# BUSTED_2x2

*sadie*

*2/3/2020*

load libraries

read in data

```
#add rate category count and order and gene for each file (can be found in file name FILE)

mtDNA_SRV_2x2_2_3_2020 <- read_csv("~/bin/mtDNA_redo/data/mtDNA_SRV_2x2_2_3_2020")
```

```
## Parsed with column specification:
## cols(
##   FILE = col_character(),
##   Sites = col_double(),
##   Sequences = col_double(),
##   BUSTED.SRV.LR = col_double(),
##   BUSTED.SRV.UNLogL = col_double(),
##   CV.SRV = col_double(),
##   CV.NSRV = col_double(),
##   BUSTED.SRV.P = col_double(),
##   BUSTED.SRV.AICc = col_double(),
##   BUSTED.SRV.treelength = col_double(),
##   srv.omega.1.rate = col_double(),
##   srv.omega.2.rate = col_double(),
##   srv.omega.1.prop = col_double(),
##   srv.omega.2.prop = col_double(),
##   srv.alpha.1.rate = col_double(),
##   srv.alpha.2.rate = col_double(),
##   srv.alpha.1.prop = col_double(),
##   srv.alpha.2.prop = col_double()
## )
```

```
mtDNA_SRV_2x2_2_3_2020 <- mtDNA_SRV_2x2_2_3_2020 %>%
  mutate(.,
         NS.rates = 2,
         S.rates =2,
         order = str_extract_all(mtDNA_SRV_2x2_2_3_2020$FILE, "\\w+(?=-)", simplify = T)[,1],
         gene = str_extract_all(mtDNA_SRV_2x2_2_3_2020$FILE, "\\w+(?=-)", simplify = T)[,2])

mtDNA_BUSTED_2x2_2_3_2020 <- read_csv("~/bin/mtDNA_redo/data/mtDNA_BUSTED_2x2_2_3_2020")
```

```
## Parsed with column specification:
## cols(
##   FILE = col_character(),
##   Sites = col_double(),
##   Sequences = col_double(),
##   BUSTED.LR = col_double(),
##   BUSTED.UNLogL = col_double(),
##   CV.NSRV = col_double(),
##   BUSTED.P = col_double(),
##   BUSTED.AICc = col_double(),
##   BUSTED.treelength = col_double(),
##   busted.omega.1.rate = col_double(),
##   busted.omega.2.rate = col_double(),
##   busted.omega.1.prop = col_double(),
##   busted.omega.2.prop = col_double()
## )
```

```
mtDNA_BUSTED_2x2_2_3_2020<- mtDNA_BUSTED_2x2_2_3_2020 %>%  mutate(., NS.rates = 2,
          S.rates = 2,
          order = str_extract_all(mtDNA_BUSTED_2x2_2_3_2020$FILE, "\\w+(?=-)", simplify = T)[,1],
            gene = str_extract_all(mtDNA_BUSTED_2x2_2_3_2020$FILE, "\\w+(?=-)", simplify = T)[,2])

#these are the orders used in the original analysis
orders_used <- read_delim("~/bin/mtDNA_redo/data/actual_orders_used.txt", delim = "\n", col_names = FALS
```

```
## Parsed with column specification:
## cols(
##   X1 = col_character()
## )
```

```
mtDNA_2x2 <- full_join(mtDNA_BUSTED_2x2_2_3_2020, mtDNA_SRV_2x2_2_3_2020, by = c("FILE", "Sites", "Sequ



#test_row <- bind_rows(mtDNA_BUSTED_2x2_2_3_2020, mtDNA_SRV_2x2_2_3_2020)
```

```
mtDNA_2x2$gene= toupper(mtDNA_2x2$gene)
mtDNA_2x2$order = toupper(mtDNA_2x2$order)

#fix some mispellings of order names
mtDNA_2x2$order[which(mtDNA_2x2$order == "CHIMAERIFORMS")]= "CHIMAERIFORMES"
mtDNA_2x2$order[which(mtDNA_2x2$order == "CARNIVORES")] <-"CARNIVORA"
mtDNA_2x2$order[which(mtDNA_2x2$order == "GASTEROSTEIFORMES")] <-"GASTEROSTEALES"
```

