Divison of Undergraduate Education

Dominik Glandorf

2022-11-11

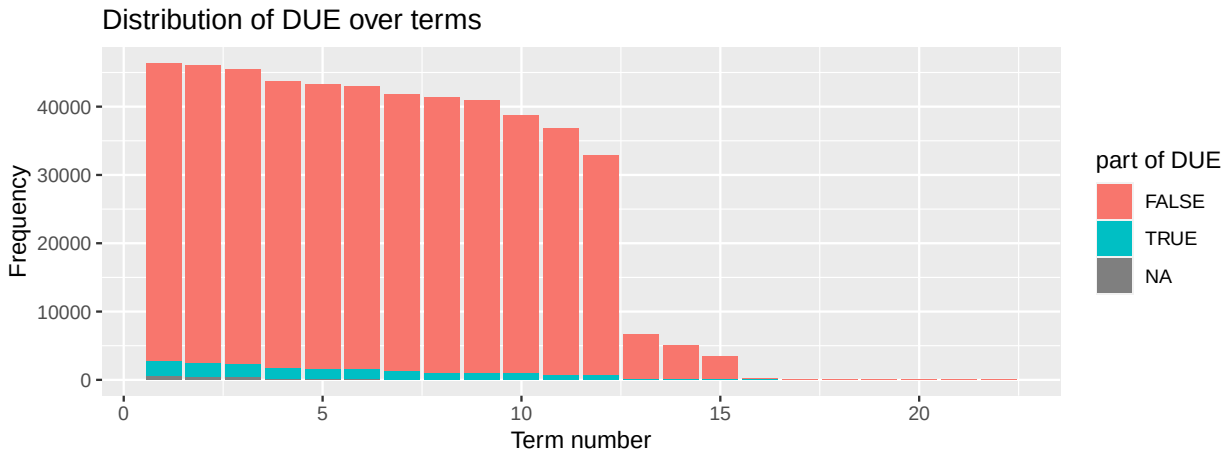
Division of Undergraduate Education

Students are assigned to one or more schools for each term (missing for 0.3% of terms). These are the schools occurring in the data:

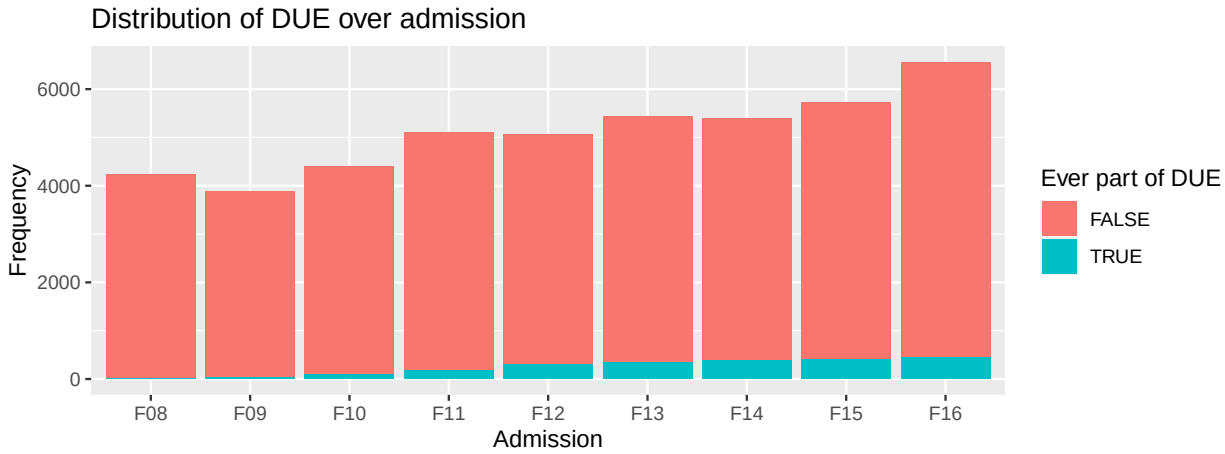
```
##  [1] "School of Biological Sciences"  "School of Humanities"
##  [3] "Program in Public Health"       "School of the Arts"
##  [5] "Pharmacy"                       "School of Social Ecology"
##  [7] "School of Engineering"          "School of Physical Sciences"
##  [9] "School of Social Sciences"      "School of Education"
## [11] "Div of Undergraduate Education" "School of Info & Computer Sci"
## [13] "School of Business"             "School of Nursing"
## [15] NA                               "Summer Session"
```

```
## [1] 0
```

