HW6

Tianqi Wu

3/5/2020

Problem 1

Section 1: Data wrangling

Import analysis1.assoc.logistic.

```
library(tidyverse)
data = read.table('analysis1.assoc.logistic',header = TRUE)
```

Remove all rows that don't correspond to testing the SNP effect (remove rows where the TEST column does not say ADD)

```
data_filter = data %>% filter(TEST=='ADD')
```

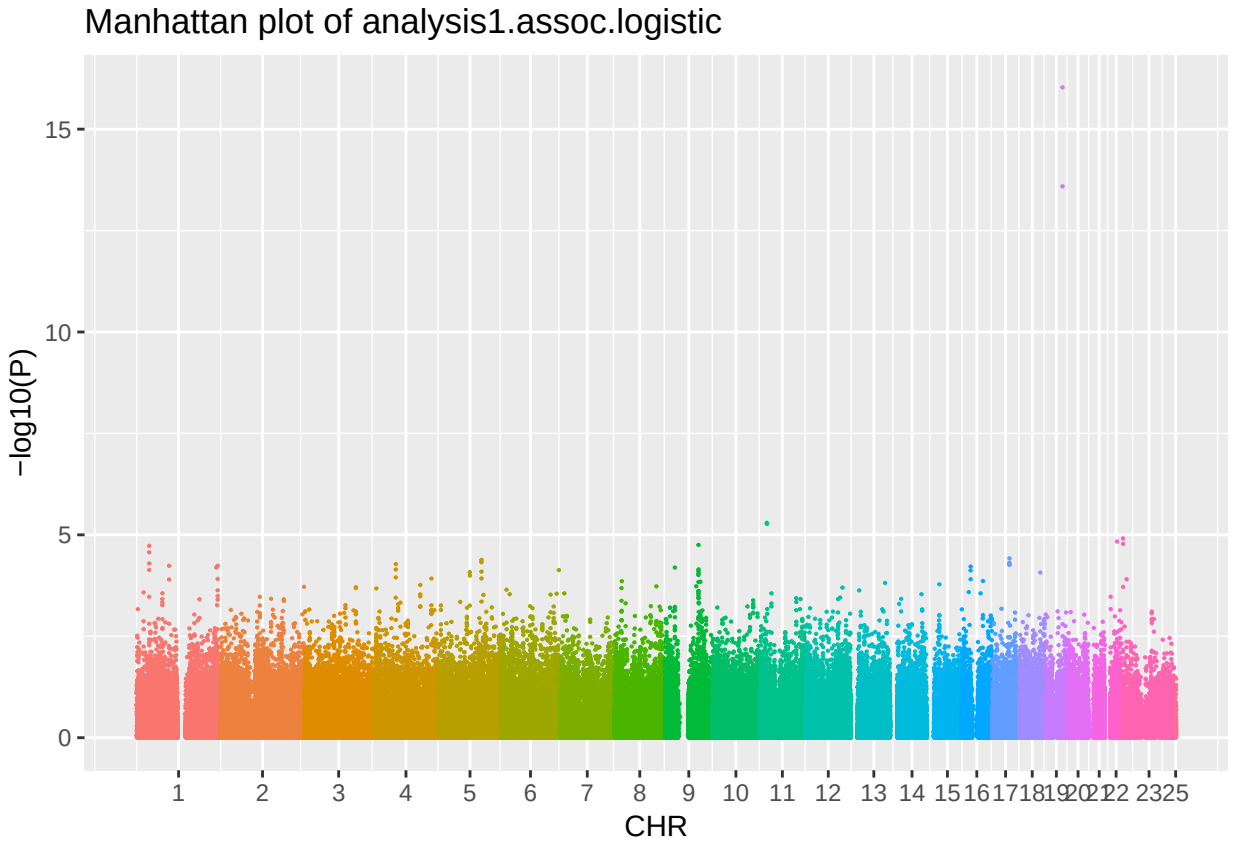
Section 2: Data visualization

```
## prepare data for ggplot
data_manh = data_filter %>%
  group_by(CHR) %>%
  summarize(BP_max=max(BP)) %>%
  mutate(BP_cum = cumsum(BP_max)-BP_max) %>%
  left_join(data_filter,by='CHR') %>%
  mutate(BP_pos = BP_cum+BP)
```

Write your own function using ggplot2 that plots a Manhattan plot using the filtered data from Section 1, Make the colors of the points alternate across chromosomes, or give each chromosome a different color

```
CHR_axis = data_manh %>%
  group_by(CHR) %>%
  summarize(BP_center=(max(BP_pos)+min(BP_pos))/2)

ggplot(data_manh,aes(BP_pos,-log10(P)))+
  geom_point(aes(color=as.factor(CHR)),size=0.1,show.legend = FALSE)+
  scale_x_continuous(label = CHR_axis$CHR, breaks= CHR_axis$BP_center)+xlab('CHR') +
  ggtitle('Manhattan plot of analysis1.assoc.logistic')
```



