

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.

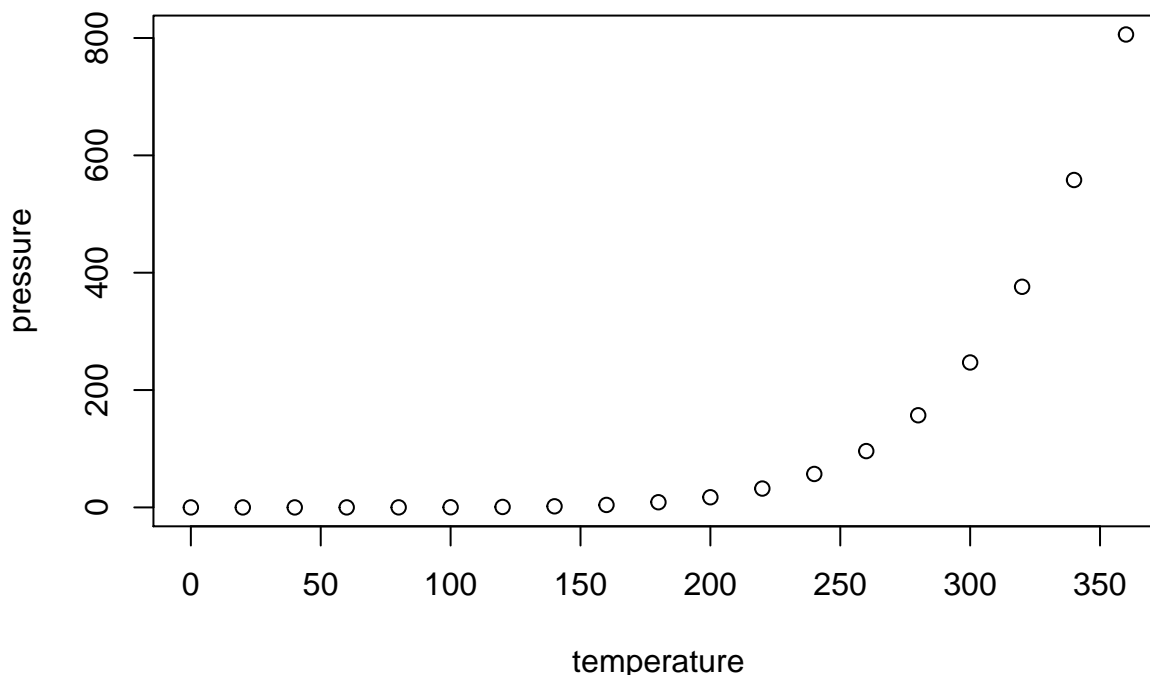
This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

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.

```
test_vector = c(1.1, 2.2, 3.3, 4.4)
test_vector
```

```
## [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'.

```
test_df = data.frame(size = c(12, 14, 12, 15), color = c("red", "blue", "green", "yellow"))
test_df
```

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

Question 2: create a function in R

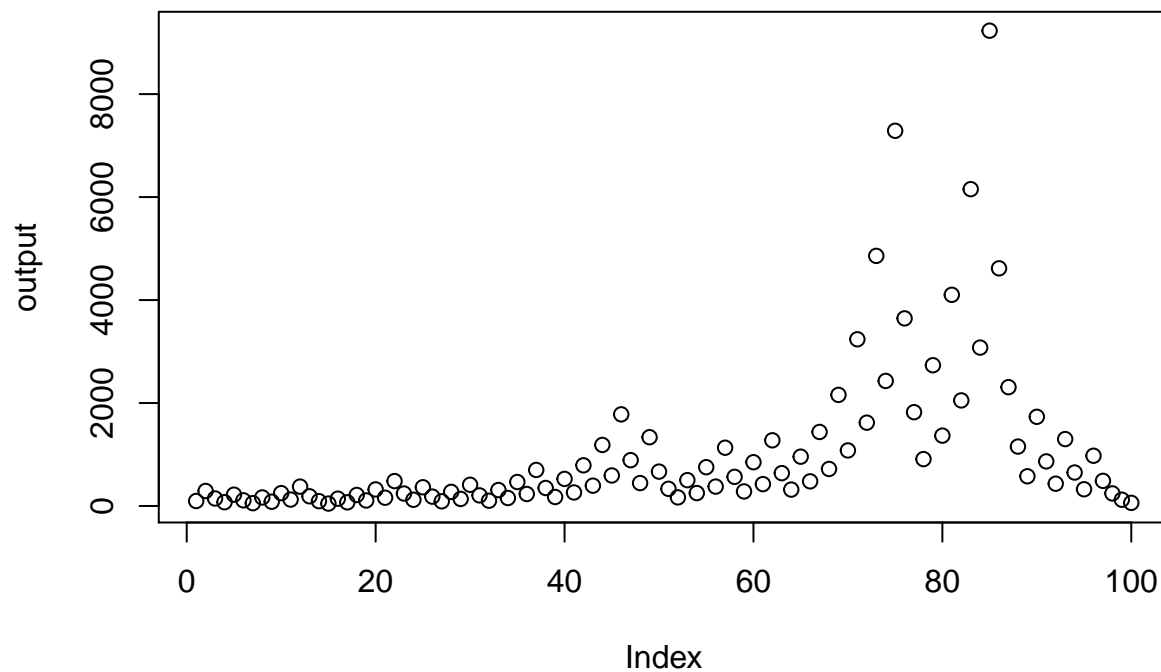
Write R code to create a function called "latz" which takes an integer as input. If the integer is even, it divides the input by two and returns the result; if the integer is odd, it multiplies the input by three, adds one, and returns the result.

```
latz = function(input){
  if(input %% 2 == 0){
    output = input / 2
  } else{
    output = input * 3 + 1
  }
  return(output)
}
```

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.

```
output = c(97, rep(NA, 99))
for(a in 1:99){
  output[a + 1] = latz(output[a])
}
plot(output)
```



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

```
mean(output)
```

```
## [1] 1022.99
```

```
var(output)
```

```
## [1] 2367794
```

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