Survey Analysis

(1) Load the datasets

```
ds <- read.csv("survey-results.csv")
nrow(ds)
```

```
## [1] 2409
```

```
colnames(ds)
```

```
##  [1] "rid"        "sid"        "qid"        "atom"       "time"
##  [6] "correct"    "experience" "education"  "total"      "ref"
## [11] "empty"      "ssid"
```

```
atomId <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 0)
atomDescription <- c("Arithmetic as Logic", "Assignment as Value",
            "Automatic Semicolon Insertion",
           "Comma Operator", "Ternary Operator", "Implict Predicate",
           "Logic as Control Flow", "Ommitted Curly Braces", "Post Increment",
           "Pre Increment")

atom_ds <- data.frame(atomId, atomDescription)
```

(2) Filter the datasets

```
subSet <- ds[ds$ref=="reddit" & ds$total == "YES",]
summary(subSet$total)
```

```
##    Length     Class      Mode
##      1930 character character
```

```
squareLengths <- tapply(subSet$time,subSet$rid,length)
completeCases <- names(squareLengths)[squareLengths==20]
ds <- ds[is.element(el = ds$rid, set = completeCases),]
dim(ds)
```

```
## [1] 1400   12
```

(3) Exploratory Data Analysis

Demographics

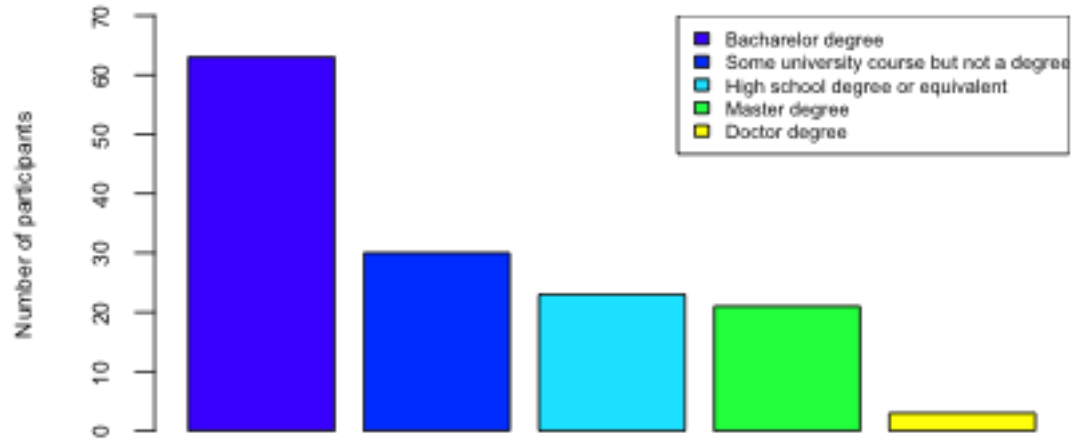
```
educationLevelIds <- c(1, 2, 3, 4, 5)
educationLevelLabels <- c("High school degree or equivalent",
                          "Some university course but not a degree",
                          "Bacharelor degree", "Master degree",
                          "Doctor degree")

education_ds <- data.frame(educationLevelIds, educationLevelLabels)
```

```
colnames(education_ds) <- c("id", "description")

tmp <- sqldf("select distinct sid, education from ds")

demEducation <- sqldf("select b.description, count(*) total
                      from tmp a, education_ds b
                      where a.education = b.id
                      group by b.description
                      order by 2 desc")

demEducation
```

Education

```
##                               description total
## 1                        Bacharelor degree    63
## 2 Some university course but not a degree    30
## 3        High school degree or equivalent    23
## 4                           Master degree    21
## 5                           Doctor degree     3
```

```
barplot(demEducation$total, col=topo.colors(5),
        ylim = c(0, 70), cex=0.7, cex.lab = 0.7, cex.axis=0.7,
        ylab="Number of participants")

legend("topright", legend=demEducation$description, fill=topo.colors(5), cex=0.6)
```



