Problem Set 2_FINAL

Neeharika , Francesca, Giacomo, Kun and Noor

01/02/2022

```
library(readxl)
data<- read_excel("~/Desktop/PS 2 -METRICS/problem_set_2.xls")

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

#adding the 4th order polynomials into the data
data$age_sq<-data$age^2
data$age_cube<-data$age^3
data$age_4<-data$age^4
data$yob_sq<-data$yob^2
data$yob_cube<-data$yob^3
data$yob_4<-data$yob^4

#running the simple OLS of the 4th order polynomial
reg1 <- glm(logearn~schooling+age+age_sq+age_cube+age_4+yob+yob_sq+yob_cube+yob_4,
data = data)
summary(reg1)

##
## Call:
## glm(formula = logearn ~ schooling + age + age_sq + age_cube +
##     age_4 + yob + yob_sq + yob_cube + yob_4, data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.8438  -0.2490   0.0057   0.2662   2.9785
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.489e+00  6.176e+00   0.241   0.8094
## schooling    1.582e-01  2.471e-03  64.043   <2e-16 ***
## age         -2.541e-01  5.405e-01  -0.470   0.6382
## age_sq       9.347e-03  1.740e-02   0.537   0.5912
## age_cube    -1.317e-04  2.467e-04  -0.534   0.5934
## age_4        6.164e-07  1.299e-06   0.474   0.6352
## yob          4.414e-01  2.524e-01   1.749   0.0803 .
```

```
## yob_sq      -1.823e-02  1.164e-02  -1.566   0.1174
## yob_cube     3.184e-04  2.344e-04   1.358   0.1744
## yob_4       -1.955e-06  1.741e-06  -1.123   0.2616
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2439494)
##
##     Null deviance: 8810.2  on 30800  degrees of freedom
## Residual deviance: 7511.4  on 30791  degrees of freedom
## AIC: 43968
##
## Number of Fisher Scoring iterations: 2
```

```
stargazer(reg1)
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-
mail: hlavac at fas.harvard.edu
## % Date and time: Lun, Feb 07, 2022 - 17:08:04
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & logearn \\
## \hline \\[-1.8ex]
##  schooling & 0.158$^{***}$ \\
##   & (0.002) \\
##   & \\
##  age & $-$0.254 \\
##   & (0.540) \\
##   & \\
##  age\_sq & 0.009 \\
##   & (0.017) \\
##   & \\
##  age\_cube & $-$0.0001 \\
##   & (0.0002) \\
##   & \\
##  age\_4 & 0.00000 \\
##   & (0.00000) \\
##   & \\
##  yob & 0.441$^{*}$ \\
##   & (0.252) \\
##   & \\
##  yob\_sq & $-$0.018 \\
##   & (0.012) \\
##   & \\
##  yob\_cube & 0.0003 \\
```

