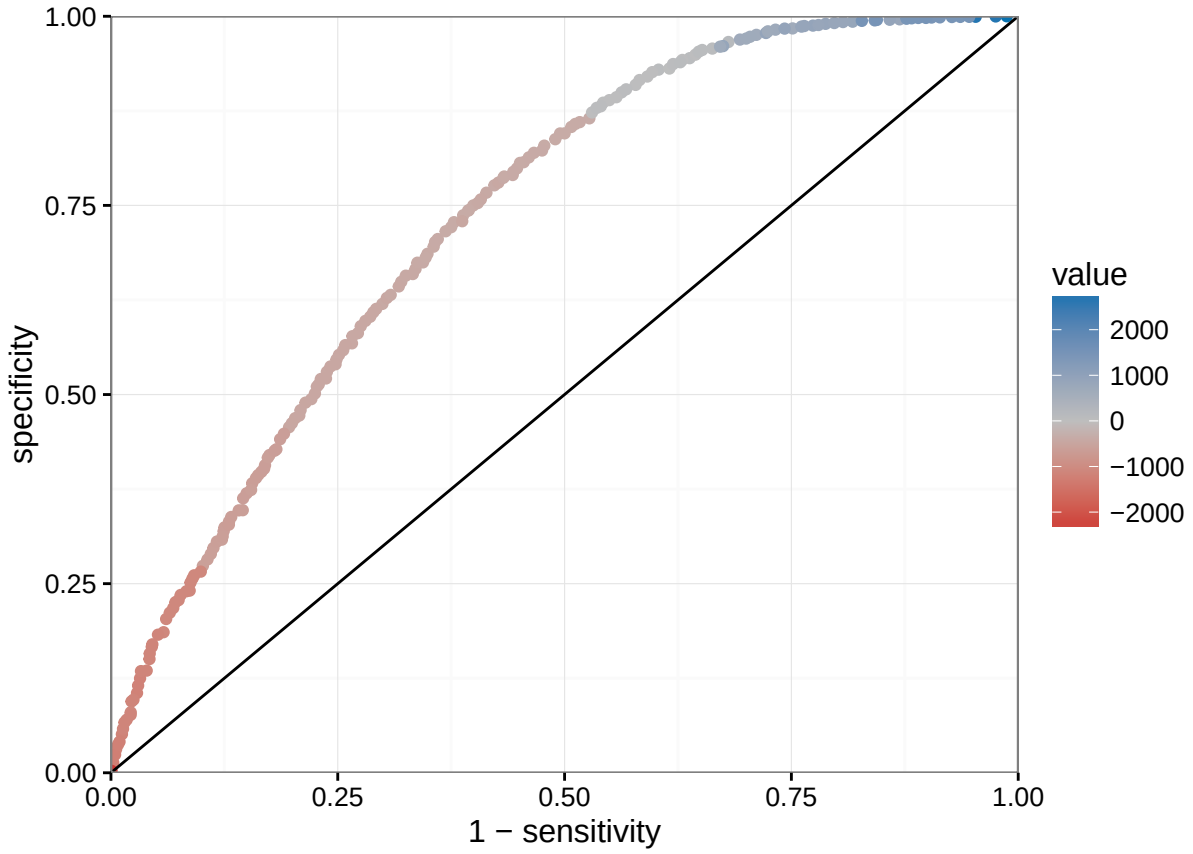
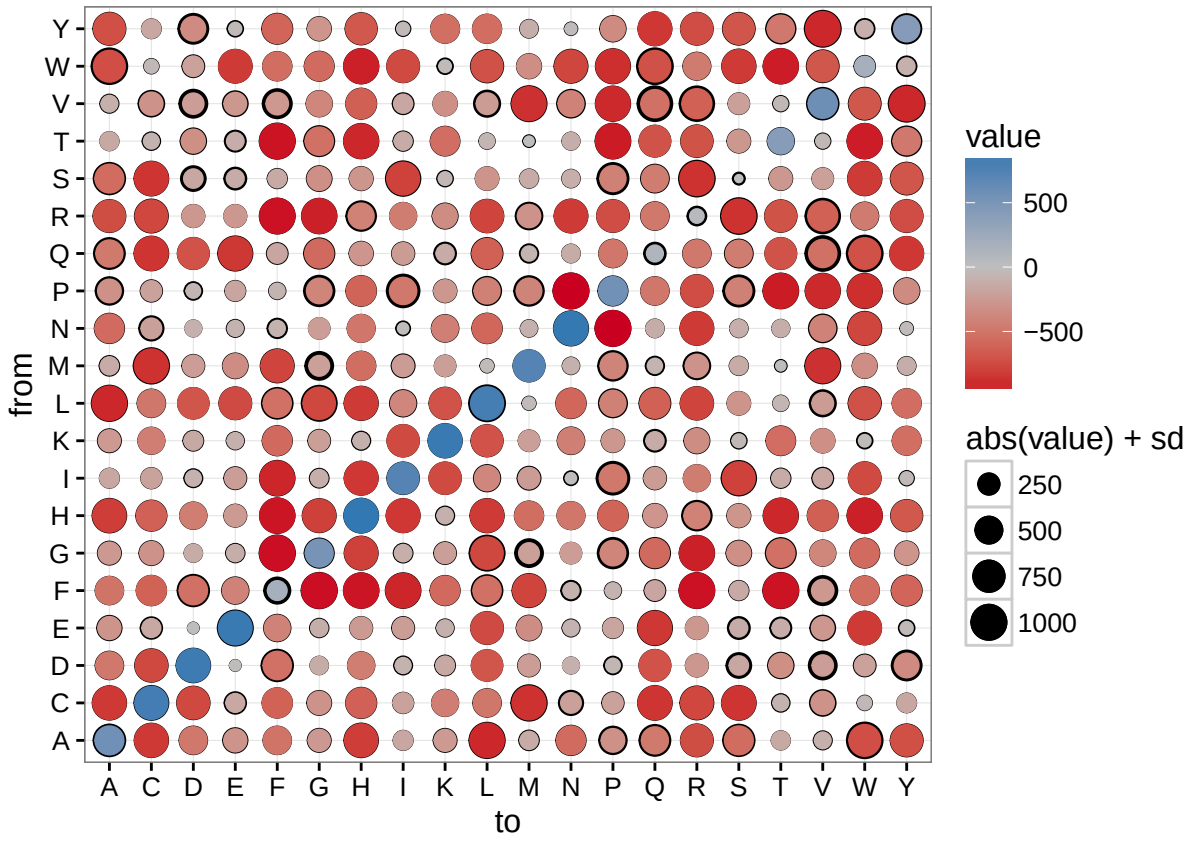
VDJdb summary statistics

```
library(ggplot2)
library(plyr)
library(reshape2)

df <- read.table("roc.txt", header=T, sep="\t")
df$sensitivity <- -df$sensitivity
df$specificity <- -df$specificity

df.2 <- read.table("solutions.txt", header=T,sep="\t")

df.3 <- merge(df, subset(df.2, parameter == "threshold"))

ggplot(df.3, aes(x=1-sensitivity, y=specificity)) +
  geom_point(aes(color = value)) +
  geom_abline(slope=1, intercept=0) +
  scale_x_continuous(limits=c(0,1), expand=c(0,0)) +
  scale_y_continuous(limits = c(0,1), expand=c(0,0)) +
