Week1

Zai Rutter

1/15/2022

Question 1

```
###### Question 1

### A
AlabamaCourt <- read_csv("AlabamaCourt.csv")
```

```
## Rows: 2926 Columns: 14
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (4): race, sex, court_action, atty
## dbl (10): person, county_num, case_year, dob_year, amountpaid, amountdue, pr...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
AC<-AlabamaCourt

### B
AC$black <- ifelse(AC$race == "B", 1,0)
AC$amountremain <- (AC$amountdue - AC$amountpaid)

### C

## African-American
  AC.Black<- AC %>%
  filter(black == 1)
  stargazer(as.data.frame(AC.Black[c("amountremain", "amountdue","amountpaid")]),
            type = "text")
```

```
##
## ===========================================================
## Statistic      N     Mean     St. Dev.   Min      Max
## -----------------------------------------------------------
## amountremain 1,366 3,568.939 10,222.090 0.000 161,076.000
## amountdue    1,366 4,040.682 10,339.030 0.000 163,642.000
## amountpaid   1,366  471.743  1,095.375  0.000 20,366.000
## -----------------------------------------------------------
```

```r
## Non-Black
  AC.NonBlack <- AC %>%
    filter(black != 1)
  stargazer(as.data.frame(AC.NonBlack[c("amountremain", "amountdue","amountpaid")]),
            type = "text")
```

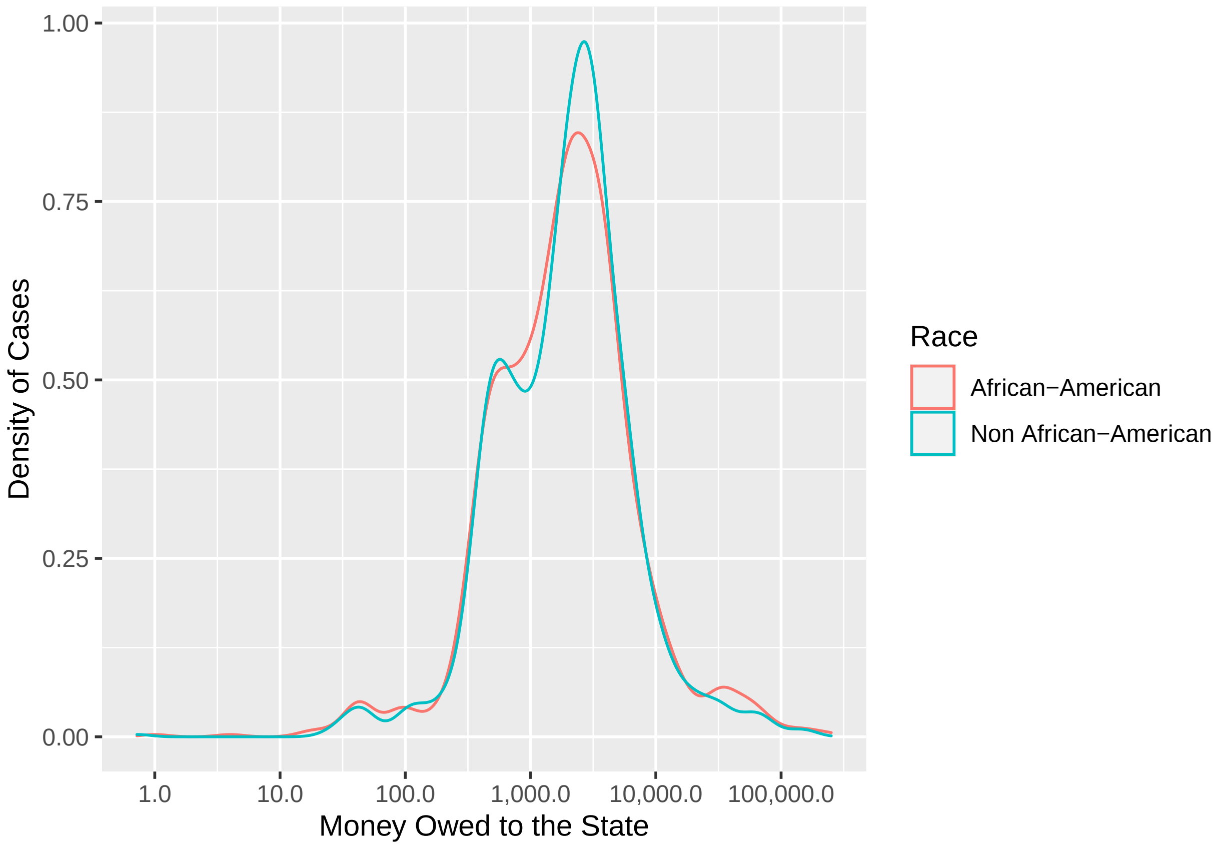
```
##
## ==========================================================
## Statistic      N     Mean     St. Dev.   Min      Max
## ----------------------------------------------------------
## amountremain 1,560 3,582.679 12,659.160 0.000 251,650.000
## amountdue    1,560 4,398.811 12,997.500 0.000 262,245.000
## amountpaid   1,560  816.132  1,834.353  0.000 40,328.000
## ----------------------------------------------------------
```

