EDA Report code

```r
col_08_09 <- read.csv("../mfrubenstein/Downloads/CollegeScorecard_Raw_Data/MERGED2008_09_PP.csv")
college_08_09 <- col_08_09 %>%
  mutate("Year" = "2008-09")

col_09_10 <- read.csv("../mfrubenstein/Downloads/CollegeScorecard_Raw_Data/MERGED2009_10_PP.csv")
college_09_10 <- col_09_10 %>%
  mutate("Year" = "2009-10")

col_10_11 <- read.csv("../mfrubenstein/Downloads/CollegeScorecard_Raw_Data/MERGED2010_11_PP.csv")
college_10_11 <- col_10_11 %>%
  mutate("Year" = "2010-11")

col_11_12 <- read.csv("../mfrubenstein/Downloads/CollegeScorecard_Raw_Data/MERGED2011_12_PP.csv")
college_11_12 <- col_11_12 %>%
  mutate("Year" = "2011-12")

col_12_13 <- read.csv("../mfrubenstein/Downloads/CollegeScorecard_Raw_Data/MERGED2012_13_PP.csv")
college_12_13 <- col_12_13 %>%
  mutate("Year" = "2012-13")

col_13_14 <- read.csv("../mfrubenstein/Downloads/CollegeScorecard_Raw_Data/MERGED2013_14_PP.csv")
college_13_14 <- col_13_14 %>%
  mutate("Year" = "2013-14")

col_14_15 <- read.csv("../mfrubenstein/Downloads/CollegeScorecard_Raw_Data/MERGED2014_15_PP.csv")
college_14_15 <- col_14_15 %>%
  mutate("Year" = "2014-15")

college_08_13 <- rbind(college_08_09, college_09_10, college_10_11, college_11_12, college_12_13, colleg
```

```
## Warning in `[<-.factor`(`*tmp*`, ri, value = c(100200L, 105200L,
## 2503400L, : invalid factor level, NA generated

## Warning in `[<-.factor`(`*tmp*`, ri, value = c(100200L, 105200L,
## 2503400L, : invalid factor level, NA generated

## Warning in `[<-.factor`(`*tmp*`, ri, value = c(100200L, 105200L,
## 2503400L, : invalid factor level, NA generated

## Warning in `[<-.factor`(`*tmp*`, ri, value = c(100200L, 105200L,
## 2503400L, : invalid factor level, NA generated
```