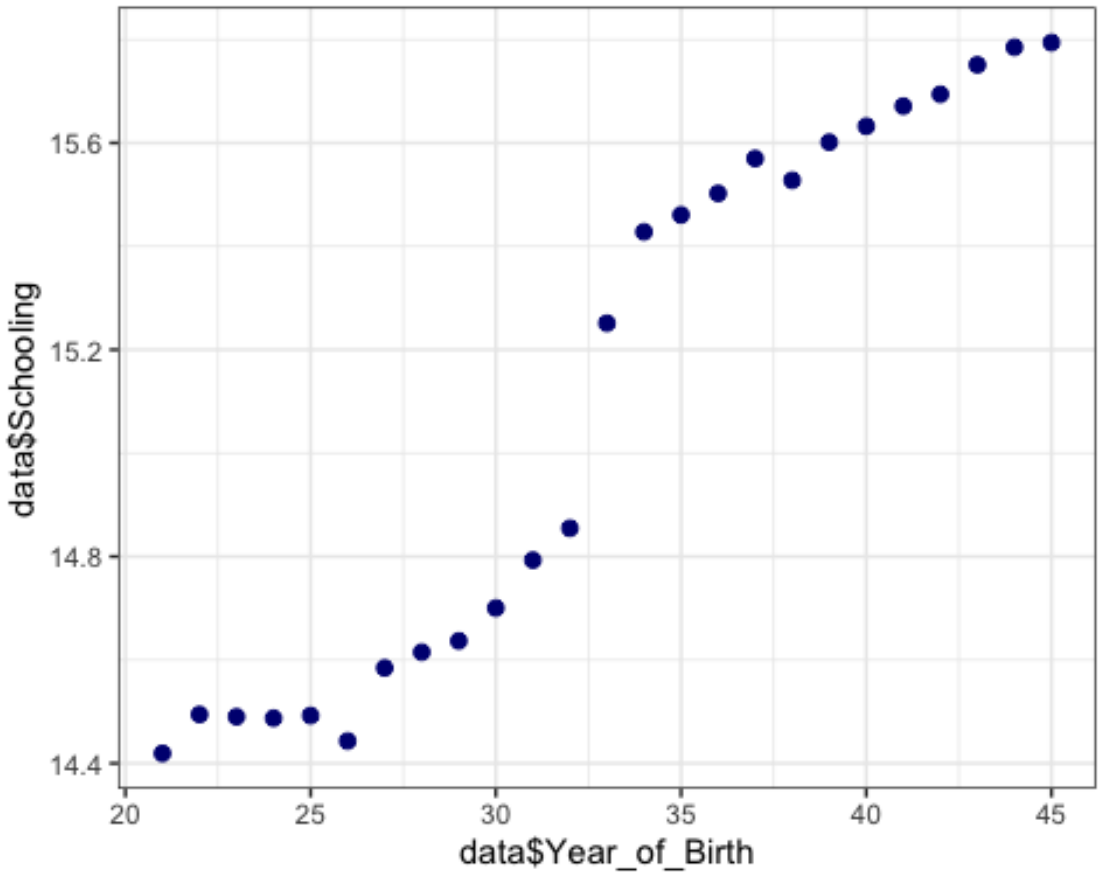
```
##   & (0.0002) \\
##   & \\
##  yob\_4 & $-$0.00000 \\
##   & (0.00000) \\
##   & \\
##  Constant & 1.489 \\
##   & (6.176) \\
##   & \\
## \hline \\[-1.8ex]
## Observations & 30,801 \\
## Log Likelihood & $-$21,973.790 \\
## Akaike Inf. Crit. & 43,967.570 \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05;
$^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

