ABCDH
https://abcdh.eu/

Distant Reading in R

Intro to Distant Reading and R

Simone Rebora & Giovanni Pietro Vitali
simone.rebora@univr.it giovannipietrovitali@gmail.com

OVERVIEW

Two-week workshop: first week analysis (with Simone); second week visualization (with Giovanni)
An overview workshop: (almost) a new subject each day
With one constant: the R programming language

Simone's week

- a bit of theory, a lot of practice
- hands-on sessions each day (self-help, discussion, experimentation...)

Workshop results (on Saturday evening)

- 5-minutes presentation: something nice and funny
- we will give you (as much as possible) time to think about ideas

DISTANT READING

Conjectures on World Literature Moretti, Franco New Left Review; Jan 1, 2000; 1, ProQuest pg. 54

FRANCO MORETTI

CONJECTURES ON WORLD LITERATURE

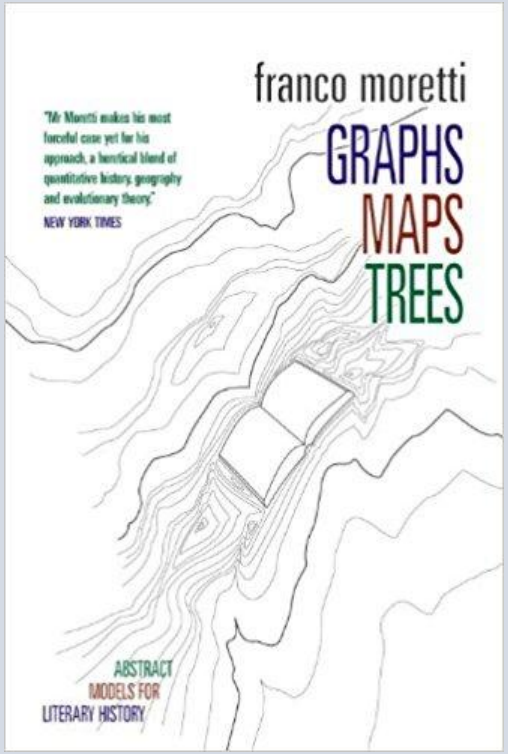
DISTANT READING

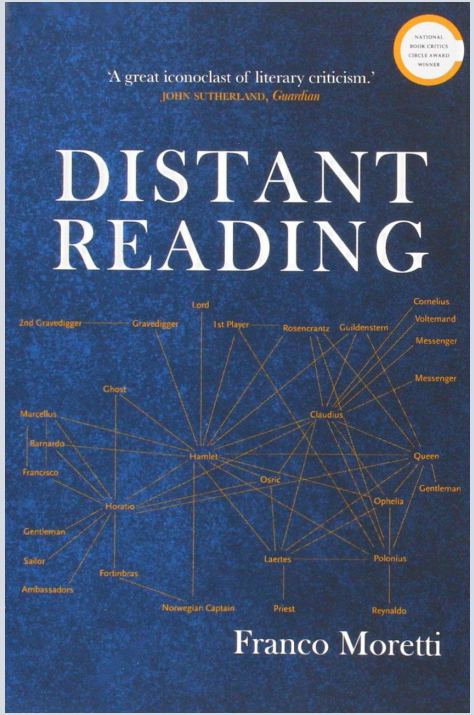
"[…] literary history will quickly become very different from what it is now: it will become 'second hand': a patchwork of other people's research, *without a single direct textual reading*. Still ambitious, and actually even more so than before (world literature!); but the ambition is now directly proportional *to the distance from the text*: the more ambitious the project, the greater must the distance be"
(Moretti, 2000)

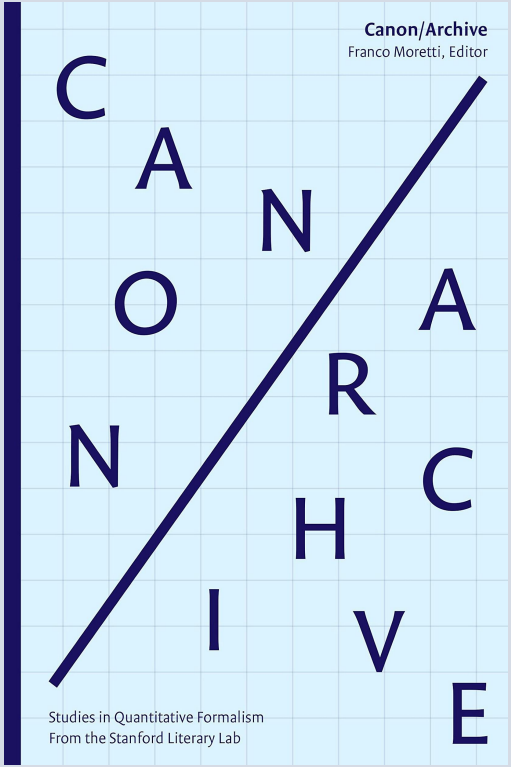
DISTANT READING

“[…] a canon of two hundred novels, for example, sounds very large for nineteenth-century Britain (and is much larger than the current one), but is still less than one per cent of the novels that were actually published: twenty thousand, thirty, more, no one really knows—and close reading won’t help here, a novel a day every day of the year would take a century or so”
(Moretti, 2005)

