European Holocaust Research Infrastructure
H2020-INFRAIA-2014-2015
GA no. 654164

D14.3

Open Analysis Toolset

Tobias Blanke, Marco Braghieri
King's College London

Start: April 2016 [M12]
Due: February 2019 [M46]
Actual: June 2019 [M50]

Document Information

Project URL	www.ehri-project.eu
Document URL	https://training.ehri-project.eu/unit/8-cultural-analytics-interactive-learning-environment-r
Deliverable	14.3 Open Analysis Toolset
Work Package	WP14
Lead Beneficiary	5 - KCL
Relevant Milestones	MS 3
Dissemination level	Public
Contact Person	Tobias Blanke, King's College London tobias.blanke@kcl.ac.uk
Abstract (for dissemination)	A range of tools that can be used by the community to analyse large-scale Holocaust Dataset. This range of tools has been implemented through an online learning environment named 'Cultural Analytics: Interactive Learning Environment in R'. Each course lesson is devoted to a single topic, providing examples, exercises, self-assessment questions and references. Users will find instructions on how to run the course either on the cloud or on their personal machines, providing them with an introduction to R programming language, text analysis, spatial modelling, statistics art and visualization. Moreover, The course's goal is to provide users with a knowledge base that will allow users to apply R-based tools to datasets of interest. Within the frame of European Holocaust Research Infrastructure project, this course allows users to acquire competences that can be used to analyse real life-based datasets, as demonstrated by the Holocaust-based examples provided during the course
Management Summary	n/a

Table of Contents

1   Introduction

The Cultural Analytics Course is an eleven lesson interactive course which has been published on the EHRI Online Course in Holocaust Studies website. The EHRI Online Course in Holocaust Studies has been devised to 'provide teachers, lectures and students with source material and background information in order to give them an overview on recent trends in historiography'.[1]

The Cultural Analytics Course has been added to existing topics within the EHRI Online Course in Holocaust Studies website and it is an interactive learning environment based on Swirl, a 'software package for the R programming language that turns the R console into an interactive learning environment'.[2] The Cultural Analytics Course provides users with an introduction to the R programming language, text analysis, spatial modelling, statistics art and visualisation with a hands-on approach. Each course lesson is devoted to a single topic, providing examples, exercises, self-assessment questions and references.
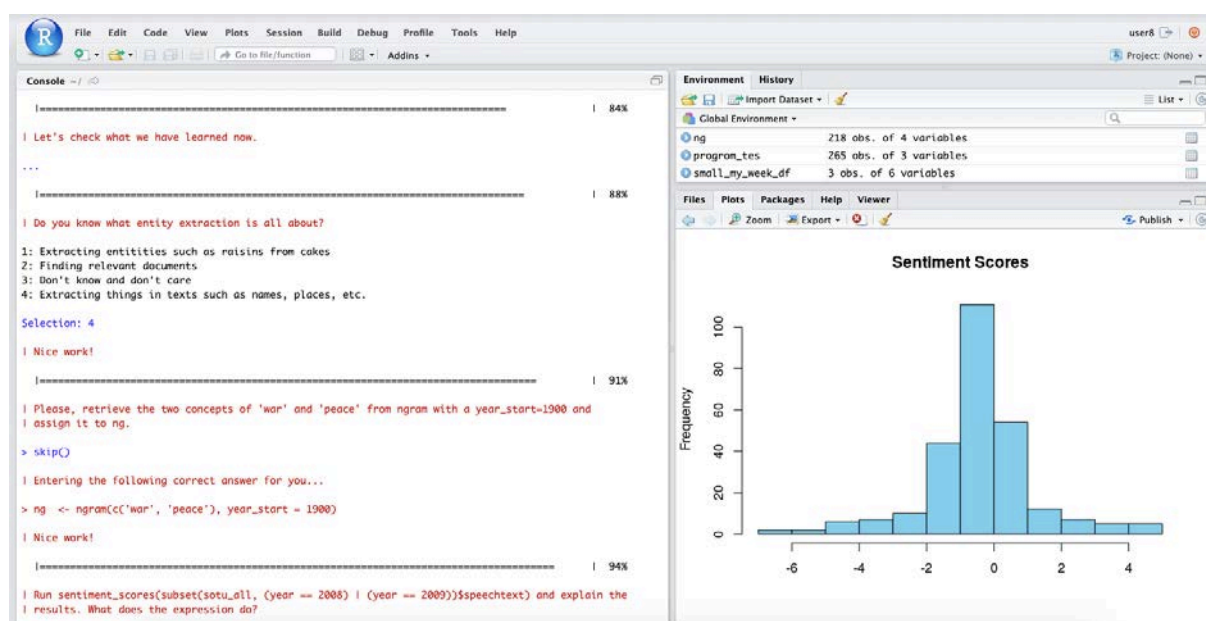

Figure 1 - A self-assessment session at the end of one of the Cultural Analytics Course lesson

The Cultural Analytics Course has been developed using Swirl, which has been designed specifically for training purposes and is based on R, 'a system for statistical computation and graphics. It consists of a language plus a run-time environment with graphics, a debugger, access to certain system functions, and the ability to run programs stored in script files'.[3] R is allows users to perform a number of functions, including 'linear and generalized linear models, nonlinear regression models, time series analysis, classical parametric and nonparametric tests, clustering and smoothing'.[4]

---

[1] Source: https://training.ehri-project.eu/overview (Accessed: 5 May 2019).

