

Group Concept Mapping in R

Haim Bar

Department of Statistics, University of Connecticut
haim.bar@uconn.edu

August 22, 2018

Introduction

R-CMap is an open-source concept mapping software, implemented in R (R Core Team 2018). It provides a graphical user interface to guide users through the concept mapping analytical process (Trochim and McLinden 2017).

This document provides information on the required format of the input data, the installation process, and the graphical and analytical capabilities of R-CMap.

R-CMap (Bar and Mentch 2017) is user-friendly and does not require any programming experience. It can be used to generate cluster maps, point rating, cluster rating maps, pattern matching, and go-zone plots. It can also be used to generate detailed reports with statistical analyses, in MicroSoft Word format, using the rtf package (Schaffer 2018). Windows users can generate Windows Metafile formatted plots which can be edited in Microsoft Word and PowerPoint, in order to manually adjust various features to achieve the best visual results. Details about the plots, reports, and features appear in the sections below.

The R-CMap homepage is <https://haim-bar.uconn.edu/software/R-CMap/>. For questions, comments, and suggestions, please contact [Haim Bar](#)

To cite R-CMap, use:

Bar, Haim, and Lucas Mentch. 2017. "R-Cmap, an Open-Source Software for Concept Mapping." *Evaluation and Program Planning* 60: 284-92. doi:10.1016/j.evalprogplan.2016.08.018

Installing and starting R-CMap

To use R-CMap, you must first install R (version $\geq 3.5.1$) (R Core Team 2018) which is available from <https://www.r-project.org/> and RStudio (version $\geq 1.1.456$) (RStudio Team 2015) which is available from <https://www.rstudio.com/products/rstudio/download/>

R-CMap requires the following packages:

- DT (Xie 2016)
- readxl (Wickham and Bryan 2017)
- rtf (Schaffer 2018)
- shiny (Chang et al. 2017)
- smacof (de Leeuw and Mair 2009)
- tools (R Core Team 2018)
- xtable (Dahl 2016)

Follow the installation instruction of R and RStudio. If you already have R and RStudio, check the version and upgrade to the latest versions, if needed. Then, start RStudio and in the Console, type the following:

```
install.packages("shiny")
install.packages("DT")
install.packages("readxl")
install.packages("rtf")
install.packages("smacof")
install.packages("xtable")
```

There are two ways to run R-CMap - locally or remotely. To run it locally, download the file <https://haim-bar.uconn.edu/wp-content/uploads/sites/1740/2018/06/RCMAP.zip> and unzip it in a folder on your computer. In the following, we call this folder RCMAPdir (although you may use a different folder name.) Then, start RStudio and use the **File > Open File...** menu to open the file server.R under RCMAPdir (or whatever name you used for the the R-CMap source folder.) Click on the small triangle to the right of the “Run App” button, and select “**Run External**”, and then click on “Run App”. Alternatively, in the RStudio console, type the following:

```
runApp('RCMAPdir', display.mode="no", launch.browser=FALSE, port=2197)
```

Replace RCMAPdir with the name of the folder where you saved the R-CMap files. At the end of this document you can see a screenshot that shows the two ways to run R-CMap locally from the RStudio user interface.

You may also run R-CMap *remotely*, which means that you do not have to download and unzip the source code on your computer. Simply type the following in the RStudio Console:

```
runUrl('https://haim-bar.uconn.edu/wp-content/uploads/sites/1740/2018/06/RCMA
P.zip')
```

After that, the R-CMap graphical user interface will appear on your default web browser and the “New Project” tab will be displayed. At the end of this document you can find a screenshot. In the sidebar on the left, click the “Browse” button and select the Concept Mapping data file (see below about the file’s structure.)

Once you have selected the data file and R-CMap has loaded it, the sidebar will show a link to “Load a different data set”, which takes you back to the New Project screen if you want to select a different Concept Mapping project. The main panel has 6 tabs at the top (Summary, Plots, Reports, Analysis, Users, and Settings.) These tabs are described in detail below.

The Input Data

The input data has to be in an Excel file, which contains five sheets, as follows. Note that the sheet and column names below are all case-sensitive. An example called DemoProject.xls is provided with the R-CMap software for reference.

