Canisius
COLLEGE
Where leaders are made

Introduction to R and graphical presentation

(DA500, Summer 2017)

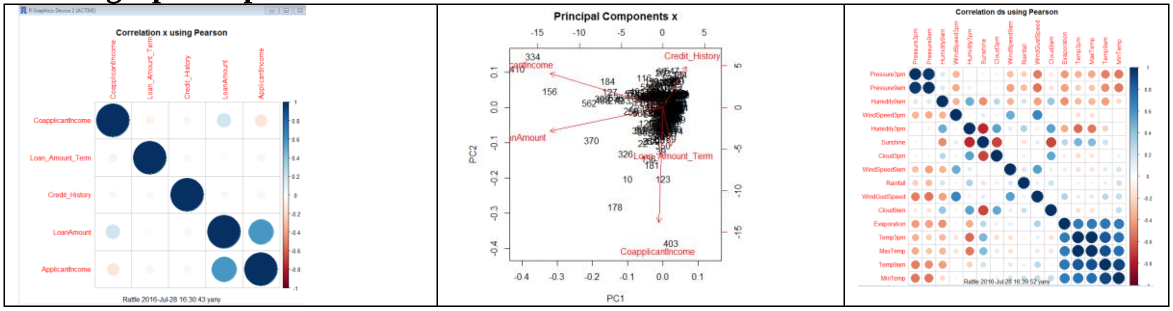
Instructor:	Paul Yan
Contact Information:	yany@canisius.edu, (716) 888-2604
Classroom:	TBA (A computer lab)
Office hours:	TBA
Prerequisites:	None
Textbook:	Financial Modeling using R, Yuxing Yan, 2016, https://www.tatepublishing.com/bookstore/book.php?w=9781681875309
Websites:	*An Introduction to R* http://canisius.edu/~yany/doc/R-intro.pdf *The R Language Definition* http://canisius.edu/~yany/doc/R-lang.pdf *My related R web site:* http://canisius.edu/~yany/R.shtml
QR codes for two more web sites	https://www.r-project.org/ http://www3.canisius.edu/~yany/R.shtml Fun question: how many lines of R codes to genreate one image?
One-line R codes	*> source("http://canisius.edu/~yany/intro2R.R")* Note: I will explain this line in week 2
Format	This is a one-credit summer course. In total, we have 10 lectures conducted in a computer lab. Hands-on is critical.
Course Description:	Students would learn how to install R, how to launch and quit R. In a sense, we assume that students know nothing about R. Then, we learn simple R concepts, such as R is case sensitive, how to assign a value to variables, how to list and remove those variables. Since manipulating data is very important, we learn how to input data from various external sources, such as text file, csv file or from an internet file. We learn simple methods to clean, manipulate data. After that, we learn how to output our data for future usage or analysis. In addition, we learn simple text analysis and machine learning. (for this part, students would use my programs). To have fun: such as get population, racial composition, mean salary, real estate prices for a given zip code, find a 9 among hundreds of 6, how to generate a QR code, randomly call a student and

	show his/her photo, visualize the top 10 most frequent words by Obama in his 2007 speech.
Several objectives:	1) Learn basic R concepts and functionalities 2) Learn simple R functions 3) Learn how to retrieve data from various sources 4) Learn simple methods to clean data 5) learn simple text analysis and machine learning 6) have fun
Capacity of the class	20 (ideally the number of students should be less than 15 since hands-on is critical)
R Software	R is open source statistical and computational software https://www.r-project.org/
Attendance Policy:	Attending classes regularly is required. Before-class preparation and in-class participation is an integral part of this course. Students are strongly encouraged to participate in class discussions and ask questions. Students are encouraged to discuss current events relevant to this course or their own experiences. Homework problems are regularly assigned.
Course Level Learning Goals:	Can use R to write simple functions, such as pv_f(), fv_f() etc. Know loops and conditions Understand how to input and output data Know how to view input data sets, such as size, dimension, sample statistics, missing values, number of unique values Know how to conduct simple data manipulation, such as remove missing value, choose a subset, merge different data sets Can draw basic graphs and more advanced graphs
College, Program and Major Learning Goals:	This course is designed to help students achieve one or more College Core, Business Program and/or Major level learning goals and objectives. You can see the specific College, Program or Major level learning goals and objectives associated with the course from this page on the College website: http://bit.ly/bcoreLG
Grade Evaluation:	Data cases 50% 2 small Quizzes 30% Class participation 20% --- Total 100%