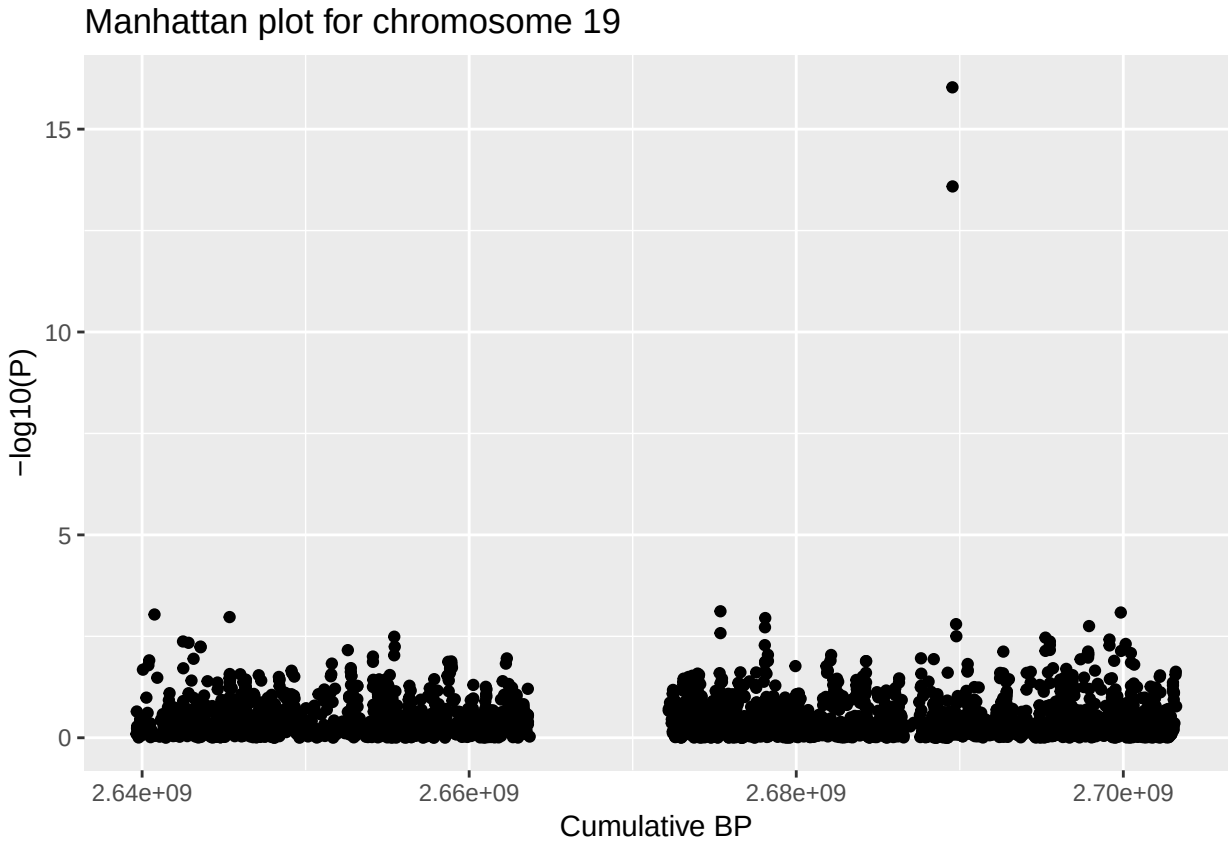
Describe the rationale behind your code in narrative prose.

First I keep all the rows where $TEST =' ADD'$ since they are the rows related to SNP effect. For the visualization, since we only have the physical position of base-pair(BP) for each chromosome, we need to combine them together so that it can be displayed in the manhattan plot. To do that, we need to find the max BP position of each chromosome and add them cumulatively. Then, adding the cumulative BP position to its original BP position would concatenate the BP position of all chromosome together. This step makes the base-pair position of all the chromosome connected one by one. Also, in order to display a scale of x-axis with CHR, we need to find the center BP position of each chromosome. Finally, we can generate the manhattan plot with the information that we get.

Section 3: Zoom in on chromosome 19

Use your code to plot a Manhattan plot only for SNPs on chromosome 19.

```
chromo_19 = filter(data_manh,CHR=='19')
ggplot(chromo_19,aes(BP_pos,-log10(P)))+
  geom_point() +xlab('Cumulative BP') + ggtitle('Manhattan plot for chromosome 19')
```


Manhattan plot for chromosome 19
15
10
5
0
−log10(P)
2.64e+09
2.66e+09
2.68e+09
2.70e+09
Cumulative BP