[2] Source: https://swirlstats.com/faq.html (Accessed: 5 May 2019).

[3] Source: https://cran.r-project.org/doc/FAQ/R-FAQ.html#What-is-R_003f (Accessed: 5 May 2019).

[4] Ibid.

Moreover, R is capable of providing 'a flexible graphical environment for creating various kinds of data presentations'.[5] The Cultural Analytics Course Swirl package is designed to run in RStudio, which is 'integrated development environment for R'[6] which includes 'a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management'.[7]

Thus, the Cultural Analytics Course exploits all of R's features, combining them into an interactive learning environment. This allows users to have a hands-on experience of different analysis, plotting and visualisation tools within the frame of a dedicated software package. The competences users will gain will then be applicable to real-life datasets, such as the ones developed and provided by the European Holocaust Research Infrastructure.
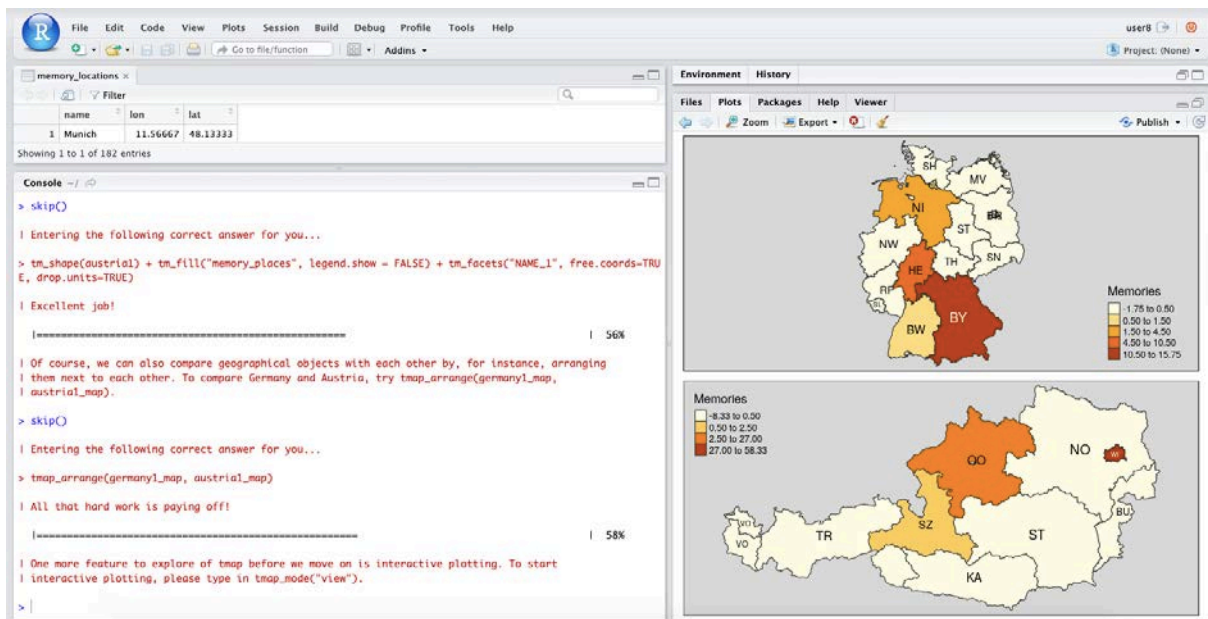

Figure 2 - A plot of 'places of memory' derived from an EHRI dataset

---

2   Cultural Analytics Course

The Cultural Analytics Course is articulated in eleven lessons, which explore the various features of the R programming language and its application within a Swirl-based environment.

Moreover, users are guided through the installation of the necessary software packages in order to execute the Cultural Analytics Course both in the cloud and on their own local machines (covering all major operating systems, such as Windows, Mac OS and Linux).

Once users have completed the install procedures, they will access an interactive environment which aims to cover all the major aspects of the R programming language and the possibilities offered by the Swirl software package. The specific focus of the online course is to provide users with a set of tools which can be deployed to analyse real-life datasets of diverse nature and to perform different actions such as text analysis, spatial modelling and visualisation practices.

