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, in breakout rooms (self-help, discussion, experimentation...)

Workshop results (on Saturday evening)

- use the hands-on sessions to think about ideas for presentation
- at the end of Friday and Saturday's sessions, we will give you some time to discuss and prepare them

DISTANT READING

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

FRANCO MORETTI

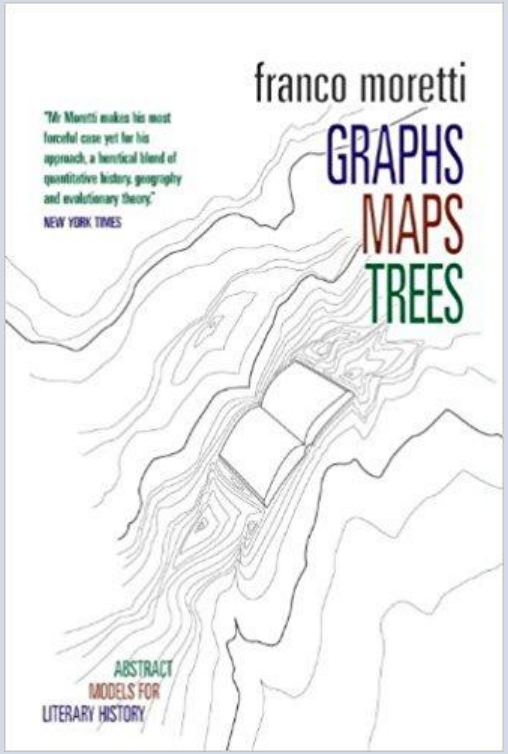
CONJECTURES ON WORLD LITERATURE

DISTANT READING

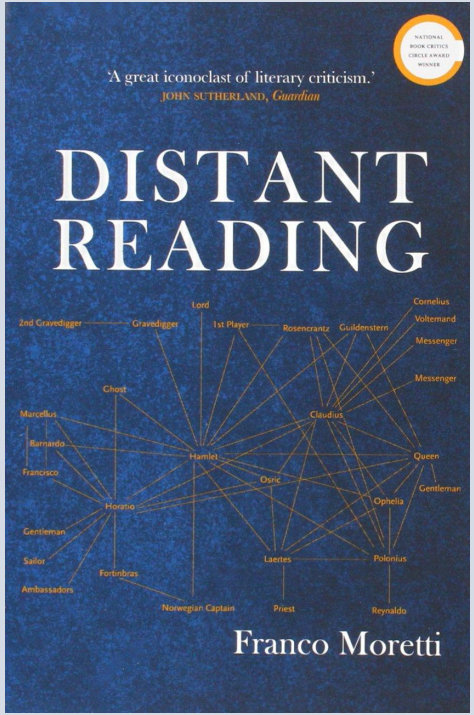
"[…] literary history will quickly become very different from what it is now: it will become 'second hand': a patchwork of other people's research, *without a single direct textual reading*. Still ambitious, and actually even more so than before (world literature!); but the ambition is now directly proportional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

