

About DO's status

Jon Guillemette

February 6, 2017

Getting a list of DO

Note that this was done with the data prior to obtaining the new pull with the MSE

We can obtain a list of DO (by student number) by using the DOfinder.R program that we will keep refining to verify inter-school transfers, but currently DOs are defined as people who aren't currently students and people who haven't graduated. A clear weak point is if someone is taking a year off or some sort of hiatus, they are counted as a DO.

Let us begin with the 36k DOs that we have found since 2010 at Dawson.

Step 1

The weirdness begins when we start tracking those DOs across different tables:

```
AdCounted<-DO[student_number %in% admission$student_number]
length(unique(AdCounted$student_number))
```

```
## [1] 20651
```

```
AdLeftout<- DO[student_number %NI% admission$student_number]
length(unique(AdLeftout$student_number))
```

```
## [1] 15313
```

Step 2

Ok. So this means that there are 15.3k students who are NOT in the admissions table, but were counted as DOs when reading from the etudiant session table. That is a bit odd, let's dig deeper into the etudiant session table to see who is doing what:

```
IDESDO<-etudiant_session[student_number %in% DO$student_number][,.(student_number, IDEtudiantSession, SPE
                                                                    TypeFrequentation)]
IDESDO[, ':= ' (COUNT=.N), by=student_number]
IDESDO[, .N, by=COUNT]
```

```
##      COUNT      N
##  1:      6 19086
##  2:      5 20395
##  3:      9  8505
##  4:      7 18200
##  5:     11  3289
##  6:      1 14947
##  7:      4  9592
##  8:      8 12928
```

```
## 9:      2  7454
## 10:     12 1944
## 11:      3 3975
## 12:     13  923
## 13:     10 5170
## 14:     15  330
## 15:     17   85
## 16:     14  700
## 17:     21  105
## 18:     19   38
## 19:     16  176
## 20:     25   25
## 21:     20   20
##      COUNT    N
```

```
IDESD01<-IDESD0[IDESD0$COUNT==1]
```

If you specifically look at where count equals 1 (line 6), you can see that there are almost 15k instances of students only having 1 ID etudiant session which means that they stayed only 1 semester.

Step 3

In that case, let's look more closely at the type of stay that these students have:

```
D0stat<- etudiant_session[student_number %in% AdLeftout$student_number][,.(student_number,TypeFrequentation)
D0stat[,.N, by=TypeFrequentation]
```

```
##      TypeFrequentation      N
## 1:                  SO 15629
## 2:                  TPA   130
## 3:                  TPL    46
```

```
D0stat[, ':= ' (COUNT = .N), by=student_number]
D0stat1<-etudiant_session[student_number %in% IDESD01$student_number][,.(student_number,TypeFrequentation)
D0stat1SO<-D0stat1[TypeFrequentation=="SO"]
```

Ok, more weirdness. Keep in mind that these are based on ID etudiant session so it is normal that the total is larger than 15.3k. It is still weird that more than 15k of students who are in the etudiant session table have SO as a type de frequentation. This should be TPL for full time or TPA for part time. SO would mean no classes. Let's double check that in the inscription table.

Step 4

```
InIDESD0 <- inscription[IDEtudiantSession %in% D0stat1SO$IDEtudiantSession]
InIDESD0[,.N,by=Note]
```

```
##      Note      N
## 1:    NA   624
```

This is the weirdest part... Our 15k or so students who have student session IDs, but almost all have SO as a status and therefore aren't taking any classes. Of these students, only 624 are actually registered for a class AND they all got NA as a grade which means they never finished the class (not sure the exact criteria to get NA) but they definitely didn't finish.

Conclusion

Let's put this in context: of the 36k dropouts that we have found, we just realised that approximately 15k of those NEVER WALKED INTO DAWSON. That means that the drop-out rate isn't approximately 36%, but rather more like 20%! I think the way everyone has been calculating drop outs is incorrect and are therefore overestimating Qc's drop-outs. The end purpose of this study was to reduce the number of drop-outs, I guess this counts :)

This also affects who we should try to help. We can't do much for the students who never walk in...

Follow-up questions:

- Sam: could this be the reason why the admission table has garbage. Students who were admitted to the college, but never attended a single class would be given a bogus admission ID?
- Of the other 21k who stayed and eventually dropped out, I think we should be able to develop a new category to see if it is for academic reasons or not. This would be like what we had done for 1 student together, but done for all.
- Let's confirm the use of SO with Derek or someone who knows...