

Journal of Visualized Experiments

Integrating computerized linguistic analysis and social network analysis to capture addiction recovery capital in an online community

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE58851R2
Full Title:	Integrating computerized linguistic analysis and social network analysis to capture addiction recovery capital in an online community
Keywords:	addiction, recovery communities, online social interaction, social network analysis, computerised linguistic analysis
Corresponding Author:	Ana-Maria Bliuc Western Sydney University Sydney, AUSTRALIA
Corresponding Author's Institution:	Western Sydney University
Corresponding Author E-Mail:	A.Bliuc@westernsydney.edu.au
Order of Authors:	Ana-Maria Bliuc Muhammad Iqbal David Best
Additional Information:	
Question	Response
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US\$2,400)
Please indicate the city, state/province, and country where this article will be filmed . Please do not use abbreviations.	Sydney, New South Wales, Australia

TITLE:

Integrating Computerized Linguistic and Social Network Analyses to Capture Addiction Recovery Capital in an Online Community

AUTHORS & AFFILIATIONS:

Ana-Maria Bliuc¹, Muhammad Iqbal², David Best³

¹School of Social Sciences and Psychology, Western Sydney University, Australia

²Institute for Sustainable Industries and Liveable Cities (ISILC), Victoria University, Melbourne, Australia

³Centre for Regional Economic and Social Research, Sheffield Hallam University, United Kingdom

Corresponding Author:

Ana-Maria Bliuc (a.bliuc@westernsydney.edu.au)

Email Addresses of Co-authors:

Muhammad Iqbal (muhammad.iqbal@vu.edu.au)

David Best (d.best@shu.ac.uk)

KEYWORDS:

online social interaction, online community, addiction, recovery, supportive networks, social identity

SUMMARY:

The article describes a novel approach for analyzing dynamic online social interactions (in an online context) exemplified by a study of an online community of recovery from alcohol and drug addiction.

ABSTRACT:

The article describes a new methodology designed with the aim of finding a comprehensive, unobtrusive, and accurate way of capturing social recovery capital development in online communities of recovery from alcohol and drug (AOD) addiction. Recovery capital was conceptualised as both engagement in the online recovery community and identification with the community. To measure recovery capital development, naturally occurring data were extracted from the social media page of a specific recovery program, with the page being set up as a resource for a face-to-face recovery program. To map engagement with the online community, social network analysis (SNA) capturing online social interaction was performed. Social interaction was measured through the linkages between the online contributors/members of the online community as represented by program clients, staff, and supporters from the broader community. To capture markers of social identification with the online community, computerised linguistic analysis of the textual data (content from posts and comments) was conducted. Recovery capital captured in this way was analysed against retention data (a proxy outcome indicator), as days spent in the (face-to-face) recovery

program. The online data extracted was linked to participant data in regards to program retention to test prediction of a key recovery outcome. This approach allowed the examination of the role of online support communities and assessment of the association between recovery capital (developed via the online community of recovery) and recovery outcomes.

INTRODUCTION:

The presented method has been designed to capture alcohol and other drugs (AOD) addiction recovery capital in online contexts. In the field of addiction, recovery capital has been defined as “the sum total of one's resources that can be brought to bear on the initiation and maintenance of substance misuse cessation”¹. Recovery capital has been primarily measured through self-reports^{2,3} in face-to-face contexts. This approach provides an alternative method of measuring recovery capital in online contexts by capturing the quality and quantity of online interactions in online communities of recovery.

Given the steady increase in the use of online resources in the form of peer-support in a range of health-related issues^{4,5}, it is necessary to develop new methods to capture the quality of these resources. Online peer support occurs in the form of social interactions in online forums and communities. Supportive social interactions in these online contexts contribute to building recovery capital, which in turn has a positive impact on the recovery process^{6,7}. The method proposed presents a number of advantages over alternative methods. Firstly, it overcomes some of the limitations involving the use of self-report measures in addiction research, particularly around recall and self-presentational biases. While self-report measures are considered to have reasonable levels of reliability and validity, they are susceptible to biases and inaccuracies. To enhance accuracy and minimize bias, it has been recognized that there is a need to increase the use of novel measures and data collection situations designed to avoid or minimize these issues⁸. By accessing data naturally occurring in contexts where people in various stages of recovery interact spontaneously, and by using analysis methods that can extract meaningful information from these data (able to capture indicators of psychological states), biases due to social desirability (self-presentational) and inaccuracies due to limitations in recall can be reduced or even eliminated. Secondly, this method is highly efficient and cost-effective, as it relies on the extraction of already existing online data (i.e., in open online forums that are publicly accessible).

Described next is the method that was applied to a study of building recovery capital in an online community established to complement a traditional, face-to-face addiction recovery program for addicts in early recovery stages. In this case, online (social media) data were linked to program retention data, but the method can be also used in cases where linkage data is not available or accessible.

PROTOCOL:

The research described here was approved by the research ethics community at Sheffield Hallam University.

1. Setup

NOTE: Please refer to the attached R script provided as **Supplementary File 1**.

1.1. Load required packages (Rfacebook⁹, dplyr¹⁰, igraph¹¹, and openxlsx¹²) in R. Packages refer to functions, datasets, or compiled code that allow users to analyze, transform, or extract data.

1.2. Load (external) retention and user data into R as a data frame from a CSV file.

NOTE: Retention data refers to the number of days in which a client participates in the offline (traditional) addiction recovery program. It was provided by the administrator of the (offline) recovery program as recorded onto a CSV file with the participant name and number of days they have been involved in the program. The participant name was replaced by the anonymous ID number prior to being imported into R.

2. Data extraction from the online community (the social page of an addiction recovery community)

NOTE: This protocol applies to a social media page, but it can be adapted to different types of online communities. In the case of the Rfacebook package, it allows the user to extract data from the social media page into R.

2.1 Create a social media (Facebook) access token by following the guide on the referenced website¹³.

2.2. Create access token in R.

2.3. Using the “getGroup” function from Rfacebook, extract data from the social media page of the community of interest (e.g., content of post, number of comments and likes for each post, a unique ID number for each post, etc.). This data is then saved as a data frame.

NOTE: A data frame is essentially a table within R used to store data.

2.4. Using the “getPosts” function from Rfacebook, along with the Post IDs extracted in step 2.3, extract data about posts likes made on the page.

2.5. Using the “getPosts” function from Rfacebook, along with the Post IDs extracted in step 2.3, extract data on the comments made on each post (e.g., user IDs of people commenting the post, when the comment was made, how many likes the post received). This data is then saved as a data frame.

2.6. Using the comment IDs extracted in step 2.5, extract data on the “comment likes” made on each post (e.g., user IDs of people liking the comment). This data is then saved as a data frame.

2.7. Combine the posts, post likes, comments, and comment likes data into one data frame.

2.8. Add a monthly breakdown (i.e., month 1 to 8).

3. Calculation of social media activity made and received by each client

3.1. Calculate the number of posts, comments, post likes and comment likes made by each client.

3.2. Calculate the number of posts, comments, post likes and comment likes received by each client.

3.3 Join the data frame of social media activity made and received by each client to the retention data frame.

3.4 Calculate the difference between posts and comments with likes and no likes.

3.5 Calculate the difference between posts with comments and no comments.

3.6 Join the likes difference data to the retention data.

3.7 Join the comments difference data to the retention data.

3.8 Calculate all the likes made by each client.

3.9 Calculate all the likes received by each client.

3.10 Identify which users did not participate in social media group (i.e., no activity).

4. Conducting social network analysis

4.1. Create an edge list. An edge list is a list of relationships within the social network, which in this case is based on 1) liking posts and comments and 2) commenting on posts. This is done by looking at two columns within the dataset. The first column contains the anonymous ID of the person making the post, while the second contains the anonymous ID of the person liking or commenting on the post.

4.2. Create a vertex list. A vertex list is a list of all individuals in the group. This is done by converting the two columns in the list of relationships into one column, and removing duplicate anonymous IDs so only the unique anonymous ID is left.

4.3. Using the “graph.data.frame” and “get.adjacency” functions in the igraph package, create graph and graph matrix objects from the edge and vertex lists.

4.4. Using the “degree” and “betweenness” functions from the igraph package, obtain the

network statistics (degree and betweenness) of the online group.

5. Conducting computerized linguistic analysis in LIWC

5.1. Export textual social media data (i.e., posts and comments) and post/comment ID column into CSV files.

5.2. Import the CSV files of textual social media data into the Linguistic Inquiry Word Count (LIWC) software.

5.3. Generate LIWC categories and save to new CSV files. Do this by clicking on “Analyze Text”, then on “Excel/CSV file”, and clicking on the column containing the posts and comments to select the text to be analyzed. After LIWC has completed analyzing the textual data, save the output as a new CSV file.

5.4. Import the LIWC results CSV file into R, and merge with existing data. The data is matched by the post/comment ID column, which exists in both LIWC and existing data frames.

5.5. Calculate total LIWC scores for each user in posts and comments, then join to the retention data.

5.6. Calculate total LIWC scores for each user in all textual data (post and comments combined), then join to the retention data.

5.7. Remove NAs from the retention data data frame.

6. Conducting regression analysis (to determine if indicators of engagement with the online community predict retention in the offline recovery program)

6.1. Define the independent variables.

6.2. Using the “lm” function in base R, conduct linear regression analysis using the retention data as the dependent variable, and LIWC categories, comments, post likes, and comment likes as independent variables.

6.3 Combine regression analysis results into one data frame.

7. Creating monthly SNA maps

7.1. Prepare data frames for SNA Maps.

7.2. Create an edge list based on monthly cumulative social media activity.

7.3. Create a vertex list based on monthly cumulative social media activity.

7.4. Create graphs and graph matrices based on monthly cumulative social media activity.

7.5. Set the layout of SNA maps based on cumulative social media activity.

7.6. Add colors based on user roles.

7.7. Create SNA maps and save them to a file.

8. Calculating monthly cumulative social media activity of the social media group

8.1. Calculate monthly cumulative social media activity by staff, clients, and other members of the social media group.

8.2. Calculate monthly cumulative social media activity by all members of the social media group.

8.3. Join the monthly cumulative social media activity data frames together.

REPRESENTATIVE RESULTS:

A detailed description of representative results obtained using this method can be found in our recent work¹⁴, which was reviewed and received full approval from the research ethics committee of the institution at which the research was conducted. In the report described here, the study investigated whether online participation in a community of recovery contributes to the recovery process through recovery capital building (as captured by increased levels and quality of online social interactions and positive identity development). In other words, the study examined whether indicators of online recovery capital developed over the eight months of online data assessed and also predicted retention in a recovery program designed for fostering community involvement for addicts in early stages of recovery.

To map how participants interacted online, social network analysis (SNA) using data extracted from the social media page (n = 609) of a recovery community was conducted. A visual representation of the social network and its evolution is presented in **Figure 1**. The figure illustrates the activity in the online community observed each month for a period of 8 months in the form of connections between all participants in the online community (i.e., commenting on posts, liking posts, and liking comments). The number of connections that an “agent” in the network has determines how central they will be in the social network. Computerized linguistic analysis was used to assess the textual data (capturing social identity markers), and linear regression analysis was conducted to determine whether the indicators of recovery capital predicted program retention. These analyses indicated that program retention was indeed predicted by: (a) levels of group validation received in the form of comment likes and all likes received on the social media page, (b) position in the social network (network centrality), and (c) group identity and achievement (as captured by the linguistic content of online communication). The results supported the argument that, overall, positive social interactions

between members of an online recovery community are supportive of the recovery process. A summary of those findings is presented below.

[Figure 1 about here]

Descriptive statistics

Participants' levels of engagement with the online community were measured by computing the contributions of all participants in the online community as number of posts, comments, and likes made by staff, clients, and broader community members. **Table 1** presents a breakdown by type of contribution (as made by each category of participant) across 8 months.

[Insert Table 1 about here]

Determinants of retention in the program

The following hypotheses were tested: (1) program retention should be associated with indicators of recovery capital development (i.e., reflected in the quantity and quality of online interaction), and (2) program retention should also be associated with indicators of identity change, (i.e., indicators of positive recovery identity development). The quantity of online interaction was indicated by the a) number of posts made, b) number of comments made, c) number of post likes received, d) number of comment likes received, and e) number of all likes received.

To determine the quality of online interaction, network structure and language content were analyzed. More specifically, degree and betweenness coefficients derived from social network analysis (SNA) and linguistic indicators of positive affect derived from computerized linguistic analysis were used. As indicators of positive identity change (as identification with the recovery community) the frequency of use of the pronoun "we" and achievement words (e.g., try, goal, win, etc.) were used. Finally, the dependent variable (retention in the program) was indicated by the total number of days spent in the program (ranging from 86 to 464 days here). As shown by the results, levels of online interaction and in-group validation (as reflected by the number of likes received for posts and comments) predicted program retention (**Table 2**). Program retention was also predicted by identification markers (as captured by the use of the pronoun "we" in posts and of achievement words in both posts and comments). Finally, where participants are situated within the social network (i.e., degree of centrality) also represents an important aspect of retention (**Table 2**).

[Insert Table 2 about here]

FIGURE & TABLE LEGENDS:

Figure 1: Monthly representations of the social network of the online community over 8 months suggest changes in the pattern of social interactions between the participants. These representations illustrate how at the start, most of the client members in the online community (clients of the offline recovery program) are mostly disconnected, and it is the program staff

and only a small number of clients who drive the online activity. However, this gradually changes, so that after 8 months, the clients are the ones most connected (therefore the most central), with the highest number of connections in the network (figure is adapted from a previous publication)¹⁴.

Table 1: Shown is the number of online contributions by type (post and comments made, likes given to posts, and likes given to comments) by members of the online community on the across 8 months. The members of the online community are classified as staff (support staff employed by the offline recovery program), clients (people in recover who are participating in the offline recovery program), and others (supporters and pro-recovery advocates from the broader community).

Table 2: Retention time as predicted by online engagement, network statistics, and linguistic categories.

DISCUSSION:

The approach described here is based on a new method of measuring how online group processes can impact retention in an addiction recovery program. Applying this method to an online community of recovery from addiction, it was found that there were four key aspects predicted program retention: being highly involved in the online community, being central in the online social network, positive affect expressed in communication with other members of the online community, and receiving validation from others for contributions to the network¹⁴. The findings obtained by using this method support existing theoretical models of recovery. That is, two key models in the recovery literature, the Social Identity Model of Recovery¹⁵ and the Social Identity Model of Cessation Maintenance¹⁶, both emphasize the importance of active participation in groups which are supportive of recovery. Both models suggest that increased identification and commitment to such groups contribute to lower future contact with using groups and consequent relapse.

As illustrated in our research, the method allowed us to map out trajectories of recovery or change of individual members of the online community¹⁴. Visualizations of the online social networks and their evolution over time can provide valuable information about the movement of members of the online community from the periphery to the center of the network and vice-versa (these movements in the network indicate changes in levels of engagement with the online community). In a 2017 study¹⁴, interviews with members of the online community who undertook the most significant changes in terms of movement from the periphery to center of the networks were conducted as a way of triangulating our findings based on SNA, computerized linguistic analysis, and regression against retention data. Future studies may focus instead on those members who became disengaged with the online community, on those who never become engaged, or on more direct measures of outcome such as substance use and reoffending. This methodology can further fine-tuned to be used in intervention programs, for example, for assessing the role of moderators in help forums.

There are currently no studies providing evidence on the benefits of the method described here

when used by itself (the method described was used in conjunction with retention data and triangulated with qualitative data from interviews with key online community members¹⁴), but this approach can provide accurate and bias-free data that can complement self-reporting and other measures in studies of addiction recovery.

This method was applied to examine online social interactions in the context of a social media page established as a complementary form of support to a standard, face-to-face recovery program. However, with minor changes, the method can be used to investigate online social interactions in other types of online communities (online forums, discussion groups, chat rooms, commentary websites, etc.). One of the key advantages of this method is that it can be adapted and applied to contexts beyond communities of addiction recovery to any online community. For example, in our own political psychology research, we use a similar method (developed from the method described here) to capture the quality of online interactions and changes in these interactions between members of far-right online communities. In effect, the method can be applied to any online community in which data in the form of connections between members (as social network linkages) and linguistic content can be extracted.

However, in accessing and working with online data, researchers need to be aware of ethical issues, some which apply to self-reporting and other types of data in general and some which are only encountered in an online environment. In the research described here (which was approved by the research ethics community at Sheffield Hallam University), consent was obtained from the organization managing the recovery program, and strict measures were taken to ensure complete anonymity of participants in the open social media page (e.g., after online and retention data matching, all identifying information was removed from the files and also no potentially self-identifying quotes were used from the publicly accessible online communication).

Close communication with the organization also ensured that the participants in the program were aware of the study and research findings, and one of the researchers met regularly with the group to explain the study and its results. In other cases, however, where online communities are not associated with specific offline programs, it may be harder to determine who should be asked for consent regarding data extraction (applicable especially in unmoderated forums, where people in recovery seek online peer support). While the general principles of ethical research will apply, researchers need to adopt a case-by-case approach to ensure that the extraction and analysis of online data does not pose any significant risks to the participants (e.g., compromising privacy).

ACKNOWLEDGMENTS:

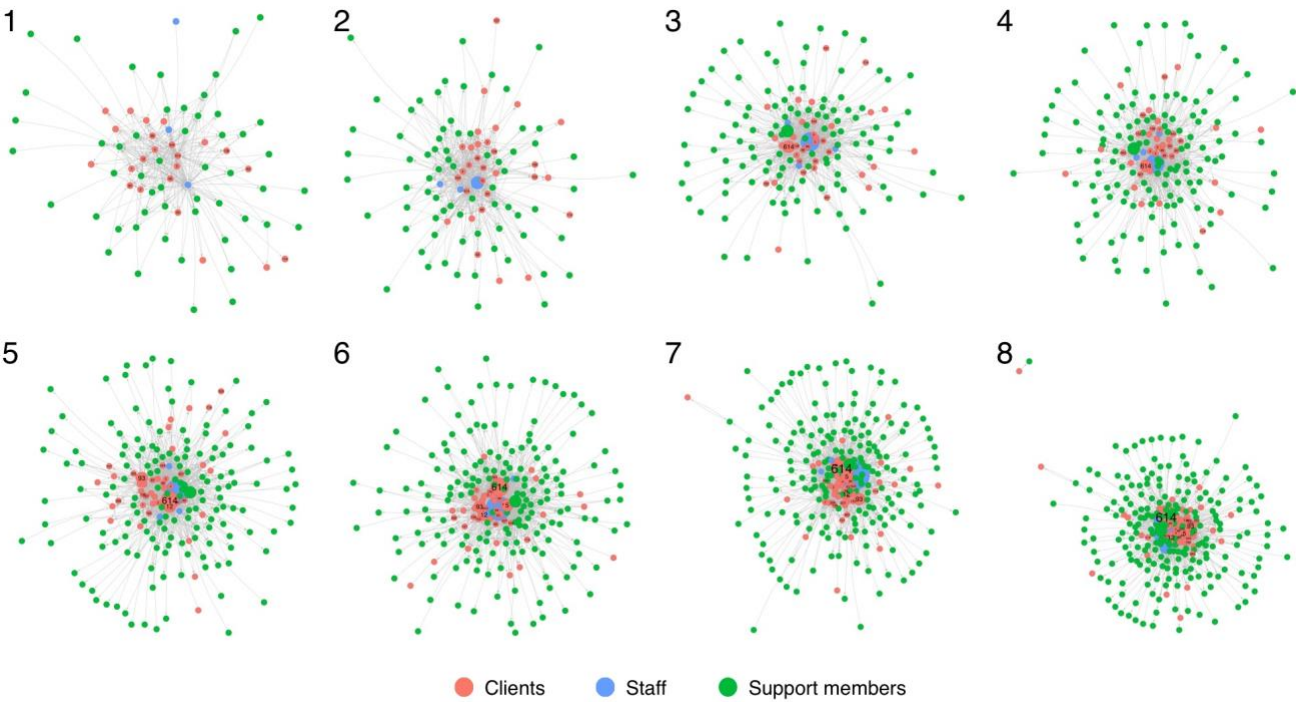
We are grateful to the clients and staff of Jobs, Friends and Houses, UK, who supported and agreed to participate in our research.

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES

1. Cloud, W., Granfield, R. Conceptualizing recovery capital: Expansion of a theoretical construct. *Substance Use and Misuse*. **43**, 1971-1986 (2008).
2. Best, D. et al. Mapping the recovery stories of drinkers and drug users in Glasgow: Quality of life and its associations with measures of recovery capital. *Drug and Alcohol Review*. **31** (3), 334-341 (2012).
3. Laudet, A. B., White, W. L. Recovery capital as prospective predictor of sustained recovery, life satisfaction, and stress among former poly-substance users. *Substance Use and Misuse*. **43** (1), 27-54 (2008).
4. Moorhead, S. A. et al. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of Medical Internet Research*. **15**, e85 (2013).
5. White, M., Dorman, S. M. Receiving social support online: implications for health education. *Health Education Research*. **16**, 693-707 (2001).
6. Best, D., Bliuc, A. M., Iqbal, M., Upton, K., Hodgkins, S. Mapping social identity change in online networks of addiction recovery. *Addiction Research and Theory*. **26** (3), 163-173 (2018).
7. Bliuc, A. M., Best, D., Beckwith, M., Iqbal, M. Online support communities in addiction recovery. *Addiction, behavioral change and social identity: The path to resilience and recovery*. 137, (2016).
8. Del Boca, F. K., Darkes, J. The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction* **98**, 1-12 (2003).
9. Barbera, P., Package 'Rfacebook', <https://cran.r-project.org/web/packages/Rfacebook/Rfacebook.pdf> (2017).
10. Wickham, H., François, R., Henry, L., Müller, K. Package 'dplyr', <https://cran.r-project.org/web/packages/dplyr/dplyr.pdf> (2018).
11. Csárdi, G. Package 'igraph', <https://cran.r-project.org/web/packages/igraph/igraph.pdf> (2018)
12. Walker, A., Braglia, L. Package 'openxlsx', <https://cran.r-project.org/web/packages/openxlsx/openxlsx.pdf> (2018)
13. How to get a Facebook access token which never expires. <<https://smashballoon.com/custom-facebook-feed/access-token/>> (2018).
14. Bliuc, A. M., Best, D., Iqbal, M., Upton, K. Building addiction recovery capital through online participation in a recovery community. *Social Science and Medicine*. **193**, 110-117 (2017).
15. Best, D. et al. Overcoming alcohol and other drug addiction as a process of social identity transition: The Social Identity Model of Recovery (SIMOR). *Addiction Research and Theory*. **24**, 111-123 (2016).
16. Frings, D., Albery, I. P. The social identity model of cessation maintenance: Formulation and initial evidence. *Addictive Behaviors*. **44**, 35-42 (2015).



Group members	Type of online contribution	Month 1	Month 2	Month 3	Month 4	Month 5
All	Posts and comments	382	388 (770)	579 (1349)	369 (1718)	530 (2248)
	Post likes given	1167	878 (2045)	1856 (3901)	1440 (5341)	1880 (7221)
	Comment likes given	784	970 (1604)	825 (2429)	171 (2600)	634 (3234)
Staff	Posts and comments	129	106 (235)	170 (405)	96 (501)	185 (686)
	Post likes given	188	147 (335)	302 (637)	209 (846)	385 (1231)
	Comment likes given	168	303 (471)	237 (708)	69 (777)	168 (945)
Clients	Posts and comments	145	155 (300)	214 (514)	132 (646)	208 (854)
	Post likes given	365	252 (617)	415 (1032)	303 (1335)	549 (1884)
	Comment likes given	143	318 (461)	235 (696)	33 (729)	143 (872)
Others	Posts and comments	108	127 (235)	195 (430)	141 (571)	137 (708)
	Post likes given	614	479 (1093)	1139 (2232)	928 (3160)	946 (4106)
	Comment likes given	473	349 (672)	353 (1025)	69 (1094)	323 (1417)

Month 6	Month 7	Month 8
581 (2829)	796 (3625)	674 (4299)
1756 (8977)	2667 (11644)	1857 (13501)
970 (4204)	825 (5029)	171 (5200)
176 (862)	227 (1089)	316 (1405)
372 (1603)	567 (2170)	511 (2681)
303 (1248)	237 (1485)	69 (1554)
286 (1140)	419 (1559)	253 (1812)
529 (2413)	898 (3311)	576 (3887)
318 (1190)	235 (1425)	33 (1458)
119 (827)	150 (977)	105 (1082)
855 (4961)	1202 (6163)	770 (6933)
349 (1766)	353 (2119)	69 (2188)

Variable	<i>B</i>	<i>SE</i>	β	<i>R</i> ²
Comment likes received	0.43	0.18	.47 [*]	0.22
Likes received (all)	0.08	0.03	.43 [*]	0.18
Comment-like difference	1.09	0.5	.43 [*]	0.19
Network degree	0.01	0	.43 [*]	0.18
LIWC We (Post)	3.89	1.76	.43 [*]	0.19
LIWC Achievement (Post)	0.56	0.26	.43 [*]	0.18
LIWC Achievement (All)	0.14	0.07	.42 [*]	0.17

Name of Material/ Equipment	Company	Catalog Number
LIWC software	Receptiviti	https://liwc.wpengine.com/
R software	n/a	https://www.r-project.org/

Comments/Description

computerised linguistic analysis software
free statistical and data visualisation software

ARTICLE AND VIDEO LICENSE AGREEMENT - UK

Title of Article:	Integrating computerized linguistic analysis and social network analysis to capture addiction recovery capital in an online community
Author(s):	Ana-Maria Bliuc, Muhammad Iqbal, David Best

Item 1: The Author elects to have the Materials be made available (as described at <http://www.jove.com/publish>) via:

☐

Standard Access

☒

Open Access

Item 2: Please select one of the following items:

☒

The Author is **NOT** a United States government employee.

☐

The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐

The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. **Defined Terms.** As used in this Article and Video License Agreement, the following terms shall have the following meanings: "**Agreement**" means this Article and Video License Agreement; "**Article**" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "**Author**" means the author who is a signatory to this Agreement; "**Collective Work**" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "**CRC License**" means the Creative Commons Attribution 3.0 Agreement (also known as CC-BY), the terms and conditions of which can be found at: <http://creativecommons.org/licenses/by/3.0/us/legalcode>; "**Derivative Work**" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "**Institution**" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "**JoVE**" means MyJoVE Corporation, a Massachusetts corporation and the publisher of The Journal of Visualized Experiments; "**Materials**" means the Article and / or the Video; "**Parties**" means the Author and JoVE; "**Video**" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion

of the Article, and in which the Author may or may not appear.

2. **Background.** The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. **Grant of Rights in Article.** In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4 and 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

4. **Retention of Rights in Article.** Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. **Grant of Rights in Video - Standard Access.** This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. **Grant of Rights in Video - Open Access.** This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats.

7. **Government Employees.** If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be

deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. **Protection of the work.** The Author(s) authorize JoVE to take steps in the Author(s) name and on their behalf if JoVE believes some third party could be infringing or might infringe the copyright of either the Author's Article and/or Video.

9. **Likeness, Privacy, Personality.** The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

10. **Author Warranties.** The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

11. **JoVE Discretion.** If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole

ARTICLE AND VIDEO LICENSE AGREEMENT - UK

discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

12. **Indemnification.** The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in JoVE or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or


decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

13. **Fees.** To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US\$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

14. **Transfer, Governing Law.** This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to be one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement is required per submission.

