College-Scorecard

Nanditha Sundararajan

12/07/2018

1. Combining the data for 2009-2013

```r
college_09_10 <- col_09_10 %>%
  mutate(GRAD_DEBT_MDN = as.numeric(as.character(GRAD_DEBT_MDN)),
         MD_EARN_WNE_P8 = as.numeric(as.character(MD_EARN_WNE_P8)),
         "Year" = "2009-10") %>%
  mutate(DEBT_TO_EARN = GRAD_DEBT_MDN/MD_EARN_WNE_P8)


college_11_12 <- col_11_12 %>%
  mutate(GRAD_DEBT_MDN = as.numeric(as.character(GRAD_DEBT_MDN)),
         MD_EARN_WNE_P8 = as.numeric(as.character(MD_EARN_WNE_P8)),
         "Year" = "2011-12") %>%
  mutate(DEBT_TO_EARN = GRAD_DEBT_MDN/MD_EARN_WNE_P8)


college_12_13 <- col_12_13 %>%
  mutate(GRAD_DEBT_MDN = as.numeric(as.character(GRAD_DEBT_MDN)),
         MD_EARN_WNE_P8 = as.numeric(as.character(MD_EARN_WNE_P8)),
         "Year" = "2012-13") %>%
  mutate(DEBT_TO_EARN = GRAD_DEBT_MDN/MD_EARN_WNE_P8)


college_13_14 <- col_13_14 %>%
  mutate(GRAD_DEBT_MDN = as.numeric(as.character(GRAD_DEBT_MDN)),
         MD_EARN_WNE_P8 = as.numeric(as.character(MD_EARN_WNE_P8)),
         "Year" = "2013-14") %>%
  mutate(DEBT_TO_EARN = GRAD_DEBT_MDN/MD_EARN_WNE_P8)


college_14_15 <- col_14_15 %>%
  mutate(GRAD_DEBT_MDN = as.numeric(as.character(GRAD_DEBT_MDN)),
         MD_EARN_WNE_P8 = as.numeric(as.character(MD_EARN_WNE_P8)),
         "Year" = "2014-15") %>%
  mutate(DEBT_TO_EARN = GRAD_DEBT_MDN/MD_EARN_WNE_P8)


college_09_15 <- rbind(college_09_10, college_11_12, college_12_13,
                       college_13_14, college_14_15)
```

2. Partitioning the data into training and test sets

```r
college_09_15_parts <- resample_partition(college_09_15,
                                          c(train = 0.6, valid = 0.2, test = 0.2))

college_09_15_train_ <- as_tibble(college_09_15_parts$train)
