BDS Project

Gina O’Riordan - East

11/16/2018

Hypotheses:

After walking through each of the deliverables during this semester, I have honed my hypothesis and research question:

I am interested in what factors predict a prior record, age of offense, and years on death row of an executed Texas inmate using the following variables as predictor variables: * countyorCountry * educationYears * codefsYes * totalVictims * femaleVictim * foreignNational * race2

Exploratory Data Analysis

Data Import & Review

```
#load libraries
library(tidyverse)
```

```
## ── Attaching packages ───────────────────────────────────────────
────────────────────────── tidyverse 1.2.1 ──
```

```
## ✔ ggplot2 3.1.0     ✔ purrr   0.2.5
## ✔ tibble  1.4.2     ✔ dplyr   0.7.8
## ✔ tidyr   0.8.1     ✔ stringr 1.3.1
## ✔ readr   1.1.1     ✔ forcats 0.3.0
```

```
## ── Conflicts ────────────────────────────────────────────────────
───────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
library(reshape)
```

```
##
## Attaching package: 'reshape'
```

```
## The following object is masked from 'package:dplyr':
##
##     rename
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, smiths
```

```
library(data.table)
```

```
##
## Attaching package: 'data.table'
```

```
## The following object is masked from 'package:reshape':
##
##     melt
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
## The following object is masked from 'package:purrr':
##
##     transpose
```

```
library(ggplot2)
library(ggcorrplot)
library(caret)
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```
library(rpart)
library(randomForest)
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
library(rminer)
library(partykit)
```

```
## Loading required package: grid
```

```
## Loading required package: libcoin
```

```
## Loading required package: mvtnorm
```

```
options(warn=-1)
```

First I will import the original data set and view it as a whole to get a feel for the variables.

```
#load data
texas <- read.csv("~/Google Drive/ND MS Data Science/Fall 2018: Behavioral Data Science/
Project/BDS-W13-TexasExecution.csv", header=T)

#review data as a whole
head(texas)
```

```
##   executionNumber inmateNumber    lastName firstName            fullName
## 1               1          592 Brooks, Jr.   Charlie Brooks, Charlie Jr.
## 2               2          670       Autry     James  Autry, James David
## 3               3          529     O'Bryan    Ronald      ronald o'bryan
## 4               4          621    Barefoot    Thomas     thomas barefoot
## 5               5          518    Skillern     Doyle      doyle dkillern
## 6               6          712       Morin   Stephen       stephen morin
##         fullName2 dateofExecution dateofExecution2 executionDecade
## 1  Charlie Brooks       12/7/1982        12/7/1982             80s
## 2     James Autry       3/14/1984        3/14/1984             80s
## 3  Ronald O'Bryan       3/31/1984        3/31/1984             80s
## 4 Thomas Barefoot      10/30/1984       10/30/1984             80s
## 5  Doyle Skillern       1/16/1985        1/16/1985             80s
## 6   Stephen Morin       3/13/1985        3/13/1985             80s
##   dateReceived dateofOffense ageatDateofOffense dateofOffense2
## 1    4/25/1978    12/14/1976                 34     12/14/1976
## 2   10/10/1980     4/20/1980                 25      4/20/1980
## 3         <NA>          <NA>                 NA           <NA>
## 4         <NA>          <NA>                 NA           <NA>
## 5         <NA>          <NA>                 NA           <NA>
## 6         <NA>          <NA>                 NA           <NA>
##   methodofExecution dateofBirth dateofBirth2 ageatExecution
## 1  Lethal Injection    9/1/1942     9/1/1942             40
## 2  Lethal Injection   9/27/1954    9/27/1954             29
## 3  Lethal Injection        <NA>         <NA>             39
## 4  Lethal Injection        <NA>         <NA>             39
## 5  Lethal Injection        <NA>         <NA>             49
## 6  Lethal Injection        <NA>         <NA>             34
##   ageatExecution2 ageReceived yearsonDeathRow countyTDCJMain
## 1              40          35               5        tarrant
## 2              29          26               3      jefferson
## 3              39          31               8         harris
## 4              39          34               5           bell
## 5              48          NA              NA        lubbock
## 6              37          NA              NA      jefferson
##   countyorCountry nativeCounty   county2 nativeState nativeCountry  sex
## 1         tarrant      tarrant   tarrant       texas          <NA> Male
## 2       jefferson       potter jefferson       texas          <NA> Male
## 3            <NA>         <NA>    harris        <NA> united states <NA>
## 4            <NA>   new iberia      bell   louisiana united states <NA>
## 5            <NA>         <NA>  live oak        <NA> united states <NA>
## 6            <NA>         <NA> jefferson        <NA> united states <NA>
##   sex2 hairColor                          eyeColor victimRaceGender  race
## 1    m     black mar (according to dps records)      White  Male Black
## 2    m     brown                             brown           female White
## 3    m      <NA>                              <NA>       white male White
## 4    m      <NA>                              <NA>       white male White
## 5    m      <NA>                              <NA>             male White
## 6    m      <NA>                              <NA>     white female White
##   race2 victimRaceMatch height weight educationYears priorOccupation
## 1 Black              No     69    150             12         Laborer
## 2 White              No     68    137              6         Laborer
## 3 White              No     NA     NA             NA            <NA>
```

```
## 4 White                No     NA     NA             NA            <NA>
## 5 White                No     NA     NA             NA            <NA>
## 6 White                No     NA     NA             NA            <NA>
##
priorRecord
## 1                    Federal  Prison, Leavenworth,  Illegal Possession of Firearms, Dis
charged 1968
## 2 5  year sentence for Assault and Attempted Robbery - 1972; 8 year sentence for  Bur
glary - 1975
## 3
        None
## 4
           1
## 5
        <NA>
## 6
           1
##   priorRecordYes juvenile federal volunteer foreignNational
## 1            Yes       No      No        No              No
## 2            Yes       No      No        No              No
## 3             No       No      No        No              No
## 4            Yes       No      No        No              No
## 5           <NA>       No      No        No              No
## 6            Yes       No      No       Yes              No
##




                              summaryofCrime
## 1 Brooks  went to a car lot under the pretense of wanting to test drive a car. A mech
anic  accompanied him on the drive. Brooks stopped to pick up a co-defendant. The  mecha
nic was put in the trunk of the car. Brooks and his co-defendant went to a  motel. The m
echanic was brought out of the trunk and taken into a motel room.  The mechanic was boun
d with coat hangers, gagged with adhesive tape, and shot in  the head, causing his deat
h. Brooks and the co-defendant fled the scene.
## 2                                                       On  April 20, 1980, Autry shot
a 43 year old female convenience store clerk between  the eyes with a .38 caliber pistol
causing her death. Autry had been arguing  with the clerk about the price of a six pack
of beer. Two witnesses were also  shot in the head. One witness was a 43 year old former
Roman Catholic priest,  who died instantly. The other witness was a Greek seaman who sur
vived the  gunshot, with serious injuries.
## 3





                                        <NA>
## 4





                                        <NA>
```

```
## 5
                                        <NA>
## 6
                                        <NA>
##          codefendants codefsYes totalVictims victims2orMore femaleVictim
## 1       Woody  Loudres       Yes            1             No           No
## 2 John  Alton Sandifer       Yes            1             No          Yes
## 3                None.        No            1             No           No
## 4                None.        No            1             No           No
## 5                    1       Yes            1             No           No
## 6                None.        No            1             No          Yes
##   totalWhite totalBlack totalLatino totalAsian totalNativeAmerican
## 1          1          0           0          0                   0
## 2          1          0           0          0                   0
## 3          1          0           0          0                   0
## 4          1          0           0          0                   0
## 5          1          0           0          0                   0
## 6          1          0           0          0                   0
##   totalOther totalMale totalFemale
## 1          0         1           0
## 2          0         0           1
## 3          0         1           0
## 4          0         1           0
## 5          0         1           0
## 6          0         0           1
##
                                                                     statementTD
CJ
## 1
                                                                             Sta
tement to the Media: I, at this very moment, have absolutely no fear of what may happen
to this body.  My fear is for Allah, God only, who has at this moment the only power to
determine if I should live or die... As a devout Muslim, I am taught and believe that th
is material life is only for the express purpose of preparing oneself for the real life
that is to come... Since becoming Muslim, I have tried to live as Allah wanted me to liv
e.
## 2
```

```
                                               This offender declined to make a last statement.

## 3 What is about to transpire in a few moments is wrong! However, we as human beings d
o make mistakes and errors. This execution is one of those wrongs yet doesn\x92t mean ou
r whole system of justice is wrong. Therefore, I would forgive all who have taken part i
n any way in my death. Also, to anyone I have offended in any way during my 39 years, I
pray and ask your forgiveness, just as I forgive anyone who offended me in any way. And
I pray and ask God\x92s forgiveness for all of us respectively as human beings. To my lo
ved ones, I extend my undying love. To those close to me, know in your hearts I love you
one and all. God bless you all and may God\x92s best blessings be always yours. Ronald
C. O\x92Bryan P.S. During my time here, I have been treated well by all T.D.C. personne
l.
## 4                                                        When asked if he had a last stateme
nt, he replied, "Yes, I do."I hope that one day we can look back on the evil that we\x92
re doing right now like the witches we burned at the stake. I want everybody to know tha
t I hold nothing against them. I forgive them all. I hope everybody I\x92ve done anythin
g to will forgive me. I\x92ve been praying all day for Carl Levin\x92s wife to drive the
bitterness from her heart because that bitterness that\x92s in her heart will send her t
o Hell just as surely as any other sin. I\x92m sorry for everything I\x92ve ever done to
anybody. I hope they\x92ll forgive me. "Sharon, tell all my friends goodbye. You know wh
o they are: Charles Bass, David Powell\x85" Then he coughed and nothing else was said.
## 5

                              I pray that my family will rejoice and will forgive, thank yo
u.
## 6

                                                                              Heavenly Fath
er, I give thanks for this time, for the time that we have been together, the fellowship
in your world, the Christian family presented to me (He called the names of the personal
witnesses.). Allow your holy spirit to flow as I know your love as been showered upon m
e. Forgive them for they know not what they do, as I know that you have forgiven me, as
I have forgiven them. Lord Jesus, I commit my soul to you, I praise you, and I thank yo
u.
##   gaveLastStatement externalStatementCheck
## 1               Yes                     No
## 2                No                     No
## 3               Yes                     No
## 4               Yes                     No
## 5               Yes                     No
```

6 Yes No
##

correctedStatements
1

Statement to the Media: I, at this very moment, have absolutely no fear of what may happen to this body. My fear is for Allah, God only, who has at this moment the only power to determine if I should live or die... As a devout Muslim, I am taught and believe that this material lif e is only for the express purpose of preparing oneself for the real life that is to com e... Since becoming Muslim, I have tried to live as Allah wanted me to live.
2

<NA>
3 What is about to transpire in a few moments is wrong! However, we as human beings d o make mistakes and errors. This execution is one of those wrongs yet doesn't mean our w hole system of justice is wrong. Therefore, I would forgive all who have taken part in a ny way in my death. Also, to anyone I have offended in any way during my 39 years, I pra y and ask your forgiveness, just as I forgive anyone who offended me in any way. And I p ray and ask God's forgiveness for all of us respectively as human beings. To my loved on es, I extend my undying love. To those close to me, know in your hearts I love you one a nd all. God bless you all and may God's best blessings be always yours. Ronald C. O'Brya n P.S. During my time here, I have been treated well by all T.D.C. personnel.
4

