

CS Last Course Analysis

Code ▾

Vanessa Gonzalez

2018-06-24

CS Grad: Yes - CS Students OtherMajor - Left CS NG - Not Graduated

Data from 2008-2018 and excludes current students and students that graduated from CS.

Import Data Set and open libraries.

Hide

```
library(readxl)
CourseDateData <- read_excel("CourseDateData.xlsx")
library(readxl)
CourseDateDataNoCS <- read_excel("CourseDateDataNoCS.xlsx")
```

Take a look at the data.

Hide

```
summary(CourseDateDataNoCS)
```

UID	Year of OriginalMajorDate	YearsFromOMD	4YG	
5YG	6YG			
Length:195	Length:195	Length:195	Length:195	L
Length:195	Length:195			
Class :character	Class :character	Class :character	Class :character	C
Class :character	Class :character			
Mode :character	Mode :character	Mode :character	Mode :character	M
Mode :character	Mode :character			
GraduationStatus	CsGrad	Nine.CSCI442	Eight.CSCI400	Seven.CSCI370
Six.CSCI406	Five.CSCI403	Five.MATH332		
Length:195	Length:195	Min. :2010	Min. :2010	Min. :2011
Min. :2010	Min. :2010	Min. :2010		M
Class :character	Class :character	1st Qu.:2012	1st Qu.:2014	1st Qu.:2012
1st Qu.:2013	1st Qu.:2014	1st Qu.:2012		1
Mode :character	Mode :character	Median :2014	Median :2015	Median :2015
Median :2015	Median :2016	Median :2014		M
		Mean :2014	Mean :2015	Mean :2014
				M

```

ean      :2015      Mean      :2016      Mean      :2014
                                3rd Qu.:2017      3rd Qu.:2016      3rd Qu.:2016      3
rd Qu.:2017      3rd Qu.:2017      3rd Qu.:2015
                                Max.      :2018      Max.      :2017      Max.      :2018      M
ax.      :2019      Max.      :2019      Max.      :2018
                                NA's      :180      NA's      :177      NA's      :182      N
A's      :172      NA's      :177      NA's      :125
    Five.CSCI306    Four.CSCI358    Four.CSCI341    Four.MATH225    Three.CSCI262    Three.MAT
H213    Two.CSCI261    Two.MATH201
    Min.      :2010    Min.      :2009    Min.      :2010    Min.      :2009    Min.      :2009    Min.      :2
009    Min.      :2009    Min.      :2010
    1st Qu.:2012    1st Qu.:2011    1st Qu.:2012    1st Qu.:2011    1st Qu.:2012    1st Qu.:2
011    1st Qu.:2011    1st Qu.:2011
    Median :2014    Median :2014    Median :2014    Median :2013    Median :2014    Median :2
013    Median :2013    Median :2013
    Mean    :2014    Mean    :2014    Mean    :2014    Mean    :2013    Mean    :2014    Mean    :2
013    Mean    :2013    Mean    :2013
    3rd Qu.:2016    3rd Qu.:2016    3rd Qu.:2016    3rd Qu.:2015    3rd Qu.:2016    3rd Qu.:2
015    3rd Qu.:2015    3rd Qu.:2015
    Max.    :2019    Max.    :2018    Max.    :2019    Max.    :2019    Max.    :2019    Max.    :2
018    Max.    :2018    Max.    :2019
    NA's    :153    NA's    :132    NA's    :123    NA's    :73    NA's    :94    NA's    :5
9    NA's    :37    NA's    :118
    Two.MATH112    One.MATH111    One.CSCI101
    Min.      :2009    Min.      :2009    Min.      :2011
    1st Qu.:2010    1st Qu.:2010    1st Qu.:2012
    Median :2012    Median :2012    Median :2014
    Mean    :2013    Mean    :2012    Mean    :2014
    3rd Qu.:2015    3rd Qu.:2015    3rd Qu.:2015
    Max.    :2018    Max.    :2018    Max.    :2018
    NA's    :28    NA's    :10    NA's    :66