college_09_15_test_  <- as_tibble(college_09_15_parts$test)
college_09_15_valid_ <- as_tibble(college_09_15_parts$valid)
```

3. Subsetting variables:

```r
college_09_15_train <- college_09_15_train_ %>%
  select(COMPL_RPY_3YR_RT, GRAD_DEBT_MDN,PCTFLOAN, PCTPELL,MD_EARN_WNE_P8,
         COSTT4_A, Year, DEBT_TO_EARN, MD_FAMINC) %>%
  mutate(COMPL_RPY_3YR_RT = as.numeric(as.character(COMPL_RPY_3YR_RT)),
         GRAD_DEBT_MDN = as.numeric(as.character(GRAD_DEBT_MDN)),
         MD_EARN_WNE_P8 = as.numeric(as.character(MD_EARN_WNE_P8)),
         COSTT4_A = as.numeric(as.character(COSTT4_A)),
         PCTFLOAN = as.numeric(as.character(PCTFLOAN)),
         PCTPELL = as.numeric(as.character(PCTPELL)),
         MD_FAMINC = as.numeric(as.character(MD_FAMINC)))
```

Correlation Coefficients and their respective graphs

```r
college_09_15_train <- college_09_15_train %>%
  select(COMPL_RPY_3YR_RT, GRAD_DEBT_MDN,
         PCTFLOAN, PCTPELL, MD_EARN_WNE_P8,
         COSTT4_A,Year, DEBT_TO_EARN,
         MD_FAMINC)


college_09_15_2<-college_09_15_train %>%
  filter(Year == "2013-14")
```

```r
#PCTPELL vs 3 YR Repayment Rate

r1<-cor.test(college_09_15_2$PCTPELL,
             college_09_15_2$COMPL_RPY_3YR_RT,
             method = "pearson",
             conf.level = 0.95)
r1
```
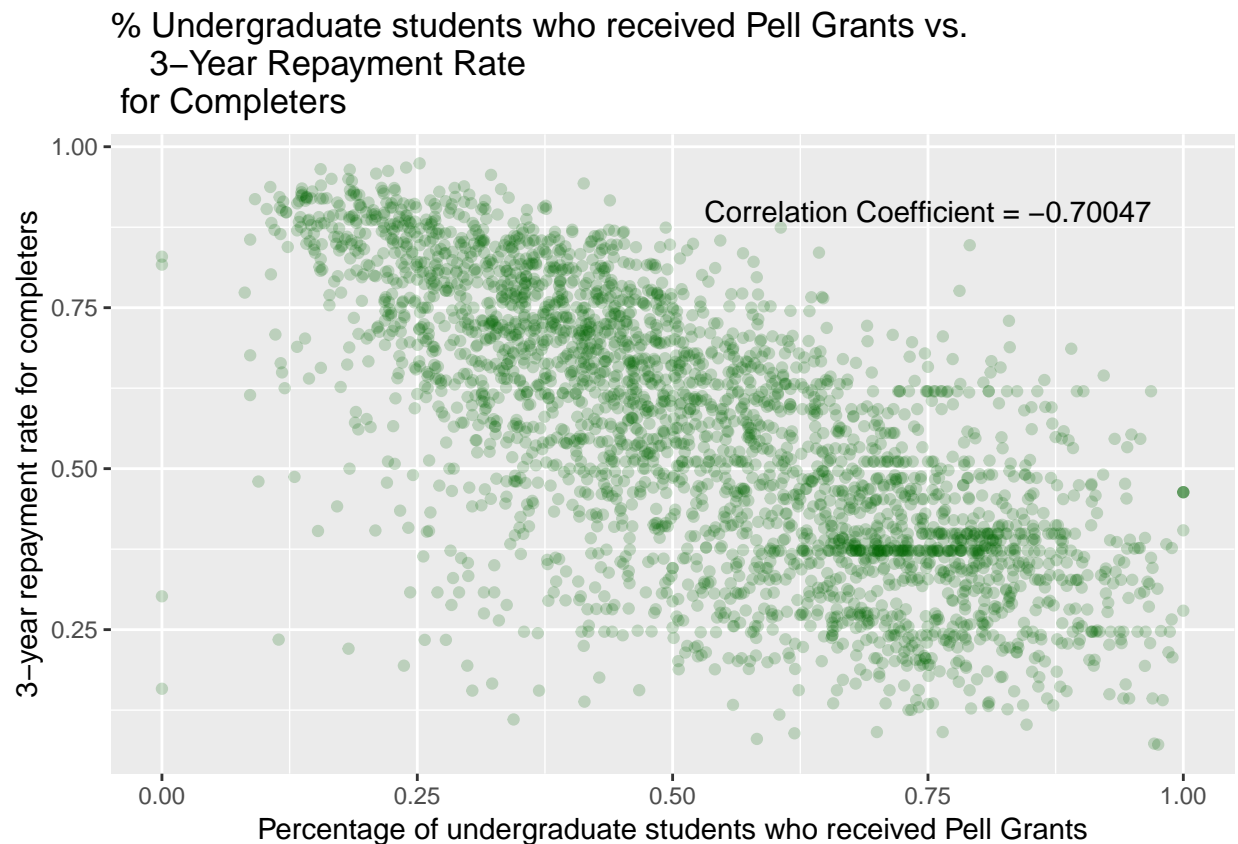
```
##
##  Pearson's product-moment correlation
##
## data:  college_09_15_2$PCTPELL and college_09_15_2$COMPL_RPY_3YR_RT
## t = -53.483, df = 2984, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.7174574 -0.6808056
## sample estimates:
##        cor
## -0.6995914
```

```r
college_09_15_2 %>%
  ggplot() +
  geom_point(aes(x = PCTPELL, y = COMPL_RPY_3YR_RT),
    color= "darkgreen", alpha = 0.2)+
  labs(title="% Undergraduate students who received Pell Grants vs.
