Deducer Quick Start Guide



Gail Chapman University of California, Los Angeles



This guide was created under the auspices of the National Science Foundation Math/Science Partnership grant, "MOBILIZE: Mobilizing for Innovative Computer Science Teaching and Learning." Co-principal Investigators: Deborah Estrin (UCLA, CENS), Mark Hansen (UCLA, CENS), Joanna Goode (University of Oregon, College of Education), Jane Margolis (UCLA, Center X), Thomas Philip (UCLA, Center X), Jody Priselac (UCLA, Center X), and Todd Ullah (LAUSD).

Introduction
Installation Instructions3
Data Files3
Console and Data Viewer4
Loading and Navigating a Data File5
Frequency Tables7
Sorting
Subsets
Spatial Data10
Creating a Shape File10
Plotting Points on a Map11
Bubble Charts
Plots and Analysis14
Bar Plot15
Contingency Tables17
Mosaic Plots
Descriptives
Histograms20
Box Plots
Transforming Data22
Text Analytics
Create a Corpus23
View Corpus
Word Counts25
Processing Text27

Introduction

Deducer is a graphical interface designed to work with R (a free, data analysis software environment for statistical computing and graphics) and allow users to perform data analysis without programming. The underlying language of R can be seen at each step, which enables students to learn about R if they are interested, but typing R commands at the command line is not required. This manual is designed to provide a brief overview of the commands that will be required to perform the analysis in the *Exploring Computer Science* curriculum. Features of Deducer are introduced in the order in which they are first used in the curriculum. For additional information related to the various features described and information on the other features of Deducer, view the **Tutorial Video** or the **Online Manual** that can be accessed through the **Data Viewer** screen. (See image below.)



Installation Instructions

The most up-to-date instructions for installing Deducer can be found at http://mobilizingcs.org/softwaretools.

Data Files

The data files for use in ECS v4.0 Unit 5 can be found at http://exploring cs.org/curriculum.

Console and Data Viewer



The **Console** is where R commands are entered if needed and where the R code that is executed when Deducer features are used appears. Many of the results for the analysis features will appear here as well. The bottom screen (input pane) is for entering the R commands; the top screen (output pane) is where the results appear.

Loading and Navigating a Data File

Click on the **Open Data** button in the **Data Viewer** and select the file to be loaded.



The file will either load directly into the **Data Viewer** or the following pop-up window will appear.

🕯 Read Delim	ited File			
Record Seperat	or Qu	iote		
Comma (,)	- D	ouble Quote (")	✓	🗸 Header
	row.names	name	longitude	latitude
1	1	1st & Alameda	-118.238	34.04 🔨
2	2	4th & Wilton	-118.313	34.06
3	3	7th & Figueroa	-118.26	34.04
4	4	8th & La Brea	-118.345	34.0 _
5	5	9th & Pacific	-118.287	33.73
6	6	Alvarado & 7th	-118.277	34.05
7	7	Broadway B	-118.221	34.07
8	8	Ballona Creek	-118.417	33.98
9	9	Eagle Rock	-118.214	34.13 🎽
<	1111			>
			Cancel	Load

Be sure that the **Record Separator** is comma and that the data view looks correct. (Note: There are a variety of possible file seperators, but all of the files in *ECS* are comma delimited. For information on other types see the online manual.)

Click Load.

The loaded file will appear in the **Data Viewer**.

