Survival Analysis of OV with Taxol, Taxotere, Cisplatin

This is for survival analysis of OV data from TCGA.

Date: 05/12/2015

450 patients have been both survival data and drug sensitivity prediction.

Time measured by month.

```
##
## Attaching package: 'boot'
##
## The following object is masked from 'package:survival':
##
##     aml
```

1. Load Data:

```
load("./surv.lv3.pred.rda")
dim(surv)
```

```
## [1] 450  16
```

```
surv$carboplatin.predict = surv$cisplatin.predict
surv$carboplatin.level = surv$cisplatin.level
head(surv)
```

```
##              contact.days death.days days death months taxol taxotere
## TCGA-25-2042          396        396  396  TRUE  13.01  TRUE    FALSE
## TCGA-25-2399          608        608  608  TRUE  19.98  TRUE    FALSE
## TCGA-29-2427         1900         NA 1900 FALSE  62.42  TRUE    FALSE
## TCGA-59-2355           NA         65   65  TRUE   2.14  TRUE    FALSE
## TCGA-59-2363          165         NA  165 FALSE   5.42  TRUE    FALSE
## TCGA-25-2391         1492       1492 1492  TRUE  49.02  TRUE     TRUE
##              carboplatin cisplatin other taxol.predict taxotere.predict
## TCGA-25-2042        TRUE     FALSE  TRUE     -2.455937        -5.514242
## TCGA-25-2399        TRUE     FALSE  TRUE     -3.603826        -5.551854
## TCGA-29-2427        TRUE      TRUE  TRUE     -2.493162        -5.089083
## TCGA-59-2355        TRUE     FALSE FALSE     -3.449695        -5.975291
## TCGA-59-2363        TRUE     FALSE FALSE     -3.116783        -4.609608
## TCGA-25-2391        TRUE     FALSE FALSE     -2.985817        -5.385037
##              cisplatin.predict taxol.level taxotere.level cisplatin.level
## TCGA-25-2042          3.744354         low           high             low
## TCGA-25-2399          3.207412        high           high          medium
## TCGA-29-2427          3.616082         low         medium          medium
## TCGA-59-2355          3.275449        high           high          medium
## TCGA-59-2363          3.032010      medium         medium            high
## TCGA-25-2391          3.475698      medium           high          medium
```

```
##              carboplatin.predict carboplatin.level
## TCGA-25-2042            3.744354                low
## TCGA-25-2399            3.207412             medium
## TCGA-29-2427            3.616082             medium
## TCGA-59-2355            3.275449             medium
## TCGA-59-2363            3.032010               high
## TCGA-25-2391            3.475698             medium
```

```
qplot(taxol.level, taxol.predict, data=surv , geom=c("boxplot", "jitter"),
    fill=taxol.level,
    xlab="", ylab="Predicted Taxol IC50")
```
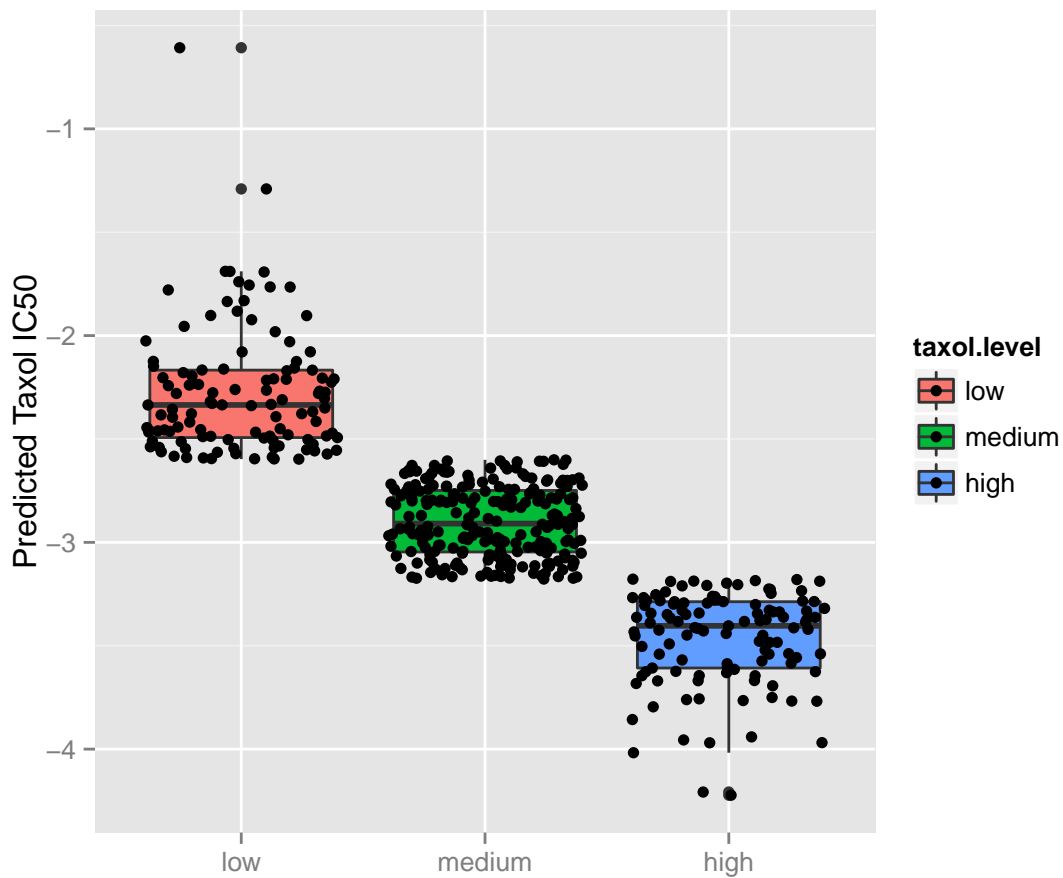

Figure 1:

```
qplot(taxotere.level, taxotere.predict, data=surv , geom=c("boxplot", "jitter"),
    fill=taxotere.level,
    xlab="", ylab="Predicted Taxotere IC50")
```

```
qplot(cisplatin.level, cisplatin.predict, data=surv , geom=c("boxplot", "jitter"),
    fill=cisplatin.level,
    xlab="", ylab="Predicted Cisplatin IC50")
```
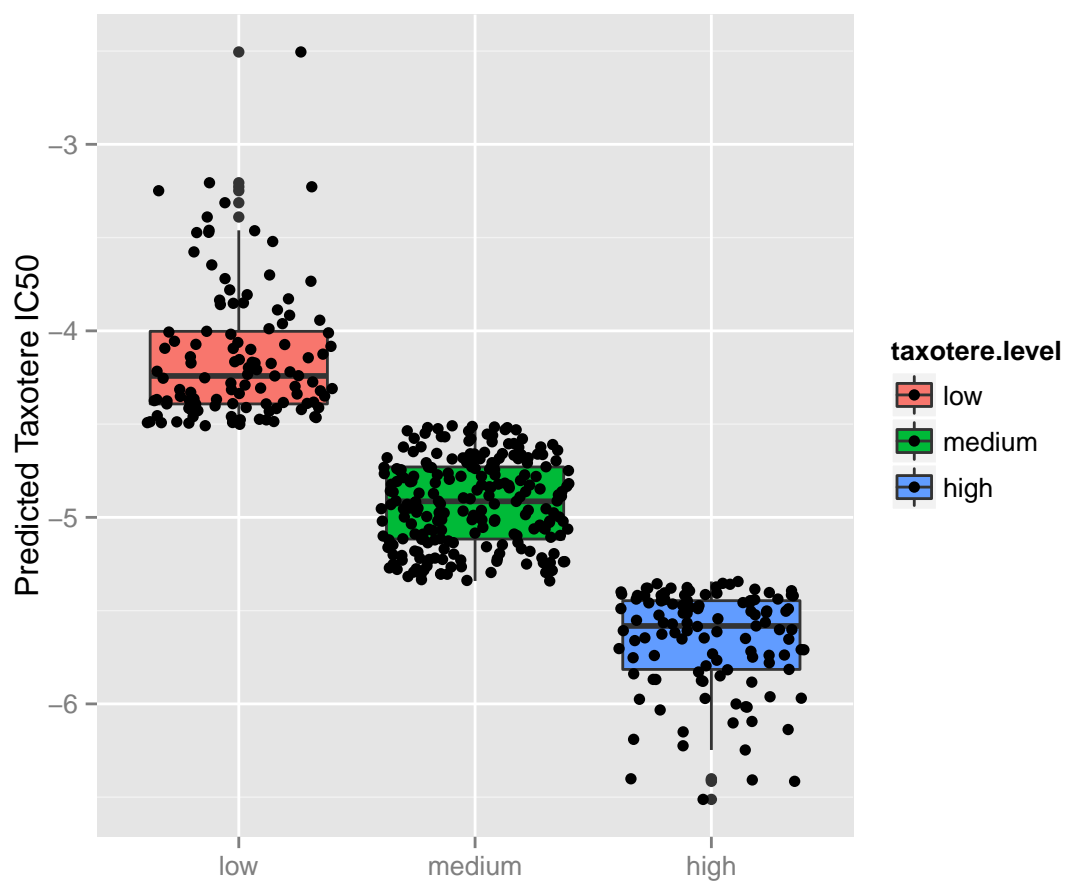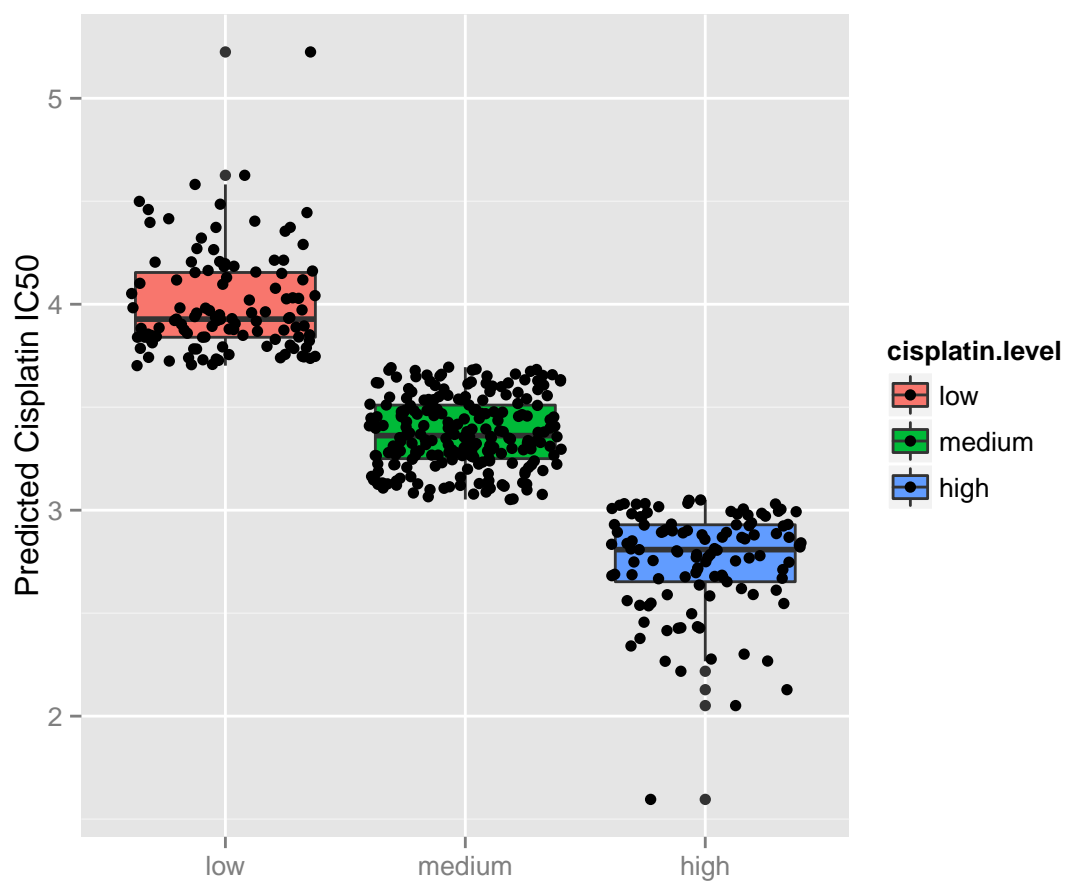
Figure 2:

Figure 3:

