

L'Institut Des Humanités Numériques : outils et ressources

École d'hiver de l'IDHN - L'IA pour les SHS, les SHS pour l'IA : regards croisés

Mardi 16 décembre 2025

Maison de la recherche SHS Annie Ernaux, Cergy, France

Dimitra Niaouri

dimitra.niaouri@ensea.fr

Marine Delaborde

marine.delaborde@cyu.fr

Mathias Quoy

mathias.quoy@cyu.fr



IDHN

institut des humanités
numériques

IDHN

- Rôle de l'IDHN à CY Cergy Paris Université
 - **Fédération** de la recherche en HN à CY
 - Environnement propice aux **collaborations**
 - **Partage de méthodes** et diffusion d'une culture scientifique numérique
 - Accompagnement et mise à disposition de **ressources**
- **Perspectives :**
 - Développement de la formation et montée en compétences (liens avec consortium)
 - Besoins de puissance de calcul
 - Investissement dans de nouveaux serveurs grâce à deux projets LT2D

L'IDHN depuis le 1er juillet 2025

5 laboratoires



Conseil de direction

- **Mathias QUOY** (ETIS) - directeur
- **Marine DELABORDE** (LT2D) - directrice adjointe
- **Marie-Stéphanie ABOUNA** (AGORA)
- **Boris BORZIC** (pôle ingénierie ETIS)
- **Sonia CASTAGNET** (EMA)
- **Basile MICHEL** (PLACE)

Ressources

Matériel et logiciels

- **Matériel :**

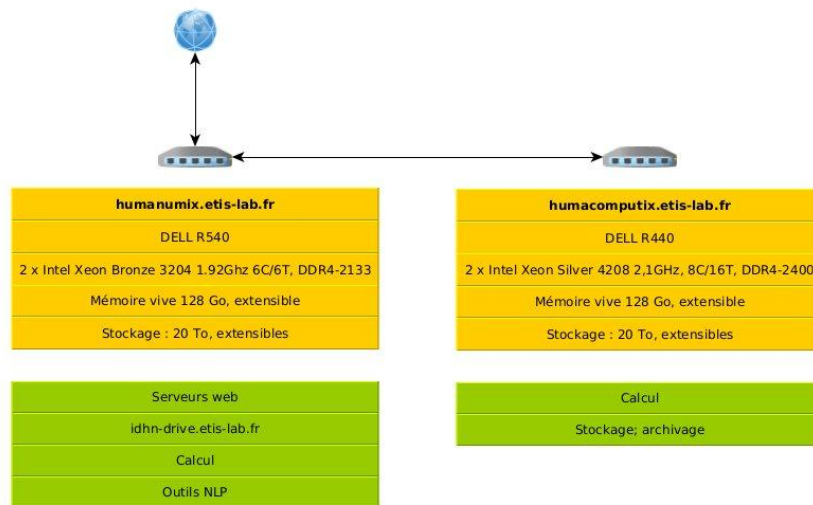
- **serveurs :**

- Humanumix (stockage)
 - Humacomputix (calculs)

- **ordinateurs fixes**

- **Outils :**

- R / RStudio
 - IRaMuTeQ
 - QGIS
 - NVivo
 - Emotaix (Tropes)
 - PyDoct / Tesseract OCR
 - INCEpTION
 - TXM



Aide à l'utilisation des outils et ressources

- **Objectif initial** : mutualiser l'effort de recrutement selon les compétences recherchées
 - Financement par le DIM : ingénieurs HN (Zakarya Despré, Jeremy Demange)
 - Développement d'outils pour des projets de recherche liés au DIM et à l'IDHN
 - API de textométrie
 - Modèles de détection de sentiments et d'ironie
 - Tests de culture politique
 - Interface graphique pour application de modèles de détection de la coréférence
 - VIKY.ai : partenariat de recherche avec la société Pertimm pour le développement d'une plateforme multi-agents pour le traitement de données linguistiques
 - <https://github.com/cyidhn>

Aide à l'utilisation des outils et ressources

- À l'heure actuelle (2025/2026) :
 - **Mission doctorale hors enseignement** : 32 jours (Dimitra Niaouri)
 - Production de documentation : rédaction de fiches pratiques et de tutoriels
 - Veille : suivi de l'évolution des outils et méthodes en HN
 - Accompagnement ponctuel personnalisé :
 - entretiens avec les chercheurs, conseils sur les outils adaptés, aide à la mise en application
 - Réalisation de rapports synthétiques des accompagnements
 - Contactez-nous si vous avez besoin d'un outil ou d'un accompagnement :
 - mathias.quoy@cyu.fr
 - marine.delaborde@cyu.fr

Découverte d'outils et de ressources

- Actions actuelles :
 - Séminaires
 - Conférences
 - Écoles thématiques
- En discussion → formations :
 - Ajout du séminaire **ONuSel** (Outils Numériques au Service de la Langue) au Catalogue des formations de l'**ED AHSS Arts, Humanités, Sciences Sociales**
 - Organisation de **formations** - en collaboration avec des consortiums **Huma-Num** ou non



Ressources

Méthodologie

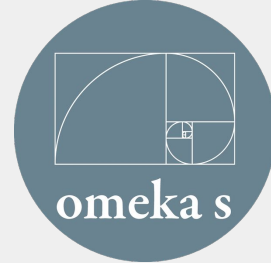
- [Guide de valorisation des données au LT2D](#) :
 - **Faire le lien avec les chercheurs**
 - Cadre : CPJ « Ressources numériques en SHS »
 - Ojectif 1 : « Valorisation des données au LT2D »
 - Mission IGE : Thomas Gervais d'Aldin (2024)
 - **Données langagières** écrites ou orales (Lexiques, Textes, Discours, Dictionnaires)
 - Objectif : veille et amélioration régulière
 - Données issues des RSN (CER + DPO)
 - Séminaire ONuSeL en 2026

1 Les enjeux de la valorisation des données de la recherche	2
1.1 Sauvegarder les travaux	2
1.2 Donner de la visibilité à ses travaux	3
1.3 Assurer la réutilisation	4
1.4 Sécuriser le financement	5
2 Préparer la constitution d'un jeu de données	6
2.1 Respecter les principes éthiques et légaux	6
2.1.1 Les principes FAIR	6
2.1.2 Le cycle de vie des données	7
2.1.3 Élaborer un Plan de Gestion des Données (PGD)	8
2.1.4 Le Règlement Général sur la Protection des données (RGPD)	8
2.2 Les ressources : des instances, des outils et des formations	10
2.2.1 À CY Cergy Paris Université	11
2.2.2 Se former à la science ouverte et au numérique	12
2.2.3 Différents formats de données langagières	13
2.2.4 Des bases de données en ligne : des catalogues de ressources	16
2.2.5 Des outils de traitement du langage	19
3 Le partage et la valorisation des données de la recherche	23
3.1 Donner accès aux données	23
3.1.1 Les entrepôts de données de la recherche	23
3.1.2 Attribuer un identifiant unique pérenne	24
3.1.3 Rendre ses recherches accessibles	24
3.2 Les conditions d'utilisation des données	25
3.2.1 La licence : un instrument juridique	25
3.2.2 Les droits liés aux œuvres	26
3.3 Documenter les données	26
3.3.1 Les métadonnées : des données à propos des données	26
3.3.2 Les <i>data papers</i> : pour une meilleure description des données	27
3.4 Des canaux de communication : pour la centralisation de l'actualité scientifique	28
3.4.1 Des listes de diffusion	28
3.4.2 Des consortiums	29
3.5 La vulgarisation des données : pour le grand public	30
3.5.1 Le référencement sur des répertoires d'expert-e-s	30
3.5.2 Des podcasts et des chroniques	30
3.5.3 Des ateliers et des conférences	31



Musée virtuel des dictionnaires

Laboratoire LT2D - Centre Jean Pruvost



- MVD : Référencement d'ouvrages lexicographiques
 - Passage d'un site statique à une base de données : réalisé par Luc Massip (LT2D)
- Omeka S : Plateforme de publication Web
 - gestion de fonds documentaires
 - système de publication multi-sites
 - gestion des métadonnées pour décrire les ressources
 - utilisation de vocabulaires du web sémantique + standard IIIF
 - modules (de connectivité)
- Sur serveur IDHN : <https://dictionnaires-lt2d.idhn.etis-lab.fr/>
- Contact : marine.delaborde@cyu.fr

INCEpTION

- Annotation multicouche
- Approche collaborative
 - gestion de cohortes d'annotateurs (rôles)
 - attribution de textes
 - adjudication et calcul d'accord inter-annotateur
- Active learning
- Serveur CORLI (consortium Huma-Num) hébergé par l'ATILF : <https://corli.huma-num.fr/fr/inception/>
- Serveur IDHN : <https://inception.idhn.etis-lab.fr/>
- Contact : marine.delaborde@cyu.fr

The screenshot displays the INCEpTION web interface, which is divided into two main panels: 'Active Learning' on the left and 'Annotation' on the right.

Active Learning Panel:

- Session:** Layer: Named entity, Terminate button.
- Recommendation:** Text: Illinois, Label: LOC, Score: 1, Delta: 1. Buttons: Accept, Reject, Skip.
- Learning History:** A table showing the history of recommendations and their status.

Source	URL	Entity	Status
Berkeley	http://www.wikidata.org/entity/Q168756	skipped	
Tesla	PER	accepted	
Tesla	PER	accepted	
Tesla	PER	accepted	
Tesla	PER	accepted	
Tesla	PER	accepted	
Science	OTH	rejected	
Tesla	PER	accepted	

Annotation Panel:

- Layer:** Surface form.
- Annotation:** A list of text segments with annotations. For example, 'Barack Hussein Obama II' is annotated with 'PER', 'born August 4, 1961' with 'DATE', 'is an' with 'occurrence', and 'American politician who served as the 44th President of the United States from 2009 to 2017' with 'POLITICIAN'.
- Identifier:** A dropdown menu showing the current identifier 'Illinois'.
- Annotation Details:** A panel showing the details of the current annotation, including the text 'Illinois Senate' and its description: 'upper chamber of the Illinois General Assembly, the legislative branch of the government of the state of Illinois in the United States'.

source : <https://inception-project.github.io/>

IRaMuTeQ

- Open-source software for statistical text analysis.
- It was created by French researcher Pierre Ratinaud at LERASS (Laboratoire d'Études et de Recherches Appliquées en Sciences Sociales).
- Based on **R** and **Python**.
- **Webpage:**
<https://pratinaud.gitpages.huma-num.fr/iramuteq-website/>



Accueil Installation Documentation Aide

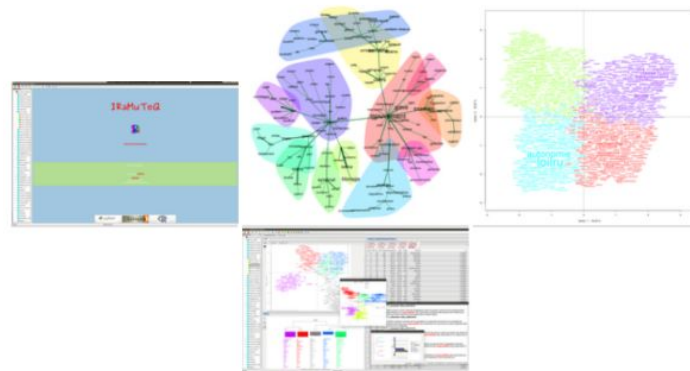
IRaMuTeQ

Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. Un logiciel libre construit avec des logiciels libres.

