Stat 696, Example Application of `knitr`

Kelso Quan

September 23, 2018

```
summary(fm1)

##
## Call:
## lm(formula = Apps ~ Private + Elite + Accept + Outstate + Room.Board +
##     Grad.Rate, data = College)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5094.5  -329.7   -22.6   226.8 10114.6
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -985.95379  204.82380  -4.814 1.78e-06 ***
## PrivateYes  -291.62328  133.28335  -2.188 0.028970 *
## EliteYes    1745.00184  151.74052  11.500  < 2e-16 ***
## Accept         1.42869    0.02024  70.601  < 2e-16 ***
## Outstate      -0.01427    0.01690  -0.845 0.398600
## Room.Board     0.16615    0.04953   3.355 0.000834 ***
## Grad.Rate      8.63483    2.94718   2.930 0.003491 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1128 on 770 degrees of freedom
## Multiple R-squared:  0.9157,Adjusted R-squared:  0.915
## F-statistic:  1394 on 6 and 770 DF,  p-value: < 2.2e-16
```

The new predicted values: newpred1: 7000, 4800, 9300 and newpred2: 2300, 57, 4600 from the data. There is a comparison of every variable correlation to others in figure 1.

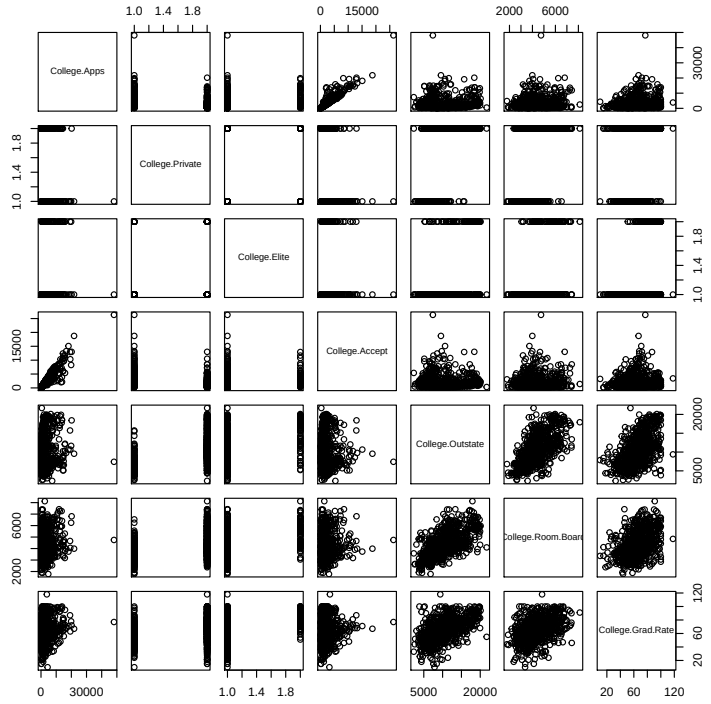

Figure 1: Matrix of variables

Table 1: Summary statistics for the ISLR College data set.

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Applicants	3,001.638	3,870.201	81	776	3,624	48,094
TotalAccepted	2,018.804	2,451.114	72	604	2,424	26,330
TotalEnrolled	779.973	929.176	35	242	902	6,392
Top10Top25FullTimeStudent	3,699.907	4,850.421	139	992	4,005	31,643
PartTimeStudent	855.299	1,522.432	1	95	967	21,836
OutofStateTution	10,440.670	4,023.016	2,340	7,320	12,925	21,700
RoomBoardPrices	4,357.526	1,096.696	1,780	3,597	5,050	8,124
BookPrices	549.381	165.105	96	470	600	2,340
PersonalExpense	1,340.642	677.071	250	850	1,700	6,800
S/F Ratio	14.090	3.958	2.500	11.500	16.500	39.800
budgetPerStudent	9,660.171	5,221.768	3,186	6,751	10,830	56,233
Grad.Rate	65.463	17.178	10	53	78	118

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-985.95	204.82	-4.81	0.00
PrivateYes	-291.62	133.28	-2.19	0.03
EliteYes	1745.00	151.74	11.50	0.00
Accept	1.43	0.02	70.60	0.00
Outstate	-0.01	0.02	-0.84	0.40
Room.Board	0.17	0.05	3.35	0.00
Grad.Rate	8.63	2.95	2.93	0.00

Table 2: Inferences from regressing number of applications on whether the college is private or public, whether the college is elite or not, acceptance rate, out of state tuition, room and board, and graduation rate.

Appendix A: Supplementary Plots

Figure 2 shows us the residuals.

```r
par(mfrow=c(2,2))
plot(fm1)
```

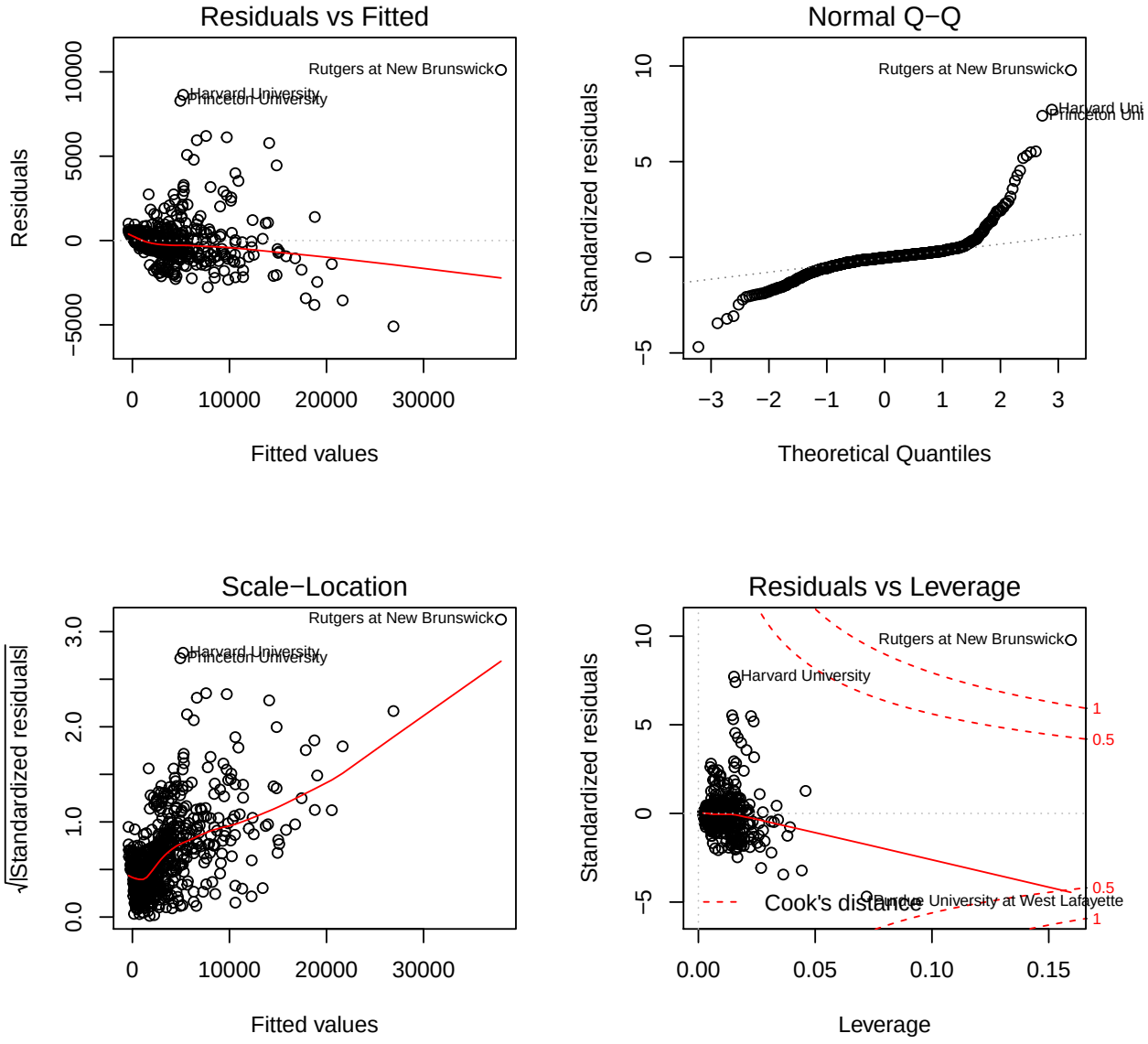

Figure 2: Supplemental figures

Appendix B: R Code

```r
% Stat 696: Knitr lab
% Illustrating knitr to present analyses of College data from ISLR text
% Packages required: knitr, xtable, stargazer, ISLR
% To use, make sure to call library(knitr) first in console
% To run and create a .tex file: knit('knitr_ClassVersion.Rnw') in R
```

```
% August 25, 2017


% To show at start of class:
% -- Step through preface briefly and show where to enter name
% -- LaTeX preface (preface.tex) vs. Knitr preface (knitr_ClassVersion.Rnw);
%    briefly delineate two approaches to report writing
% -- Code in place for regression analysis and prediction
% -- There are 6 tasks to complete with sample code and hints where needed
% -- Suggest cutting an pasting R code into console first to debug


% Preface required in the knitr RnW file
\documentclass{article}

\usepackage{rotating}
\usepackage{graphics}
\usepackage{latexsym}
\usepackage{color}
\usepackage{listings} % allows for importing code scripts into the tex file

% Approximately 1 inch borders all around
\setlength\topmargin{-.56in}
\setlength\evensidemargin{0in}
\setlength\oddsidemargin{0in}
\setlength\textwidth{6.49in}
\setlength\textheight{8.6in}

% Options for code listing; from Patrick DeJesus, October 2016
\definecolor{codegreen}{rgb}{0,0.6,0}
\definecolor{codegray}{rgb}{0.5,0.5,0.5}
\definecolor{codepurple}{rgb}{0.58,0,0.82}
\definecolor{backcolour}{rgb}{0.95,0.95,0.92}
\lstdefinestyle{mystyle}{
  backgroundcolor=\color{backcolour},   commentstyle=\color{codegreen},
  keywordstyle=\color{magenta},
  numberstyle=\tiny\color{codegray},
  stringstyle=\color{codepurple},
  basicstyle=\footnotesize,
  breakatwhitespace=false,
  breaklines=true,
  captionpos=b,
  keepspaces=true,
  numbers=left,
  numbersep=5pt,
  showspaces=false,
  showstringspaces=false,
  showtabs=false,
  tabsize=2
}
%"mystyle" code listing set
\lstset{style=mystyle}
%\lstset{inputpath=appendix/}


\title{Stat 696, Example Application of \texttt{knitr}}
\author{Kelso Quan}
\date{\today}

\begin{document}
\maketitle

% Code to start knitr
<<include=FALSE>>=
  library(knitr)
```

```
opts_chunk$set(
  concordance=TRUE
)
@

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code snippet to load in libraries and data
% THIS IS HOW R-CODE IS READ INTO LaTeX DOC WITH knitr
% Environment:
% <<...>>=
% [Code here]
% @
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
<<load data, include=FALSE>>=
# Load in libraries, load in data and set up variables
library(ISLR)
library(stargazer)
library(xtable)

rm(list=ls(all=TRUE)) # remove all previous objects from memory

# Set up data for the illustration
# For illustration purposes we will use the College data set from the ISLR text
# Create an indicator of Elite College status (see exercise 8 in Ch. 2 of ISLR text)
Elite=rep("No",nrow(College))
Elite[College$Top10perc >50]="Yes"
Elite=as.factor(Elite)
College=data.frame(College ,Elite)
numvars = length(College) # number of variables in the College data set
n = dim(College)[1]
@


% Code snippet to run the regression analysis, including prediction for two new
    Universities
<<college apps regression, echo=FALSE>>=
# Fit a model
fm1 = lm(Apps~Private+Elite+Accept+Outstate+Room.Board+Grad.Rate, data = College)
# predcit a new school not in the data set
new1 = data.frame(Private="No", Elite="No", Accept=5000, Outstate=8000, Room.Board=6000,
    Grad.Rate=0.6)
newpred1 = signif(predict(fm1, new1, interval="prediction"), 2)
new2 = data.frame(Private="Yes", Elite="Yes", Accept=1000, Outstate=16000, Room.Board=4000,
     Grad.Rate=0.90)
newpred2 = signif(predict(fm1, new2, interval="prediction"), 2)
@


%%%%%%%%%%%%
% Lab Tasks
%%%%%%%%%%%%

%% Task 1: Present an R dump of the summary of the regression fit
% [Place knitr code chunk here]
<<R dump>>=
summary(fm1)
@
% \Sexpr{...} allows us to run R-code or grab R elements inside the text.
% Try it out!  Write a sentence using \Sexpr to grab the predicted values for
% the new schools (variables newpred1 and newpred2 from above).
The new predicted values: newpred1: \Sexpr{newpred1} and newpred2: \Sexpr{newpred2} from
    the data.

%% Task 2: Insert a pairwise scatterplot into your document
% For plots, start by setting up the LaTeX figure environment,
```

```
% then place R code to knit, then set up LaTeX code to complete figure environment.
% Below I give the code for this task.  You will practice with this code in Task 5.
\begin{figure}
\begin{center}
<<echo=FALSE, out.width='4in'>>=
y = data.frame(College$Apps, College$Private, College$Elite, College$Accept, College$Outstate, College$Room.Board, College$Grad.Rate)
pairs(y)
@
\caption{Matrix of variables}
\label{matrix}
\end{center}
\end{figure}

% Write a short blurb of text to cite your figure.
There is a comparison of every variable correlation to others in figure \ref{matrix}.

%% Task 3: Use stargazer to present summary statistics of the College data set
<<descrips, r, results="asis", echo=FALSE>>=
##############
# Recall the stargazer table created in the online video:
#  stargazer(College, title="Summary statistics for the ISLR College data set.", label="descrips")
# Now try the following options to re-format your table:
#  summary.stat option: use to get rid of the sample size N column and add 25th and 75th percentiles
#  covariate.labels option: present more informative variable labels, rather than variable names
#  float.env option: force LaTeX to present table in landscape (sidewaystable)
##############
stargazer(College, title="Summary statistics for the ISLR College data set.",
          label="descrips", omit.summary.stat = "n",
          iqr = T,
          float = T,
          float.env = "sidewaystable",
          covariate.labels = c("Applicants", "TotalAccepted", "TotalEnrolled", "Top10%HSstudents",
                               "Top25%HSstudents", "FullTimeStudent", "PartTimeStudent", "OutofStateTution",
                               "RoomBoardPrices", "BookPrices", "PersonalExpense", "%Facw/PhD",
                               "%Facw/TerminalDeg", "S/F Ratio", "%AlumniDonate",
                               "budgetPerStudent", "Grad.Rate")
          )
##############

# For the curious: check out the stargazer help screen for the plethora of options!
# For example, stargazer can present a regression inferences table.  Try the following code:
#   stargazer(fm1, title="Regression inferences", dep.var.labels="Number of Applications", label="inf",
#          keep.stat="n", ci=TRUE, ci.level=0.95, single.row=TRUE,
#          covariate.labels=c("Private", "Elite", "Number Accept", "Out-of-state tuition",
#                             "Room and board", "Grad rate"))
@


%% Task 4: Create a table of predictions using xtable
% I provide the code below for a base table.
% The task is then to add additional columns to the table and create the LaTeX code using xtable.
% Note that we use results="asis" to force knitr to present the table code for compiling in LaTeX
<<predictions, results="asis", echo=FALSE>>=
# create the table and store in 'x'
```

```
univ = rbind("University 1", "University 2")
elite = rbind("No", "Yes")
gradrate = rbind(new1[,6], new2[,6])
preds = rbind(newpred1[,1], newpred2[,1])
lwr = rbind(newpred1[,2], newpred2[,2])
upr = rbind(newpred1[,3], newpred2[,3])
# Tasks:
# 1) Add columns of out of state tuition, elite status, and graduation rate to the table.
# 2) Use xtable to create and output LaTeX code for the table.  Here is the code from the online video:
#  fm.table = xtable(fm1, digits=2,
#                    caption="Inferences from regressing number of applications on whether the college is private or public,
#                    whether the college is elite or not, acceptance rate, out of state tuition, room and board, and
#                    graduation rate.",
#                    label="reginf")
#  align(fm.table) <- "|l|rrrr|"  # place vertical lines on left and right of table, and after first column of var names
#  print(fm.table)
# Note: consider the include.rownames option in the print command to remove row names.
###############
fm.table = xtable(fm1, digits=2,
                  caption="Inferences from regressing number of applications on whether the college is private or public, whether the college is elite or not, acceptance rate, out of state tuition, room and board, and graduation rate.",
                  label="reginf")
  align(fm.table) <- "|l|rrrr|"  # place vertical lines on left and right of table, and after first column of var names
  print(fm.table)
@

% Write a short blurb citing your table.


%% Task 5: Create an appendix of plots
% We will create a 2x2 graphic of regression diagnostics
\newpage
\noindent \Large{{\bf Appendix A: Supplementary Plots}}
\begin{figure}[h!]
\begin{center}
%%%%%%%%%%%%%%%%
% Here is code for the default regression diagnostics from R
% Write knitr code to present a 2x2 graphic for this appendix.
% Suggestion: use the knitr code environment from the scatterplot matrix of Task 2
%
%  par(mfrow=c(2,2))
%  plot(fm1)
%
%%%%%%%%%%%%%%%%%%%
<<graphics>>=
  par(mfrow=c(2,2))
  plot(fm1)
  @
\caption{Supplemental figures}
\label{figures}
\end{center}
\end{figure}

% Write a short blurb citing the figure and stating what it is.
Figure \ref{figures} shows us the residuals.

%% Task 6: Create an appendix of code
% Here is the LaTeX code from the online video.
% Recall that this is straight LaTeX, no knitr code chunk needed!
```

```
  \newpage
  \noindent \Large{{\bf Appendix B: R Code}}
  \lstinputlisting[language=R, caption = Knitr Lab]{knitr_lab.Rnw}


\end{document}
```

Listing 1: Knitr Lab