- ggplot(data = A, aes(x = SSC.Placement.Test.Percentage.10th.Grade, y =College.Placemen
t.Percentage)) + geom_point()
plotB + geom_smooth(formula = y ~ x, method = "lm")
```




```
#Correlation is positive at r=.26 and with a linear model r squared value of .214. The t value i
s statistically significant.
```

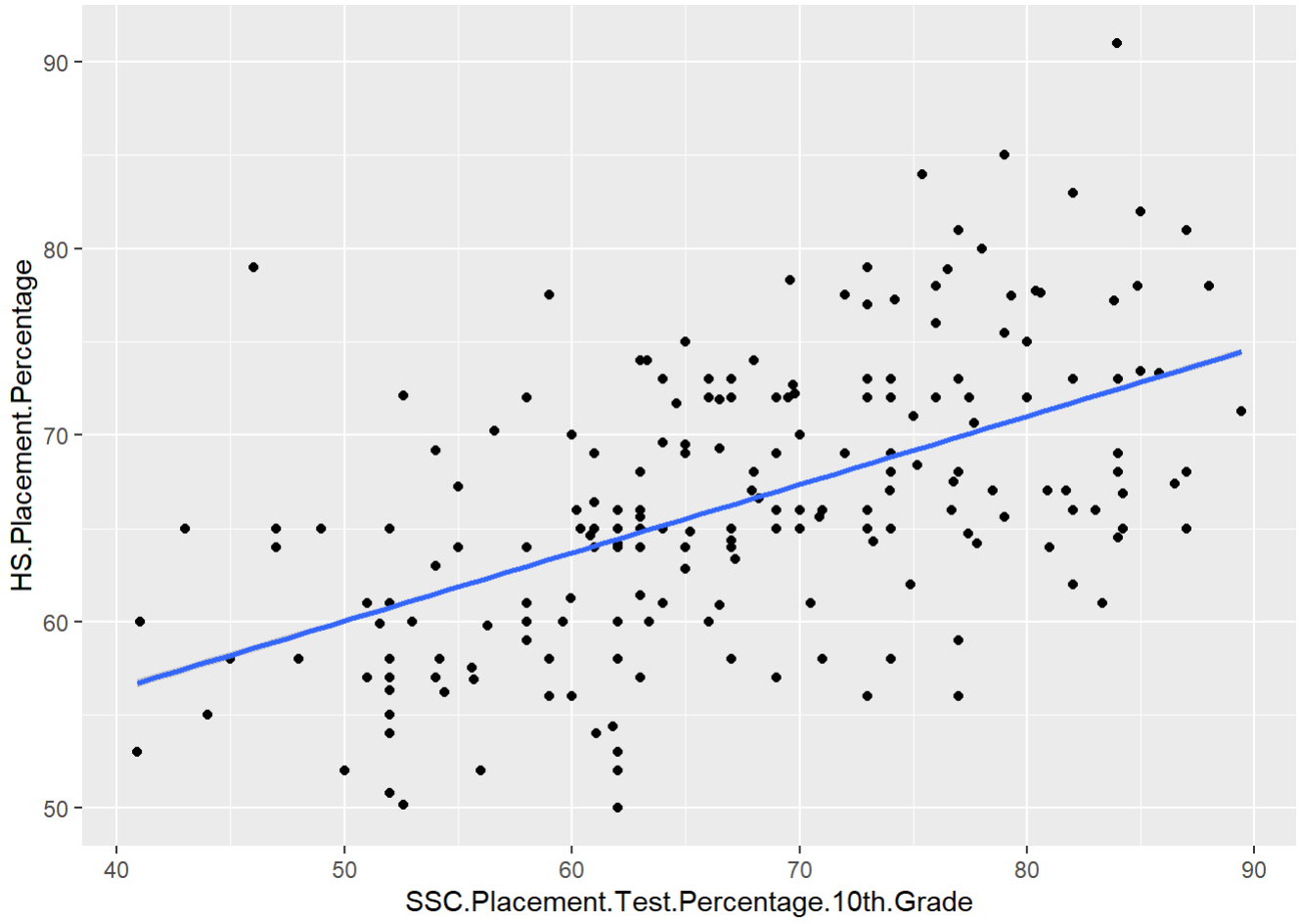
```
cor.test(A$SSC.Placement.Test.Percentage.10th.Grade, A$HS.Placement.Percentage)
```

```
##
##  Pearson's product-moment correlation
##
## data:  A$SSC.Placement.Test.Percentage.10th.Grade and A$HS.Placement.Percentage
## t = 93.209, df = 21283, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5287945 0.5478755
## sample estimates:
##      cor
## 0.538404
```

```
LM000<-lm(formula=A$SSC.Placement.Test.Percentage.10th.Grade ~ A$HS.Placement.Percentage, data=
A)
summary(LM000)
```

```
##
## Call:
## lm(formula = A$SSC.Placement.Test.Percentage.10th.Grade ~ A$HS.Placement.Percentage,
##     data = A)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -31.3084  -5.5040  -0.2571   5.9899  20.7820
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                14.726609   0.567514   25.95   <2e-16 ***
## A$HS.Placement.Percentage   0.792175   0.008499   93.21   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.103 on 21283 degrees of freedom
## Multiple R-squared:  0.2899, Adjusted R-squared:  0.2898
## F-statistic:  8688 on 1 and 21283 DF,  p-value: < 2.2e-16
```

```
library(ggplot2)
plotC <- ggplot(data = A, aes(x = SSC.Placement.Test.Percentage.10th.Grade, y =HS.Placement.Perc
entage)) + geom_point()
plotC + geom_smooth(formula = y ~ x, method = "lm")
```



```
#Correlation here is positive with r=.54, and an explain
```

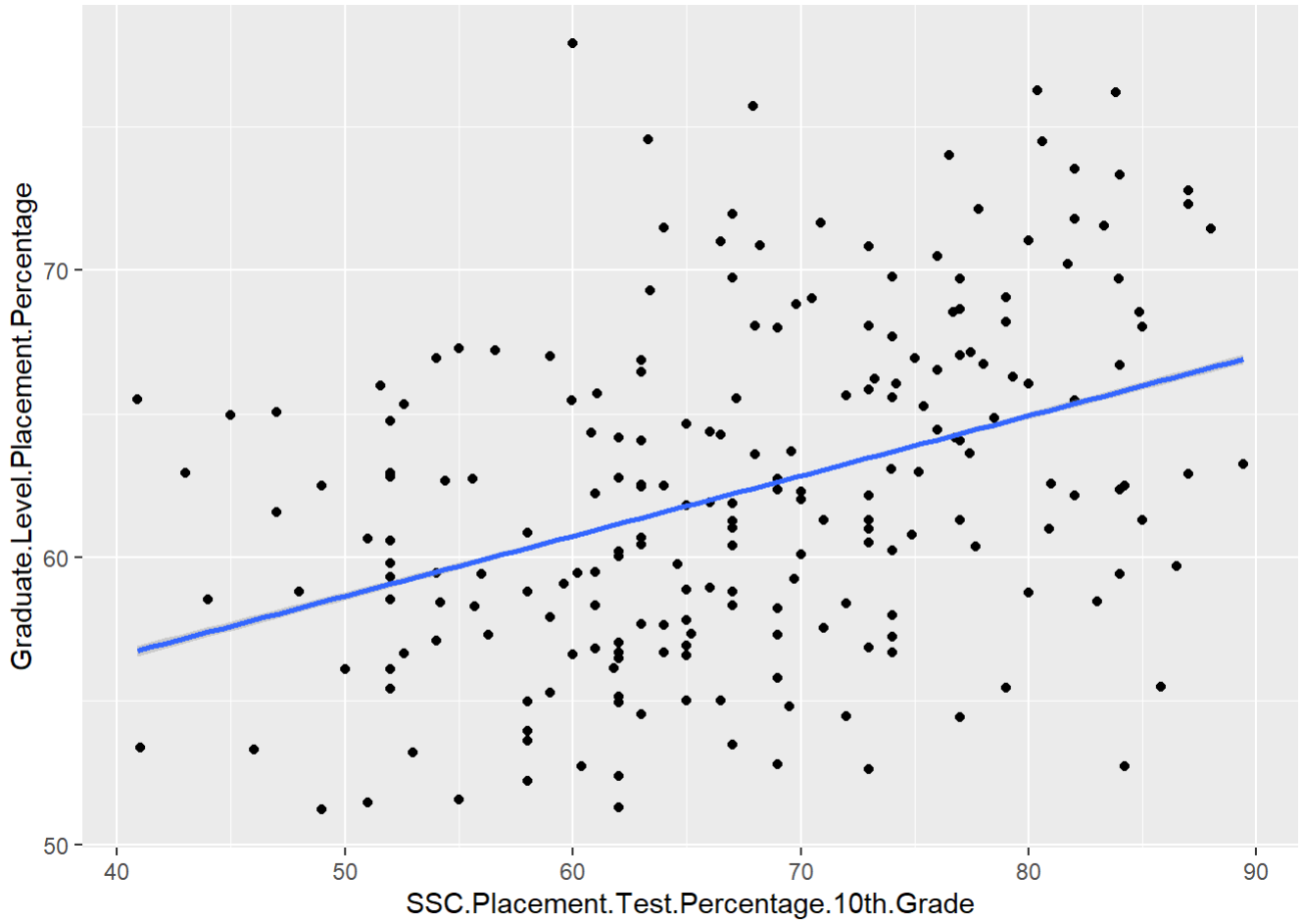
```
cor.test(A$SSC.Placement.Test.Percentage.10th.Grade,A$Graduate.Level.Placement.Percentage)
```

