RDD Assignment 1

Ben Carlton

3/5/2021

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.5     v dplyr   1.0.3
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(rdrobust)
library(rdd)
```

```
## Loading required package: sandwich
```

```
## Loading required package: lmtest
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Loading required package: AER
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

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


## The following object is masked from 'package:purrr':
##
##     some


## Loading required package: survival


## Loading required package: Formula
```

```r
library(rddensity)
library(jtools)
library(dplyr)
library(readxl)
library(ggplot2)
library(stargazer)
```

```
##
## Please cite as:


##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.


##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```r
hansen <- read_excel("C:/Users/benca/Desktop/hansen_dwi.xlsx")
hansen$dui <- ifelse(hansen$bac1 >= 0.08, 1, 0)

bac4 <- hansen%>% filter( between(bac1, 0, .15) )
bac5 <- hansen%>% filter( between(bac1, .03, .13) )
bac6 <- hansen%>% filter( between(bac1, 0.55, .105) )
```

1

https://github.com/benjcarlton/RDD

2

Hansen's paper's objective is to explore the causal effect of harsher DUI penalties on the likelihood of repeat DUI offenses. Using Washington State's administrative records on DUI stops from 1997 to 2011 and filters out instances of stops involving persons under the age of 21. With this data, he explores the likelihood of repeat offenses (recidivism) and analyzes the relationship between BAC at the time of the stop and characteristics that the people in question cannot change like gender, age, and race. He establishes BAC cutoff points of .08 (DUI) and .15 (aggrevated DUI) to see how these penalties correlate with repeat offenses within 4 years of the initial stop. Utilizing a variety of statistical checks lHe finds evidence that 10 percent increase in sanctions and punishments is associated with a 2.3 percent decline in drunk driving. and found that having a BAC over the 0.15 limit is associated with an additional 1 percentage point decline in repeat drunk driving.

3

R-code used: hansen$dui <- ifelse(hansen$bac1 >= 0.08, 1, 0)

4

Given that blood-alcohol-content is something that a person can directly control (that is people choose their amount of alcohol consumption), Hansen worries about using BAC at the time of the traffic stop since people in the dataset may have some level of control that could skew his findings. To test if this is the case, he plots the BAC levels of those at the time of the initial stop. If manipulation is occurring, one would expect to see a "bunching" (disproportionate grouping) of observations around key cutoff points (BAC of .08 and BAC of .15). Looking at BAC levels of the observations in a histogram format shows the following:

```
hansen$dui <- ifelse(hansen$bac1 >= 0.08, 1, 0)
ggplot(hansen) + geom_histogram(aes(x=bac1), bins = 1000)+
stat_bin(binwidth = .001, aes(x=bac1)) + geom_vline(aes(xintercept = 0.08, color = "Red"))
```
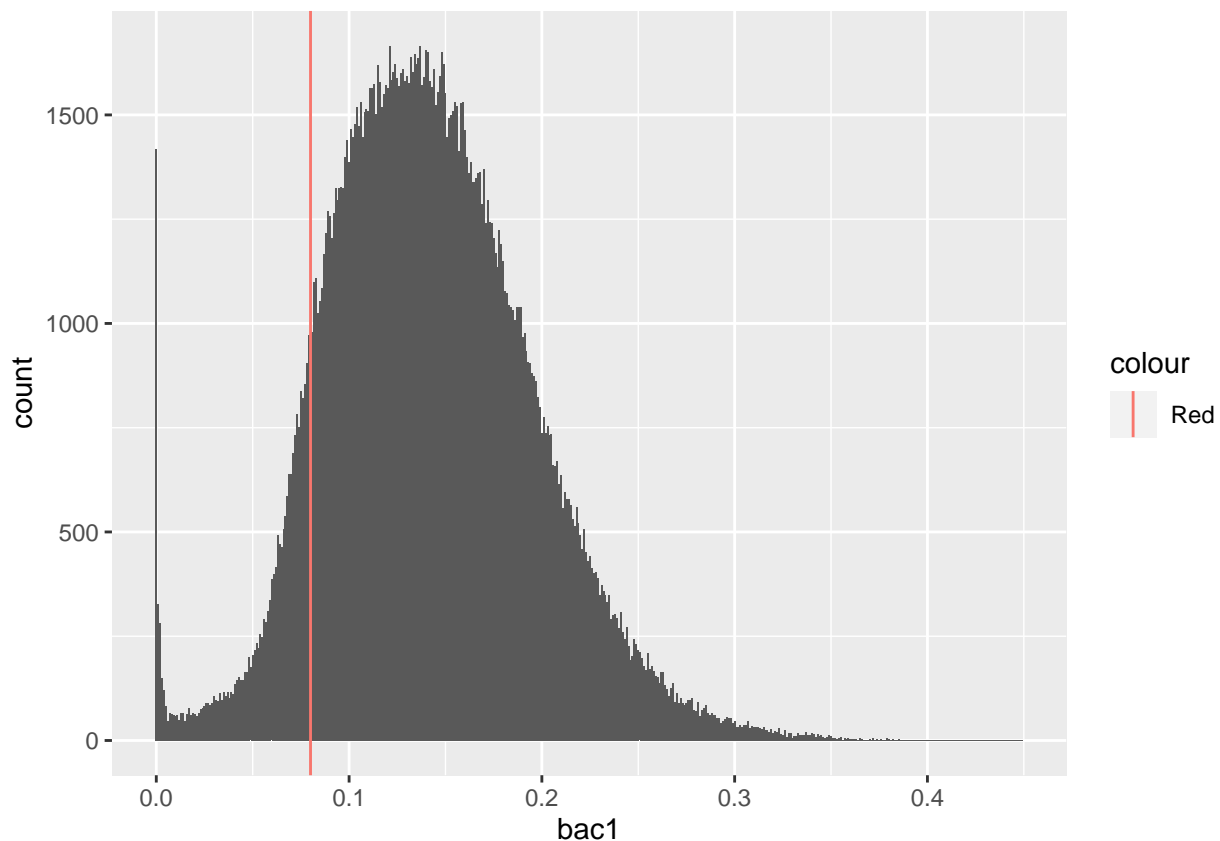

While there are some regions of the graph that appear to be "bunched together", the overall curvature is relatively smooth and does not suggest the kind of self-selection right before the .08 and .15 cutoff points that we would expect to see if people were able to exercise some level of control over their BAC at the time of the stop. Using the rddensity command in R shows the following:

```
density <- rddensity(hansen$bac1, c = 0.08)
rdplotdensity(density, hansen$bac1)
```
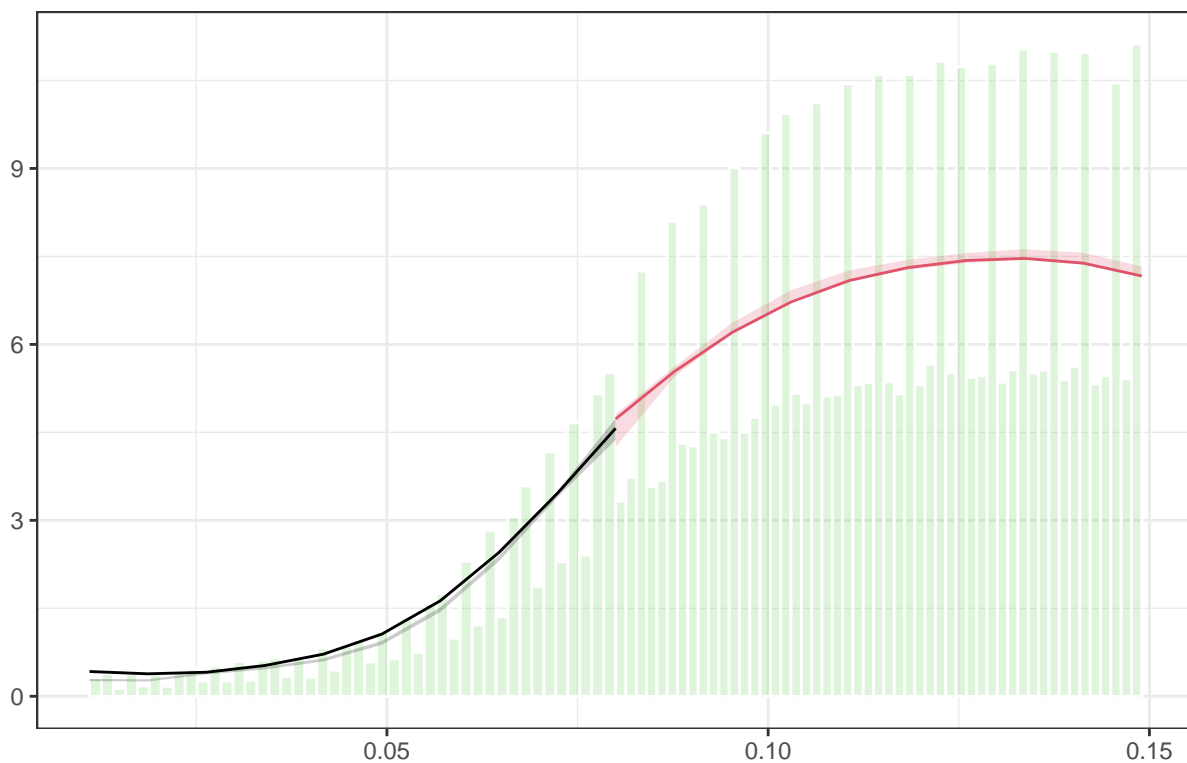
```
## $Estl
## Call: lpdensity
##
## Sample size                            23010
## Polynomial order for point estimation    (p=)   2
## Order of derivative estimated            (v=)   1
## Polynomial order for confidence interval (q=)   3
## Kernel function                          triangular
## Scaling factor                           0.107239568040194
## Bandwidth method                         user provided
##
## Use summary(...) to show estimates.
##
## $Estr
## Call: lpdensity
##
## Sample size                            191548
## Polynomial order for point estimation    (p=)   2
## Order of derivative estimated            (v=)   1
## Polynomial order for confidence interval (q=)   3
## Kernel function                          triangular
## Scaling factor                           0.892755771193669
## Bandwidth method                         user provided
##
## Use summary(...) to show estimates.
##
```

$Estplot
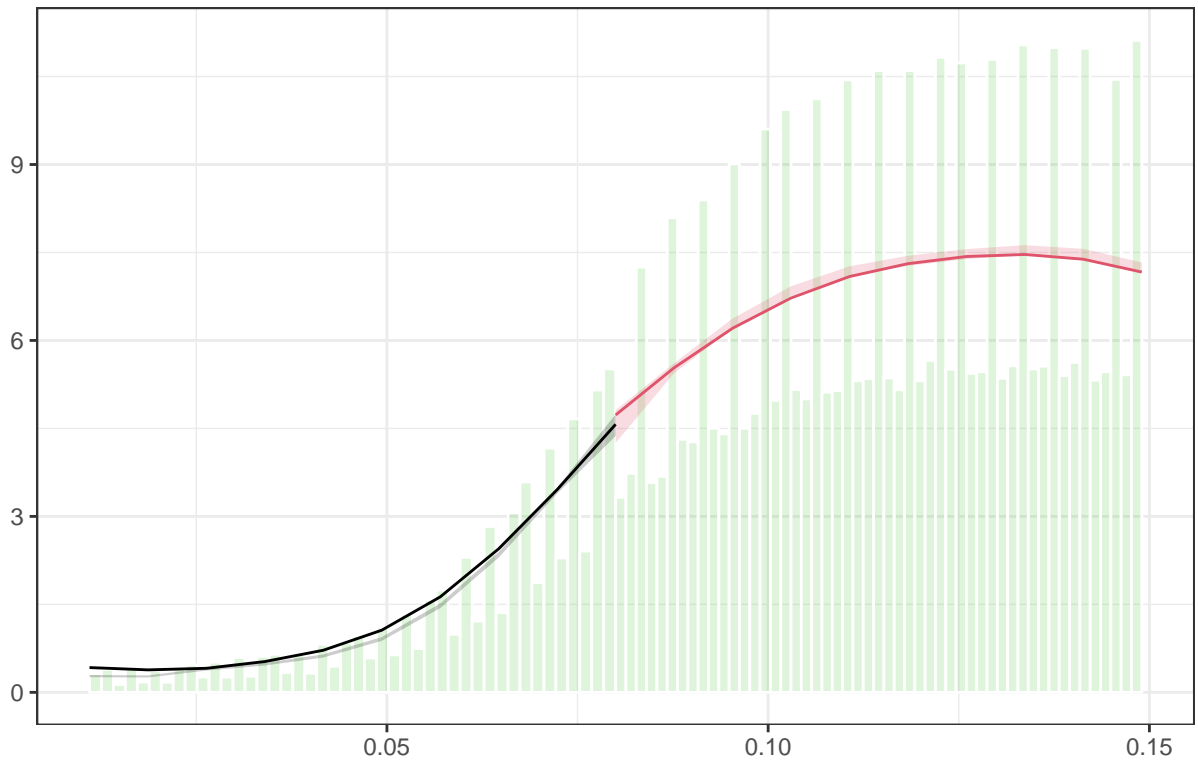
Even though the line is not perfectly continuous at the 0.08 cutoff point, the "gap" between the two lines is within the interval shown at the cutoff point, so there does not appear to be any sorting within the running variable.

5

