Compute pairwise distances

Deal with 0 counts

Similarity of allele frequency with distance?

quick look at CCR5 data

Matthew Stephens

2016-02-25

Last updated: 2017-03-06

Code version: 80315be

Load Data

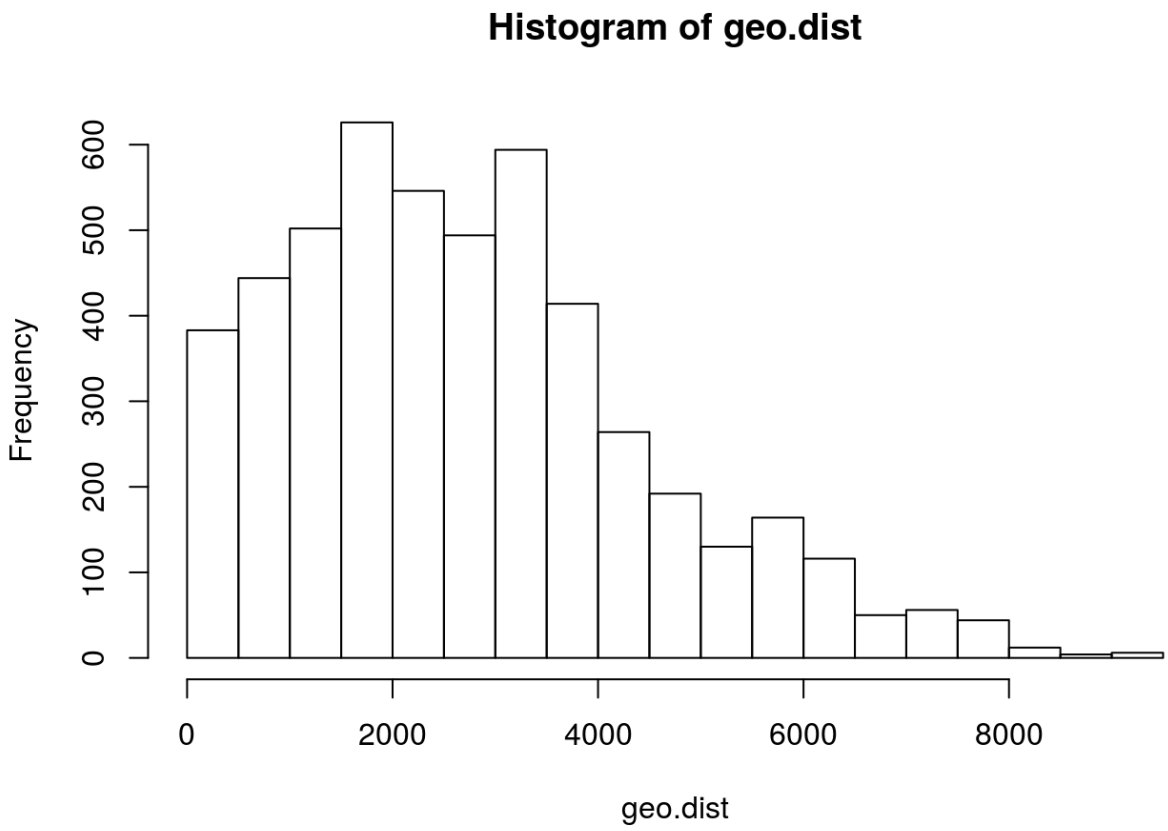
The following reads in the data, and converts longitudes >180 to corresponding negative longitudes (this avoids warnings in geosphere package later)

```
ccr5 = read.table("../data/CCR5/CCR5.freq.txt",header=TRUE)
ccr5[,1] = ifelse(ccr5[,1]>180,ccr5[,1]-360,ccr5[,1]) # changes longitudes>180 to negative
```

Compute pairwise distances

This can be done using the geosphere package. Dividing by 1000 gives distance in km.

