Report: Evaluating Writing Center Consultations

Kayla Cummings, Grant Belsterling, Vanessa Machuca

December 7, 2017

Overview

At the Pomona College Writing Center, peer Writing Partners consult with visiting students to understand, revise, edit, polish, and brainstorm for writing assignments and oral presentations.

The Writing Center is at the Smith Campus Center, pictured here.

Each Writing Partner has certain areas of expertise that reflect their unique strengths. After each consultation, students have the option to fill out an exit survey that summarizes their visit. These exit surveys serve as a mechanism for internal self-assessment. This work contains data analysis based on exit survey data and Writing Partner expertise data. Our goal was to provide Writing Center affiliates with actionable, non-obvious recommendations in order for affiliates to continue to improve their services.

In order to perform the data analysis, we needed to wrangle the data into a format that would be recognizable by statistical algorithms. A large portion of entries had missing values, so we employed several imputation methods in order to save as many observations and variables as we could. This was the most time-consuming part of our project.

For our first piece of analysis, we used clustering with the hopes of obtaining information about characteristic consultations. Clustering algorithms partition data into groups of similar observations. We developed simple dissimilarity measures that were adapted to suit our specific data. Although we didn't glean any concrete solutions or recommendations from this analysis, we will still present what we learned in the process.

Finally, we used machine learning to train models that classified consultations as successful or unsuccessful. Because all of our variables were categorial or discrete, we used random forests, which are able to interpret any kind of variable. Simply put, random forests aggregate predictions of a large number of trees, which are greedy classification models that explicitly represent the decision-making process. Our goal

was not to predict the success of future consultations, but rather to reveal which predictors have the largest impact on our models' process of deciding whether a consultation is successful. We have based our recommendations on these results.

Cleaning Data

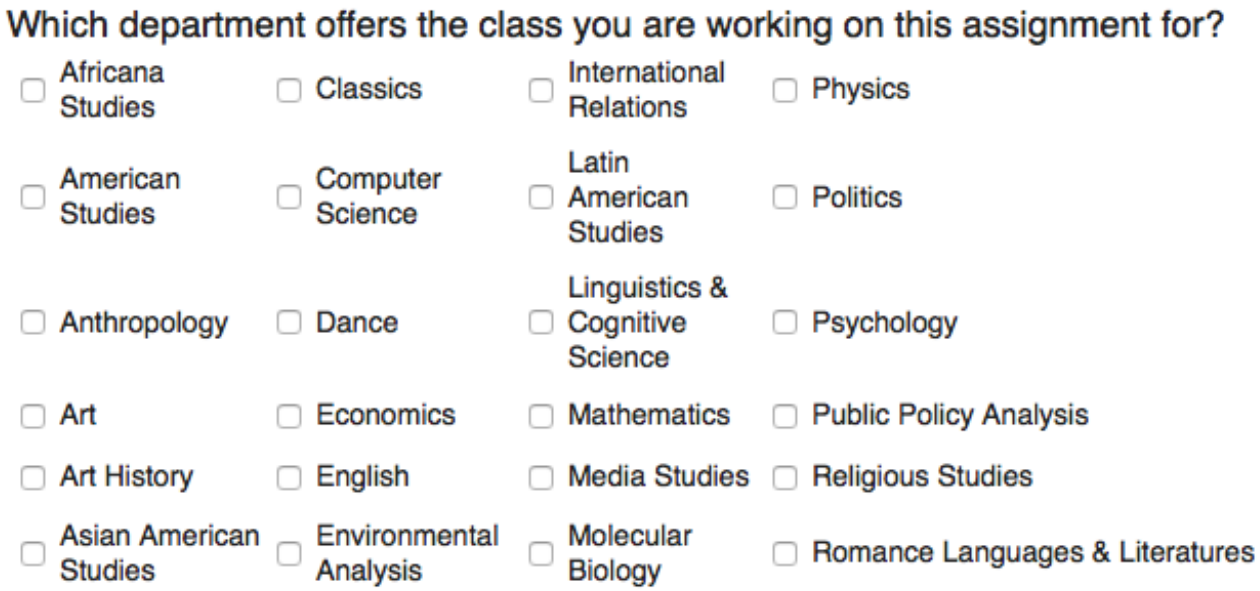
Before we can perform any analysis, we have to wrangle the data into a workable format. The Director of the Writing Center, Pam Bromley, provided us with 5 years' worth of exit survey results on Qualtrics from the school years from 2012-2013 through 2016-2017. We also received five corresponding Excel spreadsheets with Pam's manual entries describing Writing Partners' areas of expertise. This data is not publicly available, and we were given access to it for the sole purposes of this project. One challenge was that data collection was standardized at the Writing Center in the 2013-2014 school year, so the 2012-2013 data is formatted differently from the rest. We downloaded the exit survey data directly from Qualtrics and the Writing Partner expertise data directly from Pam's emails; they are included in the file path ./Report/orig_data. All data wrangling can be reproduced by running the document ./Report/reproducibleDataWrangling.R. In this report we include selected code to provide illustrative examples, but don't review everything.

Exit Surveys

After joining the five survey spreadsheets, we combined similar columns that contained answers to similar questions that were phrased differently in surveys from separate school years. For example, "How did you originally learn about the Writing Center? Check all that apply." and "How did you originally learn about the Writing Center? Check all that apply…" are the same question with different phrasing from different years, so they were united into one variable.

We also converted variables whose values were ratings (i.e. Strongly Disagree to Strongly Agree) to numerical Likert scales, so that future models would not interpret ratings as equidistant categories. Our conversions assumed an unavailable "Neutral" option with value 3. We applied the same logic to student years (Freshman through Senior became 1 through 4).

Finally, certain questions asked students to check all options that applied.



Qualtrics converted students' answers to these questions into a single variable, with each checked value separated by a comma. First, we separated the multiple answers into different columns, and then we one-hot encoded them.

To see what this process looks like, let's look at how_learned_about_1, how_learned_about_2, and how_learned_about_3. These are three variables that indicate the ways that students learned about the

Writing Center; there are three columns because students selected no more than 3 options.

```
##  [1] Other (please specify)
##  [2] From an instructor
##  [3] From another student
##  [4] From a resource fair or other student event
##  [5] From a class visit by a Writing Center representative
##  [6] From the website or online schedule
##  [7] <NA>
##  [8] Other (please specify)White/Caucasian
##  [9] From our website or online schedule
## [10] From a brochure
## 9 Levels: From a brochure ...
```

Some values were redundant, like the two separate website/online schedule options. We also see a mistake, “Other (please specify)White/Caucasian”, which we grouped in with “Other (please specify)”. To transform these three overlapping variables into seven binary variables that told us whether students learned about the Writing Center in a brochure, from a representative, etc., we did the following:

1. We grouped together redundant entries, e.g. “From our website or online schedule” and “From the website and online schedule”.
2. We one-hot encoded each variable how_learned_about_i for i $\in$ {1, 2, 3} and obtained three separate 0-1 matrices.
3. We combined these three matrices into one 0-1 matrix by adding entries that correspond to the same answer. We knew that this output matrix will be a 0-1 matrix because students don’t have the option to check the same box twice. In other words, if a student learned about the Writing Center from a brochure, then there is a 1 in exactly one of the three columns; otherwise, all three columns contain zeroes.
4. We appended these seven new binary variables to the original data frame and removed the three old categorical variables. We chose not to include “NA” as a variable; therefore, students who did not select any options had all 0-entries for these seven binary variables.

