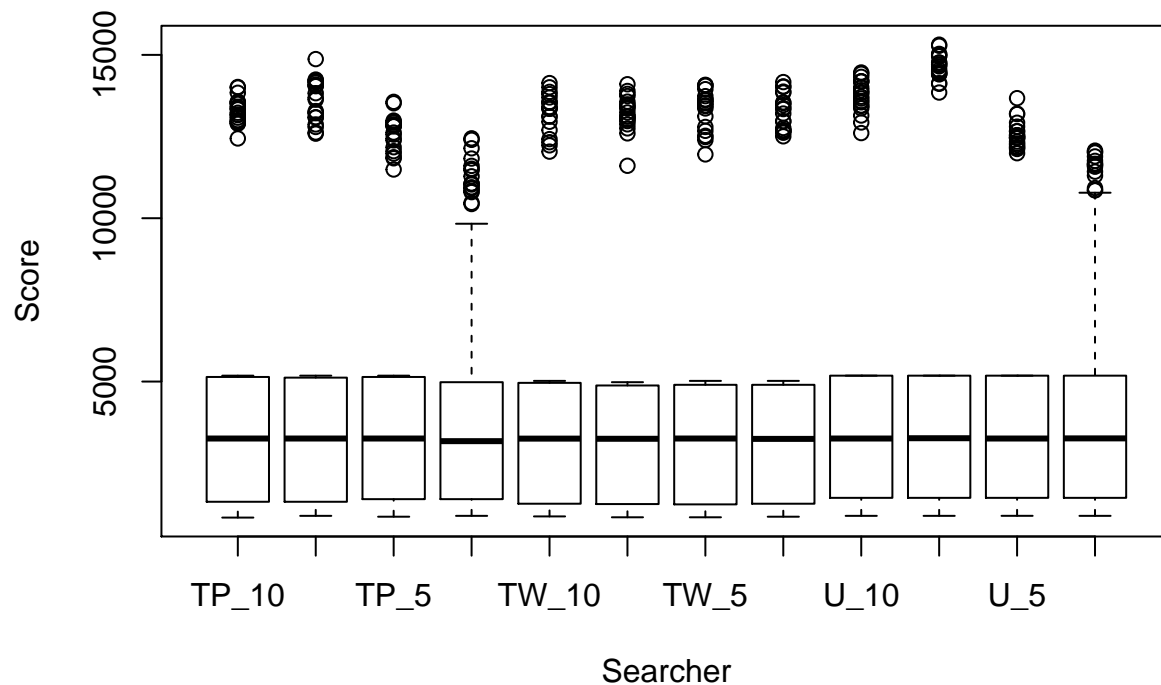


Population_Search_Analysis

Maggie Casale, Lenny Scott, & Zach Copic

March 4, 2016

```
pop_data_20_runs <- read.csv("/home/casal033/ECAI/new-simple-search-the-second/100000tries_no_1000item.csv")
plot(pop_data_20_runs$Score ~ pop_data_20_runs$Searcher,
     xlab="Searcher", ylab="Score")
```



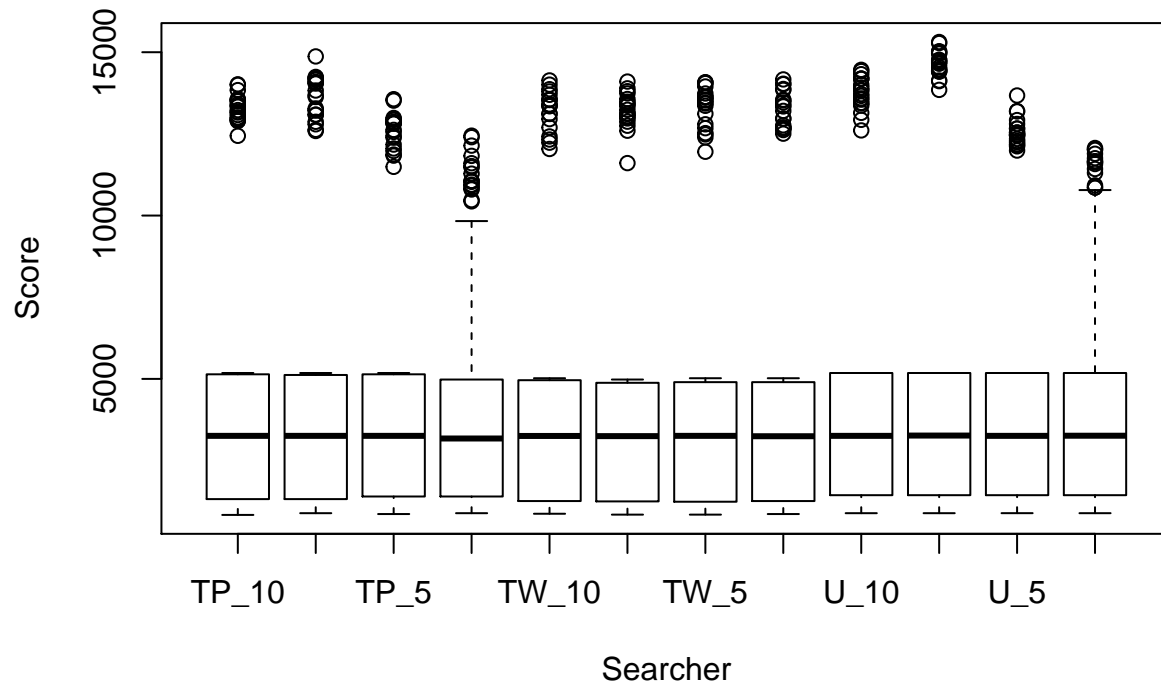
```
negs <- subset(pop_data_20_runs, Score<0)
nrow(negs)
```

```
## [1] 0
```

```
unique(negs$Problem)
```

```
## factor(0)
## Levels: K_11_20 K_11_200 K_13_20 K_13_200 K_16_20 K_16_200
```

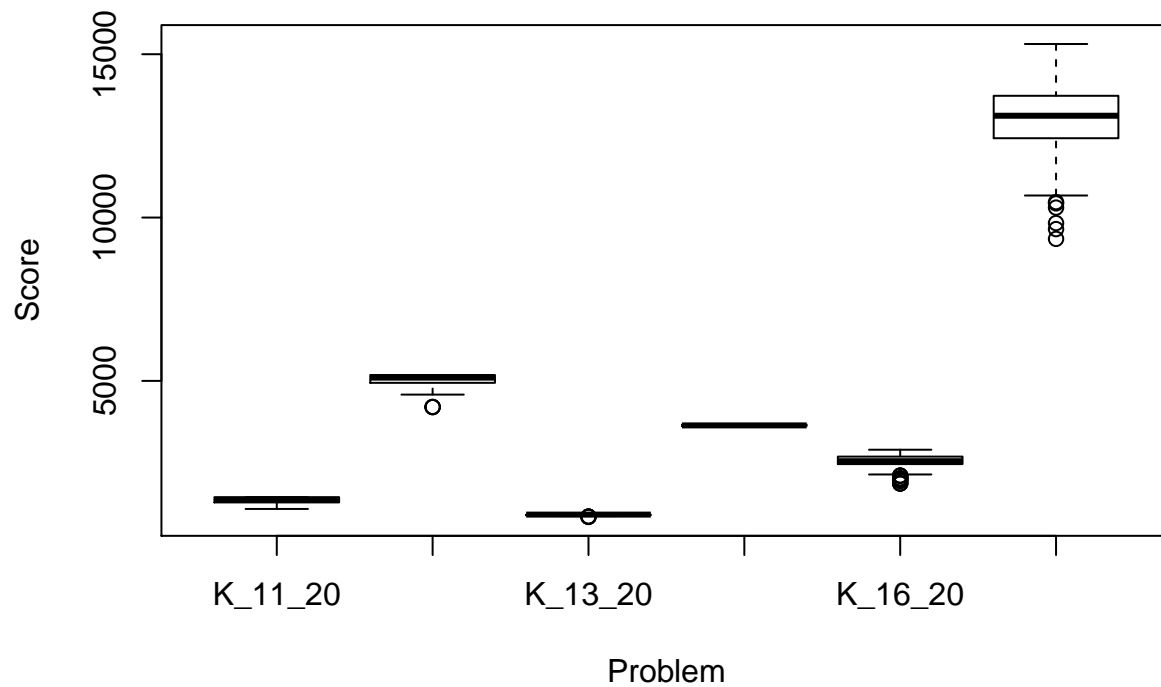
```
pop_data_20_runs$Non_negative_score = ifelse(pop_data_20_runs$Score<0, 0, pop_data_20_runs$Score)
plot(pop_data_20_runs$Non_negative_score ~ pop_data_20_runs$Searcher,
     xlab="Searcher", ylab="Score")
```



```
pairwise.wilcox.test(pop_data_20_runs$Non_negative_score, pop_data_20_runs$Search)
```