Tentative schedule

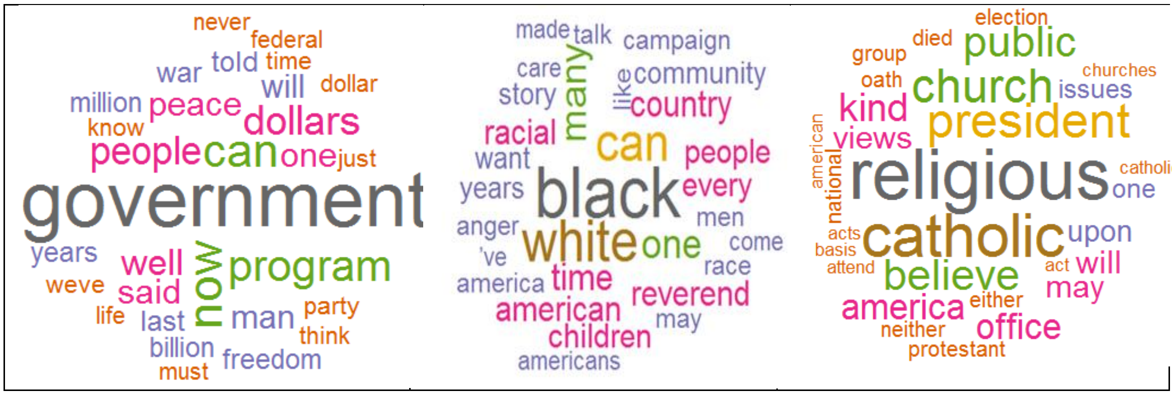
#	Date	Topics	Description
1	7/5	Chapter 17: R basics Generate your simple R functions	R installation, starting and quitting R, case sensitive, variable assignment, one command per line vs. multi commands per line, locations of various help documents Simplest one-line functions dd() to duble any input values, Format of a typical R function, pv_f(), pv_annuit() etc. One line comment vs. multiple line comments , help(pv_f)
	7/7	Chapter 18: Simple value assignment	Several ways to assign a value to a variable, viewing objectives using the ls() function, seq() function, position and keyword approach for a set of input variables
2		Chapter 19: inputting data from external files (sources) Chapter 20: simple data manipulation	Read data from a text file (read.table() , read.csv(), input data from clipboard, fixed width file (read.fwf(), read from the internet, generate and read a R data set (.rda, .RData and .rds), readRDS() Several R functions: head(), tail(), ls(), ls(pattern='my_pattern'), length(), dim(), summary(), rbind(), cbind(), matrix manipulation, etc.
		Chapter 31: introduction to R packages	Three ways to install an R package, loaded vs. preinstalled packages, using .libPaths() function, find manual for a specific R package, help(package=XML), list of all R packages, R view
3		Chapter 21: R loops Chapter 22: if-else, logic OR and logic AND	For loop, double loops, how to stop an execution, while loop, if() function, if-else function, if() and stop() pair, logic Or and logic AND
		Chapter 23: outputting to a file	Several functions: write.table(), write.csv(), save(), saveRDS(), append data to an existing data file, writing a binary file, the sink() function
4		Chapter 29: Simple Plot and Graph	Plot for a single graph, adding labels on horizontal and vertical axis, shading certain area, putting several graphs together, add Greek letters, saving a PDF file, output a high-resolution images, overlapping graphs.
		More advanced plot and graphics	ggplot2, ggcorrplot, gemnet, add trend, add text, add colors, add titles, add labels, etc. , see Appendices D, E and F.
5		Introduction to text analysis	What is text analysis? Most frequently used words, visual presentation (see Appendix A)
		Introduction to machine learning	Introduction rattle() R package, R view for machine learning, several examples, R package called rattle(), see Appendix G