```
##
##  Pearson's product-moment correlation
##
## data:  A$SSC.Placement.Test.Percentage.10th.Grade and A$Graduate.Level.Placement.Percentage
## t = 61.504, df = 21283, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3770108 0.3998252
## sample estimates:
##       cor
## 0.3884776
```

```
LM0000<-lm(formula=A$SSC.Placement.Test.Percentage.10th.Grade ~ A$Graduate.Level.Placement.Perce
ntage, data=A)
summary(LM0000)
```

```
##
## Call:
## lm(formula = A$SSC.Placement.Test.Percentage.10th.Grade ~ A$Graduate.Level.Placement.Percenta
ge,
##     data = A)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.7293  -6.6531   0.7187   6.8066  23.7885
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        22.39809    0.73330   30.54   <2e-16 ***
## A$Graduate.Level.Placement.Percentage  0.72104    0.01172   61.50   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.954 on 21283 degrees of freedom
## Multiple R-squared:  0.1509, Adjusted R-squared:  0.1509
## F-statistic:  3783 on 1 and 21283 DF,  p-value: < 2.2e-16
```

```
plotD <- ggplot(data = A, aes(x = SSC.Placement.Test.Percentage.10th.Grade, y =Graduate.Level.Pl
acement.Percentage)) + geom_point()
plotD + geom_smooth(formula = y ~ x, method = "lm")
```




```
#Here there is also a positive correlation between the two variables, with r=.39 and the account
ed for r squared value at .721. These are both statistically significant.
```

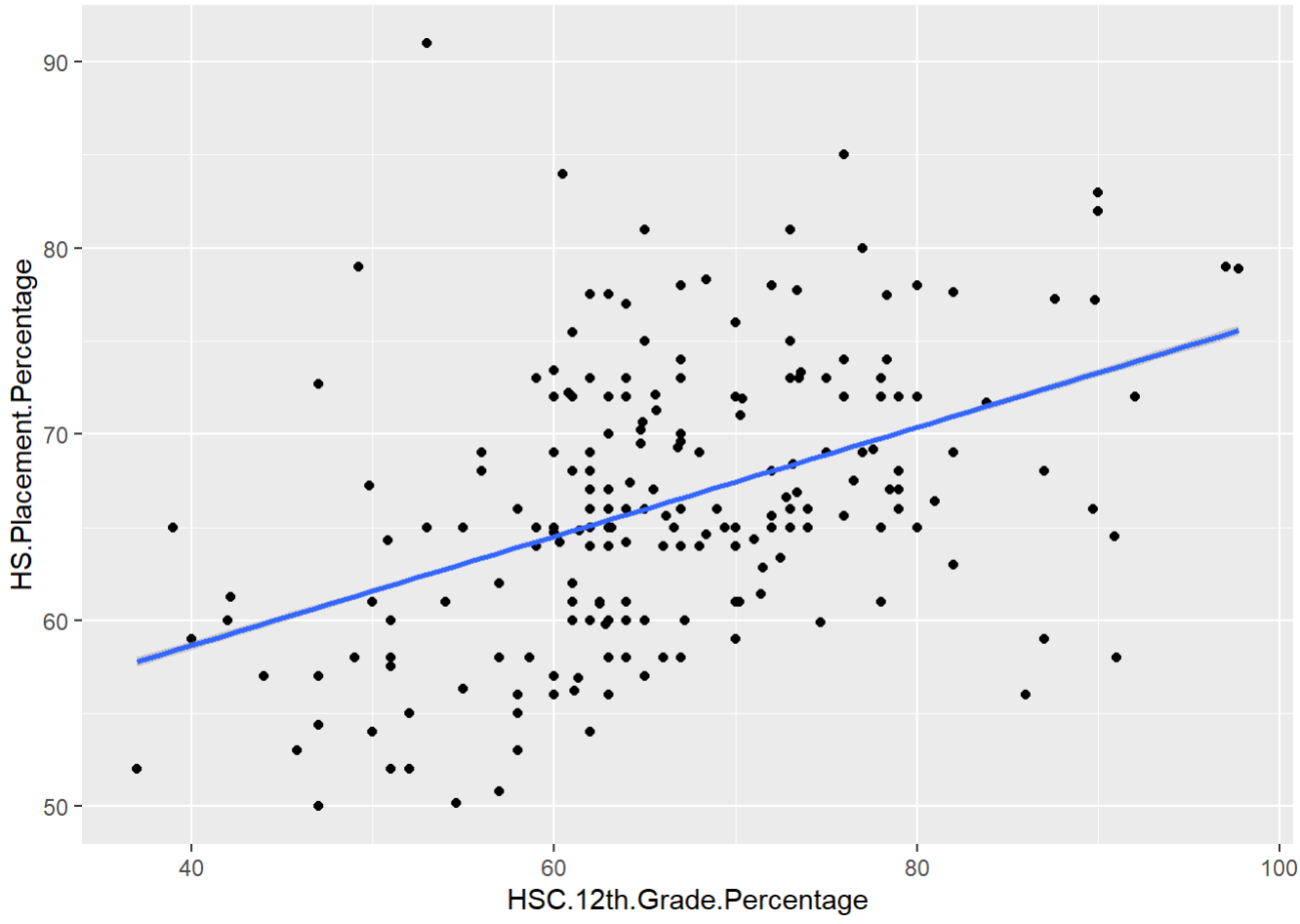
```
cor.test(A$HSC.12th.Grade.Percentage, A$HS.Placement.Percentage)
```

```
##
##  Pearson's product-moment correlation
##
## data:  A$HSC.12th.Grade.Percentage and A$HS.Placement.Percentage
## t = 70.32, df = 21283, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4232404 0.4450441
## sample estimates:
##       cor
## 0.4342058
```

```
LM00000<-lm(formula=A$HSC.12th.Grade.Percentage ~ A$HS.Placement.Percentage, data=A)
summary(LM00000)
```

```
##
## Call:
## lm(formula = A$HSC.12th.Grade.Percentage ~ A$HS.Placement.Percentage,
##     data = A)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.170  -5.667  -0.308   6.049  30.049
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                23.656323   0.610599   38.74   <2e-16 ***
## A$HS.Placement.Percentage   0.643012   0.009144   70.32   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.794 on 21283 degrees of freedom
## Multiple R-squared:  0.1885, Adjusted R-squared:  0.1885
## F-statistic:  4945 on 1 and 21283 DF,  p-value: < 2.2e-16
```

```
plotE <- ggplot(data = A, aes(x = HSC.12th.Grade.Percentage, y =HS.Placement.Percentage)) + geom
_point()
plotE + geom_smooth(formula = y ~ x, method = "lm")
```



```
#Here there is a positive correlation between the two placement tests, with r=.423, and the r squared value accounting for 64.3% of the variance. Both are statistically significant.
```

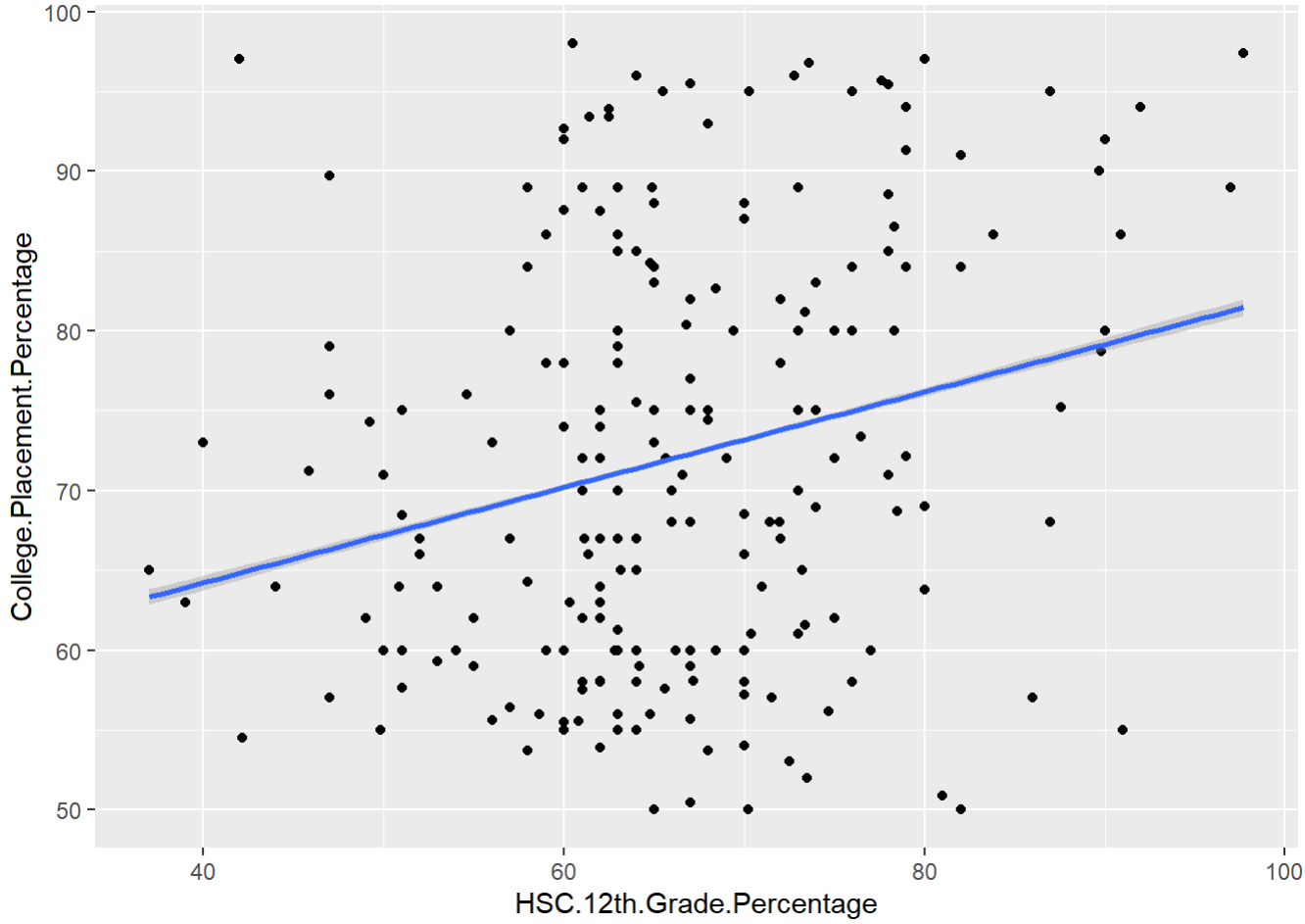
```
cor.test(A$HSC.12th.Grade.Percentage, A$College.Placement.Percentage)
```