```r
collegeuse <- select(college_08_13, Year, INSTNM, CITY, STABBR, REGION, ICLEVEL, LATITUDE, LONGITUDE, CU

collegeuse <- filter(collegeuse, CONTROL %in% c("1", "2"), ICLEVEL == "1")
collegeuse <- mutate(collegeuse, CONTROL = recode(CONTROL, "1" = "Public","2" = "Private"))

#3 Year Repayment Completers vs. Non-Completers
repay <- select(collegeuse, INSTNM, CONTROL, Year, STABBR, COMPL_RPY_3YR_RT, NONCOM_RPY_3YR_RT, LO_INC_
repay <- within(repay, {
        COMPL_RPY_3YR_RT <- as.numeric(as.character(COMPL_RPY_3YR_RT))
```

```
            NONCOM_RPY_3YR_RT <- as.numeric(as.character(NONCOM_RPY_3YR_RT))
            LO_INC_RPY_3YR_RT <- as.numeric(as.character(LO_INC_RPY_3YR_RT))
            MD_INC_RPY_3YR_RT <- as.numeric(as.character(MD_INC_RPY_3YR_RT))
            HI_INC_RPY_3YR_RT <- as.numeric(as.character(HI_INC_RPY_3YR_RT))
            PELL_RPY_3YR_RT <- as.numeric(as.character(PELL_RPY_3YR_RT))
            })
```

```
## Warning in eval(substitute(expr), e): NAs introduced by coercion
```

```
## Warning in eval(substitute(expr), e): NAs introduced by coercion
```

```
## Warning in eval(substitute(expr), e): NAs introduced by coercion
```

```
## Warning in eval(substitute(expr), e): NAs introduced by coercion
```

```
## Warning in eval(substitute(expr), e): NAs introduced by coercion
```

```
## Warning in eval(substitute(expr), e): NAs introduced by coercion
```

```
repay <- summarise(group_by(repay, Year, STABBR, CONTROL, INSTNM),
                   COMPL_RPY_3YR_RT = mean(COMPL_RPY_3YR_RT, na.rm = TRUE),
