prac2

Brian Ma

N/A

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.

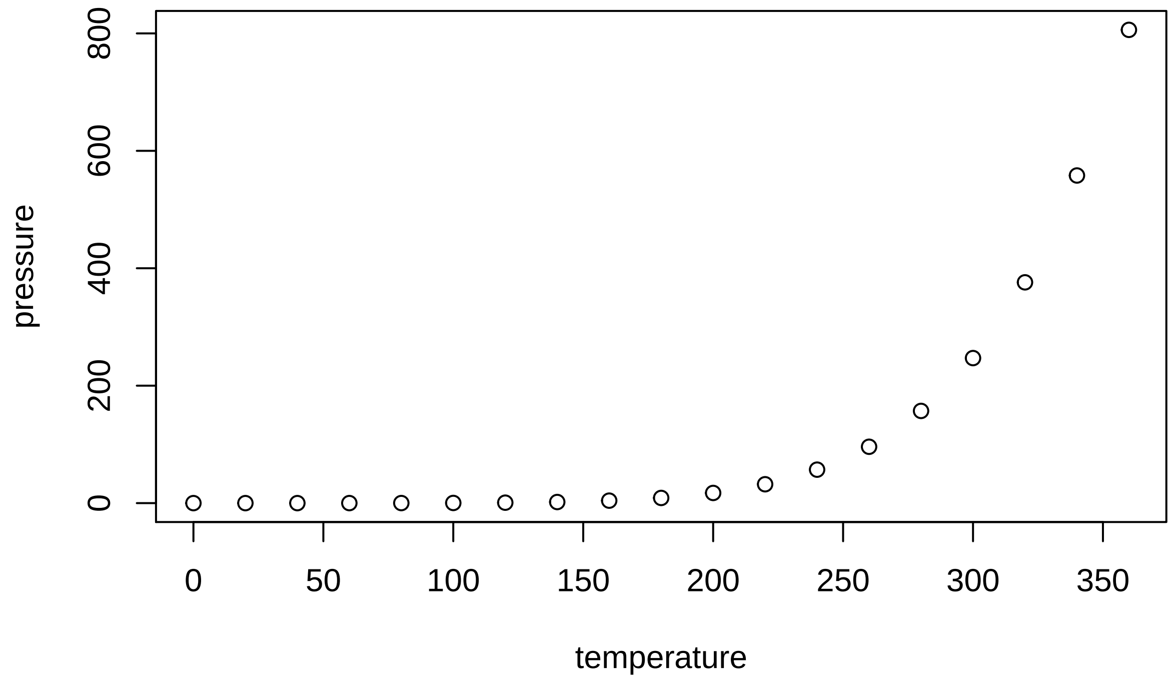
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
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.

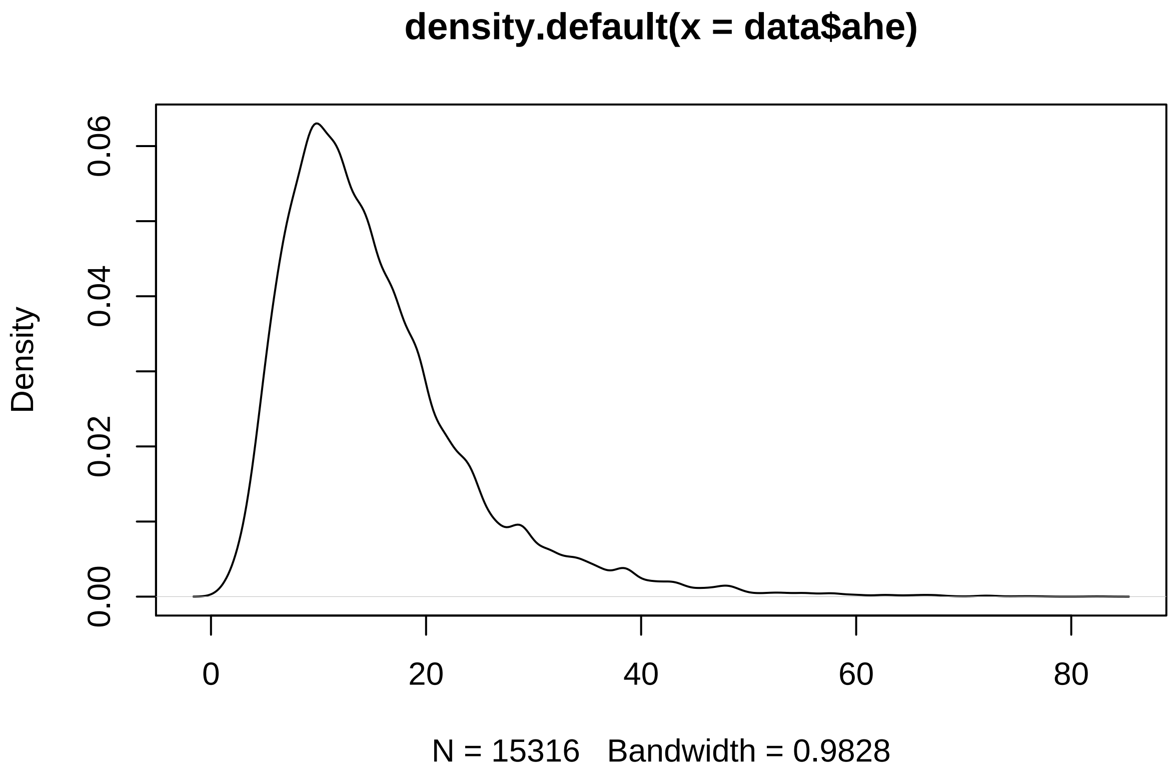
```
setwd("/Users/Brian/desktop/R/lab2")
data <- read.csv("cps92_08.csv")
summary(data)
```

```
##       year           ahe            bachelor          female      
##  Min.   :1992   Min.   : 1.314   Min.   :0.0000   Min.   :0.0000  
##  1st Qu.:1992   1st Qu.: 9.177   1st Qu.:0.0000   1st Qu.:0.0000  
##  Median :2008   Median :13.462   Median :0.0000   Median :0.0000  
##  Mean   :2000   Mean   :15.327   Mean   :0.4356   Mean   :0.4295  
##  3rd Qu.:2008   3rd Qu.:19.231   3rd Qu.:1.0000   3rd Qu.:1.0000  
##  Max.   :2008   Max.   :82.418   Max.   :1.0000   Max.   :1.0000  
##       age       
##  Min.   :25.00  
##  1st Qu.:27.00  
##  Median :30.00  
##  Mean   :29.64  
##  3rd Qu.:32.00  
##  Max.   :34.00
```

```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

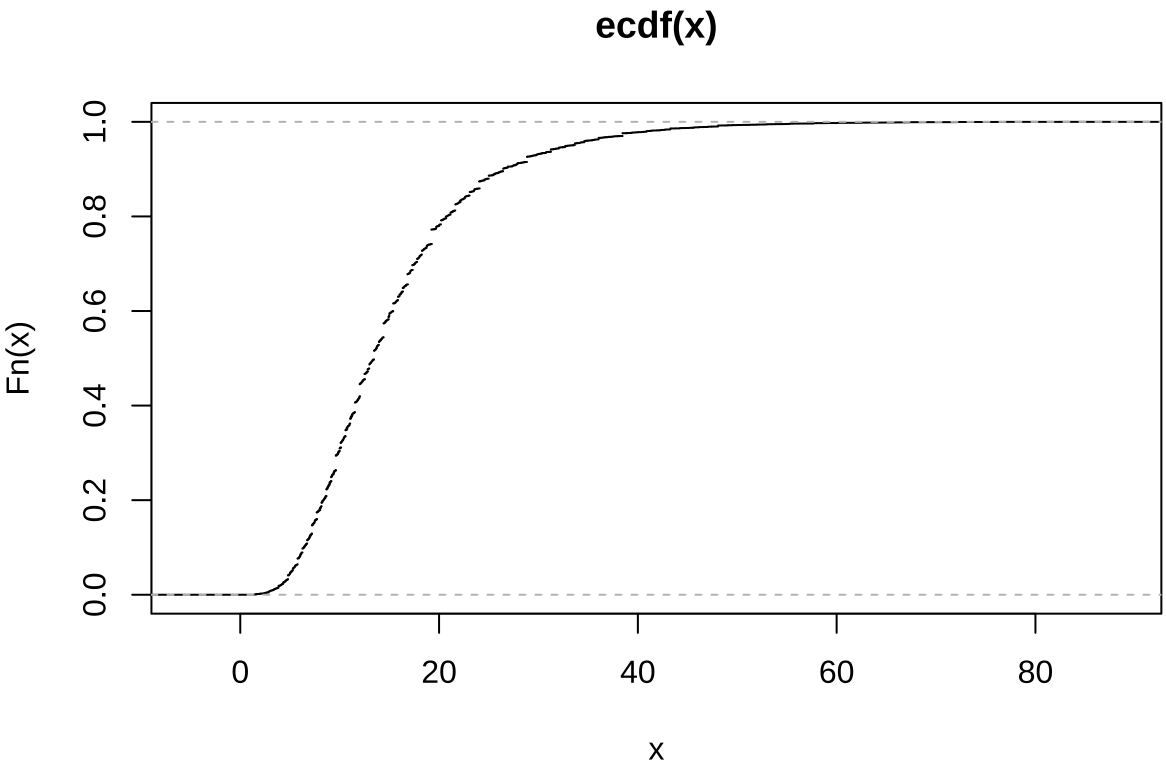
```r
data_ahe <- density(data$ahe)
plot(data_ahe)
```



