Covid-19 SNP analysis

Code

Xiaomi Liu

Hide

```
library(dplyr)
```

```
Warning: package 'dplyr' was built under R version 4.3.2
```

Hide

```
library(ggplot2)
library(adegenet)
```

```
Warning: package 'adegenet' was built under R version 4.3.3Warning: package 'ade4' was built und
er R version 4.3.3
```

Hide

```
library(LEA)
library(dartR)
```

```
Warning: package 'dartR' was built under R version 4.3.3Warning: package 'dartR.data' was built
under R version 4.3.3
```

Hide

```
library(gtools)
```

```
Warning: package 'gtools' was built under R version 4.3.3
```

Hide

```
library(tidyverse)
```

```
Warning: package 'tidyverse' was built under R version 4.3.3Warning: package 'tibble' was built
under R version 4.3.2Warning: package 'tidyr' was built under R version 4.3.3Warning: package 'r
eadr' was built under R version 4.3.3Warning: package 'purrr' was built under R version 4.3.3War
ning: package 'lubridate' was built under R version 4.3.3
```

Hide

```
library(RColorBrewer)
```

Read in data

Hide

```
map <- read.table("../cov19_snp/cov19.map", header = F)
ped <- read.table("../cov19_snp/cov19.ped", header = F)
info <- read.table("../../sample_info.txt", header = T)
```

Modify data

Hide

```
#rename chromosome
map$V1 <- "genome"
#write to output
write.table(map,file = "../cov19_snp/covid19edit.map", quote = FALSE, sep = "\t", row.names = FA
LSE, col.names = FALSE)
#assign group
ped$V1[1] <- "reference"
ped$V1[2:1931] <- info$state
#set V1 as factor
ped$V1 <- as.factor(ped$V1)
levels(ped$V1)
```

```
 [1] "AK"        "AL"        "AR"        "AZ"        "CA"        "CO"        "CT"
 [8] "DE"        "FL"        "GA"        "HI"        "IA"        "ID"        "IL"
[15] "IN"        "KS"        "KY"        "LA"        "MA"        "MD"        "ME"
[22] "MI"        "MN"        "MO"        "MS"        "MT"        "NC"        "NE"
[29] "NH"        "NJ"        "NM"        "NV"        "NY"        "OH"        "OK"
[36] "OR"        "PA"        "PR"        "reference" "RI"        "SC"        "SD"
[43] "TN"        "TX"        "UT"        "VA"        "VT"        "WA"        "WI"
[50] "WV"        "WY"
```

Hide

```
#reorder levels in Column 1
ped$V1 <- factor(ped$V1, levels = c("reference", unique(info$state) ))
ped <-ped[order(ped$V1),]
#write to output
write.table(ped,file = "../cov19_snp/covid19edit.ped", quote = FALSE, sep = "\t", row.names = FA
LSE, col.names = FALSE)
#generate geno file
ped2geno("../cov19_snp/covid19edit.ped", output.file = "../cov19_snp/covid19edit.geno")
```

```
    - number of detected individuals:   1931
    - number of detected loci:      2599

[1] "../cov19_snp/covid19edit.geno"
```

Hide

```
#use PLINK to reformat our data into raw output
system("bash -c '../plink/plink.exe --version'")
```

```
PLINK v1.90b6.20 64-bit (21 Sep 2020)
[1] 0
```