```
model_a <- lm(male ~ dui + bac1 + dui*bac1, data = hansen)
model_b <- lm(white ~ dui + bac1 + dui*bac1, data = hansen)
model_c <- lm(aged ~ dui + bac1 + dui*bac1, data = hansen)
model_d = lm(acc ~ dui + bac1 + dui*bac1, data = hansen)

stargazer(model_a, model_b, model_c, model_d,
  title="Driver demographic characteristics",
  dep.var.caption ="DUI, BAC, Interaction",
  covariate.labels = c("DUI", "BAC", "Interaction", "Constant"),
  notes.label="Significance Levels",
  type = "text"
  )
```

```
##
## Driver demographic characteristics
## =============================================================================
```

```
##                                       DUI, BAC, Interaction
##                         ------------------------------------------
##                            male     white      aged       acc
##                             (1)       (2)       (3)       (4)
## -----------------------------------------------------------------
## DUI                      0.031***   0.003   -7.787***  -0.219***
##                          (0.007)   (0.006)   (0.204)    (0.006)
##
## BAC                      0.218**    0.154*  -56.361***  -1.540***
##                          (0.108)   (0.092)   (3.038)    (0.093)
##
## Interaction             -0.311***   0.017   83.400***   2.656***
##                          (0.110)   (0.093)   (3.089)    (0.095)
##
## Constant                 0.773***  0.835***  38.571***  0.201***
##                          (0.007)   (0.006)   (0.184)    (0.006)
##
## -----------------------------------------------------------------
## Observations             214,558   214,558   214,558    214,558
## R2                       0.0001     0.001     0.013      0.021
## Adjusted R2              0.0001     0.001     0.013      0.021
## Residual Std. Error (df = 214554)   0.408    0.345     11.431     0.351
## F Statistic (df = 3; 214554)     9.428***  58.401*** 907.467*** 1,566.761***
## =================================================================
## Significance Levels                       *p<0.1; **p<0.05; ***p<0.01
```

Thinking about the methodology used by Hansen in dettermining which data to use, as well as the coefficients generated from the table above leads me to believe that is minimal bias within the covariates used as controls for this study. Hansen's use of traffic stops, many resulting in BACs well below the DUI threshold, indicates to me that the dataset is more likely than not comprised of a random sample across all of the "control" attributes.
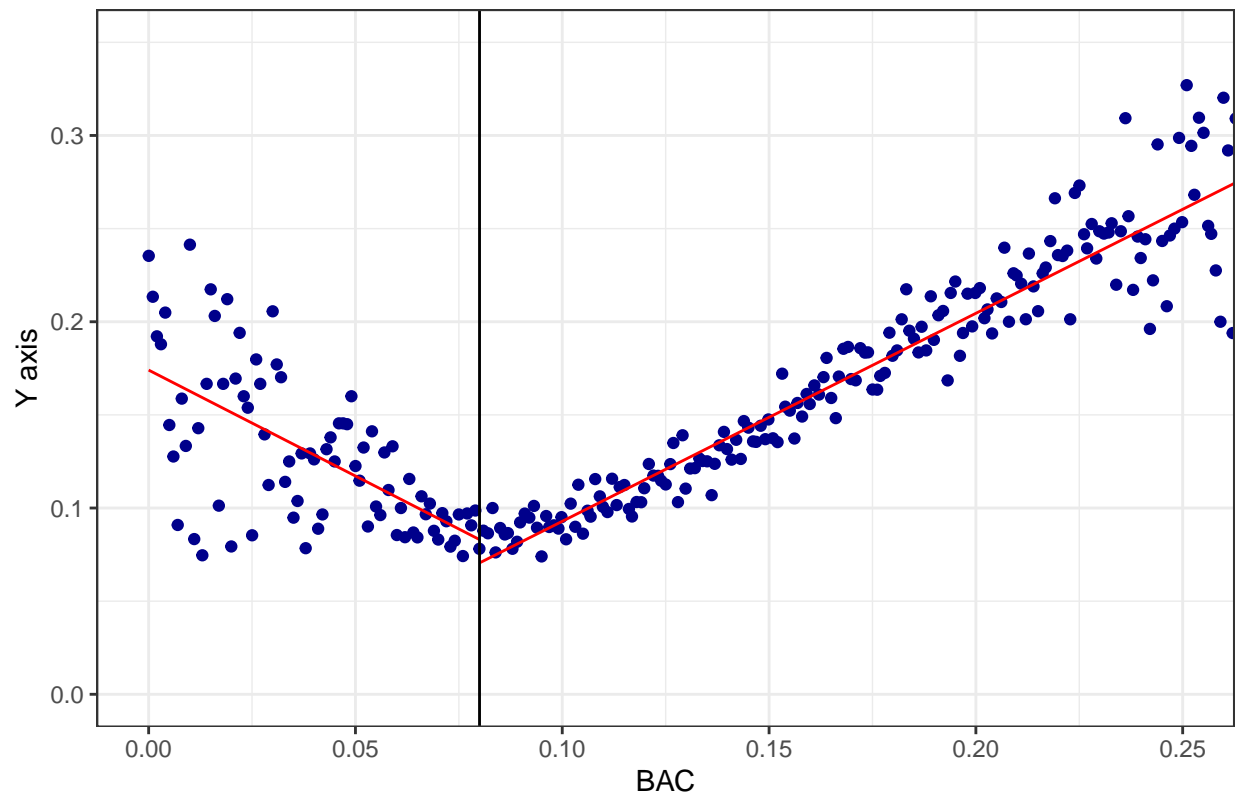
6

Panel A Linear

```
rdplot(hansen$acc, hansen$bac1, p = 1, c = 0.08, title = "Panel A. Accidents at scene",  x.label = 'BAC
```

