

# Project 1

*Aaron Niskin, Charlie Edelson, Chris Leonard*

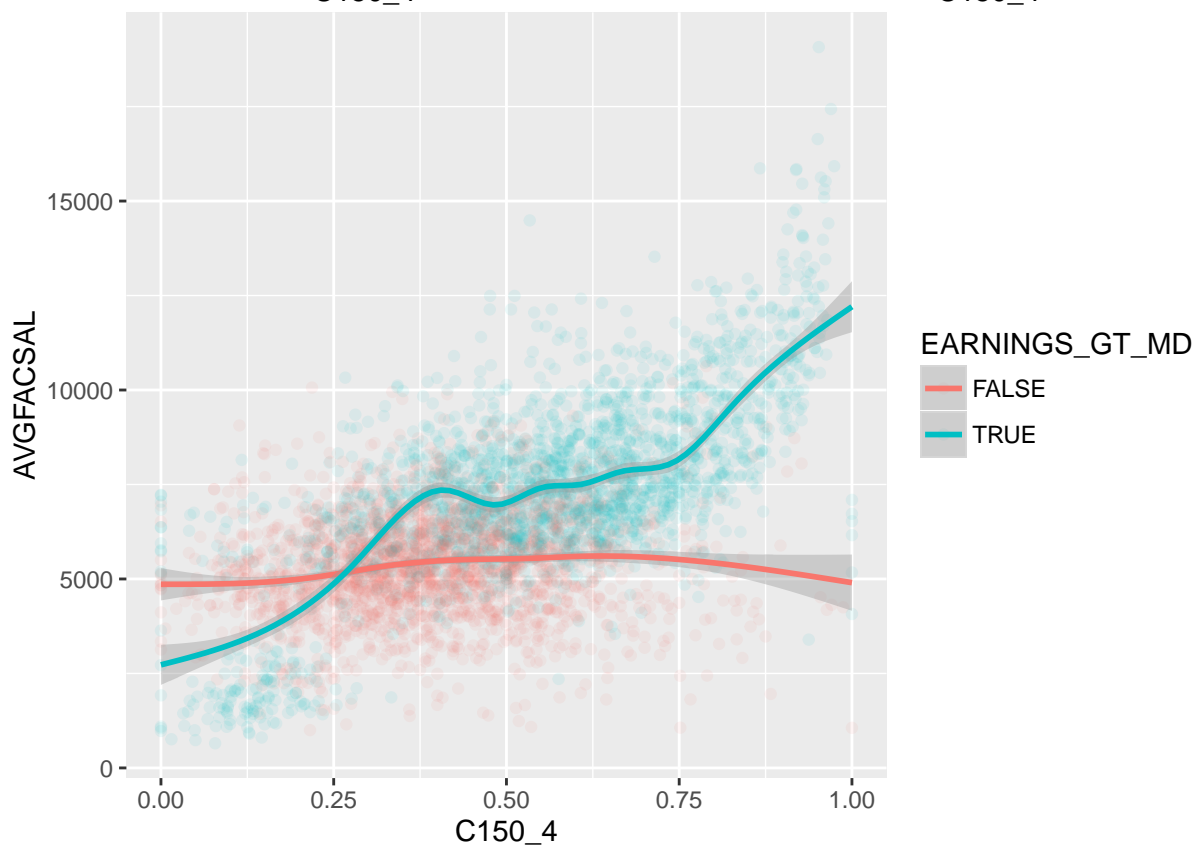
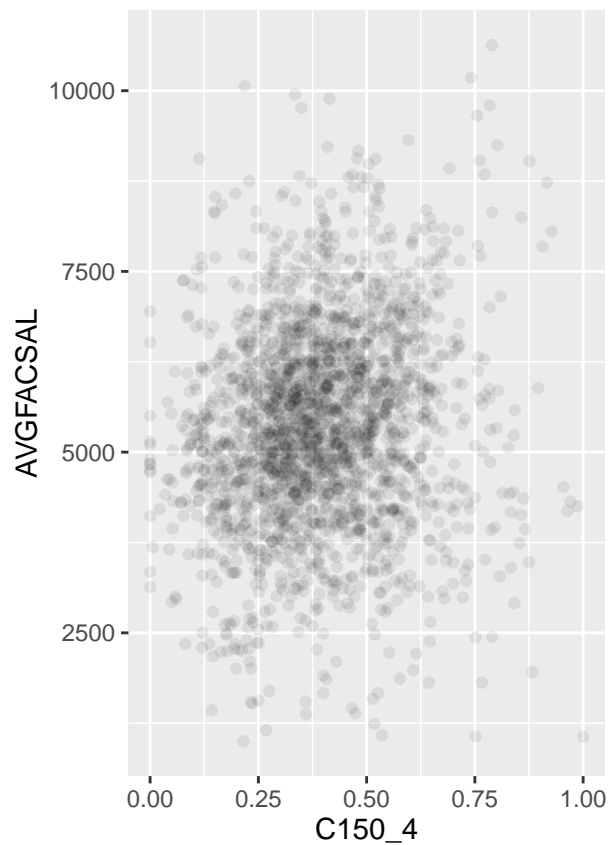
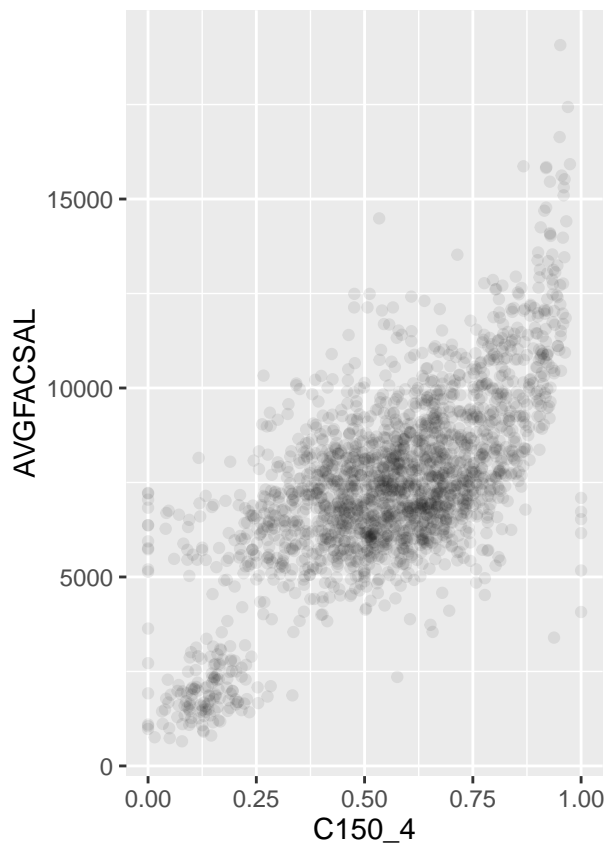
*October 26, 2016*

```
library(ggplot2)
library(gridExtra)
setwd("~/Documents/courses/newCollege/current/eda/projects/project1")
```

```
getFilesColumns <- function(years, cols, nullVals) {
  r_formatting_sucks <- function(n) {
    if(n < 10) {
      return(paste("0", n, sep=""))
    }
    else {
      return(n)
    }
  }
  getFileName <- function(startYear) {
    R_is_stupid <- paste("ignore/MERGED", startYear, "_", r_formatting_sucks((startYear + 1) %% 100), ".csv")
    return(R_is_stupid)
  }
  getThisFile <- function(year) {
    tmp <- read.csv(getFileName(year), na=nullVals, stringsAsFactors = FALSE)
    tmp <- tmp[,cols]
    tmp$startYear = year
    return(tmp)
  }
  tmpDf <- getThisFile(years[1])
  for (year in years[2:length(years)]){
    tmpDf <- rbind(tmpDf, getThisFile(year))
  }
  tmpDf$startYear <- as.factor(tmpDf$startYear)
  return(tmpDf)
}
```

```
dataStartYears <- 2010:2014
csvNullVals = c("NULL", "PrivacySuppressed")
desiredCols = c("FAMINC", "AVGFACsAL", "C150_4", "MD_EARN_WNE_P10")

DF = getFilesColumns(dataStartYears, desiredCols, csvNullVals)
rm(getFilesColumns)
rm(dataStartYears)
rm(csvNullVals)
rm(desiredCols)
```



```
## Warning: Removed 555 rows containing missing values (geom_point).
```