Hide

```
system("bash -c '../plink/plink.exe --file ../cov19_snp/covid19edit --chr-set -1 --allow-extra-c
hr --recodeA --out ../cov19_snp/covid19Raw'")
```

```
PLINK v1.90b6.20 64-bit (21 Sep 2020)          www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Note: --recodeA flag deprecated.  Use "--recode A ...".
Logging to ../cov19_snp/covid19Raw.log.
Options in effect:
  --allow-extra-chr
  --chr-set -1
  --file ../cov19_snp/covid19edit
  --out ../cov19_snp/covid19Raw
  --recode A

15731 MB RAM detected; reserving 7865 MB for main workspace.
Scanning .ped file...
0%▒▒▒1%▒▒▒1%▒▒▒1%▒▒▒1%▒▒▒1%▒▒▒1%▒▒▒1%▒▒▒1%▒▒▒1%▒▒▒1%▒▒▒1%▒▒▒1%▒▒▒1%▒▒▒1%▒▒▒1%▒▒▒1%▒▒▒1%▒▒▒2%▒▒▒2%▒▒▒2%▒▒▒2%▒▒▒
2%▒▒▒2%▒▒▒2%▒▒▒2%▒▒▒2%▒▒▒2%▒▒▒2%▒▒▒2%▒▒▒2%▒▒▒2%▒▒▒2%▒▒▒2%▒▒▒2%▒▒▒2%▒▒▒3%▒▒▒3%▒▒▒3%▒▒▒3%▒▒▒3%▒▒▒3%▒▒▒3%▒▒▒3%▒▒▒
3%▒▒▒3%▒▒▒3%▒▒▒3%▒▒▒3%▒▒▒3%▒▒▒3%▒▒▒3%▒▒▒3%▒▒▒4%▒▒▒4%▒▒▒4%▒▒▒4%▒▒▒4%▒▒▒4%▒▒▒4%▒▒▒4%▒▒▒4%▒▒▒4%▒▒▒4%▒▒▒4%▒▒▒
4%▒▒▒4%▒▒▒4%▒▒▒4%▒▒▒4%▒▒▒5%▒▒▒5%▒▒▒5%▒▒▒5%▒▒▒5%▒▒▒5%▒▒▒5%▒▒▒5%▒▒▒5%▒▒▒5%▒▒▒5%▒▒▒5%▒▒▒5%▒▒▒5%▒▒▒5%▒▒▒5%▒▒▒
5%▒▒▒6%▒▒▒6%▒▒▒6%▒▒▒6%▒▒▒6%▒▒▒6%▒▒▒6%▒▒▒6%▒▒▒6%▒▒▒6%▒▒▒6%▒▒▒6%▒▒▒6%▒▒▒6%▒▒▒6%▒▒▒6%▒▒▒6%▒▒▒7%▒▒▒7%▒▒▒7%▒▒▒
7%▒▒▒7%▒▒▒7%▒▒▒7%▒▒▒7%▒▒▒7%▒▒▒7%▒▒▒7%▒▒▒7%▒▒▒7%▒▒▒7%▒▒▒7%▒▒▒7%▒▒▒7%▒▒▒8%▒▒▒8%▒▒▒8%▒▒▒8%▒▒▒8%▒▒▒8%▒▒▒8%▒▒▒
8%▒▒▒8%▒▒▒8%▒▒▒8%▒▒▒8%▒▒▒8%▒▒▒8%▒▒▒8%▒▒▒8%▒▒▒8%▒▒▒9%▒▒▒9%▒▒▒9%▒▒▒9%▒▒▒9%▒▒▒9%▒▒▒9%▒▒▒9%▒▒▒9%▒▒▒9%▒▒▒9%▒▒▒9%▒▒▒
9%▒▒▒9%▒▒▒9%▒▒▒9%▒▒▒9%▒▒▒9%▒▒▒10%▒▒▒▒10%▒▒▒▒10%▒▒▒▒10%▒▒▒▒10%▒▒▒▒10%▒▒▒▒10%▒▒▒▒10%▒▒▒▒10%▒▒▒▒10%▒▒▒▒10%▒▒▒▒10%▒▒▒▒1
0%▒▒▒▒10%▒▒▒▒10%▒▒▒▒10%▒▒▒▒10%▒▒▒▒10%▒▒▒▒10%▒▒▒▒10%▒▒▒▒11%▒▒▒▒11%▒▒▒▒11%▒▒▒▒11%▒▒▒▒11%▒▒▒▒11%▒▒▒▒11%▒▒▒▒11%▒▒▒▒1
1%▒▒▒▒11%▒▒▒▒11%▒▒▒▒11%▒▒▒▒11%▒▒▒▒11%▒▒▒▒11%▒▒▒▒11%▒▒▒▒11%▒▒▒▒11%▒▒▒▒11%▒▒▒▒12%▒▒▒▒12%▒▒▒▒12%▒▒▒▒12%▒▒▒▒12%▒▒▒▒1
2%▒▒▒▒12%▒▒▒▒12%▒▒▒▒12%▒▒▒▒12%▒▒▒▒12%▒▒▒▒12%▒▒▒▒12%▒▒▒▒12%▒▒▒▒12%▒▒▒▒12%▒▒▒▒12%▒▒▒▒12%▒▒▒▒12%▒▒▒▒13%▒▒▒▒1
3%▒▒▒▒13%▒▒▒▒13%▒▒▒▒13%▒▒▒▒13%▒▒▒▒13%▒▒▒▒13%▒▒▒▒13%▒▒▒▒13%▒▒▒▒13%▒▒▒▒13%▒▒▒▒13%▒▒▒▒13%▒▒▒▒13%▒▒▒▒13%▒▒▒▒13%▒▒▒▒1
3%▒▒▒▒13%▒▒▒▒14%▒▒▒▒14%▒▒▒▒14%▒▒▒▒14%▒▒▒▒14%▒▒▒▒14%▒▒▒▒14%▒▒▒▒14%▒▒▒▒14%▒▒▒▒14%▒▒▒▒14%▒▒▒▒14%▒▒▒▒14%▒▒▒▒14%▒▒▒▒1
4%▒▒▒▒14%▒▒▒▒14%▒▒▒▒14%▒▒▒▒14%▒▒▒▒15%▒▒▒▒15%▒▒▒▒15%▒▒▒▒15%▒▒▒▒15%▒▒▒▒15%▒▒▒▒15%▒▒▒▒15%▒▒▒▒15%▒▒▒▒15%▒▒▒▒15%▒▒▒▒1
5%▒▒▒▒15%▒▒▒▒15%▒▒▒▒15%▒▒▒▒15%▒▒▒▒15%▒▒▒▒15%▒▒▒▒15%▒▒▒▒16%▒▒▒▒16%▒▒▒▒16%▒▒▒▒16%▒▒▒▒16%▒▒▒▒16%▒▒▒▒16%▒▒▒▒16%▒▒▒▒1
6%▒▒▒▒16%▒▒▒▒16%▒▒▒▒16%▒▒▒▒16%▒▒▒▒16%▒▒▒▒16%▒▒▒▒16%▒▒▒▒16%▒▒▒▒16%▒▒▒▒16%▒▒▒▒16%▒▒▒▒17%▒▒▒▒17%▒▒▒▒17%▒▒▒▒17%▒▒▒▒1
7%▒▒▒▒17%▒▒▒▒17%▒▒▒▒17%▒▒▒▒17%▒▒▒▒17%▒▒▒▒17%▒▒▒▒17%▒▒▒▒17%▒▒▒▒17%▒▒▒▒17%▒▒▒▒17%▒▒▒▒17%▒▒▒▒17%▒▒▒▒17%▒▒▒▒18%▒▒▒▒1