Writing Partner Expertise

We also had to parse Writing Partners’ areas of expertise into a format that our models would recognize. For school years from 2013-2014 to 2016-2017, we had information on whether Partners were minoring in anything, and whether they had expertise on lab reports, foreign languages, creative writing, and other specialties. For 2012-2013, all of this information was contained in one variable. We also had information on all of the Partners’ majors. We parsed the 2012-2013 Partners’ specialties to conform with the more specific variables, then combined these five spreadsheets into one data frame.

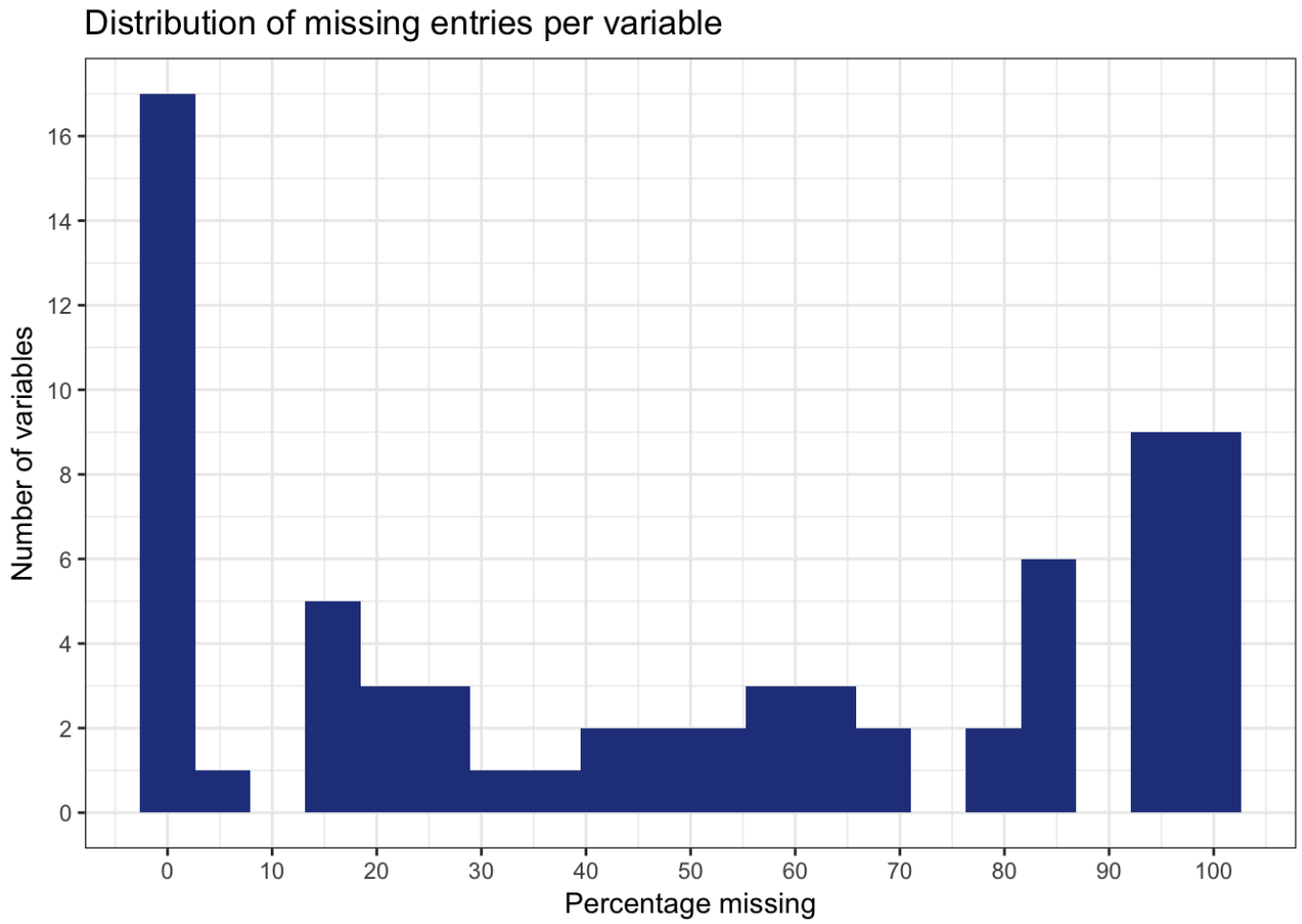
Merging

Finally, we merged the two datasets by distinct schoolyear/Partner pairs. We left-joined the Writing Partner data to the exit survey data using the school year and the Partner’s first name. Although some students worked at the Writing Center over multiple school years, we treated them as distinct for each school year in order to account for how their areas of expertise evolved over their time at Pomona. The final outcome was one data frame, with each observation corresponding to a consultation, and each variable either corresponding to exit survey question selections or information about the Writing Partner who helped them during their consultation.

Imputation

Before we could use our clean data to build statistical models, we had to figure out what to do with all of the

missing entries.



There are many ways that we can deal with NA entries. In some cases, we reinterpreted what “NA” meant and transformed entries into a separate variable value. For others, we combined columns whose similarity we missed earlier. Finally, we used more sophisticated methods to impute missing values of Likert scale variables. We will give an example of each.