4.7% of students are part of it at least once, overall in 3.1% of terms. However, 3.48% of students have been part of DUE in all of their terms whereas 4.3% have been in their first term.



The share of DUE students decreases over time. Few students (0.4%) join DUE after their first term.



The school is much often assigned for students admitted in later years.

Dropout

45.9% of students that have ever been part of DUE drop out, compared to 9.46% among those who have never been.

49.6% of students that have been part of DUE in first term drop out, compared to 9.45% among those who have not been.

60.1% of students that have always been part of DUE drop out, compared to 9.41% among those who have not always been.

How are DUE students different from others?

Do they study a shorter time at UCI?

```
num_terms = aggregate(term_num ~ mellon_id, term_features, FUN=max)
student_ever_DUE = merge(student_ever_DUE, num_terms, by="mellon_id")
is_DUE = student_ever_DUE$major_school_name_1==TRUE
dropout = student_ever_DUE$dropout==TRUE

terms_DUE_grads = mean(student_ever_DUE$term_num[is_DUE&!dropout], na.rm=T)
terms_non_DUE_grads = mean(student_ever_DUE$term_num[!is_DUE&!dropout], na.rm=T)

terms_DUE_DO = mean(student_ever_DUE$term_num[is_DUE&dropout], na.rm=T)
terms_non_DUE_DO = mean(student_ever_DUE$term_num[!is_DUE&dropout], na.rm=T)
```

Graduates both ever been at DUE or not usually study 12.3 or 12 terms. Dropouts ever been at DUE only study 4 compared to 5.3.

Is it the ones that did not declare their major or were unaffiliated in the first term?

Almost no students are unaffiliated in the first term (0.06%), neither DUE nor non-DUE students.

Among the DUE students 93.3% did not declare a major in the first term. However, among the remainder this is the case for 19.5%.

Is it international students?