🛃 Dat	🖀 Data Viewer									
File E	File Edit Window Help									
					Data labike (data.fra	a Set ame) dim 🗸			200	
Data Vi	ew Variable Vie	w								
		longitudo	latitudo	tuno	bika coup	nod coun				٦
	1 st % Alamada	110.000105	24.040175	cype bono	DINE_COURT.	241				
- 1	4th & Wilton	-118.238125	24.049173	hike route	62	241				2
- 2	7th & Figueres	110.313441	24.040299	DIRETOLLE	216	1070				
3	Oth & La Pres	-110.259003	34.049300	none	210	1979				
	Oth & Dacific	110.344041	22 725110	none	72	140				
- 5	Alugrade 9, 7th	-110.207306	33.735110	none	50	100				
	Alvaradu ot 701	-110.276021	34.030479		150	020				
	proadway B	-110.221302	34.072107	hike path	63	26				
	Eagle Bock	-110.410053	24 120124	bike route	353	181				
- 10	Eable RUCK	110.214221	34.039129	DINE FOULE	53	246				
10	Echo Park &	-118.260441	34.076076	none	121	1369				
10	Figueroa &	-118.211303	34.090838	none	93	253				
12	Florence &	-118.242545	33.974755	none	119	1526				
13	Fountain &	-118.292327	34.09553	none	110	518				
14	Giendale & P	-118.261514	34.07623	DIKE YOUCE	65	189				
15	Hollywood &	-118.338633	34.101535	none	93	13//				
16	Hoover & M	-118.283916	34.02734	bike lane	9//	/11				
1/	Idaho & Bundy	-118.4613/	34.038827	none	105	263				
18	Kittridge & D	-118.588486	34.191145	bike route	56	191				
19	LA River @	-118.270011	34.120758	bike path	164	46				
20	Lankershim	-118.370433	34.157557	none	97	213				-1
21	Lincoln & Blu	-118.427467	33.967886	bike path	35	32				
22	Long Beach	-118.211603	33.930649	none	39	224				
23	Los Feliz & R	-118.271599	34.116566	bike route	97	124				
24	Manchester	-118.287048	33.960056	bike route	73	407				
25	National & O	-118.417597	34.03239	none	42	147				
26	Santa Monic	-118.338675	34.090803	none	103	411				
27	Santa Monic	-118.436952	34.049673	bike lane	153	370				
28	Sepulveda &	-118.446908	34.050206	bike path	138	167				
29	Sunset & Hy	-118.27971	34.091905	bike lane	145	542				
30	Topanga &	-118.605738	34.171832	none	35	98				
31	Venice & Na	-118.388457	34.02958	bike lane	118	287				
32	Washington	-118.244069	34.020884	none	90	621				
33	Washington	-118.432311	33.996811	bike path	391	154				
34	Westholme	-118.433948	34.062081	bike route	52	162				
35	Westwood	-118.445277	34.063717	bike lane	220	3806				
36	Wilshire & W	-118.309193	34.061619	none	155	2303				
37	Woodman &	-118.431244	34.179927	bike path	49	133				
38	York & Ave 50	-118.20684	34.121433	none	40	180				
39										
40										
41										
47								1		4
	<								>	

Column headings are the names of the variables in the data set; horizontal row numbers indicate each entry. In the above example, there were 38 street intersections and 6 different variables in the *labike* file. Columns can be expanded (as in excel) so that the entire entry can be viewed. Scrolling can be used to view additional rows and columns for larger files.

Frequency Tables

Choose **Analysis** \rightarrow **Frequencies** from the menu bar on the **Console Window**. The following window will appear.

🛓 Run Frequencies	
labike	Run Frequencies On:
	OK Cancel

From the top left pull down menu, choose the data set on which to run frequencies. In this example, the file is *labike*. Other open files can be chosen from the pull down menu. Choose the variable(s) on which to run frequencies and add it to the **Run Frequencies On:** space by clicking on the right arrow. To remove a variable from the **Run Frequencies On:** space, use the left arrow. Click **OK**. The frequency table will appear in the output pane of the **Console** window. The table below shows a frequency table for the variable *type*.

& Consolo									
File Edit Workspa	ace Data An	alvsis Plots S	patial Text	Packages & I	Data Window	Help			
R 🖪 🔚		·**	.	X					
> frequencies(labike[c("t	ype")], r.	digits = 1)					^
\$type									
		Frequencie	 «						
			-						
Value #	of Cases	% Cumu	lative %						
l bike lane	5	13.2	13.2						
2 bike path	6	15.8	28.9						
3 bike route	7	18.4	47.4						
4 none	20	32.0	100.0						
		Case Summa	ry						
			-						
Val	id Missing	Total							
# of cases	38 0	38							
>									~
									<u></u>
🛃 start 📄	🐵 🖸 🐼	» 💷 Dis <u>con</u> .	😹 Wir	ido 📝	2 Int	🔄 Mobiliz	Deduc	😡 Inbox	🕌 2 Ja

Sorting

Choose **Data** \rightarrow **Sor**t from the menu bar on the **Console Window**. The following window will appear.

🕌 Sort Data Frame	
Image: state	Sort data by:
	OK Cancel

Choose the data set to be sorted. In this example, the file is *labike*. Other open files can be chosen from the pull down menu.