```

```
str(CourseDateDataNoCS)
```

```
Classes 'tbl_df', 'tbl' and 'data.frame':  195 obs. of  25 variables:
 $ UID                      : chr  "12972" "12973" "41647" "98022" ...
 $ Year of OriginalMajorDate: chr  "2008" "2008" "2011" "2008" ...
 $ YearsFromOMD             : chr  "9.88" "9.88" "6.88" "9.88" ...
 $ 4YG                      : chr  "No" "No" "No" "Yes" ...
 $ 5YG                      : chr  "Yes" "No" "No" "Yes" ...
 $ 6YG                      : chr  "Yes" "No" "No" "Yes" ...
 $ GraduationStatus         : chr  "Graduated" "InactiveReg" "InactiveReg" "Graduated"
 " ...
 $ CsGrad                   : chr  "OtherMajor" "NG" "NG" "OtherMajor" ...
 $ Nine.CSCI442             : num  NA NA NA NA NA ...
 $ Eight.CSCI400            : num  NA NA 2017 NA NA ...
 $ Seven.CSCI370            : num  NA NA NA NA NA ...
 $ Six.CSCI406              : num  NA NA 2018 NA NA ...
 $ Five.CSCI403             : num  NA NA 2017 NA NA ...
 $ Five.MATH332             : num  NA 2011 2015 NA NA ...
 $ Five.CSCI306             : num  NA NA 2017 NA 2011 ...
 $ Four.CSCI358             : num  NA 2012 2017 NA NA ...
 $ Four.CSCI341            : num  NA 2011 2015 NA NA ...
 $ Four.MATH225             : num  2010 2011 2015 2010 2009 ...
 $ Three.CSCI262            : num  NA 2010 2015 NA 2010 ...
 $ Three.MATH213            : num  2010 2010 2013 2009 2009 ...
 $ Two.CSCI261              : num  2010 2010 2014 2010 2009 ...
 $ Two.MATH201              : num  2011 2012 NA 2010 NA ...
 $ Two.MATH112              : num  2009 2010 2013 2009 2009 ...
 $ One.MATH111              : num  2009 2009 2013 2009 2009 ...
 $ One.CSCI101              : num  NA NA NA NA NA ...
```

Remove unnecessary columns

Hide

```
DataNoCS <- CourseDateDataNoCS[-(1:6)]
str(DataNoCS)
```

```
Classes 'tbl_df', 'tbl' and 'data.frame':  195 obs. of  19 variables:
 $ GraduationStatus: chr  "Graduated" "InactiveReg" "InactiveReg" "Graduated" ...
 $ CsGrad          : chr  "OtherMajor" "NG" "NG" "OtherMajor" ...
 $ Nine.CSCI442    : num  NA NA NA NA NA ...
 $ Eight.CSCI400   : num  NA NA 2017 NA NA ...
 $ Seven.CSCI370   : num  NA NA NA NA NA ...
 $ Six.CSCI406     : num  NA NA 2018 NA NA ...
 $ Five.CSCI403    : num  NA NA 2017 NA NA ...
 $ Five.MATH332    : num  NA 2011 2015 NA NA ...
 $ Five.CSCI306    : num  NA NA 2017 NA 2011 ...
 $ Four.CSCI358    : num  NA 2012 2017 NA NA ...
 $ Four.CSCI341    : num  NA 2011 2015 NA NA ...
 $ Four.MATH225    : num  2010 2011 2015 2010 2009 ...
 $ Three.CSCI262   : num  NA 2010 2015 NA 2010 ...
 $ Three.MATH213   : num  2010 2010 2013 2009 2009 ...
 $ Two.CSCI261     : num  2010 2010 2014 2010 2009 ...
 $ Two.MATH201     : num  2011 2012 NA 2010 NA ...
 $ Two.MATH112     : num  2009 2010 2013 2009 2009 ...
 $ One.MATH111     : num  2009 2009 2013 2009 2009 ...
 $ One.CSCI101     : num  NA NA NA NA NA ...
```

Transform data type to the appropriate type.