Appendix A: graphical presentations of correlations

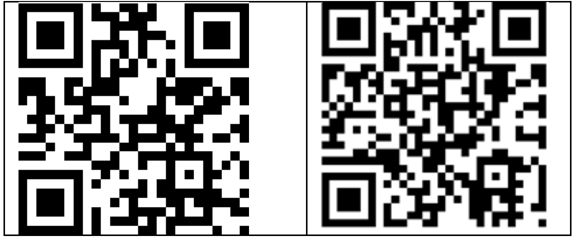


Appendix B: example of text analysis

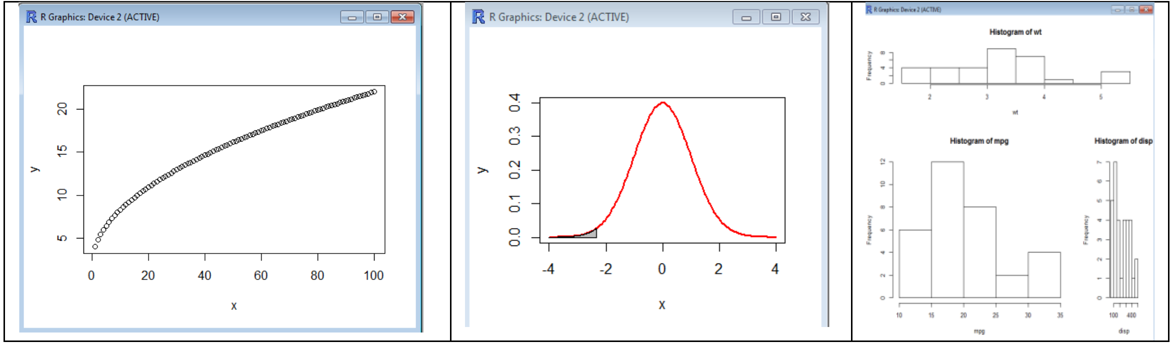
Based on Kennedy (1960), Reagan (1964) and Obama (2008) speeches, we could have the following images. In the images, the sizes of the words depend on their frequency. The higher is the frequency, the bigger is its font. For example, in the leftist image, the word "Government" is the first frequently used word. Could you guess their belonging?



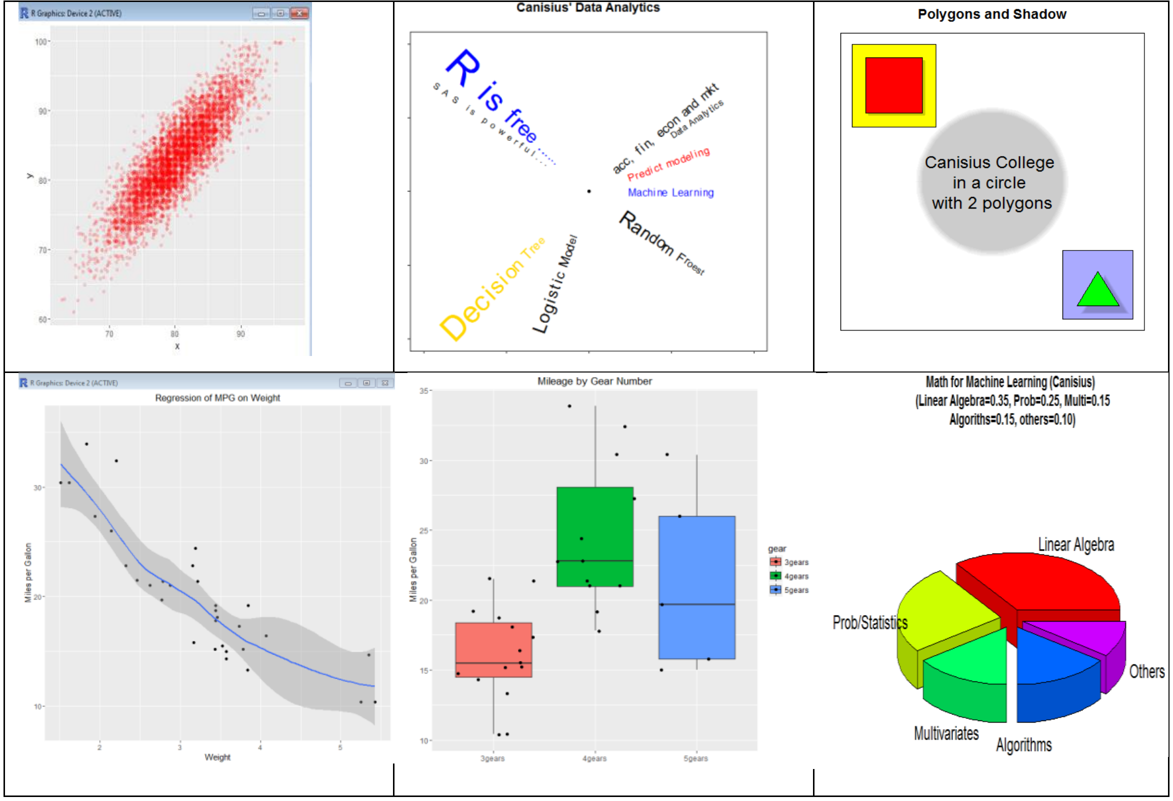
Appendix C: using 3 lines of R codes to generate the following two QR codes



