STAT 472 HW 2

Wulf, Han, Chunwei, Sean

March 14, 2019 ~ Pie Day

S1

Switcher and Persister Groups (Table S1)

Switcher Group	Number of Students	Beginning of Term	End of Term; Reflect	End of Term	Follow Up
1	160	Yes			No
2	118	Maybe			No
3	15	NA	Yes		No
4	3	NA	Maybe		No
5	38	Yes	Yes	Maybe	NA
6	123	Yes		No	NA
7	17	Maybe	Yes	Maybe	NA
8	152	Maybe		No	NA
9	34	NA	Yes	Maybe	NA
10	78	NA	Yes	No	NA
11	65	NA	Maybe	No	NA

Persister Group	Number of Students	Beginning of Term	End of Term; Reflect	End of Term	Follow Up
12	586	Yes			Yes
13	63	Maybe			Yes
14	67	NA	Yes		Yes
15	2	NA	Maybe		Yes
16	1543	Yes		Yes	NA
17	35	Yes	Maybe	Maybe	NA
18	5	Yes	No	Maybe	NA
19	1	Yes	NA	Maybe	NA
20	193	Maybe		Yes	NA
21	64	Maybe	Maybe	Maybe	NA
22	22	Maybe	No	Maybe	NA
23	3	Maybe	NA	Maybe	NA
24	1325	NA	Yes	Yes	NA
25	53	NA	Maybe	Yes	NA
26	103	NA	Maybe	Maybe	NA

S5 Table

Percentage of students that switched out of calculus by career choice and gender.

Career Choice	Gender	N	Switcher %
STEM	Male	263	10.6
	Female	223	16.1
Engineering	Male	539	3.5
	Female	249	6.4
Pre-Med	Male	199	21.6
	Female	318	33.3
Non-STEM	Male	136	36.8
	Female	126	38.1
Undecided	Male	99	26.3
	Female	114	28.1

S6 Table

Percentage of students that switched out of calculus by previous calculus experience and gender.

Previous Calculus	Gender	N	Switcher %
High School	Male	740	13
	Female	699	21.7
College	Male	99	17.2
	Female	62	14.5
None	Male	397	13.4
	Female	269	28.6

S7 Table

Percentage of students that switched out of calculus by standardized mathematics test percentile and gender.

Test Percentile	Gender	N	Switcher %	Test Percentile	Gender	N	Switcher %
90 - 100	Male	707	12.4	**40 - 49**	Male	20	26.7
	Female	523	22.6		Female	30	10
80 - 89	Male	250	14.8	**30 - 39**	Male	10	33.3
	Female	220	22.3		Female	6	0
70 - 79	Male	165	12.7	**20 - 29**	Male	3	0
	Female	157	25.5		Female	2	50
60 - 69	Male	50	18.3	**10 - 19**	Male	2	50
	Female	60	23.1		Female	2	0
50 - 59	Male	26	27.6	**0 - 9**	Male	3	100
	Female	29	20		Female	1	12.4

S8 Table

Percentage of students that switched out of calculus by aggregate measures of instruction perception and gender.

Instructor Quality Rating	Gender	N	Switcher %	Student-Centered Instruction Response	Gender	N	Switcher %
$[5.5, 6]$	Male	165	9.1	$[5.5, 6]$	Male	17	17.6
	Female	152	16.4		Female	21	38.1
$[4.5, 5.5)$	Male	613	12.2	$[4.5, 5.5)$	Male	155	12.3
	Female	464	20.3		Female	124	25
$[3.5, 4.5)$	Male	323	13.6	$[3.5, 4.5)$	Male	363	12.1
	Female	281	25.3		Female	282	22
$[2.5, 3.5)$	Male	95	24.2	$[2.5, 3.5)$	Male	393	16.3
	Female	89	29.2		Female	296	23.6
$[1.5, 2.5)$	Male	33	24.2	$[1.5, 2.5)$	Male	252	11.1
	Female	37	48.6		Female	257	23
$[1, 1.5)$	Male	7	14.3	$[1, 1.5)$	Male	56	14.3
	Female	7	57.1		Female	50	16