DISTANT READING



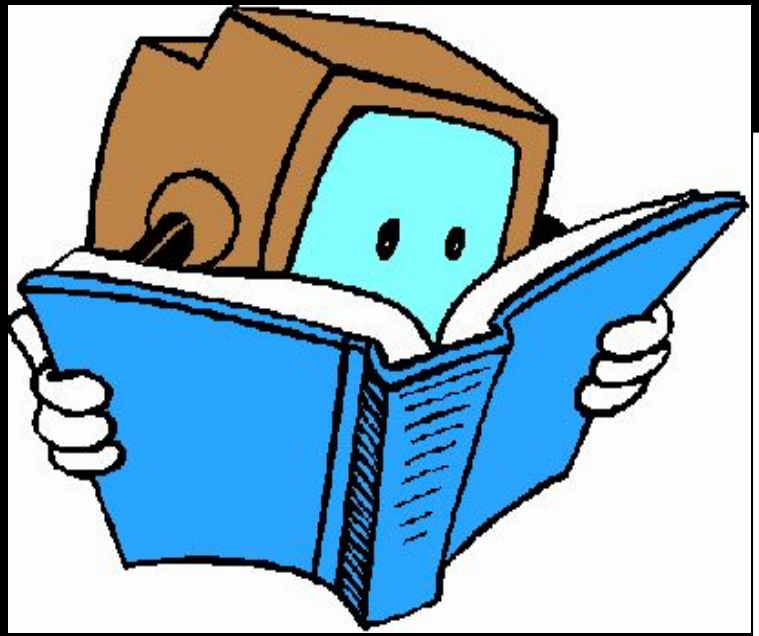


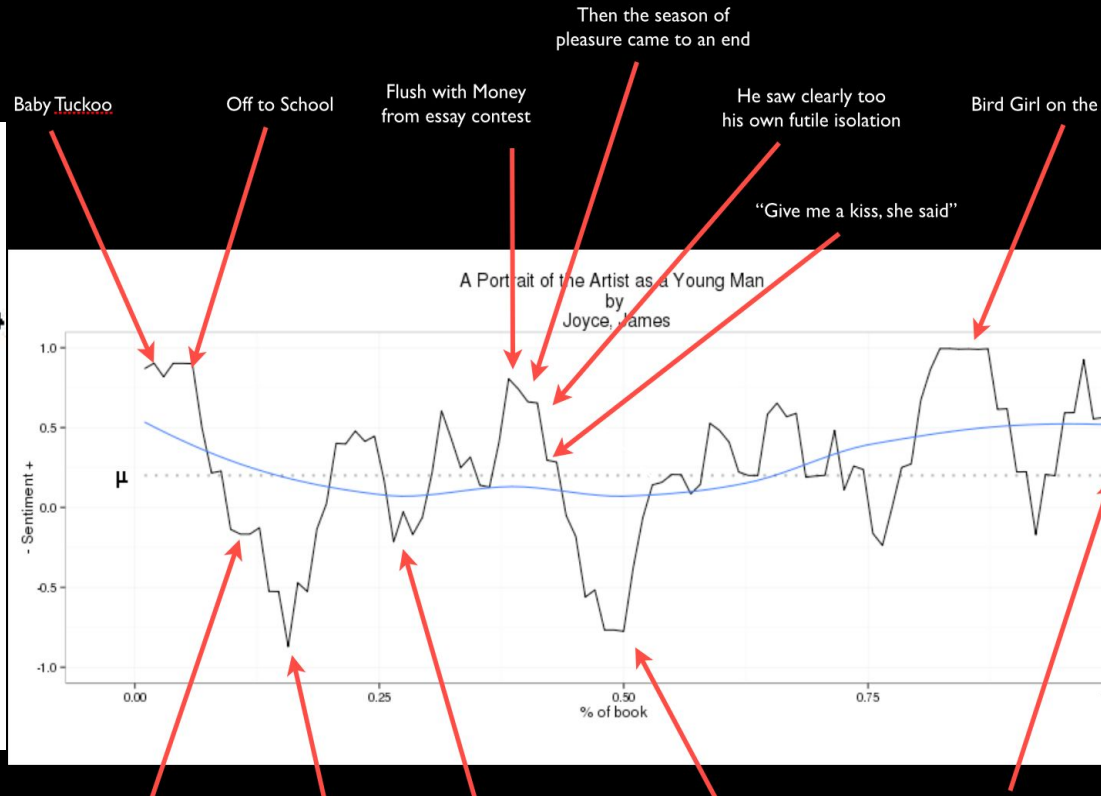


"Instead of reading texts in the traditional way – so-called close reading –, he invites to count, to graph and to map or, in other words, to visualize them" **(Jänicke et al., 2015)**