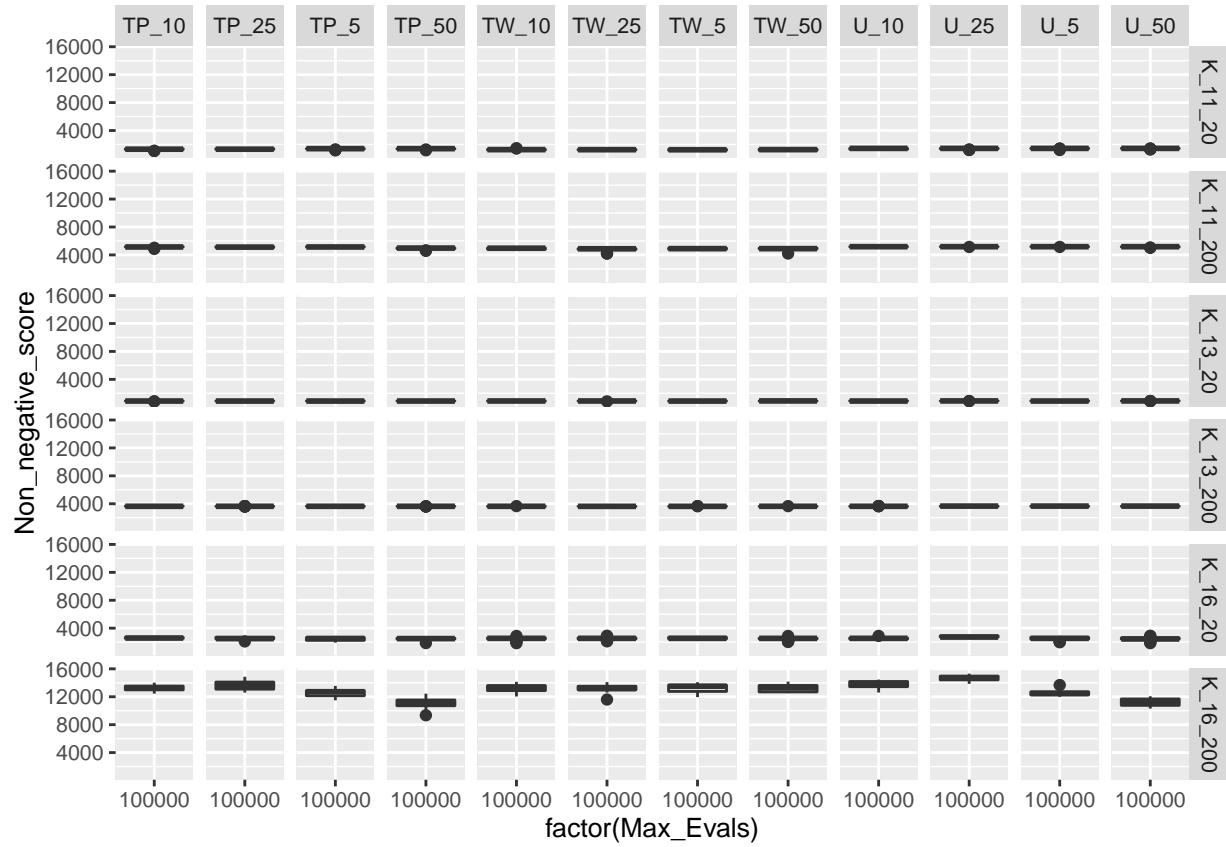
```
##
## Pairwise comparisons using Wilcoxon rank sum test
##
## data: pop_data_20_runs$Non_negative_score and pop_data_20_runs$Search
##
##      TP_10 TP_25 TP_5 TP_50 TW_10 TW_25 TW_5 TW_50 U_10 U_25 U_5
## TP_25 1    -    -    -    -    -    -    -    -    -    -
## TP_5  1    1    -    -    -    -    -    -    -    -    -
## TP_50 1    1    1    -    -    -    -    -    -    -    -
## TW_10 1    1    1    1    -    -    -    -    -    -    -
## TW_25 1    1    1    1    1    -    -    -    -    -    -
## TW_5  1    1    1    1    1    1    -    -    -    -    -
## TW_50 1    1    1    1    1    1    1    -    -    -    -
## U_10  1    1    1    1    1    1    1    1    -    -    -
## U_25  1    1    1    1    1    1    1    1    1    -    -
## U_5   1    1    1    1    1    1    1    1    1    1    -
## U_50  1    1    1    1    1    1    1    1    1    1    1
##
## P value adjustment method: holm
```

```
plot(pop_data_20_runs$Non_negative_score ~ pop_data_20_runs$Problem,
      xlab="Problem", ylab="Score")
```



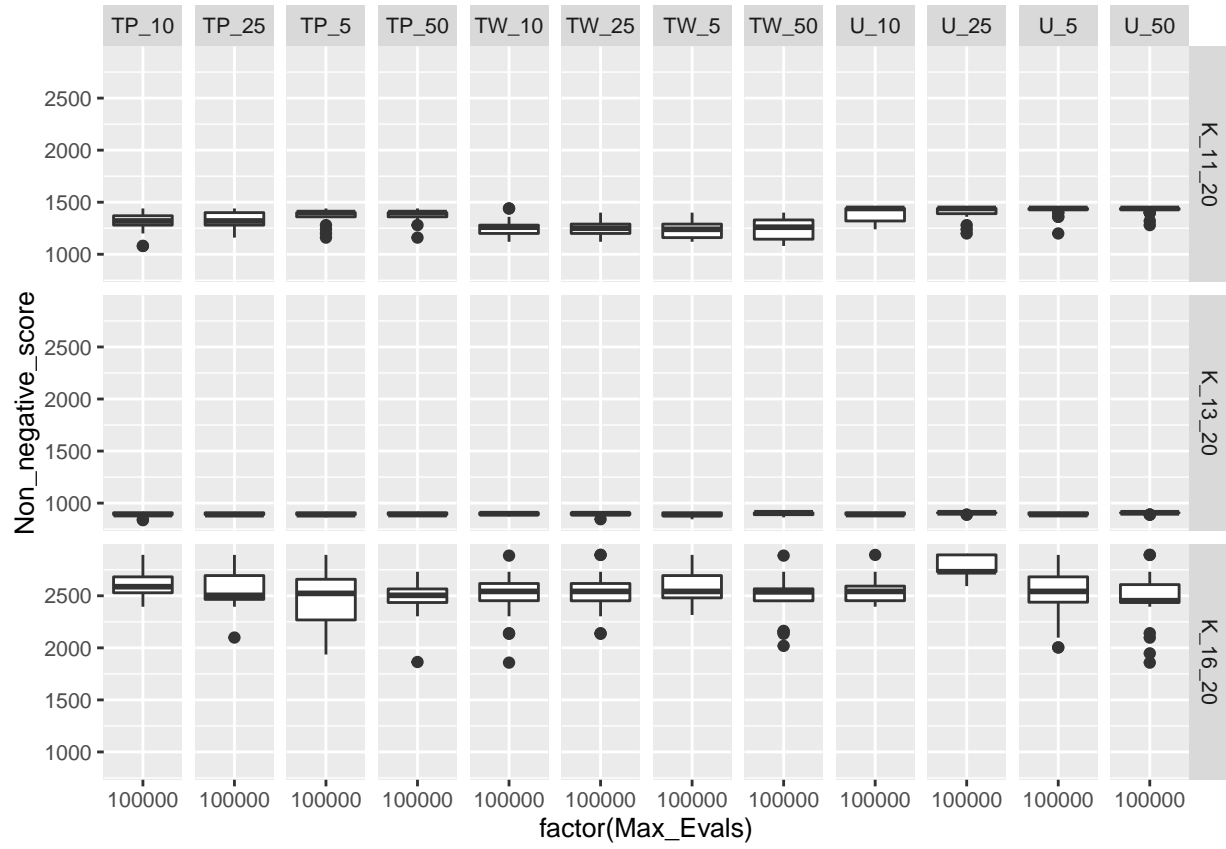
```
library("ggplot2")

ggplot(pop_data_20_runs,
       aes(x=factor(Max_Evals), y=Non_negative_score, group=Max_Evals)) +
  geom_boxplot() + facet_grid(Problem ~ Search) + theme_grey(base_size = 10)
```