Published: Sat 08 November 2025

By [Pierre Ratinaud](#)

In [Accueil](#).



Présentation

- Iramuteq est un logiciel libre distribué sous les termes de la licence GNU GPL (v2).
- Il permet de faire des analyses statistiques sur des corpus texte et sur des tableaux individus/caractères.
- Il repose sur le logiciel R (www.r-project.org) et le langage python (www.python.org)

Analyses sur des corpus texte

- Iramuteq reproduit la méthode de classification décrite par Reinert (1983, 1991) (Classification Hiérarchique Descendante sur un tableau croisant les formes pleines et des segments de texte);

IRaMuTeQ

Installation guidelines:

<https://pratinaud.gitpages.huma-num.fr/iramuteq-website/category/installation.html>

[Accueil](#)[Installation](#)[Documentation](#)[Aide](#)

Installation

Installation sur windows

Published: Sat 08 November 2025

By [Pierre Ratinaud](#)

In [Installation](#).

Translations: [english](#)

- Téléchargez et installez la dernière version de R pour Windows : <https://cran.r-project.org/bin/windows/base/R-4.5.2-win.exe>
- Téléchargez et installez la dernière version d'iramuteq pour Windows sur sourceforge : <https://sourceforge.net/projects/iramuteq/files/iramuteq-0.8alpha7/setup-iramuteq-0.8a7.exe/download>
- Au premier lancement d'iramuteq, acceptez l'installation des packages de R.

Installation sur Mac OS

- Téléchargez et installez xquartz: <https://www.xquartz.org/>

Pour les processeurs Apple Silicon (M1, M2,...)

- Téléchargez et installez la dernière version de R pour Mac OS Apple Silicon : <https://cloud.r-project.org/bin/macosx/big-sur-arm64/base/R-4.5.2-arm64.pkg>
- Téléchargez et installez la dernière version d'iramuteq pour Apple silicon : <https://sourceforge.net/projects/iramuteq/files/iramuteq-0.8alpha7/iramuteq-0.8a7-macOS-Mx.zip/download>
- Au premier lancement d'iramuteq, acceptez l'installation des packages de R.

Pour les processeurs Intel

- Téléchargez et installez la dernière version de R pour Mac OS Intel : https://cloud.r-project.org/bin/macosx/big-sur-x86_64/base/R-4.5.2-x86_64.pkg
- Téléchargez et installez la dernière version d'iramuteq pour Mac OS Intel : <https://sourceforge.net/projects/iramuteq/files/iramuteq-0.8alpha7/iramuteq-0.8a7-macOS-Intel.zip/download>
- Au premier lancement d'iramuteq, acceptez l'installation des packages de R.

Installation sur Linux

IRaMuTeQ

Documentation:

<https://pratinaud.gitpages.huma-num.fr/iramuteq-website/category/installation.html>

Preparing IRaMuTeQ input files:

<https://pratinaud.gitpages.huma-num.fr/iramuteq-website/documentation/DataFilePreparation.pdf>

Installation

Installation sur windows

Published: Sat 08 November 2025

By [Pierre Ratinaud](#)

In [Installation](#).

Translations: [english](#)

- Téléchargez et installez la dernière version de R pour Windows : <https://cran.r-project.org/bin/windows/base/R-4.5.2-win.exe>
- Téléchargez et installez la dernière version d'iramuteq pour Windows sur sourceforge : <https://sourceforge.net/projects/iramuteq/files/iramuteq-0.8alpha7/setup-iramuteq-0.8a7.exe/download>
- Au premier lancement d'iramuteq, acceptez l'installation des packages de R.

Installation sur Mac OS

- Téléchargez et installez xquartz: <https://www.xquartz.org/>

Pour les processeurs Apple Silicon (M1, M2,...)

- Téléchargez et installez la dernière version de R pour Mac OS Apple Silicon : <https://cloud.r-project.org/bin/macosx/big-sur-arm64/base/R-4.5.2-arm64.pkg>
- Téléchargez et installez la dernière version d'iramuteq pour Apple silicon : <https://sourceforge.net/projects/iramuteq/files/iramuteq-0.8alpha7/iramuteq-0.8a7-macOS-Mx.zip/download>
- Au premier lancement d'iramuteq, acceptez l'installation des packages de R.

Pour les processeurs Intel

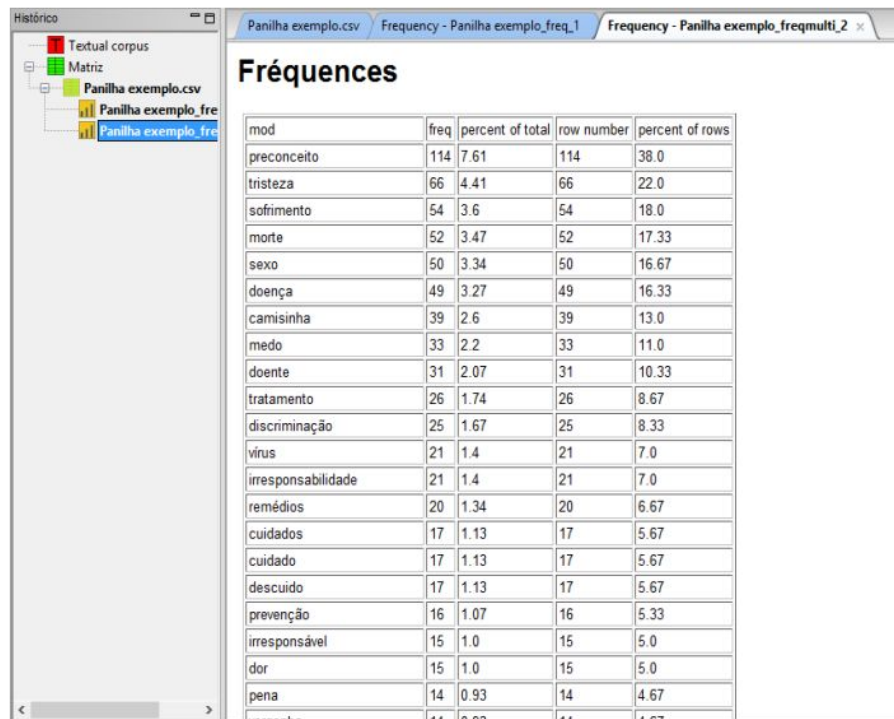
- Téléchargez et installez la dernière version de R pour Mac OS Intel : https://cloud.r-project.org/bin/macosx/big-sur-x86_64/base/R-4.5.2-x86_64.pkg
- Téléchargez et installez la dernière version d'iramuteq pour Mac OS Intel : <https://sourceforge.net/projects/iramuteq/files/iramuteq-0.8alpha7/iramuteq-0.8a7-macOS-Intel.zip/download>
- Au premier lancement d'iramuteq, acceptez l'installation des packages de R.

Installation sur Linux

IRaMuTeQ

1. Classic Lexicographical Analysis

- Counts word frequencies.
- Identifies rare words (hapax).
- Creates reduced word forms (lemmatization).
- Builds a dictionary of reduced forms.
- Useful for counting and interpreting data.



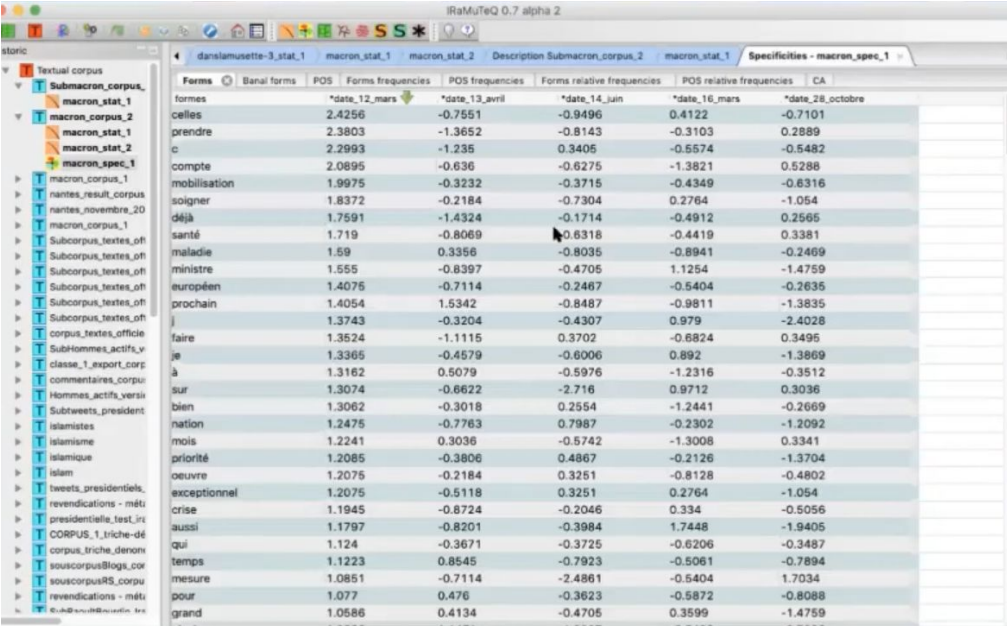
Fréquences

mod	freq	percent of total	row number	percent of rows
preconceito	114	7.61	114	38.0
tristeza	66	4.41	66	22.0
sofrimento	54	3.6	54	18.0
morte	52	3.47	52	17.33
sexo	50	3.34	50	16.67
doença	49	3.27	49	16.33
camisinha	39	2.6	39	13.0
medo	33	2.2	33	11.0
doente	31	2.07	31	10.33
tratamento	26	1.74	26	8.67
discriminação	25	1.67	25	8.33
vírus	21	1.4	21	7.0
irresponsabilidade	21	1.4	21	7.0
remédios	20	1.34	20	6.67
cuidados	17	1.13	17	5.67
cuidado	17	1.13	17	5.67
descuido	17	1.13	17	5.67
prevenção	16	1.07	16	5.33
irresponsável	15	1.0	15	5.0
dor	15	1.0	15	5.0
pena	14	0.93	14	4.67
...	...	...	...	...

