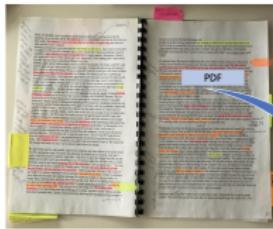


Improving the efficiency of textual analysis in R

Sophie Avard (sophie.avard@hdr.mq.edu.au)
Macquarie University

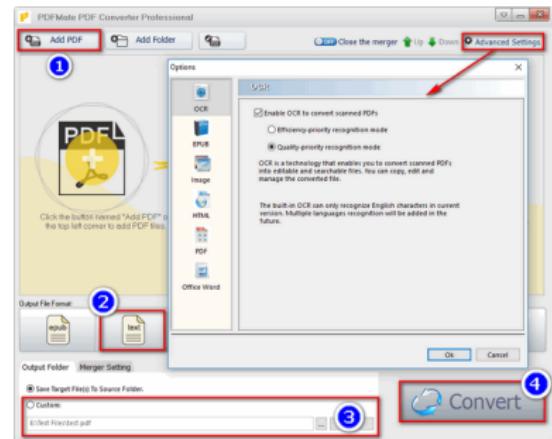


Text file

Path to folder

Path to file

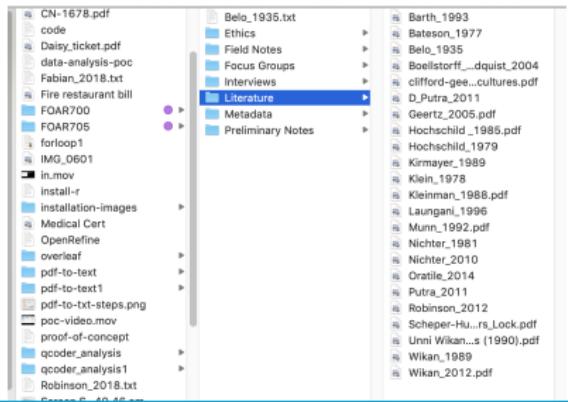
Analysis



Problem 1: Unreliable tagging.

Problem 2: Documents needed to be in a txt file format.

Problem 3: Moving the files into the correct folder was time consuming.



Improving the efficiency of textual analysis in R

Sophie Avard (sophie.avard@hdr.mq.edu.au)
Macquarie University

Solution

A lightweight and easy-to-use script to make textual analysis in R more efficient. Here's what my solution did!

The screenshot shows the RStudio interface with the following details:

- Code Editor:** Displays the `proof-of-concept.R` script. The code installs packages, reads files, and processes them using `pdftools` and `pdftoolsr` to extract text and save it to PDF and TSV files.
- Environment View:** Shows the workspace with objects like `codes_Analysts`, `new_dataframe`, `new_dataframe`, `variables`, `codes_df_path`, `datafiles`, `docs_df_path`, `file`, `file_name`, `output_file`, `output_file`, and `project_name`.
- File Explorer:** Shows the project structure with subfolders like `Analysts`, `data`, `docs`, `pdftools`, `pdftoolsr`, `proof-of-concept`, `pdftools`, and `pdftoolsr`.
- Output View:** Displays the results of the R script execution, including file paths and modification times.
- Help View:** Shows the help documentation for `pdftoolsr::read_pdf()`.

Script Content:

```
1 # proof-of-concept
2
3 install.packages("pdftools", repos = "http://cran.us.r-project.org")
4 install.packages("pdftoolsr", repos = "https://cran.us.r-project.org")
5
6 library(pdftoolsr)
7
8 File.names <- dir(path="pdfs", patterns = ".pdf", full.names=TRUE)
9
10 for(File in File.names) {
11   text <- pdf_text(file)
12   print(text)
13   pdf_file <- pdf("pdftoolsr.pdf")
14   write(text, pdf_file)
15   pdf_close(pdf_file)
16
17   output_file <- gsub(".pdf", ".txt", file)
18   print(output_file)
19   pdf_file <- pdf(output_file)
20   write(text, pdf_file)
21   pdf_close(pdf_file)
22
23   install.packages("pdftoolsr", repos = "https://cran.us.r-project.org")
24   devtools::load_all()
25   devtools::load_all()
26
27   # Natural language support but running on an English locale
28   R is a collaborative effort with many contributors.
29   Type 'contributors()' for more information.
30   'citation()' on how to cite R or R packages in publications.
31
32   Type 'demo()' for some demos, 'help()' for on-line help, or
33   'help.start()' for an HTML browser interface to help.
34   Type 'q()' to exit R.
35
36 [Workspace loaded from ~/Desktop/Sophie-Avard-Proof-of-concept]
```

