r_assignment_markdown

Tom McBrien

October 14, 2015

Loading and Merging Dataa

```r
d <- read.csv("Dataset_S1.txt", sep = ',', header = TRUE) #reading in Data
head(d)
```

```
##   start    end total.SNPs total.Bases depth unique.SNPs dhSNPs
## 1 55001 56000          0        1894  3.41           0      0
## 2 56001 57000          5        6683  6.68           2      2
## 3 57001 58000          1        9063  9.06           1      0
## 4 58001 59000          7       10256 10.26           3      2
## 5 59001 60000          4        8057  8.06           4      0
## 6 60001 61000          6        7051  7.05           2      1
##   reference.Bases  Theta     Pi Heterozygosity    X.GC Recombination
## 1             556  0.000  0.000          0.000 54.8096   0.009601574
## 2            1000  8.007 10.354          7.481 42.4424   0.009601574
## 3            1000  3.510  1.986          1.103 37.2372   0.009601574
## 4            1000  9.929  9.556          6.582 38.0380   0.009601574
## 5            1000 12.915  8.506          4.965 41.3413   0.009601574
## 6            1000  7.817  9.121          8.864 36.1361   0.009601574
##     Divergence Constraint SNPs
## 1 0.003006012          0    0
## 2 0.018018020          0    0
## 3 0.007007007          0    0
## 4 0.012012010          0    0
## 5 0.024024020          0    0
## 6 0.016016020          0    0
```

```r
rcmb <- read.delim("motif_recombrates.txt", header = TRUE) #read in motif recomb rates data
rpts <- read.delim("motif_repeats.txt", header = TRUE) #read in motif repeat rates data
rcmb$pos <- paste(rcmb$chr, rcmb$motif_start, sep="-") #making column of specific positions per chromos
rpts$pos <- paste(rpts$chr, rpts$motif_start, sep="-") #same as above with repeats file
joined <- merge(rcmb, rpts, by.x="pos", by.y="pos") #mergin
head(joined)
```

```
##               pos chr.x motif_start.x motif_end    dist recomb_start
## 1 chr1-101890123  chr1     101890123 101890136 34154.0    101855215
## 2 chr1-101890123  chr1     101890123 101890136 35717.5    101853608
## 3 chr1-101890123  chr1     101890123 101890136  9704.0    101878637
## 4 chr1-101890123  chr1     101890123 101890136  7864.5    101882213
## 5 chr1-101890123  chr1     101890123 101890136 29463.0    101859577
## 6 chr1-101890123  chr1     101890123 101890136 37189.5    101852271
##   recomb_end   recom          motif chr.y     start       end  name
## 1  101856736 0.0700 CCTCCCTAGCCAC   chr1 101890032 101890381 THE1B
## 2  101855216 0.0722 CCTCCCTAGCCAC   chr1 101890032 101890381 THE1B
## 3  101882214 0.2445 CCTCCCTAGCCAC   chr1 101890032 101890381 THE1B
```

```
## 4  101882317 0.2445 CCTCCCTAGCCAC  chr1 101890032 101890381 THE1B
## 5  101861756 0.0691 CCTCCCTAGCCAC  chr1 101890032 101890381 THE1B
## 6  101853609 0.4441 CCTCCCTAGCCAC  chr1 101890032 101890381 THE1B
##   motif_start.y
## 1     101890123
## 2     101890123
## 3     101890123
## 4     101890123
## 5     101890123
## 6     101890123
```
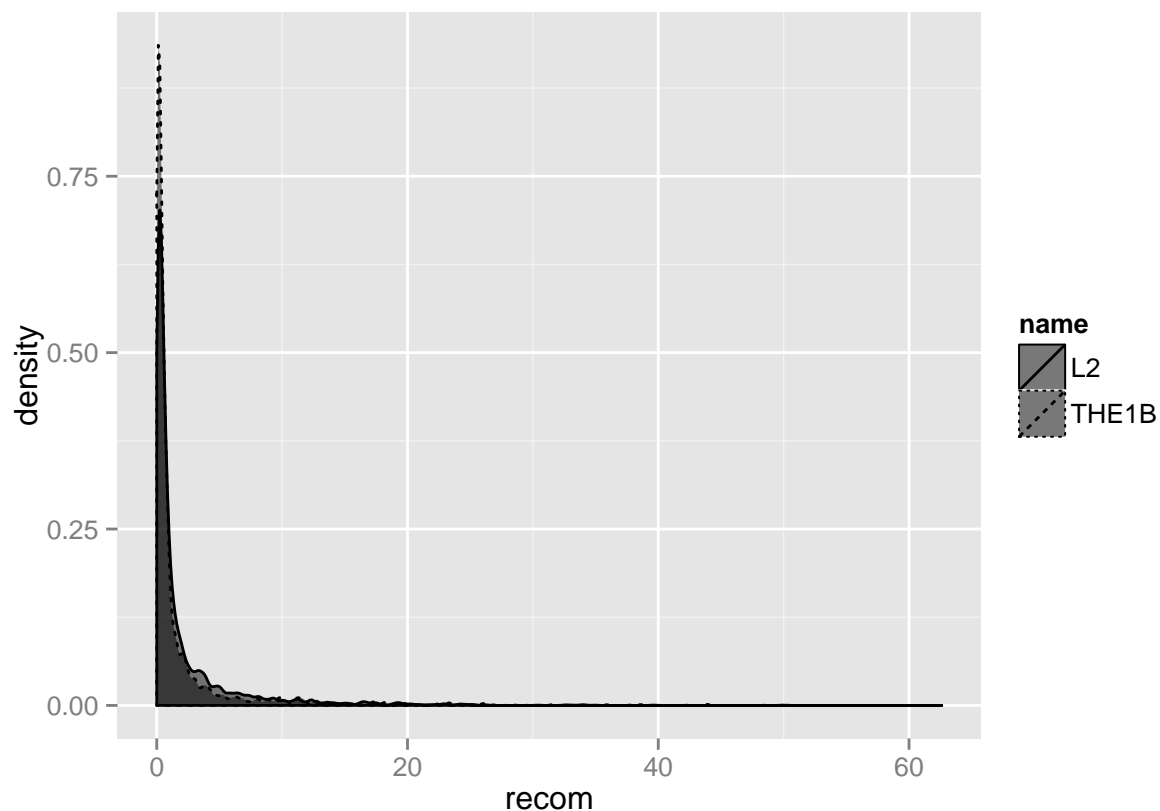
Analysing Data

```r
aggregate(joined$recom, list(motif=joined$motif), mean) #this uses the aggregate function to give two s
```

```
##          motif        x
## 1 CCTCCCTAGCCAC 1.963472
## 2 CCTCCCTGACCAC 2.138344
```

Analyzing if Distributions of Recombination Rate Differs by Motif Type

```r
library(ggplot2)
ggplot(joined) + geom_density(aes(x=recom, linetype=name), fill='black', alpha=0.5)
```

```
#This didn't show much so I will try to bin.#
joined$recom.binned <- cut(joined$recom, 5)
ggplot(joined) + geom_density(aes(x=recom.binned, linetype=name), fill='black', alpha=0.5)
```