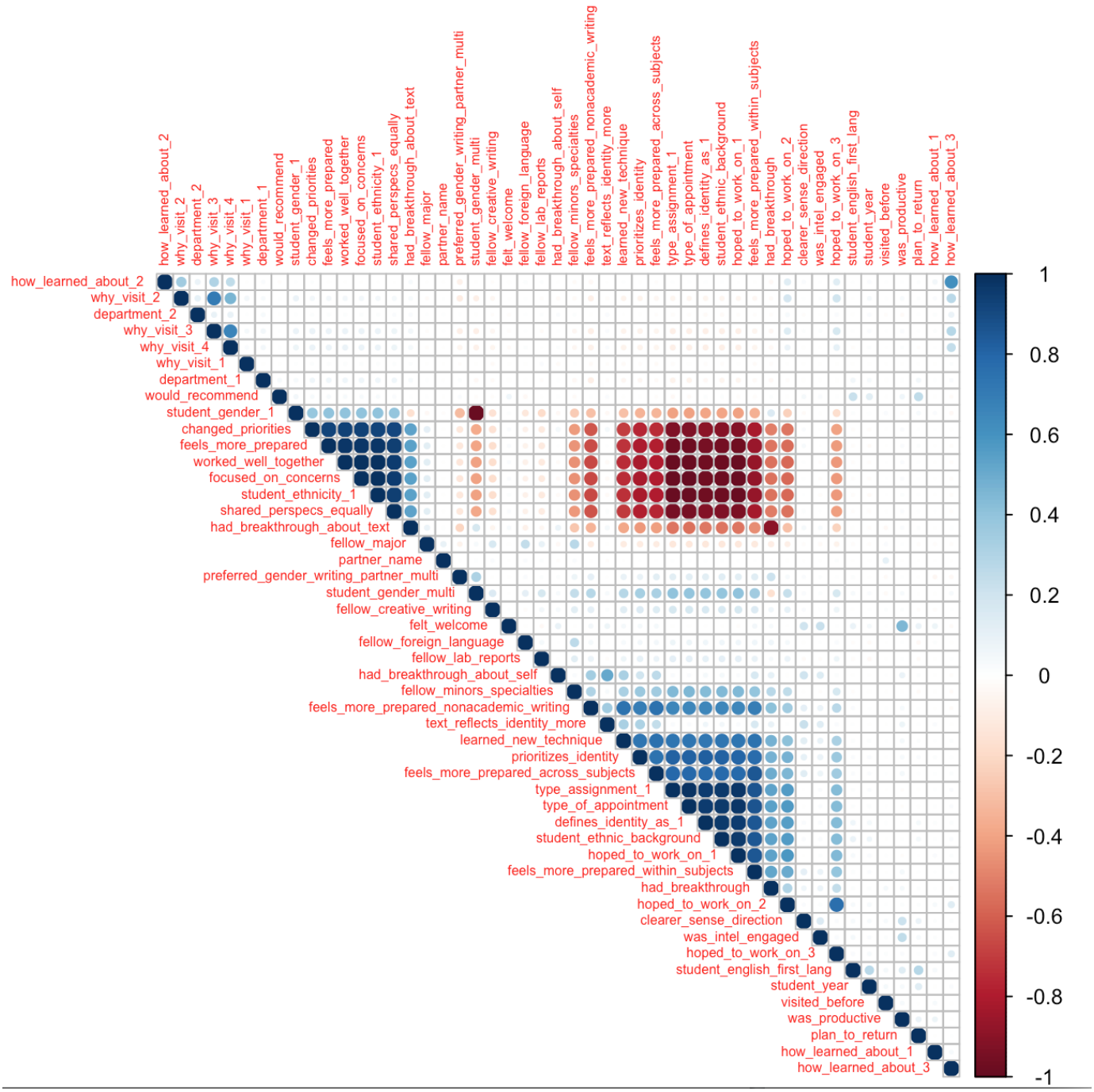
Reinterpreting “NA”

By reinterpreting “NA” as a new variable value, we can reduce the number of missing entries in a particular variable without having to get really technical. Writing Partner areas of expertise present one example. For these variables, we can interpret “NA” to mean that the Writing Partner has no expertise in that area. Then these missing values are recast to provide more meaningful information to future models.

Combining columns

Some columns expressed similar sentiments with slightly different wordings. We found that this was because some questions changed slightly or were reworded between years.

We wanted to discover correlations between NAs. For example, if one column was always NA where another column was always filled in, then it was likely the case that one question was swapped for the other at some points during the years we examined. Most of the times, these questions expressed the same idea and so we combined columns.



The correlation between NA's across columns.

For example, in the above plot, we notice clusters of blue variables, where the correlation was high. For these pairs, either both were filled out or both were NA.

More relevant to us is the red cluster. These questions were never asked in the same year. We see some pairs that obviously express the same sentiment, such as feels_prepared and feels_more_prepared_within_subjects. We combined those columns. Others in the red cluster, however, were simply questions that were added to the survey and express new ideas. We consider imputing them in the next section.

MICE

We didn't want to throw out columns with sparse coverage if they had high potential to impact the response variable. In order to do this, we wanted to use imputation methods that would ensure trustworthy values.

We used the MICE package and predictive mean matching to impute these values. We only imputed values that had more than 90% coverage in order to ensure that we had enough existing observations to accurately predict the small number of missing values.

Clustering (Attempt)

Partitioning Around Medoids (PAM) is a clustering algorithm that first searches for k representative observations called medoids. We can interpret these medoids to be characteristic Writing Center consultations, from which we can hopefully learn something about common experiences at the Writing Center. After finding the medoids, the next step is to partition the data into k clusters by assigning each observation to the nearest medoid. This notion of "distance" is based on a dissimilarity measure that we supply to the PAM algorithm. We have to build our own dissimilarity measures because regular notions of distance (such as the Euclidean norm) do not translate well to categorical and binary variables. The algorithm's objective function minimizes the sum of dissimilarities within each cluster.

Dissimilarity Measures

There are a few ways that we can extend beyond Euclidean distance. The Hamming distance between two strings is the number of indices for which the characters differ. For example, "great" and "grate" have an unnormalized Hamming distance of 3. Our clustering functions require a dissimilarity matrix A with entries A[i, j] = diss(i, j) = diss(j, i) = A[j, i].

```
hamming <- function(df)
{
  num_obs <- length(df[,1])
  A <- matrix(rep(0,num_obs^2),nrow=num_obs)
  for (i in 1:num_obs){                          # each row
    for (j in i:num_obs) {                       # each column
      A[i,j] <- A[j,i] <- sum(df[i,]!=df[j,])     # Hamming distance
    }
  }
  return(A)
}
```

This measure makes the most sense for binary variables, but we can do a little bit better. For our discrete-scale variables, we can find the absolute value of the difference between the two variable entries, and then normalize by the difference between the maximum and minimum possible values. This normalization constant ensures that the impact of this variable is equivalent to the impact of binary variables, because the value remains between 0 and 1.

