Problem Set 3 - Hong Ngoc Nguyen

Hong Ngoc Nguyen

18 March 2019

Select the entire time period, 1963-2017. Report the regression result.

```
# Select the entire time period, 1963-2017

library(tidyverse)
```

```
## -- Attaching packages -------------------------------------------------------------------------------

## v ggplot2 3.1.0       v purrr   0.3.0
## v tibble  2.0.1       v dplyr   0.8.0.1
## v tidyr   0.8.2       v stringr 1.4.0
## v readr   1.3.1       v forcats 0.4.0

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

```
data_01 <- read.csv(file="data_01.csv",head=TRUE)


attach(data_01)

# Select the male sample
data_01 <- data_01 %>%
  filter(sex == 1)

hs <- matrix(0,54,1)
col <- matrix(0,54,1)
hsdrop <- matrix(0,54,1)
somecol <- matrix(0,54,1)
hs_equi <- matrix(0,54,1)
col_equi <- matrix(0,54,1)
relsup <- matrix(0,54,1)
logw1w2 <- matrix(0,54,1)

for (i in 1:54)
{
  # quantity of high school graduates
  hs[i] <- sum(with(data_01, deduc_1 == 0 & deduc_2 == 0 & deduc_3 == 0 & deduc_4 == 0 & year==1963+i))

  w_hs <- mean(rwage[deduc_1 == 0 & deduc_2 == 0 & deduc_3 == 0 & deduc_4 == 0]) #weight

  # quantity of college graduates
  col[i] <- sum(with(data_01, (deduc_3 == 1 | deduc_4 == 1)  & year==1963+i))

  w_col <- mean(rwage[deduc_3 == 1 | deduc_4 == 1]) #weight

  # quantity of high school dropout
```

```r
  hsdrop[i] <- sum(with(data_01, deduc_1 == 1  & year==1963+i))

  w_hsdrop <- mean(rwage[deduc_1 == 1]) #weight

  # quantity of some college
  somecol[i] <- sum(with(data_01, deduc_2 == 1 & year==1963+i))

  w_somecol <- mean(rwage[deduc_2 == 1]) #weight


  # aggregate labor inputs of high school equivalent
  #hs_equi <- hs + 0.69*somecol + 0.93*hsdrop
  hs_equi[i] <- hs[i]*w_hs + 0.69*somecol[i]*w_somecol + 0.93*hsdrop[i]*w_hsdrop

  # aggregate labor inputs of college equivalent
  #col_equi <- col + 0.29*somecol - 0.05*hsdrop
  col_equi[i] <- col[i]*w_col + 0.29*somecol[i]*w_somecol - 0.05*hsdrop[i]*w_hsdrop

  # relative supply
  relsup[i] <- col_equi[i]/hs_equi[i]
  logw1w2[i] <- log(relsup[i])
}

# Calculate the relative college/high school wage ratio

# Create the five-year experience brackets
data_01 <- data_01 %>%
  mutate(expbr = ifelse(exp <= 5, 1, NA)) %>%
  mutate(expbr = ifelse(exp > 5 & exp <= 10, 2, expbr)) %>%
  mutate(expbr = ifelse(exp > 10 & exp <= 15, 3, expbr)) %>%
  mutate(expbr = ifelse(exp > 15 & exp <= 20, 4, expbr)) %>%
  mutate(expbr = ifelse(exp > 20 & exp <= 25, 5, expbr)) %>%
  mutate(expbr = ifelse(exp > 25 & exp <= 30, 6, expbr)) %>%
  mutate(expbr = ifelse(exp > 30 & exp <= 35, 7, expbr)) %>%
  mutate(expbr = ifelse(exp > 35, 8, expbr))

