Methods

Read in the data.

```
hate_df = read_csv("./data/HateCrimes.csv")
```

```
## Parsed with column specification:
## cols(
##    state = col_character(),
##    unemployment = col_character(),
##    urbanization = col_character(),
##    median_household_income = col_double(),
##    perc_population_with_high_school_degree = col_double(),
##    perc_non_citizen = col_double(),
##    gini_index = col_double(),
##    perc_non_white = col_double(),
##    hate_crimes_per_100k_splc = col_character()
## )
```

Jacy's ideas:

```
##Project ideas
hate = read.csv("/Users/jacysparks/Downloads/HateCrimes.csv")
head(hate)
dim(hate)
hate$hate_crimes_per_100k_splc = as.character(hate$hate_crimes_per_100k_splc)
hate$hate_crimes_per_100k_splc = as.numeric(hate$hate_crimes_per_100k_splc)
summary(hate)
##Four NA's for outcome
##NA for Wyoming, South Dakota, North Dakota, and Idaho
hate[,c(1,9)]
##Could remove
hate = na.omit(hate)

##3 NA's for non citizen


##Create indicators
names(hate)[names(hate)=="unemployment"] = "High.Unemployment"
names(hate)[names(hate)=="urbanization"] = "High.Urban"
names(hate)[names(hate)=="median_household_income"] = "Med.Income"
names(hate)[names(hate)=="perc_population_with_high_school_degree"] = "HS.Degree"
names(hate)[names(hate)=="perc_non_citizen"] = "Non.Citizen"
names(hate)[names(hate)=="perc_non_white"] = "Non.White"
names(hate)[names(hate)=="hate_crimes_per_100k_splc"] = "Hate.Crime"
hate$High.Unemployment = ifelse(hate$High.Unemployment=="high",1,0)
hate$High.Urban = ifelse(hate$High.Urban=="high",1,0)
```

```r
##Outcome var is skewed
hate$Hate.Crime = log(hate$Hate.Crime)
hist(hate$Hate.Crime)
##Much better

reg = lm(Hate.Crime~.-state,data=hate)
summary(reg)

pairs(hate[,4:9],lower.panel=NULL)
cor(hate[,4:9])
#Percent white and percent non-white highly correlated


##Check linearity
for(i in 4:8){
  plot(hate[,i],hate$Hate.Crime,main=colnames(hate)[i])
}
plot(reg)
```