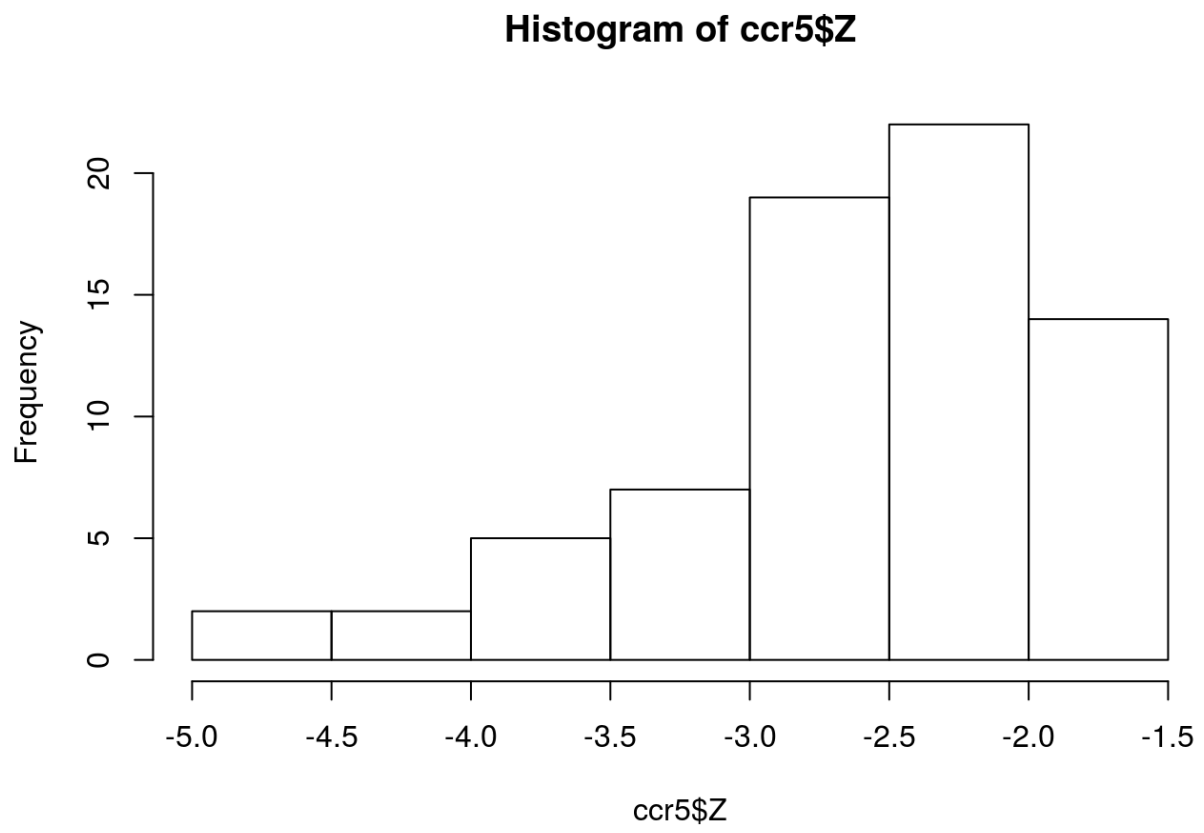
```
geo.dist = geosphere::distm(ccr5[,1:2])/1000
hist(geo.dist)
```



Deal with 0 counts

Some of the frequency estimates are 0. I deal with this by first working out the original counts (frequency * samplesize *2), and then adding a pseudocount to each allele before recomputing frequencies. The resulting column "fhat" can be transformed by log(fhat/(1-fhat)) to be something that is not entirely non-normal...

