Homework 1

Hanao Li

February 13, 2019

Question 1

```
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 3.4.4
```

```
setwd("c:/Users/36576/Desktop/gr5243")
pk <- fread('Prior Courses.csv')
grade <- fread('Grades.csv')
l2 <- fread('Knowledge Check -- Level 2.csv')
l5 <- fread('Knowledge check -- Level 5.csv')
```

Question 1a)

```
dim(grade)
```

```
## [1] 100  10
```

```
dim(l2)
```

```
## [1] 100   7
```

```
dim(l5)
```

```
## [1] 100  13
```

```
dim(pk)
```

```
## [1] 103   3
```

Question 1b)

```
pk[,.(length(unique(`Student ID`)))]
```

```
##     V1
## 1: 100
```

```
l2[,.(length(unique(`Student ID`)))]
```

```
##     V1
## 1: 100
```

```
l5[,.(length(unique(`Student ID`)))]
```

```
##     V1
## 1: 100
```

```
grade[,.(length(unique(`Student ID`)))]
```

```
##     V1
## 1: 100
```

```
# So the number of unique students is 100
```

Question 2

```
# As we can see from the previous question, the prior courses file contain more than 1 row per student.
library(DT)
```

```
## Warning: package 'DT' was built under R version 3.4.4
```

```
display.multiple.records <- function(x){
  id <- names((table(x$`Student ID`) > 1)[(table(x$`Student ID`) > 1) == TRUE])
  records <- x[`Student ID` %in% id]
  setorderv(records, "Student ID")
  return(datatable(records))
}
display.multiple.records(grade)
```

Show 10 entries Search:

Student ID	HW 1: Mechanics	HW 2: Momentum	HW 3: Gravity	HW 4: Electricity	HW 5: Magnetism	HW 6: Relativity	Midterm	Final Exam	Total Score
No data available in table									

Showing 0 to 0 of 0 entries Previous Next

```
display.multiple.records(l2)
```

Show 10 entries Search:

Student ID	Mechanics	Momentum	Gravity	Time: Mechanics	Time: Momentum	Time: Gravity
No data available in table						

Showing 0 to 0 of 0 entries Previous Next

```
display.multiple.records(l5)
```

Show 10 entries Search:

Student ID	Mechanics	Momentum	Gravity	Electricity	Magnetism	Relativity	Time: Mechanics	Time: Momentum	Tim Grav
No data available in table									

Showing 0 to 0 of 0 entries Previous Next

```
display.multiple.records(pk)
```

Show 10 entries Search:

	Student ID	Trigonometry	Calculus
1	Student 57	85	88
2	Student 57	88	86
3	Student 66	85	83
4	Student 66	86	86
5	Student 76	84	88
6	Student 76	83	87

Showing 1 to 6 of 6 entries Previous 1 Next

Question 3

Question 3a)

```
id <- names((table(pk$`Student ID`) > 1)[(table(pk$`Student ID`) > 1) == TRUE])
summarize.pk.class <- function(x){
  x$Trigonometry <- as.numeric(x$Trigonometry)
  variables <- c("Trigonometry", "Calculus")
  measured.records <- c(length(which(!is.na(x$Trigonometry))), length(which(!is.na(x$Calculus))))
  t <- 0
  for (i in 1:length(id)){
      if(sum(!is.na(x$Trigonometry[x$`Student ID` %in% id[i]])) != 0){
      t <- t + 1
    }
  }
  c <- 0
  for (i in 1:length(id)){
    if(sum(!is.na(x$Calculus[x$`Student ID` %in% id[i]])) != 0){
      c <- c + 1
    }
  }
  unique.records <- measured.records - c(t,c)
  mean.records <- c(round(mean(x$Trigonometry, na.rm = TRUE), 2), round(mean(x$Calculus, na.rm = TRUE), 2))
  std.records <- c(round(sd(x$Trigonometry, na.rm = TRUE), 2), round(sd(x$Calculus, na.rm = TRUE), 2))
  new <- data.table(variables, measured.records, unique.records, mean.records, std.records)
  datatable(new)
}
summarize.pk.class(pk)
```

```
## Warning in summarize.pk.class(pk): NAs introduced by coercion
```

Show 10 entries Search:

	variables	measured.records	unique.records	mean.records	std.records
1	Trigonometry	101	98	84.04	4.05
2	Calculus	103	100	84.36	4.07