```r
ggsave("ahe_pdf.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

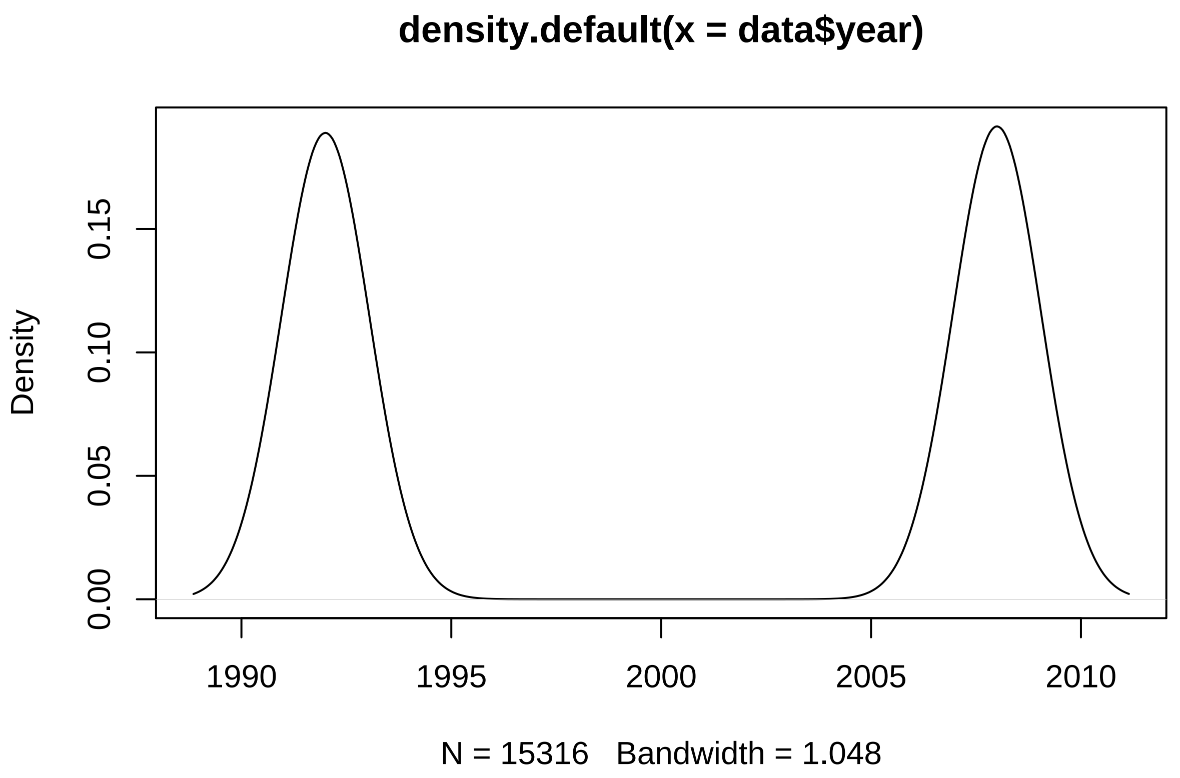
```r
plot.ecdf(data$ahe)
```



```
ggsave("ahe_ecdf.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

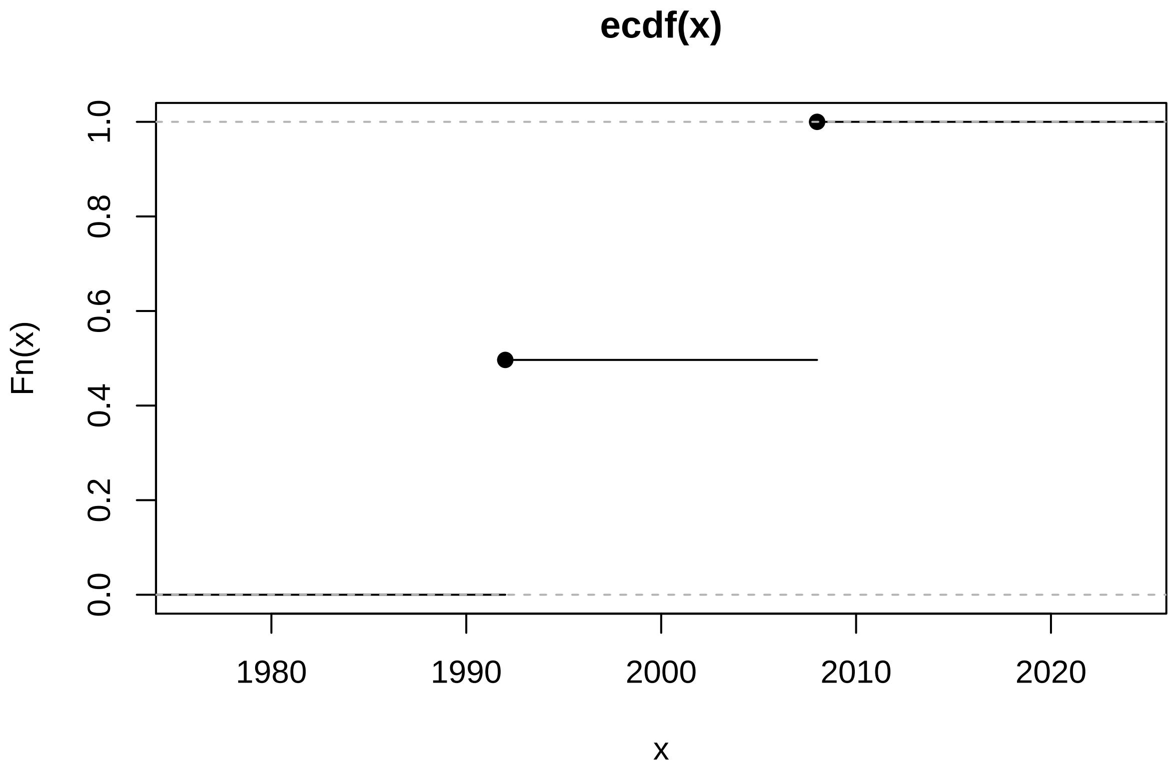
```
data_year <- density(data$year)
plot(data_year)
```



