CALIFORNIA STATE UNIVERSITY FULLERTON™

Common Data Set Preparation using R

Nov 9, 2017

CAIR Conference

Alexis Furuichi, Sunny Moon

Cal State Fullerton at a Glance

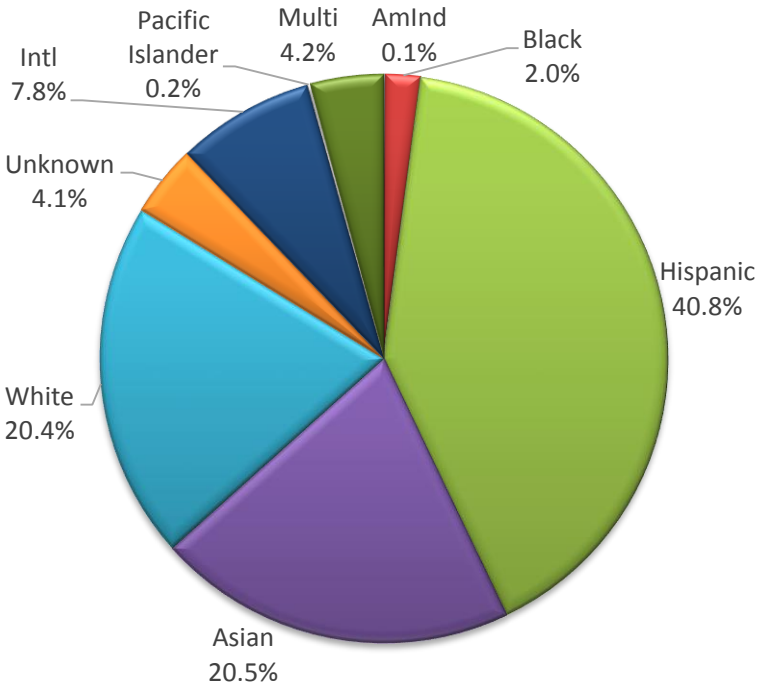
Fall 2017

- 40,439 Students Enrolled/largest in CSU
- 34,800 Undergraduate Students
- 4,437 First-Time Freshmen
- 3,755 New Transfers
- 45% UG Students - Pell Recipients
- 58% UG Students 1st gen college students

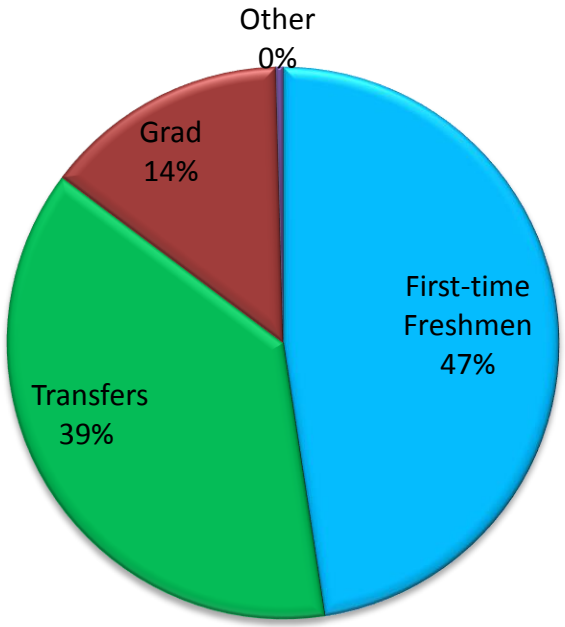
Cal State Fullerton at a Glance

Fall 2017

Ethnic Composition



Entering Characteristics Composition



Cal State Fullerton at a Glance
Fall 2017

- Average High School GPA of First-Time Freshmen : 3.63
- Average SAT of First-Time Freshmen: 1101
- Average Transfer GPA: 3.28
- Average Age of Undergraduates: 22

CALIFORNIA STATE UNIVERSITY, FULLERTON

INSTITUTIONAL RESEARCH AND ANALYTICAL STUDIES

TRENDS IN DEGREES AWARDED BY LEVEL: 1959 TO PRESENT

YEAR	BACHELOR'S DEGREES	MASTER'S DEGREES	DOCTORAL DEGREES	TOTAL DEGREES
2016-2017	8450	2038	42	10530
2015-2016	8397	1868	47	10312
2014-2015	7725	1667	62	9454
2013-2014	7451	1476	59	8986
2012-2013	7472	1566	26	9064
2011-2012	6724	1563	19	8306
2010-2011	6875	1562	11	8448
2009-2010	6481	1394	7	7882
2008-2009	6580	1421		8001
1969-1970	1750	419		2169
1968-1969	1465	297		1762
1967-1968	1182	223		1405
1966-1967	860	166		1026
1965-1966	652	124		776
1964-1965	517	47		564
1963-1964	401	8		409
1962-1963	301	0		301
1961-1962	220	0		220
1960-1961	65	0		65
TOTALS:	**223,610**	**46,932**	**273**	**270,815**

Common Data Set Initiative

- Collaborative Efforts between Publishers and the educational community.
- **Institutional Reporting**
  - Chancellor's Office
  - Federal Reporting
  - Surveys : External Organizations
    - US News & WR, College Board, NSF/GSS,
    - CGS/GRE, Peterson, Barren's, Wintergreen, Princeton Review