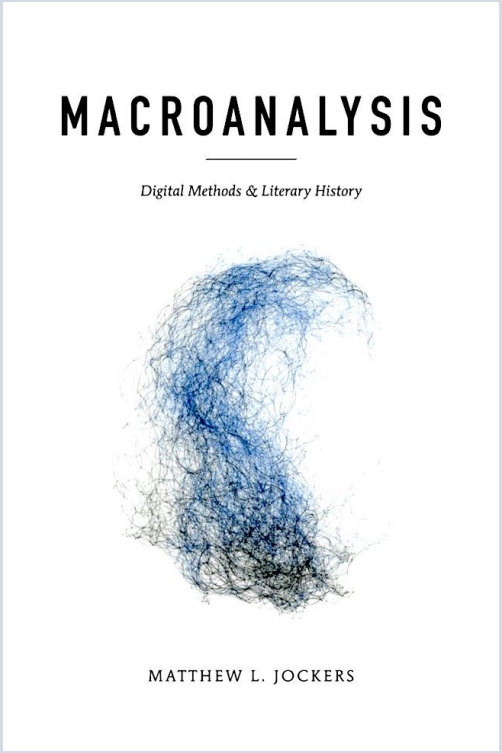
Distant Reading in R

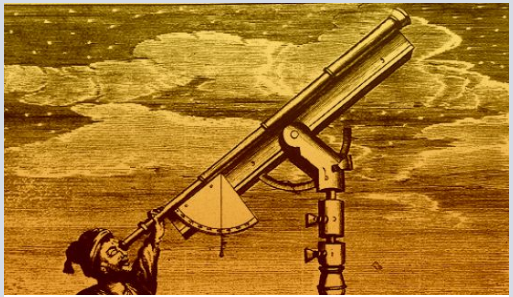
DISTANT READING





DISTANT READING OR COMPUTATIONAL LITERARY STUDIES?





Multiple definitions for (almost) the same concept

The **Stanford Literary Lab** is a research collective that applies ***Computational Criticism*** to the study of literature. The Lab is open to students and faculty at Stanford University, California, and to those from other institutions all over the world. Computational criticism is defined as applying quantitative methods for the interpretation of literature and the Lab's projects range from individual and group publications, lectures, courses, panels, and conferences where the research including Pamphlets, Projects and Technical blogs is posted onto their website. The Stanford Lab was co-founded by Franco Moretti and Matthew Jockers, whose work has been influential in developing 'Digital Humanities' into an established movement.

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Cite this page
Wikidata item

Print/export
Download as PDF
Printable version

Distant reading

From Wikipedia, the free encyclopedia

This article **may be too technical for most readers to understand**. Please help improve it to make it understandable to non-experts, without removing the technical details. *(May 2020)* *(Learn how and when to remove this template message)*

Distant reading is an approach in literary studies that applies computational methods to literary data, usually derived from large digital libraries, for the purposes of literary history and theory. While the term is collective, and is used to refer to a range of different computational methods of analysing literary data, similar approaches also include macroanalysis, cultural analytics, computational formalism, computational literary studies, quantitative literary studies, and algorithmic literary criticism.

Contents [hide]

1 History
2 Principles and practice
3 Criticisms of distant reading
4 Examples
5 See also
6 References

History [edit]

The term "distant reading" is generally attributed to Franco Moretti and his 2000 article, *Conjectures on World Literature*.[1] In the article, Moretti proposed a mode of reading which included works outside of established literary canons, which he variously termed "the great unread"[2] and, elsewhere, "the Slaughterhouse of Literature".[3] The innovation it proposed, as far as literary studies was concerned, was that the method employed samples, statistics, paratexts, and other features not often considered within the ambit of literary analysis. Moretti also established a direct opposition to the theory and methods of close reading: "One thing for sure: it cannot mean the very close reading of very few texts—