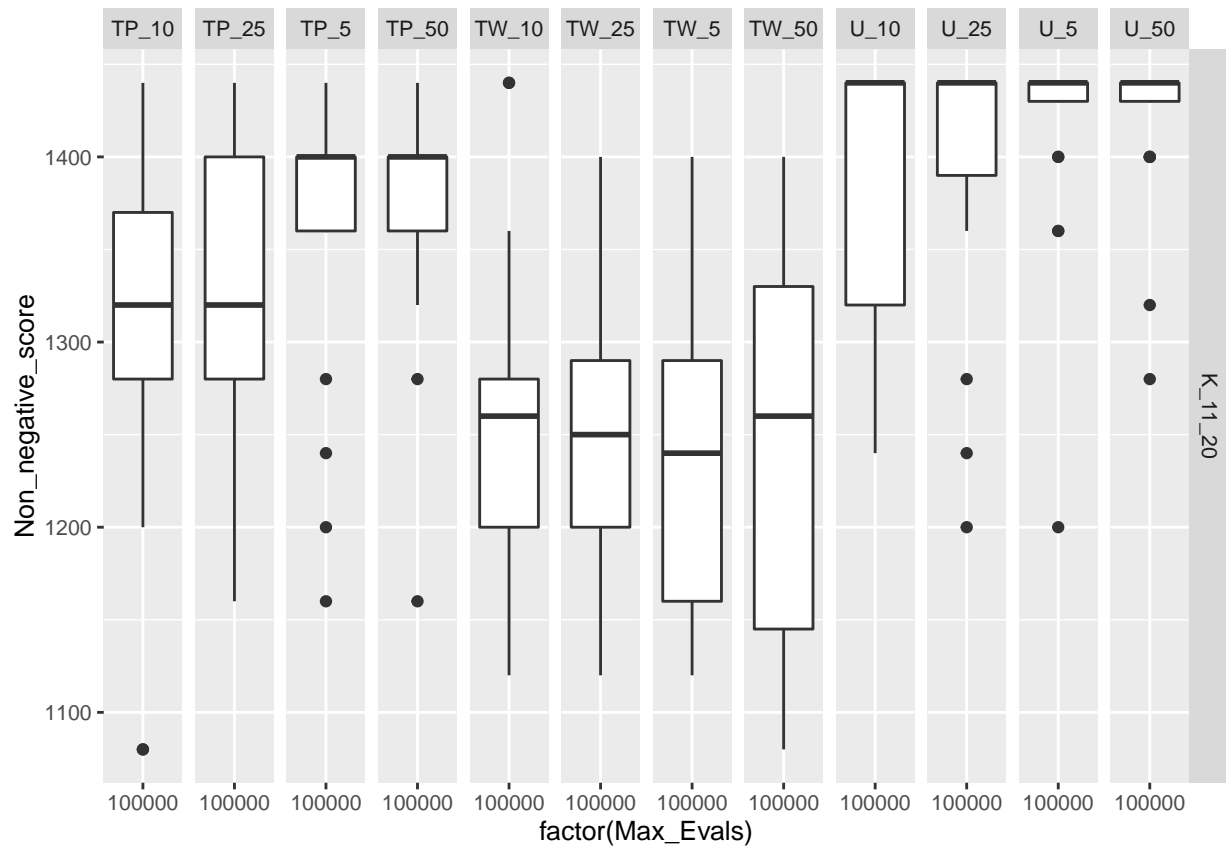


```
twenty_item_problems = subset(pop_data_20_runs, Problem=="K_11_20" | Problem=="K_13_20" | Problem=="K_16_20")
twenty_item_11 = subset(pop_data_20_runs, Problem=="K_11_20")
twenty_item_13 = subset(pop_data_20_runs, Problem=="K_13_20")
twenty_item_16 = subset(pop_data_20_runs, Problem=="K_16_20")
```

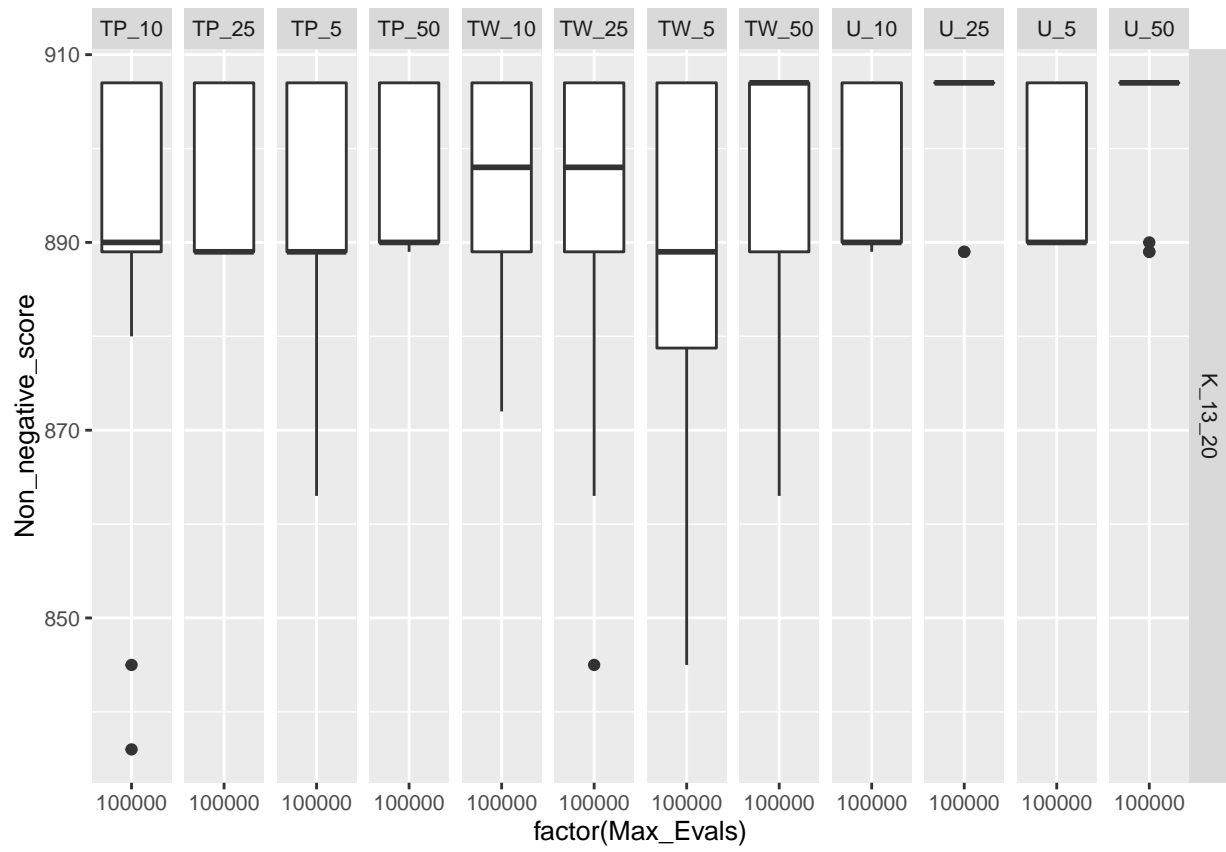
```
ggplot(twenty_item_problems, aes(factor(Max_Evals), Non_negative_score)) + geom_boxplot() + facet_grid(Problem ~ K)
```