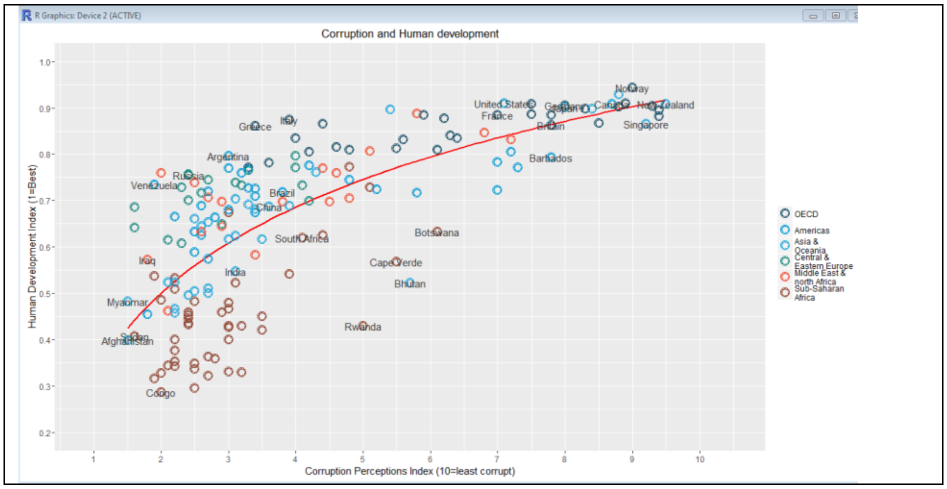
Appendix D: Simple R graphs



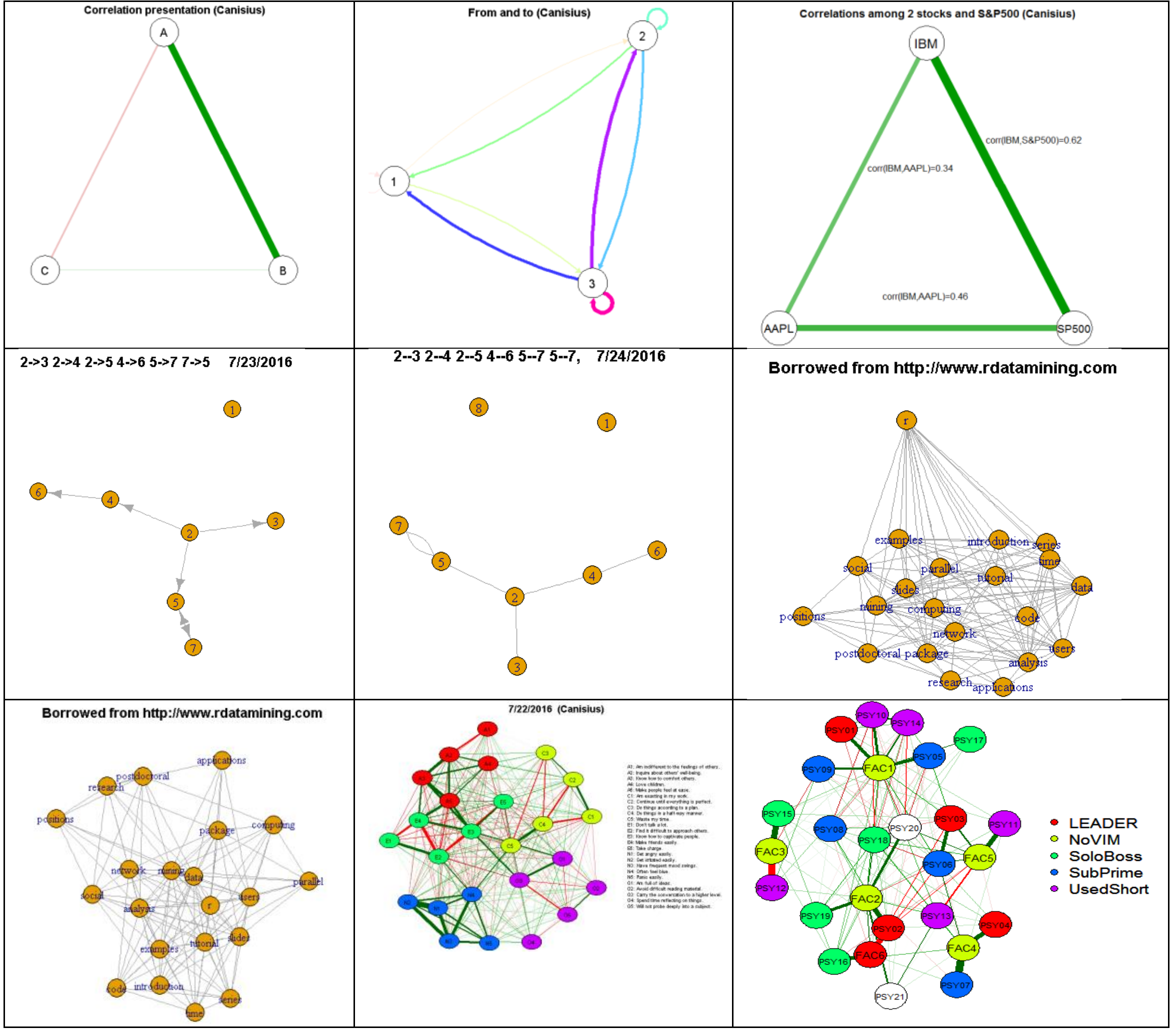
Appendix E: more graphs