```
#creating a dummy for instrument
data$LAW <- ifelse (data$yob >= 33, 1, 0)
reg6<- lm(schooling ~ LAW, data = data)
#stargazer(reg6)
reg7<- lm(logearn ~ LAW, data = data )
stargazer(reg7)
```

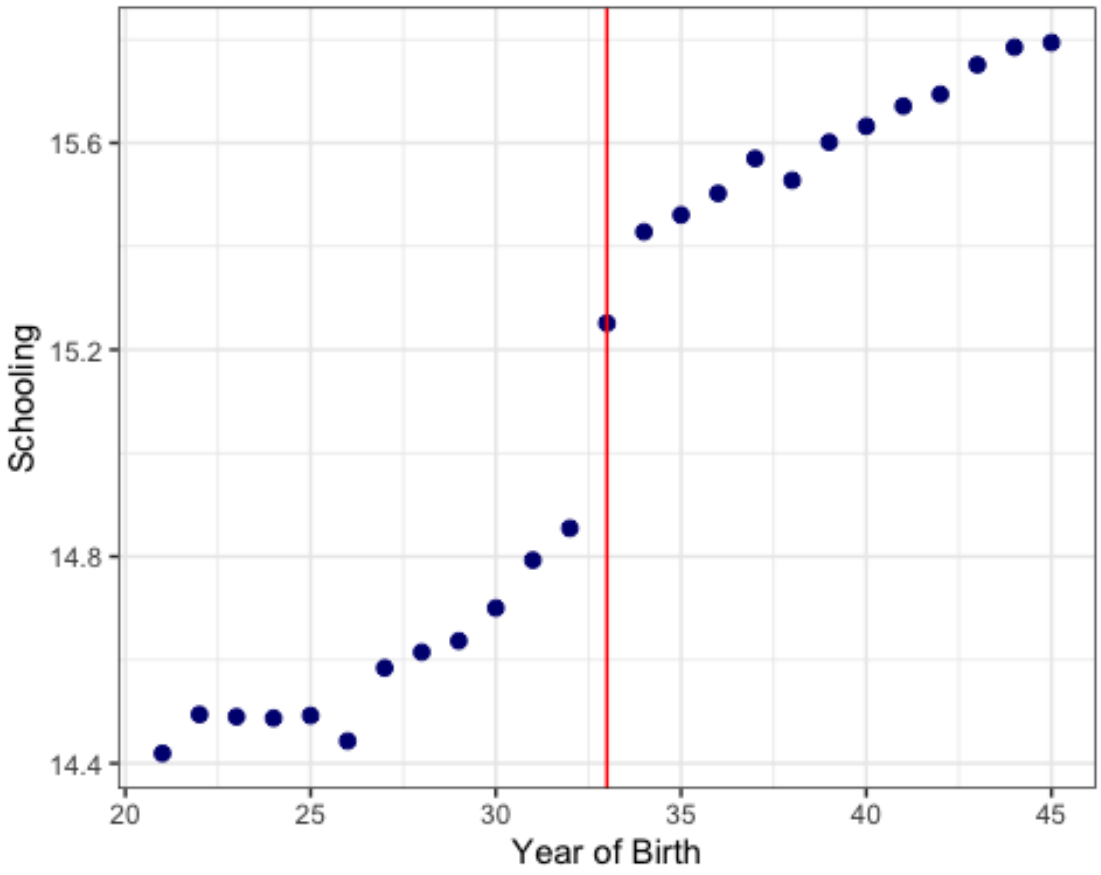
```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-
mail: hlavac at fas.harvard.edu
## % Date and time: Lun, Feb 07, 2022 - 17:08:04
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & logearn \\
## \hline \\[-1.8ex]
##  LAW & 0.162$^{***}$ \\
##   & (0.007) \\
##   & \\
##  Constant & 5.671$^{***}$ \\
##   & (0.005) \\
##   & \\
## \hline \\[-1.8ex]
## Observations & 30,801 \\
## R$^{2}$ & 0.019 \\
## Adjusted R$^{2}$ & 0.019 \\
## Residual Std. Error & 0.530 (df = 30799) \\
## F Statistic & 607.678$^{***}$ (df = 1; 30799) \\
```

```
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05;
$^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}

0.161749 /0.99034

## [1] 0.1633267

#0.1633267
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1
──

## ✓ ggplot2 3.3.5     ✓ purrr   0.3.4
## ✓ tibble  3.1.0     ✓ dplyr   1.0.5
## ✓ tidyr   1.1.3     ✓ stringr 1.4.0
## ✓ readr   1.4.0     ✓ forcats 0.5.1

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

#install.packages("ivreg")
library(ivreg)
a<-ivreg(logearn~schooling|LAW, data = data)
stargazer(a)

##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-
mail: hlavac at fas.harvard.edu
## % Date and time: Lun, Feb 07, 2022 - 17:08:13
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & logearn \\
## \hline \\[-1.8ex]
##  schooling & 0.163$^{***}$ \\
##   & (0.006) \\
##   & \\
##  Constant & 3.283$^{***}$ \\
##   & (0.095) \\
##   & \\
```

