

# 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
```

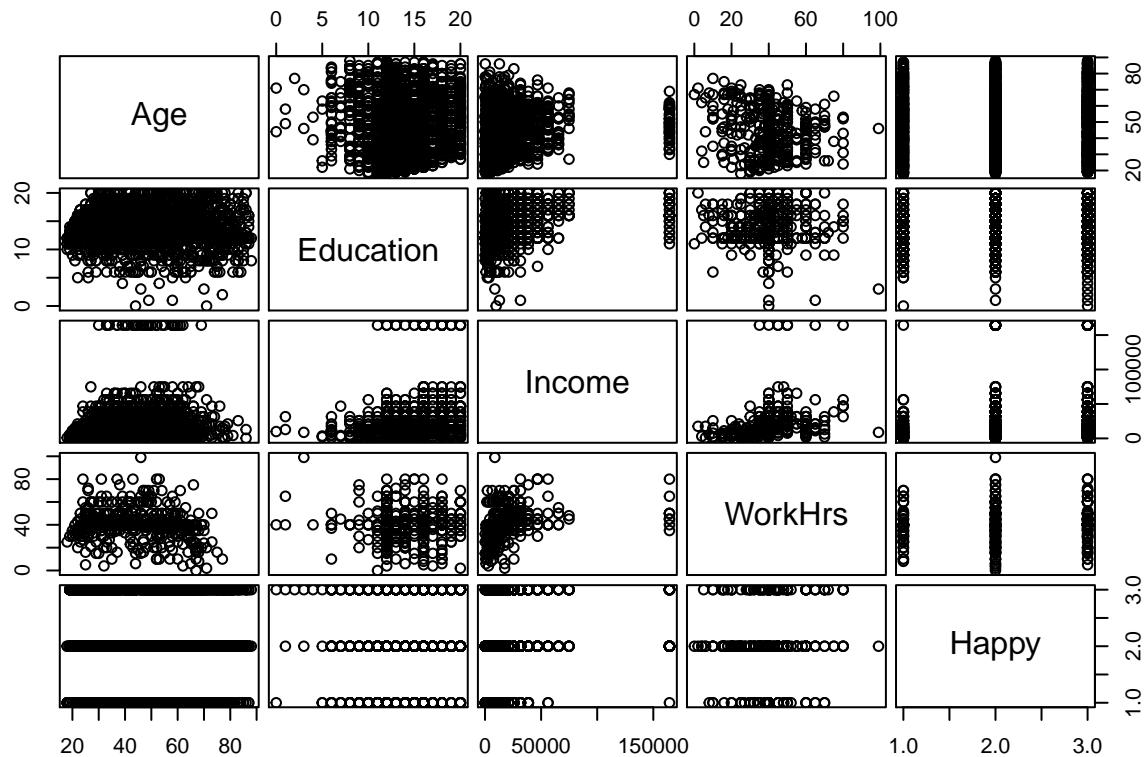
```
happiness_data$Income[happiness_data$Income == 0 | happiness_data$Income == 999998 | happiness_data$Income == 999999] <- NA
```

