Recidivism - MA415/615 Final Project

Julia Pan, Eesha Pendharkar

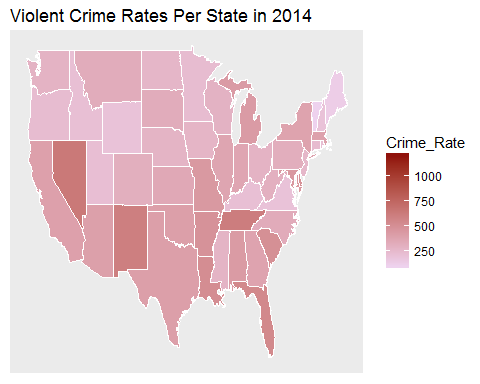
## Introduction

In this report we will use data from Broward County, Florida to examine how COMPAS, a risk assessment instrument widely used in the US, predicts recidivism (tendency to reoffend) among offenders. The COMPAS questionnaire collects information on offenders, including past criminal offenses, substance abuse, family history, social connections, etc. to produce risk scores. These risk scores are categorized into three sections: Low, Medium, and High Risk. We will look at the rates of recidivism among each of these three groups across various ethnicities to determine the efficacy of this instruments for different ethnic groups.

Additionally, we will perform a textual and sentiment analysis using various sources about the meaning of recidivism and why it is a concern in the criminal justice world. We have visualized the most commonly used words to capture the difference in sentiment and tone by the government vs the media.

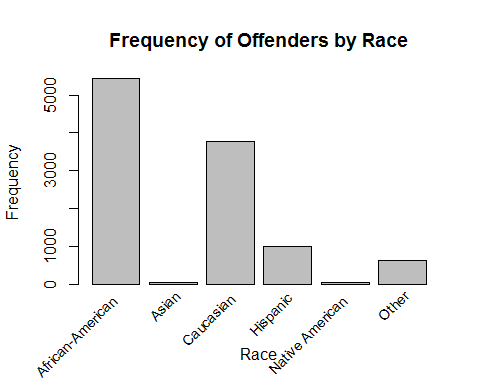
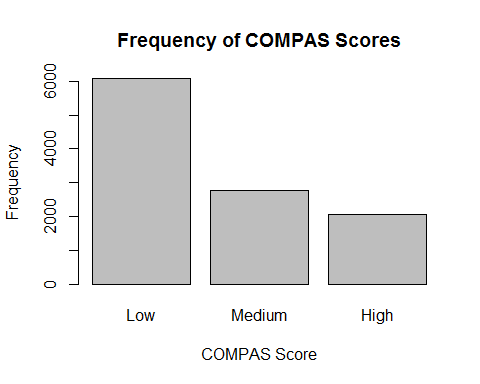
## Crime Rates in the US

A major incentive for the development of risk assessment models and algorithms in the criminal justice system was the over-saturated penal system in the US. The map below shows the rate of violent crime in each state in 2014. Darker hues indicate higher crime rates. The Shiny app will demonstrate average crime rates over a selected time period, for a wider range of crimes.



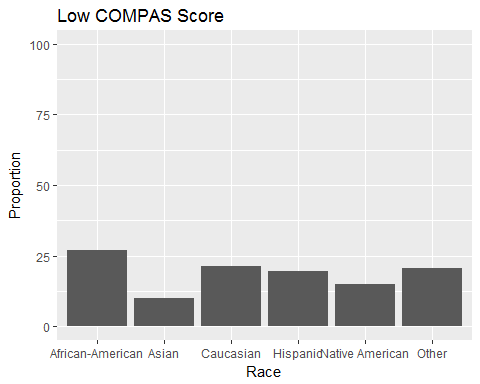
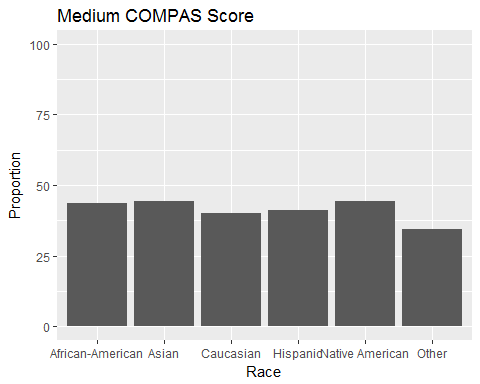
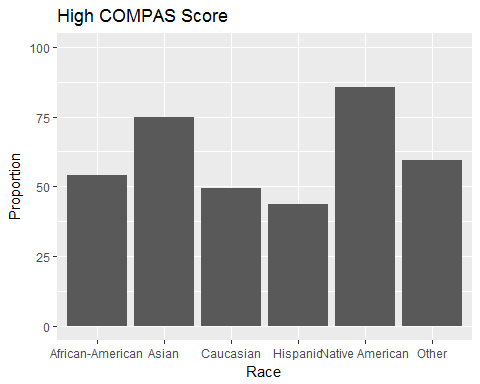
## Population Summary

Now, we will examine recidivism rates in Broward County, Florida. Before we begin our analysis, let's look at general summary information to frame our conclusions. In this dataset, we are examining 10,892 individuals. Of these, 3,652 did recidivate. The charts below demonstrate the demographics of this population by score category and race.