Yes, I do."I h ope that one day we can look back on the evil that we're doing right now like the witche s we burned at the stake. I want everybody to know that I hold nothing against them. I f orgive them all. I hope everybody I've done anything to will forgive me. I've been prayi ng all day for Carl Levin's wife to drive the bitterness from her heart because that bit terness that's in her heart will send her to Hell just as surely as any other sin. I'm s orry for everything I've ever done to anybody. I hope they'll forgive me. "Sharon, tell all my friends goodbye. You know who they are: Charles Bass, David Powell..."
5

```
              I pray that my family will rejoice and will forgive, thank you.
## 6




                      Heavenly Father, I give thanks for this time, for the time that we
have been together, the fellowship in your world, the Christian family presented to me .
Allow your holy spirit to flow as I know your love as been showered upon me. Forgive the
m for they know not what they do, as I know that you have forgiven me, as I have forgive
n them. Lord Jesus, I commit my soul to you, I praise you, and I thank you.
##   uniqueWords typeTokenRatio sentenceCount sentenceLength syllableCount
## 1          56      0.6321839             3           29.0      1.402299
## 2          NA             NA            NA             NA            NA
## 3          96      0.6000000            10           15.5      1.335484
## 4          79      0.6320000            10           12.5      1.336000
## 5          11      0.9166667             1           12.0      1.333333
## 6          48      0.6000000             4           20.0      1.312500
##   characterCount letterCount       FOG   flesch measTextLexDiversity
## 1            446         343 14.358621 58.76552             55.16418
## 2             NA          NA        NA       NA                   NA
## 3            777         598  9.296774 78.12056             83.13483
## 4            620         477  7.240000 81.12190             76.81081
## 5             64          50  8.133333 81.85500             12.00000
## 6            408         314 11.500000 75.49750             40.50000
```

```
str(texas)
```

```
## 'data.frame':    518 obs. of  72 variables:
##  $ executionNumber       : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ inmateNumber          : int  592 670 529 621 518 712 713 628 551 555 ...
##  $ lastName              : Factor w/ 421 levels "Adams","Adanandus",..: 54 16 290 24
360 272 105 264 306 341 ...
##  $ firstName             : Factor w/ 249 levels "Aaron","Adolph",..: 32 114 211 229 6
7 224 127 31 104 31 ...
##  $ fullName              : Factor w/ 505 levels "aaron fuller",..: 39 20 434 463 127
455 249 58 189 60 ...
##  $ fullName2             : Factor w/ 515 levels "Aaron Foust",..: 61 191 448 476 127
468 236 54 180 57 ...
##  $ dateofExecution       : Factor w/ 515 levels "1/10/2007","1/12/2000",..: 153 211 2
35 74 9 210 287 372 420 477 ...
##  $ dateofExecution2      : Factor w/ 515 levels "1/10/2007","1/12/2000",..: 153 211 2
35 74 9 210 287 372 420 476 ...
##  $ executionDecade       : Factor w/ 4 levels "00s","10s","80s",..: 3 3 3 3 3 3 3 3 3
3 ...
##  $ dateReceived          : Factor w/ 130 levels "1/18/2001","1/20/1999",..: 68 10 NA
NA NA NA NA NA NA NA ...
##  $ dateofOffense         : Factor w/ 127 levels "1/11/2003","1/16/2003",..: 26 66 NA
NA NA NA NA NA NA NA ...
##  $ ageatDateofOffense    : int  34 25 NA NA NA NA 20 NA 34 18 ...
##  $ dateofOffense2        : Factor w/ 127 levels "1/11/2003","1/16/2003",..: 26 66 NA
NA NA NA NA NA NA NA ...
##  $ methodofExecution     : Factor w/ 1 level "Lethal Injection": 1 1 1 1 1 1 1 1 1 1
...
##  $ dateofBirth           : Factor w/ 134 levels "04/24/0976","1/11/1974",..: 124 130
NA NA NA NA NA NA NA NA ...
##  $ dateofBirth2          : Factor w/ 134 levels "1/11/1974","1/13/1979",..: 124 130 N
A NA NA NA NA NA NA NA ...
##  $ ageatExecution        : int  40 29 39 39 49 34 24 34 43 28 ...
##  $ ageatExecution2       : int  40 29 39 39 48 37 24 34 43 28 ...
##  $ ageReceived           : int  35 26 31 34 NA NA 22 27 35 19 ...
##  $ yearsonDeathRow       : int  5 3 8 5 NA NA 2 7 8 9 ...
##  $ countyTDCJMain        : Factor w/ 89 levels "anderson","aransas",..: 77 42 36 7 54
42 8 77 77 67 ...
##  $ countyorCountry       : Factor w/ 48 levels "anderson","bailey",..: 43 24 NA NA NA
NA NA NA NA NA ...
##  $ nativeCounty          : Factor w/ 201 levels "alameda","albany",..: 180 154 NA 131
NA NA NA NA 14 NA ...
##  $ county2               : Factor w/ 95 levels "anderson","aransas",..: 83 46 37 8 55
46 9 83 83 72 ...
##  $ nativeState           : Factor w/ 44 levels "alabama","alberta",..: 37 37 NA 16 NA
NA NA NA NA NA ...
##  $ nativeCountry         : Factor w/ 11 levels "canada","dominican republic",..: NA N
A 9 9 9 9 9 9 9 9 ...
##  $ sex                   : Factor w/ 2 levels "Female","Male": 2 2 NA NA NA NA NA NA
NA NA ...
##  $ sex2                  : Factor w/ 2 levels "f","m": 2 2 2 2 2 2 2 2 2 2 ...
##  $ hairColor             : Factor w/ 7 levels "black","blonde",..: 1 4 NA NA NA NA NA
1 NA NA ...
##  $ eyeColor              : Factor w/ 8 levels "black","blue",..: 7 3 NA NA NA NA NA 3
NA NA ...
```

```
## $ victimRaceGender      : Factor w/ 131 levels " male"," white female",..: 112 32 12
5 125 55 120 55 32 55 55 ...
## $ race                  : Factor w/ 4 levels "Black","Hispanic",..: 1 4 4 4 4 4 2 1
2 4 ...
## $ race2                 : Factor w/ 5 levels "Asian","Black",..: 2 5 5 5 5 5 3 2 3 5
...
## $ victimRaceMatch       : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 2 1 1 ...
## $ height                : int  69 68 NA NA NA NA NA NA NA NA ...
## $ weight                : int  150 137 NA NA NA NA NA NA NA NA ...
## $ educationYears        : int  12 6 NA NA NA NA NA NA NA NA ...
## $ priorOccupation       : Factor w/ 190 levels "accounting","air conditioner repairm
an",..: 108 108 NA NA NA NA NA NA 137 NA ...
## $ priorRecord           : Factor w/ 72 levels "#1090018  on a 2 year sentence from H
idalgo   County for escape.",..: 30 26 37 22 NA 22 22 NA 22 22 ...
## $ priorRecordYes        : Factor w/ 2 levels "No","Yes": 2 2 1 2 NA 2 2 NA 2 2 ...
## $ juvenile              : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 2 ...
## $ federal               : Factor w/ 1 level "No": 1 1 1 1 1 1 1 1 1 1 ...
## $ volunteer             : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 2 1 1 1 2 ...
## $ foreignNational       : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
## $ summaryofCrime        : Factor w/ 134 levels "Awaiting  Information",..: 5 93 NA N
A NA NA NA NA NA NA ...
## $ codefendants          : Factor w/ 60 levels "1","Adams,  Beunka",..: 60 30 44 44 1
44 44 NA 44 NA ...
## $ codefsYes             : Factor w/ 2 levels "No","Yes": 2 2 1 1 2 1 1 NA 1 NA ...
## $ totalVictims          : int  1 1 1 1 1 1 1 1 1 1 ...
## $ victims2orMore        : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
## $ femaleVictim          : Factor w/ 2 levels "No","Yes": 1 2 1 1 1 2 1 2 1 1 ...
## $ totalWhite            : int  1 1 1 1 1 1 0 0 1 1 ...
## $ totalBlack            : int  0 0 0 0 0 0 0 1 0 0 ...
## $ totalLatino           : int  0 0 0 0 0 0 0 0 0 0 ...
## $ totalAsian            : int  0 0 0 0 0 0 1 0 0 0 ...
## $ totalNativeAmerican   : int  0 0 0 0 0 0 0 0 0 0 ...
## $ totalOther            : int  0 0 0 0 0 0 0 0 0 0 ...
## $ totalMale             : int  1 0 1 1 1 0 1 0 1 1 ...
## $ totalFemale           : int  0 1 0 0 0 1 0 1 0 0 ...
## $ statementTDCJ         : Factor w/ 444 levels " \"I've got one thing to say, get yo
ur Warden off this gurney and shut up. I am from the island of Barbados. I "| __truncate
d__,..: 348 367 384 91 249 197 169 82 262 171 ...
## $ gaveLastStatement     : Factor w/ 2 levels "No","Yes": 2 1 2 2 2 2 2 2 2 2 ...
## $ externalStatementCheck: Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 2 1 1 1 ...
## $ correctedStatements   : Factor w/ 423 levels "\"I've got one thing to say, get you
r Warden off this gurney and shut up. I am from the island of Barbados. I a"| __truncate
d__,..: 254 NA 305 374 121 52 18 279 142 21 ...
## $ uniqueWords           : int  56 NA 96 79 11 48 16 21 137 34 ...
## $ typeTokenRatio        : num  0.632 NA 0.6 0.632 0.917 ...
## $ sentenceCount         : int  3 NA 10 10 1 4 3 2 23 3 ...
## $ sentenceLength        : num  29 NA 15.5 12.5 12 ...
## $ syllableCount         : num  1.4 NA 1.34 1.34 1.33 ...
## $ characterCount        : int  446 NA 777 620 64 408 134 136 1390 228 ...
## $ letterCount           : int  343 NA 598 477 50 314 104 104 1089 170 ...
## $ FOG                   : num  14.36 NA 9.3 7.24 8.13 ...
## $ flesch                : num  58.8 NA 78.1 81.1 81.9 ...
## $ measTextLexDiversity  : num  55.2 NA 83.1 76.8 12 ...
```

The variables look to be in the correct format and class for the time being. Based on the variables I'm interested in, I will select the variables I need for analysis before continuing to review the variables. For the time being, I will leave in only executionNumber variable as an identifier since this will have no bearing on analysis besides as identification for the offenders.

```
#select only the needed variables for hypothesis testing
texas_analysis <- texas %>% select(executionNumber, priorRecordYes, ageatDateofOffense,
 yearsonDeathRow, countyorCountry, educationYears, codefsYes, totalVictims, femaleVicti
m, race2)

#review the limited dataset
str(texas_analysis)
```

```
## 'data.frame':    518 obs. of  10 variables:
##  $ executionNumber   : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ priorRecordYes    : Factor w/ 2 levels "No","Yes": 2 2 1 2 NA 2 2 NA 2 2 ...
##  $ ageatDateofOffense: int  34 25 NA NA NA NA 20 NA 34 18 ...
##  $ yearsonDeathRow   : int  5 3 8 5 NA NA 2 7 8 9 ...
##  $ countyorCountry   : Factor w/ 48 levels "anderson","bailey",..: 43 24 NA NA NA NA
NA NA NA NA ...
##  $ educationYears    : int  12 6 NA NA NA NA NA NA NA NA ...
##  $ codefsYes         : Factor w/ 2 levels "No","Yes": 2 2 1 1 2 1 1 NA 1 NA ...
##  $ totalVictims      : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ femaleVictim      : Factor w/ 2 levels "No","Yes": 1 2 1 1 1 2 1 2 1 1 ...
##  $ race2             : Factor w/ 5 levels "Asian","Black",..: 2 5 5 5 5 5 3 2 3 5 ...
```

All the variables are confirmed to be in the correct format. Then I'll create a table to count the missing values in each variable.

```
#obtain a table with the number of missing values
sapply(texas_analysis, function(d) sum(is.na(d)))
```

```
##    executionNumber     priorRecordYes ageatDateofOffense
##                  0                 18                 59
##    yearsonDeathRow    countyorCountry     educationYears
##                 17                384                 44
##          codefsYes       totalVictims       femaleVictim
##                 21                  0                  0
##              race2
##                  0
```

Note that only the countyorCountry variable is missing a significant amount of data. This may skew results; further analysis will be done later in this EDA.