Execution Results:

```
research when Walter Cannon expanded on this document is copyrighted by the American Psychological Association or one of its allied publishers. Bernard's work and identified homostasis. The contributions of Hans Selye, considered the father of stress research; Sir William Osterley Yerkes and Richard Lazarus are also discussed. Finally I discuss the work in the new millennium, research on psychological stress has expanded across disciplines ranging from physiology to medicine, chemistry, endocrinology, neuroscience, epidemiology, psychiatry, ergonomics, and psychology reflecting the complex and dynamic nature of stress. I present the results of a systematic review of the literature on stress and health, homoeostasis. Stress in health and disease is medically sociological and phisi-discipline. Having conducted stress research over the past decade specifically the most meaningful subject for humanity that I can think range from the basic sciences using the laboratory rat to epideidemiology studies with thousands of participants, I have found it - Hans Selye (1907-1972) talk about all or even study stress may use the same language but do not ice to stress is ubiquitous in modern society, and the operational definition of stress is a recurrent descriptor of negative experiences diversity of stress research related to anything from daily hassles, relationship issues, and a lack a historical stress pressures at work to health concerns and debilitating phobias. It is developed, swing article is interesting that most popular definitions would likely describe this as individual user and is not to be disseminated broadly meant to provide idents of purely psychological phenomena, yet less than 100 years ago, the health establishment was dominated by a particular type of medical model of health. Today concept of stress research, the concept of stress is pervasive in popular culture and the major developments in stress literature. Despite its prevalence, stress remains a late 19th century through to the most recent cent developments. In the process to Modern stress research was birthed out of divergent interests in the field, I developed two medicine and physiology what was introduced as a novel concept of the concept of concept in psychology. (Stress,Modernity) Stress is one physiological and the other psychological - that would include the changes in how it eventually converge post -World War II. Finally, postwar development and extraordinary conditions of stress research are described, as well as the gaps what is missing in both the popular and in the field, scientific discussion of
```

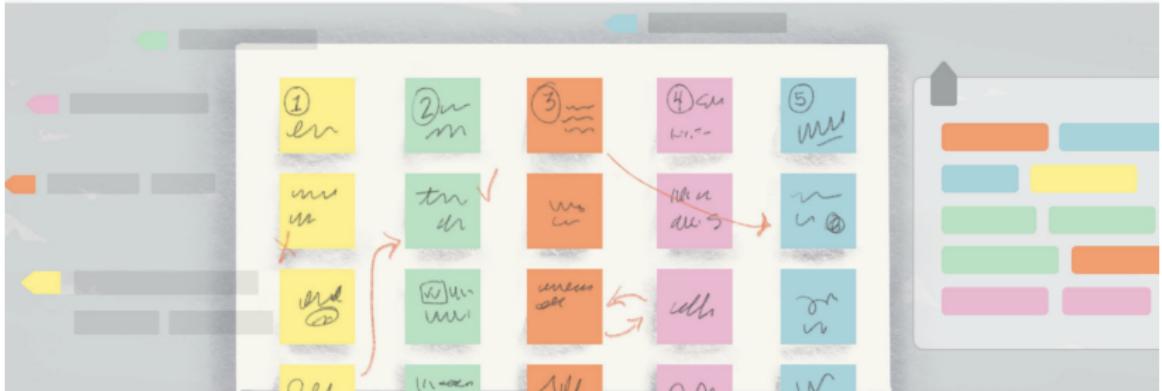
My proof of concept:

1. converts pdf's to txt files
2. Builds a path to the txt files
3. Imports the data into qcoder
4. Runs qcoder Shiny app for textual analysis

