Project Alcohol

Dominic Ulicne

4/6/2021

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.

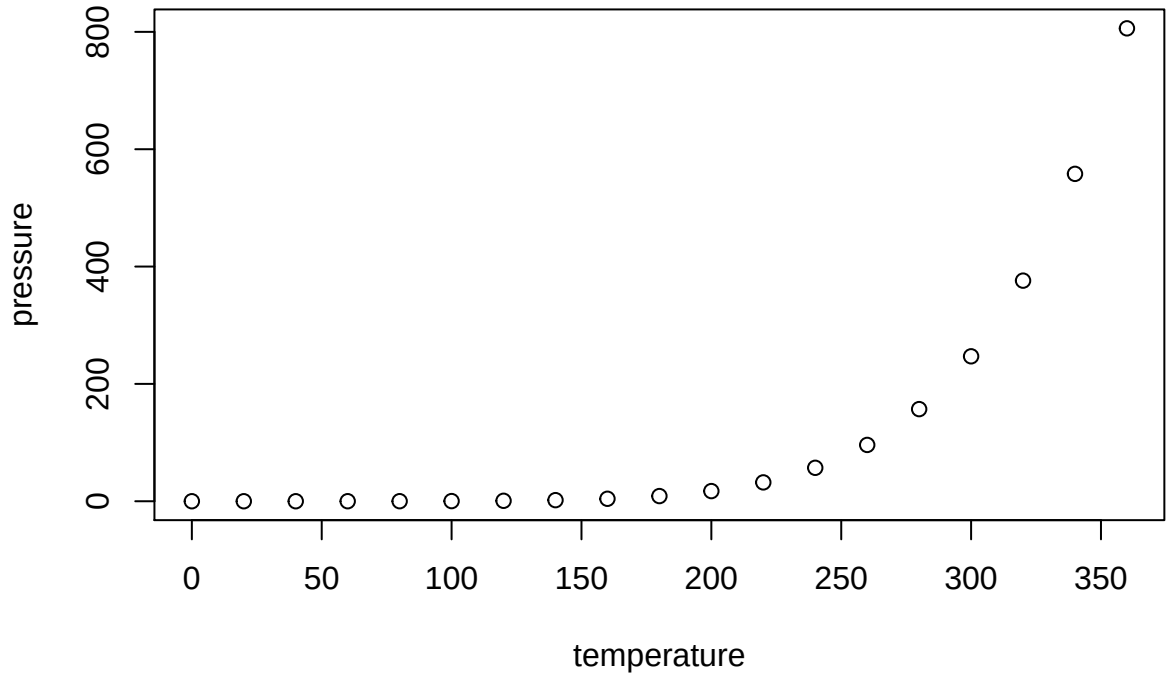
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
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.

```
math_class= read.csv("student-mat.csv")
port_class= read.csv("student-por.csv")
```

```
str(math_class)
```

```
## 'data.frame':    395 obs. of  33 variables:
##  $ school    : chr  "GP" "GP" "GP" "GP" ...
##  $ sex       : chr  "F" "F" "F" "F" ...
##  $ age       : int  18 17 15 15 16 16 16 17 15 15 ...
##  $ address   : chr  "U" "U" "U" "U" ...
##  $ famsize   : chr  "GT3" "GT3" "LE3" "GT3" ...
##  $ Pstatus   : chr  "A" "T" "T" "T" ...
##  $ Medu      : int  4 1 1 4 3 4 2 4 3 3 ...
##  $ Fedu      : int  4 1 1 2 3 3 2 4 2 4 ...
##  $ Mjob      : chr  "at_home" "at_home" "at_home" "health" ...
##  $ Fjob      : chr  "teacher" "other" "other" "services" ...
##  $ reason    : chr  "course" "course" "other" "home" ...
##  $ guardian  : chr  "mother" "father" "mother" "mother" ...
##  $ traveltime: int  2 1 1 1 1 1 1 2 1 1 ...
##  $ studytime : int  2 2 2 3 2 2 2 2 2 2 ...
##  $ failures  : int  0 0 3 0 0 0 0 0 0 0 ...
##  $ schoolsup : chr  "yes" "no" "yes" "no" ...
##  $ famsup    : chr  "no" "yes" "no" "yes" ...
```