8%▒▒▒▒18%▒▒▒▒18%▒▒▒▒18%▒▒▒▒18%▒▒▒▒18%▒▒▒▒18%▒▒▒▒18%▒▒▒▒18%▒▒▒▒18%▒▒▒▒18%▒▒▒▒18%▒▒▒▒18%▒▒▒▒18%▒▒▒▒18%▒▒▒▒18%▒▒▒▒1
8%▒▒▒▒18%▒▒▒▒19%▒▒▒▒19%▒▒▒▒19%▒▒▒▒19%▒▒▒▒19%▒▒▒▒19%▒▒▒▒19%▒▒▒▒19%▒▒▒▒19%▒▒▒▒19%▒▒▒▒19%▒▒▒▒19%▒▒▒▒19%▒▒▒▒19%▒▒▒▒1
9%▒▒▒▒19%▒▒▒▒19%▒▒▒▒19%▒▒▒▒19%▒▒▒▒19%▒▒▒▒20%▒▒▒▒20%▒▒▒▒20%▒▒▒▒20%▒▒▒▒20%▒▒▒▒20%▒▒▒▒20%▒▒▒▒20%▒▒▒▒20%▒▒▒▒20%▒▒▒▒2
0%▒▒▒▒20%▒▒▒▒20%▒▒▒▒20%▒▒▒▒20%▒▒▒▒20%▒▒▒▒20%▒▒▒▒20%▒▒▒▒20%▒▒▒▒21%▒▒▒▒21%▒▒▒▒21%▒▒▒▒21%▒▒▒▒21%▒▒▒▒21%▒▒▒▒21%▒▒▒▒2
1%▒▒▒▒21%▒▒▒▒21%▒▒▒▒21%▒▒▒▒21%▒▒▒▒21%▒▒▒▒21%▒▒▒▒21%▒▒▒▒21%▒▒▒▒21%▒▒▒▒21%▒▒▒▒21%▒▒▒▒22%▒▒▒▒22%▒▒▒▒22%▒▒▒▒22%▒▒▒▒2
2%▒▒▒▒22%▒▒▒▒22%▒▒▒▒22%▒▒▒▒22%▒▒▒▒22%▒▒▒▒22%▒▒▒▒22%▒▒▒▒22%▒▒▒▒22%▒▒▒▒22%▒▒▒▒22%▒▒▒▒22%▒▒▒▒22%▒▒▒▒22%▒▒▒▒22%▒▒▒▒2
3%▒▒▒▒23%▒▒▒▒23%▒▒▒▒23%▒▒▒▒23%▒▒▒▒23%▒▒▒▒23%▒▒▒▒23%▒▒▒▒23%▒▒▒▒23%▒▒▒▒23%▒▒▒▒23%▒▒▒▒23%▒▒▒▒23%▒▒▒▒23%▒▒▒▒23%▒▒▒▒2
3%▒▒▒▒23%▒▒▒▒23%▒▒▒▒24%▒▒▒▒24%▒▒▒▒24%▒▒▒▒24%▒▒▒▒24%▒▒▒▒24%▒▒▒▒24%▒▒▒▒24%▒▒▒▒24%▒▒▒▒24%▒▒▒▒24%▒▒▒▒24%▒▒▒▒24%▒▒▒▒2
4%▒▒▒▒24%▒▒▒▒24%▒▒▒▒24%▒▒▒▒24%▒▒▒▒24%▒▒▒▒25%▒▒▒▒25%▒▒▒▒25%▒▒▒▒25%▒▒▒▒25%▒▒▒▒25%▒▒▒▒25%▒▒▒▒25%▒▒▒▒25%▒▒▒▒25%▒▒▒▒2
5%▒▒▒▒25%▒▒▒▒25%▒▒▒▒25%▒▒▒▒25%▒▒▒▒25%▒▒▒▒25%▒▒▒▒25%▒▒▒▒25%▒▒▒▒25%▒▒▒▒26%▒▒▒▒26%▒▒▒▒26%▒▒▒▒26%▒▒▒▒26%▒▒▒▒26%▒▒▒▒2
6%▒▒▒▒26%▒▒▒▒26%▒▒▒▒26%▒▒▒▒26%▒▒▒▒26%▒▒▒▒26%▒▒▒▒26%▒▒▒▒26%▒▒▒▒26%▒▒▒▒26%▒▒▒▒26%▒▒▒▒26%▒▒▒▒27%▒▒▒▒27%▒▒▒▒27%▒▒▒▒2
7%▒▒▒▒27%▒▒▒▒27%▒▒▒▒27%▒▒▒▒27%▒▒▒▒27%▒▒▒▒27%▒▒▒▒27%▒▒▒▒27%▒▒▒▒27%▒▒▒▒27%▒▒▒▒27%▒▒▒▒27%▒▒▒▒27%▒▒▒▒27%▒▒▒▒27%▒▒▒▒2
8%▒▒▒▒28%▒▒▒▒28%▒▒▒▒28%▒▒▒▒28%▒▒▒▒28%▒▒▒▒28%▒▒▒▒28%▒▒▒▒28%▒▒▒▒28%▒▒▒▒28%▒▒▒▒28%▒▒▒▒28%▒▒▒▒28%▒▒▒▒28%▒▒▒▒28%▒▒▒▒2
8%▒▒▒▒28%▒▒▒▒28%▒▒▒▒29%▒▒▒▒29%▒▒▒▒29%▒▒▒▒29%▒▒▒▒29%▒▒▒▒29%▒▒▒▒29%▒▒▒▒29%▒▒▒▒29%▒▒▒▒29%▒▒▒▒29%▒▒▒▒29%▒▒▒▒29%▒▒▒▒2
9%▒▒▒▒29%▒▒▒▒29%▒▒▒▒29%▒▒▒▒29%▒▒▒▒29%▒▒▒▒29%▒▒▒▒30%▒▒▒▒30%▒▒▒▒30%▒▒▒▒30%▒▒▒▒30%▒▒▒▒30%▒▒▒▒30%▒▒▒▒30%▒▒▒▒30%▒▒▒▒3
0%▒▒▒▒30%▒▒▒▒30%▒▒▒▒30%▒▒▒▒30%▒▒▒▒30%▒▒▒▒30%▒▒▒▒30%▒▒▒▒30%▒▒▒▒30%▒▒▒▒31%▒▒▒▒31%▒▒▒▒31%▒▒▒▒31%▒▒▒▒31%▒▒▒▒31%▒▒▒▒3
1%▒▒▒▒31%▒▒▒▒31%▒▒▒▒31%▒▒▒▒31%▒▒▒▒31%▒▒▒▒31%▒▒▒▒31%▒▒▒▒31%▒▒▒▒31%▒▒▒▒31%▒▒▒▒31%▒▒▒▒32%▒▒▒▒32%▒▒▒▒32%▒▒▒▒3
2%▒▒▒▒32%▒▒▒▒32%▒▒▒▒32%▒▒▒▒32%▒▒▒▒32%▒▒▒▒32%▒▒▒▒32%▒▒▒▒32%▒▒▒▒32%▒▒▒▒32%▒▒▒▒32%▒▒▒▒32%▒▒▒▒32%▒▒▒▒32%▒▒▒▒32%▒▒▒▒3
2%▒▒▒▒33%▒▒▒▒33%▒▒▒▒33%▒▒▒▒33%▒▒▒▒33%▒▒▒▒33%▒▒▒▒33%▒▒▒▒33%▒▒▒▒33%▒▒▒▒33%▒▒▒▒33%▒▒▒▒33%▒▒▒▒33%▒▒▒▒33%▒▒▒▒33%▒▒▒▒3
3%▒▒▒▒33%▒▒▒▒33%▒▒▒▒33%▒▒▒▒34%▒▒▒▒34%▒▒▒▒34%▒▒▒▒34%▒▒▒▒34%▒▒▒▒34%▒▒▒▒34%▒▒▒▒34%▒▒▒▒34%▒▒▒▒34%▒▒▒▒34%▒▒▒▒3
4%▒▒▒▒34%▒▒▒▒34%▒▒▒▒34%▒▒▒▒34%▒▒▒▒34%▒▒▒▒34%▒▒▒▒35%▒▒▒▒35%▒▒▒▒35%▒▒▒▒35%▒▒▒▒35%▒▒▒▒35%▒▒▒▒35%▒▒▒▒35%▒▒▒▒35%▒▒▒▒3
5%▒▒▒▒35%▒▒▒▒35%▒▒▒▒35%▒▒▒▒35%▒▒▒▒35%▒▒▒▒35%▒▒▒▒35%▒▒▒▒35%▒▒▒▒35%▒▒▒▒36%▒▒▒▒36%▒▒▒▒36%▒▒▒▒36%▒▒▒▒36%▒▒▒▒3
```