About Presenters – Alexis

- Ph.D. Sociology
- Statistician/Project Coordinator
- Evaluation Analyst
- Data & Stat Consultant
- Research Analyst

http://asfuruichi.wixsite.com/furuichi

Evolution of CDS Preparation at CSUF

- SPSS syntax => SPSS output => Excel
- SPSS macro => R tables => LaTex => PDF output
- SPSS macro => R tables => PDF output
- R codes => R tables => PDF output

R packages

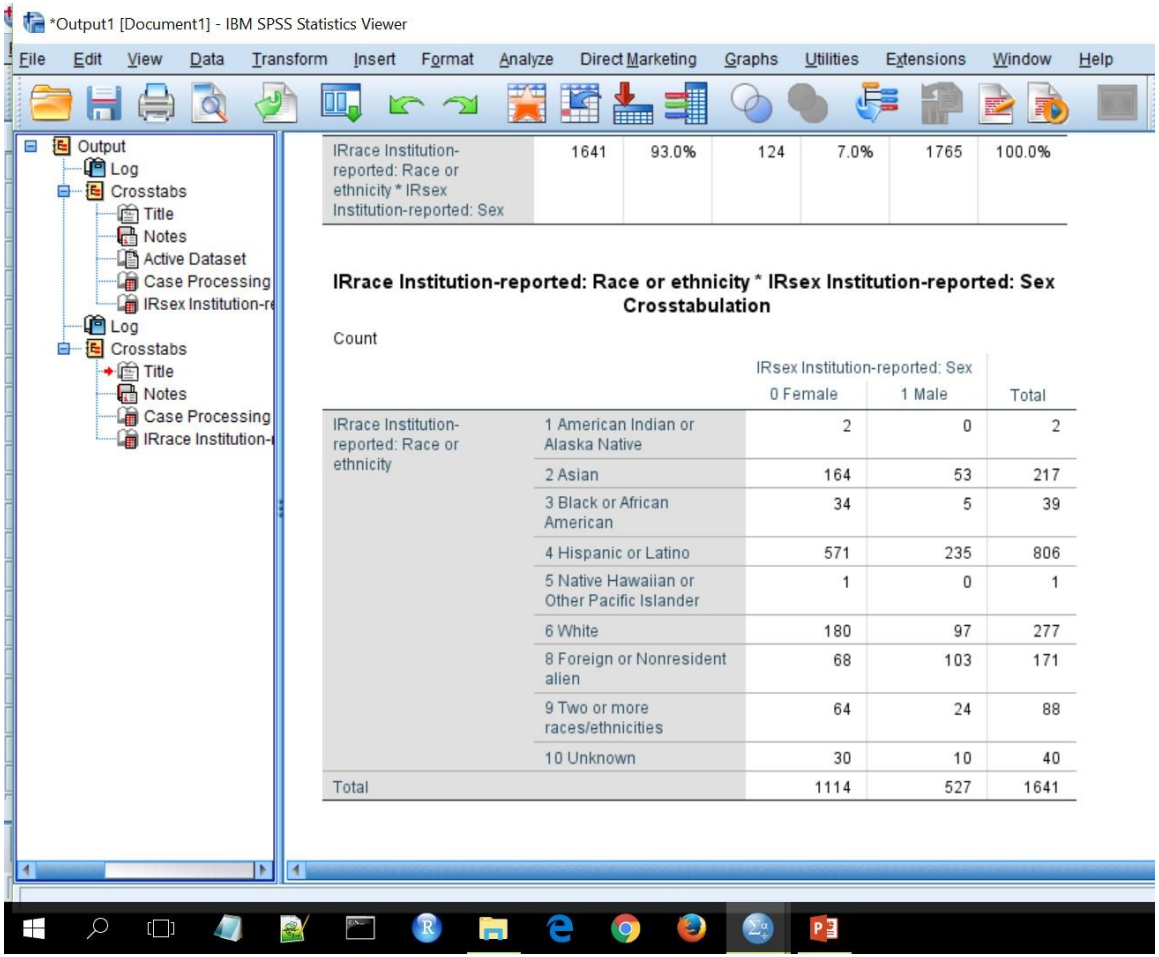
Knitr

ReportRs

SPSS Syntax

```
****C11 file.
dataset activate speronly.
select if ~missing(hsgpa).
execute.
recode hsgpa (3.75 thru highest = 1)(3.50 thru 3.74= 2)(3.25 thru 3.49 =3)
   (3.00 thru 3.24 = 4)(2.50 thru 2.99 = 5)(2.0 thru 2.49 = 6)(1.00 thru 1.99 = 7)(0 thru 0.99 = 8) into hsrank.
execute.
value labels hsrank 1 '3.75 thru highest' 2 '3.50 thru 3.74' 3 '3.25 thru 3.49' 4 '3.00 thru 3.24'
   5 '2.50 thru 2.99' 6 '2.0 thru 2.49' 7 '1.00 thru 1.99' 8 '0.00 thru 0.99'.

dataset declare C11.
aggregate /outfile = C11 /break = hsrank /freq = nu.

dataset activate C11.
save translate /outfile = 'C11.csv' /type = csv /fieldnames /cells = labels /replace.
dataset close C11.

****C12 file.
dataset activate speronly.
dataset declare C12.
aggregate /outfile = C12 /avgGPA = mean(hsgpa).

add files file = C12 /keep = avgGPA.
```