2. Correlation between taxol and taxotere, and cisplatin:

```
l = lm(surv$taxotere.predict ~ surv$taxol.predict)
plot(surv$taxol.predict, surv$taxotere.predict)
abline(l)
```
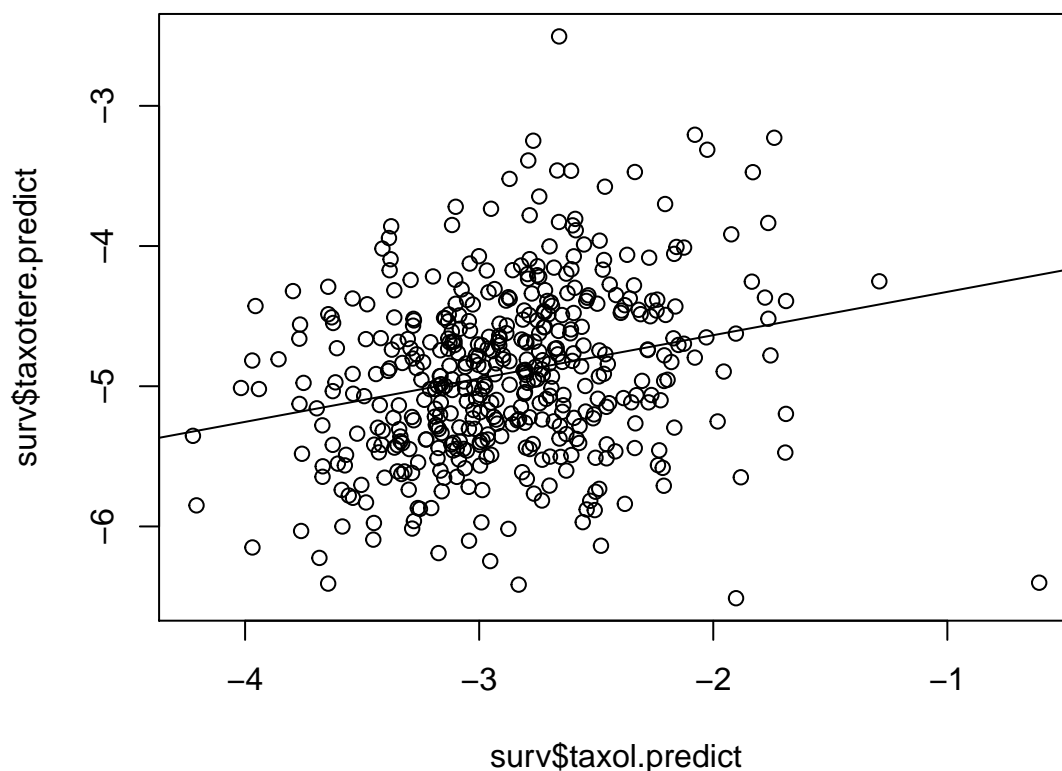

Figure 4:

```
print("IC50 prediction correlation b/w taxol and taxoltere")
```

```
## [1] "IC50 prediction correlation b/w taxol and taxoltere"
```

```
summary(lm(surv$taxol.predict ~ surv$taxotere.predict))$adj.r.squared
```

```
## [1] 0.05706013
```

```
print("IC50 prediction correlation b/w cisplatin and taxoltere")
```

```
## [1] "IC50 prediction correlation b/w cisplatin and taxoltere"
```

```
summary(lm(surv$cisplatin.predict ~ surv$taxotere.predict))$adj.r.squared
```

```
## [1] 0.1554508
```

```r
print("IC50 prediction correlation b/w taxol and cisplatin")
```

```
## [1] "IC50 prediction correlation b/w taxol and cisplatin"
```

```r
summary(lm(surv$taxol.predict ~ surv$cisplatin.predict))$adj.r.squared
```

```
## [1] 0.004594506
```

---

3. taxol:

```r
drug = "taxol"
data = GetMyData(drug)

  print("Month of Events vs IC50")
```

```
## [1] "Month of Events vs IC50"
```

```r
# the predicted sensitivity has a negative effect on survival time.
# However, the same trend is also seen in patients without such drug
# treatment, which suggest some bias/confounder effect not related to
# the drug treatment itself
  qplot(level, months, data=data[data$use, ] , geom=c("boxplot", "jitter"),
   fill=death, main="Month to Event ~ predicted IC50")
```

```r
    print("General Trend")
```

```
## [1] "General Trend"
```

```r
# General Trend were drawn with regression line and/or confidence
# intervals.  However, there is some bug in R MarkDown, and it does
# not shown.  You can manually draw it following code in Rmd file.
    ggplot(data, aes(x=IC50, y=months, fill=interaction(use, death),    color=interaction(use, death)))
```

```r
    print("Do deceased patients have weak sensitivity?")
```

```
## [1] "Do deceased patients have weak sensitivity?"
```

```r
# There is noticable increase of IC50 from live patients(green) to
# deceased ones (purple).  I didn't see a similar increase in patients
# who was not treated (from red to blue), which suggesting the observed
# effect could be real.  That's why I believe categorize by death is
# better than counting real number of months.
    qplot(death, IC50, data=data, geom=c("boxplot", "jitter"),
   fill=interaction(use, death), main="Predicted IC50 by Death of all patients",  ylab="Predicted IC50")
```
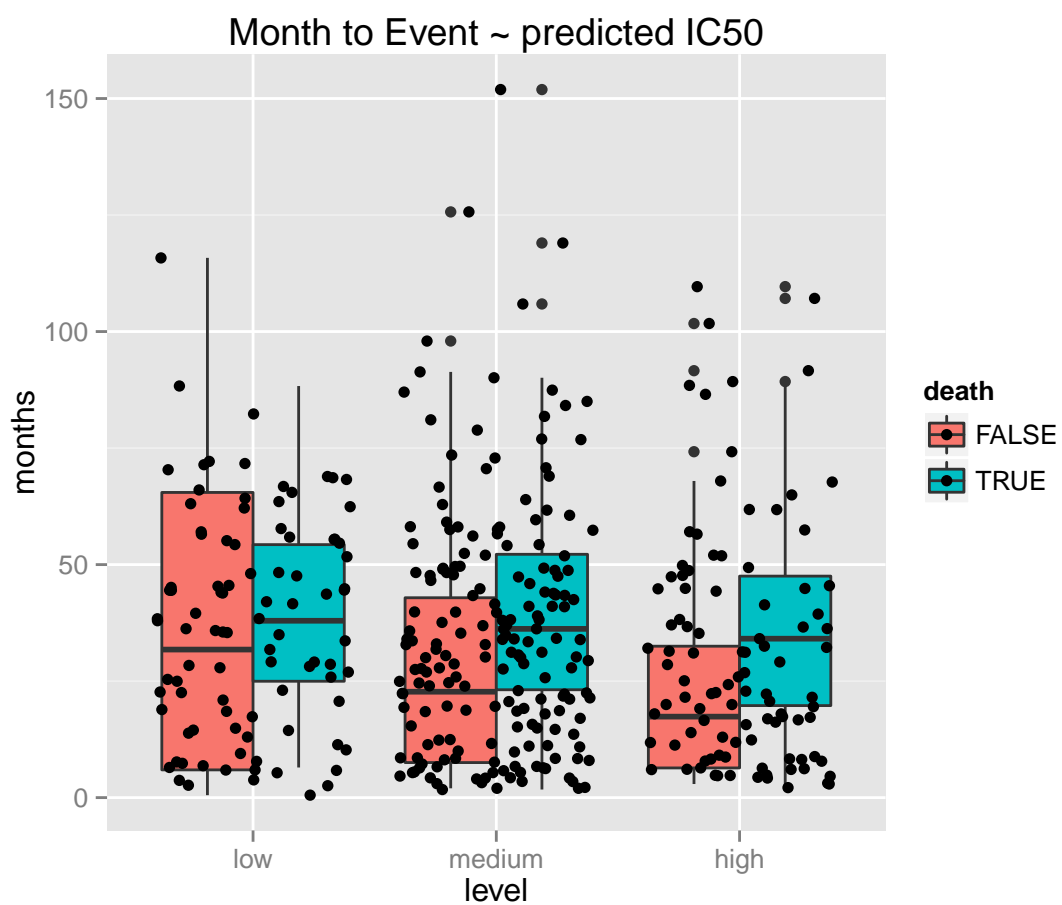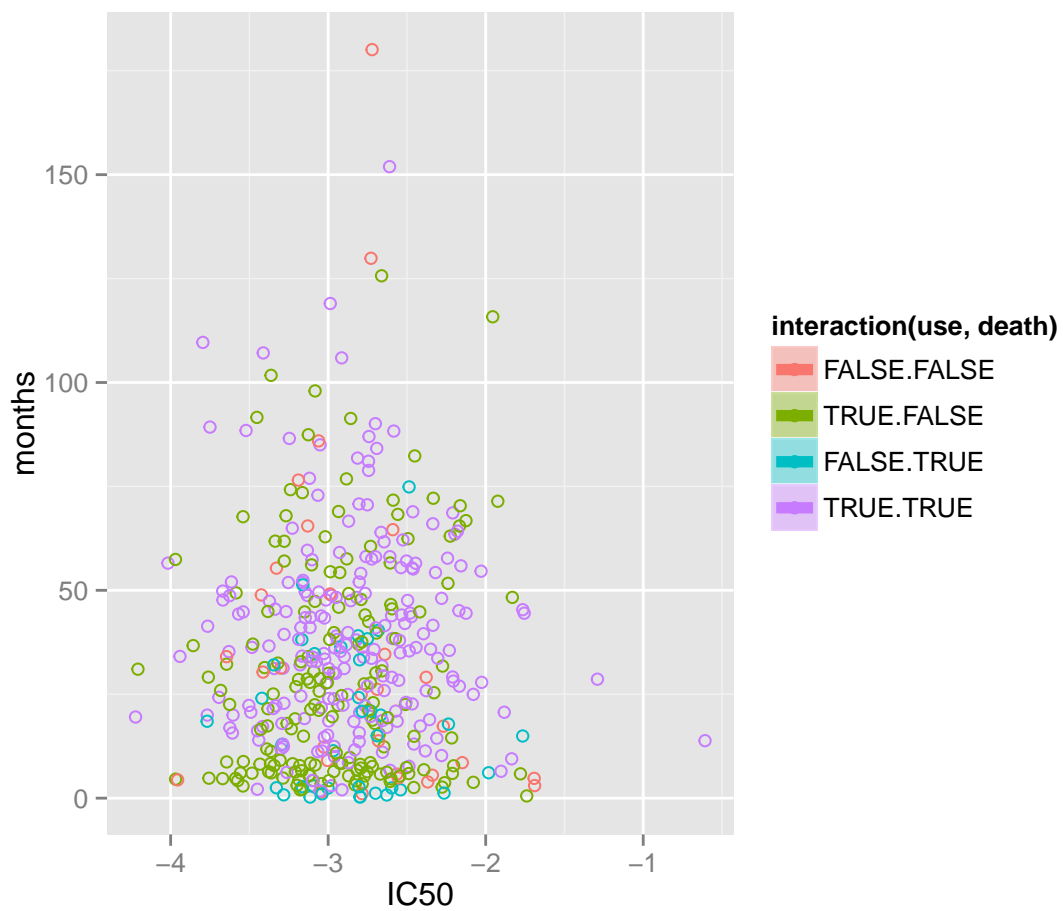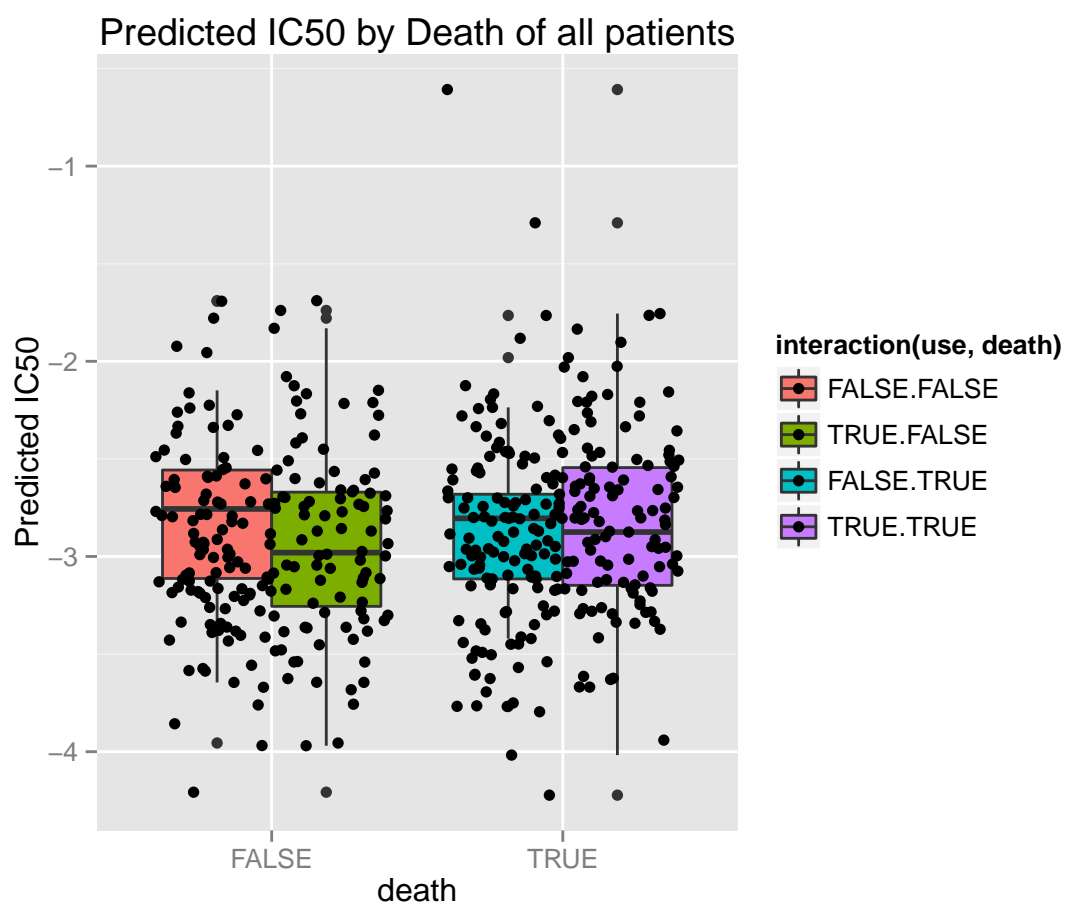
Figure 5:

Figure 6:

Figure 7:

```r
qplot(death, IC50, data=data[data$use, ], geom=c("boxplot", "jitter"),
  fill=death, main="Predicted IC50 by Death with patients who used the drug", ylab="Predicted IC50")
```
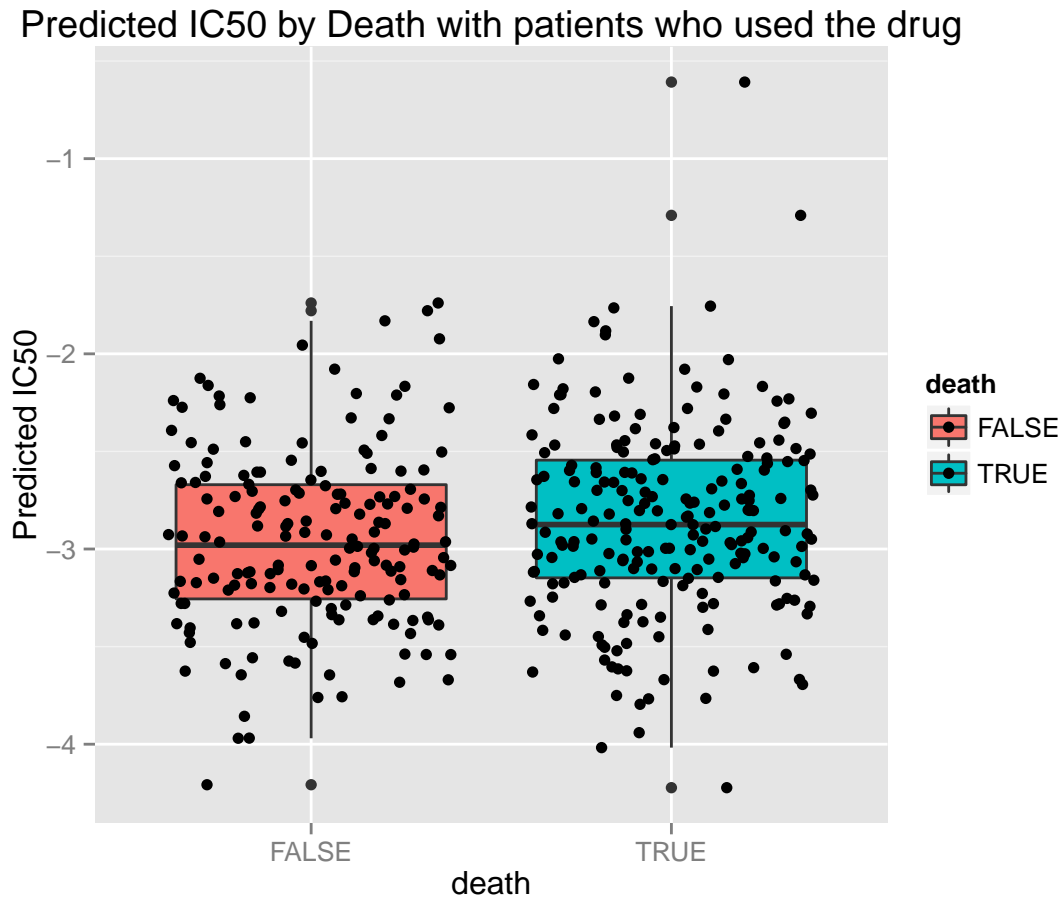

Figure 8:

```r
print("T-test for difference")
```

```
## [1] "T-test for difference"
```

```r
with(data=data[data$use, ], t.test(IC50[death], IC50[!death]))
```

```
##
##  Welch Two Sample t-test
##
## data:  IC50[death] and IC50[!death]
## t = 1.9914, df = 371.3, p-value = 0.04717
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.001216279 0.192623362
## sample estimates:
## mean of x mean of y
## -2.855512 -2.952432
```

```
    print("KM survival curves by predicted categories")
```

```
## [1] "KM survival curves by predicted categories"
```

```
    ggsurv(survfit(with(data[data$use, ], Surv(months,death) ~ level)))
```

```
## Loading required package: scales
```
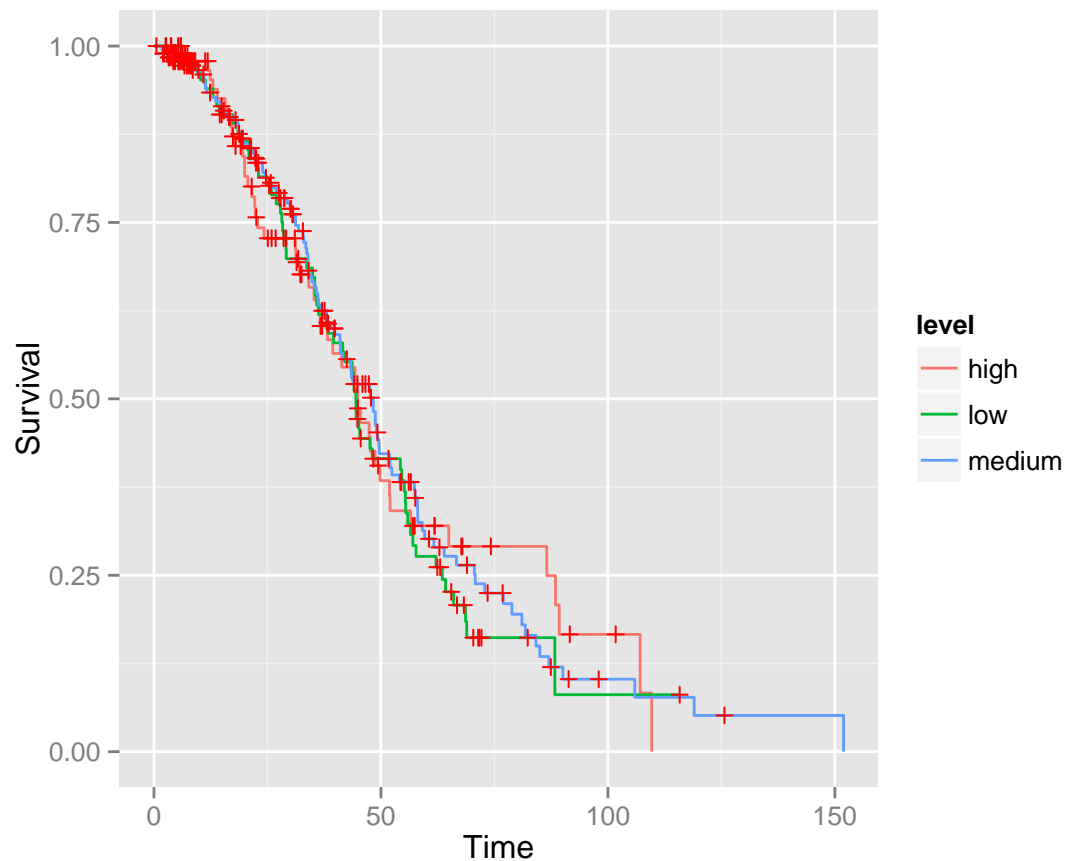

Figure 9:

```
# prediction of death using IC50 prediction and "other" drug usage
my.logit = glm(death ~ IC50 + other, data=data[data$use, ], family="binomial")
summary(my.logit)
```

```
##
## Call:
## glm(formula = death ~ IC50 + other, family = "binomial", data = data[data$use,
##     ])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7887  -0.9286   0.7162   0.8718   1.6660
##
```

```
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.9682     0.6873   1.409   0.1589
## IC50          0.5724     0.2350   2.436   0.0148 *
## otherTRUE     1.6226     0.2271   7.145 9.03e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 523.76  on 380  degrees of freedom
## Residual deviance: 464.37  on 378  degrees of freedom
## AIC: 470.37
##
## Number of Fisher Scoring iterations: 4
```

```r
cost = function(r, pi = 0) mean(abs(r-pi) > 0.5)
my.cv.err = cv.glm(data[data$use, ], my.logit, cost, K=10)
# delta, similar to prediction error
# major prediction power is in "other" drug use, and it has a negative
# effect on survival.
my.cv.err$delta[2]
```

```
## [1] 0.3141684
```

```r
# prediction using other only
my.logit2 = glm(death ~ other, data=data[data$use, ], family="binomial")
summary(my.logit)
```

```
##
## Call:
## glm(formula = death ~ IC50 + other, family = "binomial", data = data[data$use,
##     ])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7887  -0.9286   0.7162   0.8718   1.6660
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.9682     0.6873   1.409   0.1589
## IC50          0.5724     0.2350   2.436   0.0148 *
## otherTRUE     1.6226     0.2271   7.145 9.03e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 523.76  on 380  degrees of freedom
## Residual deviance: 464.37  on 378  degrees of freedom
## AIC: 470.37
##
## Number of Fisher Scoring iterations: 4
```

```
cost = function(r, pi = 0) mean(abs(r-pi) > 0.5)
my.cv.err = cv.glm(data[data$use, ], my.logit, cost, K=10)
# delta, similar to prediction error
# major prediction power is in "other" drug use, and it has a negative
# effect on survival
my.cv.err$delta[2]
```

```
## [1] 0.3060326
```