CORRESPONDING AUTHOR

Name:	Ana-Maria Bliuc		
Department:	School of Social Sciences and Psychology		
Institution:	Western Sydney University		
Title:	Dr		
Signature:		Date:	31/01/2019

Please submit a **signed** and **dated** copy of this license by one of the following three methods:

1. Upload an electronic version on the JoVE submission site
2. Fax the document to +1.866.381.2236
3. Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02140

Dear Dr. DSouza,

We have now revised our manuscript, JoVE58851 "Integrating computerized linguistic analysis and social network analysis to capture addiction recovery capital in a virtual community", according to the reviewers' feedback - please see enclosed the revised manuscript (all revisions are tracked) and the Revisions letter where we describe how all of the reviewers' comments were addressed.

We have also included a signed Author Licence Agreement and copyright permission to reuse figures from Bliuc et al. (2017).

Sincerely,

Ana-Maria Bliuc (on behalf of all the authors)

Editorial Comments	Our Response
1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues.	The manuscript was proofread and further edited.
2. Please print and sign the attached Author License Agreement - UK. Please then scan and upload the signed ALA with the manuscript files to your Editorial Manager account. As some authors are affiliated with UK institutions, can you please check whether open access is required by your funding agencies.	Please see attached.
3. Please obtain explicit copyright permission to reuse any figures from a previous publication. Explicit permission can be expressed in the form of a letter from the editor or a link to the editorial policy that allows re-prints. Please upload this information as a .doc or .docx file to your Editorial Manager account. The Figure must be cited appropriately in the Figure Legend, i.e. "This figure has been modified from [citation]."	Please see permission attached and appropriate citation of the figure was included in the Figure Legend.
4. Please revise lines 150-152, 160-162, 169-184, 203-205, and 207-209 to avoid previously published text.	All these lines have been now revised.
<p>5. Please note that your protocol will be used to generate the script for the video and must contain everything that you would like shown in the video. Software must have a GUI (graphical user interface) and software steps must be more explicitly explained ('click', 'select', etc.). Please add more specific details (e.g. button clicks for software actions, numerical values for settings, etc.) to your protocol steps.</p> <p>Please provide software screenshots as supplementary files to match each step. Some examples:</p> <p>6. 1.1.2: Please specify the online community page. Is the weblink pasted somewhere in the function?</p> <p>7. 1.1.4 and 1.1.5: Should they refer to step 1.1.3 instead of 1.1.2 as ID number is extracted during step 1.1.3?</p> <p>8. 1.1.6: Please specify the criteria for cleaning and merging data. Are the data extracted in steps 1.1.2-1.1.5 exported somewhere?</p> <p>9. 2.1: What are retention data? How are they</p>	<p>We have now included more detail for each of the steps and we attach screenshots as supplementary files.</p> <p>Some limitations of the method are described in the discussion, and the References and the other issues have been addressed.</p>

<p>obtained and imported?</p> <p>10. 3.1.1-3.1.2: These steps are unclear. Please describe how to create different lists.</p> <p>11. 4.3: Please describe how.</p> <p>12. 4.4: Please describe how to match with existing textural data.</p> <p>13. Line 148: Please describe Figure 1 in more detail. For instance, what do different panels show?</p> <p>14. Table 2: Please explain what the asterisk symbols represent.</p> <p>15. Discussion: As we are a methods journal, please also discuss critical steps within the protocol, any modifications and troubleshooting of the technique, and any limitations of the technique.</p> <p>16. References: Please do not abbreviate journal titles. If there are six or more authors, list the first author and then "et al."</p> <p>17. Table of Materials: Please sort the items in alphabetical order according to the name of material/equipment.</p>	
Reviewers' comments:	Our Response
<p>Reviewer #1:</p> <p>Manuscript Summary:</p> <p>The article did a wonderful job of explaining the how and why of an innovative method useful in evaluating recovery capital.</p> <p>Major Concerns:</p> <p>N/A</p>	<p>We thank the reviewer for the constructive feedback.</p>
<p>Minor Concerns:</p> <p>Line 39 (and abstract): the information about how the group was established is a bit confusing. I might flip it around to write "the facebook page is a resource for a face-to-face recovery group". - I might also indicate that the group involves members, staff, and community members/other and indicate what "community members" means.</p>	<p>We have integrated the feedback in the revision by changing the wording in the Long Abstract where we now say:</p> <p>"To measure recovery capital development, naturally occurring data was extracted from the Facebook page of a specific recovery program - with the Facebook page being set-up as a resource to support a face-to-face recovery program."</p> <p>and</p> <p>"Social interaction was measured through the linkages between the online contributors/members of the online community as represented by program clients, staff and supporters from the broader community".</p>
<p>Line 69: You note traditional self-report measures having biases and inaccuracies - it might be useful to note some examples of what that looks like, and then indicate how this method reduces them.</p>	<p>We have changed that statement as follows:</p> <p>"Firstly, our method overcomes some of the limitations due to the use of self-report measures in addiction research, particularly around recall and self-presentational biases. While self-report measures are</p>

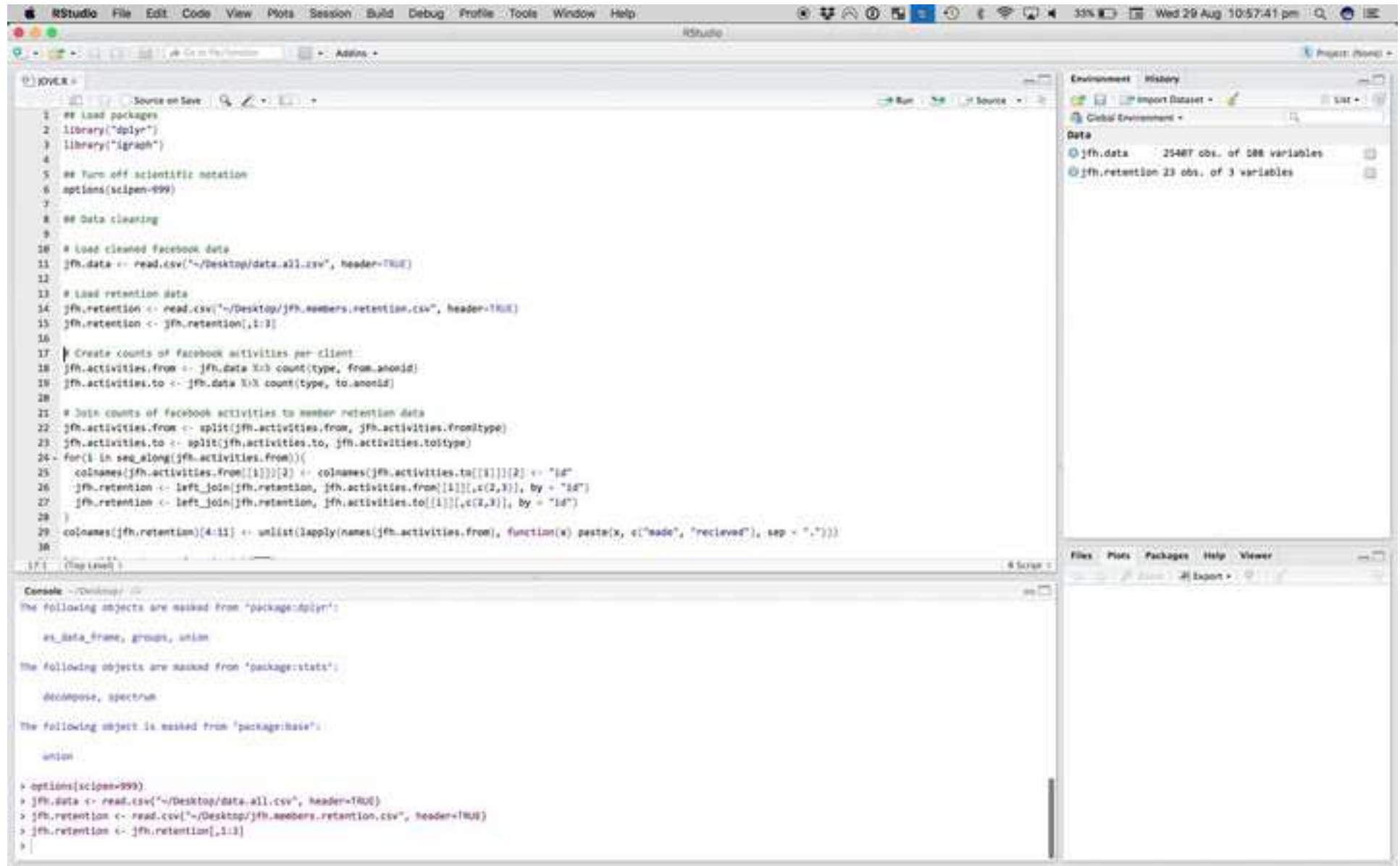
	<p>considered to have reasonable levels of reliability and validity, they are susceptible to biases and inaccuracies. To enhance accuracy and minimize bias, it has been recognized that there is a need to increase the use of novel measures and data collection situations designed to avoid or minimize these issues⁸. By accessing data naturally occurring in contexts where people in various stages of recovery interact spontaneously and by using analysis methods which can extract meaningful information from this data (able to capture indicators of psychological states), biases due to social desirability (self-presentational) and inaccuracies due to limitations in recall can be reduced or even eliminated.”</p>
Line 81: A brief sentence or two describing the basics of programming would be helpful to novice users (i.e., An R-package is....)	<p>We now say:</p> <p>“ (...)Packages refer to functions, datasets, or compiled code that allow users to analyze, transform, or extract data. In the case of the Rfacebook package, it allows the user to extract data from Facebook into R.”</p>
Line 141: "Contributes the recovery" should be "contributes to..."	This has been corrected.
Line 143: should be "...eight months, and predicted..."	This has been corrected.
Line 144: rather than "designed around..." should be "designed for fostering..."	This has been corrected.
Briefly explain what Figure 1 is showing - for those not as familiar with network analysis	<p>We have now added the following explanation of Figure 1:</p> <p>“The figure illustrates the activity in the online community observed each month for a period of 8 months, in the form of connections between all participants in the online community (i.e., commenting on posts, liking posts, and liking comments). The number of connections that an ‘agent’ in the network has will determine how central they will be in the social network.”</p>
Line 172: list: commas should be used rather than semi colons	This has been corrected.
Line 177: range of time in program, small to larger number (i.e., 86-464	This has been corrected

Line 178: remove "as the" in "as the number of likes received"	This has been corrected
Line 183: comma after "posts"	This has been corrected
Figure 1: A bit more explanation needed. "movement from periphery to center, indicating...."	We now say: "Visualizations of the online social networks and their evolution over time can provide valuable information on the movement of members of the online community from the periphery to the center of the network and vice-versa (these movements in the network indicate changes in levels of engagement with the online community)."
Line 201: remove "upon" from "impact upon retention..."	This has been corrected.
Line 212-214: This is very confusing. Do you mean the study described in this article? Above what method? Where was the retention data from? The face-to-face program? More information is needed here.	We changed the order of the paragraphs in the Discussion, so we present more specific explanations of the use of the method before this paragraph (see comment below) - the meaning of this paragraph should now be clear.
Lines 233-237: some commas are needed in that sentence, and perhaps break the sentence up into more than one sentence. Perhaps that final paragraph (lines 230-240) could be moved closer to the beginning of the discussion, as it explains how/why the method was used most fully.	We have moved this paragraph in the Discussion section (so this is now the second paragraph in the Discussion).
Reviewer #2: Manuscript Summary: This manuscript describes in more detail the methods that were used to produce findings for a study of recovery capital built through online communities (outcomes already published in another journal). It is quite an interesting method (although really it is a series of methods and analyses) and I could see it being especially useful to those working with health behaviors and online communities, as well as for those interested in doing this sort of data collection.	We thank the reviewer for the constructive feedback.
Overall, I like that the authors have spent time going into more detail for readers the methods used (than could be provided with the published report), but I found myself wondering at the aim of the article. The text starting on line 79 (under protocol) looks like step-by-step instructions for the analysis, but I was not sure whether this article is meant to be a tutorial or something else. That is, if a tutorial, I would find that more specific step-by-step instructions, rationale and output would be	We have now added significant additional information for each of the steps of the protocol, so the resulting video will provide enough information to make the method reproducible and applicable to other research. Perhaps the fact that the protocol and accompanying materials are to be used to create the script for the video article (a method paper) has not been clearly conveyed. However, the aims should be clear in the finalised iteration of the article.

<p>useful. As a tutorial there is far too little information. If not a tutorial, then I am not sure of the authors' aims because the collection of data from a social media site such as facebook is not necessarily an innovative approach as it is gaining traction among many fields of health behaviors. I struggle to make direct suggestions because the aim of the article is still not clear to me, although I think it could be useful with a clearer aim, framing, and different organization.</p>	
<p>Minor issue: Should something about institutional approval be stated/included here? (I would be curious to know if there were institutional review board issues with collecting this kind of data, especially given all the privacy issues social media sites like Facebook have come up against more recently!)</p>	<p>We have added a discussion of ethical issues in the Discussion where we now say:</p> <p>“In effect, the method can be applied to any online community where data in the form of connections between members (as social network linkages) and linguistic content can be extracted. However, in accessing and working with online data, researchers need to be aware of ethical issues, some which are applicable to self-report and other types of data more generally, and some which may only be encountered in the online environment. In the research described here, consent was obtained from the organization managing the recovery program and strict measures were taken to ensure the complete anonymity of the participants in the open Facebook page (e.g., after online and retention data matching, all identifying information was removed from the files and also no potentially self-identifying quotes were used from the publicly accessible online communication). Close communication with the organization also ensured that the participants in the program were aware of the study and the research findings, and the research team met with members of the recovery community on a number of occasions to discuss the study. In other cases, however, where online communities are not associated with specific offline programs, it might be harder to determine the entity who should be asked for consent for the data extraction (this would apply especially in unmoderated forums where people in recovery seek online peer support). While the general principles of ethical research would apply, researchers need to adopt a case-by-case approach to ensure that the extraction and analysis of online data does not pose any significant risks to the participants (e.g., compromising their privacy).”</p>
<p>Reviewer #3: Manuscript Summary: The paper presents interesting analytic methods and findings related to social support groups on</p>	<p>We thank the reviewer for the constructive feedback.</p>

<p>social media and their role on fostering support for addiction problems. However, the current version of the manuscript doesn't fully describe data collection procedures and ethical standards that were practiced during this research on vulnerable populations.</p>	<p>See response to Reviewer #2 for more detailed feedback</p>
<p>Major Concerns: My concern lies on the ethical process of data mining especially regarding the Facebook group you selected for social network analysis. You mentioned data were mined a Facebook group. Data from Face Pages are open to the public, but some Facebook Groups are set for "closed" or "secret" settings. The group participants share their issues with an expectation that the Group(s) is "closed" thus their communications won't be used for other purposes or visible to outside people (non-members). I wonder if authors reached the group admin's prior to mining the data or received permission from the Group members in terms of using their communication data for the study. There was no information how the IRB reviewed the protocol.</p>	<p>We agree with the reviewer in regards with the importance of following strict ethical guidelines in conducting this type of research. The particular Facebook page we used was open and we obtained approval to conduct the research from the organization running the recovery program (prior to receiving ethics approval from the university). A discussion of potential ethical concerns was now included in the "Discussion" section. We now say (see also response to Reviewer 2):</p> <p>"However, in accessing and working with online data, researchers need to be aware of ethical issues, some which are applicable to self-report and other types of data more generally, and some which may only be encountered in the online environment. In the research described here, consent was obtained from the organization managing the recovery program and strict measures were taken to ensure the complete anonymity of the participants in the open Facebook page (e.g., after online and retention data matching, all identifying information was removed from the files and also no potentially self-identifying quotes were used from the publicly accessible online communication). Close communication with the organization also ensured that the participants in the program were aware of the study and the research findings, and the research team met with members of the recovery community on a number of occasions to discuss the study.. In other cases, however, where online communities are not associated with specific offline programs, it might be harder to determine the entity who should be asked for consent for the data extraction (this would apply especially in unmoderated forums where people in recovery seek online peer support). While the general principles of ethical research would apply, researchers need to adopt a case-by-case approach to ensure that the extraction and analysis of online data does not pose any significant risks to the participants (e.g., compromising their privacy)."</p>
<p>It's not clear if you used data from Face Pages or Facebook Groups because you mentioned "Facebook group page" (e.g., on page 1 of 6 -the protocol section; Line 80). These two platforms have different privacy settings, which can affect</p>	<p>We have specified that we refer to Facebook – we now say:</p> <p>"(...) i.e., the Facebook page of an addiction recovery community)"</p>

study designs. Will need clarification.	
<p>Minor Concerns: Figure 1.1 ~ 1.8 present social network analysis maps, but to fully understand each figure authors should provide more information on the features and scientific interpretation of each sub-figure in Figure 1. Also, you used the term "clients" - how "clients" different from "members"?</p>	<p>We have included more detail on what Figure 1 represents (also in response to Reviewer 1). We now say in the text:</p> <p>“The figure illustrates the activity in the online community observed each month for a period of 8 months, in the form of connections between all participants in the online community (i.e., commenting on posts, liking posts, and liking comments). The number of connections that an ‘agent’ in the network has will determine how central they will be in the social network.”</p> <p>In the Figure 1 legend we now specify:</p> <p>“Figure 1. Monthly representations of the social network of the online community over 8 months suggest changes in the pattern of social interactions between the participants. Different types of participants in the online community are represented with different colors: red indicates clients (participants who are part of the offline recovery program), blue, staff members of the recovery program, and green, members of the broader community who are supportive of the clients (family, friends, pro-recovery advocates, etc.). This figure has been adapted from Bliuc, Best, Iqbal & Upton (2017).”</p>

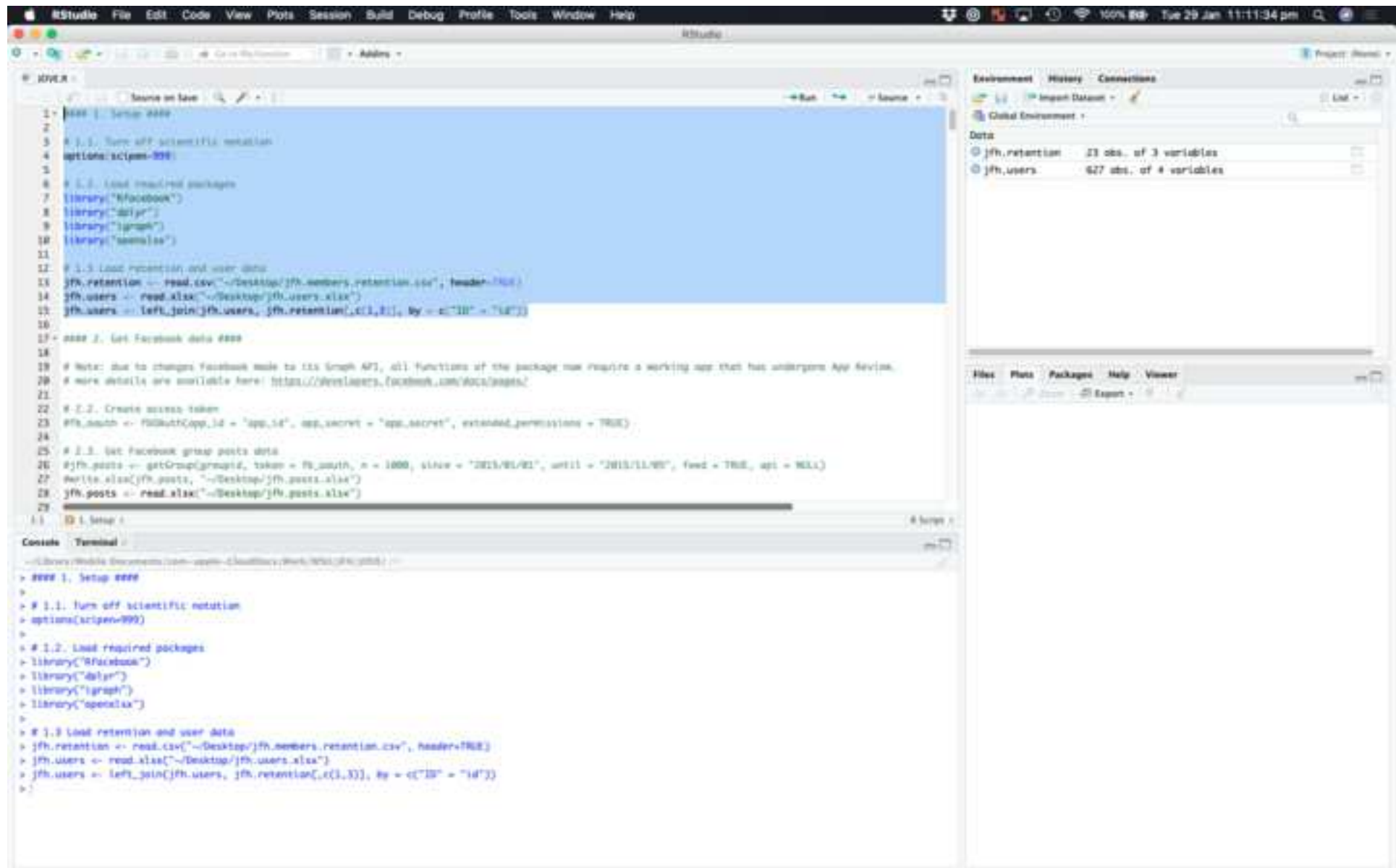


The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains an R script with the following code:


```
1 # Load packages
2 library("dplyr")
3 library("igraph")
4
5 # Turn off scientific notation
6 options(scipen=999)
7
8 # Data cleaning
9
10 # Load cleaned facebook data
11 jfh.data <- read.csv("~/Desktop/data.all.csv", header=TRUE)
12
13 # Load retention data
14 jfh.retention <- read.csv("~/Desktop/jfh.members.retention.csv", header=TRUE)
15 jfh.retention <- jfh.retention[,1:3]
16
17 # Create counts of facebook activities per client
18 jfh.activities.from <- jfh.data %>% count(type, from.anonid)
19 jfh.activities.to <- jfh.data %>% count(type, to.anonid)
20
21 # Join counts of facebook activities to member retention data
22 jfh.activities.from <- split(jfh.activities.from, jfh.activities.from$type)
23 jfh.activities.to <- split(jfh.activities.to, jfh.activities.to$type)
24 for(i in seq_along(jfh.activities.from)){
25   colnames(jfh.activities.from[[i]]$count) <- "id"
26   jfh.retention <- left_join(jfh.retention, jfh.activities.from[[i]][,c(2,3)], by = "id")
27   jfh.retention <- left_join(jfh.retention, jfh.activities.to[[i]][,c(2,3)], by = "id")
28 }
29 colnames(jfh.retention)[4:11] <- unlist(lapply(names(jfh.activities.from), function(x) paste(x, c("made", "received"), sep = ".")))
30
31 # End of script
```
- Console:** Shows the output of the script execution, including package loading messages and the final data structure:


```
> options(scipen=999)
> jfh.data <- read.csv("~/Desktop/data.all.csv", header=TRUE)
> jfh.retention <- read.csv("~/Desktop/jfh.members.retention.csv", header=TRUE)
> jfh.retention <- jfh.retention[,1:3]
>
```
- Environment:** Displays the objects in the global environment:
 - `jfh.data`: 25487 obs. of 588 variables
 - `jfh.retention`: 23 obs. of 3 variables



The screenshot displays the RStudio interface with a script editor, console, and environment pane.

Script Editor:

```
1> #### 1. Setup ####
2
3 # 1.1. Turn off scientific notation
4 options(scipen=999)
5
6 # 1.2. Load required packages
7 library("Rfacebook")
8 library("dplyr")
9 library("igraph")
10 library("openssl")
11
12 # 1.3 Load retention and user data
13 jfh.retention <- read.csv("~/Desktop/jfh_members_retention.csv", header=TRUE)
14 jfh.users <- read.xlsx("~/Desktop/jfh_users.xlsx")
15 jfh.users <- left_join(jfh.users, jfh.retention[,c(1,3)], by = c("ID" = "id"))
16
17- #### 2. Get Facebook data ####
18
19 # Note: due to changes Facebook made to its Graph API, all functions of the package now require a working app that has undergone App Review.
20 # more details are available here: https://developers.facebook.com/docs/pages/
21
22 # 2.1. Create access token
23 fb_auth <- FBAuthAppId = "app_id", app_secret = "app_secret", extended_permissions = TRUE)
24
25 # 2.2. Get Facebook group posts data
26 jfh.posts <- getGroup(groupId, token = fb_auth, n = 1000, since = "2015/01/01", until = "2015/11/01", feed = TRUE, api = NULL)
27 write.xlsx(jfh.posts, "~/Desktop/jfh_posts.xlsx")
28 jfh.posts <- read.xlsx("~/Desktop/jfh_posts.xlsx")
29
30 1. Setup
```

Console:

```
> #### 1. Setup ####
>
> # 1.1. Turn off scientific notation
> options(scipen=999)
>
> # 1.2. Load required packages
> library("Rfacebook")
> library("dplyr")
> library("igraph")
> library("openssl")
>
> # 1.3 Load retention and user data
> jfh.retention <- read.csv("~/Desktop/jfh_members_retention.csv", header=TRUE)
> jfh.users <- read.xlsx("~/Desktop/jfh_users.xlsx")
> jfh.users <- left_join(jfh.users, jfh.retention[,c(1,3)], by = c("ID" = "id"))
>
```

Environment:

Object	Class	Attributes
jfh.retention	data.frame	23 obs. of 3 variables
jfh.users	data.frame	627 obs. of 4 variables

The screenshot displays the RStudio interface with a script editor, console, and environment pane.

Script Editor: The code is organized into sections for data collection. The visible section (lines 24-52) includes:

- Section 2.3:** Retrieves Facebook group posts data using `getGroup(groupId, token = fb_auth, n = 1000, since = "2013/01/01", until = "2013/11/01", feed = TRUE, api = MSL)`. It writes the data to `~/Desktop/jfh.posts.xlsx`, reads it back, and joins it with `jfh.users` by `from_name`.
- Section 2.4:** Defines a function `getPostLikes` to retrieve post likes. It uses `lapply` to iterate over a list of post IDs, applying the function to each. The function uses `getPost(postIDlist[x], token=fb_auth@likes)` and returns a list of likes.
- Section 2.5:** Retrieves Facebook group comments data using `getGroup(groupId, token = fb_auth, n = 1000, since = "2013/01/01", until = "2013/11/01", feed = TRUE, api = MSL)`. It writes the data to `~/Desktop/jfh.posts.xlsx`, reads it back, and joins it with `jfh.users` by `from_name`.