6%   36%   36%   36%   36%   36%   36%   36%   36%   36%   36%   36%   36%   37%   37%   3
7%   37%   37%   37%   37%   37%   37%   37%   37%   37%   37%   37%   37%   37%   37%   3
7%   38%   38%   38%   38%   38%   38%   38%   38%   38%   38%   38%   38%   38%   38%   3
8%   38%   38%   38%   38%   39%   39%   39%   39%   39%   39%   39%   39%   39%   39%   3
9%   39%   39%   39%   39%   39%   39%   40%   40%   40%   40%   40%   40%   40%   40%   4
0%   40%   40%   40%   40%   40%   40%   40%   40%   40%   40%   41%   41%   41%   41%   4
1%   41%   41%   41%   41%   41%   41%   41%   41%   41%   41%   41%   41%   41%   42%   4
2%   42%   42%   42%   42%   42%   42%   42%   42%   42%   42%   42%   42%   42%   42%   4
2%   42%   43%   43%   43%   43%   43%   43%   43%   43%   43%   43%   43%   43%   43%   4
3%   43%   43%   43%   43%   44%   44%   44%   44%   44%   44%   44%   44%   44%   44%   4
4%   44%   44%   44%   44%   44%   44%   44%   45%   45%   45%   45%   45%   45%   45%   4
5%   45%   45%   45%   45%   45%   45%   45%   45%   45%   45%   45%   46%   46%   46%   4
6%   46%   46%   46%   46%   46%   46%   46%   46%   46%   46%   46%   46%   46%   47%   4
7%   47%   47%   47%   47%   47%   47%   47%   47%   47%   47%   47%   47%   47%   47%   4
7%   47%   48%   48%   48%   48%   48%   48%   48%   48%   48%   48%   48%   48%   48%   4
8%   48%   48%   48%   48%   49%   49%   49%   49%   49%   49%   49%   49%   49%   49%   4
9%   49%   49%   49%   49%   49%   49%   49%   50%   50%   50%   50%   50%   50%   50%   5
0%   50%   50%   50%   50%   50%   50%   50%   50%   50%   50%   50%   51%   51%   51%   5
1%   51%   51%   51%   51%   51%   51%   51%   51%   51%   51%   51%   51%   51%   51%   5
2%   52%   52%   52%   52%   52%   52%   52%   52%   52%   52%   52%   52%   52%   52%   5
2%   52%   52%   53%   53%   53%   53%   53%   53%   53%   53%   53%   53%   53%   53%   5
3%   53%   53%   53%   53%   53%   54%   54%   54%   54%   54%   54%   54%   54%   54%   5
4%   54%   54%   54%   54%   54%   54%   54%   54%   54%   55%   55%   55%   55%   55%   5
5%   55%   55%   55%   55%   55%   55%   55%   55%   55%   55%   55%   55%   56%   56%   5
6%   56%   56%   56%   56%   56%   56%   56%   56%   56%   56%   56%   56%   56%   56%   5
7%   57%   57%   57%   57%   57%   57%   57%   57%   57%   57%   57%   57%   57%   57%   5
7%   57%   57%   58%   58%   58%   58%   58%   58%   58%   58%   58%   58%   58%   58%   5
8%   58%   58%   58%   58%   58%   58%   59%   59%   59%   59%   59%   59%   59%   59%   5
9%   59%   59%   59%   59%   59%   59%   59%   59%   59%   60%   60%   60%   60%   60%   6
0%   60%   60%   60%   60%   60%   60%   60%   60%   60%   60%   60%   60%   61%   61%   6
1%   61%   61%   61%   61%   61%   61%   61%   61%   61%   61%   61%   61%   61%   61%   6
1%   62%   62%   62%   62%   62%   62%   62%   62%   62%   62%   62%   62%   62%   62%   6
2%   62%   62%   62%   63%   63%   63%   63%   63%   63%   63%   63%   63%   63%   63%   6
3%   63%   63%   63%   63%   63%   64%   64%   64%   64%   64%   64%   64%   64%   64%   6
4%   64%   64%   64%   64%   64%   64%   64%   64%   64%   64%   65%   65%   65%   65%   6
5%   65%   65%   65%   65%   65%   65%   65%   65%   65%   65%   65%   65%   66%   66%   6
6%   66%   66%   66%   66%   66%   66%   66%   66%   66%   66%   66%   66%   66%   66%   6
6%   67%   67%   67%   67%   67%   67%   67%   67%   67%   67%   67%   67%   67%   67%   6
7%   67%   67%   67%   68%   68%   68%   68%   68%   68%   68%   68%   68%   68%   68%   6
8%   68%   68%   68%   68%   68%   68%   69%   69%   69%   69%   69%   69%   69%   69%   6
9%   69%   69%   69%   69%   69%   69%   69%   69%   69%   69%   70%   70%   70%   70%   7
0%   70%   70%   70%   70%   70%   70%   70%   70%   70%   70%   70%   70%   70%   71%   7
1%   71%   71%   71%   71%   71%   71%   71%   71%   71%   71%   71%   71%   71%   71%   7
1%   71%   72%   72%   72%   72%   72%   72%   72%   72%   72%   72%   72%   72%   72%   7
2%   72%   72%   72%   72%   73%   73%   73%   73%   73%   73%   73%   73%   73%   73%   7
3%   73%   73%   73%   73%   73%   73%   74%   74%   74%   74%   74%   74%   74%   74%   7
4%   74%   74%   74%   74%   74%   74%   74%   74%   74%   74%   75%   75%   75%   75%   7
5%   75%   75%   75%   75%   75%   75%   75%   75%   75%   75%   75%   75%   75%   76%   7
6%   76%   76%   76%   76%   76%   76%   76%   76%   76%   76%   76%   76%   76%   76%   7
6%   76%   77%   77%   77%   77%   77%   77%   77%   77%   77%   77%   77%   77%   77%   7
7%   77%   77%   77%   77%   78%   78%   78%   78%   78%   78%   78%   78%   78%   78%   7
8%   78%   78%   78%   78%   78%   78%   78%   79%   79%   79%   79%   79%   79%   79%   7

```
9%▒▒▒▒79%▒▒▒▒79%▒▒▒▒79%▒▒▒▒79%▒▒▒▒79%▒▒▒▒79%▒▒▒▒79%▒▒▒▒79%▒▒▒▒79%▒▒▒▒79%▒▒▒▒80%▒▒▒▒80%▒▒▒▒80%▒▒▒▒80%▒▒▒▒8
0%▒▒▒▒80%▒▒▒▒80%▒▒▒▒80%▒▒▒▒80%▒▒▒▒80%▒▒▒▒80%▒▒▒▒80%▒▒▒▒80%▒▒▒▒80%▒▒▒▒80%▒▒▒▒80%▒▒▒▒80%▒▒▒▒80%▒▒▒▒80%▒▒▒▒8
1%▒▒▒▒81%▒▒▒▒81%▒▒▒▒81%▒▒▒▒81%▒▒▒▒81%▒▒▒▒81%▒▒▒▒81%▒▒▒▒81%▒▒▒▒81%▒▒▒▒81%▒▒▒▒81%▒▒▒▒81%▒▒▒▒81%▒▒▒▒81%▒▒▒▒8
