

**Enseigner R sur YouTube
en temps de COVID-19 :
retour d'expérience des webin R**

Joseph Larmarange

Janvier **2026**

Auteurs Authors

Joseph Larmarange (joseph.larmarange@ird.fr) est démographe en santé publique, directeur de recherche à l'IRD, et membre du Centre Population et Développement.

<http://orcid.org/0000-0001-7097-700X>

Citation recommandée Recommended citation

Larmarange J. « Enseigner R sur YouTube en temps de COVID-19 : retour d'expérience des webin-R », *Working Paper du Ceped*, n°62, Ceped (UMR 196, Université Paris Cité, Université Sorbonne Paris Nord, IRD, Inserm), Paris, Janvier 2026.

<https://doi.org/10.5281/zenodo.18272454> | <http://www.cephed.org/wp62>

Ceped Centre Population et Développement

UMR 196 · Université Paris Cité · Université Sorbonne Paris Nord · IRD · Inserm
45 rue des Saints-Pères 75006 Paris, France · 20 avenue George Sand 93210 Saint-Denis, France
<https://www.cephed.org/> · contact@cephed.org

Les Working Papers du Ceped constituent des **documents de travail** portant sur des recherches menées par des chercheurs du Ceped ou associés. Rédigés pour une diffusion rapide, ces papiers n'ont pas été formellement relus et édités. Certaines versions de ces documents de travail peuvent être soumises à une revue à comité de lecture. **Les droits d'auteur sont détenus par les auteurs.**

Ceped Working Papers are **working papers** on current research conducted by Ceped-affiliated researchers. Written for timely dissemination, these papers have not been formally edited or peer reviewed. Versions of these working papers are sometimes submitted for publication in peer-reviewed journals. **Copyrights are held by the authors.**

Enseigner R sur YouTube en temps de COVID-19 : retour d'expérience des webin-R

Joseph Larmarange^a

En contexte COVID-19, j'ai mis en place en novembre 2020 un webinaire hebdomadaire d'enseignement à R : les webin-R. Diffusées en direct sur YouTube, les sessions duraient de 1 h 30 à 2 h, avec partage d'écran. Les contenus reposaient sur analyse-R, un support de formation en ligne créé et maintenu depuis 2015.

Mi-mai 2021, une enquête a été menée auprès des participant·es aux webin-R (120 questionnaires complétés). Les répondant·es étaient en majorité des hommes (64 %), âgés de 40 ans ou moins (62 %). La moitié (49 %) résidait hors de France. Ils/Elles étaient mastéran·es (20 %), doctorant·es (25 %), postdoctorant·es (3 %), ingénieur·es (26 %), chercheur·es (18 %). Les disciplines étaient tout aussi variées.

En termes de participation, nous avons observé une fatigue au cours du temps, et surtout une baisse du suivi en direct au profit du différé.

Les participant·es ont souligné l'importance du visionnage en différé, la mise à disposition des scripts, le chapitrage des vidéos, l'articulation avec analyse-R, le partage d'écran, la langue française et la gratuité des contenus. Certains regrettent l'absence d'exercices ou la non-possibilité de valider des crédits d'enseignement.

L'expérience des webin-R traduit une forte demande pour des contenus d'autoformation à R, avec une visée appliquée, en français, et facilement accessibles. Le format libre participe au développement d'une science ouverte.

Mots clés : enseignement statistique, R, YouTube, enquête en ligne, COVID-19.

CONTEXTE

R dans les sociales quantitatives

Né en 1993, le langage de programmation statistique **R** (R Core Team, 2024) a connu ces deux dernières décennies un essor important dans différents champs scientifiques. Si R a d'abord plutôt été utilisé en biologie, son usage s'est rapidement élargi à la biostatistique, l'épidémiologie et les sciences sociales quantitatives.

Le développement de nouveaux outils, en particulier **RStudio** (Posit team, 2024), **knitr** (Xie et al., 2024) et la suite de *packages* constituée par le **tidyverse** (Wickham & RStudio, 2023) vont rendre R plus accessible et lui faire gagner en popularité. De fait, R est aujourd'hui intégré à la majorité des cursus universitaires de sciences sociales à forte composante

^a Ceped (Centre Population et Développement, UMR 196, Université Paris Cité, Université Sorbonne Paris Nord, IRD, Inserm), France

quantitative, notamment en démographie, MASS (mathématiques appliquées aux sciences sociales), économie, sociologie quantitative (Abbasnasab Sardareh et al., 2021; Zeitlin & Auerbach, 2019).

La gratuité de R permet de le rendre accessible aux étudiants, y compris sur leur ordinateur personnel, sans que les universités aient à s'acquitter de licences onéreuses, comme c'est le cas avec la majorité des logiciels propriétaires (Stata, SPSS, SAS...). Cela est d'autant plus importants dans les pays à faible et moyen revenu où le budget des universités et les moyens financiers des étudiants sont limités. Par ailleurs, cela facilite la réutilisation du logiciel dans la vie professionnelle future des étudiants, notamment pour celles et ceux amenés à travailler dans des structures aux moyens limités (par exemple dans des organisations non gouvernementales).

R est un logiciel libre au code ouvert et, à ce titre, offre plus de transparence que les logiciels propriétaires et facilite la reproductibilité des analyses, un enjeu majeur des statistiques modernes (Wilcox, 2017), notamment dans le contexte d'une demande croissante pour une science ouverte.

Enfin, ce logiciel étant utilisé dans des domaines scientifiques diversifiés, R facilite l'interdisciplinarité et la réalisation d'analyses conjointes par des chercheur-es de disciplines différentes.

Parcours personnel avec R

À titre personnel, j'ai fait la rencontre de R en 2005, pendant ma thèse de doctorat en démographie, notamment pour créer de nouveaux outils de lissage spatial. Cela m'a permis de développer en 2007 mon tout premier package R : **prevR** (Larmarange, 2023). Dans mon parcours académique, j'ai été confronté à une demande croissante de formations sur R ce qui m'a amené à écrire progressivement plusieurs supports de cours sur des sujets spécifiques.

En 2013, Julien Barnier a mis en ligne une *Introduction à R*¹ à destination des étudiants en sciences sociales et plus généralement d'un public peu familier avec l'informatique et l'écriture de code informatique. Ce document abordait en particulier l'installation, la prise en main de l'outil, le recodage de variables et la statistique uni- et bivariable. Il s'agissait

¹ <https://github.com/juba/intro-r> Cette introduction à R sera maintenue jusqu'en 2017.

d'un document PDF généré avec **Sweave**², un format permettant d'intégrer du code R au sein d'un document LaTeX.