IRaMuTeQ

2. Specificities & Correspondence Factor Analysis

- Links text segments with variables (e.g., gender, age, topic).
- Shows which words are specific to each variable.
- Helps compare texts according to the characterization variables



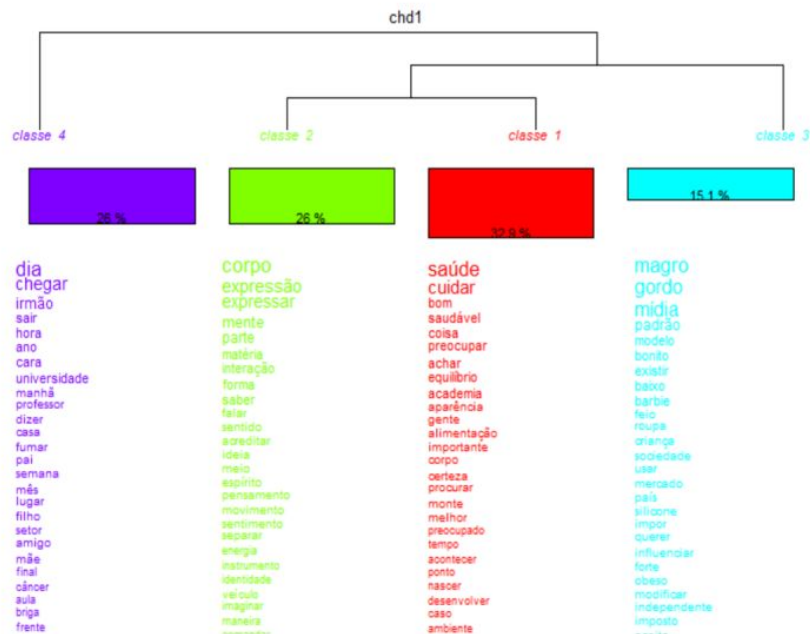
The screenshot displays the IRaMuTeQ 0.7 alpha 2 software interface. On the left, a tree view shows a project named 'danslamusette-3_stat_1' containing several sub-corpora. The main window shows a table titled 'Specificities - macron_stat_1'. The table has columns for 'Formes', 'Banal forms', 'POS', and four sets of 'Forms frequencies' and 'Forms relative frequencies' for different dates. The rows list various words and their corresponding values across these categories.

Formes	Banal forms	POS	Forms frequencies	POS frequencies	Forms relative frequencies	POS relative frequencies	CA
formes			*date_12_mars	*date_13_avril	*date_14_juin	*date_16_mars	*date_28_octobre
celles			2.4256	-0.7551	-0.9496	0.4122	-0.7101
prendre			2.3803	-1.3652	-0.8143	-0.3103	0.2889
c			2.2993	-1.235	0.3405	-0.5574	-0.5482
compte			2.0896	-0.636	-0.6275	-1.3821	0.5298
mobilisation			1.9975	-0.3232	-0.3715	-0.4349	-0.6316
soigner			1.8372	-0.2184	-0.7304	0.2764	-1.054
déjà			1.7591	-1.4324	-0.1714	-0.4912	0.2595
santé			1.719	-0.8069	0.6318	-0.4419	0.3381
maladie			1.59	0.3356	-0.8035	-0.8941	-0.2469
ministre			1.555	-0.8397	-0.4705	1.1254	-1.4759
européen			1.4075	-0.7114	-0.2467	-0.5404	-0.2635
prochain			1.4054	1.5342	-0.8487	-0.9811	-1.3835
j			1.3743	-0.3204	-0.4307	0.979	-2.4028
faire			1.3524	-1.1115	0.3702	-0.8824	0.3495
je			1.3365	-0.4579	-0.6006	0.892	-1.3869
à			1.3162	0.5079	-0.5976	-1.2316	-0.3512
sur			1.3074	-0.6622	-2.716	0.9712	0.3036
bien			1.3062	-0.3018	0.2554	-1.2441	-0.2669
nation			1.2475	-0.7763	0.7987	-0.2302	-1.2092
mois			1.2241	0.3036	-0.5742	-1.3008	0.3341
priorité			1.2085	-0.3806	0.4867	-0.2126	-1.3704
oeuvre			1.2075	-0.2184	0.3251	-0.8128	-0.4802
exceptionnel			1.2075	-0.5118	0.3251	0.2764	-1.054
crise			1.1945	-0.8724	-0.2046	0.334	-0.5056
aussi			1.1797	-0.8201	-0.3984	1.7448	-1.9405
qui			1.124	-0.3671	-0.3725	-0.6206	-0.3487
temps			1.1223	0.8545	-0.7923	-0.5061	-0.7894
mesure			1.0851	-0.7114	-2.4861	-0.5404	1.7034
pour			1.077	0.476	-0.3623	-0.5872	-0.8088
grand			1.0586	0.4134	-0.4705	0.3599	-1.4759

IRaMuTeQ

3. Descending Hierarchical Analysis (DHA)

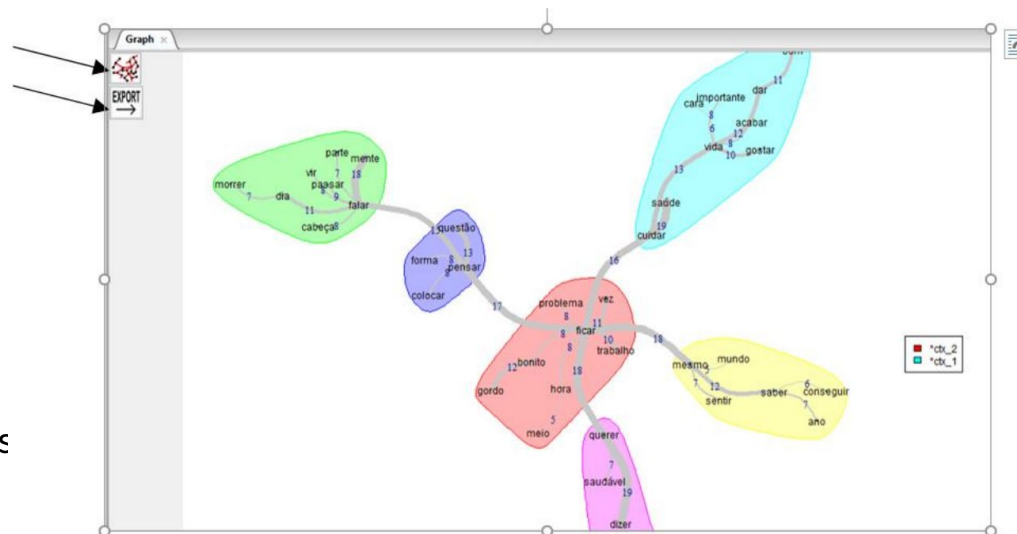
- Groups text segments based on shared vocabulary.
- Produces clusters with similar words.
- Shows the cluster structure through a dendrogram.
- Provides top words and key variables for each cluster.
- Used to identify themes, lexical fields, or social representations.



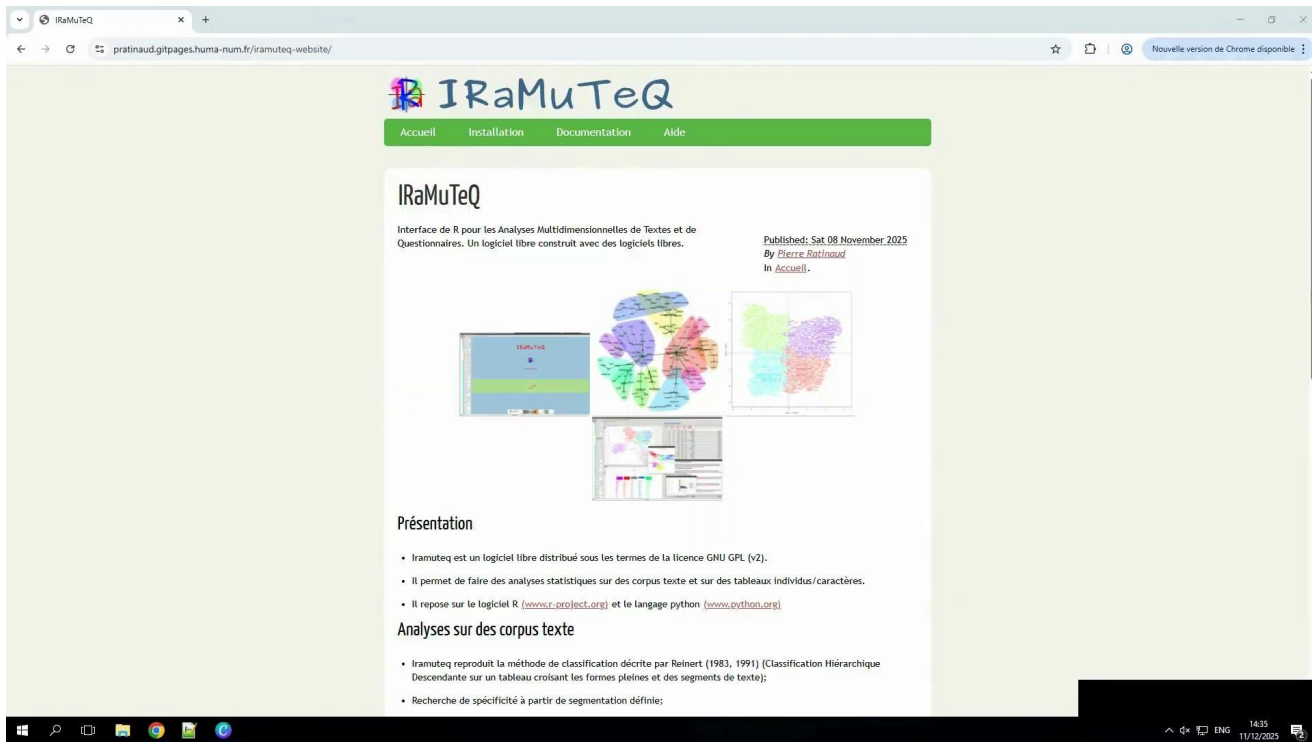
IRaMuTeQ

4. Similarity Analysis

- Shows how words are connected through co-occurrences.
- Helps visualize the structure of the text (central vs. peripheral words).
- Allows to identify the word-cooccurrences providing information about the word connectivity in a corpus.
- Useful for identifying shared content across variables.



IRaMuTeQ



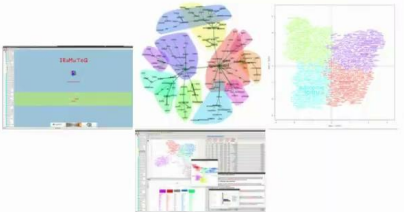
IRaMuTeQ

Accueil Installation Documentation Aide

IRaMuTeQ

Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. Un logiciel libre construit avec des logiciels libres.

Published: Sat 08 November 2025
By [Pierre Ratinaud](#)
In [Accueil](#).



Présentation

- Iramuteq est un logiciel libre distribué sous les termes de la licence GNU GPL (v2).
- Il permet de faire des analyses statistiques sur des corpus texte et sur des tableaux individus/caractères.
- Il repose sur le logiciel R (www.r-project.org) et le langage python (www.python.org)

Analyses sur des corpus texte

- Iramuteq reproduit la méthode de classification décrite par Reinert (1983, 1991) (Classification Hiérarchique Descendante sur un tableau croisant les formes pleines et des segments de texte);
- Recherche de spécificité à partir de segmentation définie;

Tropes

- Open-source software for semantic analysis
- The initial release of the Tropes software dates to 1994

Webpage: <https://www.tropes.fr/>

Why Choose Tropes?