Environment Pane: Shows the following data objects:

- `jfh.posts`: 887 obs. of 14 variables
- `jfh.retention`: 23 obs. of 3 variables
- `jfh.users`: 627 obs. of 4 variables

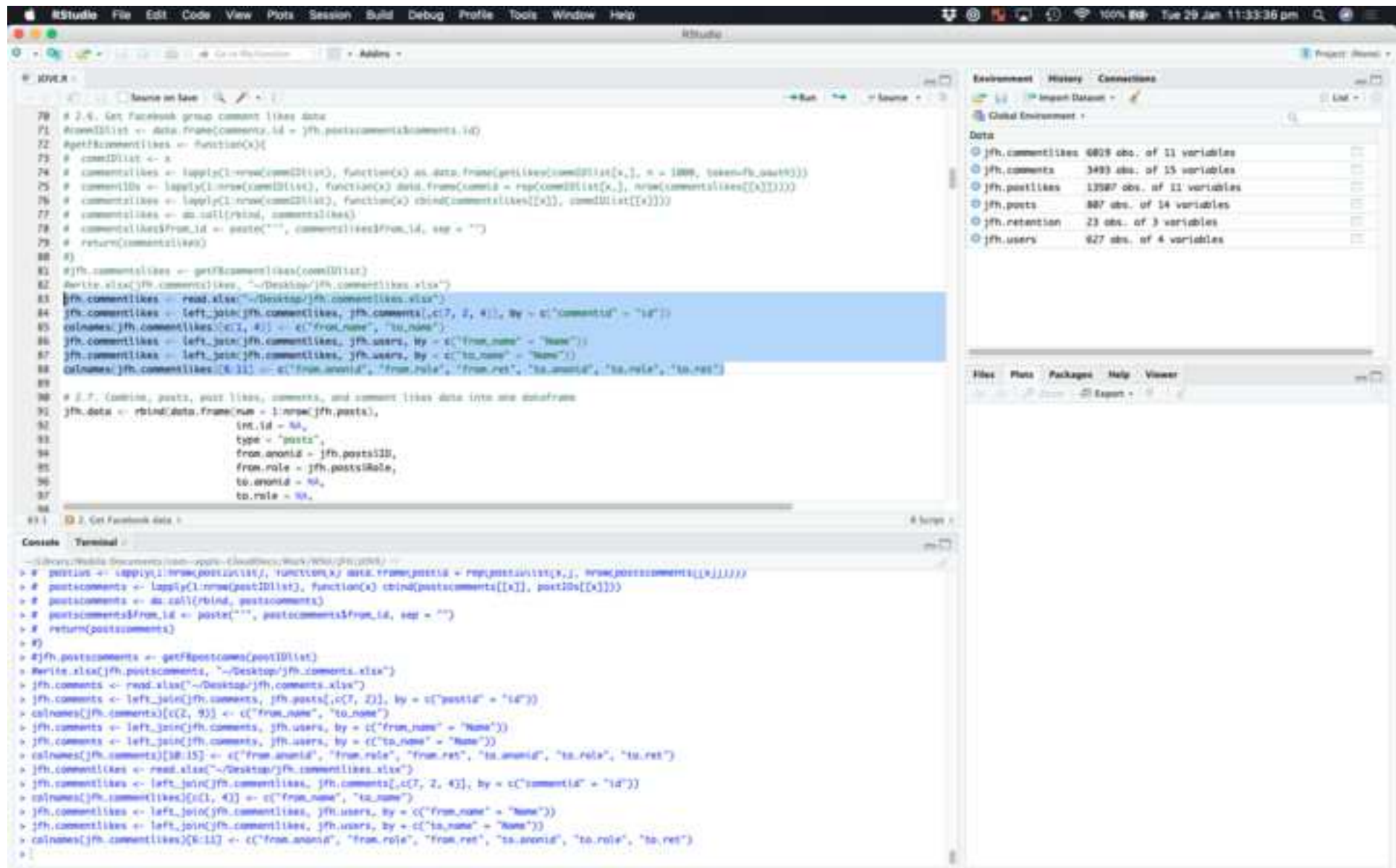
Console: Shows the execution of the code, including the loading of required packages (`library("Rfacebook")`, `library("dplyr")`, `library("igraph")`, `library("xlsx")`) and the execution of the data collection steps.

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for two functions: `getFacebookPostLikes` and `getFacebookPostComments`. The code uses `lapply` to iterate over a list of post IDs, fetching data from Facebook using the `facebook` package. The `getFacebookPostLikes` function returns a list of post likes, and the `getFacebookPostComments` function returns a list of post comments. The code also includes a section for loading retention and user data from CSV files.
- Environment Pane:** Located on the right, it shows the objects created in the global environment:
 - `jfh.postlikes`: 13987 obs. of 11 variables
 - `jfh.posts`: 887 obs. of 14 variables
 - `jfh.retention`: 23 obs. of 3 variables
 - `jfh.users`: 627 obs. of 4 variables
- Console:** Shows the execution of the code, including the loading of the `facebook` package and the execution of the `getFacebookPostLikes` function.

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for data manipulation. The code is organized into sections for getting Facebook group comments data (lines 51-68) and Facebook group comment likes data (lines 71-78). The code uses functions like `lapply`, `data.frame`, `read.xlsx`, `write.xlsx`, `left_join`, and `colnames` to process and merge data.
- Environment Pane:** Located on the right, it shows the following data objects:
 - `jfh.comments`: 3493 obs. of 13 variables
 - `jfh.postlikes`: 13587 obs. of 11 variables
 - `jfh.posts`: 887 obs. of 14 variables
 - `jfh.retention`: 29 obs. of 3 variables
 - `jfh.users`: 627 obs. of 4 variables
- Files Pane:** Located at the bottom right, it shows the file structure of the project, including folders like `data` and `scripts`.
- Console:** Located at the bottom left, it shows the execution of the R code, including the output of the `write.xlsx` function.



The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for data manipulation. The code is organized into sections: 2.4 (Facebook group comment likes data), 2.5 (Facebook post comments data), and 2.6 (Combine posts, post likes, comments, and comment likes data into one dataframe). The code uses functions like `lapply`, `do.call`, `rbind`, `read.xlsx`, and `write.xlsx` to process and combine data from various sources.
- Environment Pane:** Located on the right, it shows the global environment with the following data objects:
 - `jfh.comments`: 6829 obs. of 11 variables
 - `jfh.comments`: 3493 obs. of 15 variables
 - `jfh.postlikes`: 13987 obs. of 11 variables
 - `jfh.posts`: 887 obs. of 14 variables
 - `jfh.retention`: 23 obs. of 3 variables
 - `jfh.users`: 627 obs. of 4 variables
- Files Pane:** Located at the bottom right, it shows the file explorer with the following files:
 - `jfh.comments.xlsx`
 - `jfh.comments`
 - `jfh.postlikes.xlsx`
 - `jfh.posts`
 - `jfh.retention`
 - `jfh.users`

The screenshot displays the RStudio interface with a script editor, console, and environment pane.

Script Editor: The code combines two data frames, `jfh.posts` and `jfh.comments`, into a single data frame `jfh.data`. The code for `jfh.posts` includes variables like `int.id`, `type`, `from.anonid`, `from.role`, `to.anonid`, `to.role`, `from.ret`, `to.ret`, `date`, `week`, `month`, `text`, `comments`, `likes`, and `stringsAsFactors`. The code for `jfh.comments` includes similar variables but with different column names like `commentlikes`.

```
## 1.7. Combine posts, post likes, comments, and comment likes data into one dataframe
jfh.data = rbind.data.frame(num = 1:nrow(jfh.posts),
                             int.id = NA,
                             type = "posts",
                             from.anonid = jfh.posts$ID,
                             from.role = jfh.posts$role,
                             to.anonid = NA,
                             to.role = NA,
                             from.ret = jfh.posts$membership.length,
                             to.ret = NA,
                             date = gsub("-", "", gsub("T.*", "", jfh.posts$created.time)),
                             week = "x",
                             month = "x",
                             text = jfh.posts$message,
                             comments = jfh.posts$comments.count,
                             likes = jfh.posts$likes.count,
                             stringsAsFactors = FALSE),
                             data.frame(num = 1:nrow(jfh.comments),
                                         int.id = NA,
                                         type = "comments",
                                         from.anonid = jfh.comments$from.anonid,
                                         from.role = jfh.comments$from.role,
                                         to.anonid = jfh.comments$to.anonid,
                                         to.role = jfh.comments$to.role,
                                         from.ret = jfh.comments$from.ret,
                                         to.ret = jfh.comments$to.ret,
                                         date = gsub("-", "", gsub("T.*", "", jfh.comments$created.time)),
                                         week = "x",
                                         month = "x",
                                         text = NA,
                                         comments = NA,
                                         likes = NA,
                                         stringsAsFactors = FALSE)
```

Environment Pane: Lists the objects in the global environment:

- `jfh.commentlikes`: 6829 obs. of 11 variables
- `jfh.comments`: 3493 obs. of 15 variables
- `jfh.data`: 23826 obs. of 25 variables
- `jfh.postlikes`: 13587 obs. of 11 variables
- `jfh.posts`: 887 obs. of 14 variables
- `jfh.retention`: 23 obs. of 3 variables
- `jfh.users`: 627 obs. of 4 variables

Console: Shows the execution of the code, including the creation of the `jfh.data` object.

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for data manipulation. The visible code includes:


```
# 2.8. Add monthly breakdown (1-6, Month 1 to 6)
jfh.data$month <- ifelse(jfh.data$date == 20150429, 1,
                        ifelse(jfh.data$date == 20150527, 2,
                              ifelse(jfh.data$date == 20150624, 3,
                                    ifelse(jfh.data$date == 20150722, 4,
                                            ifelse(jfh.data$date == 20150819, 5,
                                                  ifelse(jfh.data$date == 20150916, 6,
                                                        ifelse(jfh.data$date == 20150814, 7, 8)))))))

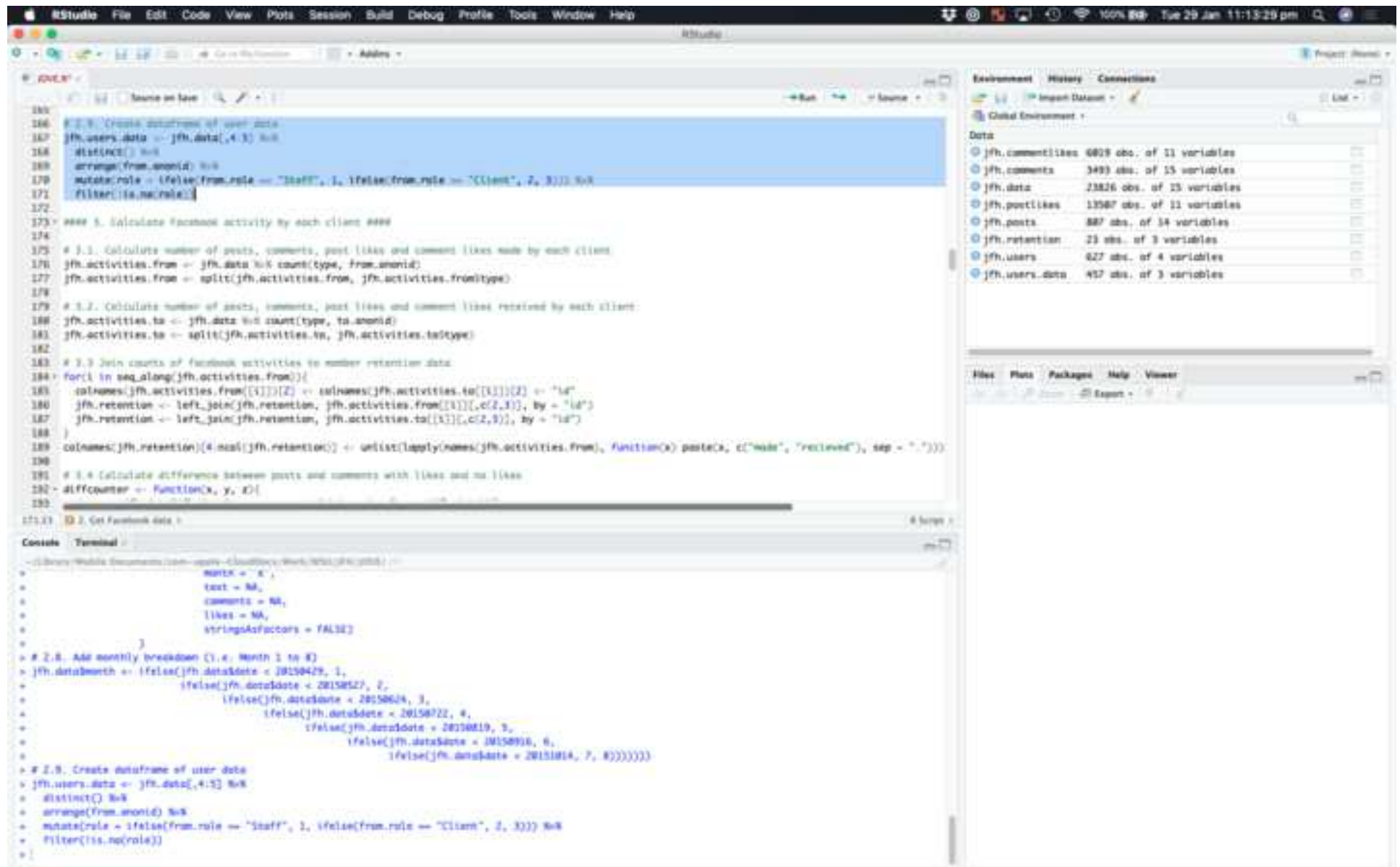
# 2.9. Create dataframe of user data
jfh.users.data <- jfh.data[,4:5] %>%
  distinct() %>%
  arrange(from.anonid) %>%
  mutate(role = ifelse(from.role == "Staff", 1, ifelse(from.role == "Client", 2, 3))) %>%
  filter(!is.na(role))

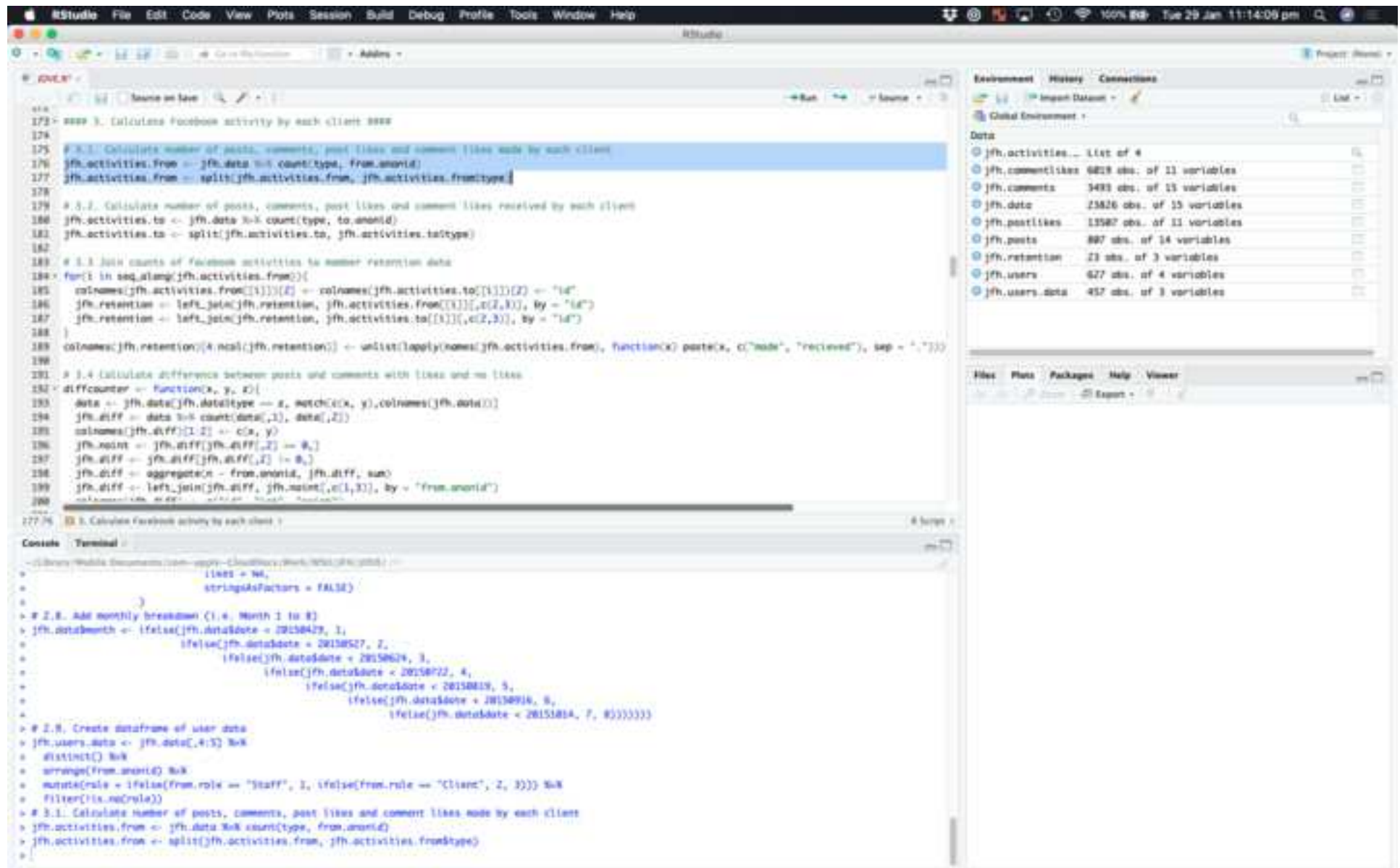
### 3. Calculate Facebook activity by each client ###

# 3.1. Calculate number of posts, comments, post likes and comment likes made by each client
jfh.activities.from <- jfh.data %>% count(type, from.anonid)
jfh.activities.from <- split(jfh.activities.from, jfh.activities.from$type)

# 3.2. Calculate number of posts, comments, post likes and comment likes received by each client
jfh.activities.to <- jfh.data %>% count(type, to.anonid)
jfh.activities.to <- split(jfh.activities.to, jfh.activities.to$type)

# 3.3 Join counts of facebook activities to member retention data
```
- Environment Panel:** Shows the loaded datasets:
 - jfh.commentlikes: 6819 obs. of 11 variables
 - jfh.comments: 349 obs. of 15 variables
 - jfh.data: 23826 obs. of 15 variables
 - jfh.postlikes: 13587 obs. of 11 variables
 - jfh.posts: 887 obs. of 14 variables
 - jfh.retention: 23 obs. of 3 variables
 - jfh.users: 627 obs. of 4 variables
- Console/Terminal:** Displays the output of the executed code, showing the creation of the 'jfh' package environment and the loading of the 'jfh' dataset.



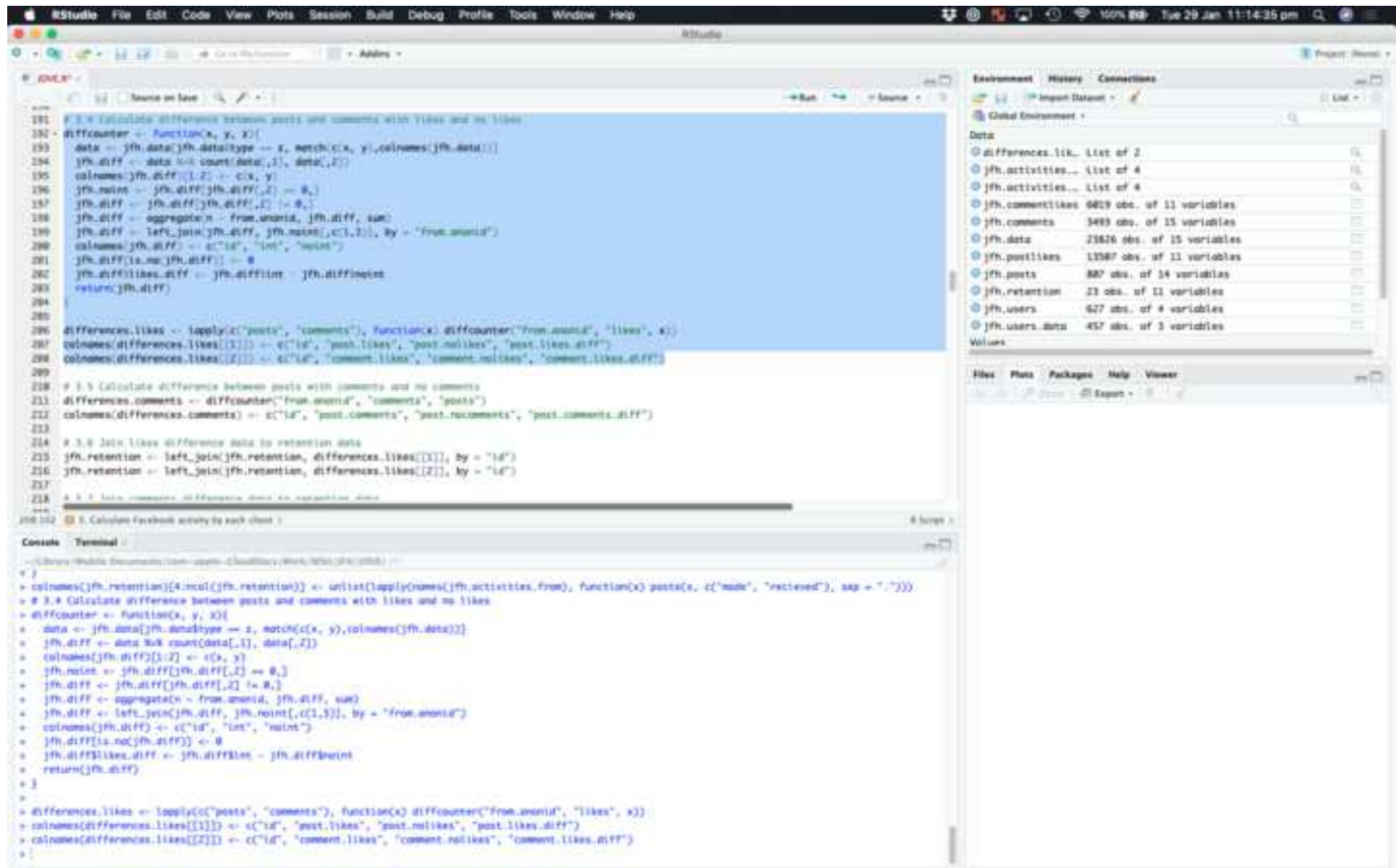


The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for data manipulation. The code includes:
 - Comments for steps 3.1, 3.2, and 3.3.
 - Functions `diffcounter` and `difference_likes` for calculating differences between posts and comments.
 - Data manipulation steps for `jfh.activities`, `jfh.retention`, and `jfh.users.data`.
- Environment Pane:** Lists the following objects:
 - `jfh.activities`: List of 4
 - `jfh.activities.to`: List of 4
 - `jfh.commentlikes`: 6819 obs. of 11 variables
 - `jfh.comments`: 3493 obs. of 13 variables
 - `jfh.data`: 23826 obs. of 15 variables
 - `jfh.postlikes`: 13587 obs. of 11 variables
 - `jfh.posts`: 887 obs. of 14 variables
 - `jfh.retention`: 23 obs. of 3 variables
 - `jfh.users`: 627 obs. of 4 variables
 - `jfh.users.data`: 457 obs. of 3 variables
- Console:** Shows the execution of the code, including the creation of `jfh.users.data` and the calculation of Facebook activity by each client.

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for data manipulation. The code includes:
 - Comments and variable assignments for `jfh.activities.to` and `jfh.activities.tp`.
 - A function `diffcounter` that calculates differences between posts and comments with likes and no likes.
 - Code to calculate Facebook activity by each client, including a `diffcounter` function and a `diffcounter` variable.
- Environment:** Shows the current environment with the following data objects:
 - `jfh.activities`: List of 4
 - `jfh.activities.to`: List of 4
 - `jfh.commentslikes`: 6819 obs. of 11 variables
 - `jfh.comments`: 3493 obs. of 13 variables
 - `jfh.data`: 23826 obs. of 15 variables
 - `jfh.postslikes`: 13587 obs. of 11 variables
 - `jfh.posts`: 887 obs. of 14 variables
 - `jfh.retention`: 23 obs. of 11 variables
 - `jfh.users`: 627 obs. of 4 variables
 - `jfh.users.data`: 457 obs. of 3 variables
- Files:** Shows the file explorer with the following files:
 - `1`
 - `4L`
- Console:** Shows the output of the R code, including the execution of the `diffcounter` function and the calculation of Facebook activity by each client.



The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for calculating differences between posts and comments with and without likes, and for joining these differences to retention data.
- Environment Pane:** Lists the objects in the global environment, including `differences.likes`, `jfh.activities`, `jfh.activities2`, `jfh.commentslikes`, `jfh.comments`, `jfh.data`, `jfh.postlikes`, `jfh.posts`, `jfh.retention`, `jfh.users`, and `jfh.users.data`.
- Console:** Shows the execution of the code, including the definition of `diffcounter` and the application of `lapply` to calculate differences.

```
# 3.4 Calculate difference between posts and comments with likes and no likes
diffcounter <- function(x, y, X){
  data <- jfh.data[jfh.data$type == x, match(c(x, y), colnames(jfh.data))]
  jfh.diff <- data %>% count(data[,1], data[,2])
  colnames(jfh.diff)[1:2] <- c(x, y)
  jfh.noint <- jfh.diff[jfh.diff[,2] == 0,]
  jfh.diff <- jfh.diff[jfh.diff[,2] != 0,]
  jfh.diff <- aggregate(n ~ from.anonid, jfh.diff, sum)
  jfh.diff <- left_join(jfh.diff, jfh.noint[,c(1,3)], by = "from.anonid")
  colnames(jfh.diff) <- c("id", "int", "noint")
  jfh.diff[is.na(jfh.diff)] <- 0
  jfh.diff$likes.diff <- jfh.diff$int - jfh.diff$noint
  return(jfh.diff)
}

differences.likes <- lapply(c("posts", "comments"), function(x) diffcounter("from.anonid", "likes", x))
colnames(differences.likes[[1]]) <- c("id", "post.likes", "post.no.likes", "post.likes.diff")
colnames(differences.likes[[2]]) <- c("id", "comment.likes", "comment.no.likes", "comment.likes.diff")

# 3.5 Calculate difference between posts with comments and no comments
differences.comments <- diffcounter("from.anonid", "comments", "posts")
colnames(differences.comments) <- c("id", "post.comments", "post.no.comments", "post.comments.diff")

# 3.6 Join likes difference data to retention data
jfh.retention <- left_join(jfh.retention, differences.likes[[1]], by = "id")
jfh.retention <- left_join(jfh.retention, differences.likes[[2]], by = "id")

# 3.7 Join comments difference data to aggregation data

# 3.8 Calculate Facebook activity by each client
```

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for data manipulation. Key sections include:
 - Calculating difference between posts with comments and no comments (lines 210-213).
 - Joining likes difference data to retention data (lines 214-217).
 - Joining comments difference data to retention data (lines 218-219).
 - Calculating all likes made by each client (lines 220-221).
 - Calculating all likes received by each client (lines 222-223).
 - Identifying which users did not participate in Facebook group (lines 224-225).
 - Section 4: Social Network Analysis (lines 226-227).
 - Creating edge list (lines 228-237).
 - Section 5: Calculate Facebook activity by each client (lines 238-239).
- Environment Pane:** Lists data objects:
 - differences.com: 78 obs. of 4 variables
 - differences.lik: list of 2
 - jfh.activities...: list of 4
 - jfh.activities...: list of 4
 - jfh.commentlikes: 6829 obs. of 11 variables
 - jfh.comments: 3493 obs. of 15 variables
 - jfh.data: 23826 obs. of 15 variables
 - jfh.postlikes: 13587 obs. of 11 variables
 - jfh.posts: 887 obs. of 14 variables
 - jfh.retention: 23 obs. of 11 variables
 - jfh.users: 627 obs. of 4 variables
 - jfh.users.data: 627 obs. of 4 variables
- Console:** Shows the execution of the code, including the definition of the `diffcounter` function and its application to the data.