```
ggsave("year_density.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

```
plot.ecdf(data$year)
```



```r
ggsave("year_cdf.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

By oberserving the density function, we know that "ahe" is continuously dis tributed, while "year" is a discrete variable.

(above are the graph of Probability distribution function & Cumulative distribution function)

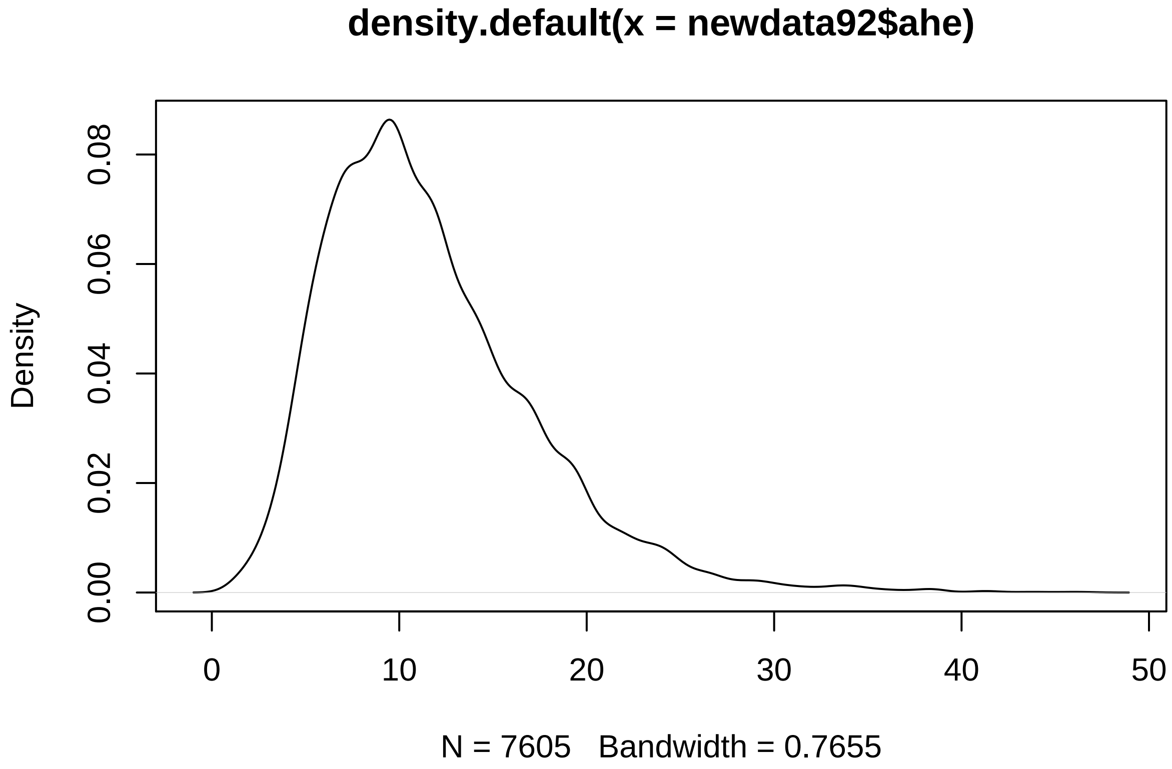
```r
summary(data$ahe)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   1.314   9.177  13.460  15.330  19.230  82.420
```