Showing 1 to 2 of 2 entries Previous 1 Next

Question 3b)

```
for (i in 1:length(id)){
  pk[pk$`Student ID` == id[i]]$Trigonometry <- max(pk$Trigonometry[pk$`Student ID` %in% id[i]])
  pk[pk$`Student ID` == id[i]]$Calculus <- max(pk$Calculus[pk$`Student ID` %in% id[i]])
}
pk$Trigonometry <- as.numeric(pk$Trigonometry)
```

```
## Warning: NAs introduced by coercion
```

```
pk$PriorKnowledgeLevel <- 0.5 * (pk$Trigonometry + pk$Calculus)
n <- which(is.na(pk$Trigonometry))
for (i in 1:length(n)){
  pk[n[i]]$PriorKnowledgeLevel <- pk[n[i]]$Calculus
}
n <- which(is.na(pk$Calculus))
for (i in 1:length(n)){
  pk[n[i]]$PriorKnowledgeLevel <- pk[n[i]]$Trigonometry
}
pk.reduced <- pk[1:100, ]
datatable(pk.reduced)
```

Show 10 entries Search:

	Student ID	Trigonometry	Calculus	PriorKnowledgeLevel
1	Student 1	87	90	88.5
2	Student 2	89	85	87
3	Student 3	87	84	85.5
4	Student 4	86	86	86
5	Student 5	82	84	83
6	Student 6	79	81	80

	Student ID	Trigonometry	Calculus	PriorKnowledgeLevel
7	Student 7	86	85	85.5
8	Student 8	84	86	85
9	Student 9	88	87	87.5
10	Student 10	79	84	81.5

Showing 1 to 10 of 100 entries Previous 1 2 3 4 5 … 10 Next

```
datatable(pk.reduced[`Student ID` %in% id])
```

Show 10 entries Search:

	Student ID	Trigonometry	Calculus	PriorKnowledgeLevel
1	Student 57	88	88	88
2	Student 66	86	86	86
3	Student 76	84	88	86

Showing 1 to 3 of 3 entries Previous 1 Next

Question 3c)

```
summarize.pk.reduced <- function(x){
  variables <- c("Trigonometry", "Calculus", "PriorKnowledgeLevel")
  measured.records <- c(length(which(!is.na(x$Trigonometry))), length(which(!is.na(x$Calculus))), length(x$PriorKnowledgeLev
el))
  unique.records <- measured.records
  mean.records <- c(round(mean(x$Trigonometry, na.rm = TRUE), 2), round(mean(x$Calculus, na.rm = TRUE), 2), round(mean(x$Pri
orKnowledgeLevel), 2))
  std.records <- c(round(sd(x$Trigonometry, na.rm = TRUE), 2), round(sd(x$Calculus, na.rm = TRUE), 2), round(sd(x$PriorKnowl
edgeLevel), 2))
  new <- data.table(variables, measured.records, unique.records, mean.records, std.records)
  datatable(new)
}
summarize.pk.reduced(pk.reduced)
```

Show 10 entries Search:

	variables	measured.records	unique.records	mean.records	std.records
1	Trigonometry	98	98	84.03	4.11
2	Calculus	100	100	84.33	4.11
3	PriorKnowledgeLevel	100	100	84.19	3.89

Showing 1 to 3 of 3 entries Previous 1 Next

Question 4

```
m <- merge(pk.reduced, l2, by = "Student ID")
m <- merge(m, l5, by = 'Student ID')
m <- merge(m, grade, by = 'Student ID')
m$ordered <- as.numeric(unlist(regmatches(m$`Student ID`, gregexpr("[0-9]+", m$`Student ID`))))
setorderv(m, cols = 'ordered')
m$ordered <- NULL
cols <- names(m)[2:31]
m[,(cols) := round(.SD,2), .SDcols=cols]
datatable(m)
```

Show 10 entries Search:

	Student ID	Trigonometry	Calculus	PriorKnowledgeLevel	Mechanics.x	Momentum.x	Gravity.x	Time: Mechanics.x	M
1	Student 1	87	90	88.5	1.44	1.07	1.1	48.37	
2	Student 2	89	85	87	1.44	1.27	1.1	42.3	