```
## \hline \\[-1.8ex]
## Observations & 30,801 \\
## R$^{2}$ & 0.138 \\
## Adjusted R$^{2}$ & 0.138 \\
## Residual Std. Error & 0.497 (df = 30799) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05;
$^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}

#to obtain the probability of someone leaving before the year 15
data$leave<-ifelse(data$schooling < 15,1,0) #if someone left school before 15 = 1
and 0 otherwise
leave_tab<-table(data$leave)
leave_tab

##
##     0     1
## 23256  7545

#to get probability
prop.table(leave_tab)#proportion table

##
##         0         1
## 0.7550404 0.2449596

#probability for each individual if LEAVE ==1  for each year
data$prob<-with(data, ave(leave,yob))

names(data)[3] <- "Year_of_Birth"
names(data)[5] <- "Schooling"
names(data)[14] <- "Probability"
#install.packages("binsreg")
library(binsreg)
binscatter1<-binsreg(data$Schooling, data$Year_of_Birth)

## Warning in binsregselect(y, x, w, deriv = deriv, bins = dots, binspos =
## binspos, : too small effective sample size for DPI selection.

## Warning in binsreg(data$Schooling, data$Year_of_Birth): DPI selection fails.
ROT
## choice used.

## Warning in binsreg(data$Schooling, data$Year_of_Birth): too small effective
## sample size for dots. # of mass points or clusters used.

## Warning in binsreg(data$Schooling, data$Year_of_Birth): dots=c(0,0) used.
```




```
binscatter1$bins_plot + geom_vline (xintercept = 33, colour = "red" ) + xlab("Year of Birth") + ylab("Schooling")
```




```
  ggtitle("Binscatter of Schooling vs Year of Birth") +
  theme(plot.title = element_text(hjust = 0.3, vjust = 0.3, face = 'italic')) +
  theme(element_line(margin(t = 3, r = 3, b = 3, l = 3, unit = "pt")))
```

```
## NULL
```

```
binscatter2<-binsreg(data$Probability, data$Year_of_Birth)
```

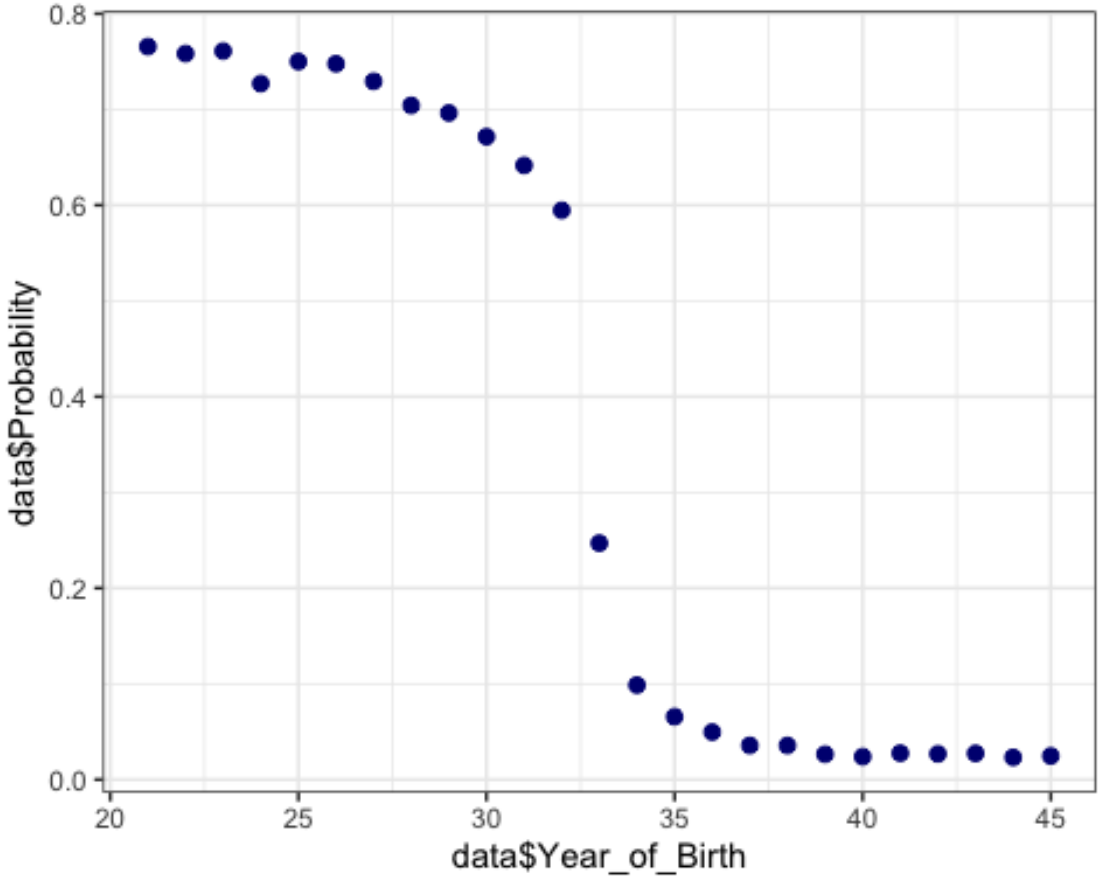
```
## Warning in binsregselect(y, x, w, deriv = deriv, bins = dots, binspos =
## binspos, : too small effective sample size for DPI selection.
```