```r
sd(data$ahe)
```

```
## [1] 8.994762
```

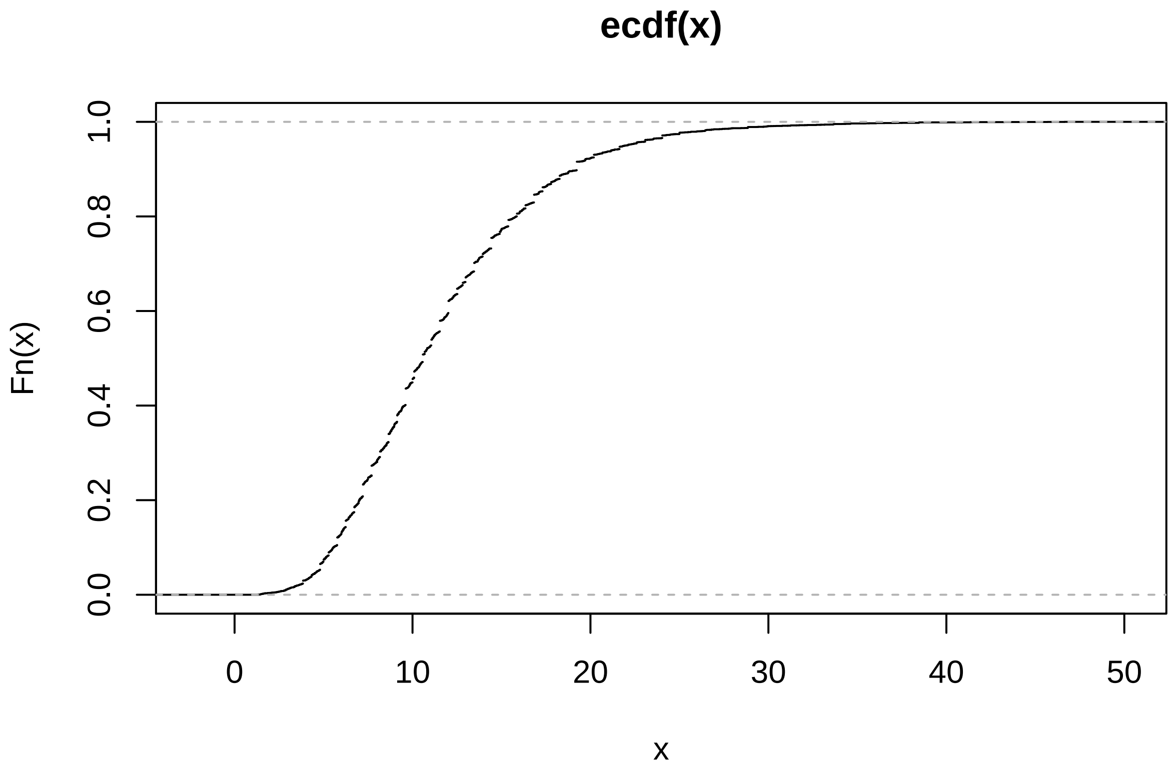
```
newdata92 <- subset(data, year == 1992)
newdata08 <- subset(data, year == 2008)
data_ahe_92 <- density(newdata92$ahe)
plot(data_ahe_92)
```



```
ggsave("ahe_92_density.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

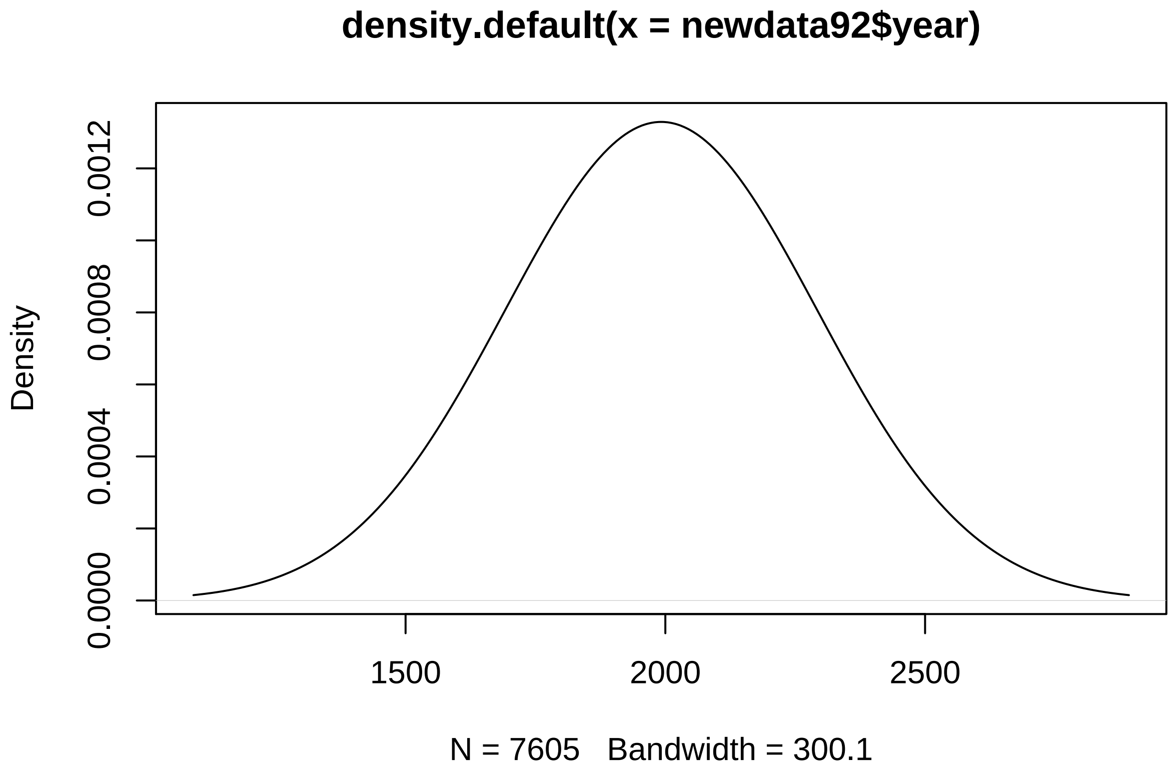
```
plot.ecdf(newdata92$ahe)
```



```
ggsave("ahe_cdf_92.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

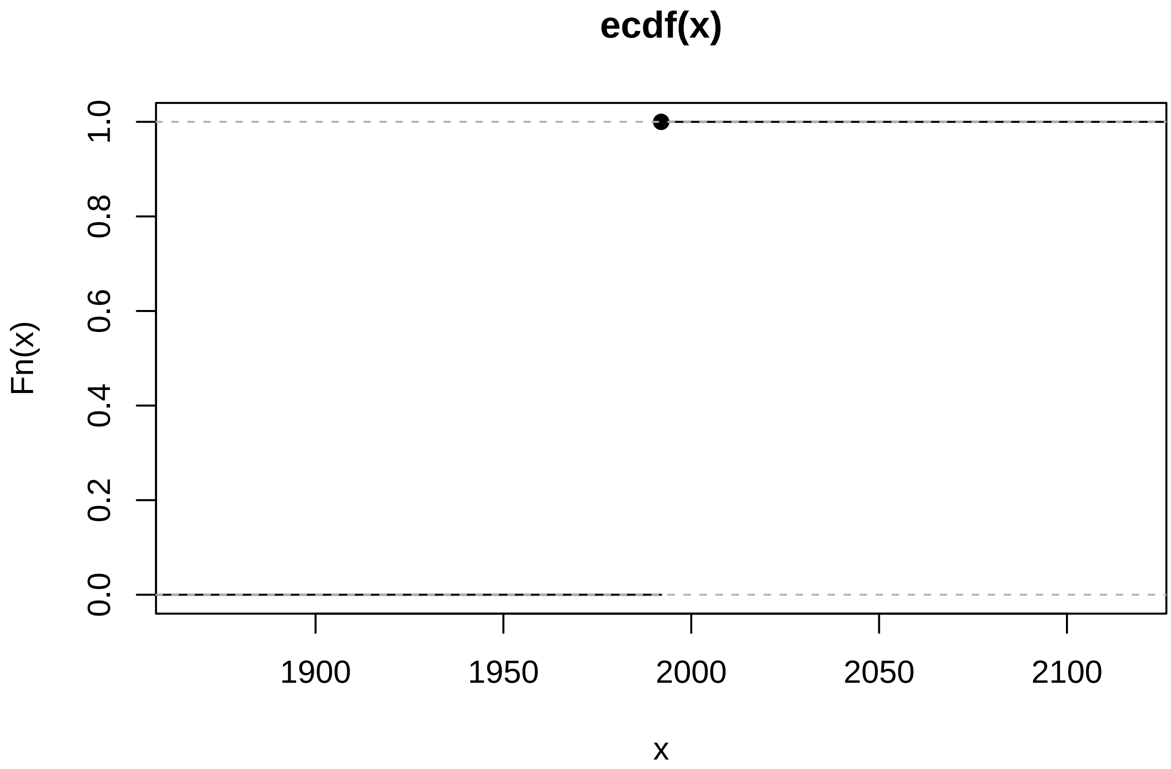
```
data_year_92 <- density(newdata92$year)
plot(data_year_92)
```