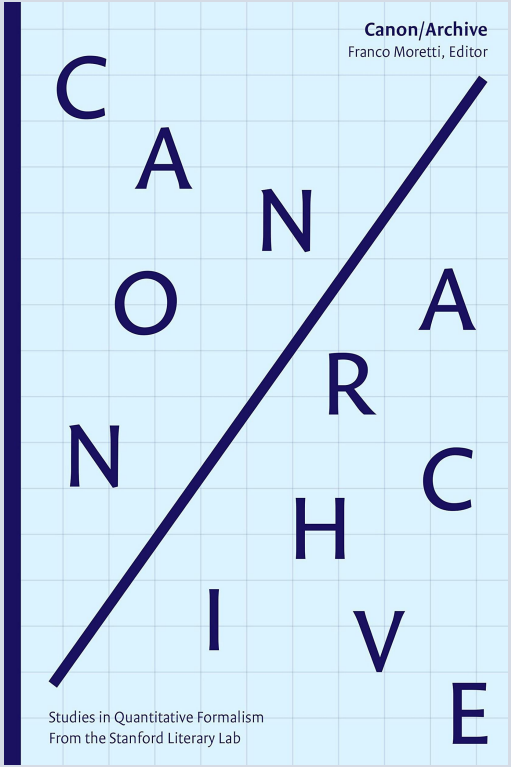
DISTANT READING

“[…] 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

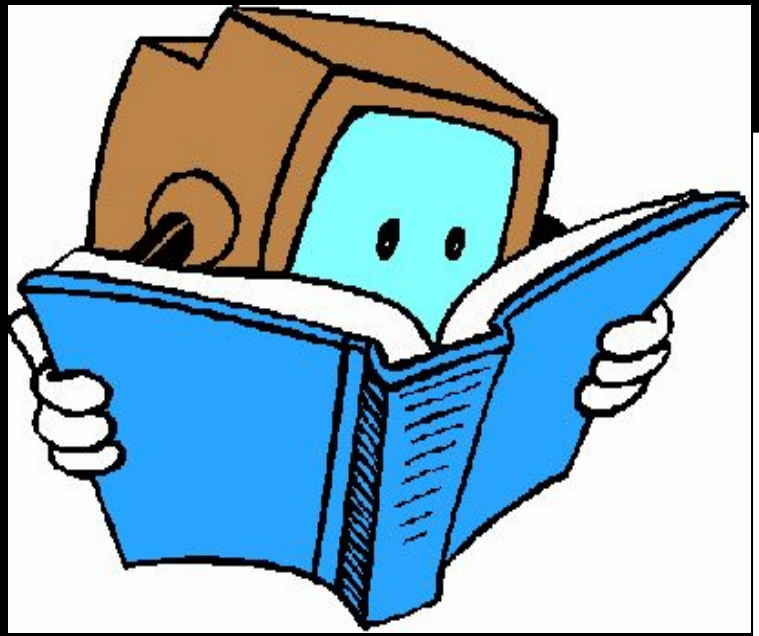


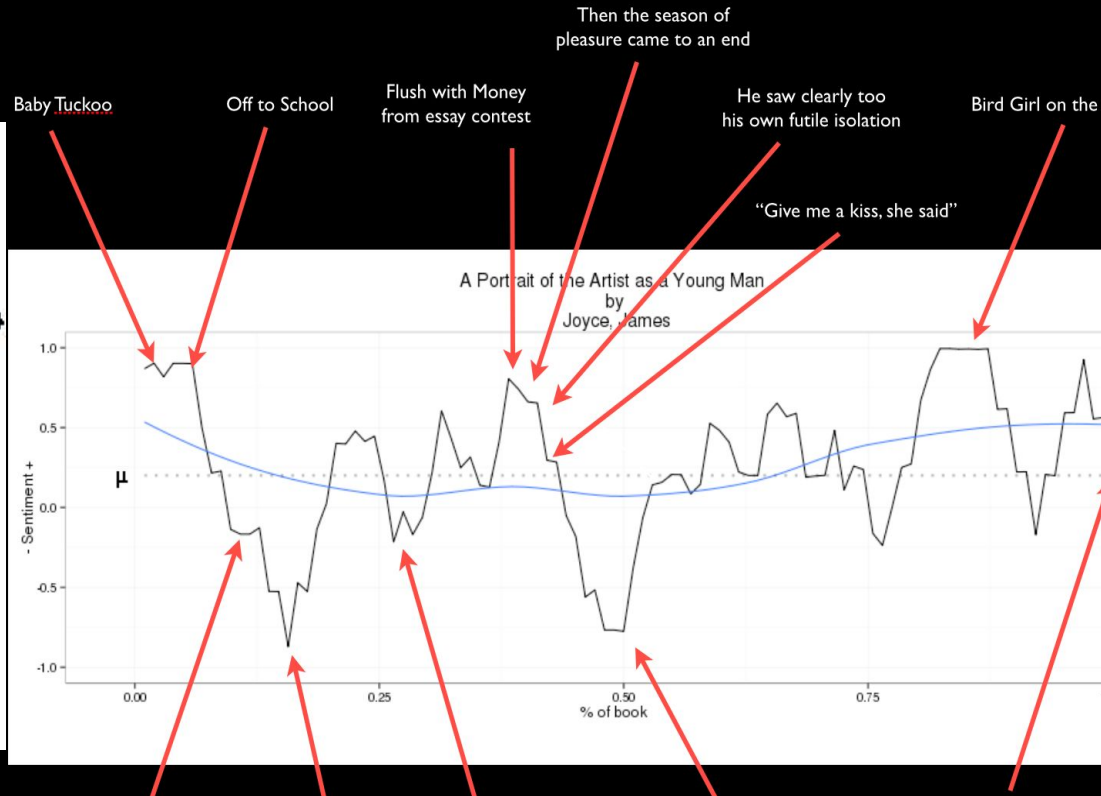


"Instead of reading texts in the traditional way – so-called close reading –, he invites to count, to graph and to map or, in other words, to visualize them" **(Jänicke et al., 2015)**

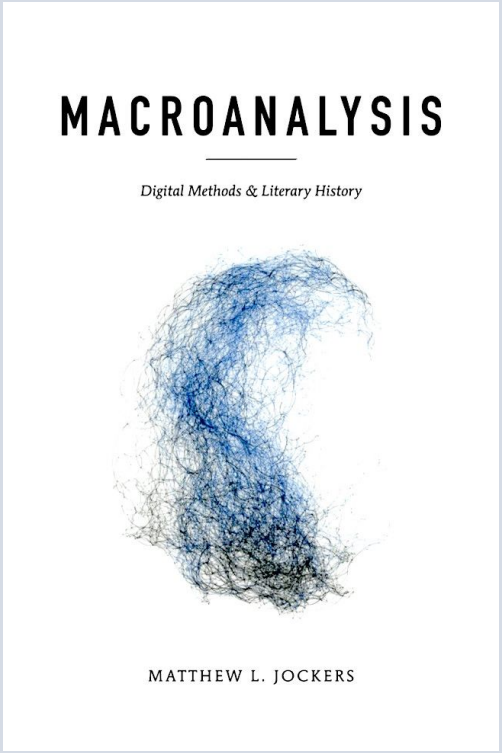
Distant Reading in R

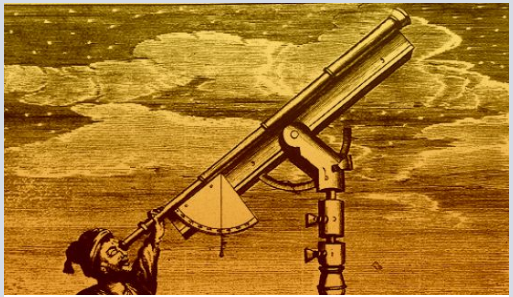
DISTANT READING





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.