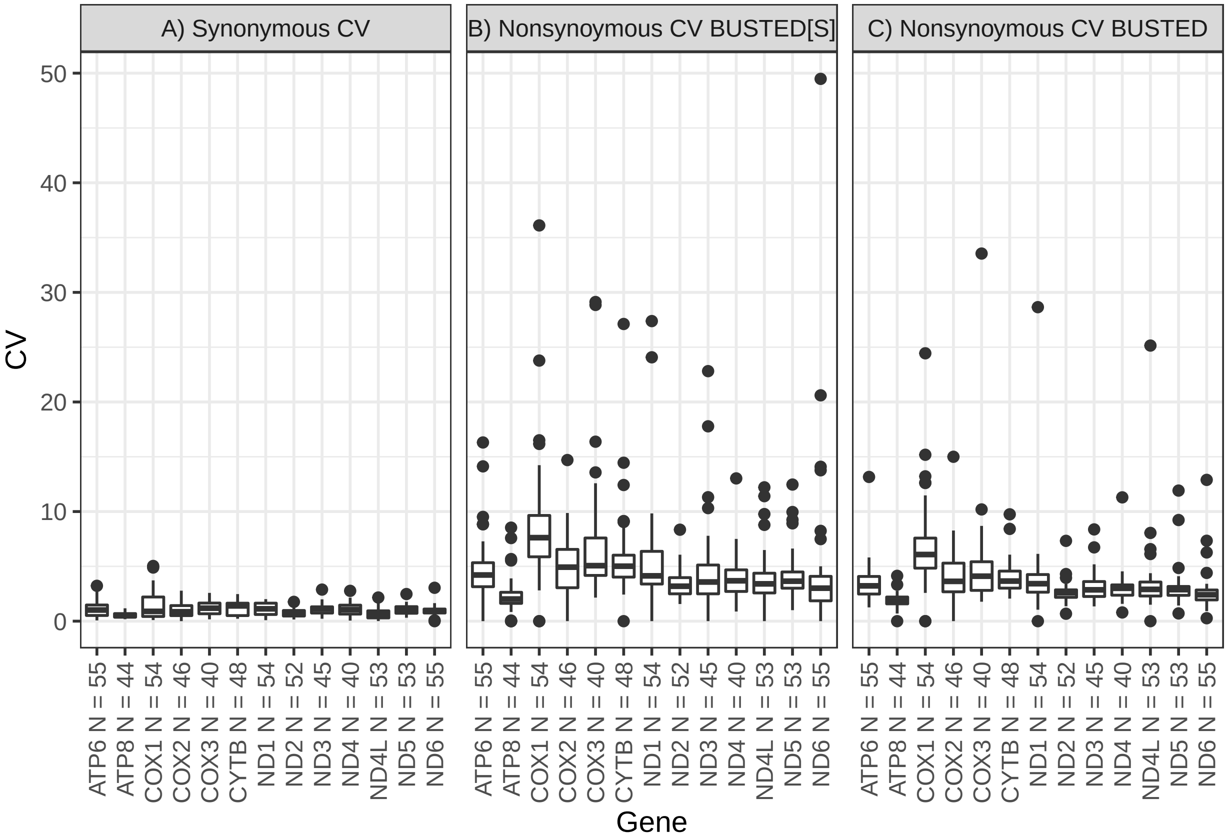
```
#filter based on orders previously used:
mtDNA_2x2 <- mtDNA_2x2 %>% filter(order %in% orders_used$X1)
```

```
syn_labels <- list("Synonymous.CV"="A) Synonymous CV",
                   "NS.CV" = "B) Nonsynoymous CV BUSTED[S]",
                   "CV.NSRV.busted" = "C) Nonsynoymous CV BUSTED")

syn_labeller <- function(variable,value){
  return(syn_labels[value])
}
```