About

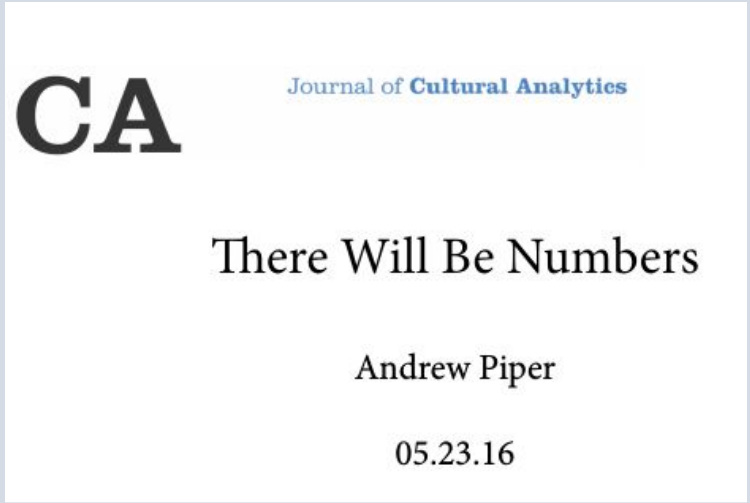
Distant Reading for European Literary History (COST Action CA16204) is a project aiming to create a vibrant and diverse network of researchers jointly developing the resources and methods necessary to change the way European literary history is written. Grounded in the Distant Reading paradigm (i.e. using computational methods of analysis for large collections of literary texts), the Action will create a shared theoretical and practical framework to enable innovative, sophisticated, data-driven, computational methods of literary text analysis across at least 10 European languages. Fostering insight into cross-national, large-scale patterns and evolutions across European literary traditions, the Action will facilitate the creation of a broader, more inclusive and better-grounded account of European literary history and cultural identity. To accomplish this, the Action will:

1. build a multilingual European Literary Text Collection (ELTeC), ultimately containing around 2,500 full-text novels in at least 10 different languages, permitting to test methods and compare results across national traditions (more);
2. establish and share best practices and develop innovative computational methods of text analysis adapted to Europe's multilingual literary traditions (more);
3. consider the consequences of such resources and methods for rethinking fundamental concepts in literary theory and history (more).

The Action will contribute to the development and distribution of methods, competencies, data, best practices, standards and tools relevant to Distant Reading research. This will not only affect the way scholars in the Humanities do research, but also the way institutions like libraries will make their holdings available to researchers in the future. The Action will foster distributed research, the systematic exchange of expertise, and the visibility of all participants, activities and resources.

If you'd like to find out more about you can join the Action's activities, see our page How to join! We also have pages with more information about the management structure of the Action as well as an interactive map showing where our Action members are located.

ADVANTAGES OF DISTANT READING



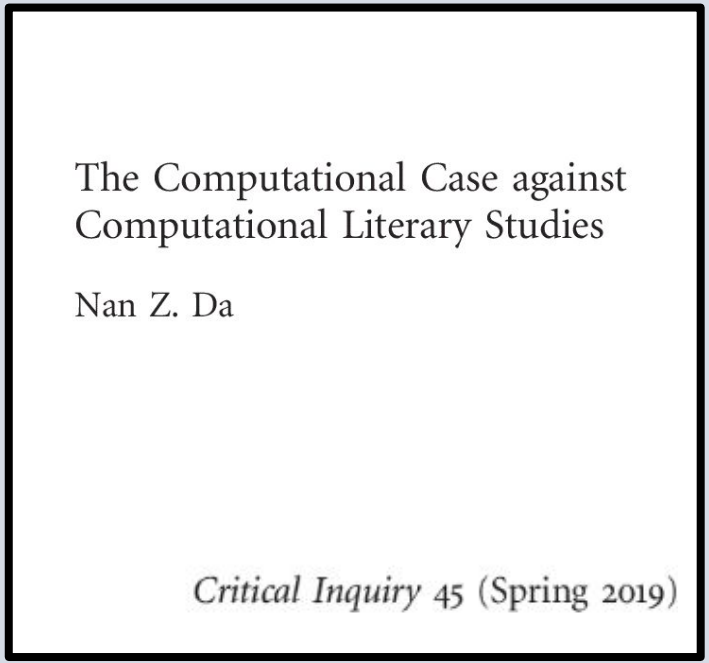
The «double revolution» of computational methods in literary studies:

- From an «impressionistic» (and frequently opportunistic) choice of case studies
- To their «representativeness»

- From an «agonistic» criticism
- To a «consensus-driven» criticism

CRITICAL ASPECTS (PRACTICAL)

- Da took 14 representative computational studies (by authors such as Underwood and Jockers)
- …and she repeated their analyses, showing errors and limitations
- but caution: this was just a “cherry picking”!