```
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

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

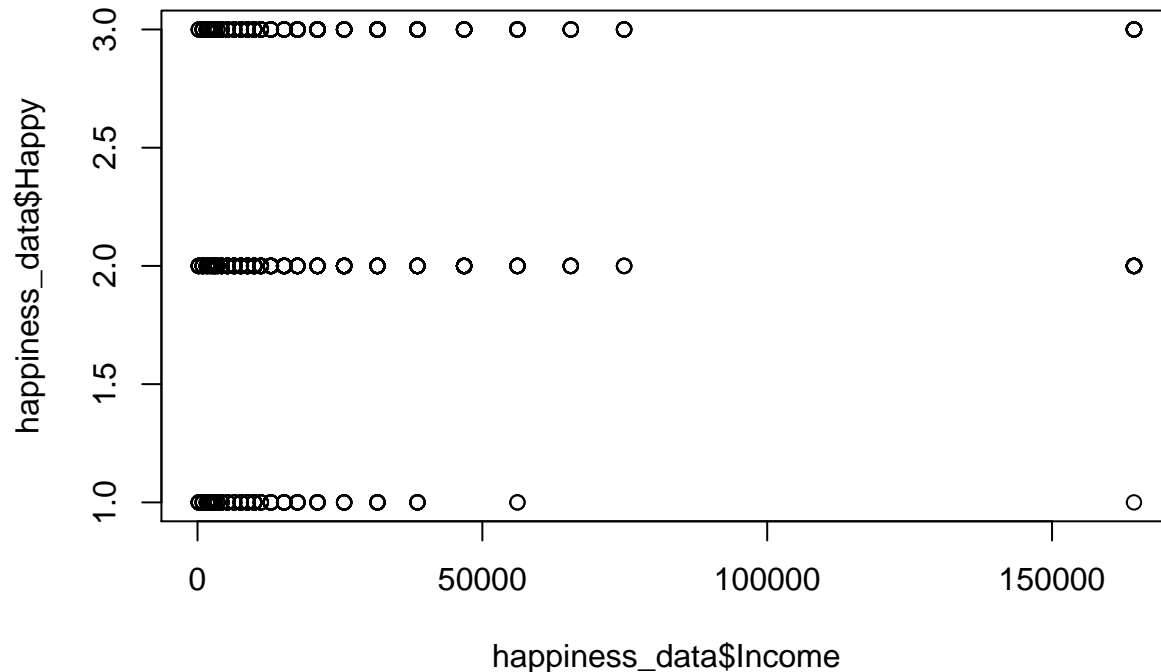


```
m1 <- lm(happiness_data$Happy ~ happiness_data$Age + happiness_data$Education + happiness_data$Income +
happiness_data$WorkHrs)
summary(m1)
```

```
##
## 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
```

```
#adding factors/levels
```

```
HappyCat <- factor(happiness_data$Happy)
levels(HappyCat)
```

```
## [1] "1" "2" "3"

#m4 <- lm(HappyCat ~ happiness_data$Income)
#summary(m4)
#plot(HappyCat ~ happiness_data$Income)

#m5 <- lm(happiness_data$Income ~ HappyCat)
#summary(m5)
#plot(happiness_data$Income ~ HappyCat)

#transforming linear income
mean_income <- mean(happiness_data$Income, na.rm=TRUE)
tIncome <- happiness_data$Income/mean_income
m6 <- lm(happiness_data$Happy ~ tIncome)
summary(m6)

##
## Call:
## lm(formula = happiness_data$Happy ~ tIncome)
##
## 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)  2.09534    0.02170   96.57 < 2e-16 ***
## tIncome      0.07334    0.01368    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

#transforming inverse income
inverse_income <- 1/(happiness_data$Income)
m7 <- lm(happiness_data$Happy ~ inverse_income)
summary(m7)

##
## Call:
## lm(formula = happiness_data$Happy ~ inverse_income)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1918 -0.1889 -0.1821  0.8099  1.3017
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.19252    0.01796 122.051 < 2e-16 ***
## inverse_income -115.64146   29.23653  -3.955 8.05e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.6167 on 1326 degrees of freedom
## (1039 observations deleted due to missingness)
## Multiple R-squared: 0.01166, Adjusted R-squared: 0.01092
## F-statistic: 15.64 on 1 and 1326 DF, p-value: 8.046e-05
```

```
library(alr3)
```

```
## Loading required package: car
```