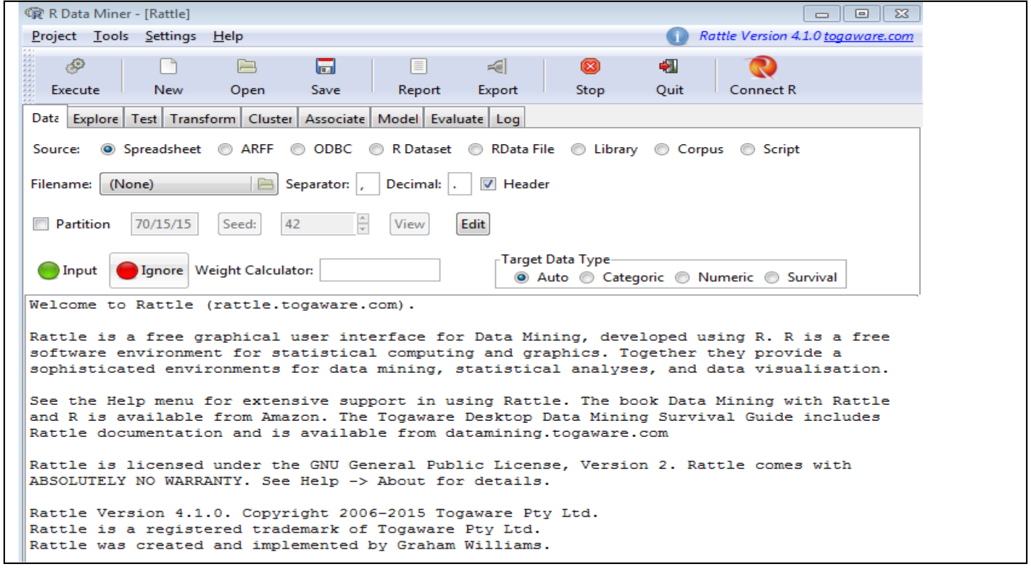
Appendix F: A graphical presentation for the correlation between corruption (Corruption Perception Index) and human development