The Code

```
setwd('C:/Data') # Change these as needed
dataSet <- read.csv('CalcData.csv')
```

```
## Creating the switch or persist column & adding it to the df
switchPersist <- rep(NA,length(dataSet[,1]))
dataSet <- cbind(dataSet,switchPersist)
```

```
## Based on S1 table, the conditions for groups 1-11
dataSet$switchPersist[dataSet$Q26 == 1 & dataSet$Q3FUS_No == "No"
  & dataSet$Q3FUS_Yes == " "] = 1
dataSet$switchPersist[dataSet$Q26 == 3 & dataSet$Q3FUS_No == "No"
  & dataSet$Q3FUS_Yes == " "] = 2
dataSet$switchPersist[is.na(dataSet$Q26) & dataSet$Q5Post == 1
  & dataSet$Q3FUS_No == "No" & dataSet$Q3FUS_Yes == " "] = 3
dataSet$switchPersist[is.na(dataSet$Q26) & dataSet$Q5Post == 3
  & dataSet$Q3FUS_No == "No" & dataSet$Q3FUS_Yes == " "] = 4
dataSet$switchPersist[dataSet$Q26 == 1 & dataSet$Q5Post == 1
  & dataSet$Q3Post == 3 & dataSet$Q3FUS_No == " " & dataSet$Q3FUS_Yes == " "] = 5
dataSet$switchPersist[dataSet$Q26 == 1 & dataSet$Q3Post == 2
  & dataSet$Q3FUS_No == " " & dataSet$Q3FUS_Yes == " "] = 6
dataSet$switchPersist[dataSet$Q26 == 3 & dataSet$Q5Post == 1
  & dataSet$Q3Post == 3 & dataSet$Q3FUS_No == " " & dataSet$Q3FUS_Yes == " "] = 7
dataSet$switchPersist[dataSet$Q26 == 3 & dataSet$Q3Post == 2
  & dataSet$Q3FUS_No == " " & dataSet$Q3FUS_Yes == " "] = 8
dataSet$switchPersist[is.na(dataSet$Q26) & dataSet$Q5Post == 1
  & dataSet$Q3Post == 3 & dataSet$Q3FUS_No == " " & dataSet$Q3FUS_Yes == " "] = 9
dataSet$switchPersist[is.na(dataSet$Q26) & dataSet$Q5Post == 1
  & dataSet$Q3Post == 2 & dataSet$Q3FUS_No == " " & dataSet$Q3FUS_Yes == " "] = 10
dataSet$switchPersist[is.na(dataSet$Q26) & dataSet$Q5Post == 3
  & dataSet$Q3Post == 2 & dataSet$Q3FUS_No == " " & dataSet$Q3FUS_Yes == " "] = 11
```

```
## Based on S1 table, the conditions for groups 12-21
dataSet$switchPersist[dataSet$Q26 == 1 & dataSet$Q3FUS_Yes == "Yes"
& dataSet$Q3FUS_No == " "] = 12
dataSet$switchPersist[dataSet$Q26 == 3 & dataSet$Q3FUS_Yes == "Yes"
  & dataSet$Q3FUS_No == " "] = 13
dataSet$switchPersist[is.na(dataSet$Q26) & dataSet$Q5Post == 1
  & dataSet$Q3FUS_Yes == "Yes" & dataSet$Q3FUS_No == " "] = 14
dataSet$switchPersist[is.na(dataSet$Q26) & dataSet$Q5Post == 3
  & dataSet$Q3FUS_Yes == "Yes" & dataSet$Q3FUS_No == " "] = 15
dataSet$switchPersist[dataSet$Q26 == 1  & dataSet$Q3Post == 1
  & dataSet$Q3FUS_Yes == " " & dataSet$Q3FUS_No == " "] = 16
dataSet$switchPersist[dataSet$Q26 == 1 & dataSet$Q5Post == 3
  & dataSet$Q3Post == 3 & dataSet$Q3FUS_Yes == " " & dataSet$Q3FUS_No == " "] = 17
dataSet$switchPersist[dataSet$Q26 == 1 & dataSet$Q5Post == 2
  & dataSet$Q3Post == 3 & dataSet$Q3FUS_Yes == " " & dataSet$Q3FUS_No == " "] = 18
dataSet$switchPersist[dataSet$Q26 == 1 & is.na(dataSet$Q5Post)
  & dataSet$Q3Post == 3 & dataSet$Q3FUS_Yes == " " & dataSet$Q3FUS_No == " "] = 19
dataSet$switchPersist[dataSet$Q26 == 3 & dataSet$Q3Post == 1
  & dataSet$Q3FUS_Yes == " " & dataSet$Q3FUS_No == " "] = 20
dataSet$switchPersist[dataSet$Q26 == 3 & dataSet$Q5Post == 3
  & dataSet$Q3Post == 3 & dataSet$Q3FUS_Yes == " " & dataSet$Q3FUS_No == " "] = 21
```

```
dataSet$switchPersist[dataSet$Q26 == 3 & dataSet$Q5Post == 2
  & dataSet$Q3Post == 3 & dataSet$Q3FUS_Yes == " " & dataSet$Q3FUS_No == " "] = 22

## groups 23-26
dataSet$switchPersist[dataSet$Q26 == 3 & is.na(dataSet$Q5Post)
  & dataSet$Q3Post == 3 & dataSet$Q3FUS_Yes == " " & dataSet$Q3FUS_No == " "] = 23
dataSet$switchPersist[is.na(dataSet$Q26) & dataSet$Q5Post == 1
  & dataSet$Q3Post == 1 & dataSet$Q3FUS_Yes == " " & dataSet$Q3FUS_No == " "] = 24
dataSet$switchPersist[is.na(dataSet$Q26) & dataSet$Q5Post == 3
  & dataSet$Q3Post == 1 & dataSet$Q3FUS_Yes == " " & dataSet$Q3FUS_No == " "] = 25
dataSet$switchPersist[is.na(dataSet$Q26) & dataSet$Q5Post == 3
  & dataSet$Q3Post == 3 & dataSet$Q3FUS_Yes == " " & dataSet$Q3FUS_No == " "] = 26
```

```
# subsetting the groups into their own df:
group1 <- subset(dataSet, switchPersist == 1)
group2 <- subset(dataSet, switchPersist == 2)
group3 <- subset(dataSet, switchPersist == 3)
group4 <- subset(dataSet, switchPersist == 4)
group5 <- subset(dataSet, switchPersist == 5)
group6 <- subset(dataSet, switchPersist == 6)
group7 <- subset(dataSet, switchPersist == 7)
group8 <- subset(dataSet, switchPersist == 8)
group9 <- subset(dataSet, switchPersist == 9)
group10 <- subset(dataSet, switchPersist == 10)
group11 <- subset(dataSet, switchPersist == 11)

# Subsetting groups again..
group12 <- subset(dataSet, switchPersist == 12)
group13 <- subset(dataSet, switchPersist == 13)
group14 <- subset(dataSet, switchPersist == 14)
group15 <- subset(dataSet, switchPersist == 15)
group16 <- subset(dataSet, switchPersist == 16)
group17 <- subset(dataSet, switchPersist == 17)
group18 <- subset(dataSet, switchPersist == 18)
group19 <- subset(dataSet, switchPersist == 19)
group20 <- subset(dataSet, switchPersist == 20)
group21 <- subset(dataSet, switchPersist == 21)
group22 <- subset(dataSet, switchPersist == 22)
group23 <- subset(dataSet, switchPersist == 23)
group24 <- subset(dataSet, switchPersist == 24)
group25 <- subset(dataSet, switchPersist == 25)
group26 <- subset(dataSet, switchPersist == 26)
```

```
## Creating the career choice grouping and adding it to the df
career <- rep(NA,length(dataSet[,1]))
dataSet <- cbind(dataSet,career)

# STEM #
dataSet$career[dataSet$Q60 == 3] = 1
dataSet$career[dataSet$Q60 == 4] = 1
dataSet$career[dataSet$Q60 == 5] = 1
dataSet$career[dataSet$Q60 == 7] = 1
dataSet$career[dataSet$Q60 == 8] = 1
dataSet$career[dataSet$Q60 == 9] = 1
```

```
# ENGINEERING #
dataSet$career[dataSet$Q60 == 6] = 2

# PRE_MED #
dataSet$career[dataSet$Q60 == 1] = 3
dataSet$career[dataSet$Q60 == 2] = 3

# NON_STEM #
dataSet$career[dataSet$Q60 == 10] = 4
dataSet$career[dataSet$Q60 == 11] = 4
dataSet$career[dataSet$Q60 == 12] = 4
dataSet$career[dataSet$Q60 == 13] = 4
dataSet$career[dataSet$Q60 == 14] = 4
dataSet$career[dataSet$Q60 == 15] = 4

# UNDECIDED #
dataSet$career[dataSet$Q60 == 16] = 5

## subset the careers
STEM <- subset(dataSet, career == 1)
ENGINEERING <- subset(dataSet, career == 2)
PRE_MED <- subset(dataSet, career == 3)
NON_STEM <- subset(dataSet, career == 4)
UNDECIDED <- subset(dataSet, career == 5)
```

```
dfQ18 <- dataSet[,c(22:29)]
dfQ18Rev <- dfQ18
dfQ18Rev[,8] <- (7 - dfQ18Rev[,8])

## Only Complete cases
dfQ18Complete <- dfQ18[complete.cases(dfQ18Rev), ]

## Calculating PCA
PC_Q18 <- prcomp(dfQ18Complete)
PC1_Q18 <- as.numeric(abs(PC_Q18$rotation[,1]))