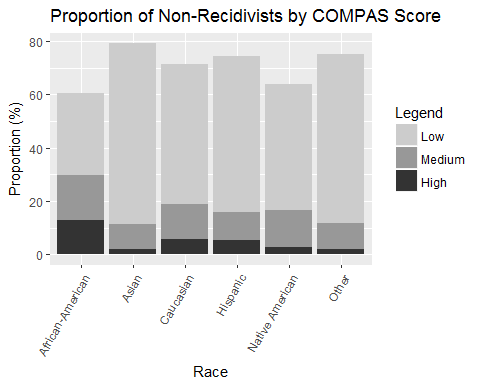
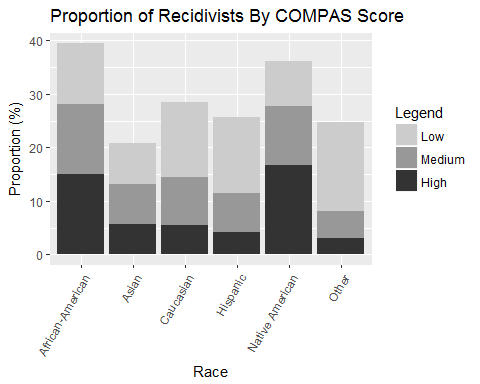
## Recidivists by Risk Score

Next, let's look at the breakdown of each risk score category by race. We can see that there is a high percentage of Asians and Native Americans who received high risk scores, though the low representation of these ethnicities in this dataset should be recognized. Seeing as African-Americans make up a majority of this dataset, the high correlation between high risk score and high rate of recidivism is significant. For medium COMPAS scores, the proportion of recidivists is relatively constant across ethnicities. Though Asians have a lower rate of recidivists among Low-risk individuals, they again make up a small portion of the total population in this dataset.



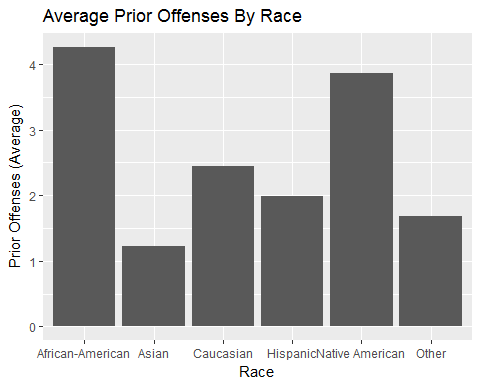
### Comparing Recidivists and Non-Recidivists

Here, we will look at the proportion of recidivists and non-recidivists through stacked bar charts. In the first chart, we can see that African-Americans and Native-Americans have the highest total rate of recidivism as well as the most individuals with High risk scores. More compelling, however, is the second chart which shows risk scores for non-recidivists. It is clear that African-Americans receive high risk scores more frequently than any other ethnicity for offenders who do not recidivate. This could indicate a racial bias in the risk assessment instrument.



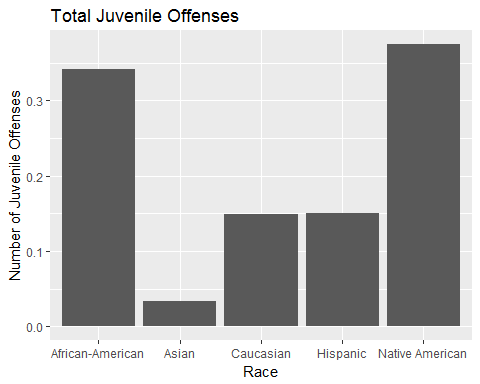
## Criminal History

After looking at general trends in recidivists by score category and race, let's take a closer look at the criminal backgrounds of these individuals to get a better understanding of their risk scores and recidivism rates. Looking at the chart of average prior offenses below, we can see that both African-Americans and Native-Americans have high numbers of prior offenses. This conclusion is consistent with the disproportionate number of African-Americans in the US criminal justice system who are incarcerated and targeted through arguably biased policing methods. Additionally, Native-Americans are another disenfranchised population in the US demonstrating high crime rates.