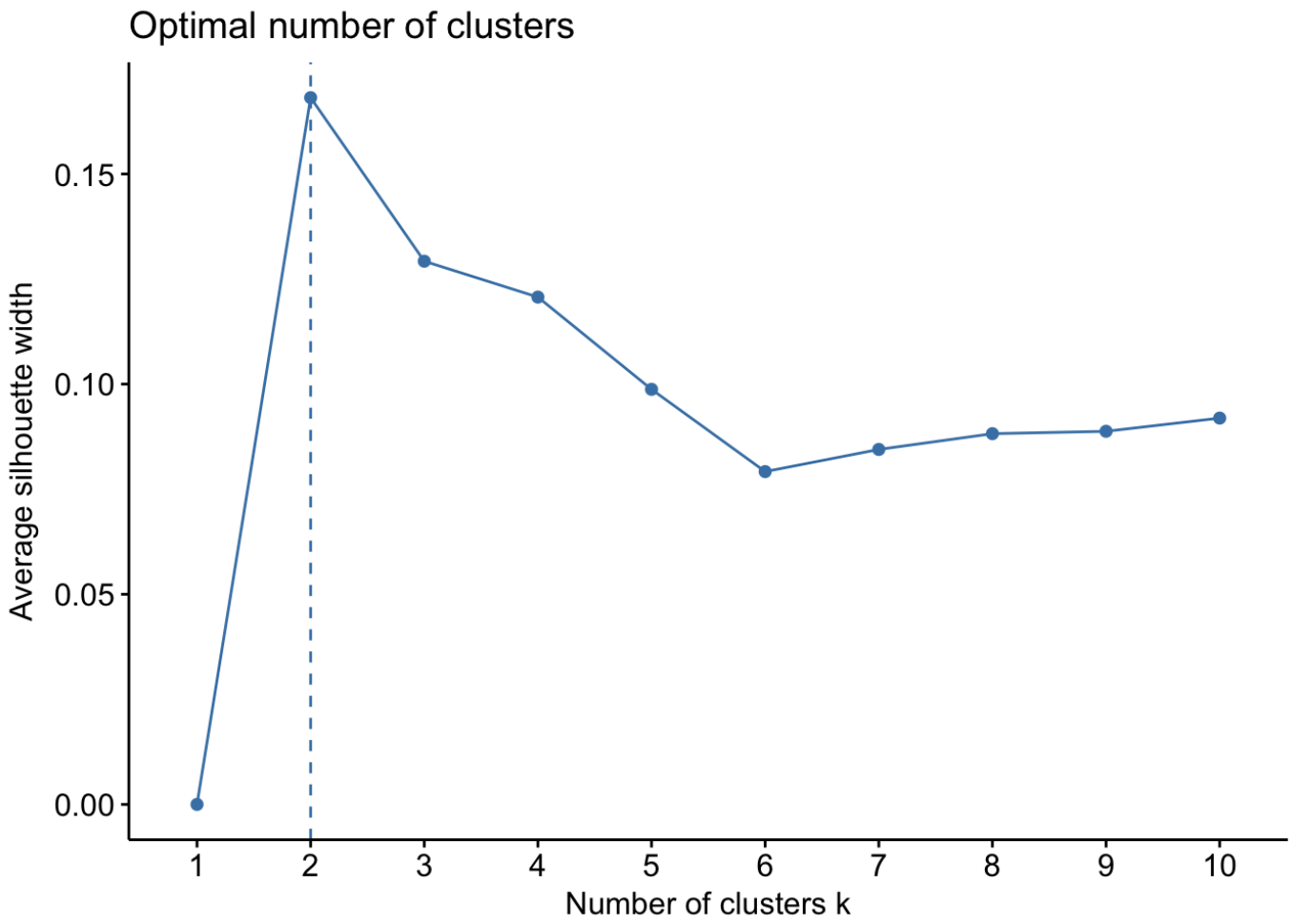
```
mod_hamming <- function(df)
{
  discrete_cols <- max_vals <- min_vals <- c()
  for (col in 1:length(df)) {
    if (length(unique(df[,col]))>2) {                  # discrete vars with max &
min values
      discrete_cols <- c(discrete_cols,col)
      max_vals <- c(max_vals,max(df[,col]))
      min_vals <- c(min_vals,min(df[,col]))
    }
  }

  num_obs <- length(df[,1])
  A <- matrix(rep(0,num_obs^2),nrow=num_obs)                          # dissimilarity
matrix
  for (i in 1:num_obs){                                                 # each row
    for (j in i:num_obs) {                                              # each colum
n
      div <- ifelse(max_vals-min_vals==0,1,max_vals-min_vals)           # don't divid
e by 0
      A[i,j] <- A[j,i] <-
        sum(df[i,-discrete_cols] != df[j,-discrete_cols])               # binary (ham
ming)
        sum(abs(df[i,discrete_cols] - df[j,discrete_cols])/div)         # discrete
    }
  }
  return(A)
}
```

Silhouette Plots to Choose the Number of Clusters

While there are many ways to estimate the optimal number of clusters, we used average silhouette width, which provides a way to visualize how well each observation fits in with its respective cluster. After we have clustered our observations, consider a single observation i. If we think about this observation's average dissimilarity a(i) to other observations within its own cluster, this value should be very small. This value will ideally grow if we measure the observation's average dissimilarity to all observations in a separate cluster. Using this notion of distance to choose the "next-closest cluster" to i with average dissimilarity b(i), the silhouette width is s(i) = b(i) − a(i)max This definition implies that $s(i) \approx 1$ means that the next-closest cluster is pretty far away, while $s(i) \approx 0$ is less great. By computing the average silhouette width, $\bar{s} = \frac{1}{n}\sum_{i = 1}^n s(i)$, over all n observations, we obtain a measure of how well our observations have clustered. If we build multiple partitions of our data using different numbers of clusters, we choose the number of clusters that yields the highest silhouette width.

Our data was a little too large for the algorithms to work quickly (they are $O(n^2)$ due to the matrix computations), so as we rebalanced the data, we took that opportunity to downsize. In addition, the silhouette function we used unfortunately defaults to a Euclidean norm and it was unclear how we could pass our own dissimilarity metric to it. However, the results from this silhouette plot will serve as a good proxy.



Results

We can treat these medoids as “characteristic observations.” Because there are so many variables, we elect to only look at the variable names for which the two consultations differed.

```
##  [1] "visited_before.Yes"
##  [2] "text_reflects_identity_more"
##  [3] "student_year"
##  [4] "gender.Female"
##  [5] "gender.Male"
##  [6] "type_of_appointment.A.drop.in.appointment."
##  [7] "type_of_appointment.A.regular.appointment.that.I.made.online."
##  [8] "type_assignment_1.A.course.paper"
##  [9] "type_assignment_1.Other..please.specify."
## [10] "learned_new_technique"
## [11] "prioritizes_identity"
## [12] "defines_identity_as_1.My.voice.or.style"
## [13] "defines_identity_as_1.Other..please.specify."
## [14] "fellow_creative_writing.None"
## [15] "fellow_creative_writing.Yes"
## [16] "visit_challenging_assn"
## [17] "visit_right_track"
## [18] "visit_routine"
## [19] "learnedfrom_resource_fair"
## [20] "hoped_style"
## [21] "hoped_workingon_oral"
```

By looking at these variables, we can only make simple comparisons; for example, one was male while the other was female, and one had visited the Writing Center before while the other hadn’t. While these

differences between characteristic consultations are vaguely interesting, there isn't much to learn by just looking at two students' experiences.