Surtout, J. Barnier a déposé son introduction sur la plateforme **GitHub** sous licence *Creative Commons*³, permettant ainsi à d'autres personnes de profiter de son travail et de le faire évoluer. Ainsi, en octobre 2013, j'ai pu partir de son document⁴ pour le compléter selon mes besoins et développer une *Introduction à l'analyse d'enquêtes avec R*⁵ en y ajoutant une série d'encadrés sur l'interface de RStudio et des chapitres consacrés à la régression logistique, à l'analyse factorielle, à la classification ascendante hiérarchique ou encore à l'analyse de séquences. Mes besoins étant différents de ceux de J. Barnier, seuls quelques éléments (notamment sur RStudio) ont ensuite été intégrés au document original de J. Barnier⁶.

Le format Sweave étant peu adapté aux gros documents et l'export PDF de l'introduction à l'analyse d'enquêtes dépassant les 200 pages, j'ai décidé en mai 2015 de transformer ces contenus sous la forme d'un site web et de créer *analyse-R : introduction à l'analyse d'enquêtes avec R et RStudio*⁷, toujours sous licence *Creative Commons*. analyse-R est écrit en **Rmarkdown** (J. J. Allaire et al., 2024) un format permettant d'encapsuler du code R au sein d'un fichier *markdown*. Il s'agit d'un format plus souple : chaque chapitre peut se lire indépendamment et les différents chapitres ne se suivent pas nécessairement. Le site a été maintenu activement de 2015 à 2022 et a ainsi évolué par l'ajout de différents chapitres ou la mise à jour de chapitres existants. Certains chapitres sont des adaptations, avec l'accord de leurs auteurs, de supports de cours écrits initialement par d'autres personnes⁸. Le développement d'analyse-R a été relativement important puisque la version PDF compte aujourd'hui environ 1400 pages. Du fait de son libre accès au format HTML, analyse-R a rapidement été bien référencé dans les moteurs de recherche et générait, en 2020, de 40 à 50 000 vues mensuelles. Pour ma part, analyse-R était devenu mon principal support de cours, les différentes formations que j'animais se

² <https://en.wikipedia.org/wiki/Sweave>

³ CC-BY-NC-SA, <https://creativecommons.org/licenses/by-nc-sa/3.0/fr/>

⁴ On parle d'un *fork* sur GitHub.

⁵ <https://github.com/larmarange/intro-r/tree/CoursM2>

⁶ <https://github.com/juba/intro-r/pull/5>

⁷ <https://larmarange.github.io/analyse-R/>, doi : 10.5281/zenodo.2081968

⁸ En particulier, analyse-R contient des contributions de Julien Barnier, Julien BiauDET, François Briatte, Milan Bouchet-Valat, Ewen Gallic, Frédérique Giraud, Joël Gombin, Mayeul Kauffmann, Christophe Lalanne et Nicolas Robette.

reposant sur une sélection de chapitres en fonction de la thématique de chaque session. En parallèle de ma pratique d'enseignement à R, j'ai également contribué à ou développé plusieurs packages R⁹.

À l'origine des webin-R : le second confinement lié à l'épidémie de Covid-19

Pour les années universitaires 2019-2020 et 2020-2021, je devais initialement animer plusieurs séances de formation intensive à R (de 3 à 5 jours) pour des publics variés (mastérent-es, doctorant-es, postdoctorant-es, ingénieur-es, enseignant-es-chercheur-es et chercheur-es) et dans des contextes divers (master de l'école des Hautes Études en Démographie, réseaux doctoraux, formations continues de l'Institut de Recherche pour le Développement, etc.), en France et en Afrique de l'Ouest. L'arrivée de l'épidémie de COVID-19 et les différentes mesures sanitaires mises en place au premier semestre 2020 ont fortement perturbé (reports, annulations) les formations dispensées au titre de l'année 2019-2020. L'annonce d'un second confinement¹⁰ en octobre 2020 m'a amené à repenser mes modalités d'intervention : au lieu de maintenir des sessions de formation intensive en présentiel, l'idée a été de les remplacer par un webinaire hebdomadaire en ligne ouvert à toutes et à tous : les *webin-R*.

Les webin-R n'ont pas pour vocation d'enseigner les statistiques proprement dites mais plutôt comment réaliser des statistiques avec les logiciels R et RStudio. Ils se focalisent sur la mise en œuvre pratique, l'objectif étant de donner les clés de départ pour permettre aux participants de pouvoir continuer à s'autoformer par la suite.

Cet article vise à partager les aspects opérationnels concernant la mise en place des webin-R mais également à partager les résultats d'une petite enquête réalisées auprès des participantes et des participants en mai et juin 2021, avant de discuter les leçons apprises de cette expérimentation.

⁹ Tels que **questionr** (Barnier et al., 2023), **labelled** (Larmarange et al., 2024), **broom.helpers** (Larmarange & Sjoberg, 2024), **GGally** (Schloerke et al., 2024), **ggstats** (Larmarange, 2024), **gtsummary** (Sjoberg et al., 2024), **lubridate**, etc.

¹⁰ Après un premier confinement du 17 mars au 11 mai 2020, la France connaîtra deux autres confinements, du 30 octobre au 15 décembre 2020 et du 3 avril au 3 mai 2021.

MISE EN PLACE DES WEBIN-R

La décision de mettre en place les webin-R s'est prise au cours de la première semaine du second confinement. Les deux premières étapes ont consisté, début novembre 2020, à mettre en place une liste de diffusion dédiée et un site web d'information.

Une ébauche de **site web** est créée le 6 novembre 2020. Le site des webin-R est hébergé sur la plateforme GitHub¹¹ et repose sur le package **distill** (Dervieux et al., 2023). Un préprogramme des dix premières séances est publié le 9 novembre. Les webin-R sont prévus tous les jeudis de 17 h à 19 h avec une première séance annoncée pour le 26 novembre 2020.

Une **liste de diffusion** est créée sur la plateforme de l'IRD le 7 novembre 2020 et validée par les services informatiques le 9 novembre. Le même jour, les modalités d'inscription à la liste de diffusion et le préprogramme sont diffusés par courriel via différents canaux et listes de diffusion. Entre le 9 et le 20 novembre, près de 400 personnes se sont inscrites à cette liste de diffusion¹².

Reste à identifier une **plateforme de diffusion**. La première idée consistait à réaliser les webin-R sur **Zoom**¹³, dans la mesure où mon université y était abonnée. Cependant, la limite applicable au nombre de participants à une réunion était de 100 personnes, soit en dessous de ce que laissait présager la participation au regard du nombre d'inscrits à la liste de diffusion. Par ailleurs, l'abonnement de l'université ne couvrait pas les fonctionnalités du type webinaire qui permettent une meilleure gestion des participant-es (par exemple en bloquant leur flux vidéo ou audio). Une plateforme comme **Canal-U**¹⁴ aurait, pour sa part, été adaptée pour la rediffusion des webin-R après le direct, mais ne permettait pas de diffuser en direct. Par ailleurs, les serveurs de Canal-U étant localisés principalement en France, la qualité de la connexion peut être mauvaise pour des personnes localisées sur un autre continent. Or, de nombreuses personnes inscrites à la liste de diffusion résidaient en Afrique de l'Ouest. En dernier ressort, une semaine avant le premier webin-R, il fut décidé de les diffuser sur **YouTube** et une chaîne dédiée

¹¹ <https://larmarange.github.io/webin-R/>

¹² Mi-septembre 2024, le nombre d'inscrits s'élevait à 936 personnes.