total <- sum(PC1_Q18)

## Weighted proprtions for each Question
propVec <- PC1_Q18/total


## Weighted mean
instruct <- apply(dfQ18Rev,1,weighted.mean,w=propVec,na.rm=TRUE)

## attaching these values tyo the df
dataSet <- cbind(dataSet,instruct)

dataSet$instruct[is.nan(dataSet$instruct)] <- NA
```

```
## SO do we do the reverse coding before or after PCA's ? (Here reverse coding after...)

dfSCP <- dataSet[,c(30:37)]
dfSCPRev <- dfSCP
dfSCPRev[,6] <- (7 - dfSCP[,6])
```

```
## Only Complete cases
dfSCPComplete <- dfSCPRev[complete.cases(dfSCPRev), ]

## Calculating PCA
PC_SCP <- prcomp(dfSCPComplete)
PC1_SCP <- as.numeric(abs(PC_SCP$rotation[,1]))
total <- sum(PC1_SCP)

## Weighted proprtions for each Question
propVecSCP <- PC1_SCP/total


## Weighted mean
SCP <- apply(dfSCPRev,1,weighted.mean,w=propVecSCP,na.rm=TRUE)

## attaching these values to the df
dataSet <- cbind(dataSet,SCP)

dataSet$SCP[is.nan(dataSet$SCP)] <- NA
```

```
prevCalc <- rep(NA,length(dataSet[,1])) # none experinece = 0
dataSet <- cbind(dataSet,prevCalc)

# hs experince = 1
dataSet$prevCalc[!is.na(dataSet$Q15_CalculusNonAPFinalGrade)
                 |!is.na(dataSet$Q17_CalculusABFinalGrade) |
                   !is.na(dataSet$Q17_CalculusBCFinalGrade)] = 1


# college experience = 2
dataSet$prevCalc[dataSet$Q18 == 1] = 2

dataSet$prevCalc[dataSet$Q18 == 2 & is.na(dataSet$prevCalc)] = 3
```

```
# Excludes 'NA's in our switchpersist column
switch <- subset(dataSet, switchPersist != 'NA')

# Only includes data which was used in the study
# (Q3_SATMath, Q7_ACTMath, Q15_CalculusNonAPFinalGrade, Q17 Questions,
# Q18, Q48, Q60, Q18 Survey questions, Q19 survey questions)
datanew <- switch[complete.cases((switch$Q3_SATMath | switch$Q7_ACTMath)
 & (switch$Q15_CalculusNonAPFinalGrade | switch$Q17_CalculusABFinalGrade
    | switch$Q17_CalculusBCFinalGrade | switch$Q18)
& switch$Q48 & switch$Q60 & (switch$Q18Post_AskedQs | switch$Q18Post_Listened
| switch$Q18Post_Applications | switch$Q18Post_Appointments)
 &(switch$Q19Post_AskQuestions | switch$Q19Post_SpecificProblems |
switch$Q19Post_WorkTogether | switch$Q19Post_Discussion |
switch$Q19Post_Presentations | switch$Q19Post_Individually |
switch$Q19Post_Lecture |
switch$Q19Post_ExplainThinking')),]

# Adjusting SAT and ACT Columns to only have correct values
dtnew <- datanew[which((datanew$Q3_SATMath%%10 == 0 &
 datanew$Q3_SATMath <= 800 & datanew$Q3_SATMath >= 200 ) |
```

```
(datanew$Q7_ACTMath <= 36 & datanew$Q7_ACTMath >= 1)),]
```

```
#####################################################################
################ dtnew has 2,266 observations ###############
#####################################################################
 # dtnew=dataSet is unnecessary here
dtnew$PSAT[dtnew$Q3_SATMath > 800] = NA
dtnew$PSAT[dtnew$Q3_SATMath <= 800]= 99
dtnew$PSAT[dtnew$Q3_SATMath <= 770]= 98
dtnew$PSAT[dtnew$Q3_SATMath <= 760]= 97
dtnew$PSAT[dtnew$Q3_SATMath <= 740]= 96
dtnew$PSAT[dtnew$Q3_SATMath <= 730]= 95
dtnew$PSAT[dtnew$Q3_SATMath <= 710]= 94
dtnew$PSAT[dtnew$Q3_SATMath <= 700]= 93
dtnew$PSAT[dtnew$Q3_SATMath <= 690]= 91
dtnew$PSAT[dtnew$Q3_SATMath <= 680]= 90
dtnew$PSAT[dtnew$Q3_SATMath <= 670]= 88
dtnew$PSAT[dtnew$Q3_SATMath <= 660]= 87
dtnew$PSAT[dtnew$Q3_SATMath <= 650]= 85
dtnew$PSAT[dtnew$Q3_SATMath <= 640]= 83
dtnew$PSAT[dtnew$Q3_SATMath <= 630]= 82
dtnew$PSAT[dtnew$Q3_SATMath <= 620]= 79
dtnew$PSAT[dtnew$Q3_SATMath <= 610]= 77
dtnew$PSAT[dtnew$Q3_SATMath <= 600]= 75
dtnew$PSAT[dtnew$Q3_SATMath <= 590]= 73
dtnew$PSAT[dtnew$Q3_SATMath <= 580]= 70
dtnew$PSAT[dtnew$Q3_SATMath <= 570]= 67
dtnew$PSAT[dtnew$Q3_SATMath <= 560]= 64
dtnew$PSAT[dtnew$Q3_SATMath <= 550]= 62
dtnew$PSAT[dtnew$Q3_SATMath <= 540]= 59
dtnew$PSAT[dtnew$Q3_SATMath <= 530]= 55
dtnew$PSAT[dtnew$Q3_SATMath <= 520]= 52
dtnew$PSAT[dtnew$Q3_SATMath <= 510]= 49
dtnew$PSAT[dtnew$Q3_SATMath <= 500]= 45
dtnew$PSAT[dtnew$Q3_SATMath <= 490]= 42
dtnew$PSAT[dtnew$Q3_SATMath <= 480]= 40
dtnew$PSAT[dtnew$Q3_SATMath <= 470]= 36
dtnew$PSAT[dtnew$Q3_SATMath <= 460]= 33
dtnew$PSAT[dtnew$Q3_SATMath <= 450]= 30
dtnew$PSAT[dtnew$Q3_SATMath <= 440]= 27
dtnew$PSAT[dtnew$Q3_SATMath <= 430]= 24
dtnew$PSAT[dtnew$Q3_SATMath <= 420]= 21
dtnew$PSAT[dtnew$Q3_SATMath <= 410]= 19
dtnew$PSAT[dtnew$Q3_SATMath <= 400]= 16
dtnew$PSAT[dtnew$Q3_SATMath <= 390]= 14
dtnew$PSAT[dtnew$Q3_SATMath <= 380]= 12
dtnew$PSAT[dtnew$Q3_SATMath <= 370]= 10
dtnew$PSAT[dtnew$Q3_SATMath <= 360]= 9
dtnew$PSAT[dtnew$Q3_SATMath <= 350]= 7
dtnew$PSAT[dtnew$Q3_SATMath <= 340]= 6
dtnew$PSAT[dtnew$Q3_SATMath <= 330]= 5
dtnew$PSAT[dtnew$Q3_SATMath <= 320]= 4
dtnew$PSAT[dtnew$Q3_SATMath <= 310]= 3
dtnew$PSAT[dtnew$Q3_SATMath <= 290]= 2
```

```
dtnew$PSAT[dtnew$Q3_SATMath <= 270]= 1
dtnew$PSAT[dtnew$Q3_SATMath <= 220]= 0
dtnew$PSAT[dtnew$Q3_SATMath > 800] = NA
dtnew$PSAT[dtnew$Q3_SATMath < 200] = NA
dtnew$PSAT[dtnew$Q3_SATMath %% 10 != 0] = NA


dtnew$PACT[dtnew$Q7_ACTMath == 1] = 1
dtnew$PACT[dtnew$Q7_ACTMath == 2] = 1
dtnew$PACT[dtnew$Q7_ACTMath == 3] = 1
dtnew$PACT[dtnew$Q7_ACTMath == 4] = 1
dtnew$PACT[dtnew$Q7_ACTMath == 5] = 1
dtnew$PACT[dtnew$Q7_ACTMath == 6] = 1
dtnew$PACT[dtnew$Q7_ACTMath == 7] = 1
dtnew$PACT[dtnew$Q7_ACTMath == 8] = 1
dtnew$PACT[dtnew$Q7_ACTMath == 9] = 1
dtnew$PACT[dtnew$Q7_ACTMath == 10] = 1
dtnew$PACT[dtnew$Q7_ACTMath == 11] = 1
dtnew$PACT[dtnew$Q7_ACTMath == 12] = 4
dtnew$PACT[dtnew$Q7_ACTMath == 13] = 7
dtnew$PACT[dtnew$Q7_ACTMath == 14] = 12
dtnew$PACT[dtnew$Q7_ACTMath == 15] = 18
dtnew$PACT[dtnew$Q7_ACTMath == 16] = 24
dtnew$PACT[dtnew$Q7_ACTMath == 17] = 30
dtnew$PACT[dtnew$Q7_ACTMath == 18] = 36
dtnew$PACT[dtnew$Q7_ACTMath == 19] = 43
dtnew$PACT[dtnew$Q7_ACTMath == 20] = 50
dtnew$PACT[dtnew$Q7_ACTMath == 21] = 56
dtnew$PACT[dtnew$Q7_ACTMath == 22] = 62
dtnew$PACT[dtnew$Q7_ACTMath == 23] = 68
dtnew$PACT[dtnew$Q7_ACTMath == 24] = 74
dtnew$PACT[dtnew$Q7_ACTMath == 25] = 79
dtnew$PACT[dtnew$Q7_ACTMath == 26] = 83
dtnew$PACT[dtnew$Q7_ACTMath == 27] = 87
dtnew$PACT[dtnew$Q7_ACTMath == 28] = 90
dtnew$PACT[dtnew$Q7_ACTMath == 29] = 92
dtnew$PACT[dtnew$Q7_ACTMath == 30] = 95
dtnew$PACT[dtnew$Q7_ACTMath == 31] = 96
dtnew$PACT[dtnew$Q7_ACTMath == 32] = 98
dtnew$PACT[dtnew$Q7_ACTMath == 33] = 99
dtnew$PACT[dtnew$Q7_ACTMath == 34] = 99
dtnew$PACT[dtnew$Q7_ACTMath == 35] = 99
dtnew$PACT[dtnew$Q7_ACTMath == 36] = 99
dtnew$PACT[dtnew$Q3_ACTMath > 36] = NA
dtnew$PACT[dtnew$Q3_ACTMath < 1] = NA


scores<-cbind(dtnew$PSAT,dtnew$PACT)

dtnew$Percentile=rowMeans (scores, na.rm = T)
```

