

A. BoxPlot for .632 errors

```
df<-read.table("632",sep="\t",header=TRUE)
library(plyr)
library(ggplot2)
plot_Data <-ddply(df,.(dfsType, RF, survival),
mutate,med=median(error),min=min(error),max=max(error),Q1=quantile(error,
1/4),Q3=quantile(error, 3/4), IQR=Q3-Q1, upper.limit=Q3+1.5*IQR, lower.limit=Q1-1.5*IQR)
p <-ggplot(data = plot_Data, aes(x = dfsType))
p <-p + geom_boxplot(aes(lower = Q1, upper = Q3, middle = med, ymin = min, ymax = max), stat=
"identity", colour="blue", outlier.colour = "black", outlier.shape = 16, outlier.size = 16)
p <-p +geom_point(data=plot_Data[plot_Data$error > plot_Data$upper.limit |plot_Data$error <
plot_Data$lower.limit,], aes(y=error))
p <-p + facet_grid( RF ~ survival, scales="free", space="free")
p <-p + theme(plot.title = element_text("0.632 error"),axis.text.x = element_text(angle = 90, hjust =
1, size = 8, colour = "grey50"),plot.title = element_text(face="bold", size=11),axis.title.x =
element_text(face="bold", size=9),axis.title.y = element_text(face="bold", size=9,
angle=90),panel.grid.major = element_blank(),panel.grid.minor = element_blank())
p <-p + scale_fill_hue(c=45, l=80)
ggsave("error.632.png",plot=p,dpi=600)
```

Snippet of 632 file:

iteration	RF	dfsType	survival	error
51	SM	high	disease-free survival	0.7857142857
52	SM	high	disease-free survival	0.7857142857
53	SM	high	disease-free survival	0.7857142857
54	SM	high	disease-free survival	0.7857142857
55	SM	high	disease-free survival	0.7142857143
101	SM	high	disease-free survival	0.7857142857
201	SM	high	disease-free survival	0.7857142857
202	SM	high	disease-free survival	0.7857142857
251	SM	high	disease-free survival	0.7857142857
252	SM	high	disease-free survival	0.7857142857
253	SM	high	disease-free survival	0.7857142857
254	SM	high	disease-free survival	0.7857142857
255	SM	high	disease-free survival	0.7857142857
256	SM	high	disease-free survival	0.7857142857
257	SM	high	disease-free survival	0.7142857143
258	SM	high	disease-free survival	0.7857142857
259	SM	high	disease-free survival	0.7857142857
260	SM	high	disease-free survival	0.7857142857
261	SM	high	disease-free survival	0.7857142857
262	SM	high	disease-free survival	0.7857142857
263	SM	high	disease-free survival	0.7857142857
301	SM	high	disease-free survival	0.7857142857
302	SM	high	disease-free survival	0.7857142857
303	SM	high	disease-free survival	0.6428571429
401	SM	high	disease-free survival	0.7857142857
402	SM	high	disease-free survival	0.7857142857
403	SM	high	disease-free survival	0.7857142857
404	SM	high	disease-free survival	0.8571428571

```

405 SM high disease-free survival 0.7857142857
406 SM high disease-free survival 0.7857142857
407 SM high disease-free survival 0.7857142857
408 SM high disease-free survival 0.7857142857

```

B. Probability estimation by log-rank test for Kaplan Meier (KM) disease-free survival distributions

In the below script, temp is a tab-delimited file, where the rows correspond to patient samples and the first column is the disease-free survival, second is the alive (1)/dead (0) binary status, and third onwards, various test conditions evaluated. Replace X below with the column number of the test condition being evaluated.

```

paste -d"\t" <(cut -f1 temp) <(cut -f2 temp) <(cut -fX temp) | awk '{if ($3 ~ /[12]/) print}'> T
echo -e "library(survival)"
cat <(paste -d"\t" <(echo -n "time<-c") <(cut -f1 T | tr '\n' ',') | sed 's/,$/)/g') <(paste -d"\t" <(echo -n
"status<-c") <(cut -f2 T | tr '\n' ',') | sed 's/,$/)/g') <(paste -d"\t" <(echo -n "treatment<-c") <(cut -f3 T
| tr '\n' ',') | sed 's/,$/)/g')
echo -e "fit <-survdiff(Surv(time, status) ~ treatment)\nfit"

```

C. KM Plots (representative example)

```

library(survival)
time<-
c(14.10,12.90,13.93,19.50,0.97,88.30,30.73,0.77,3.97,2.50,7.03,3.77,1.60,5.90,14.17,1.07,12.20,53.
83,76.83,28.77,21.57,27.90)
status<-c(1,1,1,1,0,1,1,0,0,1,1,0,0,1,1,1,1,1)
treatment<-c(1,1,1,1,1,2,2,2,2,2,1,1,1,1,1,1,2,2,2,2)
fit <-survdiff(Surv(time, status) ~ treatment)
fit

time1<-c(14.10,12.90,13.93,19.50,0.97,7.03,3.77,1.60,5.90,14.17,1.07,12.20)
status1<-c(1,1,1,1,0,1,1,0,0,1,0,0)
treatment1<-c(1,1,1,1,1,1,1,1,1,1,1)
time2<-c(88.30,30.73,0.77,3.97,2.50,53.83,76.83,28.77,21.57,27.90)
status2<-c(1,1,0,0,0,1,1,1,1,1)
treatment2<-c(2,2,2,2,2,2,2,2,2,2)
fit1 <-survfit(Surv(time1, status1) ~ 1)
fit2 <-survfit(Surv(time2, status2) ~ 1)
png("kmPlot.png")
plot(fit1,conf.int="none",col='blue',xlab='disease-free survival (months)',ylab= 'Cumulative
Survival Probability',xlim=c(0,60))
lines(fit2,conf.int="none",col='red')
legend(25,1,c('(TP53.NOTCH2.11p)WT [12]', '(TP53)M NOTCH2(M) 11p(L) [10]'),col=
c('blue','red'),bty='n',lty=1)
text(35,0.85,"pval=0.000175")
title(main='with HPV infection: OTSCC')
dev.off()

```