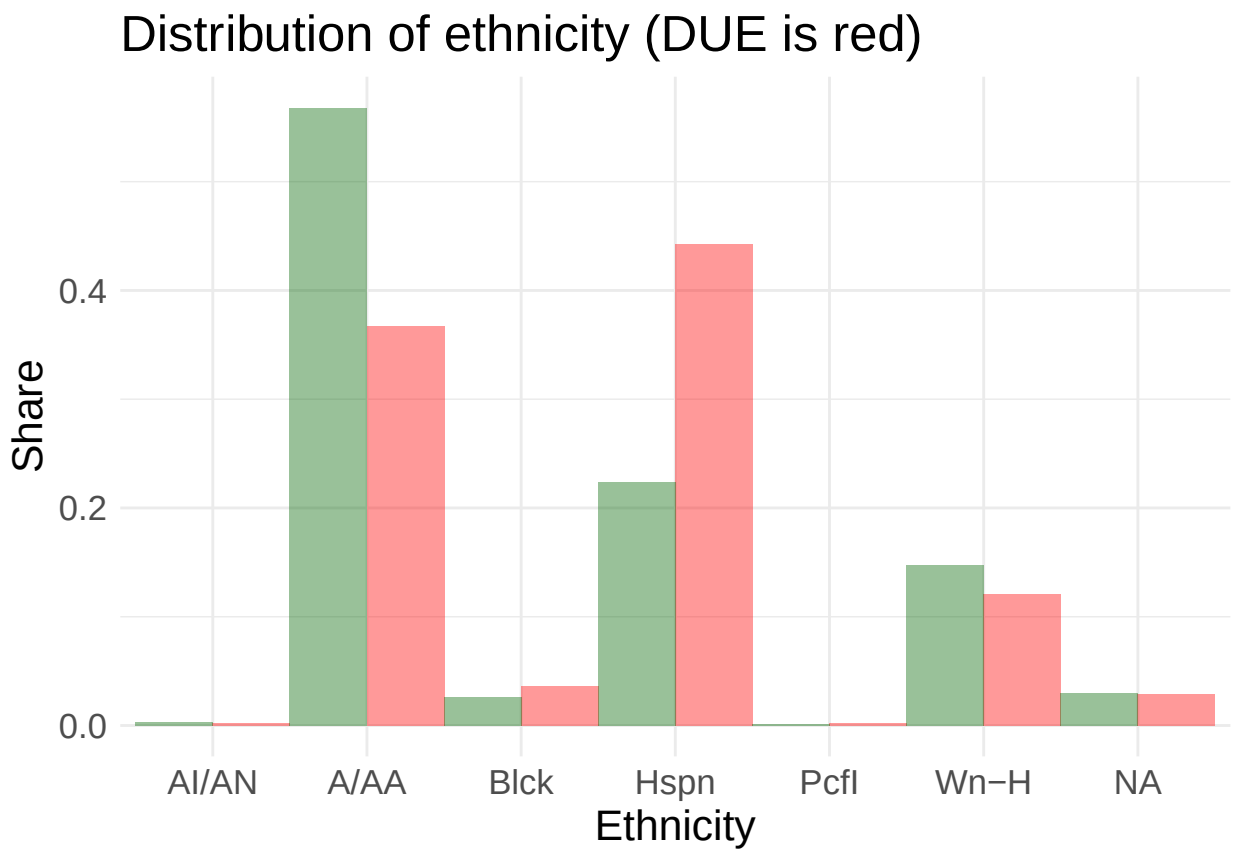
The ratio of international students only differs slightly between ever DUE and never DUE students (10.7% vs 13.4%).

Do they differ demographically?