Introduction

Providing adequate mental healthcare in non-Western low-resource settings requires understanding culturally salient expressions of mental suffering and how such expressions manifest in various contexts. While translating DSM or ICD-based screening tools to reflect local languages provides a modicum of cultural representation, this etic approach often fails to capture the unique expressions of mental suffering that occur within non-Western cultures (Kaiser et al. 2015b; Kleinman 1987). In Haiti, for instance, locally-developed screening tools that incorporated idioms of distress were associated with better specificity and cultural equivalence than screening tools that were directly translated from English to Haitian Kreyòl (Kaiser et al. 2013). The problematic imposition of Western conceptualizations of mental illness on non-Western cultures perpetuates paternalist and neo-colonial practices that exclude local ideologies at the expense of those suffering from mental distress. A growing body of research acknowledges the subjectivity of the presentation and conceptualization of mental ill-health and calls for emic research that understands non-Western explanatory models and forms of expression of mental distress. From a conceptual perspective (Summerfield 2008; Kaiser et al. 2014; Weaver 2017), Incorporating local idioms of mental distress into mental healthcare and psychiatric treatment has become an increasingly well-recognized way of prioritizing local perspectives and grassroots mental health research, as well as improving the quality of care for those suffering from mental distress.

In many non-Western settings, the terms "anxiety" and "depression" have no direct equivalences in local languages and are not well recognized by people who have had little exposure to Western medical training or formalized education (Summerfield 2008). This lack of equivalence can lead to biomedical models oversimplifying the nuances of language that describe the breadth of experiences associated with mental suffering and potentially fail to identify many who could benefit from treatment (Summerfield 2012). Psychiatry, which unlike other medical specialties, is based on disease rather than distress, can also differ between cultures rather than on biomarkers or lab tests, has the potential to provide an avenue for more culturally resonant clinical encounters through understanding



Significance

- (a) Able to manage large amounts of data
- (b) Greater flexibility and thoroughness - add codes, define codes, search data.
- (c) Reduces the time spent converting and moving files
- (d) Supports other data analysis packages in R

Improving the efficiency of textual analysis in R

Sophie Avard (sophie.avard@hdr.mq.edu.au)
Macquarie University

```
install.packages('plyr', repos = "http://cran.us.r-project.org")
install.packages('pdftools', repos = "http://cran.us.r-project.org")

library(pdftools)

file.names <- dir(path="pdfs", pattern = ".pdf", full.names=TRUE)

for (file in file.names) {
  text <- pdf_text(file)

  output.file <- gsub("pdf", "txt", file)
  print(output.file)
  print(file)
  write(text, output.file)
}

#create path to directory
#create path to txt files within directory
#copy txt files into qcoder_analysis directory
rawPath <- "/Users/sophieavard/Desktop/proof-of-concept/txts"
datafiles <- dir(rawPath, "*.txt", ignore.case = TRUE, all.files = TRUE)
file.copy(file.path(rawPath, datafiles), file_path, overwrite = TRUE)
```

```
install.packages("devtools", repos = "http://cran.us.r-project.org")
devtools::install_github("ropensci/qcoder")

library(shiny)
library(qcoder)
#create_qcoder_project("qcoder-analysis-project")

import_project_data(project = "qcoder-analysis-project")

project_name = "qcoder-analysis-project"
file_name <- "Fabian_2018.txt"
docs_df_path <- "qcoder-analysis-project/data_frames/qcoder_documents_qcoder-analysis-project"
codes_df_path <- "qcoder-analysis-project/data_frames/qcoder_codes_qcoder-analysis-project.r"
file_path <- "qcoder-analysis-project/documents"
dir("qcoder-analysis-project/documents")
```

```
new_dataframe <- readRDS(docs_df_path)

read_code_data(project_name = project_name)
codes_dataframe <- readRDS(codes_df_path)

qcode(use_md=TRUE)
```

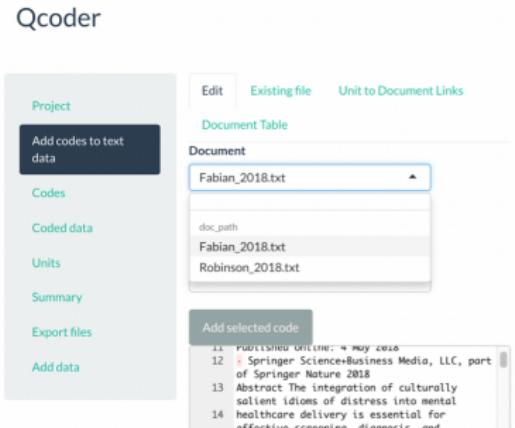
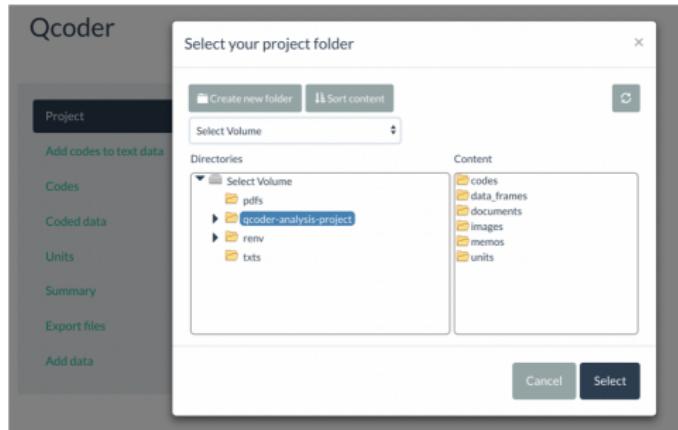
more ...