```r
### D
## Create three kernel density plots that compare the distributions
  ## of the variables
## amountremain", \amountdue", \amountpaid",
## respectively, for African-Americans and non-African-Americans.

ggplot(AC, aes(x=amountremain, colour = (black==1))) +
  geom_density() +
  scale_x_continuous(trans = 'log10',
                     breaks=c(1,10,100,1000,10000,100000),
                     labels=comma,name="Money Owed to the State") +
  ylab("Density of Cases") +
  scale_color_discrete(name="Race",
                       labels=c("African-American", "Non African-American"))
```

```
## Warning: Transformation introduced infinite values in continuous x-axis
```

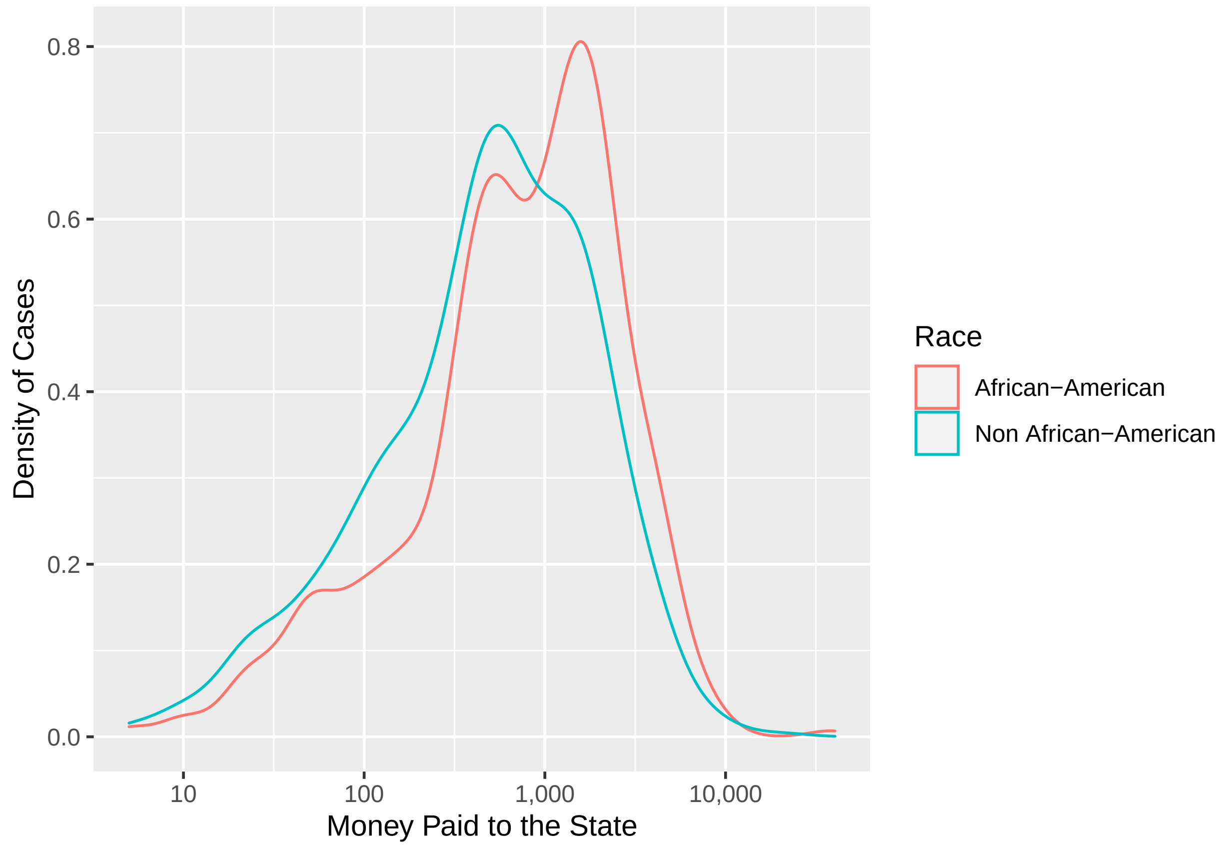
```
## Warning: Removed 657 rows containing non-finite values (stat_density).
```




```
ggplot(AC, aes(x=amountpaid, colour = (black==1))) +
  geom_density() +
  scale_x_continuous(trans = 'log10',
                     breaks=c(1,10,100,1000,10000,100000),
                     labels=comma,name="Money Paid to the State") +
  ylab("Density of Cases") +
  scale_color_discrete(name="Race",
                       labels=c("African-American", "Non African-American"))
```

```
## Warning: Transformation introduced infinite values in continuous x-axis
```

