Assignment #0 – Example Submission

Charles Nicholson

Tuesday, August 11, 2015

A good submission is clean, structured, well-documented, and self-contained.

This means:

1. heading with name, date, assignment number
2. separate problem sections
3. proliferate use of meaingful comments
4. load packages in code, i.e. using "library"
5. the script should be compilable (everything necessary to run it is included in code)

```
#Assignment #0
#Name: Charles Nicholson
#ISE 5103 Intelligent Data Analytics
#Date: mm/dd/yyyy

#required packages for this assignment
library(ggplot2)  #provides advanced graphics
```

```
## Warning: package 'ggplot2' was built under R version 3.0.3
```

Problem #1

```
#Problem 1: create a vector of strings and demonstrate access

#Problem 1(a)
x<-c('R','is','a','powerful','open source','statistical','scripting language')

x[1]    #access first element
```

```
## [1] "R"
```

```
x[4:6]  #access third thru fifth elements
```

```
## [1] "powerful"    "open source" "statistical"
```

```
#Problem 1(b)
length(x)  #determine length of x
```

```
## [1] 7
```

```
#Problem 1(c)
#create a new vector "z" from x, but insert "my favorite"
z<-c(x[1:2],"my favorite",x[4:length(x)])
z  #view the results
```

```
## [1] "R"                    "is"                   "my favorite"
## [4] "powerful"             "open source"          "statistical"
## [7] "scripting language"
```

Problem #2

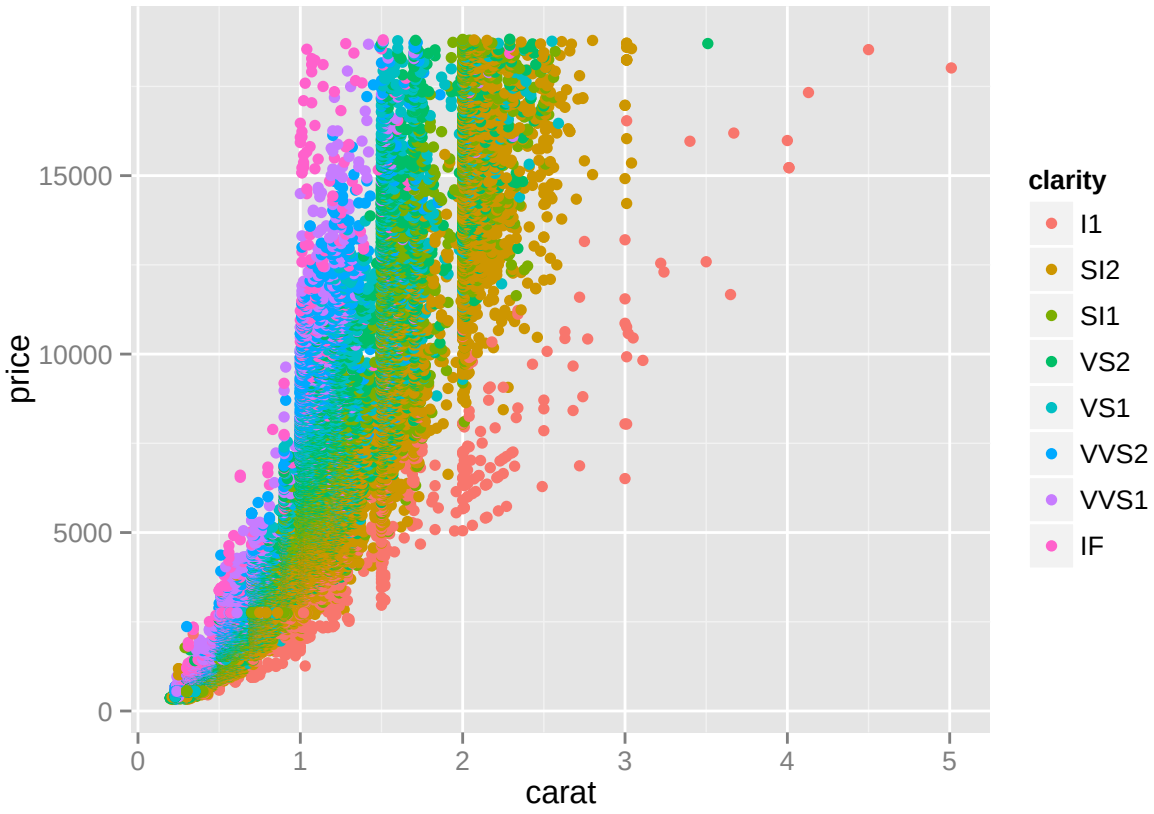
```
#Problem 2   --------------------------------------------------------------------

#the data and the command `qplot' are from the package `ggplot2'

data(diamonds)  #data frame from ggplot2
head(diamonds)  #examine first few rows
```

```
##   carat       cut color clarity depth table price    x    y    z
## 1  0.23     Ideal     E     SI2  61.5    55   326 3.95 3.98 2.43
## 2  0.21   Premium     E     SI1  59.8    61   326 3.89 3.84 2.31
## 3  0.23      Good     E     VS1  56.9    65   327 4.05 4.07 2.31
## 4  0.29   Premium     I     VS2  62.4    58   334 4.20 4.23 2.63
## 5  0.31      Good     J     SI2  63.3    58   335 4.34 4.35 2.75
## 6  0.24 Very Good     J    VVS2  62.8    57   336 3.94 3.96 2.48
```

```
#scatterplot of diamond carat and price; clarity differentiated by color
qplot(carat, price, data = diamonds, colour = clarity)
```



The figure depicts a definite relationship between price, carat size, and clarity. There seem to be a definite pattern in the frequency of integer caret sizes – for example, at 1, 2, and 3 caret values the number of observations increase. This is a reasonable hypothesis given that certain carat sizes are more marketable.

The following detailed histogram of carat sizes shows that in fact this is the case. However, the increased frequencies seem to fall around carats of 0.25, 0.5, 0.75, 1, 1.5, and 2.

```
hist(diamonds$carat,breaks=50)
```