```
## Warning in binsregselect(y, x, w, deriv = deriv, bins = dots, binspos =
## binspos, : some bins have too few distinct values of x for DPI selection.
```

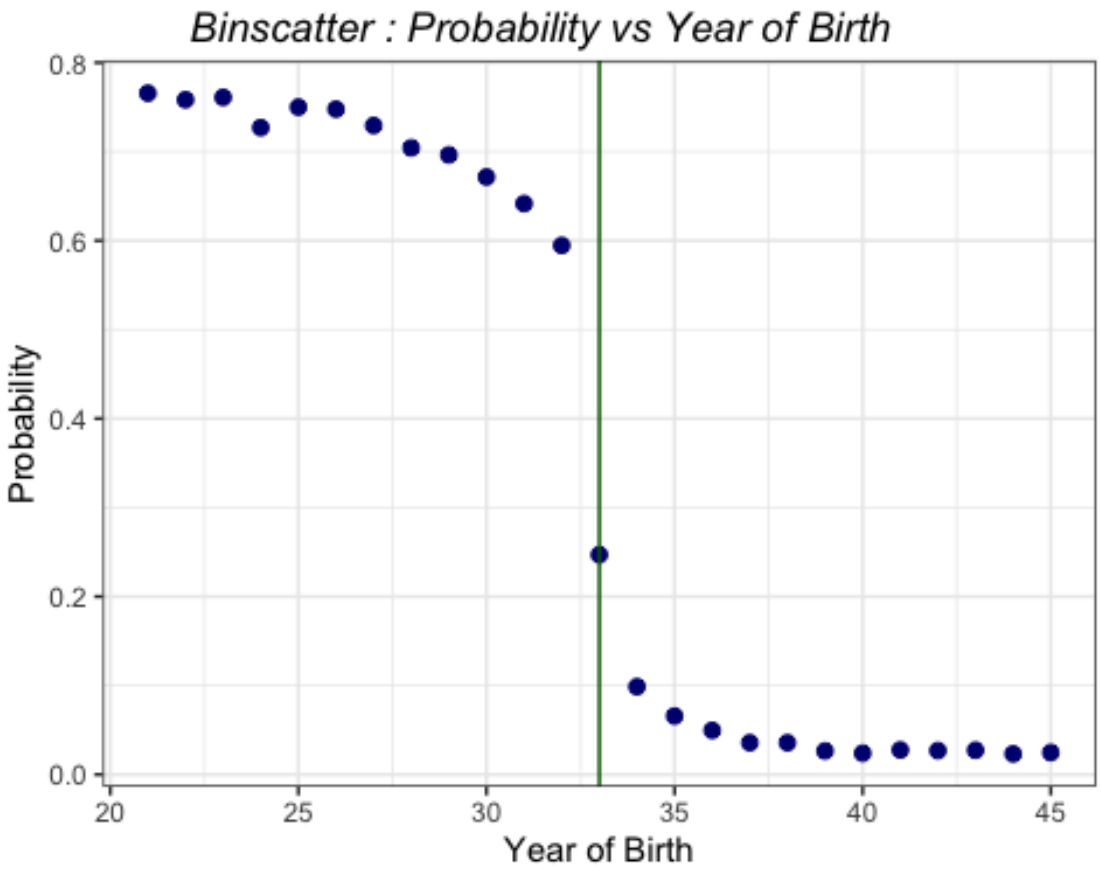
```
## Warning in binsreg(data$Probability, data$Year_of_Birth): DPI selection fails.
## ROT choice used.
```