```
ggsave("92_pdf_year.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

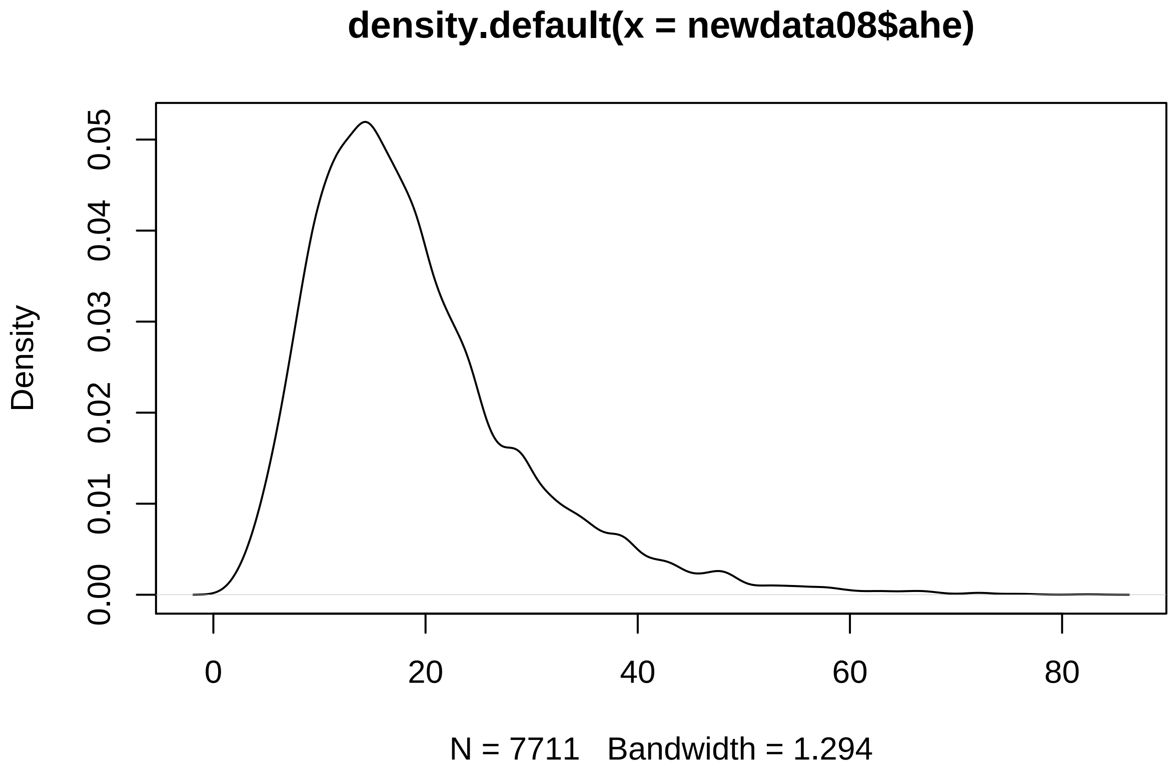
```
plot.ecdf(newdata92$year)
```



```
ggsave("year_92_cdf.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

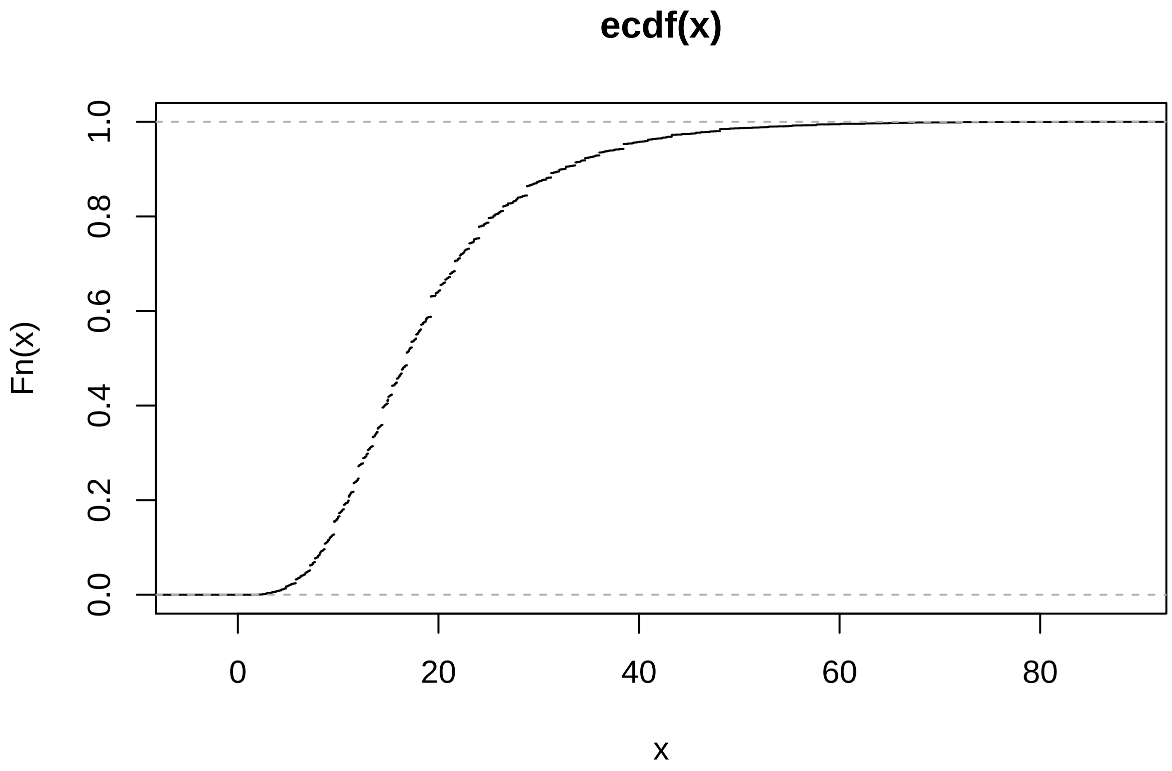
```
data_ahe_08 <- density(newdata08$ahe)
plot(data_ahe_08)
```