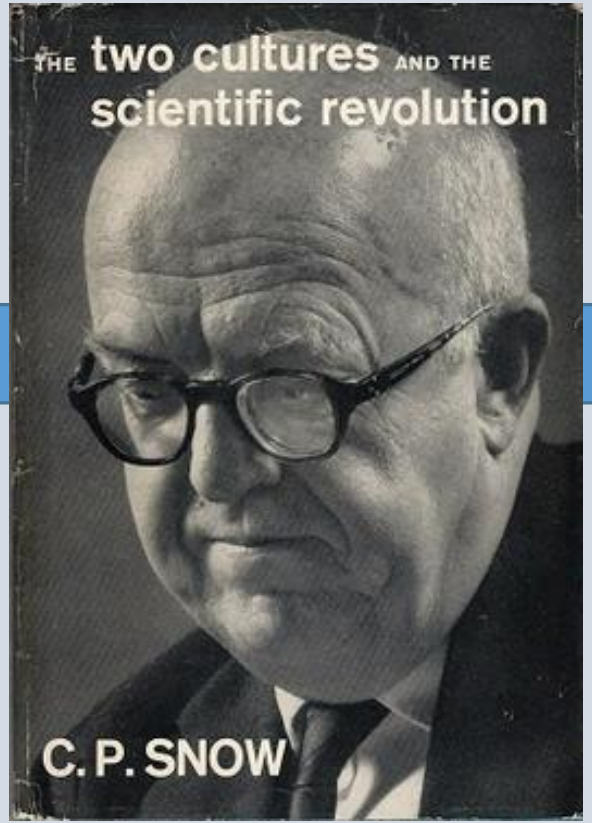


CRITICAL ASPECTS (IDEOLOGICAL)

«For the moment, a far more concrete risk than the improbable government of machines seems to concern the rise to power of people in the flesh perhaps less suitable than many machines to reason with the complexity and foresight that certain subjects would require. That is, machine designers» (Tomasin, 2017)

THE TWO CULTURES

Science

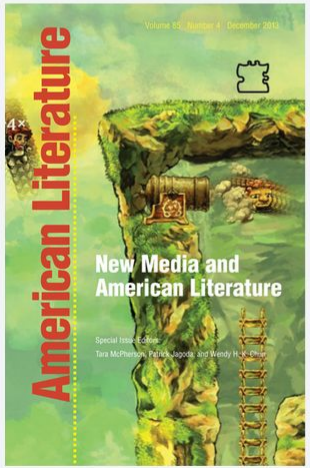


Humanities

A POSSIBLE SOLUTION?

«[Science and Humanities] can both be practiced at the same time, as two incommensurable positions, in an irresolvable yet productive tension, so that the questions, issues and approaches specific to each are capable of generating new findings, insights and realisations in the other - to the point where both of their identities are brought into question» (Hall, 2013)

Volume 85, Issue 4
1 December 2013



RESEARCH ARTICLE | DECEMBER 01 2013

Toward a Postdigital Humanities: Cultural Analytics and the Computational Turn to Data-Driven Scholarship

Gary Hall

American Literature (2013) 85 (4): 781-809.
https://doi.org/10.1215/00029831-2367337

Cite Share Permissions

What forms will literary and cultural criticism take in the twenty-first century, given the move toward open access, open data, and open government that is currently being promoted in the name of greater

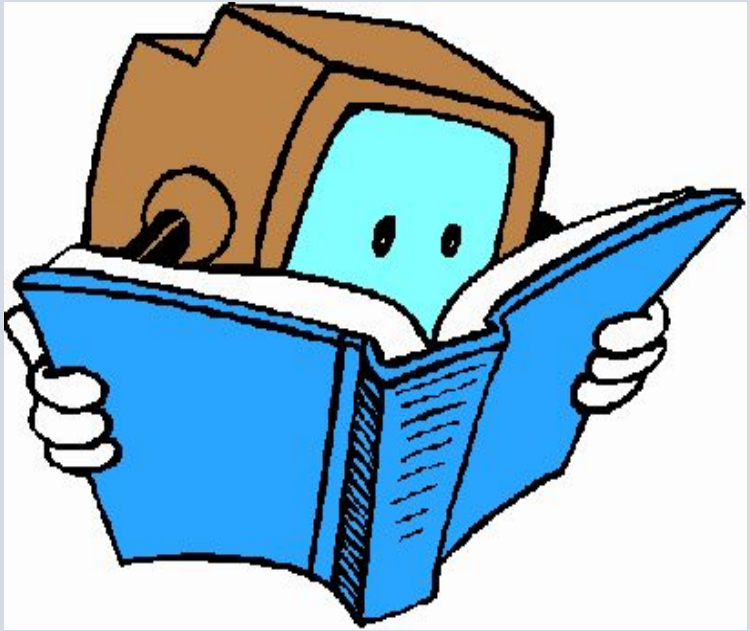
WHY USE R?

Graphical interfaces

- https://voyant-tools.org/

VS.

- Programming languages



WHY USE R?

Graphical interfaces

- https://voyant-tools.org/

VS.

- Programming languages

```
rsimone@rsimone-Inspiron-13-5378:~$ R

R version 3.4.4 (2018-03-15) -- "Someone to Lean On"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> a <- 1:10
> a
 [1]  1  2  3  4  5  6  7  8  9 10
> for(i in 1:10)
+ print(i:10)
 [1]  1  2  3  4  5  6  7  8  9 10
[1]  2  3  4  5  6  7  8  9 10
[1]  3  4  5  6  7  8  9 10
[1]  4  5  6  7  8  9 10
[1]  5  6  7  8  9 10
[1]  6  7  8  9 10
[1]  7  8  9 10
[1]  8  9 10
[1]  9 10
[1] 10
>
```

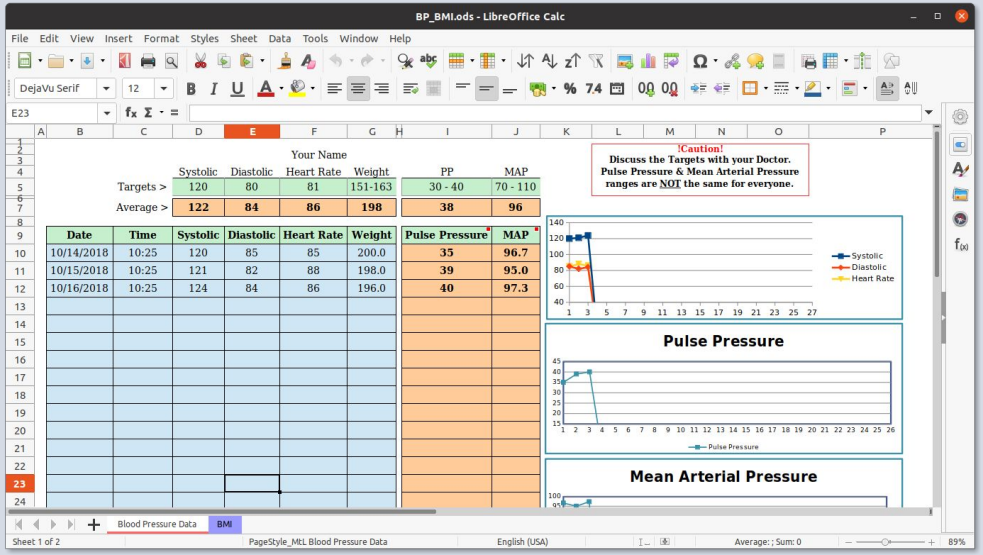
WHY USE R?

Spreadsheets

- https://www.libreoffice.org/discover/calc/

VS.

- Programming languages



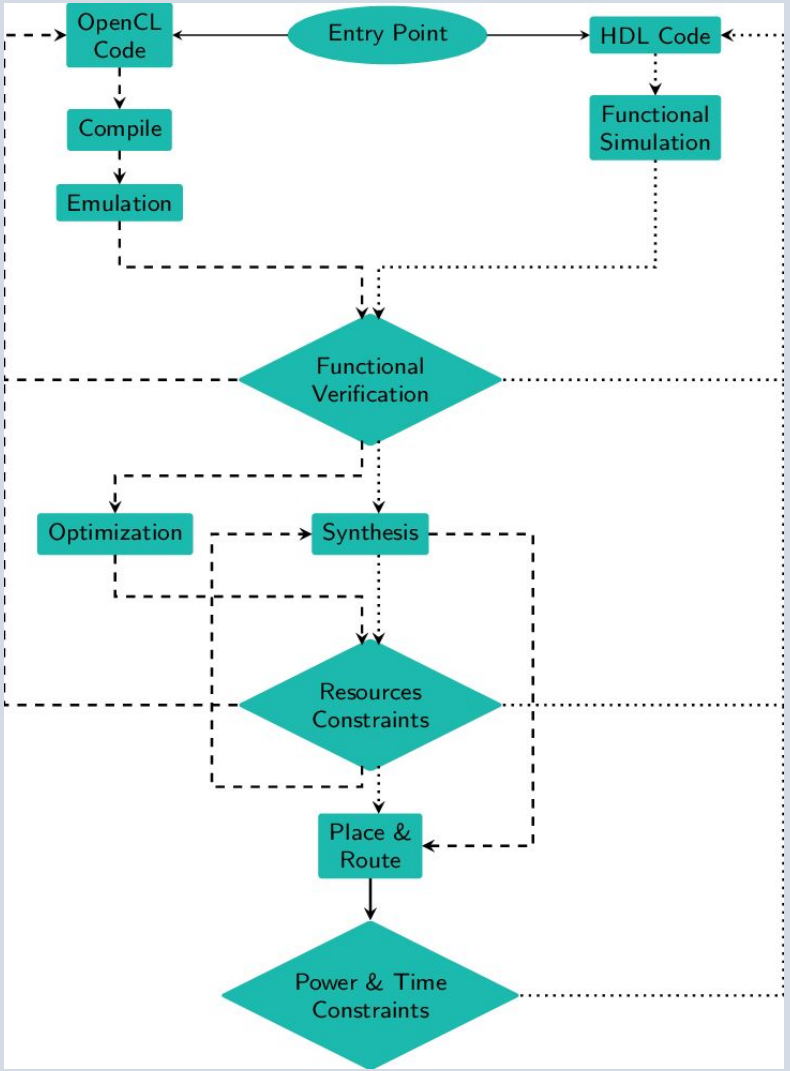
WHY USE R?