- Free Software
- No coding required
- No need for predefined conceptual categories or coding frames.
- Mainly a text-mining tool



Tropes

Installation guidelines:

<https://www.tropes.fr/download.htm>

Téléchargement du logiciel Tropes

[Sommaire](#) | [Forum](#) | [Documentation](#) | [Téléchargement](#) | [Contact](#)

TELECHARGEMENT DE TROPES V8

Tropes fonctionne avec Microsoft Windows (toutes versions depuis Windows XP).

Ce logiciel peut être utilisé gratuitement, sous réserve des conditions décrites dans son contrat de licence :

CONTRAT DE LICENCE DU LOGICIEL TROPES V8

Tropes est un logiciel développé par Pierre Molette et Agnès Landré sur la base des travaux de Rodolphe Ghiglione.
Copyright 1995-2018. Tous droits réservés.

La redistribution et l'utilisation du logiciel Tropes sont autorisées si les conditions suivantes sont remplies :

1. Le logiciel Tropes doit être distribué gratuitement.

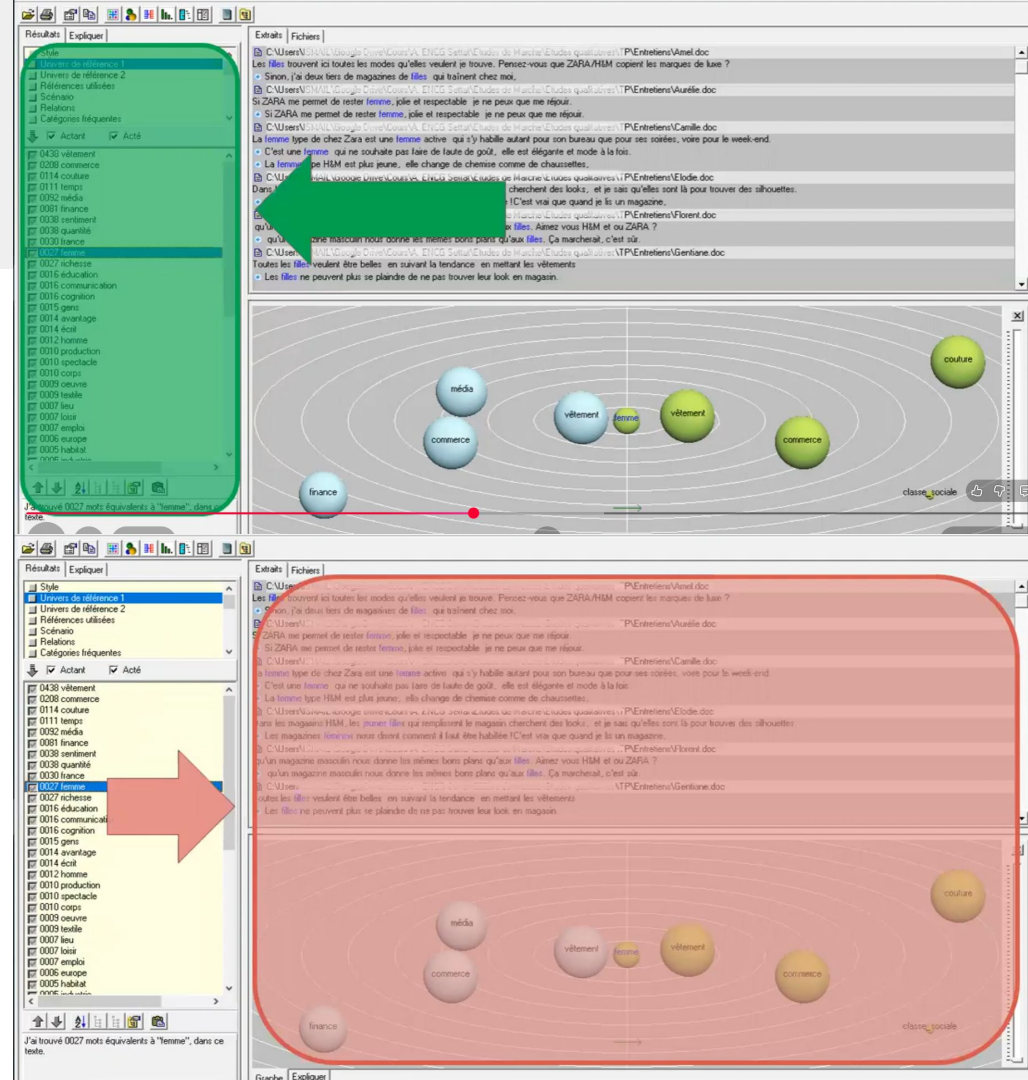
Cliquez ci-dessous pour télécharger Tropes

Dernières versions officielles de Tropes :

Version française	Version anglaise
Tropes VF 8.5 :	Tropes US 8.4.4 :
 8 décembre 2018	 28 juin 2014

Chaque langue fait l'objet d'une version différente : si vous voulez analyser des textes dans ces deux langues, vous devez installer les deux logiciels.

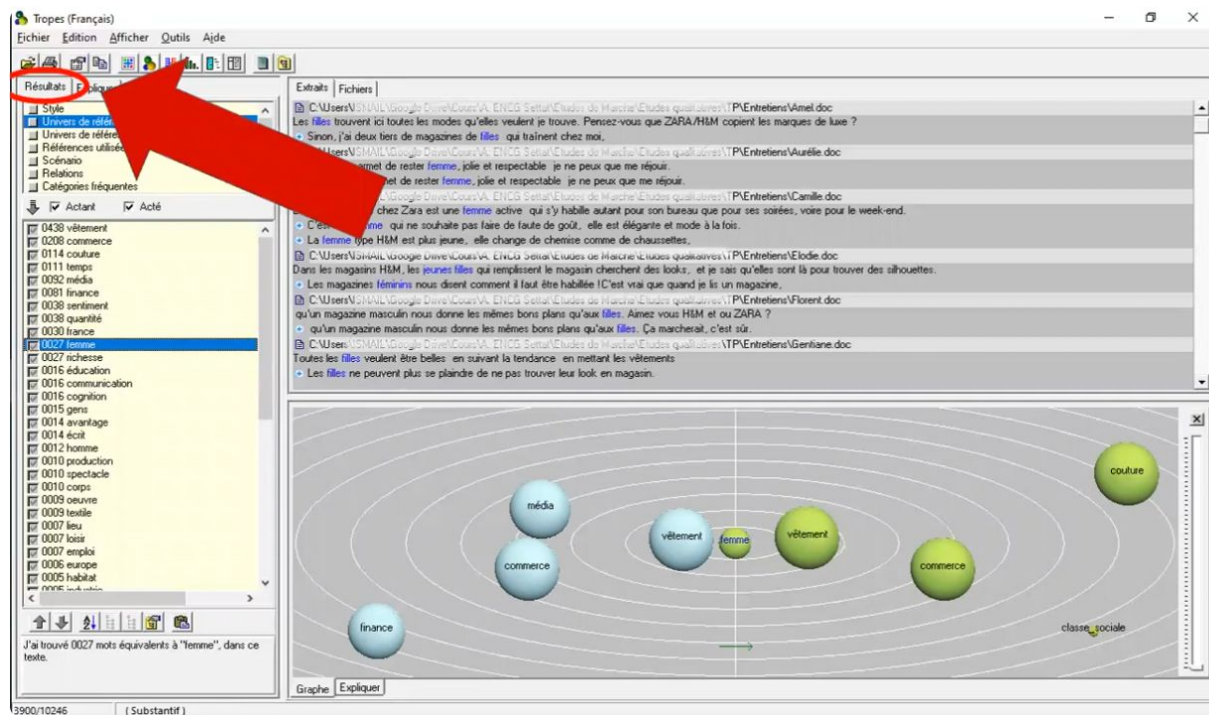
Pensez à désinstaller les versions précédentes et à sauvegarder vos scénarios avant toute mise à jour.



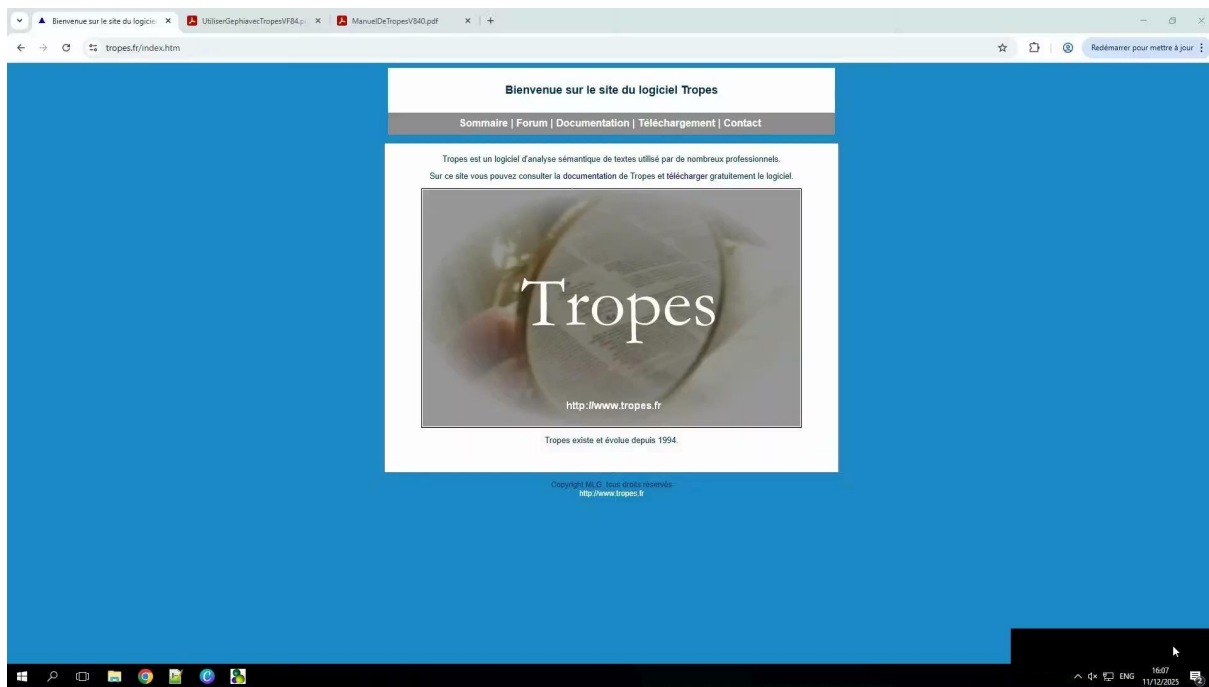
Tropes

Key Concept: Universes of Reference

- A core component of Tropes.
- A “universe of reference” is a classification or semantic universe grouping terms.
- Tropes organises words into:
 - Classes
 - Semantic universes
 - Levels (general to specific)



Tropes



R & R Studio

- R is a free software environment for statistical computing and graphics.
- Webpage: <https://www.r-project.org/>



[Home]

Download

[CRAN](#)

R Project

[About R](#)

[Logo](#)

[Contributors](#)

[What's New?](#)

[Reporting Bugs](#)

[Conferences](#)

[Search](#)

[Get Involved: Mailing Lists](#)

[Get Involved: Contributing](#)

[Developer Pages](#)

[R Blog](#)

R Foundation

[Foundation](#)

[Board](#)

[Members](#)

[Donors](#)

[Donate](#)

Help With R

[Getting Help](#)

Documentation

[Manuals](#)

[FAQs](#)

[The R Journal](#)

[Books](#)

[Certification](#)

[Other](#)

Links

[Bioconductor](#)

[R-Forge](#)

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- [R version 4.5.2 \(\[Not\] Part in a Rumble\)](#) has been released on 2025-10-31.
- The [useR! 2026](#) conference will take place in Warsaw, Poland, July 7-9.
- You can support the R Foundation with a renewable subscription as a [supporting member](#).