```
##
##  Pearson's product-moment correlation
##
## data:  A$HSC.12th.Grade.Percentage and A$College.Placement.Percentage
## t = 36.884, df = 21283, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2324440 0.2576987
## sample estimates:
##       cor
## 0.2451129
```

```
LM01<-lm(formula=A$HSC.12th.Grade.Percentage ~ A$College.Placement.Percentage, data=A)
summary(LM01)
```

```
##
## Call:
## lm(formula = A$HSC.12th.Grade.Percentage ~ A$College.Placement.Percentage,
##     data = A)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.3429  -6.1297  -0.6529   6.9669  28.1075
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   51.826542   0.399886  129.60   <2e-16 ***
## A$College.Placement.Percentage  0.201200   0.005455   36.88   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.54 on 21283 degrees of freedom
## Multiple R-squared:  0.06008,    Adjusted R-squared:  0.06004
## F-statistic:  1360 on 1 and 21283 DF,  p-value: < 2.2e-16
```

```
plotF <- ggplot(data = A, aes(x = HSC.12th.Grade.Percentage, y =College.Placement.Percentage)) +
geom_point()
plotF + geom_smooth(formula = y ~ x, method = "lm")
```



#Here there is a positive correlation between the two values, with r=.25. The Explained variation of r squared here is only just over 20% of the variance, still statistically significant but not as high as some of what we have seen between previous tests.

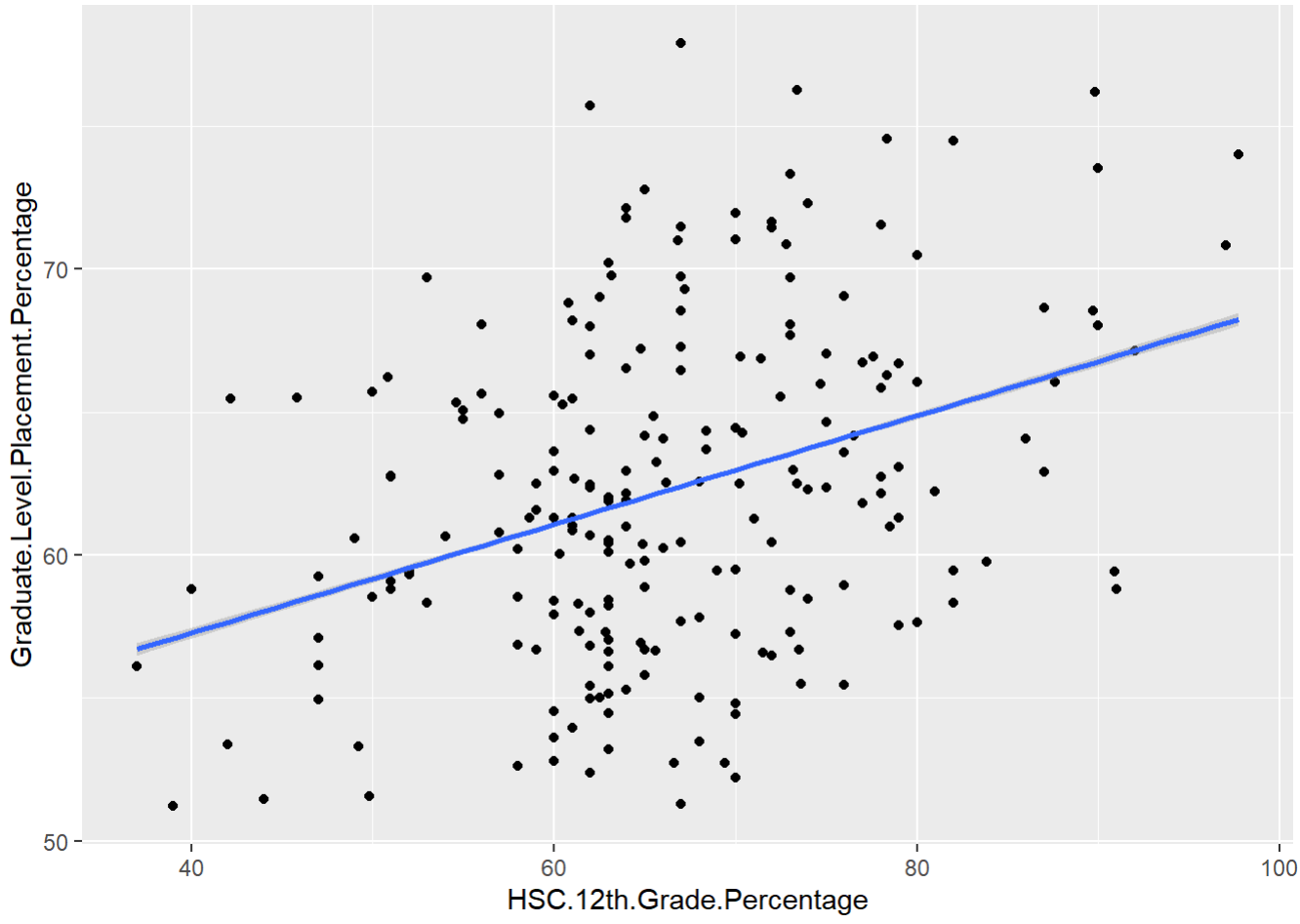
```
cor.test(A$HSC.12th.Grade.Percentage, A$Graduate.Level.Placement.Percentage)
```

```
##
##  Pearson's product-moment correlation
##
## data:  A$HSC.12th.Grade.Percentage and A$Graduate.Level.Placement.Percentage
## t = 55.366, df = 21283, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3430234 0.3665098
## sample estimates:
##       cor
## 0.3548226
```

```
LM010<-lm(formula=A$HSC.12th.Grade.Percentage ~ A$Graduate.Level.Placement.Percentage, data=A)
summary(LM010)
```

```
##
## Call:
## lm(formula = A$HSC.12th.Grade.Percentage ~ A$Graduate.Level.Placement.Percentage,
##     data = A)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -26.2955  -6.8704  -0.3995   7.0064  26.9724
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         25.05182    0.74885   33.45   <2e-16 ***
## A$Graduate.Level.Placement.Percentage  0.66285    0.01197   55.37   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.17 on 21283 degrees of freedom
## Multiple R-squared:  0.1259, Adjusted R-squared:  0.1259
## F-statistic:  3065 on 1 and 21283 DF,  p-value: < 2.2e-16
```

```
plotG <- ggplot(data = A, aes(x = HSC.12th.Grade.Percentage, y =Graduate.Level.Placement.Percent
age)) + geom_point()
plotG + geom_smooth(formula = y ~ x, method = "lm")
```



```
#Here r=.36, with a r squared in our linear model of 66.3%. Both are statistically significant.
```

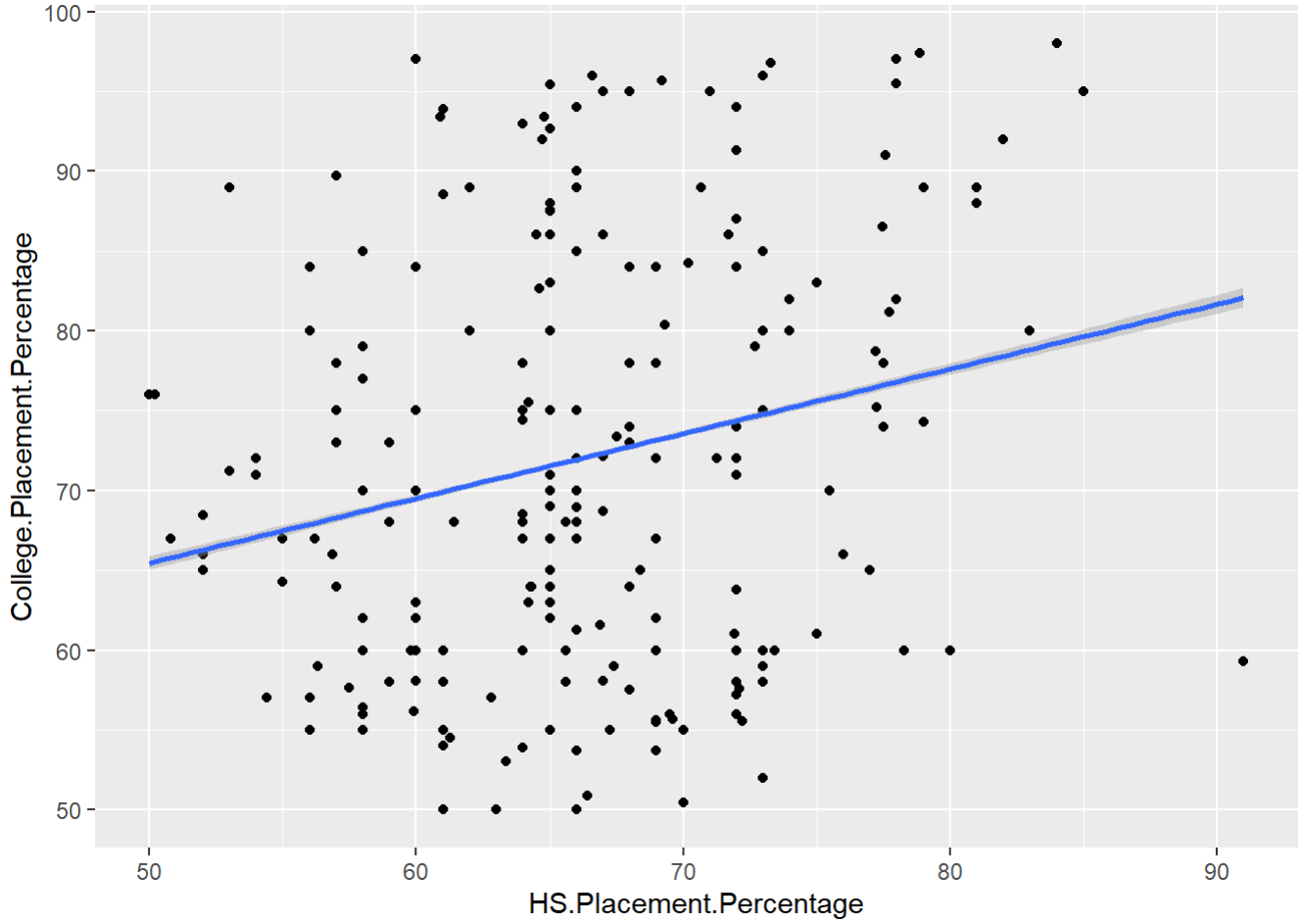
```
cor.test(A$HS.Placement.Percentage, A$College.Placement.Percentage)
```