                   NONCOM_RPY_3YR_RT = mean(NONCOM_RPY_3YR_RT, na.rm = TRUE),
                   LO_INC_RPY_3YR_RT = mean(LO_INC_RPY_3YR_RT, na.rm = TRUE),
                   MD_INC_RPY_3YR_RT = mean(MD_INC_RPY_3YR_RT, na.rm = TRUE),
                   HI_INC_RPY_3YR_RT = mean(HI_INC_RPY_3YR_RT, na.rm = TRUE),
                   PELL_RPY_3YR_RT = mean(PELL_RPY_3YR_RT, na.rm = TRUE))

repay <- filter(repay, STABBR == "MA")
repay<- mutate(repay, INSTNM = recode(INSTNM, "Massachusetts Institute of Technology" = "MIT", "Universi

repay <- filter(repay, INSTNM %in% c("MIT","Harvard University", "Brandeis University", "Boston College"
repay <- filter(repay, Year != "2008-09")

repayoverall <- select(repay, STABBR, CONTROL, INSTNM, Year, COMPL_RPY_3YR_RT, NONCOM_RPY_3YR_RT)
repayoverall <- gather(repayoverall, "COMPL_RPY_3YR_RT", "NONCOM_RPY_3YR_RT", key = "type", value = "nu
repay <- gather(repay, "LO_INC_RPY_3YR_RT", "MD_INC_RPY_3YR_RT", "HI_INC_RPY_3YR_RT", key = "type", valu

repay <- mutate(repay, type = recode(type, "LO_INC_RPY_3YR_RT" = "Low Income", "MD_INC_RPY_3YR_RT" = "Me

repay$type <- factor(repay$type, levels = c("Low Income", "Medium Income", "High Income"))
ggplot(data= repay) + geom_col(data = repay, mapping = aes(x= INSTNM, y = number, fill = type), position
```
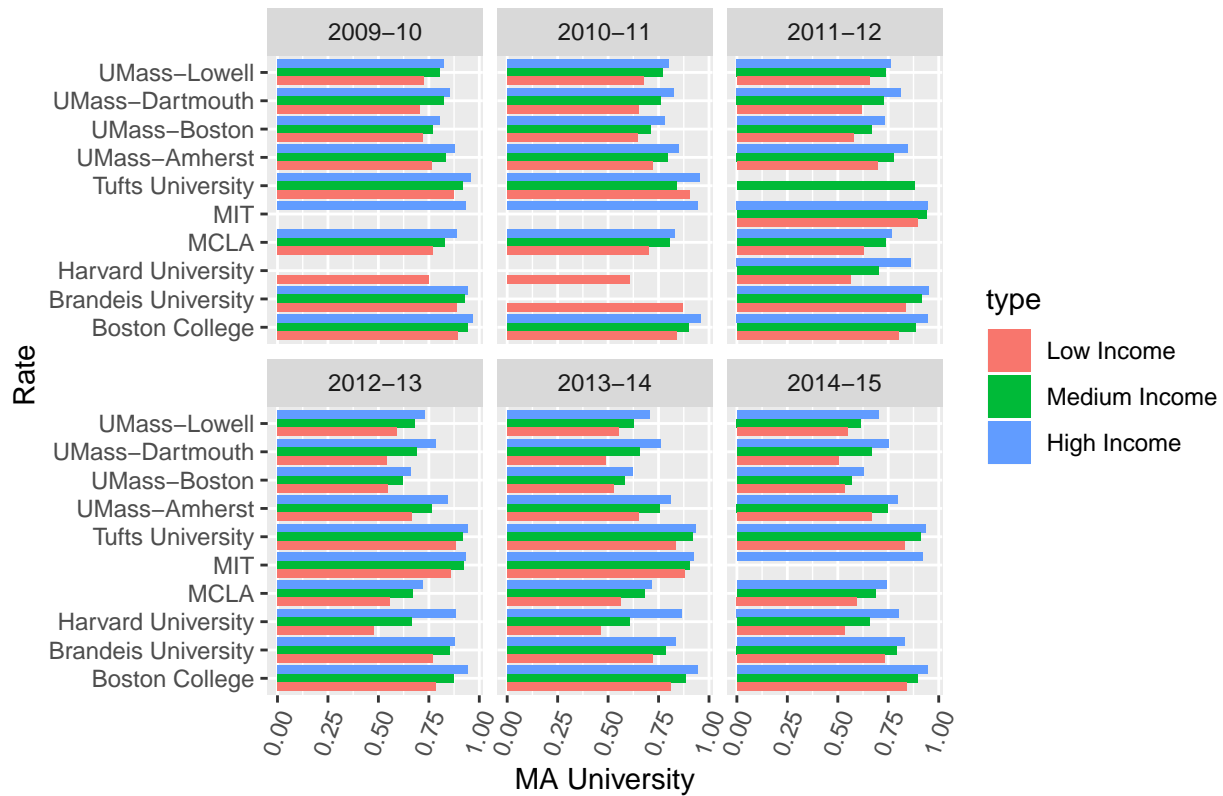
```
## Warning: Removed 14 rows containing missing values (geom_col).
```
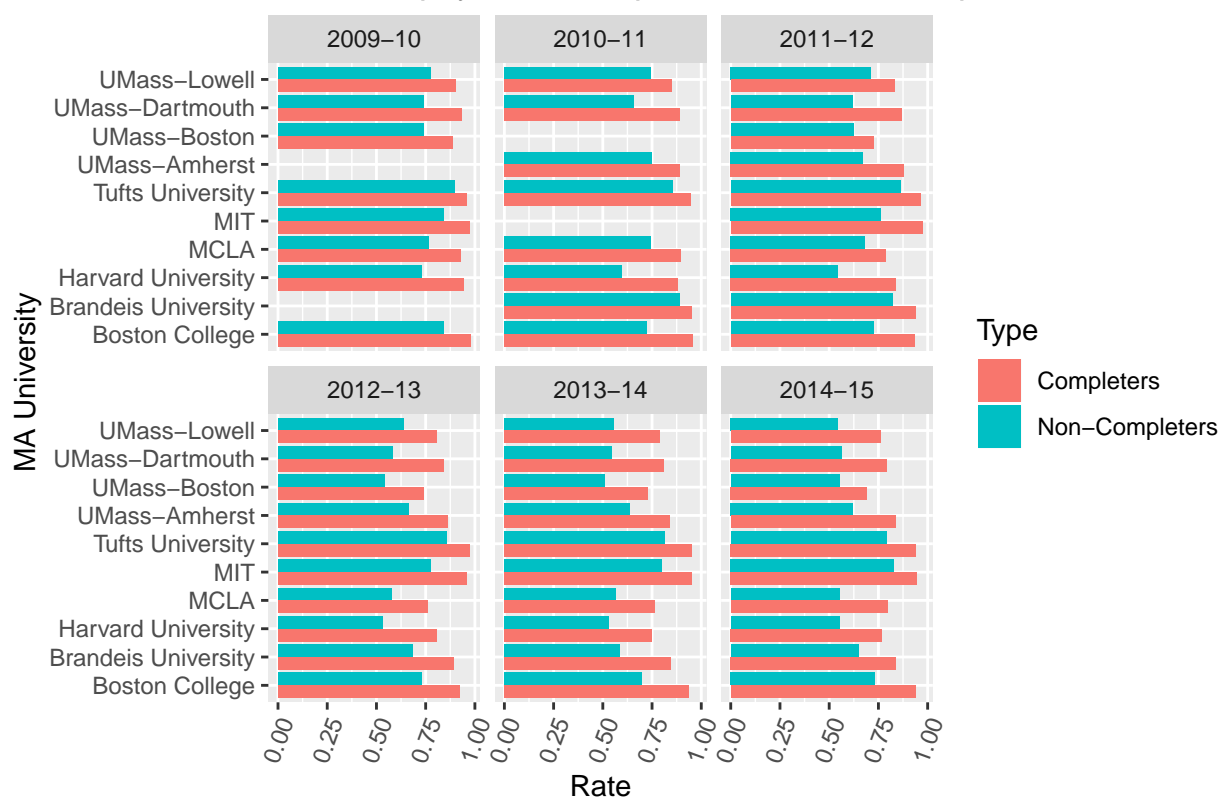
3 Year Repayment Rate by Family Income


```r
ggplot(data= repayoverall) + geom_col(data = repayoverall, mapping = aes(x= INSTNM, y = number, fill =
```