```r
ggsave("pdf_ahe_08.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

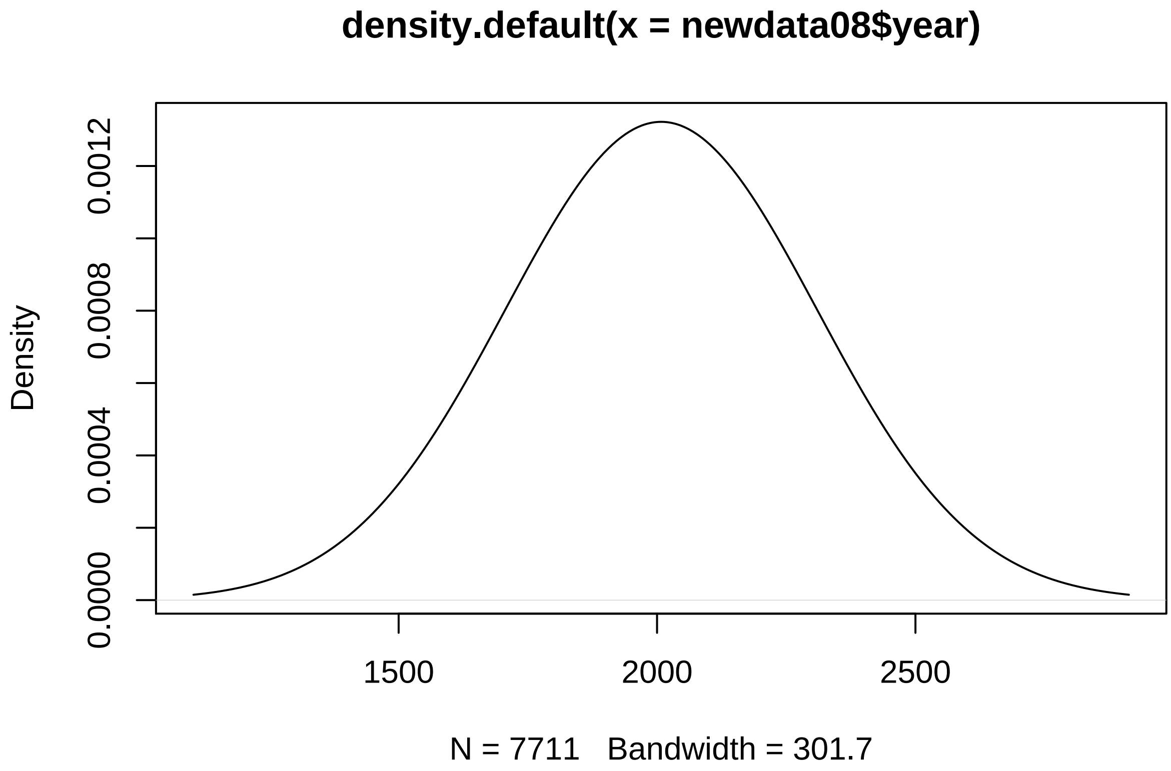
```r
plot.ecdf(newdata08$ahe)
```



```
ggsave("ahe_cdf_08.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

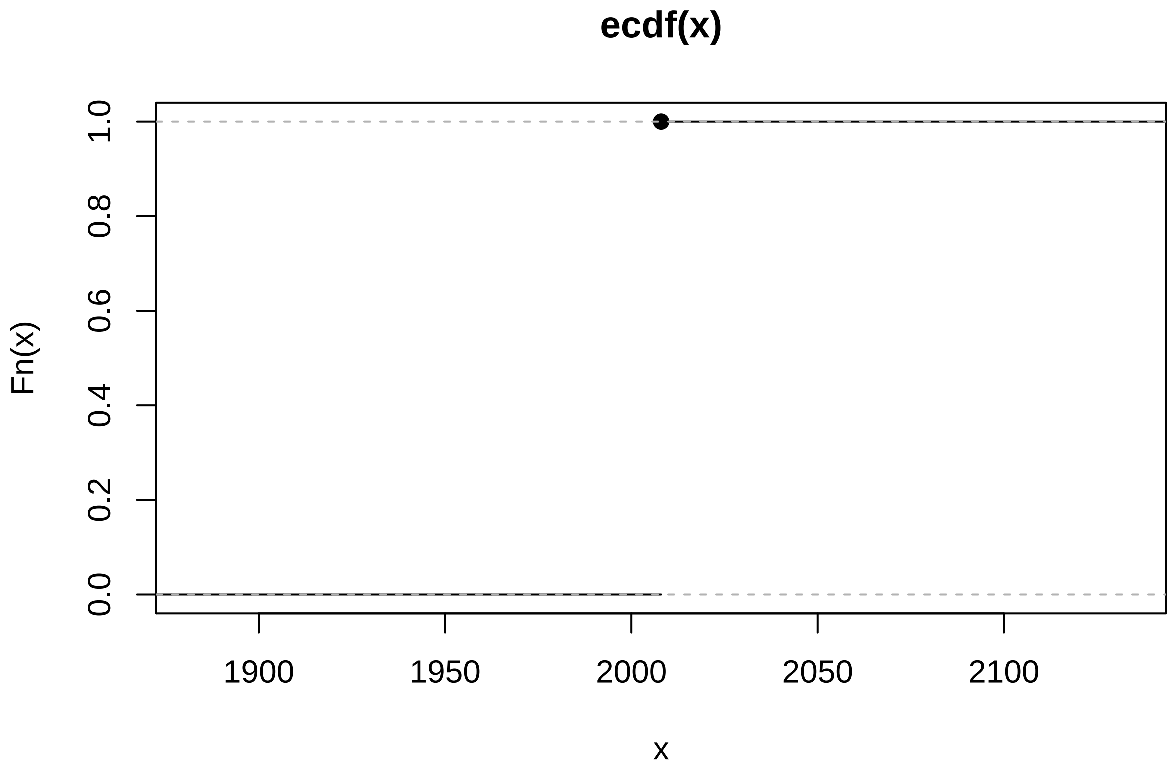
```
data_year_08 <- density(newdata08$year)
plot(data_year_08)
```



```r
ggsave("pdf_year_08.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

```r
plot.ecdf(newdata08$year)
```



```
ggsave("year_08_cdf.pdf")
```