Appendix G: Graphical presentation for association and correlation



Appendix H: R Rattle package



Source: ● Spreadsheet ○ ARFF ○ ODBC ○ R Dataset ○ RData File ○ Library ○ Corpus ○ Script

Filename: (None) Separator: , Decimal: . ☑ Header

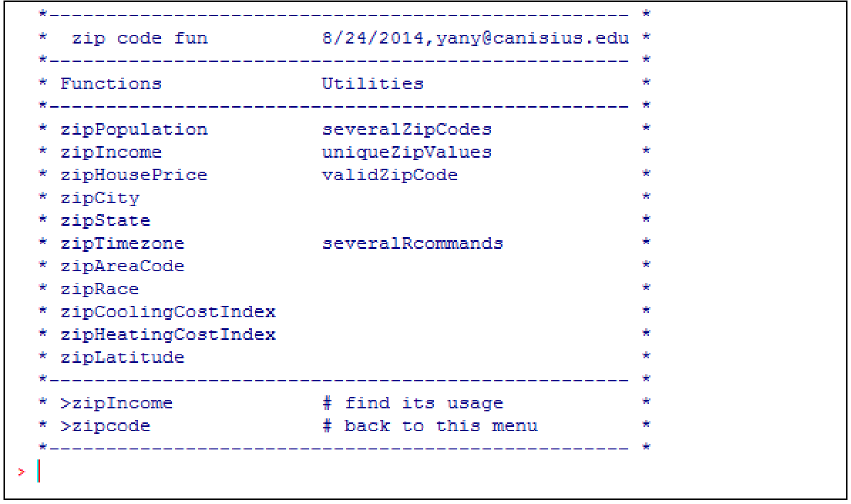
□ Partition 70/15/15 Seed: 42 View Edit

Input Ignore Weight Calculator:

Target Data Type: ● Auto ○ Categoric ○ Numeric ○ Survival

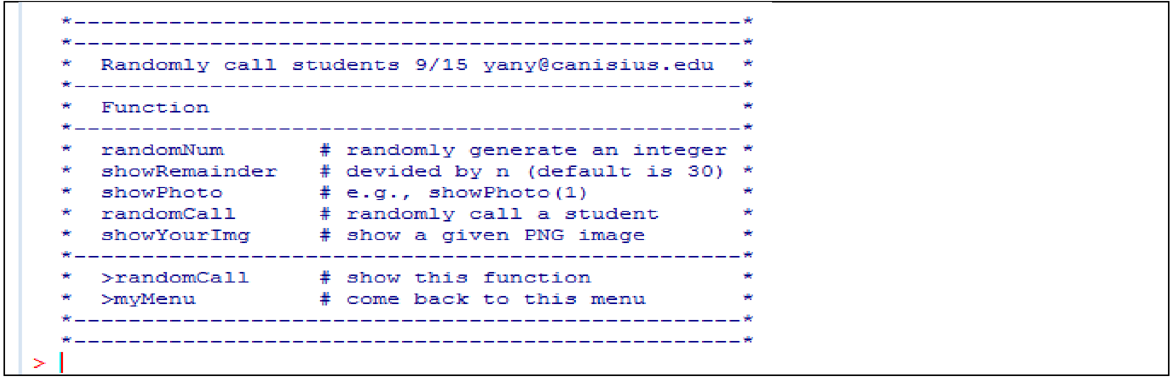
```
Welcome to Rattle (rattle.togaware.com).

Rattle is a free graphical user interface for Data Mining, developed using R. R is a free
software environment for statistical computing and graphics. Together they provide a
sophisticated environments for data mining, statistical analyses, and data visualisation.

See the Help menu for extensive support in using Rattle. The book Data Mining with Rattle
and R is available from Amazon. The Togaware Desktop Data Mining Survival Guide includes
Rattle documentation and is available from datamining.togaware.com

Rattle is licensed under the GNU General Public License, Version 2. Rattle comes with
ABSOLUTELY NO WARRANTY. See Help -> About for details.

Rattle Version 4.1.0. Copyright 2006-2015 Togaware Pty Ltd.
Rattle is a registered trademark of Togaware Pty Ltd.
Rattle was created and implemented by Graham Williams.
```