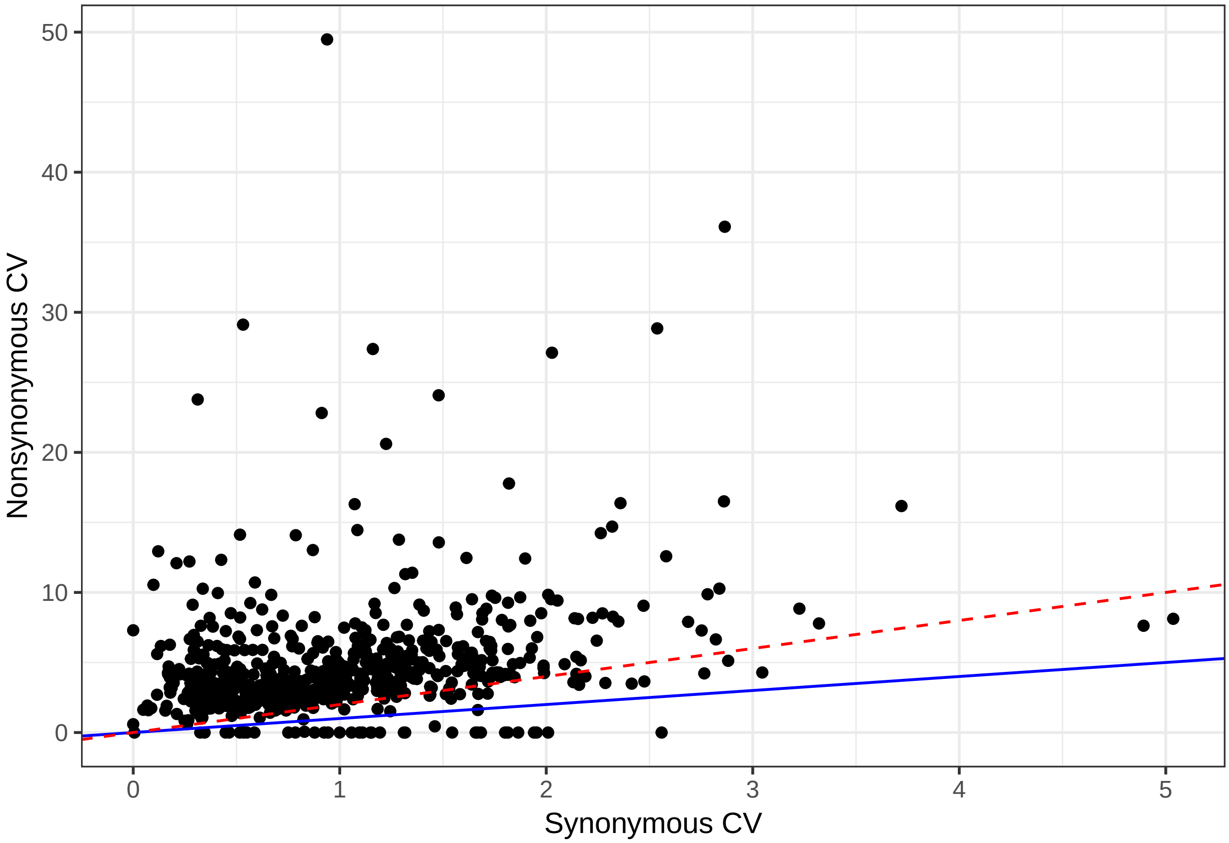
boxplots of the CVs grouped by genes

```
num_orders_per_gene = mtDNA_2x2  %>% count(gene)
gene_boxplots <- mtDNA_2x2 %>% select(CV.SRV, CV.NSRV.srv, CV.NSRV.busted,gene)
gene_boxplots <-gene_boxplots %>% melt(id.vars = "gene")




gene_boxplots %>%ggplot(aes(gene, value))+
  geom_boxplot()+ facet_grid(~variable,labeller = syn_labeller)+
  #coord_cartesian(ylim = c(0,3.5))+
  ylab("CV")+xlab("Gene")+ theme_bw()+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+
  scale_x_discrete(labels = paste(num_orders_per_gene$gene, num_orders_per_gene$n, sep = " N = "))
```

```
## Warning: The labeller API has been updated. Labellers taking `variable`and
## `value` arguments are now deprecated. See labellers documentation.

## Warning: Removed 2 rows containing non-finite values (stat_boxplot).
```



```
mtDNA_2x2 %>% ggplot()+geom_point(aes(CV.SRV, CV.NSRV.srv))+ xlab("Synonymous CV")+
  ylab("Nonsynonymous CV")+ theme_bw()+
  geom_abline(slope = 1, intercept = 0, color = 'blue') +
  geom_abline(slope = 2, intercept = 0,color='red', linetype = "dashed" )
```



```
#+
 # coord_cartesian(ylim = c(0,3.5), xlim = c(0,1.65))
```

```
source("/Volumes/GoogleDrive/My Drive/BUSTED-SRV/R/useful_functions.R")
gen.sig.table(mtDNA_2x2)
```

```
## Loading required package: xtable
```

```
##                 BUSTED-SRV
## BUSTED           No Selection  Selection
##   No Selection    0.89671362 0.03286385
##   Selection       0.05007825 0.02034429
```