S1

```
library(knitr)
library(kableExtra)

## Table S1 Persisters
library(plyr)
obs <- count(dataSet$switchPersist)
v1s <- 1:11
v2s <- obs[c(1:11), 2]
v3s <- c('Yes', 'Maybe', 'NA', 'NA', 'Yes', 'Yes', 'Maybe', 'Maybe', 'NA', 'NA', 'NA')
v4s <- c(' ', ' ', 'Yes', 'Maybe', 'Yes', ' ', 'Yes', ' ', 'Yes', 'Yes', 'Maybe' )
v5s <- c(' ', ' ', ' ', ' ', 'Maybe', 'No', 'Maybe', 'No', 'Maybe', 'No', 'No')
v6s <- c('No', 'No', 'No', 'No', 'NA', 'NA', 'NA', 'NA', 'NA', 'NA', 'NA')

## Table S1 Persisters
v1p <- seq(12,26,1)
v2p <- obs[c(12:26), 2]
v3p <- c('Yes', 'Maybe', 'NA', 'NA', rep('Yes', 4), rep('Maybe', 4), rep('NA', 3))
v4p <- c(' ', ' ', 'Yes', 'Maybe', ' ', 'Maybe', 'No',
         'NA', ' ', 'Maybe', 'No', 'NA', 'Yes', 'Maybe', 'Maybe')
v5p <- c(rep(' ', 4), 'Yes', 'Maybe', 'Maybe', 'Maybe',
         'Yes', 'Maybe', 'Maybe', 'Maybe', 'Yes', 'Yes', 'Maybe')
v6p <- c(rep('Yes', 4), rep('NA', 11))

S1s <- data.frame(cbind(v1s,v2s,v3s,v4s,v5s,v6s))
names(S1s) <- c("Switcher Group", "Number of Students",
                "Beginning of Term","End of Term; Reflect","End of Term","Follow Up")
S1p <- data.frame(cbind(v1p,v2p,v3p,v4p,v5p,v6p))
names(S1p) <- c("Persister Group", "Number of Students",
                "Beginning of Term","End of Term; Reflect","End of Term","Follow Up")
```