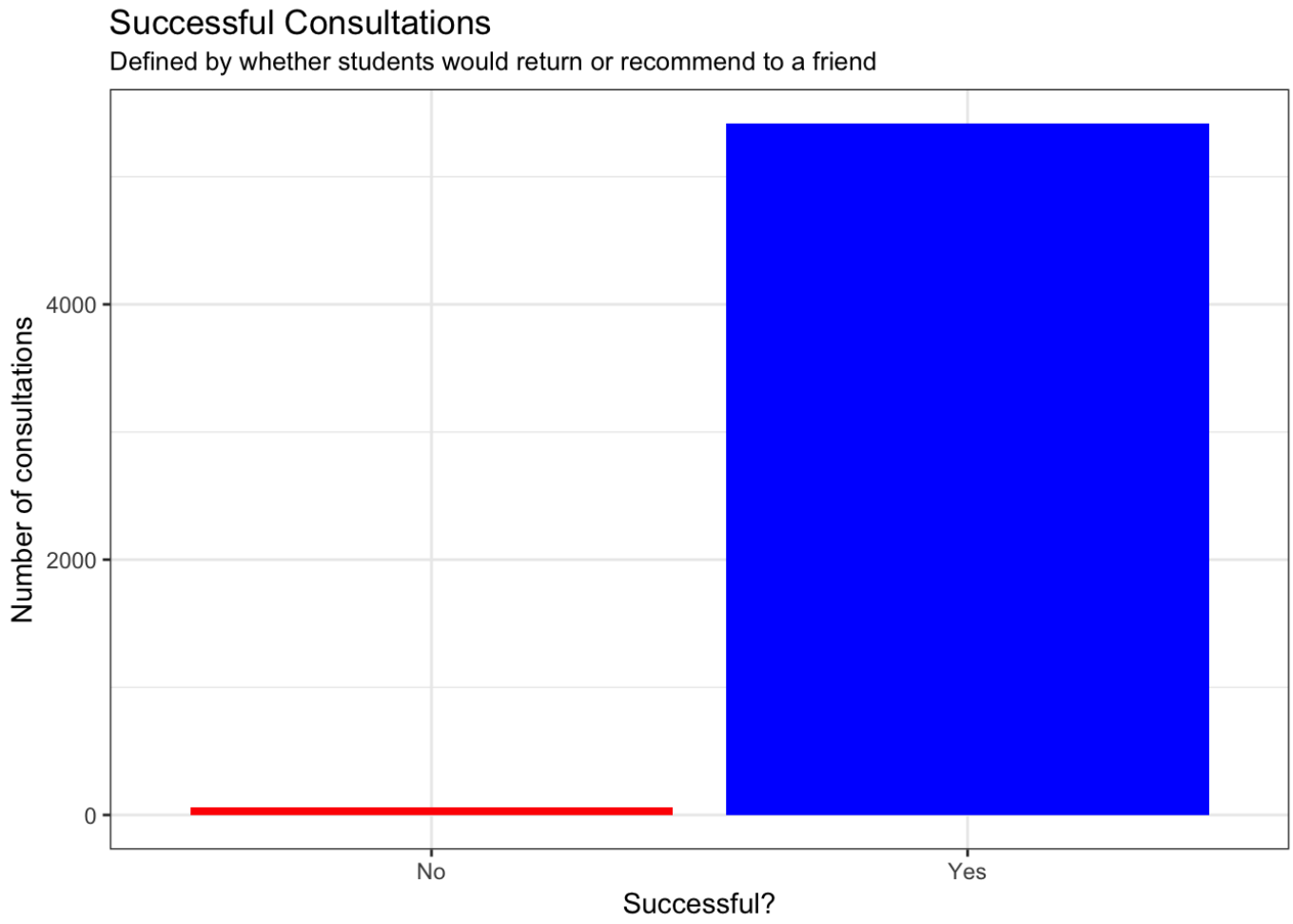
However, since we have already developed two dissimilarity measures, we take a look at the two medoids of the data when using the modified hamming dissimilarity measure. Because the process of computing the dissimilarity matrix was so slow, we sampled a rebalanced dataset that was half the size of the previous (2000 down to 1000), with a 50/50 ratio of successful/unsuccesful consultations. Again, we look at where the two consultations differed.

```
##  [1] "felt_welcome"
##  [2] "was_productive"
##  [3] "was_intel_engaged"
##  [4] "clearer_sense_direction"
##  [5] "text_reflects_identity_more"
##  [6] "gender.Male"
##  [7] "gender.Prefer.not.to.specify"
##  [8] "type_of_appointment.A.drop.in.appointment."
##  [9] "type_of_appointment.A.regular.appointment.that.I.made.online."
## [10] "type_assignment_1.A.course.paper"
## [11] "type_assignment_1.Other..please.specify."
## [12] "learned_new_technique"
## [13] "prioritizes_identity"
## [14] "defines_identity_as_1.My.voice.or.style"
## [15] "defines_identity_as_1.Other..please.specify."
## [16] "fellow_creative_writing.None"
## [17] "fellow_creative_writing.Yes"
## [18] "visit_challenging_assn"
## [19] "learnedfrom_classvisit"
## [20] "learnedfrom_instructor"
## [21] "learnedfrom_other_student"
## [22] "hoped_style"
## [23] "eth_asian"
## [24] "eth_pref_unspecified"
```

As for these two typical students, we are able to glean things such as differing ethnicities and levels of intellectual engagement. As we have discussed, it is interesting to make comparisons between these two "characteristic" consultations, but there isn't any basis for making recommendations to the Writing Center, which is our ultimate goal. After looking at these two model outputs, we decided to build classification models instead.

Imbalanced Response

When we began to build the classification models, they merely predicted that every writing center consultation was successful. This was because the majority of consultations *were* successful. Below is the distribution of positive and negative responses.



To remedy this problem, we explored several solutions. The first was to resample the negative class a few thousand times until we had equal numbers of both classes. However, this upsampling would greatly (and artifically) reduce the variance in the negative class, since you have thousands of data points derived from only a few actual negative samples.

Next we explored downsampling the positive class to reduce its number. If we did this to equalize the size of the classes using the original number of positive samples, we would only be left with a few hundred observations. This is not ideal to train a model, especially when we have thousands of (positive) observations we would be throwing out that would otherwise be perfectly good to train a model on.

We settled on a hybrid solution. We downsampled the majority positive class, included the entire minority class, and upsampled the minority class as well. The benefits of including the minority class *and* upsampling it rather than merely upsample is that on average, a boostrap includes only 2/3 of observations. We thought that each negative observation was pretty important and valuable, so we wanted to include each of them.

This solution avoided the extreme problems of drastically misrepresenting variance in the population of one of the classes. However, it introduced the problem of having slightly misrepresented variances in both of the classes. We found this tolerable and chose to proceed with this method.

Random Forests

Initial Models

However, the above thought process is an edited, streamlined version of events that led us to our final model. The actual process involved a lot of trial and error.

Below we look at our first attempt at a preliminary model using our data with our original, limited definition of sucess (that a person recommends to a friend or that they would return to the writing center). This definition of success was severely imbalanced, as we discussed above.

```
mod_data <- imp_data %>%
  mutate(successful_consultation = ifelse(plan_to_return=="No" | would_recommend=="
No","No","Yes")) %>%
  select(-c(plan_to_return,would_recommend))

mod_data$successful_consultation <- as.factor(mod_data$successful_consultation)

training <- createDataPartition(y=mod_data$successful_consultation, p=.9, list=FALS
E)
training_data <- mod_data[training,]
test_data <- mod_data[-training,]
```

```
rf_initial <- caret::train(successful_consultation ~., data=training_data, method="
rf",
                     trControl=trainControl(method="oob"),
                     ntree=100,tuneGrid=data.frame(mtry=20)) #mtry and ntree reduc
ed in order to speed runtime for the report
```

```
confusionMatrix(predict(rf_initial,newdata=test_data),test_data$successful_consulta
tion)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  No Yes
##        No    1   0
##        Yes   5 541
##
##                Accuracy : 0.9909
##                  95% CI : (0.9788, 0.997)
##     No Information Rate : 0.989
##     P-Value [Acc > NIR] : 0.44479
##
##                   Kappa : 0.2835
##  Mcnemar's Test P-Value : 0.07364
##
##             Sensitivity : 0.166667
##             Specificity : 1.000000
##          Pos Pred Value : 1.000000
##          Neg Pred Value : 0.990842
##              Prevalence : 0.010969
##          Detection Rate : 0.001828
##    Detection Prevalence : 0.001828
##       Balanced Accuracy : 0.583333
##
##        'Positive' Class : No
##
```

This is a problem! Although our test accuracy is 99%, when we group by response, we have 0% accuracy for the minority class and 100% accuracy for the majority class (because our prediction is always “Yes”).

Next we tried upsampling in isolation and downsampling in isolation. Neither method returned satisfactory results, so they are omitted in this report.