	Student ID	Trigonometry	Calculus	PriorKnowledgeLevel	Mechanics.x	Momentum.x	Gravity.x	Time: Mechanics.x	M
3	Student 3	87	84	85.5	0.74	1.47	1.1	41.71	
4	Student 4	86	86	86	1.84	1.77	1.3	43.35	
5	Student 5	82	84	83	2.44	1.87	1.8	38.88	
6	Student 6	79	81	80	1.94	2.07	1.8	21.45	
7	Student 7	86	85	85.5	2.04	1.97	1.6	49.18	
8	Student 8	84	86	85	2.04	2.17	1.6	35.41	
9	Student 9	88	87	87.5	2.34	2.57		14.38	
10	Student 10	79	84	81.5	2.34	1.77	1.8	34.39	

Showing 1 to 10 of 100 entries Previous 1 2 3 4 5 … 10 Next

Question 5

```
topic <- colnames(l2)[2:4]
nstudents <- c(sum(!is.na(l2$Mechanics)), sum(!is.na(l2$Momentum)), sum(!is.na(l2$Gravity)))
meanscore <- c(round(mean(l2$Mechanics, na.rm = TRUE), 2), round(mean(l2$Momentum, na.rm = TRUE), 2), round(mean(l2$Gravity,
na.rm = TRUE), 2))
stdscore <- c(round(sd(l2$Mechanics, na.rm = TRUE), 2), round(sd(l2$Momentum, na.rm = TRUE), 2), round(sd(l2$Gravity, na.rm
 = TRUE), 2))
percentage <- round(c(sum(l2$Mechanics >= 2, na.rm = TRUE), sum(l2$Momentum >= 2, na.rm = TRUE), sum(l2$Gravity >= 2, na.rm
 = TRUE)) / nstudents, 2)
meanmin <- c(round(mean(l2$`Time: Mechanics`, na.rm = TRUE), 2), round(mean(l2$`Time: Momentum`, na.rm = TRUE), 2), round(me
an(l2$`Time: Gravity`, na.rm = TRUE), 2))
stdmean <- c(round(sd(l2$`Time: Mechanics`, na.rm = TRUE), 2), round(sd(l2$`Time: Momentum`, na.rm = TRUE), 2), round(sd(l2
$`Time: Gravity`, na.rm = TRUE), 2))
table1 <- data.table(topic, nstudents, meanscore, stdscore, percentage, meanmin, stdmean)
datatable(table1)
```

Show 10 entries Search:

	topic	nstudents	meanscore	stdscore	percentage	meanmin	stdmean
1	Mechanics	98	2.04	0.47	0.63	37.41	9.25
2	Momentum	97	1.88	0.46	0.43	38.5	8.11
3	Gravity	94	1.6	0.44	0.17	37.67	8.43

Showing 1 to 3 of 3 entries Previous 1 Next

Question 6

```
topic <- colnames(l5)[2:7]
nstudents <- c(sum(!is.na(l5$Mechanics)), sum(!is.na(l5$Momentum)), sum(!is.na(l5$Gravity)), sum(!is.na(l5$Electricity)), su
m(!is.na(l5$Magnetism)), sum(!is.na(l5$Relativity)))
meanscore <- c(round(mean(l5$Mechanics, na.rm = TRUE), 2), round(mean(l5$Momentum, na.rm = TRUE), 2), round(mean(l5$Gravity,
na.rm = TRUE), 2), round(mean(l5$Electricity, na.rm = TRUE), 2), round(mean(l5$Magnetism, na.rm = TRUE), 2), round(mean(l5$R
elativity, na.rm = TRUE), 2))
stdscore <- c(round(sd(l5$Mechanics, na.rm = TRUE), 2), round(sd(l5$Momentum, na.rm = TRUE), 2), round(sd(l5$Gravity, na.rm
 = TRUE), 2), round(sd(l5$Electricity, na.rm = TRUE), 2), round(sd(l5$Magnetism, na.rm = TRUE), 2), round(sd(l5$Relativity,
 na.rm = TRUE), 2))
percentage <- round(c(sum(l5$Mechanics >= 5, na.rm = TRUE), sum(l5$Momentum >= 5, na.rm = TRUE), sum(l5$Gravity >= 5, na.rm
 = TRUE), sum(l5$Electricity >= 5, na.rm = TRUE), sum(l5$Magnetism >= 5, na.rm = TRUE), sum(l5$Relativity >= 5, na.rm = TRUE
)) / nstudents, 2)
meanmin <- c(round(mean(l5$`Time: Mechanics`, na.rm = TRUE), 2), round(mean(l5$`Time: Momentum`, na.rm = TRUE), 2), round(me
an(l5$`Time: Gravity`, na.rm = TRUE), 2), round(mean(l5$`Time: Electricity`, na.rm = TRUE), 2), round(mean(l5$`Time: Magneti
sm`, na.rm = TRUE), 2), round(mean(l5$`Time: Relativity`, na.rm = TRUE), 2))
stdmean <- c(round(sd(l5$`Time: Mechanics`, na.rm = TRUE), 2), round(sd(l5$`Time: Momentum`, na.rm = TRUE), 2), round(sd(l5
$`Time: Gravity`, na.rm = TRUE), 2), round(sd(l5$`Time: Electricity`, na.rm = TRUE), 2), round(sd(l5$`Time: Magnetism`, na.r
m = TRUE), 2), round(sd(l5$`Time: Relativity`, na.rm = TRUE), 2))
table2 <- data.table(topic, nstudents, meanscore, stdscore, percentage, meanmin, stdmean)
datatable(table2)
```