```
##
##  Pearson's product-moment correlation
##
## data:  A$HS.Placement.Percentage and A$College.Placement.Percentage
## t = 33.605, df = 21283, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2116742 0.2371892
## sample estimates:
##       cor
## 0.2244702
```

```
LM101<-lm(formula=A$HSC.12th.Grade.Percentage ~ A$College.Placement.Percentage, data=A)
summary(LM101)
```

```
##
## Call:
## lm(formula = A$HSC.12th.Grade.Percentage ~ A$College.Placement.Percentage,
##     data = A)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.3429  -6.1297  -0.6529   6.9669  28.1075
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   51.826542   0.399886  129.60   <2e-16 ***
## A$College.Placement.Percentage  0.201200   0.005455   36.88   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.54 on 21283 degrees of freedom
## Multiple R-squared:  0.06008,    Adjusted R-squared:  0.06004
## F-statistic:  1360 on 1 and 21283 DF,  p-value: < 2.2e-16
```

```
plotH <- ggplot(data = A, aes(x = HS.Placement.Percentage, y =College.Placement.Percentage)) + g
eom_point()
plotH + geom_smooth(formula = y ~ x, method = "lm")
```




```
#Here the correlation is r=.23, and the linear model accounts for about 20% of the variance. Bot
h are statistically significant.
```

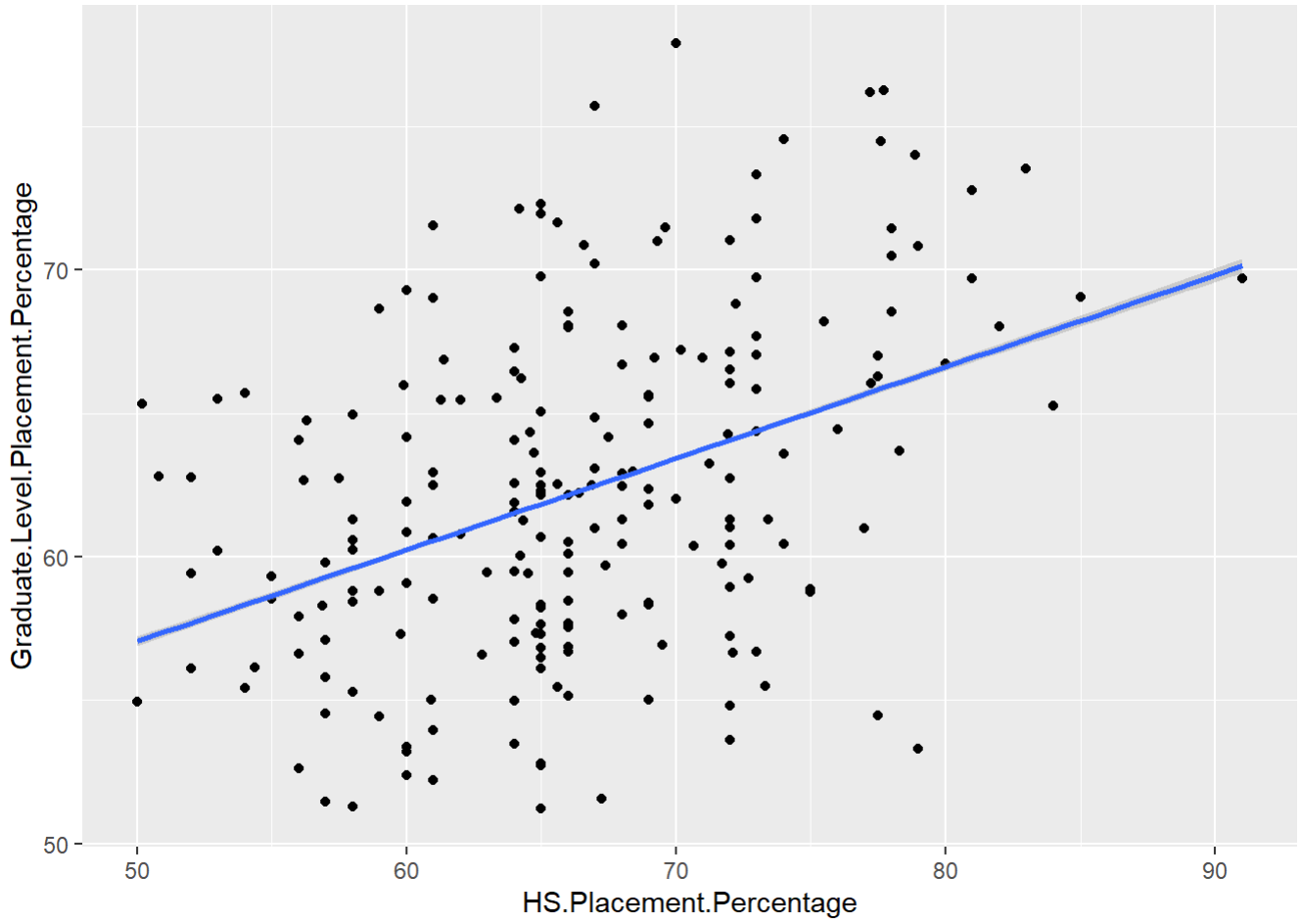
```
cor.test(A$HS.Placement.Percentage, A$Graduate.Level.Placement.Percentage)
```

```
##
##  Pearson's product-moment correlation
##
## data:  A$HS.Placement.Percentage and A$Graduate.Level.Placement.Percentage
## t = 64.119, df = 21283, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3910432 0.4135626
## sample estimates:
##       cor
## 0.4023638
```

```
LM0101<-lm(formula=A$HS.Placement.Percentage ~ A$Graduate.Level.Placement.Percentage, data=A)
summary(LM0101)
```

```
##
## Call:
## lm(formula = A$HS.Placement.Percentage ~ A$Graduate.Level.Placement.Percentage,
##     data = A)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.7192  -4.8282   0.4174   4.2478  20.8576
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         34.759213   0.495153   70.20   <2e-16 ***
## A$Graduate.Level.Placement.Percentage  0.507577   0.007916   64.12   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.721 on 21283 degrees of freedom
## Multiple R-squared:  0.1619, Adjusted R-squared:  0.1619
## F-statistic:  4111 on 1 and 21283 DF,  p-value: < 2.2e-16
```

```
plotI <- ggplot(data = A, aes(x = HS.Placement.Percentage, y =Graduate.Level.Placement.Percentag
e)) + geom_point()
plotI + geom_smooth(formula = y ~ x, method = "lm")
```