Next I will view a summary of all mean and standard deviation values for the numeric variable totalVictims.

```
#view all numeric variable means
texas_analysis %>% select(ageatDateofOffense, yearsonDeathRow, educationYears, totalVict
ims) %>%
  na.omit() %>%
  summarize_all(c("mean"))
```

```
##   ageatDateofOffense yearsonDeathRow educationYears totalVictims
## 1           26.89041        11.29452       10.17352       1.3379
```

```
#view all numeric variable standard deviations
texas_analysis %>% select(ageatDateofOffense, yearsonDeathRow, educationYears, totalVict
ims) %>%
  na.omit() %>%
  summarize_all(c("sd"))
```

```
##   ageatDateofOffense yearsonDeathRow educationYears totalVictims
## 1           7.741643        4.223452       2.106334    0.8636982
```

In reviewing these calcuations, I have a few observations:

- It seems the offenders' average age is relative low at 26.4 but there is a large standard deviation of almost 8 years, which means the offenders' ages are probably pretty varied.
- The average years on death row is 11.1 years but with a smaller deviation than age at 3.9.
- The mean years of education for offenders is 10 years with a 2 year standard deviation, putting most offenders with a high school education or lower.
- The total victim average is just above 1 with a standard deviation less than 1, so there's minimal variation between how many victims each executed prison had.

All other variables are factor or identifier values, which will be more analyzed in other ways moving forward. Next let's review these variables in a table to see the count of each response.

```
table(texas_analysis$priorRecordYes)
```

```
##
##  No Yes
## 225 275
```

This variable is quite evenly distributed.

```
table(texas_analysis$countyorCountry)
```

```
##                                                 
##                                        anderson 
##                                               1 
##                                          bailey 
##                                               1 
##                                         bandera 
##                                               1 
##                                             bee 
##                                               1 
##                                            bell 
##                                               2 
##                                           bexar 
##                                              14 
##                                           bowie 
##                                               3 
##                                        brazoria 
##                                               1 
##                                          brazos 
##                                               1 
##      brazos (on a change of venue from jasper) 
##                                               1 
##                                        cherokee 
##                                               2 
##           clay (change of venue from montague) 
##                                               1 
##                                          collin 
##                                               1 
## collin - change of venue from hutchinson county 
##                                               1 
##                                          dallas 
##                                              19 
##                                          denton 
##                                               3 
##                                         el paso 
##                                               3 
##                                       fort bend 
##                                               2 
##                                           gregg 
##                                               1 
##                                          harris 
##                                              18 
##                                         hidalgo 
##                                               1 
##                                         hopkins 
##                                               1 
##                                            hunt 
##                                               2 
##                                       jefferson 
##                                               2 
##                                         kaufman 
##                                               1 
##                                            kerr 
##                                               1 
```

```
##                                              lamar 
##                                                  1 
##                               leon c/v from walker 
##                                                  1 
##                                            liberty 
##                                                  1 
##          llano (on change of venue from hood county) 
##                                                  1 
##                                            lubbock 
##                                                  4 
##                                          matagorda 
##                                                  2 
##                                           mclennan 
##                                                  1 
##                                         montgomery 
##                                                  4 
##                                        nacogdoches 
##                                                  1 
##                                            navarro 
##                                                  1 
##                                             nueces 
##                                                  3 
##                                              pecos 
##                                                  1 
##                                               polk 
##                                                  2 
##                                             potter 
##                                                  3 
##                                            refugio 
##                                                  2 
##                                              smith 
##                                                  2 
##                                            tarrant 
##                                                 14 
##                                          tom green 
##                                                  1 
##                                             travis 
##                                                  1 
##                                          val verde 
##                                                  1 
##                                           victoria 
##                                                  1 
##                                         williamson 
##                                                  1
```

There are a significant number of counties and these will be dealt with in a moment.

```
table(texas_analysis$codefsYes)
```

```
## 
##  No Yes 
## 275 222
```

This variable is quite evenly distributed.

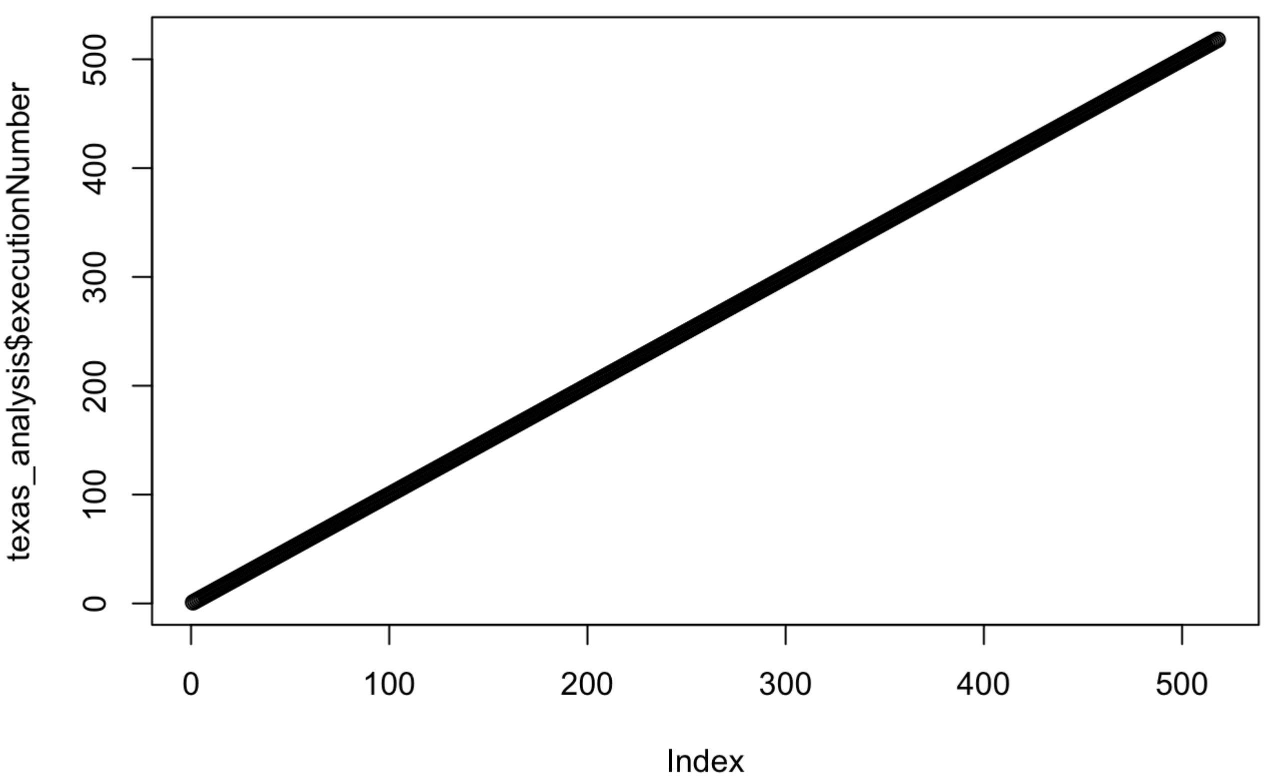
```
table(texas_analysis$race2)
```

```
## 
##           Asian           Black          Latino Native American 
##               2             193              93               2 
##           White 
##             228
```

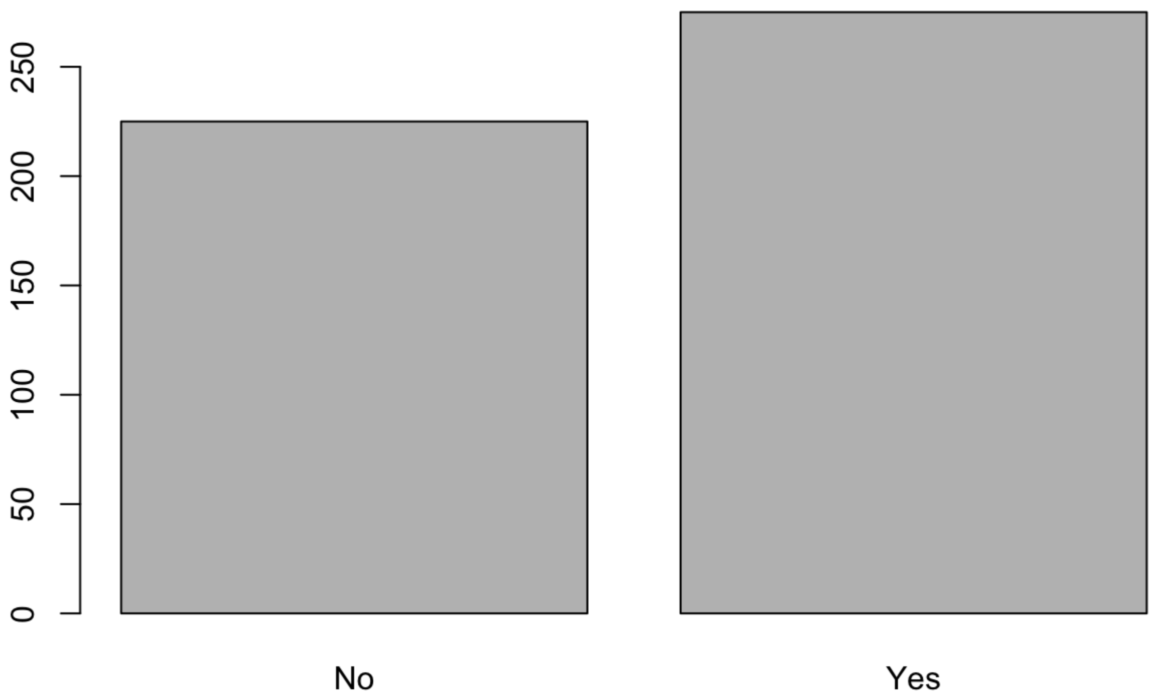
This variable has some weight towards black and white races, but this should not disturb the analysis.

Plots of All Variables

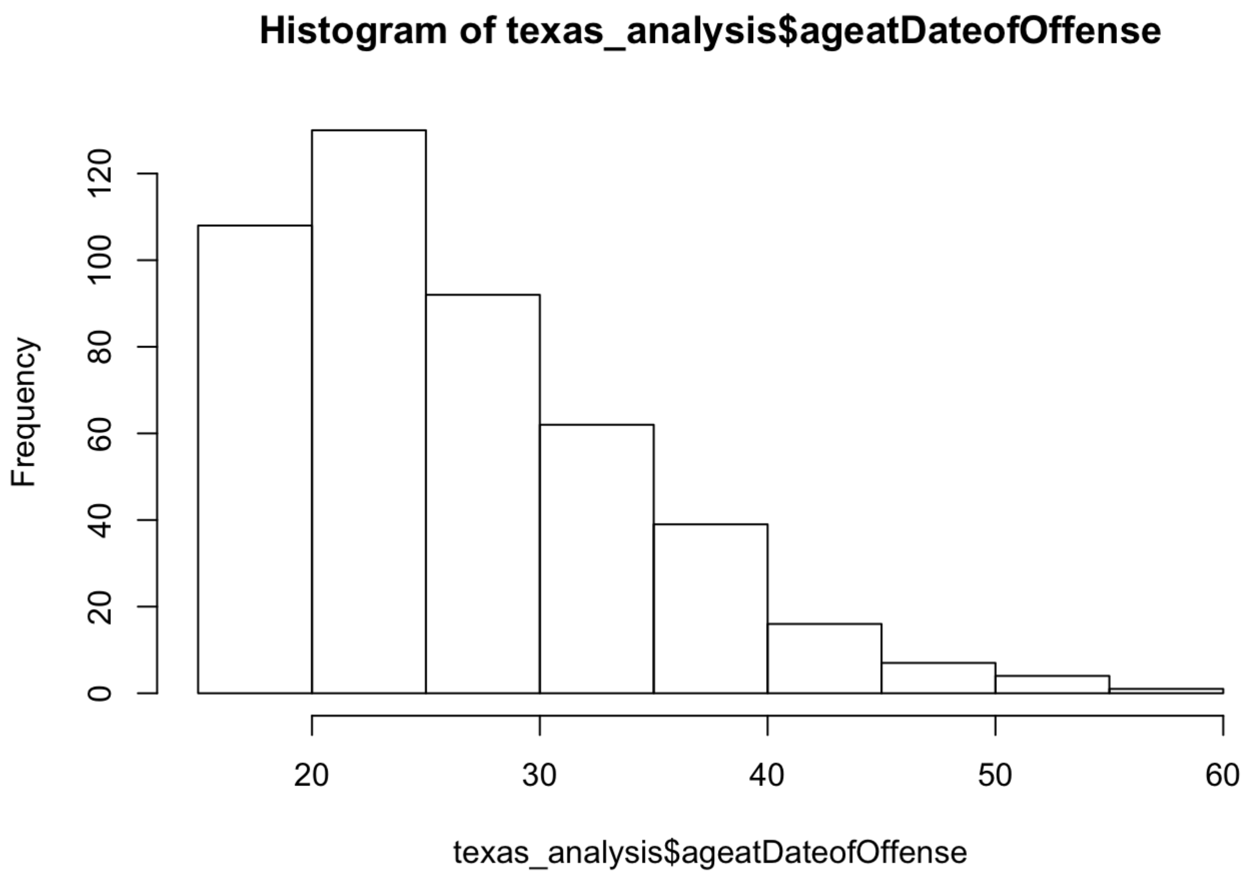
```
#execution number
plot(texas_analysis$executionNumber)
```