## Juvenile Offenses

Continuing on from prior offenses, let's also look at juvenile offenses. We can see again that African-Americans and Native-Americans are the two populations with the highest number of juvenile offenses in this population. This is indicative of inadequate support systems for minority populations that begins at adolescence and continues into adulthood.

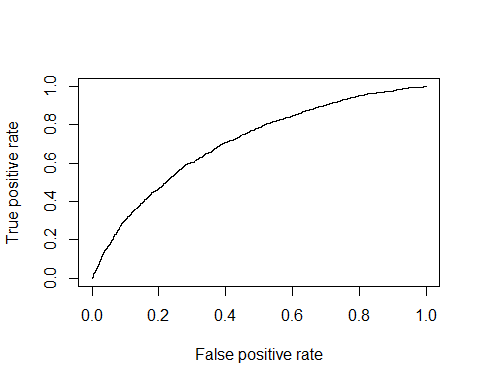


## Logistic Regression

To supplement our graphical exploration of recidivism, we created a logistic regression to predict recidivism using variables such as the decile COMPAS score, number of priors, age, race, and sex. Below, the model summary is shown as well as the ANOVA table and ROC curve. We can see that the AUC is 0.708.

##   
## Call:  
## glm(formula = is\_recid ~ ., family = binomial, data = test)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4026 -0.8950 -0.6487 1.1542 2.2180   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.659123 0.157799 -4.177 2.95e-05 \*\*\*  
## decile\_score 0.107586 0.013695 7.856 3.97e-15 \*\*\*  
## priors\_count 0.092589 0.008198 11.294 < 2e-16 \*\*\*  
## age -0.028435 0.003292 -8.638 < 2e-16 \*\*\*  
## raceAsian -0.524531 0.517337 -1.014 0.310627   
## raceCaucasian -0.048394 0.070974 -0.682 0.495336   
## raceHispanic -0.148642 0.117820 -1.262 0.207092   
## raceNative American 0.345253 0.571563 0.604 0.545810   
## raceOther -0.107363 0.149061 -0.720 0.471361   
## sexMale 0.280422 0.078254 3.583 0.000339 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 6755.7 on 5222 degrees of freedom  
## Residual deviance: 6140.4 on 5213 degrees of freedom  
## AIC: 6160.4  
##   
## Number of Fisher Scoring iterations: 4

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: is\_recid  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)   
## NULL 5222 6755.7   
## decile\_score 1 423.00 5221 6332.7 < 2.2e-16 \*\*\*  
## priors\_count 1 94.82 5220 6237.9 < 2.2e-16 \*\*\*  
## age 1 81.38 5219 6156.5 < 2.2e-16 \*\*\*  
## race 5 3.02 5214 6153.5 0.6970869   
## sex 1 13.09 5213 6140.4 0.0002973 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

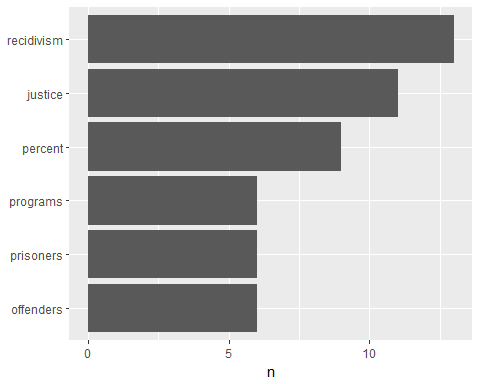


## [1] 0.7076298

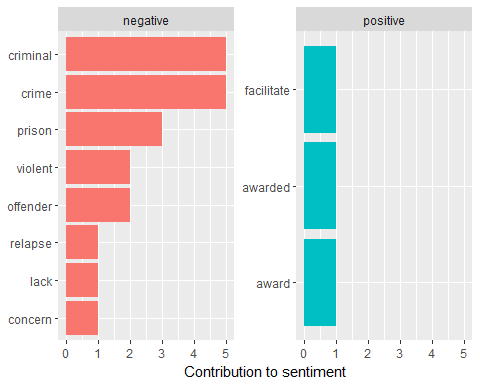
## Recidivism: Dataframe, wordcloud and charts

Here we analyzed documents from the National Institute of Justice about recidivism that explain what it means and why it is a concern in the criminal justice world. We also did sentiment analysis on the above documents and an article from the Tampa Bay Times to visualize the primary words used and the sentiment of the government vs the media.