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]

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—

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

About

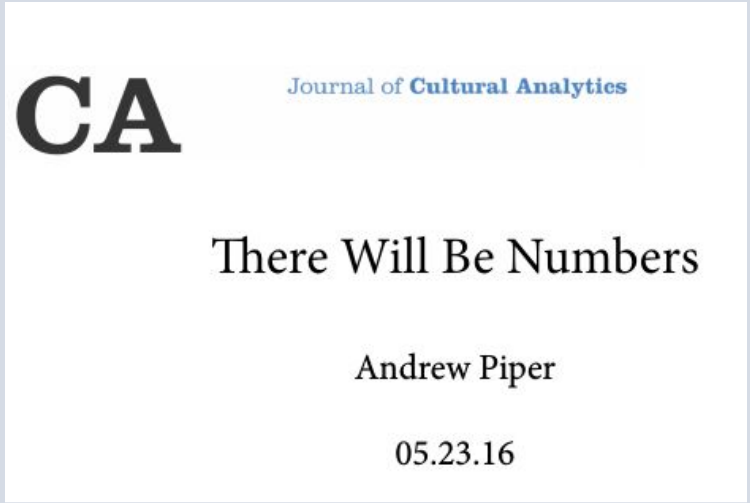
Distant Reading for European Literary History (COST Action CA16204) is a project aiming to create a vibrant and diverse network of researchers jointly developing the resources and methods necessary to change the way European literary history is written. Grounded in the Distant Reading paradigm (i.e. 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 (THEORETICAL)