Show 10 entries Search:

	topic	nstudents	meanscore	stdscore	percentage	meanmin	stdmean

	topic	nstudents	meanscore	stdscore	percentage	meanmin	stdmean
1	Mechanics	95	5.04	0.71	0.69	22.13	10.12
2	Momentum	95	4.89	0.69	0.57	33.13	9.45
3	Gravity	92	4.8	0.82	0.47	33.08	9.43
4	Electricity	98	4.91	0.74	0.67	26.46	10.87
5	Magnetism	96	4.96	0.76	0.69	36.14	9.65
6	Relativity	99	4.87	0.7	0.53	30.42	10.18

Showing 1 to 6 of 6 entries Previous 1 Next

Question 7

```
topics <- c("Mechanics 1", "Mechanics 2", "Momentum 1", "Momentum 2", "Gravity 1", "Gravity 2", "Electricity", "Magnetism",
"Relativity")
correlations <- round(c(cor(m$Mechanics.x, m$`Time: Mechanics.x`, use = 'complete.obs'), cor(m$Mechanics.y, m$`Time: Mechani
cs.y`, use = 'complete.obs'), cor(m$Momentum.x, m$`Time: Momentum.x`, use = 'complete.obs'), cor(m$Momentum.y, m$`Time: Mome
ntum.y`, use = 'complete.obs'), cor(m$Gravity.x, m$`Time: Gravity.x`, use = 'complete.obs'), cor(m$Gravity.y, m$`Time: Gravi
ty.y`, use = 'complete.obs'), cor(m$Electricity, m$`Time: Electricity`, use = 'complete.obs'), cor(m$Magnetism, m$`Time: Mag
netism`, use = 'complete.obs'), cor(m$Relativity, m$`Time: Relativity`, use = 'complete.obs')), 2)
cortable <- data.table(topics, correlations)
datatable(cortable)
```

Show 10 **entries** **Search:**

	topics	correlations
1	Mechanics 1	-0.13
2	Mechanics 2	-0.27
3	Momentum 1	-0.01
4	Momentum 2	-0.16
5	Gravity 1	-0.05
6	Gravity 2	-0.13
7	Electricity	-0.15
8	Magnetism	-0.16
9	Relativity	-0.17

Showing 1 to 9 of 9 entries Previous 1 Next

```
# I believe spendind more time on the digital system should be benefical. But from the results, we could see that spending m
ore time on the digital system is not benefical. The reason why it is not benefical might because students that know the mat
erials could finish the assignments faster while students who don't understand the materials will finish those assignements
 slower than expected.
```