```

```
    3-Year Repayment Rate \n for Completers",
    x="Percentage of undergraduate students who received Pell Grants",
    y="3-year repayment rate for completers")+
  annotate("text",
           x=0.75, y=0.9,
           label="Correlation Coefficient = -0.70047")
```

% Undergraduate students who received Pell Grants vs. 3−Year Repayment Rate for Completers
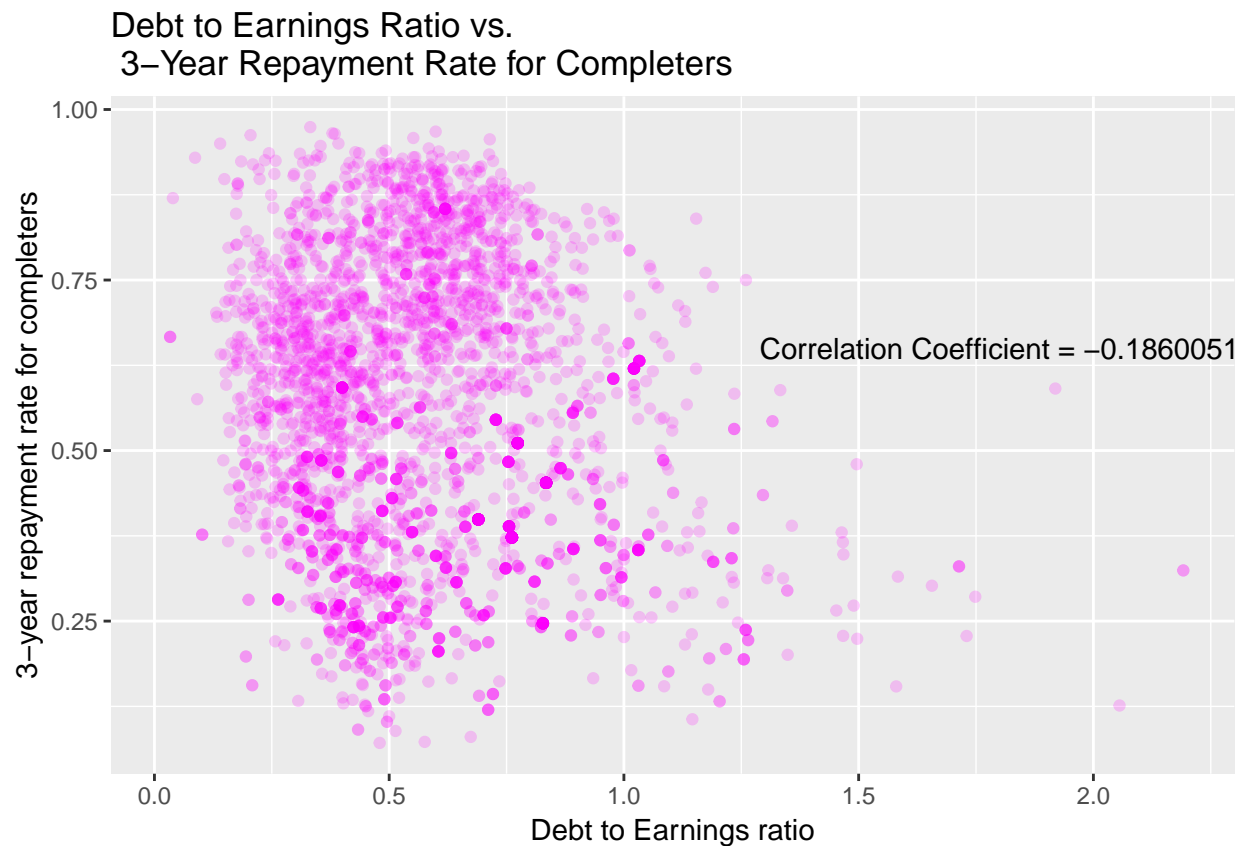


```
#DEBT-EARNINGS vs 3 yr repayment rate:
r2<-cor.test(college_09_15_2$DEBT_TO_EARN,
             college_09_15_2$COMPL_RPY_3YR_RT,
             method = "pearson",
             conf.level = 0.95)
r2
```

```
##
##  Pearson's product-moment correlation
##
## data:  college_09_15_2$DEBT_TO_EARN and college_09_15_2$COMPL_RPY_3YR_RT
## t = -9.2106, df = 3077, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.1979800 -0.1292246
## sample estimates:
##        cor
## -0.1638012
```

```
college_09_15_2 %>%
  ggplot() +
  geom_point(aes(x = DEBT_TO_EARN, y = COMPL_RPY_3YR_RT),
    color= "magenta", alpha = 0.2)+
  labs(title="Debt to Earnings Ratio vs. \n 3-Year Repayment Rate for Completers",
    x="Debt to Earnings ratio",
    y="3-year repayment rate for completers") +
  annotate("text", x=1.8, y=0.65, label="Correlation Coefficient = -0.1860051")
```


Debt to Earnings Ratio vs.