Finally we tried combining the two methods and upsampling the unsuccessful consultations while downsampling the successful consultations (the majority).

```
survey_sample <- sample_n(mod_data, 2000)

negatives <- mod_data %>% filter(successful_consultation=='No')
resampledNeg <- sample_n(negatives, 200, replace=TRUE)

survey_sample <- bind_rows(survey_sample, resampledNeg)
```

```
rf_combined <- caret::train(successful_consultation ~., data=survey_sample, method=
"rf",
                     trControl=trainControl(method="oob"),
                     ntree=100,tuneGrid=data.frame(mtry=(30)))
```

This model doesn't look too terrible. We see a more balanced ratio of predictions between successful and unsuccessful, which was our goal. The model is actually capable of predicting unsuccessful responses now!

```
rf_combined$finalModel
```

```
##
## Call:
##  randomForest(x = x, y = y, ntree = 100, mtry = param$mtry)
##                Type of random forest: classification
##                      Number of trees: 100
## No. of variables tried at each split: 30
##
##         OOB estimate of  error rate: 0.23%
## Confusion matrix:
##      No  Yes class.error
## No  217    3 0.013636364
## Yes   2 1978 0.001010101
```

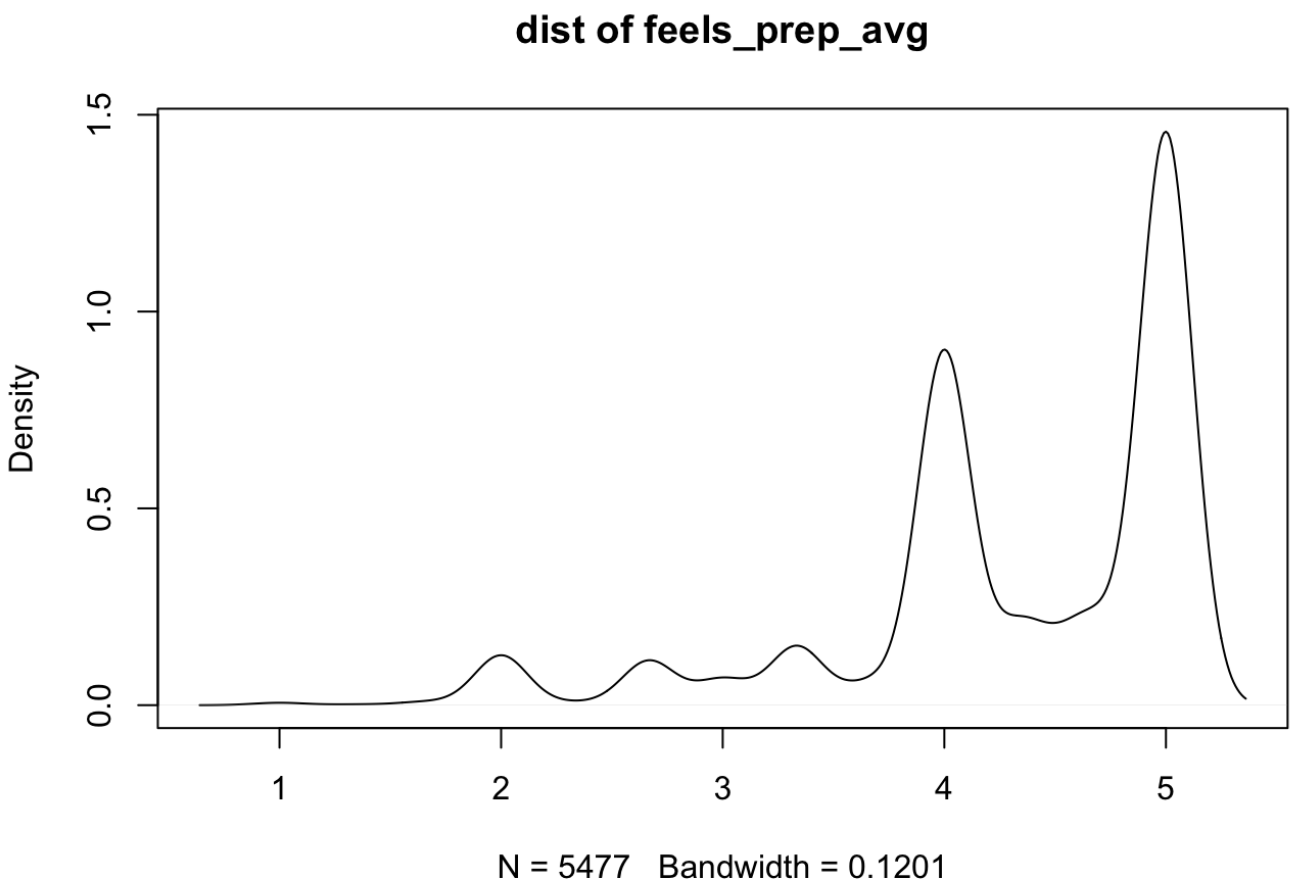
```
varImp(rf_combined)
```

```
## rf variable importance
##
##   only 20 most important variables shown (out of 323)
##
##
Overall
## feels_prep_avg
100.00
## departmentMathematics
97.01
## breakthrough
75.77
## student_year
60.18
## was_productive
49.62
## clearer_sense_direction
44.91
## prioritizes_identity
36.65
## was_intel_engaged
32.54
## type_of_appointmentAn appointment for a course with an attached Writing Fellow (
either for ID1 or for another course).   31.18
## visit_better_grade
29.69
## learned_new_technique
28.56
## text_reflects_identity_more
24.95
## visit_instructor_rec
21.51
## preferred_gender_writing_partner_multiOther
19.82
## eth_pref_unspecified
19.50
## visit_challenging_assn
19.46
## fellow_majorChemistry and Math
18.58
## genderMale
18.08
## partner_nameVanessa
17.84
## felt_welcome
17.80
```

We see that the most important variables for distinguishing between successful consultations and nonsuccessful ones are whether the student had a breakthrough, their rating on whether they felt more prepared, and their year (first, sophomore, junior, etc). The first two are obvious: we expect people that have successful consultations to feel more prepared to finish their paper. These variables seem like alternate definitions of a ‘successful consultation.’ Let’s try removing them from the predictor space and putting them as response variables instead. This way, we get to unmask more interesting predictors that lead to successful consultation; we also expand the size of our minority class by also including consultations with low values for feeling more prepared and having a breakthrough.