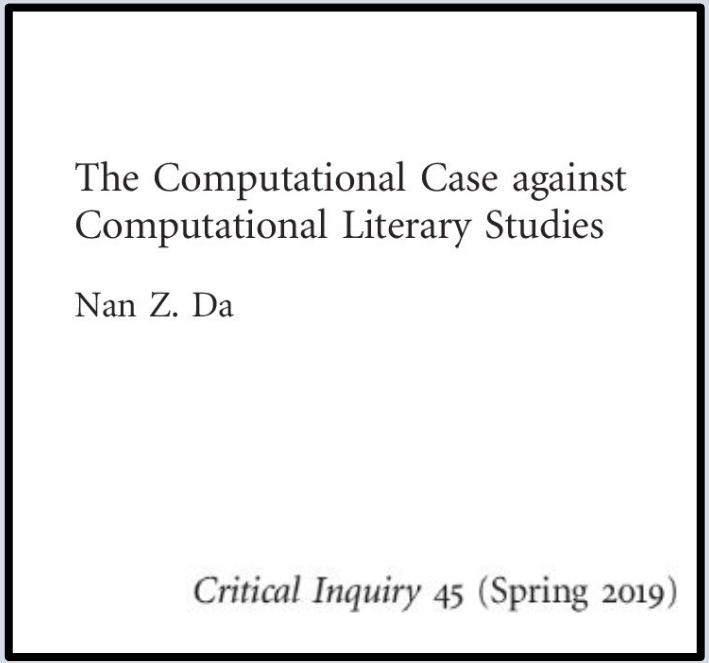
"[…] digital humanities and machine readings of text have resurrected key structuralist presuppositions"

"When we allow our algorithms to overly familiarize that which is fundamentally ambiguous, we risk turning our work, the project of literary criticism, into what Burckhardt would call explanation. This activity of explanation risks too quickly closing down the disturbing possibility of texts"

(Dobson, 2015)

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”!

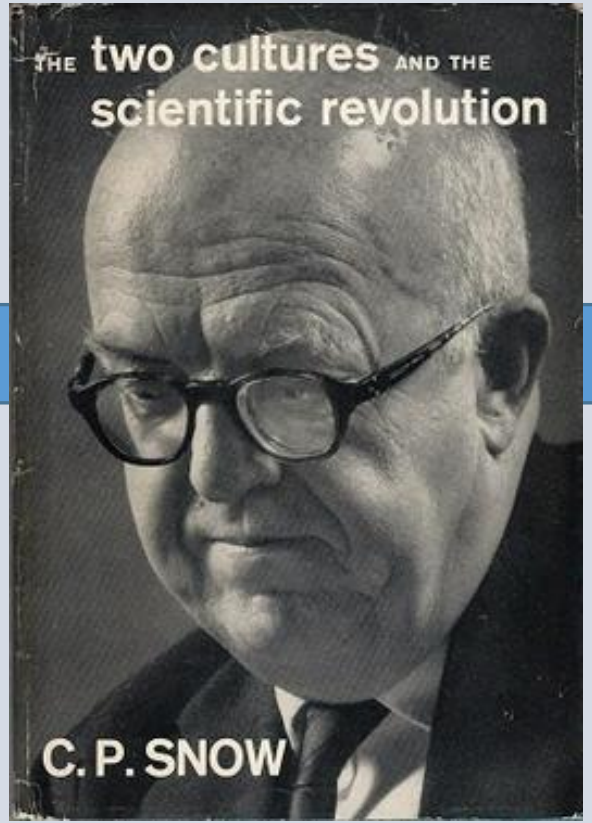


CRITICAL ASPECTS (IDEOLOGICAL)

«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)

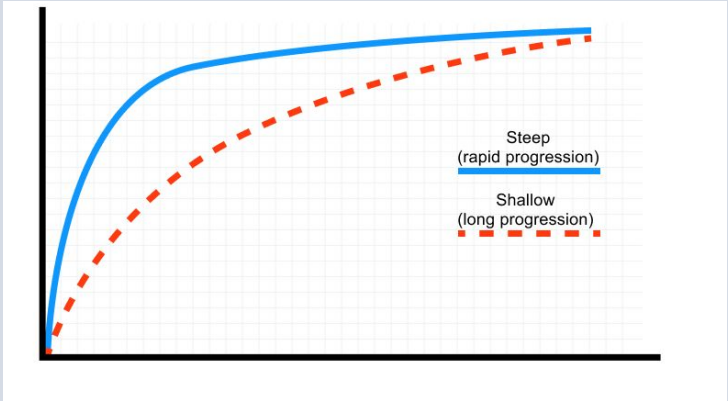
WHY R?