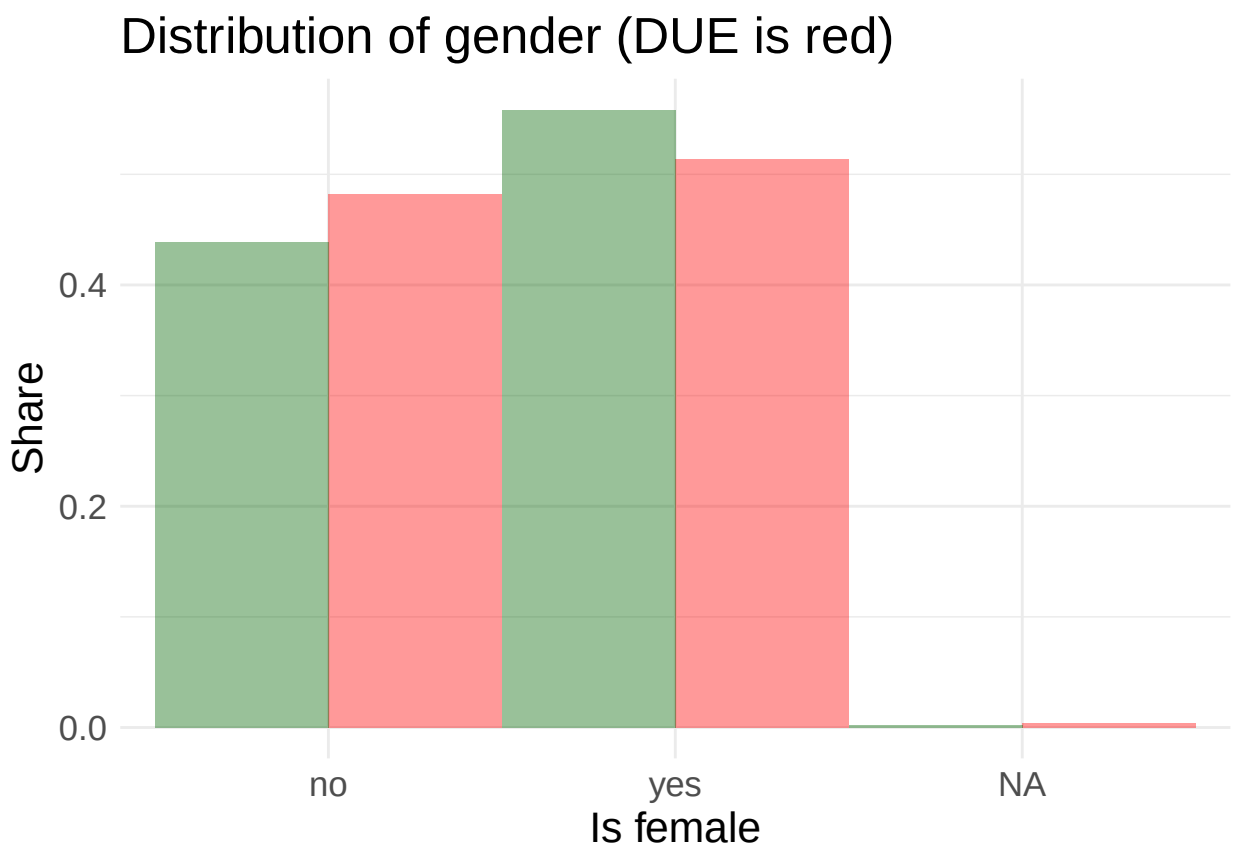
```
student_ever_DUE = merge(student_ever_DUE, bg_sub[,c('mellon_id','ethnicity','female')], by="mellon_id"

ggplot(data = student_ever_DUE, aes(x = ethnicity)) +
  geom_bar(data=subset(student_ever_DUE,major_school_name_1 == T),aes(y = (..count..)/sum(..count..)),f
  geom_bar(data=subset(student_ever_DUE,major_school_name_1 == F),aes(y = (..count..)/sum(..count..)),f
  theme_minimal() +
  labs(title = "Distribution of ethnicity (DUE is red)",
       x = "Ethnicity",
       y = "Share") +
  theme(text = element_text(size = 16)) +
  scale_x_discrete(label=abbreviate)
```

```
## Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(count)` instead.
```




```
ggplot(data = student_ever_DUE, aes(x = female)) +
  geom_bar(data=subset(student_ever_DUE,major_school_name_1 == T),aes(y = (..count..)/sum(..count..)),f
  geom_bar(data=subset(student_ever_DUE,major_school_name_1 == F),aes(y = (..count..)/sum(..count..)),f
  theme_minimal() +
  labs(title = "Distribution of gender (DUE is red)",
       x = "Is female",
       y = "Share") +
  theme(text = element_text(size = 16)) +
  scale_x_discrete(label=abbreviate)
```



Do they perform differently in pre uni scores?

