



HAL
open science

Initier la communauté de recherche à l'analyse d'enquêtes avec R et RStudio: retours d'expériences d'analyse-R et des webin-R

Joseph Larmarange

► To cite this version:

Joseph Larmarange. Initier la communauté de recherche à l'analyse d'enquêtes avec R et RStudio: retours d'expériences d'analyse-R et des webin-R. Semaine Data SHS 2021, CDSP; Plateforme Universitaire de Données des Grands-Moulins, Dec 2021, Paris, France. ird-04058396

HAL Id: ird-04058396

<https://ird.hal.science/ird-04058396>

Submitted on 4 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



webin-R

un séminaire sur l'analyse d'enquêtes avec R et RStudio

Initier la communauté de recherche à l'analyse d'enquêtes avec R et RStudio : retours d'expériences d'analyse-R et des webin-R

Joseph Larmarange

Semaine Data SHS

"Traiter et analyser les données en Sciences Humaines et Sociales"

Paris | 10 décembre 2021

avec l'appui de



D'où je parle ...

- › Démographe en santé publique
- › Chercheur à l'IRD (Ceped, IRD, Université de Paris, Inserm)
- › Terrains africains dans le champ du VIH et des maladies infectieuses

- › Utilisateur de R depuis une quinzaine d'années
- › Développement et contribution à plusieurs packages (prevR, questionr, labelled, broom.helpers, GGally, lubridate, gtsummary...)

- › Enseigne R depuis plus de dix ans
- › Public : master, doctorant·e·s, post-docs, chercheurs & EC
- › Format : principalement des stages intensifs de 2 à 5 jours pleins (sans TP)
- › Pas d'enseignement à la statistique mais à la réalisation de statistiques avec R
- › Présentation de la mise en œuvre pour donner les clés de départ (autoformation)

Est-ce pertinent d'enseigner R ?

- › R est un logiciel **libre** (*open science*, reproductibilité...)
- › R & R Studio sont des logiciels **gratuits** (accès pour les étudiants, réutilisation ...)
- › R est un logiciel de statistique **puissant**
- › R possède un **moteur graphique avancé**
- › R permet d'**automatiser** analyses et rapports
- › De très nombreuses **ressources** sont disponibles (notamment en anglais)

- › R est un langage de programmation avec un **coût d'entrée**
- › **Courbe d'apprentissage** difficile au début (il faut maîtriser les bases)...
- › ... alors que beaucoup se tournent vers R pour des **besoins avancés**

Faut-il enseigner R à des débutants ?

- › Il faut distinguer enseigner :
 - » la statistique comme discipline (connaissance mathématiques requises)
 - » la pratique des statistiques (savoir quoi utiliser et interpréter les résultats)
 - » l'utilisation d'un logiciel de statistique **ET** la manipulation de données

- › Pour des premières années qui découvrent la statistique :
 - » l'apprentissage du logiciel ne doit pas primer sur celui de la pratique des stats
 - » quel cursus majoritaire après la licence ? feront-ils plus que de la stat descriptive ?
 - » l'investissement dans R est-il « rentable » à moyen terme ?
 - » d'autres logiciels libres et gratuits sont disponibles :
 - > <http://trideux.cloud/>
 - > <https://www.cdc.gov/epiinfo/>
 - > <https://www.jamovi.org/> (pour un premier contact avec R ?)

À des utilisateurs avancés ?

- › R devient un **investissement rentable**
- › Il est essentiel de savoir travailler avec des **scripts** (quel que soit le logiciel)
- › Une connaissance préalable de la statistique permet de se focaliser sur l'outil

Des difficultés possibles :

- › Désapprendre ce que l'on maîtrise déjà avec un autre outil
 - » gérer la frustration, revenir aux bases, ne pas maîtriser de suite la stat avancée
- › Décalage temporel entre formation et application à ses propres données
 - » apprendre à s'autonomiser,
 - » acquérir les bases pour s'autoformer ensuite avec les ressources en ligne
- › Découvrir et savoir penser la manipulation/gestion de données

De nouveaux outils qui facilitent les choses...

› **R Studio**

- › une interface intégrée de développement sous R
- › intégration des rapports automatisés (R markdown, knitr, pandoc...)