🖆 Sort Data Frame		×
labike Filter: name longitude latitude type ped_count_pm	Sort data by: bike_count_pm Increasing Image: Count_pm Increasing	
	OK Cancel]

Choose the variable to sort and add it to the **Sort data by:** space by clicking on the right arrow. To remove a variable from the **Sort data by:** space, use the left arrow. Choose ascending or descending order by clicking on the appropriate button. Click **OK**. The sorted file will appear in the **Data Viewer**.

Subsets

Choose **Data→Subset** from the menu bar on the **Console Window**. The following window will appear.

Data Subset	
Iabike Filter: name longitude latitude type bike_count_pm ped_count_pm Subset Name: <auto></auto>	Subset Expression Recent: Logical Functions Logical Operators is.character is.logical is.complex is.double is.numeric is.vector is.factor is.factor is.finite Not
	OK Reset Cancel

Choose the data set from which to create a subset. In this example, the file is *labike*. Other open files can be chosen from the pull down menu.

Data Subset	X
labike Filter: name longitude latitude type bike_count_pm ped_count_pm Subset Name: Bike Greater	Subset Expression bike_count_pm >= ped_count_pm Recent: Logical Functions is.character is.logical is.complex is.oduble is.numeric is.factor is.factor is.finite
	OK Reset Cancel

Enter the expression for the desired subset in the **Subset Expression** space. The subset desired for this example is those intersections where the bike count was greater than or equal to the pedestrian count. Note that to create a subset on non-numeric variables (e.g., *type*) the values of the variable must be enclosed with quotation marks (e.g., *type* == "none"). Include a name for the subset in the **Subset Name:** space. Deducer will automatically provide a name if one is not supplied. Click **OK**. The subset will appear as a new entry in the pull down menu of the **Data Viewer**. (The original data remains unchanged.)

Spatial Data

Creating a Shape File

In order for Deducer to use the latitude and longitude coordinates to plot points on a map, the data file must be converted to a shape file.

Choose **Spatial**→**Convert data.frame** from the menu bar on the **Console Window**. The following window will appear.



Add the variable *latitude* to the **Latitude** space and the variable *longitude* to the **Longitude** space by using the right arrows. Add a name for the new spatial file to the **New data name** space. Deducer will automatically provide a name if one is not supplied. Click **Run**. The shape file will appear as a new entry in the pull down menu of the **Data Viewer**. (The original data remains unchanged.)

_									
🍝 Da	ta Viewer								X
File E	Edit Help							 	_
_					Data	a Set			_
					(sp-p) labike2	*		-net	KI.
_									
Data V	iew Variable Vi	ew Coordinate	s						
	name	longitude	latitude	type	bike_coun	ped_coun			
1	1st & Alameda	-118.238125	34.049175	none	62	241			^
2	4th & Wilton	-118.313441	34.06713	bike route	48	87			
3	7th & Figueroa	-118.259883	34.049388	none	216	1979			
4	8th & La Brea	-118.344641	34.060446	none	72	272			
5	9th & Pacific	-118.287306	33.735118	none	58	160			
6	Alvarado & 7th	-118.276821	34.056479	none	150	625			-
7	Broadway B	-118.221302	34.072107	none	63	26			
8	Ballona Creek	-118.416653	33.986428	bike path	353	181			1
9	Eagle Rock	-118.214221	34.139124	bike route	53	246			-
10	Echo Park &	-118.260441	34.078078	none	121	1369			-
11	Figueroa &	-118.211303	34.090838	none	93	253			-
12	Florence &	-118.242545	33.974755	none	119	1526			-
13	Fountain &	-118.292327	34.09553	none	110	518			-
14	Glendale & P	-118.261514	34.07623	bike route	65	189			-
15	Hollywood &	-118.338633	34.101535	none	93	1377			
16	Hoover & M	-118.283916	34.02734	bike lane	977	711			-
17	Idaho & Bundy	-118.46137	34.038827	none	105	263			-
18	Kittridge & D	-118.588486	34.191145	bike route	56	191			-
19	LA River @	-118.270011	34.120758	bike path	164	46			-
20	Lankershim	-118.370433	34.157557	none	97	213			-
21	Lincoln & Blu	-118.427467	33.967886	bike path	35	32			-
22	Long Beach	-118.211603	33.930649	none	39	224			
23	Los Feliz & R	-118.271599	34.116566	bike route	97	124			
24	Manchastor	110 307040	22 NENNES	biko routo	79	407			
	<							>	4

Note that the new file has an (sp-p) designation and includes a new tab for coordinates (the latitude and longitude).