- **Statements** must have two columns: StatementID and Statement. The statement IDs (first column) are number from 1 to N, where N is the number of statements. This sheet must have the field names (“StatementID” and “Statement”) in the first line.
- **SortedCards** records the statements in each pile, per sorter. The first column in the User ID (again, sequential number from 1 to M, where M is the number of participants). The second column contains a name of the pile, as given by the sorter. It may be blank, but the column must be included in the sheet. The next columns contain the card numbers that person j put in pile k. Note that each row in this file may have a different number of non-empty cells.
- **Demographics** contains information about the sorters. The first column must be named SorterID and must contain the IDs of all the sorters. The other columns will contain information specific to the experiment. The first row must contain the variable names.
- **Ratings** contains ratings given to each card (statement) by each sorter. The first column must be named UserID and must contain the IDs of all the sorters. The second column must be called StatementID and it contains the ID of each card. So, if there are 10 sorters and 80 statements, there will be 800 rows in the sheet, plus one row for the header. The rest of the columns will contain any rating variable that the sorters were instructed to rate. In the provided example, we have Feasibility and Importance. It is recommended that all rating variables are all on the same scale (e.g. 1-5).
- **RatingsScale** contains the allowed ranges of rating variables. The first column must contain “Variable”, “Min”, and “Max”. The subsequent columns contain the name of the rating variable in the first row, and the minimum and maximum values of each variable, in lines 2 and 3, respectively.

Most common data-entry problems occur because of one of the following reasons:

- Statements from the same sorter are entered to more than one pile.
- StatementID or UserID are not consecutive numbers
- Sheet names are not as specified above
- Variable names in the “Ratings” sheet are different from the variable names in the “RatingsScale” sheet.

To detect data-entry problems, a simple command-line script is provided with R-CMap (**checkDataFile.R**). Once you have completed entering the Concept Mapping experiment data into an Excel file using the required structure, open RStudio, and in the top menu choose **File > Open File...** and open the checkDataFile.R in your RCMAPdir folder. In the console, type

```
source('checkDataFile.R')
```

A file browser will open, and you will be asked to select the Excel file that you created. The program will look for structural errors in the data. Fix any problems in the data file before running R-CMap.

R-CMap tabs

The Summary tab

This tab includes a short summary of the dataset currently being used. It contains the following:

- Data file name
- Number of raters
- Number of statements

- Description
- Rating variables

The Plots tab

Point Map (MDS)

Displays the two-dimensional representation of the distances between statements, as obtained from the MDS (multi-dimensional scaling) algorithm.

Statements are labeled on the plot using their number. To see the corresponding text, put the mouse over a point. This will also show the average rating of the selected statement, for the rating variable selected from the drop-down menu in the sidebar.

Clusters

Displays the two-dimensional MDS plot, with points grouped into clusters. The number of clusters is determined by the user in the Settings tab (see below). There are two types of cluster display - Rays, where each point in a cluster is connected to the cluster's center; and Polygons, where clusters appear as convex polygons, with smooth corners. The default cluster names are number 1, ..., L where L is the selected number of clusters. However, the user can choose more descriptive cluster names via the Settings tab. As in the Point Map (MDS) plot, hovering over a point shows the corresponding statement's text and the average rating of the selected variable.

Univariate Rating - by Statement

Similar to the Point Map (MDS), except that the average rating of each point is represented by a vertical bar. The bars are color-coded according to the cluster to which each statement belongs.

The sidebar also contains a drop-down called Group, where the user can choose to display the average ratings for a subset of the sorters. Defining subsets of sorters is explained in the description of the Users tab below.

Univariate Rating - by Cluster

A combination of the polygon cluster plot and the Univariate Rating plot, except that the average ratings are displayed graphically for each cluster (as smoothed, multi-layered polygons).

As in the previous plot, it is possible to choose a subset of sorters via the drop down menu "Group" in the sidebar.

Bar Chart

Displays a bar chart for a rating variable, where each bar corresponds to a cluster. If more than one rating variable is available in the data the user may select up to four variables to be included in this graph, in a 2x2 array. Selecting the rating variables is done via drop down menus in the sidebar.

Here, too, it is possible to choose a subset of sorters via the drop down menu "Group" in the sidebar.