```
#Here the correlation is r=.40, and the r squared value accounts for 51% of the variance in the
 model. Both are statistically significant.
```

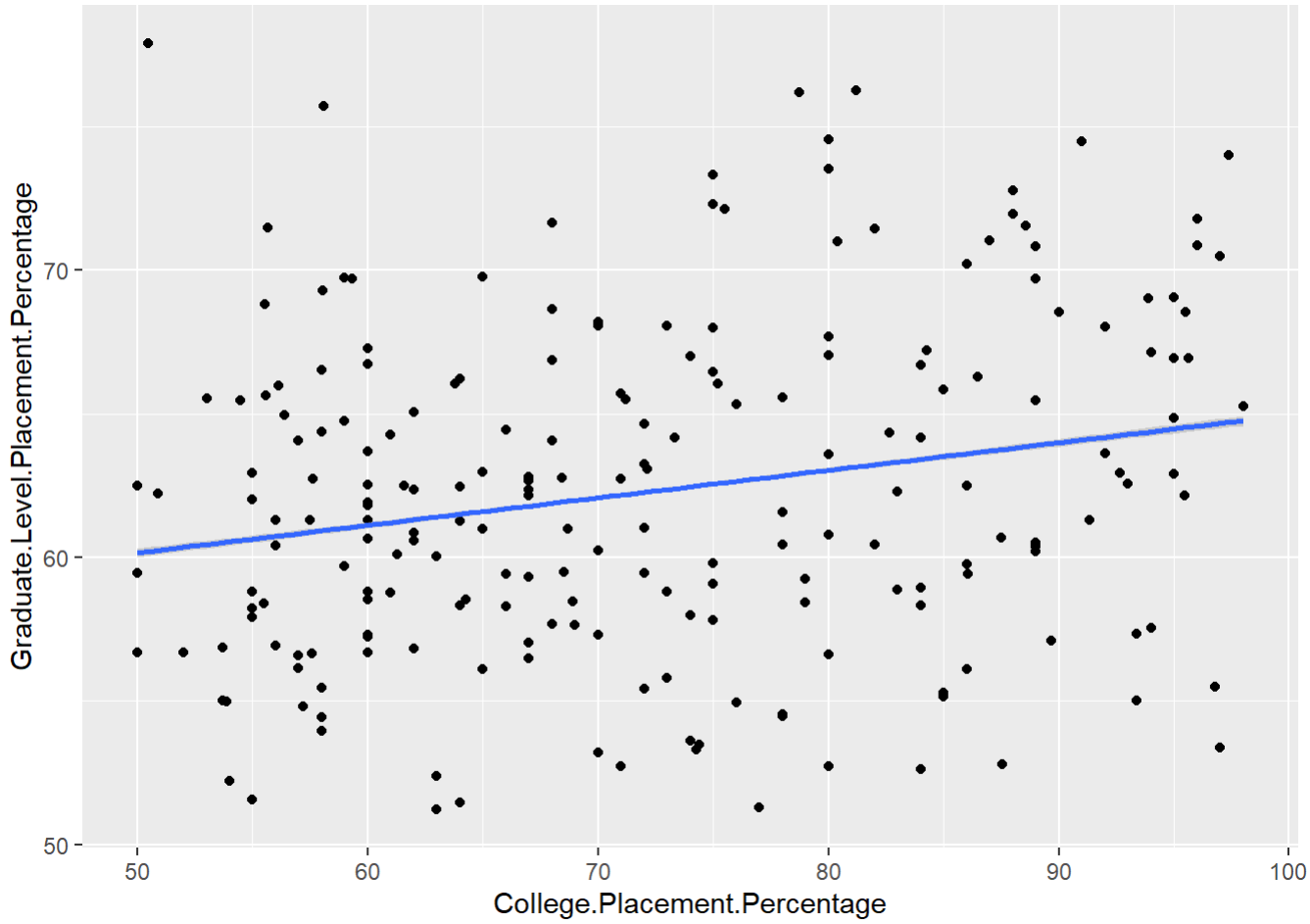
```
cor.test(A$College.Placement.Percentage, A$Graduate.Level.Placement.Percentage)
```

```
##
##  Pearson's product-moment correlation
##
## data:  A$College.Placement.Percentage and A$Graduate.Level.Placement.Percentage
## t = 32.596, df = 21283, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2052215 0.2308128
## sample estimates:
##       cor
## 0.2180547
```

```
LM01010<-lm(formula=A$College.Placement.Percentage ~ A$Graduate.Level.Placement.Percentage, data
=A)
summary(LM01010)
```

```
##
## Call:
## lm(formula = A$College.Placement.Percentage ~ A$Graduate.Level.Placement.Percentage,
##     data = A)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.368 -10.813  -1.495  10.702  29.310
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         41.19429    0.95230   43.26   <2e-16 ***
## A$Graduate.Level.Placement.Percentage  0.49626    0.01522   32.60   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.93 on 21283 degrees of freedom
## Multiple R-squared:  0.04755,    Adjusted R-squared:  0.0475
## F-statistic:  1062 on 1 and 21283 DF,  p-value: < 2.2e-16
```

```
plotJ <- ggplot(data = A, aes(x = College.Placement.Percentage, y =Graduate.Level.Placement.Perc
entage)) + geom_point()
plotJ + geom_smooth(formula = y ~ x, method = "lm")
```




```
#Here there is a positive correlation of r=.22, and the explained variance of the linear model a
ccounts for about 50% of the variance.
```

```
cor.test(A$SSC.Placement.Test.Percentage.10th.Grade, A$Reported.Salary)
```

```
##
##  Pearson's product-moment correlation
##
## data:  A$SSC.Placement.Test.Percentage.10th.Grade and A$Reported.Salary
## t = 4.279, df = 14650, p-value = 1.89e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.01914909 0.05149309
## sample estimates:
##        cor
## 0.03533034
```

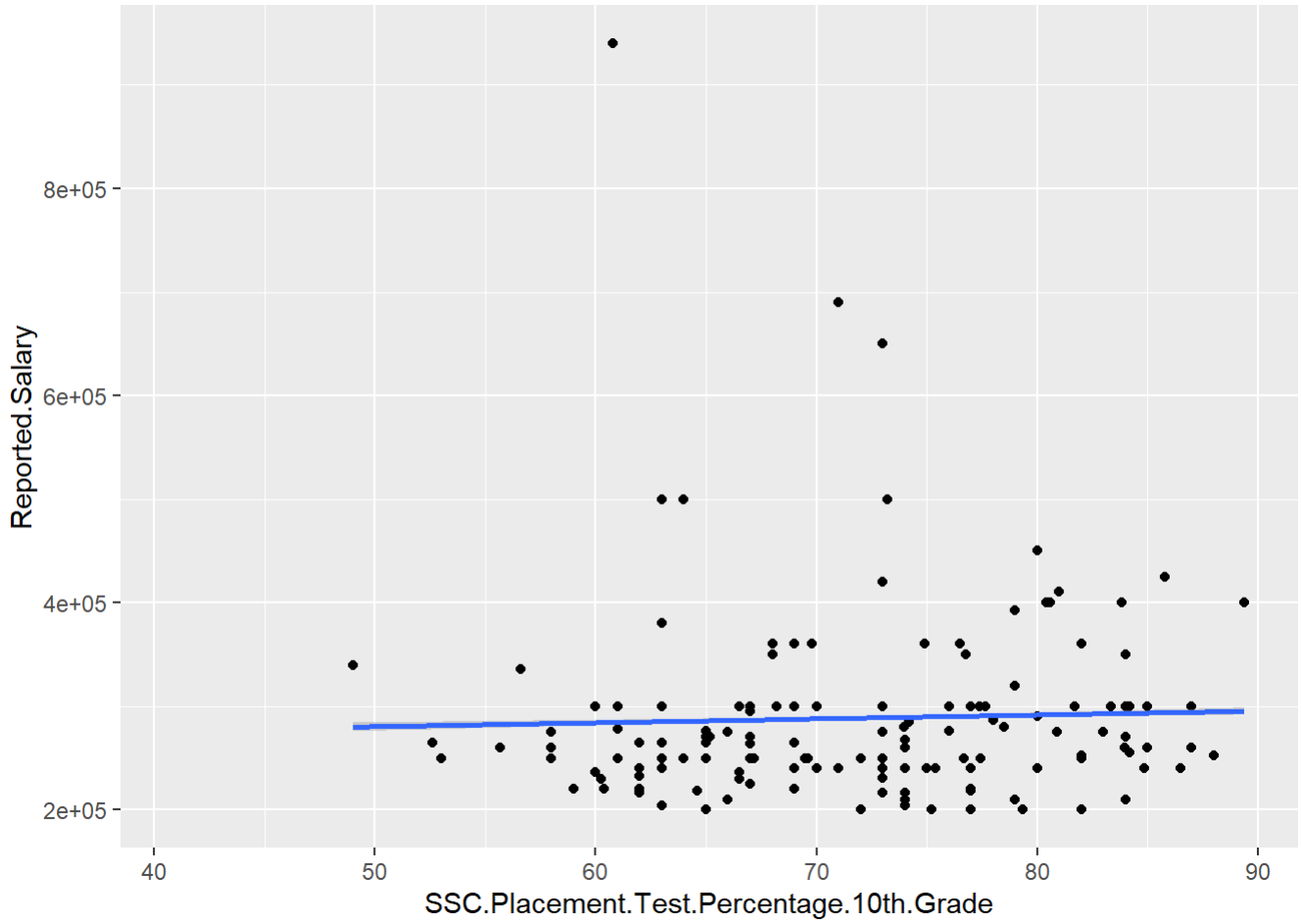
```
LM020<-lm(formula=A$SSC.Placement.Test.Percentage.10th.Grade ~ A$Reported.Salary, data=A)
summary(LM020)
```

```
##
## Call:
## lm(formula = A$SSC.Placement.Test.Percentage.10th.Grade ~ A$Reported.Salary,
##     data = A)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.8907  -6.6435   0.4882   6.4147  17.3117
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       7.077e+01  2.335e-01 303.025  < 2e-16 ***
## A$Reported.Salary 3.295e-06  7.700e-07   4.279 1.89e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.681 on 14650 degrees of freedom
##   (6633 observations deleted due to missingness)
## Multiple R-squared:  0.001248,   Adjusted R-squared:  0.00118
## F-statistic: 18.31 on 1 and 14650 DF,  p-value: 1.89e-05
```

```
plotK <- ggplot(data = A, aes(x = SSC.Placement.Test.Percentage.10th.Grade, y =Reported.Salary))
+ geom_point()
plotK + geom_smooth(formula = y ~ x, method = "lm")
```

```
## Warning: Removed 6633 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 6633 rows containing missing values (geom_point).
```




```
#Here there is a positive correlation of r=.03, and the explained variance is basically non-exis
tent. While there is statistical significance, this is because of the very large n value. There
 is no practical significance.
```