Appendix I: zip code fun



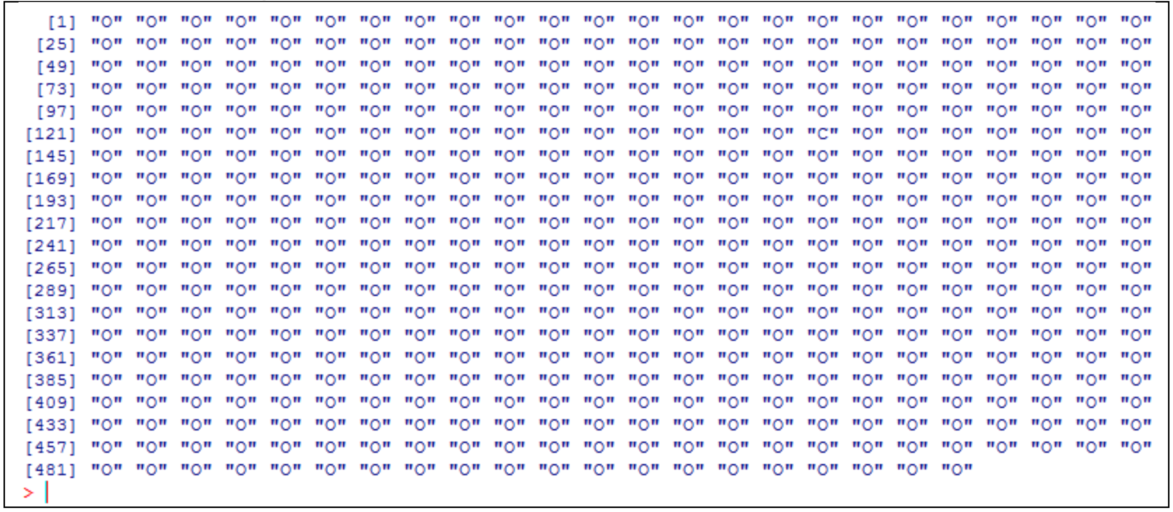
```
  *--------------------------------------------------- *
  *  zip code fun          8/24/2014,yany@canisius.edu *
  *--------------------------------------------------- *
  * Functions              Utilities                   *
  *--------------------------------------------------- *
  * zipPopulation          severalZipCodes             *
  * zipIncome              uniqueZipValues             *
  * zipHousePrice          validZipCode                *
  * zipCity                                            *
  * zipState                                           *
  * zipTimezone            severalRcommands            *
  * zipAreaCode                                        *
  * zipRace                                            *
  * zipCoolingCostIndex                                *
  * zipHeatingCostIndex                                *
  * zipLatitude                                        *
  *--------------------------------------------------- *
  * >zipIncome             # find its usage            *
  * >zipcode               # back to this menu         *
  *--------------------------------------------------- *
> 
```

Appendix J: randomly call students and show their photo on the screen



```
  *--------------------------------------------------*
  *--------------------------------------------------*
  *  Randomly call students 9/15 yany@canisius.edu   *
  *--------------------------------------------------*
  *  Function                                        *
  *--------------------------------------------------*
  *  randomNum         # randomly generate an integer *
  *  showRemainder     # devided by n (default is 30) *
  *  showPhoto         # e.g., showPhoto(1)          *
  *  randomCall        # randomly call a student      *
  *  showYourImg       # show a given PNG image       *
  *--------------------------------------------------*
  *  >randomCall       # show this function           *
  *  >myMenu           # come back to this menu       *
  *--------------------------------------------------*
  *--------------------------------------------------*
> 
```