## Joining, by = "word"

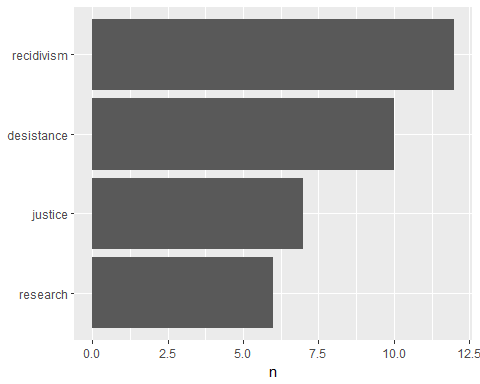
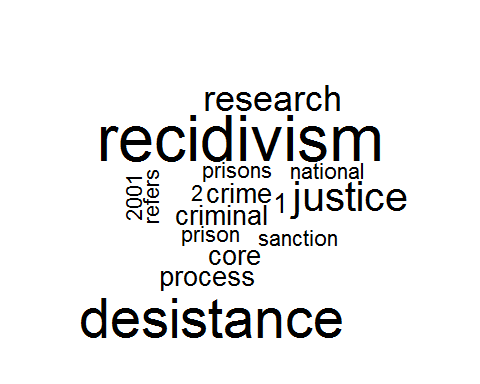


## Selecting by n

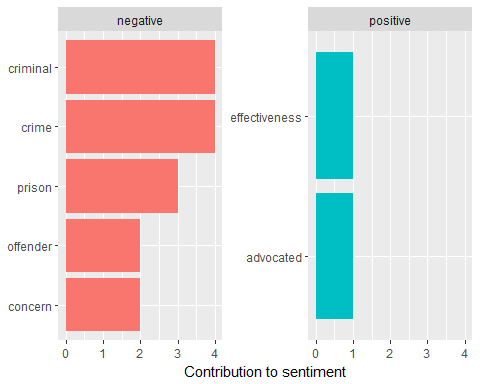


## Why is recidivism a core criminal justice system concern?

## Joining, by = "word"

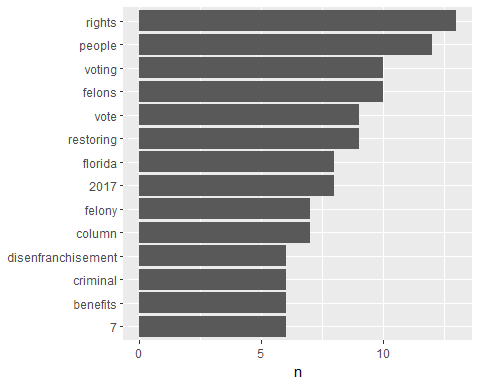
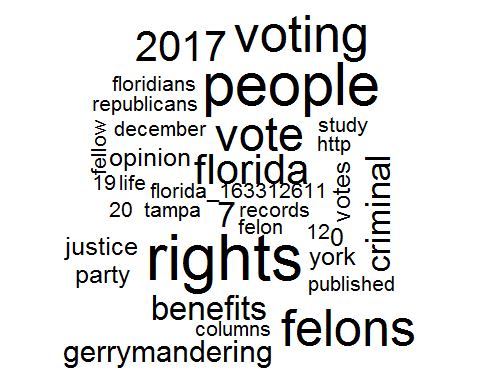


## Selecting by n

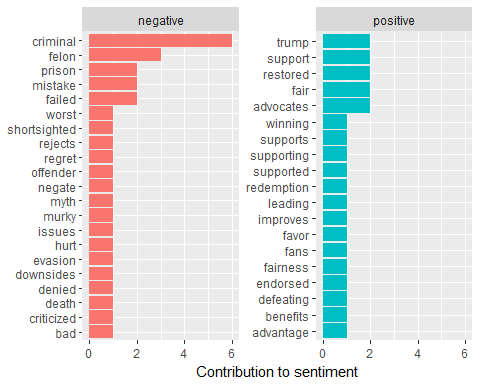


## Tampa Bay Times: Sentiment Analysis

## Joining, by = "word"



## Selecting by n



## Summary

Thus we saw through the analysis that the contributing words here are very different from those that contributed to sentiment in the official NIJ documents

However, the number of negative words in the article was much more, although "Trump" was considered a positive word even though it was a proper noun and referred to the President

## Twitter

Here we pulled 100 tweets about recidivism from Twitter and turned them into a data frame.

