

Other Majors

Code ▾

Vanessa Gonzalez

2018-06-24

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

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

Import Data Set

Hide

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

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
Mean :2015	Mean :2016	Mean :2014		M
		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.	: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	:2
013	Median	:2013	Median	:2013					
Mean	:2014	Mean	:2014	Mean	:2014	Mean	:2013	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.	:2
018	Max.	:2018	Max.	:2019					
NA's	:153	NA's	:132	NA's	:123	NA's	:73	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				

[Hide](#)

```
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 ...
```

[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 ...
```

[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 ...
```

Hide

```
dfCourseDatesNoCS <- as.data.frame(DataNoCS)
```

Hide

```
JustOM<-subset(dfCourseDatesNoCS, CsGrad == 'OtherMajor')
head(JustOM)
```

	GraduationStatus <fctr>	CsGrad <fctr>	Nine.CSCI442 <dbl>	Eight.CSCI400 <dbl>	Seven.CSCI370 <dbl>	Six.CSCI406 <dbl>
1	Graduated	OtherMajor	NA	NA	NA	NA
4	Graduated	OtherMajor	NA	NA	NA	NA
5	Graduated	OtherMajor	NA	NA	NA	NA
7	Graduated	OtherMajor	NA	NA	NA	NA
13	Graduated	OtherMajor	2011	NA	NA	2011
14	Graduated	OtherMajor	NA	NA	NA	NA

6 rows | 1-8 of 19 columns

Hide

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

```
'data.frame': 57 obs. of 11 variables:
 $ Nine.CSCI442 : num NA NA NA NA 2011 ...
 $ Eight.CSCI400: num NA NA NA NA NA NA NA NA NA ...
 $ Seven.CSCI370: num NA NA NA NA NA NA NA NA NA ...
 $ Six.CSCI406 : num NA NA NA NA 2011 ...
 $ Five.CSCI403 : num NA NA NA NA NA NA NA NA NA ...
 $ Five.CSCI306 : num NA NA 2011 NA 2011 ...
 $ Four.CSCI358 : num NA NA NA NA 2009 ...
 $ Four.CSCI341 : num NA NA NA 2012 2010 ...
 $ Three.CSCI262: num NA NA 2010 NA 2010 ...
 $ Two.CSCI261 : num 2010 2010 2009 2010 2009 ...
 $ One.CSCI101 : num NA NA NA NA NA ...
```

Replace NA with 0

Hide

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

	Nine.CSCI442	Eight.CSCI400	Seven.CSCI370	Six.CSCI406	Five.CSCI403	Five.CSCI306
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	2011
7	0	0	0	0	0	0
13	2011	0	0	2011	0	2011
14	0	0	0	0	0	0

6 rows | 1-8 of 11 columns

Hide

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

	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
4	0	0	0	0	0	0
5	0	0	0	0	0	2011
7	0	0	0	0	0	0
13	2011	0	0	2011	0	2011
14	0	0	0	0	0	0

6 rows | 1-8 of 12 columns

Hide

```
summary(JustOM)
```

```

  Nine.CSCI442      Eight.CSCI400      Seven.CSCI370      Six.CSCI406      Five.CSCI403
Five.CSCI306      Four.CSCI358
Min.   :  0.00  Min.   :  0.00  Min.   :  0.00  Min.   :  0.0  Min.   :  0.
00  Min.   :  0.0  Min.   :  0.0
1st Qu.:  0.00  1st Qu.:  0.00  1st Qu.:  0.00  1st Qu.:  0.0  1st Qu.:  0.
00  1st Qu.:  0.0  1st Qu.:  0.0
Median :  0.00  Median :  0.00  Median :  0.00  Median :  0.0  Median :  0.
00  Median :  0.0  Median :  0.0
Mean   :  70.65  Mean   :  70.74  Mean   :  35.39  Mean   : 212.1  Mean   :  70.
75  Mean   : 317.9  Mean   : 635.4
3rd Qu.:  0.00  3rd Qu.:  0.00  3rd Qu.:  0.00  3rd Qu.:  0.0  3rd Qu.:  0.
00  3rd Qu.:  0.0  3rd Qu.:2010.0
Max.   :2016.00  Max.   :2017.00  Max.   :2017.00  Max.   :2017.0  Max.   :2017.
00  Max.   :2017.0  Max.   :2018.0

  Four.CSCI341  Three.CSCI262  Two.CSCI261  One.CSCI101  LastClassOM
Min.   :  0  Min.   :  0  Min.   :  0  Min.   :  0.0  Length:57
1st Qu.:  0  1st Qu.:  0  1st Qu.:2009  1st Qu.:  0.0  Class :character
Median :  0  Median :  0  Median :2011  Median :  0.0  Mode  :character
Mean   : 459  Mean   : 918  Mean   :1729  Mean   : 988.9
3rd Qu.:  0  3rd Qu.:2011  3rd Qu.:2013  3rd Qu.:2013.0
Max.   :2019  Max.   :2019  Max.   :2018  Max.   :2018.0
```

Hide

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

Hide

```
tbLastClassOM <- table(JustOM$LastClassOM)
tbLastClassOM
```

```
Eight.CSCI400  Five.CSCI306  Four.CSCI341  Four.CSCI358  Nine.CSCI442  One.CSCI101
Six.CSCI406  Three.CSCI262  Two.CSCI261
           2           2           5           6           6           5
3           7           21
```

Hide

```
dfLastClassOM <- as.data.frame(tbLastClassOM)
dfLastClassOM
```

Var1 <fctr>	Freq <int>
Eight.CSCI400	2
Five.CSCI306	2
Four.CSCI341	5
Four.CSCI358	6
Nine.CSCI442	6
One.CSCI101	5
Six.CSCI406	3
Three.CSCI262	7
Two.CSCI261	21
9 rows	

Hide


```
p<-ggplot(data=dfLastClassOM, aes(x= reorder(Var1, Freq), y=Freq)) +
  geom_bar(stat="identity", fill="navyblue") +
  labs(title = "Last Highest CS Course Taken by Other Major Students", x = "Course", y
= "Frequency") +
  ylim(0,45) +
  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
```