```
##  $ paid      : chr  "no" "no" "yes" "yes" ...
##  $ activities: chr  "no" "no" "no" "yes" ...
##  $ nursery   : chr  "yes" "no" "yes" "yes" ...
##  $ higher    : chr  "yes" "yes" "yes" "yes" ...
##  $ internet  : chr  "no" "yes" "yes" "yes" ...
##  $ romantic  : chr  "no" "no" "no" "yes" ...
##  $ famrel    : int  4 5 4 3 4 5 4 4 4 5 ...
##  $ freetime  : int  3 3 3 2 3 4 4 1 2 5 ...
##  $ goout     : int  4 3 2 2 2 2 4 4 2 1 ...
##  $ Dalc      : int  1 1 2 1 1 1 1 1 1 1 ...
##  $ Walc      : int  1 1 3 1 2 2 1 1 1 1 ...
##  $ health    : int  3 3 3 5 5 5 3 1 1 5 ...
##  $ absences  : int  6 4 10 2 4 10 0 6 0 0 ...
##  $ G1        : int  5 5 7 15 6 15 12 6 16 14 ...
##  $ G2        : int  6 5 8 14 10 15 12 5 18 15 ...
##  $ G3        : int  6 6 10 15 10 15 11 6 19 15 ...
```

```
str(port_class)
```

```
## 'data.frame':    649 obs. of  33 variables:
##  $ school    : chr  "GP" "GP" "GP" "GP" ...
##  $ sex       : chr  "F" "F" "F" "F" ...
##  $ age       : int  18 17 15 15 16 16 16 17 15 15 ...
##  $ address   : chr  "U" "U" "U" "U" ...
##  $ famsize   : chr  "GT3" "GT3" "LE3" "GT3" ...
##  $ Pstatus   : chr  "A" "T" "T" "T" ...
##  $ Medu      : int  4 1 1 4 3 4 2 4 3 3 ...
##  $ Fedu      : int  4 1 1 2 3 3 2 4 2 4 ...
##  $ Mjob      : chr  "at_home" "at_home" "at_home" "health" ...
##  $ Fjob      : chr  "teacher" "other" "other" "services" ...
##  $ reason    : chr  "course" "course" "other" "home" ...
##  $ guardian  : chr  "mother" "father" "mother" "mother" ...
##  $ traveltime: int  2 1 1 1 1 1 1 2 1 1 ...
##  $ studytime : int  2 2 2 3 2 2 2 2 2 2 ...
##  $ failures  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ schoolsup : chr  "yes" "no" "yes" "no" ...
##  $ famsup    : chr  "no" "yes" "no" "yes" ...
##  $ paid      : chr  "no" "no" "no" "no" ...
##  $ activities: chr  "no" "no" "no" "yes" ...
##  $ nursery   : chr  "yes" "no" "yes" "yes" ...
##  $ higher    : chr  "yes" "yes" "yes" "yes" ...
##  $ internet  : chr  "no" "yes" "yes" "yes" ...
##  $ romantic  : chr  "no" "no" "no" "yes" ...
##  $ famrel    : int  4 5 4 3 4 5 4 4 4 5 ...
##  $ freetime  : int  3 3 3 2 3 4 4 1 2 5 ...
##  $ goout     : int  4 3 2 2 2 2 4 4 2 1 ...
##  $ Dalc      : int  1 1 2 1 1 1 1 1 1 1 ...
##  $ Walc      : int  1 1 3 1 2 2 1 1 1 1 ...
##  $ health    : int  3 3 3 5 5 5 3 1 1 5 ...
##  $ absences  : int  4 2 6 0 0 6 0 2 0 0 ...
##  $ G1        : int  0 9 12 14 11 12 13 10 15 12 ...
##  $ G2        : int  11 11 13 14 13 12 12 13 16 12 ...
##  $ G3        : int  11 11 12 14 13 13 13 13 17 13 ...
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
glimpse(math_class)
```

```
## Rows: 395
## Columns: 33
## $ school     <chr> "GP", "GP", "GP", "GP", "GP", "GP", "GP", "GP", "GP", "G...
## $ sex        <chr> "F", "F", "F", "F", "F", "M", "M", "F", "M", "M", "F", "...
## $ age        <int> 18, 17, 15, 15, 16, 16, 16, 17, 15, 15, 15, 15, 15, 15, ...
## $ address    <chr> "U", "U", "U", "U", "U", "U", "U", "U", "U", "U", "U", "...
## $ famsize    <chr> "GT3", "GT3", "LE3", "GT3", "GT3", "LE3", "LE3", "GT3", ...
## $ Pstatus    <chr> "A", "T", "T", "T", "T", "T", "T", "A", "A", "T", "T", "...
## $ Medu       <int> 4, 1, 1, 4, 3, 4, 2, 4, 3, 3, 4, 2, 4, 4, 2, 4, 4, 3, 3,...
## $ Fedu       <int> 4, 1, 1, 2, 3, 3, 2, 4, 2, 4, 4, 1, 4, 3, 2, 4, 4, 3, 2,...
## $ Mjob       <chr> "at_home", "at_home", "at_home", "health", "other", "ser...
## $ Fjob       <chr> "teacher", "other", "other", "services", "other", "other...
## $ reason     <chr> "course", "course", "other", "home", "home", "reputation...
## $ guardian   <chr> "mother", "father", "mother", "mother", "father", "mothe...
## $ traveltime <int> 2, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 3, 1, 2, 1, 1, 1, 3, 1,...
## $ studytime  <int> 2, 2, 2, 3, 2, 2, 2, 2, 2, 2, 2, 3, 1, 2, 3, 1, 3, 2, 1,...
## $ failures   <int> 0, 0, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3,...
## $ schoolsup  <chr> "yes", "no", "yes", "no", "no", "no", "no", "yes", "no",...
## $ famsup     <chr> "no", "yes", "no", "yes", "yes", "yes", "no", "yes", "ye...
## $ paid       <chr> "no", "no", "yes", "yes", "yes", "yes", "no", "no", "yes...
## $ activities <chr> "no", "no", "no", "yes", "no", "yes", "no", "no", "no", ...
## $ nursery    <chr> "yes", "no", "yes", "yes", "yes", "yes", "yes", "yes", "...
## $ higher     <chr> "yes", "yes", "yes", "yes", "yes", "yes", "yes", "yes", ...
## $ internet   <chr> "no", "yes", "yes", "yes", "no", "yes", "yes", "no", "ye...
## $ romantic   <chr> "no", "no", "no", "yes", "no", "no", "no", "no", "no", "...
## $ famrel     <int> 4, 5, 4, 3, 4, 5, 4, 4, 4, 5, 3, 5, 4, 5, 4, 4, 3, 5, 5,...
## $ freetime   <int> 3, 3, 3, 2, 3, 4, 4, 1, 2, 5, 3, 2, 3, 4, 5, 4, 2, 3, 5,...
## $ goout      <int> 4, 3, 2, 2, 2, 2, 4, 4, 2, 1, 3, 2, 3, 3, 2, 4, 3, 2, 5,...
## $ Dalc       <int> 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2,...
## $ Walc       <int> 1, 1, 3, 1, 2, 2, 1, 1, 1, 1, 2, 1, 3, 2, 1, 2, 2, 1, 4,...
## $ health     <int> 3, 3, 3, 5, 5, 5, 3, 1, 1, 5, 2, 4, 5, 3, 3, 2, 2, 4, 5,...
## $ absences   <int> 6, 4, 10, 2, 4, 10, 0, 6, 0, 0, 0, 4, 2, 2, 0, 4, 6, 4, ...
## $ G1         <int> 5, 5, 7, 15, 6, 15, 12, 6, 16, 14, 10, 10, 14, 10, 14, 1...
## $ G2         <int> 6, 5, 8, 14, 10, 15, 12, 5, 18, 15, 8, 12, 14, 10, 16, 1...
## $ G3         <int> 6, 6, 10, 15, 10, 15, 11, 6, 19, 15, 9, 12, 14, 11, 16, ...
```

```r
glimpse(port_class)
```

```
## Rows: 649
## Columns: 33
## $ school     <chr> "GP", "GP", "GP", "GP", "GP", "GP", "GP", "GP", "GP", "G...
## $ sex        <chr> "F", "F", "F", "F", "F", "M", "M", "F", "M", "M", "F", "...
## $ age        <int> 18, 17, 15, 15, 16, 16, 16, 17, 15, 15, 15, 15, 15, 15, ...
## $ address    <chr> "U", "U", "U", "U", "U", "U", "U", "U", "U", "U", "U", "...
## $ famsize    <chr> "GT3", "GT3", "LE3", "GT3", "GT3", "LE3", "LE3", "GT3", ...
## $ Pstatus    <chr> "A", "T", "T", "T", "T", "T", "T", "A", "A", "T", "T", "...
## $ Medu       <int> 4, 1, 1, 4, 3, 4, 2, 4, 3, 3, 4, 2, 4, 4, 2, 4, 4, 3, 3,...
## $ Fedu       <int> 4, 1, 1, 2, 3, 3, 2, 4, 2, 4, 4, 1, 4, 3, 2, 4, 4, 3, 2,...
## $ Mjob       <chr> "at_home", "at_home", "at_home", "health", "other", "ser...
## $ Fjob       <chr> "teacher", "other", "other", "services", "other", "other...
## $ reason     <chr> "course", "course", "other", "home", "home", "reputation...
## $ guardian   <chr> "mother", "father", "mother", "mother", "father", "mothe...
## $ traveltime <int> 2, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 3, 1, 2, 1, 1, 1, 3, 1,...
## $ studytime  <int> 2, 2, 2, 3, 2, 2, 2, 2, 2, 2, 2, 3, 1, 2, 3, 1, 3, 2, 1,...
## $ failures   <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3,...
## $ schoolsup  <chr> "yes", "no", "yes", "no", "no", "no", "no", "yes", "no",...
## $ famsup     <chr> "no", "yes", "no", "yes", "yes", "yes", "no", "yes", "ye...
## $ paid       <chr> "no", "no", "no", "no", "no", "no", "no", "no", "no", "n...
## $ activities <chr> "no", "no", "no", "yes", "no", "yes", "no", "no", "no", ...
## $ nursery    <chr> "yes", "no", "yes", "yes", "yes", "yes", "yes", "yes", "...
## $ higher     <chr> "yes", "yes", "yes", "yes", "yes", "yes", "yes", "yes", ...
## $ internet   <chr> "no", "yes", "yes", "yes", "no", "yes", "yes", "no", "ye...
## $ romantic   <chr> "no", "no", "no", "yes", "no", "no", "no", "no", "no", "...
## $ famrel     <int> 4, 5, 4, 3, 4, 5, 4, 4, 4, 5, 3, 5, 4, 5, 4, 4, 3, 5, 5,...
## $ freetime   <int> 3, 3, 3, 2, 3, 4, 4, 1, 2, 5, 3, 2, 3, 4, 5, 4, 2, 3, 5,...
## $ goout      <int> 4, 3, 2, 2, 2, 2, 4, 4, 2, 1, 3, 2, 3, 3, 2, 4, 3, 2, 5,...
## $ Dalc       <int> 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2,...
## $ Walc       <int> 1, 1, 3, 1, 2, 2, 1, 1, 1, 1, 2, 1, 3, 2, 1, 2, 2, 1, 4,...
## $ health     <int> 3, 3, 3, 5, 5, 5, 3, 1, 1, 5, 2, 4, 5, 3, 3, 2, 2, 4, 5,...
## $ absences   <int> 4, 2, 6, 0, 0, 6, 0, 2, 0, 0, 2, 0, 0, 0, 0, 6, 10, 2, 2...
## $ G1         <int> 0, 9, 12, 14, 11, 12, 13, 10, 15, 12, 14, 10, 12, 12, 14...
## $ G2         <int> 11, 11, 13, 14, 13, 12, 12, 13, 16, 12, 14, 12, 13, 12, ...
## $ G3         <int> 11, 11, 12, 14, 13, 13, 13, 13, 17, 13, 14, 13, 12, 13, ...
```

```r
table(math_class$Dalc)
```

```
##
##   1   2   3   4   5
## 276  75  26   9   9
```

```r
table(math_class$G3)
```

```
##
##  0  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
## 38  1  7 15  9 32 28 56 47 31 31 27 33 16  6 12  5  1
```

```r
table(port_class$Dalc)
```

```
##
##   1   2   3   4   5
## 451 121  43  17  17
```

```r
table(port_class$G3)
```

```
##
##   0   1   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19
##  15   1   1   3  10  35  35  97 104  72  82  63  49  36  29  15   2
```

```r
library(caTools)
```

```
## Warning: package 'caTools' was built under R version 4.0.3
```

```r
set.seed(123)
split = sample.split(math_class$G3, SplitRatio = 0.8)
training_set = subset(math_class, split == TRUE)
test_set = subset(math_class, split == FALSE)
```

```r
regressor = lm(formula = G3 ~ famrel + freetime + goout + Dalc + Walc + health + absences,
               data = training_set)
summary(regressor)
```

```
##
## Call:
## lm(formula = G3 ~ famrel + freetime + goout + Dalc + Walc + health +
##     absences, data = training_set)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.2502  -2.0045   0.2599   3.0435   8.7349
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.31033    1.59945   7.071 1.02e-11 ***
## famrel       0.37444    0.30353   1.234  0.21827
## freetime     0.37922    0.27250   1.392  0.16503
## goout       -0.70786    0.25892  -2.734  0.00662 **
## Dalc        -0.30520    0.38034  -0.802  0.42291
## Walc         0.25002    0.27319   0.915  0.36082
## health      -0.46867    0.18606  -2.519  0.01228 *
## absences     0.02670    0.03225   0.828  0.40832
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.516 on 310 degrees of freedom