```
## Warning in binsreg(data$Probability, data$Year_of_Birth): too small effective
## sample size for dots. # of mass points or clusters used.
```

```
## Warning in binsreg(data$Probability, data$Year_of_Birth): dots=c(0,0) used.
```




```
binscatter2$bins_plot + geom_vline(xintercept = 33, colour = "darkgreen") +
xlab("Year of Birth") +
  ylab("Probability")+
  ggtitle("Binscatter : Probability vs Year of Birth") +
  theme(plot.title = element_text(hjust = 0.3 , vjust = 0.3, face = 'italic')) +
  theme(element_line(margin(t = 3, r = 3, b = 3, l = 3, unit = "pt")))
```



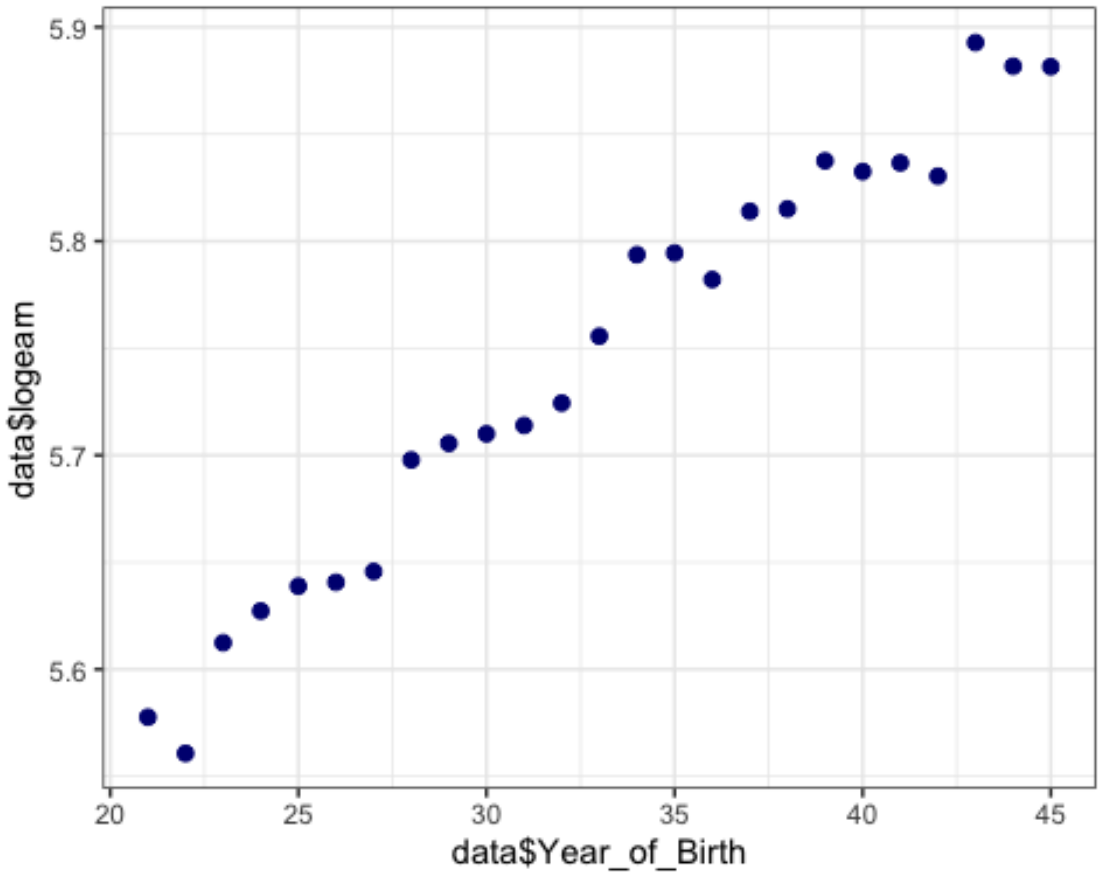