PROS:

- a single tool for many different tasks
- a tool that can be adapted for doing (almost) everything you can think of
- a tool supported by a wide community of researchers and developers

CONS:

- the “steep learning curve” of programming languages



PROGRAMMING LANGUAGES

Simon: “How many times does the word “whale” appear in *Moby Dick* by Herman Melville?”

The computer: “It appears 123 times, dear Simon!”

...encoding

...decoding

Simon:
00100100101000100101
01010001010101001010
11101010110101010110
101011010101

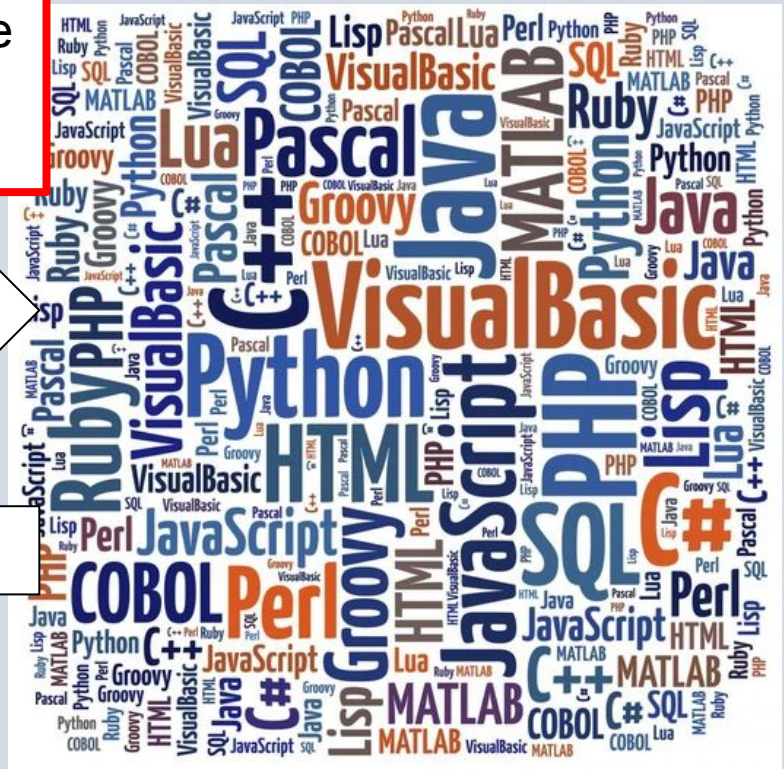
The computer:
11101010100101010101
01010101010101010101
0101010, dear Simon!

PROGRAMMING LANGUAGES

How many times does the word “whale” appear in *Moby Dick*?

It appears in this new iPhone you have to buy, dear Simon!

PROGRAMMING LANGUAGES



PROGRAMMING LANGUAGES

*Low-level programming languages

```
fib:
    mov edx, [esp+8]
    cmp edx, 0
    ja @f
    mov eax, 0
    ret

    @@:
    cmp edx, 2
    ja @f
    mov eax, 1
    ret

    @@:
    push ebx
    mov ebx, 1
    mov ecx, 1

    @@:
        lea eax, [ebx+ecx]
        cmp edx, 3
        jbe @f
        mov ebx, ecx
        mov ecx, eax
        dec edx
    jmp @b
```

[Assembly]

PROGRAMMING LANGUAGES

*High-level programming languages

```c
unsigned fib(unsigned n) {
    if (!n)
        return 0;
    else if (n <= 2)
        return 1;
    else {
        unsigned a, c;
        for (a = c = 1; ; --n) {
            c += a;
            if (n <= 3) return c;
            a = c - a;
        }
    }
}
```

[C]

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: