

stats-101a-project-markdown

Naren Akurati

3/11/2018

```
#data cleanup
```

```
happiness_data <- read.table("Happiness.txt", header = TRUE)
```

```
head(happiness_data)
```

```
##   Household Health OwnHome Instagram Marital Sex Age Children Education
## 3         2      2        0         2      1  1  72         2         16
## 4         4      2        1         0      1  2  43         4         12
## 5         3      1        0         1      1  2  55         2         18
## 6         2      0        1         1      1  2  53         2         14
## 7         3      4        1         0      1  1  50         2         14
## 8         2      2        0         1      1  2  23         3         11
```

```
##   JobSat Income WorkHrs Happy
```

```
## 3      0      0      -1      1
## 4      0  5265      -1      2
## 5      3   936      15      1
## 6      0      0      -1      1
## 7      0 164382      -1      2
## 8      2   7605      30      1
```

```
happiness_data$Household[happiness_data$Household == 8 | happiness_data$Household == 9] <- NA
```

```
happiness_data$Health[happiness_data$Health == 8 | happiness_data$Health == 9 | happiness_data$Health == 10] <- NA
```

```
happiness_data$Health[happiness_data$Health == 1] <- 400
```

```
happiness_data$Health[happiness_data$Health == 2] <- 300
```

```
happiness_data$Health[happiness_data$Health == 3] <- 2
```

```
happiness_data$Health[happiness_data$Health == 4] <- 1
```

```
happiness_data$Health[happiness_data$Health == 400] <- 4
```

```
happiness_data$Health[happiness_data$Health == 300] <- 3
```

```
happiness_data$OwnHome[happiness_data$OwnHome == 0 | happiness_data$OwnHome == 8 | happiness_data$OwnHome == 9] <- NA
```

```
happiness_data$Instagram[happiness_data$Instagram == 0 | happiness_data$Instagram == 8 | happiness_data$Instagram == 9] <- NA
```

```
happiness_data$Marital[happiness_data$Marital == 9] <- NA
```

```
happiness_data$Age[happiness_data$Age == 89 | happiness_data$Age == 98 | happiness_data$Age == 99] <- NA
```

```
happiness_data$Children[happiness_data$Children == 8 | happiness_data$Children == 9] <- NA
```

```
happiness_data$Education[happiness_data$Education == 97 | happiness_data$Education == 98 | happiness_data$Education == 99] <- NA
```

```
happiness_data$JobSat[happiness_data$JobSat == 0 | happiness_data$JobSat == 8 | happiness_data$JobSat == 9] <- NA
```

```
happiness_data$Income[happiness_data$Income == 0 | happiness_data$Income == 999998 | happiness_data$Income == 999999] <- NA
```

```
happiness_data$WorkHrs[happiness_data$WorkHrs == -1 | happiness_data$WorkHrs == 998 | happiness_data$WorkHrs == 999] <- NA
```

```
happiness_data$Happy[happiness_data$Happy == 0 | happiness_data$Happy == 8 | happiness_data$Happy == 9] <- NA
```

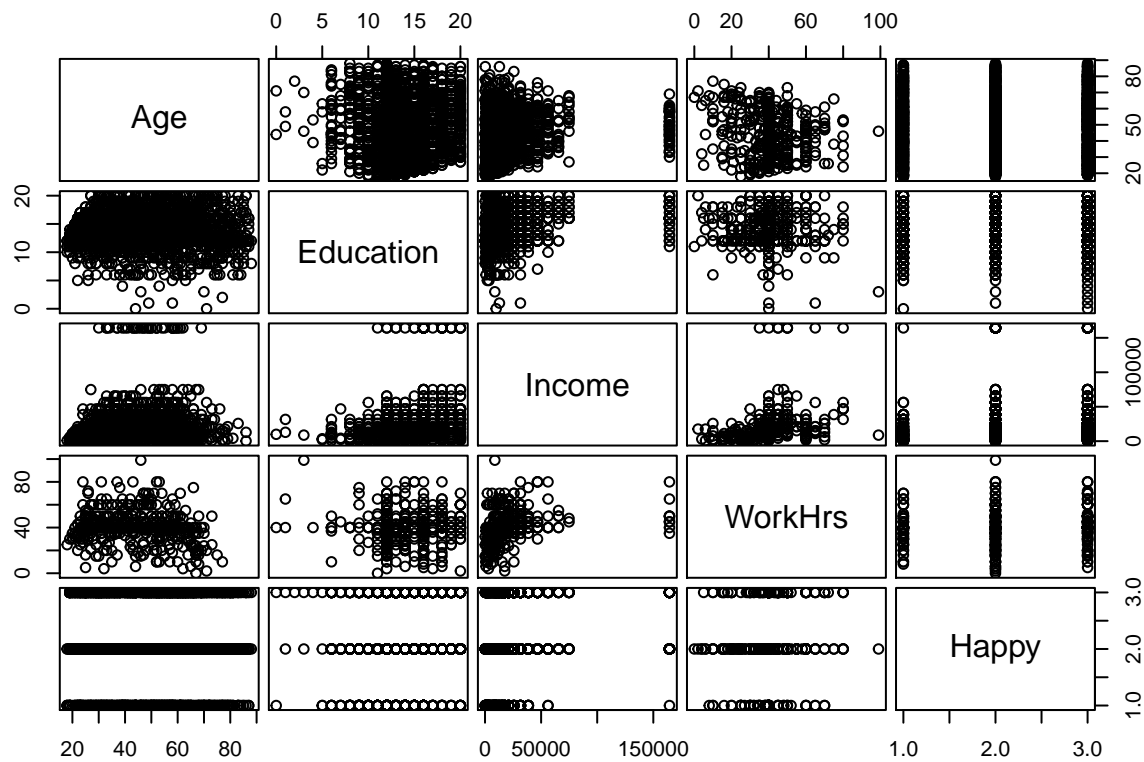
```

happiness_data$Happy[happiness_data$Happy == 1] <- 100
happiness_data$Happy[happiness_data$Happy == 3] <- 1
happiness_data$Happy[happiness_data$Happy == 100] <- 3

#happiness_data <- happiness_data[,-1]
#happiness_data

pairs(happiness_data[, -c(1,2,3,4,5,6,8,10)], gap=0.4,cex.labels=1.5)