The screenshot shows the RStudio interface with a script editor on the left and the Environment pane on the right. The script editor contains R code for data manipulation, including joining data, calculating likes, and creating edge and vertex lists. The Environment pane lists various data objects and their dimensions.

```

214 # 3.6 Join likes difference data to retention data
215 jfh.retention <- left_join(jfh.retention, differences.likes[[1]], by = "id")
216 jfh.retention <- left_join(jfh.retention, differences.likes[[2]], by = "id")
217
218 # 3.7 Join comments difference data to retention data
219 jfh.retention <- left_join(jfh.retention, differences.comments, by = "id")
220
221 # 3.8 Calculate all likes made by each client
222 jfh.retention$all.likes.made <- jfh.retention$postlikes.made + jfh.retention$commentlikes.made
223
224 # 3.9 Calculate all likes received by each client
225 jfh.retention$all.likes.recieved <- jfh.retention$postlikes.recieved + jfh.retention$commentlikes.recieved
226
227 # 3.10 Identify which users did not participate in Facebook group (no activity)
228 jfh.retention$no.activity <- as.numeric(apply(1:nrow(jfh.retention), function(x) {false(sum(jfh.retention[x, which(grep("made", colnames(jfh.rete
229
230 #### 4. Social Network Analysis ####
231
232 # 4.1. Create edge list
233 edges.all <- jfh.data[jfh.data$type != "posts", c(4,6)] %>%
234   mutate(relname = paste(from.anonid, " to ", to.anonid, sep = "")) %>%
235   filter(!is.na(from.anonid)) %>%
236   filter(!is.na(to.anonid))
237 edges <- edges.all %>% distinct()
238 edges <- left_join(edges, edges.all %>% count(relname), by = "relname")
239 colnames(edges) <- c("from", "to", "relname", "weight")
240
241 # 4.2. Create vertex list
242
243 244 77 5. Calculate Facebook activity by each client.

```

Environment

- Global Environment +
- Data
 - differences.com: 78 obs. of 4 variables
 - differences.lik: list of 2
 - jfh.activities...: list of 4
 - jfh.activities...: list of 4
 - jfh.commentlikes: 6829 obs. of 11 variables
 - jfh.comments: 3493 obs. of 15 variables
 - jfh.data: 23826 obs. of 15 variables
 - jfh.postlikes: 13587 obs. of 11 variables
 - jfh.posts: 887 obs. of 14 variables
 - jfh.retention: 23 obs. of 17 variables
 - jfh.users: 627 obs. of 4 variables
 - jfh.users.data: 627 obs. of 3 variables

Console

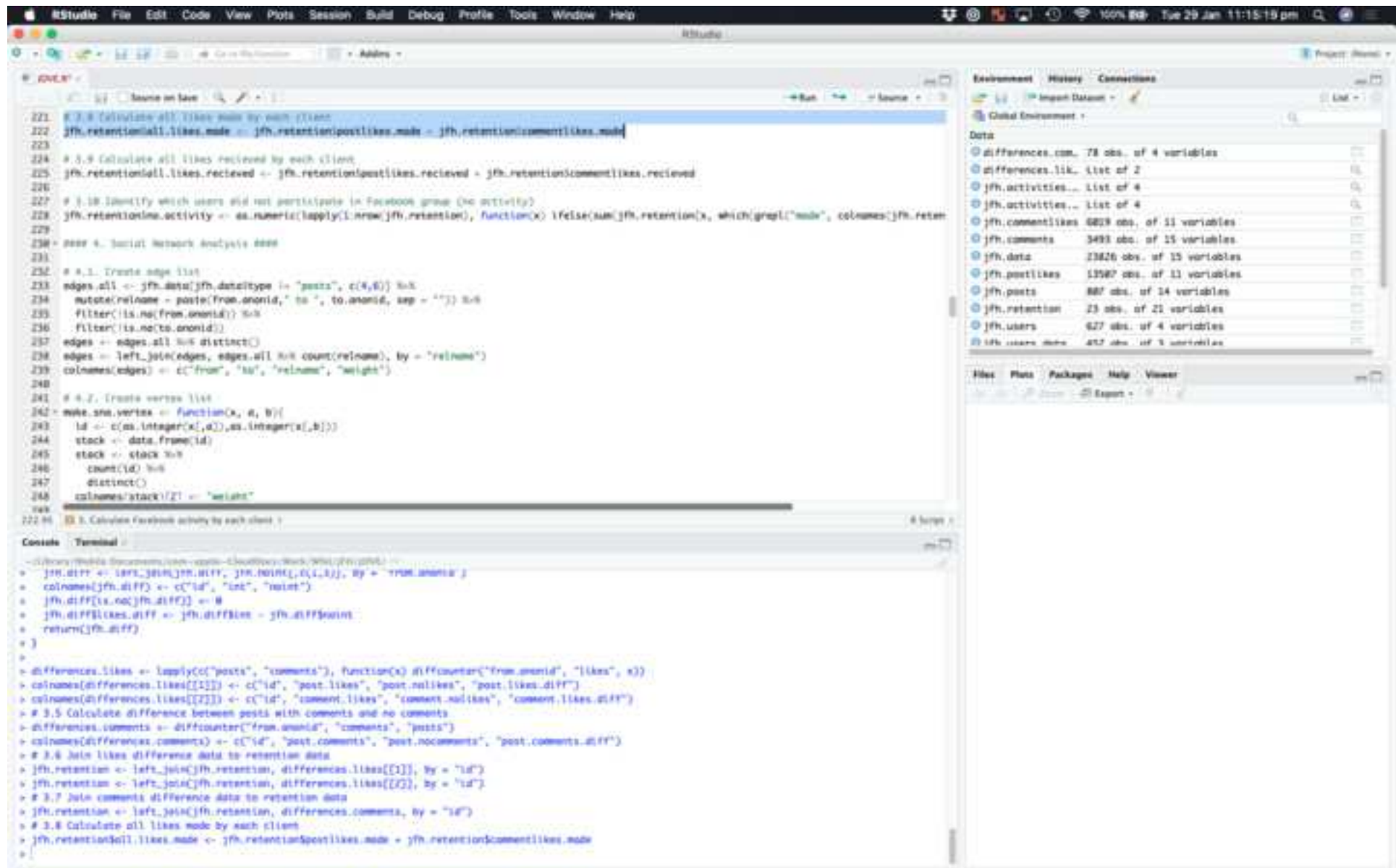
```

> colnames(jfh.diff)[1:2] <- c("x", "y")
> jfh.noint <- jfh.diff[jfh.diff[,2] == 0,]
> jfh.diff <- jfh.diff[jfh.diff[,2] != 0,]
> jfh.diff <- aggregate(x = from.anonid, jfh.diff, sum)
> jfh.diff <- left_join(jfh.diff, jfh.noint[,c(1,3)], by = "from.anonid")
> colnames(jfh.diff) <- c("id", "like", "noint")
> jfh.diff[is.na(jfh.diff)] <- 0
> jfh.diff$likes.diff <- jfh.diff$like - jfh.diff$noint
> return(jfh.diff)
>
> differences.likes <- lapply(c("posts", "comments"), function(x) diffcounters("from.anonid", "likes", x))
> colnames(differences.likes[[1]]) <- c("id", "post.likes", "post.no.likes", "post.likes.diff")
> colnames(differences.likes[[2]]) <- c("id", "comment.likes", "comment.no.likes", "comment.likes.diff")
> # 3.5 Calculate difference between posts with comments and no comments
> differences.comments <- diffcounters("from.anonid", "comments", "posts")
> colnames(differences.comments) <- c("id", "post.comments", "post.no.comments", "post.comments.diff")
> # 3.6 Join likes difference data to retention data
> jfh.retention <- left_join(jfh.retention, differences.likes[[1]], by = "id")
> jfh.retention <- left_join(jfh.retention, differences.likes[[2]], by = "id")
>

```

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for calculating differences between posts and comments with and without likes. The code includes comments in Chinese and uses functions like `left_join`, `lapply`, and `diffcounter`. Line numbers 218 through 228 are visible.
- Environment Panel:** Located on the right, it lists the following data objects:
 - `differences.com`: 78 obs. of 4 variables
 - `differences.lik`: list of 2
 - `jfh.activities...`: list of 4
 - `jfh.activities...`: list of 4
 - `jfh.commentlikes`: 6829 obs. of 11 variables
 - `jfh.comments`: 3493 obs. of 15 variables
 - `jfh.data`: 23826 obs. of 15 variables
 - `jfh.postlikes`: 13587 obs. of 11 variables
 - `jfh.posts`: 887 obs. of 14 variables
 - `jfh.retention`: 23 obs. of 28 variables
 - `jfh.users`: 627 obs. of 4 variables
 - `jfh.users.data`: 627 obs. of 3 variables
- Console:** Shows the execution of the code, including the definition of the `diffcounter` function and the calculation of differences for posts and comments.



```
# 3.8 Calculate all likes made by each client
jfh.retention$all.likes.made <- jfh.retention$postlikes.made + jfh.retention$commentlikes.made

# 3.9 Calculate all likes received by each client
jfh.retention$all.likes.received <- jfh.retention$postlikes.received + jfh.retention$commentlikes.received

# 3.10 Identify which users did not participate in Facebook group (no activity)
jfh.retention$online.activity <- as.numeric(apply(1:nrow(jfh.retention), function(x) {ifelse(sum(jfh.retention[x, which(grep("made", colnames(jfh.rete

### 4. Social Network Analysis ###

# 4.1. From edge list
edges.all <- jfh.data[jfh.data$type == "posts", c(4,6)] %>%
  mutate(relname = paste(from.anonid, " to ", to.anonid, sep = " ")) %>%
  filter(!is.na(from.anonid)) %>%
  filter(!is.na(to.anonid))
edges.all <- edges.all %>% distinct()
edges <- left_join(edges, edges.all %>% count(relname), by = "relname")
colnames(edges) <- c("from", "to", "relname", "weight")

# 4.2. From vertex list
make_sna_vertices <- function(x, a, b){
  id <- c(as.integer(x[,a]), as.integer(x[,b]))
  stack <- data.frame(id)
  stack <- stack %>%
    count(id) %>%
    distinct()
  colnames(stack)/21 <- "weights"
}

# 5. Calculate Facebook activity by each client.
```

Environment History Connections

Global Environment +

Data

- differences.com 78 obs. of 4 variables
- differences.lik. list of 2
- jfh.activities... list of 4
- jfh.activities... list of 4
- jfh.commentlikes 6829 obs. of 11 variables
- jfh.comments 3493 obs. of 15 variables
- jfh.data 23826 obs. of 15 variables
- jfh.postlikes 13587 obs. of 11 variables
- jfh.posts 887 obs. of 14 variables
- jfh.retention 23 obs. of 21 variables
- jfh.users 627 obs. of 4 variables
- jfh.users.data 627 obs. of 3 variables

Files Plots Packages Help Viewer

Console Terminal

```
# jfh.diff <- lapply(jfh.diff, jfh.names[,c(1,2)], by = "from.anonid")
# colnames(jfh.diff) <- c("id", "int", "weight")
# jfh.diff[is.na(jfh.diff)] <- 0
# jfh.diff$likes.diff <- jfh.diff$likes - jfh.diff$weight
# return(jfh.diff)

# differences.likes <- lapply(c("posts", "comments"), function(x) diffcounter("from.anonid", "likes", x))
# colnames(differences.likes[[1]]) <- c("id", "post.likes", "post.no.likes", "post.likes.diff")
# colnames(differences.likes[[2]]) <- c("id", "comment.likes", "comment.no.likes", "comment.likes.diff")
# # 3.5 Calculate difference between posts with comments and no comments
# differences.comments <- diffcounter("from.anonid", "comments", "posts")
# colnames(differences.comments) <- c("id", "post.comments", "post.no.comments", "post.comments.diff")
# # 3.6 Join likes difference data to retention data
# jfh.retention <- left_join(jfh.retention, differences.likes[[1]], by = "id")
# jfh.retention <- left_join(jfh.retention, differences.likes[[2]], by = "id")
# # 3.7 Join comments difference data to retention data
# jfh.retention <- left_join(jfh.retention, differences.comments, by = "id")
# # 3.8 Calculate all likes made by each client
# jfh.retention$all.likes.made <- jfh.retention$postlikes.made + jfh.retention$commentlikes.made
```

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for data manipulation. The visible code includes:


```

224 # 3.4 Calculate all likes received by each client
225 jfh.retention$all.likes.received <- jfh.retention$postlikes.received + jfh.retention$commentlikes.received
226
227 # 3.5 Identify which users did not participate in Facebook group (no activity)
228 jfh.retention$no.activity <- as.numeric(apply(1:nrow(jfh.retention), function(x) ifelse(sum(jfh.retention[x, which(grep("made", colnames(jfh.reten
229
230 ## 4. Social Network Analysis ##
231
232 # 4.1. Create edge list
233 edges.all <- jfh.data[jfh.data$type != "posts", c(4,6)] %>%
234   mutate(reaname = paste(from.anonid, " to ", to.anonid, sep = "")) %>%
235   filter(!is.na(from.anonid)) %>%
236   filter(!is.na(to.anonid))
237 edges.all <- edges.all %>% distinct()
238 edges <- left_join(edges, edges.all %>% count(reaname), by = "reaaname")
239 colnames(edges) <- c("from", "to", "reaaname", "weight")
240
241 # 4.2. Create vertex list
242 make.sns.vertices <- function(x, a, b){
243   id <- c(as.integer(x[,a]), as.integer(x[,b]))
244   stack <- data.frame(id)
245   stack <- stack %>%
246     count(id) %>%
247     distinct()
248   colnames(stack)[2] <- "weight"
249   return(stack)
250 }
251
252 ## 5. Calculate Facebook activity by each client.
253
254 # 5.1
255
256 # 5.2
257
258 # 5.3
259
260 # 5.4
261
262 # 5.5
263
264 # 5.6
265
266 # 5.7
267
268 # 5.8
269
270 # 5.9
271
272 # 5.10
273
274 # 5.11
275
276 # 5.12
277
278 # 5.13
279
280 # 5.14
281
282 # 5.15
283
284 # 5.16
285
286 # 5.17
287
288 # 5.18
289
290 # 5.19
291
292 # 5.20
293
294 # 5.21
295
296 # 5.22
297
298 # 5.23
299
300 # 5.24
301
302 # 5.25
303
304 # 5.26
305
306 # 5.27
307
308 # 5.28
309
310 # 5.29
311
312 # 5.30
313
314 # 5.31
315
316 # 5.32
317
318 # 5.33
319
320 # 5.34
321
322 # 5.35
323
324 # 5.36
325
326 # 5.37
327
328 # 5.38
329
330 # 5.39
331
332 # 5.40
333
334 # 5.41
335
336 # 5.42
337
338 # 5.43
339
340 # 5.44
341
342 # 5.45
343
344 # 5.46
345
346 # 5.47
347
348 # 5.48
349
350 # 5.49
351
352 # 5.50
353
354 # 5.51
355
356 # 5.52
357
358 # 5.53
359
360 # 5.54
361
362 # 5.55
363
364 # 5.56
365
366 # 5.57
367
368 # 5.58
369
370 # 5.59
371
372 # 5.60
373
374 # 5.61
375
376 # 5.62
377
378 # 5.63
379
380 # 5.64
381
382 # 5.65
383
384 # 5.66
385
386 # 5.67
387
388 # 5.68
389
390 # 5.69
391
392 # 5.70
393
394 # 5.71
395
396 # 5.72
397
398 # 5.73
399
400 # 5.74
401
402 # 5.75
403
404 # 5.76
405
406 # 5.77
407
408 # 5.78
409
410 # 5.79
411
412 # 5.80
413
414 # 5.81
415
416 # 5.82
417
418 # 5.83
419
420 # 5.84
421
422 # 5.85
423
424 # 5.86
425
426 # 5.87
427
428 # 5.88
429
430 # 5.89
431
432 # 5.90
433
434 # 5.91
435
436 # 5.92
437
438 # 5.93
439
440 # 5.94
441
442 # 5.95
443
444 # 5.96
445
446 # 5.97
447
448 # 5.98
449
450 # 5.99
451
452 # 5.100
453
454 # 5.101
455
456 # 5.102
457
458 # 5.103
459
460 # 5.104
461
462 # 5.105
463
464 # 5.106
465
466 # 5.107
467
468 # 5.108
469
470 # 5.109
471
472 # 5.110
473
474 # 5.111
475
476 # 5.112
477
478 # 5.113
479
480 # 5.114
481
482 # 5.115
483
484 # 5.116
485
486 # 5.117
487
488 # 5.118
489
490 # 5.119
491
492 # 5.120
493
494 # 5.121
495
496 # 5.122
497
498 # 5.123
499
500 # 5.124
501
502 # 5.125
503
504 # 5.126
505
506 # 5.127
507
508 # 5.128
509
510 # 5.129
511
512 # 5.130
513
514 # 5.131
515
516 # 5.132
517
518 # 5.133
519
520 # 5.134
521
522 # 5.135
523
524 # 5.136
525
526 # 5.137
527
528 # 5.138
529
530 # 5.139
531
532 # 5.140
533
534 # 5.141
535
536 # 5.142
537
538 # 5.143
539
540 # 5.144
541
542 # 5.145
543
544 # 5.146
545
546 # 5.147
547
548 # 5.148
549
550 # 5.149
551
552 # 5.150
553
554 # 5.151
555
556 # 5.152
557
558 # 5.153
559
560 # 5.154
561
562 # 5.155
563
564 # 5.156
565
566 # 5.157
567
568 # 5.158
569
570 # 5.159
571
572 # 5.160
573
574 # 5.161
575
576 # 5.162
577
578 # 5.163
579
580 # 5.164
581
582 # 5.165
583
584 # 5.166
585
586 # 5.167
587
588 # 5.168
589
590 # 5.169
591
592 # 5.170
593
594 # 5.171
595
596 # 5.172
597
598 # 5.173
599
600 # 5.174
601
602 # 5.175
603
604 # 5.176
605
606 # 5.177
607
608 # 5.178
609
610 # 5.179
611
612 # 5.180
613
614 # 5.181
615
616 # 5.182
617
618 # 5.183
619
620 # 5.184
621
622 # 5.185
623
624 # 5.186
625
626 # 5.187
627
628 # 5.188
629
630 # 5.189
631
632 # 5.190
633
634 # 5.191
635
636 # 5.192
637
638 # 5.193
639
640 # 5.194
641
642 # 5.195
643
644 # 5.196
645
646 # 5.197
647
648 # 5.198
649
650 # 5.199
651
652 # 5.200
653
654 # 5.201
655
656 # 5.202
657
658 # 5.203
659
660 # 5.204
661
662 # 5.205
663
664 # 5.206
665
666 # 5.207
667
668 # 5.208
669
670 # 5.209
671
672 # 5.210
673
674 # 5.211
675
676 # 5.212
677
678 # 5.213
679
680 # 5.214
681
682 # 5.215
683
684 # 5.216
685
686 # 5.217
687
688 # 5.218
689
690 # 5.219
691
692 # 5.220
693
694 # 5.221
695
696 # 5.222
697
698 # 5.223
699
700 # 5.224
701
702 # 5.225
703
704 # 5.226
705
706 # 5.227
707
708 # 5.228
709
710 # 5.229
711
712 # 5.230
713
714 # 5.231
715
716 # 5.232
717
718 # 5.233
719
720 # 5.234
721
722 # 5.235
723
724 # 5.236
725
726 # 5.237
727
728 # 5.238
729
730 # 5.239
731
732 # 5.240
733
734 # 5.241
735
736 # 5.242
737
738 # 5.243
739
740 # 5.244
741
742 # 5.245
743
744 # 5.246
745
746 # 5.247
747
748 # 5.248
749
750 # 5.249
751
752 # 5.250
753
754 # 5.251
755
756 # 5.252
757
758 # 5.253
759
760 # 5.254
761
762 # 5.255
763
764 # 5.256
765
766 # 5.257
767
768 # 5.258
769
770 # 5.259
771
772 # 5.260
773
774 # 5.261
775
776 # 5.262
777
778 # 5.263
779
780 # 5.264
781
782 # 5.265
783
784 # 5.266
785
786 # 5.267
787
788 # 5.268
789
790 # 5.269
791
792 # 5.270
793
794 # 5.271
795
796 # 5.272
797
798 # 5.273
799
800 # 5.274
801
802 # 5.275
803
804 # 5.276
805
806 # 5.277
807
808 # 5.278
809
810 # 5.279
811
812 # 5.280
813
814 # 5.281
815
816 # 5.282
817
818 # 5.283
819
820 # 5.284
821
822 # 5.285
823
824 # 5.286
825
826 # 5.287
827
828 # 5.288
829
830 # 5.289
831
832 # 5.290
833
834 # 5.291
835
836 # 5.292
837
838 # 5.293
839
840 # 5.294
841
842 # 5.295
843
844 # 5.296
845
846 # 5.297
847
848 # 5.298
849
850 # 5.299
851
852 # 5.300
853
854 # 5.301
855
856 # 5.302
857
858 # 5.303
859
860 # 5.304
861
862 # 5.305
863
864 # 5.306
865
866 # 5.307
867
868 # 5.308
869
870 # 5.309
871
872 # 5.310
873
874 # 5.311
875
876 # 5.312
877
878 # 5.313
879
880 # 5.314
881
882 # 5.315
883
884 # 5.316
885
886 # 5.317
887
888 # 5.318
889
890 # 5.319
891
892 # 5.320
893
894 # 5.321
895
896 # 5.322
897
898 # 5.323
899
900 # 5.324
901
902 # 5.325
903
904 # 5.326
905
906 # 5.327
907
908 # 5.328
909
910 # 5.329
911
912 # 5.330
913
914 # 5.331
915
916 # 5.332
917
918 # 5.333
919
920 # 5.334
921
922 # 5.335
923
924 # 5.336
925
926 # 5.337
927
928 # 5.338
929
930 # 5.339
931
932 # 5.340
933
934 # 5.341
935
936 # 5.342
937
938 # 5.343
939
940 # 5.344
941
942 # 5.345
943
944 # 5.346
945
946 # 5.347
947
948 # 5.348
949
950 # 5.349
951
952 # 5.350
953
954 # 5.351
955
956 # 5.352
957
958 # 5.353
959
960 # 5.354
961
962 # 5.355
963
964 # 5.356
965
966 # 5.357
967
968 # 5.358
969
970 # 5.359
971
972 # 5.360
973
974 # 5.361
975
976 # 5.362
977
978 # 5.363
979
980 # 5.364
981
982 # 5.365
983
984 # 5.366
985
986 # 5.367
987
988 # 5.368
989
990 # 5.369
991
992 # 5.370
993
994 # 5.371
995
996 # 5.372
997
998 # 5.373
999
1000 # 5.374
1001
1002 # 5.375
1003
1004 # 5.376
1005
1006 # 5.377
1007
1008 # 5.378
1009
1010 # 5.379
1011
1012 # 5.380
1013
1014 # 5.381
1015
1016 # 5.382
1017
1018 # 5.383
1019
1020 # 5.384
1021
1022 # 5.385
1023
1024 # 5.386
1025
1026 # 5.387
1027
1028 # 5.388
1029
1030 # 5.389
1031
1032 # 5.390
1033
1034 # 5.391
1035
1036 # 5.392
1037
1038 # 5.393
1039
1040 # 5.394
1041
1042 # 5.395
1043
1044 # 5.396
1045
1046 # 5.397
1047
1048 # 5.398
1049
1050 # 5.399
1051
1052 # 5.400
1053
1054 # 5.401
1055
1056 # 5.402
1057
1058 # 5.403
1059
1060 # 5.404
1061
1062 # 5.405
1063
1064 # 5.406
1065
1066 # 5.407
1067
1068 # 5.408
1069
1070 # 5.409
1071
1072 # 5.410
1073
1074 # 5.411
1075
1076 # 5.412
1077
1078 # 5.413
1079
1080 # 5.414
1081
1082 # 5.415
1083
1084 # 5.416
1085
1086 # 5.417
1087
1088 # 5.418
1089
1090 # 5.419
1091
1092 # 5.420
1093
1094 # 5.421
1095
1096 # 5.422
1097
1098 # 5.423
1099
1100 # 5.424
1101
1102 # 5.425
1103
1104 # 5.426
1105
1106 # 5.427
1107
1108 # 5.428
1109
1110 # 5.429
1111
1112 # 5.430
1113
1114 # 5.431
1115
1116 # 5.432
1117
1118 # 5.433
1119
1120 # 5.434
1121
1122 # 5.435
1123
1124 # 5.436
1125
1126 # 5.437
1127
1128 # 5.438
1129
1130 # 5.439
1131
1132 # 5.440
1133
1134 # 5.441
1135
1136 # 5.442
1137
1138 # 5.443
1139
1140 # 5.444
1141
1142 # 5.445
1143
1144 # 5.446
1145
1146 # 5.447
1147
1148 # 5.448
1149
1150 # 5.449
1151
1152 # 5.450
1153
1154 # 5.451
1155
1156 # 5.452
1157
1158 # 5.453
1159
1160 # 5.454
1161
1162 # 5.455
1163
1164 # 5.456
1165
1166 # 5.457
1167
1168 # 5.458
1169
1170 # 5.459
1171
1172 # 5.460
1173
1174 # 5.461
1175
1176 # 5.462
1177
1178 # 5.463
1179
1180 # 5.464
1181
1182 # 5.465
1183
1184 # 5.466
1185
1186 # 5.467
1187
1188 # 5.468
1189
1190 # 5.469
1191
1192 # 5.470
1193
1194 # 5.471
1195
1196 # 5.472
1197
1198 # 5.473
1199
1200 # 5.474
1201
1202 # 5.475
1203
1204 # 5.476
1205
1206 # 5.477
1207
1208 # 5.478
1209
1210 # 5.479
1211
1212 # 5.480
1213
1214 # 5.481
1215
1216 # 5.482
1217
1218 # 5.483
1219
1220 # 5.484
1221
1222 # 5.485
1223
1224 # 5.486
1225
1226 # 5.487
1227
1228 # 5.488
1229
1230 # 5.489
1231
1232 # 5.490
1233
1234 # 5.491
1235
1236 # 5.492
1237
1238 # 5.493
1239
1240 # 5.494
1241
1242 # 5.495
1243
1244 # 5.496
1245
1246 # 5.497
1247
1248 # 5.498
1249
1250 # 5.499
1251
1252 # 5.500
1253
1254 # 5.501
1255
1256 # 5.502
1257
1258 # 5.503
1259
1260 # 5.504
1261
1262 # 5.505
1263
1264 # 5.506
1265
1266 # 5.507
1267
1268 # 5.508
1269
1270 # 5.509
1271
1272 # 5.510
1273
1274 # 5.511
1275
1276 # 5.512
1277
1278 # 5.513
1279
1280 # 5.514
1281
1282 # 5.515
1283
1284 # 5.516
1285
1286 # 5.517
1287
1288 # 5.518
1289
1290 # 5.519
1291
1292 # 5.520
1293
1294 # 5.521
1295
1296 # 5.522
1297
1298 # 5.523
1299
1300 # 5.524
1301
1302 # 5.525
1303
1304 # 5.526
1305
1306 # 5.527
1307
1308 # 5.528
1309
1310 # 5.529
1311
1312 # 5.530
1313
1314 # 5.531
1315
1316 # 5.532
1317
1318 # 5.533
1319
1320 # 5.534
1321
1322 # 5.535
1323
1324 # 5.536
1325
1326 # 5.537
1327
1328 # 5.538
1329
1330 # 5.539
1331
1332 # 5.540
1333
1334 # 5.541
1335
1336 # 5.542
1337
1338 # 5.543
1339
1340 # 5.544
1341
1342 # 5.545
1343
1344 # 5.546
1345
1346 # 5.547
1347
1348 # 5.548
1349
1350 # 5.549
1351
1352 # 5.550
1353
1354 # 5.551
1355
1356 # 5.552
1357
1358 # 5.553
1359
1360 # 5.554
1361
1362 # 5.555
1363
1364 # 5.556
1365
1366 # 5.557
1367
1368 # 5.558
1369
1370 # 5.559
1371
1372 # 5.560
1373
1374 # 5.561
1375
1376 # 5.562
1377
1378 # 5.563
1379
1380 # 5.564
1381
1382 # 5.565
1383
1384 # 5.566
1385
1386 # 5.567
1387
1388 # 5.568
1389
1390 # 5.569
1391
1392 # 5.570
1393
1394 # 5.571
1395
1396 # 5.572
1397
1398 # 5.573
1399
1400 # 5.574
1401
1402 # 5.575
1403
1404 # 5.576
1405
1406 # 5.577
1407
1408 # 5.578
1409
1410 # 5.579
1411
1412 # 5.580
1413
1414 # 5.581
1415
1416 # 5.582
1417
1418 # 5.583
1419
1420 # 5.584
1421
1422 # 5.585
1423
1424 # 5.586
1425
1426 # 5.587
1427
1428 # 5.588
1429
1430 # 5.589
1431
1432 # 5.590
1433
1434 # 5.591
1435
1436 # 5.592
1437
1438 # 5.593
1439
1440 # 5.594
1441
1442 # 5.595
1443
1444 # 5.596
1445
1446 # 5.597
1447
1448 # 5.598
1449
1450 # 5.599
1451
1452 # 5.600
1453
1454 # 5.601
1455
1456 # 5.602
1457
1458 # 5.603
1459
1460 # 5.604
1461
1462 # 5.605
1463
1464 # 5.606
1465
1466 # 5.607
1467
1468 # 5.608
1469
1470 # 5.609
1471
1472 # 5.610
1473
1474 # 5.611
1475
1476 # 5.612
1477
1478 # 5.613
1479
1480 # 5.614
1481
1482 # 5.615
1483
1484 # 5.616
1485
1486 # 5.617
1487
1488 # 5.618
1489
1490 # 5.619
1491
1492 # 5.620
1493
1494 # 5.621
1495
1496 # 5.622
1497
1498 # 5.623
1499
1500 # 5.624
1501
1502 # 5.625
1503
1504 # 5.626
1505
1506 # 5.627
1507
1508 # 5.628
1509
1510 # 5.629
1511
1512 # 5.630
1513
1514 # 5.631
1515
1516 # 5.632
1517
1518 # 5.633
1519
1520 # 5.634
1521
1522 # 5.635
1523
1524 # 5.636
1525
1526 # 5.637
1527
1528 # 5.638
1529
1530 # 5.639
1531
1532 # 5.640
1533
1534 # 5.641
1535
1536 # 5.642
1537
1538 # 5.643
1539
1540 # 5.644
1541
1542 # 5.645
1543
1544 # 5.646
1545
1546 # 5.647
1547
1548 # 5.648
1549
1550 # 5.649
1551
1552 # 5.650
1553
1554 # 5.651
1555
1556 # 5.652
1557
1558 # 5.653
1559
1560 # 5.654
1561
1562 # 5.655
1563
1564 # 5.656
1565
1566 # 5.657
1567
1568 # 5.658
1569
1570 # 5.659
1571
1572 # 5.660
1573
1574 # 5.661
1575
1576 # 5.662
1577
1578 # 5.663
1579
1580 # 5.664
1581
1582 # 5.665
1583
1584 # 5.666
1585
1586 # 5.667
1587
1588 # 5.668
1589
1590 # 5.669
1591
1592 # 5.670
1593
1594 # 5.671
1595
1596 # 5.672
1597
1598 # 5.673
1599
1600 # 5.674
1601
1602 # 5.675
1603
1604 # 5.676
1605
1606 # 5.677
1607
1608 # 5.678
1609
1610 # 5.679
1611
1612 # 5.680
1613
1614 # 5.681
1615
1616 # 5.682
1617
1618 # 5.683
1619
1620 # 5.684
1621
1622 # 5.685
1623
1624 # 5.686
1625
1626 # 5.687
1627
1628 # 5.688
1629
1630 # 5.689
1631
1632 # 5.690
1633
1634 # 5.691
1635
1636 # 5.692
1637
1638 # 5.693
1639
1640 # 5.694
1641
1642 # 5.695
1643
1644 # 5.696
1645
1646 # 5.697
1647
1648 # 5.698
1649
1650 # 5.699
1651
1652 # 5.700
1653
1654 # 5.701
1655
1656 # 5.702
1657
1658 # 5.703
1659
1660 # 5.704
1661
1662 # 5.705
1663
1664 # 5.706
1665
1666 # 5.707
1667
1668 # 5.708
1669
1670 # 5.709
1671
1672 # 5.710
1673
1674 # 5.711
1675
1676 # 5.712
1677
1678 # 5.713
1679
1680 # 5.714
1681
1682 # 5.715
1683
1684 # 5.716
1685
1686 # 5.717
1687
1688 # 5.718
1689
1690 # 5.719
1691
1692 # 5.720
1693
1694 # 5.721
1695
1696 # 5.722
1697
1698 # 5.723
1699
1700 # 5.724
1701
1702 # 5.725
1703
1704 # 5.726
1705
1706 # 5.727
1707
1708 # 5.728
1709
1710 # 5.729
1711
1712 # 5.730
1713
1714 # 5.731
1715
1716 # 5.732
1717
1718 # 5.733
1719
1720 # 5.734
1721
1722 # 5.735
1723
1724 # 5.736
1725
1726 # 5.737
1727
1728 # 5.738
1729
1730 # 5.739
1731
1
```