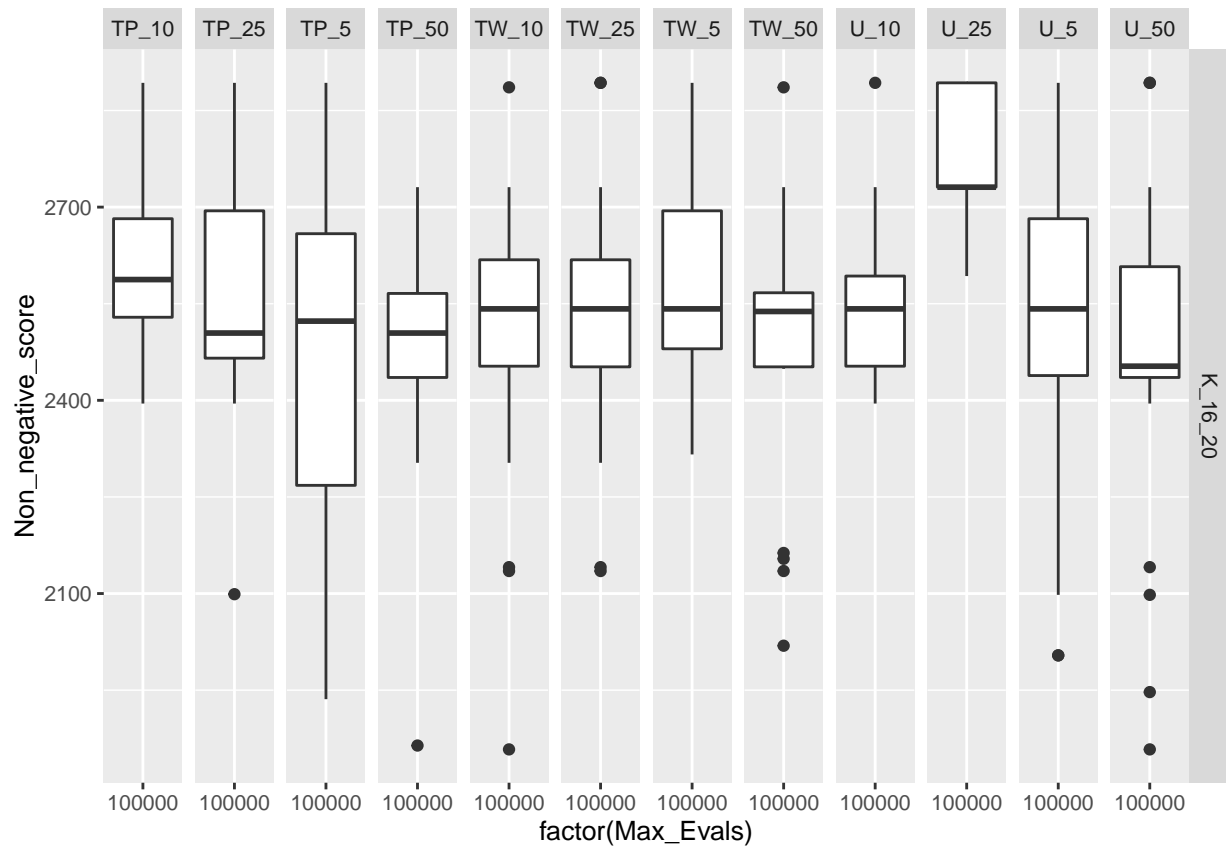
```
ggplot(twenty_item_11, aes(factor(Max_Evals), Non_negative_score)) + geom_boxplot() + facet_grid(Problem ~ K, scales = 'y')
```



```
ggplot(twenty_item_13, aes(factor(Max_Evals), Non_negative_score)) + geom_boxplot() + facet_grid(Problem ~ K_11_20)
```



```
ggplot(twenty_item_16, aes(factor(Max_Evals), Non_negative_score)) + geom_boxplot() + facet_grid(Problem ~ K_13_20)
```



```

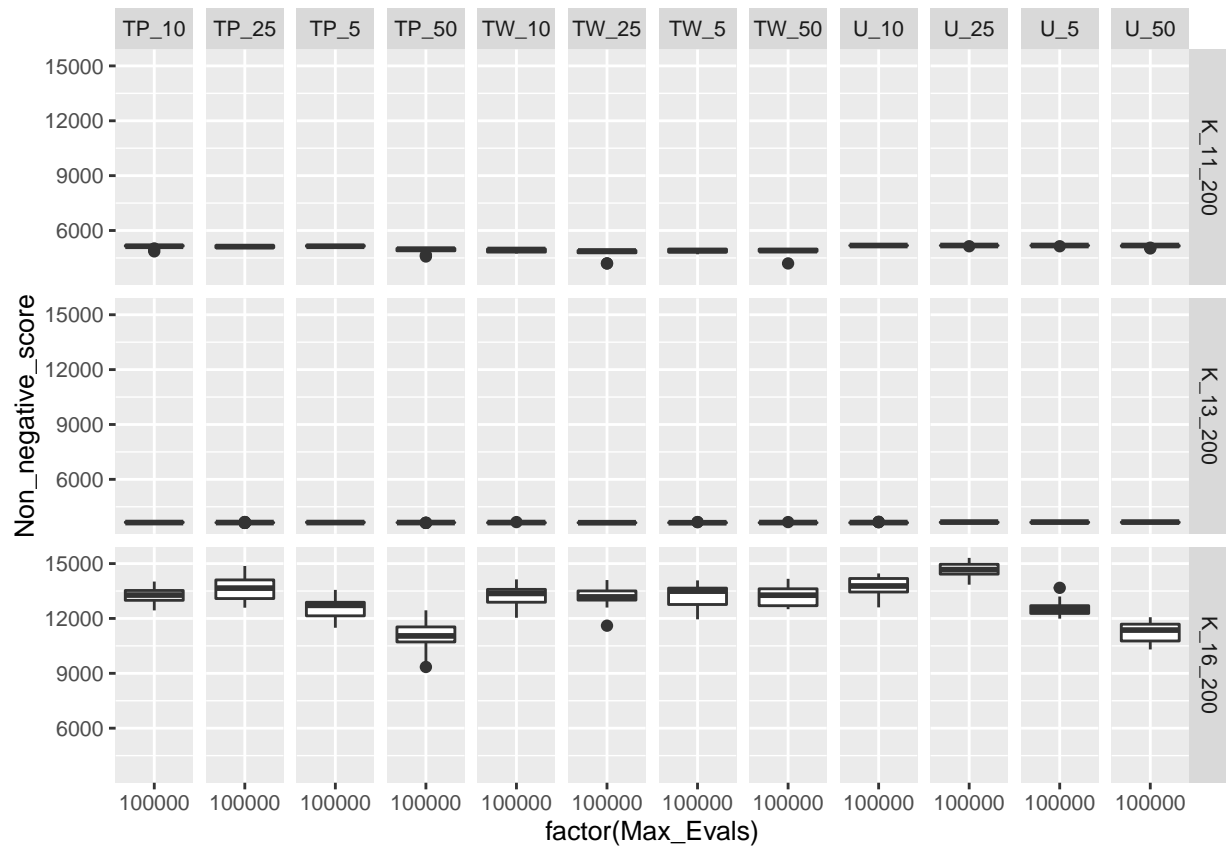
twohundred_item_problems = subset(pop_data_20_runs, Problem=="K_11_200" | Problem=="K_13_200" | Problem=="K_16_200")
twohundred_item_11 = subset(pop_data_20_runs, Problem=="K_11_200")
twohundred_item_13 = subset(pop_data_20_runs, Problem=="K_13_200")
twohundred_item_16 = subset(pop_data_20_runs, Problem=="K_16_200")

```

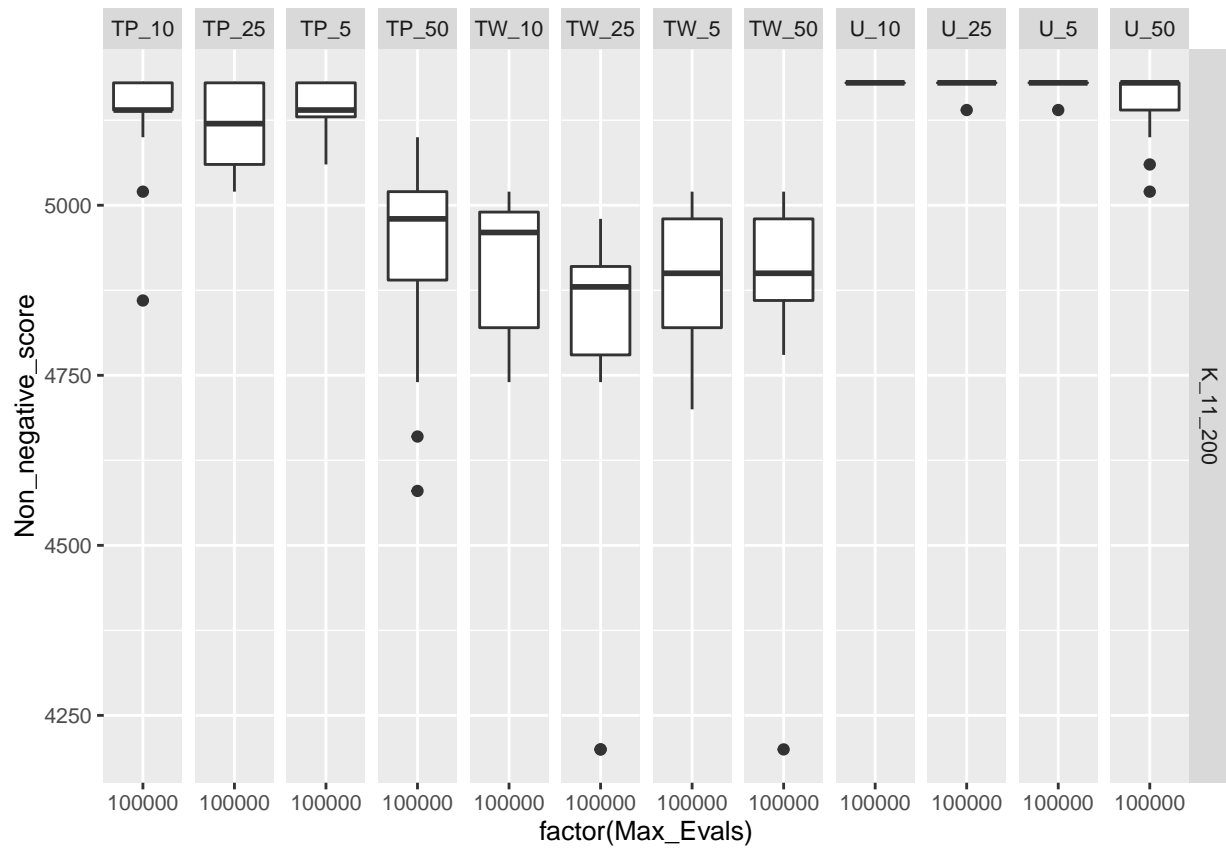
```

ggplot(twohundred_item_problems, aes(factor(Max_Evals), Non_negative_score)) + geom_boxplot() + facet_g

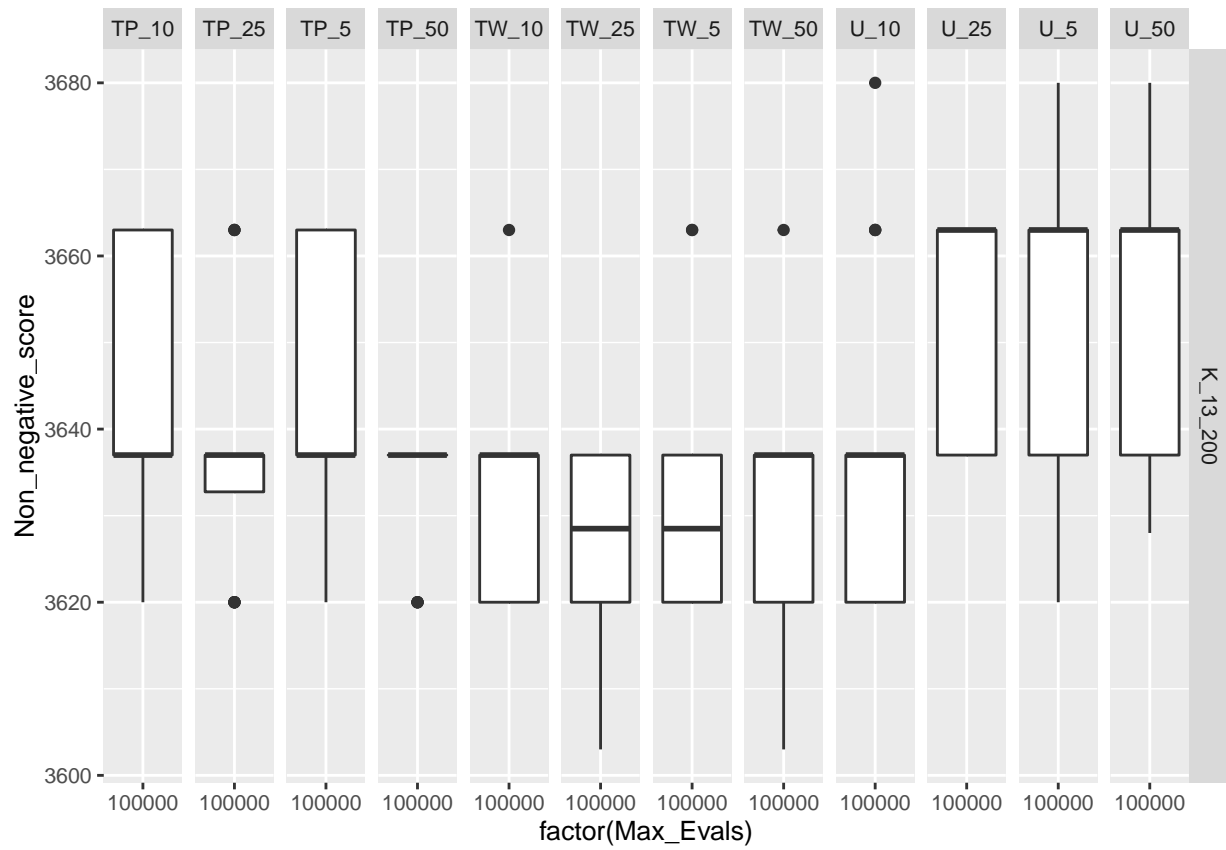
```

