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)
```

head ## 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] "562,279,256"
```

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

```
## [1] "613,475,184"
```

```
write.csv(combi_tri, file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capstone Pr
rm(combi_tri)
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] "685,925,800"
```

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

```
## [1] "738,071,984"
```

```
write.csv(combi_tri, file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capstone Pr
rm(combi_tri)
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] "491,684,000"
```

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

```
## [1] "550,957,496"
```

```
write.csv(combi_tri, file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capstone Pr
rm(combi_tri)
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] "580,366,688"
```

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

```
## [1] "642,861,416"
```

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

quads

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

```
## [1] "1,138,489,480"
```

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

```
## [1] "1,193,230,568"
```

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

```
## [1] "1,191,585,912"
```

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

```
## [1] "1,246,405,152"
```

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

```
## [1] "1,322,382,928"
```

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

```
## [1] "1,409,559,064"
```

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

```
## [1] "1,429,409,216"
```

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

```
## [1] "1,517,629,136"
```

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

sen things with quads

quins

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

```
## [1] "1,276,509,416"
```

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

```
## [1] "1,331,632,608"
```

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

```
## [1] "1,311,080,808"
```

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

```
## [1] "1,366,217,640"
```

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

```
## [1] "2,008,492,416"
```

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

```
## [1] "2,105,312,928"
```

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

```
## [1] "2,055,939,128"
```

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

```
## [1] "2,153,011,528"
```

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

```
## Warning in rm(combi_tri): object 'combi_tri' not found
```

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.