# college/high school wage ratio

m_wageratio <- matrix(0,54,8)
m_share <- matrix(0,54,8)
m_weight <- matrix(0,8,1)
m_wwage <- matrix(0,54,8)
relwage <- matrix(0,54,1)
logx1x2 <- matrix(0,54,1)

attach(data_01)
```

```
## The following objects are masked from data_01 (pos = 3):
##
##     age, asecflag, asecwt, asecwth, classwly, cpi, cpsid, cpsidp,
##     deduc_1, deduc_2, deduc_3, deduc_4, dfemale, drace_1, drace_2,
##     educ, exp, fullpart, hflag, incwage, indly, lrwage, month,
##     pernum, race, rwage, schlcoll, serial, sex, top_incwage,
##     wkswork, wkswork1, wkswork2, X, year
```

```r
for (i in 1:54)
{
  for (j in 1:8)
  {
    #male

    # ratio of the average weekly wage of college graduates to the average weekly wage of high school g
    m_wageratio[i,j] <- mean(rwage[(deduc_3 == 1 | deduc_4 == 1) & year==1963+i & expbr==j & sex==1])/me

    # The fixed weight for each cell is the cell's average share of total employment over the period
    m_share[i,j] <- sum(wkswork[year==1963+i & expbr==j & sex==1])/sum(wkswork[year==1963+i])
    m_weight[j] <- mean(m_share[1:54,j])

    m_wwage[i,j] <- m_wageratio[i,j] * m_weight[j] #weighted wage

    #female
    #f_wageratio[i,j] <- mean(rwage[(deduc_3 == 1 | deduc_4 == 1) & year==1963+i & expbr==j & sex==2])/
    #f_share[i,j] <- sum(wkswork[year==1963+i & expbr==j & sex==2])/sum(wkswork[year==1963+i])
    #f_weight[j] <- mean(f_share[1:54,j])

    #f_wwage[i,j] <- f_wageratio[i,j] * f_weight[j] #weighted wage
  }

  # The overall college/high school wage ratio

  #relwage[i] <- sum(m_wwage[i,1:8])+sum(f_wwage[i,1:8])
  relwage[i] <- sum(m_wwage[i,1:8])
  logx1x2[i] <- log(relwage[i])
}

# time trend

time=array(1:54,dim = c(54,1))

# the regression of the Equation (19) in Katz and Murphy (1992)
m1 <- lm(logw1w2 ~ logx1x2 + time)
coef(m1)
```

```
## (Intercept)     logx1x2        time
## -1.62320754 -0.02512482  0.05401754
```

```r
summary(m1)
```

```
##
## Call:
## lm(formula = logw1w2 ~ logx1x2 + time)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.06155 -0.01459 -0.00560  0.02080  0.06967
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.623208   0.070560  -23.00 2.47e-15 ***
```

```
## logx1x2     -0.025125   0.025120   -1.00     0.33
## time         0.054018   0.003255   16.59 9.21e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03564 on 19 degrees of freedom
##   (32 observations deleted due to missingness)
## Multiple R-squared:  0.9896, Adjusted R-squared:  0.9885
## F-statistic: 906.7 on 2 and 19 DF,  p-value: < 2.2e-16
```

Select the same time period as Katz and Murphy (1992), 1963-1987

```r
# Select the same time period as Katz and Murphy (1992), 1963-1987

library(tidyverse)
data_01 <- read.csv(file="data_01.csv",head=TRUE)

attach(data_01)
```

```
## The following objects are masked from data_01 (pos = 3):
##
##     age, asecflag, asecwt, asecwth, classwly, cpi, cpsid, cpsidp,
##     deduc_1, deduc_2, deduc_3, deduc_4, dfemale, drace_1, drace_2,
##     educ, exp, fullpart, hflag, incwage, indly, lrwage, month,
##     pernum, race, rwage, schlcoll, serial, sex, top_incwage,
##     wkswork, wkswork1, wkswork2, X, year
```

```
## The following objects are masked from data_01 (pos = 4):
##
##     age, asecflag, asecwt, asecwth, classwly, cpi, cpsid, cpsidp,
##     deduc_1, deduc_2, deduc_3, deduc_4, dfemale, drace_1, drace_2,
##     educ, exp, fullpart, hflag, incwage, indly, lrwage, month,
##     pernum, race, rwage, schlcoll, serial, sex, top_incwage,
##     wkswork, wkswork1, wkswork2, X, year
```

