Self-regulated Learning

INFO 5200 Learning Analytics: Week 9 Homework

[[Cole Walsh, 4399966]]

In this homework, you will learn how to mine a clickstream dataset for correlational patterns. You are given the SRL survey responses and the most recent export of the edx clickstream log for this course.

The SRL responses are recoded to range from 0 (low) to 4 (high) for each strategy index.

Learning Objectives:

1. Exploring response distributions of survey data
2. Merging survey with behavioral data
3. Engineering features that could represent SRL strategies
4. Checking if any behavioral features predict survey responses using a linear model

Part 1: Explore responses and merge data

Question 1: Plot a boxplot for the responses for each SRL strategy in the dataset (6 in total).

```
# Import cleaned up SRL data
srl = read.csv("info5200.6.srl.csv")

##################################################
####### BEGIN INPUT: Plot the distribution #######
##################################################

library(reshape2)
```
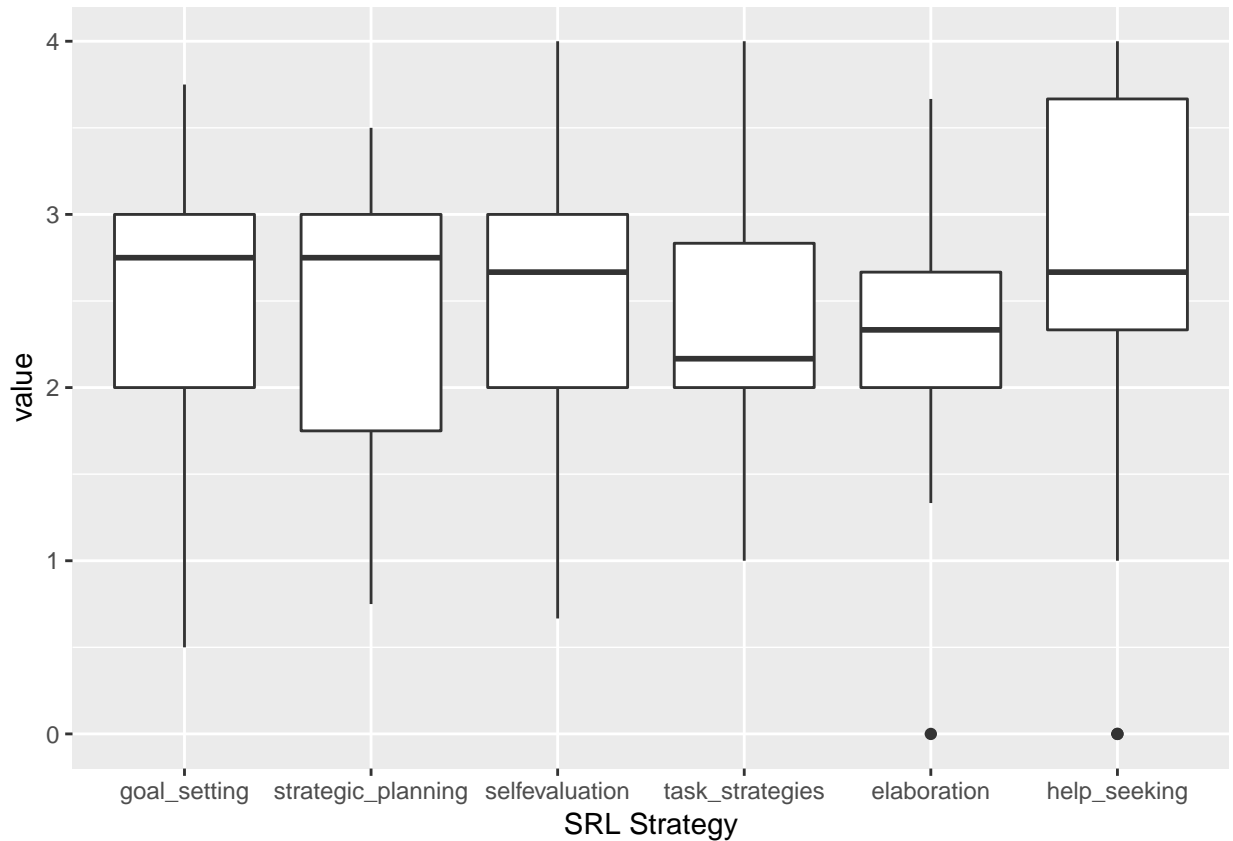
```
##
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':
##
##     smiths
```

```
srl_melted <- melt(srl, id.vars = c('survey_id'))
ggplot(srl_melted, aes(x = variable, y = value)) +
  geom_boxplot() +
  xlab('SRL Strategy')
```

```
#################################################
#################################################
```