› **Le tidyverse et le pipe**

- › penser/écrire le code en étapes logiques
- › des outils simples et avancés de manipulation de données

› **gtsummary**

- › des tableaux statistiques publiables en l'état
- › tris à plat, tableaux croisés, tableaux personnalisés, tests, intervalles de confiance
- › mise en forme des résultats de modèles (voir aussi *GGally::ggcoef_model()* et *broom.helpers*)

› **ggplot2**

- › une grammaire des graphiques qui incite à penser ses représentations visuelles
- › *esquisse* pour un premier contact et une exploration interactive des données

› des **ressources en ligne** : l'exemple d'analyse-R et des webin-R

analyse-R

analyse-R Manipuler ▾ Analyser ▾ Approfondir ▾ Index ▾ PDF webin-R 

Rechercher



analyse-R

Introduction à l'analyse d'enquêtes avec R et RStudio

Dernière mise à jour : 10 mai 2021

DOI [10.5281/zenodo.4746202](https://doi.org/10.5281/zenodo.4746202)

Contributeurs

Par ordre alphabétique :

Julien Barnier, Julien Biaudet, François Briatte, Milan Bouchet-Valat, Ewen Gallic, Frédérique Giraud, Joël Gombin, Mayeul Kauffmann, Christophe Lalanne, Joseph Larmarange, Nicolas Robette.

Création et Maintenance :

Joseph Larmarange — <http://joseph.larmarange.net>

- › Créé en 2015
- › À partir d'un support de cours de Julien Barnier
- › Complété / Actualisé au fil des années selon temps/besoins
- › Écosystème : *tidyverse*
- › Licence Creative Commons
- › Auteurs multiples
- › ~ 1400 pages (version PDF)
- › ~ 50 000 pages vues / mois

<https://larmarange.github.io/analyse-R/>

DOI: 10.5281/zenodo.2081968

Webin-R : contexte

- › Annulation de plusieurs formations/stages pendant le 1^{er} confinement
- › Octobre 2020 : anticipation de l'annulation de plusieurs stages à venir
- › Forte demande de partenaires africains

- › Décision de mettre en place un webinaire en visioconférence
- › Basé sur les contenus déjà développés pour analyse-R
- › Séances de 1 à 2 h, avec partage d'écran et réalisation du code en live
- › Séances archivées, accessibles & contenus sous licence Creative Commons

Chaîne YouTube dédiée

The screenshot shows the YouTube channel page for 'webin-R'. At the top, there is a search bar with the text 'Rechercher' and a microphone icon. Below the search bar is a banner for the channel with the text 'webin-R' and 'un séminaire sur l'analyse d'enquêtes avec R et RStudio'. The channel name 'webin-R' is displayed with a profile picture and '753 abonnés'. There are two buttons: 'PERSONNALISER LA CHAÎNE' and 'GÉRER LES VIDÉOS'. The navigation menu includes 'ACCUEIL', 'VIDÉOS', 'PLAYLISTS', 'CHAÎNES', 'DISCUSSION', and 'À PROPOS'. The main content area is divided into 'Diffusions en direct à venir' and 'Diffusions en direct terminées'. The 'à venir' section shows a live stream titled 'webin-R #19 : Trajectoires de soins, un exemple de données longitudinales (3)' scheduled for 10/06/2021 17:00. The 'terminées' section shows five past live streams with their titles, view counts, and upload dates.

Premium FR

Rechercher

webin-R

un séminaire sur l'analyse d'enquêtes avec R et RStudio

webin-R

webin-R

753 abonnés

PERSONNALISER LA CHAÎNE GÉRER LES VIDÉOS

ACCUEIL VIDÉOS PLAYLISTS CHAÎNES DISCUSSION À PROPOS

Diffusions en direct à venir

webin-R #19 : Trajectoires de soins, un exemple de données longitudinales (3)

webin-R • Planifié pour le 10/06/2021 17:00

Ressources du webin-R : <https://larmarange.github.io/webin-R/posts/2021-06-10-trajectoires-de-soins-3/>

#R #RStudio #Rprogramming #Rtutorial #RStats #statistiques #DataScience...