The screenshot displays the RStudio interface with a script editor on the left, a console at the bottom, and an environment pane on the right.

Script Editor (Left):

```

227 # 3.10 Identify which users did not participate in Facebook group (no activity)
228 jfh.retention.no.activity <- as.numeric(apply(1:nrow(jfh.retention), function(x) ifelse(sum(jfh.retention[x, which(grep("made", colnames(jfh.retention))) > 0, 0, 1]))
229
230 ### 4. Social Network Analysis ###
231
232 # 4.1. Create edge list
233 edges.all <- jfh.data[jfh.data$type != "posts", c(4,6)] %>%
234 mutate(relname = paste(from.anonid, " to ", to.anonid, sep = "")) %>%
235 filter(!is.na(from.anonid)) %>%
236 filter(!is.na(to.anonid))
237 edges <- edges.all %>% distinct()
238 edges <- left_join(edges, edges.all %>% count(relname), by = "relname")
239 colnames(edges) <- c("from", "to", "relname", "weight")
240
241 # 4.2. Create vertex list
242 make.sna.vertices <- function(x, a, b){
243   id <- c(as.integer(x[,a]), as.integer(x[,b]))
244   stack <- data.frame(id)
245   stack <- stack %>%
246     count(id) %>%
247     distinct()
248   colnames(stack)[2] <- "weight"
249   return(stack)
250 }
251 vertices <- make.sna.vertices(edges, 1, 2)
252
253 # 4.3. Create graph and graph matrix objects
254 graph <- graph.data.frame(vertices, directed = TRUE)
255
256 ### 5. Calculate Facebook activity by each client ###
257
258 # 5.1
259 > differences.likes <- lapply(c("posts", "comments"), function(x) diffcounter("from.anonid", "likes", x))
260 > colnames(differences.likes[[1]]) <- c("id", "post.likes", "post.no.likes", "post.likes.diff")
261 > colnames(differences.likes[[2]]) <- c("id", "comment.likes", "comment.no.likes", "comment.likes.diff")
262 # 5.2 Calculate difference between posts with comments and no comments
263 > differences.comments <- diffcounter("from.anonid", "comments", "posts")
264 > colnames(differences.comments) <- c("id", "post.comments", "post.no.comments", "post.comments.diff")
265 # 5.3 Join likes difference data to retention data
266 > jfh.retention <- left_join(jfh.retention, differences.likes[[1]], by = "id")
267 > jfh.retention <- left_join(jfh.retention, differences.likes[[2]], by = "id")
268 # 5.4 Join comments difference data to retention data
269 > jfh.retention <- left_join(jfh.retention, differences.comments, by = "id")
270 # 5.5 Calculate all likes made by each client
271 > jfh.retention$all.likes.made <- jfh.retention$postlikes.made + jfh.retention$commentlikes.made
272 # 5.6 Calculate all likes received by each client
273 > jfh.retention$all.likes.received <- jfh.retention$postlikes.received + jfh.retention$commentlikes.received
274 # 5.7 Identify which users did not participate in Facebook group (no activity)
275 > jfh.retention.no.activity <- as.numeric(apply(1:nrow(jfh.retention), function(x) ifelse(sum(jfh.retention[x, which(grep("made", colnames(jfh.retention))) > 0, 0, 1]))
276

```

Environment Pane (Right):

Global Environment

Data

- differences.com: 78 obs. of 4 variables
- differences.lik: list of 2
- jfh.activities: list of 4
- jfh.activities: list of 4
- jfh.commentlikes: 6829 obs. of 11 variables
- jfh.comments: 3493 obs. of 15 variables
- jfh.data: 23826 obs. of 15 variables
- jfh.postlikes: 13587 obs. of 11 variables
- jfh.posts: 887 obs. of 14 variables
- jfh.retention: 23 obs. of 23 variables
- jfh.users: 627 obs. of 4 variables
- jfh.users.data: 627 obs. of 3 variables

Console (Bottom):

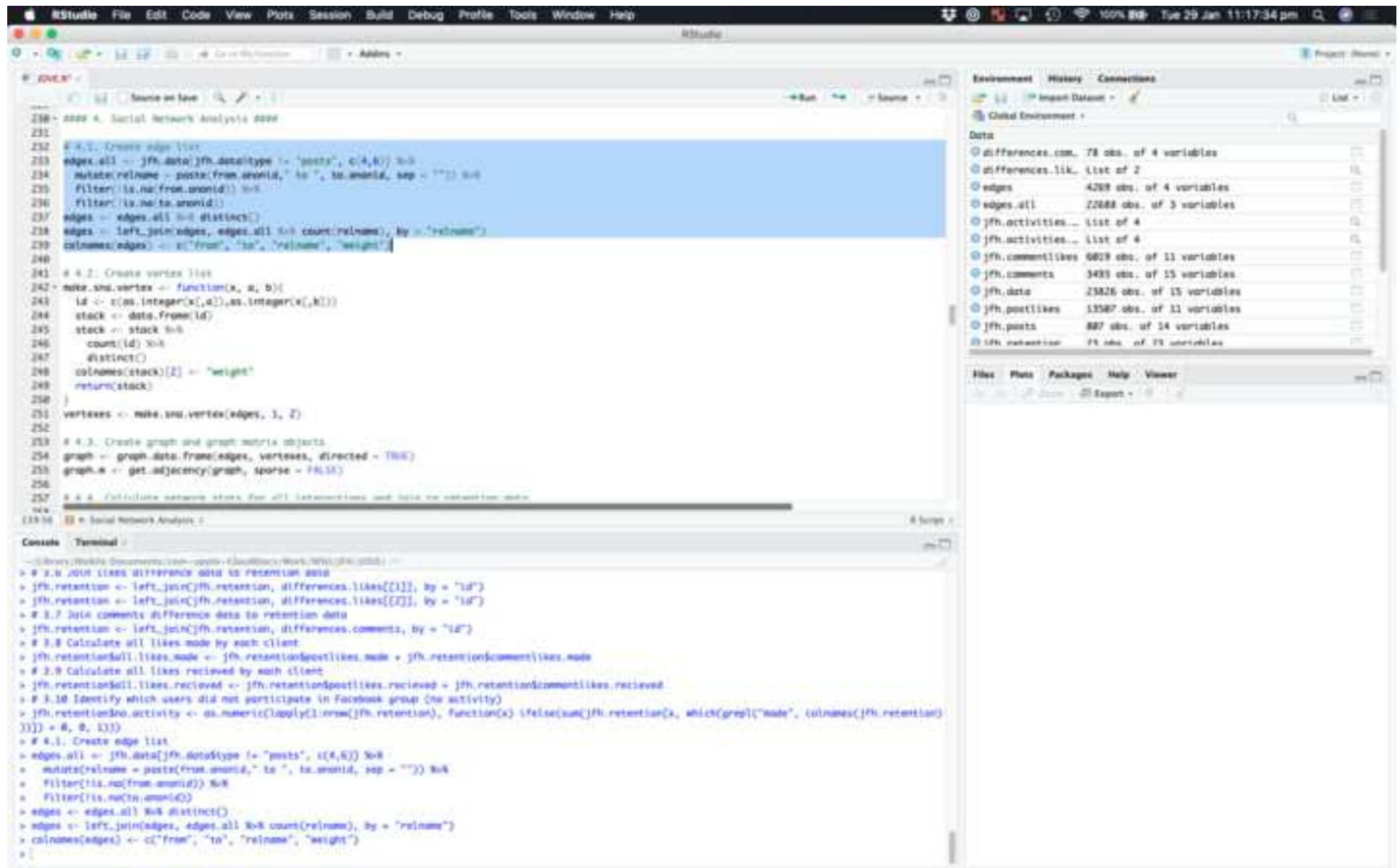
```

> jfh.retention.no.activity <- as.numeric(apply(1:nrow(jfh.retention), function(x) ifelse(sum(jfh.retention[x, which(grep("made", colnames(jfh.retention))) > 0, 0, 1]))
>
> differences.likes <- lapply(c("posts", "comments"), function(x) diffcounter("from.anonid", "likes", x))
> colnames(differences.likes[[1]]) <- c("id", "post.likes", "post.no.likes", "post.likes.diff")
> colnames(differences.likes[[2]]) <- c("id", "comment.likes", "comment.no.likes", "comment.likes.diff")
> # 5.2 Calculate difference between posts with comments and no comments
> differences.comments <- diffcounter("from.anonid", "comments", "posts")
> colnames(differences.comments) <- c("id", "post.comments", "post.no.comments", "post.comments.diff")
> # 5.3 Join likes difference data to retention data
> jfh.retention <- left_join(jfh.retention, differences.likes[[1]], by = "id")
> jfh.retention <- left_join(jfh.retention, differences.likes[[2]], by = "id")
> # 5.4 Join comments difference data to retention data
> jfh.retention <- left_join(jfh.retention, differences.comments, by = "id")
> # 5.5 Calculate all likes made by each client
> jfh.retention$all.likes.made <- jfh.retention$postlikes.made + jfh.retention$commentlikes.made
> # 5.6 Calculate all likes received by each client
> jfh.retention$all.likes.received <- jfh.retention$postlikes.received + jfh.retention$commentlikes.received
> # 5.7 Identify which users did not participate in Facebook group (no activity)
> jfh.retention.no.activity <- as.numeric(apply(1:nrow(jfh.retention), function(x) ifelse(sum(jfh.retention[x, which(grep("made", colnames(jfh.retention))) > 0, 0, 1]))
>

```


The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for data manipulation and graph creation. Key sections include:
 - Identifying users who did not participate in Facebook group activity (lines 227-228).
 - Creating an edge list from a data frame (lines 233-238).
 - Creating a vertex list (lines 241-250).
 - Creating graph and graph matrix objects (lines 253-254).
 - Calculating Facebook activity by each client (lines 258-260).
- Environment Pane:** Lists the following data objects:
 - differences.com: 78 obs. of 4 variables
 - differences.lik: list of 2
 - jfh.activities...: list of 4
 - jfh.activities...: list of 4
 - jfh.commentlikes: 6829 obs. of 11 variables
 - jfh.comments: 3493 obs. of 15 variables
 - jfh.data: 23826 obs. of 15 variables
 - jfh.postlikes: 13587 obs. of 11 variables
 - jfh.posts: 887 obs. of 14 variables
 - jfh.retention: 23 obs. of 23 variables
 - jfh.users: 627 obs. of 4 variables
 - jfh.users.data: 627 obs. of 3 variables
- Console:** Shows the execution of the code, including the creation of the 'differences.likes' object and the calculation of difference between posts with and without comments.



The screenshot shows the RStudio interface with a script editor on the left, a console at the bottom, and an environment pane on the right.

Script Editor:

```

241 # 4.2. Create vertex list
242 make_sna_vertex <- function(x, a, b){
243   id <- c(as.integer(a),a),as.integer(x[,b]))
244   stack <- data.frame(id)
245   stack <- stack %>%
246     count(id) %>%
247     distinct()
248   colnames(stack)[2] <- "weight"
249   return(stack)
250 }
251 vertices <- make_sna_vertex(edges, 1, 2)
252
253 # 4.3. Create graph and graph matrix objects
254 graph <- graph_data.frame(edges, vertices, directed = TRUE)
255 graph_m <- get_adjacency(graph, sparse = FALSE)
256
257 # 4.4. Calculate network stats for all interactions and join to retention data
258 vertexes$degree <- degree(graph)
259 vertexes$betweenness <- betweenness(graph)
260 vertexes$id <- as.integer(vertexes$id)
261 jfh$retention <- left_join(jfh$retention, vertexes, by = "id")
262
263 # 4.5. Calculate network stats only for post-comment interactions
264 edges_postcome <- jfh$data[jfh$data$type == "comments", c(4,6)] %>%
265   mutate(reaname = paste(from.anonid, " to ", to.anonid, sep = "")) %>%
266   filter(!is.na(from.anonid)) %>%
267   filter(!is.na(to.anonid)) %>%
268   distinct()
269
270 #####
271 # Social Network Analysis 2

```

Environment Pane:

Object	Size	Variables
edges	4269 obs.	4 variables
edges.all	22688 obs.	3 variables
jfh.activities	List of 4	
jfh.activities	List of 4	
jfh.commentlikes	6819 obs.	11 variables
jfh.comments	3493 obs.	15 variables
jfh.data	23826 obs.	15 variables
jfh.postlikes	13567 obs.	11 variables
jfh.posts	887 obs.	14 variables
jfh.retention	23 obs.	23 variables
jfh.users	927 obs.	4 variables
jfh.users.data	457 obs.	3 variables
vertexes	457 obs.	2 variables

Console:

```

####
# 4.1. Create edge list
> edges.all <- jfh.data[jfh.data$type != "posts", c(4,6)] %>%
+   mutate(reaname = paste(from.anonid, " to ", to.anonid, sep = "")) %>%
+   filter(!is.na(from.anonid)) %>%
+   filter(!is.na(to.anonid)) %>%
+   edges <- edges.all %>% distinct()
+ edges <- left_join(edges, edges.all %>% count(reaname), by = "reaname")
+ colnames(edges) <- c("from", "to", "reaname", "weight")
# 4.2. Create vertex list
> make_sna_vertex <- function(x, a, b){
+   id <- c(as.integer(x[,a]),as.integer(x[,b]))
+   stack <- data.frame(id)
+   stack <- stack %>%
+     count(id) %>%
+     distinct()
+   colnames(stack)[2] <- "weight"
+   return(stack)
+ }
> vertexes <- make_sna_vertex(edges, 1, 2)
>

```

The screenshot shows the RStudio interface with a script editor on the left, a console at the bottom, and an environment pane on the right.

Script Editor (Lines 252-288):

```

252 # 4.3. Create graph and graph matrix objects
253 graph <- graph.data.frame(edges, vertices, directed = TRUE)
254 graph.m <- get.adjacency(graph, sparse = FALSE)
255
256 # 4.4. Calculate network stats for all interactions and join to retention data
257 vertexes.degree <- degree(graph)
258 vertexes.betweenness <- betweenness(graph)
259 vertexes.id <- as.integer(vertexes$id)
260 jfh.retention <- left_join(jfh.retention, vertexes, by = "id")
261
262 # 4.5. Calculate network stats only for post-comment interactions
263 edges.postcomm <- jfh.data[jfh.data$type == "comments", c(4,6)] %>%
264   mutate(reiname = paste(from.anonid, " to ", to.anonid, sep = "")) %>%
265   filter(!is.na(from.anonid)) %>%
266   filter(!is.na(to.anonid)) %>%
267   distinct()
268
269 vertexes.postcomm <- make.sna.vertex(edges.postcomm, 1, 2)
270 graph.postcomm <- graph.data.frame(edges.postcomm, vertexes.postcomm, directed = TRUE)
271 graph.m.postcomm <- get.adjacency(graph.postcomm, sparse = FALSE)
272 vertexes.postcomm.degree.postcomm <- degree(graph.postcomm)
273 vertexes.postcomm.betweenness.postcomm <- betweenness(graph.postcomm)
274 vertexes.postcomm.id <- as.integer(vertexes.postcomm$id)
275 jfh.retention <- left_join(jfh.retention, vertexes.postcomm, by = "id")
276
277 ### 5. Conduct computerized linguistic analysis in LINC ###
278
279 # 5.1. Export textual data along with user ID and text data type (post/comment) into a csv file to run through LINC
280

```

Environment Pane:

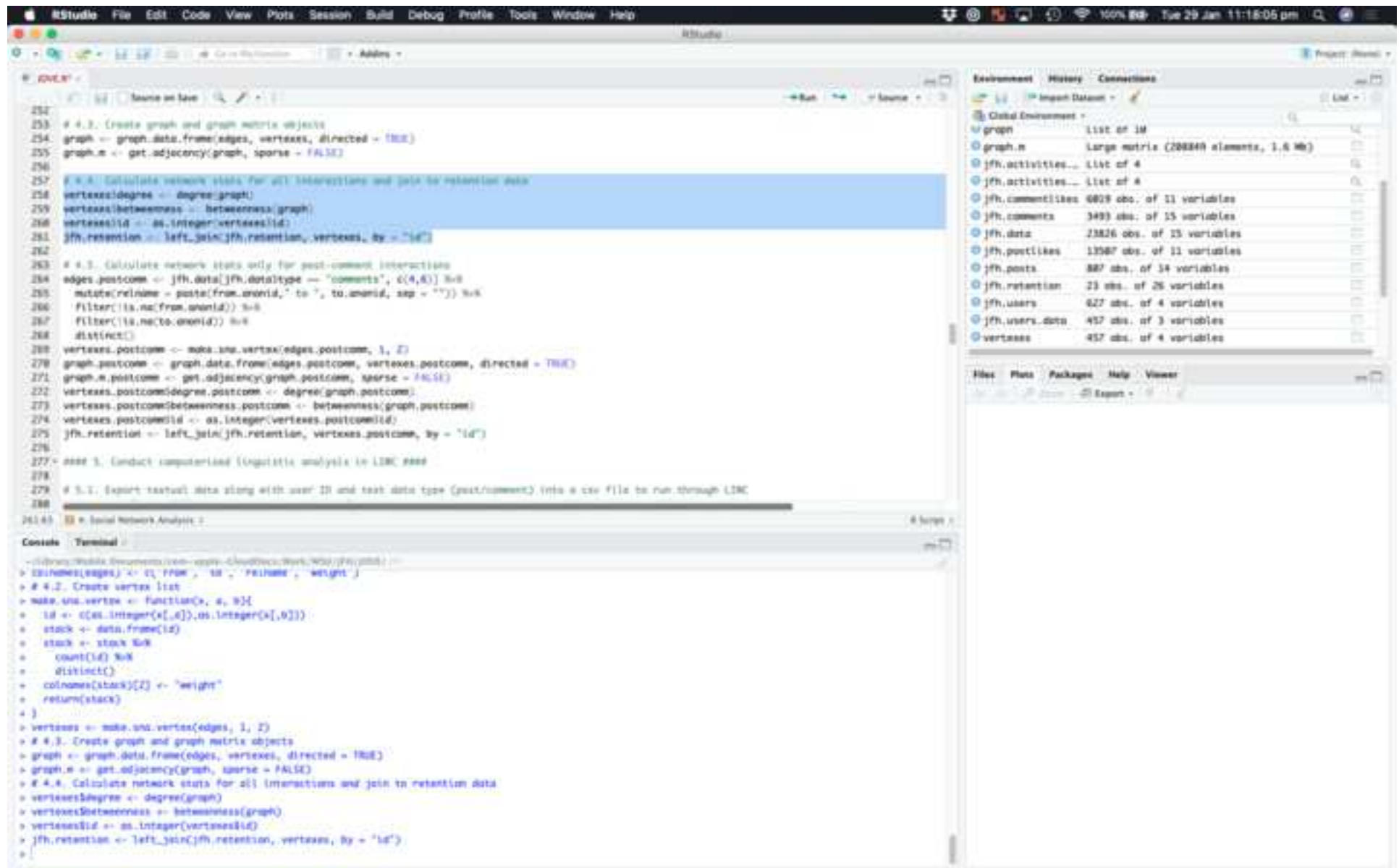
Object	Class	Size
edges	data.frame	4269 obs. of 4 variables
edges.all	data.frame	22688 obs. of 3 variables
graph	igraph	list of 18
graph.m	matrix	large matrix (288849 elements, 1.6 Mb)
jfh.activities...	list	list of 4
jfh.activities...	list	list of 4
jfh.commentlikes	data.frame	6819 obs. of 11 variables
jfh.comments	data.frame	3493 obs. of 13 variables
jfh.data	data.frame	23826 obs. of 15 variables
jfh.postlikes	data.frame	13547 obs. of 11 variables
jfh.posts	data.frame	887 obs. of 14 variables
jfh.retention	data.frame	23 obs. of 23 variables
jfh.users	data.frame	827 obs. of 4 variables

Console (Lines 255-48):

```

255 # Social Network Analysis 2
48
> library(devtools)
> source("https://raw.githubusercontent.com/jfh/jfh")
> mutate(reiname = paste(from.anonid, " to ", to.anonid, sep = "")) %>%
> filter(!is.na(from.anonid)) %>%
> filter(!is.na(to.anonid)) %>%
> edges <- edges.all %>% distinct()
> edges <- left_join(edges, edges.all %>% count(reiname, by = "reiname")
> colnames(edges) <- c("from", "to", "reiname", "weight")
> # 4.2. Create vertex list
> make.sna.vertex <- function(x, a, b){
>   id <- c(as.integer(x[,a]), as.integer(x[,b]))
>   stack <- data.frame(id)
>   stack <- stack %>%
>     count(id) %>%
>     distinct()
>   colnames(stack)[1] <- "weight"
>   return(stack)
> }
> vertexes <- make.sna.vertex(edges, 1, 2)
> # 4.3. Create graph and graph matrix objects
> graph <- graph.data.frame(edges, vertexes, directed = TRUE)
> graph.m <- get.adjacency(graph, sparse = FALSE)
>

```

The screenshot shows an RStudio interface with a script editor on the left, a console at the bottom, and an environment pane on the right.

