Net hourly wages across multiple countries Mc Donald's

Code

Nishanth

25 July, 2018, 17:31

1 Importing data into R

1.1 Viewing top rows in CSV

Hide

```
data <- read.csv("data/BigMac-NetHourlyWage.csv")

head(data,5)
```

```
##     Country Big.Mac.Price.... Net.Hourly.Wage....
## 1 Argentina              1.78                 3.3
## 2 Australia              3.84                14.0
## 3    Brazil              4.91                 4.3
## 4   Britain              3.48                13.9
## 5    Canada              4.00                12.8
```

1.2 viewing the structure of the dataset

Hide

```
str(data)
```

```
## 'data.frame':    27 obs. of  3 variables:
##  $ Country           : Factor w/ 27 levels "Argentina","Australia",..: 1 2 3 4 5 6
7 8 9 10 ...
##  $ Big.Mac.Price....  : num  1.78 3.84 4.91 3.48 4 3.34 1.95 3.43 4.9 3.33 ...
##  $ Net.Hourly.Wage....: num  3.3 14 4.3 13.9 12.8 3.1 3 5.1 17.7 3 ...
```

1.3 Basic Summary stats of dataset

Hide

```
summary(data)
```

```
##       Country   Big.Mac.Price.... Net.Hourly.Wage....
##  Argentina: 1   Min.   :1.780     Min.   : 1.300
##  Australia: 1   1st Qu.:2.475     1st Qu.: 3.100
##  Brazil   : 1   Median :3.330     Median : 5.100
##  Britain  : 1   Mean   :3.349     Mean   : 7.726
##  Canada   : 1   3rd Qu.:3.785     3rd Qu.:13.150
##  Chile    : 1   Max.   :6.560     Max.   :22.600
##  (Other)  :21
```

1.4 seperating numerical and categorical variables from the dataset

Hide

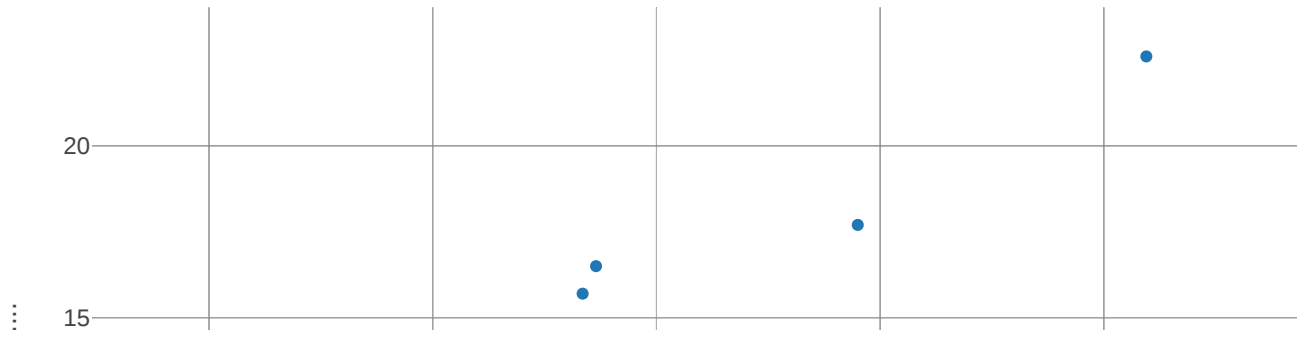
```
variable_types <- sapply(data, is.factor)
numerical_data <- data[,!(variable_types)]
categorical_data <- data[,(variable_types)]
```