Switcher and Persister Groups (Table S1)

```
kable(S1s, format = 'latex', align=rep('c', 6), booktabs = T,
linesep ="") %>% kable_styling(latex_options = c('striped', 'condensed', 'bordered'),
full_width = F, stripe_color = '#E8E8E8') %>%
column_spec(2:3, width= '20mm', border_left = T) %>%
  column_spec(1, width = '14mm') %>% column_spec(4, width = '24mm',border_left = T)%>%
  column_spec(5:6, width = '12mm',border_left = T)
```

Switcher Group	Number of Students	Beginning of Term	End of Term; Reflect	End of Term	Follow Up
1	160	Yes			No
2	118	Maybe			No
3	15	NA	Yes		No
4	3	NA	Maybe		No
5	38	Yes	Yes	Maybe	NA
6	123	Yes		No	NA
7	17	Maybe	Yes	Maybe	NA
8	152	Maybe		No	NA
9	34	NA	Yes	Maybe	NA
10	78	NA	Yes	No	NA
11	65	NA	Maybe	No	NA

```
kable(S1p, format = 'latex', align=rep('c', 6), booktabs = T, linesep ="") %>%
  kable_styling(latex_options = c('striped', 'condensed', 'bordered'),
full_width = F, stripe_color = '#E8E8E8') %>%
column_spec(2:3, width= '20mm', border_left = T) %>%
  column_spec(1, width = '14mm') %>% column_spec(4, width = '24mm',border_left = T) %>%
  column_spec(5:6, width = '12mm',border_left = T)
```

Persister Group	Number of Students	Beginning of Term	End of Term; Reflect	End of Term	Follow Up
12	586	Yes			Yes
13	63	Maybe			Yes
14	67	NA	Yes		Yes
15	2	NA	Maybe		Yes
16	1543	Yes		Yes	NA
17	35	Yes	Maybe	Maybe	NA
18	5	Yes	No	Maybe	NA
19	1	Yes	NA	Maybe	NA
20	193	Maybe		Yes	NA
21	64	Maybe	Maybe	Maybe	NA
22	22	Maybe	No	Maybe	NA
23	3	Maybe	NA	Maybe	NA
24	1325	NA	Yes	Yes	NA
25	53	NA	Maybe	Yes	NA
26	103	NA	Maybe	Maybe	NA

S5 Table

```
# creating an indicator variable for switch persist where 1 == switc, 0 == persist
SWP <- rep(NA,nrow(dtnew))
dtnew <- cbind(dtnew,SWP)

for(i in seq(1,11,1)){
dtnew$SWP[dtnew$switchPersist == i] = 1
}

for(i in seq(12,26,1)){
dtnew$SWP[dtnew$switchPersist == i] = 0
}

# Subsetting by career choices
STEM <- subset(dtnew, career == 1)
ENGINEERING <- subset(dtnew, career == 2)
PRE_MED <- subset(dtnew, career == 3)
NON_STEM <- subset(dtnew, career == 4)
UNDECIDED <- subset(dtnew, career == 5)


p1 <- count(STEM$Q48 == 1 & STEM$SWP == 1)[2,2]/count(STEM$Q48 == 1)[2,2]
p2 <- count(STEM$Q48 == 2 & STEM$SWP == 1)[2,2]/count(STEM$Q48 == 2)[2,2]
p3 <- count(ENGINEERING$Q48 == 1 & ENGINEERING$SWP == 1)[2,2]/count(ENGINEERING$Q48 == 1)[2,2]
p4 <- count(ENGINEERING$Q48 == 2 & ENGINEERING$SWP == 1)[2,2]/count(ENGINEERING$Q48 == 2)[2,2]
```

```
p5 <- count(PRE_MED$Q48 == 1 & PRE_MED$SWP == 1)[2,2]/count(PRE_MED$Q48 == 1)[2,2]
p6 <- count(PRE_MED$Q48 == 2 & PRE_MED$SWP == 1)[2,2]/count(PRE_MED$Q48 == 2)[2,2]
p7 <- count(NON_STEM$Q48 == 1 & NON_STEM$SWP == 1)[2,2]/count(NON_STEM$Q48 == 1)[2,2]
p8 <- count(NON_STEM$Q48 == 2 & NON_STEM$SWP == 1)[2,2]/count(NON_STEM$Q48 == 2)[2,2]
p9 <- count(UNDECIDED$Q48 == 1 & UNDECIDED$SWP == 1)[2,2]/count(UNDECIDED$Q48 == 1)[2,2]
p10 <- count(UNDECIDED$Q48 == 2 & UNDECIDED$SWP == 1)[2,2]/count(UNDECIDED$Q48 == 2)[2,2]


# Creating columns
S5col1 <- c("STEM", " " , "Engineering" , " " , "Pre-Med" ,
            " " , "Non-STEM" , " " , "Undecided" , " ")
S5col2 <- c("Male" , "Female" , "Male" , "Female" , "Male" ,
            "Female" ,"Male" , "Female" ,"Male" , "Female")
S5col3 <- c( count(STEM$Q48 ==2)[1,2],count(STEM$Q48 ==2)[2,2],
count(ENGINEERING$Q48 ==2)[1,2],count(ENGINEERING$Q48 ==2)[2,2] , count(PRE_MED$Q48 ==2)[1,2],count(PRE_
count(NON_STEM$Q48 ==2)[2,2] , count(UNDECIDED$Q48 ==2)[1,2],count(UNDECIDED$Q48 ==2)[2,2])
S5col4 <- round(c(p1,p2,p3,p4,p5,p6,p7,p8,p9,p10),digits = 3)*100

S5 <- data.frame(cbind(S5col1,S5col2,S5col3,S5col4)) # add first row names
```