Question 2: Now merge the clickstream data with the SRL survey data using the `survey_id` identifier. Note that not everyone in the clickstream data has survey records. You only need to keep students who have survey data. Thus, you should end up with all clickstream records for students who filled out the survey: 29 students and 65,407 rows total.

```r
# Import edx clickstream log
cl = readRDS("info5200.6.clickstream.rds")

#################################################
####### BEGIN INPUT: Merge Datasets     ########
#################################################

df <- left_join(srl, cl, by = 'survey_id')
```

```
## Warning: Column `survey_id` joining factor and character vector, coercing
## into character vector
```

```r
length(unique(df$survey_id))
```

```
## [1] 29
```

```r
nrow(df)
```

```
## [1] 65407
```

```r
#################################################
#################################################
```

Part 2: Engineering Features

Once again you get to practice feature engineering. By now you know what the clickstream data looks like thanks to the prediction project. Unlike in the project, however, you do not need to worry this time about the timing. Your goal is to come up with behavioral indicators of self-regulated learning. You know what the survey responses measure (check out the actual questions) and you know what the clickstream records represent (check out the corresponding pages in the course).

Question 3: Before coding up variables, describe in words ONE OR MORE (sequences/patterns of) actions that you think might be indicative of EACH SRL strategy. It will be easier for some than for others. At the end, you can describe some additional "general" activity features.

```r
#################################################
####### BEGIN INPUT: Plan features ############
#################################################


# goal setting
# 1. Total number of sessions
# 2. Average time per session

# strategic planning
# 1. Number of times displaying side bar
# 2. Number of times visiting syllabus

# self-evaluation
# 1. Number of of times a previous module is navigated to
# 2. Number of times checking grades

# task strategies
# 1. Median time between sessions
# 2. Number of times a user checks the next unit in a sequnce

# elaboration
# 1. Number of times a user jumps between units
# 2. Number of times returning to main course page

# help seeking
# 1. Total number of clicks
# 2. Number of times visiting discussion forums

# general features
# 1. Problems submitted
# 2. Average number of clicks per session


#################################################
#################################################
```

Question 4: For each student, engineer the features you described above using the clickstream data. Combine all features into a dataset that has 1 row per student (29 rows here) with all features you created and the SRL data.

```r
###################################################
####### BEGIN INPUT: Feature engineering #########
###################################################

# goal setting features

# Define new session as being at least 1h after previous click

df.TimeStamp <- df %>%
  group_by(survey_id) %>%
  arrange(timestamp) %>%
  mutate(Time.diff = replace_na(timestamp -lag(timestamp), 3601)) %>%
  group_by(survey_id) %>%
  mutate(New.Session = Time.diff > 3600,
         Session.Number = cumsum(New.Session))

gs = df.TimeStamp %>%
  group_by(survey_id) %>%
  summarize(Num.Sessions = max(Session.Number)) %>%
  left_join(., df.TimeStamp %>%
              group_by(survey_id, Session.Number) %>%
              summarize(Session.Length = max(timestamp) - min(timestamp)) %>%
              group_by(survey_id) %>%
              summarize(Avg.Session.Length = mean(Session.Length)), by = 'survey_id')

# strategic planning features

sp = df %>%
  group_by(survey_id) %>%
  summarize(Side.Bar.Displays = sum(event_type ==
                                    paste('edx.bi.course.upgrade.sidebarupsell.',
                                          'displayed', sep = ''), na.rm = TRUE),
            Syllabus.visits = sum(referer ==
                                  paste('https://edge.edx.org/courses/course-v1:',
'CornellX+INFO5200+2019_Spring/courseware/2d815b2e787344838a1509c7a5861d2d/',
'6fffcdccb3b84a8cbc79c173cbbe20e8/1?activate_block_id=block-v1%3ACornellX%2BINFO5200%',
'2B2019_Spring%2Btype%40vertical%2Bblock%404337136773ca4198aa081242fcbf56d2', sep = ''),
na.rm = TRUE))

# self-evaluation features

se = df %>%
  group_by(survey_id) %>%
  summarize(Prev.Unit = sum(event_type == 'seq_prev', na.rm = TRUE),
            Num.Check.Grades = sum(referer ==
                                   paste('https://edge.edx.org/courses/course-v1:',
'CornellX+INFO5200+2019_Spring/progress', sep = ''), na.rm = TRUE))

# task strategies features
```