Pattern Matching

Also known as a "ladder graph", this plot provides a visual comparison between cluster averages, using two rating variables. Average cluster ratings for each variable are displayed on two parallel vertical axes and joined by a line, making it easy to compare how each cluster is viewed by the sorters in terms of the two variables. For example, a cluster of statements may be seen as the most important (on average) among all clusters, but perhaps less feasible than other clusters. Plotting the "ladder graph" by groups (subsets of sorters) may be used to see whether the observed bivariate patterns

are consistent across sub-populations. This plot is particularly useful to detect overall differences between rating variables, by cluster.

Go-Zone

A bivariate plot showing on Cartesian axes the ratings of two variables. Each statement is represented by a point, allowing to see very easily how points are evaluated by users on both axes, simultaneously. For example, if the variables are Importance and Feasibility, points in the upper-right corner are considered highly important and also very feasible. Each point is also labeled by the cluster to which it belongs, allowing to see which clusters rate higher on each dimension. The plot is divided into quadrants, defined by the mean ratings along each axis. Two dotted vertical lines and two dotted horizontal lines represent the mean plus/minus one standard deviation.

It is possible to select a subset of the clusters, using the check-boxes in the sidebar. When a subset of clusters is selected, the quadrants are defined by the mean ratings in the selected clusters. However, for reference, dashed white lines (vertical and horizontal) display the mean ratings across all clusters. Furthermore, the plot can be drawn for specific groups of sorters.

Go-zones are very useful for planning purposes, as they highlight which specific statements rank highest on two dimensions, thus helping stakeholders to prioritize action-items, either by individual statements or by cluster. The ability to view this plot by group can be particularly helpful in planning and evaluating.

Dendrogram

Following the MDS step, a hierarchical clustering is performed on the two-dimensional representation of the data. In the process, components (statements, or sets of statements) are joined iteratively with their nearest neighbor. The nearest neighbor may be another statement, or a group of statements which were joined into one group in a previous iteration. A dendrogram depicts the hierarchical clustering process. “Leaves” in this binary-tree diagram correspond to the statements, and branches represent the nearest neighbor connections made in each iteration. The length of an edge in a dendrogram is a function of the dis-similarity between joined components.

Put the mouse over the numbers at the bottom of the dendrogram to view the corresponding statement.

Notes

To save a plot, check the “Include in report” box at the bottom of the sidebar. Then, use the “Save Report” option in the Settings tab to export all the selected plots to a Microsoft Word (.docx) file. It is possible to save the same type of plot using multiple configurations, such as number of clusters, and selected subgroups.

The Reports Tab

Three types of reports are available, to summarize Sorter, Statement, or Cluster level data. In all three reports, it is possible to obtain summaries for all the participants in the pile-sorting experiment, or for a subset, by defining sub-groups via the Settings tab, and choosing the subset in the Group drop-down menu.

As is the case for the plots, each of the reports below can be saved to a final report in a Microsoft Word format, by clicking on the “Include in report” check-box in the sidebar.

Sorters

The Sorters report shows for each sorter the number of statements sorted, and the number of piles to which they were placed. Note that some sorters may leave some cards unsorted.

Statements

This report shows all the statements with their mean rating and the standard deviation. It also shows to which cluster each statement belongs. The rating variable can be selected from a drop-down menu in the sidebar.

The sidebar also contains a drop-down menu to select a subgroup. It is also possible to choose how to sort the data. The two options are by statement ID or by mean rating (from largest to smallest). The sidebar also contains a check-box which allows the user to view the report by cluster.

Clusters

This option shows summaries for each cluster, including the number of statements in each cluster, the average squared-distance to the cluster’s center (SSq/N), the mean and the standard deviation of the selected rating variable. The rating variable, as well as the group of sorters can be selected via the sidebar.

The Analysis Tab

Two cluster-level analyses are currently available:

- ANOVA (analysis of variance) - test whether all the clusters have the same mean rating or not. A small p-value indicates that at least one cluster has a mean rating which is significantly different.
- Tukey - a pair-wise comparison between clusters. With this test it is possible to determine which pairs are different. Tukey’s test controls the probability of Type I error across all comparisons.