```
## Warning: Removed 8 rows containing missing values (geom_col).
```

3 Year Repayment Completers vs. Non-Completers

```
repaystate <- select(collegeuse, INSTNM, CONTROL, Year, STABBR, COMPL_RPY_3YR_RT, NONCOM_RPY_3YR_RT, LO_
repaystate <- within(repaystate, {
        COMPL_RPY_3YR_RT <- as.numeric(as.character(COMPL_RPY_3YR_RT))
        NONCOM_RPY_3YR_RT <- as.numeric(as.character(NONCOM_RPY_3YR_RT))
        LO_INC_RPY_3YR_RT <- as.numeric(as.character(LO_INC_RPY_3YR_RT))
        MD_INC_RPY_3YR_RT <- as.numeric(as.character(MD_INC_RPY_3YR_RT))
        HI_INC_RPY_3YR_RT <- as.numeric(as.character(HI_INC_RPY_3YR_RT))
        PELL_RPY_3YR_RT <- as.numeric(as.character(PELL_RPY_3YR_RT))
        })
```

```
## Warning in eval(substitute(expr), e): NAs introduced by coercion
```

```
## Warning in eval(substitute(expr), e): NAs introduced by coercion
```

```
## Warning in eval(substitute(expr), e): NAs introduced by coercion
```

```
## Warning in eval(substitute(expr), e): NAs introduced by coercion
```

```
## Warning in eval(substitute(expr), e): NAs introduced by coercion
```

```
## Warning in eval(substitute(expr), e): NAs introduced by coercion
```

```
repaystate <- summarise(group_by(repaystate, Year, STABBR, CONTROL, INSTNM),
                COMPL_RPY_3YR_RT = mean(COMPL_RPY_3YR_RT, na.rm = TRUE),
                NONCOM_RPY_3YR_RT = mean(NONCOM_RPY_3YR_RT, na.rm = TRUE),
                LO_INC_RPY_3YR_RT = mean(LO_INC_RPY_3YR_RT, na.rm = TRUE),
                MD_INC_RPY_3YR_RT = mean(MD_INC_RPY_3YR_RT, na.rm = TRUE),
                HI_INC_RPY_3YR_RT = mean(HI_INC_RPY_3YR_RT, na.rm = TRUE),
```

```
                  PELL_RPY_3YR_RT = mean(PELL_RPY_3YR_RT, na.rm = TRUE))

repaystate <- filter(repaystate, Year == "2014-15")
repaystate <- gather(repaystate, "COMPL_RPY_3YR_RT", "NONCOM_RPY_3YR_RT", key = "type", value = "number

ggplot(data= repaystate) + geom_col(data = repaystate, mapping = aes(x= STABBR, y = number, fill = type
```

```
## Warning: Removed 1108 rows containing missing values (geom_col).
```


3 Year Repayment Completers vs. Non–Completers Across States, 2014–