Data Analysis Assignment 1

STAT 897: Applied Data Mining and Statistical Learning

August, 2017

The aim of this exercise is just to get you familiar with creating .rmd and .pdf documents in R Markdown that you will submit for the data analysis assignments. I recommend you use RStudio for these assignments.

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.

If you are using Rstudio, 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. 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
```

You can also embed plots, for example:
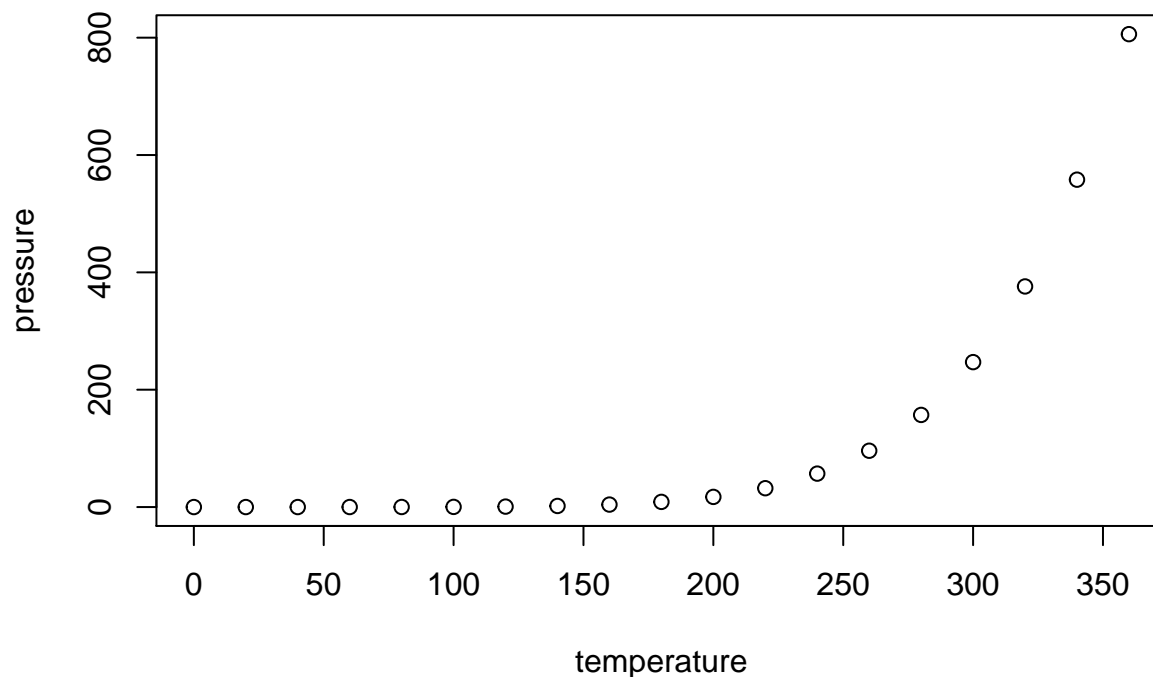
Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Question 1: create a vector and dataframe

Write your own R code to create a vector with entries 1.1, 2.2, 3.3, 4.4.

embed r code here as above

```r
v1 = c(1.1, 2.2, 3.3, 4.4)
v1
```

```
## [1] 1.1 2.2 3.3 4.4
```

Now create an R dataframe with columns size and color, where size is the vector 12, 14, 12, 15 and color is 'red', 'blue', 'green', 'yellow'.

embed r code here

```r
m1 = data.frame(size=c(12, 14, 12, 15), color=c('red', 'blue', 'green', 'yellow'))
m1
```

```
##   size  color
## 1   12    red
## 2   14   blue
## 3   12  green
## 4   15 yellow
```

Question 2: create a function in R

Write R code to create a function called "latz" which takes an integer as input. If the integer is even, it divides the input by two and returns the result; if the integer is odd, it multiplies the input by three, adds one, and returns the result.

embed r code here

```r
latz <- function(x) {
  if (is.even(x)) {
    # return value for even
    return(x/2)
  }
  # else return value for odd
  return(x*3+1)
}

is.even <- function(x) x %% 2 == 0
```

Question 3: call it

Now write a loop that starts with the number 97, and calls your function "latz"" 99 times, *using the output of each call as the input of the next*, and plot the resulting 100 numbers.

embed r code here

```r
latz_out <- numeric(0)
for (pos in seq(from=97,length.out=100)){
```

```
  if(pos == 97) {
    latz_out <- c(latz_out, 97)
  }
  else {
    latz_out <- c(latz_out, latz(tail(latz_out, n=1)))
  }
}
```

Compute the mean and variance of these 100 numbers using the built-in R functions.

embed r code here

```
print(paste('Mean: ', mean(latz_out)))
```

```
## [1] "Mean:  1022.99"
```

```
print(paste('Variance: ', var(latz_out)))
```

```
## [1] "Variance:  2367794.43424242"
```

Finally submit BOTH your .rmd file and the resulting .pdf file with Canvas as Data Analysis Assignment 1.