4

```r
ts = df.TimeStamp %>%
  filter(New.Session == TRUE) %>%
  group_by(survey_id) %>%
  summarize(Med.Time.Btw.Sessions = median(Time.diff)) %>%
  left_join(., df %>%
              group_by(survey_id) %>%
              summarize(Next.Unit = sum(event_type == 'seq_next', na.rm = TRUE)),
            by = 'survey_id')

# elaboration features

el = df %>%
  group_by(survey_id) %>%
  summarize(Jump.Unit = sum(event_type == 'seq_goto', na.rm = TRUE),
            Main.Page.Visits = sum(referer ==
                                    paste('https://edge.edx.org/courses/course-v1:',
'CornellX+INFO5200+2019_Spring/course/', sep = ''), na.rm = TRUE))

# help seeking features

hs = df %>%
  group_by(survey_id) %>%
  summarize(N.Clicks = n(),
            Forum.Visits = sum(referer ==
                                paste('https://edge.edx.org/courses/course-v1:',
'CornellX+INFO5200+2019_Spring/discussion/forum/', sep = ''), na.rm = TRUE))

# general features

ge = df %>%
  group_by(survey_id) %>%
  summarize(Num.Problems.Sub = sum(event_type == 'edx.grades.problem.submitted',
                                    na.rm = TRUE)) %>%
  left_join(., df.TimeStamp %>%
              group_by(survey_id, Session.Number) %>%
              summarize(N.Clicks = n()) %>%
              group_by(survey_id) %>%
              summarize(N.Clicks.Session = mean(N.Clicks)), by = 'survey_id')

# combine into one dataset

# When I left joined the cl and srl datasets, I
# renamed it df, thus I have changed cl to df here

dat = df %>%
    select(survey_id, goal_setting:help_seeking) %>%
    unique %>%
    left_join(gs) %>%
    left_join(sp) %>%
    left_join(se) %>%
    left_join(ts) %>%
    left_join(el) %>%
    left_join(hs) %>%
```

```
    left_join(ge)
```

```
## Joining, by = "survey_id"
## Joining, by = "survey_id"
## Joining, by = "survey_id"
## Joining, by = "survey_id"
## Joining, by = "survey_id"
## Joining, by = "survey_id"
## Joining, by = "survey_id"
```

```
##############################################
##############################################
```

Part 3: Explore SRL Association

There are many options for how to check if the behavioral features are associated with SRL strategies. We will use a very simple method: linear regression predicting each SRL index (self-report) with the relevant behavioral features. (Feel free to try out more complex ideas.)

Question 5: For each SRL index (goal_setting, strategic_planning, etc.) as the outcome, fit TWO linear regression models (`lm()`): one that only has the relevant features you describe above, and another that has all of the features.

For example, if you created two features specifically to indicate help seeking (H1 and H2), then you fit one model `help_seeking ~ H1 + H2` and a second model `help_seeking ~ H1 + H2 + all_other_features`. For each one, you should output the `summary()` of the `lm()` object. **Do not use the survey_id and the other SRL measures as predictors in the model!** So if you predict help_seeking then do not have goal_setting etc. in the model.

```
#####################################################
####### BEGIN INPUT: Fit Regression Models #########
#####################################################

# goal setting
summary(lm(goal_setting ~ Num.Sessions + Avg.Session.Length, dat))
```

```
##
## Call:
## lm(formula = goal_setting ~ Num.Sessions + Avg.Session.Length,
##     data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5833 -0.6171  0.1841  0.5587  1.3039
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         2.096e+00  6.304e-01   3.325  0.00264 **
## Num.Sessions        4.629e-03  3.391e-03   1.365  0.18394
## Avg.Session.Length -7.096e-05  2.261e-04  -0.314  0.75615
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

6