DÉFINIR UN RAPPEL

Diffusions en direct terminées

Titre	Vues	Diffusé il y a
webin-R #18 : Trajectoires de soins, un exemple de...	137 vues	1 semaine
webin-R #17 : Trajectoires de soins, un exemple de...	249 vues	2 semaines
webin-R #16 : Analyse de séquences	317 vues	3 semaines
webin-R #15 : Analyse de survie	458 vues	1 mois
webin-R #14 : Exemples de graphiques avancés (2)	370 vues	1 mois

Au 9 décembre 2021

- › 1200 abonnés
- › 24 vidéos
- › 36 940 vues
- › Toutes les vidéos sont chapitrées
- › Renvoi vers un site complémentaire dédié

<https://www.youtube.com/c/webinR>

Site dédié webin-R

Séances

June 10, 2021
Joseph Larmarange

Trajectoires de soins : un exemple de données longitudinales (3)

ANALYSER MANIPULER AVANCÉ

webin-R #19 du 10 juin 2021



May 24, 2021
Joseph Larmarange

Trajectoires de soins : un exemple de données longitudinales (2)

ANALYSER MANIPULER AVANCÉ

webin-R #18 du 25 mai 2021



May 19, 2021
Joseph Larmarange

Trajectoires de soins : un exemple de données longitudinales

ANALYSER MANIPULER AVANCÉ

webin-R #17 du 19 mai 2021



May 6, 2021
Joseph Larmarange

Analyse de séquences

ANALYSER AVANCÉ

webin-R #16 du 6 mai 2021



RESTEZ INFORMÉ-E

Envoyer un email avec pour objet "subscribe webin-R" à sympa@listes.ird.fr



CATEGORIES

- Articles (22)
- Analyser (15)
- Avancé (7)
- Débutant (5)
- FAQ (3)
- Intermédiaire (7)
- Manipuler (9)

- › Hébergé gratuitement sur GitHub
- › Développé avec R / RStudio / Git
- › Utilise Distill / Rmarkdown

- › Intégration des vidéos
- › Données des exemples
- › Scripts R / Rmd des séances
- › Liens vers les chapitres analyse-R
- › Liens vers ressources additionnelles

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

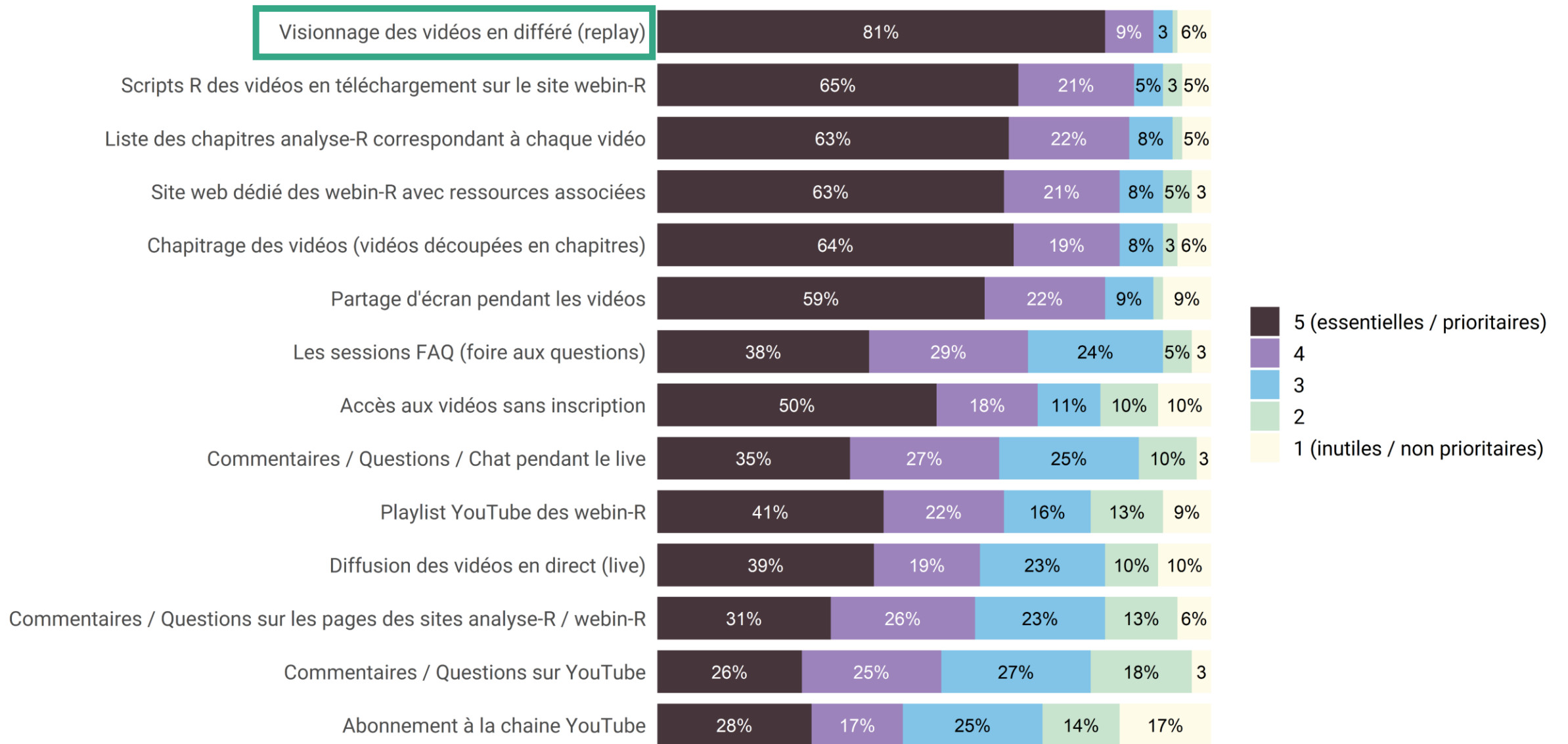
Enquête webin-R

- › Réalisation d'une petite enquête auprès des utilisatrices et utilisateurs des webin-R
- › Questionnaire en ligne et anonyme réalisé avec LimeSurvey
- › Promotion de l'enquête via la liste de diffusion des webin-R (~770 abonnés)
- › 115 questionnaires complets collectés entre le 17 mai et le 3 juin 2021

Profil des participant·e·s

- › **Sexe** : enquête : 66% d'hommes
- › **Âge**
 - » Enquête : de 24 à 71 ans, 35 ans en médiane
- › **International** : 40% vivent à l'étranger / 54% nés à l'étranger (principalement Afrique francophone)
- › **Situation académique**
 - » Mastorant·e·s (19%), Doctorant·e·s (26%), Post-doctorant·e·s (3%), Ingénieur·e·s (27%), Chercheur·e·s & Enseignant·e·s chercheur·e·s (17%), Médecins / Pharmacien·ne·s (6%)
- › **Disciplines**
 - » Santé publique (18%), Démographie (14%), Sociologie (14%), Épidémiologie (14%), Statistique (8%), Économie (7%), Médecine (4%), Géographie (3%), Écologie (3%)
 - » mais aussi anthropologie, neurosciences, hydrologie, sciences politiques, sciences de l'éducation, sciences de la données, psychologie...

Fonctionnalités les plus importantes



Visionnage des vidéos en différé

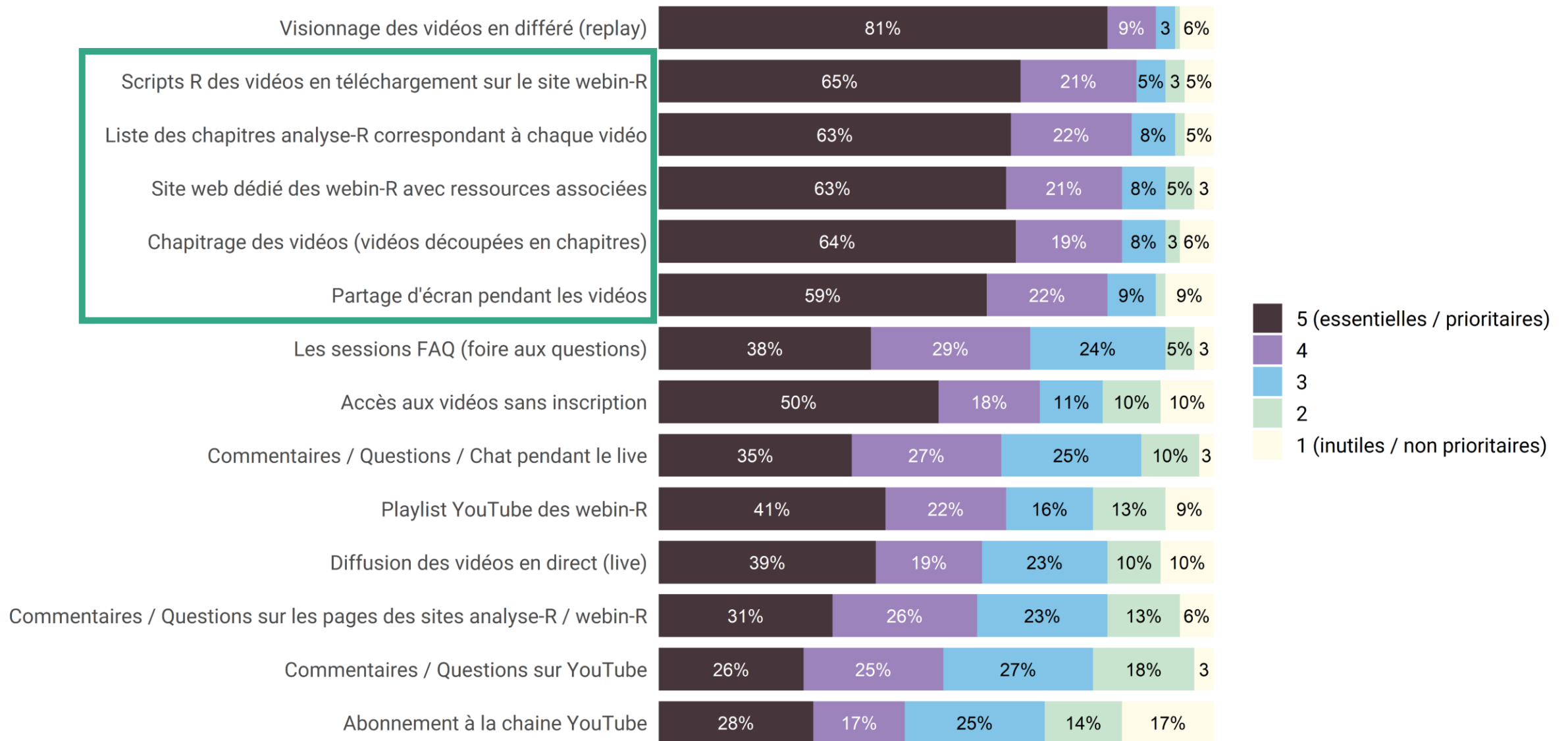
« Le choix direct versus replay est idéal car il n'est pas toujours possible d'être libre lors du créneau dédié au live. »

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

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

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

Fonctionnalités les plus importantes



Ressources, Chapitrage & Partage d'écran

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

81% déclarent utiliser régulièrement analyse-R

« La mise à disposition des scripts permettent de refaire l'exercice. »

« 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). »

« Les manipulations en direct sont un point fort. »

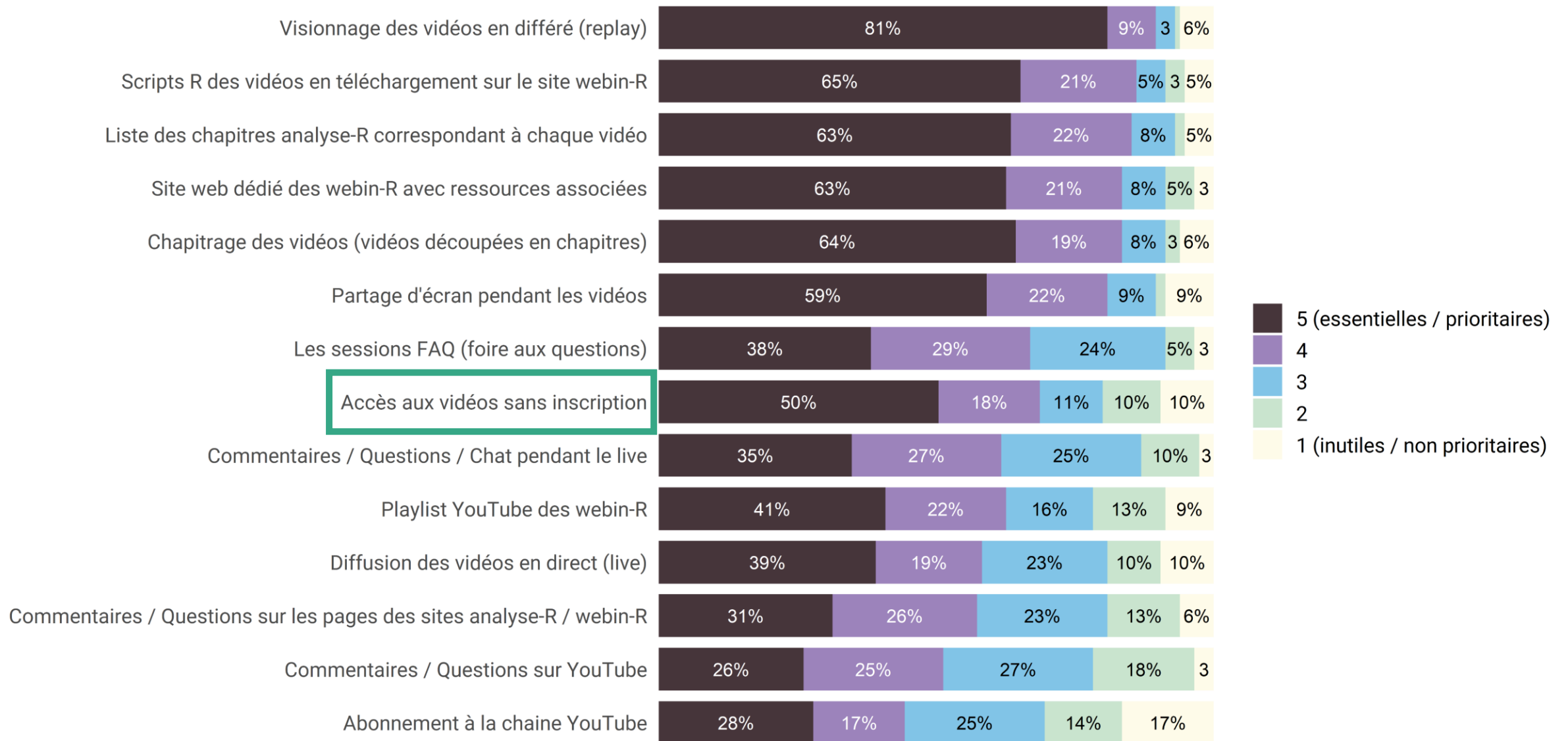
« Vue pratique directement sur Rstudio ! »

« Voir R ! »

« 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. »

La vidéo permet de voir la non-linéarité du processus d'écriture du code.

Fonctionnalités les plus importantes



Accessibilité

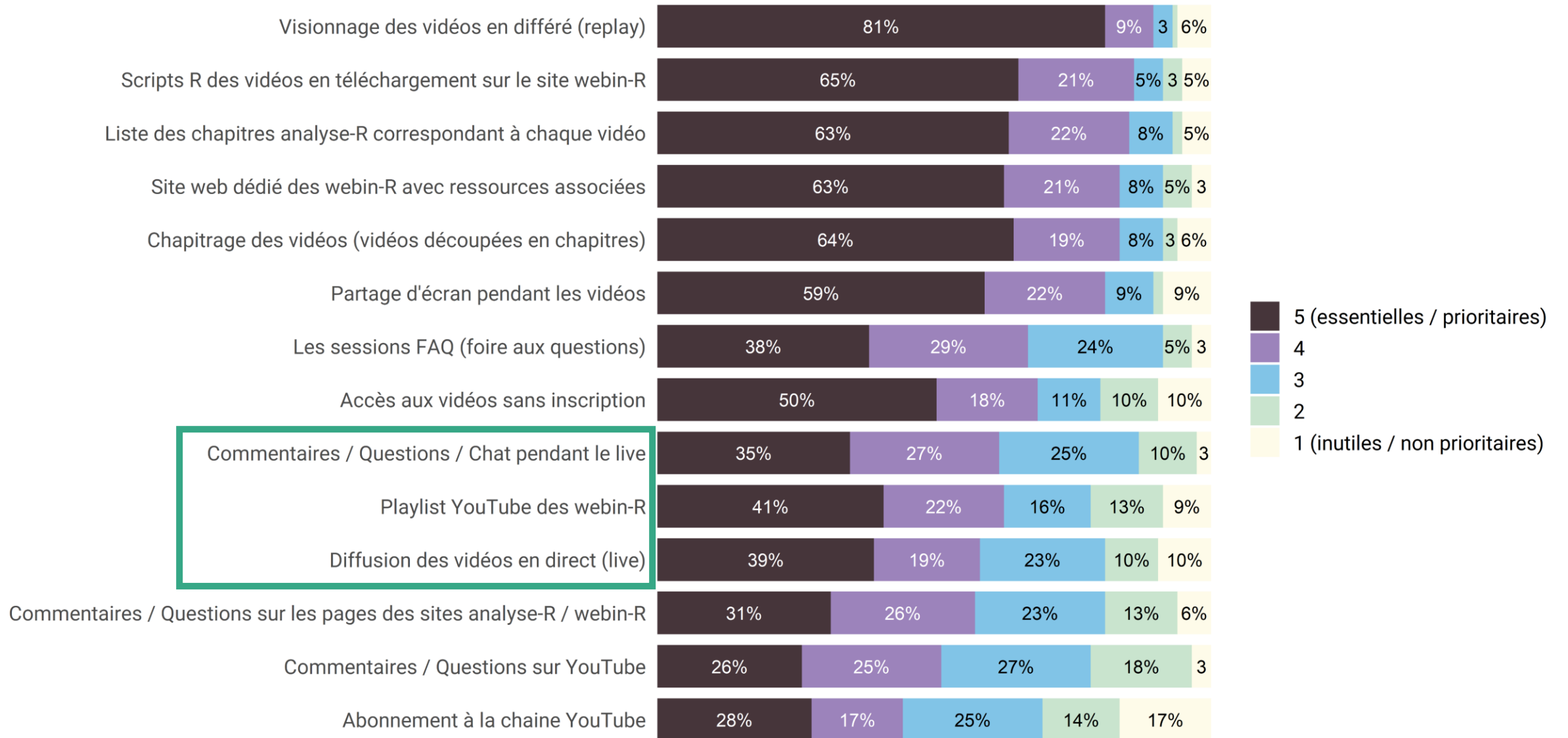
« C'est bien que les vidéos soient toujours disponibles, en accès libre, le fait qu'on puisse y revenir dès qu'on en a besoin. »

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

« Merci pour cette opportunité que vous nous donnez d'apprendre Gratuitement. »

*« 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. »*

Fonctionnalités les plus importantes



Diffusion en direct

« J'ai beaucoup apprécié les vidéos en direct car ça nous permet de poser nos questions. »

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

« Le format live avec d'éventuelles apartés rendent le contenu plus vivant. »

« La diffusion en direct amène un dynamisme et une régularité. »

Attentes exprimées / Points faibles

- › Hétérogènes du fait de l'hétérogénéité du public
 - » notamment en matière de thématiques, de disciplines
 - » pas assez spécifique / trop spécifique
 - » plus de statistiques de base / plus de statistiques avancées
- › En partie hors cadre initial (cours de stats, construction de l'objet...)
- › L'absence d'exercices, de travaux pratiques
- › La non-possibilité de valider des crédits d'enseignements
- › Durée des vidéos

Éléments de Discussion

- › Forte demande pour des contenus d'**autoformation à R** en **français**, accessible **gratuitement** et avec une visée appliquée
- › Intérêt de supports génériques librement accessibles en matière d'autoformation, en particulier au second et au troisième cycle (et au-delà), complémentaires des enseignements spécifiques et disciplinaires
- › Valorisation des enseignements sous des formats ouverts et librement accessibles est essentiel dans une perspective de **science ouverte**.
- › Recenser / Référencer les différentes initiatives permet de mutualiser (cf. <https://rzine.fr/>)
- › Les contenus en ligne ne remplacent pas mais complètent le présentiel
 - » en particulier au moment de la mise en œuvre des enseignements avec ses propres données
- › Pré-apprentissage en ligne possible avec complément en présentiel (e.g. Master HED)



webin-R

un séminaire sur l'analyse d'enquêtes avec R et RStudio

Merci

À bientôt pour un nouveau webin-R

Plus d'informations sur

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



webin-R

un séminaire sur l'analyse d'enquêtes avec R et RStudio

#22

Tableaux statistiques avec gtsummary

Lundi 13 décembre à 17h sur [YouTube](#)