```
## [1] "Mass points detected in the running variable."
```
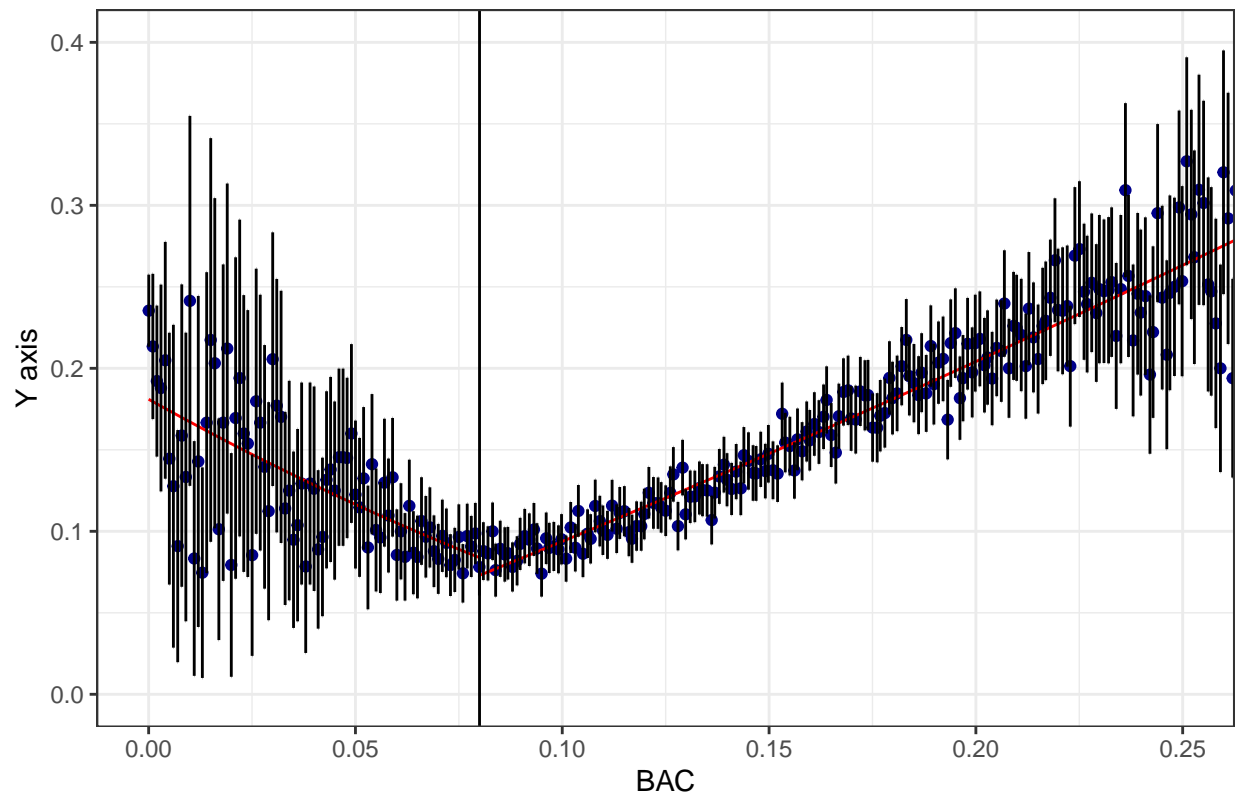
Panel A. Accidents at scene

Panel A Quadratic fit

```
rdplot(hansen$acc, hansen$bac1, p = 2, c = 0.08, ci = 95, title = "Panel A. Accidents at scene", x.label
```

```
## [1] "Mass points detected in the running variable."
```

Panel A. Accidents at scene

While not an exact match, I do think that I was able to come pretty close to replicating Panel A. I am working on figuring out how to group several of these types of plots together in the format that Hansen found

Panel B Linear

```
rdplot(hansen$male, hansen$bac1, p =1, c = 0.08, title = "Panel B. Male", x.label = 'BAC', kernel = "tr
```

```
## [1] "Mass points detected in the running variable."
```

Panel B. Male
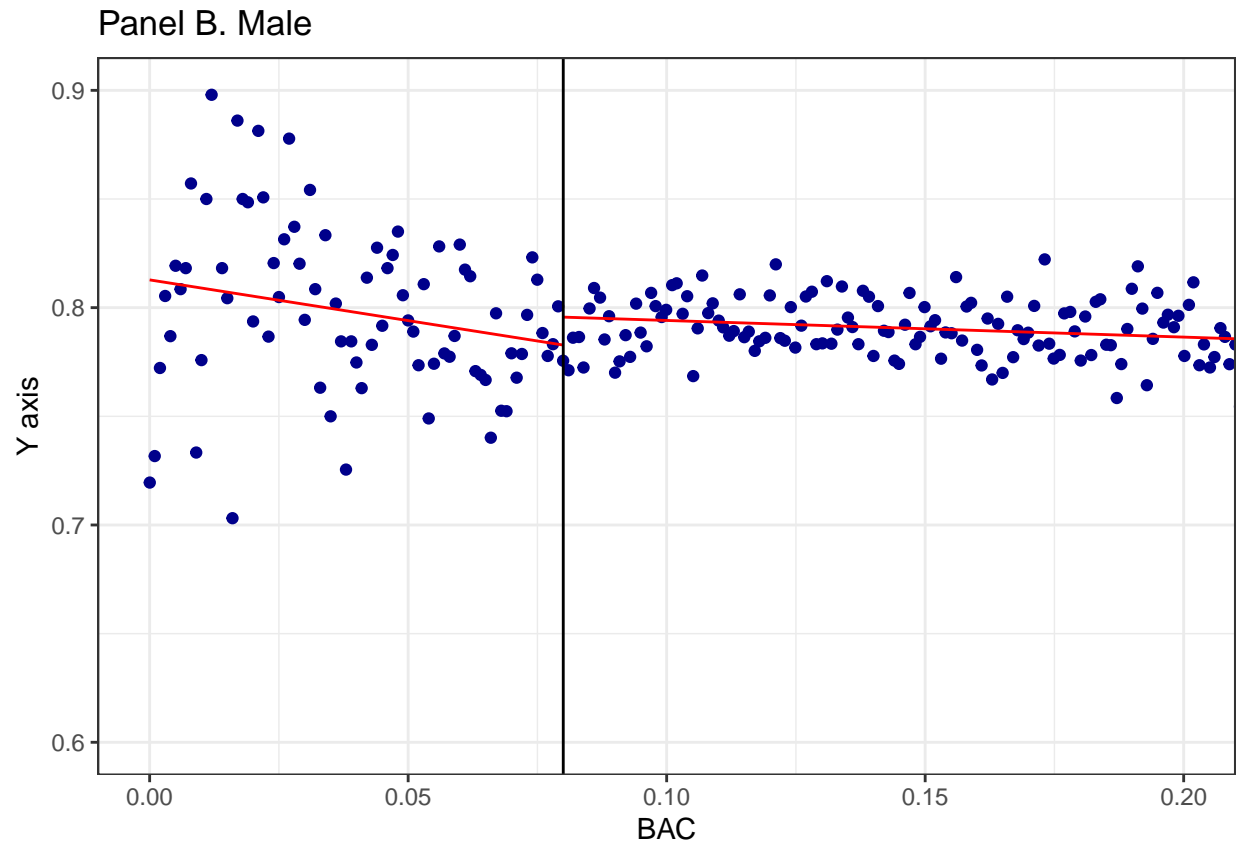
While the axis min&max are not the same, I believe it is due to: 1. Not having the entire dataset that he used; and 2. the number of bins that Hansen used in his paper being different from the number generated in Panel B.

Panel B Quandratic fit