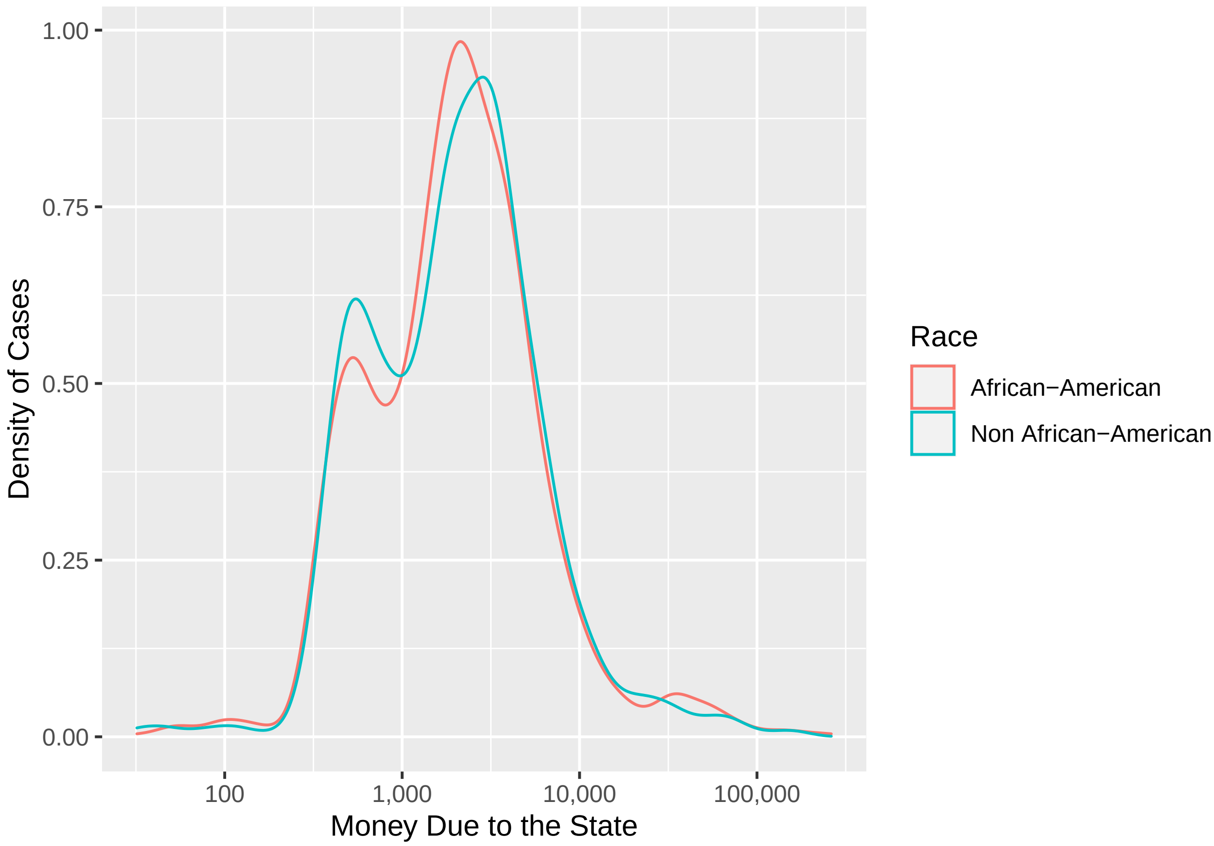
```
## Warning: Removed 1343 rows containing non-finite values (stat_density).
```




```r
ggplot(AC, aes(x=amountdue, colour = (black==1))) +
  geom_density() +
  scale_x_continuous(trans = 'log10',
                     breaks=c(1,10,100,1000,10000,100000),
                     labels=comma,name="Money Due to the State") +
  ylab("Density of Cases") +
  scale_color_discrete(name="Race",
                       labels=c("African-American", "Non African-American"))
```

```
## Warning: Transformation introduced infinite values in continuous x-axis
```

```
## Warning: Removed 69 rows containing non-finite values (stat_density).
```




```
### E

# In the graph "Paid", it appears that there is significantly higher peak that African Americans
# have paid higher sums. The lines are slighly dissimilar prior to that mark with more non blacks
# have paid slightly less sums. After that $1000 mark African Americans seem to pay more than the rest
# In the graph "Due" the graphs are generally similar with slight differences.
# Around $1000 non-blacks have been charged more where as Afrian Americans have more been charged aroun
# 5,000 mark.
# In the graph "owed" non blacks owe more than African Americans.

# According to these numbers it appears that while non-blacks owe more, they have also paid less.
# The most telling graph would "Money due" where there is a slight difference in amount charged to Afri
# I think more data about income disparity could build a more telling story about racist policies in LF
```

Question 2

```
CollegeBasketball <- read_csv("CollegeBasketball.csv")
```

```
## Rows: 241 Columns: 9
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): Date, Favorite, Underdog