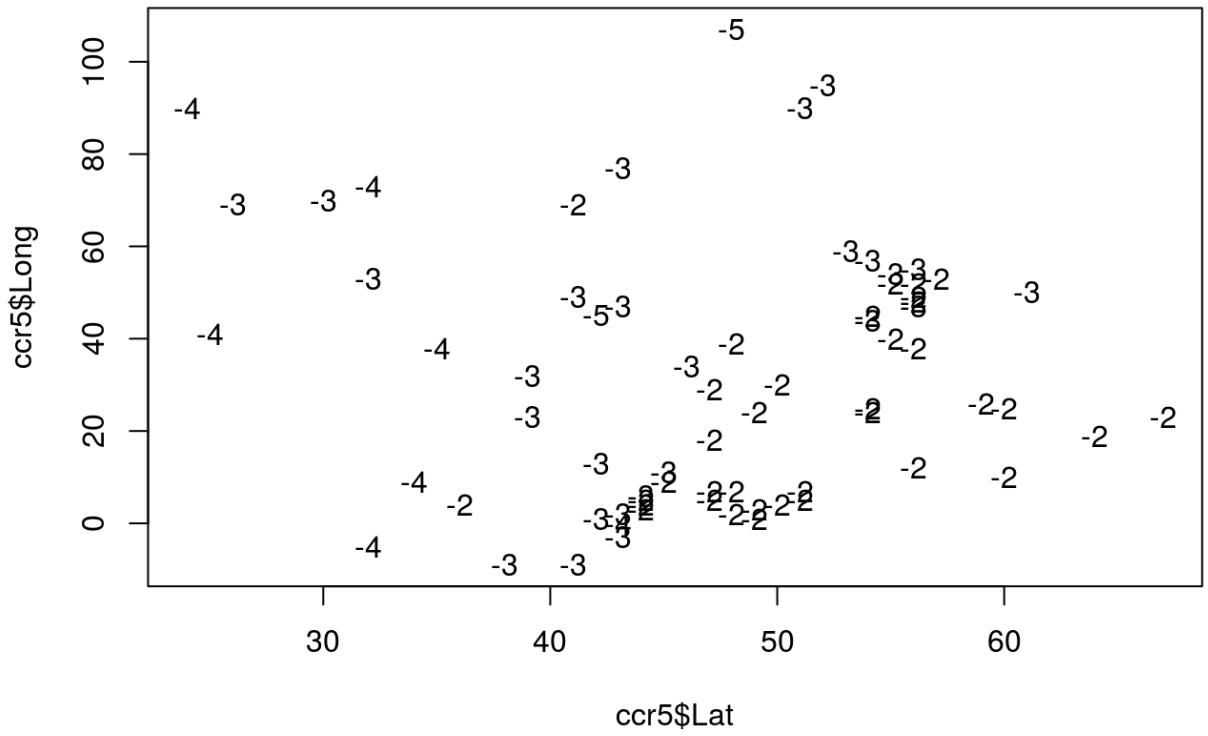
```
ccr5$count = round(ccr5$Freq* ccr5$SampleSize * 2)
ccr5$fhat = (ccr5$count+1)/(ccr5$SampleSize*2+2)
ccr5$Z = log(ccr5$fhat/(1-ccr5$fhat))
hist(ccr5$Z)
```



Similarity of allele frequency with distance?

Here we take a quick look to check that Z has some kind of spatial trend.

```
plot(x=ccr5$Lat,y=ccr5$Long,type="n")
text(round(ccr5$Z),x=ccr5$Lat,y=ccr5$Long)
```



Session information

```
sessionInfo()
```

```
R version 3.3.2 (2016-10-31)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 14.04.5 LTS

locale:
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
 [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
 [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
 [9] LC_ADDRESS=C               LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] workflowr_0.4.0    rmarkdown_1.3.9004

loaded via a namespace (and not attached):
 [1] Rcpp_0.12.9     lattice_0.20-34 digest_0.6.12   rprojroot_1.2
 [5] grid_3.3.2      backports_1.0.5 git2r_0.18.0    magrittr_1.5
 [9] evaluate_0.10   stringi_1.1.2   geosphere_1.5-5 rstudioapi_0.6
[13] sp_1.2-4        whisker_0.3-2   tools_3.3.2     stringr_1.2.0
[17] yaml_2.1.14     htmltools_0.3.5 knitr_1.15.1
```

This R Markdown (http://rmarkdown.rstudio.com) site was created with workflowr (https://github.com/jdblischak/workflowr)