```
rdplot(hansen$male, hansen$bac1, p =2, c = 0.08, title = "Panel B. Male", x.label = 'BAC', kernel = "tr
```

```
## [1] "Mass points detected in the running variable."
```
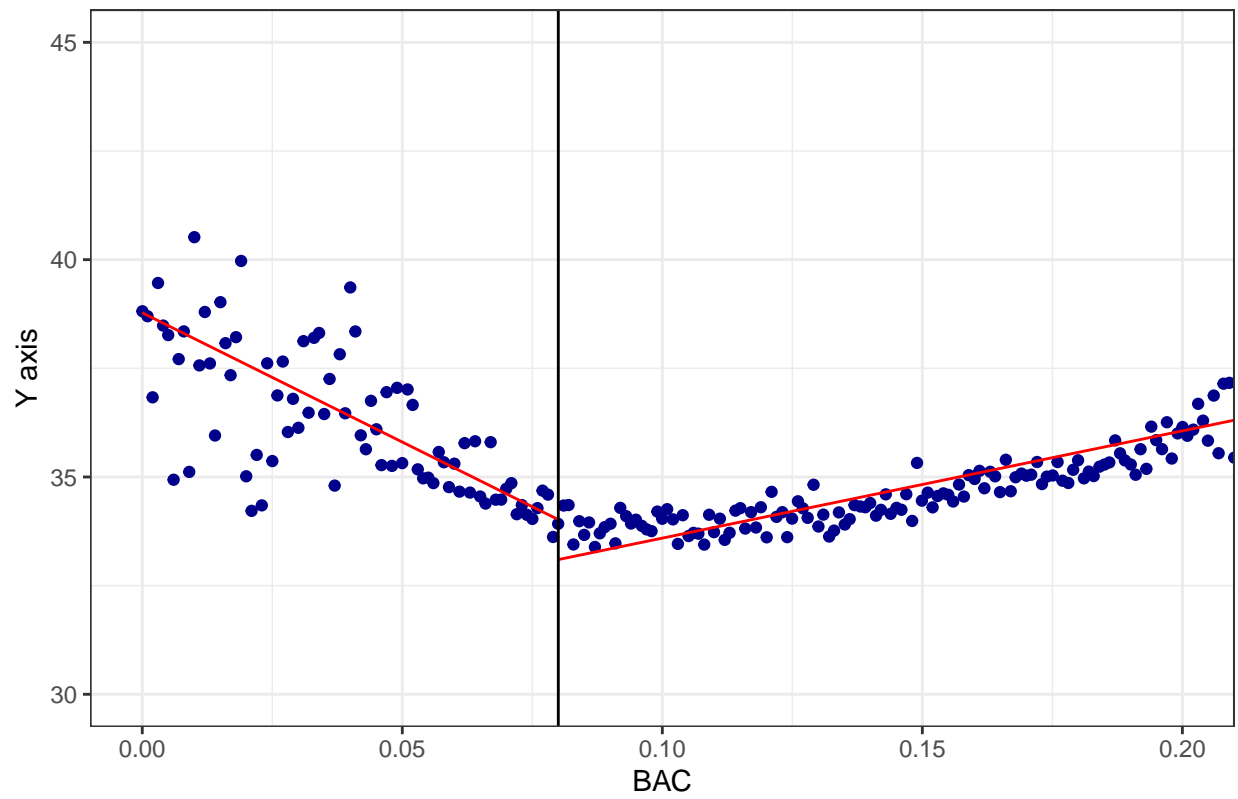
Panel B. Male
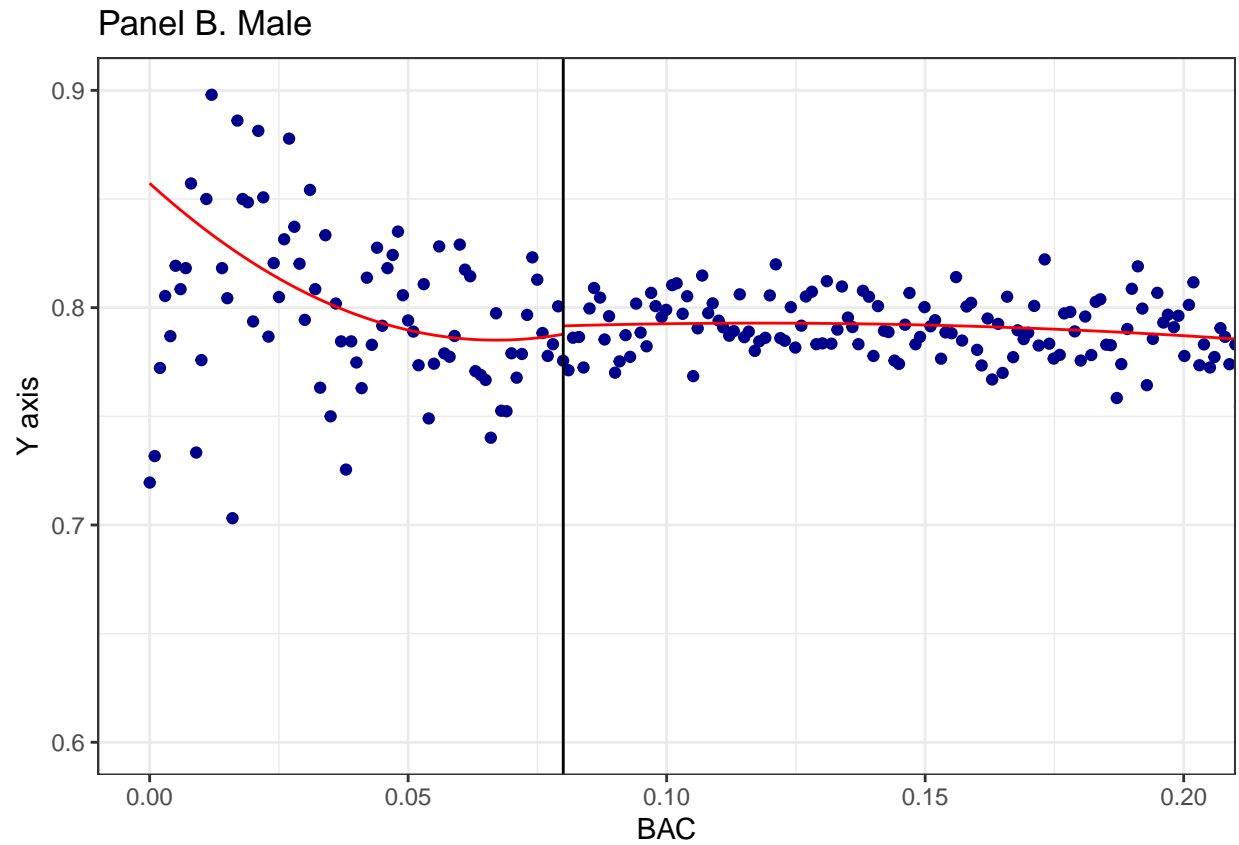
Panel C Linear

