Stat 123 Final Project

Jonathan Wilson

January 28, 2019

```
#Read in Domains
read.domain <- function(letter){
  path <- sprintf("Data.and.Domains/Domains Form%s.csv", letter)
  domain <- data.frame()
  assign(paste(letter, domain, sep = ""), read.table(path,
                                                     sep=",",
                                                     header=TRUE,
                                                     stringsAsFactors=FALSE))
}

domainA <- read.domain("A")
domainB <- read.domain("B")

#Read in Forms
read.form <- function(letter){
  path <- sprintf("Data.and.Domains/Form%s.csv", letter)
  form <- read.table(path, sep=",", header=FALSE, stringsAsFactors=FALSE)

  key <- head(form, 1)
  key[1] <- NULL

  student.questions <- tail(form, -1)
  student.id <- student.questions[,1]
  student.id <- as.integer(student.id)
  student.id <- data.frame(student.id)
  student.questions[,1] <- NULL
  result <- list("id" = student.id, "key" = key, "questions" = student.questions)
  return(result)

}

#Create key, student ids and answers for A and B
form.info <- read.form("A")
keyA <- form.info$key
student.answersA <- form.info$questions
student.IDs.A <- form.info$id
form.info <- read.form("B")
keyB <- form.info$key
student.answersB <- form.info$questions
student.IDs.B <- form.info$id

#Get overall Scores:

get.scores <- function(df, letter){
    datalist <- list()
    for(i in 1:nrow(df)){
    #scoresA <- student.answersA[i,] == keyA
```

```
    if(letter == "A"){
      dat <- as.integer(df[i,] == keyA)
    }
    else{
      dat <- as.integer(df[i,] == keyB)
    }
    datalist[[i]] <- dat
    }
  scores <- do.call(rbind, datalist)
  return(scores)
}

scoresA <- get.scores(student.answersA, "A")
scoresA <- data.frame(scoresA)

scoresB <- get.scores(student.answersB, "B")
scoresB <- data.frame(scoresB)

#Generate tables to query from merging domains with student scores for A
table.a <- data.frame()
table.a <- cbind(student.IDs.A, "A", scoresA, stringsAsFactors=FALSE)
names(table.a)[1] <- "Student"
names(table.a)[2] <- "Form"
names(table.a)[3:152] <- 1:150
table.a <- melt(table.a, id=c("Student", "Form"))
names(table.a)[3] <- "question"
names(table.a)[4] <- "score"
names(domainA)[3] <- "question"
table.a <- merge(x = table.a, y = domainA[ ,c("question", "Domain")], by="question", all.x=TRUE)
table.a$question <- as.integer(as.character(table.a$question))#Get rid of the factor

#Generate tables to query from merging domains with student scores for B
table.b <- data.frame()
table.b <- cbind(student.IDs.B, "B", scoresB, stringsAsFactors=FALSE)
names(table.b)[1] <- "Student"
names(table.b)[2] <- "Form"
names(table.b)[3:152] <- 1:150
table.b <- melt(table.b, id=c("Student", "Form"))
names(table.b)[3] <- "question"
names(table.b)[4] <- "score"
names(domainB)[3] <- "question"
table.b <- merge(x = table.b, y = domainB[ ,c("question", "Domain")], by="question", all.x=TRUE)
table.b$question <- as.integer(as.character(table.b$question))#Get rid of the factor

full.table <- rbind(table.b, table.a)

#Section A: Student Scores
#student table
student.table <- full.table %>% group_by(Student, Form) %>%
  summarize(Overall.Score = sum(score), Overall.Percentage = (sum(score)/150)*100) %>%
  select(Student, Form, Overall.Score, Overall.Percentage)

#domain score table
```

```r
d1 <- 30
d2 <- 35
d3 <- 30
d4 <- 30
d5 <- 25
domain.table <- full.table %>% group_by(Student, Domain) %>%
  summarize(Domain.Score = sum(score)) #, Domain.Percentage = (sum(score)/150)*100
domain.table <- domain.table %>% select(Student, Domain, Domain.Score)
domain.table <- spread(domain.table, key = Domain, value = Domain.Score)
names(domain.table)[2] <- "Domain.1.Score"
names(domain.table)[3] <- "Domain.2.Score"
names(domain.table)[4] <- "Domain.3.Score"
names(domain.table)[5] <- "Domain.4.Score"
names(domain.table)[6] <- "Domain.5.Score"

#Domain percentage table
domain.percentage <- data.frame()
percentage.row <- data.frame(d1, d2, d3, d4, d5)
names(percentage.row) <- c("Domain.1.Percentage",
                           "Domain.2.Percentage",
                           "Domain.3.Percentage",
                           "Domain.4.Percentage",
                           "Domain.5.Percentage")
for(i in 1:99){
  for(j in 2:6){
    if(j==2){
      percentage.row[1,1] <- (domain.table[i,j]/d1)*100
    }
    else if(j==3){
      percentage.row[1,2] <- (domain.table[i,j]/d2)*100
    }
    else if(j==4){
      percentage.row[1,3] <- (domain.table[i,j]/d3)*100
    }
    else if(j==5){
      percentage.row[1,4] <- (domain.table[i,j]/d4)*100
    }
    else if(j==6){
      percentage.row[1,5] <- (domain.table[i,j]/d5)*100
    }

  }
  domain.percentage <- rbind(domain.percentage, percentage.row)
}
domain.percentage <- cbind.data.frame(domain.table$Student,domain.percentage)
names(domain.percentage)[1] <- "Student"

#Table sorted by Student ID:
Section.A.Table1 <- merge(student.table, domain.table, by="Student")
Section.A.Table1 <- merge(Section.A.Table1, domain.percentage, by="Student")
#Table sorted by highest score:
Section.A.Table2 <- Section.A.Table1 %>% select(Student,
                                                Form,
```

```
                                            Overall.Percentage,
                                            Overall.Score,
                                            Domain.1.Percentage,
                                            Domain.2.Percentage,
                                            Domain.3.Percentage,
                                            Domain.4.Percentage,
                                            Domain.5.Percentage,
                                            Domain.1.Score,