```



```

m1 <- lm(happiness_data$Happy ~ happiness_data$Age + happiness_data$Education + happiness_data$Income + 
summary(m1)

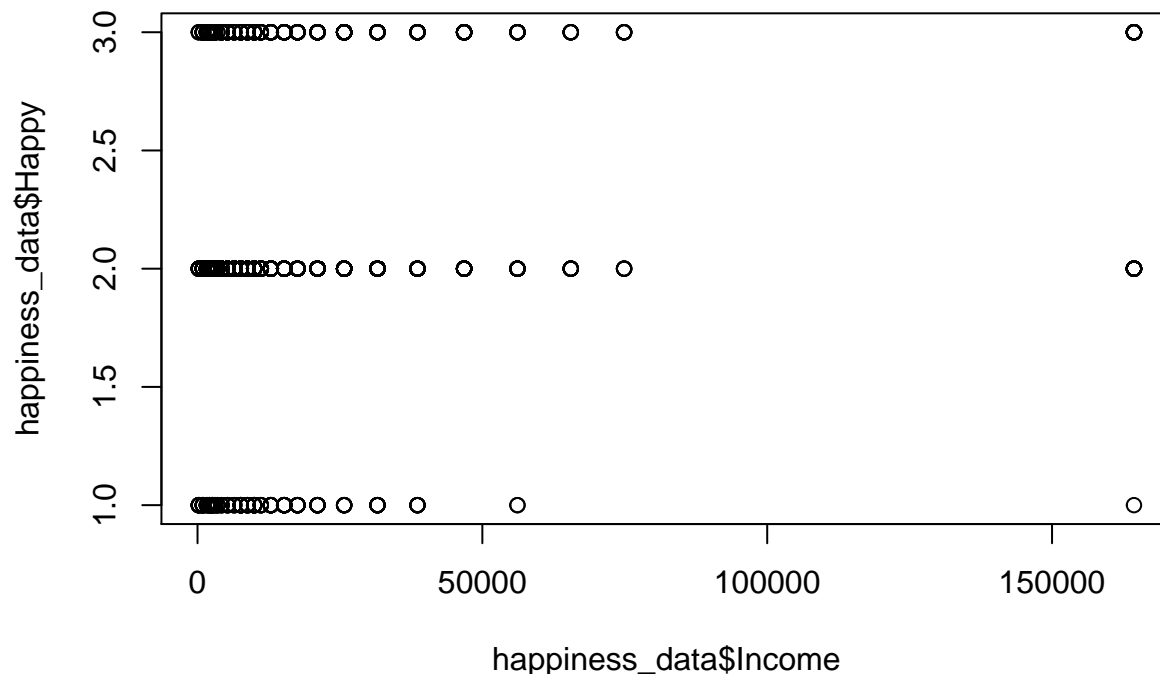
```

```

##
## Call:
## lm(formula = happiness_data$Happy ~ happiness_data$Age + happiness_data$Education + 
##     happiness_data$Income + happiness_data$WorkHrs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5164 -0.2267 -0.1453  0.7486  0.9651
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.915e+00  2.240e-01   8.546 3.04e-16 ***
## happiness_data$Age    2.136e-03  2.470e-03   0.865   0.388
## happiness_data$Education -2.974e-03  1.052e-02  -0.283   0.778
## happiness_data$Income  1.950e-06  1.412e-06   1.381   0.168
## happiness_data$WorkHrs  4.098e-03  2.552e-03   1.606   0.109

```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6347 on 384 degrees of freedom
## (1978 observations deleted due to missingness)
## Multiple R-squared:  0.01724,    Adjusted R-squared:  0.007001
## F-statistic: 1.684 on 4 and 384 DF,  p-value: 0.1529
m2 <- lm(happiness_data$Happy ~ happiness_data$Income)
plot(happiness_data$Happy ~ happiness_data$Income)
```



```
summary(m2)
```

```
##
## Call:
## lm(formula = happiness_data$Happy ~ happiness_data$Income)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6110 -0.1944 -0.1302  0.7579  0.9039
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.095e+00  2.170e-02  96.57  < 2e-16 ***
## happiness_data$Income 3.137e-06  5.852e-07   5.36  9.78e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6137 on 1326 degrees of freedom
## (1039 observations deleted due to missingness)
## Multiple R-squared:  0.02121,    Adjusted R-squared:  0.02047
## F-statistic: 28.73 on 1 and 1326 DF,  p-value: 9.783e-08
```