```
rdplot(hansen$aged, hansen$bac1, p =1, c = 0.08, title = "Panel C. Age", x.label = 'BAC', kernel = "tria
```

[1] "Mass points detected in the running variable."
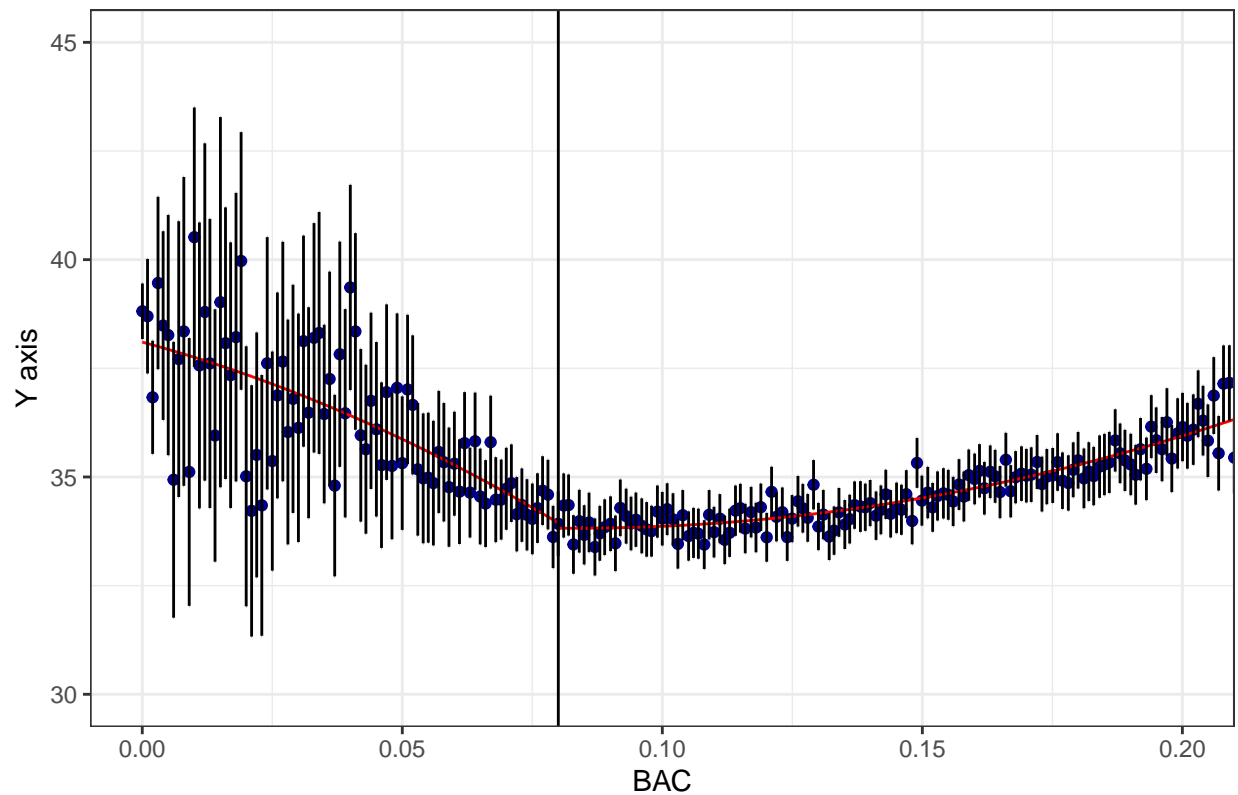
Panel C. Age

Panel C Quadratic fit

```
rdplot(hansen$aged, hansen$bac1, p =2, c = 0.08, ci = 95, title = "Panel C. Age", x.label = 'BAC', kern
```

```
## [1] "Mass points detected in the running variable."
```

Panel C. Age

I have not found a way to filter the aged data to convert it into a decimal like Hansen used in his paper. In the spirit of the assignment, I wanted to try this out on my own using R rather than using Excel/Access to manipulate the original data.

Panel D

```
rdplot(hansen$white, hansen$bac1, p =1, c = 0.08, title = "Panel D. White", x.label = 'BAC', kernel = "
```