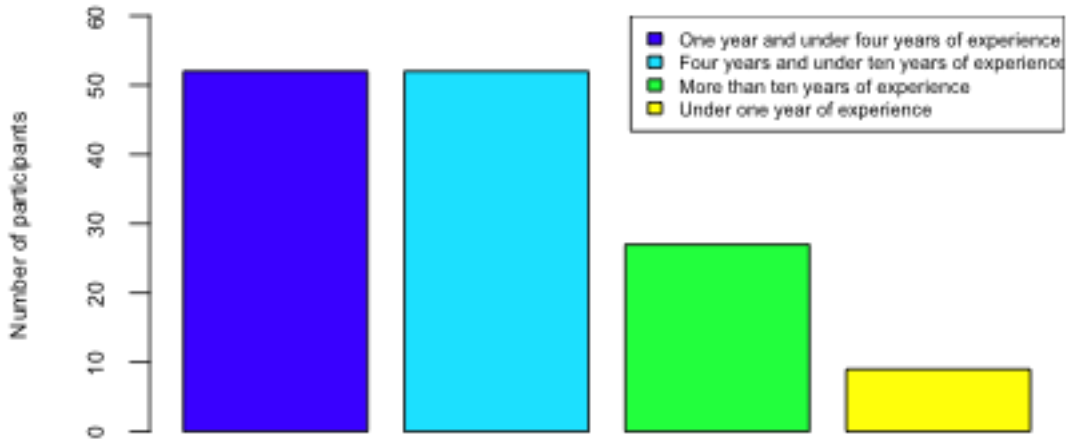

```
experienceLevelIds <- c(1, 2, 3, 4)
experienceLevelLabels <- c("Under one year of experience",
```

```
                          "One year and under four years of experience",
                          "Four years and under ten years of experience",
                          "More than ten years of experience")

experience_ds <- data.frame(experienceLevelIds, experienceLevelLabels)

colnames(experience_ds) <- c("id", "description")

tmp <- sqldf("select distinct sid, experience from ds")
demExperience <- sqldf("select experience, count(*) total
                       from tmp group by experience order by 1")

demExperience <- sqldf("select a.experience, b.description, a.total
                        from demExperience a, experience_ds b
                        where a.experience = b.id
                        order by 3 desc")
demExperience
```

Experience

```
##   experience                                 description total
## 1          2  One year and under four years of experience    52
## 2          3 Four years and under ten years of experience    52
## 3          4            More than ten years of experience    27
## 4          1                 Under one year of experience     9
```

```
barplot(demExperience$total, col=topo.colors(4),
        ylim = c(0, 60), cex=0.7, cex.lab = 0.7, cex.axis=0.7,
        ylab="Number of participants")

legend("topright", legend=demExperience$description, fill=topo.colors(4), cex=0.6)
```



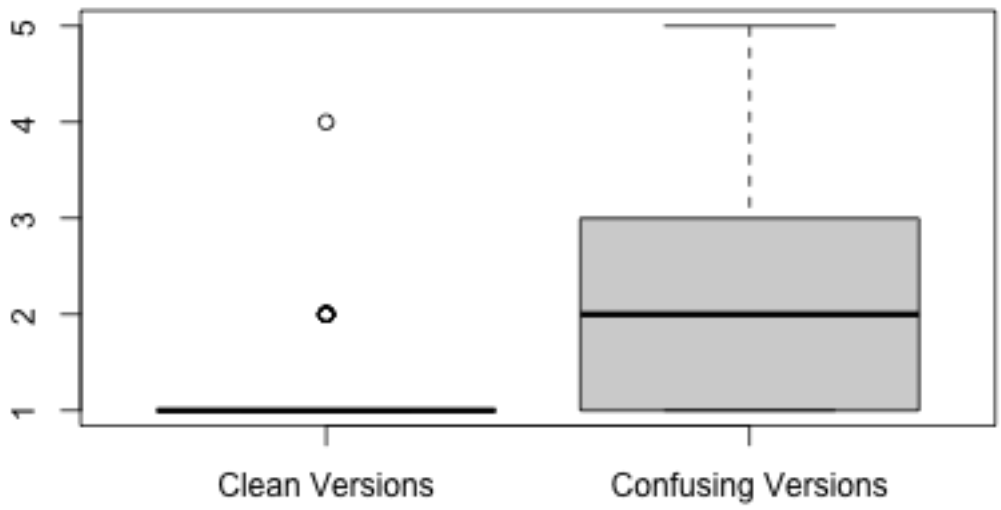
Total number of correct answers (Table III)

```
codeWithAtoms <- sqldf("select qid, count(*) confuseCode, avg(time) timeConfuseCode
                        from ds
                        where atom == 'YES' and correct = 'CORRECT'
                        group by qid")

codeWithoutAtoms <- sqldf("select qid, count(*) cleanCode, avg(time) timeCleanCode