## dbl (6): Favorite3, Underdog3, PredictedDifference, PredictedPoints, ActualD...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
basketball<-CollegeBasketball

### A

basketball$super <- (basketball$PredictedDifference - basketball$ActualDifference)
mean(abs(basketball$super))
```

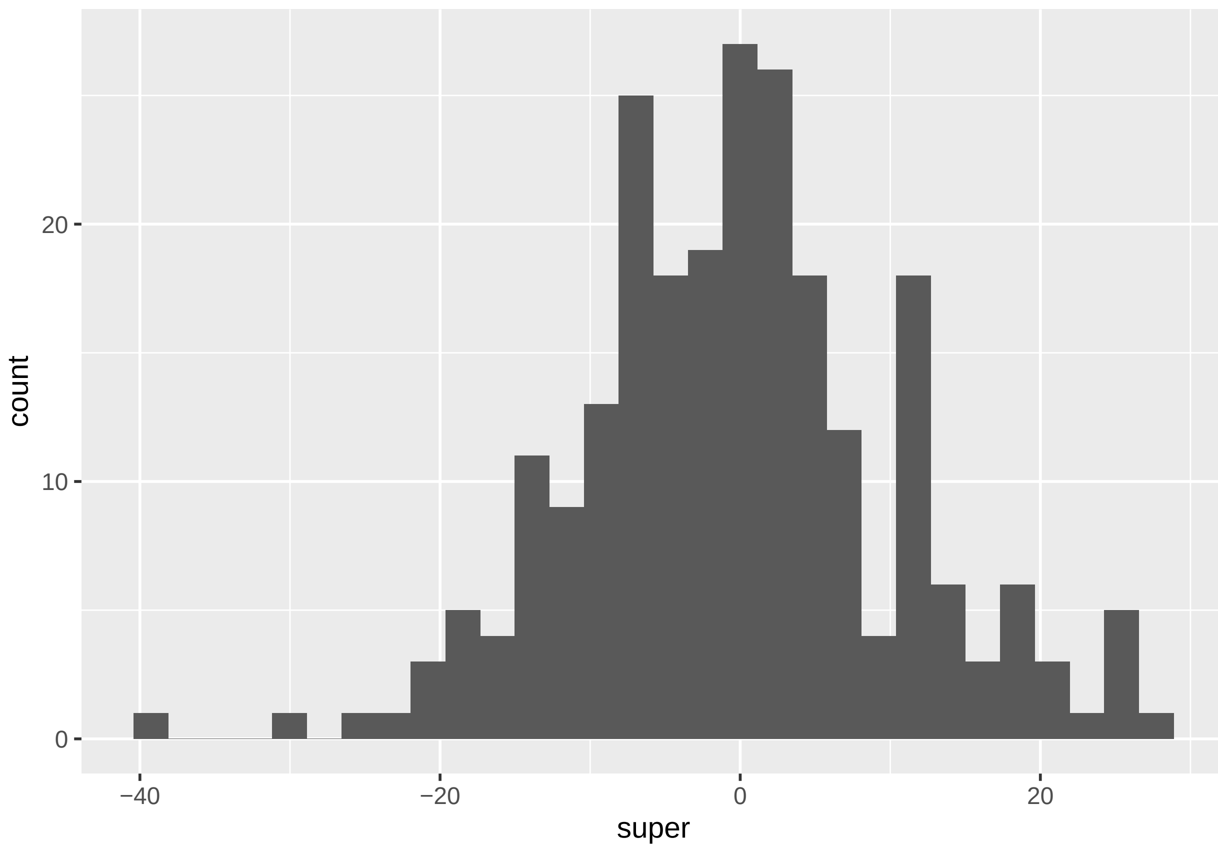
```
## [1] 8.26971
```

```r
abs(mean(basketball$super))
```

```
## [1] 0.1659751
```

```r
ggplot(basketball, aes(x=super)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```r
### B

basketball$Differential <- rep(NA, length(basketball$Favorite))

for(i in 1:length(basketball$Differential)){
```

```r
  basketball$Differential[i] <-if(basketball$super[i] < 0 ) {
    "W"
  } else if (basketball$super[i] == 0)  {
    "E"
  } else if (basketball$super[i] > 0 )  {
    "L"
  } else {}

}

mean(basketball$Differential=="W")
```

```
## [1] 0.5020747
```

```r
mean(basketball$Differential=="E")
```

```
## [1] 0.02904564
```

```r
mean(basketball$Differential=="L")
```

```
## [1] 0.4688797
```

```r
### C
basketball$PointDiffer <- (basketball$PredictedPoints - basketball$ActualPoints)

### D

basketball$ptsdummy <- rep(NA, length(basketball$Favorite))

for(i in 1:length(basketball$ptsdummy)){

  basketball$ptsdummy[i] <- if(basketball$PointDiffer[i] == 0 ) {
    "T"
  } else if(basketball$PointDiffer[i] > 0 )  {
    "F"
  }
  else if(basketball$PointDiffer[i] < 0) {
    "M"
  }
  else {}
}

mean(basketball$ptsdummy=="F")
```

```
## [1] 0.5726141
```

```r
mean(basketball$ptsdummy=="T")
```

```
## [1] 0.04149378
```

```r
mean(basketball$ptsdummy=="M")
```

```
## [1] 0.3858921
```

```r
mean(basketball$PointDiffer == 0)
```

```
## [1] 0.04149378
```

```r