```
binscatter3<-binsreg(data$logearn, data$Year_of_Birth)
```

```
## Warning in binsregselect(y, x, w, deriv = deriv, bins = dots, binspos =
## binspos, : too small effective sample size for DPI selection.
```

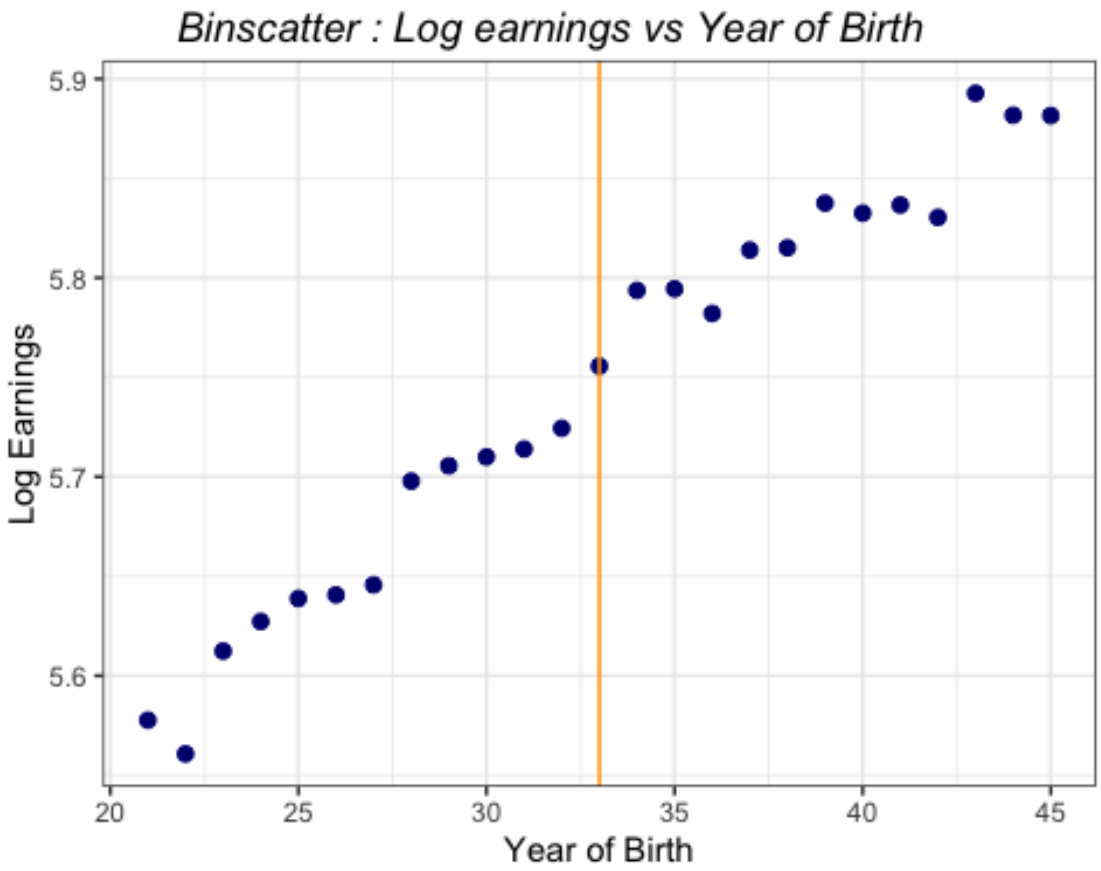
```
## Warning in binsreg(data$logearn, data$Year_of_Birth): DPI selection fails. ROT
## choice used.
```