1%▒▒▒▒81%▒▒▒▒81%▒▒▒▒82%▒▒▒▒82%▒▒▒▒82%▒▒▒▒82%▒▒▒▒82%▒▒▒▒82%▒▒▒▒82%▒▒▒▒82%▒▒▒▒82%▒▒▒▒82%▒▒▒▒82%▒▒▒▒82%▒▒▒▒8
2%▒▒▒▒82%▒▒▒▒82%▒▒▒▒82%▒▒▒▒82%▒▒▒▒82%▒▒▒▒83%▒▒▒▒83%▒▒▒▒83%▒▒▒▒83%▒▒▒▒83%▒▒▒▒83%▒▒▒▒83%▒▒▒▒83%▒▒▒▒83%▒▒▒▒8
3%▒▒▒▒83%▒▒▒▒83%▒▒▒▒83%▒▒▒▒83%▒▒▒▒83%▒▒▒▒83%▒▒▒▒83%▒▒▒▒83%▒▒▒▒83%▒▒▒▒84%▒▒▒▒84%▒▒▒▒84%▒▒▒▒84%▒▒▒▒84%▒▒▒▒8
4%▒▒▒▒84%▒▒▒▒84%▒▒▒▒84%▒▒▒▒84%▒▒▒▒84%▒▒▒▒84%▒▒▒▒84%▒▒▒▒84%▒▒▒▒84%▒▒▒▒84%▒▒▒▒84%▒▒▒▒84%▒▒▒▒85%▒▒▒▒85%▒▒▒▒85%▒▒▒▒8
5%▒▒▒▒85%▒▒▒▒85%▒▒▒▒85%▒▒▒▒85%▒▒▒▒85%▒▒▒▒85%▒▒▒▒85%▒▒▒▒85%▒▒▒▒85%▒▒▒▒85%▒▒▒▒85%▒▒▒▒85%▒▒▒▒85%▒▒▒▒85%▒▒▒▒8
6%▒▒▒▒86%▒▒▒▒86%▒▒▒▒86%▒▒▒▒86%▒▒▒▒86%▒▒▒▒86%▒▒▒▒86%▒▒▒▒86%▒▒▒▒86%▒▒▒▒86%▒▒▒▒86%▒▒▒▒86%▒▒▒▒86%▒▒▒▒86%▒▒▒▒8
6%▒▒▒▒86%▒▒▒▒86%▒▒▒▒87%▒▒▒▒87%▒▒▒▒87%▒▒▒▒87%▒▒▒▒87%▒▒▒▒87%▒▒▒▒87%▒▒▒▒87%▒▒▒▒87%▒▒▒▒87%▒▒▒▒87%▒▒▒▒87%▒▒▒▒8
7%▒▒▒▒87%▒▒▒▒87%▒▒▒▒87%▒▒▒▒87%▒▒▒▒87%▒▒▒▒87%▒▒▒▒88%▒▒▒▒88%▒▒▒▒88%▒▒▒▒88%▒▒▒▒88%▒▒▒▒88%▒▒▒▒88%▒▒▒▒88%▒▒▒▒88%▒▒▒▒8
8%▒▒▒▒88%▒▒▒▒88%▒▒▒▒88%▒▒▒▒88%▒▒▒▒88%▒▒▒▒88%▒▒▒▒88%▒▒▒▒88%▒▒▒▒88%▒▒▒▒89%▒▒▒▒89%▒▒▒▒89%▒▒▒▒89%▒▒▒▒89%▒▒▒▒89%▒▒▒▒8
9%▒▒▒▒89%▒▒▒▒89%▒▒▒▒89%▒▒▒▒89%▒▒▒▒89%▒▒▒▒89%▒▒▒▒89%▒▒▒▒89%▒▒▒▒89%▒▒▒▒89%▒▒▒▒89%▒▒▒▒89%▒▒▒▒90%▒▒▒▒90%▒▒▒▒90%▒▒▒▒9
0%▒▒▒▒90%▒▒▒▒90%▒▒▒▒90%▒▒▒▒90%▒▒▒▒90%▒▒▒▒90%▒▒▒▒90%▒▒▒▒90%▒▒▒▒90%▒▒▒▒90%▒▒▒▒90%▒▒▒▒90%▒▒▒▒90%▒▒▒▒90%▒▒▒▒9
0%▒▒▒▒91%▒▒▒▒91%▒▒▒▒91%▒▒▒▒91%▒▒▒▒91%▒▒▒▒91%▒▒▒▒91%▒▒▒▒91%▒▒▒▒91%▒▒▒▒91%▒▒▒▒91%▒▒▒▒91%▒▒▒▒91%▒▒▒▒91%▒▒▒▒9
1%▒▒▒▒91%▒▒▒▒91%▒▒▒▒91%▒▒▒▒92%▒▒▒▒92%▒▒▒▒92%▒▒▒▒92%▒▒▒▒92%▒▒▒▒92%▒▒▒▒92%▒▒▒▒92%▒▒▒▒92%▒▒▒▒92%▒▒▒▒92%▒▒▒▒9
2%▒▒▒▒92%▒▒▒▒92%▒▒▒▒92%▒▒▒▒92%▒▒▒▒92%▒▒▒▒92%▒▒▒▒93%▒▒▒▒93%▒▒▒▒93%▒▒▒▒93%▒▒▒▒93%▒▒▒▒93%▒▒▒▒93%▒▒▒▒93%▒▒▒▒93%▒▒▒▒9
3%▒▒▒▒93%▒▒▒▒93%▒▒▒▒93%▒▒▒▒93%▒▒▒▒93%▒▒▒▒93%▒▒▒▒93%▒▒▒▒93%▒▒▒▒93%▒▒▒▒93%▒▒▒▒94%▒▒▒▒94%▒▒▒▒94%▒▒▒▒94%▒▒▒▒94%▒▒▒▒9
4%▒▒▒▒94%▒▒▒▒94%▒▒▒▒94%▒▒▒▒94%▒▒▒▒94%▒▒▒▒94%▒▒▒▒94%▒▒▒▒94%▒▒▒▒94%▒▒▒▒94%▒▒▒▒94%▒▒▒▒94%▒▒▒▒94%▒▒▒▒95%▒▒▒▒95%▒▒▒▒9
5%▒▒▒▒95%▒▒▒▒95%▒▒▒▒95%▒▒▒▒95%▒▒▒▒95%▒▒▒▒95%▒▒▒▒95%▒▒▒▒95%▒▒▒▒95%▒▒▒▒95%▒▒▒▒95%▒▒▒▒95%▒▒▒▒95%▒▒▒▒95%▒▒▒▒9
5%▒▒▒▒96%▒▒▒▒96%▒▒▒▒96%▒▒▒▒96%▒▒▒▒96%▒▒▒▒96%▒▒▒▒96%▒▒▒▒96%▒▒▒▒96%▒▒▒▒96%▒▒▒▒96%▒▒▒▒96%▒▒▒▒96%▒▒▒▒96%▒▒▒▒9
6%▒▒▒▒96%▒▒▒▒96%▒▒▒▒96%▒▒▒▒96%▒▒▒▒97%▒▒▒▒97%▒▒▒▒97%▒▒▒▒97%▒▒▒▒97%▒▒▒▒97%▒▒▒▒97%▒▒▒▒97%▒▒▒▒97%▒▒▒▒97%▒▒▒▒9
7%▒▒▒▒97%▒▒▒▒97%▒▒▒▒97%▒▒▒▒97%▒▒▒▒97%▒▒▒▒97%▒▒▒▒97%▒▒▒▒98%▒▒▒▒98%▒▒▒▒98%▒▒▒▒98%▒▒▒▒98%▒▒▒▒98%▒▒▒▒98%▒▒▒▒98%▒▒▒▒9
8%▒▒▒▒98%▒▒▒▒98%▒▒▒▒98%▒▒▒▒98%▒▒▒▒98%▒▒▒▒98%▒▒▒▒98%▒▒▒▒98%▒▒▒▒98%▒▒▒▒98%▒▒▒▒99%▒▒▒▒99%▒▒▒▒99%▒▒▒▒99%▒▒▒▒99%▒▒▒▒9
9%▒▒▒▒99%▒▒▒▒99%▒▒▒▒99%▒▒▒▒99%▒▒▒▒99%▒▒▒▒99%▒▒▒▒99%▒▒▒▒99%▒▒▒▒99%▒▒▒▒99%▒▒▒▒99%▒▒▒▒99%▒▒▒▒99%▒▒▒▒100%
.ped scan complete (for binary autoconversion).
Performing single-pass .bed write (2599 variants, 1931 samples).
0%▒▒▒1%▒▒▒2%▒▒▒3%▒▒▒4%▒▒▒5%▒▒▒6%▒▒▒7%▒▒▒8%▒▒▒9%▒▒▒10%▒▒▒▒11%▒▒▒▒12%▒▒▒▒13%▒▒▒▒14%▒▒▒▒15%▒▒▒▒16%▒▒▒▒17%▒▒▒▒18%▒▒▒▒1
9%▒▒▒▒20%▒▒▒▒21%▒▒▒▒22%▒▒▒▒23%▒▒▒▒24%▒▒▒▒25%▒▒▒▒26%▒▒▒▒27%▒▒▒▒28%▒▒▒▒29%▒▒▒▒30%▒▒▒▒31%▒▒▒▒32%▒▒▒▒33%▒▒▒▒34%▒▒▒▒3
5%▒▒▒▒36%▒▒▒▒37%▒▒▒▒38%▒▒▒▒39%▒▒▒▒40%▒▒▒▒41%▒▒▒▒42%▒▒▒▒43%▒▒▒▒44%▒▒▒▒45%▒▒▒▒46%▒▒▒▒47%▒▒▒▒48%▒▒▒▒49%▒▒▒▒50%▒▒▒▒5
1%▒▒▒▒52%▒▒▒▒53%▒▒▒▒54%▒▒▒▒55%▒▒▒▒56%▒▒▒▒57%▒▒▒▒58%▒▒▒▒59%▒▒▒▒60%▒▒▒▒61%▒▒▒▒62%▒▒▒▒63%▒▒▒▒64%▒▒▒▒65%▒▒▒▒66%▒▒▒▒6
7%▒▒▒▒68%▒▒▒▒69%▒▒▒▒70%▒▒▒▒71%▒▒▒▒72%▒▒▒▒73%▒▒▒▒74%▒▒▒▒75%▒▒▒▒76%▒▒▒▒77%▒▒▒▒78%▒▒▒▒79%▒▒▒▒80%▒▒▒▒81%▒▒▒▒82%▒▒▒▒8
3%▒▒▒▒84%▒▒▒▒85%▒▒▒▒86%▒▒▒▒87%▒▒▒▒88%▒▒▒▒89%▒▒▒▒90%▒▒▒▒91%▒▒▒▒92%▒▒▒▒93%▒▒▒▒94%
--file: ../cov19_snp/covid19Raw-temporary.bed +
../cov19_snp/covid19Raw-temporary.bim + ../cov19_snp/covid19Raw-temporary.fam
written.
2599 variants loaded from .bim file.
1931 samples (0 males, 0 females, 1931 ambiguous) loaded from .fam.
Ambiguous sex IDs written to ../cov19_snp/covid19Raw.nosex .
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 1931 founders and 0 nonfounders present.
Calculating allele frequencies... 0%▒▒▒1%▒▒▒2%▒▒▒3%▒▒▒4%▒▒▒5%▒▒▒6%▒▒▒7%▒▒▒8%▒▒▒9%▒▒▒10%▒▒▒▒11%▒▒▒▒12%▒▒▒▒1
3%▒▒▒▒14%▒▒▒▒15%▒▒▒▒16%▒▒▒▒17%▒▒▒▒18%▒▒▒▒19%▒▒▒▒20%▒▒▒▒21%▒▒▒▒22%▒▒▒▒23%▒▒▒▒24%▒▒▒▒25%▒▒▒▒26%▒▒▒▒27%▒▒▒▒28%▒▒▒▒2
9%▒▒▒▒30%▒▒▒▒31%▒▒▒▒32%▒▒▒▒33%▒▒▒▒34%▒▒▒▒35%▒▒▒▒36%▒▒▒▒37%▒▒▒▒38%▒▒▒▒39%▒▒▒▒40%▒▒▒▒41%▒▒▒▒42%▒▒▒▒43%▒▒▒▒44%▒▒▒▒4
5%▒▒▒▒46%▒▒▒▒47%▒▒▒▒48%▒▒▒▒49%▒▒▒▒50%▒▒▒▒51%▒▒▒▒52%▒▒▒▒53%▒▒▒▒54%▒▒▒▒55%▒▒▒▒56%▒▒▒▒57%▒▒▒▒58%▒▒▒▒59%▒▒▒▒60%▒▒▒▒6
1%▒▒▒▒62%▒▒▒▒63%▒▒▒▒64%▒▒▒▒65%▒▒▒▒66%▒▒▒▒67%▒▒▒▒68%▒▒▒▒69%▒▒▒▒70%▒▒▒▒71%▒▒▒▒72%▒▒▒▒73%▒▒▒▒74%▒▒▒▒75%▒▒▒▒76%▒▒▒▒7
7%▒▒▒▒78%▒▒▒▒79%▒▒▒▒80%▒▒▒▒81%▒▒▒▒82%▒▒▒▒83%▒▒▒▒84%▒▒▒▒85%▒▒▒▒86%▒▒▒▒87%▒▒▒▒88%▒▒▒▒89%▒▒▒▒90%▒▒▒▒91%▒▒▒▒92%▒▒▒▒9
3%▒▒▒▒94%▒▒▒▒95%▒▒▒▒96%▒▒▒▒97%▒▒▒▒98%▒▒▒▒99%▒▒▒▒▒ done.
Total genotyping rate is 0.991728.
2599 variants and 1931 samples pass filters and QC.
Note: No phenotypes present.
--recode A to ../cov19_snp/covid19Raw.raw ... 0%▒▒▒1%▒▒▒2%▒▒▒3%▒▒▒4%▒▒▒5%▒▒▒6%▒▒▒7%▒▒▒8%▒▒▒9%▒▒▒10%▒▒▒▒1
```

```
1%□□□12%□□□13%□□□14%□□□15%□□□16%□□□17%□□□18%□□□19%□□□20%□□□21%□□□22%□□□23%□□□24%□□□25%□□□26%□□□2
7%□□□28%□□□29%□□□30%□□□31%□□□32%□□□33%□□□34%□□□35%□□□36%□□□37%□□□38%□□□39%□□□40%□□□41%□□□42%□□□4
3%□□□44%□□□45%□□□46%□□□47%□□□48%□□□49%□□□50%□□□51%□□□52%□□□53%□□□54%□□□55%□□□56%□□□57%□□□58%□□□5
9%□□□60%□□□61%□□□62%□□□63%□□□64%□□□65%□□□66%□□□67%□□□68%□□□69%□□□70%□□□71%□□□72%□□□73%□□□74%□□□7
5%□□□76%□□□77%□□□78%□□□79%□□□80%□□□81%□□□82%□□□83%□□□84%□□□85%□□□86%□□□87%□□□88%□□□89%□□□90%□□□9
1%□□□92%□□□93%□□□94%□□□95%□□□96%□□□97%□□□98%□□□99%□□□done.
[1] 0
```