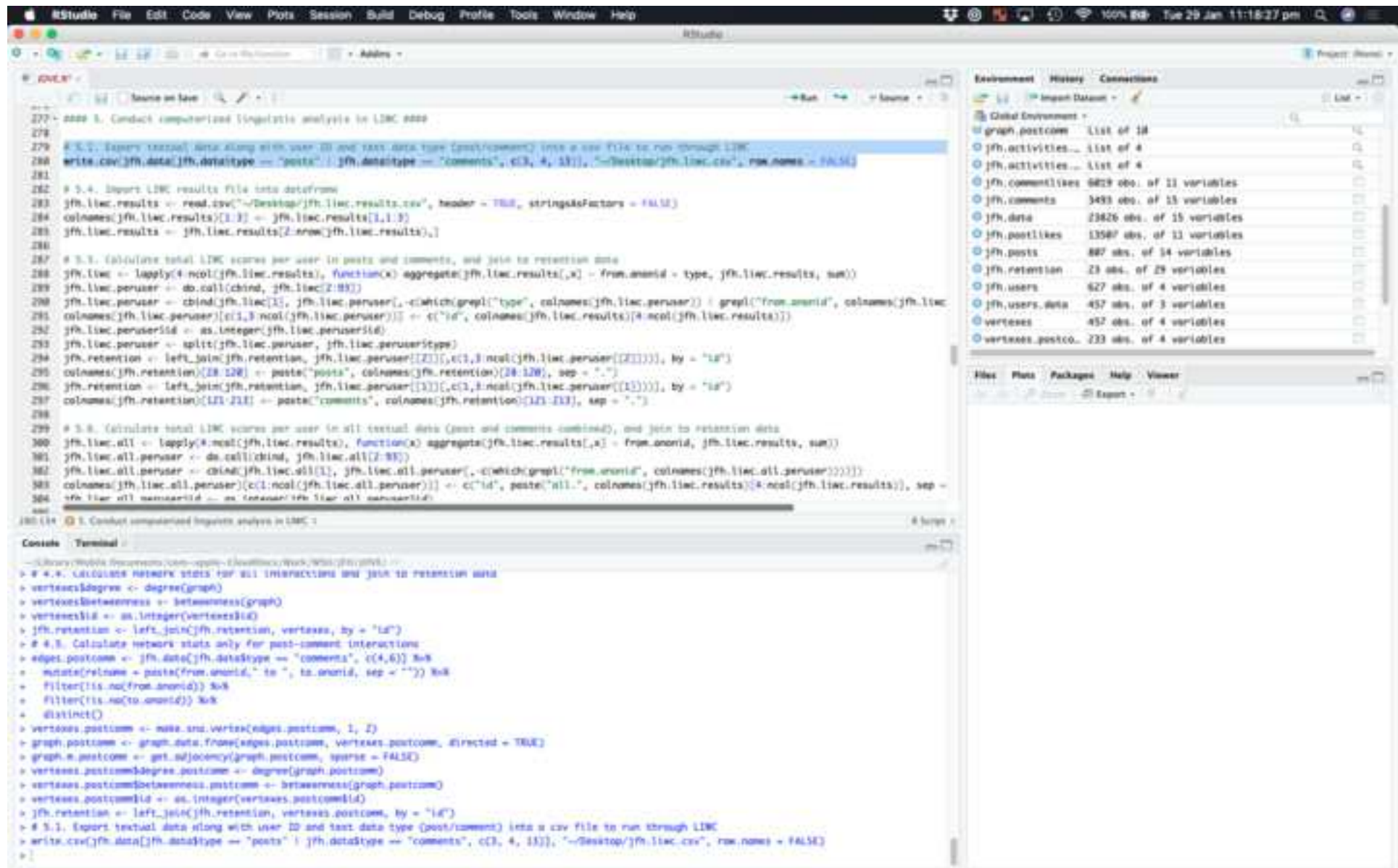
Script Editor: The code is a series of R commands for network analysis. It includes comments in Spanish and English. Key steps include:

- Filtering data for post-comments interactions (lines 263-268).
- Creating a graph object from edges and vertices (lines 269-273).
- Calculating network statistics like degree and betweenness (lines 274-276).
- Exporting data to a CSV file (lines 277-280).
- Importing the CSV file back into a dataframe (lines 281-284).
- Calculating total LINC scores per user (lines 285-288).
- Aggregating results by user ID (lines 289-290).

Environment Pane: Lists the objects in the global environment:

- graph.postcomm: List of 18
- jfh.activities: List of 4
- jfh.activities: List of 4
- jfh.comments: 5829 obs. of 11 variables
- jfh.data: 23826 obs. of 15 variables
- jfh.postlikes: 13587 obs. of 11 variables
- jfh.posts: 887 obs. of 14 variables
- jfh.retention: 23 obs. of 29 variables
- jfh.users: 627 obs. of 4 variables
- jfh.users.data: 457 obs. of 3 variables
- vertices.postco: 233 obs. of 4 variables

Console: Shows the execution of the code, with output for the graph object and the final LINC scores.



The screenshot displays the RStudio interface with a script editor on the left, a console at the bottom, and an environment pane on the right. The script editor contains R code for data manipulation and network analysis. The console shows the execution of the code, and the environment pane lists the objects created during the session.

```
## 5. Conduct computerized linguistic analysis in LINC ##
## 5.1. Export textual data along with user ID and text data type (post/comment) into a csv file to run through LINC
write.csv(jfh.data[jfh.data$type == "posts" | jfh.data$type == "comments", c(3, 4, 13)], "~/Desktop/jfh.linc.csv", row.names = FALSE)

## 5.2. Import LINC results file into dataframe
jfh.linc.results <- read.csv("~/Desktop/jfh.linc.results.csv", header = TRUE, stringsAsFactors = FALSE)
colnames(jfh.linc.results)[1:3] <- jfh.linc.results[1:3]
jfh.linc.results <- jfh.linc.results[2:nrow(jfh.linc.results),]

## 5.3. Calculate total LINC scores per user in posts and comments, and join to retention data
jfh.linc <- lapply(4:ncol(jfh.linc.results), function(x) aggregate(jfh.linc.results[,x] ~ from.anonid ~ type, jfh.linc.results, sum))
jfh.linc.peruser <- do.call(cbind, jfh.linc[2:93])
jfh.linc.peruser <- cbind(jfh.linc[,1], jfh.linc.peruser[,c(which(grepl("type", colnames(jfh.linc.peruser)) | grepl("from.anonid", colnames(jfh.linc
colnames(jfh.linc.peruser)[c(1,3:ncol(jfh.linc.peruser))] <- c("id", colnames(jfh.linc.results)[4:ncol(jfh.linc.results)])
jfh.linc.peruserid <- as.integer(jfh.linc.peruser[,1])
jfh.linc.peruser <- split(jfh.linc.peruser, jfh.linc.peruserid)
jfh.retention <- left_join(jfh.retention, jfh.linc.peruser[[1]][,c(1,3:ncol(jfh.linc.peruser[[1]]))], by = "id")
colnames(jfh.retention)[28:128] <- paste("posts", colnames(jfh.retention)[28:128], sep = ".")
jfh.retention <- left_join(jfh.retention, jfh.linc.peruser[[1]][,c(1,3:ncol(jfh.linc.peruser[[1]]))], by = "id")
colnames(jfh.retention)[121:211] <- paste("comments", colnames(jfh.retention)[121:211], sep = ".")

## 5.4. Calculate total LINC scores per user in all textual data (post and comments combined), and join to retention data
jfh.linc.all <- lapply(4:ncol(jfh.linc.results), function(x) aggregate(jfh.linc.results[,x] ~ from.anonid, jfh.linc.results, sum))
jfh.linc.all.peruser <- do.call(cbind, jfh.linc.all[2:93])
jfh.linc.all.peruser <- cbind(jfh.linc.all[,1], jfh.linc.all.peruser[,c(which(grepl("from.anonid", colnames(jfh.linc.all.peruser)))]
colnames(jfh.linc.all.peruser)[c(1:ncol(jfh.linc.all.peruser))] <- c("id", paste("all.", colnames(jfh.linc.results)[4:ncol(jfh.linc.results)], sep =
~> jfh.linc.all.peruserid <- as.integer(jfh.linc.all.peruser[,1])

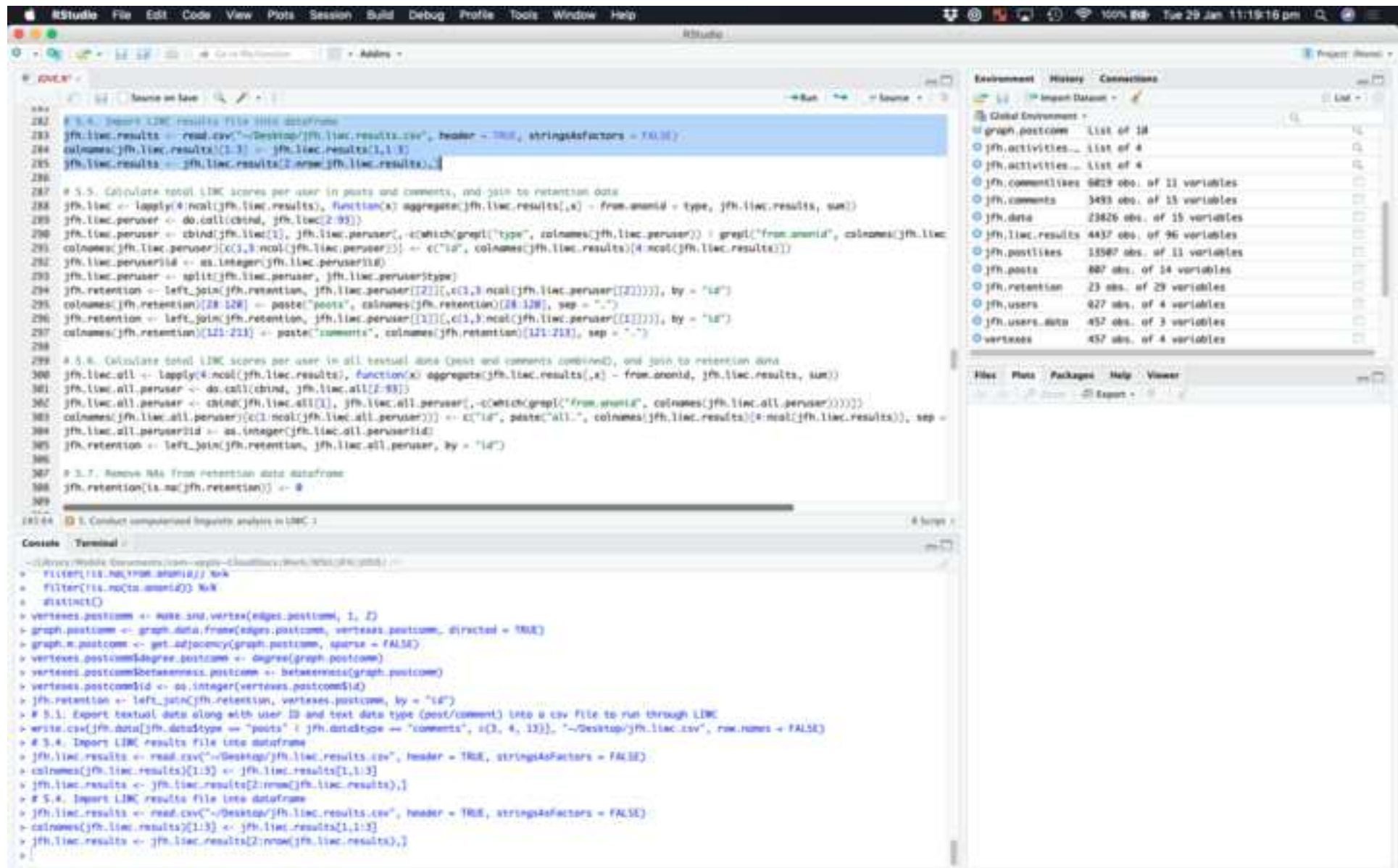
## 5.5. Conduct computerized linguistic analysis in LINC 1
```

The console shows the execution of the code, and the environment pane lists the objects created during the session.

```
## 5.4. Calculate network stats for all interactions and join to retention data
vertices$degree <- degree(graph)
vertices$betweenness <- betweenness(graph)
vertices$id <- as.integer(vertices$id)
jfh.retention <- left_join(jfh.retention, vertices, by = "id")
## 5.5. Calculate network stats only for post-comment interactions
edges.postcomm <- jfh.data[jfh.data$type == "comments", c(4,6)] %>%
  mutate(retname = paste(from.anonid, " to ", to.anonid, sep = "")) %>%
  filter(!is.na(from.anonid)) %>%
  filter(!is.na(to.anonid)) %>%
  distinct()
vertices.postcomm <- make_snn_vertices(edges.postcomm, 1, 2)
graph.postcomm <- graph_data_frame(edges.postcomm, vertices.postcomm, directed = TRUE)
graph.e.postcomm <- get_adjacency(graph.postcomm, sparse = FALSE)
vertices.postcomm$degree.postcomm <- degree(graph.postcomm)
vertices.postcomm$betweenness.postcomm <- betweenness(graph.postcomm)
vertices.postcomm$id <- as.integer(vertices.postcomm$id)
jfh.retention <- left_join(jfh.retention, vertices.postcomm, by = "id")
## 5.6. Export textual data along with user ID and text data type (post/comment) into a csv file to run through LINC
write.csv(jfh.data[jfh.data$type == "posts" | jfh.data$type == "comments", c(3, 4, 13)], "~/Desktop/jfh.linc.csv", row.names = FALSE)
```

The environment pane lists the objects created during the session:

- Global Environment +
- graph.postcomm List of 18
- jfh.activities... List of 4
- jfh.activities... List of 4
- jfh.commentlikes 9829 obs. of 11 variables
- jfh.comments 3493 obs. of 15 variables
- jfh.data 23826 obs. of 15 variables
- jfh.postlikes 13587 obs. of 11 variables
- jfh.posts 887 obs. of 14 variables
- jfh.retention 23 obs. of 29 variables
- jfh.users 627 obs. of 4 variables
- jfh.users.data 457 obs. of 3 variables
- vertices 457 obs. of 4 variables
- vertices.postcomm 233 obs. of 4 variables



```
## 5.3. Import LINC results file into dataframe
282 jfh.linc.results <- read.csv("~/Desktop/jfh.linc.results.csv", header = TRUE, stringsAsFactors = FALSE)
283
284 colnames(jfh.linc.results)[1:3] <- jfh.linc.results[1,1:3]
285 jfh.linc.results <- jfh.linc.results[2:nrow(jfh.linc.results),]
286
## 5.5. Calculate total LINC scores per user in posts and comments, and join to retention data
287
288 jfh.linc <- lapply(4:ncol(jfh.linc.results), function(x) aggregate(jfh.linc.results[,x] ~ from.anonid ~ type, jfh.linc.results, sum))
289 jfh.linc.peruser <- do.call(cbind, jfh.linc[2:93])
290 jfh.linc.peruser <- cbind(jfh.linc[1], jfh.linc.peruser[, -c(which(grep("type", colnames(jfh.linc.peruser)) ~ grep("from.anonid", colnames(jfh.linc
291 colnames(jfh.linc.peruser)[c(1,3):ncol(jfh.linc.peruser))] <- c("id", colnames(jfh.linc.results)[4:ncol(jfh.linc.results)])
292 jfh.linc.peruserid <- as.integer(jfh.linc.peruser[[1]])
293 jfh.linc.peruser <- split(jfh.linc.peruser, jfh.linc.peruserid)
294 jfh.retention <- left_join(jfh.retention, jfh.linc.peruser[[2]], c(1,3) ~ ncol(jfh.linc.peruser[[2]])), by = "id")
295 colnames(jfh.retention)[28:128] <- paste("posts", colnames(jfh.retention)[28:128], sep = "-")
296 jfh.retention <- left_join(jfh.retention, jfh.linc.peruser[[1]], c(1,3) ~ ncol(jfh.linc.peruser[[1]])), by = "id")
297 colnames(jfh.retention)[121:211] <- paste("comments", colnames(jfh.retention)[121:211], sep = "-")
298
## 5.6. Calculate total LINC scores per user in all textual data (post and comments combined), and join to retention data
299
300 jfh.linc.all <- lapply(4:ncol(jfh.linc.results), function(x) aggregate(jfh.linc.results[,x] ~ from.anonid, jfh.linc.results, sum))
301 jfh.linc.all <- do.call(cbind, jfh.linc.all[2:93])
302 jfh.linc.all.peruser <- cbind(jfh.linc.all[1], jfh.linc.all.peruser[, -c(which(grep("from.anonid", colnames(jfh.linc.all.peruser))))])
303 colnames(jfh.linc.all.peruser)[c(1:ncol(jfh.linc.all.peruser))] <- c("id", paste("all.", colnames(jfh.linc.results)[4:ncol(jfh.linc.results)]), sep =
304 jfh.linc.all.peruserid <- as.integer(jfh.linc.all.peruser[[1]])
305 jfh.retention <- left_join(jfh.retention, jfh.linc.all.peruser, by = "id")
306
## 5.7. Remove NAs from retention data dataframe
307
308 jfh.retention[is.na(jfh.retention)] <- 0
309
## 5.8. Conduct computerized linguistic analysis in LINC 1
310
```

Environment History Connections

Global Environment	
graph.postcomm	List of 18
jfh.activities...	List of 4
jfh.activities...	List of 4
jfh.commentlikes	6829 obs. of 11 variables
jfh.comments	3493 obs. of 15 variables
jfh.data	23826 obs. of 15 variables
jfh.linc.results	4437 obs. of 96 variables
jfh.postlikes	11587 obs. of 11 variables
jfh.posts	887 obs. of 14 variables
jfh.retention	23 obs. of 29 variables
jfh.users	927 obs. of 4 variables
jfh.users.data	457 obs. of 3 variables
vertices	457 obs. of 4 variables

Files Plots Packages Help Viewer

Console Terminal

```
- cd /Users/tyler/Documents/comp-apply-4/analysis/09inc/09inc-jfh-pb01 -
> filter(is.na(from.anonid)) %>%
> filter(is.na(to.anonid)) %>%
> distinct()
> vertices.postcomm <- make_sna(vertices(edges.postcomm, 1, 2))
> graph.data.frame(edges.postcomm, vertices.postcomm, directed = TRUE)
> graph.m.postcomm <- get_adjacency(graph.postcomm, sparse = FALSE)
> vertices.postcomm$degree.postcomm <- degree(graph.postcomm)
> vertices.postcomm$betaness.postcomm <- betweenness(graph.postcomm)
> vertices.postcomm$id <- as.integer(vertices.postcomm$id)
> jfh.retention <- left_join(jfh.retention, vertices.postcomm, by = "id")
> ## 5.1. Export textual data along with user ID and text data type (post/comment) into a csv file to run through LINC
> write.csv(jfh.data[jfh.datatype == "posts" | jfh.datatype == "comments", c(1, 4, 13)], "~/Desktop/jfh.linc.csv", row.names = FALSE)
> ## 5.4. Import LINC results file into dataframe
> jfh.linc.results <- read.csv("~/Desktop/jfh.linc.results.csv", header = TRUE, stringsAsFactors = FALSE)
> colnames(jfh.linc.results)[1:3] <- jfh.linc.results[1,1:3]
> jfh.linc.results <- jfh.linc.results[2:nrow(jfh.linc.results),]
> ## 5.4. Import LINC results file into dataframe
> jfh.linc.results <- read.csv("~/Desktop/jfh.linc.results.csv", header = TRUE, stringsAsFactors = FALSE)
> colnames(jfh.linc.results)[1:3] <- jfh.linc.results[1,1:3]
> jfh.linc.results <- jfh.linc.results[2:nrow(jfh.linc.results),]
>
```



```

287 # 5.5. Calculate total LINC scores per user in posts and comments, and join to retention data
288 jfh.linc <- lapply(4:ncol(jfh.linc.results), function(x) aggregate(jfh.linc.results[,x] ~ from.anonid ~ type, jfh.linc.results, sum))
289 jfh.linc.peruser <- do.call(cbind, jfh.linc[2:93])
290 jfh.linc.peruser <- cbind(jfh.linc[1], jfh.linc.peruser[, c(which(grepl("type", colnames(jfh.linc.peruser)) | grepl("from.anonid", colnames(jfh.linc
291 colnames(jfh.linc.peruser)[c(1,5):ncol(jfh.linc.peruser))] <- c("id", colnames(jfh.linc.results)[4:ncol(jfh.linc.results)])
292 jfh.linc.peruserid <- as.integer(jfh.linc.peruserid)
293 jfh.linc.peruser <- split(jfh.linc.peruser, jfh.linc.peruserid)
294 jfh.retention <- left_join(jfh.retention, jfh.linc.peruser[[2]][,c(1,5):ncol(jfh.linc.peruser[[2]]), by = "id"])
295 colnames(jfh.retention)[28:128] <- paste("posts", colnames(jfh.retention)[28:128], sep = ".")
296 jfh.retention <- left_join(jfh.retention, jfh.linc.peruser[[1]][,c(1,5):ncol(jfh.linc.peruser[[1]]), by = "id"])
297 colnames(jfh.retention)[121:213] <- paste("comments", colnames(jfh.retention)[121:213], sep = ".")
298
299 # 5.6. Calculate total LINC scores per user in all (actual) data (post and comments combined), and join to retention data
300 jfh.linc.all <- lapply(4:ncol(jfh.linc.results), function(x) aggregate(jfh.linc.results[,x] ~ from.anonid, jfh.linc.results, sum))
301 jfh.linc.all.peruser <- do.call(cbind, jfh.linc.all[2:93])
302 jfh.linc.all.peruser <- cbind(jfh.linc.all[1], jfh.linc.all.peruser[, c(which(grepl("from.anonid", colnames(jfh.linc.all.peruser))) | grepl("from.anonid", colnames(jfh.linc.all.peruser)))])
303 colnames(jfh.linc.all.peruser)[c(1,ncol(jfh.linc.all.peruser))] <- c("id", paste("all.", colnames(jfh.linc.results)[4:ncol(jfh.linc.results)]), sep =
304 jfh.linc.all.peruserid <- as.integer(jfh.linc.all.peruserid)
305 jfh.retention <- left_join(jfh.retention, jfh.linc.all.peruser, by = "id")
306
307 # 5.7. Remove NAs from retention data dataframe
308 jfh.retention[is.na(jfh.retention)] <- 0
309
310 ### 6. Regression analysis ###
311
312 # 6.1. Define independent variables
313 IVs <- c("commentlikes", "received",
314         "all.likes.received")
315
316 # 6.2. Conduct computerized logistic analysis in LINC 1
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

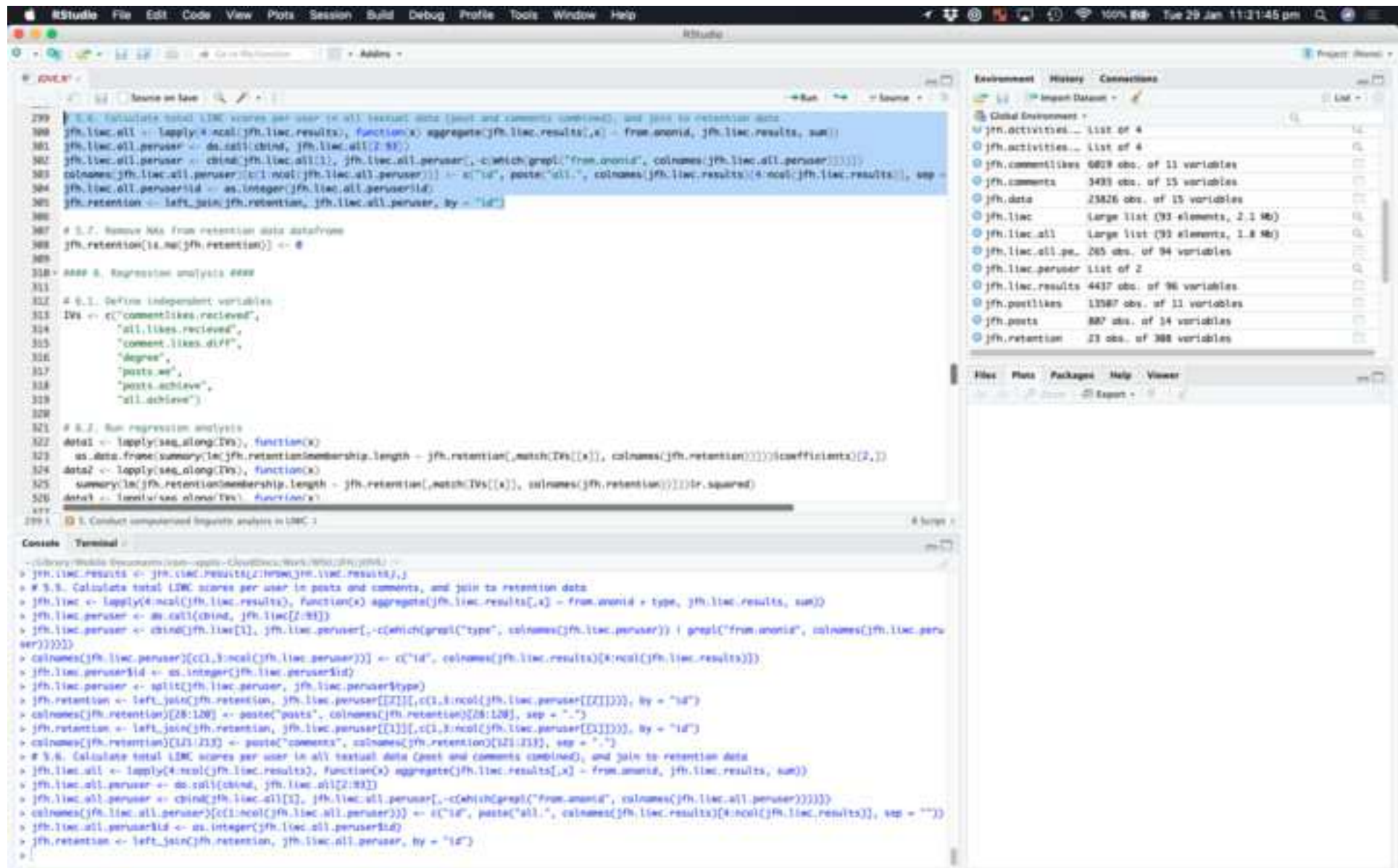
Environment: History Connections

Global Environment +

- graph.m.postcomm num (1:455, 1:455) 1 1 1 1 1 1 1 1 1 ...
- graph.postcomm List of 10
- jfh.activities... List of 4
- jfh.activities... List of 4
- jfh.commentlikes 6809 obs. of 11 variables
- jfh.comments 3493 obs. of 15 variables
- jfh.data 23826 obs. of 25 variables
- jfh.linc Large list (93 elements, 2.1 Mb)
- jfh.linc.peruser List of 2
- jfh.linc.results 4437 obs. of 96 variables
- jfh.postlikes 13587 obs. of 11 variables
- jfh.posts 887 obs. of 14 variables
- jfh.retention 23 obs. of 215 variables

Files Plots Packages Help Viewer

Export +



The screenshot displays the RStudio interface with a script editor on the left, an environment pane on the right, and a console at the bottom.

Script Editor:

```
# 5.6. Calculate total LMC scores per user in all textual data (posts and comments combined), and join to retention data
jfh.lmc.all <- lapply(4:ncol(jfh.lmc.results), function(x) aggregate(jfh.lmc.results[,x] ~ from.anonid, jfh.lmc.results, sum))
jfh.lmc.all.peruser <- do.call(cbind, jfh.lmc.all[2:93])
jfh.lmc.all.peruser <- cbind(jfh.lmc.all[1], jfh.lmc.all.peruser[, -c(which(grep("from.anonid", colnames(jfh.lmc.all.peruser))))])
colnames(jfh.lmc.all.peruser)[c(1:ncol(jfh.lmc.all.peruser))] <- c("id", paste("all.", colnames(jfh.lmc.results)[4:ncol(jfh.lmc.results)], sep =
jfh.lmc.all.peruserid <- as.integer(jfh.lmc.all.peruserid)
jfh.retention <- left_join(jfh.retention, jfh.lmc.all.peruser, by = "id")

# 5.7. Remove NAs from retention data dataframe
jfh.retention[is.na(jfh.retention)] <- 0

## 5. Regression analysis ##

# 5.1. Define independent variables
IVs <- c("commentlikes.recieved",
        "all.likes.recieved",
        "comment.likes.diff",
        "degree",
        "posts.w",
        "posts.achieve",
        "all.achieve")