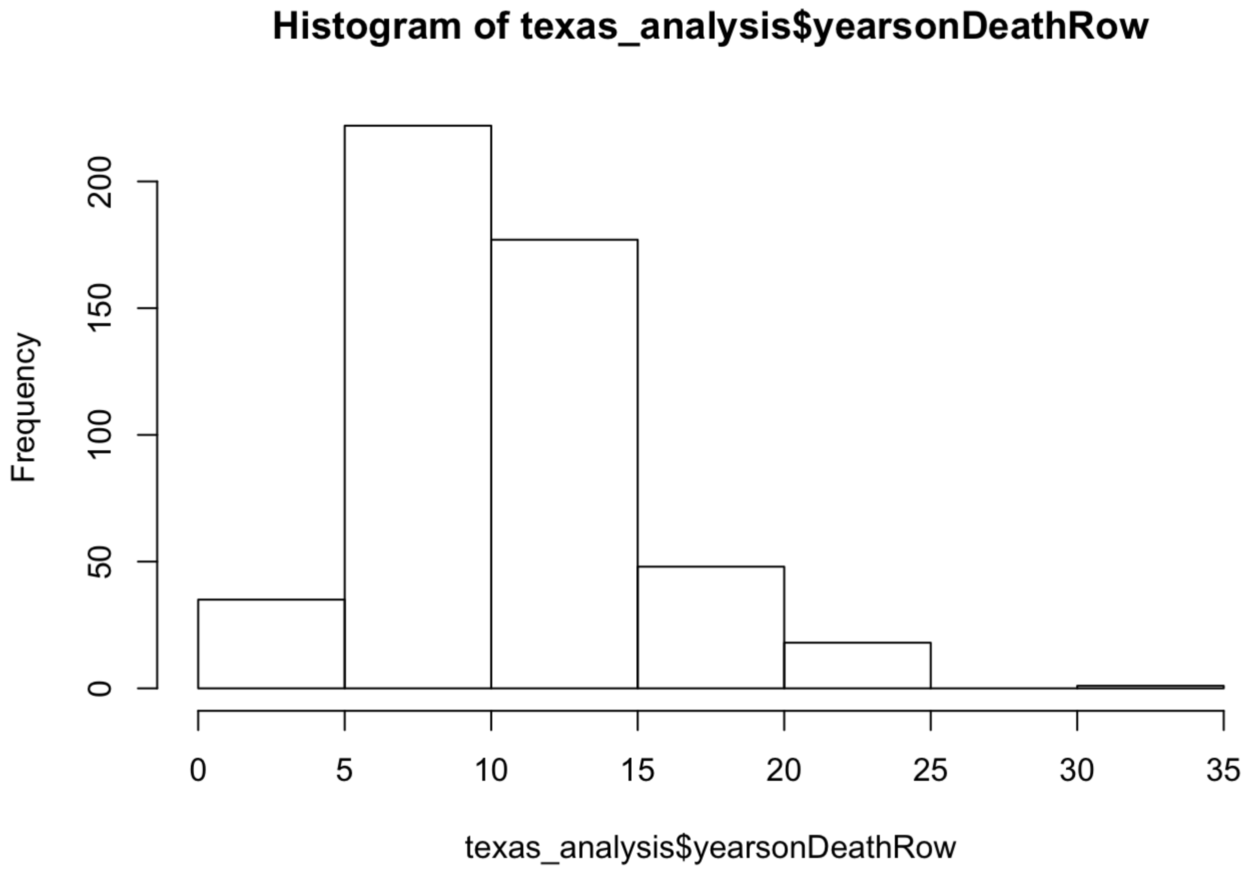

```
#prior record
plot(texas_analysis$priorRecordYes)
```



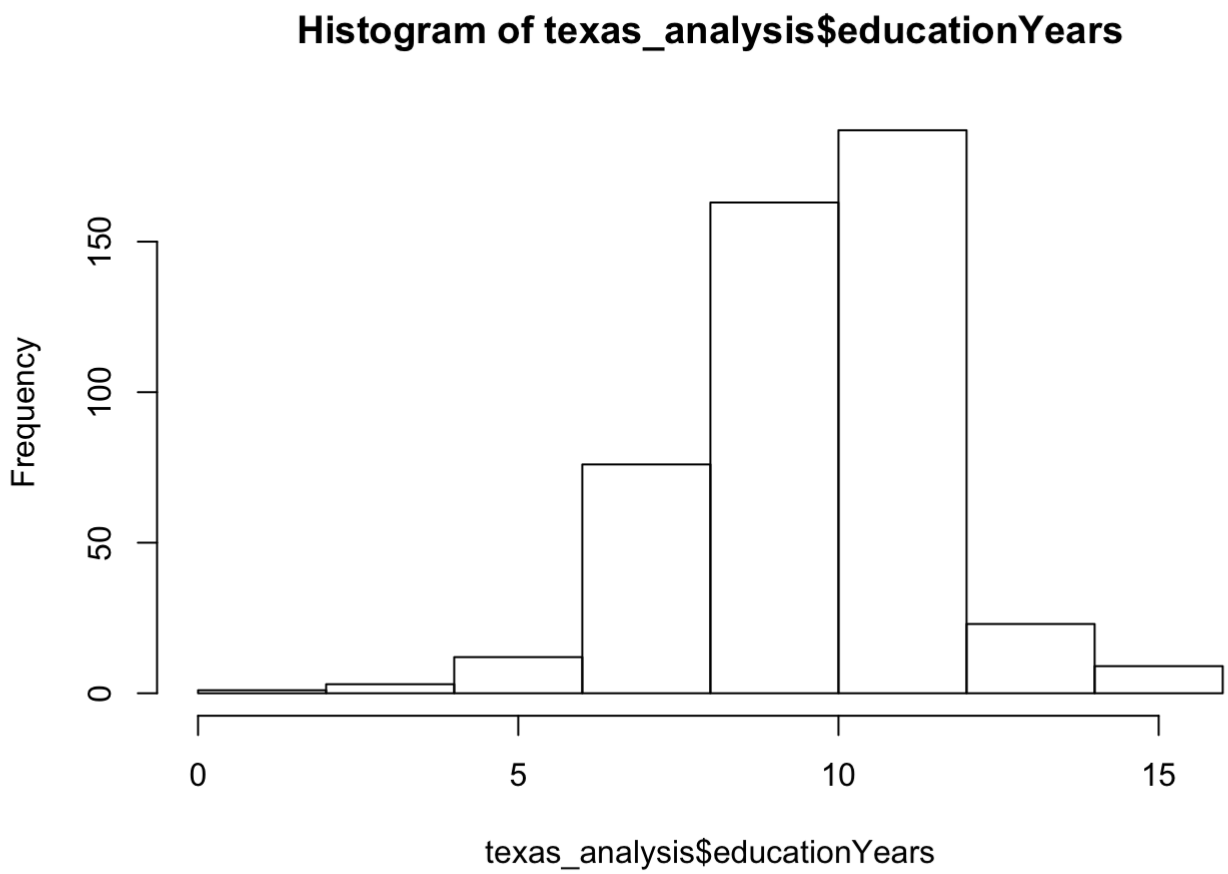

```
#age at date of offense
hist(texas_analysis$ageatDateofOffense)
```



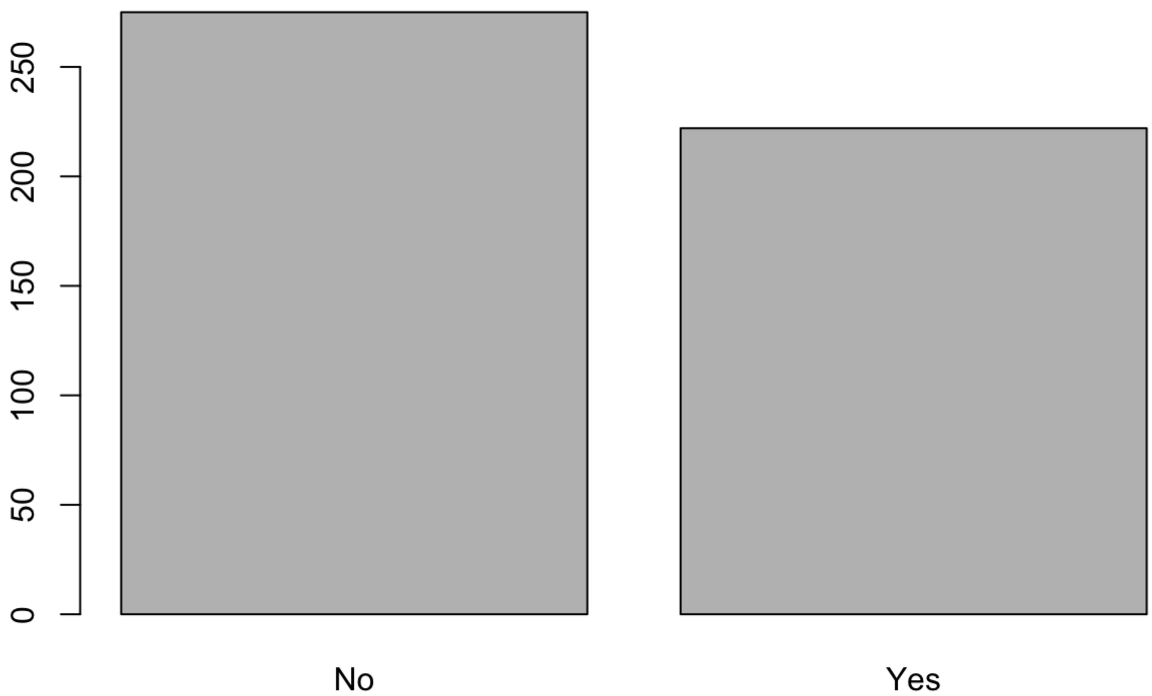

```
#years on death row
hist(texas_analysis$yearsonDeathRow)
```