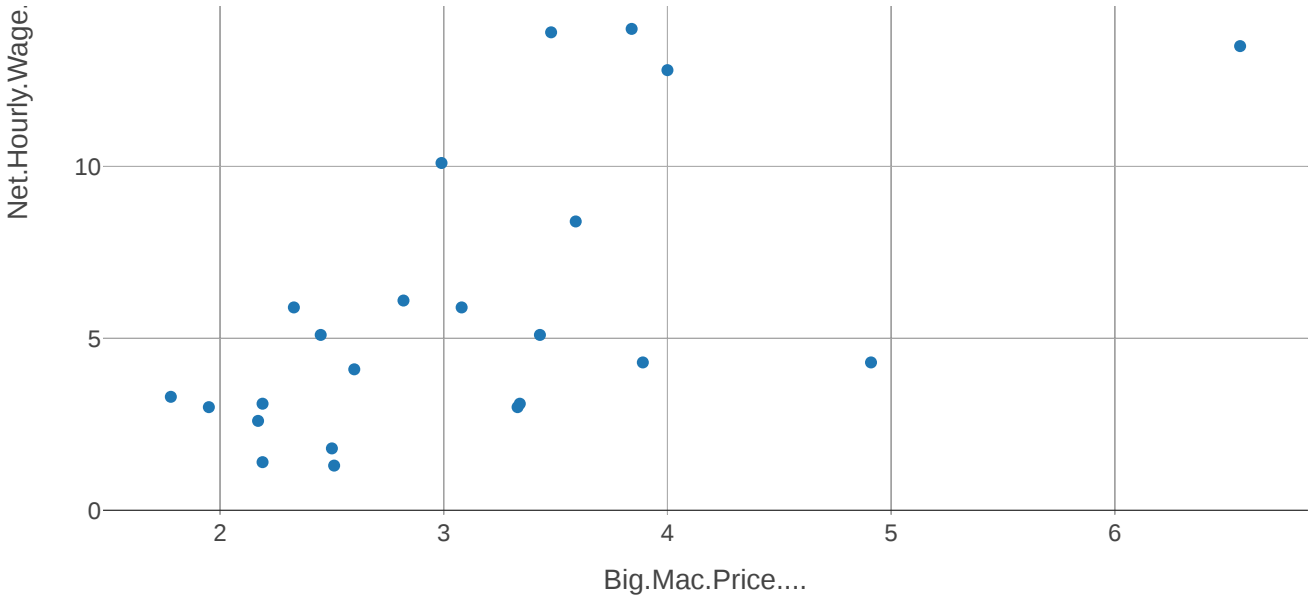
2 Grapical representaion of the data

2.1 scatter plot for net hourly wages and Big Mac price

Hide

```
library(plotly)
p <- plot_ly(data = numerical_data, x = ~ Big.Mac.Price...., y = ~Net.Hourly.Wage....)
p
```