---

4. taxotere:

```
drug = "taxotere"
data = GetMyData(drug)

  print("Month of Events vs IC50")
```

```
## [1] "Month of Events vs IC50"
```

```
  qplot(level, months, data=data[data$use, ] , geom=c("boxplot", "jitter"),
    fill=death, main="Month to Event ~ predicted IC50")
```

```
## Warning in loop_apply(n, do.ply): Removed 1 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 1 rows containing missing values
## (geom_point).
```

```
  print("General Trend")
```

```
## [1] "General Trend"
```

```
    ggplot(data, aes(x=IC50, y=months, fill=interaction(use, death),    color=interaction(use, death)))
```

```
    print("Do deceased patients have weak sensitivity?")
```

```
## [1] "Do deceased patients have weak sensitivity?"
```

```
  qplot(death, IC50, data=data, geom=c("boxplot", "jitter"),
    fill=interaction(use, death), main="Predicted IC50 by Death of all patients",  ylab="Predicted IC50")
```

```
  qplot(death, IC50, data=data[data$use, ], geom=c("boxplot", "jitter"),
    fill=death, main="Predicted IC50 by Death with patients who used the drug",  ylab="Predicted IC50")
```
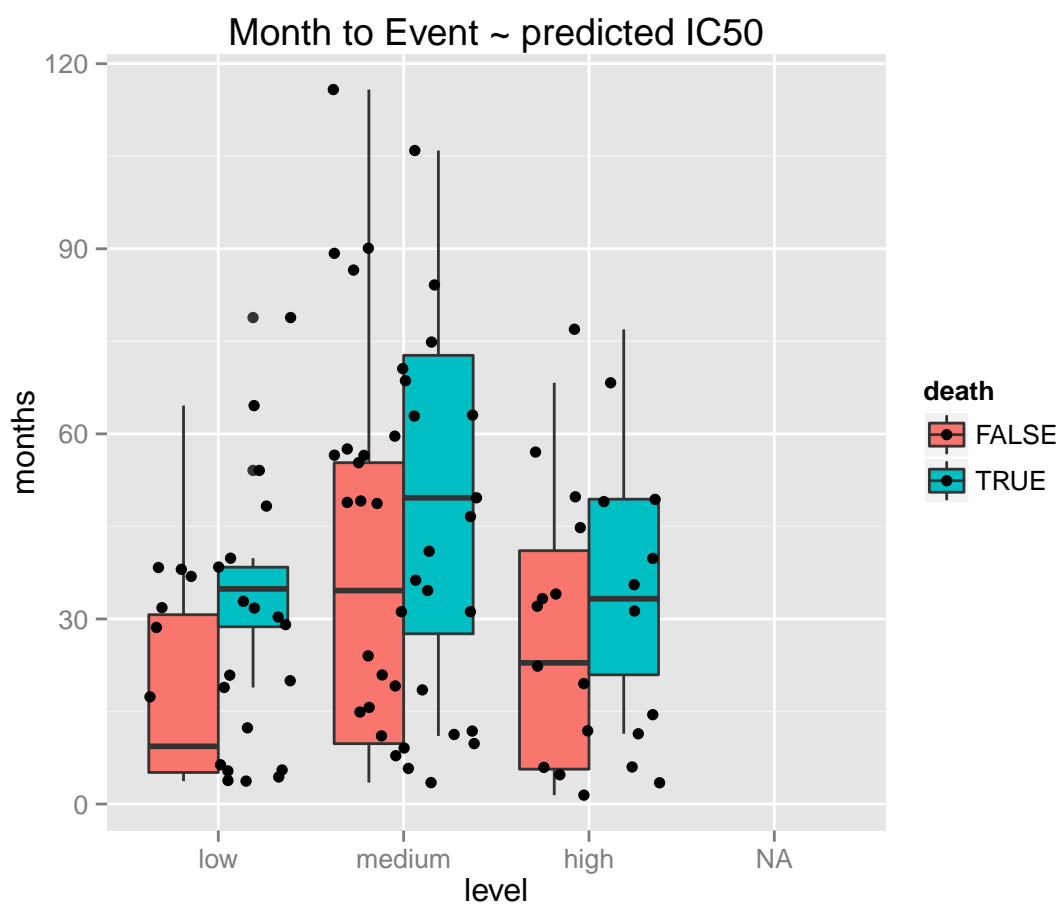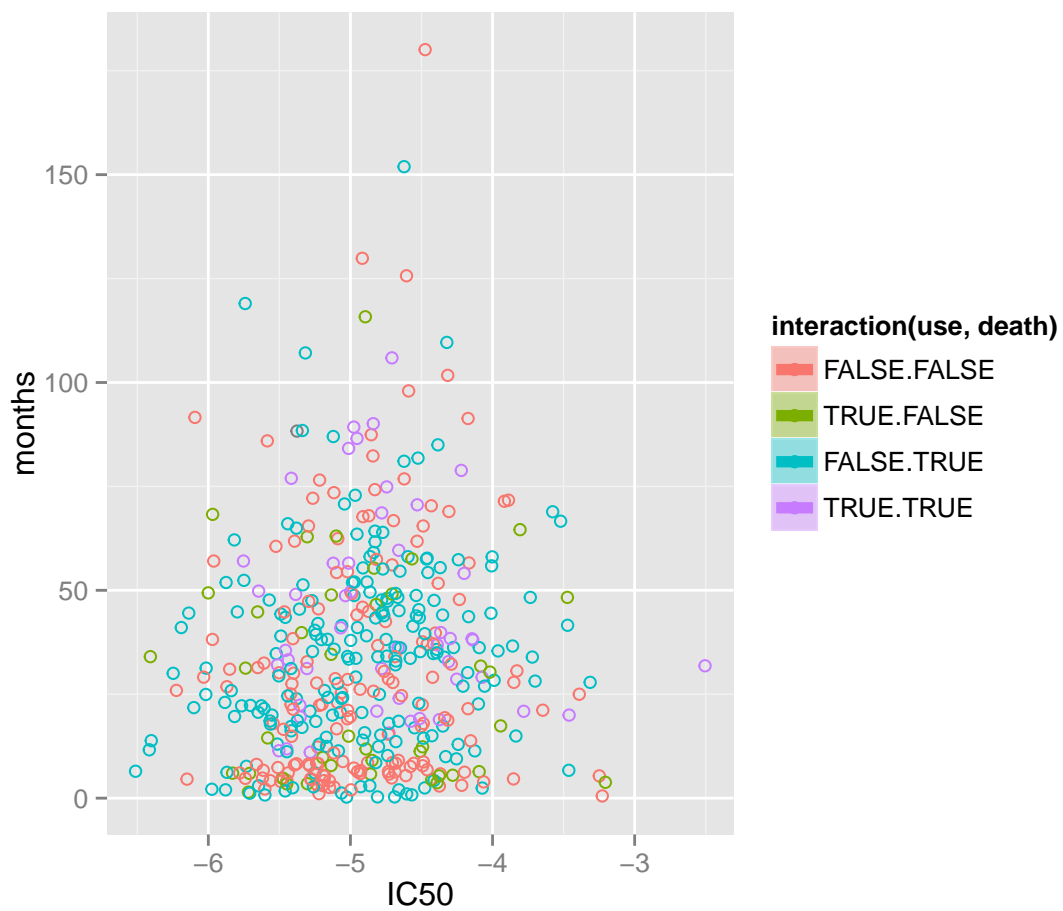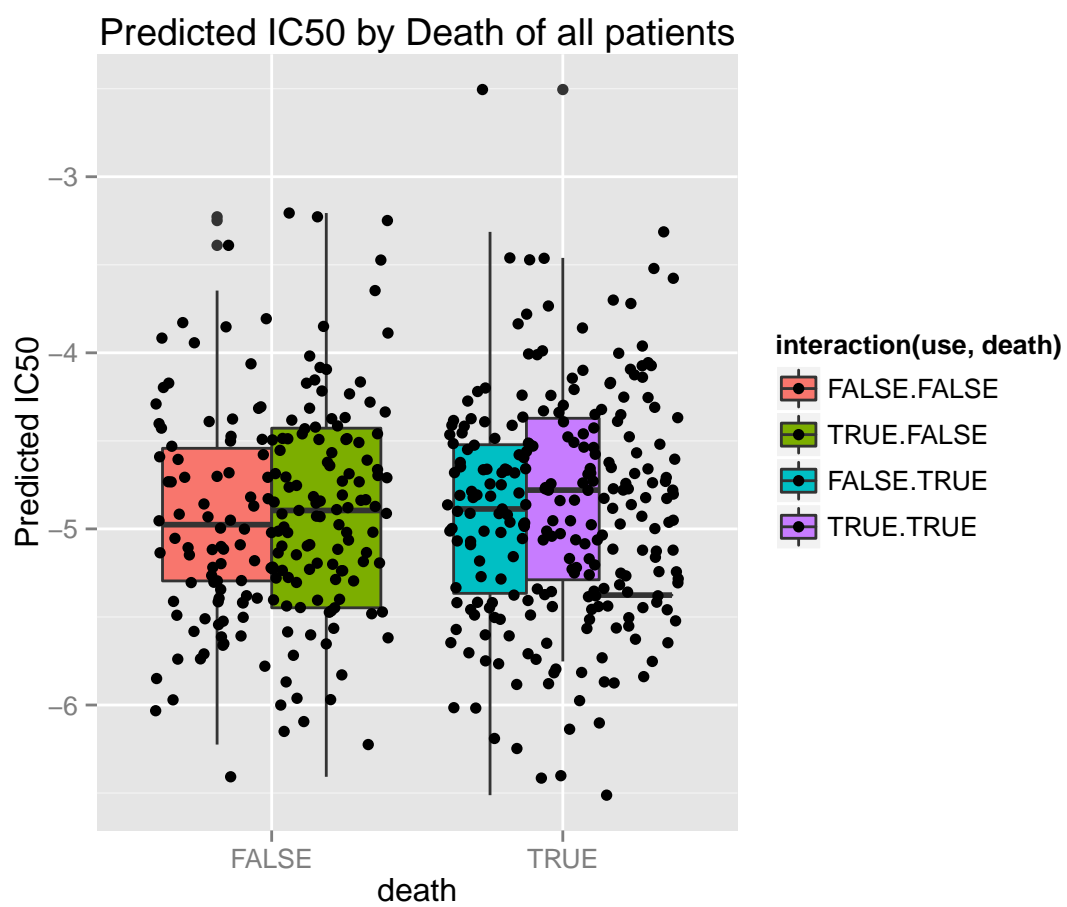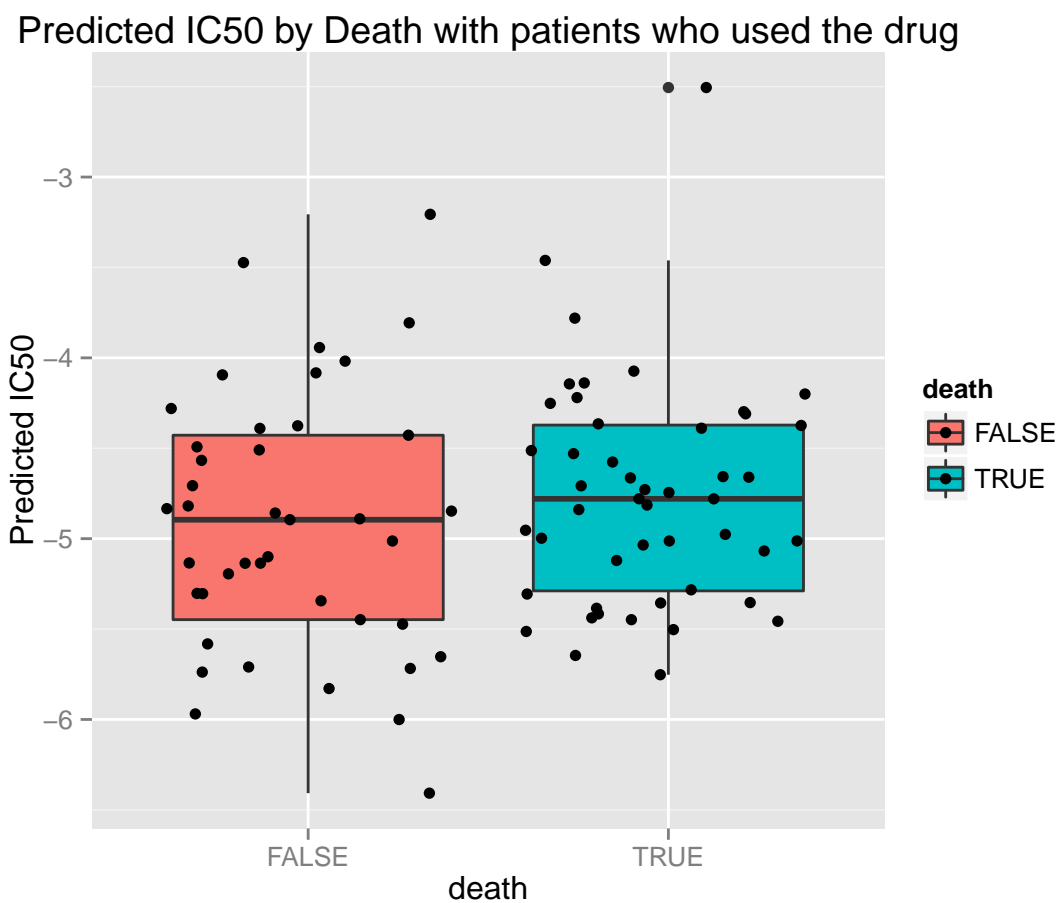
Figure 10:

Figure 11:

Figure 12:

Figure 13:

```
## Warning in loop_apply(n, do.ply): Removed 1 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 1 rows containing missing values
## (geom_point).
```

```r
print("T-test for difference")
```

```
## [1] "T-test for difference"
```

```r
with(data=data[data$use, ], t.test(IC50[death], IC50[!death]))
```

```
##
##  Welch Two Sample t-test
##
## data:  IC50[death] and IC50[!death]
## t = 1.1062, df = 79.322, p-value = 0.272
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1267661  0.4439548
## sample estimates:
## mean of x mean of y
## -4.761204 -4.919798
```

```r
print("KM survival curves by predicted categories")
```

```
## [1] "KM survival curves by predicted categories"
```

```r
ggsurv(survfit(with(data[data$use, ], Surv(months,death) ~ level)))
```
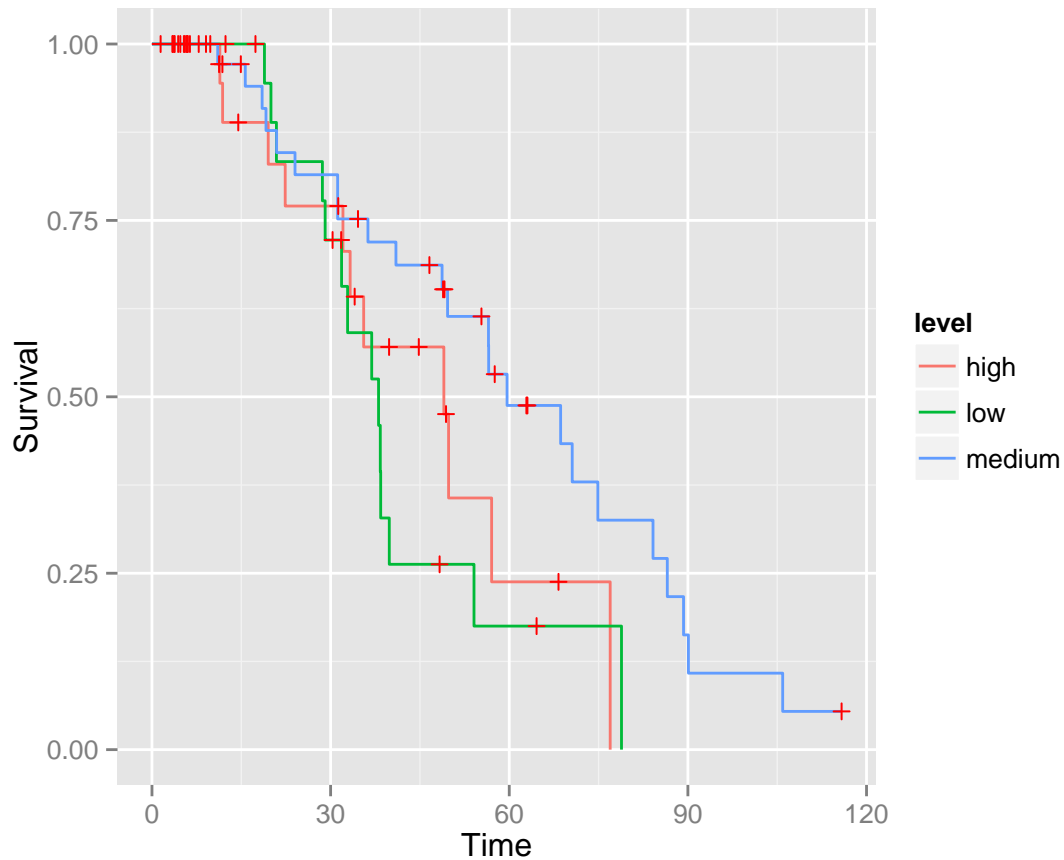
5. cisplatin:

```
drug = "cisplatin"
data = GetMyData(drug)

  print("Month of Events vs IC50")
```

```
## [1] "Month of Events vs IC50"
```

```
  qplot(level, months, data=data[data$use, ] , geom=c("boxplot", "jitter"),
   fill=death, main="Month to Event ~ predicted IC50")
```

```
## Warning in loop_apply(n, do.ply): Removed 1 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 1 rows containing missing values
## (geom_point).
```

```
  print("General Trend")
```

```
## [1] "General Trend"
```
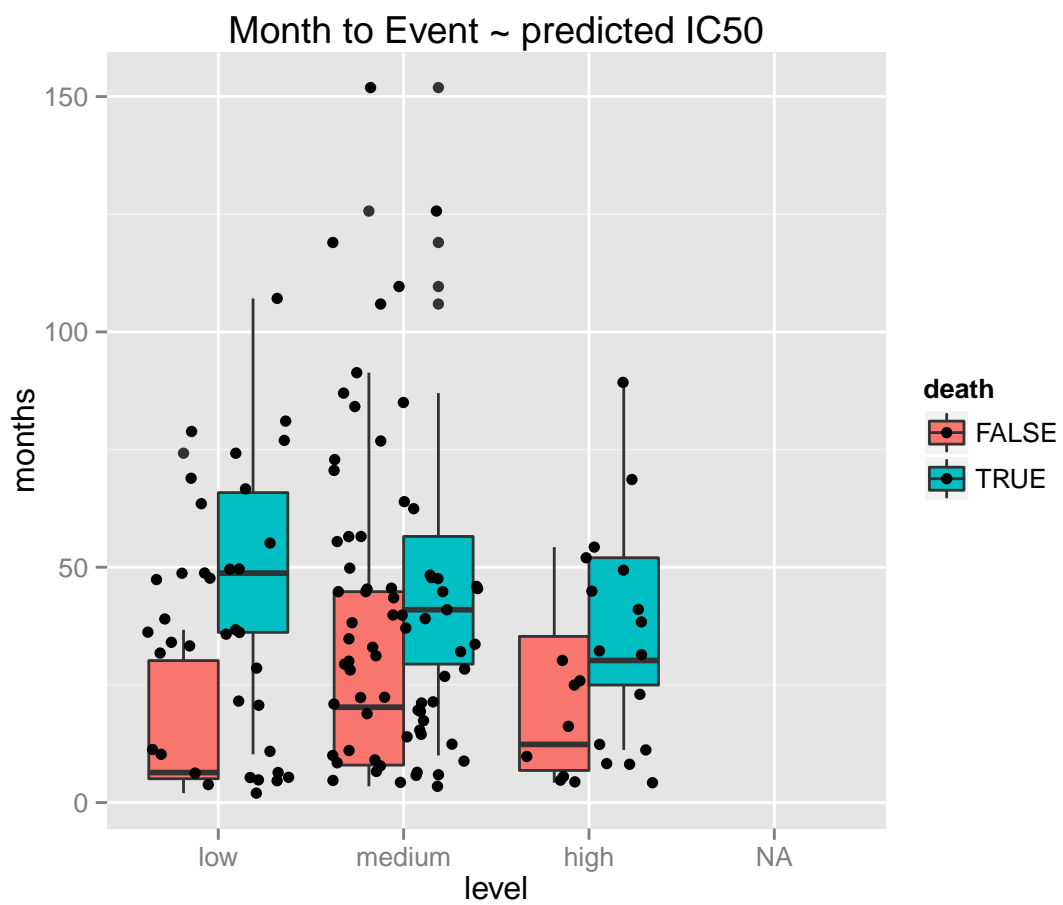
Figure 14:

```
ggplot(data, aes(x=IC50, y=months, fill=interaction(use, death),  color=interaction(use, death))) + g
```


Figure 15:

```
print("Do deceased patients have weak sensitivity?")
```

[1] "Do deceased patients have weak sensitivity?"

```
qplot(death, IC50, data=data, geom=c("boxplot", "jitter"),
  fill=interaction(use, death), main="Predicted IC50 by Death of all patients",  ylab="Predicted IC50")
```

```
qplot(death, IC50, data=data[data$use, ], geom=c("boxplot", "jitter"),
  fill=death, main="Predicted IC50 by Death with patients who used the drug",  ylab="Predicted IC50")
```

Warning in loop_apply(n, do.ply): Removed 1 rows containing non-finite
values (stat_boxplot).

Warning in loop_apply(n, do.ply): Removed 1 rows containing missing values
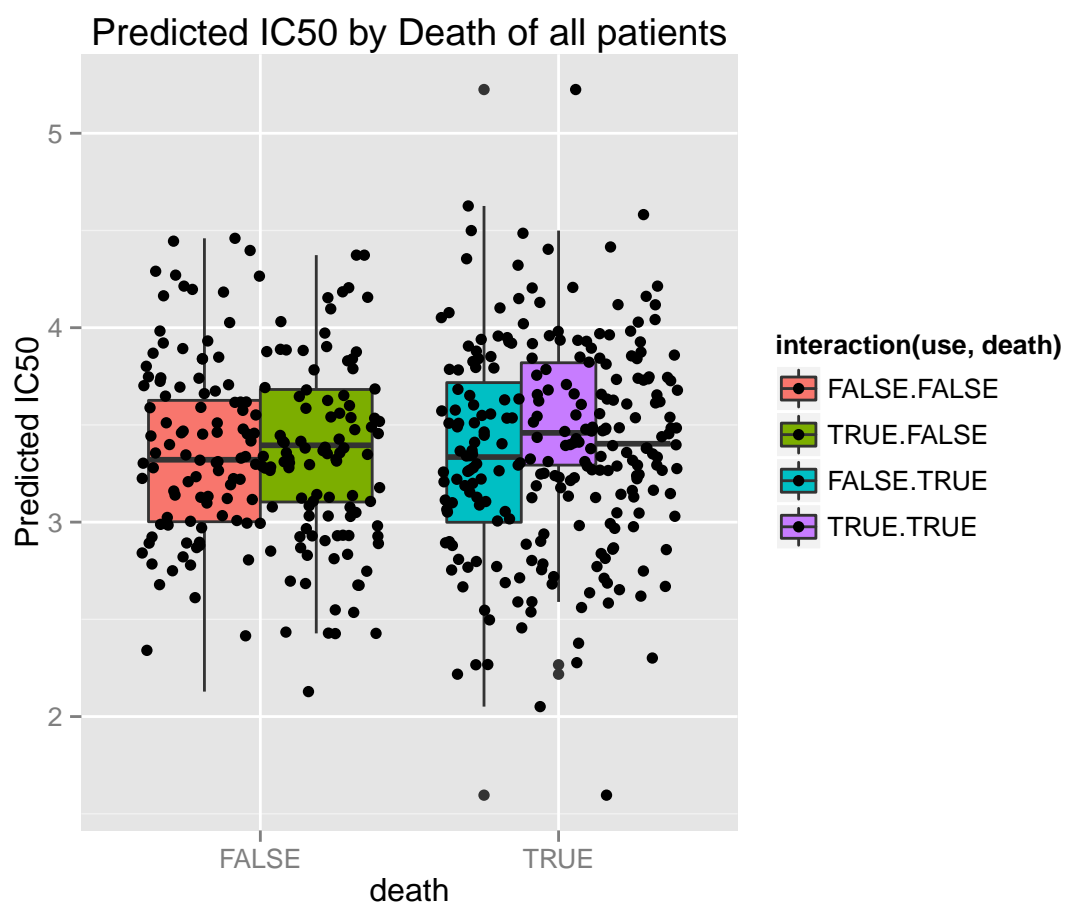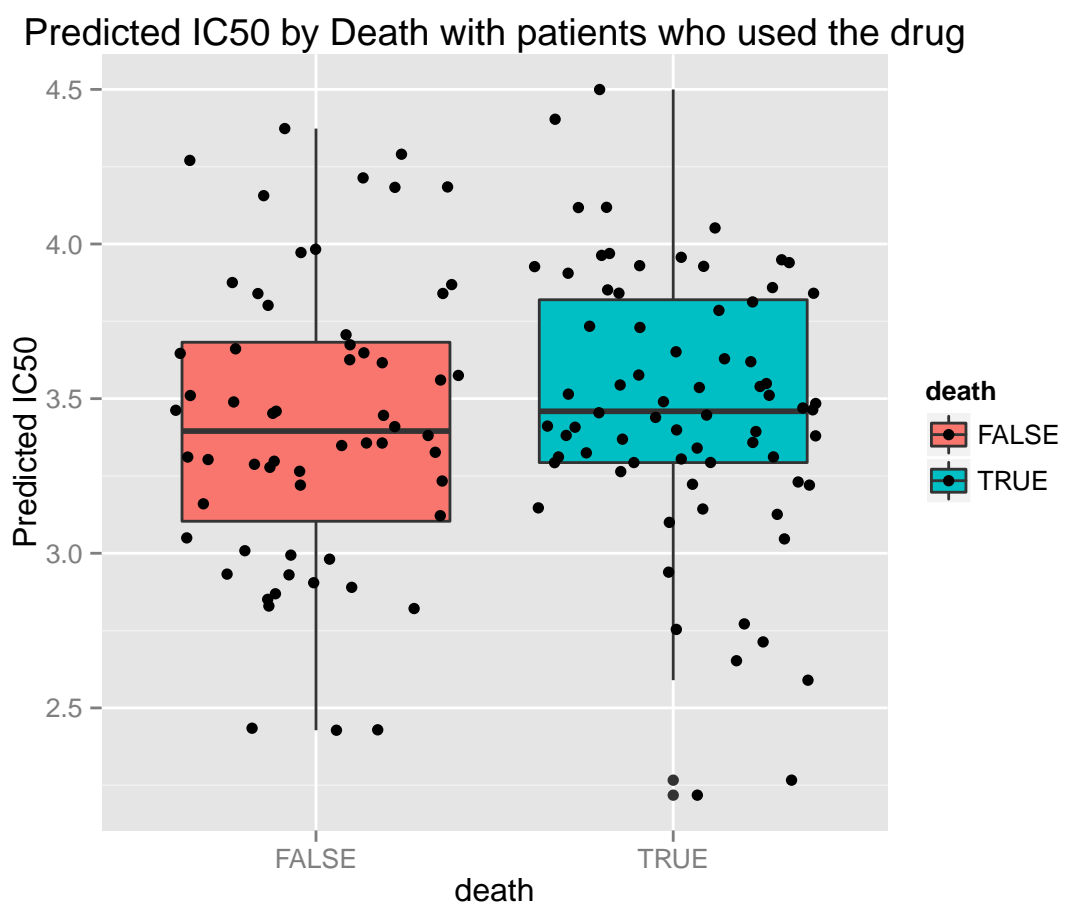## (geom_point).