Read in edited SNP data

Hide

```
# now import data into adegenet using the function read.PLINK()
# Windows users will want to add the argument parallel = FALSE
covidSNPs <- read.PLINK("../cov19_snp/covid19Raw.raw", parallel = F)
```

```
 Reading PLINK raw format into a genlight object...


 Reading loci information...

 Reading and converting genotypes...
..
 Building final object...

...done.
```

Hide

```
# check to see if it worked by just executing the object name
names(covidSNPs)
```

```
 [1] "gen"        "n.loc"      "ind.names"  "loc.names"  "loc.all"    "chromosome" "position"
 [8] "ploidy"     "pop"        "strata"     "hierarchy"  "other"
```

Hide

```
covidSNPs$ploidy <- as.integer(rep(1, 1931))
# save
#save.image(file = "./SNP_analysis.rdata")
```

Sanity check

Hide

```
#use nInd() to check the number of individuals
nInd(covidSNPs)
```

```
[1] 1931
```

Hide

```
# use nLoc() to check the number of SNPs (loci)
nLoc(covidSNPs)
```

```
[1] 2599
```

Hide

```
#pop() returns a factor indicating the population of each individual
#use table() to produce a table of the output of pop()
table(pop(covidSNPs))
```

```
       AK        AL        AR        AZ        CA        CO        CT        DE        FL 
       50        16        50        50        50        50        50        50        50 
       GA        HI        IA        ID        IL        IN        KS        KY        LA 
       50        10        50        50        50        34         3         5        50 
       MA        MD        ME        MI        MN        MO        MS        MT        NC 
       50        50        50        50        50        50        44         4        50 
       NE        NH        NJ        NM        NV        NY        OH        OK        OR 
       50         5        50        50        50        50        21        13        50 
       PA        PR reference        RI        SC        SD        TN        TX        UT 
       50        35         1         7        50        16        11        50        50 
       VA        VT        WA        WI        WV        WY 
       50         4        50        50         2        50 
```

Hide

```
#remove groups with only 1 individual
covidSNPs <- subset(covidSNPs, covidSNPs$pop != "reference" & covidSNPs$pop != "FL")
table(pop(covidSNPs))
```

```
AK AL AR AZ CA CO CT DE GA HI IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC NE NH NJ NM NV NY 
50 16 50 50 50 50 50 50 50 10 50 50 50 34  3  5 50 50 50 50 50 50 44  4 50 50  5 50 50 50 50 
OH OK OR PA PR RI SC SD TN TX UT VA VT WA WI WV WY 
21 13 50 50 35  7 50 16 11 50 50 50  4 50 50  2 50 
```

DAPC

DAPC analysis with state as group

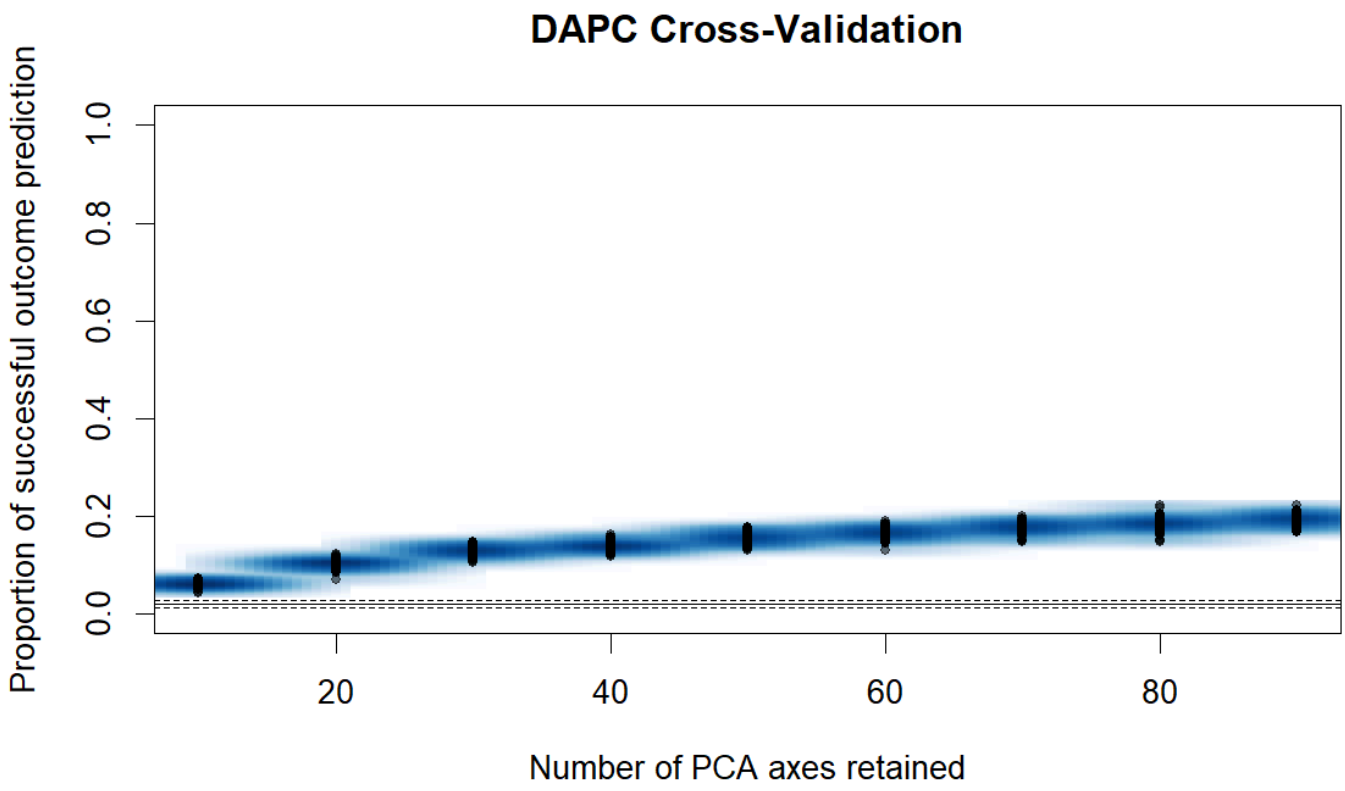
Hide

```
#change genlight object to a matrix
covidSNPsMat <- as.matrix(covidSNPs)
# replace any missing genotypes with the mean
covidSNPsImpute <- na.replace(covidSNPsMat, 0, na.rm = TRUE)
#use pop() to get group membership information for all individuals
group <- pop(covidSNPs)
#now use xvalDapc() with the group membership info from above, and
#a maximum of 100 PCs and 100 repetitions
xval<-xvalDapc(covidSNPsImpute, grp = group, n.pca.max=100, n.rep=100, parallel = "multicore", n
cpus = 8)
#look at the items two to six in the output
xval[2:6]
```

```
$`Median and Confidence Interval for Random Chance`
      2.5%        50%      97.5%
0.01436781 0.02052345 0.02708827

$`Mean Successful Assignment by Number of PCs of PCA`
        10         20         30         40         50         60         70         80
0.06102281 0.10204658 0.12926474 0.13867179 0.15536037 0.16511666 0.17612022 0.18371017
        90
0.19302710

$`Number of PCs Achieving Highest Mean Success`
[1] "90"

$`Root Mean Squared Error by Number of PCs of PCA`
       10        20        30        40        50        60        70        80        90
0.9390014 0.8979951 0.8707858 0.8613778 0.8447108 0.8349504 0.8239554 0.8163831 0.8070508

$`Number of PCs Achieving Lowest MSE`
[1] "90"
```

Hide

```
covidDAPC <- xval[7]$DAPC
#plot
par(mfrow = c(2,2))
```

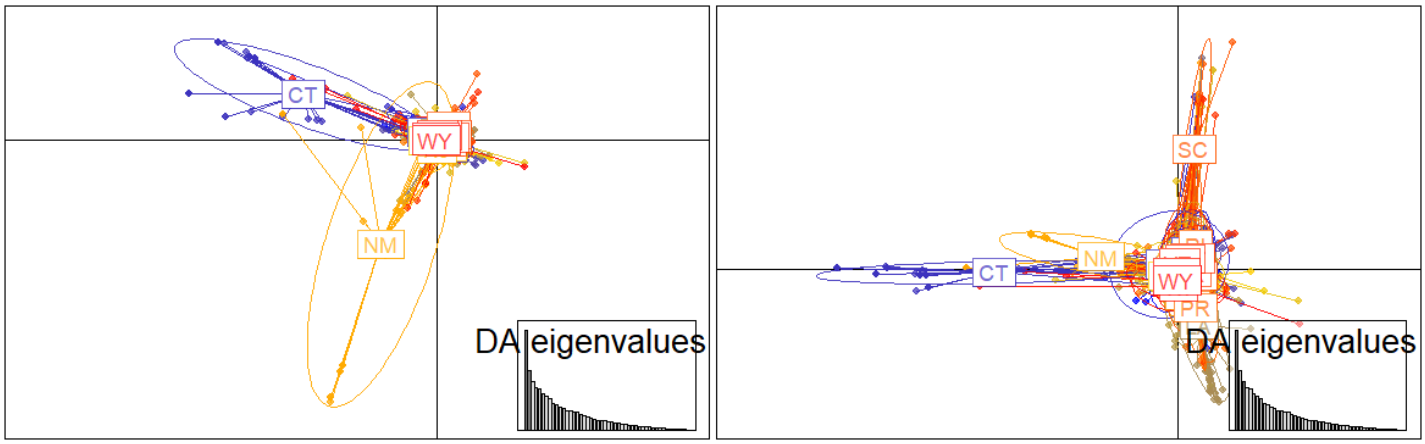


Hide

```
scatter(covidDAPC, xax = 1, yax = 2) #use discriminant functions 1 for the x-axis & 2 for the y-axis
scatter(covidDAPC, xax = 1, yax = 3) #use discriminant functions 1 for the x-axis & 3 for the y-axis
```