```r
# Select the male sample
data_01 <- data_01 %>%
  filter(sex == 1)

hs <- matrix(0,24,1)
col <- matrix(0,24,1)
hsdrop <- matrix(0,24,1)
somecol <- matrix(0,24,1)
hs_equi <- matrix(0,24,1)
col_equi <- matrix(0,24,1)
relsup <- matrix(0,24,1)
logw1w2 <- matrix(0,24,1)

for (i in 1:24)
{
  # quantity of high school graduates
  hs[i] <- sum(with(data_01, deduc_1 == 0 & deduc_2 == 0 & deduc_3 == 0 & deduc_4 == 0 & year==1963+i))

  w_hs <- mean(rwage[deduc_1 == 0 & deduc_2 == 0 & deduc_3 == 0 & deduc_4 == 0]) #weight
```

```r
  # quantity of college graduates
  col[i] <- sum(with(data_01, (deduc_3 == 1 | deduc_4 == 1)  & year==1963+i))

  w_col <- mean(rwage[deduc_3 == 1 | deduc_4 == 1]) #weight

  # quantity of high school dropout
  hsdrop[i] <- sum(with(data_01, deduc_1 == 1  & year==1963+i))

  w_hsdrop <- mean(rwage[deduc_1 == 1]) #weight

  # quantity of some college
  somecol[i] <- sum(with(data_01, deduc_2 == 1 & year==1963+i))

  w_somecol <- mean(rwage[deduc_2 == 1]) #weight


  # aggregate labor inputs of high school equivalent
  #hs_equi <- hs + 0.69*somecol + 0.93*hsdrop
  hs_equi[i] <- hs[i]*w_hs + 0.69*somecol[i]*w_somecol + 0.93*hsdrop[i]*w_hsdrop

  # aggregate labor inputs of college equivalent
  #col_equi <- col + 0.29*somecol - 0.05*hsdrop
  col_equi[i] <- col[i]*w_col + 0.29*somecol[i]*w_somecol - 0.05*hsdrop[i]*w_hsdrop

  # relative supply
  relsup[i] <- col_equi[i]/hs_equi[i]
  logw1w2[i] <- log(relsup[i])
}

# Calculate the relative college/high school wage ratio

# Create the five-year experience brackets
data_01 <- data_01 %>%
  mutate(expbr = ifelse(exp <= 5, 1, NA)) %>%
  mutate(expbr = ifelse(exp > 5 & exp <= 10, 2, expbr)) %>%
  mutate(expbr = ifelse(exp > 10 & exp <= 15, 3, expbr)) %>%
  mutate(expbr = ifelse(exp > 15 & exp <= 20, 4, expbr)) %>%
  mutate(expbr = ifelse(exp > 20 & exp <= 25, 5, expbr)) %>%
  mutate(expbr = ifelse(exp > 25 & exp <= 30, 6, expbr)) %>%
  mutate(expbr = ifelse(exp > 30 & exp <= 35, 7, expbr)) %>%
  mutate(expbr = ifelse(exp > 35, 8, expbr))

# college/high school wage ratio

m_wageratio <- matrix(0,24,8)
m_share <- matrix(0,24,8)
m_weight <- matrix(0,8,1)
m_wwage <- matrix(0,24,8)
relwage <- matrix(0,24,1)
logx1x2 <- matrix(0,24,1)

for (i in 1:24)
{
```

```r
  for (j in 1:8)
  {
    #male
    m_wageratio[i,j] <- mean(rwage[(deduc_3 == 1 | deduc_4 == 1) & year==1963+i & expbr==j & sex==1])/me
    m_share[i,j] <- sum(wkswork[year==1963+i & expbr==j & sex==1])/sum(wkswork[year==1963+i])
    m_weight[j] <- mean(m_share[1:24,j])

    m_wwage[i,j] <- m_wageratio[i,j] * m_weight[j] #weighted wage

    #female
    #f_wageratio[i,j] <- mean(rwage[(deduc_3 == 1 | deduc_4 == 1) & year==1963+i & expbr==j & sex==2])/
    #f_share[i,j] <- sum(wkswork[year==1963+i & expbr==j & sex==2])/sum(wkswork[year==1963+i])
    #f_weight[j] <- mean(f_share[1:24,j])