Figure 16:

Figure 17:

```
    print("T-test for difference")
```

```
## [1] "T-test for difference"
```

```
    with(data=data[data$use, ], t.test(IC50[death], IC50[!death]))
```

```
##
##  Welch Two Sample t-test
##
## data:  IC50[death] and IC50[!death]
## t = 0.61666, df = 122.09, p-value = 0.5386
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1085659  0.2068077
## sample estimates:
## mean of x mean of y
##  3.472380  3.423259
```

```
    print("KM survival curves by predicted categories")
```

```
## [1] "KM survival curves by predicted categories"
```

```
    ggsurv(survfit(with(data[data$use, ], Surv(months,death) ~ level)))
```

---

6. carboplatin using cisplatin predictions:

```
drug = "carboplatin"
data = GetMyData(drug)

  print("Month of Events vs IC50")
```

```
## [1] "Month of Events vs IC50"
```

```
  qplot(level, months, data=data[data$use, ] , geom=c("boxplot", "jitter"),
   fill=death, main="Month to Event ~ predicted IC50")
```

```
  print("General Trend")
```

```
## [1] "General Trend"
```

```
  ggplot(data, aes(x=IC50, y=months, fill=interaction(use, death),  color=interaction(use, death))) + g
```
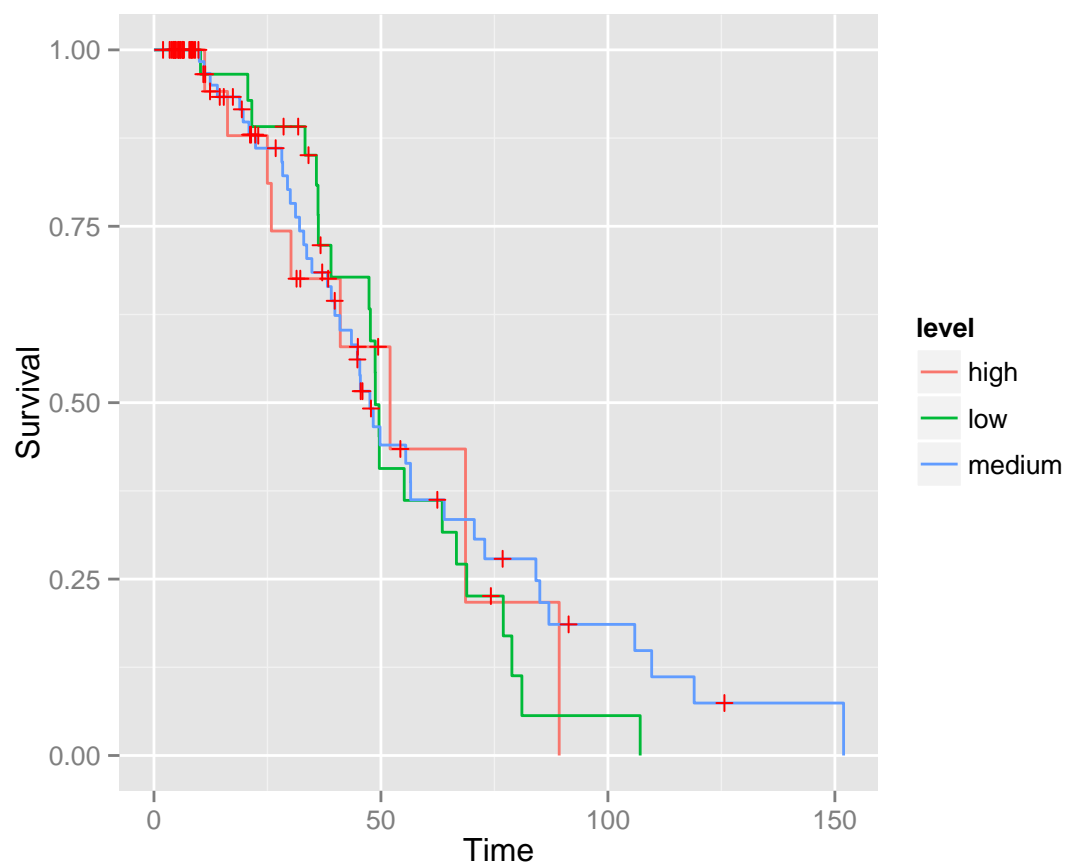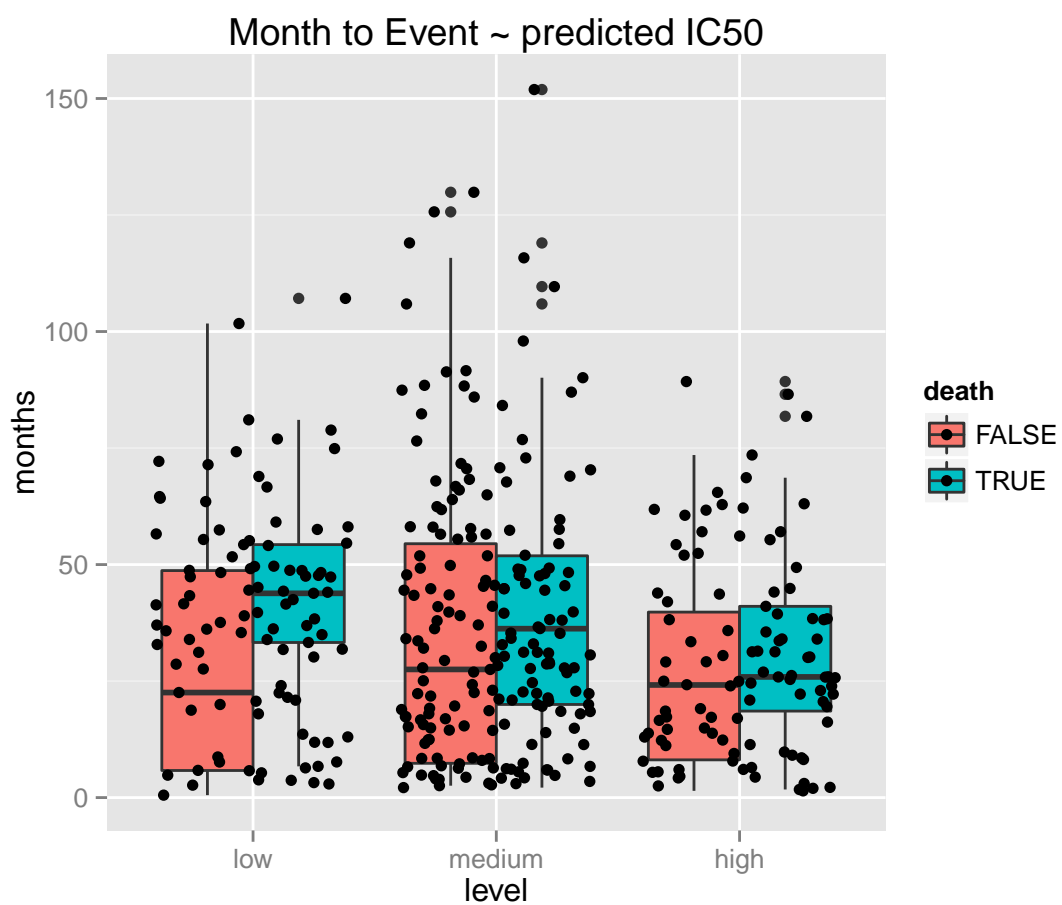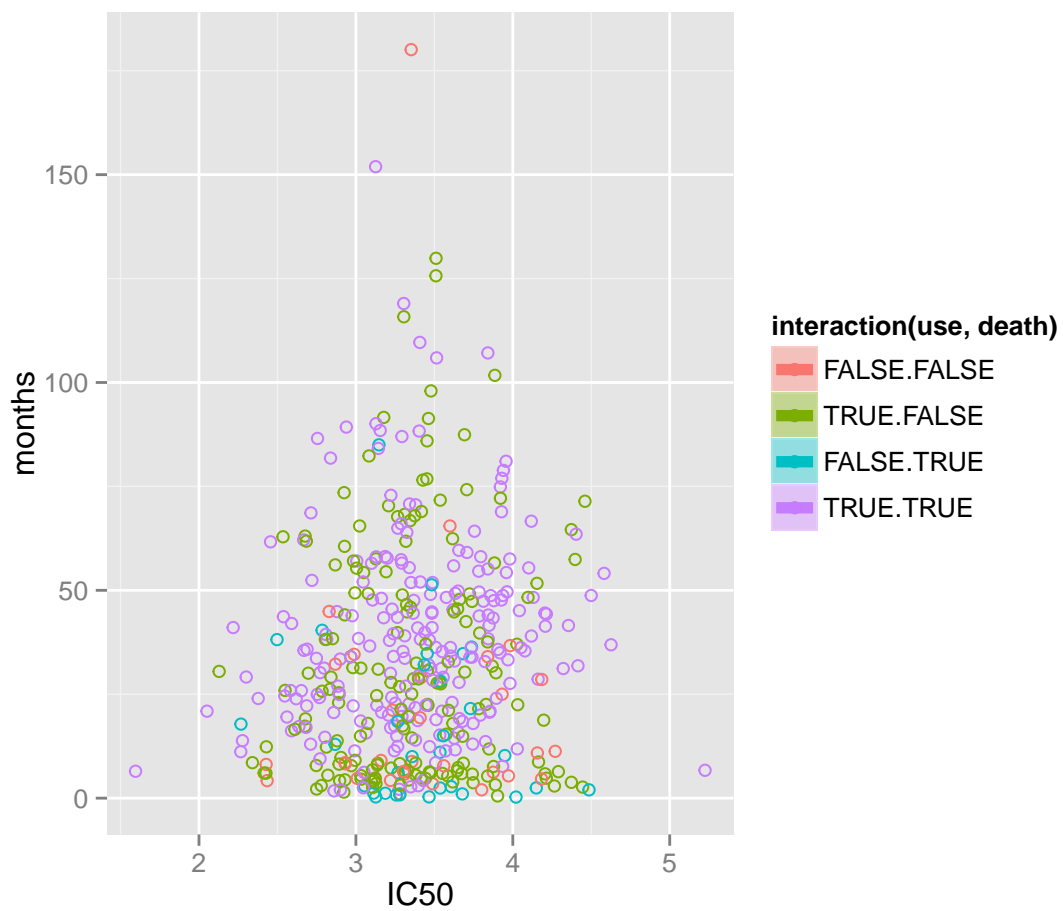
Figure 18:

Figure 19:

Figure 20:

```
    print("Do deceased patients have weak sensitivity?")
```

[1] "Do deceased patients have weak sensitivity?"

```
  qplot(death, IC50, data=data, geom=c("boxplot", "jitter"),
   fill=interaction(use, death), main="Predicted IC50 by Death of all patients",  ylab="Predicted IC50")
```
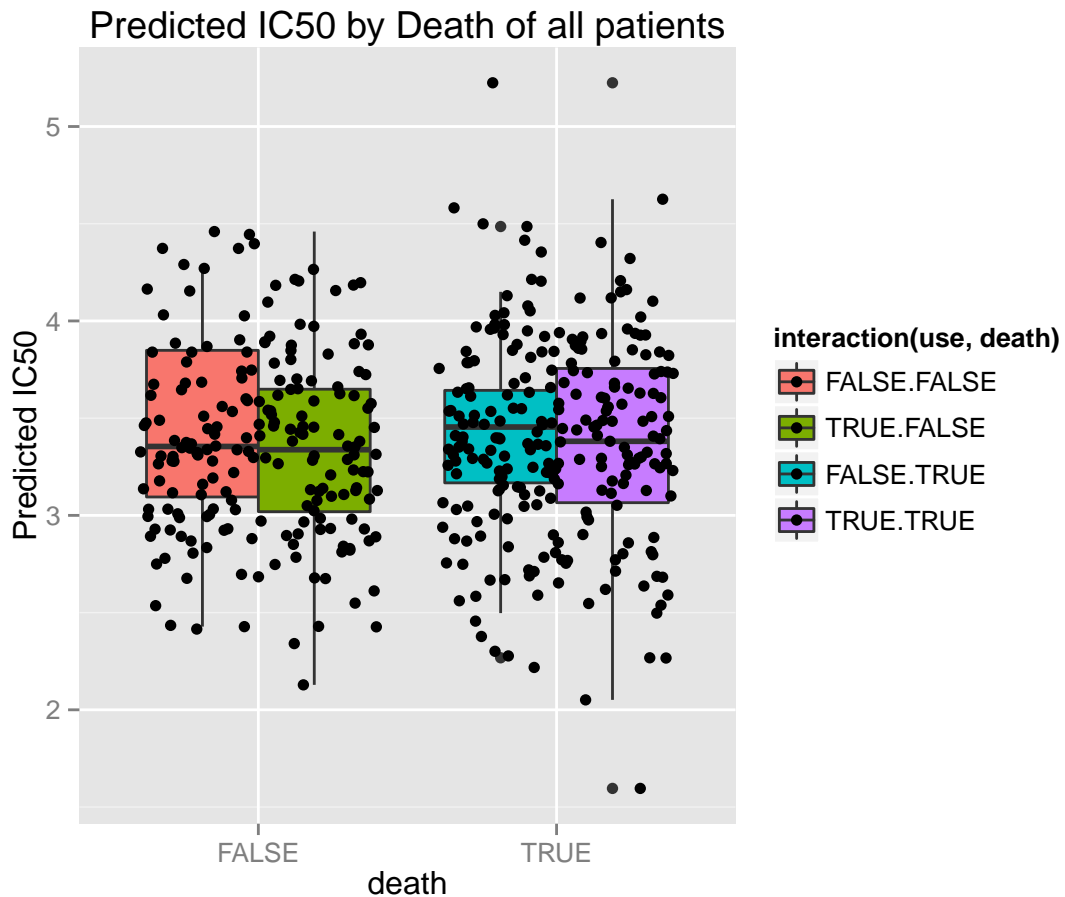

Figure 21:

```
  qplot(death, IC50, data=data[data$use, ], geom=c("boxplot", "jitter"),
   fill=death, main="Predicted IC50 by Death with patients who used the drug",  ylab="Predicted IC50")
```

```
    print("T-test for difference")
```

[1] "T-test for difference"

```
    with(data=data[data$use, ], t.test(IC50[death], IC50[!death]))
```


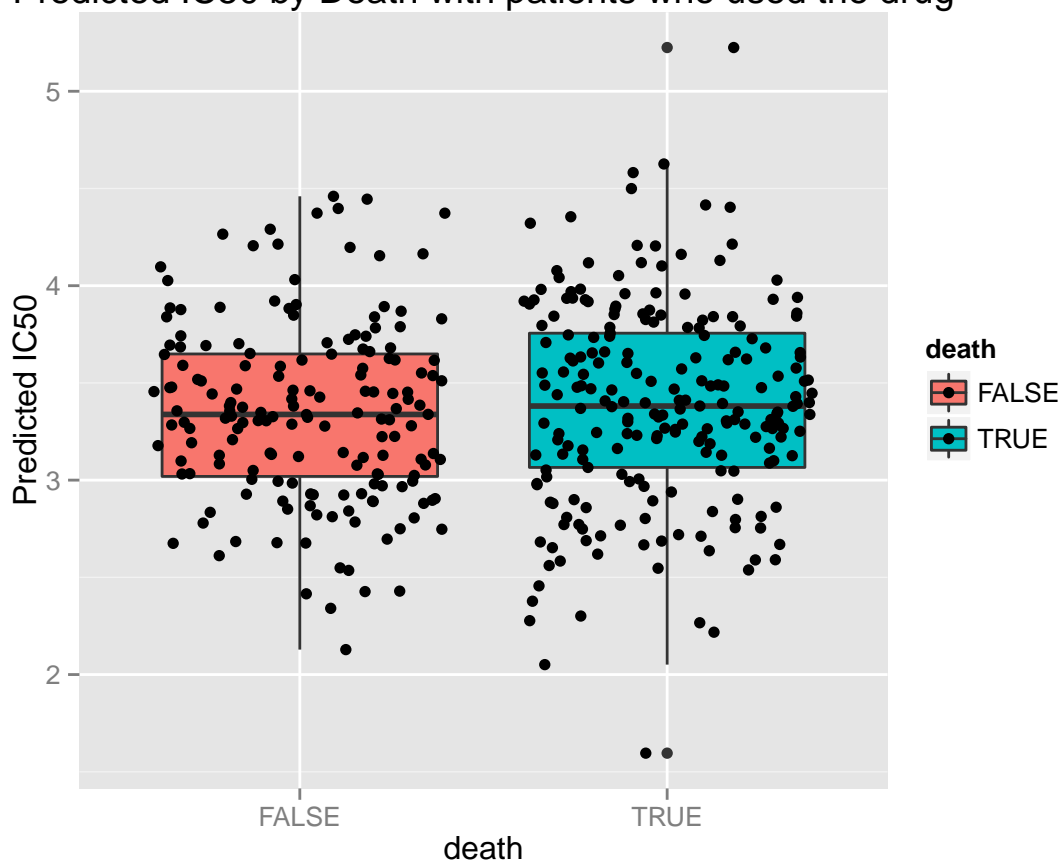## 	Welch Two Sample t-test

Figure 22:

```
## 
## data:  IC50[death] and IC50[!death]
## t = 0.74265, df = 376.13, p-value = 0.4582
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.06154412  0.13624927
## sample estimates:
## mean of x mean of y
##  3.386050  3.348698
```

```
    print("KM survival curves by predicted categories")
```

```
## [1] "KM survival curves by predicted categories"
```

```
    ggsurv(survfit(with(data[data$use, ], Surv(months,death) ~ level)))
```
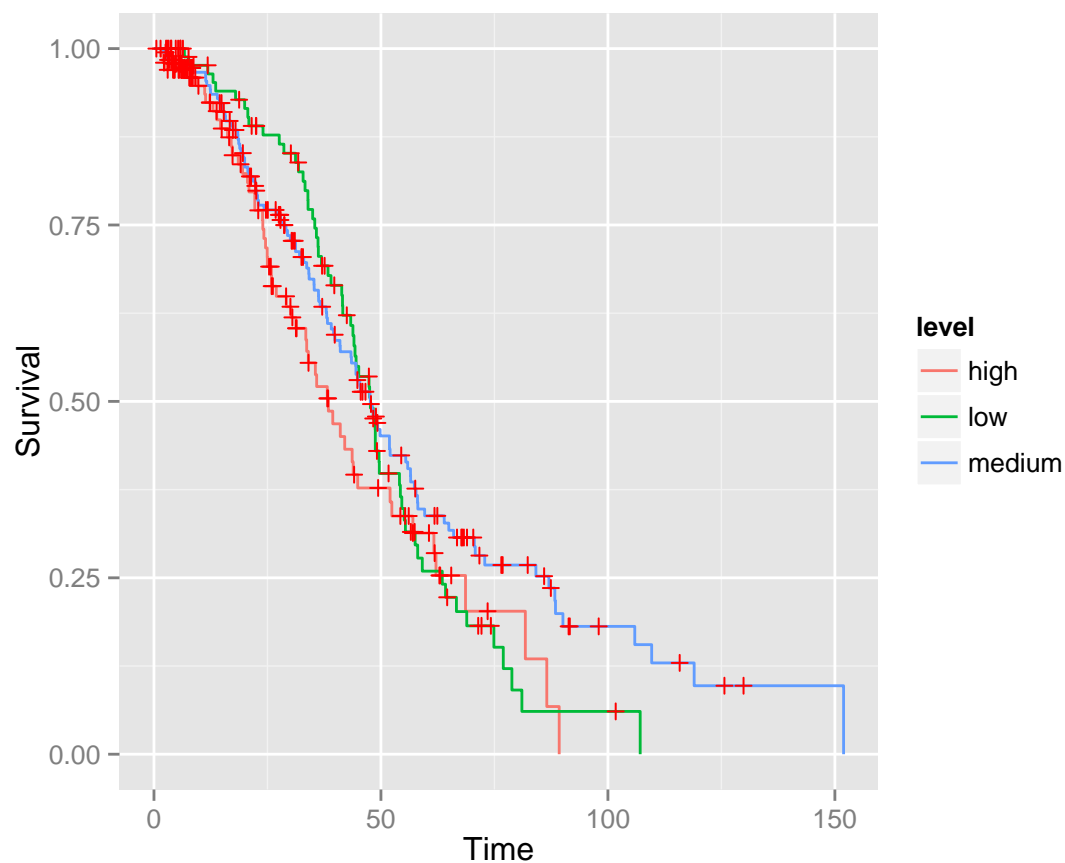

Figure 23:

7. taxol and carboplatin (using cisplatin prediction):

```
carbotaxol.data = with(surv, surv[which(taxol & carboplatin & !cisplatin & !taxotere ) , ])
level = paste(carbotaxol.data$cisplatin.level, carbotaxol.data$taxol.level, sep="|")
carbotaxol.data$level = level
w = which(level == "high|high")
data = cbind(carbotaxol.data[w,], cat="both high")
w = which(grepl("high", level))
data = rbind(data, cbind(carbotaxol.data[w,], cat="at least one high"))
w = which(level == "low|low")
data = rbind(data, cbind(carbotaxol.data[w,], cat="both low"))
w = which(!grepl("high", level))
data = rbind(data, cbind(carbotaxol.data[w,], cat="neither high"))
data$cat = factor(data$cat, levels = c("both low", "neither high", "at least one high", "both high"))

ggsurv(survfit(with(data, Surv(months,death) ~ cat)))
```
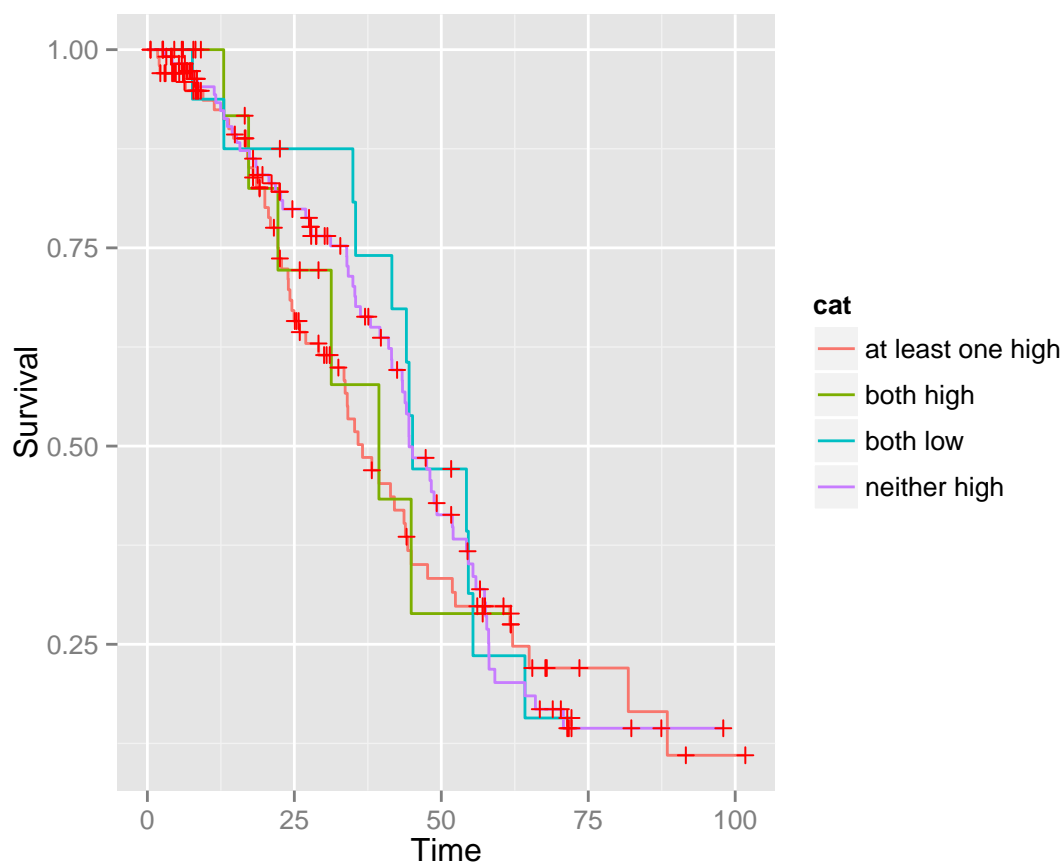