## Multiple R-squared:  0.0487, Adjusted R-squared:  0.02722
## F-statistic: 2.267 on 7 and 310 DF,  p-value: 0.02901
```

```r
regressor = lm(formula = G3 ~ famrel + freetime + goout + Walc + health + absences,
               data = training_set)
summary(regressor)
```

```
##
## Call:
## lm(formula = G3 ~ famrel + freetime + goout + Walc + health +
##     absences, data = training_set)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.2344  -1.9057   0.3188   2.9930   8.6653
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.26187    1.59739   7.050 1.16e-11 ***
## famrel       0.37627    0.30335   1.240  0.21577
## freetime     0.33983    0.26789   1.269  0.20555
## goout       -0.69723    0.25843  -2.698  0.00736 **
## Walc         0.12524    0.22449   0.558  0.57733
## health      -0.47178    0.18592  -2.538  0.01165 *
## absences     0.02522    0.03218   0.784  0.43369
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.514 on 311 degrees of freedom
## Multiple R-squared:  0.04673,    Adjusted R-squared:  0.02834
## F-statistic: 2.541 on 6 and 311 DF,  p-value: 0.02043
```

```r
regressor = lm(formula = G3 ~ famrel + freetime + goout + Walc + health,
               data = training_set)
summary(regressor)
```

```
##
## Call:
## lm(formula = G3 ~ famrel + freetime + goout + Walc + health,
##     data = training_set)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.3425  -1.8324   0.4069   2.8799   8.6177
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  11.4754     1.5730   7.295 2.48e-12 ***
## famrel        0.3665     0.3029   1.210  0.22715
## freetime      0.3157     0.2659   1.187  0.23614
## goout        -0.6953     0.2583  -2.692  0.00748 **
## Walc          0.1451     0.2229   0.651  0.51553
## health       -0.4725     0.1858  -2.543  0.01147 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 4.511 on 312 degrees of freedom
## Multiple R-squared:  0.04484,    Adjusted R-squared:  0.02954
## F-statistic:  2.93 on 5 and 312 DF,  p-value: 0.01335
```

```
regressor = lm(formula = G3 ~ famrel + freetime + goout + health,
               data = training_set)
summary(regressor)
```

```
##
## Call:
## lm(formula = G3 ~ famrel + freetime + goout + health, data = training_set)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.3793  -1.8177   0.4957   2.9042   8.7074
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  11.6715     1.5425   7.567 4.31e-13 ***
## famrel        0.3295     0.2972   1.108  0.26850
## freetime      0.3129     0.2657   1.178  0.23976
## goout        -0.6225     0.2325  -2.677  0.00782 **
## health       -0.4549     0.1836  -2.477  0.01378 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.507 on 313 degrees of freedom
## Multiple R-squared:  0.04355,    Adjusted R-squared:  0.03132
## F-statistic: 3.563 on 4 and 313 DF,  p-value: 0.007359
```

```
regressor = lm(formula = G3 ~ freetime + goout + health,
               data = training_set)
summary(regressor)
```

```
##
## Call:
## lm(formula = G3 ~ freetime + goout + health, data = training_set)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.6557  -1.8011   0.3814   2.9471   8.6378
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.7849     1.1711  10.917  < 2e-16 ***
## freetime      0.3395     0.2647   1.283  0.20050
## goout        -0.6144     0.2325  -2.643  0.00863 **
## health       -0.4364     0.1830  -2.385  0.01765 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.508 on 314 degrees of freedom
```

```
## Multiple R-squared:  0.03979,    Adjusted R-squared:  0.03062
## F-statistic: 4.337 on 3 and 314 DF,  p-value: 0.005161
```

```r
regressor = lm(formula = G3 ~ goout + health,
               data = training_set)
summary(regressor)
```

```
##
## Call:
## lm(formula = G3 ~ goout + health, data = training_set)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.2329  -1.7785   0.4422   3.1174   8.7948
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13.6172     0.9759  13.954   <2e-16 ***
## goout        -0.5478     0.2269  -2.415   0.0163 *
## health       -0.4182     0.1826  -2.291   0.0227 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.513 on 315 degrees of freedom
## Multiple R-squared:  0.03476,    Adjusted R-squared:  0.02863
## F-statistic: 5.672 on 2 and 315 DF,  p-value: 0.003803
```