Each lesson is devised as a hands-on experience, demanding direct participation from users to advance in the course and provides them with self-assessment questions in order to establish if tools and commands has been fully understood. Moreover, the Swirl-based environment allows users to run the environment as many times as necessary, thus providing continuous guidance in future applications.
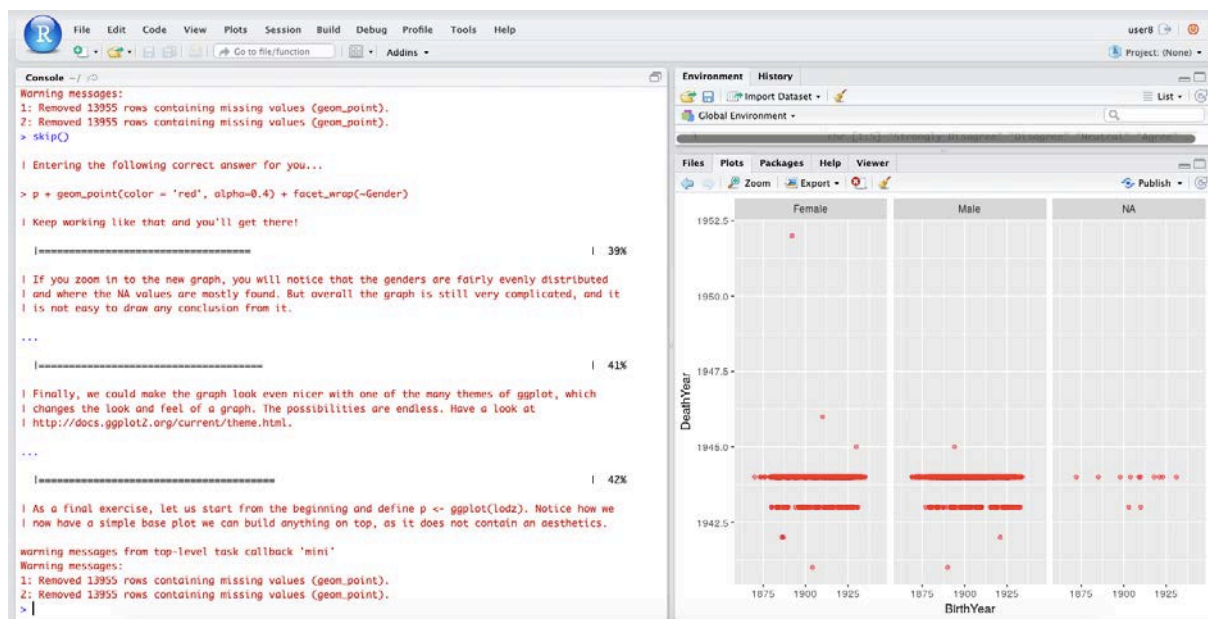

Figure 3 - Plot derived from an EHRI dataset on the Lodz Ghetto Population

Within the EHRI Online Course in Holocaust Studies [website](#), the Cultural Analytics content has been organised in accordance with previous topics and each course page focuses on a single topic, as detailed below:

- **Lesson 1** provides a first introduction to the R language and its main features such as strings, Booleans, variables, functions, vectors and plotting.
- **Lesson 2** completes the first phase of introduction to the R language by focusing on matrix and data frames function.
- **Lesson 3** focuses on advanced R constructs such as lists, tables, control structures, loops, apply functions and user-defined functions. The second part of the lesson regards real-life dataset analysis, with specific attention to the application of mean, median, summary and structure functions.
- **Lesson 4** provides an introduction to plots and data frame visualisations and their application to real-life datasets.
- **Lesson 5** focuses on different aspects and tools used in text analysis such as word count, word cloud generation, document transformation, topic modelling, frequency and association discovery
- **Lesson 6** focuses on advance technique topic modelling, providing hands-on applications on life-based datasets, including one hosted by a partner of the European Holocaust Research Infrastructure.
- **Lesson 7** introduces spatial modelling, mapping and more advanced Geographic Information System mapping, such as location clustering, point pattern analysis, quadrat counting, Choropleth mapping
- **Lesson 8** focuses on historical spatial analysis, working on a real life-based datasets centred around places of memory as intended by Nora (1996),[8] introducing tools such as spatial elaboration and interactive plotting. Finally, the lesson explores historical Geographic Information Systems and mapping.
- **Lesson 9** introduces statistics, especially inferential statistics, and which tools to implement in order to explore datasets such as visualisations (histograms, density plots, likert scale), descriptive (standard deviation, empirical cumulative distribution function) and inferential statistic techniques.
- **Lesson 10** continues to focus on statistical inference of a single value, exploring the role of p-value and z-score and then moves on to explore statistical inference of the relationship of two or more values, introducing statistical tests such as Pearson and Spearman correlation, z-test, t-test, chi-square, Analysis of Variance (ANOVA) and linear regression. The lesson finally focuses on inferential statistics and probability distribution.
- **Lesson 11** focuses on the grammar of graphs and its application to real life-based datasets, focusing on both graphic formatting and aesthetics.

Finally, each page provides users with an printout of the whole lesson, which was generated in order to provide further guidance to users.

---

[8] Nora, P. (1996) 'From lieux de mémoire to realms of memory', *Realms of memory: Rethinking the French past*. Columbia University Press New York, 1, pp. 1–21.