fourth_process

Michael Pearson

12/21/2017

R Markdown

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.

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 within the document. You can embed an R code chunk like this:

```
combi_bi <-  read.csv(file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capstone
combi_bi <- data.table(combi_bi)
print(prettyNum(object.size(combi_bi), big.mark = ",", scientific = FALSE))
```

```
## [1] "94,056,056"
```

```
combi_bi <- setorder(setDT(combi_bi), word1, -n) [,indx :=seq_len(.N), by = word1] [indx <= 5 & n > 2]
write.csv(combi_bi, file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capstone Pr
rm(combi_bi)
sen_combi_bi <-  read.csv(file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capst
sen_combi_bi <- data.table(sen_combi_bi)
print(prettyNum(object.size(sen_combi_bi), big.mark = ",", scientific = FALSE))
```

```
## [1] "83,219,840"
```

```
sen_combi_bi <- setorder(setDT(sen_combi_bi), word1, -n) [,indx :=seq_len(.N), by = word1] [indx <= 5 &
write.csv(sen_combi_bi, file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capstone
rm(sen_combi_bi)
```

trigrams now

```
combi_tri <-  read.csv(file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capstone
combi_tri <- data.table(combi_tri)
print(prettyNum(object.size(combi_tri), big.mark = ",", scientific = FALSE))
```

```
## [1] "506,272,640"
```

```
combi_tri <- setorder(setDT(combi_tri), bigrams, -n) [,indx :=seq_len(.N), by = bigrams] [indx <= 5 & n
setkey(combi_tri, bigrams)
print(prettyNum(object.size(combi_tri), big.mark = ",", scientific = FALSE))
```

```
## [1] "353,878,128"
```

```
write.csv(combi_tri, file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capstone Pr
rm(combi_tri)
sen_combi_tri <-  read.csv(file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capst
sen_combi_tri <- data.table(sen_combi_tri)
print(prettyNum(object.size(sen_combi_tri), big.mark = ",", scientific = FALSE))
```

```
## [1] "471,566,328"
```

```
sen_combi_tri <- setorder(setDT(sen_combi_tri), bigrams, -n) [,indx :=seq_len(.N), by = bigrams] [indx
setkey(sen_combi_tri, bigrams)
print(prettyNum(object.size(sen_combi_tri), big.mark = ",", scientific = FALSE))
```

```
## [1] "325,917,216"
```

```
write.csv(sen_combi_tri, file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capsto
rm(sen_combi_tri)
```

quads

```
combi_quad <-  read.csv(file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capstone
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote =
## quote, : cols = 4 != length(data) = 5
```

```
combi_quad <- data.table(combi_quad)
print(prettyNum(object.size(combi_quad), big.mark = ",", scientific = FALSE))
```

```
## [1] "1,203,252,560"
```

```
combi_quad <- setorder(setDT(combi_quad), trigrams, -n) [,indx :=seq_len(.N), by = trigrams] [indx <= 5
setkey(combi_quad, trigrams)
print(prettyNum(object.size(combi_quad), big.mark = ",", scientific = FALSE))
```

```
## [1] "988,690,136"
```

```
write.csv(combi_quad, file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capstone
rm(combi_quad)
```

sen things with quads

```
combi_sen_quad <-  read.csv(file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Cap
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote =
## quote, : cols = 4 != length(data) = 5
```

```
combi_sen_quad <- data.table(combi_sen_quad)
print(prettyNum(object.size(combi_sen_quad), big.mark = ",", scientific = FALSE))
```

```
## [1] "1,154,016,296"
```

```
combi_sen_quad <- setorder(setDT(combi_sen_quad), trigrams, -n) [,indx :=seq_len(.N), by = trigrams] [i
setkey(combi_sen_quad, trigrams)
print(prettyNum(object.size(combi_sen_quad), big.mark = ",", scientific = FALSE))
```

```
## [1] "943,693,768"
```

```
write.csv(combi_sen_quad, file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capst
rm(combi_sen_quad)
```

quins

```
combi_quin <-  read.csv(file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capstone
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote =
## quote, : cols = 4 != length(data) = 6
```

```
combi_quin <- data.table(combi_quin)
print(prettyNum(object.size(combi_quin), big.mark = ",", scientific = FALSE))
```

```
## [1] "1,696,206,480"
```

```
combi_quin <- setorder(setDT(combi_quin), quadgrams, -n) [,indx :=seq_len(.N), by = quadgrams] [indx <=
setkey(combi_quin, quadgrams)
print(prettyNum(object.size(combi_quin), big.mark = ",", scientific = FALSE))
```

```
## [1] "1,769,153,064"
```

```
write.csv(combi_quin, file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capstone
rm(combi_quin)
combi_sen_quin <-  read.csv(file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Caps
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote =
## quote, : cols = 4 != length(data) = 6
```

```
combi_sen_quin <- data.table(combi_sen_quin)
print(prettyNum(object.size(combi_sen_quin), big.mark = ",", scientific = FALSE))
```

```
## [1] "3,202,375,640"
```

```
combi_sen_quin <- setorder(setDT(combi_sen_quin), quadgrams, -n) [,indx :=seq_len(.N), by = quadgrams]
setkey(combi_sen_quin, quadgrams)
print(prettyNum(object.size(combi_sen_quin), big.mark = ",", scientific = FALSE))
```

```
## [1] "3,335,889,608"
```

```
write.csv(combi_sen_quin, file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capst
rm(combi_sen_quin)
```

stuff here

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.