### E

# Prob W | M
# the prob. that fav won by more points than expected *when* More pionts were scored than expected
mean(basketball$ptsdummy=="M" & basketball$Differential=="W")/mean(basketball$ptsdummy=="M")
```

```
## [1] 0.4301075
```

```r
# Prob L | M
# Fav earned less points than expected *when* more points were scored than expected
mean(basketball$ptsdummy=="M" & basketball$Differential=="L")/ mean(basketball$ptsdummy=="M")
```

```
## [1] 0.5483871
```

```r
# Prob W | F
# the prob. that fav won by more points than expected *when* less points were scored than expected

mean( basketball$ptsdummy=="F"& basketball$Differential=="W") / mean(basketball$Differential=="W")
```

```
## [1] 0.6363636
```

```r
# Prob L | F
# Fav earned less points than expected *when* fav won by more points than expected

mean(basketball$ptsdummy=="F" & basketball$Differential=="L") / mean(basketball$ptsdummy=="F")
```

```
## [1] 0.4057971
```

```r
### F
# Write a paragraph or two, in which you make conclusions about whether the
# evidence is consistent with my theory based on the data being summarized in
# the previous parts of this question.

# There is some consistency, as shown below by the numbers.

# Fav get more pionts when less combined points
0.6363636
```

```
## [1] 0.6363636
```

```r
# the prob. that fav won by more points than expected *when* More pionts were scored than expected
0.4301075
```

```
## [1] 0.4301075
```

```r
# Fav earned less points than expected *when* fav won by more points than expected
0.4057971
```

```
## [1] 0.4057971
```

```r
# W = Fav won by more points than expected

# E = Fav won by exact points they were expected

# L = Fav earned fewer points than expected

# M = More combined points were scored than expected

# T = combined points were exact as expected

# F = Less combined pionts were scored than expected
```