## [1] "Using direct authentication"

## [1] 1

## [[1]]  
## [1] "ToomeyWright: @austalgia Solution: Ensure they go to school by fining parents if they don't; swim/footy after school and a decent https://t.co/rZ5otTuo3z"  
##   
## [[2]]  
## [1] "Sonofabutcher22: RT @SarahW\_P: While recidivism rates for gang members in Canada is high, Trudeau thinks people who may have been born into terrorism and ra"  
##   
## [[3]]  
## [1] "gerrybuddy: RT @SarahW\_P: While recidivism rates for gang members in Canada is high, Trudeau thinks people who may have been born into terrorism and ra"  
##   
## [[4]]  
## [1] "seve22xmen: RT @SarahW\_P: While recidivism rates for gang members in Canada is high, Trudeau thinks people who may have been born into terrorism and ra"  
##   
## [[5]]  
## [1] "clinton\_grubb: RT @kygov: Expert panel recommends steps to control prison costs and reduce crime - Recommendations would reduce recidivism, cut future pri"  
##   
## [[6]]  
## [1] "Jmansbox: RT @SarahW\_P: While recidivism rates for gang members in Canada is high, Trudeau thinks people who may have been born into terrorism and ra"  
##   
## [[7]]  
## [1] "missionbasedma: RT @ACLU\_Mass: Our elected prosecutors can help break the cycle of addiction, incarceration, and recidivism by taking a public health appro"  
##   
## [[8]]  
## [1] "Geoffsabtchange: Video: The Impact of Realignment on Recidivism - Public Policy Institute of California https://t.co/b4WLO2E14t"  
##   
## [[9]]  
## [1] "Leah\_Garabedian: RT @thejakehorowitz: South Dakota cuts juvenile commitments in half since 2014, with no apparent impact on crime or recidivism https://t.co"  
##   
## [[10]]  
## [1] "Leah\_Garabedian: RT @SDGovDaugaard: The juvenile justice report offers data to analyze the success of these reforms. We are seeing fewer juveniles committed"  
##   
## [[11]]  
## [1] "MarcALevin: RT @PatNolan4Justic: MI prison bureaucracy blocks free books for inmates. There have been no reports of security issues. So, my hunch is th"  
##   
## [[12]]  
## [1] "CarolM39: @oldhippiebroad @kwilli1046 Despite drug treatments, recidivism rate is very high, just like for sex offenders. https://t.co/glwXvYUNAp"  
##   
## [[13]]  
## [1] "tech\_52: @Evan\_McMullin @JRubinBlogger Is it not understandable, when facing a wholly unfit POTUS, that ones previous views https://t.co/ZAIZxX5ATs"  
##   
## [[14]]  
## [1] "punchinghawk: RT @\_littleemermaid: Norway literally has the lowest recidivism rate in the world. Whyyy arent we learning? https://t.co/RQmwJUIpoh"  
##   
## [[15]]  
## [1] "IndivisibleIL17: RT @danielbiss: Too many formerly incarcerated individuals aren't set up for success after being released from prison. Our plan will help r"  
##   
## [[16]]  
## [1] "Shon\_Is\_Here: I think some prison facilities were designed for recidivism...\n\nFor profit business..."  
##   
## [[17]]  
## [1] "dporterJCPS: RT @arneduncan: Thank you @JohnBKing for being such a passionate advocate for prison education!\nMore education= less recidivism, less crime"  
##   
## [[18]]  
## [1] "kygov: Expert panel recommends steps to control prison costs and reduce crime - Recommendations would reduce recidivism, c https://t.co/2X8cLGbS8q"  
##   
## [[19]]  
## [1] "fra\_gee\_lay: RT @SarahW\_P: While recidivism rates for gang members in Canada is high, Trudeau thinks people who may have been born into terrorism and ra"  
##   
## [[20]]  
## [1] "ImeldaMarcos2: @helenprejean Of course they are. Higher recidivism rates mean more prisoners back in jail, more profits for private prisons."  
##   
## [[21]]  
## [1] "KendraWrites: Rapists are rarely one offs. They have insanely high recidivism rates. https://t.co/cg6ejk4i0S"  
##   
## [[22]]  
## [1] "helentaylorjoha: RT @helentaylorjoha: @Evan\_McMullin \"I am the high priestess of sorcery and recidivism!\" https://t.co/kzZyDWMENN"  
##   
## [[23]]  
## [1] "buissongourmand: RT @RepKarenBass: Thanks @RepSteveChabot for visiting me in LA to tour programs aimed at reducing recidivism and crime like @HomeboyInd and"  
##   
## [[24]]  
## [1] "stephenseymour9: RT @ReintPuzzle: 'Triple bunking': $120m for overcrowded WA prisons? Focus needed on reducing recidivism and building capacity for people t"  
##   
## [[25]]  
## [1] "Region7Success: RT @MichiganDOC: Recently, Dir. @HeidiWashington spoke at the @amprog #unlockingopportunity conference to discuss the innovative ways the d"