Hide

```
compoplot(covidDAPC)
```



Save plot

Hide

```
pdf("./plot/DAPC.pdf", width = 8, height = 5)
par(mfrow = c(2,2))
scatter(covidDAPC, xax = 1, yax = 2)
scatter(covidDAPC, xax = 1, yax = 3)
```

Hide

```
dev.off()
```

```
null device
          1
```

Hide

```
pdf("./plot/DAPC_composition.pdf", width = 10, height = 4)
compoplot(covidDAPC)
```

Hide

```
dev.off()
```

```
null device
          1
```

Clustering with no prior group

Hide

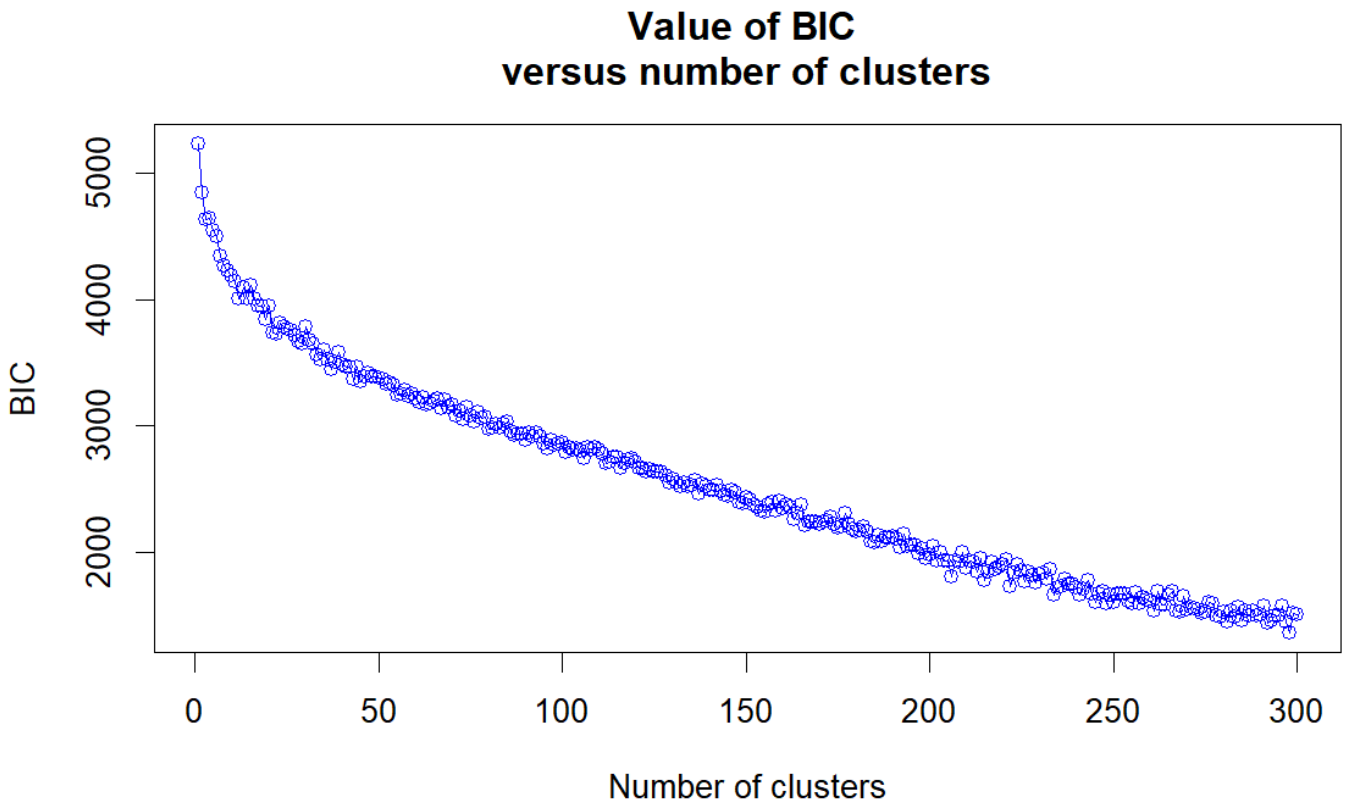
```
#run find.clusters() on the raw data
#use max.n.clust = 100
cluster <- find.clusters(covidSNPs, max.n.clust = 300, n.pca = 200)
```

```
Warning: did not converge in 100000 iterations
```

```
Choose the number of clusters (>=2):
```