    #f_wwage[i,j] <- f_wageratio[i,j] * f_weight[j] #weighted wage
  }

  #relwage[i] <- sum(m_wwage[i,1:8])+sum(f_wwage[i,1:8])
  relwage[i] <- sum(m_wwage[i,1:8])
  logx1x2[i] <- log(relwage[i])
}

# time trend

time=array(1:24,dim = c(24,1))

# the regression of the Equation (19) in Katz and Murphy (1992)
m1 <- lm(logw1w2 ~ logx1x2 + time)
coef(m1)
```

```
## (Intercept)     logx1x2        time
##  -0.3199296   0.4607456  -0.0250828
```

```r
summary(m1)
```

```
##
## Call:
## lm(formula = logw1w2 ~ logx1x2 + time)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.86571 -0.16359  0.05595  0.22571  0.39734
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.31993    0.49714  -0.644   0.5268
## logx1x2      0.46075    0.21717   2.122   0.0459 *
## time        -0.02508    0.02496  -1.005   0.3263
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3416 on 21 degrees of freedom
## Multiple R-squared:   0.32,  Adjusted R-squared:  0.2552
## F-statistic: 4.941 on 2 and 21 DF,  p-value: 0.01744
```

This result is different from Katz and Murphy (1992) since we select only the male sample while they combined both males and females.

Select the later period, 1988-2017

```
# Select the later period, 1988-2017

library(tidyverse)
data_01 <- read.csv(file="data_01.csv",head=TRUE)


attach(data_01)
```

```
## The following objects are masked from data_01 (pos = 3):
##
##     age, asecflag, asecwt, asecwth, classwly, cpi, cpsid, cpsidp,
##     deduc_1, deduc_2, deduc_3, deduc_4, dfemale, drace_1, drace_2,
##     educ, exp, fullpart, hflag, incwage, indly, lrwage, month,
##     pernum, race, rwage, schlcoll, serial, sex, top_incwage,
##     wkswork, wkswork1, wkswork2, X, year
```

```
## The following objects are masked from data_01 (pos = 4):
##
##     age, asecflag, asecwt, asecwth, classwly, cpi, cpsid, cpsidp,
##     deduc_1, deduc_2, deduc_3, deduc_4, dfemale, drace_1, drace_2,
##     educ, exp, fullpart, hflag, incwage, indly, lrwage, month,
##     pernum, race, rwage, schlcoll, serial, sex, top_incwage,
##     wkswork, wkswork1, wkswork2, X, year
```

```
## The following objects are masked from data_01 (pos = 5):
##
##     age, asecflag, asecwt, asecwth, classwly, cpi, cpsid, cpsidp,
##     deduc_1, deduc_2, deduc_3, deduc_4, dfemale, drace_1, drace_2,
##     educ, exp, fullpart, hflag, incwage, indly, lrwage, month,
##     pernum, race, rwage, schlcoll, serial, sex, top_incwage,
##     wkswork, wkswork1, wkswork2, X, year
```

```
# Select the male sample
data_01 <- data_01 %>%
  filter(sex == 1)

hs <- matrix(0,30,1)
col <- matrix(0,30,1)
hsdrop <- matrix(0,30,1)
somecol <- matrix(0,30,1)
hs_equi <- matrix(0,30,1)
col_equi <- matrix(0,30,1)
relsup <- matrix(0,30,1)
logw1w2 <- matrix(0,30,1)

for (i in 1:30)
{
  # quantity of high school graduates
  hs[i] <- sum(with(data_01, deduc_1 == 0 & deduc_2 == 0 & deduc_3 == 0 & deduc_4 == 0 & year==1987+i))
```