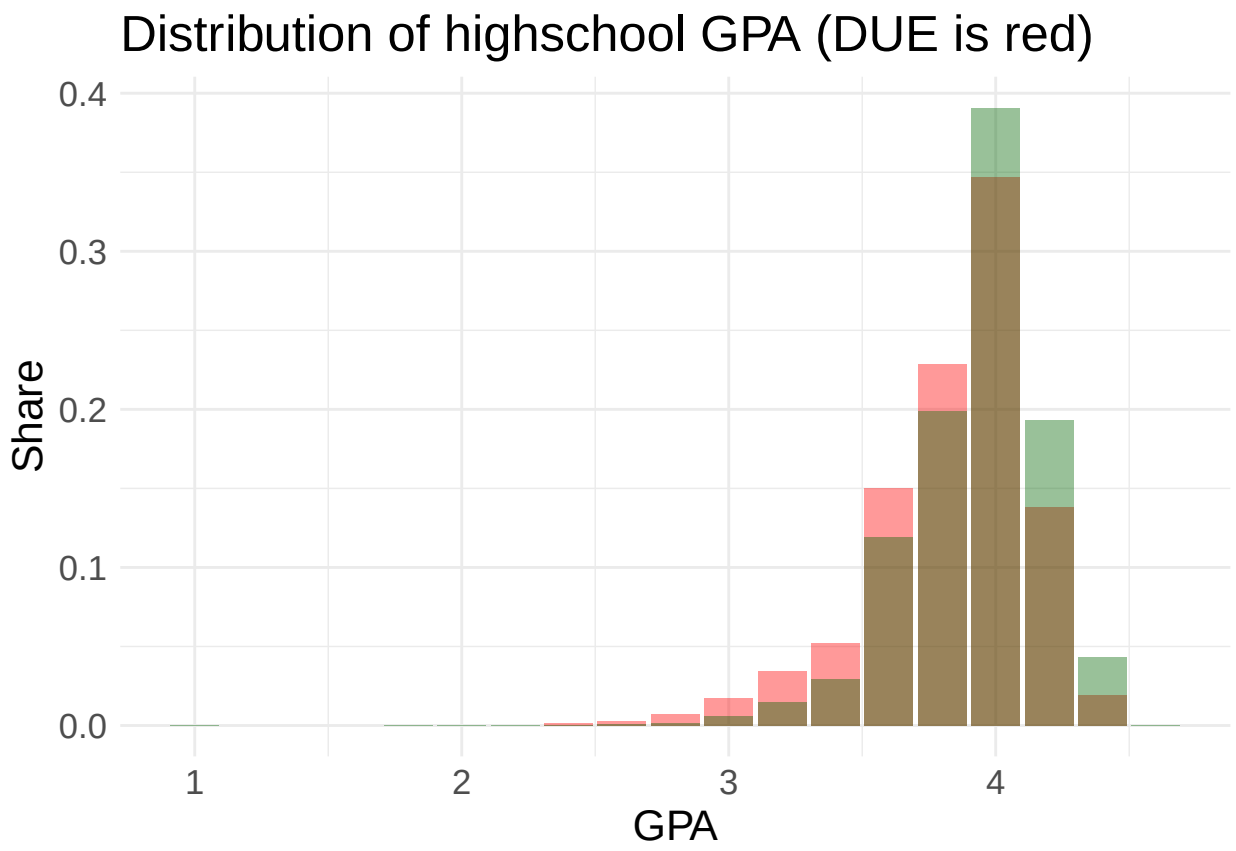
```
student_ever_DUE = merge(student_ever_DUE, students[,c('mellon_id','hs_gpa','toefl_score','uc_total_sco
```

Highschool GPA

```
ggplot(data = student_ever_DUE, aes(x = round(hs_gpa*5)/5)) +
  geom_bar(data=subset(student_ever_DUE,major_school_name_1 == T),aes(y = (..count..)/sum(..count..)),f
  geom_bar(data=subset(student_ever_DUE,major_school_name_1 == F),aes(y = (..count..)/sum(..count..)),f
  theme_minimal() +
  labs(title = "Distribution of highschool GPA (DUE is red)",
       x = "GPA",
       y = "Share") +
  theme(text = element_text(size = 16))
```

```
## Warning: Removed 129 rows containing non-finite values (`stat_count()`).
```

```
## Warning: Removed 12396 rows containing non-finite values (`stat_count()`).
```