  scale_color_gradient2(low = "#ca0020", mid="grey", high = "#0571b0", midpoint = 0) +
  theme_bw()
```




```
df.s <- subset(df.2, parameter=="substitution")
df.s <- ddply(df.s, .(from, to), summarize,
              sd = sd(value), value = mean(value))
```

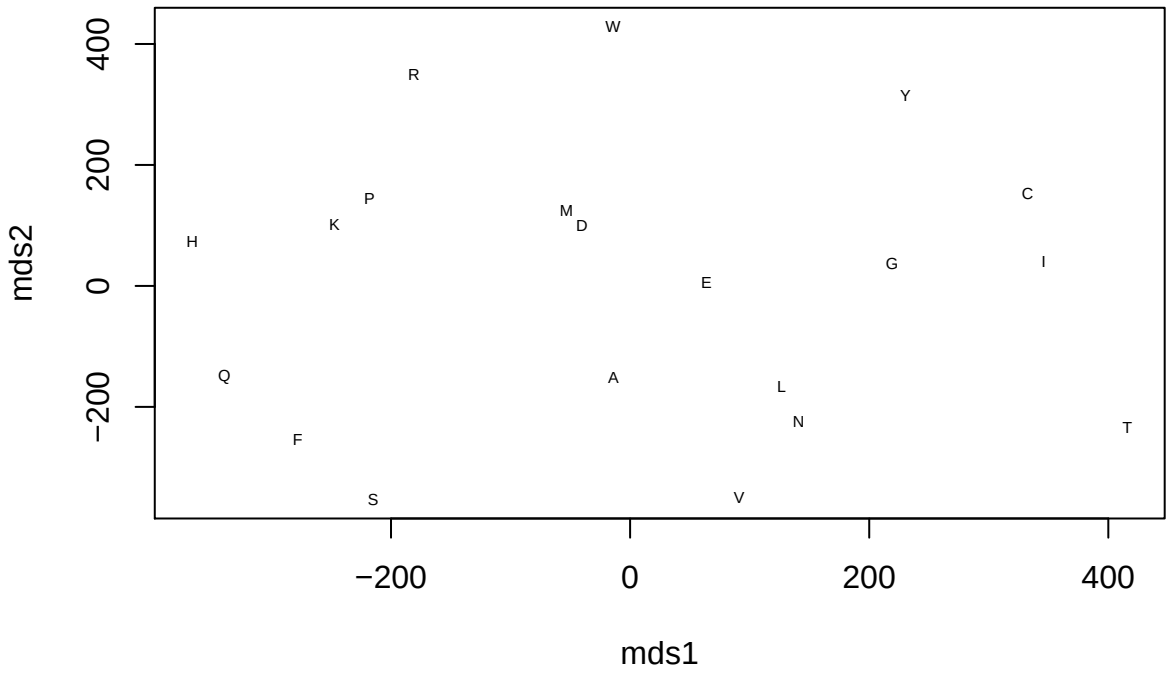
```
ggplot(df.s, aes(x=to, y=from)) +
  geom_point(aes(size=abs(value) + sd), color="black") + geom_point(aes(color=value, size=abs(value))) +
  scale_color_gradient2(low = "#ca0020", mid="grey", high = "#0571b0", midpoint = 0) + theme_bw()
```




```
library(MASS)
df.ss <- dcast(df.s, from ~ to)
rnames <- df.ss$from
df.ss$from <- NULL
df.ss <- as.matrix(df.ss)
rownames(df.ss) <- rnames

cont_aa <- c("A","D","G","L","N","R","S","T","Y")

mds <- isoMDS(-as.dist(df.ss), k = 2)
```

```
## initial  value 32.743811
## final  value 32.743506
## converged
```

```
x <- mds$points[,1]
y <- mds$points[,2]
plot(x, y, xlab="mds1", ylab="mds2", type = "n")
text(x, y, labels = rownames(df.ss), cex=.5)#col = cc_final,
```




```
df.pw <- subset(df.2, parameter=="position_weight")

ggplot(df.pw) + geom_line(aes(x=as.numeric(from) - 6, group=id, y=value), color="blue", alpha=0.3) +
  geom_vline(xintercept = 0, linetype="dashed", color="red") +
  scale_x_continuous(limits = c(-5,5), breaks=-5:5) + theme_bw()
```