```
## 
## Residual standard error: 0.7808 on 26 degrees of freedom
## Multiple R-squared:  0.08368,    Adjusted R-squared:  0.0132
## F-statistic: 1.187 on 2 and 26 DF,  p-value: 0.3211
```

```r
summary(lm(goal_setting ~ . - survey_id - strategic_planning - selfevaluation - task_strategies
           - elaboration - help_seeking, dat))
```

```
## 
## Call:
## lm(formula = goal_setting ~ . - survey_id - strategic_planning -
##     selfevaluation - task_strategies - elaboration - help_seeking,
##     data = dat)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8228 -0.2256  0.1082  0.2109  0.6165
## 
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           2.776e+00  8.454e-01   3.283  0.00544 **
## Num.Sessions         -1.765e-03  5.601e-03  -0.315  0.75730
## Avg.Session.Length    1.311e-04  1.855e-04   0.707  0.49113
## Side.Bar.Displays     7.385e-04  5.829e-03   0.127  0.90098
## Syllabus.visits      -5.138e-03  2.798e-03  -1.836  0.08767 .
## Prev.Unit             1.130e-01  4.275e-02   2.643  0.01928 *
## Num.Check.Grades      1.734e-02  5.922e-03   2.927  0.01103 *
## Med.Time.Btw.Sessions -2.565e-05 1.786e-05  -1.436  0.17298
## Next.Unit            -2.758e-02  1.172e-02  -2.354  0.03369 *
## Jump.Unit             3.433e-02  1.037e-02   3.311  0.00515 **
## Main.Page.Visits      4.911e-03  2.156e-03   2.278  0.03895 *
## N.Clicks             -2.664e-03  8.351e-04  -3.190  0.00655 **
## Forum.Visits         -2.091e-02  1.017e-02  -2.056  0.05896 .
## Num.Problems.Sub      1.445e-01  6.884e-02   2.099  0.05442 .
## N.Clicks.Session      1.088e-02  5.164e-02   0.211  0.83615
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.5181 on 14 degrees of freedom
## Multiple R-squared:  0.7828, Adjusted R-squared:  0.5655
## F-statistic: 3.603 on 14 and 14 DF,  p-value: 0.01121
```

```r
# strategic planning
summary(lm(strategic_planning ~ Side.Bar.Displays + Syllabus.visits, dat))
```

```
## 
## Call:
## lm(formula = strategic_planning ~ Side.Bar.Displays + Syllabus.visits,
##     data = dat)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.78400 -0.49905  0.09114  0.39449  1.17628
```

```
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.253609   0.297064   7.586 4.72e-08 ***
## Side.Bar.Displays 0.001700   0.001676   1.014    0.320
## Syllabus.visits  -0.002351   0.003418  -0.688    0.498
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7039 on 26 degrees of freedom
## Multiple R-squared:  0.04911,    Adjusted R-squared:  -0.02404
## F-statistic: 0.6714 on 2 and 26 DF,  p-value: 0.5196
```

```r
summary(lm(strategic_planning ~ . - survey_id - goal_setting - selfevaluation - task_strategies
           - elaboration - help_seeking, dat))
```