Improving the efficiency of textual analysis in R

Sophie Avard (sophie.avard@hdr.mq.edu.au)
Macquarie University

... back to figure

If the script is successful, the qcoder app should automatically open on the users computer. The user then should be able to open their project directory and begin tagging their data.



Improving the efficiency of textual analysis in R

Sophie Avard (sophie.avard@hdr.mq.edu.au)
Macquarie University

research when Walter Cannon expanded on this document is copyrighted by the American Psychological Association or one of its allied publishers. Bernhard's work and identified homoeostasis. The contributions of Hans Selye, considered the father of stress research; Sir William Oster; Yerkes and Dodson; and Richard Lazarus are also discussed. Finally I discuss how, in the new millennium, research on psychological stress has expanded across disciplines ranging from physiology to medicine, chemistry, endocrinology, neurosciences, epidemiology, psychiatry, epigenetics, and psychology, reflecting the complexity of the construct both theoretically and biologically. (Fretress) Keywords: stress, homoeostasis, history, health, hormones. Stress in health and disease is medically, sociologically, and philo-discipline. Having conducted stress research over the past decade I opine the most meaningful subject for humanity that I can think ranging from the basic sciences using the laboratory rat to episode of biology studies with thousands of participants, I have found it – Hans Selye (1907–1982) increasingly obvious that those who talk about or even study stress may use the same language but do not necessarily share the same. That reference to stress is ubiquitous in modern society, and the operational definition of stress. This may be due in part to the term stressfulness is a recurrent descriptor of negative experience. Diversity of stress research and application has been increasing rapidly in the last few decades. In this article I will highlight four areas that have increased our understanding of how the concept of stress pressures at work to health concerns and debilitating phobias. It is developed. To address this knowledge gap, the following article is interesting that most popular definitions would likely describe. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly. Meant to provide a brief history of stress research for students of purely psychological phenomena, yet less than 100 years ago, the health psychology, those new in the field of stress research, and term stress as a psychological phenomenon did not exist. Today those who are curious about the evolution of stress research, the concept of stress is pervasive in popular as well academic. The following article outlines the major developments in stress literature. Despite its prevalence, stress remains an elusive con-research beginning in the late 19th century through to the most recent con-re research developments. In the process of tracing the historical events that led to Modern stress research was birthed out of divergent interests in the subdiscipline of stress research in psychology, I developed two medicine and physiology before it was introduced as a novel distinct orientations regarding the development of the concept of concept in psychology. #stress, #modernity#Stress research eventually evolved to-in-stress—one physiological and the other psychological—that would elude the consideration of cognitive processes that influence how eventually converge post-World War II. Finally, postwar developments in individual responses to both ordinary and extraordinary conditions of stress research are described, as well as the current advancements of daily life. Perhaps what is missing in both the popular and in the field, scientific discussion of

You must open the application in browser to download data.

Code_id	code	code-description
1	stress	defines or describes stress
2	women	mentions women specifically
3	women's group	mentions women groups or organisations
4	emotions	discusses or refers to any emotion
5	idioms of distress	discusses or makes reference to an idiom of distress
6	Interesting	interesting point to come back to later
7	gender	mentions gender
8	modernity	mentions modernity or capitalism
9	temporality	mentions or makes reference to temporality
10	idioms	

Showing 1 to 10 of 12 entries

Previous 1 2 Next

Limitations

- ▶ Cannot add notes or memos to the data during analysis
- ▶ Due to a bug in the qcoder package, data must be coded and downloaded in one session
- ▶ Cannot export the highlights from qcoder



Proof of concept repository:

<https://github.com/MQ-FOAR705/Sophie-Avard-Proof-of-Concept>