[Hide](#)

```
DataNoCS$GraduationStatus <- as.factor(DataNoCS$GraduationStatus)
DataNoCS$CsGrad <- as.factor(DataNoCS$CsGrad)
str(DataNoCS)
```

```
Classes 'tbl_df', 'tbl' and 'data.frame': 195 obs. of 19 variables:
 $ GraduationStatus: Factor w/ 2 levels "Graduated","InactiveReg": 1 2 2 1 1 2 1 2 2
 2 ...
 $ CsGrad           : Factor w/ 2 levels "NG","OtherMajor": 2 1 1 2 2 1 2 1 1 1 ...
 $ Nine.CSCI442     : num NA NA NA NA NA ...
 $ Eight.CSCI400    : num NA NA 2017 NA NA ...
 $ Seven.CSCI370    : num NA NA NA NA NA ...
 $ Six.CSCI406      : num NA NA 2018 NA NA ...
 $ Five.CSCI403     : num NA NA 2017 NA NA ...
 $ Five.MATH332     : num NA 2011 2015 NA NA ...
 $ Five.CSCI306     : num NA NA 2017 NA 2011 ...
 $ Four.CSCI358     : num NA 2012 2017 NA NA ...
 $ Four.CSCI341     : num NA 2011 2015 NA NA ...
 $ Four.MATH225     : num 2010 2011 2015 2010 2009 ...
 $ Three.CSCI262    : num NA 2010 2015 NA 2010 ...
 $ Three.MATH213    : num 2010 2010 2013 2009 2009 ...
 $ Two.CSCI261      : num 2010 2010 2014 2010 2009 ...
 $ Two.MATH201      : num 2011 2012 NA 2010 NA ...
 $ Two.MATH112      : num 2009 2010 2013 2009 2009 ...
 $ One.MATH111      : num 2009 2009 2013 2009 2009 ...
 $ One.CSCI101      : num NA NA NA NA NA ...
```

Remove Math Courses to understand better when students leave the CS program.

Hide

```
CourseDates <- DataNoCS[-(1:2)]
CourseDates <- CourseDates[-(6)]
CourseDates <- CourseDates[-(9)]
CourseDates <- CourseDates[-(10)]
CourseDates <- CourseDates[-(11:13)]
str(CourseDates)
```

```
Classes 'tbl_df', 'tbl' and 'data.frame': 195 obs. of 11 variables:
 $ Nine.CSCI442 : num NA NA NA NA NA ...
 $ Eight.CSCI400: num NA NA 2017 NA NA ...
 $ Seven.CSCI370: num NA NA NA NA NA ...
 $ Six.CSCI406  : num NA NA 2018 NA NA ...
 $ Five.CSCI403 : num NA NA 2017 NA NA ...
 $ Five.CSCI306 : num NA NA 2017 NA 2011 ...
 $ Four.CSCI358 : num NA 2012 2017 NA NA ...
 $ Four.CSCI341 : num NA 2011 2015 NA NA ...
 $ Three.CSCI262: num NA 2010 2015 NA 2010 ...
 $ Two.CSCI261  : num 2010 2010 2014 2010 2009 ...
 $ One.CSCI101  : num NA NA NA NA NA ...
```

Transform data set into data frame.

[Hide](#)

```
dfCourseDates <- as.data.frame(CourseDates)
```

Replace NA with 0

[Hide](#)

```
dfCourseDates[is.na(dfCourseDates)] <- 0
head(dfCourseDates)
```

	Nine.CSCI442 <dbl>	Eight.CSCI400 <dbl>	Seven.CSCI370 <dbl>	Six.CSCI406 <dbl>	Five.CSCI403 <dbl>	Five.CSCI306 <dbl>
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	2017	0	2018	2017	2017
4	0	0	0	0	0	0
5	0	0	0	0	0	2011
6	0	0	0	0	0	0

6 rows | 1-8 of 11 columns

Find highest level last class each student took.