```
##
## Call:
## lm(formula = strategic_planning ~ . - survey_id - goal_setting -
##     selfevaluation - task_strategies - elaboration - help_seeking,
##     data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.39735 -0.25190 -0.00993  0.34297  0.72892
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          2.010e+00  1.191e+00   1.687    0.114
## Num.Sessions         2.680e-03  7.892e-03   0.340    0.739
## Avg.Session.Length   2.520e-04  2.613e-04   0.964    0.351
## Side.Bar.Displays   -6.341e-03  8.213e-03  -0.772    0.453
## Syllabus.visits     -1.875e-03  3.943e-03  -0.476    0.642
## Prev.Unit            6.713e-02  6.023e-02   1.115    0.284
## Num.Check.Grades     7.054e-03  8.344e-03   0.845    0.412
## Med.Time.Btw.Sessions -1.509e-05 2.516e-05  -0.599    0.558
## Next.Unit           -2.285e-02  1.651e-02  -1.384    0.188
## Jump.Unit            1.061e-02  1.461e-02   0.727    0.479
## Main.Page.Visits     3.835e-03  3.038e-03   1.262    0.227
## N.Clicks            -1.536e-03  1.177e-03  -1.305    0.213
## Forum.Visits         1.073e-03  1.434e-02   0.075    0.941
## Num.Problems.Sub     1.346e-01  9.700e-02   1.387    0.187
## N.Clicks.Session    -3.126e-03  7.276e-02  -0.043    0.966
##
## Residual standard error: 0.73 on 14 degrees of freedom
## Multiple R-squared:  0.4493, Adjusted R-squared:  -0.1014
## F-statistic: 0.8158 on 14 and 14 DF,  p-value: 0.6457
```

```r
# self-evaluation
summary(lm(selfevaluation ~ Prev.Unit + Num.Check.Grades, dat))
```

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

```
## lm(formula = selfevaluation ~ Prev.Unit + Num.Check.Grades, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7926 -0.4230 -0.1065  0.5553  1.4081
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.446958   0.263221   9.296 9.47e-10 ***
## Prev.Unit       -0.002402   0.046788  -0.051    0.959
## Num.Check.Grades 0.004907   0.005124   0.958    0.347
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8451 on 26 degrees of freedom
## Multiple R-squared:  0.03411,    Adjusted R-squared:  -0.04019
## F-statistic: 0.4591 on 2 and 26 DF,  p-value: 0.6369
```

```r
summary(lm(selfevaluation ~ . - survey_id - goal_setting - strategic_planning - task_strategies
           - elaboration - help_seeking, dat))
```

```
##
## Call:
## lm(formula = selfevaluation ~ . - survey_id - goal_setting -
##     strategic_planning - task_strategies - elaboration - help_seeking,
##     data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.72890 -0.27697  0.07139  0.52633  0.87798
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          1.843e+00  1.442e+00   1.278   0.2220
## Num.Sessions         3.208e-03  9.557e-03   0.336   0.7421
## Avg.Session.Length  -8.073e-05  3.165e-04  -0.255   0.8024
## Side.Bar.Displays   -9.717e-03  9.946e-03  -0.977   0.3452
## Syllabus.visits      8.740e-04  4.775e-03   0.183   0.8574
## Prev.Unit           -1.733e-02  7.294e-02  -0.238   0.8156
## Num.Check.Grades     1.993e-02  1.010e-02   1.972   0.0687 .
## Med.Time.Btw.Sessions -5.173e-05  3.047e-05 -1.698   0.1117
## Next.Unit           -1.863e-02  1.999e-02  -0.932   0.3671
## Jump.Unit           -6.919e-03  1.769e-02  -0.391   0.7016
## Main.Page.Visits     2.958e-03  3.679e-03   0.804   0.4348
## N.Clicks            -2.076e-03  1.425e-03  -1.457   0.1672
## Forum.Visits        -1.129e-02  1.736e-02  -0.650   0.5260
## Num.Problems.Sub     2.594e-01  1.175e-01   2.208   0.0444 *
## N.Clicks.Session     1.113e-01  8.810e-02   1.263   0.2271
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.884 on 14 degrees of freedom
## Multiple R-squared:  0.431,  Adjusted R-squared:  -0.1381
## F-statistic: 0.7574 on 14 and 14 DF,  p-value: 0.6949
```

```r
# task strategies
summary(lm(task_strategies ~ Med.Time.Btw.Sessions + Next.Unit, dat))
```

```
## 
## Call:
## lm(formula = task_strategies ~ Med.Time.Btw.Sessions + Next.Unit,
##     data = dat)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.52828 -0.30053 -0.07186  0.45115  1.46560
## 
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           2.629e+00  2.284e-01  11.510 1.05e-11 ***
## Med.Time.Btw.Sessions -9.410e-06  6.896e-06  -1.365    0.184
## Next.Unit             -4.247e-03  9.682e-03  -0.439    0.665
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.6422 on 26 degrees of freedom
## Multiple R-squared:  0.08066,    Adjusted R-squared:  0.009947
## F-statistic: 1.141 on 2 and 26 DF,  p-value: 0.3351
```

