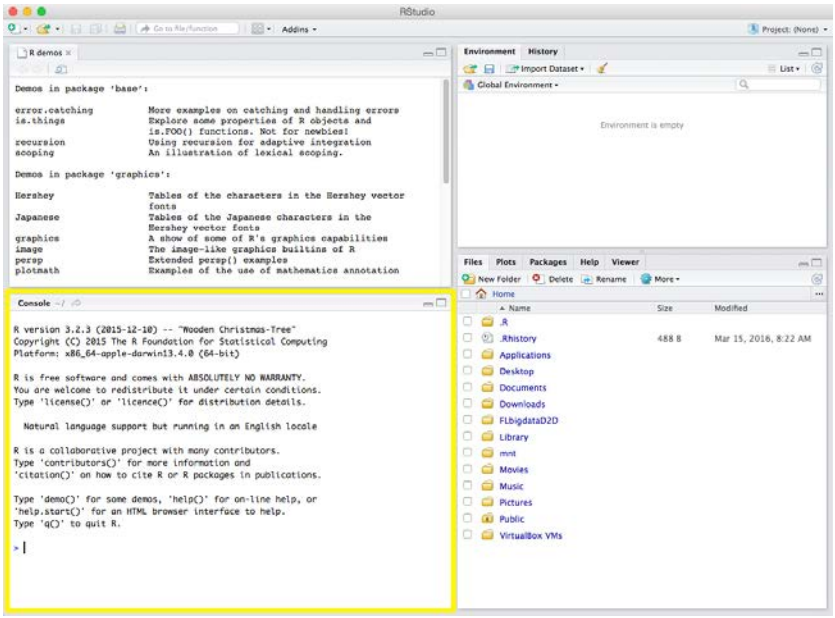
Instructions: Grouping customers by responsiveness to telemarketing offer

In this exercise you will use RStudio and H2O to explore our bank customer dataset.

1) Open RStudio. Enter each command in the steps below, one line at a time, into your RStudio console.



2) Load the R packages. Each time we open RStudio we need to load our packages for that session.

```
library(h2o)
```

3) Start the H2O server locally.

```
localH2o = h2o.init(ip = "127.0.0.1", port = 54321)
```

4) Import the dataset.

   a) Load the file path into a variable.

```
filePath = "~/FLbigdataStats/bank_customer_data.csv"
```


b) Load the dataset and save it to the local handle 'market_data'.

```
market_data <- h2o.uploadFile(filePath,
                              destination_frame = "",
                              parse = T,
                              header = T,
                              sep = ",",
                              na.strings = c("unknown"),
                              progressBar = FALSE,
                              parse_type = "CSV")
```

5) Create a new dataset that excludes the response variable and the 11th feature.

```
cluster_data <- market_data[,-11]
```

Check there are 19 variables in the dataset.

6) Fit 3 clusters to these data using the k-means algorithm in R/H2O.

```
cluster_model <- h2o.kmeans(training_frame =  cluster_data,
                            k = 3,
                            standardize = T,
                            init = "Random"
                            )
```

7) Examine the model.

```
summary(cluster_model)
```

8) Repeat steps 6 and 7 using 2, 5 and 10 clusters.

9) Examine the cluster means in H2O Flow and consider:

   c) How many clusters do you think best fits these data?

   d) What are the major characteristics of each cluster?

   e) What are the major differences between the clusters?