                        from ds
                        where atom == 'NO' and correct = 'CORRECT'
                        group by qid")

codeWithAtoms["atomId"] = codeWithAtoms$qid %% 10
codeWithoutAtoms["atomId"] = codeWithoutAtoms$qid %% 10

merged <- sqldf("select c.atomDescription as Atom,
                        a.confuseCode as 'Confusing Versions',
                        b.cleanCode as 'Clean Versions',
                        (b.cleanCode * 100 / a.confuseCode) -100 as 'Delta (%)'
                from codeWithAtoms a, codeWithoutAtoms b, atom_ds c
                where a.atomId = b.atomId and a.atomId = c.atomId
                order by 4 desc")

xtable(merged)
```

```
## % latex table generated in R 4.2.0 by xtable 1.8-4 package
## % Fri Jul 29 09:07:12 2022
## \begin{table}[ht]
## \centering
```

```
## \begin{tabular}{rlrrr}
##   \hline
##  & Atom & Confusing Versions & Clean Versions & Delta (\%) \\
##   \hline
## 1 & Comma Operator &  28 &  65 & 132 \\
##   2 & Automatic Semicolon Insertion &  32 &  68 & 112 \\
##   3 & Post Increment &  48 &  64 &  33 \\
##   4 & Ommitted Curly Braces &  47 &  58 &  23 \\
##   5 & Assignment as Value &  56 &  68 &  21 \\
##   6 & Implict Predicate &  58 &  68 &  17 \\
##   7 & Logic as Control Flow &  41 &  48 &  17 \\
##   8 & Ternary Operator &  60 &  66 &  10 \\
##   9 & Pre Increment &  50 &  53 &   6 \\
##   10 & Arithmetic as Logic &  64 &  63 &  -2 \\
##    \hline
## \end{tabular}
## \end{table}
```

```
wrongAnswers = ds[ds$correct == 'WRONG', ]

wrongAnswersByStudentTreatment <- aggregate(rid~sid+atom,
                                            data = wrongAnswers,
                                            FUN=length)

boxplot(wrongAnswersByStudentTreatment$rid ~ wrongAnswersByStudentTreatment$atom
        , ylab = "", xlab = "", main = "",
        names = c("Clean Versions", "Confusing Versions"))
```