But what constitutes a ‘low value’?

```
plot(density(mod_data$feels_prep_avg), main='dist of feels_prep_avg')
```



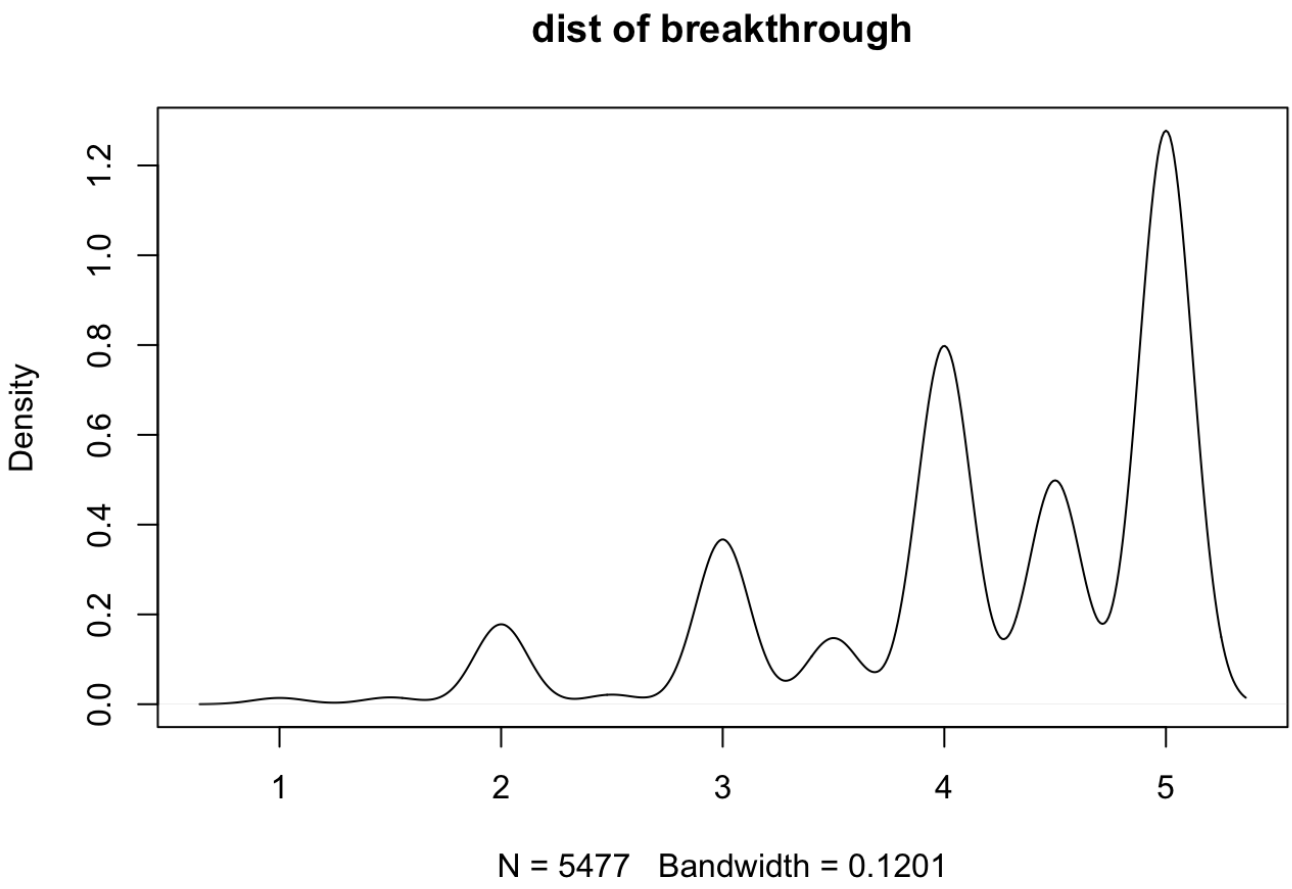

```
table(mod_data$feels_prep_avg)
```

```
##
##                 1 1.33333333333333               1.5 1.66666666666667
##                10                  3                  1                 14
##                 2 2.33333333333333 2.66666666666667                  3
##               209                 11                186               107
## 3.33333333333333               3.5 3.66666666666667                  4
##               244                  4                 87              1484
## 4.33333333333333               4.5 4.66666666666667                  5
##               293                113                311              2400
```

We notice a very long tail of observations where the student did not feel more prepared. It’s reasonable to consider these ‘unsuccessful’ since they so strong deviate from the norm of rating preparation as a 4 or a 5. Let’s set all values with feels_prep_avg < 3.5 as unsuccessful.

We similarly examine the distribution of having a breakthrough.

```
plot(density(mod_data$breakthrough), main='dist of breakthrough')
```




```
table(mod_data$breakthrough)
```

```
##
##    1  1.5    2  2.5    3  3.5    4  4.5    5
##   23   25  294   35  606  243 1318  823 2110
```

Let's set breakthrough <= 2.5 as unsuccessful.

Let's execute that by creating a new variable with a looser definition of 'success.'

```
mod_data <- mod_data %>% mutate(was_successful_loose = case_when(
  successful_consultation == 'No' ~ 'No',
  feels_prep_avg <= 3.5 ~ 'No',
  breakthrough <= 2.5 ~ 'No',
  TRUE ~ 'Yes'
))

looseData <- mod_data %>% select(-successful_consultation, -feels_prep_avg, -breakt
hrough)
```

```
table(mod_data$was_successful_loose)
```

```
##
##   No  Yes
##  987 4490
```

Our classes are naturally more balanced now, without even having to downsample or upsample!

Modified Model Results

However, we can achieve an even better balance between the classes by applying the hybrid downsampling/upsampling method we arrived at in the previous section. We downsample the positive class, take the entirety of the negative class (since bootstrapping will only get on avg 2/3 of the data and we want to include the entire set at least once), and also include a bootstrap of the negative samples to further increase their number.

```
negatives <- looseData %>% filter(was_successful_loose=='No')
positives <- looseData %>% filter(was_successful_loose=='Yes')
resampledNeg <- sample_n(negatives, 500, replace=TRUE)
sampledPos <- sample_n(positives, 2000, replace=FALSE)

survey_sample <- bind_rows(sampledPos, resampledNeg, negatives)
```

```
rf_loose <- caret::train(was_successful_loose ~., data=survey_sample, method="rf",
                     trControl=trainControl(method="oob"),
                     ntree=100,tuneGrid=data.frame(mtry=15))
```

```
rf_loose$finalModel
```

```
##
## Call:
##  randomForest(x = x, y = y, ntree = 100, mtry = param$mtry)
##                Type of random forest: classification
##                      Number of trees: 100
## No. of variables tried at each split: 15
##
##         OOB estimate of  error rate: 12.07%
## Confusion matrix:
##       No  Yes class.error
## No  1296  191   0.1284465
## Yes  230 1770   0.1150000
```

```
varImp(rf_loose)
```

```
## rf variable importance
##
##   only 20 most important variables shown (out of 321)
##
##                                                                 Overall
## learned_new_technique                                           100.000
## text_reflects_identity_more                                      88.723
## clearer_sense_direction                                          40.202
## was_intel_engaged                                                37.028
## was_productive                                                   31.458
## prioritizes_identity                                             26.727
## felt_welcome                                                     11.751
## student_year                                                     11.544
## eth_white                                                         9.388
## defines_identity_as_1My voice or style                            8.822
## visit_right_track                                                 7.889
## genderMale                                                        7.339
## visit_general                                                     7.291
## learnedfrom_classvisit                                            7.225
## learnedfrom_instructor                                            7.068
## eth_asian                                                         6.916
## student_english_first_langNo, English is not my first language.   6.812
## visit_challenging_assn                                            6.787
## defines_identity_as_1Other (please specify)                       6.712
## visited_beforeYes                                                 6.684
```

We see that the forest isn't classifying everything as a yes, which is good. The confusion matrix is relatively evenly split between the Yes/No confusion and the No/Yes confusion.

When we examine variable importance, we see that learned_new_technique, text_reflects_identity_more, and was_intel_engaged were the most important predictors. Interestingly, these are all more important than was_productive. It's reassuring that our analysis has uncovered predictors that were non-obvious.