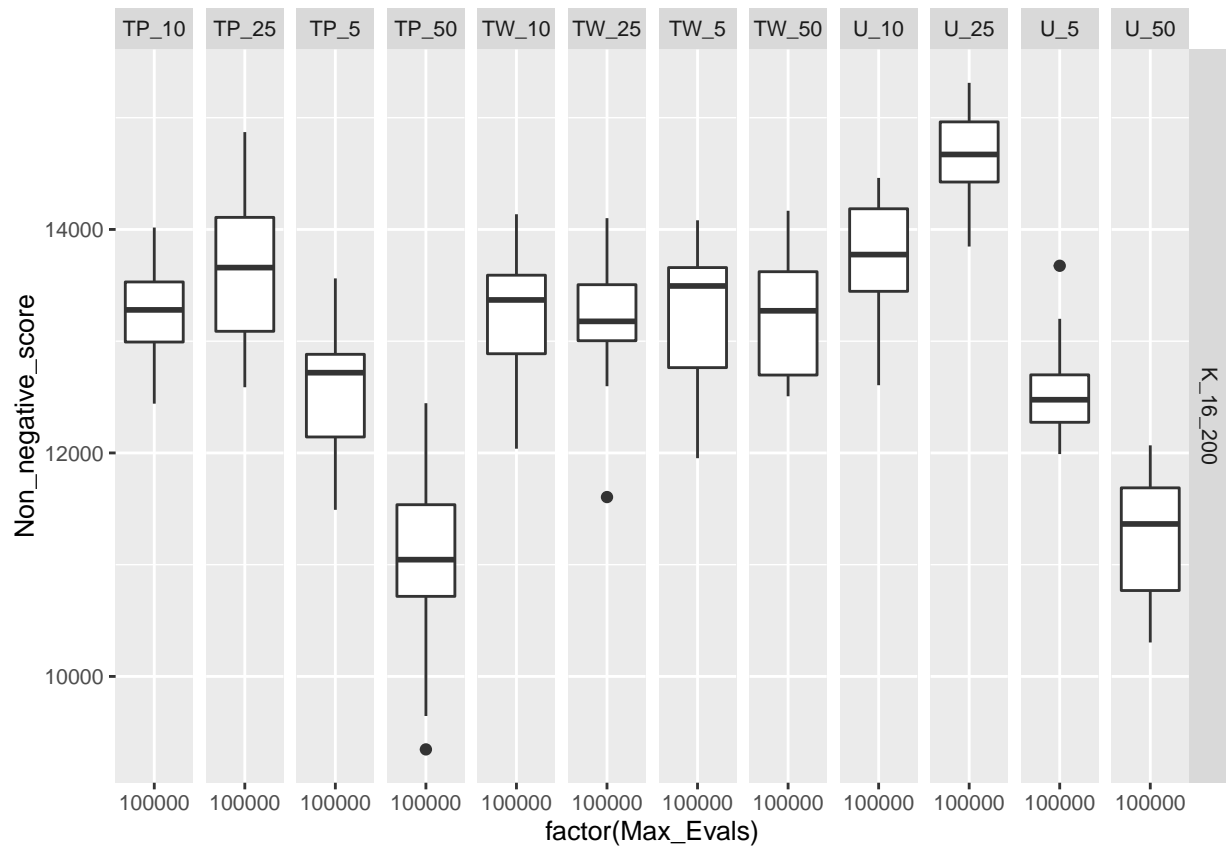
```
ggplot(twohundred_item_11, aes(factor(Max_Evals), Non_negative_score)) + geom_boxplot() + facet_grid(Pr
```



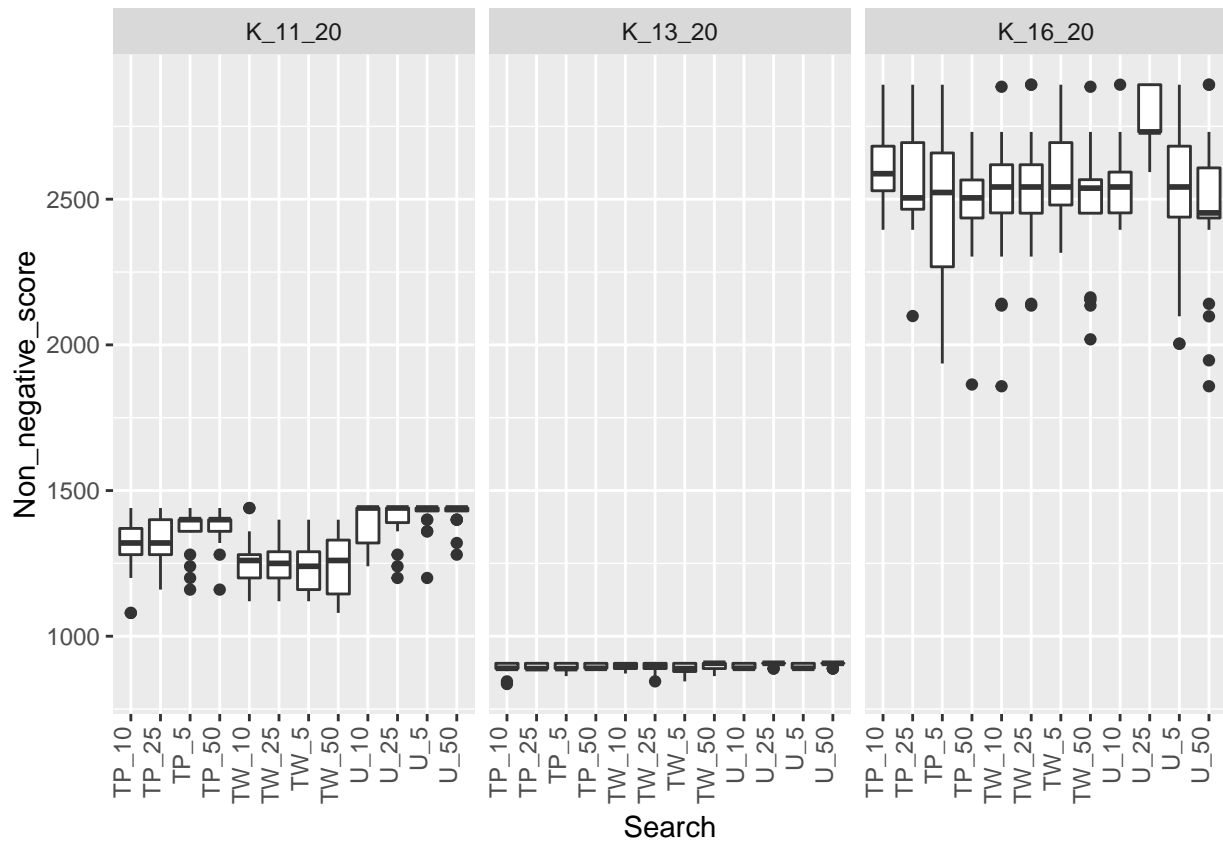
```
ggplot(twohundred_item_13, aes(factor(Max_Evals), Non_negative_score)) + geom_boxplot() + facet_grid(Pr
```