¹³ <https://zoom.us/>

¹⁴ <https://www.canal-u.tv/>

est créée¹⁵ : la diffusion est gratuite ; la plateforme dispose d'une bande-passante élevée pour la retransmission à l'international ; la plateforme propose un outil de discussion (*chat*) en temps réel ; les vidéos diffusées en direct sont de suite disponibles en rediffusion ; une licence *Creative Commons* est disponible permettant de réutiliser les vidéos dans d'autres contextes et sur d'autres plateforme.

Une fois une plateforme de d'hébergement identifiée, nous devons avoir un **logiciel de diffusion en continu** (*streaming*) pour capter les flux vidéo (caméra et capture d'écran) puis les envoyer vers YouTube. Il existe plusieurs outils libres et gratuits, tels que **OBS Studio**¹⁶, mais qui requièrent une carte graphique avancée dont je ne disposais pas sur mon ordinateur portable professionnel. Il est possible de diffuser sur YouTube une visioconférence Zoom mais avec une qualité d'images limitée à 720p, l'ajout d'un logo Zoom (*watermark*) et à condition que l'option soit activée (ce qui n'était pas le cas pour la licence de l'université). Reste des outils en ligne dédiés tels que **StreamYard**¹⁷ ou **Restream**¹⁸. En novembre 2020, ces deux plateformes proposaient une offre gratuite mais avec un flux limité en 720p et l'ajout d'un logo (*watermark*). Après discussion, la direction du Ceped (mon unité de recherche) a souscrit un abonnement payant à StreamYard (468 USD par an), permettant une diffusion en 1080p (améliorant grandement la lisibilité des captures d'écran), d'effectuer un habillage graphique et un montage simple en temps réel, ainsi que le suivi en temps réel du forum de discussion YouTube (*chat*) et la possibilité d'incruster à l'image des commentaires choisis.

La première séance des webin-R sera diffusée comme prévue le 26 novembre 2020 et ils seront ensuite diffusés toutes les semaines (hors vacances scolaires) jusqu'au 24 juin 2021, soient 24 séances d'une durée d'une heure et demie à deux heures. Cela correspond à 21 webin-R numérotés (de #01 à #21) diffusés en direct sur YouTube plus trois sessions de type FAQ (foire aux questions) réalisées sur Zoom (pour permettre une plus grande interactivité avec les participants) puis déposées sur YouTube.

¹⁵ <https://www.youtube.com/@webin-R>

¹⁶ <https://obsproject.com/>

¹⁷ <https://streamyard.com/>

¹⁸ <https://restream.io/>

En termes d'organisation :

- › le programme prévisionnel est affiché et actualisé sur le site dédié ;
- › les séances sont annoncées via la liste de diffusion dédiée et via YouTube (ce qui notifie les abonnés à la chaîne) ;
- › après diffusion, les séances sont chapitrées sur YouTube pour faciliter la navigation et un lien est ajouté dans la description de la vidéo vers le site web dédié ;
- › sur le site dédié, la vidéo de chaque webin-R est intégrée et sont ajoutées différentes ressources : données des exemples, scripts R de la séance, liens vers les chapitres analyse-R correspondant aux contenus évoqués dans la séance ;
- › les vidéos des webin-R sont ensuite intégrées dans les chapitres analyse-R correspondants.

La participation aux différentes sessions a évolué au cours du temps : lors de la toute première session le 26 novembre 2020, en plein confinement, le nombre de spectateurs simultanés était en moyenne de 144 personnes avec un pic à 172 ; lors de la séance du 14 janvier 2021, la moyenne n'était plus que de 43 personnes avec un pic à 57 ; lors de la séance du 29 avril 2021, une moyenne de 24 avec un pic à 28 ; et le 24 juin 2021, une moyenne de 3 avec un pic à 6. Plusieurs pistes peuvent être avancées : une certaine fatigue parmi les participantes et les participants ; des thématiques plus avancées pour les dernières vidéos, intéressant moins de monde, par rapport aux premières séances destinées plutôt à un public débutant ; la fin du confinement et, enfin, une tendance accrue à regarder les vidéos en différé plutôt qu'en direct. Ainsi, la vidéo du 24 juin 2021, malgré son pic de spectateurs simultanés à seulement 6 personnes, a enregistré 179 vues dans le mois suivant sa diffusion.

Au 30 juin 2021, la chaîne YouTube comptait 24 vidéos, 822 abonnés et totalisait 22 400 vues.

ENQUÊTE AUPRÈS DES PARTICIPANTES ET PARTICIPANTS

Collecte

Mi-mai 2021, une petite enquête a été menée auprès des participants et des participantes aux webin-R. Un court questionnaire a été développé pour une passation anonyme en ligne sur la plateforme **LimeSurvey** (Limesurvey GmbH, 2024) du Ceped¹⁹. La promotion de l'enquête a été effectuée sur la liste de diffusion des webin-R qui comptait alors environ 770 abonnés.

Le questionnaire, les données et le script R **quarto** (J. Allaire et al., 2024) permettant de reproduire les tables et figures présentées ci-après sont librement disponibles sur GitHub²⁰.

Profil des participant-es

Au total, 120 questionnaires complets ont été collectés entre le 17 mai et le 10 juin 2021, correspondant à un taux de participation de l'ordre de 16 % (120/770).

Les principales caractéristiques de l'échantillon sont rapportées dans la Table 1. Les personnes ayant participé à l'enquête sont principalement des hommes (64 % vs 36 %), âgés de moins de 40 ans (62 %). Sans surprise, la distribution par âge est en cohérence avec la situation académique : 20 % sont mastéran·tes, 25 % doctoran·tes, 3 % postdoctoran·tes. Les webin-R ont également attiré des personnes déjà en poste : 26 % sont ingénieur·es d'étude ou de recherche, 18 % chercheur·es ou enseignant·es-chercheur·es. Dans la catégorie « autres », on peut notamment dénombrer 6 médecins (5 %).

La répartition par discipline traduit mes propres champs scientifiques (je suis démographe de santé publique), les réseaux dans lesquels je m'inscris (et où la communication initiale à propos des webin-R a été réalisée) et mon institution d'appartenance (l'IRD). Cependant, le profil disciplinaire des participant·es reste varié : les principales disciplines représentées sont la santé publique, la démographie, la sociologie et l'épidémiologie, suivies par l'économie, la statistique, l'écologie et la géographie. Certaines disciplines ne sont représentées que par une ou

¹⁹ <https://enquetes.ceped.org/>

²⁰ <https://github.com/larmarange/enquete-webin-R/>

deux personnes : sciences de l'éducation, anthropologie, sciences politiques, psychologie, psychiatrie, géosciences, météorologie, biologie, neurosciences, sciences de l'alimentation.