News via Mastodon



R Foundation

The R Foundation is happy to be a partner on the project

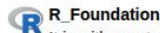
"Enabling the next generation of R contributors"

awarded a large grant (£500k) under the Research Software Maintenance Fund run by the UK's Software Sustainability Institute.

Announcement: software.ac.uk/ssi-awards-fund...

Congratulations to the project team!

Dec 10, 2023



R Foundation

It is with great sadness that we learned of the death John Fox this week. He was the author of numerous books on applied statistics and accompanying [#RStats](#) packages including car, effects, Rcmdr, ivreg, and matlib. John was a member of the R Foundation until 2023.

Nov 28, 2023



R Foundation

R version 4.5.2 "[Not] Part in a Rumble" (source version) has been released. You can find it in cran.r-project.org/src/base/R-4/, or wait for CRAN to be updated.

Oct 31, 2023

R & R Studio

Installation:

<https://cran.r-project.org/mirrors.html>

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux \(Debian, Fedora/Redhat, Ubuntu\)](#)
- [Download R for macOS](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2025-10-31, [Not] Part in a Rumble) [R-4.5.2.tar.gz](#), read [what's new](#) in the latest version.
- The CRAN directory [src-base-prepare](#) contains R alpha, beta, and rc releases as daily snapshots in time periods before a planned release.
- Between releases, the same directory [src-base-prepare](#) contains snapshots of current patched and development versions. Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Alternatively, daily snapshots are [available here](#).
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#).

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

Supporting CRAN

- CRAN operations, most importantly hosting, checking, distributing, and archiving of R add-on packages for various platforms, crucially rely on technical, emotional, and financial support by the R community.

Please consider making [financial contributions](#) to the R Foundation for Statistical Computing.

What are R and CRAN?

R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the [R project homepage](#) for further information.

CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN [mirror](#) nearest to you to minimize network load.

R & R Studio

Installation:

- <https://posit.co/downloads/>

Why use R Studio?

- Makes working with R easier
- Standardizes keyboard shortcuts across Mac, Windows, and Linux
- Provides a consistent user experience

DOWNLOAD

RStudio IDE

The most popular coding environment for R, built with love by Posit.

Used by millions of people weekly, the RStudio integrated development environment (IDE) is a set of tools built to help you be more productive with R and Python. It includes a console, syntax-highlighting editor that supports direct code execution. It also features tools for plotting, viewing history, debugging and managing your workspace.

If you're a professional data scientist and want guidance on adopting open-source tools at your organization, don't hesitate to [book a call with us.](#)

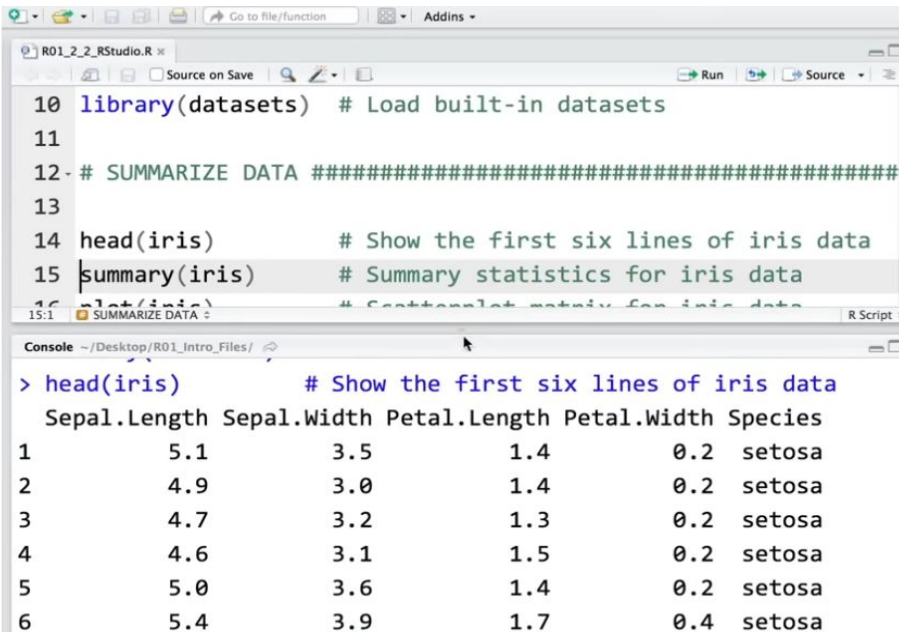
DOWNLOAD RSTUDIO

DOWNLOAD RSTUDIO SERVER

[RStudio Desktop](#)

R & R Studio

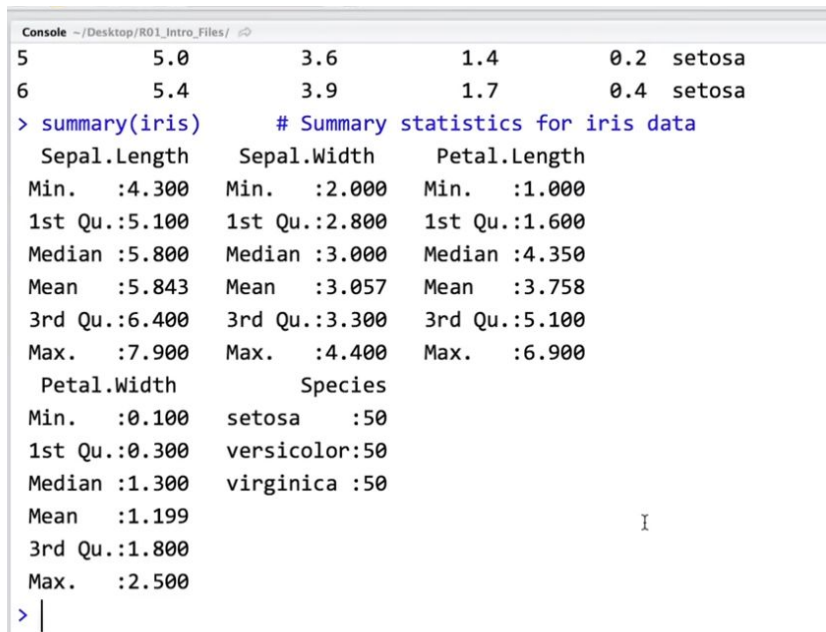
Loading and printing datasets



```
R01_2_2_RStudio.R
10 library(datasets) # Load built-in datasets
11
12 # SUMMARIZE DATA #####
13
14 head(iris) # Show the first six lines of iris data
15 summary(iris) # Summary statistics for iris data
16 plot(iris) # Scatterplot matrix for iris data

Console ~/Desktop/R01_Intro_Files/
> head(iris) # Show the first six lines of iris data
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4          0.2  setosa
2          4.9         3.0          1.4          0.2  setosa
3          4.7         3.2          1.3          0.2  setosa
4          4.6         3.1          1.5          0.2  setosa
5          5.0         3.6          1.4          0.2  setosa
6          5.4         3.9          1.7          0.4  setosa
```

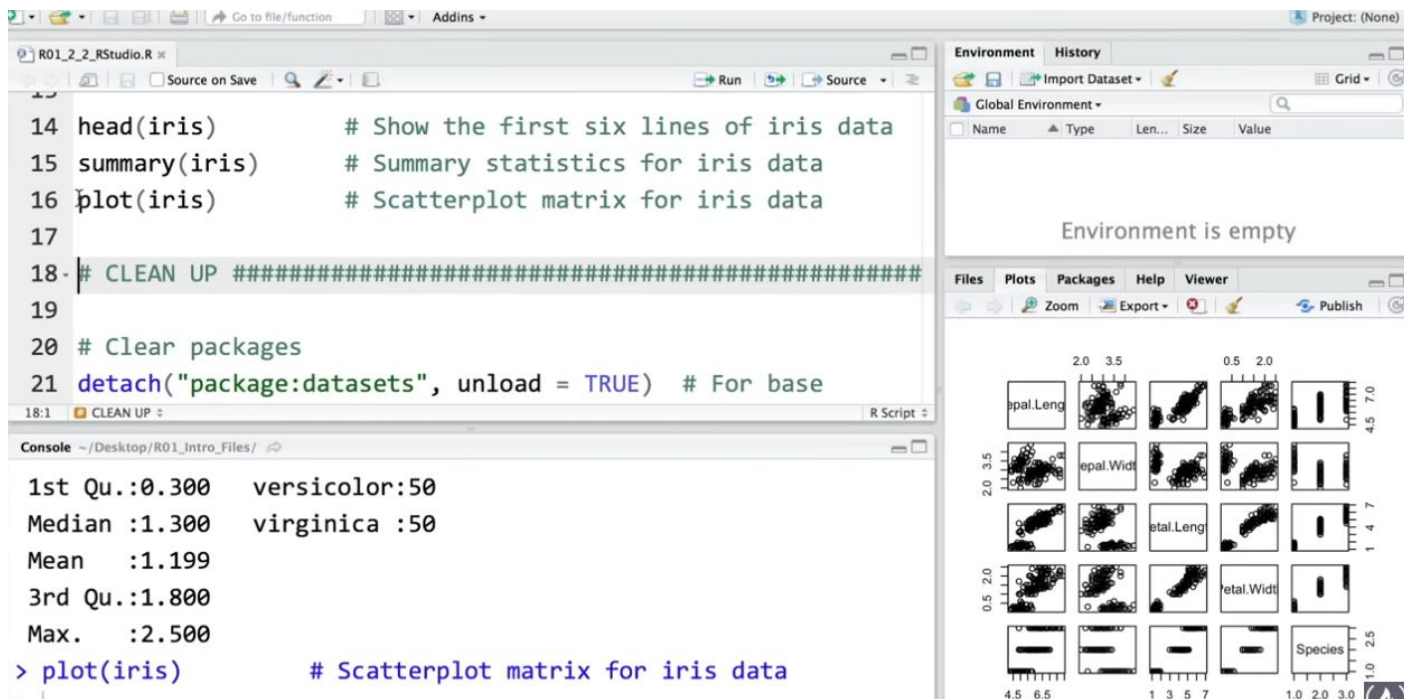
Descriptive statistics



```
Console ~/Desktop/R01_Intro_Files/
5          5.0          3.6          1.4          0.2  setosa
6          5.4          3.9          1.7          0.4  setosa
> summary(iris) # Summary statistics for iris data
  Sepal.Length Sepal.Width Petal.Length
Min.   :4.300  Min.   :2.000  Min.   :1.000
1st Qu.:5.100  1st Qu.:2.800  1st Qu.:1.600
Median :5.800  Median :3.000  Median :4.350
Mean   :5.843  Mean   :3.057  Mean   :3.758
3rd Qu.:6.400  3rd Qu.:3.300  3rd Qu.:5.100
Max.   :7.900  Max.   :4.400  Max.   :6.900
  Petal.Width Species
Min.   :0.100  setosa   :50
1st Qu.:0.300  versicolor:50
Median :1.300  virginica :50
Mean   :1.199
3rd Qu.:1.800
Max.   :2.500
> |
```

R & R Studio

Plotting datasets

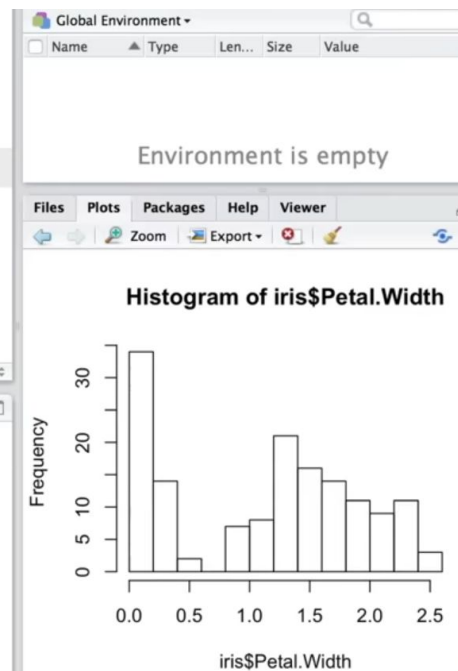


R & R Studio

Histograms

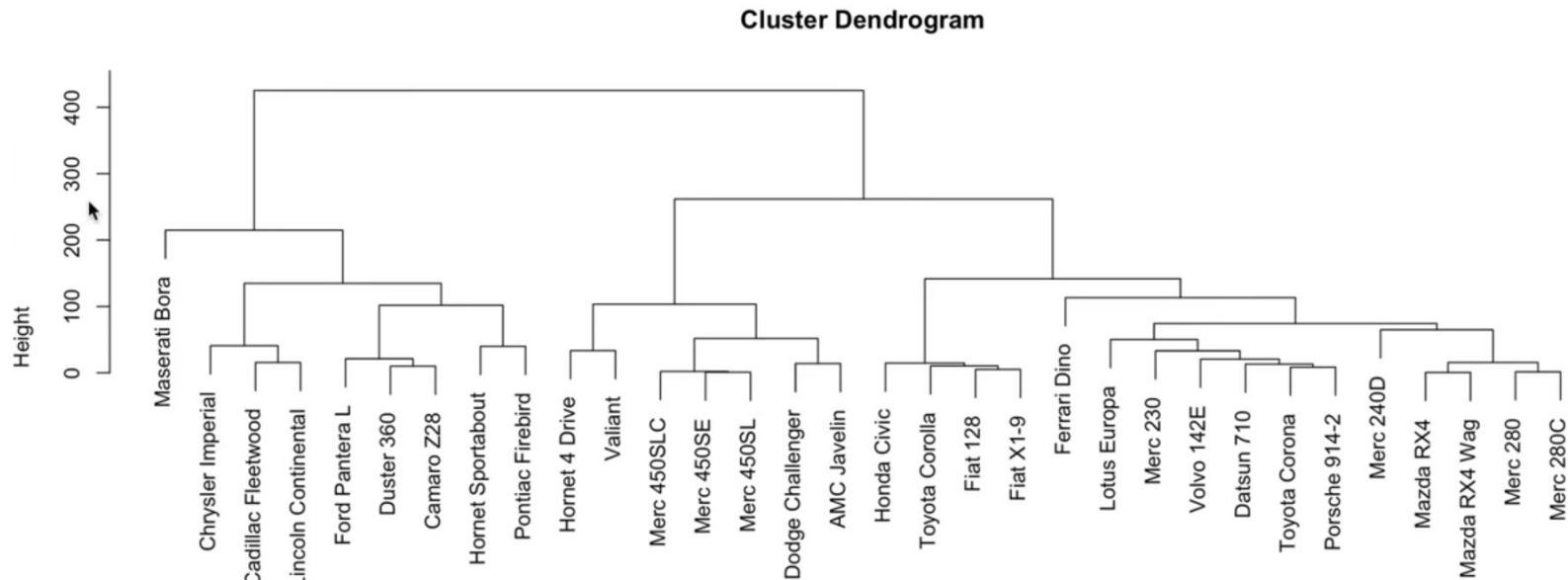
```
20  
29 # Histograms for each species using options  
30 hist(iris$Petal.Width [iris$Species == "setosa"],  
31     xlim = c(0, 3),  
32     breaks = 9,  
33     main = "Petal Width for Setosa",  
34     xlab = "",  
35     col = "red")  
36  
37 hist(iris$Petal.Width [iris$Species == "versicolor"]
```

```
32:14 HISTOGRAM BY GROUP  
Console ~/Desktop/R01_Intro_Files/ R Script  
6      5.4      3.9      1.7      0.4 setosa  
> hist(iris$Sepal.Length)  
> hist(iris$Sepal.Width)  
> hist(iris$Petal.Length)  
> hist(iris$Petal.Width)  
> par(mfrow = c(3, 1))  
> head(iris)
```



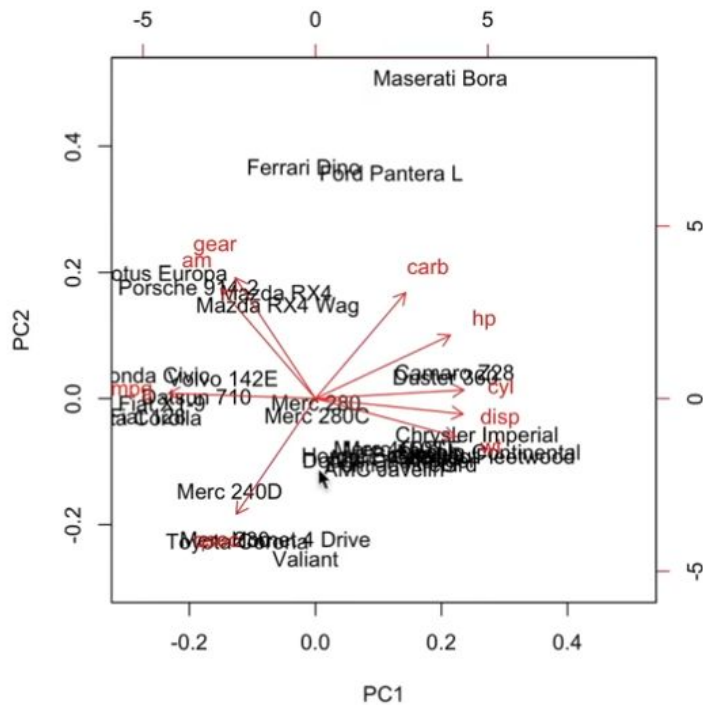
R & R Studio

Hierarchical Clustering

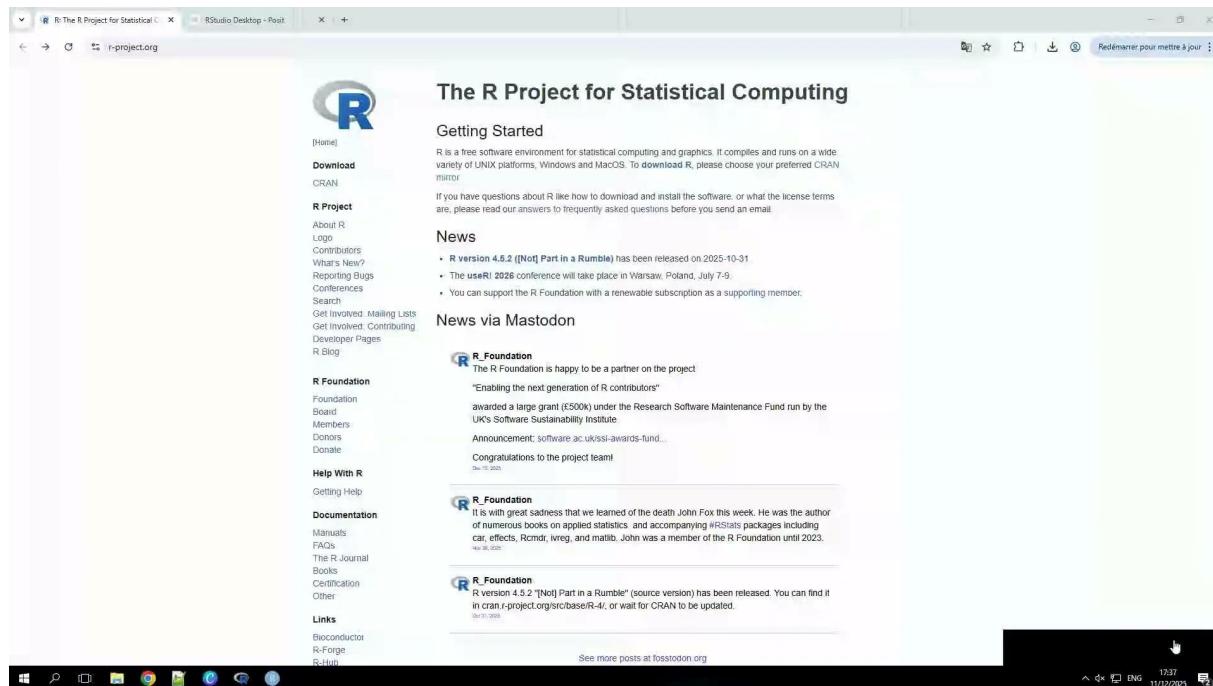


R & R Studio

Principal Component Analysis (PCA)

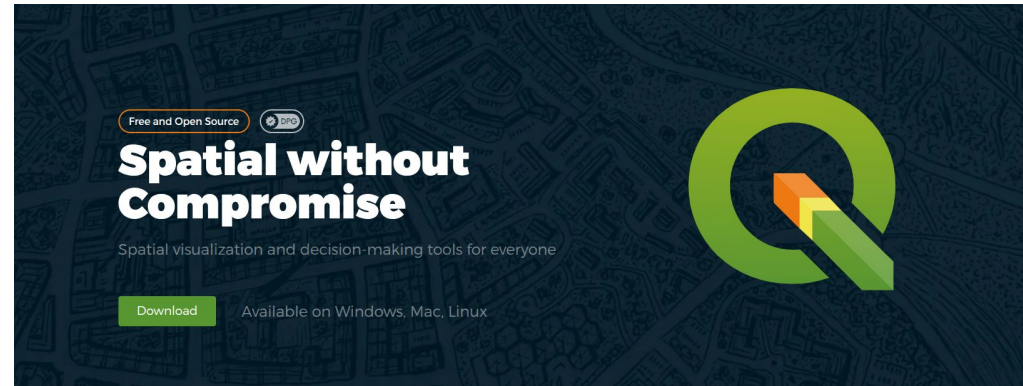


R & R Studio



QGIS

- QGIS is an open-source tool for creating, editing, visualizing, analyzing, and sharing geospatial data.
- Runs on Mac, Windows, Linux, and can be deployed in cloud environments.
- Offers a rich plugin ecosystem to extend its mapping, analysis, and data management capabilities.
- Supports Python scripting (PyQGIS) for automation and advanced customization.
- **Webpage:** <https://qgis.org/>



Installation

- Project
- Community
- Resources
- Documentation
- Installation guide
- Windows
- Linux
- Mac OS X / macOS
- FreeBSD
- OpenBSD
- Container images
- QGIS Testing warning
- Installing from Source
- Testing
- Roadmap
- Reports
- Resources Hub
- Books
- Contact and Support

on GitHub

QGIS Installers

QGIS is available on Windows, macOS, Linux, FreeBSD, OpenBSD, Android and iOS.