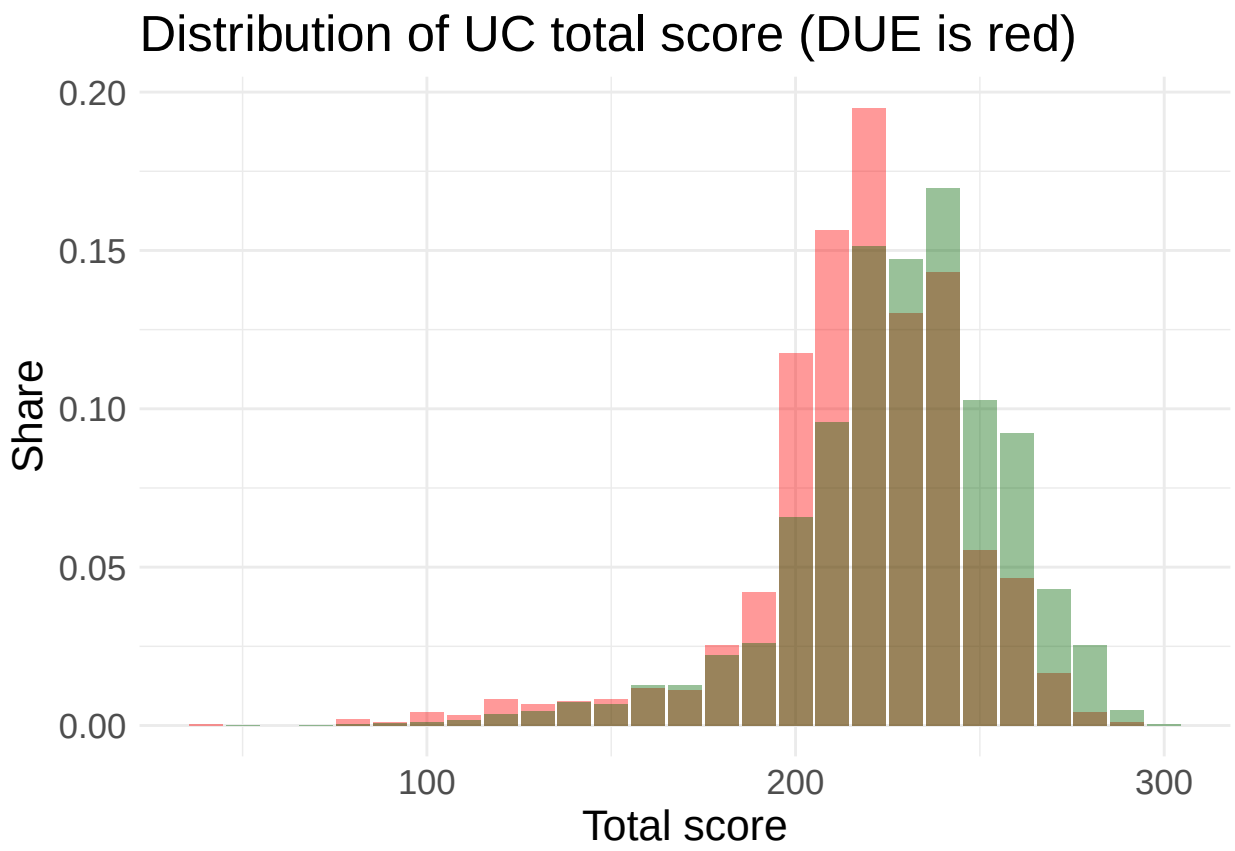


UC total score

```
ggplot(data = student_ever_DUE, aes(x = round(uc_total_score/10)*10)) +
  geom_bar(data=subset(student_ever_DUE,major_school_name_1 == T),aes(y = (..count..)/sum(..count..)),f
  geom_bar(data=subset(student_ever_DUE,major_school_name_1 == F),aes(y = (..count..)/sum(..count..)),f
  theme_minimal() +
  labs(title = "Distribution of UC total score (DUE is red)",
       x = "Total score",
       y = "Share") +
  theme(text = element_text(size = 16))
```

```
## Warning: Removed 135 rows containing non-finite values (`stat_count()`).
```

```
## Warning: Removed 12413 rows containing non-finite values (`stat_count()`).
```