Plotting Points on a Map

Choose **Spatial**→**Spatial plot builder** from the menu bar on the **Console Window**. The following window will appear.



Double click on the **Points** button. The following window will appear.

🛎 plot	
Spatial points	✓
Туре	1.0 : open circle 💌
Size	1.0
Color	Set Colour
	OK Cancel

From the pull down menu for **Spatial points**, select the file to be plotted. From the other pull down menus, choose the type of point, the size and the color.

				📓 Choose Colour 🛛 🔀
🖆 plot		📥 plot	×	Swatches HSB RGB
Spatial points	×	Spatial points	×	
Туре	1.0 : open circle 💌	Туре	1.0 : open circle	Recent:
Size	12.0 : plus square 13.0 : target 14.0 : triange square	Size	2.0	
Color	15.0 : solid square 16.0 : solid circle 17.0 : solid triangle	Color	Set Colour	
	18.0 : solid diamond 19.0 : solid circle (big)			Preview Sample Text Sample Text
[OK Cancel		OK Cancel	CK Cancel Reset

Click OK.

The following window will appear. (Note that there seems to be a collection of points near Los Angeles.)



Use + to zoom in to the desired level (- to zoom out). Click and drag the mouse to center the map as desired.



The default map type is **open street map**. To change from the default, choose **Bing aerial images** under **Map types**. Right clicking on **Points** in the **Components** pane provides options to edit, toggle, or remove this map.

Bubble Charts

Choose **Spatial** → **Spatial plot builder** from the menu bar on the **Console Window**. Double click on the **Bubble** button. The following window will appear.

labike2	▼	
Filter: name longitude latitude type		Point size
bike_count_pm ped_count_pm	Minimum size	0.01
	Maximum size	0.05
	Color	Set Colour

Choose the desired file from the pull down menu. Choose the variable to plot and use the right arrow to place it in the **Point size** space. Choose the size for the smallest bubble and the largest bubble and set the color. Click **OK**. Zoom to the desired level. Right clicking on **Bubble** in the **Components** pane provides options to make changes or remove the bubble points.



Plots and Analysis

The **Variable View** tab in the **Data Viewer** provides a list of the variables with their type and factor levels for the categorical variables.

📥 Da	ıta Viewer			
File I	Edit Window Help			
Data			Data Set labike (data.frame) dim	2
Data				
	Variable	Туре	Factor Levels	
1	name	Factor) 1st & Alameda; (2) 4th & Wilton; (3) 7th & Figueroa; (4)	
3	latitude	Double		
4	type	Factor) bike lane; (2) bike path; (3) bike route; (4) none;	
5	bike_count_pm	Integer		
6	ped_count_pm	Integer		
7				

Bar Plot

Choose **Plots**→**Plot Builder** from the menu bar on the **Console Window**. The following window will appear.

💰 Plot Builder	
File Tools Window	
Templates Geometric Elements Statistics Scales Facets Coordinates Other	
Image Image <th< th=""><th></th></th<>	
<	>
Drag a component from aboveORSelect a plot type:Digits of the product of the produc	
Run Rese	et Cancel

Choose **bar** from **Select a plot type:**. The following window will appear.

<u>ه</u>	X
survey	
year effort homework grades	Factor
	Colour by
	OK Cancel

Choose the desired file from the pull down menu. Choose the variable from which to produce the bar plot and add it to the **Factor** space by clicking on the right arrow. Click **OK**. The following bar plot was produced from the variable *effort*.

Plot Builder Fle Tools Window	
Templates Geometric Elements Statistics Scales Facets Coordinates Other	
Image: pie Image:	
<u><</u>	>
000 -	Components
500	bar
400	
200 -	
0 -	
Trying best to do well in school Could try a little hander effort	
Run Res	et Cancel

The bar plot shows the frequency of each of the possible answers for *effort*. The size of the viewing window may need to be adjusted; this can be done by dragging the bottom right corner.

To add color, double click on **Bar** in the **Components** pane. The menu will appear to the right. Choose **count** in the **Colour By** space. (Alternatively, color could have been added in the original bar plot pop up window.)