```
cor.test(A$HSC.12th.Grade.Percentage, A$Reported.Salary)
```

```
##
##  Pearson's product-moment correlation
##
## data:  A$HSC.12th.Grade.Percentage and A$Reported.Salary
## t = 9.3255, df = 14650, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.06070233 0.09289569
## sample estimates:
##       cor
## 0.07681903
```

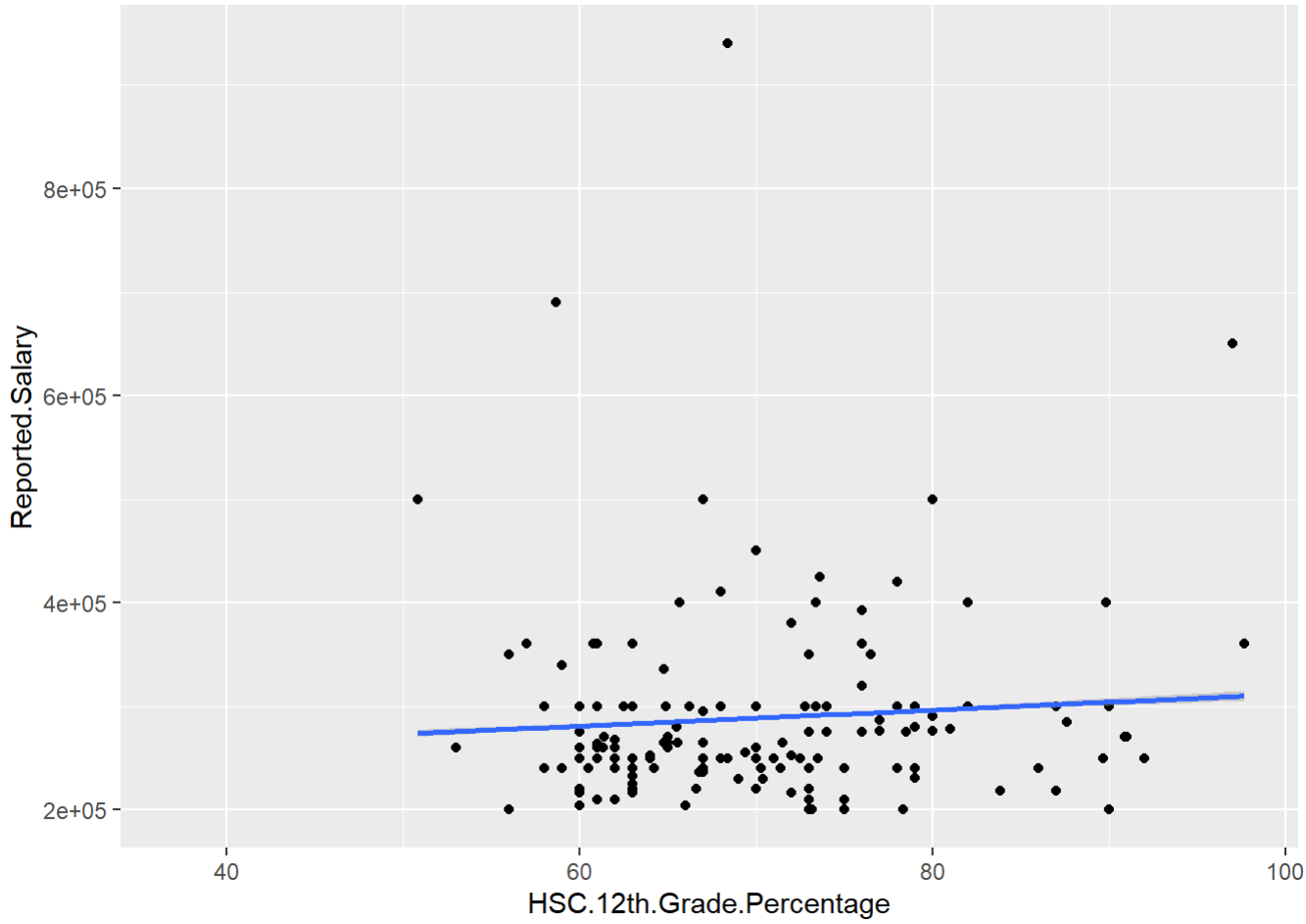
```
LM030<-lm(formula=A$HSC.12th.Grade.Percentage ~ A$Reported.Salary, data=A)
summary(LM030)
```

```
##
## Call:
## lm(formula = A$HSC.12th.Grade.Percentage ~ A$Reported.Salary,
##     data = A)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.717  -7.091  -1.822   5.564  27.226
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       6.771e+01  2.494e-01 271.491   <2e-16 ***
## A$Reported.Salary 7.668e-06  8.223e-07   9.326   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.271 on 14650 degrees of freedom
##   (6633 observations deleted due to missingness)
## Multiple R-squared:  0.005901,   Adjusted R-squared:  0.005833
## F-statistic: 86.97 on 1 and 14650 DF,  p-value: < 2.2e-16
```

```
plotL <- ggplot(data = A, aes(x = HSC.12th.Grade.Percentage, y =Reported.Salary)) + geom_point()
plotL + geom_smooth(formula = y ~ x, method = "lm")
```

```
## Warning: Removed 6633 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 6633 rows containing missing values (geom_point).
```



#Here there is a positive correlation of r=.077, and the explained variance is basically non-existent. While there is statistical significance, this is because of the very large n value. There is no practical significance.

```
cor.test(A$HS.Placement.Percentage, A$Reported.Salary)
```

```
##
##  Pearson's product-moment correlation
##
## data:  A$HS.Placement.Percentage and A$Reported.Salary
## t = -2.3331, df = 14650, p-value = 0.01966
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.035453382 -0.003080986
## sample estimates:
##         cor
## -0.01927223
```

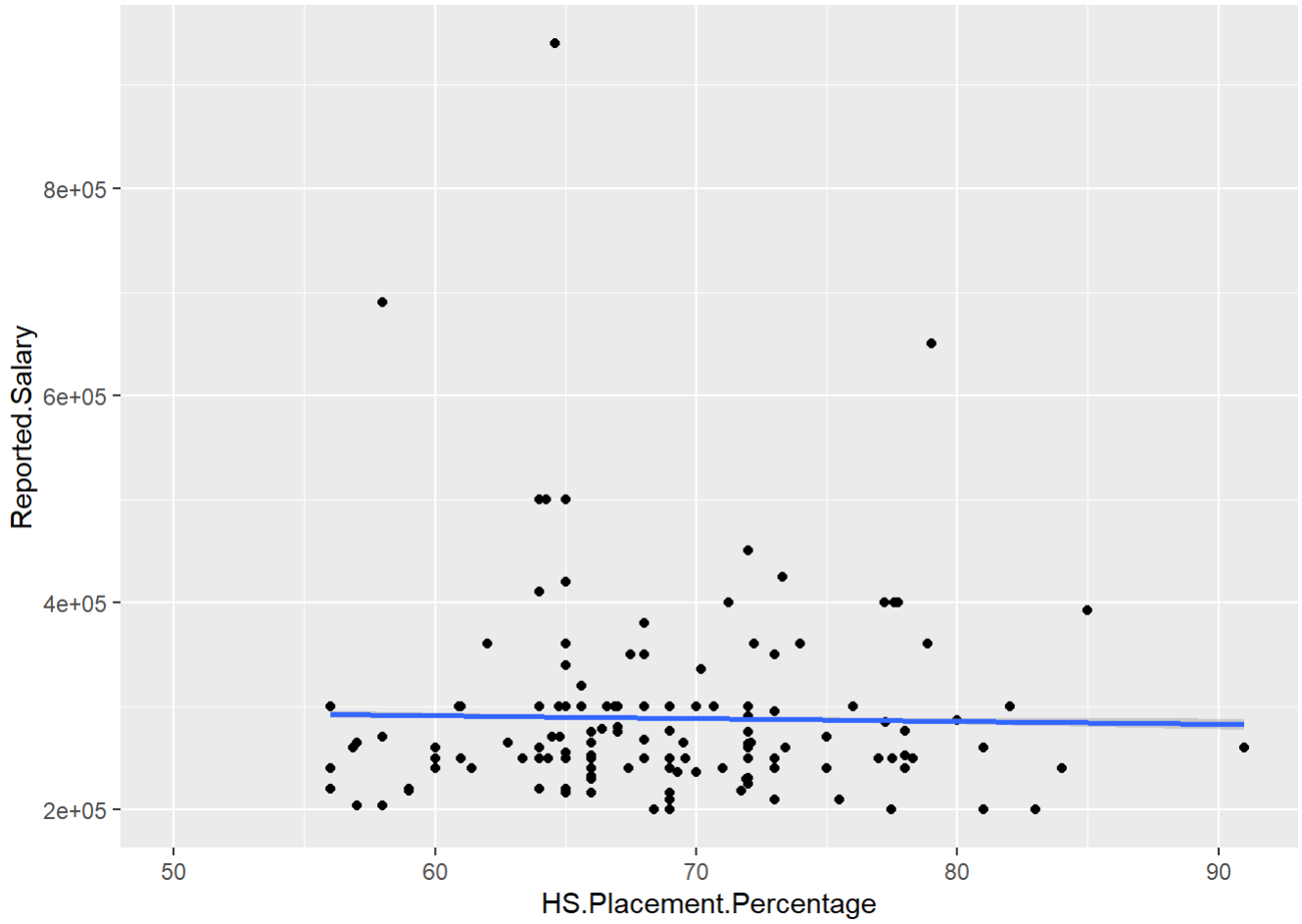
```
LM040<-lm(formula=A$HS.Placement.Percentage ~ A$Reported.Salary, data=A)
summary(LM040)
```