La majorité (82 %) des participant·es à l'enquête sont des utilisatrices et des utilisateurs réguliers du site analyse-R, 13 % déclarent le connaître sans l'utiliser et seulement 5 % ne le connaissent pas. Enfin, les webin-R ont attiré une forte population étrangère : 52 % des participant·es à l'enquête sont né·es à l'étranger et 49 % résident à l'étranger au moment de l'enquête, principalement dans des pays d'Afrique francophone.

Table 1. Profil des participant·es (n = 120)

Groupe d'âges	
24-29	25,0 % (30)
30-39	36,7 % (44)
40-49	22,5 % (27)
50 et plus	15,8 % (19)
Genre	
femme	35,8 % (43)
homme	64,2 % (77)
Lieu de naissance	
France	47,5 % (57)
Afrique	47,5 % (57)
autre	5,0 % (6)
Lieu de résidence	
France	60,8 % (73)
Afrique	35,8 % (43)
autre	3,3 % (4)
Statut académique	
étudiant·e en M1 (master première année)	4,2 % (5)
étudiant·e en M2 (master deuxième année)	15,8 % (19)
doctorant·e	25,0 % (30)
postdoctorant·e	3,3 % (4)
ingénieur·e d'études	15,0 % (18)
ingénieur·e de recherche	10,8 % (13)
enseignant·e-chercheur·e (MCF ou professeur)	10,0 % (12)
chercheur·e (CR ou DR)	7,5 % (9)
autre	8,3 % (10)

Discipline	
santé publique	17,5 % (21)
démographie	14,2 % (17)
épidémiologie	13,3 % (16)
sociologie	13,3 % (16)
économie	7,5 % (9)
statistique	7,5 % (9)
géographie	4,2 % (5)
écologie	3,3 % (4)
sciences de l'éducation	1,7 % (2)
anthropologie	0,8 % (1)
sciences politiques	0,8 % (1)
autre	15,8 % (19)
Connaissance / Utilisation du site analyse-R	
non	5,0 % (6)
oui, et je le visite régulièrement	81,7 % (98)
oui, mais je ne l'utilise pas	13,3 % (16)

Univers : participant-es à l'enquête webin-R (n=120).

Pour réaliser des tableaux statistiques univariés avec gtsummary (Sjoberg et al., 2021, 2024), voir <https://larmarange.github.io/guide-R/analyses/statistique-univariee.html>.

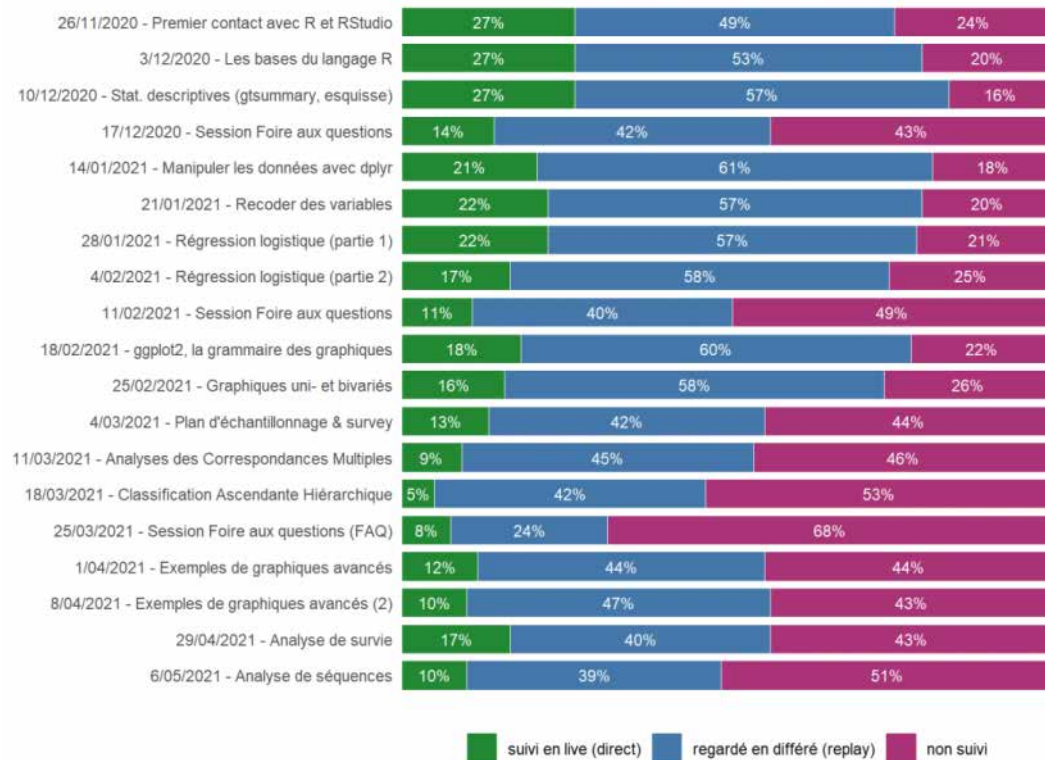
Mode de suivi des webin-R

Il a été demandé à chaque participant-e et pour chaque webin-R si ce webin-R avait été suivi et, le cas échéant, en direct ou en différé (Figure 1).

Les premières sessions sont celles qui ont été les plus suivies : de la présentation des outils à la régression logistique en passant par les statistiques uni- et bivariées et la réalisation de graphiques avec **ggplot2** (Wickham et al., 2024). Le suivi chute fortement pour les sessions suivantes abordant des notions statistiques plus complexes : données pondérées et plan d'échantillonnage complexe, analyses des correspondances multiples, classification ascendante hiérarchique, analyse de survie, analyse de séquences. Les trois sessions de type FAQ sont celles ayant été le moins suivies. Nous retrouvons la baisse croissante du suivi des séances en direct, comme suggérée dans la section précédente à partir du nombre moyen de spectateur-es simultanés dans les statistiques YouTube.

Le mode de suivi majoritaire parmi les participant-es à l'enquête reste la consultation en différé (*replay*) qui représente 40 à 60 % des suivis pour les sessions ordinaires (hors FAQ).

Figure 1. Mode suivi des webin R du 26 novembre 2020 au 6 mai 2021



Univers : participant-es à l'enquête webin-R (n=120).

La palette de couleur choisie (« *bright* ») repose sur les travaux de Paul Tol (Tol, 2021) et est adaptée aux personnes daltoniennes. Les palettes développées par P. Tol sont disponibles, entre autres, dans le package **khroma** (Frerebeau et al., 2024).

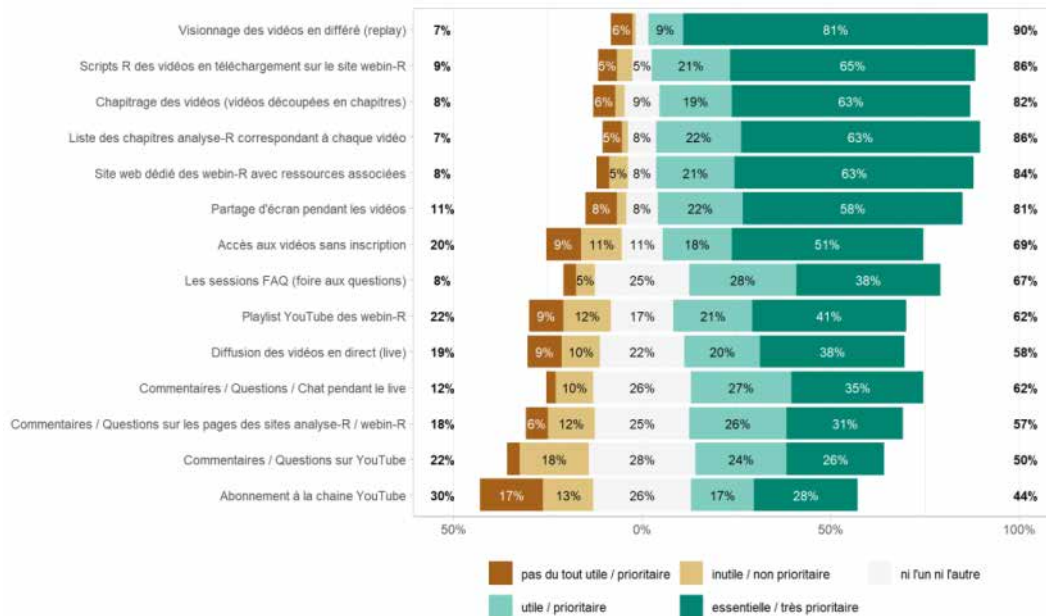
Pour réaliser graphiques en barres, voir <https://larmarange.github.io/guide-R/analyses/statistique-bivariee.html> et notamment la fonction `stat_prop()` du package **ggstats** (Larmarange, 2024) pour les étiquettes représentant les pourcentages.

Fonctionnalités les plus importantes

Les participant-es à l'enquête étaient invités à noter différentes fonctionnalités sur une échelle de Likert à cinq valeurs allant de « inutile / non prioritaire » à « essentielle / prioritaire ». Les résultats sont présentés sur la Figure 2 dont les modalités ont été ordonnées selon la mention majoritaire (Balinski & Laraki, 2007), la méthode des groupes insatisfaits étant utilisée pour départager les égalités (Fabre, 2021). La présentation des résultats est complétée avec des verbatims issues des

commentaires libres des questionnaires, ainsi que ceux d'une pré-enquête réalisée auprès de la même population en février-mars 2021.

Figure 2. Fonctionnalités les plus importantes selon les participant·es à l'enquête



Univers : participant·es à l'enquête webin-R (n=120).

Pour la réalisation de graphiques centrés en barres, appelés communément graphiques de type Likert, voir <https://larmarange.github.io/guide-R/analyses/likert.html> et notamment la fonction `gglikert()` du package `ggstats` (Larmarange, 2024).

La première fonctionnalité plébiscitée par les participant·es à l'enquête est la possibilité de pouvoir visionner les vidéos en différé et de les consulter à la demande.

Le choix direct versus replay est idéal car il n'est pas toujours possible d'être libre lors du créneau dédié au live.

(chercheur en écologie)

Le replay permet de faire avec son agenda, de suivre plus confortablement et en ayant le temps nécessaire de réécrire soi-même les commandes.

(médecin de santé publique)

C'est important d'avoir les sessions conservées pour pouvoir les visionner plusieurs semaines/mois après leur diffusion.

(ingénieure de recherche en géosciences)

J'ai eu pas mal de soucis qui ne m'ont plus permis de continuer à suivre chaque jeudi. Mais les replays me permettent de me remettre à jour.

(mastérent en épidémiologie)

Les vidéos stockées sur YouTube sont une idée fantastique, car cela peut atteindre beaucoup de personnes en dehors du webin-R.

(postdoctorante en anthropologie)

Viennent ensuite la mise à disposition des ressources (comme les scripts des séances) et l'articulation avec analyse-R.

La mise à disposition des scripts permet de rafraichir la mémoire surtout pour refaire des exercices après.

(doctorant en nutrition)

C'est une bonne ressource en complément d'analyse-R.

(ingénieur de recherche en santé publique)

Le découpage précis des vidéos en chapitres permet de retrouver l'information recherchée sans regarder la totalité de la vidéo, qui peut parfois être un peu longue.

(enseignante-chercheuse en épidémiologie)

Le partage d'écran pendant les vidéos est également fortement apprécié. Plusieurs participant-es y reviennent dans leurs commentaires libres, soulignant comment la vidéo, à l'inverse d'un contenu statique, permet de « voir R », de visualiser la non-linéarité du processus d'écriture du code.

Les manipulations en direct sont un point fort.

(enseignante-chercheuse en sciences de l'éducation)

L'apprentissage « par-dessus l'épaule », c'est-à-dire en commentant en direct les lignes de codes que l'on peut retrouver sur le net par ailleurs.

(enseignante-chercheuse en sociologie)

Le libre accès gratuitement aux contenus et aux vidéos, sans inscription préalable, est un élément essentiel pour près de 70 % des répondants.

C'est bien que les vidéos soient toujours disponibles, en accès libre, le fait qu'on puisse y revenir dès qu'on en a besoin.

(doctorante en sociologie)

Point fort : aspect gratuit et accès libre aux différents matériels de la formation.

(ingénieur de recherche statisticien)

Merci pour cette opportunité que vous nous donnez d'apprendre gratuitement.

(doctorant en économie)

Les sessions FAQ, moins suivies que les autres webin-R, sont « ni utiles ni inutiles » pour un quart des répondants. Néanmoins, pour celles et ceux qui ont participé, elles répondent à des besoins, en particulier pour les participant·es résidant à l'étranger (85 % d'opinion favorable contre 55 %).

Surtout lors des FAQ où il y a de vrais échanges entre les auditeurs et le formateur.

(chercheur en santé publique)

Les foires aux questions nous permettent de mieux comprendre.

(doctorant en géographie)

Les autres fonctionnalités, telles que la diffusion en direct, les différents espaces de commentaires, l'existence d'une liste de lecture sur YouTube ou encore la possibilité de s'abonner à la chaîne, sont moins plébiscitées. Pour autant, même si une majorité consulte les webin-R en différé, plusieurs ont souligné le dynamisme apporté par la diffusion en directe.

La diffusion en direct amène un dynamisme et une régularité qui à mon sens font aussi le succès de cette excellente initiative, en plus du contenu brillamment présenté.

(ingénieur d'études en sociologie)

Le format live avec d'éventuelles apartés rend le contenu plus vivant. La possibilité de poser des questions en live et les foires aux questions dynamise également le webin-R.

(ingénieur de recherche en santé publique)

Même si suivi en différé, les réponses aux questions au live restent intéressantes.

(mastérante en épidémiologie)

Dans les commentaires libres, plusieurs participant-es reviennent sur la complémentarité entre les webin-R et analyse-R.

On pourrait penser que la consultation d'analyse-R est suffisante pour tirer parti de ce cours. Il n'en est rien, car les webin-R permettent de : prendre aisément contact avec la matière qu'il est ensuite possible de revoir et approfondir au moyen d'analyse-R ; se familiariser avec une foule de détails relativement à l'utilisation de RStudio fournis implicitement et explicitement durant les présentations.

(enseignant-chercheur en démographie)

Enfin, la question de la langue française (en opposition à l'anglais) est revenue plusieurs fois.

Autre point très fort : c'est la langue. Pour moi, le fait que toute la documentation (sur le site) et tous les webin-R soient en français m'a permis d'entrer beaucoup plus facilement dans l'apprentissage du logiciel.

(enseignante-chercheuse en démographie)

Attentes exprimées et points faibles

Les participant-es à l'enquête étaient également invité-es à partager, sous forme de commentaires libres, des points d'amélioration et des thématiques à aborder.

En matière de thématiques, les réponses sont très hétérogènes, reflétant des besoins spécifiques à chaque discipline (et souvent en dehors du champ de compétences du formateur), des éléments en-dehors du domaine des webin-R (e.g. des cours théoriques en statistique, la construction de l'objet de recherche, etc.), mais également, selon le niveau de maîtrise du logiciel, plus de contenus pour débutant-es ou, au contraire, plus de contenus avancés.

Une demande récurrente a été de proposer, en complément des webin-R, des exercices pratiques avec leurs corrigés.

Il faudrait beaucoup plus d'exercices pour nous améliorer.

(mastérante en démographie)

Proposer de petits exercices corrigés à réaliser à la fin de la séance pour s'entraîner directement ?

(enseignante-chercheuse en sociologie)

Proposer des exercices, 15 jours à 3 semaines à l'avance, qui pourraient être corrigés lors d'une séance ?

(chargée de projet en épidémiologie)

Je comprends qu'il soit difficile d'envisager une partie de mise en pratique, à cause du format des webin-R, mais c'est un élément qui me manque. Cela est toutefois dû à ma situation particulière. Je n'ai pas encore mes données, j'apprends donc avant de mettre en pratique, et les choses peuvent parfois me sembler abstraites. Il est bien sûr possible de pratiquer en dehors des webin-R, avec les bases de données fournies par certains packages, mais la raison pour laquelle je préfère regarder les vidéos en live est que je préfère le fait d'être accompagnée en apprenant. C'est pour cela que j'ai du mal à reprendre seule les vidéos.

(doctorante en sciences de l'éducation)

La durée des vidéos a été mentionnée à plusieurs reprises dans les pistes d'amélioration. Si pour certaines et certains, le format d'une heure et demie à deux heures semblait adapté, d'autres auraient préférés des vidéos plus courtes. Il semble que cette perception soit liée au mode de suivi des webin-R. Si le format proposé était plus adapté à une session en direct, c'est moins le cas pour un visionnage en différé où des vidéos plus courtes et plus centrées sur un sujet précis semblent plus appréciés.

J'ai apprécié la durée « normale » des vidéos : pas trop longues, pas trop courtes.

(doctorant en démographie)

Réduire un peu la durée : se fixer par exemple une heure et quart maximum.

(médecin)

Je me demandais si une heure ne serait pas une durée optimale.

(chercheur en écologie)

Peut-être réduire la durée (et encore, pour les premiers ça allait). Le contenu est parfois dense et demande du temps pour pénétrer et de la réactivité.

(doctorante en économie)

Par ailleurs, j'ai été contacté par plusieurs doctorant-es pour me demander s'il leur était possible de valider des crédits d'enseignement au titre de leur participation au webin-R. Cependant, les règles de chaque école doctorale étant spécifique et le format / la plateforme ne permettant pas de contrôler la participation effective, cette possibilité n'a pas pu être offerte aux apprenant-es.

DISCUSSION, LEÇONS APPRISSES ET PERSPECTIVES

L'expérience des webin-R traduit une forte demande pour des contenus d'autoformation à R, avec une visée appliquée, en français, et facilement accessibles gratuitement.

Les webin-R se sont mis en place dans un contexte social particulier : l'épidémie de COVID-19 et les différentes mesures sanitaires qui en ont découlé telles que les confinements. De nombreuses activités réalisées jusque-là en présentiel ont dû s'adapter et la mise en place d'un webinaire hebdomadaire en direct a rapidement suscité de l'intérêt et un engouement.

Cela a été rendu possible, dans un laps de temps relativement court, grâce à l'existence d'outils de captation et de diffusion relativement simples, l'appui financier de mon unité de recherche (via l'acquisition de licences logiciels), mais aussi et surtout à la préexistence de la plateforme analyse-R. En effet, le fait de disposer de supports de cours déjà développés a été essentiel : cela a permis de proposer rapidement un programme de formation et de disposer de suffisamment de contenus pour alimenter un webinaire hebdomadaire. À l'inverse, certaines demandes des participant-es n'ont pu être satisfaites, notamment la création et la mise en ligne d'exercices corrigés. Ces derniers n'existent pas sur analyse-R et leur développement aurait été particulièrement chronophage, d'autant plus que l'écriture d'analyse-R ou l'organisation des webin-R est une activité additionnelle à ma charge de travail courante en tant que chercheur.

Un des résultats inattendus de l'enquête est le nombre de participant-es déjà dans une position académique (ingénieur-es, enseignant-es-chercheur-es, chercheur-es). Cela reflète les besoins de formation continue à R ainsi que le développement croissant de cours sur R dans les cursus universitaires nécessitant que les enseignant-es se mettent à jour.

Un grand merci ! J'ai pu m'initier à R tout en proposant un nouvel enseignement à des étudiantes de L3 de sociologie sur deux semestres. J'avais un peu peur qu'ils et elles rencontrent des difficultés du fait de l'utilisation d'un langage de programmation. Mais j'y suis allée très progressivement et lentement en m'inspirant de la documentation d'analyse-R tout en la personnalisant et en travaillant à partir d'autres bases de données.

(enseignante-chercheuse en démographie)

Le découpage par thème, le début « en douceur » et la montée progressive vers des notions plus complexes, avec en permanence des retours sur des choses que l'on a vues précédemment ainsi que

le déroulement en direct, comme si on faisait nous-même l'analyse d'un fichier de données, tout cela fait de ce webin-R un outil indispensable. Ce webin-R [permet] à des néophytes de se familiariser avec un outil qui n'est pas évident au premier abord. J'ai pourtant essayé de nombreuses fois de me mettre à R mais cela n'a jamais été concluant. Ce webin-R me permet de m'accrocher et j'y reviens souvent ainsi que sur le site analyse-R.

(ingénieure statisticienne)

Un grand merci pour ces webin-R qui me permettent de passer progressivement à R après plus de 25 ans de pratique de SAS.

(ingénieur de recherche en épidémiologie)

Les webin-R, mais également analyse-R, sont directement accessibles en ligne, sans inscription. Cela facilite la participation en supprimant les étapes liées à une inscription. Mais surtout cela rend l'ensemble des contenus visibles et indexables par les moteurs de recherche, générant des visites et amenant de nouveaux visiteurs. Ainsi, malgré une activité beaucoup plus faible depuis juin 2021, la chaîne des webin-R continue de générer 2 500 à 3 000 vues par mois (avec des creux entre juillet et septembre) et voit son nombre d'abonnés augmenter progressivement.

Le choix d'une licence libre, en l'occurrence *Creative Commons*, est également un élément crucial à ne pas négliger. C'est bien parce que J. Barnier a choisi ce type de licence en 2013 pour son *Introduction à R* que j'ai pu à l'époque repartir librement de son travail pour en développer ma propre version qui, de fil en aiguille, est un des ancêtres des webin-R. Le même choix de licence pour les webin-R permet de ne pas les enfermer sur leur plateforme de diffusion (ici YouTube) et de faciliter leur libre réutilisation. Ainsi, les webin-R ont pu être intégrés dans les chapitres d'analyse-R, ont fait l'objet d'un archivage sur la plateforme HAL²¹, ou ont été repris et intégrés à la plateforme d'*e-formation* de mon institut, l'IRD.

Toute personne est libre de réutiliser ce matériel en fonction de ses besoins. Personnellement, j'utilise les webin-R pour des formats d'enseignement hybrides, par exemple pour les étudiants de master suivant le parcours de l'école des Hautes Études en Démographie. Ces étudiants viennent de plusieurs universités, partout en France, et suivent, en plus de leurs cours de master, un parcours additionnel comportant, en première année, une mise à niveau à R. L'école organise à plusieurs moments de l'année des ateliers en présentiel dont la durée est limitée. Nous avons dès lors opté pour une formule hybride : les étudiants sont

²¹ cf. <https://cv.hal.science/joseph-larmarange>

invités à suivre en amont une partie des webin-R, à leur rythme, puis nous organisons un cours collectif en présentiel. Ce cours peut *de facto* aller plus loin dans les notions abordées du fait que plusieurs éléments ont déjà été entrevus par les étudiant-es en distanciel.

Le format des webin-R, à savoir un webinaire hebdomadaire en direct, repose sur un engagement important, tant en termes de temps que d'énergie, du formateur et des apprenant-es, rendu d'une certaine manière possible par le contexte de mise en place (second confinement en France). Comme nous l'avons déjà mentionné, une fatigue s'est progressivement instaurée avec de moins en moins de participants jusqu'à la fin de la première vague de webin-R en juin 2021. Cela est peut-être lié en partie à la fin du confinement et à une baisse du temps disponible. En décembre 2021, j'ai proposé deux nouvelles sessions en direct. Si l'annonce sur la liste de diffusion a été suivie de retours spontanés positifs d'ancien-nes participant-es, la participation au direct est restée limitée : 31 personnes en moyenne pour la première vidéo, 18 pour la seconde. Les vues ont augmenté par la suite, de manière continue, pour atteindre en septembre 2024 respectivement plus de 3 800 et 2 500 vues.

Cela m'a amené à repenser le format des futures vidéos. D'une part, mon énergie et mes disponibilités étant limitées, revenir sur des thématiques déjà abordées ne présente qu'un faible intérêt à titre individuel. D'autre part, je me suis lancé dans un autre chantier, à savoir *guide-R : guide pour l'analyse de données d'enquêtes avec R*²², une nouvelle version alternative à *analyse-R*. En effet, une partie des contenus d'*analyse-R* commence à être datés : de nouveaux packages et outils sont sortis, R intègre maintenant de manière native un *pipe* (un opérateur permettant d'enchaîner les appels de fonctions), etc. De plus, avec le temps, *analyse-R* a grandi de manière insuffisamment structurée, en souhaitant recenser différentes approches pour réaliser une même chose au risque de créer de la confusion. Bien qu'il reprenne une grande partie des contenus d'*analyse-R*, *guide-R* est plus structuré, avec des choix plus restreints sur les familles de package introduits. L'ensemble du code R est également actualisé. J'ai choisi de ne pas couvrir l'introduction à la prise en main du logiciel R et de RStudio dans la mesure où J. Barnier a développé une excellente *Introduction à R et au tidyverse*²³, suite depuis 2017 à sa précédente *Introduction à R*. Enfin, *guide-R* repose maintenant

²² <https://larmarange.github.io/guide-R/>

²³ <https://juba.github.io/tidyverse/>

sur **quarto** (J. Allaire et al., 2024) ce qui offre plus de fonctionnalités pour le site web.

En décembre 2023, j'ai tourné une nouvelle vidéo à propos des prédictions, contrastes et effets marginaux, une occasion d'aborder de nouveaux contenus, en lien avec un des nouveaux chapitres créés pour guide-R. Cette fois-ci, la vidéo n'a pas été diffusée en live, même si elle a été tournée dans les conditions d'un direct (ce qui évite de procéder à un montage post-tournage) puis déposée sur Youtube. Le format est un peu plus court : une heure. Le nombre de vues atteintes est similaire aux précédentes vidéos : 313 vues en 7 jours. Avec le recul, il me semble que, sur le long terme, il serait plus pertinent de développer des formats courts (20-30 minutes), sur des sujets plus précis et qui pourraient ainsi mieux s'intégrer à un site web comme guide-R.

Le besoin de « voir le code » exprimé par plusieurs participants m'a amené à expérimenter de nouveaux outils de visualisation, notamment les *flipbooks* (Reynolds, 2019; Reynolds et al., 2021) qui correspondent à des diaporamas permettant de visualiser chaque étape du code et ses conséquences sur le résultat final. J'en ai développé plusieurs pour guide-R²⁴.

De manière générale, l'expérience d'analyse-R, webin-R et guide-R met en lumière l'intérêt de développer des supports de formation génériques, complémentaires des enseignements spécifiques et disciplinaires ; gratuitement accessibles, pouvant être librement réutilisés dans des formats de cours variés et permettant aux individus de pouvoir continuer à s'autoformer. En effet, il est difficile pour les apprenant-es de retenir l'ensemble des notions et outils abordés lors d'une formation en présentiel ou d'un webinaire. Pouvoir revenir sur ces différents contenus, après coup, à son rythme, et en fonction de ses besoins est essentiel. C'est au moment où l'on est confronté à une question spécifique pendant ses analyses que l'on est le plus réceptif. Plus globalement, le recours à des licences libres permet de construire des contenus en partant d'éléments déjà construits par d'autres et participe pleinement au développement d'une science ouverte.

²⁴ Voir par exemple https://larmarange.github.io/guide-R/analyses/ressources/flipbook-geom_bar-dodge.html.

RÉFÉRENCES

- Abbasnasab Sardareh, S., Brown, G. T. L., & Denny, P. (2021). Comparing four contemporary statistical software tools for introductory data science and statistics in the social sciences. *Teaching Statistics*, 43(51), S157-S172. <https://doi.org/10.1111/test.12274>
- Allaire, J. J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., Iannone, R., Dunning, A., Yasumoto, A., Schloerke, B., Sievert, C., Ryan, D., Aust, F., Allen, J., Seo, J., ... Krewinkel, A. (2024). *rmarkdown* : Dynamic Documents for R (Version 2.28) [Logiciel]. <https://doi.org/10.32614/CRAN.package.rmarkdown>
- Allaire, J., Teague, C., Scheidegger, C., Xie, Y., & Derbieux, C. (2024). *Quarto* (Version 1.4) [Logiciel]. <https://doi.org/10.5281/zenodo.5960048>
- Balinski, M., & Laraki, R. (2007). A theory of measuring, electing, and ranking. *Proceedings of the National Academy of Sciences*, 104(21), 8720-8725. <https://doi.org/10.1073/pnas.0702634104>
- Barnier, J., Briatte, F., & Larmarange, J. (2023). *questionr* : Functions to Make Surveys Processing Easier (Version 0.7.8) [Logiciel]. <https://doi.org/10.32614/CRAN.package.questionr>
- Dervieux, C., Allaire, J. J., Iannone, R., Hill, A. P., Xie, Y., Posit Software PBC, Google LLC, Williams, N., Demchenko, D., The Polymer Authors, Csárdi, G., & Seo, J. (2023). *distill* : « R Markdown » Format for Scientific and Technical Writing (Version 1.6) [Logiciel]. <https://doi.org/10.32614/CRAN.package.distill>
- Fabre, A. (2021). Tie-breaking the highest median : Alternatives to the majority judgment. *Social Choice and Welfare*, 56(1), 101-124. <https://doi.org/10.1007/s00355-020-01269-9>
- Frerebeau, N., Lebrun, B., Arel-Bundock, V., Stervbo, U., Université Bordeaux Montaigne, & CNRS. (2024). *khroma* : Colour Schemes for Scientific Data Visualization (Version 1.14.0) [Logiciel]. <https://doi.org/10.32614/CRAN.package.khroma>
- Larmarange, J. (2023). *prevR* : Estimating Regional Trends of a Prevalence from a DHS and Similar Surveys (Version 5.0.0) [Logiciel]. <https://doi.org/10.32614/CRAN.package.prevR>
- Larmarange, J. (2024). *ggstats* : Extension to « ggplot2 » for Plotting Stats (Version 0.6.0) [Logiciel]. <https://doi.org/10.32614/CRAN.package.ggstats>
- Larmarange, J., Ludecke, D., Wickham, H., Bojanowski, M., & Briatte, F. (2024). *labelled* : Manipulating Labelled Data (Version 2.13.0) [Logiciel]. <https://doi.org/10.32614/CRAN.package.labelled>
- Larmarange, J., & Sjöberg, D. D. (2024). *broom.helpers* : Helpers for Model Coefficients Tibbles (Version 1.17.0) [Logiciel]. <https://doi.org/10.32614/CRAN.package.broom.helpers>
- Limesurvey GmbH. (2024). *LimeSurvey* : An Open Source survey tool (Version 6.4.4) [Logiciel]. Limesurvey GmbH. <http://www.limesurvey.org>
- Posit team. (2024). *RStudio* : Integrated Development Environment for R [Logiciel]. Posit Software, PBC. <https://www.posit.co/>
- R Core Team. (2024). *R* : A Language and Environment for Statistical Computing [Logiciel]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reynolds, E. (2019, août 8). *Flipbooks* : Presenting code and its output side-by-side and step-by-step. https://evamaerey.github.io/little_flipbooks_library/about/what_the_flipbook
- Reynolds, E., Aden-Buie, G., & Tanaka, E. (2021). *flipbookr* : Parses Code, Creates Partial Code Builds, Delivers Code Movie (Version 0.1.0) [Logiciel]. <https://doi.org/10.32614/CRAN.package.flipbookr>
- Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., Elberg, A., Toomet, O., Crowley, J., Hofmann, H., & Wickham, H. (2024). *GGally* : Extension to « ggplot2 » (Version 2.2.1) [Logiciel]. <https://doi.org/10.32614/CRAN.package.GGally>
- Sjöberg, D. D., Larmarange, J., Curry, M., Lavery, J., Whiting, K., Zabor, E. C., Bai, X., Drill, E., Flynn, J., Hannum, M., Lobaugh, S., Pileggi, S., Tin, A., & Wainberg, G. Z. (2024). *gtsummary* :

Presentation-Ready Data Summary and Analytic Result Tables (Version 2.0.2) [Logiciel].
<https://doi.org/10.32614/CRAN.package.gtsummary>

Sjoberg, D. D., Whiting, K., Curry, M., Lavery, J. A., & Larmarange, J. (2021). The R Journal : Reproducible Summary Tables with the gtsummary Package. *The R Journal*, 13(1), 570-580.
<https://doi.org/10.32614/RJ-2021-053>

Tol, P. (2021). Colour Scheme (Version 3.2). <https://personal.sron.nl/~pault/>

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., Brand, T. van den, & Posit PBC. (2024). *ggplot2 : Create Elegant Data Visualisations Using the Grammar of Graphics* (Version 3.5.1) [Logiciel].
<https://doi.org/10.32614/CRAN.package.ggplot2>

Wickham, H. & RStudio. (2023). *tidyverse : Easily Install and Load the « Tidyverse »* (Version 2.0.0) [Logiciel]. <https://doi.org/10.32614/CRAN.package.tidyverse>

Wilcox, R. (2017). *Modern Statistics for the Social and Behavioral Sciences : A Practical Introduction*, Second Edition (2e éd.). Chapman and Hall/CRC.
<https://doi.org/10.1201/9781315154480>

Xie, Y., Sarma, A., Vogt, A., Andrew, A., Zvoleff, A., Al-Zubaidi, A., Simon, A., Atkins, A., Wolen, A., Manton, A., Yasumoto, A., Baumer, B., Diggs, B., Zhang, B., Yapparov, B., Pereira, C., Dervieux, C., Hall, D., Hugh-Jones, D., ... Posit Software, PBC. (2024). *knitr : A General-Purpose Package for Dynamic Report Generation in R* (Version 1.48) [Logiciel].
<https://doi.org/10.32614/CRAN.package.knitr>

Zeitlin, W., & Auerbach, C. (2019). *Basic Statistics for the Behavioral and Social Sciences Using R*. Oxford University Press.