S6 Table

```
S6col2 <- c("Male","Female","Male","Female","Male","Female")
S6col1 <- c("High School" , " " , "College" , " " , "None" , " ")

S6col3 <- c( count(dtnew$Q48 == 1 & dtnew$prevCalc == 1)[2,2] ,
             count(dtnew$Q48 == 2 & dtnew$prevCalc == 1)[2,2] ,
             count(dtnew$Q48 == 1 & dtnew$prevCalc == 2)[2,2]  ,
             count(dtnew$Q48 == 2 & dtnew$prevCalc == 2)[2,2] ,
             count(dtnew$Q48 == 1 & dtnew$prevCalc == 3)[2,2] ,
             count(dtnew$Q48 == 2 & dtnew$prevCalc == 3)[2,2])

p1_6 <- count(dtnew$Q48 == 1 & dtnew$prevCalc == 1 & dtnew$SWP == 1)[2,2] / count(dtnew$Q48 == 1 & dtnew
p2_6 <- count(dtnew$Q48 == 2 & dtnew$prevCalc == 1 & dtnew$SWP == 1)[2,2] / count(dtnew$Q48 == 2 & dtnew
p3_6 <- count(dtnew$Q48 == 1 & dtnew$prevCalc == 2 & dtnew$SWP == 1)[2,2] / count(dtnew$Q48 == 1 & dtnew
p4_6 <- count(dtnew$Q48 == 2 & dtnew$prevCalc == 2 & dtnew$SWP == 1)[2,2] / count(dtnew$Q48 == 2 & dtnew
p5_6 <- count(dtnew$Q48 == 1 & dtnew$prevCalc == 3 & dtnew$SWP == 1)[2,2] / count(dtnew$Q48 == 1 & dtnew
p6_6 <- count(dtnew$Q48 == 2 & dtnew$prevCalc == 3 & dtnew$SWP == 1)[2,2] / count(dtnew$Q48 == 2 & dtnew

S6col4 <- round(100*c(p1_6,p2_6,p3_6,p4_6,p5_6,p6_6), digits = 1)

S6 <- data.frame(cbind(S6col1,S6col2,S6col3,S6col4))
```

```
names(S5) <- c('Career Choice', 'Gender', 'N', 'Switcher %')

kable(S5, format = 'latex', booktabs = T, linesep = c('','\\addlinespace'),
      align = c('c', 'l','c','c')) %>% kable_styling(latex_options = 'striped',
stripe_color = '#E8E8E8') %>% column_spec(2:4, border_left = T)  %>%
  row_spec(0, bold = T) %>% column_spec(1, bold = T)
```

Career Choice	Gender	N	Switcher %
STEM	Male	263	10.6
	Female	223	16.1
Engineering	Male	539	3.5
	Female	249	6.4
Pre-Med	Male	199	21.6
	Female	318	33.3
Non-STEM	Male	136	36.8
	Female	126	38.1
Undecided	Male	99	26.3
	Female	114	28.1

```
# row_spec(1, background = '#A4D3EE') %>% row_spec(3, background = '#A4D3EE')
```

```
names(S6) <- c('Previous Calculus', 'Gender', 'N', 'Switcher %')
kable(S6, format = 'latex', booktabs = T, linesep = c('','\\addlinespace'),
      align = c('c', 'l','c','c')) %>% kable_styling(latex_options = 'striped',
stripe_color = '#E8E8E8') %>% column_spec(2:4, border_left = T)  %>%
  row_spec(0, bold = T) %>% column_spec(1, bold = T)
```

Previous Calculus	Gender	N	Switcher %
High School	Male	740	13
	Female	699	21.7
College	Male	99	17.2
	Female	62	14.5
None	Male	397	13.4
	Female	269	28.6