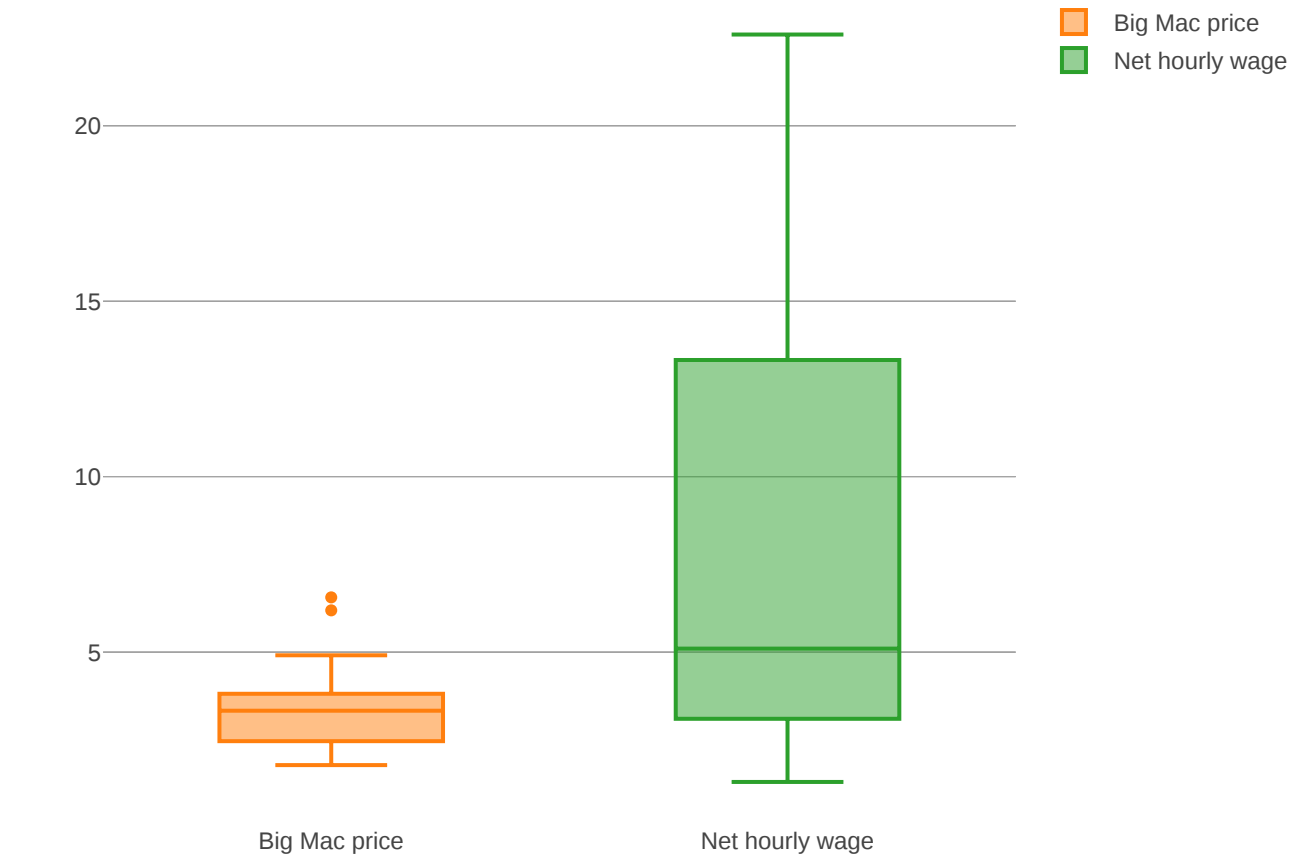


As price of the BIg Mac increases net hourly wages are also increases, there is a postive relationship between Big Mac prices and Net hourly wages

2.2 Box plot for Outlier analysis

Hide

```
p <- plot_ly(type = "box") %>%
  add_boxplot(y = numerical_data$Big.Mac.Price.... , name = "Big Mac price")  %>%
  add_boxplot(y = numerical_data$Net.Hourly.Wage.... , name = "Net hourly wage")
p
```



we see couple of outliers in the data for Big Mac

3 preprocessing

3.1 extracting the outlier points, rows and removing them

Hide

```
outlier_points <- sapply(numerical_data, function(x) boxplot(x,plot=FALSE)$out)
outlier_points
```

```
## $Big.Mac.Price....
## [1] 6.56 6.19
##
## $Net.Hourly.Wage....
## numeric(0)
```

Hide

```
outlier_rows<- which(data$Big.Mac.Price....%in% outlier_points$Big.Mac.Price....)
outlier_rows
```

```
## [1] 22 23
```

Hide

```
data_cleaned <- data[-c(outlier_rows),]
```

There are couple of outliers in Big Mac price and no outliers in net hourly wages

3.2 correlation between BIg Mac and Net Hourly Wage

Hide

```
cor(x = data$Big.Mac.Price....,y = data$Net.Hourly.Wage....)
```

```
## [1] 0.717055
```

correlation between BIg Mac and Net Hourly wage is strong and postively correlated

4 Building the regression model

4.1 applying linear regression to the Mac Donald's data

Hide

```
model.lm <- lm(Net.Hourly.Wage.... ~ Big.Mac.Price.... ,data = data_cleaned)
summary(model.lm)
```

```
##
## Call:
## lm(formula = Net.Hourly.Wage.... ~ Big.Mac.Price...., data = data_cleaned)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.4727 -2.7873 -0.3057  2.4957  7.2248
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -4.9411     3.1612  -1.563 0.131697
## Big.Mac.Price....   3.8114     0.9826   3.879 0.000759 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.107 on 23 degrees of freedom
## Multiple R-squared:  0.3955, Adjusted R-squared:  0.3692
## F-statistic: 15.05 on 1 and 23 DF,  p-value: 0.0007594
```

4.2 viewing diagnostic plots of linear regression

Hide

```
par(mfrow=c(2,2))
plot(model.lm)
```