```
  w_hs <- mean(rwage[deduc_1 == 0 & deduc_2 == 0 & deduc_3 == 0 & deduc_4 == 0]) #weight

  # quantity of college graduates
  col[i] <- sum(with(data_01, (deduc_3 == 1 | deduc_4 == 1)  & year==1987+i))

  w_col <- mean(rwage[deduc_3 == 1 | deduc_4 == 1]) #weight

  # quantity of high school dropout
  hsdrop[i] <- sum(with(data_01, deduc_1 == 1  & year==1987+i))

  w_hsdrop <- mean(rwage[deduc_1 == 1]) #weight

  # quantity of some college
  somecol[i] <- sum(with(data_01, deduc_2 == 1 & year==1987+i))

  w_somecol <- mean(rwage[deduc_2 == 1]) #weight


  # aggregate labor inputs of high school equivalent
  #hs_equi <- hs + 0.69*somecol + 0.93*hsdrop
  hs_equi[i] <- hs[i]*w_hs + 0.69*somecol[i]*w_somecol + 0.93*hsdrop[i]*w_hsdrop

  # aggregate labor inputs of college equivalent
  #col_equi <- col + 0.29*somecol - 0.05*hsdrop
  col_equi[i] <- col[i]*w_col + 0.29*somecol[i]*w_somecol - 0.05*hsdrop[i]*w_hsdrop

  # relative supply
  relsup[i] <- col_equi[i]/hs_equi[i]
  logw1w2[i] <- log(relsup[i])
}

# Calculate the relative college/high school wage ratio

# Create the five-year experience brackets
data_01 <- data_01 %>%
  mutate(expbr = ifelse(exp <= 5, 1, NA)) %>%
  mutate(expbr = ifelse(exp > 5 & exp <= 10, 2, expbr)) %>%
  mutate(expbr = ifelse(exp > 10 & exp <= 15, 3, expbr)) %>%
  mutate(expbr = ifelse(exp > 15 & exp <= 20, 4, expbr)) %>%
  mutate(expbr = ifelse(exp > 20 & exp <= 25, 5, expbr)) %>%
  mutate(expbr = ifelse(exp > 25 & exp <= 30, 6, expbr)) %>%
  mutate(expbr = ifelse(exp > 30 & exp <= 35, 7, expbr)) %>%
  mutate(expbr = ifelse(exp > 35, 8, expbr))

# college/high school wage ratio

m_wageratio <- matrix(0,30,8)
m_share <- matrix(0,30,8)
m_weight <- matrix(0,8,1)
m_wwage <- matrix(0,30,8)
relwage <- matrix(0,30,1)
logx1x2 <- matrix(0,30,1)
```

```r
for (i in 1:30)
{
  for (j in 1:8)
  {
    #male

    # ratio of the average weekly wage of college graduates to the average weekly wage of high school g
    m_wageratio[i,j] <- mean(rwage[(deduc_3 == 1 | deduc_4 == 1) & year==1987+i & expbr==j & sex==1])/me

    # The fixed weight for each cell is the cell's average share of total employment over the period
    m_share[i,j] <- sum(wkswork[year==1987+i & expbr==j & sex==1])/sum(wkswork[year==1987+i])
    m_weight[j] <- mean(m_share[1:30,j])

    m_wwage[i,j] <- m_wageratio[i,j] * m_weight[j] #weighted wage

    #female
    #f_wageratio[i,j] <- mean(rwage[(deduc_3 == 1 | deduc_4 == 1) & year==1987+i & expbr==j & sex==2])/
    #f_share[i,j] <- sum(wkswork[year==1987+i & expbr==j & sex==2])/sum(wkswork[year==1987+i])
    #f_weight[j] <- mean(f_share[1:30,j])

    #f_wwage[i,j] <- f_wageratio[i,j] * f_weight[j] #weighted wage
  }

  # The overall college/high school wage ratio

  #relwage[i] <- sum(m_wwage[i,1:8])+sum(f_wwage[i,1:8])
  relwage[i] <- sum(m_wwage[i,1:8])
  logx1x2[i] <- log(relwage[i])
}