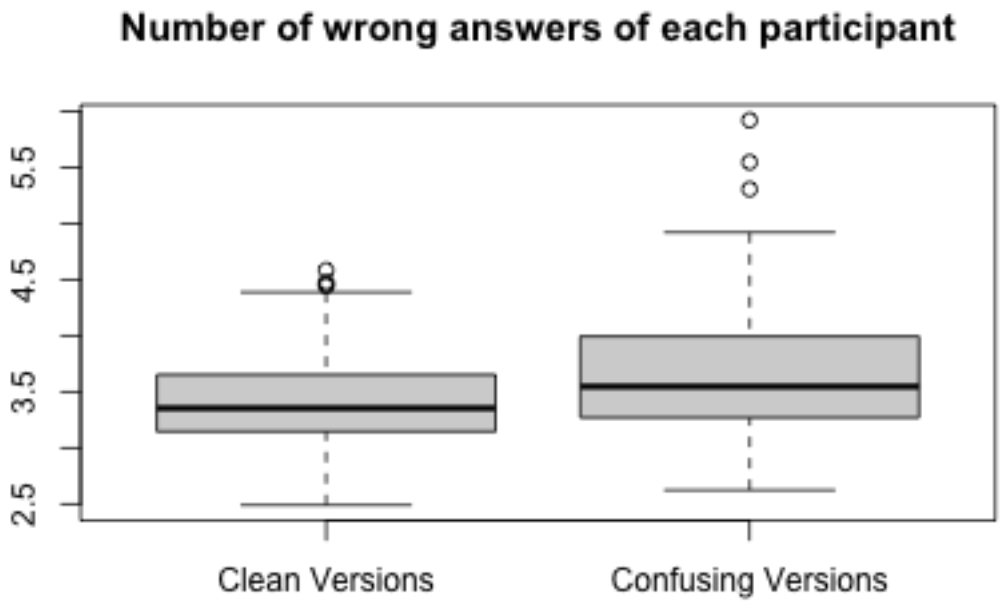


Average time for correct answers (Table IV)

```
merged <- sqldf("select c.atomDescription as Atom,
                        a.timeConfuseCode as 'Confuse Code',
                        b.timeCleanCode as 'Clean Code',
                        (b.timeCleanCode * 100 / a.timeConfuseCode) -100 as 'Delta (%)'
                from codeWithAtoms a, codeWithoutAtoms b, atom_ds c
                where a.atomId = b.atomId and a.atomId = c.atomId
                order by 4")
xtable(merged)
```

```
## % latex table generated in R 4.2.0 by xtable 1.8-4 package
## % Fri Jul 29 09:07:12 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{rlrrr}
##   \hline
##  & Atom & Confuse Code & Clean Code & Delta (\%) \\
##   \hline
## 1 & Comma Operator & 87.67 & 20.84 & -76.23 \\
##   2 & Logic as Control Flow & 108.94 & 51.07 & -53.12 \\
##   3 & Automatic Semicolon Insertion & 46.08 & 22.04 & -52.17 \\
##   4 & Ommitted Curly Braces & 48.85 & 30.00 & -38.58 \\
##   5 & Implict Predicate & 36.24 & 24.01 & -33.75 \\
##   6 & Post Increment & 28.70 & 25.67 & -10.56 \\
##   7 & Assignment as Value & 52.47 & 48.95 & -6.71 \\
##   8 & Ternary Operator & 41.80 & 39.34 & -5.90 \\
##   9 & Arithmetic as Logic & 28.82 & 37.20 & 29.06 \\
##   10 & Pre Increment & 30.71 & 42.45 & 38.19 \\
##    \hline
## \end{tabular}
## \end{table}
```

```
correctAnswers = ds[ds$correct == 'CORRECT', ]

correctAnswersByStudentTreatment <- aggregate(time~sid+atom,
                                              data = correctAnswers,
                                              FUN=mean)

boxplot(log(correctAnswersByStudentTreatment$time) ~ correctAnswersByStudentTreatment$atom
        , ylab = "", xlab = "", main = "Number of wrong answers of each participant",
        names = c("Clean Versions", "Confusing Versions"))
```




```
prop.table(table(ds$correct,ds$atom),margin = 2)
```

```
##
##                   NO       YES
##   CORRECT 0.8871429 0.6914286
##   WRONG   0.1128571 0.3085714
```

```
chisq.test(ds$correct,ds$atom)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  ds$correct and ds$atom
## X-squared = 79.437, df = 1, p-value < 2.2e-16
```

(4) Regression Analysis (correctness~atom + experience + education)

```
experience <- as.factor(ds$experience)
education <- as.factor(ds$education)
contrr<-matrix(c(rep(1,4),c(1/2,1/2,-1/2,-1/2),c(1,-1,0,0),c(0,0,1,-1)),byrow=TRUE,nrow=4)
contrasts(experience)<-solve(contrr)[,2:4]

ds$atom <- as.factor(ds$atom)
ds$correct <- as.factor(ds$correct)

mod <- glm(ds$correct~ds$atom+experience+education,family = "binomial")
summary(mod)
```

```
##
## Call:
```

```
## glm(formula = ds$correct ~ ds$atom + experience + education,
##     family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2740  -0.7579  -0.4997  -0.4014   2.3422
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.16306    0.20559 -10.521  < 2e-16 ***
## ds$atomYES   1.28111    0.14650   8.745  < 2e-16 ***
## experience1  0.82857    0.16797   4.933 8.11e-07 ***
## experience2  0.79443    0.25847   3.074  0.00212 **
## experience3  0.19794    0.20901   0.947  0.34362
## education2   0.02462    0.23495   0.105  0.91656
## education3   0.29480    0.20543   1.435  0.15128
## education4   0.46094    0.24952   1.847  0.06471 .
## education5   0.27158    0.50686   0.536  0.59210
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1441.7  on 1399  degrees of freedom
## Residual deviance: 1332.0  on 1391  degrees of freedom
## AIC: 1350
##
## Number of Fisher Scoring iterations: 4
```

```
car::Anova(mod, type=3)
```

```
## Analysis of Deviance Table (Type III tests)
##
## Response: ds$correct
##            LR Chisq Df Pr(>Chisq)
## ds$atom      84.819  1  < 2.2e-16 ***
## experience   24.181  3   2.29e-05 ***
## education     5.542  4     0.2361
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Impact considering Developer Experience

```
for(i in 1:4){
  print(paste("Experience ", i ))
  print(prop.table(table(ds$correct[ds$experience==i],ds$atom[ds$experience==i]),margin=2))
  print(chisq.test(table(ds$correct[ds$experience==i],ds$atom[ds$experience==i])))
}
```

```
## [1] "Experience  1"
##
##                  NO       YES
##   CORRECT 0.8444444 0.4222222
##   WRONG   0.1555556 0.5777778
##
```

```
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(ds$correct[ds$experience == i], ds$atom[ds$experience ==     i])
## X-squared = 15.502, df = 1, p-value = 8.24e-05
##
## [1] "Experience  2"
##
##                  NO       YES
##   CORRECT 0.8692308 0.6615385
##   WRONG   0.1307692 0.3384615
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(ds$correct[ds$experience == i], ds$atom[ds$experience ==     i])
## X-squared = 30.082, df = 1, p-value = 4.141e-08
##
## [1] "Experience  3"
##
##                   NO        YES
##   CORRECT 0.90384615 0.72307692
##   WRONG   0.09615385 0.27692308
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(ds$correct[ds$experience == i], ds$atom[ds$experience ==     i])
## X-squared = 26.817, df = 1, p-value = 2.237e-07
##
## [1] "Experience  4"
##
##                  NO       YES
##   CORRECT 0.9037037 0.7777778
##   WRONG   0.0962963 0.2222222
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(ds$correct[ds$experience == i], ds$atom[ds$experience ==     i])
## X-squared = 7.0812, df = 1, p-value = 0.00779
```

Impact considering Developer Education

```
for(i in 1:5){
  print(paste("Education ", i ))
  print(prop.table(table(ds$correct[ds$education==i],ds$atom[ds$education==i]),margin=2))
  print(chisq.test(table(ds$correct[ds$education==i],ds$atom[ds$education==i])))
}
```

```
## [1] "Education  1"
##
##                   NO        YES
##   CORRECT 0.92173913 0.69565217
##   WRONG   0.07826087 0.30434783
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
```

```
## data:  table(ds$correct[ds$education == i], ds$atom[ds$education ==     i])
## X-squared = 17.565, df = 1, p-value = 2.777e-05
##
## [1] "Education  2"
##
##                  NO       YES
##   CORRECT 0.8933333 0.7266667
##   WRONG   0.1066667 0.2733333
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(ds$correct[ds$education == i], ds$atom[ds$education ==     i])
## X-squared = 12.476, df = 1, p-value = 0.0004123
##
## [1] "Education  3"
##
##                  NO       YES
##   CORRECT 0.8793651 0.6761905
##   WRONG   0.1206349 0.3238095
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(ds$correct[ds$education == i], ds$atom[ds$education ==     i])
## X-squared = 36.45, df = 1, p-value = 1.566e-09
##
## [1] "Education  4"
##
##                  NO       YES
##   CORRECT 0.8666667 0.6761905
##   WRONG   0.1333333 0.3238095
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(ds$correct[ds$education == i], ds$atom[ds$education ==     i])
## X-squared = 9.7492, df = 1, p-value = 0.001794
##
## [1] "Education  5"
##
##                  NO       YES
##   CORRECT 0.8666667 0.7333333
##   WRONG   0.1333333 0.2666667
## Warning in chisq.test(table(ds$correct[ds$education == i], ds$atom[ds$education
## == : Aproximação do qui-quadrado pode estar incorreta
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(ds$correct[ds$education == i], ds$atom[ds$education ==     i])
## X-squared = 0.20833, df = 1, p-value = 0.6481
```

Impact of Individual Atoms and Hypotheses Testing

```r
chiTest <- c()
oddsRatio <- c()
oddsRatioTest <- c()
ci25 <- c()
ci975 <- c()
mannWhitneyTest <- c()
cliffDelta <- c()
for(i in 1:10){
  subSet <- ds[is.element(el = ds$qid, set = c(i,i+10)),]
  print(atomDescription[i])

  tableCorrectness <- table(subSet$correct,subSet$atom)
  tableTime <- aggregate(ds$time, by=list(atom=ds$atom), FUN=sum)

  print(tableCorrectness)
  print(tableTime)

  test <- chisq.test(tableCorrectness)
  print(test)

  chiTest[i] <- format.pval(test$p.value)
  oddsRatio[i] <- odds.ratio(tableCorrectness)$OR
  oddsRatioTest[i] <- format.pval(odds.ratio(tableCorrectness)$p)
  ci25[i] <- odds.ratio(tableCorrectness, level=0.95)$"2.5 %"
  ci975[i] <- odds.ratio(tableCorrectness, level=0.95)$"97.5 %"

  mannWhitneyTest[i] <- format.pval(wilcox.test(subSet$time~as.factor(subSet$atom))$p.value)
  cliffDelta[i] <- cliff.delta(subSet$time~as.factor(subSet$atom))$estimate

  experience <- as.factor(subSet$experience)
  contrr<-matrix(c(rep(1,4),c(1/2,1/2,-1/2,-1/2),c(1,-1,0,0),c(0,0,1,-1)),byrow=TRUE,nrow=4)
  contrasts(experience)<-solve(contrr)[,2:4]

  mod <- glm(subSet$correct ~ subSet$atom +  experience,family="binomial")
  print(summary(mod))
  print(prop.table(table(subSet$correct,subSet$experience),margin = 2))
}
```

```
## [1] "Arithmetic as Logic"
##
##           NO YES
##   CORRECT 63  64
##   WRONG    7   6
##   atom        x
## 1   NO 24669.70
## 2  YES 30821.46
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tableCorrectness
## X-squared = 0, df = 1, p-value = 1
##
##
## Call:
```

```
## glm(formula = subSet$correct ~ subSet$atom + experience, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.5239  -0.4750  -0.4165  -0.3768   2.3176
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.195722   0.436544  -5.030 4.91e-07 ***
## subSet$atomYES -0.208164   0.607117  -0.343    0.732
## experience1     0.416695   0.733776   0.568    0.570
## experience2    -0.141108   1.180049  -0.120    0.905
## experience3     0.005588   0.906594   0.006    0.995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 86.548  on 139  degrees of freedom
## Residual deviance: 85.817  on 135  degrees of freedom
## AIC: 95.817
##
## Number of Fisher Scoring iterations: 5
##
##
##                    1          2          3          4
##   CORRECT 0.88888889 0.88461538 0.92307692 0.92592593
##   WRONG   0.11111111 0.11538462 0.07692308 0.07407407
## [1] "Assignment as Value"
##
##           NO YES
##   CORRECT 68  56
##   WRONG    2  14
##   atom        x
## 1   NO 24669.70
## 2  YES 30821.46
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tableCorrectness
## X-squared = 8.5383, df = 1, p-value = 0.003477
##
##
## Call:
## glm(formula = subSet$correct ~ subSet$atom + experience, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.8003  -0.5414  -0.2823  -0.1879   2.8445
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -3.52018    0.75652  -4.653 3.27e-06 ***
## subSet$atomYES  2.18189    0.78860   2.767  0.00566 **
```

```
## experience1     0.68175    0.62236   1.095  0.27333
## experience2     0.04643    0.93786   0.050  0.96052
## experience3    -0.33362    0.82736  -0.403  0.68677
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 99.507  on 139  degrees of freedom
## Residual deviance: 86.443  on 135  degrees of freedom
## AIC: 96.443
##
## Number of Fisher Scoring iterations: 6
##
##
##                    1          2          3          4
##   CORRECT 0.77777778 0.86538462 0.92307692 0.88888889
##   WRONG   0.22222222 0.13461538 0.07692308 0.11111111
## [1] "Automatic Semicolon Insertion"
##
##           NO YES
##   CORRECT 68  32
##   WRONG    2  38
##   atom        x
## 1   NO 24669.70
## 2  YES 30821.46
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tableCorrectness
## X-squared = 42.875, df = 1, p-value = 5.835e-11
##
##
## Call:
## glm(formula = subSet$correct ~ subSet$atom + experience, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4023  -0.7246  -0.1463   0.9680   3.1705
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -3.6646     0.8465  -4.329 1.50e-05 ***
## subSet$atomYES   4.4184     1.0454   4.226 2.37e-05 ***
## experience1      1.5939     0.7601   2.097   0.0360 *
## experience2      3.3287     1.3569   2.453   0.0142 *
## experience3      1.1158     0.7245   1.540   0.1235
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 167.52  on 139  degrees of freedom
## Residual deviance: 104.25  on 135  degrees of freedom
```

```
## AIC: 114.25
##
## Number of Fisher Scoring iterations: 7
##
##
##                   1         2         3         4
##   CORRECT 0.4444444 0.7500000 0.6538462 0.8518519
##   WRONG   0.5555556 0.2500000 0.3461538 0.1481481
## [1] "Comma Operator"
##
##           NO YES
##   CORRECT 65  28
##   WRONG    5  42
##   atom        x
## 1   NO 24669.70
## 2  YES 30821.46
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tableCorrectness
## X-squared = 41.51, df = 1, p-value = 1.173e-10
##
##
## Call:
## glm(formula = subSet$correct ~ subSet$atom + experience, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8303  -0.4500  -0.3159   0.8191   2.3360
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -2.4822     0.5161  -4.809 1.51e-06 ***
## subSet$atomYES   3.1589     0.5571   5.671 1.42e-08 ***
## experience1      1.0339     0.5701   1.813   0.0698 .
## experience2      0.5478     0.9525   0.575   0.5652
## experience3     -0.6762     0.6213  -1.088   0.2764
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 178.68  on 139  degrees of freedom
## Residual deviance: 124.07  on 135  degrees of freedom
## AIC: 134.07
##
## Number of Fisher Scoring iterations: 5
##
##
##                   1         2         3         4
##   CORRECT 0.3333333 0.6538462 0.7307692 0.6666667
##   WRONG   0.6666667 0.3461538 0.2692308 0.3333333
## [1] "Ternary Operator"
##
```

```
##           NO YES
##   CORRECT 66  60
##   WRONG    4  10
##   atom        x
## 1   NO 24669.70
## 2  YES 30821.46
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tableCorrectness
## X-squared = 1.9841, df = 1, p-value = 0.159
##
##
## Call:
## glm(formula = subSet$correct ~ subSet$atom + experience, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3479  -0.4064  -0.3331  -0.2303   2.7428
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -2.5416     0.5876  -4.325 1.52e-05 ***
## subSet$atomYES   0.7521     0.6662   1.129   0.2589
## experience1      2.2710     0.7450   3.048   0.0023 **
## experience2      2.0928     0.8140   2.571   0.0101 *
## experience3      0.1217     1.2531   0.097   0.9227
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 91.023  on 139  degrees of freedom
## Residual deviance: 73.740  on 135  degrees of freedom
## AIC: 83.74
##
## Number of Fisher Scoring iterations: 6
##
##
##                    1          2          3          4
##   CORRECT 0.44444444 0.88461538 0.96153846 0.96296296
##   WRONG   0.55555556 0.11538462 0.03846154 0.03703704
## [1] "Implict Predicate"
##
##           NO YES
##   CORRECT 68  58
##   WRONG    2  12
##   atom        x
## 1   NO 24669.70
## 2  YES 30821.46
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tableCorrectness
```

```
## X-squared = 6.4286, df = 1, p-value = 0.01123
##
##
## Call:
## glm(formula = subSet$correct ~ subSet$atom + experience, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7262  -0.4777  -0.2853  -0.1902   2.5383
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -7.4953   514.2636  -0.015   0.9884
## subSet$atomYES    1.8906     0.7891   2.396   0.0166 *
## experience1      -7.8982  1028.5263  -0.008   0.9939
## experience2     -16.5270  2057.0524  -0.008   0.9936
## experience3      -0.9149     0.8265  -1.107   0.2683
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 91.023  on 139  degrees of freedom
## Residual deviance: 78.736  on 135  degrees of freedom
## AIC: 88.736
##
## Number of Fisher Scoring iterations: 17
##
##
##                    1          2          3          4
##   CORRECT 1.00000000 0.86538462 0.94230769 0.85185185
##   WRONG   0.00000000 0.13461538 0.05769231 0.14814815
## [1] "Logic as Control Flow"
##
##           NO YES
##   CORRECT 48  41
##   WRONG   22  29
##   atom        x
## 1   NO 24669.70
## 2  YES 30821.46
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tableCorrectness
## X-squared = 1.1104, df = 1, p-value = 0.292
##
##
## Call:
## glm(formula = subSet$correct ~ subSet$atom + experience, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3336  -0.9912  -0.8623   1.3478   1.5578
##
```

```
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.62015    0.30152  -2.057   0.0397 *
## subSet$atomYES  0.40565    0.36155   1.122   0.2619
## experience1     0.39686    0.44376   0.894   0.3711
## experience2     0.75213    0.74016   1.016   0.3095
## experience3    -0.08443    0.49632  -0.170   0.8649
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 183.64  on 139  degrees of freedom
## Residual deviance: 180.86  on 135  degrees of freedom
## AIC: 190.86
##
## Number of Fisher Scoring iterations: 4
##
##
##                   1         2         3         4
##   CORRECT 0.4444444 0.6538462 0.6538462 0.6296296
##   WRONG   0.5555556 0.3461538 0.3461538 0.3703704
## [1] "Ommitted Curly Braces"
##
##           NO YES
##   CORRECT 58  47
##   WRONG   12  23
##   atom        x
## 1   NO 24669.70
## 2  YES 30821.46
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tableCorrectness
## X-squared = 3.8095, df = 1, p-value = 0.05096
##
##
## Call:
## glm(formula = subSet$correct ~ subSet$atom + experience, family = "binomial")
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.03855  -0.81387  -0.59259  -0.03331   2.29896
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -1.5527     0.3500  -4.436 9.14e-06 ***
## subSet$atomYES   0.9182     0.4235   2.168   0.0302 *
## experience1      1.3164     0.5200   2.531   0.0114 *
## experience2      0.7189     0.7596   0.946   0.3439
## experience3      0.7157     0.7085   1.010   0.3124
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 157.45  on 139  degrees of freedom
## Residual deviance: 145.65  on 135  degrees of freedom
## AIC: 155.65
##
## Number of Fisher Scoring iterations: 4
##
##
##                   1         2         3         4
##   CORRECT 0.5555556 0.6730769 0.7884615 0.8888889
##   WRONG   0.4444444 0.3269231 0.2115385 0.1111111
## [1] "Post Increment"
##
##           NO YES
##   CORRECT 64  48
##   WRONG    6  22
##   atom        x
## 1   NO 24669.70
## 2  YES 30821.46
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tableCorrectness
## X-squared = 10.045, df = 1, p-value = 0.001528
##
##
## Call:
## glm(formula = subSet$correct ~ subSet$atom + experience, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9873  -0.7436  -0.4835  -0.3492   2.3781
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.350308   0.443670  -5.297 1.17e-07 ***
## subSet$atomYES  1.622375   0.536151   3.026  0.00248 **
## experience1     0.533662   0.673987   0.792  0.42848
## experience2     0.008071   1.184068   0.007  0.99456
## experience3    -0.299122   0.652548  -0.458  0.64667
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 140.11  on 139  degrees of freedom
## Residual deviance: 126.22  on 135  degrees of freedom
## AIC: 136.22
##
## Number of Fisher Scoring iterations: 5
##
##
##                   1         2         3         4
```

```
##   CORRECT 0.8888889 0.7307692 0.8461538 0.8148148
##   WRONG   0.1111111 0.2692308 0.1538462 0.1851852
## [1] "Pre Increment"
##
##           NO YES
##   CORRECT 53  50
##   WRONG   17  20
##   atom        x
## 1   NO 24669.70
## 2  YES 30821.46
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tableCorrectness
## X-squared = 0.14694, df = 1, p-value = 0.7015
##
##
## Call:
## glm(formula = subSet$correct ~ subSet$atom + experience, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0899  -0.8496  -0.8175   1.2675   2.2872
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.13714    0.34052  -3.339  0.00084 ***
## subSet$atomYES  0.04157    0.39777   0.105  0.91676
## experience1     1.19001    0.54610   2.179  0.02932 *
## experience2     0.58252    0.73676   0.791  0.42915
## experience3     1.61521    0.79936   2.021  0.04332 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 161.70  on 139  degrees of freedom
## Residual deviance: 153.29  on 135  degrees of freedom
## AIC: 163.29
##
## Number of Fisher Scoring iterations: 5
##
##
##                    1          2          3          4
##   CORRECT 0.55555556 0.69230769 0.71153846 0.92592593
##   WRONG   0.44444444 0.30769231 0.28846154 0.07407407
```

```
analysis_df <- data.frame(atomDescription, chiTest, oddsRatio,
                          oddsRatioTest, ci25, ci975, mannWhitneyTest, cliffDelta)

analysis_df <- analysis_df[order(-oddsRatio),]

colnames(analysis_df) <- c("Atom", "ChiTest",
                           "Odds Ratio Correctness",
```

```
                           "p-value", "CI 2.5%", "CI 97.5%",
                           "Wilcox Test (Time)",
                           "Cliff Delta")

xtable(analysis_df)
```

```
## % latex table generated in R 4.2.0 by xtable 1.8-4 package
## % Fri Jul 29 09:07:13 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{rllrlrrrlr}
##   \hline
##  & Atom & ChiTest & Odds Ratio Correctness & p-value & CI 2.5\% & CI 97.5\% & Wilcox Test (Time) & C
##   \hline
## 3 & Automatic Semicolon Insertion & 5.8352e-11 & 39.33 & 2.4351e-12 & 9.21 & 356.62 & 0.09802 & -0.1
##   4 & Comma Operator & 1.1727e-10 & 19.02 & 1.5812e-11 & 6.60 & 68.22 & 5.8249e-08 & -0.53 \\
##   2 & Assignment as Value & 0.0034775 & 8.39 & 0.0024331 & 1.81 & 79.12 & 0.82681 & 0.02 \\
##   6 & Implict Predicate & 0.01123 & 6.95 & 0.008974 & 1.46 & 66.56 & 0.0029039 & -0.29 \\
##   9 & Post Increment & 0.0015279 & 4.83 & 0.0012328 & 1.73 & 15.74 & 0.12006 & 0.15 \\
##   5 & Ternary Operator & 0.15896 & 2.73 & 0.15712 & 0.74 & 12.57 & 0.24073 & -0.12 \\
##   8 & Ommitted Curly Braces & 0.050962 & 2.35 & 0.050056 & 1.00 & 5.77 & 0.92529 & -0.01 \\
##   7 & Logic as Control Flow & 0.292 & 1.54 & 0.29198 & 0.73 & 3.28 & 7.5862e-05 & -0.39 \\
##   10 & Pre Increment & 0.70147 & 1.25 & 0.70181 & 0.55 & 2.85 & 0.0062981 & 0.27 \\
##   1 & Arithmetic as Logic & 1 & 0.84 & 1 & 0.22 & 3.12 & 0.01723 & 0.23 \\
##    \hline
## \end{tabular}
## \end{table}
```