In both cases, it is possible to change the rating variable, and/or to select a subgroup of sorters to which the analysis will be applied.

The Users Tab

The users tab uses the DT package (Xie 2016) to show all the sorters and their demographic data in an interactive table. Below each column header there is a selection box, where it is possible to select a subset of values. For numerical variables, the selection can be done by sliding the end-points, and for categorical variables the selection is done by choosing one or more levels. The table can be sorted by clicking on the up/down arrows next to the column names. The table can also be searched by using the box above the column names.

Once a subset of users has been selected, it can be saved by entering a group name in the sidebar and clicking on “Create Group”. To delete an existing group, select it in the Group drop-down in the sidebar and click “Remove Group”. The “All” group cannot be deleted. A group that has been created will now appear in all the plots, reports, and analyses which have the Group drop-down option.

The Settings Tab

This tab has four options, which can be selected by clicking on the appropriate Setting radio button:

- **Number of Clusters:** use the drop-down menu to change the number of clusters.
- **Cluster Names:** when choosing this option, the main panel shows the cluster selected in the sidebar drop-down menu using the “Rays” viewing option. Below the plot there is a list of all the statements in the cluster. If the sorters provided a label for each pile, the top-10 labels will appear at the bottom of the page. The top-10 are determined by simply counting how many time each label appeared, summing over all the cards in the cluster. Obviously, different sorters will use different labels but looking at the most common terms should help the R-CMap user decide on a proper name for the cluster.
To change the name of a cluster, enter in the “Cluster Name” box in the sidebar, and click the Change button.
- **Save Project:** this option allows to save the Concept Mapping project so that it can be retrieved later with all the user’s selections and preferred settings (e.g. number of cluster, plots/reports/analysis to be included in the final report.) Simply enter a brief description of the project (optional) and a file name. The project will be stored in a file with an .RData extension
To retrieve a saved project at a later time, click on the “Select a concept mapping file” Browse button in the New Project tab, and select the .RData file.
- **Save Report:** to save all the selected plots, reports, and analyses to a Word document, click on “Save Report” and choose a file name. For Windows users only - if you want to be able to edit the plots manually in Word or PowerPoint, check the “Editable plots” box. Click on the “Download Report” button. Generating the report may take a few seconds, depending on how many plots were selected. The process is a bit slower when the “Editable Plots” option is enabled. If the “Editable plots” is used (Windows users), each individual plot (in wmf format) will be saved in the home-directory in files called RCMaPPlotNNNN.wmf where NNNN will be incremented automatically from 0001 to 9999. The report file will contain the name of the home directory and the range of RCMaPPlot files that correspond to the report.

Visit the R-CMap online website to see a report generated by R-CMap. Note that R-CMap generated a Word document, but for the purpose of this demonstration it was saved in a pdf format.

Quitting the program

To quit the program, press the Quit checkbox in the sidebar. You will be prompted to confirm this action. If you clicked it by mistake, simply uncheck the “Quit” box. If you want to save your plots, analyses, or report, you must do that before quitting the program.

Note that the Quit action stops the Shiny server, and in some browsers it closes the tab where the R-CMap interface was displayed, but in some browsers the browser’s tab has to be closed manually.

Citations

Bar, Haim, and Lucas Mentch. 2017. "R-Cmap, an Open-Source Software for Concept Mapping." *Evaluation and Program Planning* 60: 284–92. doi:[10.1016/j.evalprogplan.2016.08.018](https://doi.org/10.1016/j.evalprogplan.2016.08.018).

Chang, Winston, Joe Cheng, JJ Allaire, Yihui Xie, and Jonathan McPherson. 2017. *Shiny: Web Application Framework for R*. <https://CRAN.R-project.org/package=shiny>.

Dahl, David B. 2016. *Xtable: Export Tables to Latex or Html*. <https://CRAN.R-project.org/package=xtable>.

de Leeuw, Jan, and Patrick Mair. 2009. "Multidimensional Scaling Using Majorization: SMACOF in R." *Journal of Statistical Software* 31 (3): 1–30. <http://www.jstatsoft.org/v31/i03/>.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

RStudio Team. 2015. *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc. <http://www.rstudio.com/>.