Spreadsheets

- https://www.libreoffice.org/discover/calc/

VS.

- Programming languages



WHY USE R?

Spreadsheets

- https://www.libreoffice.org/discover/calc/

VS.

- Programming languages

```
#### Intro Notes & Comments ####~
invisible( expression( `~
{ <= put the brace here to reset the auto indenting...~
  ~
  Base <- function()~
  {       <^~~~~~~~~~~~~~~~ Use the function as a header and nesting marker for the comments~
        that show up in the jump-menu.~
        --->8----~
  }~
  ~
  External <- function()~
  {~
    If we used a function similar to:~
      api_idea <- function() {~
        ~
        some_api_example <- function( nested ) {~
          stopifnot( some required check here )~
        }~
        ~
        print("Cut and paste this into RStudio to see the code-chunk quick-jump structure.")~
        return converted object~
      }~
    ~
    #### Code. ####~
    ^~~~~~~~~~~~~~~~~~~~~~~~~~ <= Notice that this comment section isnt in the jump menu!~
                                  Putting an apostrophe in isn't causes RStudio to parse as text~
                                  and needs to be matched prior to nested structure working again.~
    api_idea2 <- function() {~
    ~
    } # That isn't in the jump-menu, but the one below is...~
    ~
    api_idea3 <- function() {~
    ~
    }~
    ~
  }~
    ~
    # Just to show api_idea isn't really there...~
    print( api_idea )~
    }') )~
####~
~
#### Code. ####~
foo <- function() {~
  print( "The above did not evaluate and cause an error!")~
}~
~
foo()~
```

Intro Notes & Comments
- Base
- External
  - api_idea
    - some_api_example
  - api_idea3

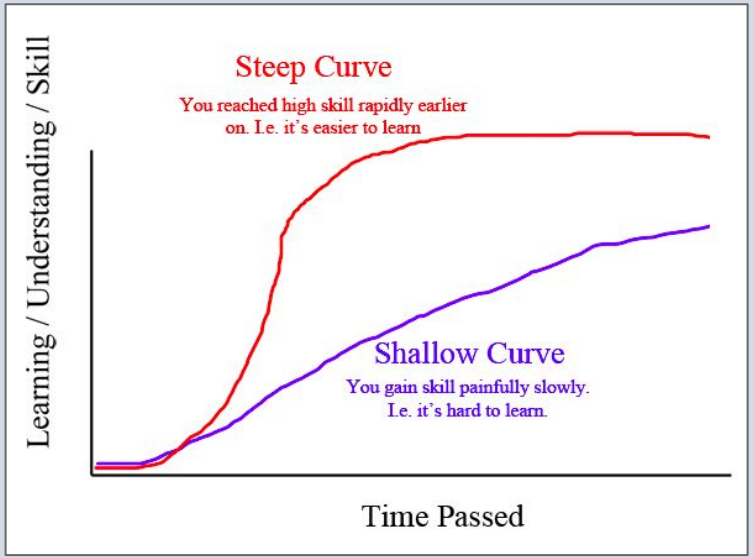
Code.
- foo

WHY USE R?

Steep learning curve!

…but then…

- full freedom in doing what you like
- possibility to fully document complex workflows



WHY USE R?

R VS.Python

…and many others!

It’s a matter of community

…more than of intrinsic qualities

GITHUB

…talking about community…

All materials for this Workshop are available via GitHub (the main online hub for programming)

Link:
https://github.com/ABC-DH/Distant_Reading_in_R

THE “R” PROGRAMMING LANGUAGE

- A (high-level) programming language for statistics and graphics
- A free software (licence GNU GPL)
- Created in 1995 by
  - Ross Ihaka (New Zealand)
  - and Robert Gentleman (Canada)
- The name? See the initials of the creators

R “RAW”

```
rsimone@rsimone-Inspiron-13-5378:~$ R

R version 3.4.4 (2018-03-15) -- "Someone to Lean On"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> a <- 1:10
> a
 [1]  1  2  3  4  5  6  7  8  9 10
> for(i in 1:10)
+ print(i:10)
 [1]  1  2  3  4  5  6  7  8  9 10
[1]  2  3  4  5  6  7  8  9 10
[1]  3  4  5  6  7  8  9 10
[1]  4  5  6  7  8  9 10
[1]  5  6  7  8  9 10
[1]  6  7  8  9 10
[1]  7  8  9 10
[1]  8  9 10
[1]  9 10
[1] 10
>
```