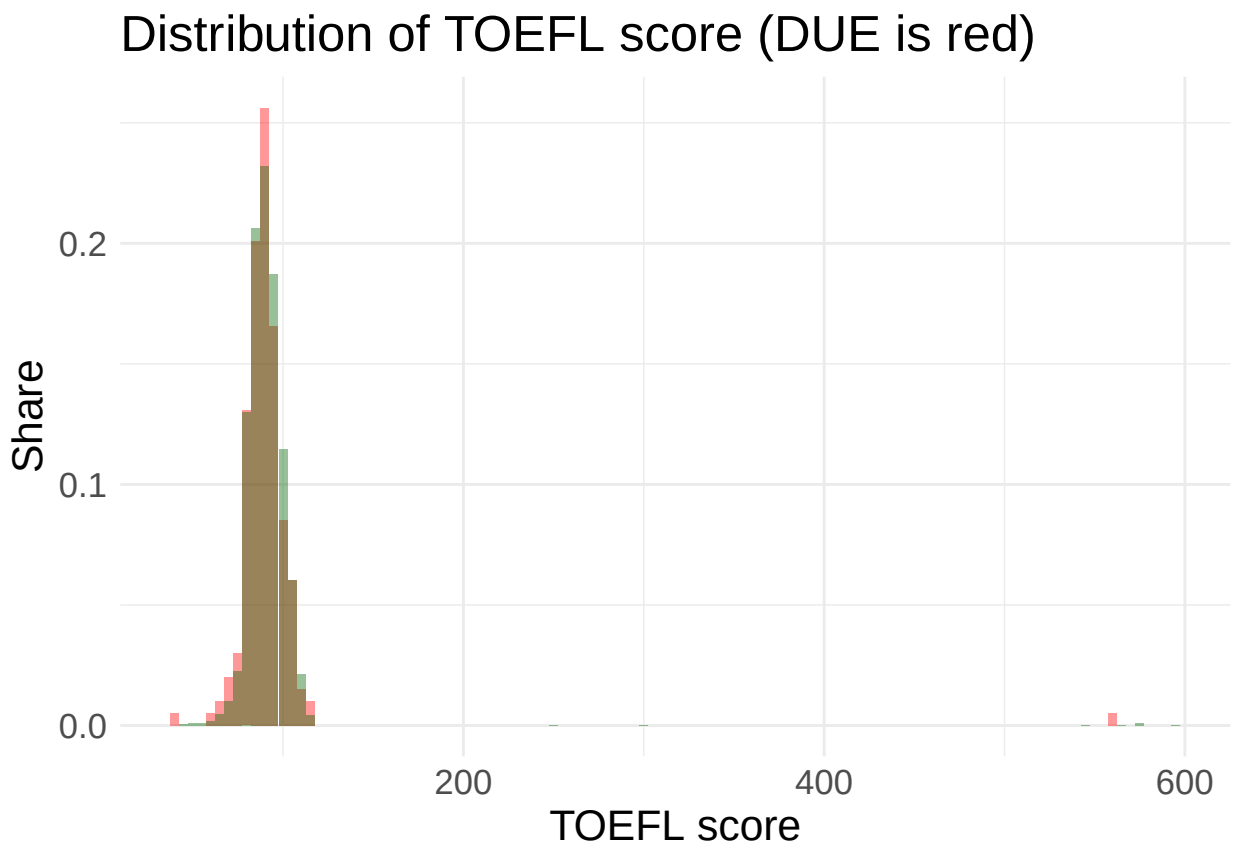


TOEFL score

```
ggplot(data = student_ever_DUE, aes(x = round(toefl_score/5)*5)) +
  geom_bar(data=subset(student_ever_DUE,major_school_name_1 == T),aes(y = (..count..)/sum(..count..)),f
  geom_bar(data=subset(student_ever_DUE,major_school_name_1 == F),aes(y = (..count..)/sum(..count..)),f
  theme_minimal() +
  labs(title = "Distribution of TOEFL score (DUE is red)",
       x = "TOEFL score",
       y = "Share") +
  theme(text = element_text(size = 16))
```

```
## Warning: Removed 1976 rows containing non-finite values (`stat_count()`).
```

```
## Warning: Removed 39284 rows containing non-finite values (`stat_count()`).
```