We recommend installing the Long Term Release (LTR) packages since our LTR versions undergo more rigorous testing before release and receive at least one year of bug fix updates.

See also [the main installers page](#).

To evaluate the upcoming release and to allow non-developers to support development we also provide testing packages (note the [warning](#)).

In the feature frozen phase preceding a release (see [Release schedule](#)) these packages are effectively **pre-releases, which we strongly urge users to test**.

Windows

There are two options for installing on Windows.

Offline (Standalone) installers

Standalone installers include everything QGIS needs in a single download. Once you have the installer, no internet is required to complete the installation. When a new release is available, you need to download the complete installer again in order to upgrade. For beginners, the standalone installer is probably the easiest way to install QGIS.

Release	Version	Package
Latest Release	3.44.5 Solothurn	Installer Checksum

INSTALLATION DOWNLOADS

[ALL RELEASES](#)

[SOURCES](#)

Download for Windows



The OSGeo4W installer is recommended for regular users or organization deployments. *It allows to have several QGIS versions place, and to keep each component up-to-date individually without having to download the whole package.*

Since QGIS 3.20 we only ship 64-bit Windows executables.

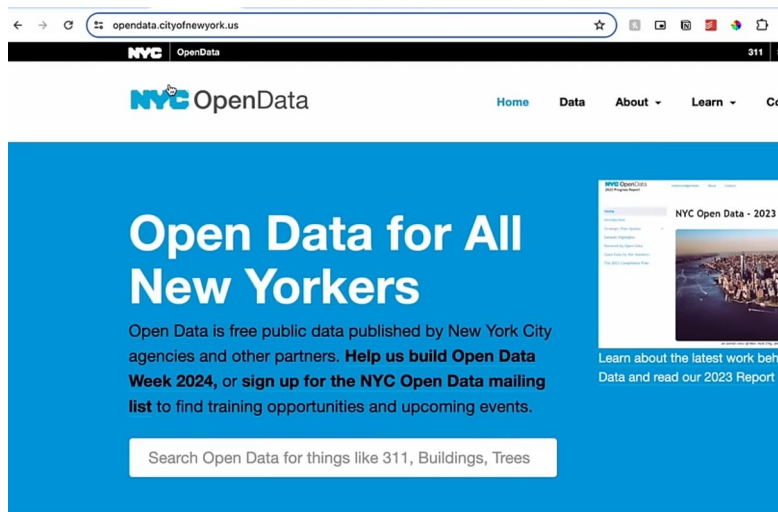
Download for macOS

Download for Linux

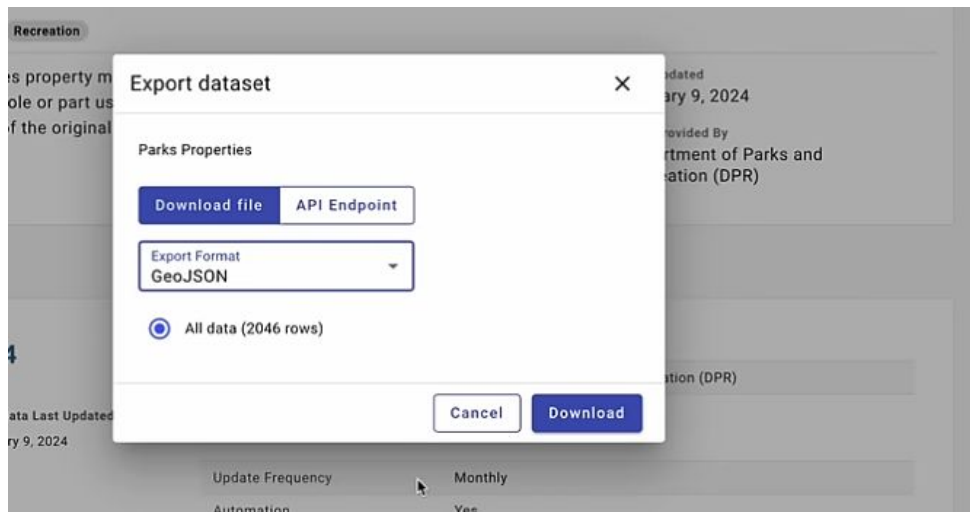


QGIS

Exporting data



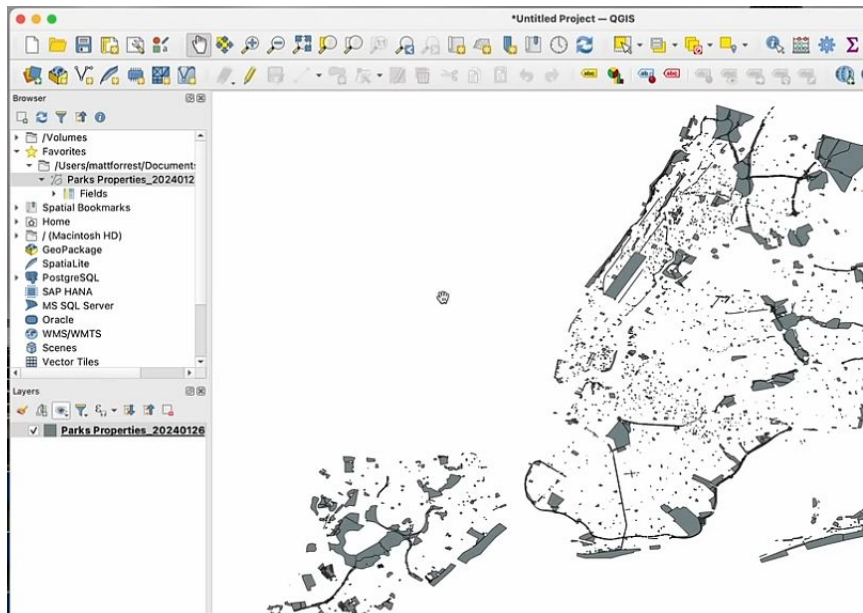
The screenshot shows the NYC OpenData website. The browser address bar displays 'opendata.cityofnewyork.us'. The website header includes the 'NYC OpenData' logo and navigation links for 'Home', 'Data', 'About', 'Learn', and 'Contact'. A large blue banner features the text 'Open Data for All New Yorkers' and a sub-header 'Open Data is free public data published by New York City agencies and other partners. Help us build Open Data Week 2024, or sign up for the NYC Open Data mailing list to find training opportunities and upcoming events.' Below the banner is a search bar with the placeholder text 'Search Open Data for things like 311, Buildings, Trees'. To the right of the banner, there is a section titled 'NYC Open Data - 2023' with a small image of a city skyline and a link to 'Learn about the latest work behind Open Data and read our 2023 Report'.



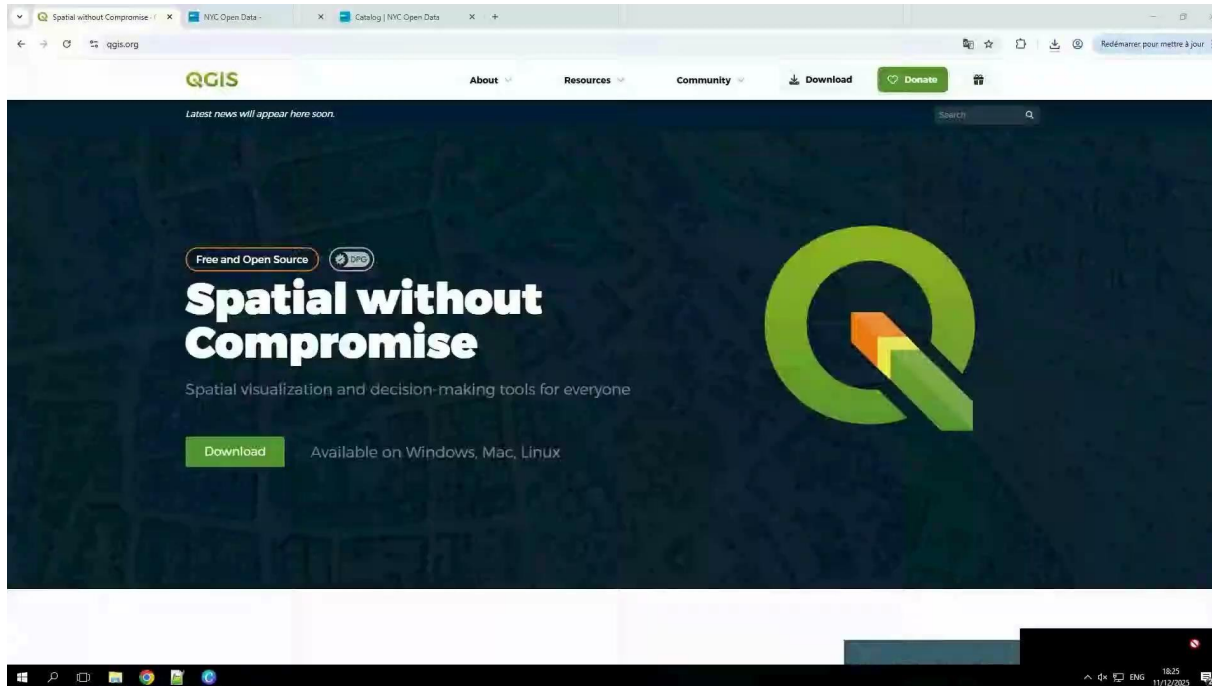
The screenshot shows a 'Recreation' dataset page with an 'Export dataset' dialog box open. The dialog box has a title bar with a close button (X). Inside, the 'Parks Properties' section is active, showing two tabs: 'Download file' (selected) and 'API Endpoint'. Below the tabs is a dropdown menu for 'Export Format' set to 'GeoJSON'. Underneath, there is a radio button selection for 'All data (2046 rows)'. At the bottom right of the dialog are 'Cancel' and 'Download' buttons. The background of the page shows a table with columns like 'Update Frequency' and 'Automation', and a 'Last Updated' date of 'January 9, 2024'.

QGIS

Adding Layers



QGIS



Tesseract

- Open-source Optical Character Recognition (OCR) engine.
- Provides libtesseract (library) and a command-line program.
- Supports 100+ languages and UTF-8.
- **Webpage:** <https://github.com/tesseract-ocr/tesseract>

Running Tesseract

Basic [command line usage](#):

```
tesseract imagename outputbase [-l lang] [--oem ocrenginemode] [--psm pagesegmode] [configfile]
```

For more information about the various command line options use `tesseract --help` or `man tesseract`.