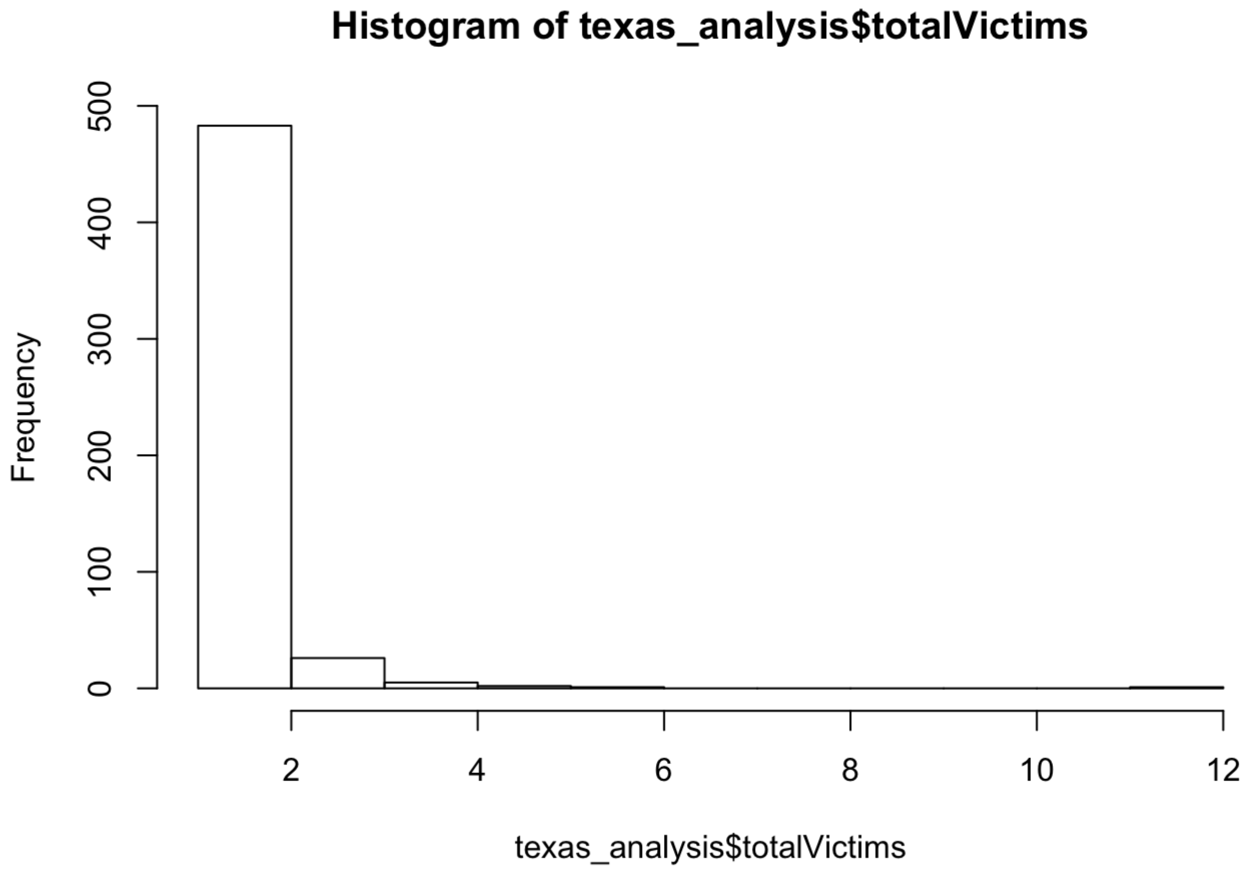

```
#education years
hist(texas_analysis$educationYears)
```



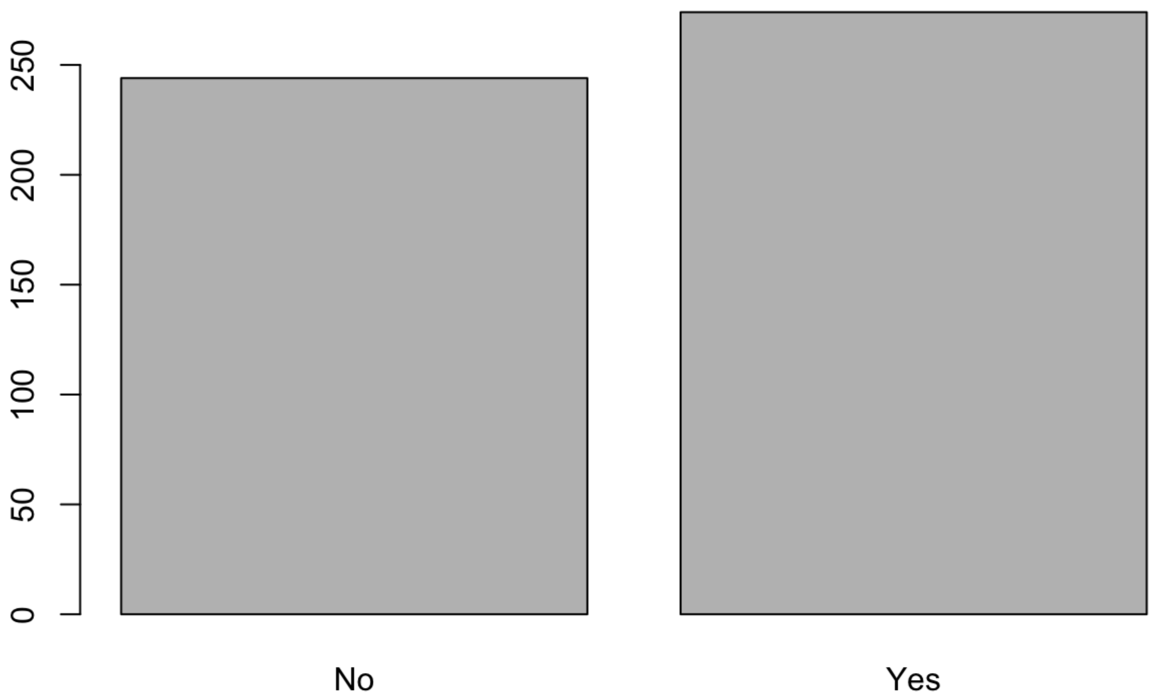

```
#codefendanats
plot(texas_analysis$codefsYes)
```



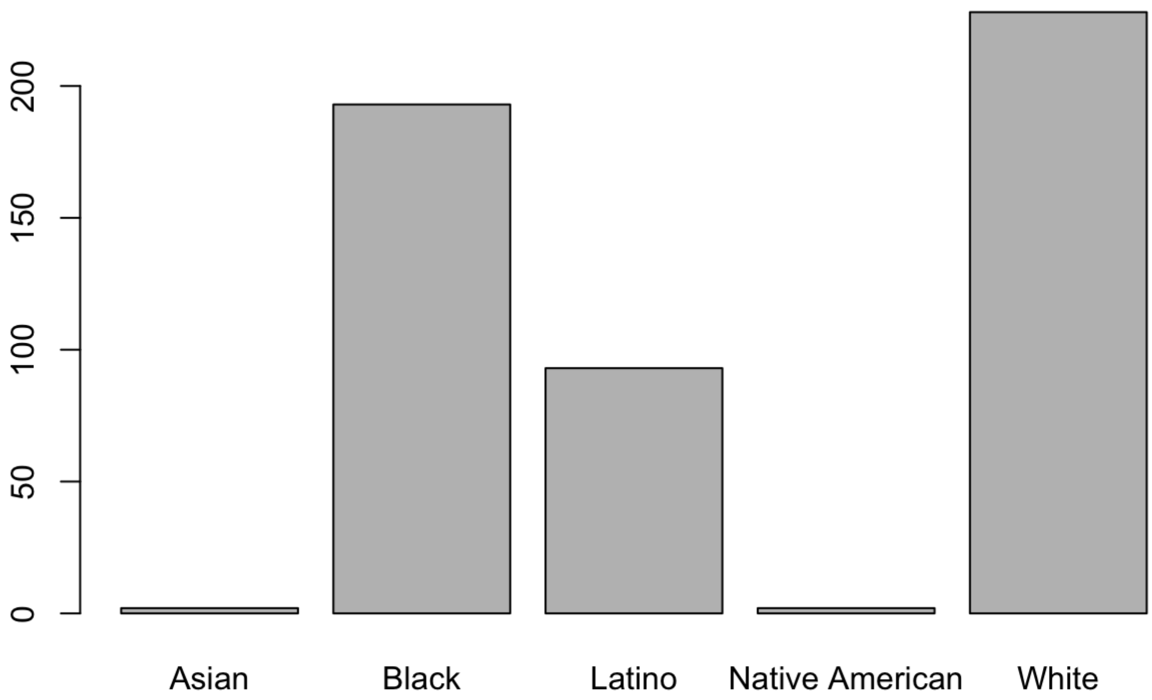

```
#total victims
hist(texas_analysis$totalVictims)
```




```
#female victims
plot(texas_analysis$femaleVictim)
```




```
#race
plot(texas_analysis$race2)
```



Note that the age at date of offense and years on death row variables are skewed to the right, while the years of education is skewed to the left.

Pre-Processing

Since the counties are listed as factors, but my goal will be comparing offenders from metropolitan v. rural counties, this variable needs to be turned into a binary variable. Based on geographic research done online, I looked at the top 10 metropolitian areas (in order of decreasing size) and matched them to the following counties:

1. Dallas county = Dallas
2. Tarrant county = Fort Worth
3. Harris county = Houston
4. Bexar county = San Antonio
5. Travis county = Austin
6. N/A = Mission
7. El Paso = El Paso
8. Nueces = Corpus Christi
9. N/A = Brownsville
10. Bell County = Temple
11. N/A = Beaumont

Dallas and Fort Worth are considered the same metropolitan area but are in separate counties, hence they are both listed in the #1 spot here. The N/A counties are ones not listed in the data set. The code below creates a new variable in the texas_analysis subset of data to use in hypothesis testing; it will be a binary variable marking if the county is rural or metropolitan. For the counties that are listed here, I will add them into the new 'metroArea' variable as 'yes' in the below code.

```
#create the metroArea variable
texas_analysis$metroArea <- ifelse(
  texas_analysis$countyorCountry == 'dallas' |
  texas_analysis$countyorCountry == 'tarrant'|
  texas_analysis$countyorCountry == 'harris' |
  texas_analysis$countyorCountry == 'bexar' |
  texas_analysis$countyorCountry == 'travis' |
  texas_analysis$countyorCountry == 'el paso' |
  texas_analysis$countyorCountry == 'nueces' |
  texas_analysis$countyorCountry == 'bell',
  'yes', 'no')
texas_analysis$metroArea <- as.factor(texas_analysis$metroArea)

#review breakdown
table(texas_analysis$metroArea)
```

```
##
##  no yes
##  60  74
```

Now it is clear that the amount of offenders from a metropolitan county are only a bit more prevalent than those form rural counties. This will make analysis easier. Before moving on, we'll remove the countyorCounty variable from the dataset.

```
#remove county variable
texas_analysis <- texas_analysis %>% select(-countyorCountry)

#review dataset
glimpse(texas_analysis)
```

```
## Observations: 518
## Variables: 10
## $ executionNumber    <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, ...
## $ priorRecordYes     <fct> Yes, Yes, No, Yes, NA, Yes, Yes, NA, Yes, Y...
## $ ageatDateofOffense <int> 34, 25, NA, NA, NA, NA, 20, NA, 34, 18, NA,...
## $ yearsonDeathRow    <int> 5, 3, 8, 5, NA, NA, 2, 7, 8, 9, 6, 4, NA, N...
## $ educationYears     <int> 12, 6, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ codefsYes          <fct> Yes, Yes, No, No, Yes, No, No, NA, No, NA, ...
## $ totalVictims       <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1...
## $ femaleVictim       <fct> No, Yes, No, No, No, Yes, No, Yes, No, No, ...
## $ race2              <fct> Black, White, White, White, White, White, L...
## $ metroArea          <fct> yes, no, NA, NA, NA, NA, NA, NA, NA, NA, NA...
```

Finally, the identifier variable remains: executionNumber. It will not be reviewed in this analysis as it has no bearing on the hypothesis testing besides identifying an inmate from another.