# time trend

time=array(1:30,dim = c(30,1))

# the regression of the Equation (19) in Katz and Murphy (1992)
m1 <- lm(logw1w2 ~ logx1x2 + time)
coef(m1)
```

```
## (Intercept)     logx1x2        time
##  -3.2799081  -0.4154597   0.1020731
```

```r
summary(m1)
```

```
##
## Call:
## lm(formula = logw1w2 ~ logx1x2 + time)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.47745 -0.35762 -0.06617  0.27860  0.72654
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.27991    0.64200  -5.109 2.27e-05 ***
```

```
## logx1x2     -0.41546    0.23769  -1.748 0.091842 .
## time         0.10207    0.02545   4.011 0.000429 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3841 on 27 degrees of freedom
## Multiple R-squared:  0.6814, Adjusted R-squared:  0.6578
## F-statistic: 28.87 on 2 and 27 DF,  p-value: 1.968e-07
```

The result from the period 1988-2017 shows that the relative (college/high school) supply has a negative impact on the relative (college/high school) wage ratio. This is in stark contrast to the result from the earlier period (1963-1987), which indicates that the relative supply has a positive and significant impact on the relative wage ratio.

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

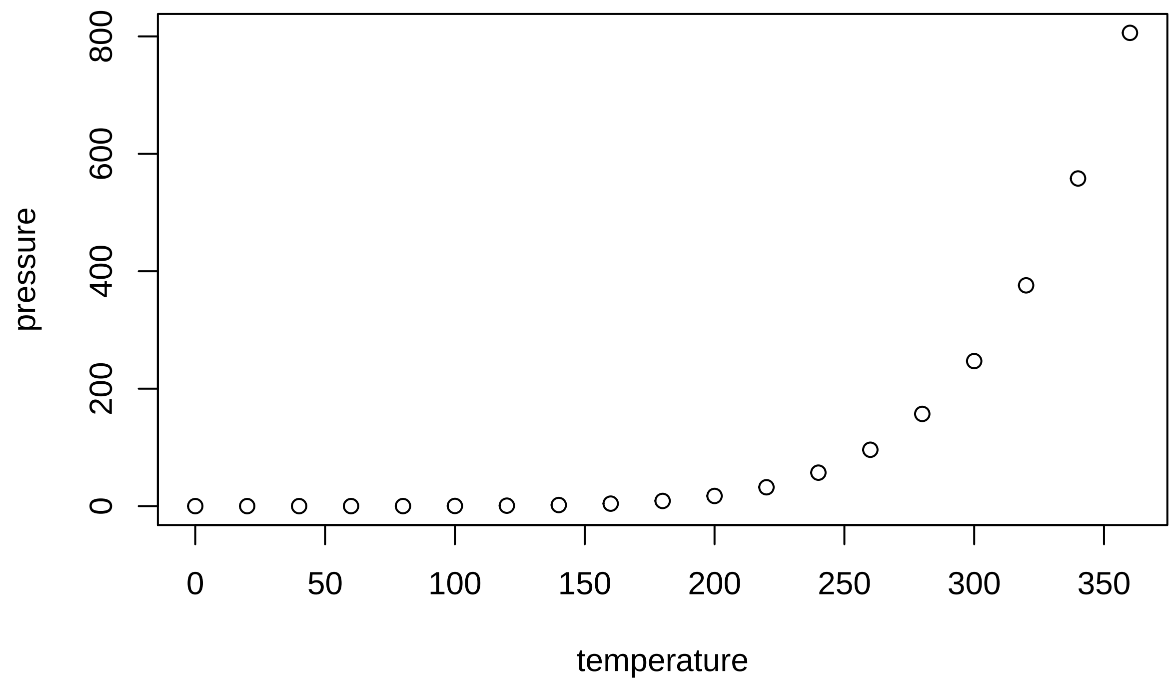
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```r
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00  
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.