```
## [1] "Mass points detected in the running variable."
```
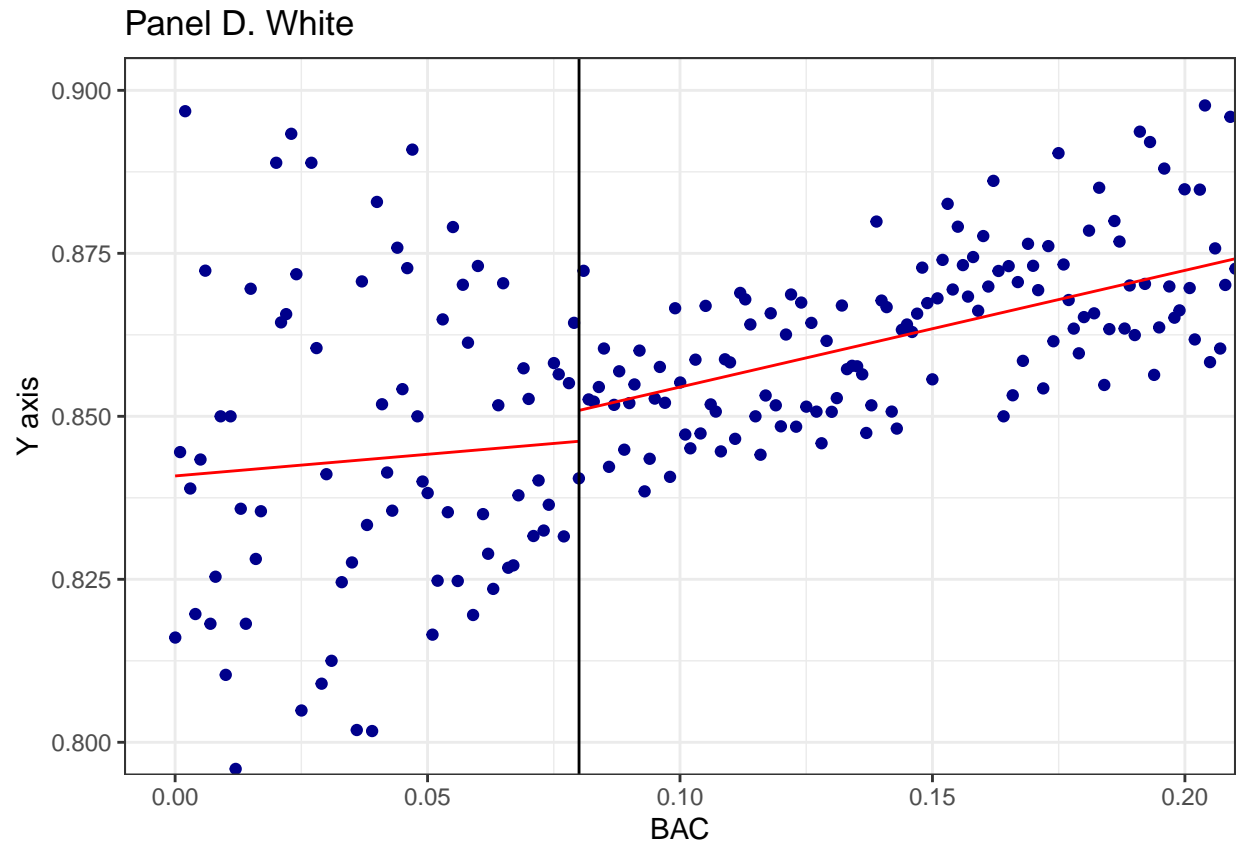
Panel D. White

I find a larger differential at the cutoff point than Hansen showed in his paper. I am not entirely sure if this is due to how the bins were set in my code versus his methodology or not, but I will continue to work on figuring out how to customize panels like these.

7

```
model_1 <- lm(recidivism ~ dui + bac1, data = hansen)
model_2 <- lm(recidivism ~ dui + bac1 + dui*bac1, data = hansen)
model_3 <- lm(recidivism ~ dui + bac1 + I(bac1^2), data = hansen)
model_d = lm(acc ~ dui + bac1 + dui*bac1, data = hansen)

stargazer(model_1, model_2, model_3,
  title="Driver demographic characteristics",
  dep.var.caption ="Recidivism",
  covariate.labels = c("DUI", "BAC", "Interaction", "Quadratic"),
  notes.label="Significance Levels",
  type = "text"
  )
```

```
##
## Driver demographic characteristics
## ================================================================================
##                                            Recidivism
##           -------------------------------------------------------------------------
```

```
##                                                     recidivism
##                                   (1)                    (2)                    (3)
## ---------------------------------------------------------------------------------------
## DUI                            -0.030***              -0.047***              -0.023***
##                                 (0.003)                (0.006)                (0.003)
##
## BAC                             0.315***               0.031                  0.119*
##                                 (0.016)                (0.086)                (0.066)
##
## Interaction                                            0.294***
##                                                        (0.087)
##
## Quadratic                                                                     0.585***
##                                                                               (0.193)
##
## Constant                        0.099***               0.115***               0.108***
##                                 (0.002)                (0.005)                (0.004)
##
## ---------------------------------------------------------------------------------------
## Observations                    214,558                214,558                214,558
## R2                               0.002                  0.002                  0.002
## Adjusted R2                      0.002                  0.002                  0.002
## Residual Std. Error     0.322 (df = 214555)      0.322 (df = 214554)      0.322 (df = 214554)
## F Statistic          205.834*** (df = 2; 214555) 141.031*** (df = 3; 214554) 140.297*** (df = 3; 21455
## =======================================================================================
## Significance Levels                                             *p<0.1; **p<0.05; ***p<0
```
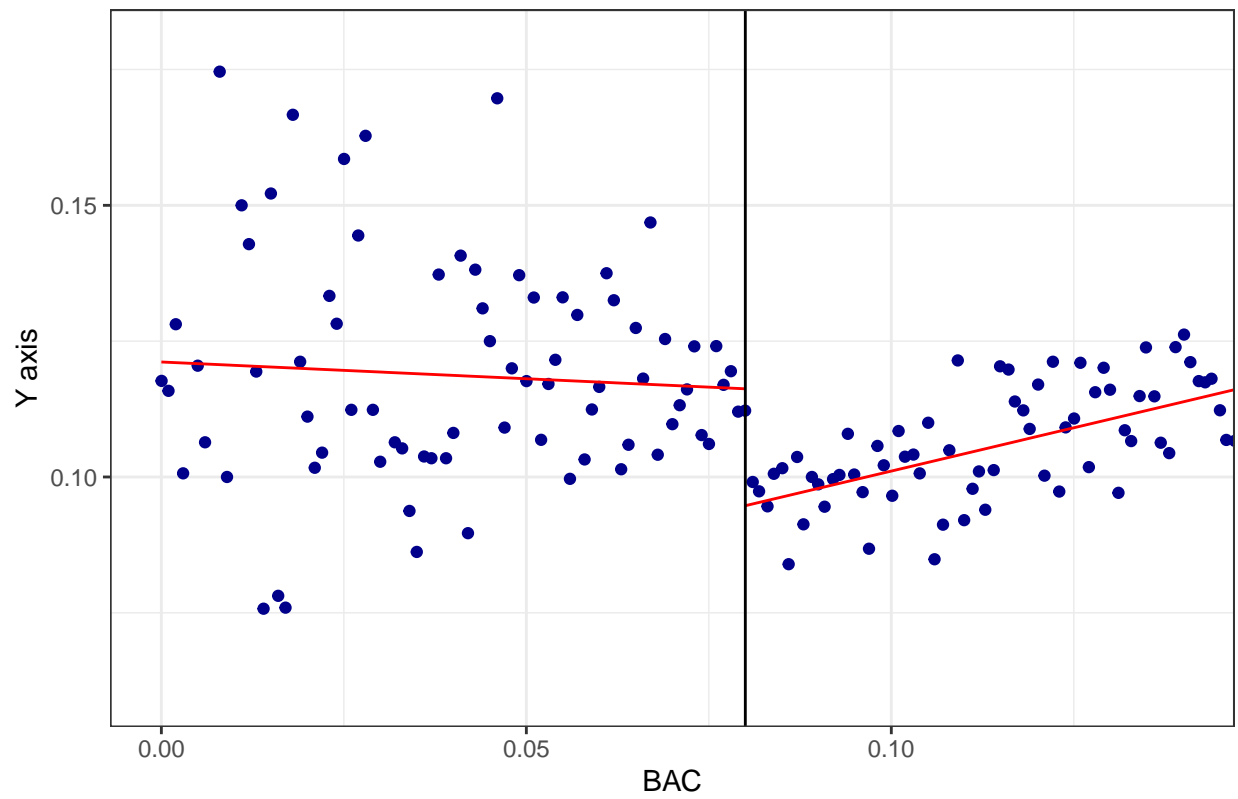
8