S7 Table

```
## Subsetting into the 10 ranges first ##
Per1 <- subset(dtnew, Percentile <= 100 & Percentile >= 90)
Per2 <- subset(dtnew, Percentile < 90 & Percentile >= 80)
Per3 <- subset(dtnew, Percentile < 80 & Percentile >= 70)
Per4 <- subset(dtnew, Percentile < 70 & Percentile >= 60)
Per5 <- subset(dtnew, Percentile < 60 & Percentile >= 50)
Per6 <- subset(dtnew, Percentile < 50 & Percentile >= 40)
Per7 <- subset(dtnew, Percentile < 40 & Percentile >= 30)
Per8 <- subset(dtnew, Percentile < 30 & Percentile >= 20)
Per9 <- subset(dtnew, Percentile < 20 & Percentile >= 10)
Per10 <- subset(dtnew, Percentile < 9 )

S7col1 <- c("90 - 100" , " " , "80 - 89" , " ", "70 - 79" , " " , "60 - 69" , " ", "50 - 59" , " " , "40

S7col2 <- c(rep(c("Male","Female"),10))
```

```
S7col3 <- c( count(Per1$Q48 == 1)[2,2],count(Per1$Q48 == 2)[2,2] , count(Per2$Q48 == 1)[2,2],count(Per2$
```

```
p1_7 <- count(Per1$Q48 == 1 & Per1$SWP == 1)[2,2]/count(Per1$Q48 == 1)[2,2]
p2_7 <- count(Per1$Q48 == 2 & Per1$SWP == 1)[2,2]/count(Per1$Q48 == 2)[2,2]

p3_7 <- count(Per2$Q48 == 1 & Per2$SWP == 1)[2,2]/count(Per2$Q48 == 1)[2,2]
p4_7 <- count(Per2$Q48 == 2 & Per2$SWP == 1)[2,2]/count(Per2$Q48 == 2)[2,2]

p5_7 <- count(Per3$Q48 == 1 & Per3$SWP == 1)[2,2]/count(Per3$Q48 == 1)[2,2]
p6_7 <- count(Per3$Q48 == 2 & Per3$SWP == 1)[2,2]/count(Per3$Q48 == 2)[2,2]

p7_7 <- count(Per4$Q48 == 1 & Per4$SWP == 1)[2,2]/count(Per4$Q48 == 1)[2,2]
p8_7 <- count(Per4$Q48 == 2 & Per4$SWP == 1)[2,2]/count(Per4$Q48 == 2)[2,2]

p9_7 <- count(Per5$Q48 == 1 & Per5$SWP == 1)[2,2]/count(Per5$Q48 == 1)[2,2]
p10_7 <- count(Per5$Q48 == 2 & Per5$SWP == 1)[2,2]/count(Per5$Q48 == 2)[2,2]

p11_7 <- count(Per6$Q48 == 1 & Per6$SWP == 1)[2,2]/count(Per6$Q48 == 1)[2,2]
p12_7 <- count(Per6$Q48 == 2 & Per6$SWP == 1)[2,2]/count(Per6$Q48 == 2)[2,2]

p13_7 <- count(Per7$Q48 == 1 & Per7$SWP == 1)[2,2]/count(Per7$Q48 == 1)[2,2]
p14_7 <- count(Per7$Q48 == 2 & Per7$SWP == 1)[2,2]/count(Per7$Q48 == 2)[2,2]

p15_7 <- count(Per8$Q48 == 1 & Per8$SWP == 1)[2,2]/count(Per8$Q48 == 1)[2,2]
p16_7 <- count(Per8$Q48 == 2 & Per8$SWP == 1)[2,2]/count(Per8$Q48 == 2)[2,2]

p17_7 <- count(Per9$Q48 == 1 & Per9$SWP == 1)[2,2]/count(Per9$Q48 == 1)[2,2]
p18_7 <- count(Per9$Q48 == 2 & Per9$SWP == 1)[2,2]/count(Per9$Q48 == 2)[2,2]

p19_7 <- count(Per10$Q48 == 1 & Per10$SWP == 1)[2,2]/count(Per10$Q48 == 1)[2,2]
p20_7 <- count(Per10$Q48 == 2 & Per10$SWP == 1)[2,2]/count(Per10$Q48 == 2)[2,2]

## Making 'NA' observations 0 and combining into column

S7col4 <- round(100 * c(p1_7,p2_7,p3_7,p4_7,p5_7,p6_7,p8_7,p9_7,p10_7
        ,p11_7,p12_7,p13_7,p14_7,0,0,p17_7,p18_7,0,p20_7), digits = 1)

S7 <- data.frame(cbind(S7col1,S7col2,S7col3,S7col4))
```

```
S7$S7col4[S7$S7col4 == 'NA'] = 0
S7new <- data.frame(c(S7[1:10,], S7[11:20,]))

names(S7new) <- c('Test Percentile', 'Gender', 'N', 'Switcher %','Test Percentile', 'Gender', 'N', 'Swi

kable(S7new, format = 'latex', booktabs = T, linesep = c('','\\addlinespace'),
      align = c('c', 'l','c','c','c', 'l','c','c')) %>%
  kable_styling(latex_options = 'striped', stripe_color = '#E8E8E8') %>%
  column_spec(2:8, border_left = T)  %>% row_spec(0, bold = T) %>%
  column_spec(c(1,5), bold = T) %>% column_spec(4, border_right = T)
```

Test Percentile	Gender	N	Switcher %	Test Percentile	Gender	N	Switcher %
90 - 100	Male	707	12.4	**40 - 49**	Male	20	26.7
	Female	523	22.6		Female	30	10
80 - 89	Male	250	14.8	**30 - 39**	Male	10	33.3
	Female	220	22.3		Female	6	0
70 - 79	Male	165	12.7	**20 - 29**	Male	3	0
	Female	157	25.5		Female	2	50
60 - 69	Male	50	18.3	**10 - 19**	Male	2	50
	Female	60	23.1		Female	2	0
50 - 59	Male	26	27.6	**0 - 9**	Male	3	100
	Female	29	20		Female	1	12.4

