Developing Data Products - Shiny App Course Project

Emanuele Pellichero 7/29/2016

Child's Adult Height Predictor

1. Introduction

This shiny app was created for the final project of the Coursera Developing Data Products class from Johns Hopkins University. As part of this assignement a web app was created to predict an outcome using a number of predictors from a known dataset.

This Application predicts the adult's height of a child based on the height of the parents and the gender of the child. The predictor function is based on the Galton dataset, which is a dataset that was used by Galton in 1885 to study the correlation between the parent's height and their children.

2. The GaltonFamilies dataset

The Galton dataset contains data from 934 adult children born to 205 fathers and mothers. The dataset has 8 variables,

- *family*: family ID, a factor with levels 001-204
- *father*: height of father
- *mother*: height of mother
- *midparentHeight*: mid-parent height, calculated as (father + 1.08mother)/2
- *children*: number of children in this family
- *childNum*: number of this child within family. Children are listed in decreasing order of height for boys followed by girls
- *child gender*: a factor with levels female male
- *childHeight*: height of child

here below is the database summary:

```
FALSE      family        father          mother      midparentHeight
FALSE  185    : 15   Min.   :62.0   Min.   :58.00   Min.   :64.40
FALSE  066    : 11   1st Qu.:68.0   1st Qu.:63.00   1st Qu.:68.14
FALSE  120    : 11   Median :69.0   Median :64.00   Median :69.25
FALSE  130    : 11   Mean   :69.2   Mean   :64.09   Mean   :69.21
FALSE  166    : 11   3rd Qu.:71.0   3rd Qu.:65.88   3rd Qu.:70.14
FALSE  097    : 10   Max.   :78.5   Max.   :70.50   Max.   :75.43
FALSE  (Other):865
FALSE     children         childNum         gender     childHeight
FALSE  Min.   : 1.000   Min.   : 1.000   female:453   Min.   :56.00
FALSE  1st Qu.: 4.000   1st Qu.: 2.000   male  :481   1st Qu.:64.00
FALSE  Median : 6.000   Median : 3.000                Median :66.50
FALSE  Mean   : 6.171   Mean   : 3.586                Mean   :66.75
FALSE  3rd Qu.: 8.000   3rd Qu.: 5.000                3rd Qu.:69.70
FALSE  Max.   :15.000   Max.   :15.000                Max.   :79.00
FALSE
```

3. The predictor function

The child's adult height was predicted by fitting a linear regression model to the Galton dataset, using *childHeight* as the outcome and *midparentHeight* and *gender* as predictors. Other variables from the dataset were omitted in the linear regression as they appeared to not have a positive effect on the accuracy of the prediction (See Fig.2 below) The regression model is shown below:

```
fit <- lm(childHeight ~ midparentHeight + gender,data=GaltonFamilies)

summary(fit)
```

```
FALSE
FALSE Call:
FALSE lm(formula = childHeight ~ midparentHeight + gender, data = GaltonFamilies)
FALSE
FALSE Residuals:
FALSE     Min      1Q  Median      3Q     Max
FALSE -9.5317 -1.4600  0.0979  1.4566  9.1110
FALSE
FALSE Coefficients:
FALSE                 Estimate Std. Error t value Pr(>|t|)
FALSE (Intercept)     16.51410    2.73392    6.04 2.22e-09 ***
FALSE midparentHeight  0.68702    0.03944   17.42  < 2e-16 ***
FALSE gendermale       5.21511    0.14216   36.69  < 2e-16 ***
FALSE ---
FALSE Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
FALSE
FALSE Residual standard error: 2.17 on 931 degrees of freedom
FALSE Multiple R-squared:  0.6332,  Adjusted R-squared:  0.6324
FALSE F-statistic: 803.6 on 2 and 931 DF,  p-value: < 2.2e-16
```

Figure 1 - Chart with the regression model:

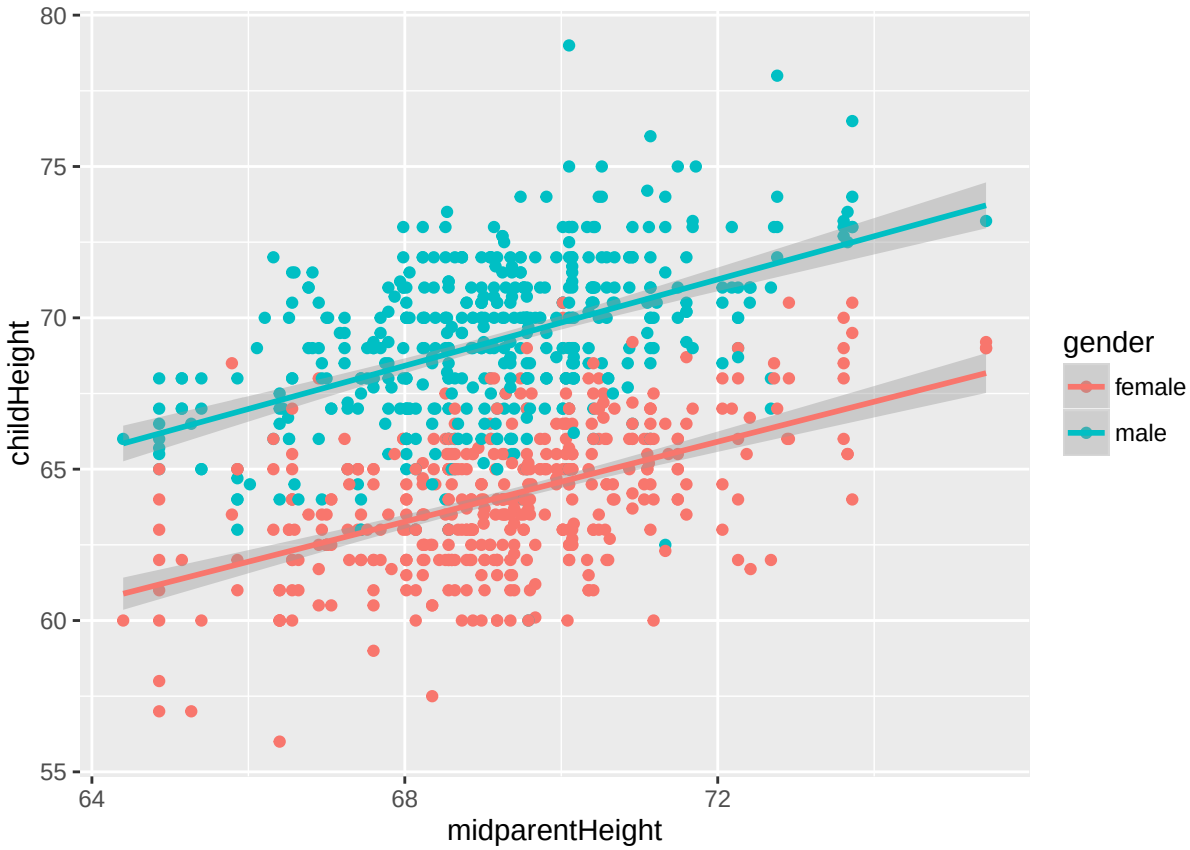
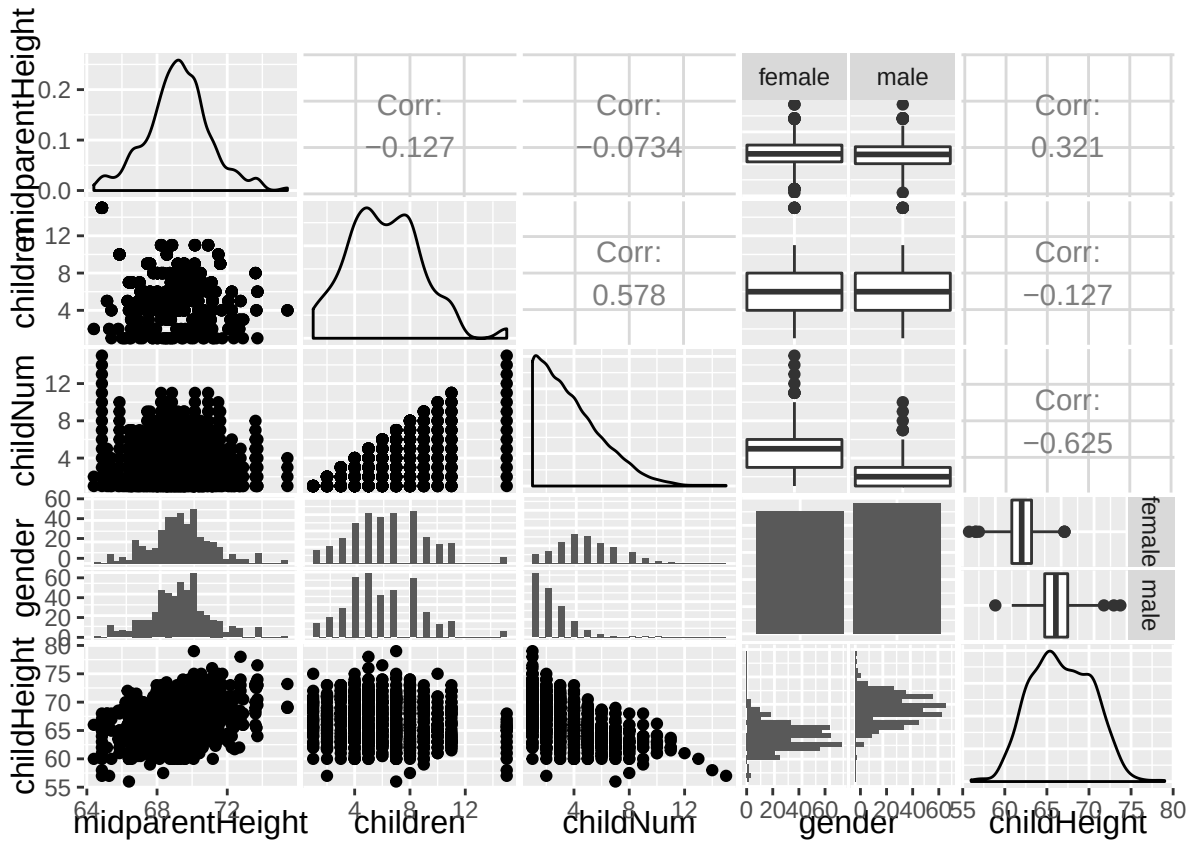


Figure 2 - Pair Wise plot for the variables in the dataset. *family* was omitted as it is just an ID and *father* and *mother* were omitted as they are obviously cooreleted to *midparentHeight*.



4. The shiny app

The shiny app is available here