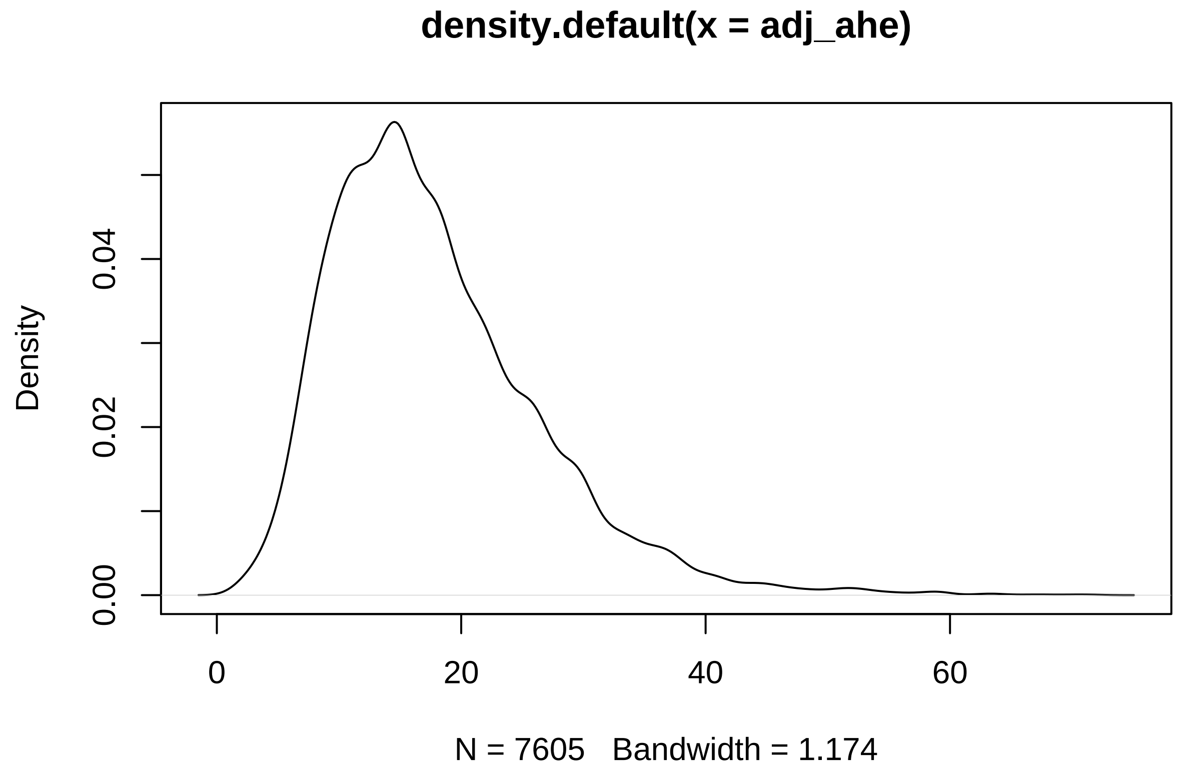
```
## Saving 6.5 x 4.5 in image
```

We cannot directly compare earnings in 1992 and 2008 because they are nomin al values, and doesn't reflect real earnigs.Inflations rate, CPI in 1992 = 140.3, 2018 =215.2

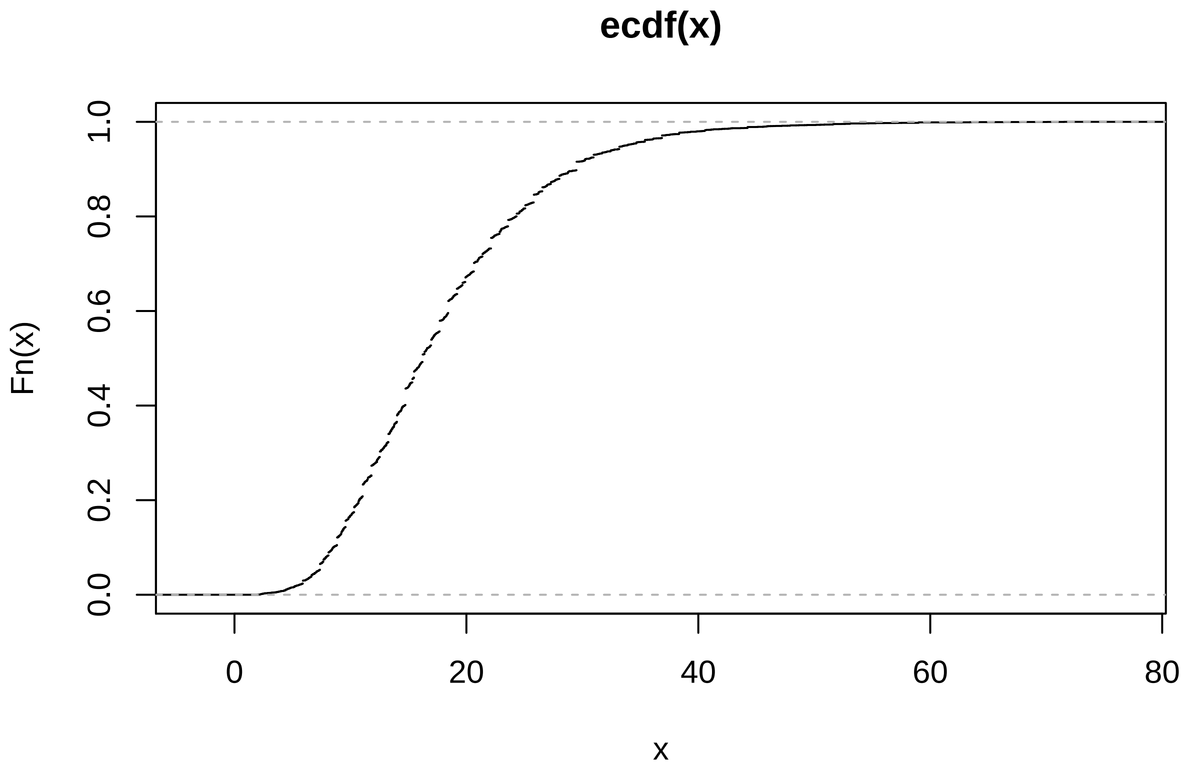
```
adj_ahe <-newdata92$ahe/140.3*215.2
sd(adj_ahe)
```

```
## [1] 8.609951
```