```
## Warning in binsreg(data$logearn, data$Year_of_Birth): too small effective
sample
## size for dots. # of mass points or clusters used.
```

```
## Warning in binsreg(data$logearn, data$Year_of_Birth): dots=c(0,0) used.
```




```
binscatter3$bins_plot + geom_vline(xintercept = 33, colour = "orange") +
ggtitle("Binscatter : Log earnings vs Year of Birth") +
  theme(plot.title = element_text(hjust = 0.3 , vjust = 0.3, face = 'italic')) +
theme(plot.caption = element_text("Yob = 33")) +
  theme(element_line(margin(t = 3, r = 3, b = 3, l = 3, unit = "pt"))) +
  xlab("Year of Birth") +
  ylab("Log Earnings")
```




```
  summary(data)

##       age           mainsamp      Year_of_Birth      logearn
##  Min.   :33.00   Min.   :0.0000   Min.   :21.00   Min.   :1.787
##  1st Qu.:44.00   1st Qu.:1.0000   1st Qu.:31.00   1st Qu.:5.484
##  Median :50.00   Median :1.0000   Median :37.00   Median :5.760
##  Mean   :49.02   Mean   :0.7995   Mean   :35.98   Mean   :5.784
##  3rd Qu.:54.00   3rd Qu.:1.0000   3rd Qu.:42.00   3rd Qu.:6.076
##  Max.   :60.00   Max.   :1.0000   Max.   :45.00   Max.   :8.992
##    Schooling         age_sq        age_cube          age_4
##  Min.   :10.00   Min.   :1089   Min.   : 35937   Min.   : 1185921
##  1st Qu.:15.00   1st Qu.:1936   1st Qu.: 85184   1st Qu.: 3748096
##  Median :15.00   Median :2500   Median :125000   Median : 6250000
##  Mean   :15.31   Mean   :2448   Mean   :124293   Mean   : 6406839
##  3rd Qu.:16.00   3rd Qu.:2916   3rd Qu.:157464   3rd Qu.: 8503056
##  Max.   :21.00   Max.   :3600   Max.   :216000   Max.   :12960000
##      yob_sq        yob_cube         yob_4              LAW
##  Min.   : 441   Min.   : 9261   Min.   : 194481   Min.   :0.0000
##  1st Qu.: 961   1st Qu.:29791   1st Qu.: 923521   1st Qu.:0.0000
##  Median :1369   Median :50653   Median :1874161   Median :1.0000
##  Mean   :1338   Mean   :51122   Mean   :1997051   Mean   :0.6961
##  3rd Qu.:1764   3rd Qu.:74088   3rd Qu.:3111696   3rd Qu.:1.0000
##  Max.   :2025   Max.   :91125   Max.   :4100625   Max.   :1.0000
##      leave        Probability
##  Min.   :0.000   Min.   :0.02322
```

```
##  1st Qu.:0.000   1st Qu.:0.02686
##  Median :0.000   Median :0.03574
##  Mean   :0.245   Mean   :0.24496
##  3rd Qu.:0.000   3rd Qu.:0.64158
##  Max.   :1.000   Max.   :0.76562
```