```
rdplot(hansen$recidivism, hansen$bac1, p = 1, c = 0.08, title = "Panel A. All offenders - linear", x.la
```

```
## [1] "Mass points detected in the running variable."
```
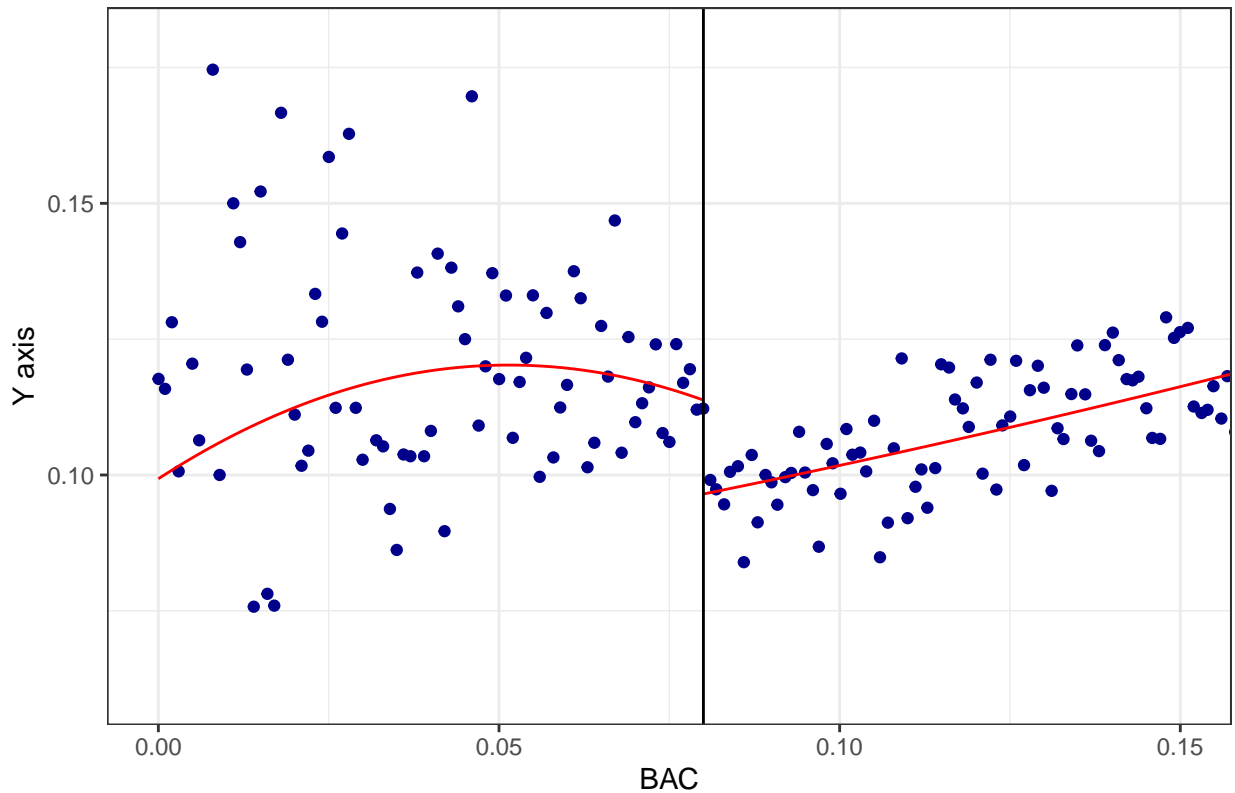
Panel A. All offenders – linear


```
rdplot(hansen$recidivism, hansen$bac1, p = 2, c = 0.08, title = "Panel A. All offenders - quadratic", x
```

```
## [1] "Mass points detected in the running variable."
```

Panel A. All offenders – quadratic

9

This assignment was very eye-opening for me. I had always figured that software packages like R were tailored to lowering the learning curve for getting research together in standard formats. Also seeing the steps that are required to ensure that things that are usually assumed to be true like unbiased sampling and randomized controls are actually fit the assumptions made. My hypothesis going into this assignment was that there would be a causal relationship between penalties and likelihood of re-offending. As someone who has himself gotten a DUI and has worked with alcoholics for years it is my experience that once the threat of punishment becomes a real factor with real consequences, some people will immediately change their behavior.

Going back to Hansen's work, I agree with his findings of a 2% drop in 4 year recidivism with harsher DUI punishments. Going step by step through this process, the methodology he employed was thorough in ensuring that the factors he employed allowed for his analysis on recidivism to likely be unbiased. The findings of decreases in crime when the costs of committing those crimes (in this case harsher punishments) are perfectly in line with the cost-benefit analysis that we all employ. I also appreciate the depth of his breakdowns (prior offenses at time of first DUI for example) and the consideration of other factors like addiction and mandated treatment methods for those convicted. "'