From this, we can already imagine some concrete suggestions to writing fellows to increase the likelihood of having a positive consultation. First, teach a concrete writing/revising technique; this is the single most important predictor of a consultation's success. Further, prompt the writer on thoughts and ideas that reflect their identity, and encourage the writer to write in a way that is honest to their identity (text_reflects_identity_more was a surprising second most powerful predictor). Finally, engage the writer intellectually. Perhaps this is an effect of the first two practices.

Models based on Area

Now we'd like to focus our analysis to answer Pam's questions, "Are STEM consultations effective?" and "How can STEM consultations be more effective?" We can attempt to address this by first asking, "How do important consultation-success predictors change among consultations for projects in STEM, the social sciences, and the humanities?" In order to answer this question, we need to make some philosophical judgments about whether certain fields fall into STEM/social sciences/humanities categories. We determined whether different consultations fell under these categories based on the "department" variable, which is the academic area that most closely fits the content of the consultation according to the student. From there, we defined an "area" variable that categorized observations into STEM, social sciences, humanities, ID1, and miscellaneous datasets. Because this significantly reduced the size of the datasets we were using to train our models, we removed categorial variables with too many levels so that the models would not make decisions based on random noise. We also defined a new variable that identified whether Writing Partners had STEM expertise, based on their majors and whether they had lab report training.

STEM Model

Before building each area model, we rebalanced the data. Since there are only 204 STEM observations, we didn't downsample the majority class. However, we did resample some of the minority response observations to mitigate the effect of the response imbalance.

```
positives <- stem %>% filter(was_successful_loose=='Yes')

negatives <- stem %>% filter(was_successful_loose=='No')
resampledNeg <- sample_n(negatives, 40, replace=TRUE)

stem_sample <- bind_rows(positives, resampledNeg, negatives)
```

Now let's take a look at the training accuracy as well as the most important variables of the model.

```
##     No Yes class.error
## No  51   0           0
## Yes  0 193           0
```

```
##                      variable importance
## 1            attached_partner  100.00000
## 2                student_year   80.90792
## 3                   eth_white   60.94373
## 4                  genderMale   31.47568
## 5                 school_year   31.32576
## 6      visit_challenging_assn   30.32772
## 7           hoped_polish_oral   29.52600
## 8                 hoped_style   22.94051
## 9 text_reflects_identity_more   22.53656
```

Unfortunately, the small sample size precludes us from making any recommendations that shouldn't be taken without a grain of salt. This high-variance model indicates that the most important variables are whether the consultation is with a Writing Partner associated to the course that the assignment is for, the student's school year, and whether the student was white. Rather than indicating to us that these factors determine consultation success, we have only learned summarizing information, e.g. the most important variable tells us that many STEM consultations only happen because they are required. One interesting output from this model is that none of the STEM-related variables made an appearance, including whether the Writing Partner had STEM expertise (including their major and lab report training level), so perhaps this hints that these factors are not as important as we might think.

Humanities Model

Because there are significantly more appointments with an emphasis on the humanities (1262 observations), we can afford to downsample the majority class now. We perform the same steps, building a Random Forest and examining the most important variables.

```
positives <- humanities %>% filter(was_successful_loose=='Yes')
sampledPos <- sample_n(positives, 800, replace=FALSE)

negatives <- humanities %>% filter(was_successful_loose=='No')
resampledNeg <- sample_n(negatives, 400, replace=TRUE)

hum_sample <- bind_rows(sampledPos, resampledNeg, negatives)
```

```
##      No Yes class.error
## No  416   0     0.00000
## Yes   1 799     0.00125
```

```
##
variable
## 1
learned_new_technique
## 2
text_reflects_identity_more
## 3
felt_welcome
## 4
clearer_sense_direction
## 5
fellow_foreign_languageSpanish
## 6
student_year
## 7
attached_partner
## 8 type_of_appointmentAn appointment for a course with an attached Writing Fellow
(either for ID1 or for another course).
## 9
prioritizes_identity
##   importance
## 1  100.00000
## 2   57.92983
## 3   46.71964
## 4   46.69619
## 5   34.47294
## 6   31.16232
## 7   27.85176
## 8   27.52454
## 9   26.79778
```

Important variables such as whether students felt welcome and whether they had a clearer sense of direction with their assignment indicate that satisfaction stems not only from the results of the assignment but also from how the students felt during the process. Variables based on identity and intellectual engagement appear above producitivity level. These results indicate that humanities students value the experiential aspects of coming into the Writing Center and engaging with their writing on a deeper level.

Social Sciences Model

Finally, there were 1769 consultations for us to build a model.

```
##      No  Yes class.error
## No  618    0           0
## Yes   0 1100           0
```

```
##                       variable importance
## 1                 student_year  100.00000
## 2              was_productive   54.03895
## 3 text_reflects_identity_more   50.87379
## 4       learned_new_technique   45.99195
## 5                  school_year   42.54762
## 6           visit_better_grade   37.31436
## 7     clearer_sense_direction   32.61140
## 8                    eth_asian   32.32114
## 9        prioritizes_identity   32.08299
```

Interestingly, student year and productivity level are the most important variables. Perhaps the former means that students in the social sciences feel more at home in their writing as they progress through their educations, while the second indicates a mental paradigm shift from humanities to social sciences. Students in humanities seem to prioritize the writing process, while students in the social sciences hope to engage deeply with the content of their writing.

Conclusions

Perhaps the most interesting and unexpected insight to emerge from our analysis is the importance of a writer's identity. Across all of our models, the response to the question "I feel my work reflects my identity more after this consultation" was in the top three predictors of a successful consultation. We were surprised that this variable consistently provided more predictive value than questions like "I feel this consultation was productive."

Perhaps this is merely a reflection of the concerns of a typical Pomona College student. The institution offers countless lenses for examining and understanding identity. Understanding the implications of identity has become paramount in our current political and cultural moment; the vantage point that Pomona College offers during these times is unique and valuable, as we see in students' thoughtfulness in response to even mundane tasks such as an exit survey.

Our findings may have implications in the praxis of the Writing Center. Our results flout Barthes' declaration that the author is dead. Our results indicate the exact opposite: that attention to the identity of the author is critical in the success of the writing process. Students that mind their identity and the way their writing reflects, refracts, or undermines it are more likely to find success and satisfaction in writing.

A practical suggestion from this is to priotize the writer rather than the writing. Instead of considering how to make the text better in isolation, perhaps Writing Partners should consider how to enable the writer to express themselves. This may be counterintuitive: when asked for help writing a paper, shouldn't we prioritize the paper? Perhaps not. This requires taking a step back from the writing in order to see the writing process and ultimately, the writer.

Another practical suggestion is to consistently teach students new writing techniques when they visit the Writing Center. This was universally the most predictive variable.

An area of future work is to examine the interactions between the variables on how a student defines and priotizies identity and the variables about whether the text reflected their identity more and how this interaction relates to a successful consultation overall. It would be interesting to know if students that express certain identities tend to feel more successful in the writing process if their finished text more directly reflects their identity. Through this analysis, we've identified that the majority of Pomona College students feel this way, but it would be interesting and useful to explore which, if any, subsets of students are [illegible]ted and rewarded by expressions of identity.