Contingency Tables

Choose **Analysis** \rightarrow **Contingency Tables** from the menu bar on the **Console Window**. The following window will appear.

survey	Row
Filter:	Cels
	 Colunn
	Stratify By
	Run Reset Cancel

To create a contingency table to relate the answers to two different survey questions add the row variable and column variable to the appropriate spaces and click **Run**. The results will appear in the **Console Window**. The example below was produced from the *survey* data set with *grades* as the row and *effort* as the column.

🛎 Console							
File Edit Workspace Data	Analysis Plots Spat	ial Text Packag	es & Data 🛛 Windov	/ Help			
2 🖬 🖾 🧐 🔇	- 🕒 🛃 🖿						
> print(tables,prop.r=	T,prop.c=T,prop.	t=F)					~
					===		
	Table: gra	ies by effort					
	effort	-					
grade	s Trying best	to do well in	school				
			-				
A Count	1		311				
Row %	1	6	57.609%				
Column	*		47.773% l-				
B Count			241				
Row %	i	4	13.345%				
Column	÷	3	87.020%				
			-				
C Count	1		84				
Row %	1	3	36.842%				
toiumn	< 		.2.903*				
Less than a C Count			15				
Row %	i	3	80.612%				
Column	%		2.304%				
			-				
Column Tota	<u>ц</u>		651				
Column	8	5	50.348%				~
							-
		-				-	
🛃 start 🔰 🎯 🖬 🕼	🥪 🎽 🖅 Discon	👗 Windo	🖉 2 Int 🔹	Mobiliz	Deduc	😡 Inbox	🛃 2 Ja

Mosaic Plots

Choose **Plots** \rightarrow **Interactive** \rightarrow **Mosai***c* from the menu bar on the **Console Window**. The following window will appear.

🛓 imosaic		X
Survey	Type	Vars
		Run Cancel

Choose the desired file from the pull down menu. Select the two variables to compare and add them to the **vars** space by clicking on the right arrow. The first variable listed will be the *x*-direction and the second will be the *y*-direction. Click **Run**. The example mosaic plot compares *grades* to *effort*. (Note that the titles of the various columns may not be correctly aligned.)



Descriptives

Choose **Analysis** \rightarrow **Descriptives** from the menu bar on the **Console Window**. The following window will appear.

cdc	Descriptives of:	
Filter:		
age		
grade		
gender		
iisp_latino		
ace		
height		
veigno		
general_neath]
nours_sleep	Charles Dur	
asunma	Duratury by:	
odraet		
eat helt		
unscreen		
tescribe weight		

Choose the desired file from the pull down menu. Select the variables for which to create descriptives and add them to the **Descriptives of:** space by clicking on the right arrow. Click **Continue**.

Descriptives Functions Skew Kurtoda St. Deviation Mean Vaid N	Run Descriptives Medan 25th Percentile 75th Percentile Primum Maximum
	Custom
	Run Reset Cancel

Select the function(s) desired (e.g., mean) and add it to the **Run Descriptives** space by clicking on the right arrow. Click **Run**. The descriptives will appear in the **Console Window**.

🗟 Console 📃 🔲	\times
File Edit Workspace Data Analysis Plots Spatial Text Packages & Data Window Help	
Large Sample	^
Test Statistic DF p-value Effect Size est. Lower (%)	
Chi Squared 90 3 <0.001 Cramer's ∀ 0.264 0.206 (2.5)	
Lest upper (%) Chi Smared D 35 (97.5)	
> remove(tables)	
> print(load("C:/Documents and Settings/Gail Chapman/My Documents/Into The Loop/ECS Version 4.0/Supplemental	
Materials/Data Files/Unit 5 data files/cdc.rda"))	
The following data objects have been loaded:	
[L] "CGC" > descriptive table(vers = d(weight beight) datas cdc	
<pre>/ descriptions.camf(vals = descript), descript(), deca = description = descriptio</pre>	
<pre>\$`strata: all cases `</pre>	
Median 25th Percentile 75th Percentile Minimum Maximum	
weight 65.32 56.70 77.11 34.47 180.99	
height 1.68 1.63 1.78 1.27 2.11	Ξ
	¥.
	^
🛃 Start 🛛 🕲 🔽 😔 🎽 🕼 Discon 👗 Windo 🖉 2 Int 🔹 Mobiliz 🔯 Deduc 😡 Inbox 🚮 2 Ja	

Histograms

Choose histogram from the Plot Builder.