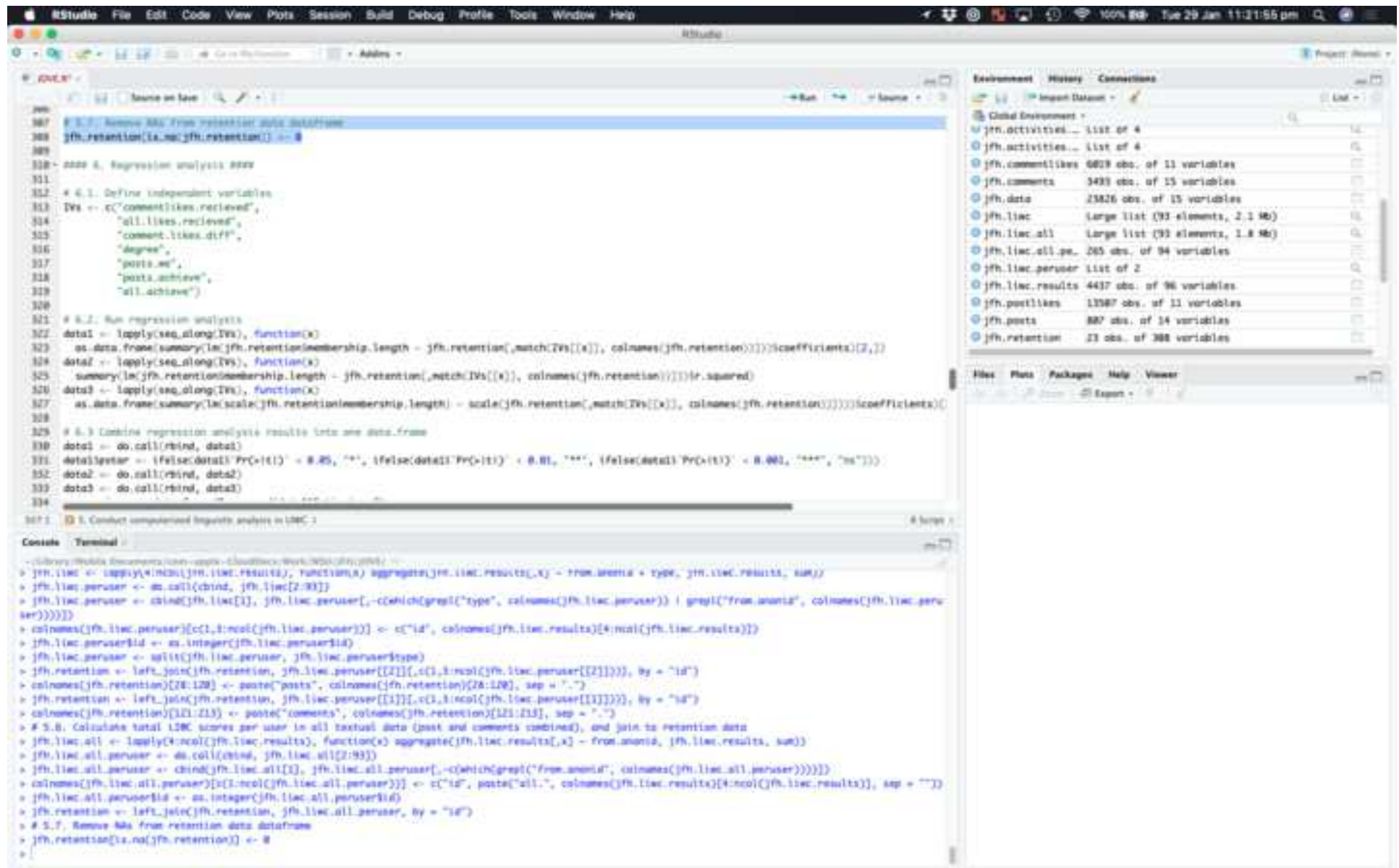
# 5.2. Run regression analysis
data1 <- lapply(seq_along(IVs), function(x)
  as.data.frame(summary(lm(jfh.retention[,match(IVs[x]], colnames(jfh.retention))])$coefficients)[2,])
data2 <- lapply(seq_along(IVs), function(x)
  summary(lm(jfh.retention[,match(IVs[x]], colnames(jfh.retention))])$r.squared)
data3 <- lapply(seq_along(IVs), function(x)
  # 5.3. Conduct completed logistic analysis in LMC 1
```

Environment Pane:

- Global Environment +
- jfh.activities... list of 4
- jfh.activities... list of 4
- jfh.commentlikes 6819 obs. of 11 variables
- jfh.comments 3439 obs. of 15 variables
- jfh.data 23826 obs. of 15 variables
- jfh.lmc large list (93 elements, 2.1 Mb)
- jfh.lmc.all large list (93 elements, 1.8 Mb)
- jfh.lmc.all.pe 265 obs. of 94 variables
- jfh.lmc.peruser list of 2
- jfh.lmc.results 4437 obs. of 96 variables
- jfh.postlikes 13587 obs. of 11 variables
- jfh.posts 887 obs. of 14 variables
- jfh.retention 23 obs. of 388 variables

Console:

```
~/Library/Mobile Documents (Mac) - Apple - CloudDocs/Work/9800/jfh/jfh4 <-
> jfh.lmc.results <- jfh.lmc.results[2:1000,jfh.lmc.results[,]]
> # 5.6. Calculate total LMC scores per user in posts and comments, and join to retention data
> jfh.lmc <- lapply(4:ncol(jfh.lmc.results), function(x) aggregate(jfh.lmc.results[,x] ~ from.anonid + type, jfh.lmc.results, sum))
> jfh.lmc.peruser <- do.call(cbind, jfh.lmc[2:93])
> jfh.lmc.peruser <- cbind(jfh.lmc[1], jfh.lmc.peruser[, -c(which(grep("type", colnames(jfh.lmc.peruser)) | grep("from.anonid", colnames(jfh.lmc.peruser)))))
> colnames(jfh.lmc.peruser)[c(1,3:ncol(jfh.lmc.peruser))] <- c("id", colnames(jfh.lmc.results)[4:ncol(jfh.lmc.results)])
> jfh.lmc.peruserid <- as.integer(jfh.lmc.peruserid)
> jfh.lmc.peruser <- split(jfh.lmc.peruser, jfh.lmc.peruserid)
> jfh.retention <- left_join(jfh.retention, jfh.lmc.peruser[[1]][,c(1,3:ncol(jfh.lmc.peruser[[1]]))], by = "id")
> colnames(jfh.retention)[28:120] <- paste("posts", colnames(jfh.retention)[28:120], sep = ".")
> jfh.retention <- left_join(jfh.retention, jfh.lmc.peruser[[1]][,c(1,3:ncol(jfh.lmc.peruser[[1]]))], by = "id")
> colnames(jfh.retention)[121:133] <- paste("comments", colnames(jfh.retention)[121:133], sep = ".")
> # 5.6. Calculate total LMC scores per user in all textual data (posts and comments combined), and join to retention data
> jfh.lmc.all <- lapply(4:ncol(jfh.lmc.results), function(x) aggregate(jfh.lmc.results[,x] ~ from.anonid, jfh.lmc.results, sum))
> jfh.lmc.all.peruser <- do.call(cbind, jfh.lmc.all[2:93])
> jfh.lmc.all.peruser <- cbind(jfh.lmc.all[1], jfh.lmc.all.peruser[, -c(which(grep("from.anonid", colnames(jfh.lmc.all.peruser)))))
> colnames(jfh.lmc.all.peruser)[c(1:ncol(jfh.lmc.all.peruser))] <- c("id", paste("all.", colnames(jfh.lmc.results)[4:ncol(jfh.lmc.results)], sep = ""))
> jfh.lmc.all.peruserid <- as.integer(jfh.lmc.all.peruserid)
> jfh.retention <- left_join(jfh.retention, jfh.lmc.all.peruser, by = "id")
>
```



The screenshot displays the RStudio interface with a script editor on the left, a console at the bottom, and an environment pane on the right.

Script Editor (Lines 310-337):

```

310- ### 6. Regression analysis ###
311-
312- # 6.1. Define independent variables
313- IVs <- c("commentlikes_received",
314-         "all_likes_received",
315-         "comment_likes_diff",
316-         "degree",
317-         "posts_wrt",
318-         "posts_achieve",
319-         "all_achieve")
320-
321- # 6.2. Run regression analysis
322- data1 <- lapply(seq_along(IVs), function(x)
323-   as.data.frame(summary(lm(jfh.retentionmembership.length ~ jfh.retention[,match(IVs[[x]], colnames(jfh.retention))])$coefficients[2,])
324-   data2 <- lapply(seq_along(IVs), function(x)
325-     summary(lm(jfh.retentionmembership.length ~ jfh.retention[,match(IVs[[x]], colnames(jfh.retention))])$r.squared)
326-   data3 <- lapply(seq_along(IVs), function(x)
327-     as.data.frame(summary(lm(scale(jfh.retentionmembership.length) ~ scale(jfh.retention[,match(IVs[[x]], colnames(jfh.retention))])$coefficients[2,])
328-
329- # 6.3 Combine regression analysis results into one data.frame
330- data1 <- do.call(rbind, data1)
331- data1$estor <- ifelse(data1$Pr<=0.01) ~ 0.05, "+", ifelse(data1$Pr<=0.01) ~ 0.01, "****", ifelse(data1$Pr<=0.01) ~ 0.001, "****", "ns"))
332- data2 <- do.call(rbind, data2)
333- data3 <- do.call(rbind, data3)
334- regressions <- data.frame(B = round(data1$Estimate, 2),
335-                          SE = round(data1$Std. Error, 2),
336-                          Beta = paste(round(data3, 2), data1$estor, sep = ""),
337-                          R2 = round(data2, 2))
338-
339- ### 6. Regression analysis 1

```

Environment Pane:

- Global Environment +
 - jfh.liac.all.pe. 205 obs. of 94 variables
 - jfh.liac.peruser list of 2
 - jfh.liac.results 4437 obs. of 96 variables
 - jfh.postlikes 53507 obs. of 11 variables
 - jfh.posts 887 obs. of 14 variables
 - jfh.retention 23 obs. of 308 variables
 - jfh.users 627 obs. of 4 variables
 - jfh.users.data 437 obs. of 3 variables
 - vertices 457 obs. of 4 variables
 - vertices.postco. 233 obs. of 4 variables
- Values

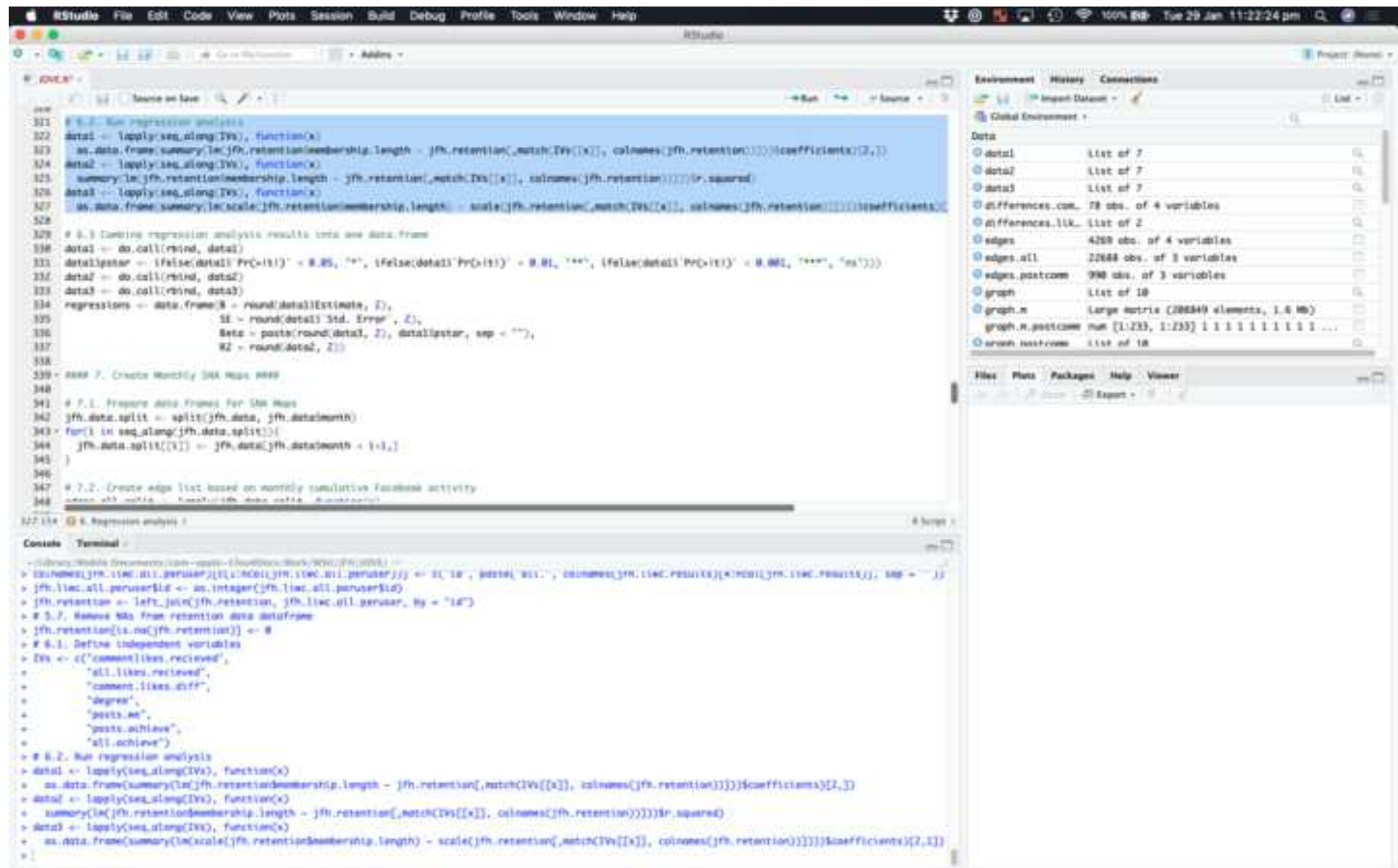
1	41
IVs	chr [1:7] "commentlikes_received" "all_likes_"

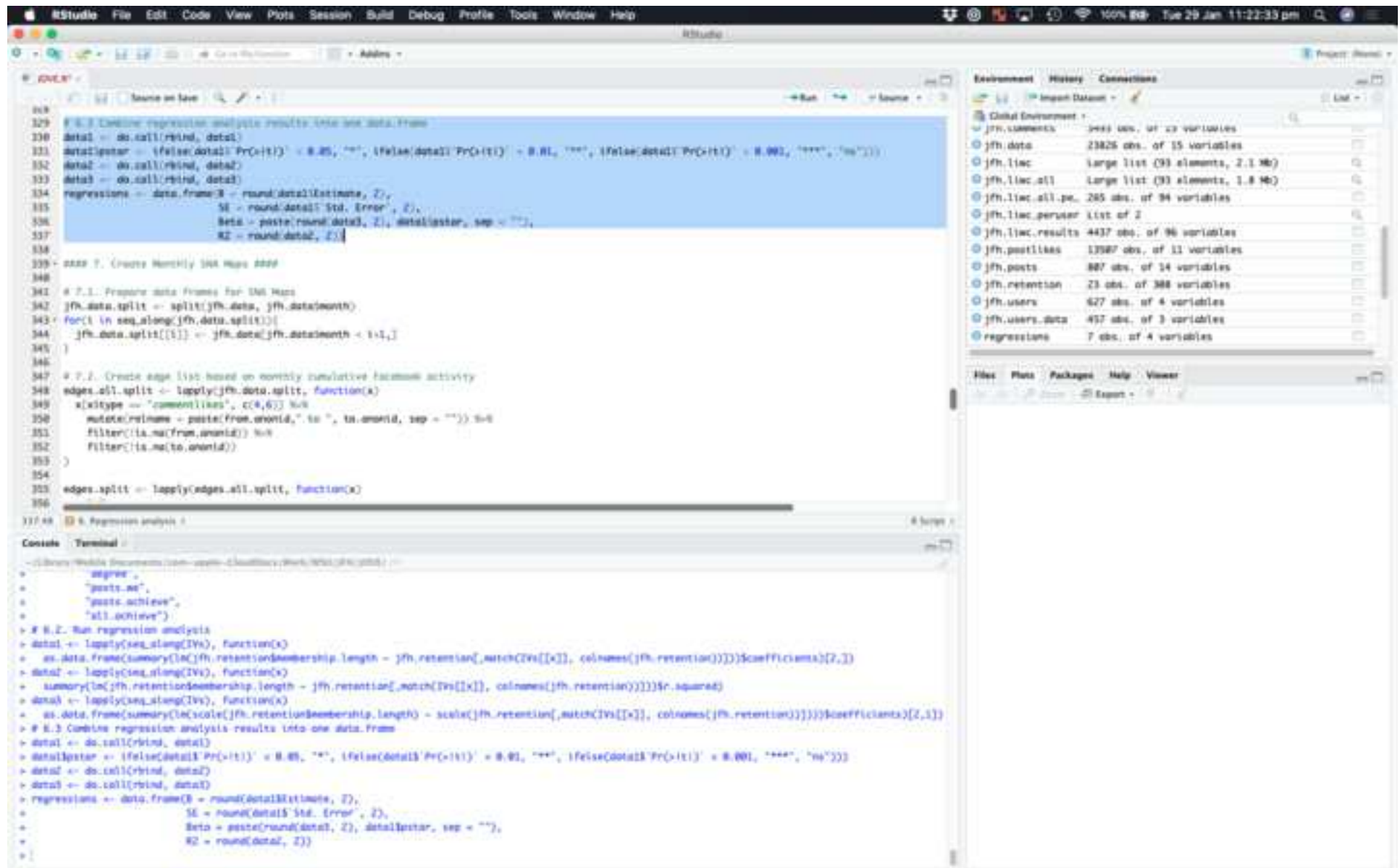
Console:

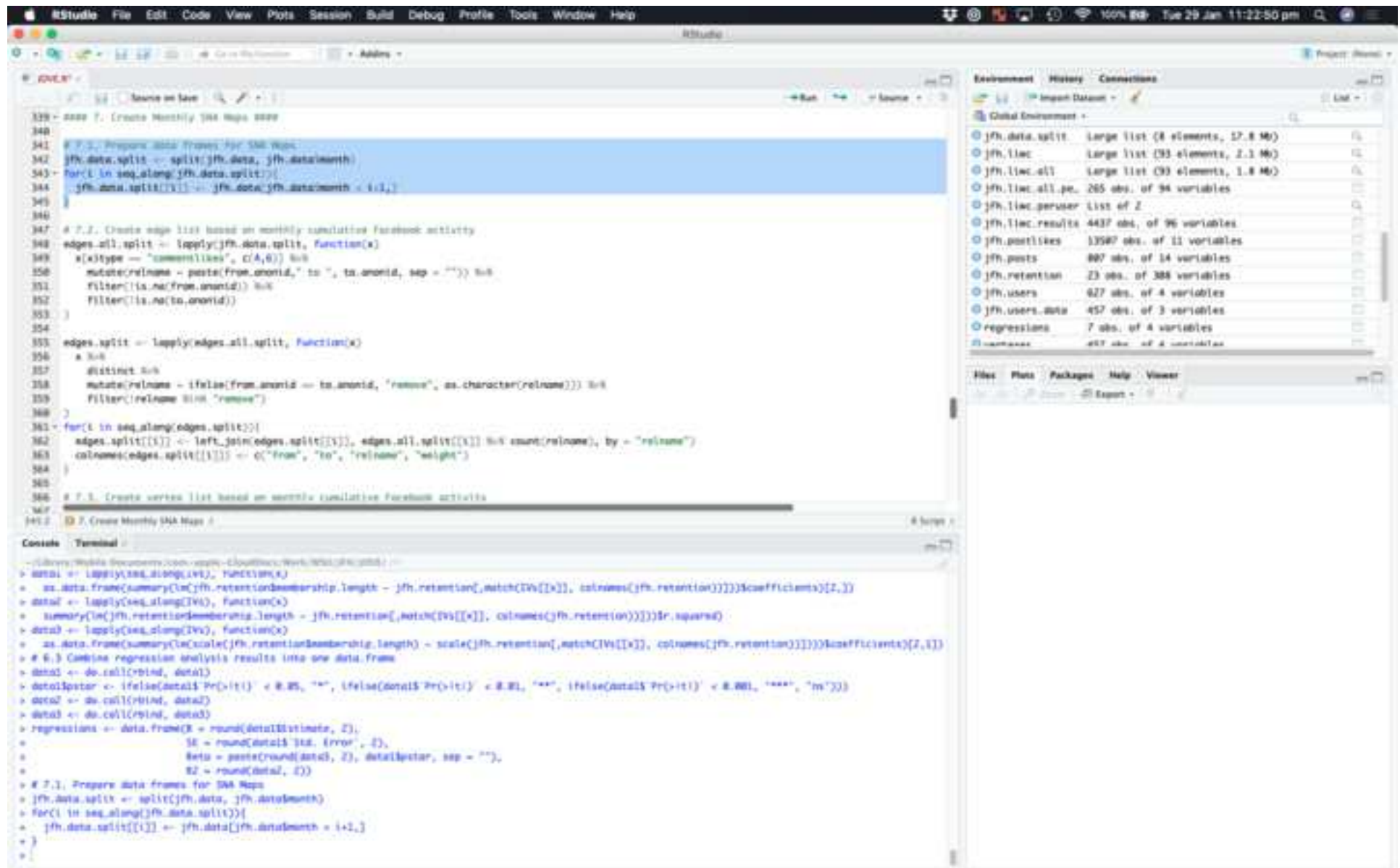
```

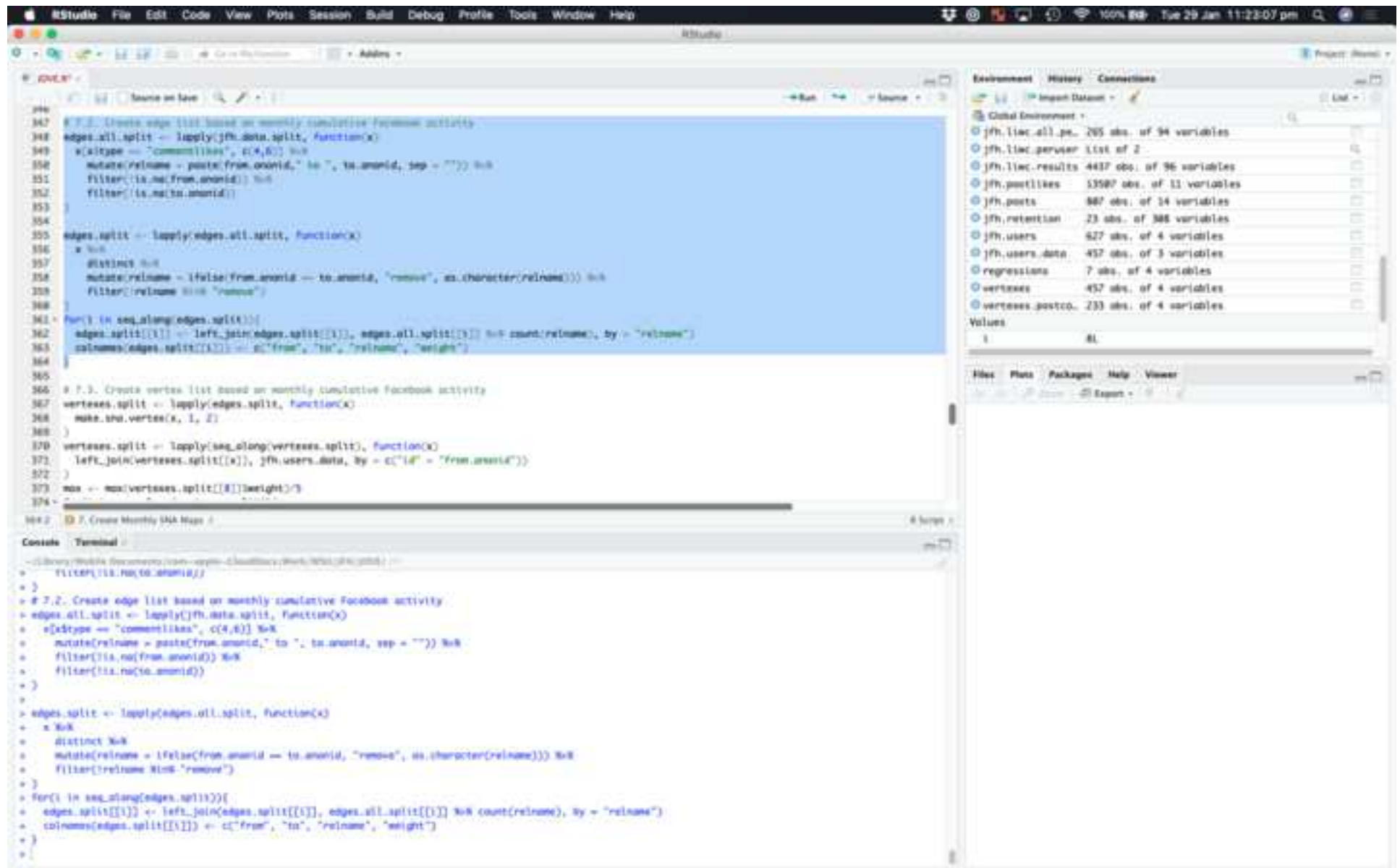
> colnames(jfh.retention)[28:100] <- paste0("posts_", colnames(jfh.retention)[28:100], sep = ".")
> jfh.retention <- left_join(jfh.retention, jfh.liac.peruser[[1]][,c(3,lincol(jfh.liac.peruser[[1]]))], by = "id")
> colnames(jfh.retention)[121:213] <- paste0("comments_", colnames(jfh.retention)[121:213], sep = ".")
> # 5.6. Calculate total LIAC scores per user in all textual data (post and comments combined), and join to retention data
> jfh.liac.all <- lapply(4:ncol(jfh.liac.results), function(x) aggregate(jfh.liac.results[,x] ~ from.anonid, jfh.liac.results, sum))
> jfh.liac.all.peruser <- do.call(rbind, jfh.liac.all[2:100])
> jfh.liac.all.peruser <- cbind(jfh.liac.all[1], jfh.liac.all.peruser[,c(which(grep("from.anonid", colnames(jfh.liac.all.peruser))))])
> colnames(jfh.liac.all.peruser)[c(1:ncol(jfh.liac.all.peruser))] <- c("id", paste("all.", colnames(jfh.liac.results)[4:ncol(jfh.liac.results)], sep = "."))
> jfh.liac.all.peruser$id <- as.integer(jfh.liac.all.peruser$id)
> jfh.retention <- left_join(jfh.retention, jfh.liac.all.peruser, by = "id")
> # 5.7. Remove NAs from retention data dataframe
> jfh.retention[is.na(jfh.retention)] <- 0
> # 6.1. Define independent variables
> IVs <- c("commentlikes_received",
+       "all_likes_received",
+       "comment_likes_diff",
+       "degree",
+       "posts_wrt",
+       "posts_achieve",
+       "all_achieve")
>

```









```

# 7.3. Create vertex list based on monthly cumulative Facebook activity
vertices.split <- lapply(edges.split, function(x)
  make sna.vertex(x, 1, 2)
)
vertices.split <- lapply(seq_along(vertices.split), function(x)
  left_join(vertices.split[[x]], jfh.users.data, by = c("id" = "from_userid"))
)
max <- max(vertices.split[[1]]$weight)/5
for(i in seq_along(vertices.split)){
  vertices.split[[i]]$weight2 <- ifelse(vertices.split[[i]]$weight < max*1, 1,
    ifelse(vertices.split[[i]]$weight < max*2, 2,
      ifelse(vertices.split[[i]]$weight < max*3, 3,
        ifelse(vertices.split[[i]]$weight < max*4, 4, 5)))
  vertices.split[[i]]$label <- ifelse(vertices.split[[i]]$role == 2, vertices.split[[i]]$id, "")
}

# 7.4. Create graphs and graph matrices based on monthly cumulative Facebook activity
graph.split <- lapply(seq_along(edges.split), function(x)
  graph.data.frame(edges.split[[x]], vertices.split[[x]], directed = TRUE)
)
graph.m.split <- lapply(graph.split, function(x)
  get.adjacency(x, sparse = FALSE)
)

# 7.5. Set layout of 346 maps based on cumulative Facebook activity
layout.split <- lapply(graph.split, function(x) layout.fruchterman.reingold(x))