[Hide](#)

```
LastClass <- colnames(dfCourseDates)[max.col(dfCourseDates,ties.method="first")]
dfCourseDates$LastClass <- LastClass
head(dfCourseDates)
```

	Nine.CSCI442 <dbl>	Eight.CSCI400 <dbl>	Seven.CSCI370 <dbl>	Six.CSCI406 <dbl>	Five.CSCI403 <dbl>	Five.CSCI306 <dbl>
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	2017	0	2018	2017	2017
4	0	0	0	0	0	0
5	0	0	0	0	0	2011

6	0	0	0	0	0	0
---	---	---	---	---	---	---

6 rows | 1-8 of 12 columns

Hide

```
summary(dfCourseDates)
```

Nine.CSCI442		Eight.CSCI400		Seven.CSCI370		Six.CSCI406		Five.CSCI403	
Five.CSCI306		Four.CSCI358		Four.CSCI341					
Min. :	0.0	Min. :	0	Min. :	0.0	Min. :	0.0	Min. :	0.0
1st Qu.:	0.0	1st Qu.:	0	1st Qu.:	0.0	1st Qu.:	0.0	1st Qu.:	0.0
Median :	0.0	Median :	0	Median :	0.0	Median :	0.0	Median :	0.0
Mean :	154.9	Mean :	186	Mean :	134.3	Mean :	237.7	Mean :	186.1
3rd Qu.:	0.0	3rd Qu.:	0	3rd Qu.:	0.0	3rd Qu.:	0.0	3rd Qu.:	0.0
Max. :	2018.0	Max. :	2017	Max. :	2018.0	Max. :	2019.0	Max. :	2019.0
Three.CSCI262		Two.CSCI261		One.CSCI101		LastClass			
Min. :	0	Min. :	0	Min. :	0	Length:195			
1st Qu.:	0	1st Qu.:	2010	1st Qu.:	0	Class :character			
Median :	2010	Median :	2012	Median :	2012	Mode :character			
Mean :	1043	Mean :	1631	Mean :	1332				
3rd Qu.:	2014	3rd Qu.:	2015	3rd Qu.:	2015				
Max. :	2019	Max. :	2018	Max. :	2018				

Open additional libraries

Hide

```
library(caret)
library(lattice)
library(ggplot2)
```

Create table from data frame.

Hide

```
tbLastClass <- table(dfCourseDates$LastClass)
tbLastClass
```

```

Eight.CSCI400  Five.CSCI306  Five.CSCI403  Four.CSCI341  Four.CSCI358  Nine.CSCI442
One.CSCI101  Seven.CSCI370  Six.CSCI406
      10           12           4           20           25           25
18           1           8
Three.CSCI262  Two.CSCI261
      21           51

```

Create a data frame from table for graphing.

[Hide](#)

```

dfLastClass <- as.data.frame(tbLastClass)
dfLastClass

```

Var1	Freq
<fctr>	<int>
Eight.CSCI400	10
Five.CSCI306	12
Five.CSCI403	4
Four.CSCI341	20
Four.CSCI358	25
Nine.CSCI442	25
One.CSCI101	18
Seven.CSCI370	1
Six.CSCI406	8
Three.CSCI262	21
1-10 of 11 rows	<div> Previous 1 2 Next </div>

Graph frequency of last CS class taken by students that left the program.

[Hide](#)


```
p<-ggplot(data=dfLastClass, aes(x= reorder(Var1, Freq), y=Freq)) +
  geom_bar(stat="identity", fill="steelblue") +
  labs(title = "Last Highest CS Course Taken by Not Grad. and Other Major Students", x
= "Course", y = "Frequency") +
  ylim(0,60) +
  theme(axis.text.x=element_text(angle=90, hjust=1)) +
  geom_text(aes(label = Freq), hjust=-.5 , position = position_dodge(width = 1), inher
it.aes = TRUE) +
  coord_flip()
p
```