Choose the desired file from the pull down menu. Add the variable from which to create the histogram by selecting it and clicking on the right arrow. The histogram below is based on *height*.



To add color, double click on **histogram** in the **Components** pane. The menu will appear to the right. Choose **count** in the **Colour By** space. (Alternatively, color could have been added in the original histogram pop up window.)



Box Plots

Choose **box plot** from the **Plot Builder**. The following window will appear.

Filter:	<u> </u>
age	Variable
grade	
gender	
hisp_latino	
race	
height	
weight	
general_health	Factor
nours_sleep	
astrima	
depressed kalmat	
nemet	
seat_oex	
describe metable	Participation (
uescribe_vieignic	Dodge by
est relad	
drink soda	
drink mile	
days everrise 20	
days exercise 60	
number teams	
hours ty	
hours videogame	
days_smoking	
fights	

Choose the desired file from the pull down menu. Choose a numeric variable and add it to the **Variable** space by clicking the right arrow. Choose a categorical variable and add it to the **Factor** space by clicking the right arrow. Click **OK**. The box plot below was created from the *Men* subset of the *cdc* data set with *height* as **Variable** and *gender* as **Factor**.

Plot Builder ile Tools Window			
Templates Geometric	Itements Statistics Scales Facets Coor	dinates Other	
template pie histogram	bergizate bar boxplot scatter lin	state temptate temptate temptate temptate	
<			
2.1 -		i	Components
2.0 -			boxplot
1.9 -			
tugi 1.7 -			
1.5 -			
1.4 -			
	Female gender	Male	
		Run	Reset Cancel

By using the original *cdc* file without subsetting, side by side box plots can be created as indicated below.

S Plot Builder
He Tools Window
Templates Geometric Elements Statistics Scales Facets Coordinates Other
Image Image <td< td=""></td<>
٤
Components
Female Male NA
Run Reset Cancel

Transforming Data

To convert data in a particular column to another format (e.g., convert meters to inches), choose **Data** \rightarrow **Transform** from the menu bar on the **Console Window**. The following window will appear.

dc .	~		Variables to Transfo	m		
Filter:						Target
98	^					
rade						
ender						
isp_latino						
sce						
eight						
eight						
eneral_health						
ours_sleep						
sthma						
epressed						
elmet						
eat_belt					~	
unscreen						
escribe_weight			Plot			
at_fruit						
at_salad		Transron	macion			
rink_soda		::: Se	lect Transformation		~	
rink_mik						
ays_exercise_20	×					

Add the desired variable to the **Variables to Transform** space. Choose the appropriate transformation from the pull down menu or choose **Enter Function...** under the **Custom** option. (Scroll down to the bottom of the pull down menu list.)

🛎 Transform Variables				×
cdc	~	Variables to Transform		
Fiter:		heightheight.tr		Li Target
ope grade grade http://www.prace/ race weight grand/path grand/path grand/path grand/path authons set/path set/		Flot		
drink_mik days_exercise_20 days_exercise_60	39.37*x		~	
		Run	Reset	Cancel

The transformation shown above converts *height* to inches in the *cdc* file. Click **Run**. The transformed variable will appear as the last column in the **Data Viewer**. (Scroll to see the last column.)