# 7.6. Add information based on vertex labels

# 7.7. Create Monthly SNA Maps

```

Environment

Object	Class	Attributes
jfh.linc.results	data.frame	4437 obs. of 36 variables
jfh.postlikes	data.frame	13567 obs. of 11 variables
jfh.posts	data.frame	887 obs. of 14 variables
jfh.relation	data.frame	23 obs. of 388 variables
jfh.users	data.frame	627 obs. of 4 variables
jfh.users.data	data.frame	457 obs. of 3 variables
regressions	data.frame	7 obs. of 4 variables
vertices	data.frame	457 obs. of 4 variables
vertices.postco	data.frame	233 obs. of 4 variables
vertices.split	list	list of 2

Values

Object	Value
1	1
2	2
3	3
4	4
5	5

Files

Plots

Packages

Help

Viewer

Export

The screenshot shows the RStudio interface with a script editor, a console, and an environment pane.

Script Editor:

```

382 # 7.4. Create graphs and graph matrices based on monthly cumulative Facebook activity
383 graph.split <- lapply(seq_along(edges.split), function(x)
384   graph.data.frame(edges.split[[x]], vertices.split[[x]], directed = TRUE)
385 )
386 graph.m.split <- lapply(graph.split, function(x)
387   get.adjacency(x, sparse = FALSE)
388 )
389
390 # 7.5. Set layout of SNA maps based on cumulative Facebook activity
391 layout.split <- lapply(graph.split, function(x) layout.fruchterman.reingold(x))
392
393 # 7.6. Add colours based on user roles
394 cols <- c("red", "blue", "green")
395 for(i in seq_along(graph.split)){
396   V(graph.split[[i]])$color <- cols[V(graph.split[[i]])$role]
397 }
398
399 # 7.7. Create SNA maps and save to file
400 for(i in seq_along(graph.split)){
401   png(file = paste0("/Users/muhammadiqbal/Desktop/SNA maps/", i, ".png", sep = ""), width = 5000, height = 5000)
402   plot(graph.split[[i]], layout=layout.split[[i]],
403         edge.arrow.size = 3,
404         edge.width = 6,
405         vertex.size = V(graph.split[[i]])$weight2*4.5,
406         vertex.frame.color="black",
407         vertex.label = V(graph.split[[i]])$label,
408         vertex.label.cex = V(graph.split[[i]])$weight2*4,
409         main = paste0("SNA Map for ", i))
410 }
411
412 # 7.8. Create Monthly SNA Maps

```

Environment Pane:

Object	Class	Size
graph.m.split	Large list (8 elements, 2 Mb)	
graph.postcom	List of 10	
graph.split	Large list (8 elements, 1.2 Mb)	
jfh.activities...	List of 4	
jfh.activities...	List of 4	
jfh.commentlikes	6829 obs. of 11 variables	
jfh.comments	3439 obs. of 15 variables	
jfh.data	23826 obs. of 15 variables	
jfh.data.split	Large list (8 elements, 17.8 Mb)	
jfh.liwc	Large list (33 elements, 2.1 Mb)	
jfh.liwc.all	Large list (33 elements, 1.8 Mb)	
jfh.liwc.all.pe	265 obs. of 94 variables	
jfh.liwc.maxsize	List of 2	

Console:

```

> make_sna_vertices(x, i, L)
>
> vertices.split <- lapply(seq_along(vertices.split), function(x)
+   left_join(vertices.split[[x]], jfh.users.data, by = c("id" = "facebook_id"))
+ )
> max <- max(vertices.split[[1]]$weight)*5
> for(i in seq_along(vertices.split)){
+   vertices.split[[i]]$weight2 <- ifelse(vertices.split[[i]]$weight < max*1, 1,
+   ifelse(vertices.split[[i]]$weight < max*2, 2,
+   ifelse(vertices.split[[i]]$weight < max*3, 3,
+   ifelse(vertices.split[[i]]$weight < max*4, 4, 5)))
+   vertices.split[[i]]$label <- ifelse(vertices.split[[i]]$role == 2, vertices.split[[i]]$id, "")
+ }
> # 7.4. Create graphs and graph matrices based on monthly cumulative Facebook activity
> graph.split <- lapply(seq_along(edges.split), function(x)
+   graph.data.frame(edges.split[[x]], vertices.split[[x]], directed = TRUE)
+ )
> graph.m.split <- lapply(graph.split, function(x)
+   get.adjacency(x, sparse = FALSE)
+ )
>

```

The screenshot shows the RStudio interface with a script editor on the left, a console at the bottom, and an environment pane on the right.

Script Editor:

```

288 # 7.5. Set layout of SNA maps based on cumulative Facebook activity
289 layout.split <- lapply(graph.split, function(x) layout.fruchterman.reingold(x))
290
291 # 7.6. Add colours based on user roles
292 cols <- c("M00KFF", "M00000", "M00000")
293 for(i in seq_along(graph.split)){
294   V(graph.split[[i]])$color <- cols[V(graph.split[[i]])$role]
295 }
296
297 # 7.7. Create SNA map and save to file
298 for(i in seq_along(graph.split)){
299   png(file = paste0("~/Users/muhammadipul/Desktop/SNA maps/", i, ".png", sep = ""), width = 1000, height = 500)
300   plot <- plot(graph.split[[i]], layout=layout.split[[i]],
301     edge.arrow.size = 3,
302     edge.width = 8,
303     vertex.size = V(graph.split[[i]])$weight*4,
304     vertex.frame.color="black",
305     vertex.label = V(graph.split[[i]])$label,
306     vertex.label.cex = V(graph.split[[i]])$weight*4,
307     vertex.label.family = "Helvetica",
308     vertex.label.color= "black"
309   )
310   dev.off()
311 }
312
313 # 7.8. Calculate monthly cumulative Facebook activity of the Facebook group
314
315 # 7.9. Create Monthly SNA Maps

```

Console:

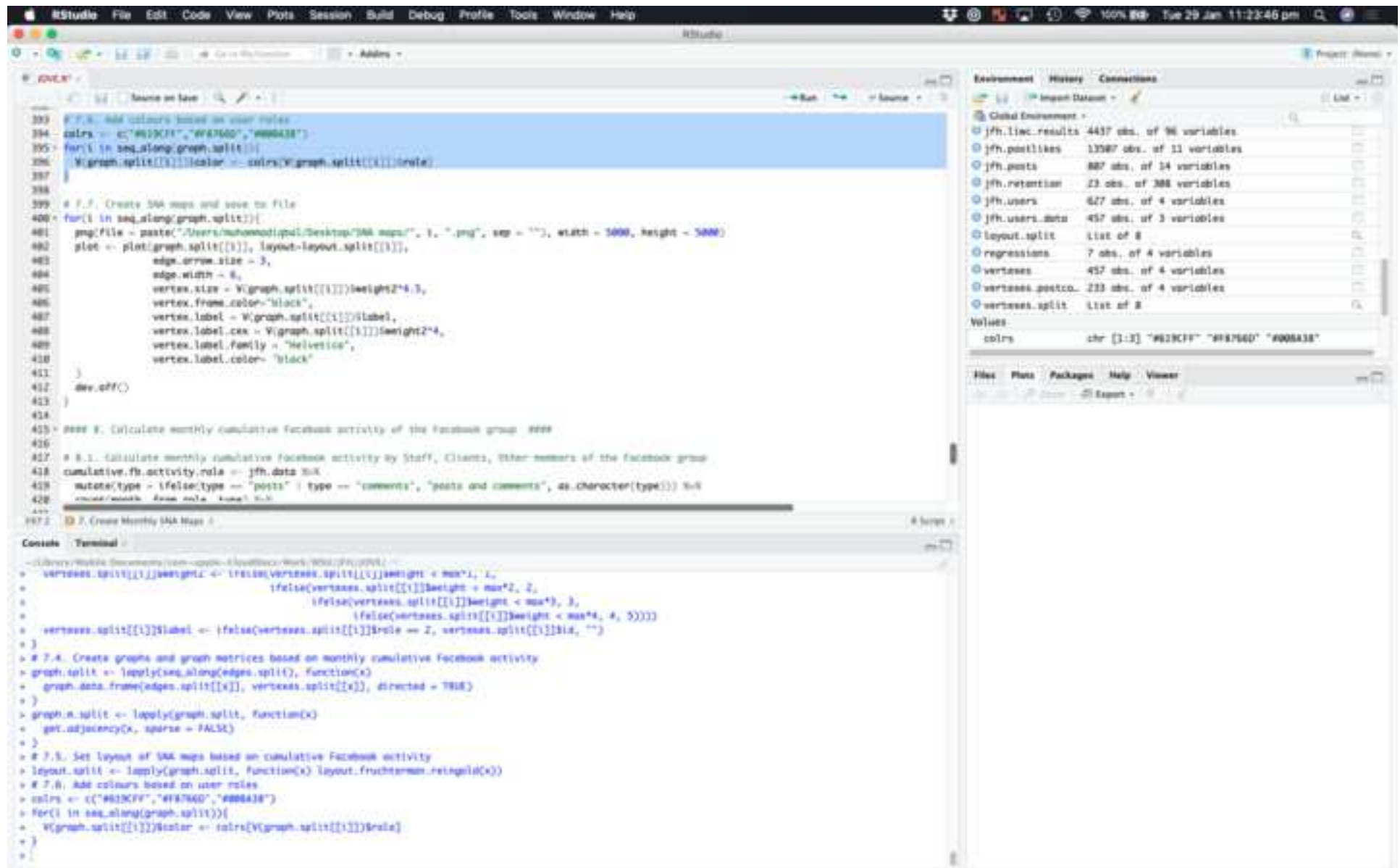
```

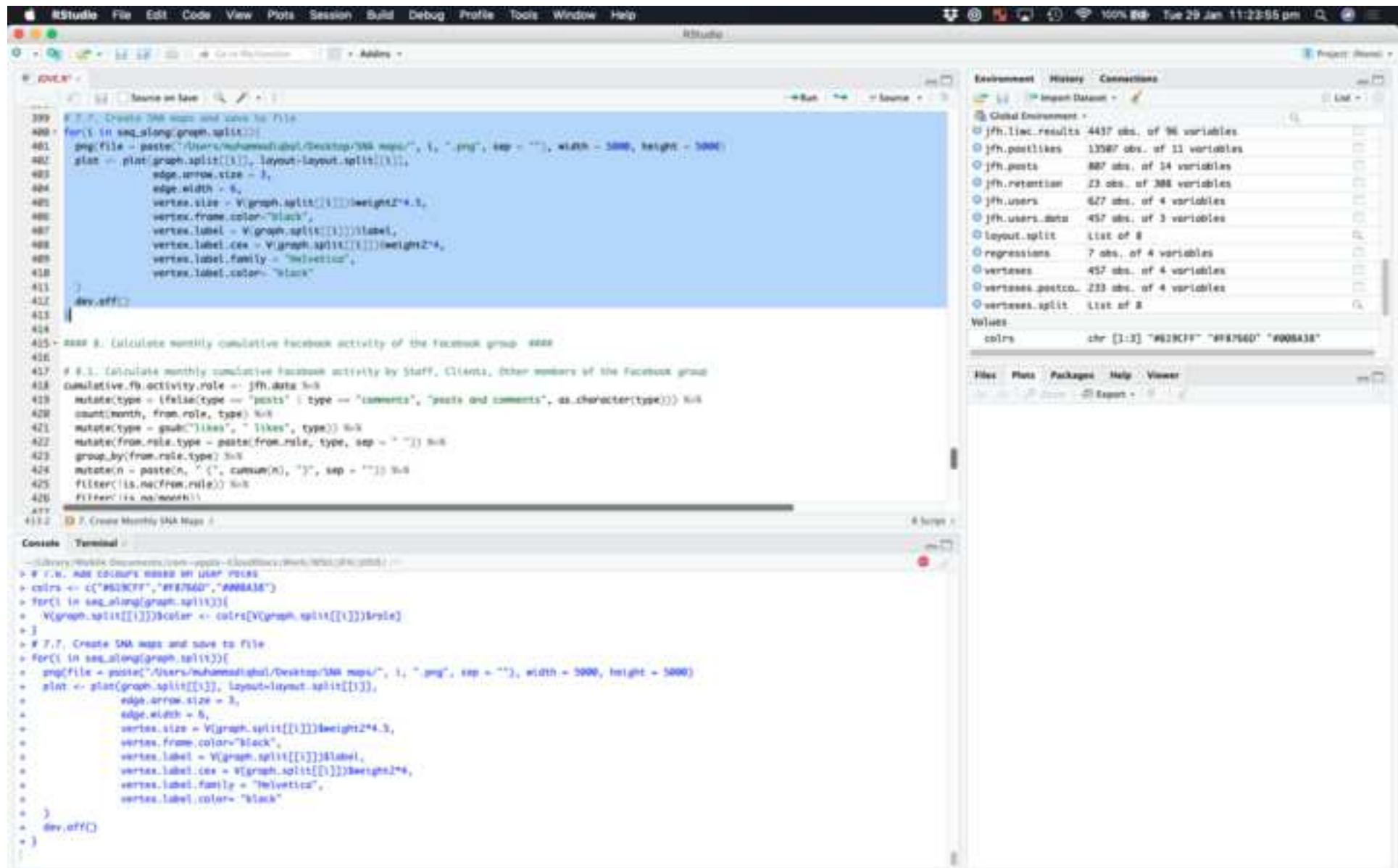
~/Library/Mobile Documents/com-apple-iCloudDocs/Desktop/jfh/jfh2012
> vertices.split <- lapply(seq_along(vertices.split), function(x){
+   left_join(vertices.split[[x]], jfh.users.data, by = c("id" = "from_userid"))
+ })
> max <- max(vertices.split[[8]]$weight)/5
> for(i in seq_along(vertices.split)){
+   vertices.split[[i]]$weight2 <- ifelse(vertices.split[[i]]$weight < max*1, 1,
+     ifelse(vertices.split[[i]]$weight < max*2, 2,
+       ifelse(vertices.split[[i]]$weight < max*3, 3,
+         ifelse(vertices.split[[i]]$weight < max*4, 4, 5)))
+   )
+   vertices.split[[i]]$label <- ifelse(vertices.split[[i]]$role == "Z", vertices.split[[i]]$id, "")
+ }
> # 7.4. Create graphs and graph matrices based on monthly cumulative Facebook activity
> graph.split <- lapply(seq_along(edges.split), function(x){
+   graph.data.frame(edges.split[[x]], vertices.split[[x]], directed = TRUE)
+ })
> graph.w.split <- lapply(graph.split, function(x){
+   get.adjacency(x, sparse = FALSE)
+ })
> # 7.5. Set layout of SNA maps based on cumulative Facebook activity
> layout.split <- lapply(graph.split, function(x) layout.fruchterman.reingold(x))
>

```

Environment:

Object	Class	Size
jfh.data	data.frame	234/9 obs. of 15 variables
jfh.data.split	list	Large list (8 elements, 17.8 Mb)
jfh.liac	list	Large list (93 elements, 2.1 Mb)
jfh.liac.all	list	Large list (93 elements, 1.8 Mb)
jfh.liac.all.pe	data.frame	265 obs. of 94 variables
jfh.liac.peruser	list	List of 2
jfh.liac.results	data.frame	4437 obs. of 96 variables
jfh.postlikes	data.frame	13587 obs. of 11 variables
jfh.posts	data.frame	887 obs. of 14 variables
jfh.retention	data.frame	23 obs. of 388 variables
jfh.users	data.frame	627 obs. of 4 variables
jfh.users.data	data.frame	457 obs. of 3 variables
layout.split	list	List of 8





```

# 7.7. Create SNA maps and save to file
for(i in seq_along(graph.split)){
  png(file = paste0("~/Users/muhammadibrahim/Desktop/SNA maps/", i, ".png", sep = ""), width = 5000, height = 5000)
  plot(graph.split[[i]], layout=layout.split[[i]],
        edge.arrow.size = 3,
        edge.width = 5,
        vertex.size = V(graph.split[[i]])$weight2*4.5,
        vertex.frame.color="black",
        vertex.label = V(graph.split[[i]])$label,
        vertex.label.cex = V(graph.split[[i]])$weight2*4,
        vertex.label.family = "helvetica",
        vertex.label.color= "black"
  )
  dev.off()
}

#### 8. Calculate monthly cumulative facebook activity of the facebook group ####

# 8.1. Calculate monthly cumulative facebook activity by Staff, Clients, Other members of the Facebook group
cumulative_fb_activity_role <- jfh.data %>%
  mutate(type = ifelse(type == "posts" | type == "comments", "posts and comments", as.character(type))) %>%
  count(month, from.role, type) %>%
  mutate(type = gsub("likes", "likes", type)) %>%
  mutate(from.role.type = paste(from.role, type, sep = " ")) %>%
  group_by(from.role.type) %>%
  mutate(n = paste(n, " (" , cumsum(n), ")", sep = "")) %>%
  filter(!is.na(from.role)) %>%
  filter(!is.na(month))

# 8.2. Create Monthly SNA Maps

```

```

# 7.6. Add colours based on user roles
colrs <- c("#81C784", "#F8766D", "#A08A38")
for(i in seq_along(graph.split)){
  V(graph.split[[i]])$color <- colrs[V(graph.split[[i]])$role]
}

# 7.7. Create SNA maps and save to file
for(i in seq_along(graph.split)){
  png(file = paste0("~/Users/muhammadibrahim/Desktop/SNA maps/", i, ".png", sep = ""), width = 5000, height = 5000)
  plot <- plot(graph.split[[i]], layout=layout.split[[i]],
        edge.arrow.size = 3,
        edge.width = 5,
        vertex.size = V(graph.split[[i]])$weight2*4.5,
        vertex.frame.color="black",
        vertex.label = V(graph.split[[i]])$label,
        vertex.label.cex = V(graph.split[[i]])$weight2*4,
        vertex.label.family = "helvetica",
        vertex.label.color= "black"
  )
  dev.off()
}

```

Environment History Connections

Global Environment +

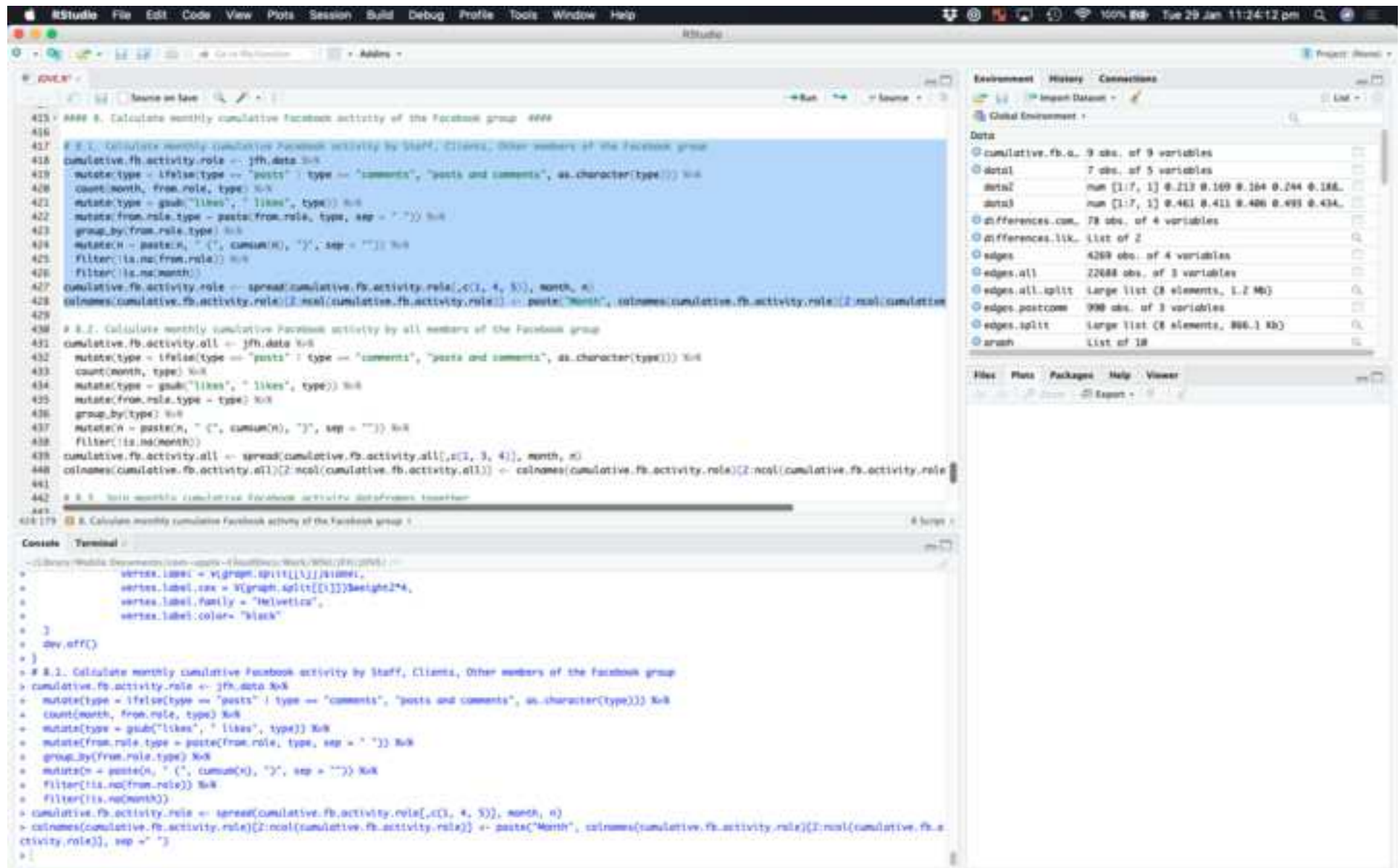
- jfh.lin.results 4437 obs. of 96 variables
- jfh.postlikes 13567 obs. of 11 variables
- jfh.posts 887 obs. of 14 variables
- jfh.retention 23 obs. of 388 variables
- jfh.users 627 obs. of 4 variables
- jfh.users.data 457 obs. of 3 variables
- layout.split list of 8
- regressions 7 obs. of 4 variables
- vertices 457 obs. of 4 variables
- vertices.postco 233 obs. of 4 variables
- vertices.split list of 8

Values

colrs chr [1:3] "#81C784" "#F8766D" "#A08A38"

Files **Plots** **Packages** **Help** **Viewer**

Export



```

# 8.2. Calculate monthly cumulative Facebook activity by all members of the Facebook group
cumulative.fb.activity.all <- jfh.data %>%
  mutate(type = ifelse(type == "posts" | type == "comments", "posts and comments", as.character(type))) %>%
  count(month, type) %>%
  mutate(type = glue("likes", " likes", type)) %>%
  mutate(from.role.type = type) %>%
  group_by(type) %>%
  mutate(n = paste(n, " (" , cumsum(n), ")", sep = "")) %>%
  filter(!is.na(month))
cumulative.fb.activity.all <- spread(cumulative.fb.activity.all[,c(1, 3, 4)], month, n)
colnames(cumulative.fb.activity.all)[2:ncol(cumulative.fb.activity.all)] <- colnames(cumulative.fb.activity.role)[2:ncol(cumulative.fb.activity.role)]

# 8.3. Join monthly cumulative Facebook activity dataframes together
cumulative.fb.activity <- bind_rows(cumulative.fb.activity.all, cumulative.fb.activity.role)

# 8.4. Calculate monthly cumulative Facebook activity by all members of the Facebook group
cumulative.fb.activity.all <- jfh.data %>%
  mutate(type = ifelse(type == "posts" | type == "comments", "posts and comments", as.character(type))) %>%
  count(month, type) %>%
  mutate(type = glue("likes", " likes", type)) %>%
  mutate(from.role.type = type) %>%
  group_by(type) %>%
  mutate(n = paste(n, " (" , cumsum(n), ")", sep = "")) %>%
  filter(!is.na(month))
cumulative.fb.activity.all <- spread(cumulative.fb.activity.all[,c(1, 3, 4)], month, n)
colnames(cumulative.fb.activity.all)[2:ncol(cumulative.fb.activity.all)] <- colnames(cumulative.fb.activity.role)[2:ncol(cumulative.fb.activity.role)]

```

Environment History Connections

Global Environment +

Data

- cumulative.fb.a 3 obs. of 9 variables
- cumulative.fb.a 9 obs. of 9 variables
- data1 7 obs. of 5 variables
- data2 num [1:7, 1] 0.213 0.189 0.184 0.244 0.188
- data3 num [1:7, 1] 0.461 0.411 0.406 0.493 0.434
- differences.com 78 obs. of 4 variables
- differences.lik list of 2
- edges 4268 obs. of 4 variables
- edges.all 22688 obs. of 3 variables
- edges.all.split Large list (8 elements, 1.2 Mb)
- edges.postcom 998 obs. of 3 variables
- edges.split Large list (8 elements, 868.1 Kb)

Files Plots Packages Help Viewer

Export

```

433 count(month, type) %>%
434 mutate(type = gsub("likes", "likes", type)) %>%
435 mutate(from.role.type = type) %>%
436 group_by(type) %>%
437 mutate(n = paste(n, "(", cumsum(n), ")", sep = "")) %>%
438 filter(!is.na(month))
439 cumulative.fb.activity.all <- spread(cumulative.fb.activity.all[,c(1, 3, 4)], month, n)
440 colnames(cumulative.fb.activity.all)[2:ncol(cumulative.fb.activity.all)] <- colnames(cumulative.fb.activity.role)[2:ncol(cumulative.fb.activity.role)]
441
442 # 8.3. Join monthly cumulative Facebook activity dataframes together
443 cumulative.fb.activity <- bind_rows(cumulative.fb.activity.all, cumulative.fb.activity.role)
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460

```

441-44: 8. Calculate monthly cumulative Facebook activity of the Facebook group

```

# library(Mobile Documents)
# group_by(FROM.role.type) %>%
# mutate(n = paste(n, "(", cumsum(n), ")", sep = "")) %>%
# filter(!is.na(from.role)) %>%
# filter(!is.na(month))
# cumulative.fb.activity.role <- spread(cumulative.fb.activity.role[,c(1, 4, 5)], month, n)
# colnames(cumulative.fb.activity.role)[2:ncol(cumulative.fb.activity.role)] <- paste("Month", colnames(cumulative.fb.activity.role)[2:ncol(cumulative.fb.a
ctivity.role)], sep = " ")
# 8.2. Calculate monthly cumulative Facebook activity by all members of the Facebook group
# cumulative.fb.activity.all <- jfb.data %>%
# mutate(type = ifelse(type == "posts" | type == "comments", "posts and comments", as.character(type))) %>%
# count(month, type) %>%
# mutate(type = gsub("likes", "likes", type)) %>%
# mutate(from.role.type = type) %>%
# group_by(type) %>%
# mutate(n = paste(n, "(", cumsum(n), ")", sep = "")) %>%
# filter(!is.na(month))
# cumulative.fb.activity.all <- spread(cumulative.fb.activity.all[,c(1, 3, 4)], month, n)
# colnames(cumulative.fb.activity.all)[2:ncol(cumulative.fb.activity.all)] <- colnames(cumulative.fb.activity.role)[2:ncol(cumulative.fb.activity.role)]
# 8.3. Join monthly cumulative Facebook activity dataframes together
# cumulative.fb.activity <- bind_rows(cumulative.fb.activity.all, cumulative.fb.activity.role)
#

```

Environment History Connections

Global Environment

Data

- cumulative.fb.a 12 obs. of 9 variables
- cumulative.fb.a 3 obs. of 9 variables
- cumulative.fb.a 9 obs. of 9 variables
- data1 7 obs. of 5 variables
- data2 num [1:7, 1] 0.213 0.169 0.164 0.244 0.188
- data3 num [1:7, 1] 0.461 0.411 0.406 0.493 0.434
- differences.com 78 obs. of 4 variables
- differences.lik List of 2
- edges 4268 obs. of 4 variables
- edges.all 22688 obs. of 3 variables
- edges.all.split large list (8 elements, 1.2 Mb)
- edges.nostrum 990 obs. of 3 variables

Files Plots Packages Help Viewer

Export



[Click here to access/download](#)
Supplemental Coding Files
JOVE.R