Hide

```
100
```



No optimal K found, skip DAPC without prior group

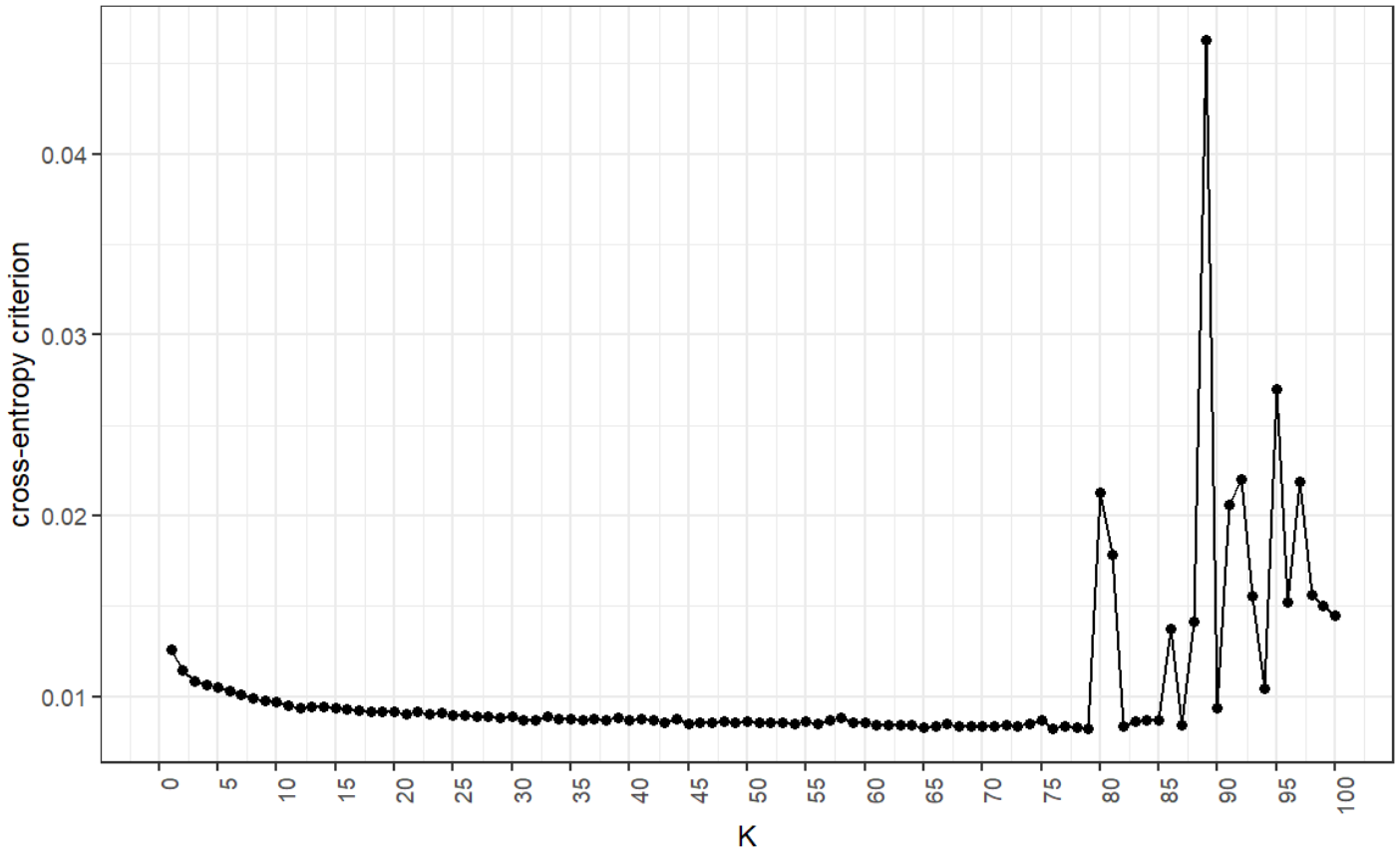
Admixture modeling

Hide

```
#save.image(file = "./SNP_analysis.rdata")
```

Hide

```
# plot cross-entroypy criterion for each K
ces_sum <-data.frame()
ce_sum <- summary(aa_res)$crossEntropy %>%
  as.data.frame()
colnames(ce_sum) <- 1:100
ce_sum <- t(ce_sum) %>%
  as.data.frame()
ce_sum$K <- 1:100
ggplot(ce_sum) +
  geom_point(aes(x = K, y = mean)) +
  geom_path(aes(x = K, y = mean)) +
  scale_x_continuous(limits=c(0,100), breaks=seq(0,100, by = 5)) +
  labs(x = "K", y = "cross-entropy criterion")+
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 0.95))
ggsave(filename = "./plot/cross_entroypy.pdf", width = 8, height = 3, units = "in")
```



Hide

```
# choose k =5
# Pick the K where cross-entropy stops dropping sharply
cross.entropy(aa_res,5)
```

```
          K = 5
run 1 0.01053063
```

Hide

```
q_k5 <- Q(aa_res, 5, 1) %>% as.data.frame()
# assign group
sample_size <- info %>% group_by(state) %>% summarise("size" = n())
sample_size <- arrange(sample_size, desc(size))
q_k5$state <- ped$V1
# set state as factor
q_k5$state <- as.factor(q_k5$state)
levels(q_k5$state)
```

```
 [1] "reference" "AK"        "AL"        "AR"        "AZ"        "CA"        "CO"
 [8] "CT"        "DE"        "FL"        "GA"        "HI"        "IA"        "ID"
[15] "IL"        "IN"        "KS"        "KY"        "LA"        "MA"        "MD"
[22] "ME"        "MI"        "MN"        "MO"        "MS"        "MT"        "NC"
[29] "NE"        "NH"        "NJ"        "NM"        "NV"        "NY"        "OH"
[36] "OK"        "OR"        "PA"        "PR"        "RI"        "SC"        "SD"
[43] "TN"        "TX"        "UT"        "VA"        "VT"        "WA"        "WI"
[50] "WV"        "WY"
```

Hide

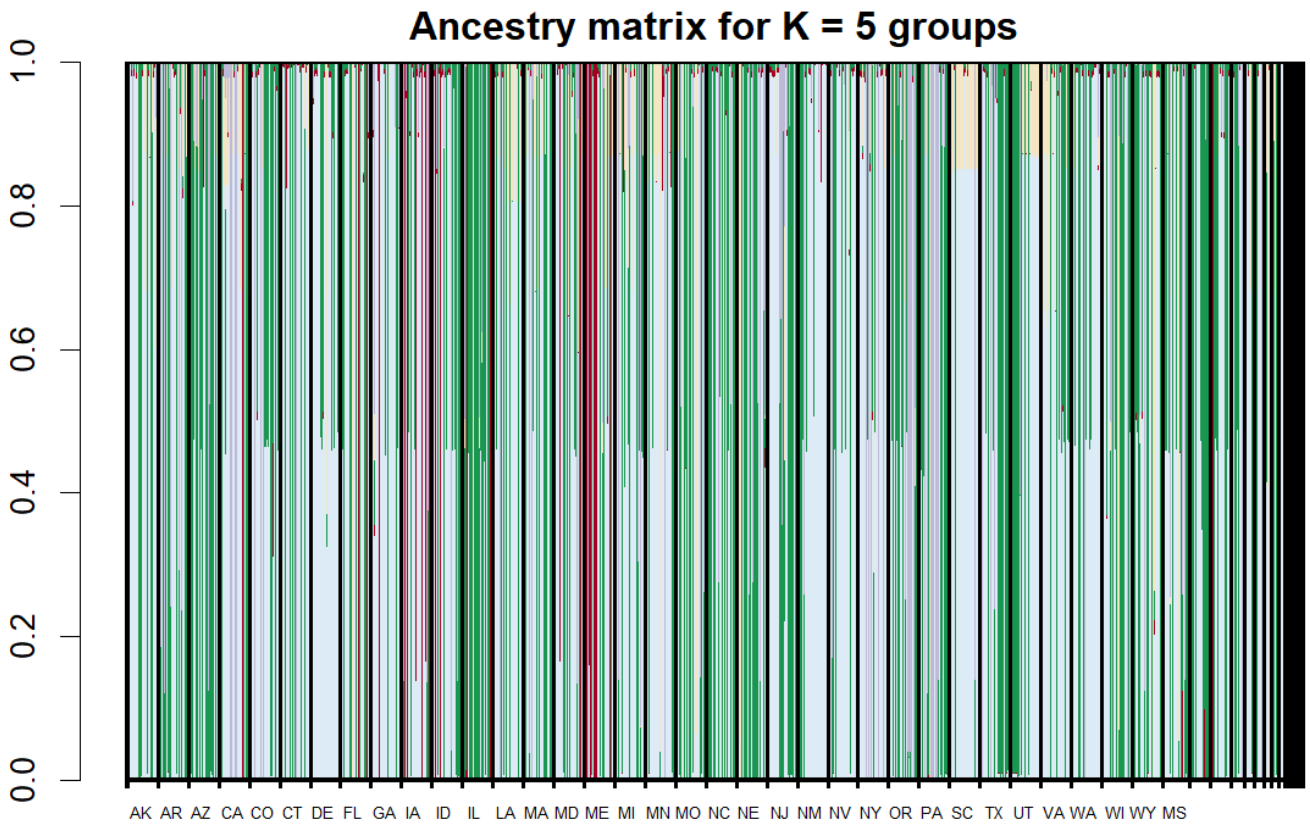
```
# reorder based on state
q_k5$state <- factor(q_k5$state, levels = c("reference", as.character(sample_size$state)))
q_k5 <-q_k5[order(q_k5$state),]
# remove last column
q_k5$state <- NULL
# store x break info
sample_size <- rbind(c("WUHAN", 1), sample_size)
sample_size$size <- as.numeric(sample_size$size)
sample_size$xbreak <- NA
for(i in 1:nrow(sample_size)){
  sample_size$xbreak[i] <- sum(sample_size$size[1:i])
}


#plot
n <- 5
qual_col_pals = brewer.pal.info
col_vector = unlist(mapply(brewer.pal, qual_col_pals$maxcolors, rownames(qual_col_pals)))
colors <- sample(col_vector, n)

gap <- (0.845-0.04)/33
par(mar = c(2, 2, 2, 2))
barplot(t(q_k5), border = NA, space = 0, col = colors,
        ylab = "Ancestry proportions", xlab = "Individuals",
        main = "Ancestry matrix for K = 5 groups", xaxt = 'n')
# add boxes around each sampling site
abline(v = c(0, sample_size$xbreak), lwd = 2)
```

Hide

```
segments(0, 0, 1931, 0, lwd = 3)
segments(0, 1, 1931, 1, lwd = 3)
```

Hide

```
for (i in 2:35){
  mtext(as.character(sample_size$state[i]), 1, adj = 0.04 + (i-2)*gap, cex = 0.5)
}
mtext("MS", 1, adj = 0.87, cex = 0.5)
```



Hide

```
pdf("./plot/Admixture.pdf", width = 18, height = 5)
gap <- (0.845-0.04)/33
par(mar = c(2, 2, 2, 2))
barplot(t(q_k5), border = NA, space = 0, col = colors,
        ylab = "Ancestry proportions", xlab = "Individuals",
        main = "Ancestry matrix for K = 5 groups", xaxt = 'n')
abline(v = c(0, sample_size$xbreak), lwd = 2)
segments(0, 0, 1931, 0, lwd = 3)
segments(0, 1, 1931, 1, lwd = 3)
for (i in 2:35){
  mtext(as.character(sample_size$state[i]), 1, adj = 0.04 + (i-2)*gap, cex = 0.5)
}
mtext("MS", 1, adj = 0.87, cex = 0.5)
dev.off()
```

```
null device
          1
```

Hide

```
save.image(file = "./SNP_analysis.rdata")
```