Finally, we check for zero variance in the data, which could alter the models.

```
#original data
nearZeroVar(texas_analysis)
```

```
## integer(0)
```

```
#no zero variance variables found
```

There are no variables with near zero variance, so the dataset is ready for analysis. First, however, we'll visualize the variables.

Analysis

Create a dataset that does not include the three identifier variables.

```
#remove identifying variables
texas_noid <- texas_analysis %>% select(-executionNumber)
head(texas_noid)
```

```
##   priorRecordYes ageatDateofOffense yearsonDeathRow educationYears
## 1            Yes                 34               5             12
## 2            Yes                 25               3              6
## 3             No                 NA               8             NA
## 4            Yes                 NA               5             NA
## 5           <NA>                 NA              NA             NA
## 6            Yes                 NA              NA             NA
##   codefsYes totalVictims femaleVictim race2 metroArea
## 1       Yes            1           No Black       yes
## 2       Yes            1          Yes White        no
## 3        No            1           No White      <NA>
## 4        No            1           No White      <NA>
## 5       Yes            1           No White      <NA>
## 6        No            1          Yes White      <NA>
```

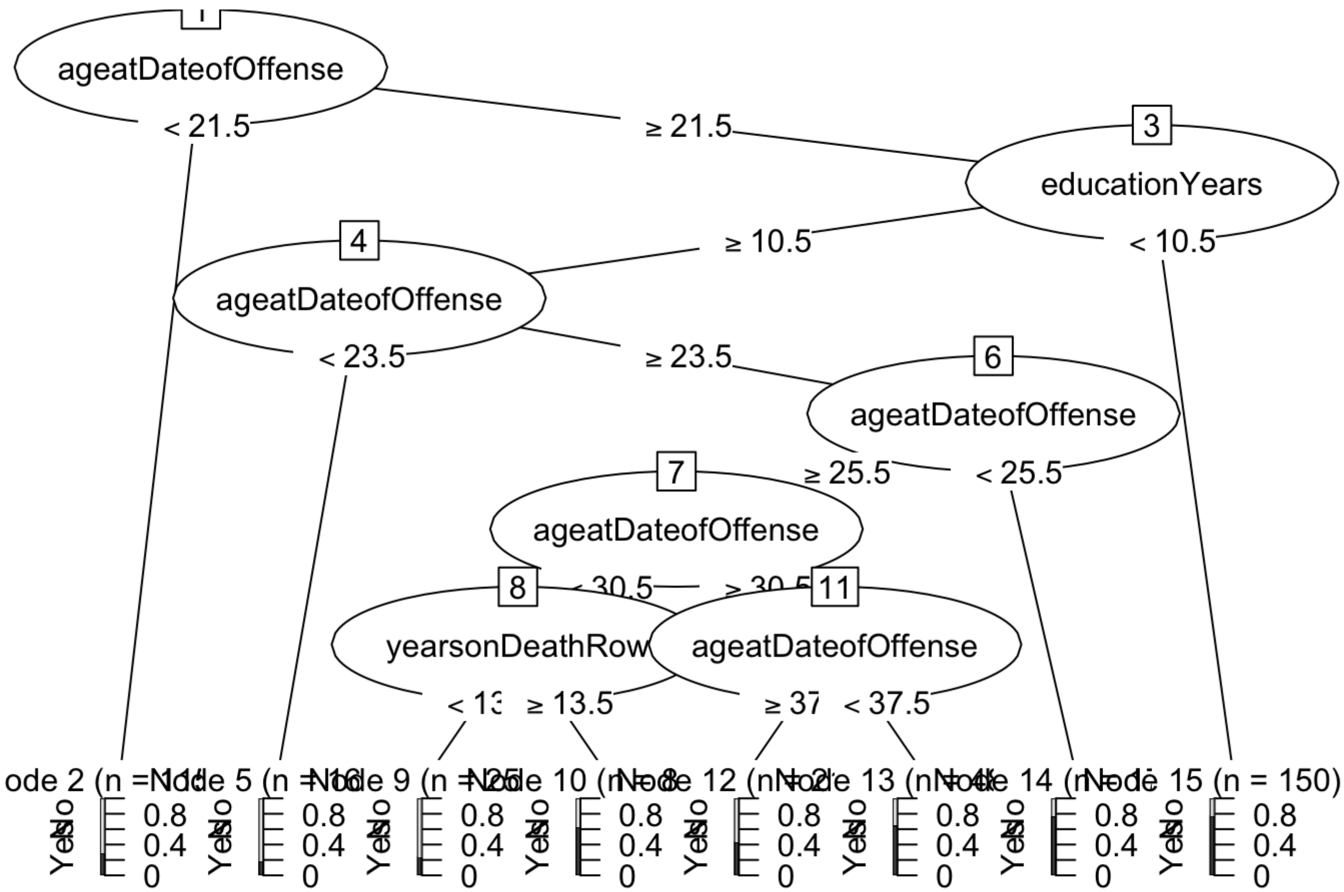
PriorRecordYes

Prep the dataset by setting the seed and splitting into test and training data.

```
#set the seed
set.seed(1842)

#save the number of rows in the dataset
n <- nrow(texas_noid)

#create the test_data dataset of 20% of the dataset number of rows
test_data <- sample.int(n, size = round(0.2*n))

#create the traning_data dataset as the remaining 80% of the dataset number of rows
training_data <- texas_noid[-test_data, ]
```

Build the first tree predicing priorRecordYes with all variables.

```
#first tree predicting variable with all the dataset variables
tree_prior_1 <- rpart(priorRecordYes~.,data=training_data)
plot(as.party(tree_prior_1))
```

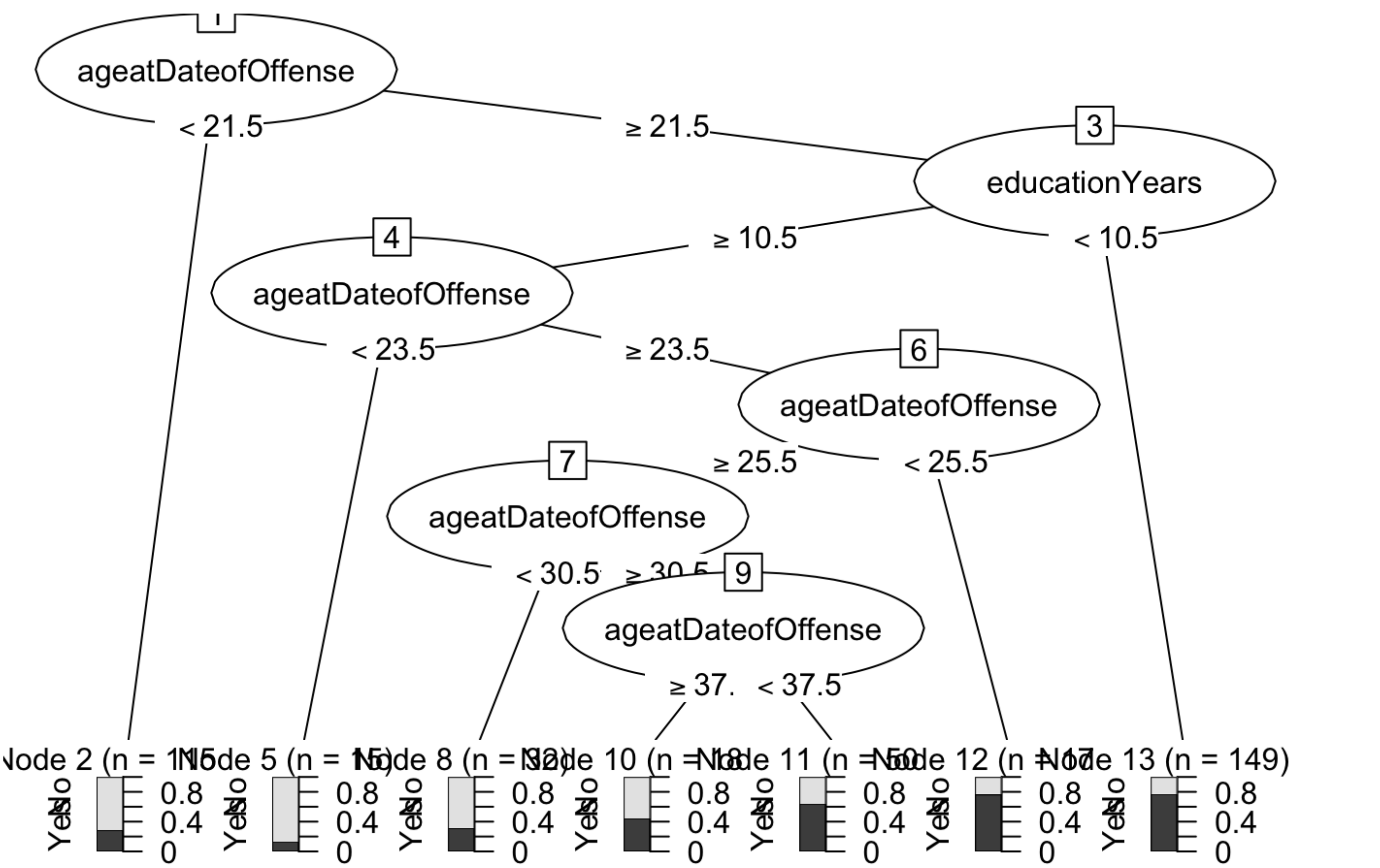



```
#CP / relative error of the first tree
printcp(tree_prior_1)
```

```
##
## Classification tree:
## rpart(formula = priorRecordYes ~ ., data = training_data)
##
## Variables actually used in tree construction:
## [1] ageatDateofOffense educationYears     yearsonDeathRow
##
## Root node error: 184/400 = 0.46
##
## n=400 (14 observations deleted due to missingness)
##
##         CP nsplit rel error  xerror     xstd
## 1 0.266304      0   1.00000 1.00000 0.054174
## 2 0.027174      1   0.73370 0.79348 0.052329
## 3 0.016304      5   0.61957 0.76087 0.051845
## 4 0.010870      6   0.60326 0.75000 0.051670
## 5 0.010000      7   0.59239 0.78261 0.052174
```

Build the second tree predicting priorRecordYes with just the top variables from the first tree.

```
#second tree predicting variable with the top predictor variables
tree_prior_2 <- rpart(priorRecordYes~educationYears+ageatDateofOffense,
                data=training_data)
plot(as.party(tree_prior_2))
```



```
#CP / relative error of the second tree
printcp(tree_prior_2)
```

```
##
## Classification tree:
## rpart(formula = priorRecordYes ~ educationYears + ageatDateofOffense,
##     data = training_data)
##
## Variables actually used in tree construction:
## [1] ageatDateofOffense educationYears
##
## Root node error: 183/396 = 0.46212
##
## n=396 (18 observations deleted due to missingness)
##
##         CP nsplit rel error  xerror     xstd
## 1 0.267760      0   1.00000 1.00000 0.054215
## 2 0.030055      1   0.73224 0.73224 0.051452
## 3 0.010929      5   0.60656 0.67760 0.050431
## 4 0.010000      6   0.59563 0.67760 0.050431
```

Then create a random forest with all variables as predictors, creating 100 trees.

```
#create a formula with the variables being analyzed
prior_formula_1 <- as.formula(priorRecordYes~., data(texas_noid))

#bagged set of trees with mtry=10 predictors, ntree=100
prior_part3 <- randomForest(prior_formula_1, data=training_data, mtry=10,
                      ntree=100,
                      na.action = na.roughfix)
prior_part3
```

```
##
## Call:
##  randomForest(formula = prior_formula_1, data = training_data,      mtry = 10, ntree
= 100, na.action = na.roughfix)
##                Type of random forest: classification
##                      Number of trees: 100
## No. of variables tried at each split: 8
##
##         OOB estimate of  error rate: 36.47%
## Confusion matrix:
##      No Yes class.error
## No  107  77   0.4184783
## Yes  74 156   0.3217391
```

```
#table of importance
importance(prior_part3)
```

```
##                     MeanDecreaseGini
## ageatDateofOffense        70.318379
## yearsonDeathRow           45.805566
## educationYears            35.468783
## codefsYes                  8.761624
## totalVictims              11.428513
## femaleVictim               8.064099
## race2                     18.243396
## metroArea                  4.917035
```