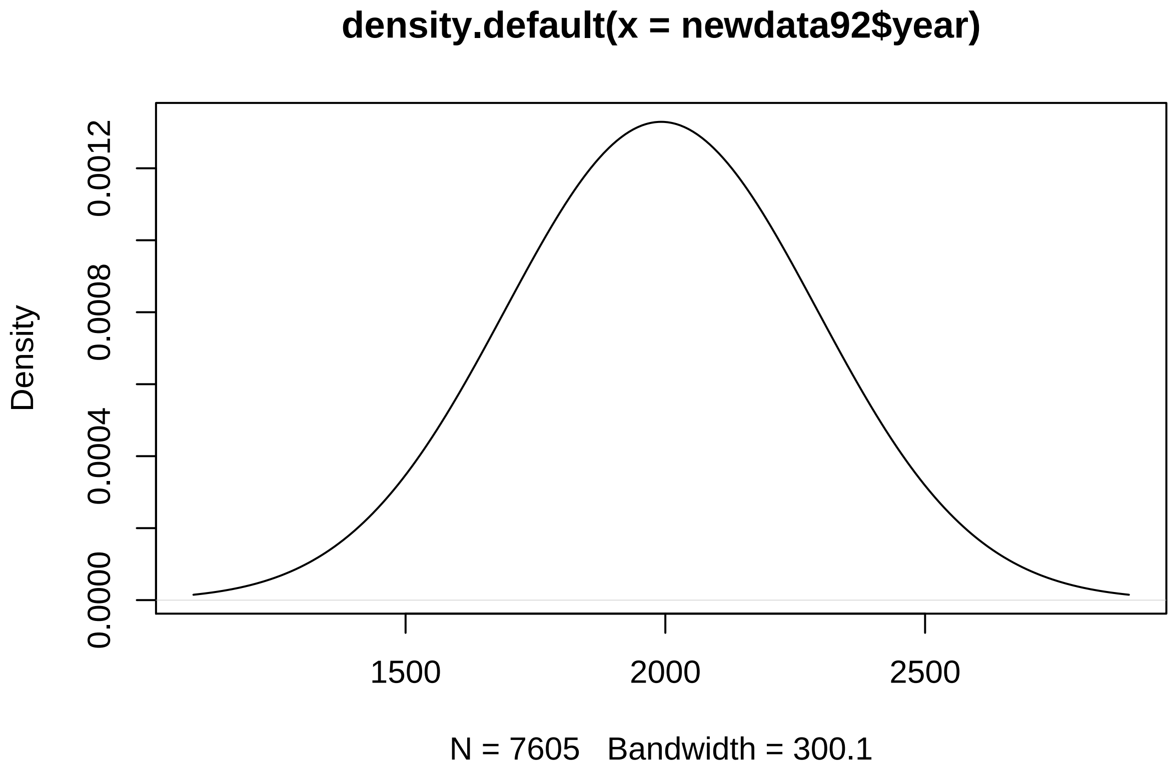
```
data_adjahe <- density(adj_ahe)
plot(data_adjahe)
```



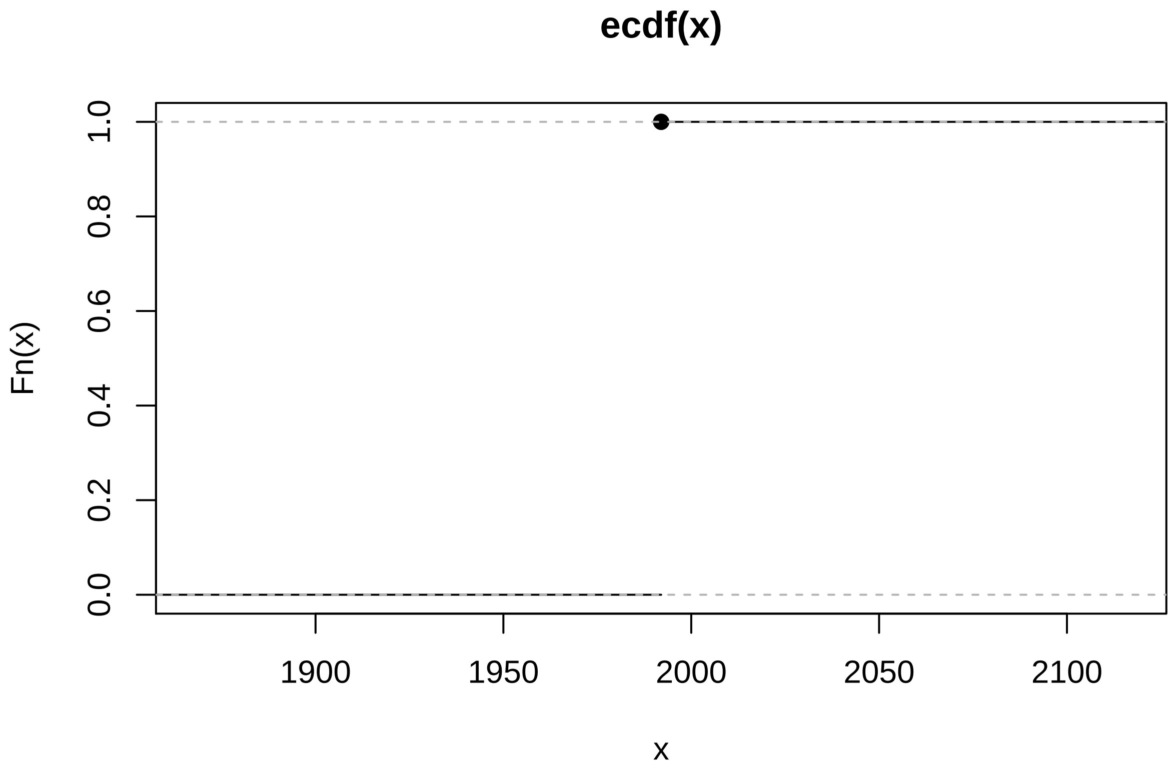
```r
plot.ecdf(adj_ahe)
```



```
data_year_adjahe <- density(newdata92$year)
plot(data_year_adjahe)
```



```
plot.ecdf(newdata92$year)
```



# the difference between the average earnings in 2008 and 1992 (measured in 2008 dollars)

```
Diff_bw_ahe_in_2008_1992 = mean(newdata08$ahe) - mean(adj_ahe)
Diff_bw_ahe_in_2008_1992
```

```
## [1] 1.142922
```

#account for the education level of the workers. # Compute D_hs the difference between the average earnings in2008 and #1992 (measured in 2008 dollars) for high school graduates. # ComputeD_cthedifferencebetweentheaverageearningsin2008and #1992 (measured in 2008 dollars) for college graduates. # Compute g92 the gap between earnings of college and high school #graduates in 1992 (measured in 2008 dollars). # Computeg g08 the gap between earnings of college and high school #graduates in 2008 (measured in 2008 dollars).

```
newdata92hs <- subset(newdata92, bachelor == 0)
newdata92c <- subset(newdata92, bachelor == 1)
newdata08hs <- subset(newdata08, bachelor == 0)
newdata08c <- subset(newdata08, bachelor == 1)
D_hs <- mean(newdata92hs$ahe) - mean(newdata08hs$ahe)
D_c <- mean(newdata92c$ahe) - mean(newdata08c$ahe)
g92 <- mean(newdata92c$ahe) - mean(newdata92hs$ahe)
g08 <- mean(newdata08c$ahe) - mean(newdata08hs$ahe)

D_hs
```

```
## [1] -5.348237
```

```
D_c
```

```
## [1] -8.706751
```

```
g92
```

```
## [1] 4.21808
```

```
g08
```

```
## [1] 7.576594
```