Schaffer, Michael E. 2018. *Rtf: Rich Text Format (Rtf) Output*. <https://CRAN.R-project.org/package=rtf>

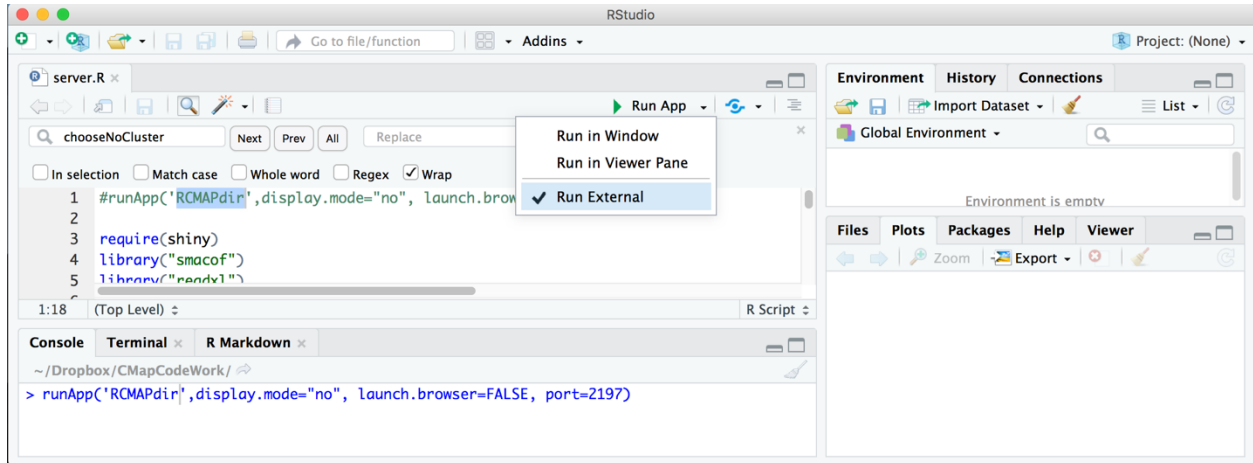
Trochim, William M., and Daniel McLinden. 2017. "Introduction to a Special Issue on Concept Mapping." *Evaluation and Program Planning* 60: 166–75. doi:<https://doi.org/10.1016/j.evalprogplan.2016.10.006>.

Wickham, Hadley, and Jennifer Bryan. 2017. *Readxl: Read Excel Files*. <https://CRAN.R-project.org/package=readxl>.

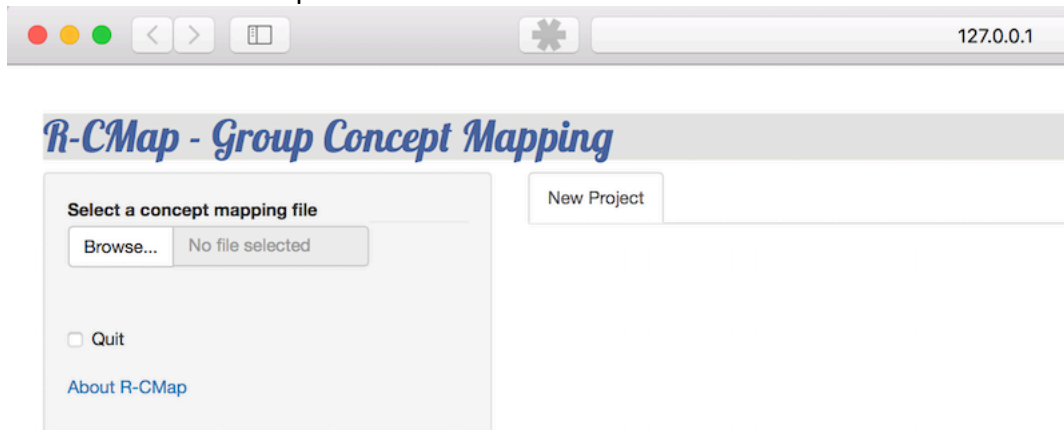
Xie, Yihui. 2016. *DT: A Wrapper of the Javascript Library 'Datatables'*. <https://CRAN.R-project.org/package=DT>.