```
##
## Call:
## lm(formula = A$HS.Placement.Percentage ~ A$Reported.Salary, data = A)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.8328  -3.8026  -0.7253   3.7275  22.2209
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)        6.913e+01  1.747e-01 395.606   <2e-16 ***
## A$Reported.Salary -1.344e-06  5.761e-07  -2.333   0.0197 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.495 on 14650 degrees of freedom
##   (6633 observations deleted due to missingness)
## Multiple R-squared:  0.0003714,  Adjusted R-squared:  0.0003032
## F-statistic: 5.443 on 1 and 14650 DF,  p-value: 0.01966
```

```
plotM <- ggplot(data = A, aes(x = HS.Placement.Percentage, y =Reported.Salary)) + geom_point()
plotM + geom_smooth(formula = y ~ x, method = "lm")
```

```
## Warning: Removed 6633 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 6633 rows containing missing values (geom_point).
```




```
#Here there is a negative, yet small correlation between the two variables, with r=-.19, and the r squared value is non-existent.
```

```
cor.test(A$College.Placement.Percentage, A$Reported.Salary)
```

```
##
##  Pearson's product-moment correlation
##
## data:  A$College.Placement.Percentage and A$Reported.Salary
## t = 21.933, df = 14650, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1625845 0.1939396
## sample estimates:
##       cor
## 0.1783073
```

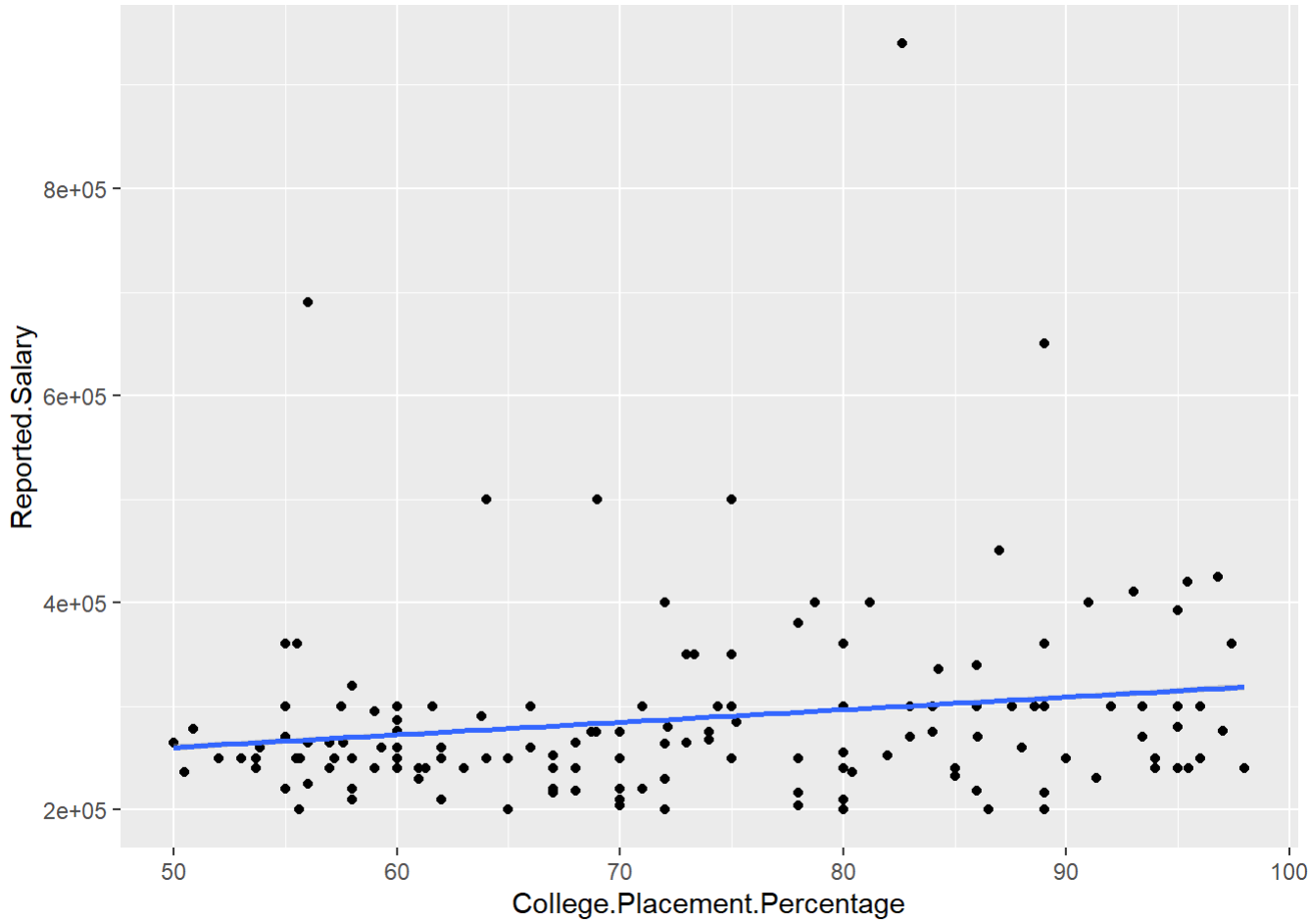
```
LM050<-lm(formula=A$College.Placement.Percentage ~ A$Reported.Salary, data=A)
summary(LM050)
```

```
##
## Call:
## lm(formula = A$College.Placement.Percentage ~ A$Reported.Salary,
##     data = A)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -27.751 -12.291  -1.099  10.629  26.036
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       6.568e+01  3.622e-01  181.31   <2e-16 ***
## A$Reported.Salary 2.619e-05  1.194e-06   21.93   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.46 on 14650 degrees of freedom
##   (6633 observations deleted due to missingness)
## Multiple R-squared:  0.03179,    Adjusted R-squared:  0.03173
## F-statistic: 481.1 on 1 and 14650 DF,  p-value: < 2.2e-16
```

```
plotN <- ggplot(data = A, aes(x = College.Placement.Percentage, y =Reported.Salary)) + geom_poin
t()
plotN + geom_smooth(formula = y ~ x, method = "lm")
```

```
## Warning: Removed 6633 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 6633 rows containing missing values (geom_point).
```




```
#Here there is a negative, yet small correlation between the two variables, with r=-.18, and the r squared value is non-existent.
```

```
cor.test(A$Graduate.Level.Placement.Percentage, A$Reported.Salary)
```

```
##
##  Pearson's product-moment correlation
##
## data:  A$Graduate.Level.Placement.Percentage and A$Reported.Salary
## t = 21.515, df = 14650, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1592721 0.1906648
## sample estimates:
##       cor
## 0.1750129
```

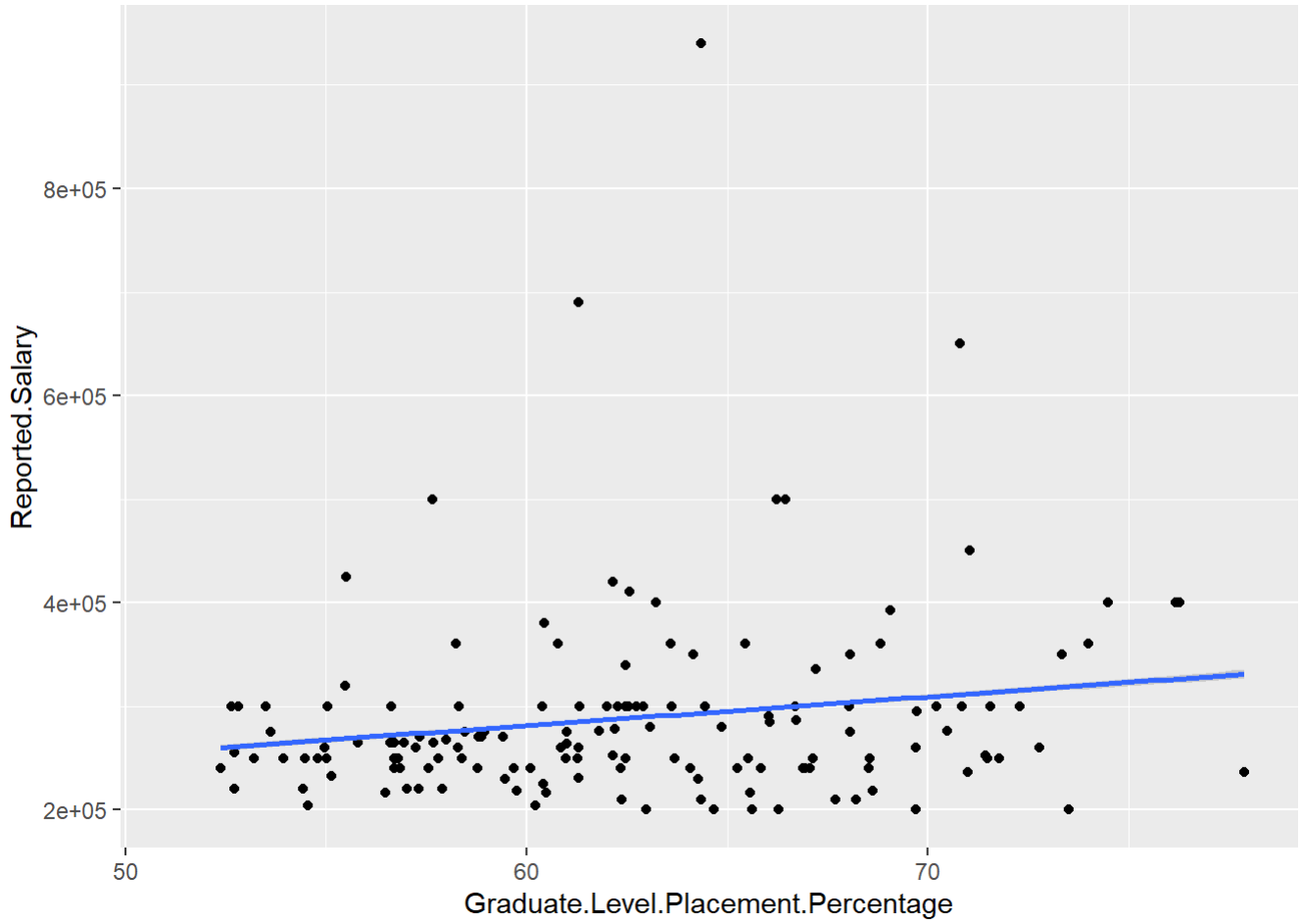
```
LM060<-lm(formula=A$Graduate.Level.Placement.Percentage ~ A$Reported.Salary, data=A)
summary(LM060)
```