RSTUDIO: A GRAPHICAL INTERFACE FOR R

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

Source

Console ~/

```
R version 3.4.4 (2018-03-15) -- "Someone to Lean On"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> a <- 1:10
> a
 [1]  1  2  3  4  5  6  7  8  9 10
>
```

Environment | **History**

Import Dataset | List

Global Environment

Values

a int [1:10] 1 2 3 4 5 6 7 8 9 10

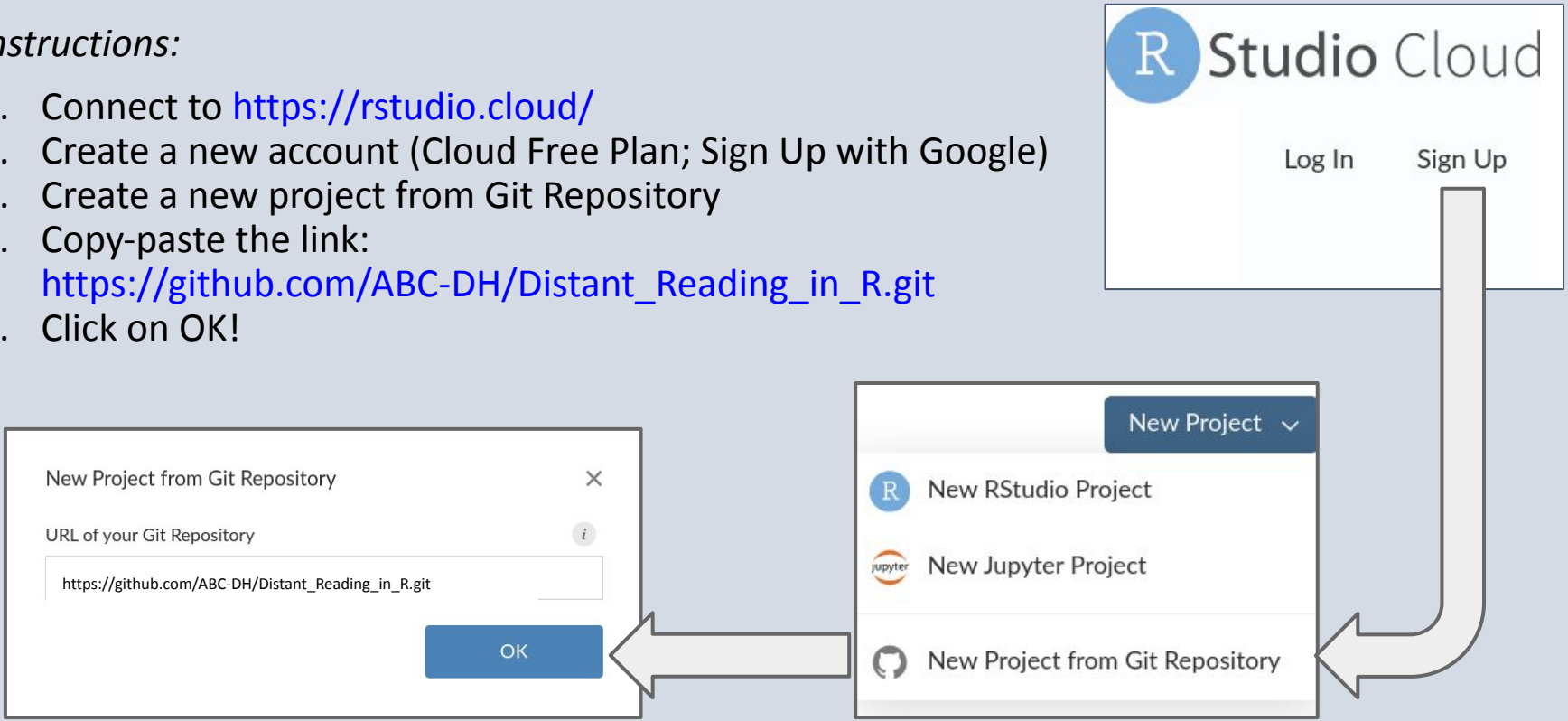
Files | **Plots** | **Packages** | **Help** | **Viewer**

New Folder | Delete | Rename | More

Home

Name	Size	Modified
.Rhistory	21.6 KB	Oct 24, 2018, 4:2
anaconda3		
AnacondaProjects		
bin		
cache		
Calibre Library		
Desktop		
dh2018-word-vector-workshops		
diffDirs	13.2 KB	Jun 13, 2018, 12:

RSTUDIO CLOUD

Instructions:

1. Connect to https://rstudio.cloud/
2. Create a new account (Cloud Free Plan; Sign Up with Google)
3. Create a new project from Git Repository
4. Copy-paste the link: https://github.com/ABC-DH/Distant_Reading_in_R.git
5. Click on OK!