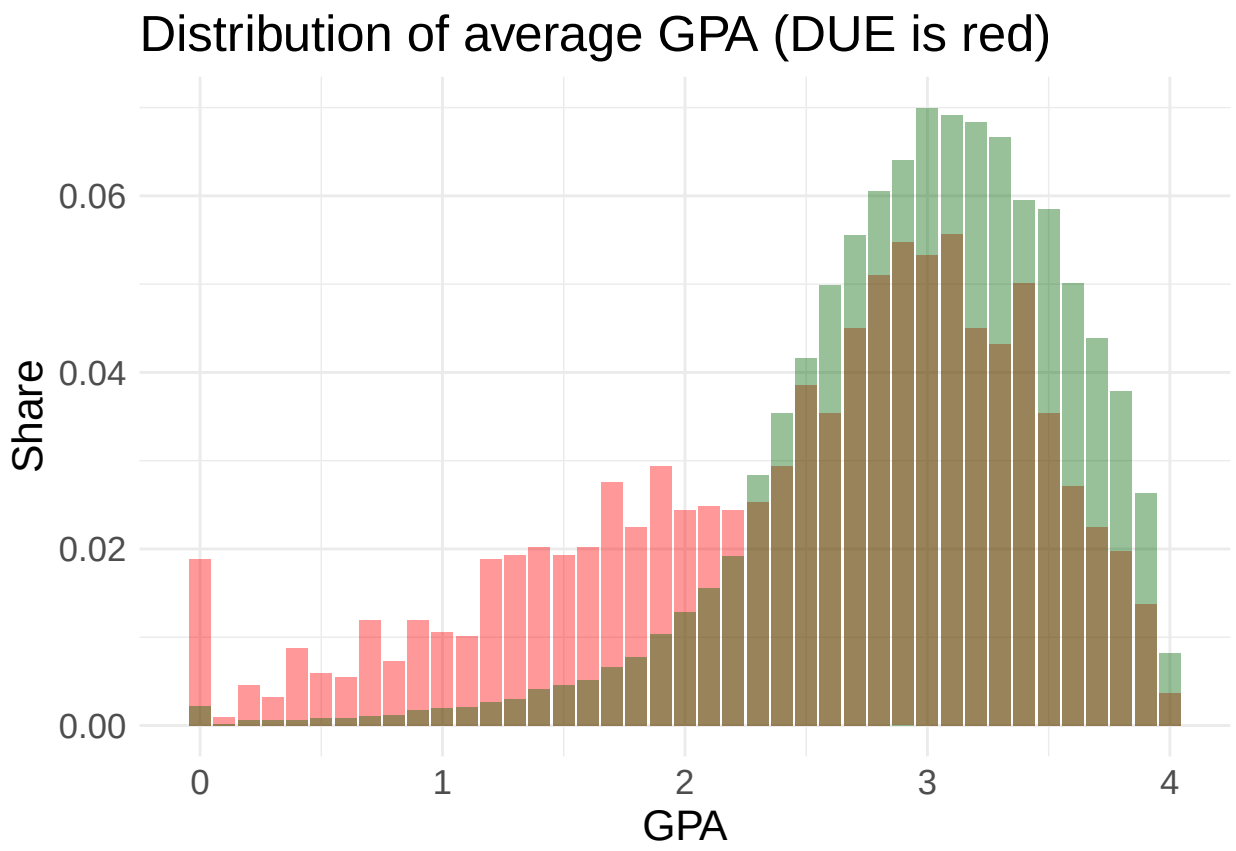


Do they perform differently in first term?

```
student_gpas = aggregate(gpa_term ~ mellon_id, terms, FUN=mean)

student_ever_DUE = merge(student_ever_DUE, student_gpas[,c('mellon_id','gpa_term')], by="mellon_id")

ggplot(data = student_ever_DUE, aes(x = round(gpa_term*10)/10)) +
  geom_bar(data=subset(student_ever_DUE,major_school_name_1 == T),aes(y = (..count..)/sum(..count..)),fi
  geom_bar(data=subset(student_ever_DUE,major_school_name_1 == F),aes(y = (..count..)/sum(..count..)),fi
  theme_minimal() +
  labs(title = "Distribution of average GPA (DUE is red)",
       x = "GPA",
       y = "Share") +
  theme(text = element_text(size = 16))
```


Distribution of average GPA (DUE is red)
Share
0.06
0.04
0.02
0.00
0
1
2
3
4
GPA