3−Year Repayment Rate for Completers

```
#COST4A vs 3 Yr repayment rates:
r3<-cor.test(college_09_15_2$COSTT4_A,
          college_09_15_2$COMPL_RPY_3YR_RT,
          method = "pearson",
          conf.level = 0.95)
r3
```

```
##
##  Pearson's product-moment correlation
##
## data:  college_09_15_2$COSTT4_A and college_09_15_2$COMPL_RPY_3YR_RT
## t = 13.666, df = 2064, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2480199 0.3271250
## sample estimates:
```

```
##         cor
## 0.2880638
```

```
college_09_15_2 %>%
  ggplot() +
  geom_point(aes(x = COSTT4_A, y = COMPL_RPY_3YR_RT),
    color= "#660099", alpha = 0.2 )+
  xlim(c(0, 120000))+
  labs(title="Cost of Attending the Institute vs.\n 3-Year Repayment Rate for Completers",
    x="Cost of attending the institute",
    y="3-year repayment rate for completers")+
  annotate("text",x=95000, y=0.65, label="Correlation Coefficient = 0.3033664")
```
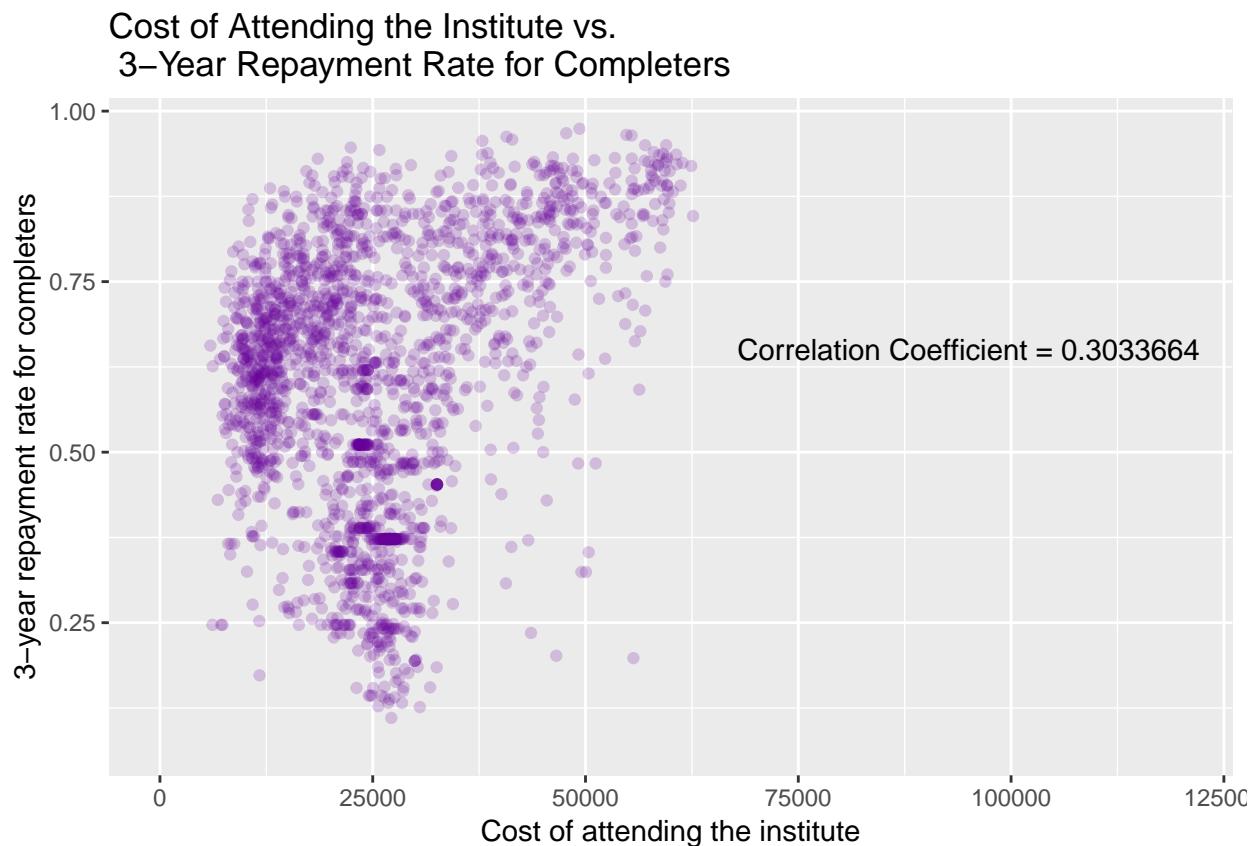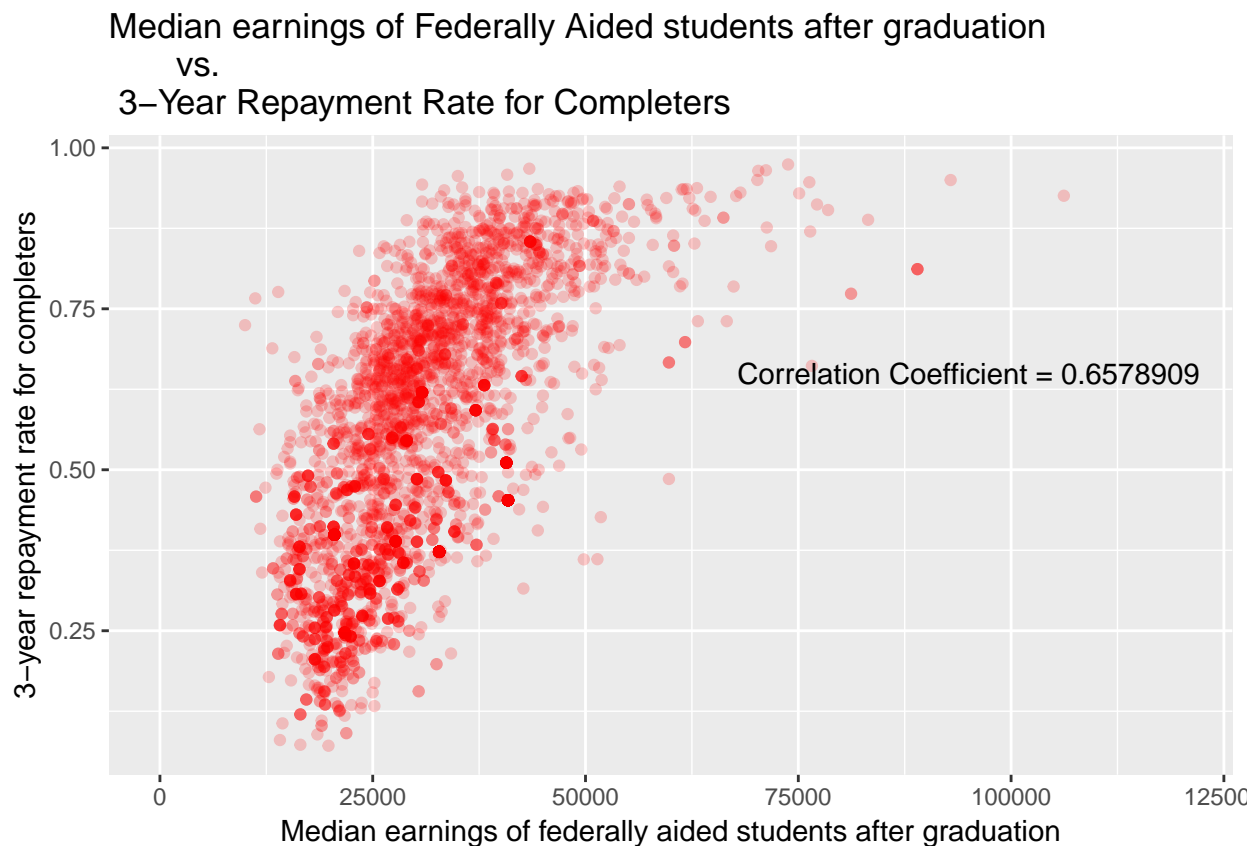


```