```r
summary(lm(task_strategies ~ . - survey_id - goal_setting - strategic_planning - selfevaluation
           - elaboration - help_seeking, dat))
```

```
## 
## Call:
## lm(formula = task_strategies ~ . - survey_id - goal_setting -
##     strategic_planning - selfevaluation - elaboration - help_seeking,
##     data = dat)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.73521 -0.22363  0.00665  0.16913  0.95616
## 
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           2.594e+00  9.041e-01   2.869   0.0124 *
## Num.Sessions          -1.498e-03  5.990e-03  -0.250   0.8062
## Avg.Session.Length     8.823e-05  1.984e-04   0.445   0.6633
## Side.Bar.Displays      3.032e-03  6.234e-03   0.486   0.6342
## Syllabus.visits       -2.135e-03  2.993e-03  -0.713   0.4874
## Prev.Unit              3.660e-02  4.572e-02   0.801   0.4368
## Num.Check.Grades       8.952e-03  6.333e-03   1.414   0.1793
## Med.Time.Btw.Sessions -2.864e-05  1.910e-05  -1.500   0.1559
## Next.Unit             -1.588e-02  1.253e-02  -1.267   0.2258
## Jump.Unit             -3.150e-03  1.109e-02  -0.284   0.7805
## Main.Page.Visits      -2.509e-03  2.306e-03  -1.088   0.2950
## N.Clicks              -6.096e-05  8.932e-04  -0.068   0.9465
## Forum.Visits          -2.409e-03  1.088e-02  -0.221   0.8280
```

```
## Num.Problems.Sub      7.845e-02  7.362e-02    1.066    0.3046
## N.Clicks.Session      2.109e-02  5.522e-02    0.382    0.7082
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5541 on 14 degrees of freedom
## Multiple R-squared:  0.6315, Adjusted R-squared:  0.2629
## F-statistic: 1.714 on 14 and 14 DF,  p-value: 0.1626
```

```r
# elaboration
summary(lm(elaboration ~ Jump.Unit + Main.Page.Visits, dat))
```

```
##
## Call:
## lm(formula = elaboration ~ Jump.Unit + Main.Page.Visits, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.15416 -0.34241  0.03326  0.39284  1.27546
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.6638297  0.3922294   6.792  3.3e-07 ***
## Jump.Unit         0.0057954  0.0082876   0.699   0.4906
## Main.Page.Visits -0.0008420  0.0004493  -1.874   0.0722 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7413 on 26 degrees of freedom
## Multiple R-squared:  0.1532, Adjusted R-squared:  0.08811
## F-statistic: 2.353 on 2 and 26 DF,  p-value: 0.115
```

```r
summary(lm(elaboration ~ . - survey_id - goal_setting - strategic_planning - selfevaluation
           - task_strategies - help_seeking, dat))
```

```
##
## Call:
## lm(formula = elaboration ~ . - survey_id - goal_setting - strategic_planning -
##     selfevaluation - task_strategies - help_seeking, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.42807 -0.22693  0.04698  0.24463  1.16101
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        5.863e-01  1.259e+00   0.466    0.649
## Num.Sessions       1.142e-02  8.343e-03   1.368    0.193
## Avg.Session.Length 1.207e-04  2.763e-04   0.437    0.669
## Side.Bar.Displays -1.167e-02  8.682e-03  -1.344    0.200
## Syllabus.visits    9.100e-05  4.168e-03   0.022    0.983
## Prev.Unit          4.292e-02  6.367e-02   0.674    0.511
## Num.Check.Grades   1.429e-02  8.820e-03   1.621    0.127
```

```
## Med.Time.Btw.Sessions -1.525e-05  2.660e-05  -0.573    0.576
## Next.Unit                5.921e-03  1.745e-02   0.339    0.739
## Jump.Unit                5.318e-03  1.544e-02   0.344    0.736
## Main.Page.Visits         3.520e-03  3.211e-03   1.096    0.292
## N.Clicks                -1.828e-03  1.244e-03  -1.469    0.164
## Forum.Visits             3.145e-04  1.515e-02   0.021    0.984
## Num.Problems.Sub         1.434e-01  1.025e-01   1.399    0.184
## N.Clicks.Session         7.454e-02  7.691e-02   0.969    0.349
##
## Residual standard error: 0.7717 on 14 degrees of freedom
## Multiple R-squared:  0.5059, Adjusted R-squared:  0.01184
## F-statistic: 1.024 on 14 and 14 DF,  p-value: 0.4826
```