S8 Table

```
In1 <- dtnew[dtnew$instruct >= 5.5,]
In2 <- dtnew[dtnew$instruct >= 4.5 & dtnew$instruct < 5.5 ,]
In3 <- dtnew[dtnew$instruct >= 3.5 & dtnew$instruct < 4.5 ,]
In4 <- dtnew[dtnew$instruct >= 2.5 & dtnew$instruct < 3.5 ,]
In5 <- dtnew[dtnew$instruct >= 1.5 & dtnew$instruct < 2.5 ,]
In6 <- dtnew[dtnew$instruct >= 1 & dtnew$instruct < 1.5 ,]


SCP1 <- dtnew[dtnew$SCP >= 5.5 & dtnew$SCP < 6,]
SCP2 <- dtnew[dtnew$SCP >= 4.5 & dtnew$SCP < 5.5,]
SCP3 <- dtnew[dtnew$SCP >= 3.5 & dtnew$SCP < 4.5,]
SCP4 <- dtnew[dtnew$SCP >= 2.5 & dtnew$SCP < 3.5,]
SCP5 <- dtnew[dtnew$SCP >= 1.5 & dtnew$SCP < 2.5,]
SCP6 <- dtnew[dtnew$SCP >= 1 & dtnew$SCP < 1.5,]

S8col1 <- c("[5.5, 6]","","[4.5, 5.5)","","[3.5, 4.5)","","[2.5, 3.5)","","[1.5, 2.5)","","[1, 1.5)",""
S8col1 <- c("$[5.5, 6]$","","$[4.5, 5.5)$","","$[3.5, 4.5)$","","$[2.5, 3.5)$","","$[1.5, 2.5)$","","$[

S8col2 <- rep(c("Male","Female"),6)
S8col3_INSTRUCT <- c(
  count(In1$Q48 == 1)[2,2],
  count(In1$Q48 == 2)[2,2],
  count(In2$Q48 == 1)[2,2],
  count(In2$Q48 == 2)[2,2],
  count(In3$Q48 == 1)[2,2],
  count(In3$Q48 == 2)[2,2],
  count(In4$Q48 == 1)[2,2],
  count(In4$Q48 == 2)[2,2],
  count(In5$Q48 == 1)[2,2],
  count(In5$Q48 == 2)[2,2],
  count(In6$Q48 == 1)[2,2],
  count(In6$Q48 == 2)[2,2]
)

S8col3_SCP <- c(
  count(SCP1$Q48 == 1)[2,2],
```

```
  count(SCP1$Q48 == 2)[2,2],
  count(SCP2$Q48 == 1)[2,2],
  count(SCP2$Q48 == 2)[2,2],
  count(SCP3$Q48 == 1)[2,2],
  count(SCP3$Q48 == 2)[2,2],
  count(SCP4$Q48 == 1)[2,2],
  count(SCP4$Q48 == 2)[2,2],
  count(SCP5$Q48 == 1)[2,2],
  count(SCP5$Q48 == 2)[2,2],
  count(SCP6$Q48 == 1)[2,2],
  count(SCP6$Q48 == 2)[2,2]
)

S8col4_INSTRUCT <- round(c(
  count(In1$Q48 == 1 & In1$SWP == 1)[2,2]/count(In1$Q48 == 1)[2,2],
  count(In1$Q48 == 2 & In1$SWP == 1)[2,2]/count(In1$Q48 == 2)[2,2],
  count(In2$Q48 == 1 & In2$SWP == 1)[2,2]/count(In2$Q48 == 1)[2,2],
  count(In2$Q48 == 2 & In2$SWP == 1)[2,2]/count(In2$Q48 == 2)[2,2],
  count(In3$Q48 == 1 & In3$SWP == 1)[2,2]/count(In3$Q48 == 1)[2,2],
  count(In3$Q48 == 2 & In3$SWP == 1)[2,2]/count(In3$Q48 == 2)[2,2],
  count(In4$Q48 == 1 & In4$SWP == 1)[2,2]/count(In4$Q48 == 1)[2,2],
  count(In4$Q48 == 2 & In4$SWP == 1)[2,2]/count(In4$Q48 == 2)[2,2],
  count(In5$Q48 == 1 & In5$SWP == 1)[2,2]/count(In5$Q48 == 1)[2,2],
  count(In5$Q48 == 2 & In5$SWP == 1)[2,2]/count(In5$Q48 == 2)[2,2],
  count(In6$Q48 == 1 & In6$SWP == 1)[2,2]/count(In6$Q48 == 1)[2,2],
  count(In6$Q48 == 2 & In6$SWP == 1)[2,2]/count(In6$Q48 == 2)[2,2]
)*100, digits = 1)

S8col4_SCP <- round(c(
  count(SCP1$Q48 == 1 & SCP1$SWP == 1)[2,2]/count(SCP1$Q48 == 1)[2,2],
  count(SCP1$Q48 == 2 & SCP1$SWP == 1)[2,2]/count(SCP1$Q48 == 2)[2,2],
  count(SCP2$Q48 == 1 & SCP2$SWP == 1)[2,2]/count(SCP2$Q48 == 1)[2,2],
  count(SCP2$Q48 == 2 & SCP2$SWP == 1)[2,2]/count(SCP2$Q48 == 2)[2,2],
  count(SCP3$Q48 == 1 & SCP3$SWP == 1)[2,2]/count(SCP3$Q48 == 1)[2,2],
  count(SCP3$Q48 == 2 & SCP3$SWP == 1)[2,2]/count(SCP3$Q48 == 2)[2,2],
  count(SCP4$Q48 == 1 & SCP4$SWP == 1)[2,2]/count(SCP4$Q48 == 1)[2,2],
  count(SCP4$Q48 == 2 & SCP4$SWP == 1)[2,2]/count(SCP4$Q48 == 2)[2,2],
  count(SCP5$Q48 == 1 & SCP5$SWP == 1)[2,2]/count(SCP5$Q48 == 1)[2,2],
  count(SCP5$Q48 == 2 & SCP5$SWP == 1)[2,2]/count(SCP5$Q48 == 2)[2,2],
  count(SCP6$Q48 == 1 & SCP6$SWP == 1)[2,2]/count(SCP6$Q48 == 1)[2,2],
  count(SCP6$Q48 == 2 & SCP6$SWP == 1)[2,2]/count(SCP6$Q48 == 2)[2,2]
)*100, digits = 1)

S8_INSTUCT <- data.frame(cbind(S8col1,S8col2,S8col3_INSTRUCT,S8col4_INSTRUCT))
S8_SCP <- data.frame(cbind(S8col1,S8col2,S8col3_SCP,S8col4_SCP))
```

```
S8 <- data.frame(cbind(S8_INSTUCT,S8_SCP))
names(S8) <- c("Instructor Quality Rating", 'Gender', 'N', 'Switcher %',"Student-Centered Instruction R

kable(S8, format = 'markdown', booktabs = T, linesep = c('','\\addlinespace')) %>% kable_styling('stripe
```

Instructor Quality Rating	Gender	N	Switcher %	Student-Centered Instruction Response	Gender	N	Switcher %

Instructor Quality Rating	Gender	N	Switcher %	Student-Centered Instruction Response	Gender	N	Switcher %
[5.5, 6]	Male	165	9.1	[5.5, 6]	Male	17	17.6
	Female	152	16.4		Female	21	38.1
[4.5, 5.5)	Male	613	12.2	[4.5, 5.5)	Male	155	12.3
	Female	464	20.3		Female	124	25
[3.5, 4.5)	Male	323	13.6	[3.5, 4.5)	Male	363	12.1
	Female	281	25.3		Female	282	22
[2.5, 3.5)	Male	95	24.2	[2.5, 3.5)	Male	393	16.3
	Female	89	29.2		Female	296	23.6
[1.5, 2.5)	Male	33	24.2	[1.5, 2.5)	Male	252	11.1
	Female	37	48.6		Female	257	23
[1, 1.5)	Male	7	14.3	[1, 1.5)	Male	56	14.3
	Female	7	57.1		Female	50	16

```
# kable(S8_INSTUCT, format = 'latex', booktabs = T, linesep = c('','\\addlinespace'), escape = F, align

# kable(S8, format = 'latex', booktabs = T, linesep = c('','\\addlinespace'), escape = F, align = c('c'
```