SPSS Output



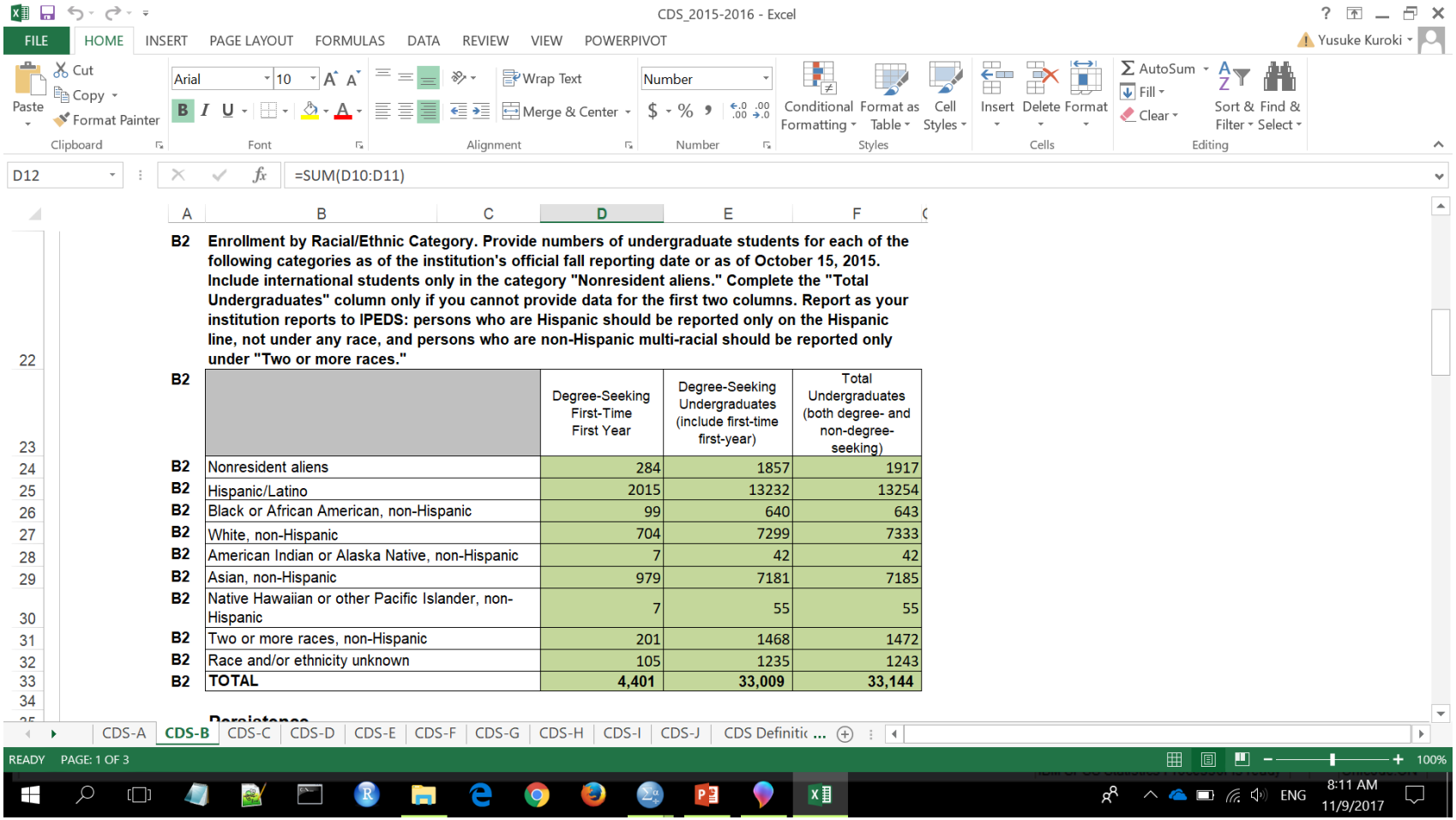
IRrace Institution-reported: Race or ethnicity * IRsex Institution-reported: Sex	1641	93.0%	124	7.0%	1765	100.0%

IRrace Institution-reported: Race or ethnicity * IRsex Institution-reported: Sex Crosstabulation

Count

		IRsex Institution-reported: Sex		
		0 Female	1 Male	Total
IRrace Institution-reported: Race or ethnicity	1 American Indian or Alaska Native	2	0	2
	2 Asian	164	53	217
	3 Black or African American	34	5	39
	4 Hispanic or Latino	571	235	806
	5 Native Hawaiian or Other Pacific Islander	1	0	1
	6 White	180	97	277
	8 Foreign or Nonresident alien	68	103	171
	9 Two or more races/ethnicities	64	24	88
	10 Unknown	30	10	40
Total		1114	527	1641

CDS Excel Template



SPSS Macro

```
5   define !prepCDS (dig2Yr = !cmdend)
6   cd !quote(!concat('M:\sfuruichi\surveys\CDS\20',!dig2Yr,'-2018\output')).
7
```

```
159  ****B1 file.
160  match files file =  !quote(!concat('M:\iras\sper\sper',!dig2Yr,'4.sav')) /in = sperfile
161    /table = !quote(!concat('M:\IRAS\Historical Data\ERSR094',!dig2Yr,'4.sav')) /in= racefile
162    /by ssn
163     /keep = cwid ssn sex age eth eth4mr eth1 eth2 eth3 eth4 eth5 multrace ipedsethnic rescode enstat le
164     degobj stustan acadplan1 units categor5 yrtrm1 race1 to race5 ftpt.
165  dataset name sperrace.
166  execute.
167
168  dataset activate sperrace.
169  if (eth4mr = 4) and (eth = '6') Race5 = 624.
170  if (eth4mr = 4) and (eth = 'H') Race5 = 601
```

```
587  dataset activate I3.
588  save translate /type = csv /outfile = 'I3.csv'.
589  dataset close B1.
590
591  dataset close all.
592
593  !enddefine.
```