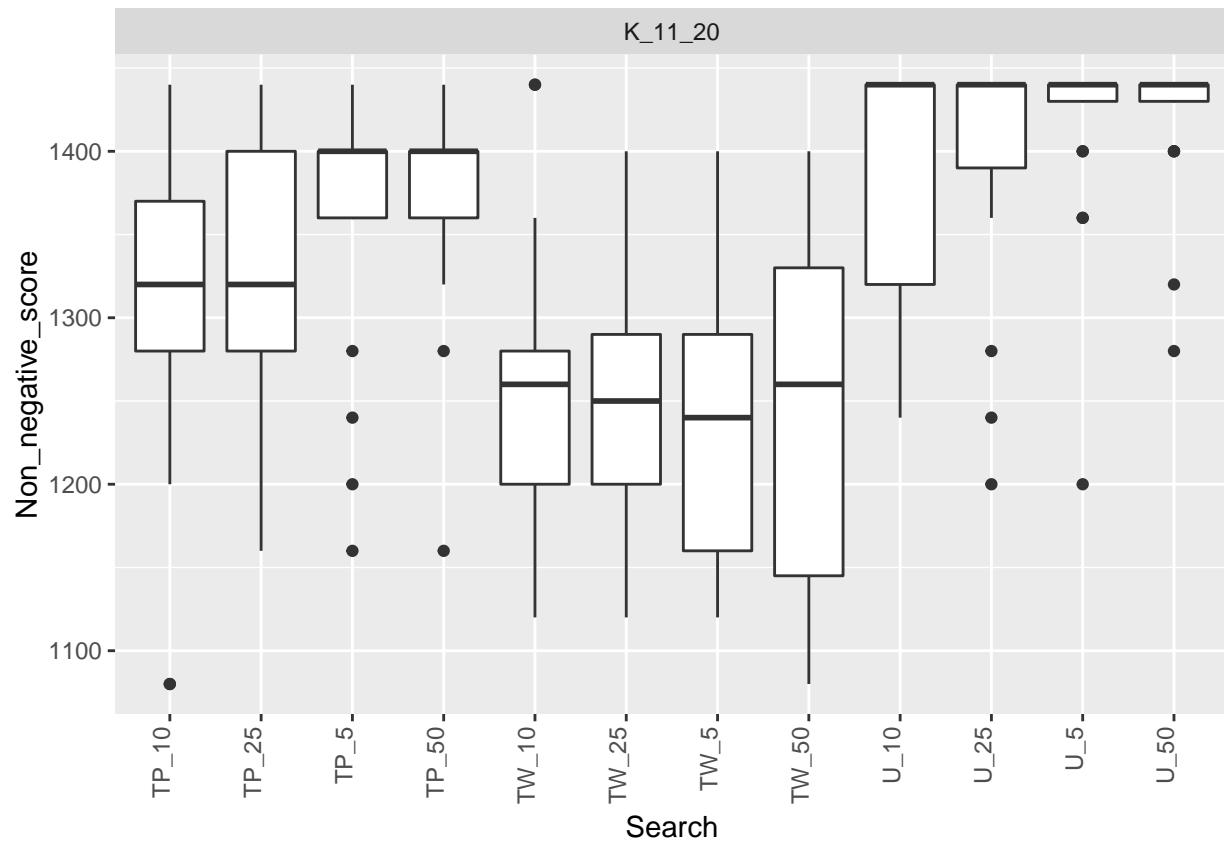
```
ggplot(twohundred_item_16, aes(factor(Max_Evals), Non_negative_score)) + geom_boxplot() + facet_grid(Pr
```



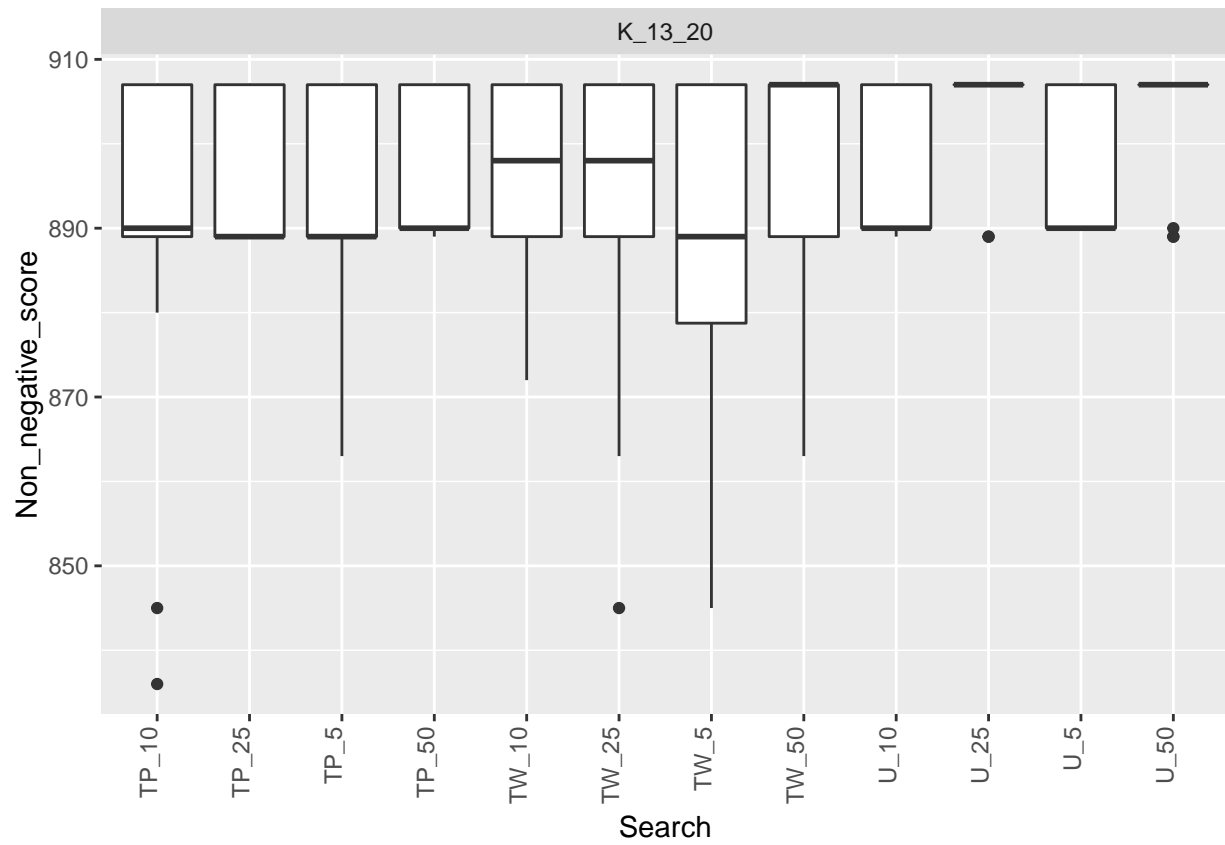
```
ggplot(twenty_item_problems, aes(Search, Non_negative_score)) + geom_boxplot() + facet_grid(. ~ Problem
```