. Ed	Niteday Male								
cu	c whom hep								
				Data	Set				
۲¥.				cdc (data.frame) din(1 🗸				6
									-
ta Vier	// Variable Wew								
	drink_mik days_exe	days_exe	number_t	hours_tv	hours_vid	days_smo	fights	height.tr	
1	1 to 3 glasses 5 days	5 days	0 teams	2 hours per	2 hours per	0 days	0 times	66.9289999	
2	3 diasses pe 7 days	3 days	1 team	2 hours per	Less than 1	0 days	1 time	68.8975	
3	4 or more gl 7 days	5 days	1 team	Less than 1	No playing v		0 times	70.866	
4	2 glasses pe 0 days	7 days	0 teams	1 hour per day	1 hour per day	0 days	0 times	57.8738999	
5	1 glass per 10 days	1 day	0 teams	3 hours per	4 hours per	1 or 2 days	0 times	72.0471	
6	2 glasses pe 3 days	4 days	3 or more te	Less than 1	No playing y	0 days	0 times	66.1416	
7	Did not drink2 days	1 day	0 teams	3 hours per	1 hour per day	6 to 9 days	1 time	64.1730999	
8	Did not drink 0 days	0 days	0 teams	2 hours per	4 hours per	0 days	0 times	64.9605	
9	Did not drink 1 day	0 days	1 team	3 hours per	2 hours per	0 days	0 times	61.0235	
10	1 dass per 2 davs	3 days	1 team	Less than 1	4 hours per	0 days	0 times	61.8109	
11	1 to 3 glasses 10 days	0 days	1 team	4 hours per	Less than 1	0 days	0 times	61.8109	
12	1 to 3 glasses 4 days	7 days	0 teams	2 hours per	1 hour per day	1 or 2 days	0 times	66.1416	
13	1 to 3 glasses 0 days	1 day	0 teams	2 hours per	Less than 1	0 days	2 or 3 times	64.9605	
14	4 to 6 glasses 10 days	7 days	1 team	3 hours per	S or more h	0 days	4 or 5 times	70.866	
15	4 to 6 glasses 7 days	2 days	0 teams	Less than 1	3 hours per	1 or 2 days	0 times	70.866	
16	Did not drink 4 days	4 days	3 or more te	2 hours per	No playing v	0 days	2 or 3 times	70.866	
17	1 to 3 glasses 5 days	5 days	2 teams	2 hours per	Less than 1	0 days	0 times	66.1416	
18	Did not dripk 0 days	0 days	Obeens	Less than 1	No playing y	All 30 days	6 or 7 times	64.9605	
19	1 to 3 glasses 4 days	0 days	1 team	4 hours per	1 hour per day	20 to 29 days	12 or more t	64,9605	
20	1 to 3 diasses 10 days	n days	Oteans	4 hours per	5 or more h	0 days	Otimes	74.0155999	
21	4 to 6 glasses 7 days	4 days	Oteams	2 hours per	Less than 1		0 times	68.8975	
22	1 to 3 diasses 7 days	1 day	Obeans	Less than 1	No playing v	all an days	12 or more t	74,803	
23	Did not drink7 days	7 days	1 team	1 hour per day	No playing y	0 days	0 times	70,866	
24	1 to 3 glasses 2 days	3 days	0 teams	Less than 1	Less than 1	0 days	0 times	66.1416	
25	Did not drink	6 days	Obeams	2 hours per	2 hours per	0 days	0 times	62,992	
26	2 glasses pe 5 days	5 days	3 or more te	1 hour per day	Less than 1	0 days	0 times	66.1416	
27	1 to 3 glasses 7 days	7 days	2 teams	3 hours per	Less than 1	0 days	0 times	74.0155999	
28	2 glasses pe 3 days	2 days	0 teams	3 hours per	No playing y	0 days	0 times	66.1416	
29	Did not drink 7 days	7 days	0 teams	Less than 1	No playing y	All 30 days	4 or 5 times	66.1416	
30	Did not drink 7 days	7 days	1 team	2 hours per	Less than 1	1 or 2 days	0 times	66.1416	
31	4 to 6 glasses 3 days	3 days	1 team	2 hours per	No playing y	0 days	0 times	68.1100999	
32	3 glasses pe 7 days	7 days	1 team	3 hours per	1 hour per day	0 days	2 or 3 times	10	
33	1 to 3 glasses 5 days	4 days	2 teams	Less than 1	5 or more h	0 days	0 times	66.1416	
34	1 glass per 1 day	5 days	1 team	1 hour per day	2 hours per	0 days	0 times	66.9289999	
35	Did not drink6 days	6 days	2 teans	5 or more h	Less than 1	0 days	0 times	72.8345	
36	3 glasses pe 5 days	5 days	0 teams	2 hours per	1 hour per day	0 days	2 or 3 times	100.00	
37	1 to 3 glasses 2 days	0 days	0 teams	5 or more h	No playing y	0 days	2 or 3 times	68.1100999	
38	1 to 3 glasses 10 days	2 days	0 teams	5 or more h	Less than 1	0 days	0 times	66.9289999	
39	1 to 3 diasses 10 days	1 day	Oteams	2 hours per	2 hours per	All 30 days	1 time	72.0471	
40	4 to 6 glasses 10 days	1 day	0 teams	1 hour per day	No playing y	All 30 days	0 times	65.1416	

Text Analytics

Create a Corpus

In order for Deducer to perform advanced analytics on a file of text, the file must be converted to a "corpus". Choose **Text→Extract Corpus from Dataframe** from the menu bar on the **Console Window**. The following window will appear.