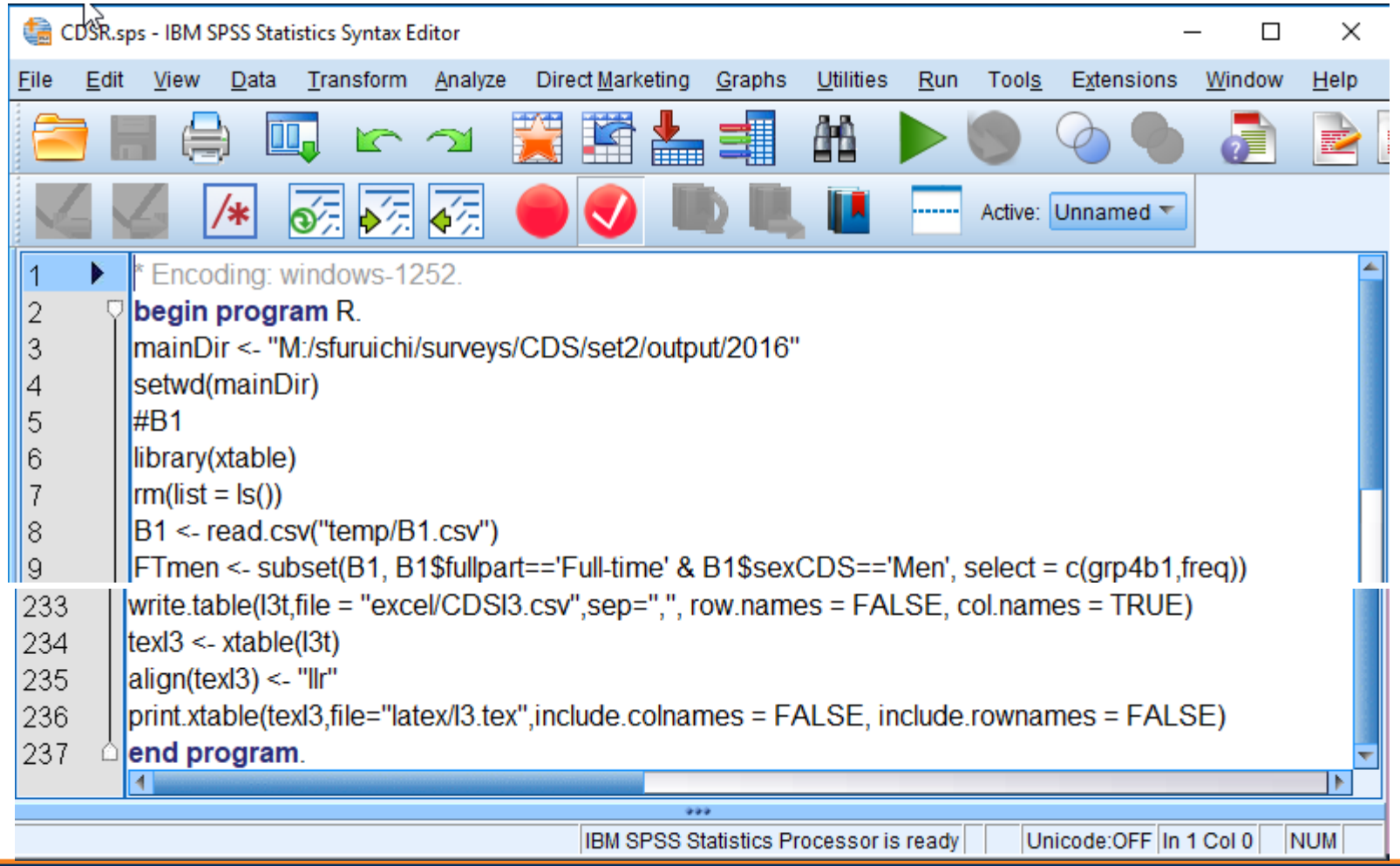
IBM SPSS Statistics Processor is ready | Unicode:OFF | In 593 Col 11 | NUM

Running SPSS Macro

```
16
17  ***create base files.
18  insert file = 'M:\sfuruichi\surveys\CDS\2017-2018\CDS.sps'.
19  !prepCDS dig2Yr = 17.
20
21  insert file =  'M:\sfuruichi\surveys\CDS\set2\program\CDSsecJ.sps'.
22  !prepCDSsecJ dig2Yr = 16.
23
24  insert file =  'M:\sfuruichi\surveys\CDS\set2\program\CDSsecF.sps'.
25  !prepCDSsecF dig2Yr = 16.
26
27  dataset close all.
28
29  ***create outputs.
30  insert file = 'M:\sfuruichi\surveys\CDS\set2\program\CDSR.sps'.
31  insert file = 'M:\sfuruichi\surveys\CDS\set2\program\CDSRsecJ.sps'.
32  insert file = 'M:\sfuruichi\surveys\CDS\set2\program\CDSRsecF.sps'.
33
```