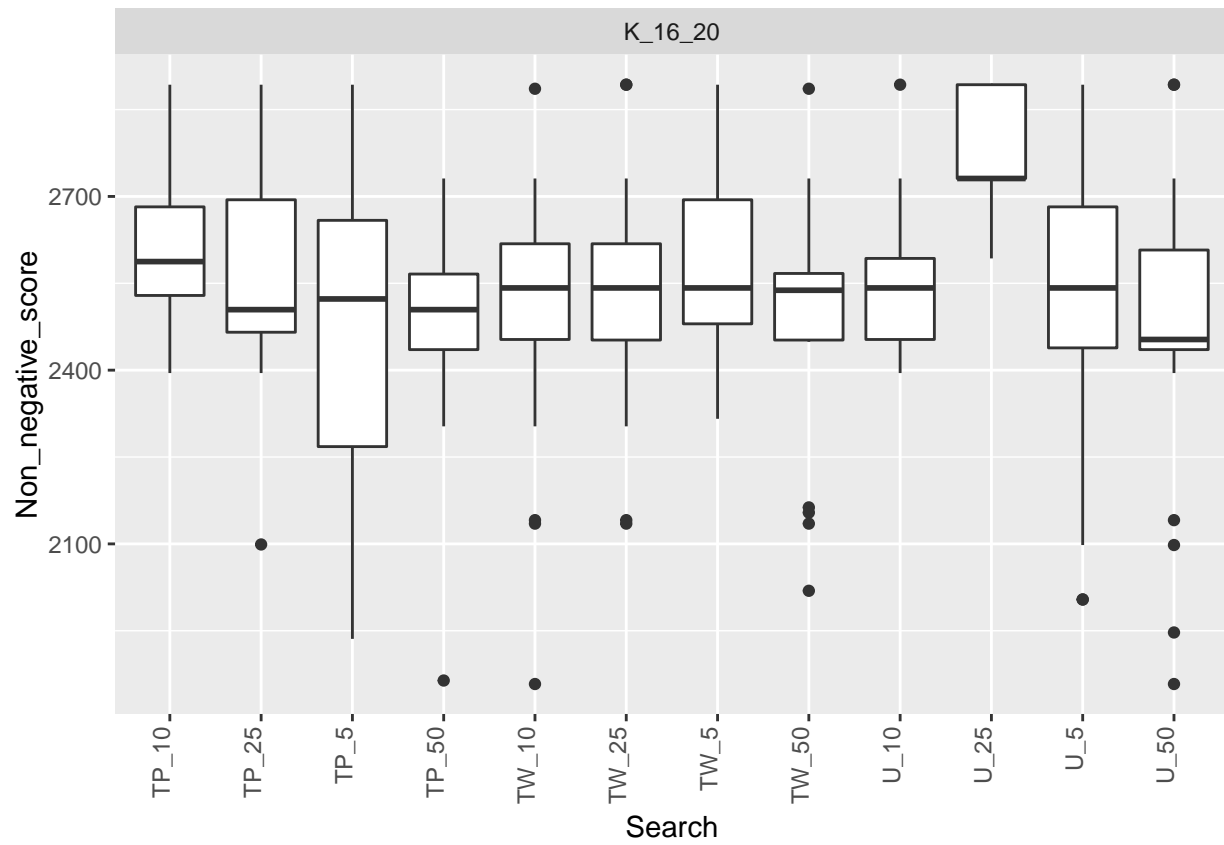
```
ggplot(twenty_item_11, aes(Search, Non_negative_score)) + geom_boxplot() + facet_grid(. ~ Problem) + theme_minimal()
```



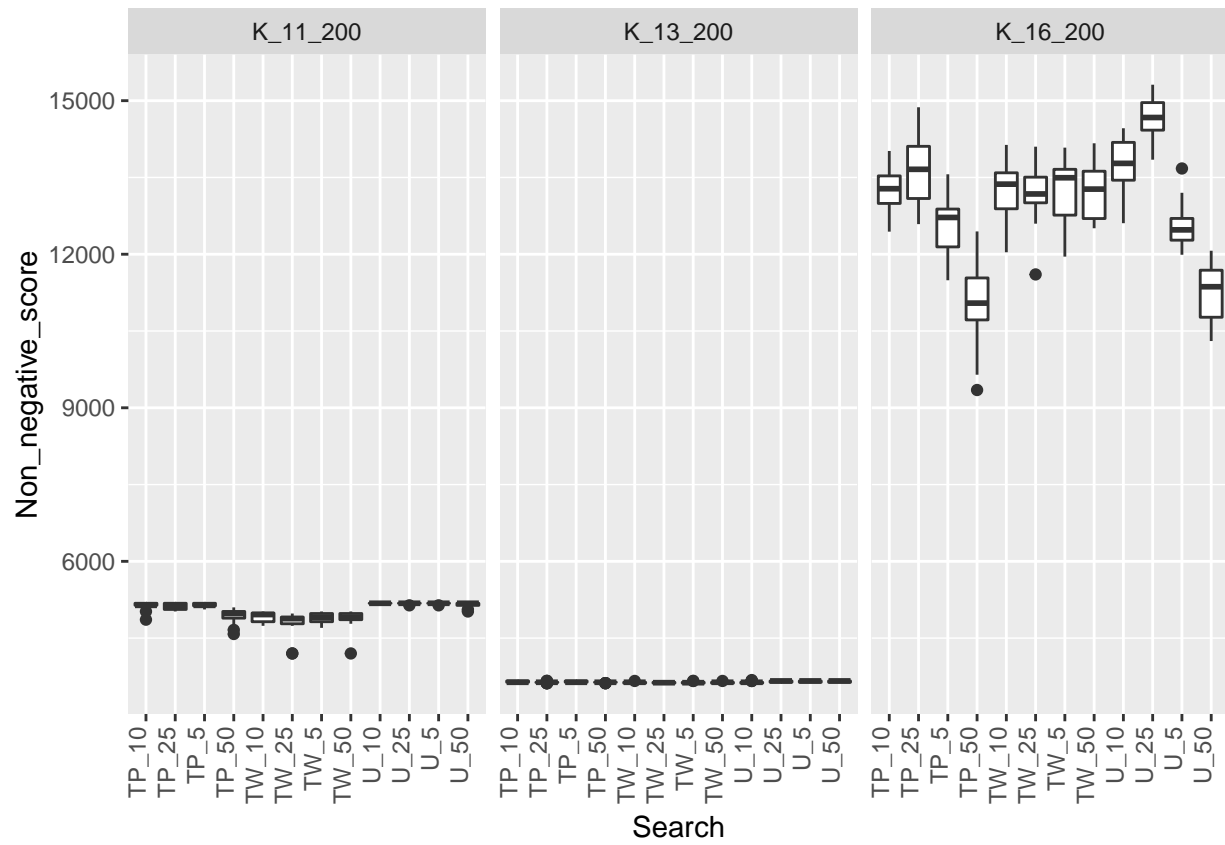
```
ggplot(twenty_item_13, aes(Search, Non_negative_score)) + geom_boxplot() + facet_grid(. ~ Problem) + theme_minimal()
```



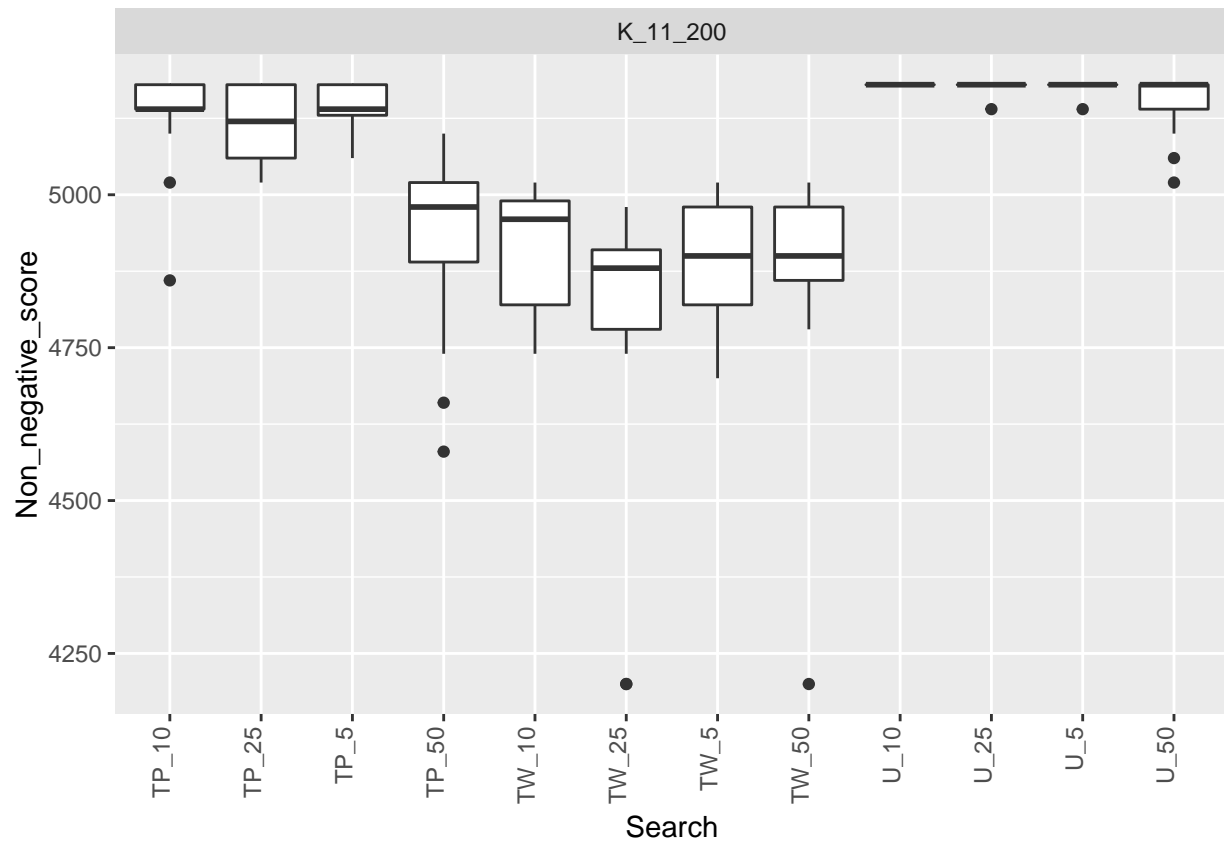
```
ggplot(twenty_item_16, aes(Search, Non_negative_score)) + geom_boxplot() + facet_grid(. ~ Problem) + theme_minimal()
```