```
##
## Call:
## lm(formula = A$Graduate.Level.Placement.Percentage ~ A$Reported.Salary,
##     data = A)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.064  -4.670  -0.637   4.181  15.891
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       5.940e+01  1.554e-01  382.35   <2e-16 ***
## A$Reported.Salary 1.102e-05  5.122e-07   21.52   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.775 on 14650 degrees of freedom
##   (6633 observations deleted due to missingness)
## Multiple R-squared:  0.03063,    Adjusted R-squared:  0.03056
## F-statistic: 462.9 on 1 and 14650 DF,  p-value: < 2.2e-16
```

```
plotO <- ggplot(data = A, aes(x = Graduate.Level.Placement.Percentage, y =Reported.Salary)) + ge
om_point()
plotO + geom_smooth(formula = y ~ x, method = "lm")
```

```
## Warning: Removed 6633 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 6633 rows containing missing values (geom_point).
```




```
#Here there is a positive correlation between the two variables, with r=.18, and a very small am
ount of explained variance with no practical significance.
```

```
LM1<-lm(formula = A$Reported.Salary ~  A$College.Placement.Percentage + A$Graduate.Level.Placeme
nt.Percentage, data=A)
summary(LM1)
```

```
##
## Call:
## lm(formula = A$Reported.Salary ~ A$College.Placement.Percentage +
##     A$Graduate.Level.Placement.Percentage, data = A)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -118960  -50675  -14831   13408  638591
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         84483.50    8247.16   10.24   <2e-16 ***
## A$College.Placement.Percentage        952.12      57.22   16.64   <2e-16 ***
## A$Graduate.Level.Placement.Percentage  2148.31     133.49   16.09   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 90860 on 14649 degrees of freedom
##   (6633 observations deleted due to missingness)
## Multiple R-squared:  0.04861,    Adjusted R-squared:  0.04848
## F-statistic: 374.3 on 2 and 14649 DF,  p-value: < 2.2e-16
```

```
#Means of Groups
library(dplyr)
aggregate(A$SSC.Placement.Test.Percentage.10th.Grade, by=list(A$Gender), FUN=mean)
```

```
##   Group.1        x
## 1       F 68.31092
## 2       M 66.75252
```

```
#Gender<- c("Female","Male")
#Mean <- c(68.31092, 66.75252)
#DF1<- data.frame(Gender, Mean, stringsAsFactors=TRUE)
#DF1
#barplot(DF1)
```

```
aggregate(A$HSC.12th.Grade.Percentage, by=list(A$Gender), FUN=mean)
```

```
##   Group.1        x
## 1       F 66.64684
## 2       M 66.16165
```

```
aggregate(A$HS.Placement.Percentage, by=list(A$Gender), FUN=mean)
```

```
##   Group.1        x
## 1       F 68.09000
## 2       M 65.42986
```

```
aggregate(A$College.Placement.Percentage, by=list(A$Gender), FUN=mean)
```

```
##   Group.1        x
## 1       F 70.59066
## 2       M 72.92612
```

```
aggregate(A$Graduate.Level.Placement.Percentage, by=list(A$Gender), FUN=mean)
```

```
##   Group.1        x
## 1       F 64.64355
## 2       M 60.98489
```

```
#aggregate(A$Reported.Salary, by=list(A$Gender), FUN=mean) #Unsure how to fix this?
```

```
#By placement Track
aggregate(A$SSC.Placement.Test.Percentage.10th.Grade, by=list(A$HS.Track), FUN=mean)
```

```
##     Group.1        x
## 1 Comm&Mgmt 65.95559
## 2    Others 64.22727
## 3  Sci&Tech 71.18932
```

```
aggregate(A$HSC.12th.Grade.Percentage, by=list(A$HS.Track), FUN=mean)
```

```
##     Group.1        x
## 1 Comm&Mgmt 67.18538
## 2    Others 60.18182
## 3  Sci&Tech 65.38559
```

```
aggregate(A$HS.Placement.Percentage, by=list(A$HS.Track), FUN=mean)
```

```
##     Group.1        x
## 1 Comm&Mgmt 66.20310
## 2    Others 60.66636
## 3  Sci&Tech 67.84424
```

```
aggregate(A$College.Placement.Percentage, by=list(A$HS.Track), FUN=mean)
```

```
##     Group.1        x
## 1 Comm&Mgmt 71.98524
## 2    Others 72.49455
## 3  Sci&Tech 72.31051
```

```
aggregate(A$College.Placement.Percentage, by=list(A$HS.Track), FUN=mean)
```

```
##     Group.1        x
## 1 Comm&Mgmt 71.98524
## 2    Others 72.49455
## 3  Sci&Tech 72.31051
```

```
#By College Track
aggregate(A$SSC.Placement.Test.Percentage.10th.Grade, by=list(A$College.Track), FUN=mean)
```

```
##   Group.1        x
## 1 Mkt&Fin 68.96167
## 2  Mkt&HR 65.20874
```

```
aggregate(A$HSC.12th.Grade.Percentage, by=list(A$College.Track), FUN=mean)
```

```
##   Group.1        x
## 1 Mkt&Fin 68.67058
## 2  Mkt&HR 63.38063
```

```
aggregate(A$College.Placement.Percentage, by=list(A$College.Track), FUN=mean)
```

```
##   Group.1        x
## 1 Mkt&Fin 74.88550
## 2  Mkt&HR 68.58274
```

```
aggregate(A$Graduate.Level.Placement.Percentage, by=list(A$College.Track), FUN=mean)
```

```
##   Group.1        x
## 1 Mkt&Fin 62.82567
## 2  Mkt&HR 61.58663
```

```
#By 10th Grade Location
aggregate(A$SSC.Placement.Test.Percentage.10th.Grade, by=list(A$SSC.10th.Grade.Location), FUN=me
an)
```

```
##   Group.1        x
## 1 Central 66.14388
## 2  Others 68.66202
```

```
aggregate(A$HSC.12th.Grade.Percentage, by=list(A$SSC.10th.Grade.Location), FUN=mean)
```

```
##   Group.1        x
## 1 Central 67.70931
## 2  Others 64.72071
```

```
aggregate(A$College.Placement.Percentage, by=list(A$SSC.10th.Grade.Location), FUN=mean)
```

```
##   Group.1        x
## 1 Central 72.33293
## 2  Others 71.82828
```

```
aggregate(A$Graduate.Level.Placement.Percentage, by=list(A$SSC.10th.Grade.Location), FUN=mean)
```

```
##   Group.1        x
## 1 Central 61.83129
## 2  Others 62.80182
```

```
#By 12th Grade Location
aggregate(A$SSC.Placement.Test.Percentage.10th.Grade, by=list(A$HSC.12th.Grade.Location), FUN=me
an)
```

```
##   Group.1        x
## 1 Central 66.39964
## 2  Others 67.88290
```

```
aggregate(A$HSC.12th.Grade.Percentage, by=list(A$HSC.12th.Grade.Location), FUN=mean)
```

```
##   Group.1        x
## 1 Central 66.59857
## 2  Others 66.16298
```

```
aggregate(A$HS.Placement.Percentage, by=list(A$HSC.12th.Grade.Location), FUN=mean)
```

```
##   Group.1        x
## 1 Central 65.75381
## 2  Others 66.76542
```

```
aggregate(A$College.Placement.Percentage, by=list(A$HSC.12th.Grade.Location), FUN=mean)
```

```
##   Group.1        x
## 1 Central 71.45369
## 2  Others 72.51534
```

```
aggregate(A$Graduate.Level.Placement.Percentage, by=list(A$HSC.12th.Grade.Location), FUN=mean)
```

```
##   Group.1        x
## 1 Central 61.62262
## 2  Others 62.69855
```

```
#By Placement
aggregate(A$SSC.Placement.Test.Percentage.10th.Grade, by=list(A$Placement.Status), FUN=mean)
```

```
##      Group.1        x
## 1 Not Placed 57.54403
## 2     Placed 71.72149
```

```
#Far Apart
```

```
aggregate(A$HSC.12th.Grade.Percentage, by=list(A$Placement.Status), FUN=mean)
```

```
##      Group.1        x
## 1 Not Placed 58.39552
## 2     Placed 69.92655
```

```
aggregate(A$HS.Placement.Percentage, by=list(A$Placement.Status), FUN=mean)
```

```
##      Group.1        x
## 1 Not Placed 61.13418
## 2     Placed 68.74054
```

```
#A bit closer
```

```
aggregate(A$College.Placement.Percentage, by=list(A$Placement.Status), FUN=mean)
```

```
##      Group.1        x
## 1 Not Placed 69.58791
## 2     Placed 73.23804
```

```
#Even closer, and more competitive at this level
```

```
aggregate(A$Graduate.Level.Placement.Percentage, by=list(A$Placement.Status), FUN=mean)
```

```
##      Group.1        x
## 1 Not Placed 61.61284
## 2     Placed 62.57939
```

```
#This is really close, really, really close.
```