R (3.3.3) in SPSS (v.24)

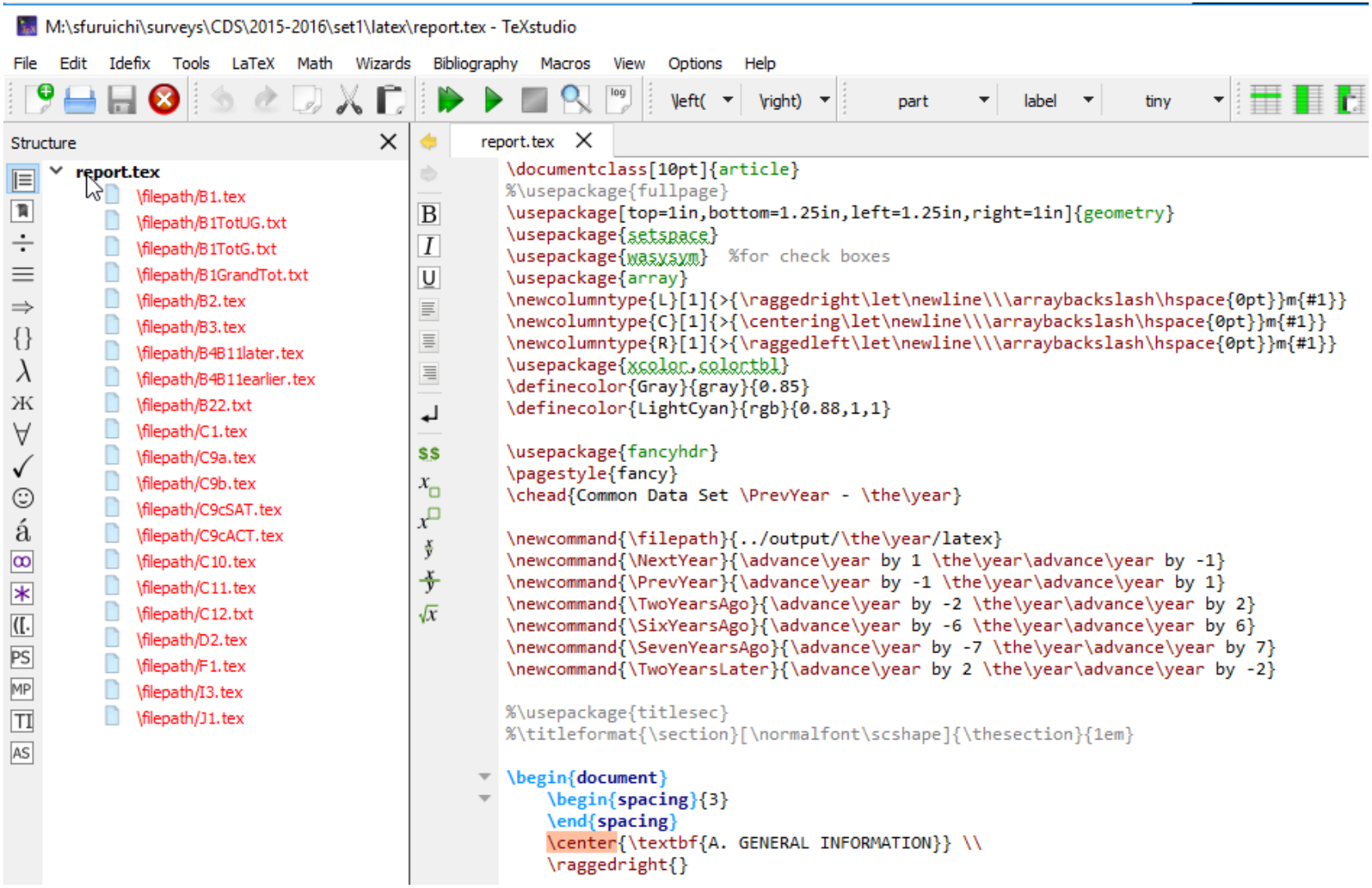
Need to install R Essentials for SPSS



LaTex Codes

```
\center{\textbf{B. ENROLLMENT AND PERSISTENCE}} \\
\raggedright{}
\begin{doublespace}
\textbf{B1. Institutional Enrollment - Men and Women} \\
\end{doublespace}
    Provide numbers of students for each of the following categories as of the inst
    students formerly designated as “first professional” in the graduate cells.\\
    \begin{spacing}{1.5}
    \end{spacing}
    \begin{table}[!htbp]
        \centering
        \input{\filepath/B1.tex} \\
    \end{table}
    \begin{spacing}{1.5}
    \end{spacing}
    \begin{tabular}{lr}
    Total all undergraduates: & \underline{\input{\filepath/B1TotUG.txt}} \\
    Total all graduate: & \underline{\input{\filepath/B1TotG.txt}}  \\
    GRAND TOTAL ALL STUDENTS: & \underline{\input{\filepath/B1GrandTot.txt}}  \\
    \end{tabular}   \\
```

LaTex Preamble



LaTex Output

B. ENROLLMENT AND PERSISTENCE

B1. Institutional Enrollment - Men and Women

Provide numbers of students for each of the following categories as of the institutions official fall reporting date or as of October 15, 2016. Note: Report students formerly designated as first professional in the graduate cells.