Create a random forest with the top variables from the first random forest as predictors, creating 100 trees.

```
#create a formula with the variables being analyzed
prior_formula_2 <- as.formula(priorRecordYes~ageatDateofOffense+educationYears+race2, da
ta(texas_noid))

#bagged set of trees with mtry=10 predictors, ntree=100
prior_part4 <- randomForest(prior_formula_2, data=training_data, mtry=10,
                      ntree=100,
                      na.action = na.roughfix)
prior_part4
```

```
##
## Call:
##  randomForest(formula = prior_formula_2, data = training_data,      mtry = 10, ntree
= 100, na.action = na.roughfix)
##                Type of random forest: classification
##                      Number of trees: 100
## No. of variables tried at each split: 3
##
##         OOB estimate of  error rate: 35.02%
## Confusion matrix:
##      No Yes class.error
## No  116  68   0.3695652
## Yes  77 153   0.3347826
```

```
#table of importance
importance(prior_part4)
```

```
##                     MeanDecreaseGini
## ageatDateofOffense         88.93478
## educationYears             47.97663
## race2                      26.84647
```

Random forest using the pruned decision tree variables.

```
#create a formula with the variables being analyzed
prior_formula_3 <- as.formula(priorRecordYes~educationYears+ageatDateofOffense, data(tex
as_noid))

#bagged set of trees with mtry=10 predictors, ntree=100
prior_part5 <- randomForest(prior_formula_3, data=training_data, mtry=10,
                      ntree=100,
                      na.action = na.roughfix)
prior_part5
```

```
##
## Call:
##  randomForest(formula = prior_formula_3, data = training_data,      mtry = 10, ntree
= 100, na.action = na.roughfix)
##                Type of random forest: classification
##                      Number of trees: 100
## No. of variables tried at each split: 2
##
##         OOB estimate of  error rate: 36.23%
## Confusion matrix:
##      No Yes class.error
## No  107  77   0.4184783
## Yes  73 157   0.3173913
```

```
#table of importance
importance(prior_part5)
```

```
##                      MeanDecreaseGini
## educationYears               45.02127
## ageatDateofOffense           82.87973
```

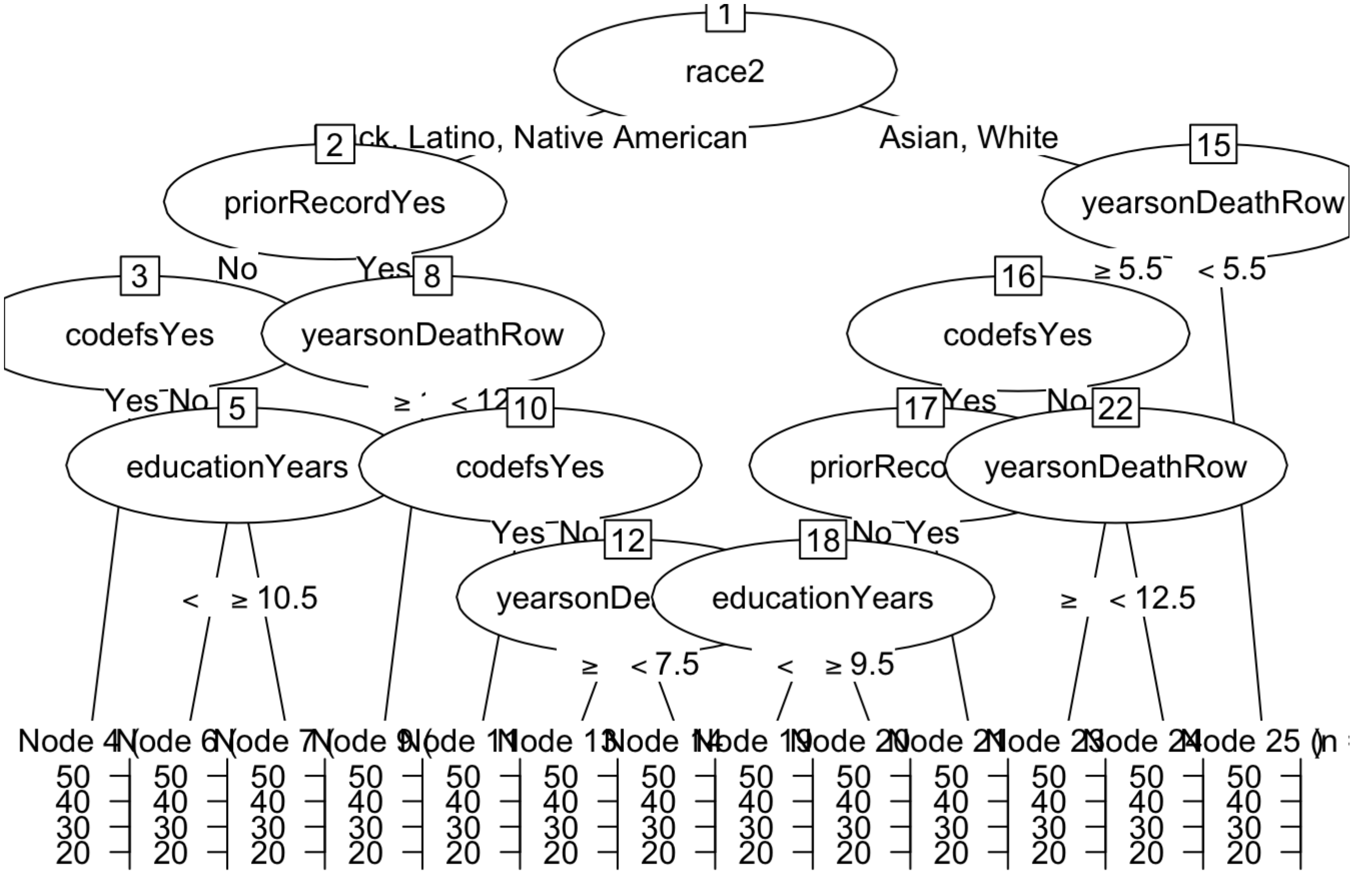
ageatDateofOffense

The training and test data have already been created.

```
#set the seed
set.seed(1842)
```

Build the first tree predicing AgeatDateofOffense with all variables.

```
#first tree predicting variable with all the dataset variables
tree_age_1 <- rpart(ageatDateofOffense~.,data=training_data)
plot(as.party(tree_age_1))
```

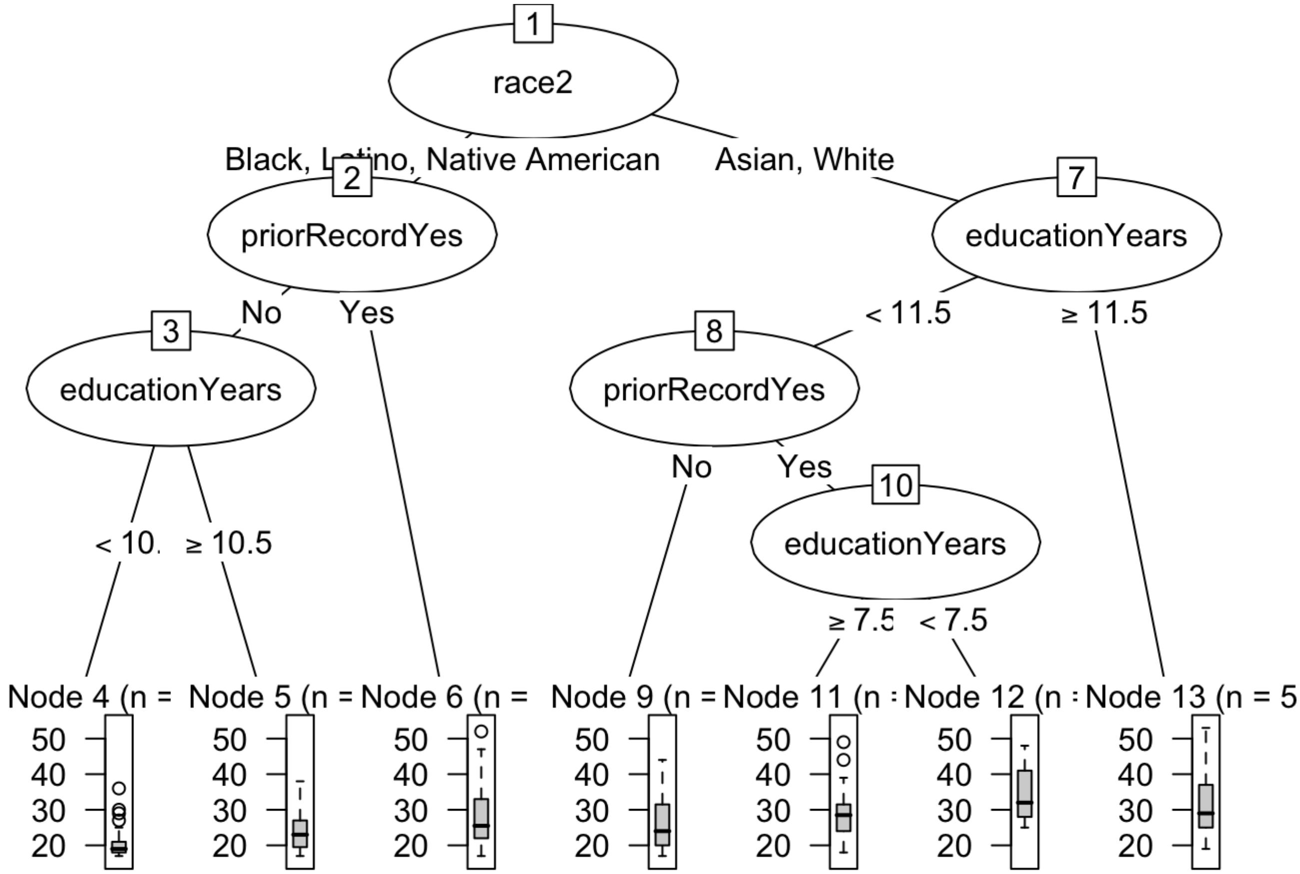



```
#CP / relative error of the first tree
printcp(tree_age_1)
```

```
##
## Regression tree:
## rpart(formula = ageatDateofOffense ~ ., data = training_data)
##
## Variables actually used in tree construction:
## [1] codefsYes       educationYears  priorRecordYes  race2
## [5] yearsonDeathRow
##
## Root node error: 21347/368 = 58.007
##
## n=368 (46 observations deleted due to missingness)
##
##          CP nsplit rel error  xerror     xstd
## 1  0.075196      0   1.00000 1.00668 0.079218
## 2  0.069706      1   0.92480 0.97534 0.077375
## 3  0.020499      2   0.85510 0.86651 0.069935
## 4  0.016044      3   0.83460 0.92908 0.073370
## 5  0.015458      4   0.81856 0.90437 0.072911
## 6  0.014004      6   0.78764 0.90411 0.072904
## 7  0.012772      7   0.77364 0.89102 0.073078
## 8  0.012745     10   0.73532 0.89595 0.073786
## 9  0.012508     11   0.72258 0.89595 0.073786
## 10 0.010000     12   0.71007 0.89072 0.073438
```