# Median Earnings vs 3 YR Repayment Rate:
r4<-corr_coeff<-cor.test(college_09_15_2$MD_EARN_WNE_P8,
                         college_09_15_2$COMPL_RPY_3YR_RT,
                         method = "pearson",
                         conf.level = 0.95)
r4
```

```
##
##  Pearson's product-moment correlation
##
## data:  college_09_15_2$MD_EARN_WNE_P8 and college_09_15_2$COMPL_RPY_3YR_RT
## t = 48.019, df = 3104, p-value < 2.2e-16
```

```
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6322090 0.6725898
## sample estimates:
##       cor
## 0.652863
```

```
college_09_15_2 %>%
  ggplot() +
  geom_point(aes(x = MD_EARN_WNE_P8, y = COMPL_RPY_3YR_RT),
    color="red", alpha = 0.2)+
  xlim(c(0, 120000))+
  labs(title="Median earnings of Federally Aided students after graduation
       vs.\n 3-Year Repayment Rate for Completers",
    x="Median earnings of federally aided students after graduation",
    y="3-year repayment rate for completers")+
  annotate("text", x=95000, y=0.65, label="Correlation Coefficient = 0.6578909")
```


Median earnings of Federally Aided students after graduation
vs.
3−Year Repayment Rate for Completers

```
#Median Family Income vs 3 YR Repayment Rate:
r5<-cor.test(college_09_15_2$MD_FAMINC,
             college_09_15_2$COMPL_RPY_3YR_RT,
             method = "pearson",
             conf.level = 0.95)
r5
```

```
##
```

```
##  Pearson's product-moment correlation
##
## data:  college_09_15_2$MD_FAMINC and college_09_15_2$COMPL_RPY_3YR_RT
## t = 68.8, df = 3221, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7570417 0.7850232
## sample estimates:
##       cor
## 0.7714051
```

```
college_09_15_2 %>%
  ggplot() +
  geom_point(aes(x =MD_FAMINC, y = COMPL_RPY_3YR_RT),
    color="blue", alpha = 0.2) +
  xlim(c(0, 120000))+
  labs(title="Median Family Income vs.3-Year Repayment Rate for Completers",
    x="Median Family Income",
    y="3-year repayment rate for completers")+
  annotate("text",x=95000, y=0.65,label="Correlation Coefficient = 0.7743165")
```


Median Family Income vs.3−Year Repayment Rate for Completers
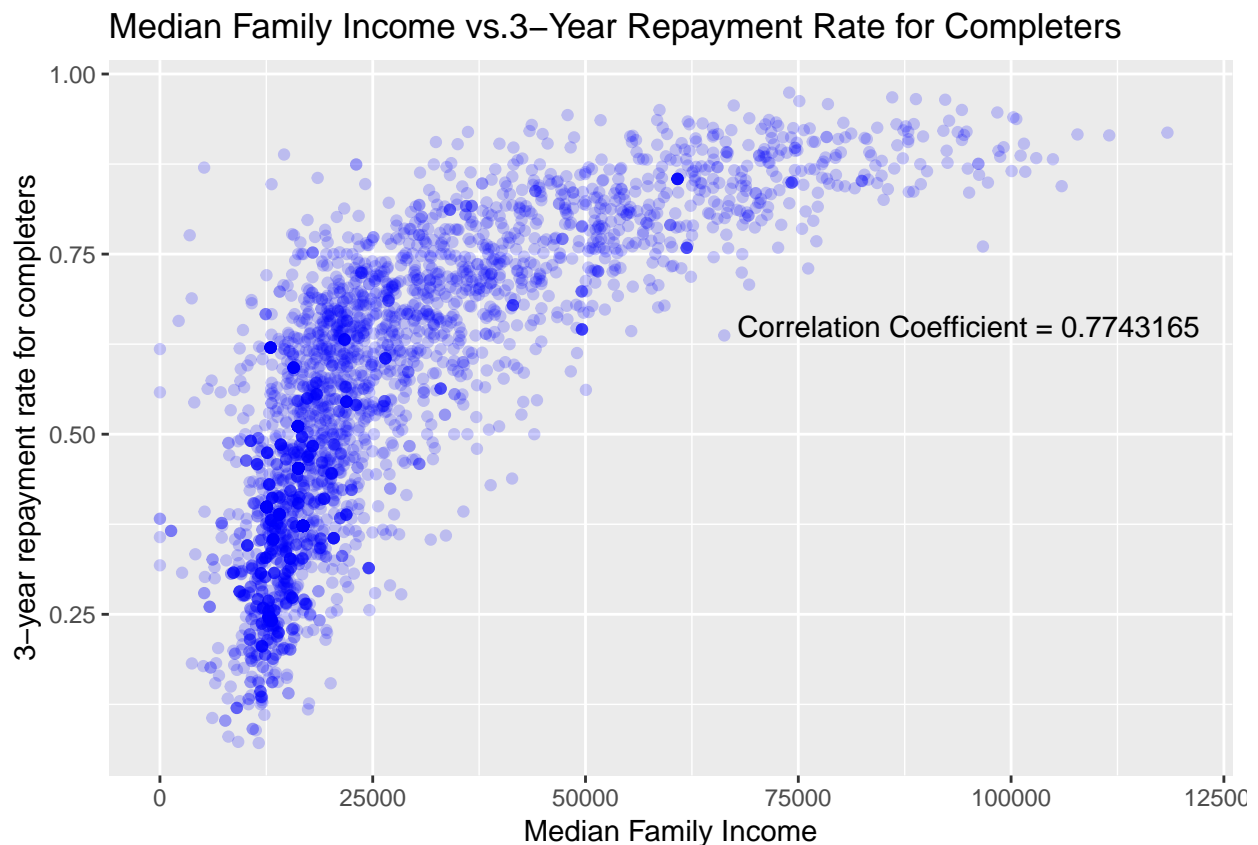
```
# PCTFLOAN vs 3 YR Repayment Rate:

r6<-cor.test(college_09_15_2$PCTFLOAN,
            college_09_15_2$COMPL_RPY_3YR_RT,
```

```
            method = "pearson",
            conf.level = 0.95)
r6
```

```
##
##  Pearson's product-moment correlation
##
## data:  college_09_15_2$PCTFLOAN and college_09_15_2$COMPL_RPY_3YR_RT
## t = -13.07, df = 2984, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.2663477 -0.1984871
## sample estimates:
##        cor
## -0.2327007
```

```r
college_09_15_2 %>%
  ggplot() +
  geom_point(aes(x = PCTFLOAN, y = COMPL_RPY_3YR_RT),
    color="#E69F00" , alpha = 0.2 )+
  labs(title="Percent of all undergraduate students receiving a federal student
    loan vs.\n 3-Year Repayment Rate for Completers",
    x="Percent of all undergraduate students receiving a federal student loan",
    y="3-year repayment rate for completers") +
  annotate("text", x=0.19, y=0.95,label="Correlation Coefficient = -0.2236313 ")
```


Percent of all undergraduate students receiving a federal student loan vs. 3−Year Repayment Rate for Completers