Figure 24:

```
qplot(cat, months, data=data , geom=c("boxplot", "jitter"),
      fill=death, main="Month to Event ~ predicted sensitivity")
```

```
qplot(cat, cisplatin.predict, data=data, geom=c("boxplot", "jitter"),
   fill=cat, main="Predicted Carboplatin Sensitivity by Death",
   xlab="", ylab="Predicted Cisplatin Sensitivity")
```
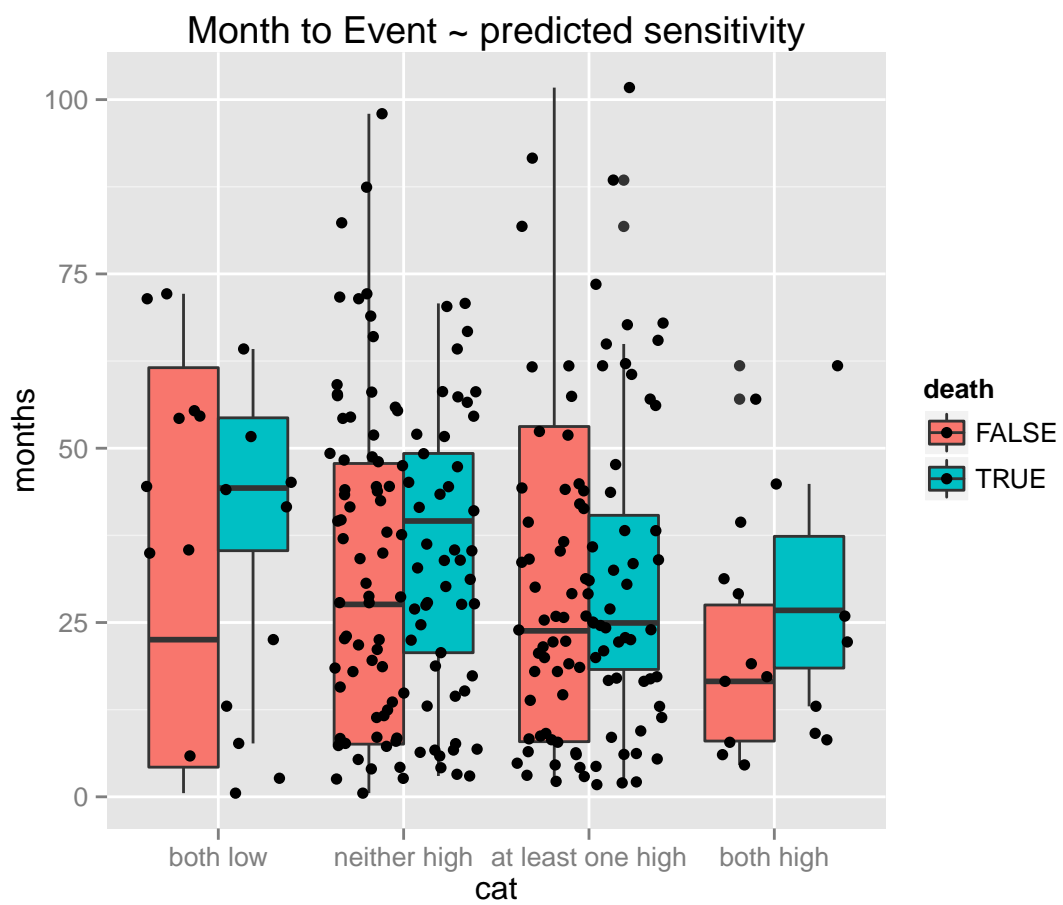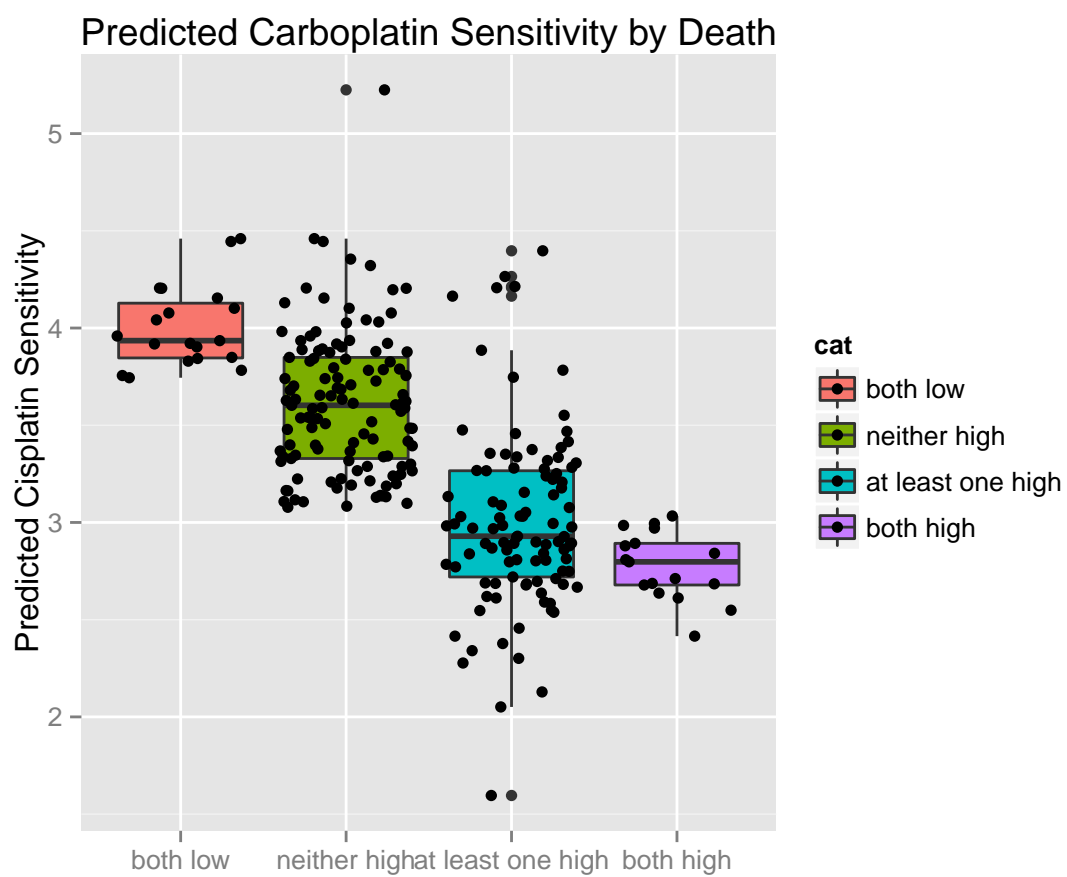
Figure 25:

Figure 26:

```
qplot(cat, taxol.predict, data=data, geom=c("boxplot", "jitter"),
    fill=cat, main="Predicted Taxol Sensitivity by Death",
    xlab="", ylab="Predicted Taxol Sensitivity")
```
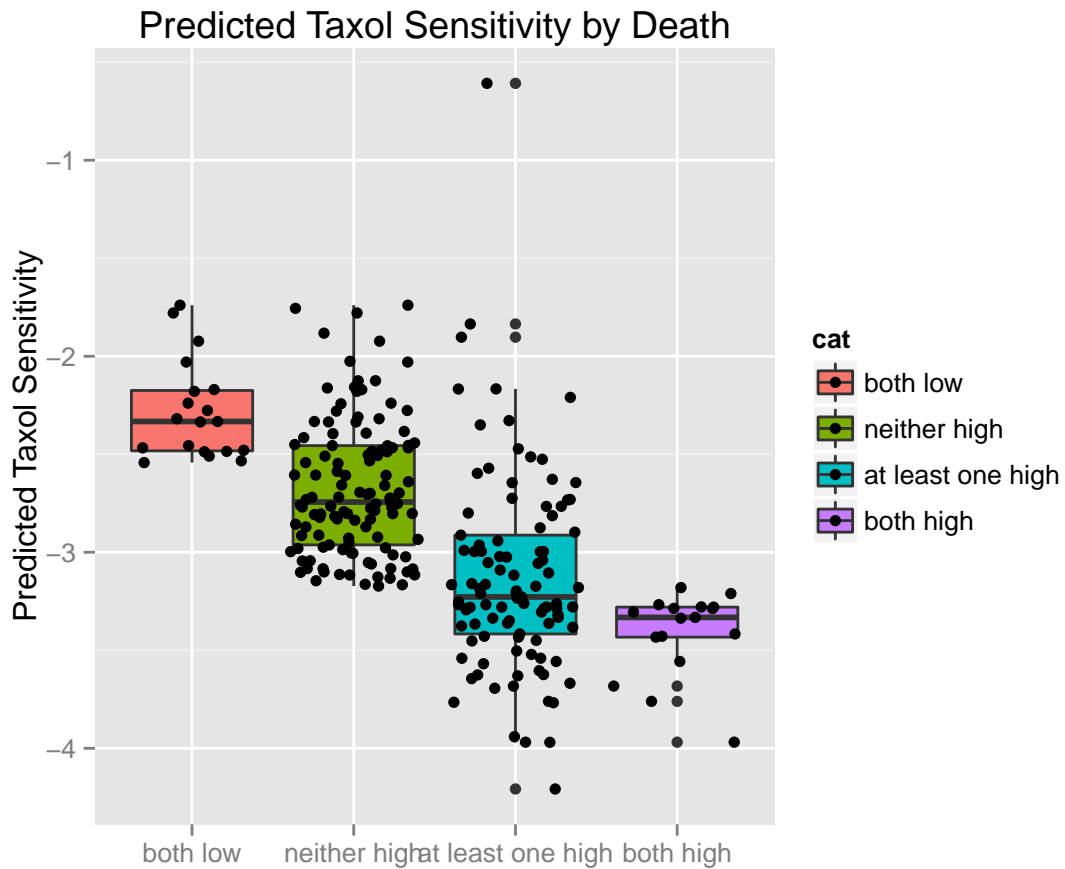

Figure 27:

```
my.logit = glm(death ~ taxol.predict + cisplatin.predict + other, data=carbotaxol.data, family="binomial
summary(my.logit)
```

```
##
## Call:
## glm(formula = death ~ taxol.predict + cisplatin.predict + other,
##     family = "binomial", data = carbotaxol.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8737  -1.0146   0.6650   0.9665   1.6420
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)         3.3342     1.4317   2.329  0.01987 *
## taxol.predict       0.9755     0.3220   3.029  0.00245 **
## cisplatin.predict  -0.3103     0.2896  -1.071  0.28400
## otherTRUE           1.3358     0.2986   4.473 7.72e-06 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 299.99  on 217  degrees of freedom
## Residual deviance: 271.36  on 214  degrees of freedom
## AIC: 279.36
##
## Number of Fisher Scoring iterations: 4
```