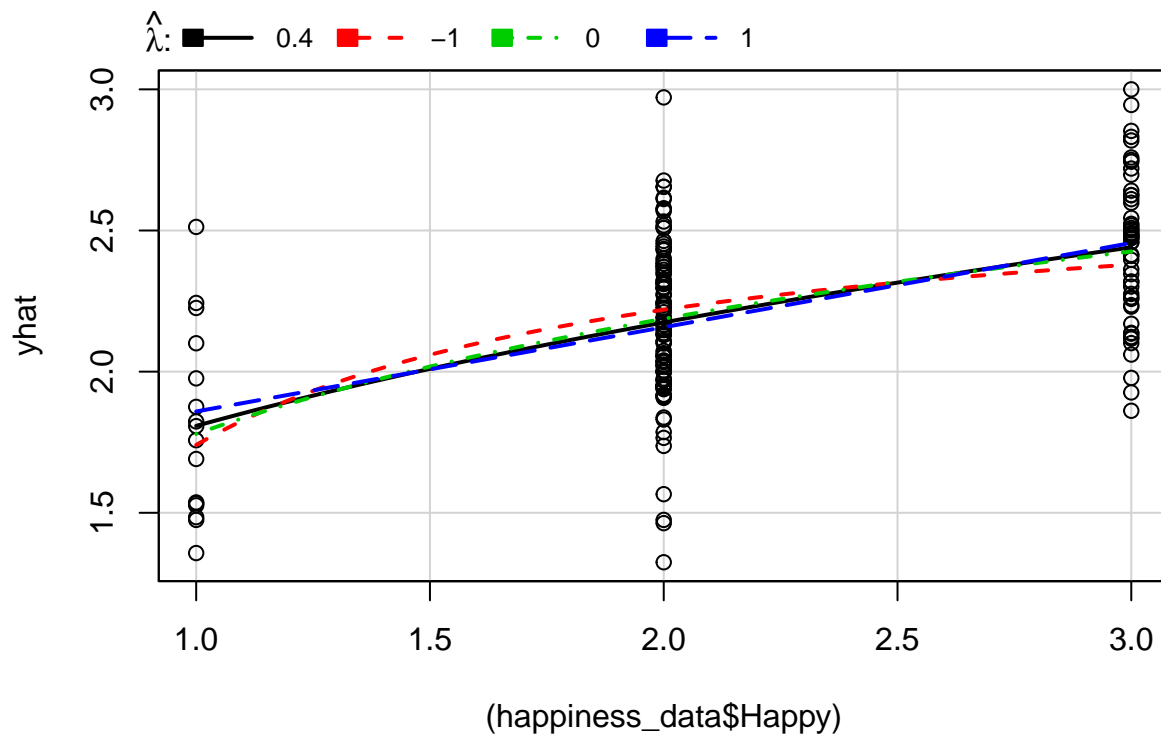
```
## Warning: package 'car' was built under R version 3.4.3
```

```
m1 <- lm((happiness_data$Happy) ~ factor(happiness_data$JobSat) + factor(happiness_data$Household) + fa
inverse.response.plot(m1, key=TRUE)
```

```
## Warning: 'inverse.response.plot' is deprecated.
```

```
## Use 'inverseResponsePlot' instead.
```

```
## See help("Deprecated") and help("alr3-deprecated").
```



```
##      lambda      RSS
## 1  0.396029 12.62671
## 2 -1.000000 13.09307
## 3  0.000000 12.66673
## 4  1.000000 12.70891
```

```
m2 <- lm((happiness_data$Happy)^0.396 ~ factor(happiness_data$JobSat) + factor(happiness_data$Household) +
summary(m2)
```

```
##
```

```
## Call:
```

```
## lm(formula = (happiness_data$Happy)^0.396 ~ factor(happiness_data$JobSat) +
##     factor(happiness_data$Household) + factor(happiness_data$OwnHome) +
##     happiness_data$Instagram + factor(happiness_data$Marital) +
##     happiness_data$Children + happiness_data$Education + factor(happiness_data$JobSat) +
```

```

##      happiness_data$Income)
##
## Residuals:
##      Min        1Q      Median        3Q        Max
## -0.42922 -0.07632 -0.00066  0.10275  0.28342
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   1.437e+00  1.015e-01  14.154 < 2e-16
## factor(happiness_data$JobSat)2 -2.026e-02  3.402e-02  -0.596  0.552445
## factor(happiness_data$JobSat)3 -7.703e-02  3.610e-02  -2.134  0.034568
## factor(happiness_data$JobSat)4 -6.828e-02  6.801e-02  -1.004  0.317088
## factor(happiness_data$JobSat)5 -1.423e-01  5.366e-02  -2.652  0.008917
## factor(happiness_data$JobSat)6 -2.960e-01  7.744e-02  -3.822  0.000197
## factor(happiness_data$JobSat)7 -7.765e-02  1.470e-01  -0.528  0.598234
## factor(happiness_data$Household)2 -4.743e-02  3.287e-02  -1.443  0.151175
## factor(happiness_data$Household)3 -1.058e-02  4.249e-02  -0.249  0.803668
## factor(happiness_data$Household)4 -1.954e-01  8.817e-02  -2.216  0.028286
## factor(happiness_data$Household)5 -8.682e-02  1.472e-01  -0.590  0.556135
## factor(happiness_data$Household)6 -1.814e-01  1.059e-01  -1.712  0.088974
## factor(happiness_data$OwnHome)2  3.689e-02  2.654e-02   1.390  0.166594
## factor(happiness_data$OwnHome)3  8.115e-02  1.483e-01   0.547  0.585169
## happiness_data$Instagram        -2.169e-02  2.714e-02  -0.799  0.425376
## factor(happiness_data$Marital)2 -2.232e-01  5.854e-02  -3.812  0.000204
## factor(happiness_data$Marital)3 -9.841e-02  3.802e-02  -2.588  0.010638
## factor(happiness_data$Marital)4 -1.541e-01  8.703e-02  -1.771  0.078747
## factor(happiness_data$Marital)5 -9.380e-02  3.764e-02  -2.492  0.013842
## happiness_data$Children          -2.539e-03  9.523e-03  -0.267  0.790147
## happiness_data$Education          5.854e-03  4.913e-03   1.191  0.235436
## happiness_data$Income            6.106e-07  4.021e-07   1.519  0.131076
##
## (Intercept)                ***
## factor(happiness_data$JobSat)2
## factor(happiness_data$JobSat)3      *
## factor(happiness_data$JobSat)4
## factor(happiness_data$JobSat)5      **
## factor(happiness_data$JobSat)6      ***
## factor(happiness_data$JobSat)7
## factor(happiness_data$Household)2
## factor(happiness_data$Household)3
## factor(happiness_data$Household)4 *
## factor(happiness_data$Household)5
## factor(happiness_data$Household)6 .
## factor(happiness_data$OwnHome)2
## factor(happiness_data$OwnHome)3
## happiness_data$Instagram
## factor(happiness_data$Marital)2      ***
## factor(happiness_data$Marital)3      *
## factor(happiness_data$Marital)4      .
## factor(happiness_data$Marital)5      *
## happiness_data$Children
## happiness_data$Education
## happiness_data$Income
## ---

```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1406 on 143 degrees of freedom
## (2202 observations deleted due to missingness)
## Multiple R-squared:  0.3064, Adjusted R-squared:  0.2045
## F-statistic: 3.008 on 21 and 143 DF,  p-value: 5.709e-05
```

```
plot(happiness_data$Happy ~ factor(happiness_data$Children))
```

