# Thesis Venn Code

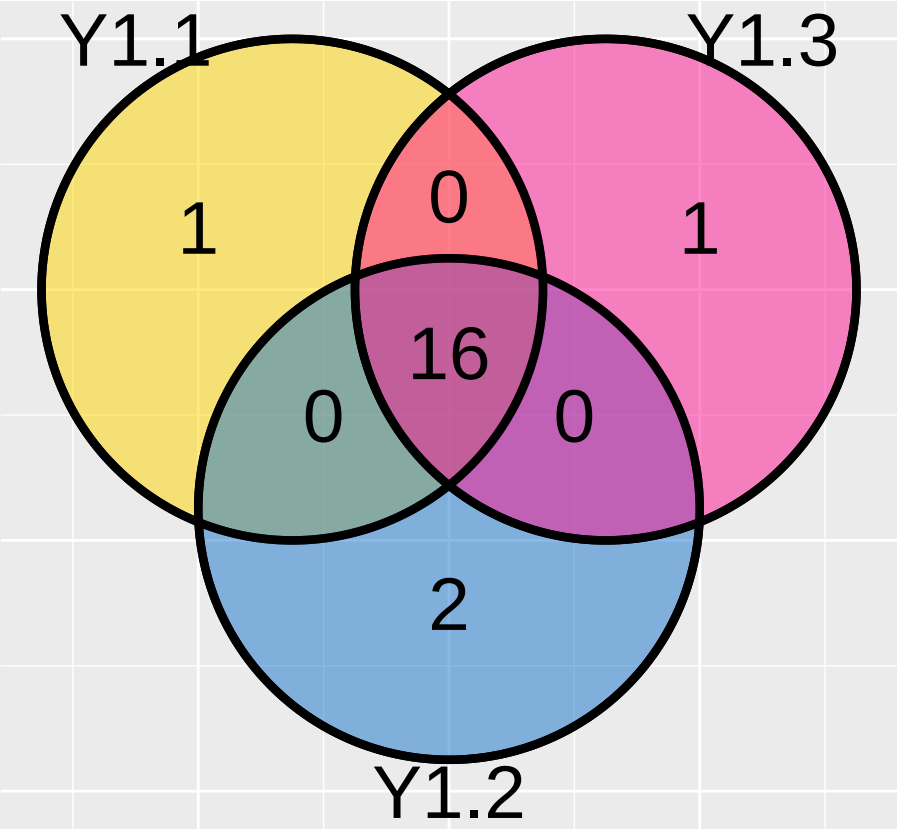
Max Barclay

2023-01-30

```
#Reads the SNP csv files for each strains snippy
Y1_1 <- read.csv(file="R_data/Y1.1.csv", header=TRUE) %>% filter(FTYPE != "")
Y1_2 <- read.csv(file="R_data/Y1.2.csv", header=TRUE) %>% filter(FTYPE != "")
Y1_3 <- read.csv(file="R_data/Y1.3.csv", header=TRUE) %>% filter(FTYPE != "")

#Puts strain csv's into data frames
Y1_1Df <- data.frame(Y1_1)
Y1_2Df <- data.frame(Y1_2)
Y1_3Df <- data.frame(Y1_3)

#Filters the strain data frames by the chromosomal position; column 2
Y1_1Df_pos <- Y1_1Df[,2, drop=TRUE]
Y1_2Df_pos <- Y1_2Df[,2, drop=TRUE]
Y1_3Df_pos <- Y1_3Df[,2, drop=TRUE]
```

Takes snippy csv output from each strain and filters to remove N/A entries. Then these csv files are put into data frames to be filtered by chromosomal position to ensure multiple SNPs for a gene are not inflating the gene mutation number.

```
#Puts all three strain chromosomal SNPS position values into one list
POS_list <- list(
  Y1.1 = Y1_1Df_pos,
  Y1.2 = Y1_2Df_pos,
  Y1.3 = Y1_3Df_pos
)

#Creates a venn diagram of the three strains
ggvenn(Venn(POS_list[1:3]))
```




The three strains data frames based on chromosomal position were merged into a list and then a venn diagram was created using the list values.