Build the second tree predicting AgeatDateofOffense with just the top variables from the first tree.

```
#second tree predicting variable with the top three predictor variables
tree_age_2 <- rpart(ageatDateofOffense~race2+priorRecordYes+educationYears,
                data=training_data)
plot(as.party(tree_age_2))
```




```
#CP / relative error of the second tree
printcp(tree_age_2)
```

```
##
## Regression tree:
## rpart(formula = ageatDateofOffense ~ race2 + priorRecordYes +
##     educationYears, data = training_data)
##
## Variables actually used in tree construction:
## [1] educationYears priorRecordYes race2
##
## Root node error: 21347/368 = 58.007
##
## n=368 (46 observations deleted due to missingness)
##
##         CP nsplit rel error  xerror     xstd
## 1 0.075196      0   1.00000 1.00888 0.079593
## 2 0.069706      1   0.92480 0.99549 0.082998
## 3 0.015078      2   0.85510 0.88061 0.071310
## 4 0.010036      3   0.84002 0.94822 0.077029
## 5 0.010000      6   0.80991 0.92791 0.075588
```

Then create a random forest with all variables as predictors, creating 100 trees.

```
#create a formula with the variables being analyzed
age_formula_1 <- as.formula(ageatDateofOffense~., data(texas_noid))

#bagged set of trees with mtry=10 predictors, ntree=100
age_part3 <- randomForest(age_formula_1, data=training_data, mtry=10,
                      ntree=100,
                      na.action = na.roughfix)
age_part3
```

```
##
## Call:
##  randomForest(formula = age_formula_1, data = training_data, mtry = 10,      ntree =
100, na.action = na.roughfix)
##                Type of random forest: regression
##                      Number of trees: 100
## No. of variables tried at each split: 8
##
##           Mean of squared residuals: 51.66314
##                     % Var explained: 0.43
```

```
#table of importance
importance(age_part3)
```

```
##                  IncNodePurity
## priorRecordYes        1744.251
## yearsonDeathRow       5420.332
## educationYears        4174.053
## codefsYes             1384.494
## totalVictims          1004.059
## femaleVictim          1027.639
## race2                 2411.062
## metroArea              623.000
```

Create a random forest with the top variables from the first random forest as predictors, creating 100 trees.

```
#create a formula with the variables being analyzed
age_formula_2 <- as.formula(ageatDateofOffense~educationYears+race2+priorRecordYes, data
(texas_noid))

#bagged set of trees with mtry=10 predictors, ntree=100
age_part4 <- randomForest(age_formula_2, data=training_data, mtry=10,
                      ntree=100,
                      na.action = na.roughfix)
age_part4
```

```
##
## Call:
##  randomForest(formula = age_formula_2, data = training_data, mtry = 10,      ntree =
100, na.action = na.roughfix)
##                Type of random forest: regression
##                      Number of trees: 100
## No. of variables tried at each split: 3
##
##           Mean of squared residuals: 51.0901
##                     % Var explained: 1.53
```

```
#table of importance
importance(age_part4)
```

```
##                IncNodePurity
## educationYears      3174.170
## race2               2054.520
## priorRecordYes      1621.262
```

Random forest using the pruned decision tree variables.

```
#create a formula with the variables being analyzed
age_formula_3 <- as.formula(ageatDateofOffense~race2+priorRecordYes+codefsYes, data(texa
s_noid))

#bagged set of trees with mtry=10 predictors, ntree=100
age_part5 <- randomForest(age_formula_3, data=training_data, mtry=10,
                      ntree=100,
                      na.action = na.roughfix)

age_part5
```

```
##
## Call:
##  randomForest(formula = age_formula_3, data = training_data, mtry = 10,      ntree =
100, na.action = na.roughfix)
##                Type of random forest: regression
##                      Number of trees: 100
## No. of variables tried at each split: 3
##
##           Mean of squared residuals: 44.34038
##                     % Var explained: 14.54
```

```
#table of importance
importance(age_part5)
```

```
##                  IncNodePurity
## race2                 1987.747
## priorRecordYes        1533.457
## codefsYes             1001.323
```

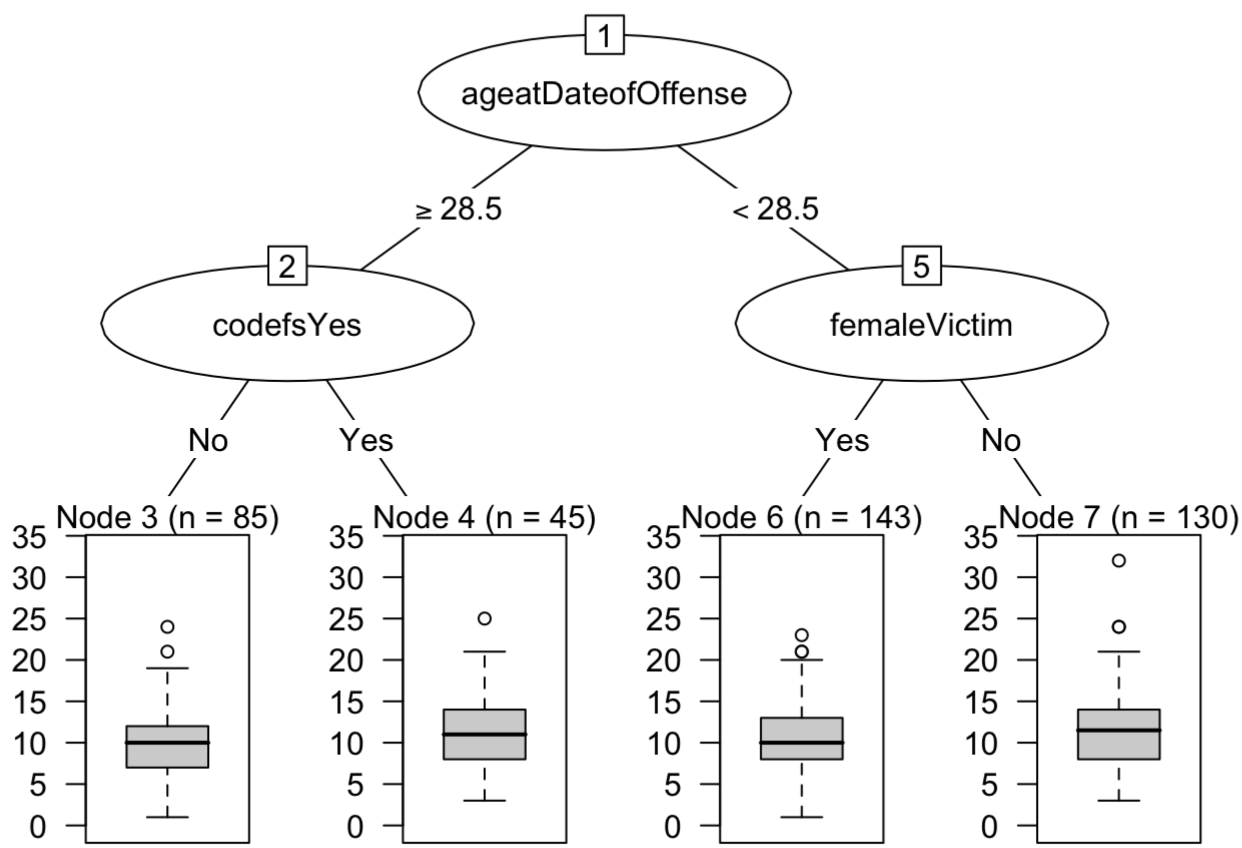
yearsonDeathRow

The training and test data have already been created.

```
#set the seed
set.seed(1842)
```

Build the first tree predicing sentenceCount with all variables.

```
#first tree predicting variable with all the dataset variables
tree_years_1 <- rpart(yearsonDeathRow~.,data=training_data)
plot(as.party(tree_years_1))
```




```
#CP / relative error of the first tree
printcp(tree_years_1)
```

```
##
## Regression tree:
## rpart(formula = yearsonDeathRow ~ ., data = training_data)
##
## Variables actually used in tree construction:
## [1] ageatDateofOffense codefsYes          femaleVictim
##
## Root node error: 7377.9/403 = 18.307
##
## n=403 (11 observations deleted due to missingness)
##
##         CP nsplit rel error xerror     xstd
## 1 0.010356      0   1.00000 1.0048 0.096232
## 2 0.010000      3   0.96893 1.1037 0.100807
```

Build the second tree predicting sentenceCount with just the top variables from the first tree.

```
#second tree predicting variable with the top three predictor variables
tree_years_2 <- rpart(yearsonDeathRow~ageatDateofOffense+codefsYes+femaleVictim,
                data=training_data)

#CP / relative error of the second tree
printcp(tree_years_2)
```

```
##
## Regression tree:
## rpart(formula = yearsonDeathRow ~ ageatDateofOffense + codefsYes +
##     femaleVictim, data = training_data)
##
## Variables actually used in tree construction:
## [1] ageatDateofOffense codefsYes          femaleVictim
##
## Root node error: 7377.9/403 = 18.307
##
## n=403 (11 observations deleted due to missingness)
##
##         CP nsplit rel error xerror     xstd
## 1 0.011419      0   1.00000 1.0075 0.096843
## 2 0.010887      7   0.92007 1.0569 0.098340
## 3 0.010000      8   0.90918 1.0547 0.098572
```

Then create a random forest with all variables as predictors, creating 100 trees.

```
#create a formula with the variables being analyzed
years_formula_1 <- as.formula(yearsonDeathRow~., data(texas_noid))

#bagged set of trees with mtry=10 predictors, ntree=100
years_part3 <- randomForest(years_formula_1, data=training_data, mtry=10,
                      ntree=100,
                      na.action = na.roughfix)
years_part3
```

```
##
## Call:
##  randomForest(formula = years_formula_1, data = training_data,      mtry = 10, ntree
= 100, na.action = na.roughfix)
##                Type of random forest: regression
##                      Number of trees: 100
## No. of variables tried at each split: 8
##
##           Mean of squared residuals: 18.84644
##                     % Var explained: -5.65
```

```
#table of importance
importance(years_part3)
```

```
##                     IncNodePurity
## priorRecordYes          289.4638
## ageatDateofOffense     2239.9106
## educationYears         1428.9110
## codefsYes               418.9085
## totalVictims            402.7893
## femaleVictim            344.7897
## race2                   568.6063
## metroArea               258.0482
```

Create a random forest with the top variables from the first random forest as predictors, creating 100 trees.

```
#create a formula with the variables being analyzed
years_formula_2 <- as.formula(yearsonDeathRow~ageatDateofOffense+educationYears+race2, d
ata(texas_noid))

#bagged set of trees with mtry=10 predictors, ntree=100
years_part4 <- randomForest(years_formula_2, data=training_data, mtry=10,
                      ntree=100,
                      na.action = na.roughfix)
years_part4
```

```
##
## Call:
##  randomForest(formula = years_formula_2, data = training_data,      mtry = 10, ntree
= 100, na.action = na.roughfix)
##                Type of random forest: regression
##                      Number of trees: 100
## No. of variables tried at each split: 3
##
##           Mean of squared residuals: 22.07825
##                     % Var explained: -23.76
```

```
#table of importance
importance(years_part4)
```

```
##                      IncNodePurity
## ageatDateofOffense      2399.789
## educationYears          1782.550
## race2                    689.892
```

Random forest using the pruned decision tree variables.

```
#create a formula with the variables being analyzed
years_formula_3 <- as.formula(yearsonDeathRow~ageatDateofOffense+codefsYes+femaleVictim,
 data(texas_noid))

#bagged set of trees with mtry=10 predictors, ntree=100
years_part5 <- randomForest(years_formula_3, data=training_data, mtry=10,
                      ntree=100,
                      na.action = na.roughfix)
years_part5
```

```
##
## Call:
##  randomForest(formula = years_formula_3, data = training_data,      mtry = 10, ntree
= 100, na.action = na.roughfix)
##                Type of random forest: regression
##                      Number of trees: 100
## No. of variables tried at each split: 3
##
##           Mean of squared residuals: 19.36584
##                     % Var explained: -8.56
```

```
#table of importance
importance(years_part5)
```

```
##                      IncNodePurity
## ageatDateofOffense     2357.2415
## codefsYes               425.2537
## femaleVictim            396.6434
```