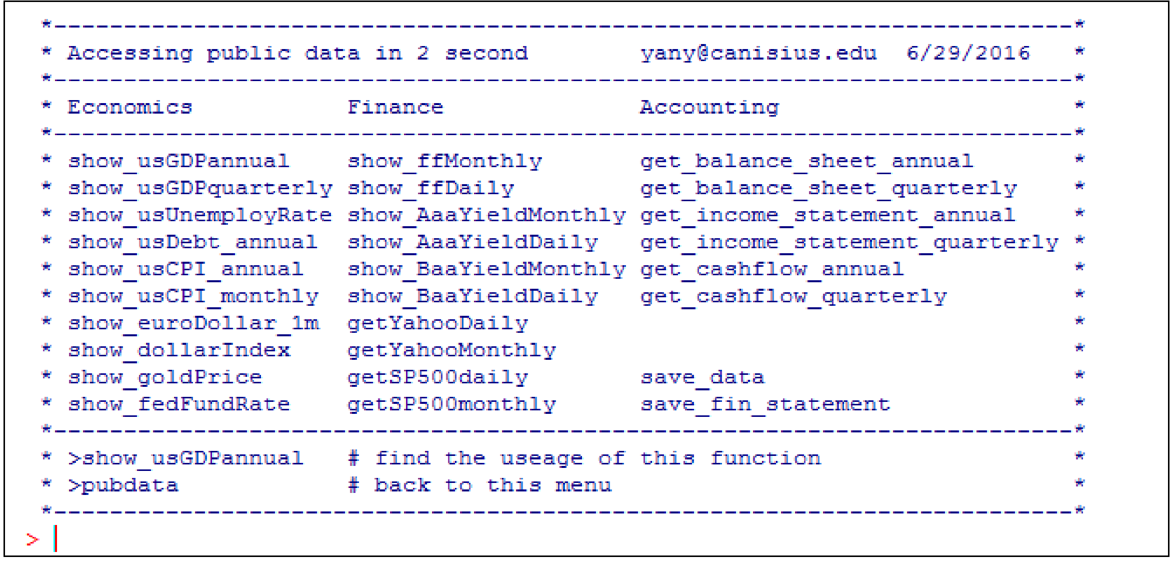
Appendix K: find C among hundreds O



```
  [1] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
 [25] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
 [49] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
 [73] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
 [97] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
[121] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "C" "O" "O" "O" "O" "O" "O" "O"
[145] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
[169] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
[193] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
[217] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
[241] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
[265] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
[289] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
[313] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
[337] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
[361] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
[385] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
[409] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
[433] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
[457] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
[481] "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O" "O"
> 
```

Note: here is the logic. First, generate 500 O, then generate a random number between 1 and 500. Then, replace O at the location with a C. If type solution(), the true location would be revealed. In addition, this game count how long (in second) and user could choose the number of O's such as 100. The default value is 500.

Appendix L: get public data in 2 second



```
*-------------------------------------------------------------------------*
* Accessing public data in 2 second        yany@canisius.edu  6/29/2016   *
*-------------------------------------------------------------------------*
* Economics           Finance              Accounting                     *
*-------------------------------------------------------------------------*
* show_usGDPannual    show_ffMonthly       get_balance_sheet_annual       *
* show_usGDPquarterly show_ffDaily         get_balance_sheet_quarterly    *
* show_usUnemployRate show_AaaYieldMonthly get_income_statement_annual    *
* show_usDebt_annual  show_AaaYieldDaily   get_income_statement_quarterly *
* show_usCPI_annual   show_BaaYieldMonthly get_cashflow_annual            *
* show_usCPI_monthly  show_BaaYieldDaily   get_cashflow_quarterly         *
* show_euroDollar_1m  getYahooDaily                                       *
* show_dollarIndex    getYahooMonthly                                     *
* show_goldPrice      getSP500daily        save_data                      *
* show_fedFundRate    getSP500monthly      save_fin_statement             *
*-------------------------------------------------------------------------*
* >show_usGDPannual   # find the useage of this function                  *
* >pubdata            # back to this menu                                 *
*-------------------------------------------------------------------------*
>
```