                                            Domain.2.Score,
                                            Domain.3.Score,
                                            Domain.4.Score,
                                            Domain.5.Score)
Section.A.Table2 <- Section.A.Table2[order(Section.A.Table2$Overall.Percentage, decreasing = TRUE), ]

#Boxplot:
plot.data <- Section.A.Table1 %>% select(Domain.1.Percentage,
                                         Domain.2.Percentage,
                                         Domain.3.Percentage,
                                         Domain.4.Percentage,
                                         Domain.5.Percentage)

#Section B: Questions
#Order by question number
tmpA <- full.table %>% group_by(question, Form) %>% filter(Form == "A") %>%
  summarize(Question.Percentage = (sum(score)/50)*100) %>%
  select(Form, question, Question.Percentage)
tmpB <- full.table %>% group_by(question, Form) %>% filter(Form == "B") %>%
  summarize(Question.Percentage = (sum(score)/49)*100) %>%
  select(Form, question, Question.Percentage)
question.table1 <- rbind(tmpA, tmpB)

#order by question percentage
question.table2 <- question.table1 %>% select(Question.Percentage, Form, question)
question.table2 <- question.table2[order(question.table2$Question.Percentage, decreasing = TRUE), ]

#Formate the %

Section.A.Table1$Overall.Percentage <- sprintf("%s%%",
                                               format(round(student.table$Overall.Percentage, 1),
                                                      nsmall = 1))
Section.A.Table1$Domain.1.Percentage <- sprintf("%s%%",
                                                format(round(domain.percentage$Domain.1.Percentage, 1),
                                                       nsmall = 1))
Section.A.Table1$Domain.2.Percentage <- sprintf("%s%%",
                                                format(round(domain.percentage$Domain.2.Percentage, 1),
                                                       nsmall = 1))
Section.A.Table1$Domain.3.Percentage <- sprintf("%s%%",
                                                format(round(domain.percentage$Domain.3.Percentage, 1),
                                                       nsmall = 1))
Section.A.Table1$Domain.4.Percentage <- sprintf("%s%%",
                                                format(round(domain.percentage$Domain.4.Percentage, 1),
                                                       nsmall = 1))
Section.A.Table1$Domain.5.Percentage <- sprintf("%s%%",
```

```
                                                format(round(domain.percentage$Domain.5.Percentage, 1),
                                                       nsmall = 1))

Section.A.Table2$Overall.Percentage <- sprintf("%s%%",
                                               format(round(Section.A.Table2$Overall.Percentage, 1),
                                                      nsmall = 1))
Section.A.Table2$Domain.1.Percentage <- sprintf("%s%%",
                                                format(round(Section.A.Table2$Domain.1.Percentage, 1),
                                                       nsmall = 1))
Section.A.Table2$Domain.2.Percentage <- sprintf("%s%%",
                                                format(round(Section.A.Table2$Domain.2.Percentage, 1),
                                                       nsmall = 1))
Section.A.Table2$Domain.3.Percentage <- sprintf("%s%%",
                                                format(round(Section.A.Table2$Domain.3.Percentage, 1),
                                                       nsmall = 1))
Section.A.Table2$Domain.4.Percentage <- sprintf("%s%%",
                                                format(round(Section.A.Table2$Domain.4.Percentage, 1),
                                                       nsmall = 1))
Section.A.Table2$Domain.5.Percentage <- sprintf("%s%%",
                                                format(round(Section.A.Table2$Domain.5.Percentage, 1),
                                                       nsmall = 1))

question.table1$Question.Percentage <- sprintf("%s%%",
                                               format(round(question.table1$Question.Percentage, 1),
                                                      nsmall = 1))
question.table2$Question.Percentage <- sprintf("%s%%",
                                               format(round(question.table2$Question.Percentage, 1),
                                                      nsmall = 1))
```

Section B: Question Analysis

Questions Sorted by Question

```
kable(question.table1, "latex", booktabs = TRUE, longtable = TRUE, caption = "Questions", col.names = c
  kable_styling(latex_options = c("hold_position", "repeat_header"))
```

Questions Sorted by Question Percentage

```
kable(question.table2, "latex",
      booktabs = TRUE,
      longtable = TRUE,
      caption = "Questions sorted from high to low percentage",
      col.names = c("Percentage",  "Form", "Question")) %>%
  kable_styling(latex_options = c("hold_position", "repeat_header"))
```

Section A: Student Scores

Note: You may have to zoom in on the tables

```
kable(Section.A.Table1, "latex", booktabs = T, caption = "Student scores") %>%
  kable_styling(latex_options = c("scale_down"))
```

```
kable(Section.A.Table2, "latex", booktabs = T, caption = "Student scores sorted by highest percentage") %>%
  kable_styling(latex_options = c("scale_down"))
```

```
#Commented out to suppress output
#boxplot(plot.data, main='Distribution of Student Percents by Domain Number',
 #       ylab ='Percentage (%)', xlab ='Domain', names = c("1", "2", "3", "4", "5"))
```