🕌 Extract Corpus	×
CATwitter	
Filter:	
created created.date.format username longitude latitude search_term message	
Save Corpus As:	
Save	Cancel

Choose the desired data set from the pull down menu. Choose the variable that contains the text data to be analyzed.

🕌 Extract Co	rpus	
CATwitter		~
Filter:		
created created.date.fc username longitude latitude search_term	rmat	
message		
Save Corpus As		
message.corpu:	5	
S	ave	Cancel

In the example above, *CATwitter* is the data set and *message* is the text variable. The name of the corpus appears in the **Save Corpus As:** space; this name can be revised if desired. Click **Save**.

View Corpus

Choose **Text**→**View Corpus** from the menu bar on the **Console Window**. The following window will appear.



Choose the corpus to be viewed from the pull down menu. The **Doc#** column indicates the element in the vector (in other words, the survey entry number or row number in the data set). The **Text** column shows the first several words of the associated text and the **Document Text**: pane shows the full text for the highlighted element. Enter a number in the **Go To:** space to go directly to a particular element.

Word Counts

Choose Text \rightarrow View Frequency Data \rightarrow Frequency Totals List from the menu bar on the Console Window. The following window will appear.

Term Frequency 🛛 🗙
Source Data: message.corpus 🗸
View As: Frequency Totals List 💌
View Options:
Sort: by frequency
 descending ascending
Save Frequencies as Variable:
message.corpus.term_freq Save
Term Frequency O Document Frequency
Use: Filtering:
O Top 100 Frequency ≥ 0
O Top 20 %
O All Terms
Close View

From the pull down menu for **Source Data:**, choose the file to be analyzed. From the **View As:** pull down menu, choose **Frequency Totals List**, **Bar Chart**, or **Word Cloud**. The view options are by **frequency** or **alphanumerically** and in either ascending or descending order.

Frequencies can be run on the *x* most appearing words in the list, the top *x* percent, or the entire list. Frequencies can also be filtered so that only words appearing more than a certain number of times will be listed.

Saving the file of frequency data (by entering a name in the **Save Frequencies as Variable:** space) makes the resulting frequencies also appear in the **Data Viewer**. Below are examples of the **Frequency List** and **Bar Chart** views.



Deducer Quick Start Guide

The menu for the **Word Cloud** view appears below.

Term Frequency
Source Data: message.corpus 🗸
View As: Word Cloud
View Options:
Min Font Size: 0.25 💌 Max Font Size: 4.0 💌
Coloring: Black
Randomly Rotate Terms
Term Frequency O Document Frequency Use: Filtering:
O Top 100 Frequency ≥ 0
• Top 20 %
O All Terms
Close View

Note that you can vary the font size, change the coloring and randomly rotate the terms in the word cloud.



Processing Text

Choose **Text** \rightarrow **Preprocess corpus** from the menu bar on the **Console Window**. The following window will appear.

Image: Construction Construction Image: Construction	Actions: To Low	
Y To Lower Case X Y Remove Punctuation X Y Remove Numbers X Y Remove Stop Words X Q Remove Whitespace X Stem Words X	To Low	
Y Remove Punctuation × Y Remove Numbers × Y Remove Stop Words × Remove Whitespace × Stem Words ×		er Case
Remove Numbers Remove Stop Words Remove Whitespace Stem Words	Remove Pund	tuation
Remove Stop Words	Remove N	lumbers 🚊
Remove Whitespace	Remove Stop	Words 💌 🛖
Stem Words	Remove Whit	tespace
	Sterr	n Words

From the **Source Corpus** pull down menu choose the file to be processed. Four **Actions** are checked by default. These can be unchecked in order to perform only one or two processes at a time.

mes	sage.corpus 🗸
Actions:	To Lower Case
	Remove Punctuation
	Remove Numbers
	Remove Stop Words 💌 🛖
	Remove Whitespace
	Stem Words
Save Corp	us As: message corpus processed

The processed corpus is saved by default and the name appears in the **Save Corpus As**: space. A new name can be provided. The original corpus will remain intact. (Note: When creating a frequency list, bar chart, or word cloud, choose the appropriate **Source Corpus** from the pull down menu because the program will default to the first corpus in the list.) The processed corpus can be viewed by returning to the **Console Window**, and choosing **Text**->**View Corpus** from the menu bar.