```r
# help seeking
summary(lm(help_seeking ~ N.Clicks + Forum.Visits, dat))
```

```
##
## Call:
## lm(formula = help_seeking ~ N.Clicks + Forum.Visits, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9299 -0.2822  0.1042  0.4979  1.4074
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.2314298  0.6437808   3.466  0.00185 **
## N.Clicks     0.0001655  0.0003047   0.543  0.59177
## Forum.Visits 0.0092455  0.0144101   0.642  0.52675
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.082 on 26 degrees of freedom
## Multiple R-squared:  0.05118,    Adjusted R-squared:  -0.0218
## F-statistic: 0.7013 on 2 and 26 DF,  p-value: 0.5051
```

```r
summary(lm(help_seeking ~ . - survey_id - goal_setting - strategic_planning - selfevaluation
           - task_strategies - elaboration, dat))
```

```
##
## Call:
## lm(formula = help_seeking ~ . - survey_id - goal_setting - strategic_planning -
##     selfevaluation - task_strategies - elaboration, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2514 -0.2385  0.1096  0.3572  1.1868
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.034e+00  1.571e+00   1.294   0.2166
## Num.Sessions      3.351e-03  1.041e-02   0.322   0.7523
## Avg.Session.Length -6.272e-05  3.448e-04  -0.182   0.8583
```

```
## Side.Bar.Displays      -1.475e-02  1.083e-02  -1.362   0.1948
## Syllabus.visits         8.437e-03  5.202e-03   1.622   0.1271
## Prev.Unit               1.419e-01  7.945e-02   1.786   0.0958 .
## Num.Check.Grades        2.640e-02  1.101e-02   2.399   0.0310 *
## Med.Time.Btw.Sessions  -2.601e-05  3.320e-05  -0.783   0.4464
## Next.Unit              -3.635e-02  2.178e-02  -1.669   0.1173
## Jump.Unit               4.146e-02  1.927e-02   2.152   0.0494 *
## Main.Page.Visits        1.015e-02  4.008e-03   2.532   0.0239 *
## N.Clicks               -3.865e-03  1.552e-03  -2.490   0.0260 *
## Forum.Visits           -1.457e-02  1.891e-02  -0.770   0.4539
## Num.Problems.Sub        1.963e-01  1.280e-01   1.534   0.1472
## N.Clicks.Session        6.478e-02  9.598e-02   0.675   0.5107
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.963 on 14 degrees of freedom
## Multiple R-squared:  0.5954, Adjusted R-squared:  0.1907
## F-statistic: 1.471 on 14 and 14 DF,  p-value: 0.2396
```

```
#################################################
#################################################
```

Question 6: Describe what you found. Which SRL strategies, if any, were you able to predict with which features? Were there any surprises? Which self-reported SRL strategy were you able to predict best with all features (look at the `Multiple R-squared`)?

Using a nominal significance level of alpha = 0.05, none of the tests failed to reject the null hypothesis that there was no relationship between the variables. With the full model, there were some significant relationships between variables and goal setting and help-seeking, though I didn't check for any colinearity among the variables or anything. I was a little surprised that none of my variables appeared important, though with a sample size of 29 I guess this wasn't too surprising. I was best able to predict goal setting with an R^2 of aboout 0.78.

Self-reflection (ungraded)

Briefly summarize your experience on this homework. What was easy, what was hard, what did you learn?

I felt like this was the longest homework I've worked on (~4-5 hours). The most difficult part was coming up with good features. The implementation and modeling was straightforward.

Submit Homework

This is the end of the homework. Please **Knit a PDF report** that shows both the R code and R output and upload it on the EdX platform. Alternatively, you can Knit it as a "doc", open it in Word, and save that as a PDF.

Important: Be sure that all your code is visible. If the line is too long, it gets cut off. If that happens, organize your code on several lines.