Examples can be found in the [documentation](#).

Tesseract OCR

coverage [passed 12 new defects](#) [CodeQL](#) [passing](#) [oss-fuzz](#) [fuzzing](#)
license [Apache-2.0](#) [download](#) [all releases](#)

Table of Contents

- [Tesseract OCR](#)
 - [About](#)
 - [Brief history](#)
 - [Installing Tesseract](#)
 - [Running Tesseract](#)
 - [For developers](#)
 - [Support](#)
 - [License](#)
 - [Dependencies](#)
 - [Latest Version of README](#)

About

This package contains an **OCR engine** - `libtesseract` and a **command line program** - `tesseract`.

Tesseract 4 adds a new neural net (LSTM) based [OCR engine](#) which is focused on line recognition, but also still supports the legacy Tesseract OCR engine of Tesseract 3 which works by recognizing character patterns. Compatibility with Tesseract 3 is enabled by using the Legacy OCR Engine mode (`--oem 0`). It also needs [traineddata](#) files which support the legacy engine, for example those from the [tessdata](#) repository.

Stefan Weil is the current lead developer. Ray Smith was the lead developer until 2017. The maintainer is Zdenko Podobny. For a list of contributors see [AUTHORS](#) and GitHub's log of [contributors](#).

Tesseract has **unicode (UTF-8) support**, and can **recognize more than 100 languages** "out of the box".

Tesseract supports [various image formats](#) including PNG, JPEG and TIFF.

Tesseract supports **various output formats**: plain text, hOCR (HTML), PDF, invisible-text-only PDF, TSV, ALTO and

Tesseract

The screenshot shows the GitHub repository for Tesseract, an Open Source OCR Engine. The page is viewed in Google Chrome. The repository is public and has 1,597 watchers, 1,844 forks, and 71,318 stars. The main repository is located at [tesseract-ocr/tesseract](https://github.com/tesseract-ocr/tesseract). The repository is managed by [dependabot\[bot\]](#) and [stweil](#). The repository contains 11 branches and 67 tags. The repository is licensed under Apache-2.0. The repository is a C++ project that provides a command-line interface and a library for OCR. The repository is used by many projects, including [Apache Tika](#), [Apache PDFBox](#), and [Apache PDFBox](#). The repository is also used by many commercial products, including [Adobe Acrobat](#), [Microsoft Office](#), and [Google Docs](#). The repository is a good example of a well-maintained open-source project.

Repository Details:

- Repository: [tesseract-ocr/tesseract](#)
- License: Apache-2.0
- Stars: 71.3k
- Forks: 1.8k
- Watchers: 1.6k
- Issues: 432
- Pull requests: 33
- Actions: 1
- Projects: 1
- Wiki: 1
- Security: 1
- Insights: 1

Recent Commits:

Commit	Author	Message	Time
112435	dependabot[bot] and stweil	Bump mikemzenjak/junit-report from 4 to 6	last week
112435	dependabot[bot] and stweil	Bump mikemzenjak/junit-report from 4 to 6	last week
112435	dependabot[bot] and stweil	Bump mikemzenjak/junit-report from 4 to 6	last week
112435	dependabot[bot] and stweil	Bump mikemzenjak/junit-report from 4 to 6	last week
112435	dependabot[bot] and stweil	Bump mikemzenjak/junit-report from 4 to 6	last week
112435	dependabot[bot] and stweil	Bump mikemzenjak/junit-report from 4 to 6	last week
112435	dependabot[bot] and stweil	Bump mikemzenjak/junit-report from 4 to 6	last week
112435	dependabot[bot] and stweil	Bump mikemzenjak/junit-report from 4 to 6	last week
112435	dependabot[bot] and stweil	Bump mikemzenjak/junit-report from 4 to 6	last week
112435	dependabot[bot] and stweil	Bump mikemzenjak/junit-report from 4 to 6	last week

Files:

- [.github](#)
- [cmake](#)
- [doc](#)
- [include/tesseract](#)
- [java](#)
- [m4](#)
- [nsis](#)
- [snap](#)
- [src](#)
- [tessdata](#)
- [test @ 232f118](#)
- [unittest](#)
- [.clang-format](#)
- [.gitattributes](#)
- [.gitignore](#)

About:

Tesseract Open Source OCR Engine (main repository)

[tesseract-ocr.github.io/](#)

[machine-learning](#) [ocr](#) [tesseract](#) [tesseract-ocr](#) [tesseract-ocr-engine](#)

[Readme](#)

[Apache-2.0 license](#)

[Contributing](#)

[Cite this repository](#)

[Activity](#)

[Custom properties](#)

[71.3k stars](#)

[1.7k watching](#)

[10.4k forks](#)

[Report repository](#)

Releases: 34

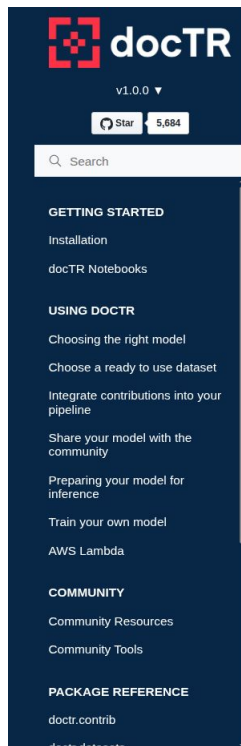
[5.5.1](#) (Latest) on May 25

[~ 32 releases](#)

Contributors: 190

Python docTR

- Open-source Document Text Recognition library by Mindee.
- Built on PyTorch for modern deep-learning OCR.
- Provides text detection + text recognition in documents.
- Extracts and structures textual information from scanned or photographed documents.
- **Webpage:** <https://mindee.github.io/doctr/index.html>



**docTR: Document Text
Recognition**

State-of-the-art Optical Character Recognition made seamless & accessible to anyone, powered by PyTorch



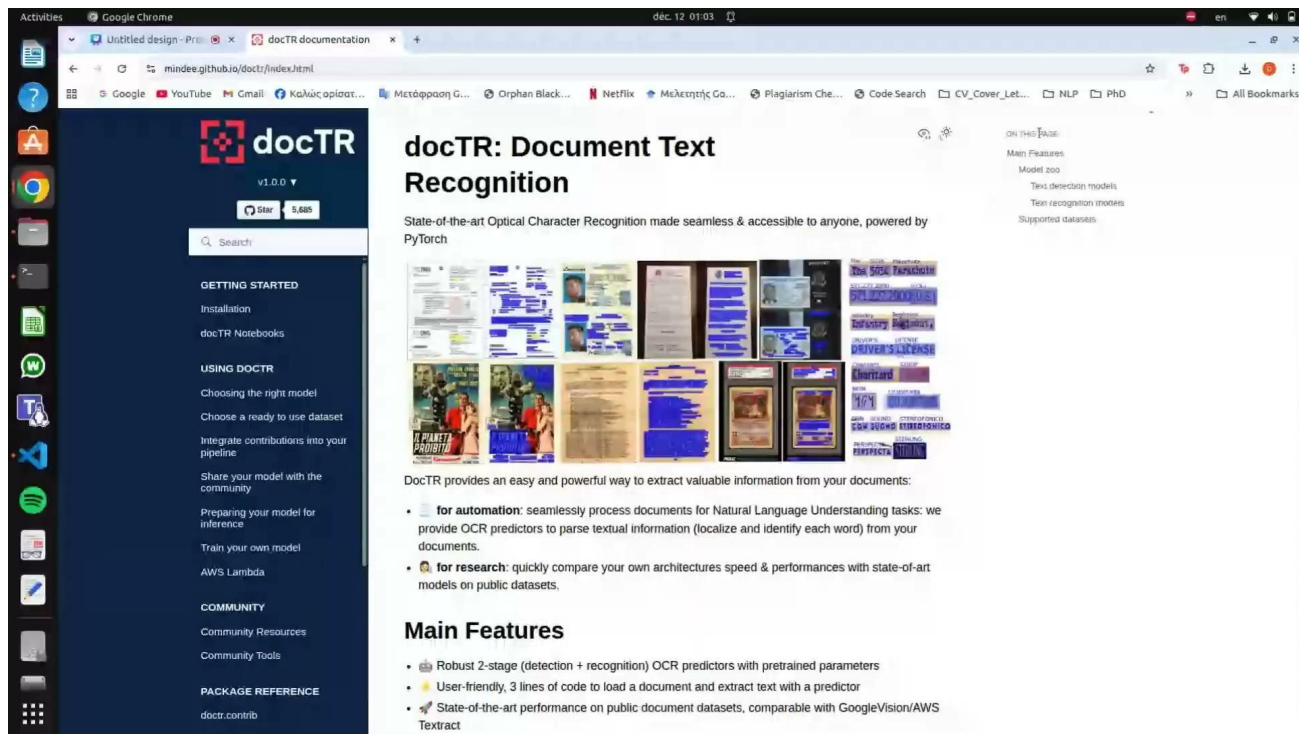
DocTR provides an easy and powerful way to extract valuable information from your documents:

-  **for automation:** seamlessly process documents for Natural Language Understanding tasks: we provide OCR predictors to parse textual information (localize and identify each word) from your documents.
-  **for research:** quickly compare your own architectures speed & performances with state-of-art models on public datasets.

Main Features

- 🤖 Robust 2-stage (detection + recognition) OCR predictors with pretrained parameters
- ⚡ User-friendly, 3 lines of code to load a document and extract text with a predictor
- 🚀 State-of-the-art performance on public document datasets, comparable with GoogleVision/AWS

Python docTR



The screenshot shows the docTR documentation website in a Google Chrome browser. The browser's address bar shows the URL `mindee.github.io/doctr/index.html`. The website has a dark blue header with the docTR logo and version `v1.0.0`. A sidebar on the left contains navigation links under the headings **GETTING STARTED**, **USING DOCTR**, and **COMMUNITY**. The main content area features the title **docTR: Document Text Recognition** and a subtitle: **State-of-the-art Optical Character Recognition made seamless & accessible to anyone, powered by PyTorch**. Below the subtitle is a grid of 12 images showing various documents and text extraction results. To the right of the grid is a sidebar with links: **ON THIS PAGE**, **Main Features**, **Model zoo**, **Text detection models**, **Text recognition models**, and **Supported datasets**. Below the grid, a paragraph states: **DocTR provides an easy and powerful way to extract valuable information from your documents:**. This is followed by two bullet points:

- **for automation:** seamlessly process documents for Natural Language Understanding tasks: we provide OCR predictors to parse textual information (localize and identify each word) from your documents.
- **for research:** quickly compare your own architectures speed & performances with state-of-art models on public datasets.

 Below this is the **Main Features** section, which includes three bullet points:

- Robust 2-stage (detection + recognition) OCR predictors with pretrained parameters
- User-friendly, 3 lines of code to load a document and extract text with a predictor
- State-of-the-art performance on public document datasets, comparable with GoogleVision/AWS Textract

Merci pour votre attention !



<https://cyidhn.cyu.fr/accueil>