Question 8

```
m$`Homework Average` <- round(rowMeans(subset(m, select = c(23:28))), 2)
variablename <- c(colnames(m)[23:32])
measuredvalue <- c(sum(!is.na(m$`HW 1: Mechanics`)), sum(!is.na(m$`HW 2: Momentum`)), sum(!is.na(m$`HW 3: Gravity`)), sum(!i
s.na(m$`HW 4: Electricity`)), sum(!is.na(m$`HW 5: Magnetism`)), sum(!is.na(m$`HW 6: Relativity`)), sum(!is.na(m$Midterm)), s
um(!is.na(m$`Final Exam`)), sum(!is.na(m$`Total Score`)), sum(!is.na(m$`Homework Average`)))
meanscores <- round(c(mean(m$`HW 1: Mechanics`, na.rm = TRUE), mean(m$`HW 2: Momentum`, na.rm = TRUE), mean(m$`HW 3: Gravity
`, na.rm = TRUE), mean(m$`HW 4: Electricity`, na.rm = TRUE), mean(m$`HW 5: Magnetism`, na.rm = TRUE), mean(m$`Homework Avera
ge`, na.rm = TRUE), mean(m$Midterm, na.rm = TRUE), mean(m$`Final Exam`, na.rm = TRUE), mean(m$`Total Score`, na.rm = TRUE),
 mean(m$`Homework Average`, na.rm = TRUE)), 2)
stdscores <- round(c(sd(m$`HW 1: Mechanics`, na.rm = TRUE), sd(m$`HW 2: Momentum`, na.rm = TRUE), sd(m$`HW 3: Gravity`, na.r
m = TRUE), sd(m$`HW 4: Electricity`, na.rm = TRUE), sd(m$`HW 5: Magnetism`, na.rm = TRUE), sd(m$`Homework Average`, na.rm =
TRUE), sd(m$Midterm, na.rm = TRUE), sd(m$`Final Exam`, na.rm = TRUE), sd(m$`Total Score`, na.rm = TRUE), sd(m$`Homework Aver
age`, na.rm = TRUE)), 2)
table3 <- data.table(variablename, measuredvalue, meanscores, stdscores)
datatable(table3)
```

Show 10 **entries**

Search:

	variablename	measuredvalue	meanscores	stdscores
1	HW 1: Mechanics	100	88.25	7.43
2	HW 2: Momentum	100	92.76	6.05
3	HW 3: Gravity	100	82.41	6.58
4	HW 4: Electricity	100	87.32	5.83
5	HW 5: Magnetism	100	88	5.89
6	HW 6: Relativity	100	87.74	3.48
7	Midterm	100	84.82	7.51
8	Final Exam	100	84.85	8.04
9	Total Score	100	86	3.94
10	Homework Average	100	87.74	3.48

Showing 1 to 10 of 10 entries Previous 1 Next

Question 9

```
outcomes <- c("Homework Average", "Midterm Score", "Final Exam Score", "Total Score")
Prior_Knowledge <- round(c(cor(m$`Homework Average`, m$PriorKnowledgeLevel), cor(m$Midterm, m$PriorKnowledgeLevel), cor(m$`F
inal Exam`, m$PriorKnowledgeLevel), cor(m$`Total Score`, m$PriorKnowledgeLevel)), 2)
l5[is.na(l5)] <- 0
m$`Total Time` <- round(l5$`Time: Mechanics` + l5$`Time: Momentum` + l5$`Time: Gravity` + l5$`Time: Electricity` + l5$`Time:
Magnetism` + l5$`Time: Relativity`, 2)
Total_Minutes <- round(c(cor(m$`Homework Average`, m$`Total Time`), cor(m$Midterm, m$`Total Time`), cor(m$`Final Exam`, m$`T
otal Time`), cor(m$`Total Score`, m$`Total Time`)), 2)
l5$sumzero <- (l5$Mechanics == 0) + (l5$Momentum == 0) + (l5$Gravity == 0) + (l5$Electricity == 0) + (l5$Magnetism == 0) +
 (l5$Relativity == 0)
m$avgscore <- round((l5$Mechanics + l5$Momentum + l5$Gravity + l5$Electricity + l5$Magnetism + l5$Relativity) / (6 - l5$sumz
ero), 2)
Average_Score <- round(c(cor(m$`Homework Average`, m$avgscore), cor(m$Midterm, m$avgscore), cor(m$`Final Exam`, m$avgscore),
cor(m$`Total Score`, m$avgscore)), 2)
table4 <- data.table(outcomes, Prior_Knowledge, Total_Minutes, Average_Score)
datatable(table4)
```

Show 10 entries Search:

	outcomes	Prior_Knowledge	Total_Minutes	Average_Score
1	Homework Average	-0.06	-0.08	0.5
2	Midterm Score	0.23	0.05	-0.09
3	Final Exam Score	0.33	0.03	0.16
4	Total Score	0.31	0.02	0.22

Showing 1 to 4 of 4 entries Previous 1 Next

Question 10

From the results we get from question 9, we could see that the digital system student used isn't very helpful. It has a little impact on the Homework, Midterm, Final and Total Score. The average score has a realtive high impact to homework average and also some impact towards the final exam and total score. The prior knonledge has a positive effect on the midterm, final and total score. We can recommend students to preview the course materials to have some prior understanding of the concepts and they can improve their score by doing this and do well in the knowledge check test.