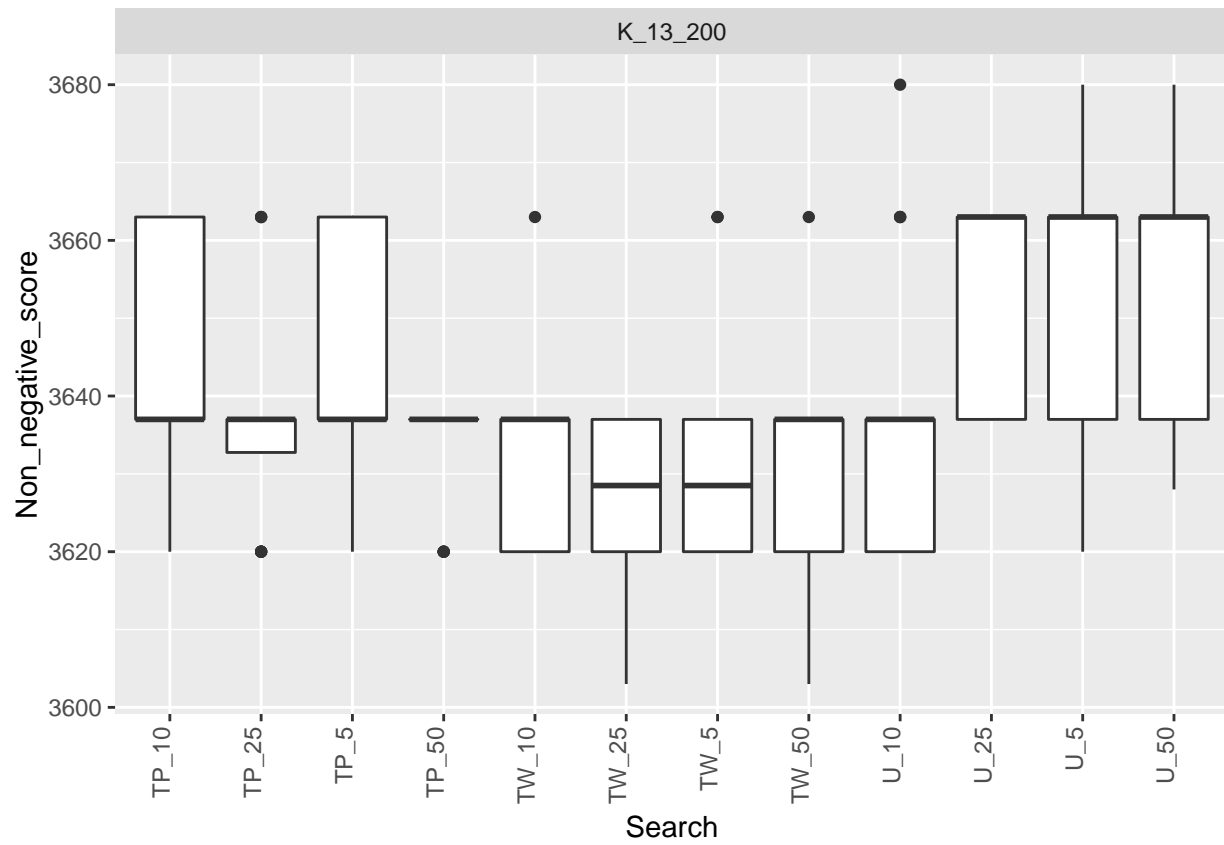
```
ggplot(twohundred_item_problems, aes(Search, Non_negative_score)) + geom_boxplot() + facet_grid(. ~ Problem)
```

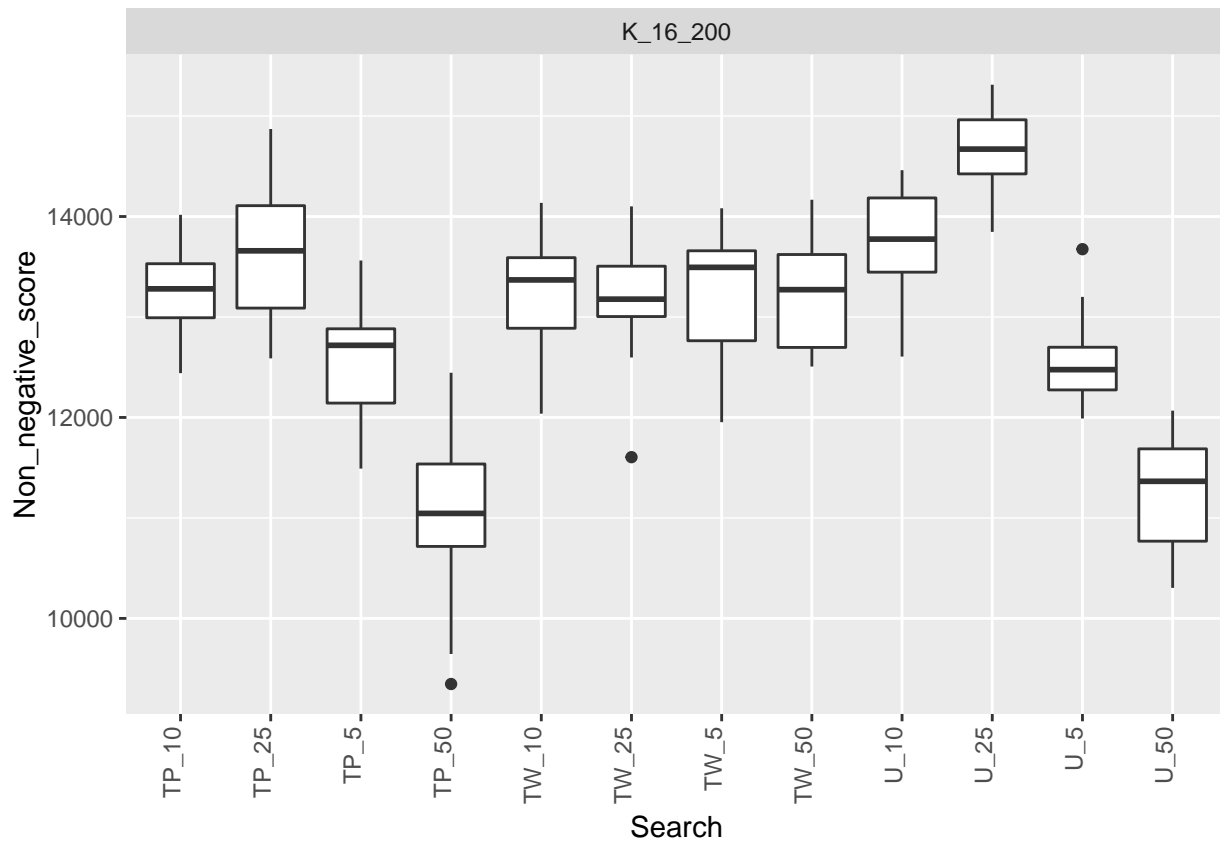
```
ggplot(twohundred_item_11, aes(Search, Non_negative_score)) + geom_boxplot() + facet_grid(. ~ Problem)+
```



```
ggplot(twohundred_item_13, aes(Search, Non_negative_score)) + geom_boxplot() + facet_grid(. ~ Problem)+
```



```
ggplot(twohundred_item_16, aes(Search, Non_negative_score)) + geom_boxplot() + facet_grid(. ~ Problem)+
```



```
{r} # thousand_item_problem_16 = subset(pop_data_20_runs, Problem=="K_16_200")
# ggplot(thousand_item_problem_16, aes(factor(Max_Evals), Non_negative_score))
# + geom_boxplot() + facet_grid(Problem ~ Search) + theme_grey(base_size = 10) #
```

```
# https://cran.r-project.org/web/packages/rpart.plot/rpart.plot.pdf
library("rpart")
library("rpart.plot")

rp <- rpart(Non_negative_score ~ Search + Problem + Max_Evals, data=pop_data_20_runs)
rp
```

```
## n= 1440
##
## node), split, n, deviance, yval
##      * denotes terminal node
##
## 1) root 1440 24240680000 4405.633
##    2) Problem=K_11_20,K_11_200,K_13_20,K_13_200,K_16_20 1200 2784022000 2691.198
##      4) Problem=K_11_20,K_13_20,K_16_20 720 364775000 1591.686
##        8) Problem=K_11_20,K_13_20 480 25402590 1114.356 *
##          9) Problem=K_16_20 240 11277300 2546.346 *
```

```
##      5) Problem=K_11_200,K_13_200 480    243181300  4340.465 *
##      3) Problem=K_16_200 240    293788800 12977.810 *
```

```
rpart.plot(rp, type=4, extra=100, Margin=0.0001)
```