Screenshots

Starting R-CMap:



Main screen of R-CMap:



The Plots tab



R-CMap - Group Concept Mapping

Plot type

- Point Map (MDS)
- Clusters
- Univariate Rating - by Statement
- Univariate Rating - by Cluster
- Bar Chart
- Pattern Matching
- GoZone
- Dendrogram

Rating variable

Feasibility

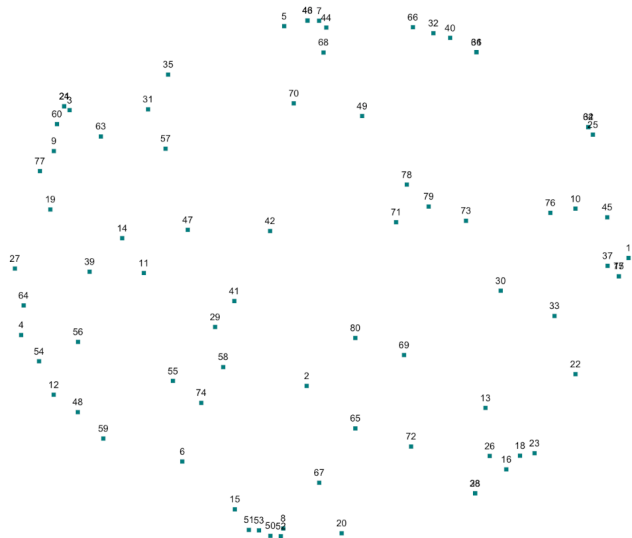
Include in report

Quit

[About R-CMap](#)

Summary **Plots** Reports Analysis Users Settings

55 : reduce middle management -- make organizational structure more horizontal (average rating=3.600)



The Reports tab

R-Map - Group Concept Mapping

Summary | Plots | **Reports** | Analysis | Users | Settings

Summary of statements, with Feasibility rating statistics for group All

ID	Statement	Mean	SD	Cluster ID
32	improve client satisfaction	4.60	0.52	1
66	take a "total quality service" approach to working with the client	4.00	0.82	1
73	investigate emerging partnership or new program opportunities	3.60	1.07	1
79	find out what our other similar programs are planning to do in our area	3.60	0.97	1
70	close facilities that are too old or are underutilized	3.50	1.43	1
71	develop better strategies to learn what other programs are doing	3.30	1.16	1
5	friendlier program managers	3.10	0.99	1
7	improve cleanliness of offices and program locations	3.10	0.88	1
37	improve targeting of messages to appropriate clients	3.10	0.74	1
68	find better ways to manage high-traffic (peak) periods in the facilities	3.10	0.99	1
36	perform regular client satisfaction surveys	3.00	0.94	1
61	conduct ongoing client focus groups to assure continuous feedback	3.00	0.82	1
10	conduct program effectiveness analysis for all major current programs	2.90	1.10	1
40	establish a client satisfaction/complaints telephone hotline	2.90	1.10	1
46	reduce the clutter of reading material and program information in the public areas of the main facility	2.90	1.20	1

The Analysis tab

R-Map - Group Concept Mapping

Summary | Plots | Reports | **Analysis** | Users | Settings

Comparing clusters (Tukey's test) for group All - Feasibility

	Difference	Lower CI	Upper CI	Adjusted p-value
2-1	0.27	-0.08	0.61	0.16
3-1	0.82	0.49	1.14	0.00
3-2	0.55	0.21	0.88	0.00

The Users tab

R-Map - Group Concept Mapping

Select Users

Group: All

Reset

Enter name of new group

Create Group Remove Group

Quit

[About R-Map](#)

Summary Plots Reports Analysis **Users** Settings

Search:

	CompanySize	Years
1	large	4.2
2	small	9.9
3	small	8.2
4	medium	6.8
5	small	7.2
6	large	8.6
7	small	1.2
8	small	2.3
9	large	0.5
10	medium	4.6

Showing 1 to 10 of 10 entries

The Settings tab

R-Map - Group Concept Mapping

Setting

Number of Clusters

Cluster Names

Save Project

Save Report

Number of clusters: 3

Quit

[About R-Map](#)

Summary Plots Reports Analysis Users **Settings**

Dendrogram showing hierarchical clustering of 80 items, with three clusters highlighted in red.