	FTmen	FTwomen	PTmen	PTwomen
First-time Fresh	1873	2414	63	51
Other Fresh	853	924	71	63
Other Deg	9160	11812	2724	3001
Total Deg	11886	15150	2858	3115
Other UG	28	41	18	48
Total UG	11914	15191	2876	3163
FT Grad	428	776	419	550
Other Grad	705	843	991	1080
Other cred			5	7
Total Grad	1133	1619	1415	1637

Total all undergraduates:	33144
Total all graduate:	5804
GRAND TOTAL ALL STUDENTS:	38948

CDS Output Components

- Headers
- Texts
- Tables
- In-text numbers

Headers

Document Title

B. ENROLLMENT AND PERSISTENCE

Institutional Effectiveness

Fall 2016

Page Header

Headings

Graduation Rates

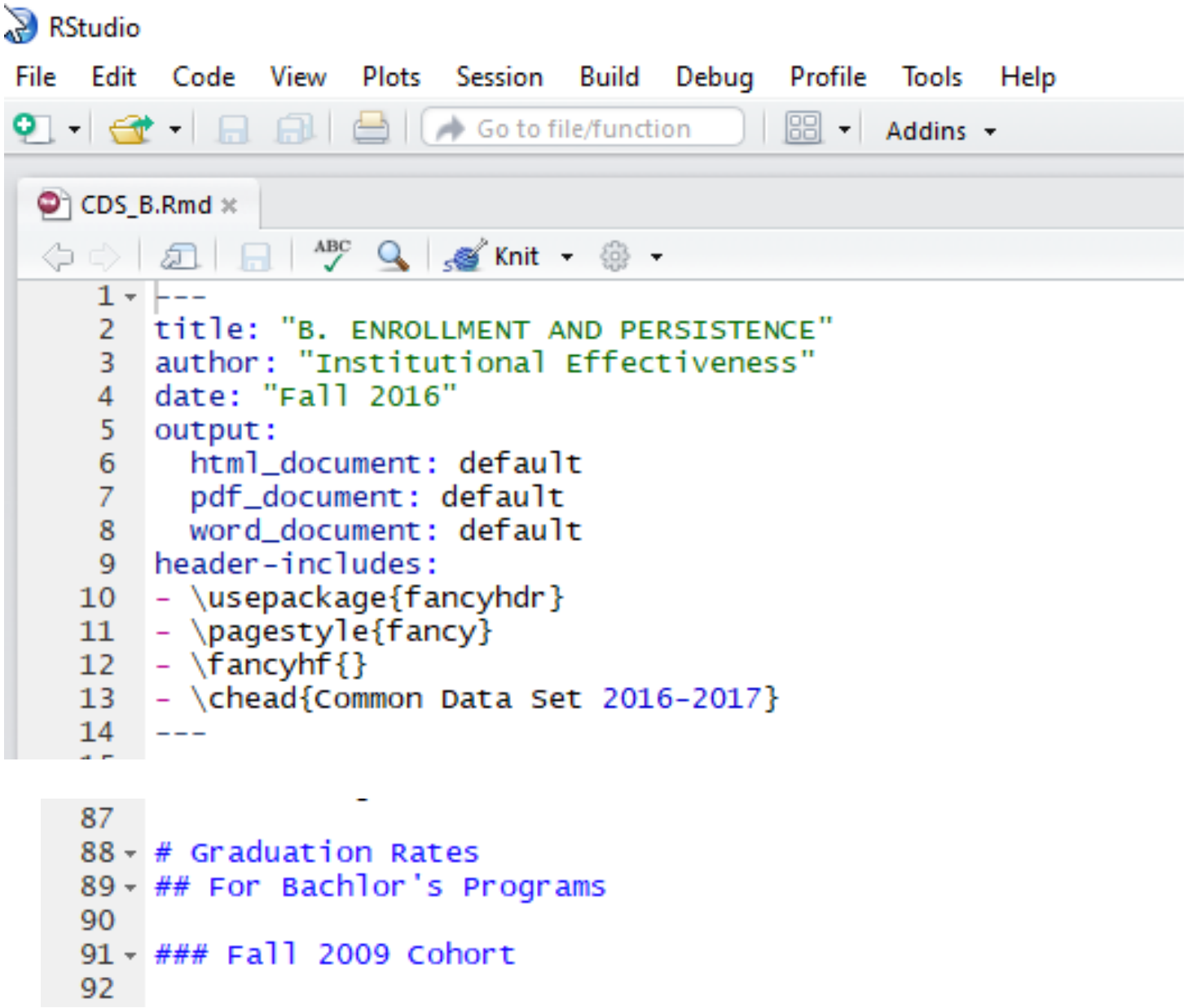
For Bachlor's Programs

Fall 2009 Cohort

Document Title

Page Header

Headings



B1. Institutional Enrollment - Men and Women

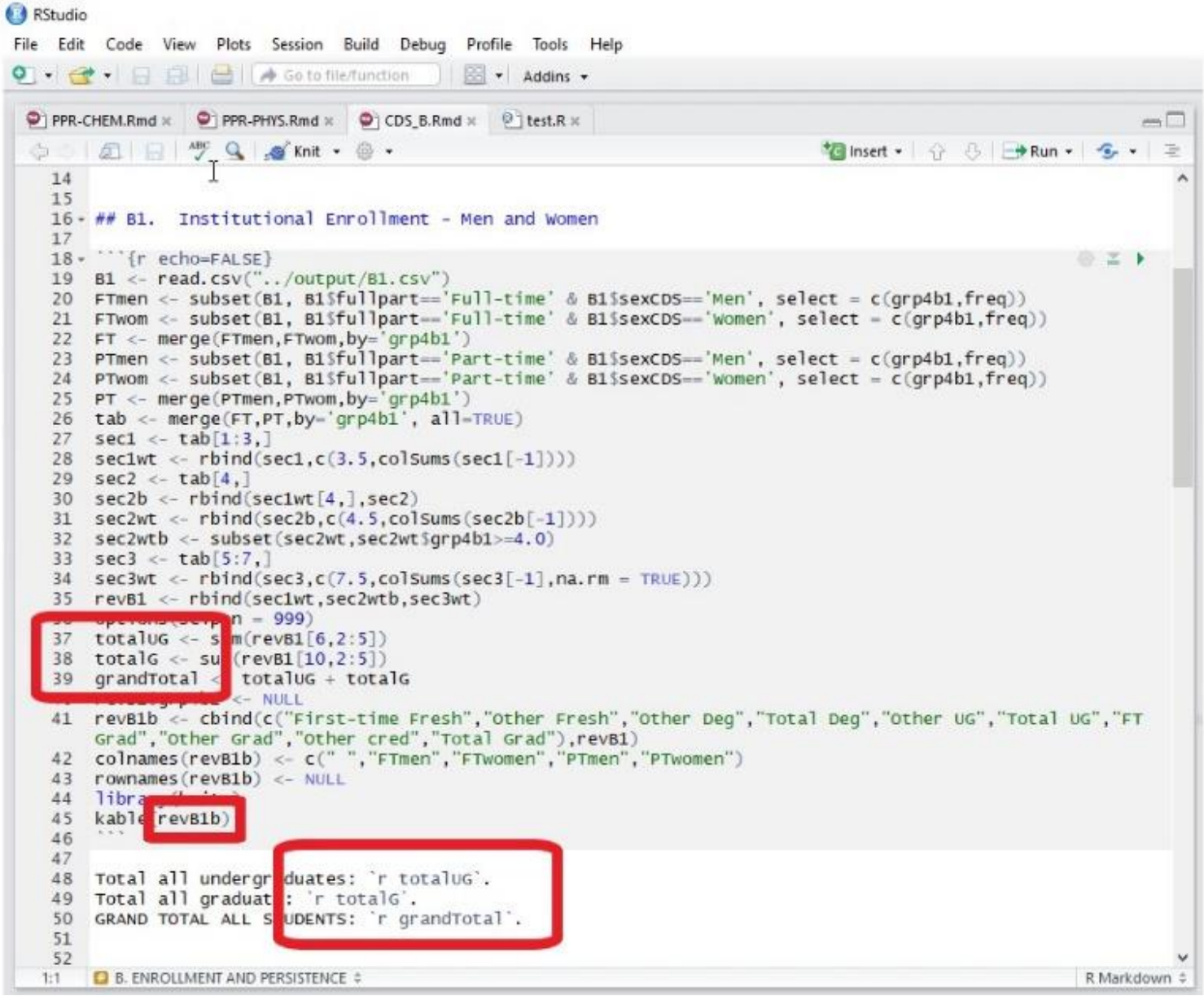
Table

X.	FTmen	FTwomen	PTmen	PTwomen
First-time Fresh	1764	2510	68	84
Other Fresh	759	879	80	60
Other Deg	9670	12343	2941	3258
Total Deg	12193	15732	3089	3402
Other UG	20	49	30	61
Total UG	12213	15781	3119	3463
FT Grad	353	792	382	527
Other Grad	601	867	956	1177
Other cred	NA	NA	1	3
Total Grad	954	1659	1339	1707

Total all undergraduates: 34576.
Total all graduate: 5659.
GRAND TOTAL ALL STUDENTS: 40235.

In-text numbers

Texts


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Addins
PPR-CHEM.Rmd
PPR-PHYS.Rmd
CDS_B.Rmd
test.R
Knit
Insert
Run
B1. Institutional Enrollment - Men and Women
```{r echo=FALSE}
B1 <- read.csv("../output/B1.csv")
FTmen <- subset(B1, B1$fullpart=='Full-time' & B1$sexCDS=='Men', select = c(grp4b1,freq))
FTwom <- subset(B1, B1$fullpart=='Full-time' & B1$sexCDS=='Women', select = c(grp4b1,freq))
FT <- merge(FTmen,FTwom,by='grp4b1')
PTmen <- subset(B1, B1$fullpart=='Part-time' & B1$sexCDS=='Men', select = c(grp4b1,freq))
PTwom <- subset(B1, B1$fullpart=='Part-time' & B1$sexCDS=='Women', select = c(grp4b1,freq))
PT <- merge(PTmen,PTwom,by='grp4b1')
tab <- merge(FT,PT,by='grp4b1', all=TRUE)
sec1 <- tab[1:3,]
sec1wt <- rbind(sec1,c(3.5,colSums(sec1[-1])))
sec2 <- tab[4,]
sec2b <- rbind(sec1wt[4,],sec2)
sec2wt <- rbind(sec2b,c(4.5,colSums(sec2b[-1])))
sec2wtb <- subset(sec2wt,sec2wt$grp4b1>=4.0)
sec3 <- tab[5:7,]
sec3wt <- rbind(sec3,c(7.5,colSums(sec3[-1],na.rm = TRUE)))
revB1 <- rbind(sec1wt,sec2wtb,sec3wt)
totalUG <- sum(revB1[6,2:5])
totalG <- sum(revB1[10,2:5])
grandTotal <- totalUG + totalG
revB1b <- cbind(c("First-time Fresh","Other Fresh","Other Deg","Total Deg","Other UG","Total UG","FT Grad","Other Grad","Other cred","Total Grad"),revB1)
colnames(revB1b) <- c(" ","FTmen","FTwomen","PTmen","PTwomen")
rownames(revB1b) <- NULL
kable(revB1b)
```
Total all undergraduates: `r totalUG`.
Total all graduate: `r totalG`.
GRAND TOTAL ALL STUDENTS: `r grandTotal`.
B. ENROLLMENT AND PERSISTENCE
R Markdown

Another Example - PPR

7. Appendices to the Self-Study

Upon request, the Office of Institutional Research and Analytical Studies will provide the data for Tables 1-9 that you will need for your review and analysis. The completed tables should be placed in the appendix, and the narrative and analyses should be woven into the self-study itself.

APPENDIX I. UNDERGRADUATE DEGREE PROGRAMS

TABLE 1. Undergraduate Program Applications, Admissions, and Enrollments
For each undergraduate degree program, a table will be provided with the number of student applications, number of students admitted, percent admitted, the number of new enrollments, and the percentage of new enrollments. Percentage of students enrolled is the number of students enrolled divided by the number of students admitted or the yield rate.

TABLE 1-A. First-time Freshmen: Program Applications, Admissions, and Enrollments

Academic Year	# Applied	# Admitted	% Admitted	# Enrolled	% Enrolled
2012-2013					
2013-2014					
2014-2015					
2015-2016					
2016-2017					

TABLE 1-B. Upper Division Transfers: Program Applications, Admissions, and Enrollments

Academic Year	# Applied	# Admitted	% Admitted	# Enrolled	% Enrolled
2012-2013					
2013-2014					
2014-2015					
2015-2016					
2016-2017					

PPR Tables Physics (BA, BS, MS)

Office of Assessment and Institutional Effectiveness

Fall 2017

APPENDIX I. UNDERGRADUATE DEGREE PROGRAMS

TABLE 1. Undergraduate Program Applications, Admissions, and Enrollments

TABLE 1-A. First-time Freshmen: Program Applications, Admissions, and Enrollments

AY	applied	admitted	pctAdmitted	enrolled	pctEnrolled
2012-2013	95	83	87.4	16	19.3
2013-2014	132	99	75.0	12	12.1
2014-2015	109	77	70.6	8	10.4
2015-2016	121	93	76.9	17	18.3
2016-2017	158	107	67.7	14	13.1

TABLE 1-B. Upper Division Transfers: Program Applications, Admissions, and Enrollments

AY	applied	admitted	pctAdmitted	enrolled	pctEnrolled
2012-2013	25	17	68.0	1	5.88
2013-2014	33	11	33.3	3	27.27
2014-2015	39	21	53.8	7	33.33
2015-2016	37	21	56.8	6	28.57
2016-2017	46	31	67.4	12	38.71

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

PPR-NURS.Rmd | PPR-PHYS.Rmd*

Knit | Insert | Run

```
---
title: "PPR Tables Physics (BA, BS, MS)"
author: "Office of Assessment and Institutional Effectiveness"
date: "Fall 2017"
output:
  word_document:
    reference_docx: word-styles-reference-01.docx
---

# APPENDIX I. UNDERGRADUATE DEGREE PROGRAMS
## TABLE 1. Undergraduate Program Applications, Admissions, and Enrollments
### TABLE 1-A. First-time Freshmen: Program Applications, Admissions, and Enrollments

```{r echo=FALSE}
t1Aadm <- read.csv("T1AappliedAdmitted.csv")
t1Aenr <- read.csv("T1Aenrolled.csv")
t1A <- merge(t1Aadm,t1Aenr,by="AY", all = TRUE)
t1A$pctEnrolled <- format((t1A$enrolled/t1A$admitted)*100,digits=2,nsmall=1)
t1A$pctAdmitted <- format((t1A$admitted/t1A$applied)*100,digits=3,nsmall=1)
library(knitr)
kable(t1A)
```


TABLE 1-B. Upper Division Transfers: Program Applications, Admissions, and Enrollments

```{r echo=FALSE}
t1Badm <- read.csv("T1BappliedAdmitted.csv")
t1Benr <- read.csv("T1Benrolled.csv")
t1B <- merge(t1Badm,t1Benr,by="AY", all=TRUE)
t1B$pctEnrolled <- format((t1B$enrolled/t1B$admitted)*100,digits=2,nsmall=1)
t1B$pctAdmitted <- format((t1B$admitted/t1B$applied)*100,digits=3,nsmall=1)
library(knitr)
kable(t1B)
```


TABLE 2. Undergraduate Program Enrollment in FTES

TABLE 2-A. Undergraduate Program Enrollment in FTES

```{r echo=FALSE}
t2A <- read.csv("T2A.csv")
is.num <- sapply(t2A, is.numeric)
```
```

10:1 | # APPENDIX I. UNDERGRADUATE DEGREE PROGRAMS | R Markdown

Thank you!

http://www.fullerton.edu/analyticalstudies/
Sunny Moon: hmoon@fullerton.edu
Alexis Furuichi: sfuruichi@fullerton.edu
http://asfuruichi.wixsite.com/furuichi