

# Médiatisation des candidat.e.s à l'élection présidentielle 2022 dans l'audiovisuel

Nicolas Hervé  
Institut National de l'Audiovisuel - Service de la Recherche  
nherve@ina.fr

17 mars 2022 - v1.2

## Principaux points

- nous observons la médiatisation de l'élection présidentielle sur les médias audiovisuels en analysant le nombre de fois que les noms des candidat.e.s sont cités, cette mesure est différente du temps d'antenne du CSA (page 5)
- selon cette mesure, différente du temps d'antenne, TPMP (figure 13) et CNews (figure 35) accordent à l'automne 2021 une place écrasante à l'extrême-droite sur leur antenne, et dans une moindre mesure BFMTV et LCI
- une modélisation permet d'établir qu'une fois gommé l'effet de l'actualité propre de chaque candidat.e, la sur-représentation de Zemmour dans TPMP est de 34 points de pourcentage, 22 points sur CNews et 17 points sur BFMTV (par rapport à France 2, voir page 17)
- on observe un tassement des écarts entre les médias sur la période d'équité qui s'est ouverte au 1er janvier avec une tendance à l'harmonisation des proportions de citations des candidat.e.s
- par rapport à l'élection de 2017, on note une baisse significative de la médiatisation de la présidentielle en cours (figures 7 à 11) mais, même s'il est probable que la pandémie ait un rôle dans ce fait, rien ne permet de l'affirmer. Cette baisse est particulièrement marquée dans les JT du soir pour la couverture de la primaire de droite.
- cette étude a vocation à être mise à jour avec de nouvelles données au fur et à mesure de l'avancée de la campagne, ainsi qu'avec de nouveaux axes d'analyse<sup>1</sup>

---

1. Voir les différentes versions en ligne : <http://www.herve.name/presidentielle>

# Table des matières

<b>1</b>	<b>Approche et corpus</b>	<b>2</b>
1.1	Méthodologie . . . . .	2
1.2	Qualité de la transcription . . . . .	3
<b>2</b>	<b>Ampleur de la médiatisation de la présidentielle</b>	<b>4</b>
2.1	Nombre de citations des candidat.e.s . . . . .	4
2.2	Citations vs. temps d'antenne . . . . .	5
2.3	Comparaison avec 2017 . . . . .	7
<b>3</b>	<b>Citation des familles politiques</b>	<b>9</b>
3.1	Par période médiatique . . . . .	10
3.2	Par semaine . . . . .	10
3.3	Par chaîne / station . . . . .	11
<b>4</b>	<b>Automne 2021, modélisations</b>	<b>14</b>
4.1	Approche naïve . . . . .	15
4.2	Textométrie . . . . .	16
4.3	Régression linéaire . . . . .	17
4.4	Régression de Poisson . . . . .	19
<b>5</b>	<b>Début 2022, période d'équité, modélisations</b>	<b>22</b>
5.1	Approche naïve . . . . .	22
5.2	Régression linéaire . . . . .	22

## 1 Approche et corpus

Cette étude est un travail en cours largement ouvert à la critique et aux suggestions.

### 1.1 Méthodologie

Notre corpus est constitué des transcriptions de 4 types de créneaux à la télévision et à la radio. Nous avons les tranches horaires de 6h00 à minuit pour les canaux d'information en continue, les informations du soir sur les chaînes généralistes et les matinales radio. Nous ajoutons les tranches d'infotainment du soir, du lundi au vendredi, sur C8 (correspondant à TPMP) et sur TMC (correspondant à Quotidien). Nous transcrivons les créneaux décrits dans le tableau 1 à partir du 30 août 2021. On ne tient pas compte des éventuelles modifications de grilles de programme, notamment les débats ou soirées spéciales. La transcription des flux audio est assurée à l'aide du logiciel développé par le laboratoire du LIUM de l'Université du Mans (Tomashenko et al., 2016) déployé sur nos serveurs.

Une fois les transcriptions obtenues, on se contente de compter le nombre d'occurrences des noms de candidat.e.s, sous la forme la plus simple : leur nom de famille (voir tableau 2). Des tests ont été réalisés avec les formes plus complexes [prénom + nom] ou [monsieur/madame + nom] et les proportions relatives de citations sont conservées. On compte donc, dans ce qui est prononcé, les mentions aux noms des candidat.e.s, mais pas leur présence en plateau ou à l'image. Contrairement aux règles du CSA (nouvellement Arcom), on regarde donc de qui on parle et non qui parle. Cela signifie également que nous ne savons pas de quelle manière un.e candidat.e est évoqué.e. Des propos aussi bien positifs que négatifs peuvent être tenus et sont comptés indifféremment. Un point important à noter, nous ne pouvons pas transcrire le nom du candidat Anasse Kazib qui n'est pas dans le dictionnaire du logiciel de transcription que nous utilisons (datant de 2016/2017). De même, la transcription de Jean Lasalle est trop aléatoire (Lasalle → la salle). Ces deux candidats ne sont pas inclus dans l'étude.

Nous avons également choisi d'inclure dans la liste des candidat.e.s les principaux protagonistes des primaires ainsi que certain.e.s candidat.e.s putatifs même s'ils ne l'ont finalement pas été. Ce nombre est une bonne approximation, même s'il est imparfait, du temps consacré à l'antenne à parler de l'élection. Pour cette élection, Emmanuel Macron est président en exercice. Son nom peut donc être cité à l'antenne pour ces deux aspects qui sont indissociables. Le cas des présidents en exercice a toujours été un sujet pour le

	Chaîne / Station	Émission	Créneau
Info continu	BFM TV	Tranche horaire	6h00→0h00
	CNews (iTélé)	Tranche horaire	6h00→0h00
	LCI	Tranche horaire	6h00→0h00
	franceinfo:(TV)	Tranche horaire	6h00→0h00
	France Info (radio)	Tranche horaire	6h00→0h00
JT soir	TF1	JT de 20h00	19h55→21h05
	France 2	JT de 20h00	19h55→21h05
	France 3	JT de 19/20	19h25→20h05
	M6	JT du soir	19h40→20h15
Matinale	France Inter	Matinale	4h55→9h05
	Europe 1	Matinale	4h55→9h05
	RMC	Matinale	5h55→9h05
	RTL	Matinale	6h55→9h35
Div.	C8	TPMP, lun.→ven.	19h00→21h20
	TMC	Quotidien, lun.→ven.	19h15→21h25

TABLE 1 – Corpus des créneaux TV et radio

Élection	Liste des candidat.e.s
2022	Zemmour, Le Pen, Philippot, Dupont-Aignan, Ciotti, Barnier, Pécresse, Bertrand, Macron, Hidalgo, Taubira, Montebourg, Jadot, Rousseau, Mélenchon, Roussel, Poutou, Arthaud

TABLE 2 – Noms des candidat.e.s en 2022

temps d’antenne, le CSA ayant des règles bien particulières pour savoir ce qui doit être décompté, et l’objet de critiques régulières de la part des opposants politiques. Pour un éclairage sur ces règles et les différentes périodes, on peut se référer à (Meunier, 2021) ou (Geoffroy et al., 2022).

## 1.2 Qualité de la transcription

Comme tous les logiciels de ce type, la transcription est basée sur un dictionnaire de termes et sur un modèle de la langue française. Les résultats obtenus sont de bonne qualité mais ne peuvent pas être parfaits. Les erreurs peuvent être dues à deux raisons principales : les mauvaises conditions acoustiques ou des termes employées qui sont inconnus du logiciel. Afin d’évaluer la qualité de la transcription, nous comparons le nombre d’occurrences des noms des candidats trouvé par le logiciel du LIUM et par un second logiciel, commercial, développé par la société Vocapia<sup>2</sup>. Ce dernier est seulement utilisé sur une petite sous-partie de notre corpus. La comparaison est ainsi faite sur 1 600 heures issues de TF1, France 2, France Inter, France Info et BFM TV du 30 août 2021 au 25 janvier 2022.

Candidat.e	LIUM	Vocapia	$\Delta$	écart	Candidat.e	LIUM	Vocapia	$\Delta$	écart
Zemmour	5 656	5 839	-183	-3.24%	Hidalgo	2 098	2 202	-104	-4.96%
Le Pen	2 863	2 833	30	1.05%	Taubira	760	763	-3	-0.39%
Philippot	53	65	-12	-22.6%	Montebourg	524	521	3	0.57%
Dupont-Aignan	156	156	0	0.00%	Jadot	1 989	1 946	43	2.16%
Ciotti	1 126	1 173	-47	-4.17%	Rousseau	729	748	-19	-2.61%
Barnier	732	758	-26	-3.55%	Mélenchon	2 002	2 023	-21	-1.05%
Pécresse	3 045	3 061	-16	-0.52%	Roussel	418	453	-35	-8.37%
Bertrand	2 167	2 185	-18	-0.83%	Poutou	65	55	10	15.4%
Macron	9 598	10 439	-841	-8.76%	Arthaud	40	37	3	7.5%

TABLE 3 – Comparaison des logiciels LIUM et Vocapia, sur une partie du corpus

Les résultats sont présentés dans le tableau 3. Il y a en moyenne un écart de 4,8% entre les deux

2. <https://www.vocapia.com/>

logiciels, sans que l'on puisse toutefois dire lequel a raison. L'établissement d'une validation manuelle n'est pas envisagée pour l'instant. Globalement le logiciel du LIUM trouve un peu moins de citations de noms de candidat.e.s que celui de Vocapia. L'écart est le plus significatifs pour celles et ceux qui sont très peu cités. Parmi les principaux candidats, Emmanuel Macron est sous-évalué de l'ordre de 8.8%, Pécresse de 0.5%, Zemmour de 3.2% ou Mélenchon de 1%.

## 2 Ampleur de la médiatisation de la présidentielle

### 2.1 Nombre de citations des candidat.e.s

Sur la figure 1 nous présentons le nombre brut d'occurrences des noms de tou.te.s les candidat.e.s détectées sur les différents type de créneaux chaque semaine. Le volume transcrit étant constant sur la période, ce nombre brut donne une indication globale de l'ampleur du traitement politique lié à la présidentielle.

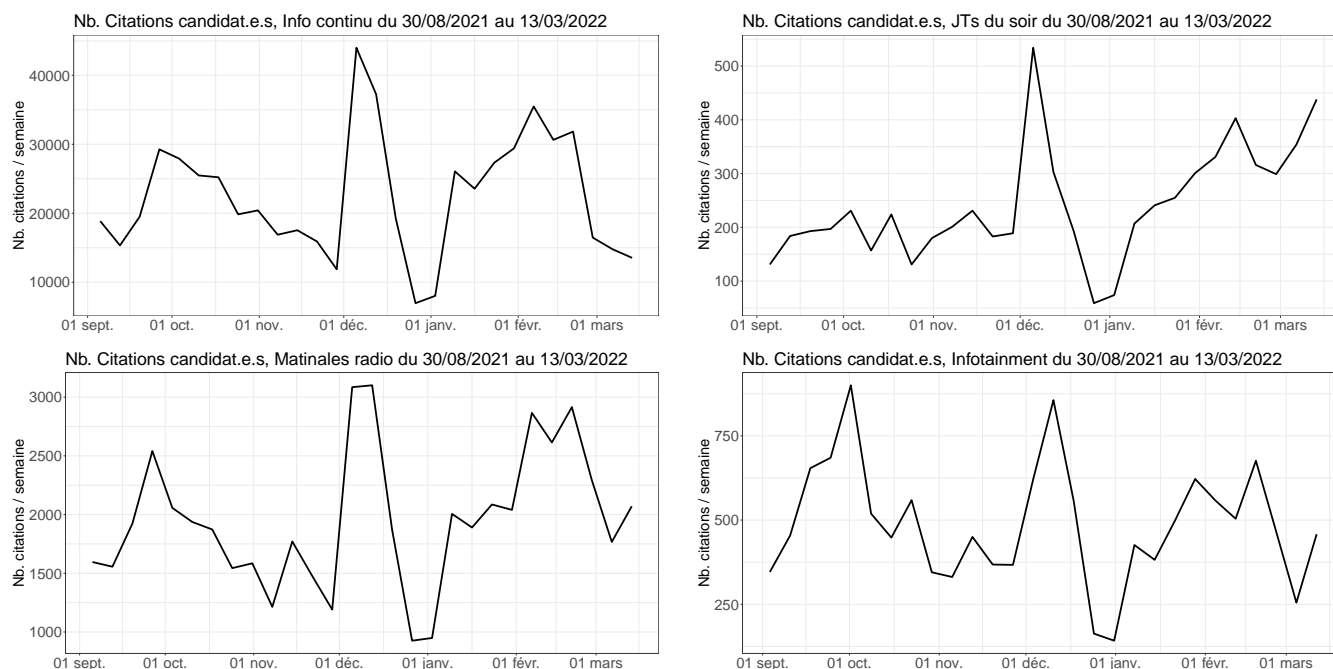


FIGURE 1 – Nombre brut de citations des candidat.e.s par semaine

Le volume horaire transcrit est significativement différent entre les JT et les chaînes d'information en continu. Les chiffres bruts ne permettent donc pas de comparer ces différentes plages médiatiques. Sur la figure 2 les nombres d'occurrences sont donc normalisés par rapport au niveau moyen pour chaque type de créneau. On peut ainsi les comparer entre eux et voir que globalement la temporalité médiatique est la même pour les chaînes d'information en continu, les JT du soir et les matinales radio. Seule l'infotainment se démarque un peu à l'automne avec une plus forte présence de noms de candidat.e.s à l'antenne. On observe également le creux de la période des fêtes de fin d'année où les politiques sont moins cités. Dernièrement, probablement avec le traitement médiatique de la guerre en Ukraine, on note une différenciation selon les type de chaînes et une plus forte implication des chaînes historique avec les JT du soir sur la présidentielle. Il faudra voir si cette tendance se confirme dans les 15 prochaines jours.

## Citations candidat.e.s, du 30/08/2021 au 13/03/2022

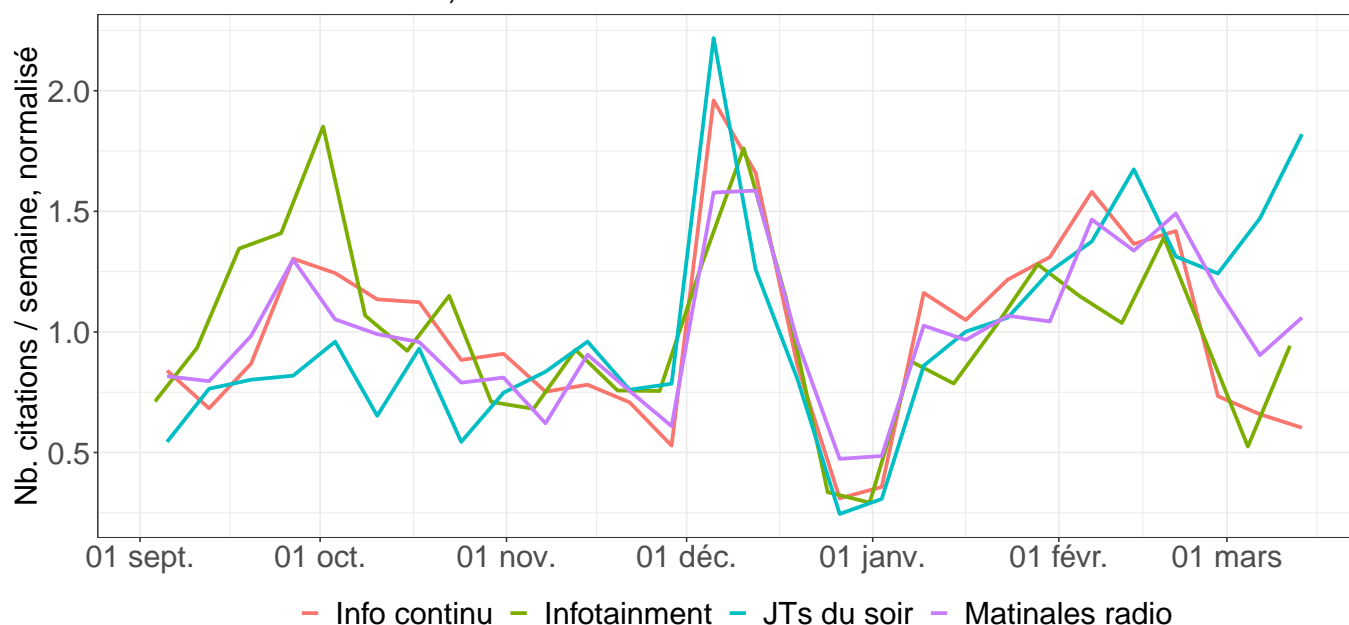


FIGURE 2 – Nombre de citations normalisé des candidat.e.s par semaine

## 2.2 Citations vs. temps d'antenne

Nous avons déjà insisté sur le fait que compter le nombre de citations des noms de candidat.e.s à l'antenne n'est pas la même chose que le temps d'antenne politique tel que défini par le CSA / Arcom. On peut toutefois penser qu'il y a une corrélation entre ces deux données. Une première source d'information est le relevé manuel fait par Claire Sécaïl (Sécaïl, 2022) pour l'émission TPMP sur la fin de l'année 2021. On constate sur la figure 3 une similitude très nette pour cette émission entre les deux approches de comptage.

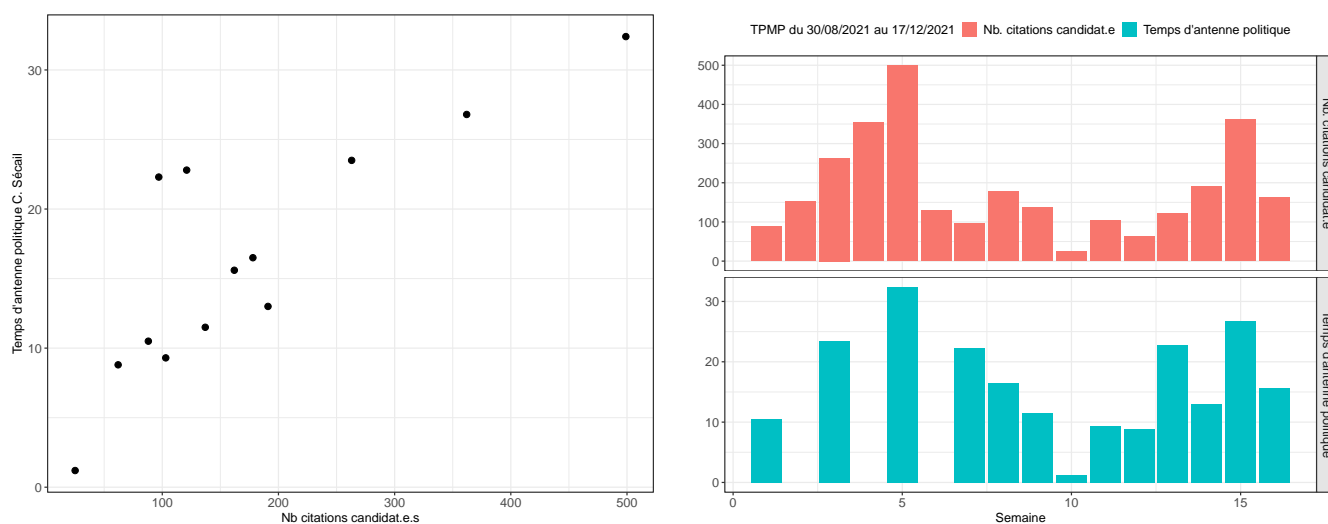


FIGURE 3 – TPMP - Nombre de citations de candidat.e.s par semaine vs. temps d'antenne politique décompté par Claire Sécaïl

Une seconde source de données provient des temps d'antenne que les chaînes remontent au CSA / Arcom. Ces données sont en semi open-data (disponibles mais difficilement exploitables automatiquement)<sup>3</sup>. Benjamin Guinaudeau a mis en place une procédure pour y accéder plus simplement<sup>4</sup>. On distingue deux périodes dans les données du CSA. À l'automne 2021, c'est le régime de routine qui est en place avec un relevé des temps d'antenne des personnalités politiques : "Conformément aux dispositions de l'article 13 de la loi du 30 septembre 1986, L'Arcom publie régulièrement les tableaux relatifs aux temps de parole

3. <https://www.csa.fr/csapluralisme/tableau>

4. <https://github.com/benjaminquinaudeau/csatimer>

des personnalités politiques relevés dans les journaux et bulletins d'information, les magazines et les autres émissions des programmes sur les antennes des télévisions et des radios". Une synthèse est présentée sur la figure 4. À partir du 1er janvier on bascule dans la période électorale et ce sont les candidat.e.s et leurs soutiens qui sont chronométrés.

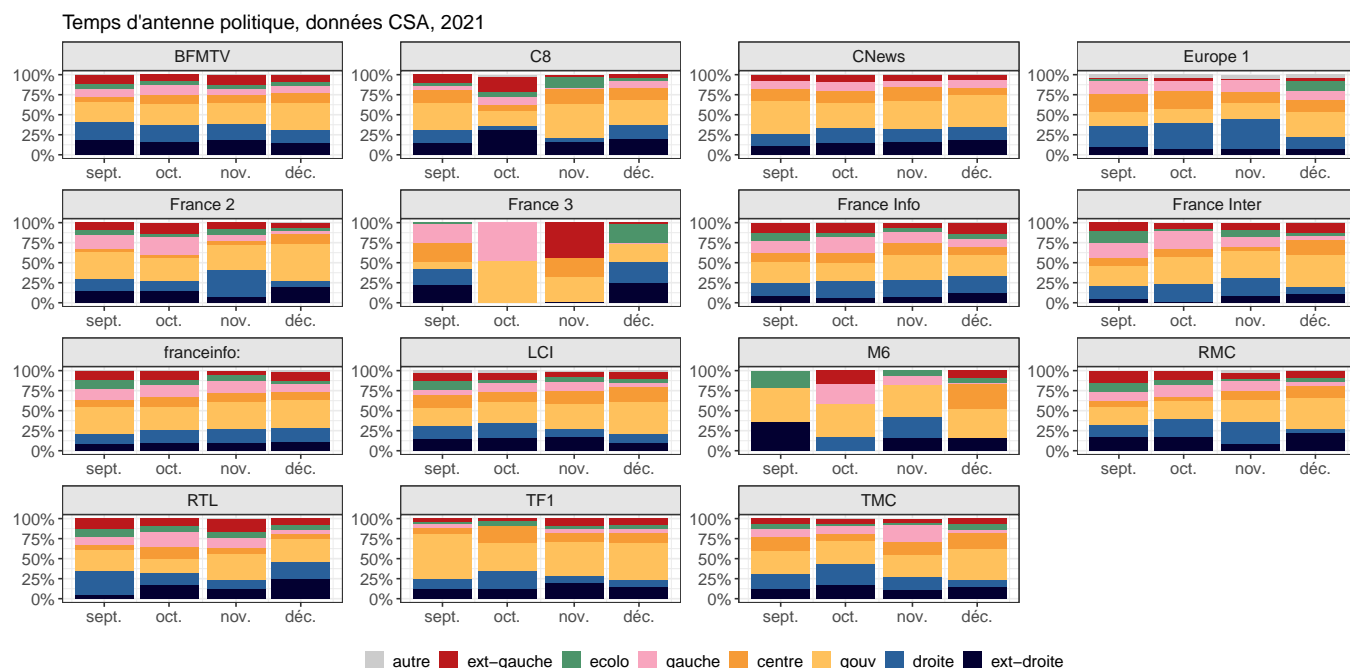


FIGURE 4 – Temps d'antenne des différentes familles politiques à l'automne 2021

Les données du CSA pour la période de routine sont fournies pour 3 catégories principales de type de programmes et agrégées par mois. De notre côté, nous avons transcrit les créneaux d'information pour la plupart des chaînes et deux émissions de divertissement. La concordance entre les deux corpus est plus difficile à établir que dans le cas de TPMP. On peut déjà voir sur la figure 5 la forte disparité des types de programmes pour lesquels les chaînes ont fait état de temps d'antenne de personnalités politiques. Ainsi, il y a de fortes chances que Quotidien sur TMC soit classé comme *Magazine*, alors que TPMP sur C8 se retrouve dans la catégorie *Autre*. La faible part de la catégorie *Information* sur LCI interroge également. Ces fortes disparités sont probablement liées à un cadrage trop souple et des interprétations différentes de la part des chaînes dans leurs remontés de données. On voit sur la figure 6 que toute corrélation entre le temps d'antenne politique et le nombre de citations des candidat.e.s est beaucoup trop hasardeuse.



FIGURE 5 – Proportion des temps d’antenne politique CSA par type de programme à l’automne 2021

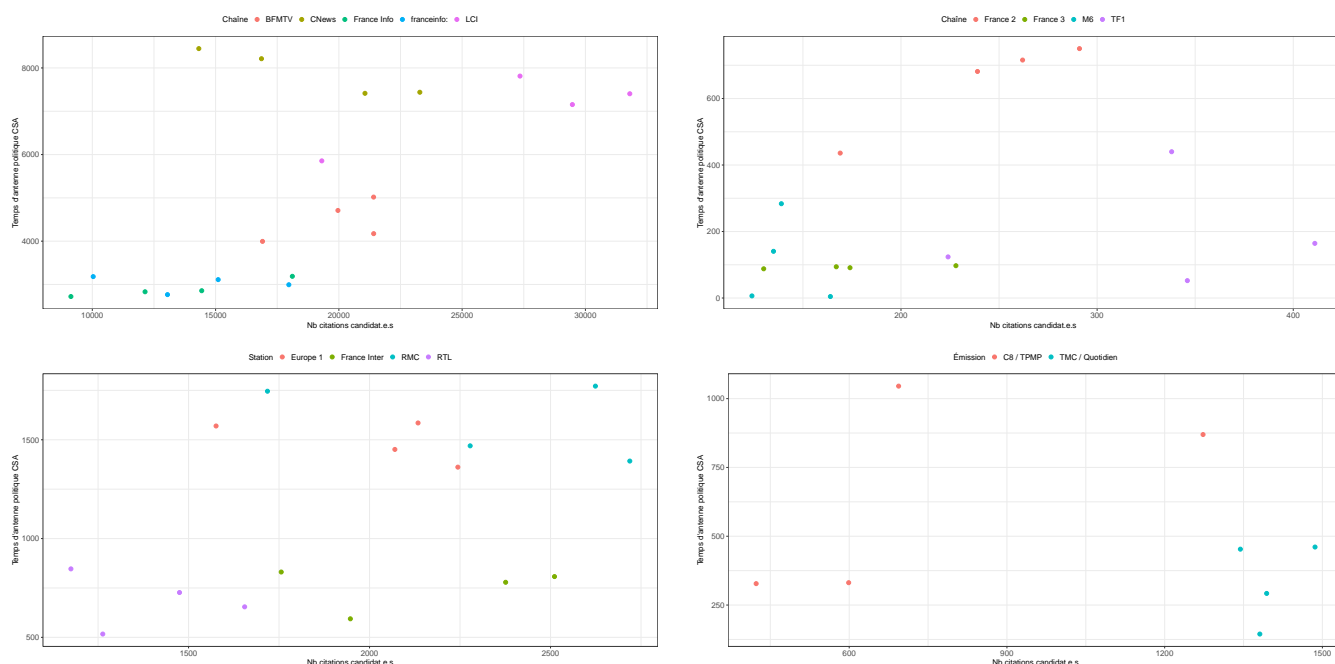


FIGURE 6 – Nombre de citations de candidat.e.s pour les 4 derniers mois de 2021 vs. temps d’antenne politique CSA (tous types de programmes)

### 2.3 Comparaison avec 2017

Contexte : question posée par Clémence Bauduin, journaliste du Parisien, dans son article du 12 janvier 2022 : « *C’est une dépression politique* » : quand le Covid-19 éclipse l’élection présidentielle (Bauduin, 2022). Peut-on dire que la pandémie de Covid19 tend à minorer la médiatisation de l’élection présidentielle 2022 ? La guerre en Ukraine prend le relais de la médiatisation du Covid et il est, à l’heure où ces lignes sont écrites, impossible de dire quel impact elle va avoir sur la médiatisation de l’élection.

Pour y répondre, il faut idéalement un autre point de référence pour lequel la seule variable qui change soit l’existence de la pandémie. C’est bien évidemment impossible, on se contente donc de comparer la médiatisation actuelle de l’élection présidentielle avec la précédente de 2017 en sachant que les candidat.e.s ne sont pas les mêmes, que les familles politiques ont évolué et que le président sortant est, cette fois-ci, candidat lui-même.

Nous reprenons le décompte des candidats auquel nous ajoutons également les termes *présidentielle* et *primaire*.

Élection	Liste des candidat.e.s
2017	Le Pen, Dupont-Aignan, Asselineau, Fillon, Sarkozy, Juppé, Cheminade, Lasalle, Macron, Bayrou, Valls, Hollande, Montebourg, Hamon, Jadot, Rivasi, Duflot, Mélenchon, Poutou, Arthaud

TABLE 4 – Noms des candidat.e.s en 2017

Sur les figures 7 et 8 on fait apparaître les périodes des primaires des écologistes (en vert), de la droite (en bleu) et de la gauche (en rouge). Pour l'élection de 2022, nous avons choisi la *primaire populaire* pour la gauche. La période de l'élection proprement dite est en gris.

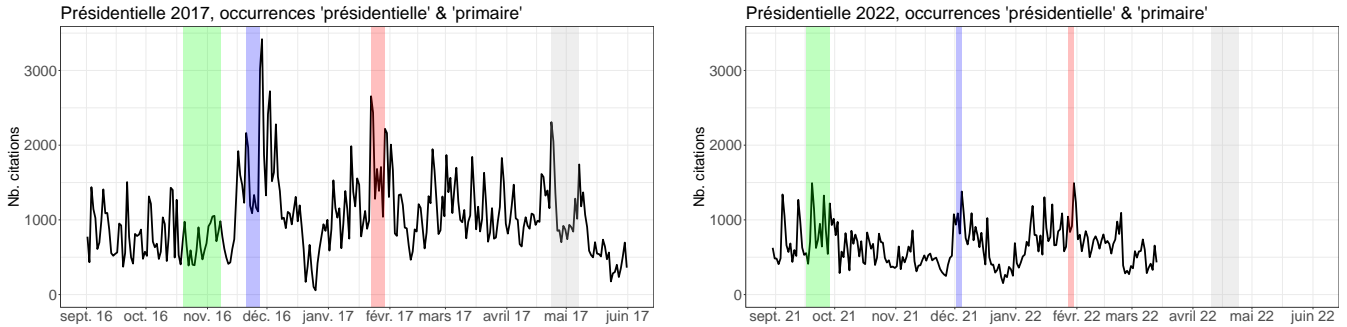


FIGURE 7 – Termes *présidentielle* et *primaire*, sur tous les médias de notre corpus

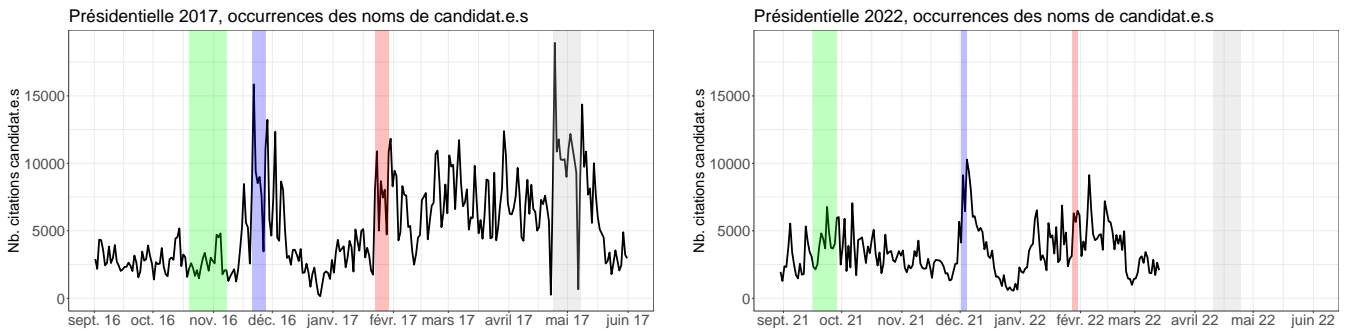


FIGURE 8 – Nom des candidat.e.s, sur tous les médias de notre corpus

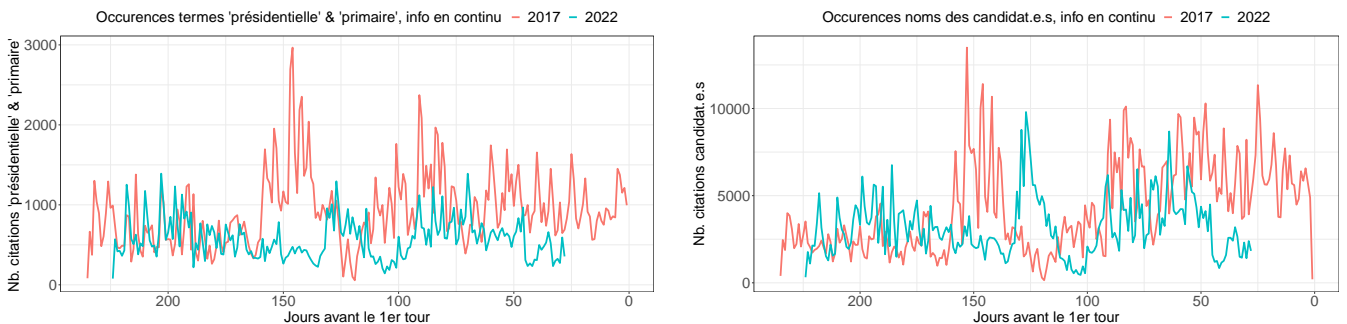


FIGURE 9 – Sur tous les médias d'information en continu

Pour comparer les deux élections entre elles, on prend comme point de référence temporel le jour du premier tour. La date étant légèrement différente entre 2017 et 2022, les courbes sur les figures 9 à 11 semblent légèrement décalées. Le creux de Noël est bien visible ainsi que le pic lié à la primaire de la droite. Puisque nous avons transcrit exactement le même volume horaire en 2017 et 2022, le nombre brut d'occurrences des noms de candidat.e.s et des termes *présidentielle* et *primaire* peuvent être directement utilisés pour comparer l'ampleur de la médiatisation de l'élection de 2022 par rapport à celle de 2017. La



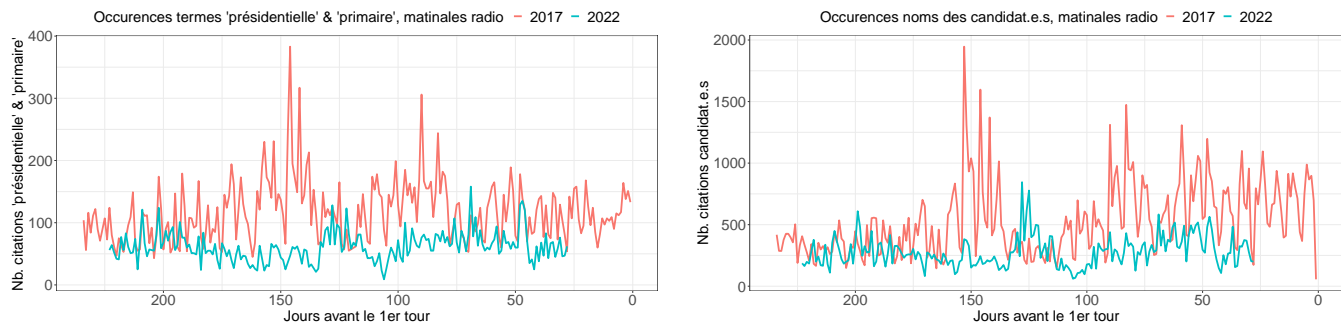


FIGURE 10 – Dans les matinales radio

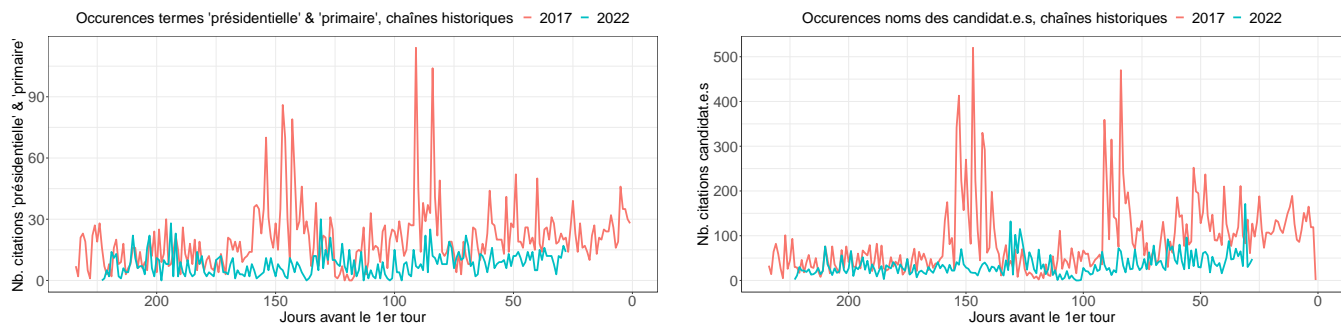


FIGURE 11 – Sur les TV généralistes

pandémie de Covid19 est loin d'être le seul facteur explicatif potentiel des différences que nous observons sur ces figures. Les candidat.e.s, leurs actions de campagne, la situation économique, les thèmes de campagne, l'intérêt des français pour l'élection sont autant de facteurs, fortement intriqués, qui pèsent dans les choix éditoriaux des rédactions dans leurs choix. On observe globalement que les différents médias ont un niveau de traitement de cette élection plus faible qu'il y a 5 ans. Ce constat est particulièrement marqué pour les différentes primaires, que l'on considère les occurrences des noms de candidat.e.s ou les termes *présidentielle* et *primaire*. De plus, on observe très nettement une chute de la médiatisation fin février avec le début de l'invasion de l'Ukraine par la Russie, bien visible pour les médias d'information en continu.

### 3 Citation des familles politiques

Afin d'observer les tendances politiques, nous regroupons les candidat.e.s selon la grande famille politique à laquelle on peut raisonnablement les rattacher. Les familles politiques utilisées sont celles du groupe d'experts Chapel Hill (Jolly et al., 2022). La subjectivité de ce classement reste de la seule responsabilité de l'auteur. Toutefois, pour celles et ceux que ce choix de regroupement peut troubler, les résultats non agrégés, par candidat, sont présentés en annexes (page 25).

Pour chaque chaîne, on définit alors la proportion de citations du nom de chaque candidat.e par rapport au nombre total de citations. Ce pourcentage indique donc la part qu'occupe un.e candidat.e dans ce qui est prononcé sur une antenne et qui concerne la politique. Il est indépendant du temps d'antenne global dévolu à l'élection présidentielle sur chaque chaîne. La place accordé à l'élection ainsi que la répartition des candidat.e.s qui sont évoqués sont deux décisions éditoriales.

Famille	Liste des candidat.e.s
extrême-droite	Zemmour, Le Pen, Philippot, Dupont-Aignan
droite	Ciotti, Barnier, Péresse, Bertrand
centre	Macron
gauche	Hidalgo, Taubira, Montebourg
écologistes	Jadot, Rousseau
extrême-gauche	Mélenchon, Roussel, Poutou, Arthaud

TABLE 5 – Familles politiques des candidat.e.s

### 3.1 Par période médiatique

Nous reprenons la présentation des résultats que Claire Sécaïl a faite (Sécaïl, 2022) en regroupant les proportions de citations de candidat.e.s par grande période de la campagne. Ces périodes sont définies par le CSA et impliquent, pour chacune d’entre elle, des règles différentes et de plus en plus strictes sur le décompte du temps d’antenne et de parole des candidat.e.s à mesure qu’on se rapproche de la date de l’élection. Nous rappelons ici que la mesure que nous proposons n’est pas similaire à celle du temps d’antenne, mais elle est complémentaire.

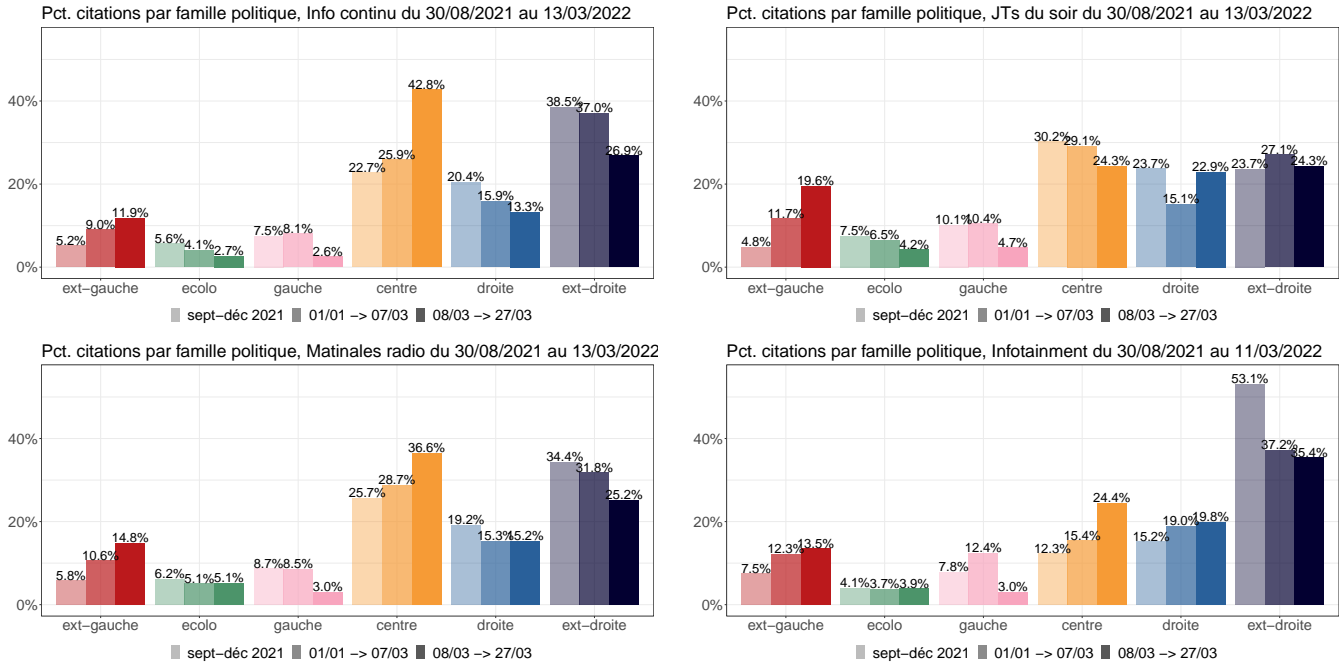


FIGURE 12 – Proportion moyenne de citations par famille politique et par période

On note sur ces premiers résultats, présentés dans la figure 12, des profils différents selon les types d’émissions qui sont observées, notamment sur les citations du centre (Macron, y compris dans son activité de président) et de l’extrême-droite. Il faut noter que pour l’instant nous ne disposons que d’une semaine de données pour la troisième période. Pour l’extrême-droite on tend à voir une baisse du traitement médiatique sur les chaînes d’information en continue et sur les créneaux d’infotainment. Là encore, il faut voir dans les 15 prochains jours si cette tendance se confirme, probablement liée au règle du CSA.

### 3.2 Par semaine

Parler ou non d’un.e candidat.e à l’antenne est une décision éditoriale. Elle relève du choix de la rédaction et est, *a priori*, influencée par l’actualité de la campagne : entrée en campagne, meeting, présentation du programme, sondages, faits divers, ralliement, ... Autant d’éléments qui sont très difficilement quantifiables. Pour analyser l’importance qu’une chaîne choisit d’accorder à telle ou tel candidat.e, il faut donc un point de comparaison. Nous regardons donc comment les médias se comportent les uns par rapport aux autres avec une granularité temporelle ramenée à la semaine afin de tenir compte de l’actualité (figures 13 et 14). Le détail par candidat est donnée en annexe (32).

Les primaires écologiste et de la droite apparaissent comme des moments privilégiés pour l’apparition des noms des candidat.e.s de ces familles politiques à l’antenne. Globalement, le centre, la droite et l’extrême-droite sont très nettement plus citées que les familles politiques de gauche. Les JT du soir, avec le volume horaire le plus contraint, semblent être les plus équilibrés. En revanche l’infotainment accorde une place écrasante à l’extrême-droite.

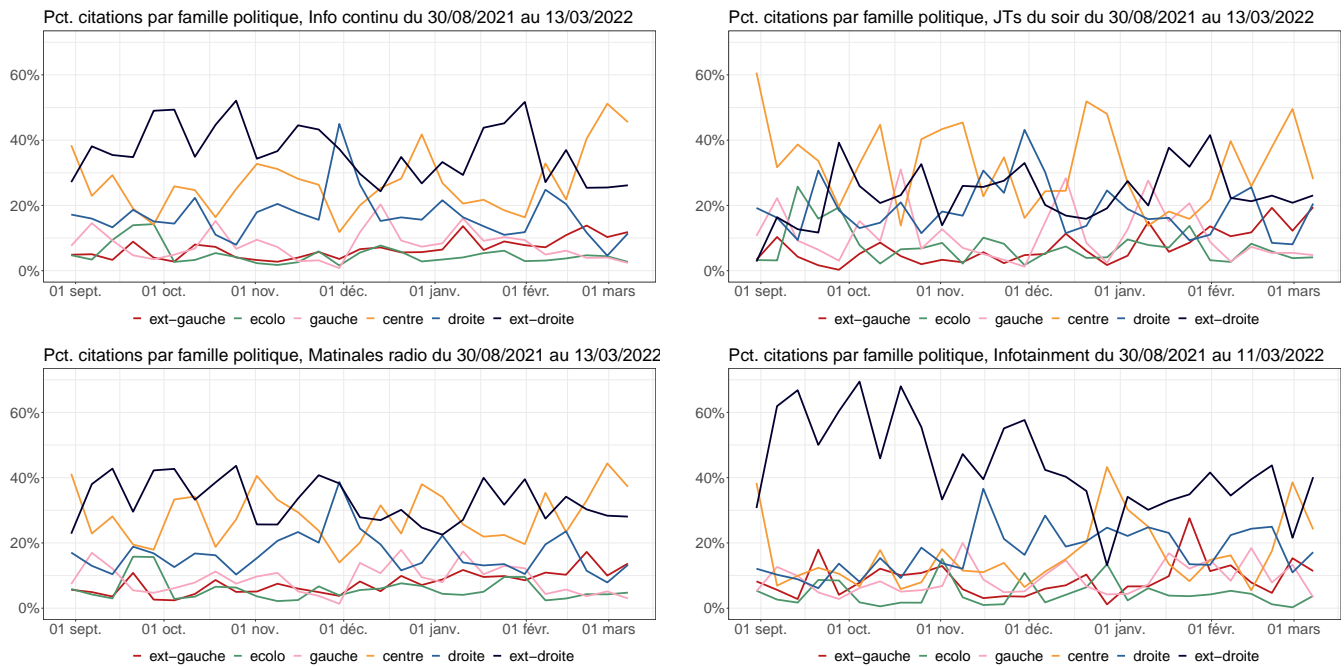


FIGURE 13 – Proportion moyenne de citations par famille politique et par semaine

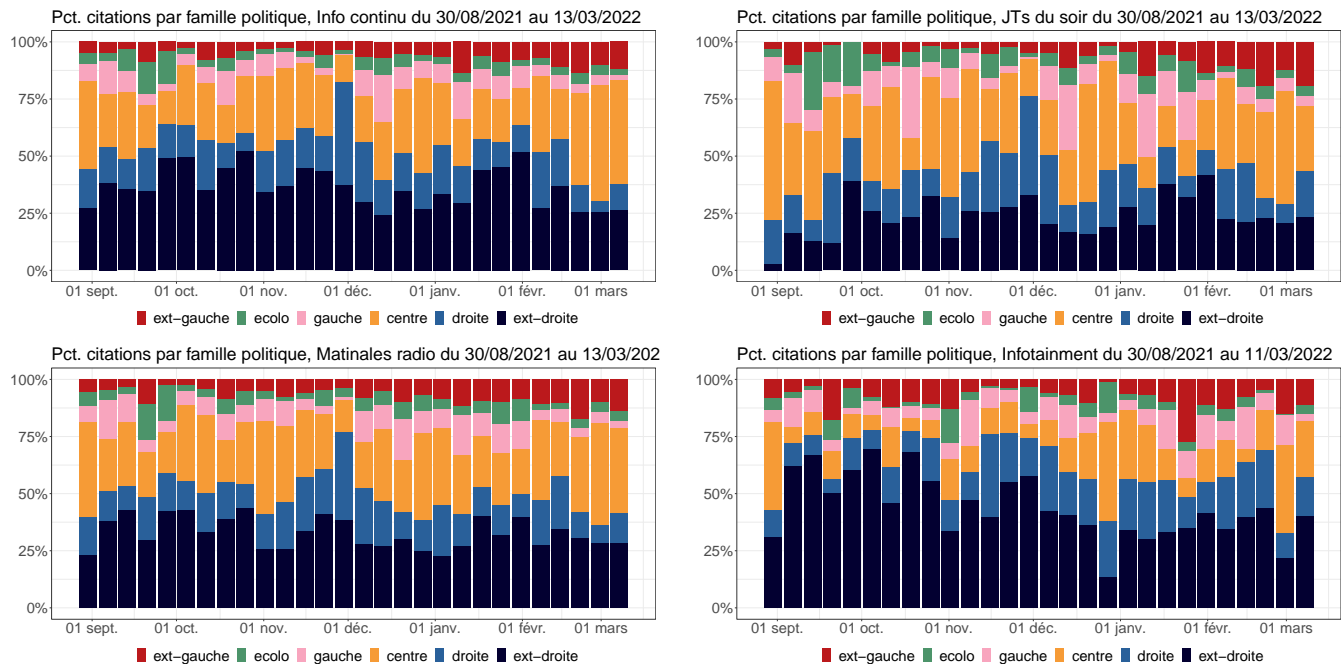


FIGURE 14 – Proportion moyenne de citations par famille politique et par semaine

### 3.3 Par chaîne / station

Les résultats très atypiques de l'infotainment nous poussent finalement à regarder ce que donnent les proportions pour chaque média afin de voir si d'autres particularités existent. Pour TPMP, on atteint 70% des citations sur plusieurs mois à l'automne. Ce constat rejoint celui fait par Claire Sécaïl et son décompte manuel du temps d'antenne des familles politiques sur cette émission (Sécaïl, 2022). On constate également que les chaînes d'information en continu privée sortent également du lot avec une forte présence de l'extrême-droite dans les candidat.e.s cités à l'antenne. Sans atteindre les scores de TPMP, la chaîne CNews présente également un plateau autour de 60% pour l'extrême-droite à l'automne. BFMTV n'est pas très loin derrière avec toutefois une couverture plus importante, proportionnellement, de la primaire de la droite.

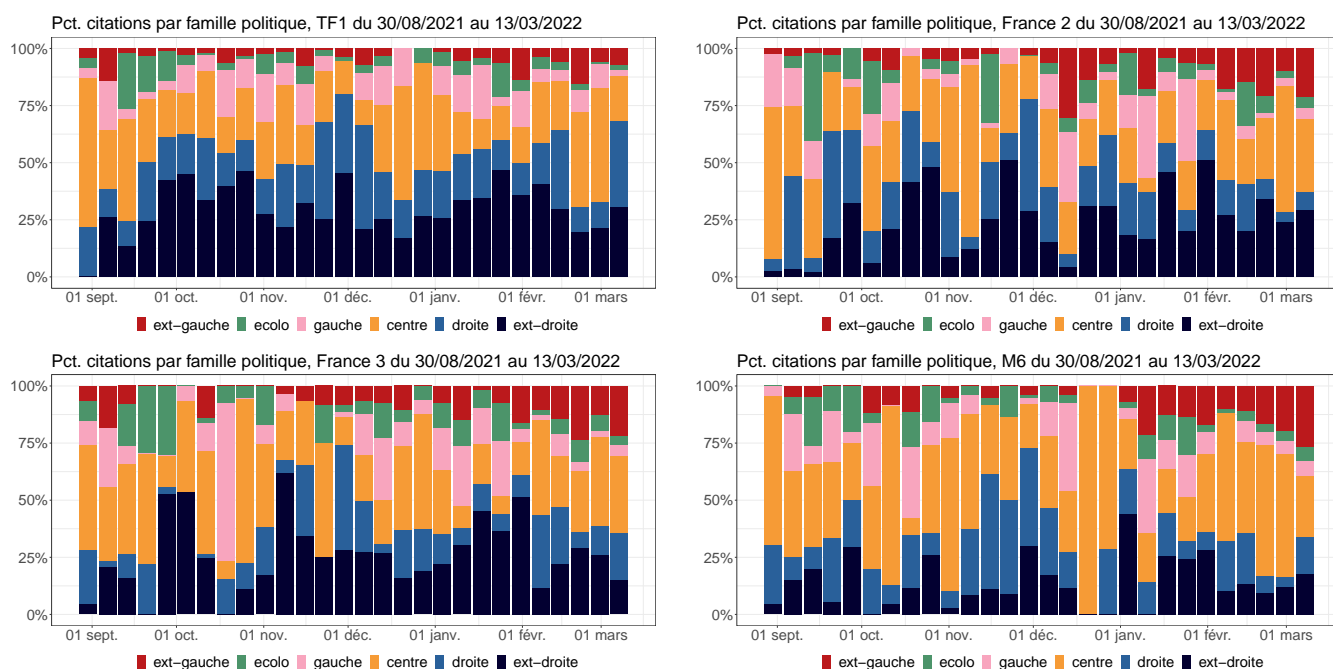


FIGURE 15 – Proportion moyenne de citations par famille politique pour les JT du soir

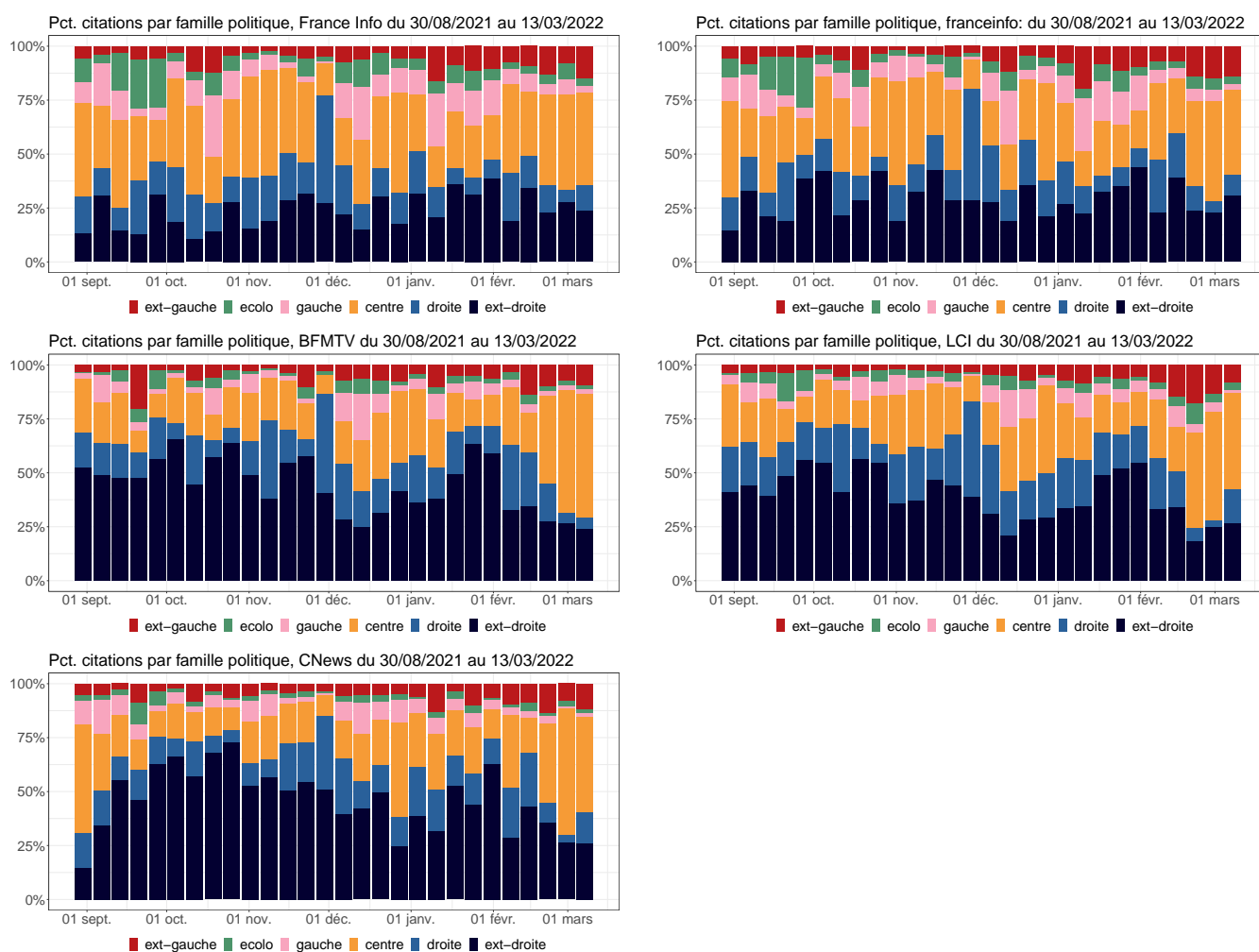


FIGURE 16 – Proportion moyenne de citations par famille politique pour les médias d'information en continu

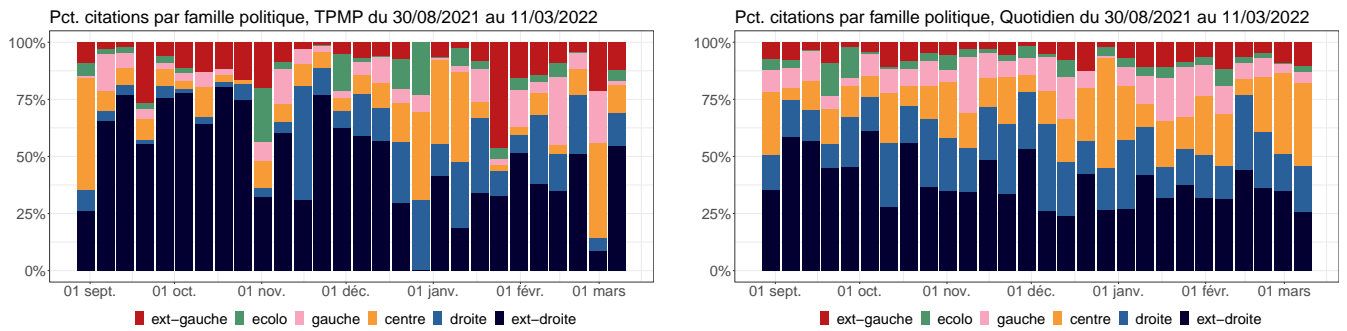


FIGURE 17 – Proportion moyenne de citations par famille politique pour les émission d'infotainment

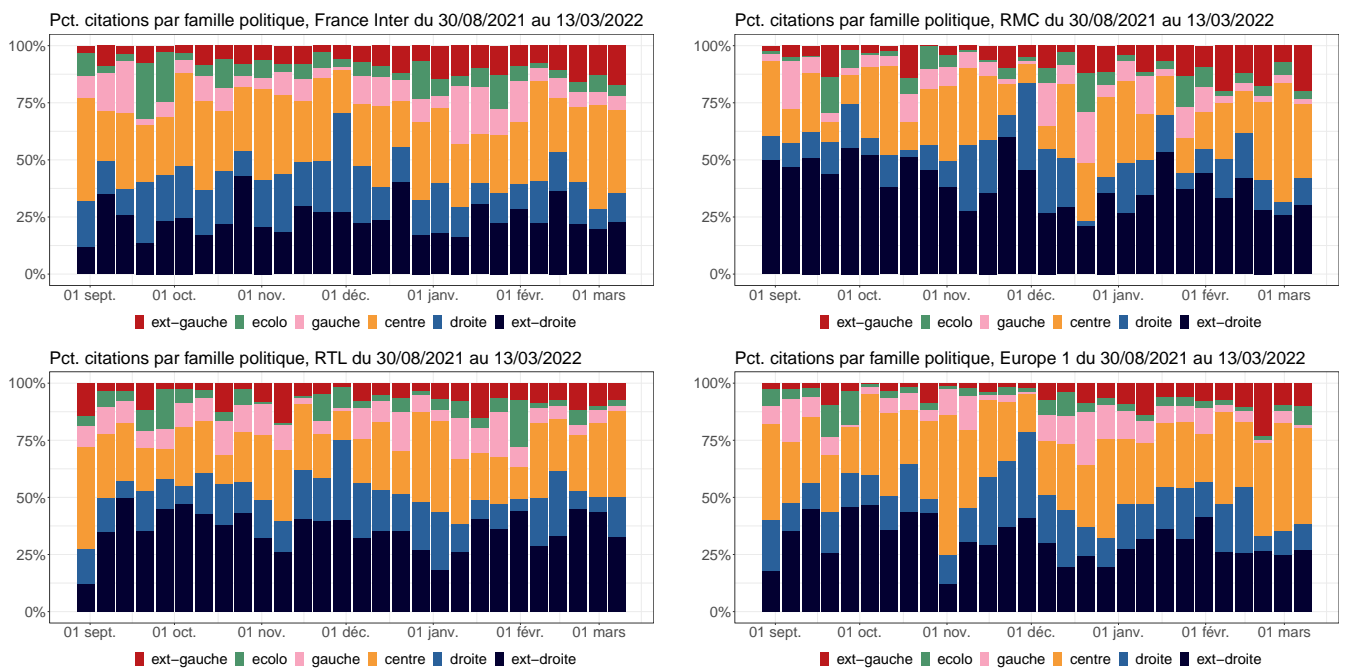


FIGURE 18 – Proportion moyenne de citations par famille politique pour les matinales radio

## 4 Automne 2021, modélisations

Quelle est l'influence des médias sur le nombre de citations des candidat.e.s sur leur antenne ? On a déjà pu voir sur le long terme l'influence de ces médias ainsi que de leurs actionnaires et des journalistes sur les invités politiques qu'ils reçoivent dans leurs émissions (Cagé et al., 2022). Ces influences sont, en théorie et en partie, contre-balançées par les règles de pluralité interne à chaque média fixées par le législateur et que le CSA/Arcom doit faire respecter. Ces règles ne disent rien en revanche des sujets qui sont abordés à l'antenne (et c'est fort heureux !) et qui dépendent autant de l'actualité que du cadrage éditorial du média. Nous cherchons à modéliser ce phénomène pour en comprendre les déterminants. On se concentre pour l'instant sur la pré-campagne en excluant la période de Noël qui est toujours un peu particulière d'un point de vue éditorial.

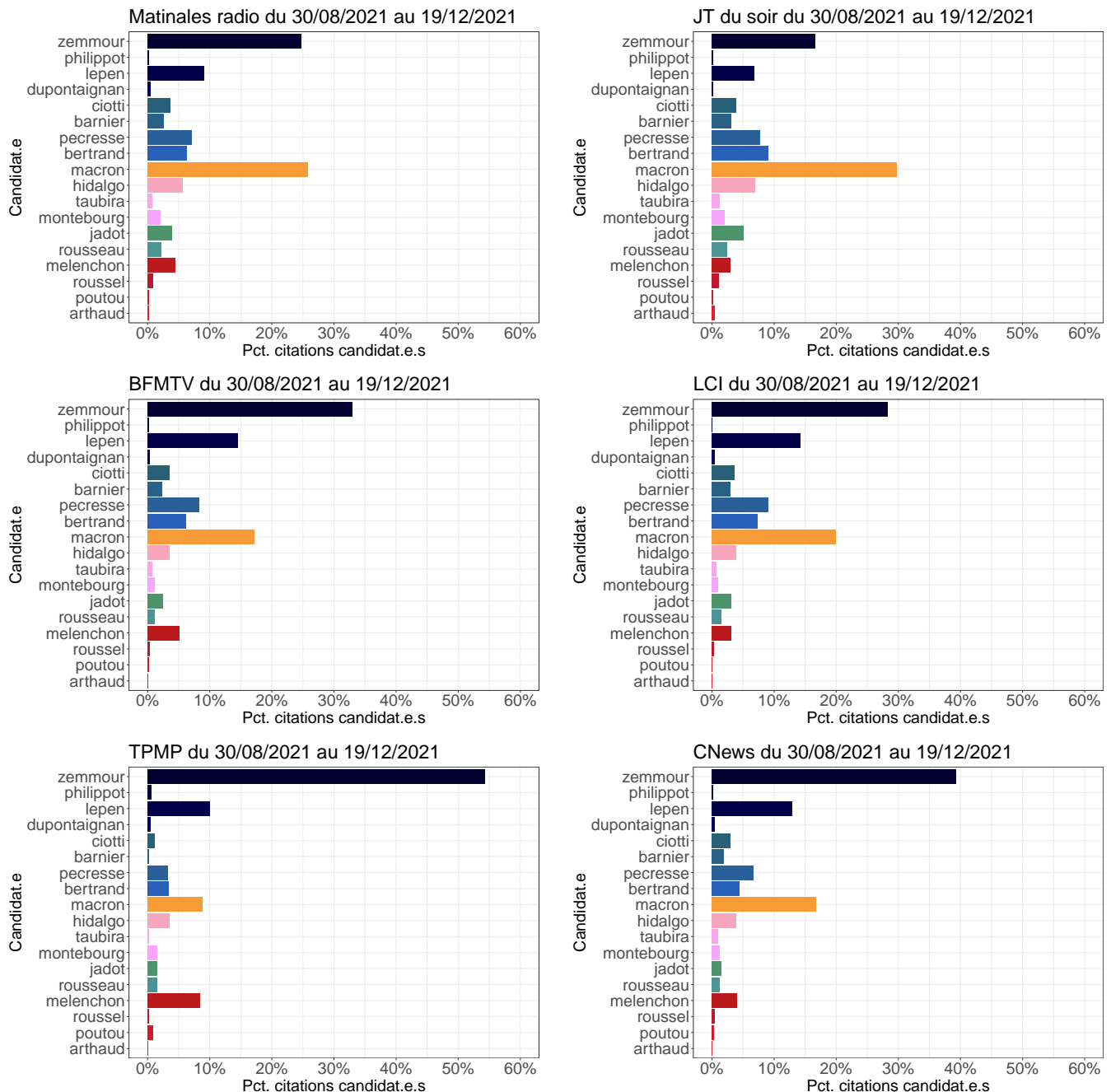


FIGURE 19 – Proportion moyenne de citations par candidat.e à l'automne 2021

On dispose du nombre de citations des différents candidats à l'élection présidentielle sur chaque chaîne TV et station de radio depuis septembre 2021 (voir table 1). Une synthèse sur la période est présentée sur la figure 19. On considère une seule émission ou tranche horaire par média. La durée transcrite pour chaque média/émission varie fortement, donc le nombre de mots transcrits est très variable d'un média à l'autre. On

cherche à savoir si un média est biaisé politiquement, c'est à dire si le nombre de citations d'une candidat est anormal par rapport aux autres médias. Outre un biais politique, ce nombre de citations peut également dépendre de l'actualité. Les données dont nous disposons sont donc les suivantes :

- nb de médias / émissions : 15 (dont 2 lun → ven)
- nb de semaines complètes : 16 (112 jours, 30 août 2021 → 19 déc. 2021)
- nb candidat.e.s : 18 (zemmour, lepen, philippot, dupontaignan, ciotti, barnier, pecresse, bertrand, macron, taubira, hidalgo, montebourg, jadot, rousseau, melenchon, rousssel, poutou, arthaud)
- nb total d'observations au niveau jour : 29 088 (dont 15 373 avec  $nb = 0$ )
- nb total d'observations agrégées au niveau de la semaine : 4 320 (dont 1 029 avec  $nb = 0$ )

## 4.1 Approche naïve

On ne considère que le temps politique pour chaque média, représenté par le nombre total de candidat.e.s cité.e.s à l'antenne, et les proportions relatives de citation de chacun.e d'entre eux. Ceci permet de gommer les deux sources de variabilités que sont la durée de l'émission et la part du traitement de l'élection présidentielle dans les choix éditoriaux. La volatilité de ce pourcentage de citation des candidat.e.s peut être un peu plus forte pour les émissions relativement courtes de notre corpus (comme M6).

On représente sur la figure 20 l'écart à la moyenne, en point de pourcentage, pour chaque semaine / chaîne / candidat.



FIGURE 20 – Différence de points de pourcentage par rapport à la moyenne des chaînes, chaque semaine, automne 2021

On distingue plusieurs blocs avec les constats déjà fait sur les données brutes. Les médias privés d'information en continu, radio et infotainment surmédiasent l'extrême-droite et sous-représentent Macron à l'inverse de l'audiovisuel public et des JT du soir sur les chaînes généralistes.

On peut utiliser un simple score de  $X^2$  pour comparer les distributions de chaque chaîne des pourcentages par rapport à la distribution moyenne des chaînes. Sur la figure 21, avec cette première analyse, on voit que les colonnes de TPMP, France2, France3 et M6 et la ligne Zemmour sortent du lot et ont les distributions les plus éloignées de la moyenne. La plus forte "anomalie" concerne le couple Zemmour/TPMP.

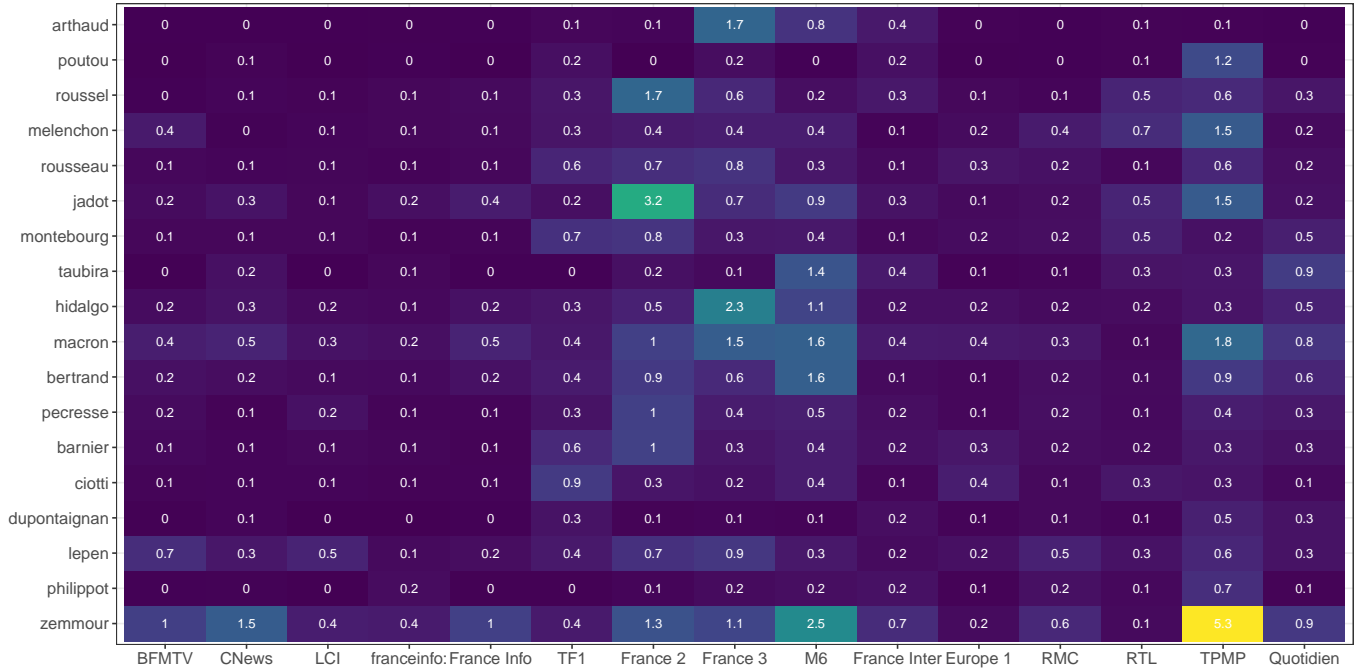


FIGURE 21 – Scores  $X^2$  média  $\times$  candidat

## 4.2 Textométrie

Nous présentons rapidement deux approches classiques en textométrie : l'analyse factorielle de correspondances (AFC) (Benzécri, 1973) et le calcul de spécificités (Lebart and Salem, 1994). On considère dans les deux cas le nombre de citations de chaque candidat.e par chaîne et par semaine.

L'AFC<sup>5</sup> est présentée sur la figure 22, on rappelle que l'interprétation de ce type de graphique est délicate, notamment pour les distances entre les différents points.

Le calcul de spécificités est souvent employé en analyse de corpus. On sait toutefois que les valeurs ont vite tendance à saturer lorsque le nombre de documents devient trop important, ce qui est notre cas ici. On présente donc les résultats<sup>6</sup> sur la figure 23 uniquement à but illustratif. Les valeurs sont automatiquement limitées à  $\pm 300$ .

5. Nous utilisons notre package R *myCA*

6. Nous utilisons le package R *textometry* du logiciel TXM



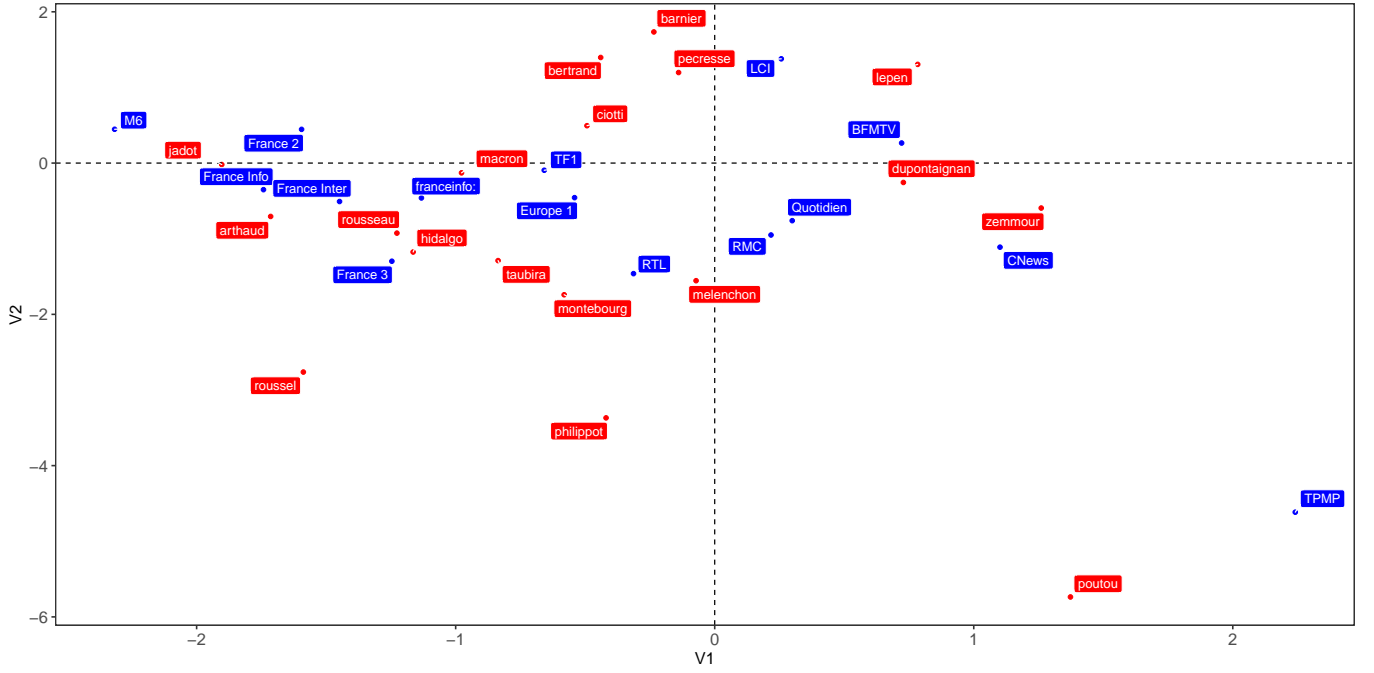


FIGURE 22 – AFC chaînes/candidats sur les données de l'automne 2021

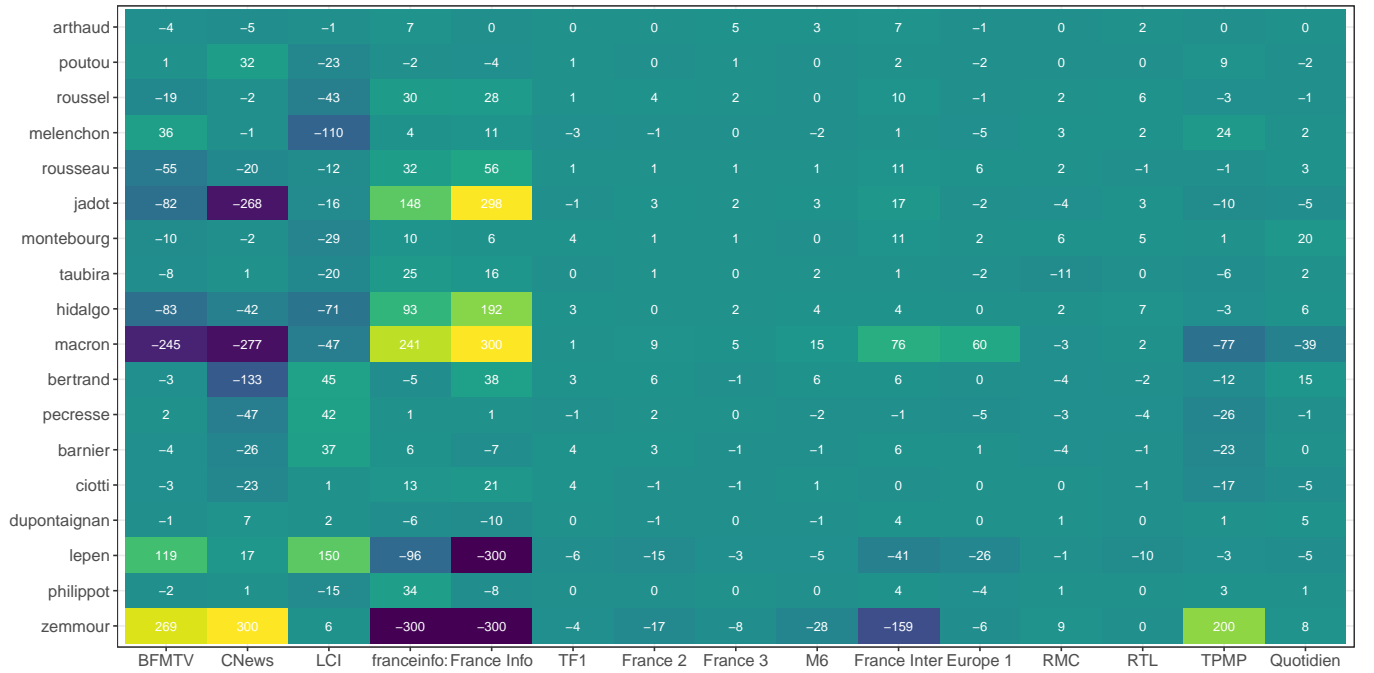


FIGURE 23 – Spécificités chaînes/candidats sur les données de l'automne 2021

### 4.3 Régression linéaire

L'idée est ici de modéliser la proportion respective de citation de chaque candidat.e à l'antenne en tenant compte des deux principaux facteurs explicatifs : l'actualité particulière de chaque candidat.e et les préférences potentielles des chaînes. Pour cette première modélisation, on ne regarde que les proportions, en faisant donc abstraction de l'importance du traitement politique sur chaque chaîne ou de la longueur des créneaux considérés. Une modélisation plus poussée sera présentée dans la prochaine section, mais ce premier modèle a l'avantage d'être simple et facilement interprétable.

Nous proposons comme modèle additif la régression linéaire avec effets fixes suivante :

$$pct_{m,c,t} = \theta_{w,c} + \eta_{m,c} + \epsilon_{m,c,w,t} \quad (1)$$

—  $pct_{m,c,t}$  pourcentage de citations du candidat  $c$  par le média  $m$  pour la période  $t$  (jour ou semaine)

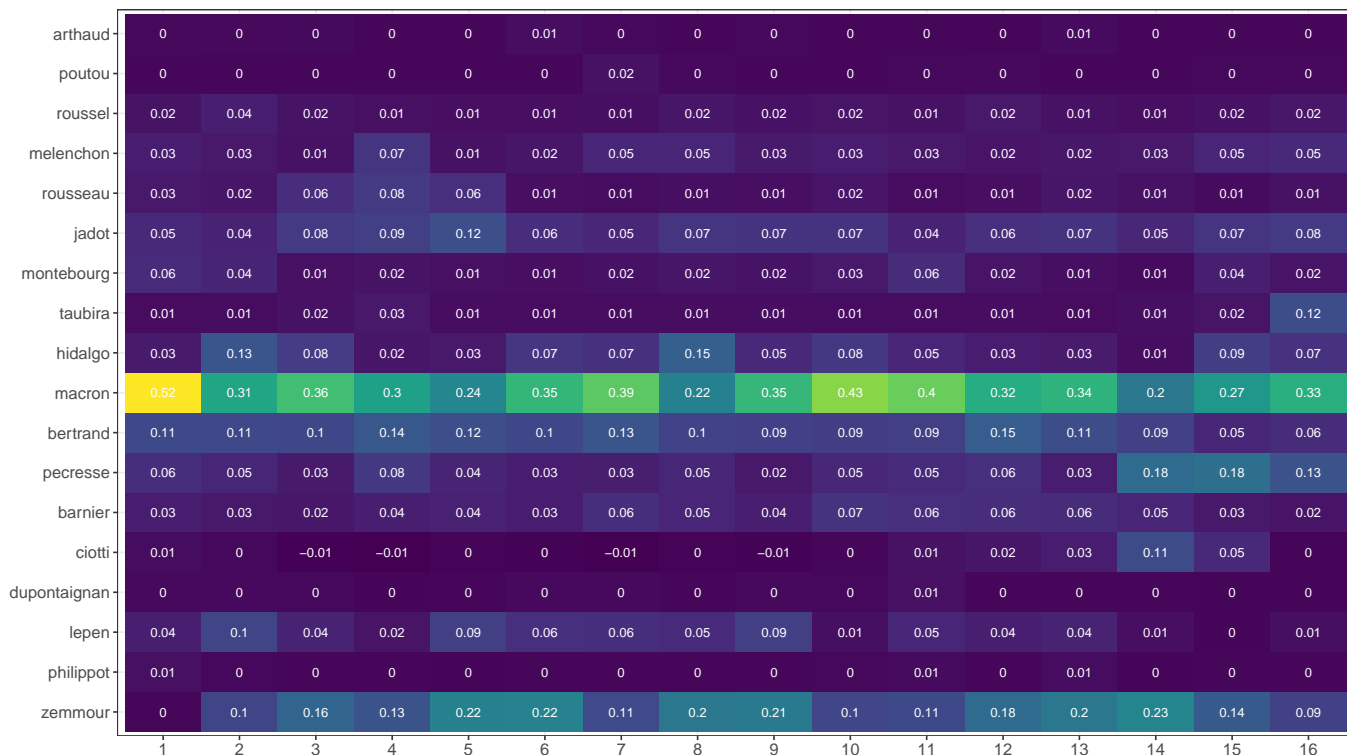


FIGURE 24 – Modèle linéaire 6 -  $\theta_{w,c}$  coeff. semaine  $\times$  candidat, automne 2021

- $\theta_{w,c}$  : actualité particulière d'un.e candidat.e pour une semaine donnée qui peut justifier un surcroît ou une baisse globale de sa médiatisation
- $\eta_{m,c}$  : affinité particulière d'un.e candidat.e et d'un média
- $\epsilon$  : terme d'erreur

On considère donc que le pourcentage de citations d'un.e candidat.e est décomposable en deux éléments distincts : son actualité de la semaine  $\theta_{w,c}$  (du lundi au dimanche) et les préférences du média  $\eta_{m,c}$ . C'est ce deuxième effet fixe qui nous intéresse ici, permettant de gommer l'actualité des candidats pour ne retenir que les affinités chaîne/candidat.e.

L'estimation du modèle peut être faite avec les données au niveau jour ou au niveau semaine. Sur les 1 616 jour  $\times$  média, 45 n'ont strictement aucun candidat cité, le nombre d'observations est donc 28 278. Au niveau semaine, on a en revanche bien toutes les 4 320 observations. On utilise France 2 comme chaîne de référence pour l'estimation.

	Modèles jour			Modèles semaine		
	(1)	(2)	(3)	(4)	(5)	(6)
EF semaine $\times$ candidat	✓		✓	✓		✓
EF média $\times$ candidat		✓	✓		✓	✓
Observations	28 278	28 278	28 278	4 320	4 320	4 320
Coefficients	288	270	540	288	270	540
R2	0.5134	0.5247	<b>0.5707</b>	0.8002	0.806	<b>0.8929</b>
Adjusted-R2	0.5084	0.5201	<b>0.5624</b>	0.7859	0.7931	<b>0.8776</b>

TABLE 6 – Modèles de régression linéaire, automne 2021

Comme on le voit dans la table 6, les modèles 3 et 6, qui cumulent les deux types d'effet fixe, permettent une meilleure modélisation de nos données. La variabilité des pourcentages de citation étant plus faible lorsqu'on agrège les données par semaine, le  $R^2$  est meilleur sur le modèle 6 que sur le 3, mais cela n'influence que très peu les coefficients obtenus. On présente les coefficients du modèle 6 dans les figures 24 et 25.

Les effets fixes  $\theta_{w,c}$  de la figure 24 correspondent donc à l'actualité spécifique de chaque candidat.e pour les 16 semaines du corpus de l'automne 2021. On y voit naturellement la forte présence de Macron, président en exercice mais également en seconde position celle de Zemmour qui a été fortement médiatisé à la rentrée.

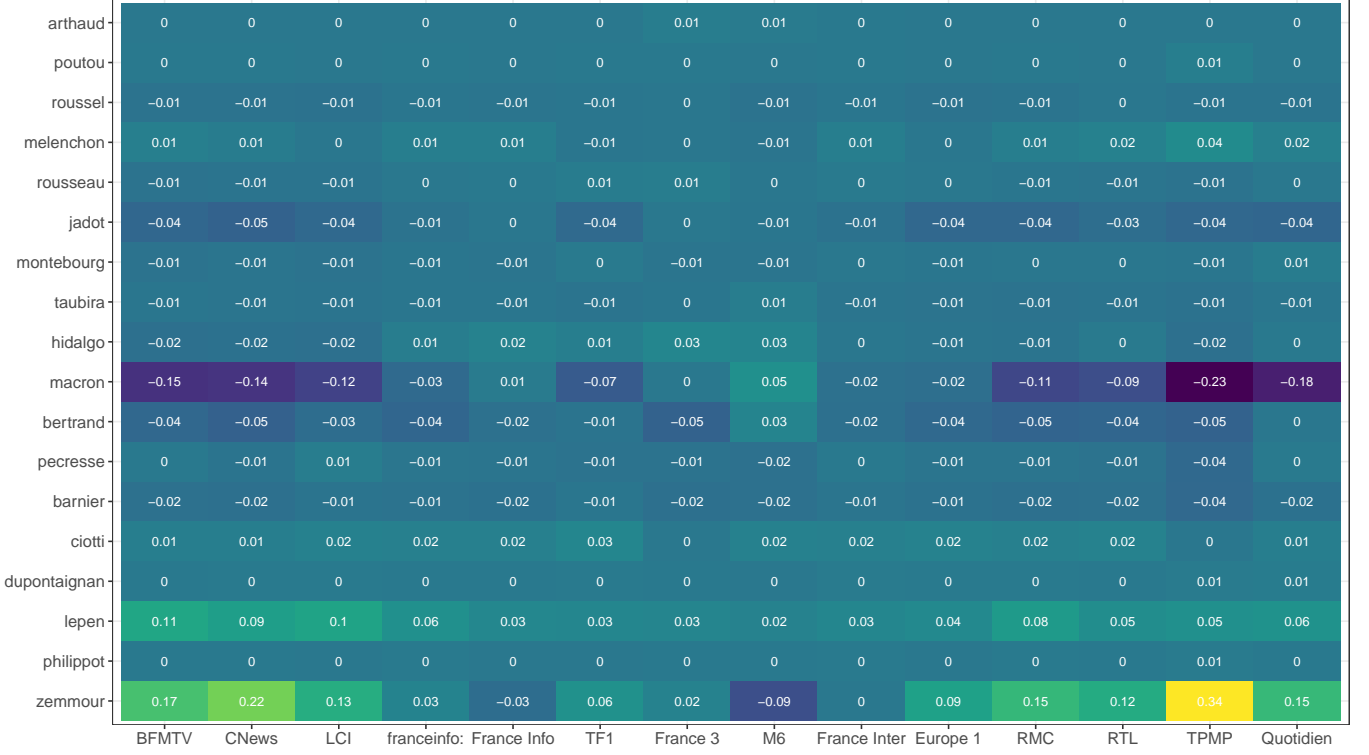


FIGURE 25 – Modèle linéaire 6 -  $\eta_{m,c}$  coeff. média  $\times$  candidat (réf. France 2), automne 2021

On peut également observé une inversion entre Bertrand et Pécresse sur les dernières semaine du corpus, au moment de la primaire de la droite début décembre 2021.

Ce qui va plus nous intéresser sont les effets fixes  $\eta_{m,c}$  de la figure 25 qui présentent les différence de pourcentage de citation des candidat.e.s par les médias, relativement à France 2. Les chiffre de ce graphique ne sont pas des données brutes, mais bien des écarts. Nous ne disposons pas d'un distribution de référence, calculée selon les critères du CSA, qui indiquerait la répartition idéale des temps d'antenne (et par extrapolation des citations). Puisque cette information n'existe pas, et ne peut pas exister, nous devons nous contenter d'informations relatives. Ainsi, on peut toutefois dire qu'une fois l'effet de l'actualité des candidat.e.s gommée, TPMP a accorder 34 points de pourcentage de citation en plus à Zemmour que France 2 et que cette même émission a sous-cité Macron de 23 points de pourcentage. France Inter est au même niveau que France 2. Ainsi dans l'ordre, la sur-médiatisation de Zemmour à l'automne, en terme de citation à l'antenne, est le fait de TPMP (0.34), CNews (0.22) et BFMTV (0.17). Suivent ensuite Quotidien, RMC et RTL. La deuxième candidate ayant une préférence marquée de la part de certains médias est Le Pen sur BFMTV (0.11), LCI (0.10) et CNews (0.09).

#### 4.4 Régression de Poisson

Puisque nous sommes sur une étude longitudinale visant à compter un nombre d'évènements, il peut être plus judicieux d'utiliser une régression de Poisson pour modéliser notre corpus. Nous ajoutons donc les effets fixes nécessaires pour tenir compte des données brutes et des facteurs explicatifs potentiels de leur variabilité. On propose alors la modélisation suivante :

$$\log(nb_{m,c,d}) = \lambda_m + \gamma_w + \mu_{d,m} + \theta_{w,c} + \eta_{m,c} + \epsilon_{m,c,d,w} \quad (2)$$

On cherche donc à estimer  $nb_{m,c,d}$  le nombre de citations du candidat  $c$  par le média  $m$  pour le jour  $d$ . On a les effets fixes suivants :

- $\lambda_m$  : lié au média, pour capturer, notamment, la typologie des émissions, la longueur des transcriptions et la propension à citer des candidat.e.s, c'est à dire à parler de l'actualité politique
- $\mu_{d,m}$  : lié au jour de la semaine, certaines des émission de notre corpus ne sont pas diffusées ou ont des durées différentes le week-end
- $\gamma_w$  : lié à la semaine afin de tenir compte de l'actualité politique globale d'une semaine, quel que soit le candidat

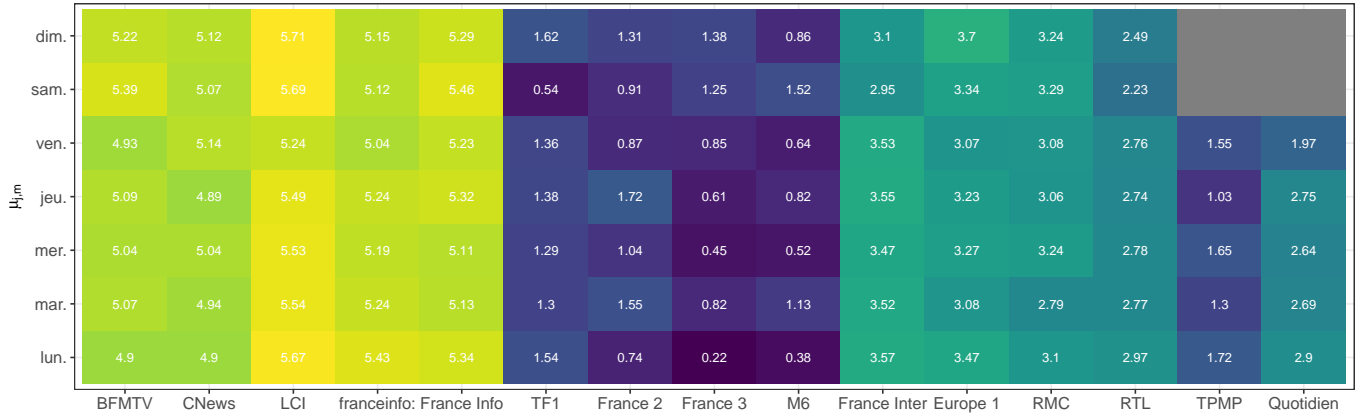


FIGURE 26 – Modèle Poisson 5 -  $\mu_{d,m}$  coeff. jour de la semaine  $\times$  média, automne 2021

- $\theta_{w,c}$  : actualité particulière d'un.e candidat.e pour une semaine donnée qui peut justifier un surcroît ou une baisse globale de la médiatisation
- $\eta_{m,c}$  : affinité particulière d'un candidat et d'un média

Nous estimons ce modèle avec différentes configurations des effets fixes présentés dans la table 7. Les seuls effets fixes qui nous intéressent sont, là encore, les  $\eta_{m,c}$ . Pour l'estimation, on utilise les références suivantes : Macron et la première semaine de notre corpus. Dans la suite, seuls les résultats du modèle 5, qui contient tous nos effets fixes pertinents, seront analysés dans une prochaine version de l'étude.

	Modèles				
	(1)	(2)	(3)	(4)	(5)
EF média			✓		
EF semaine		✓			
EF jour de la semaine $\times$ média		✓			✓
EF média $\times$ candidat	✓	✓		✓	✓
EF semaine $\times$ candidat			✓	✓	✓
Observations	29 088	29 088	29 088	29 088	29 088
Coefficients	270	375	303	540	630
Pseudo-R2	0.8550	0.8730	0.8956	0.9057	<b>0.9085</b>
AIC	492 019	437 857	369 275	339 252	<b>331 000</b>

TABLE 7 – Modèles de régression Poisson, automne 2021

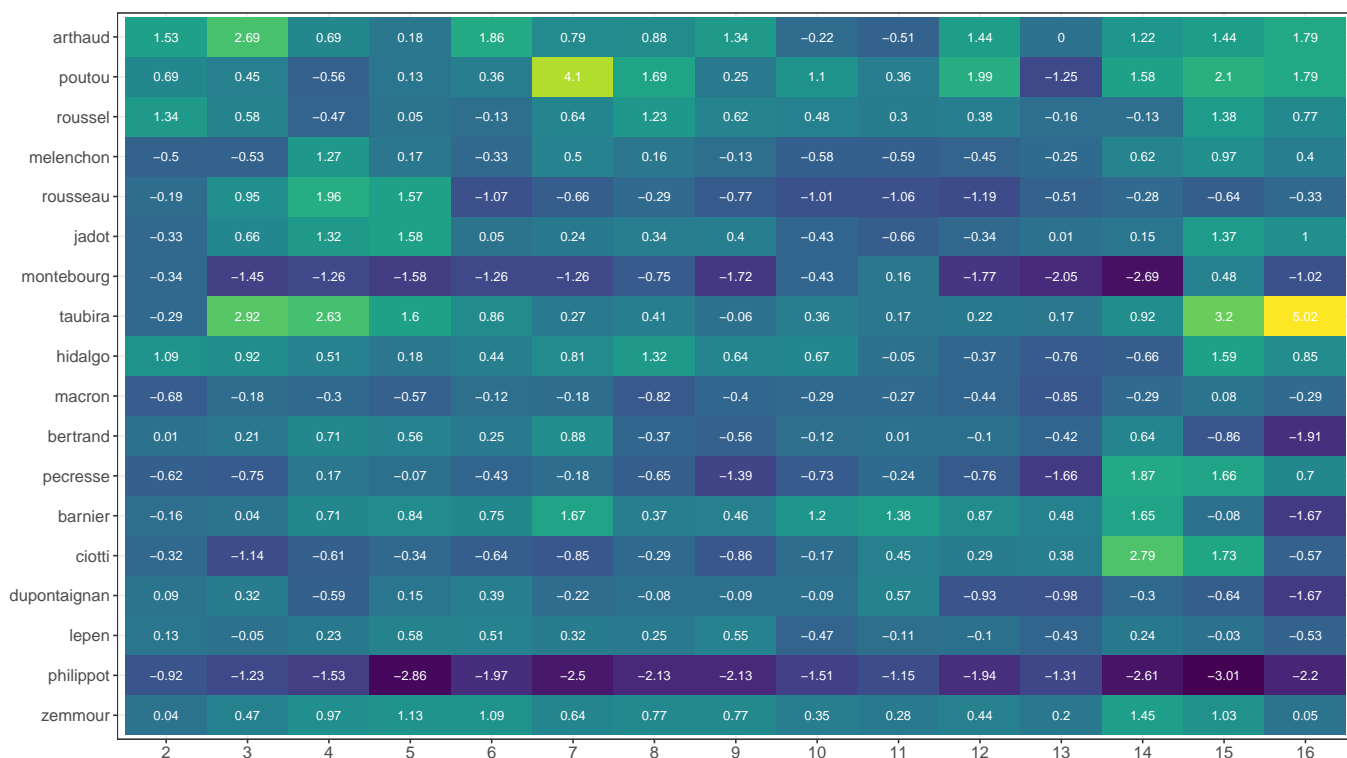


FIGURE 27 – Modèle Poisson 5 -  $\theta_{w,c}$  coeff. semaine  $\times$  candidat (réf semaine 1), automne 2021

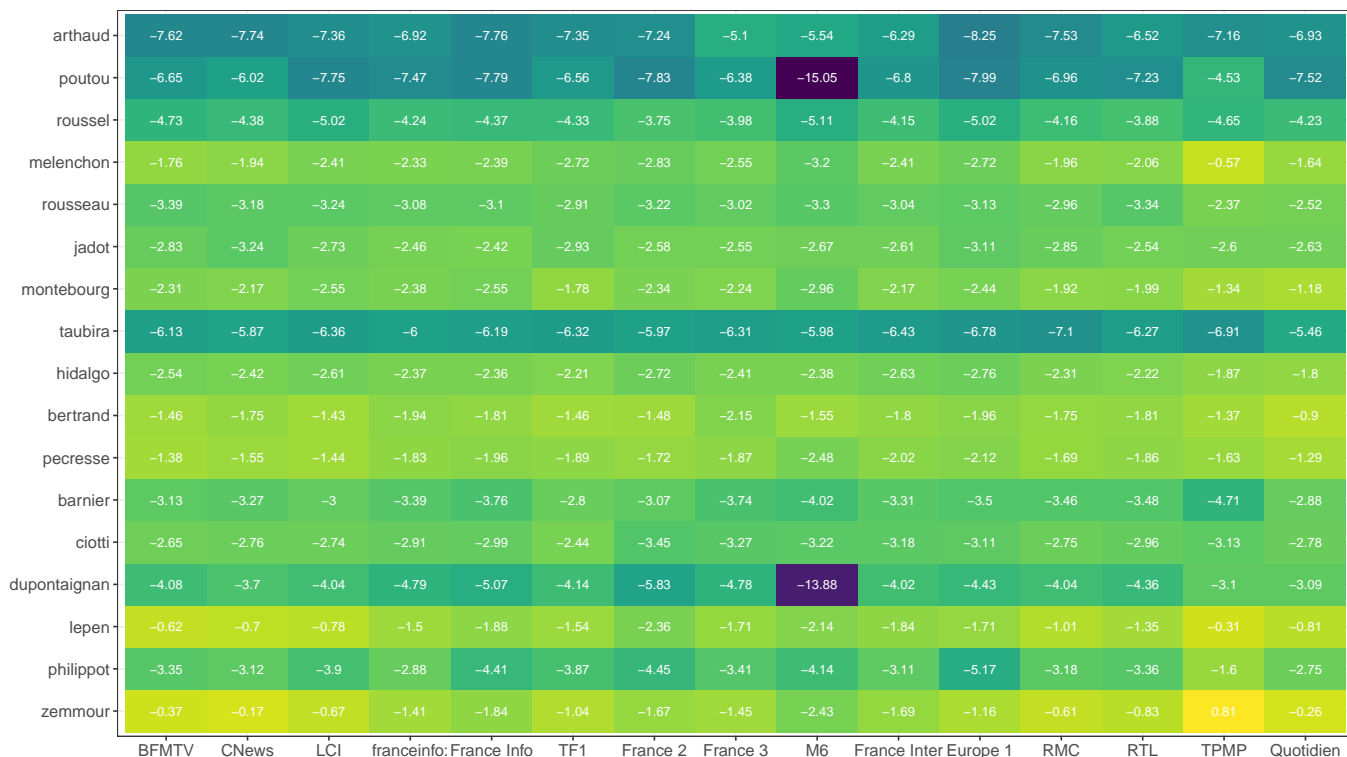


FIGURE 28 – Modèle Poisson 5 -  $\eta_{m,c}$  coeff. média  $\times$  candidat (réf. Macron), automne 2021

## 5 Début 2022, période d'équité, modélisations

Nous reprenons la visualisation des écarts à la moyenne et la régression linéaire pour la période d'équité du début de l'année 2022 (techniquement nous prenons les semaines complètes allant donc du 3 janvier 2022 au 6 mars 2022).

### 5.1 Approche naïve

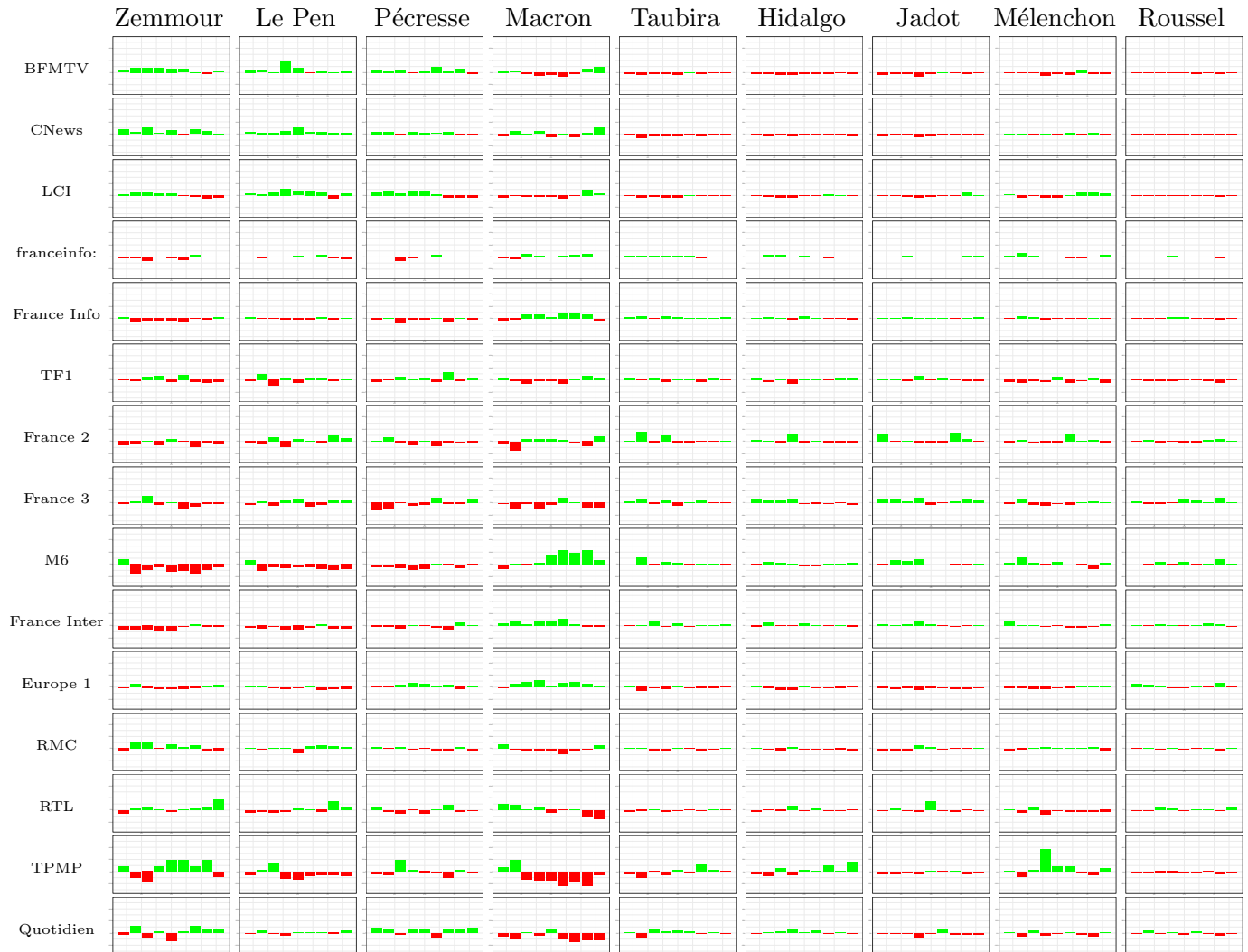


FIGURE 29 – Différence de points de pourcentage par rapport à la moyenne des chaînes, chaque semaine du 3 janvier 2022 au 6 mars 2022 (période d'équité)

### 5.2 Régression linéaire

	Modèles jour			Modèles semaine		
	(1)	(2)	(3)	(4)	(5)	(6)
EF semaine $\times$ candidat	✓		✓	✓		✓
EF média $\times$ candidat		✓	✓		✓	✓
Observations	16 200	16 200	16 200	2 430	2 430	2 430
Coefficients	162	270	540	162	270	414
R2	0.5684	0.5441	<b>0.5957</b>	0.8781	0.8276	<b>0.9105</b>
Adjusted-R2	0.5641	0.5364	<b>0.5851</b>	0.8694	0.806	<b>0.8921</b>

TABLE 8 – Modèles de régression linéaire du 3 janvier 2022 au 6 mars 2022 (période d'équité)

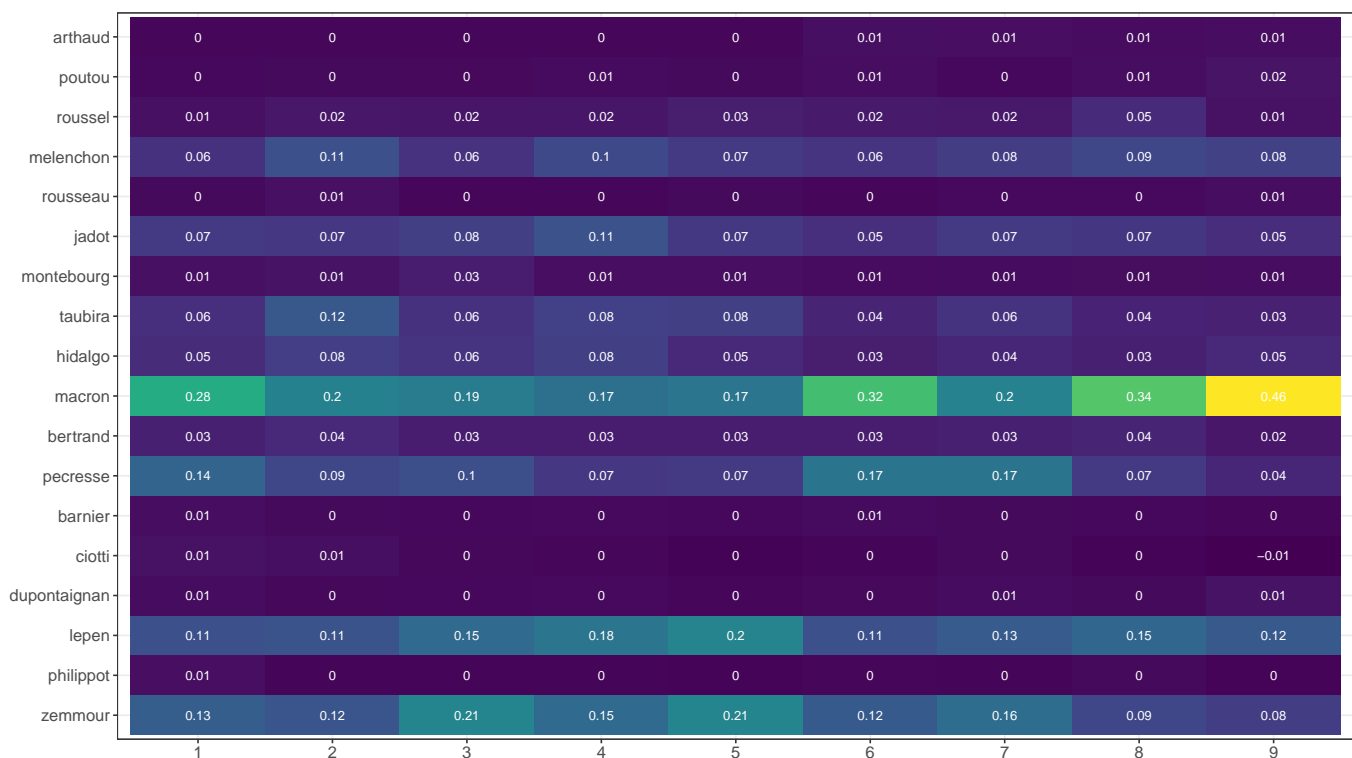


FIGURE 30 – Modèle 6 -  $\theta_{w,c}$  coeff. semaine  $\times$  candidat, du 3 janvier 2022 au 6 mars 2022 (période d'équité)

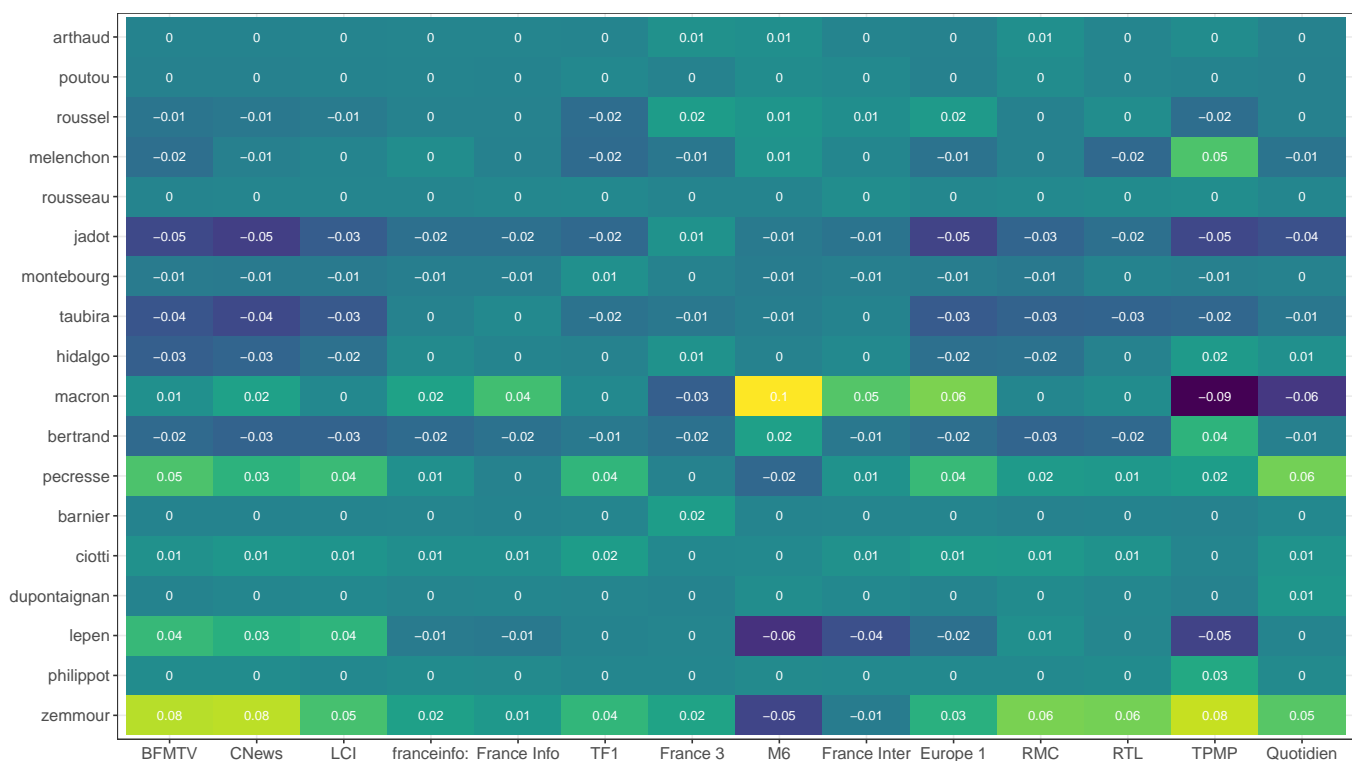


FIGURE 31 – Modèle 6 -  $\eta_{m,c}$  coeff. média  $\times$  candidat (ref France 2), du 3 janvier 2022 au 6 mars 2022 (période d'équité)

## Remerciements

Les travaux présentés dans ce document ont donné lieu à des échanges et des remarques de la part de Julia Cagé, Moritz Hengel, Camille Urvoy, Claire Sécaïl et Benjamin Guinaudeau. Je tiens à les remercier pour le temps qu'ils m'ont consacré. Toutes les erreurs, et oublis sont du seul fait de l'auteur.

## Références

- Clémence Bauduin. «c'est une dépression politique» : quand le covid-19 éclipse l'élection présidentielle. *Le Parisien*, 2022. [Voir en ligne].
- Jean-Paul Benzécri. L'Analyse des Correspondances. In *L'analyse Des Données*, volume 2. Dunod Paris, 1973.
- Julia Cagé, Moritz Hengel, Nicolas Hervé, and Camille Urvoy. Hosting Media Bias : Evidence from the Universe of French Broadcasts, 2002-2020. In *CEPR Workshop on Media, Technology, Politics, and Society*, 2022. [Voir en ligne].
- Romain Geoffroy, Raphaëlle Aubert, and Gary Dagorn. Election présidentielle 2022 : visualisez les temps de parole de chaque candidat dans les médias. *Le Monde*, 2022. [Voir en ligne].
- Seth Jolly, Ryan Bakker, Liesbet Hooghe, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen, and Milada Anna Vachudova. Chapel hill expert survey trend file, 1999–2019. 75 :102420, 2022. ISSN 0261-3794. [Voir en ligne].
- Ludovic Lebart and André Salem. *Statistique textuelle*. Paris : Dunod, 1994.
- Quentin Meunier. Tout comprendre aux règles d'équilibre du temps de parole dans la campagne présidentielle. *La Revue des Médias*, 2021. [Voir en ligne].
- Claire Sécaïl. L'élection présidentielle 2022 vue par Cyril Hanouna. 1. La pré-campagne (automne 2021), 2022. [ Voir en ligne ].
- Natalia Tomashenko, Kévin Vythelingum, Anthony Rousseau, and Yannick Estève. LIUM ASR systems for the 2016 multi-genre broadcast arabic challenge. In *IEEE Workshop on Spoken Language Technology*, 2016. [ Voir en ligne ].

## Historique

Version	Date	Auteur(s)	Modifications
1.0	16 janvier 2022	NH	résultats préliminaires
1.1	06 février 2022	NH	mise à jour des données, suppression des candidats trop peu cités, suppression de France 24, ajout de Quotidien, nouveaux types de graphes, qualité de la transcription
1.2	17 mars 2022	NH	mise à jour des données, ajout des données du CSA, modélisation économétrique



## Annexes

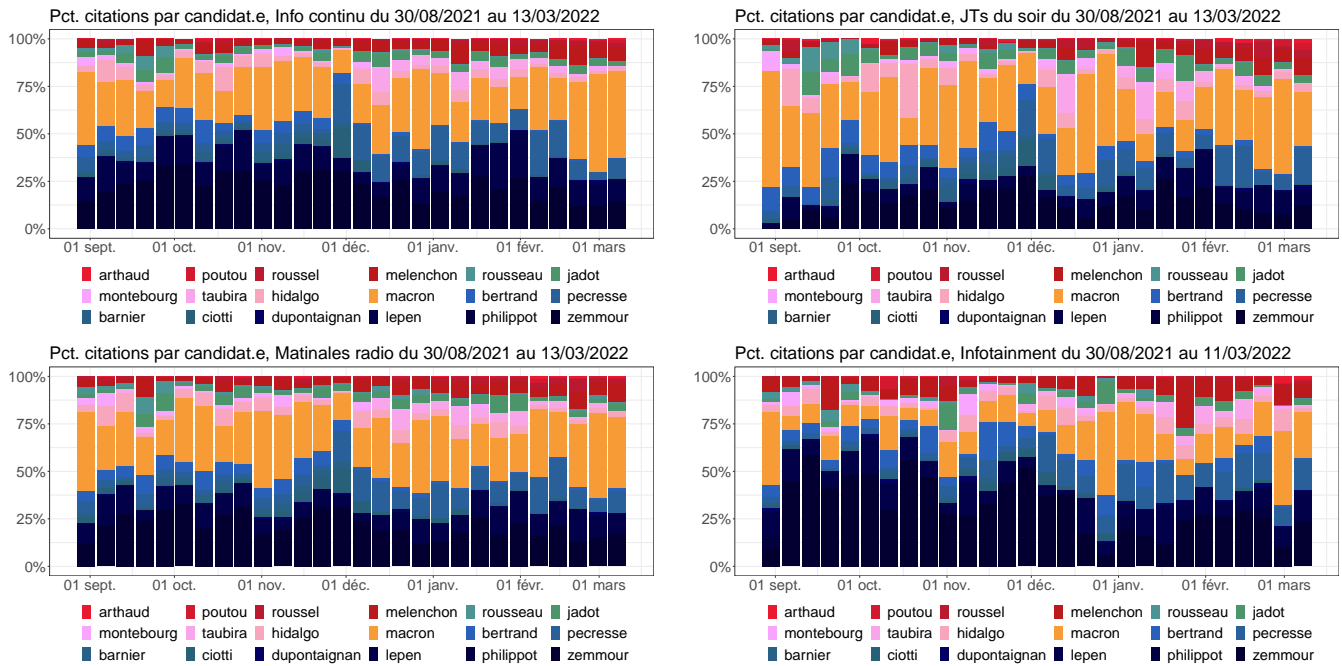


FIGURE 32 – Proportion moyenne de citations par candidat et par semaine

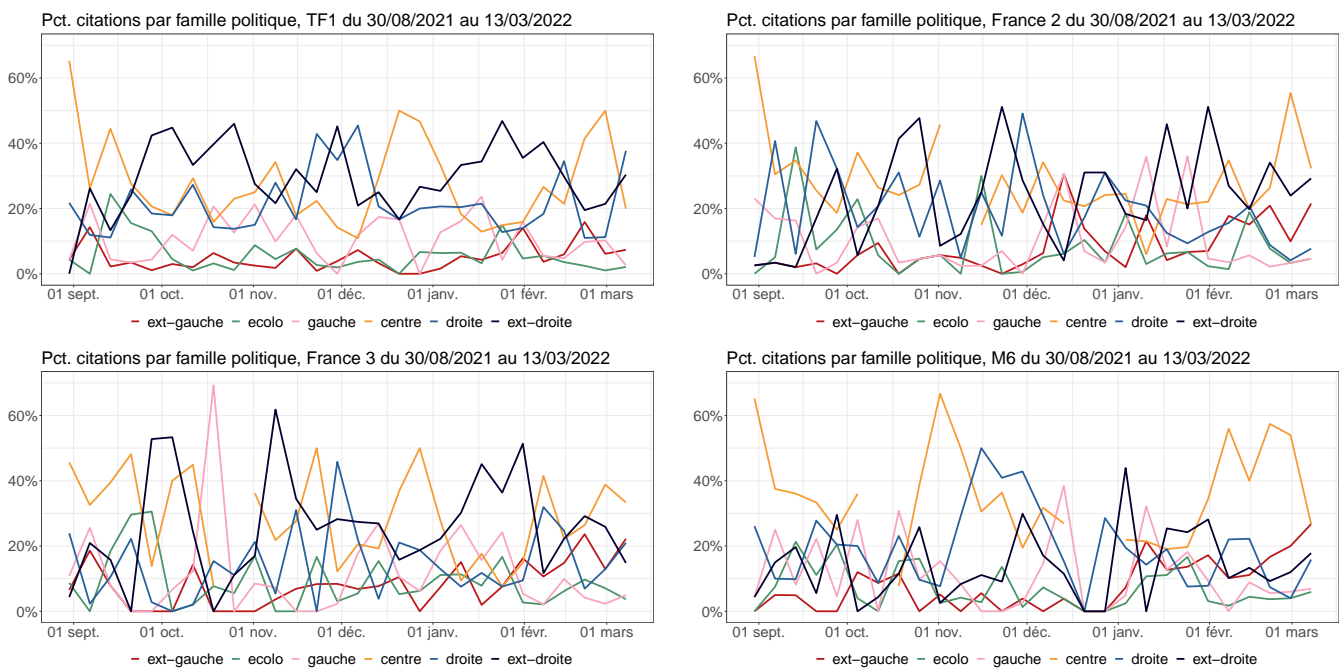


FIGURE 33 – Proportion moyenne de citations par famille politique pour les JT du soir

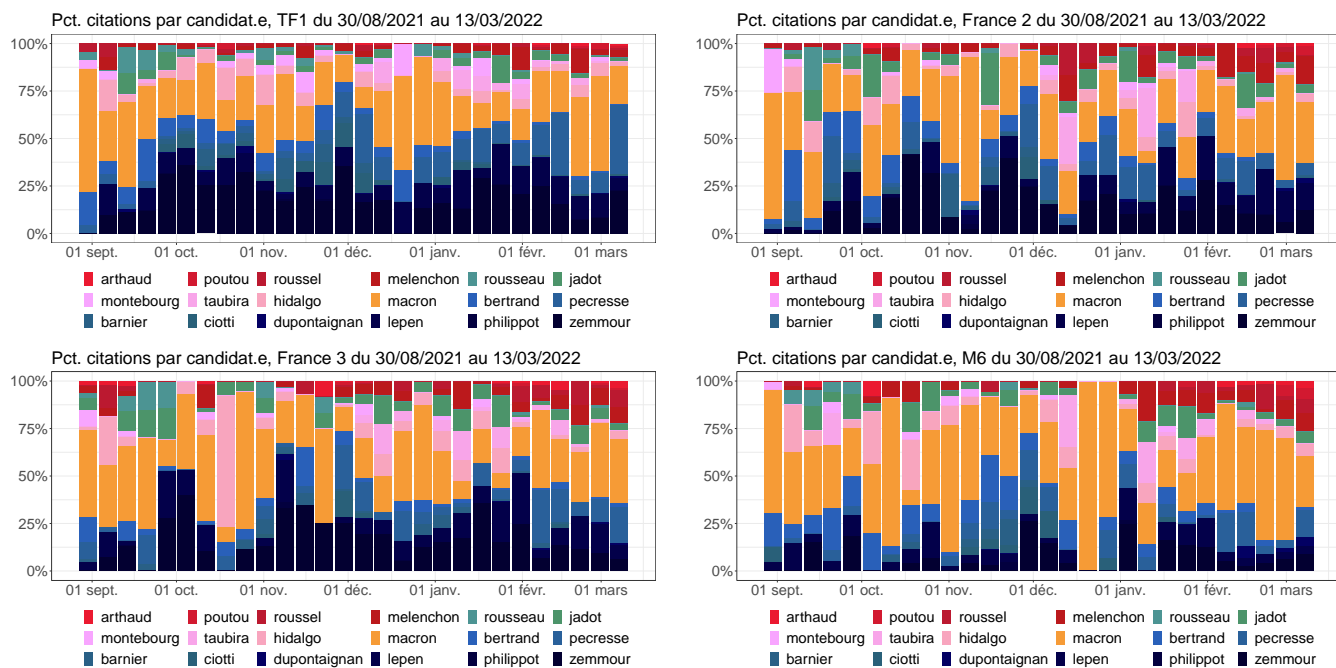


FIGURE 34 – Proportion moyenne de citations par candidat.e pour les JT du soir

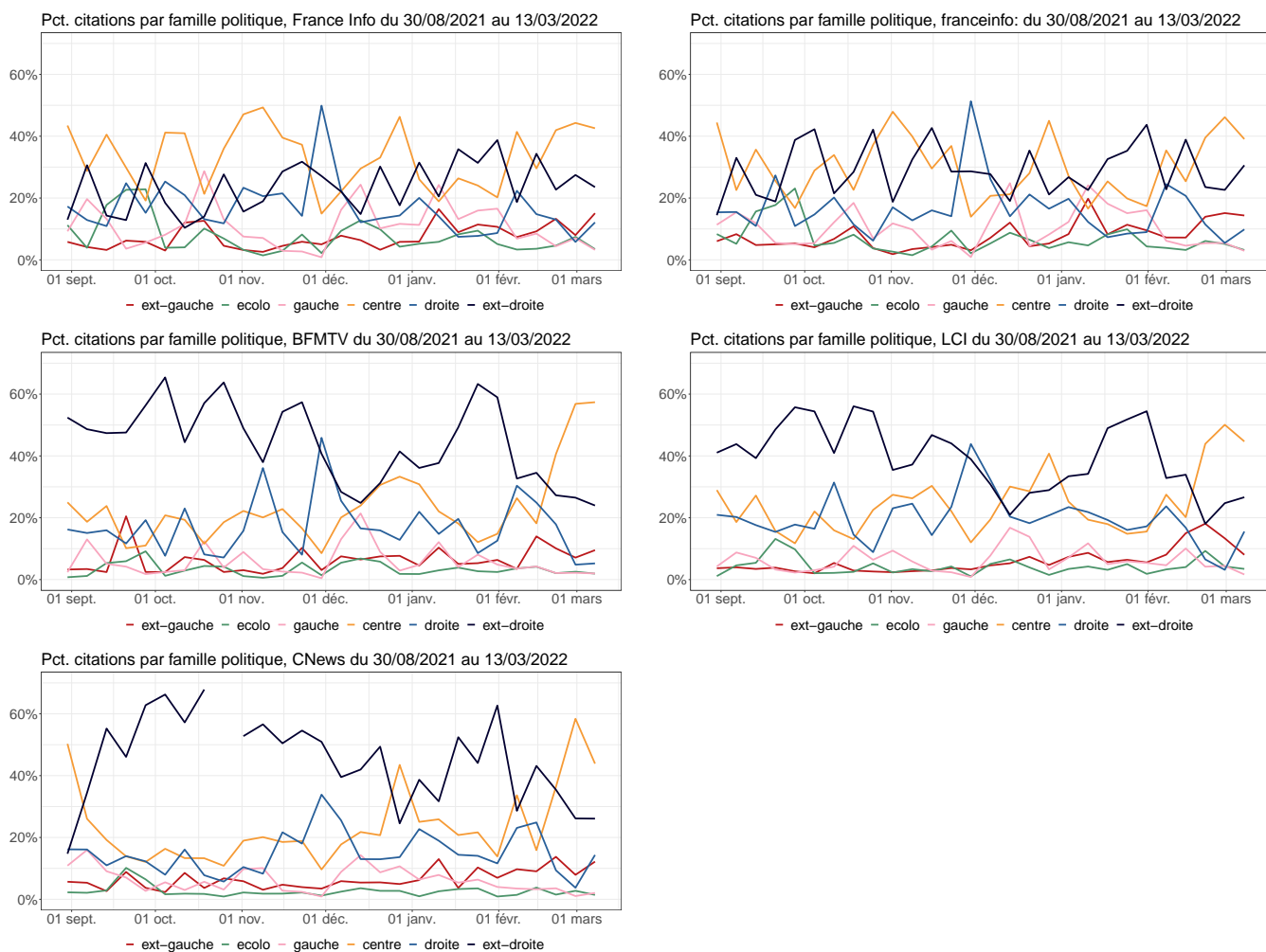


FIGURE 35 – Proportion moyenne de citations par famille politique pour les médias d'information en continu

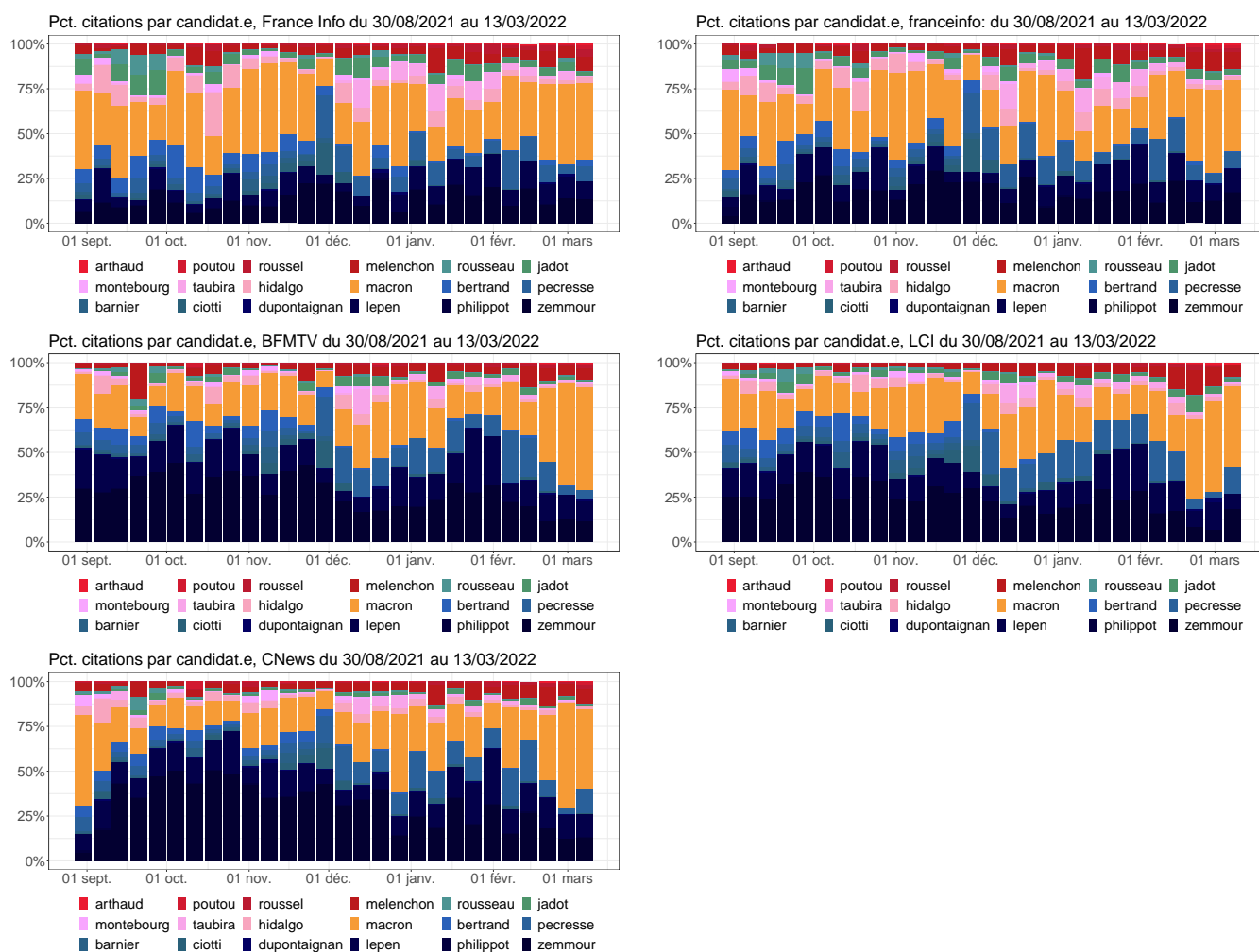


FIGURE 36 – Proportion moyenne de citations par candidat.e pour les médias d'information en continu

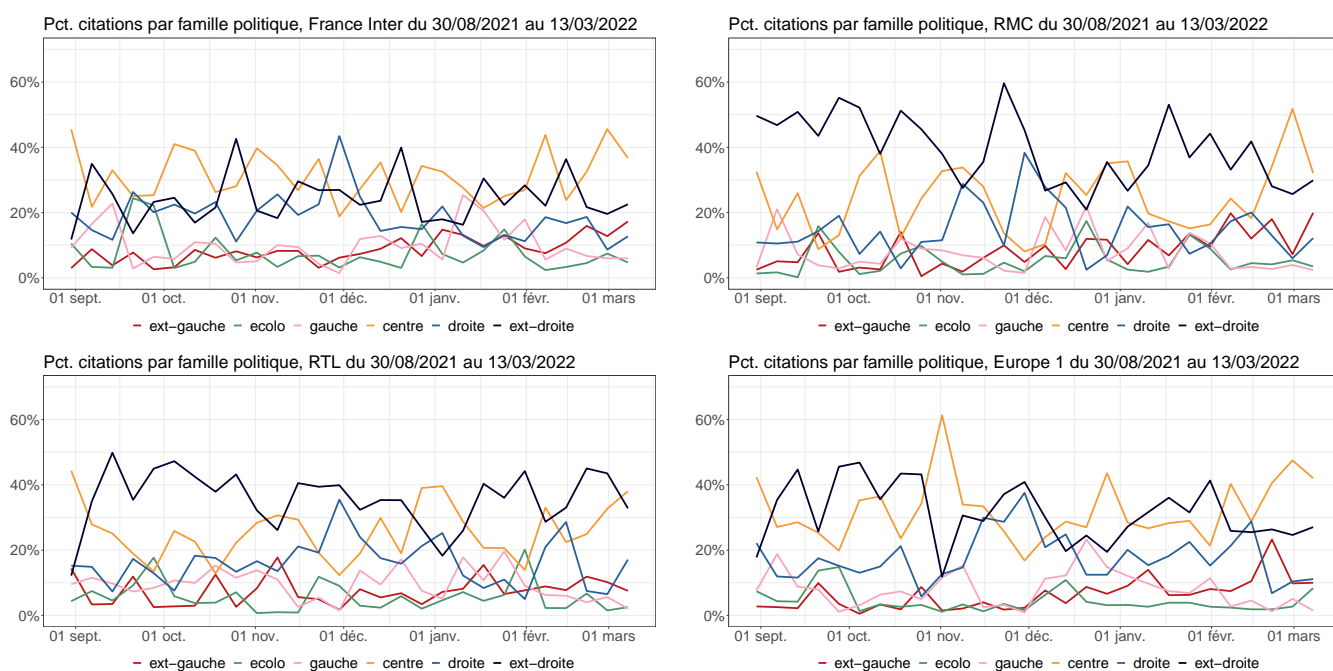


FIGURE 37 – Proportion moyenne de citations par famille politique pour les matinales radio

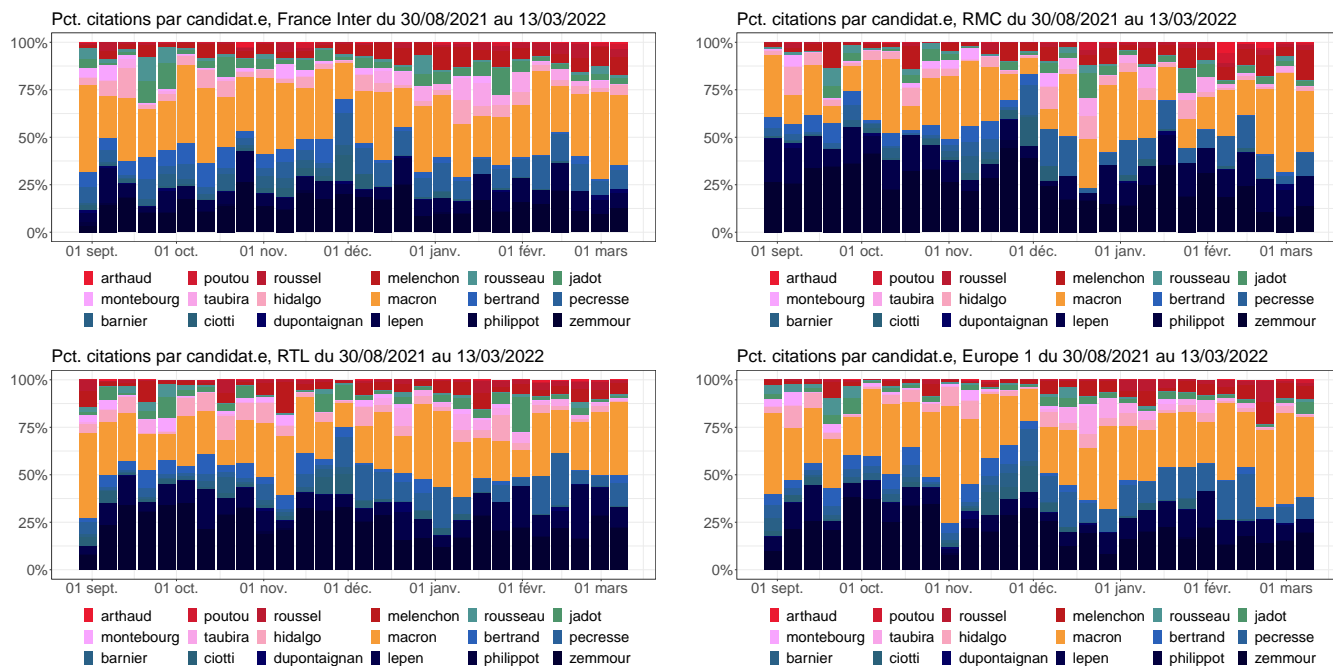


FIGURE 38 – Proportion moyenne de citations par candidat.e pour les matinales radio

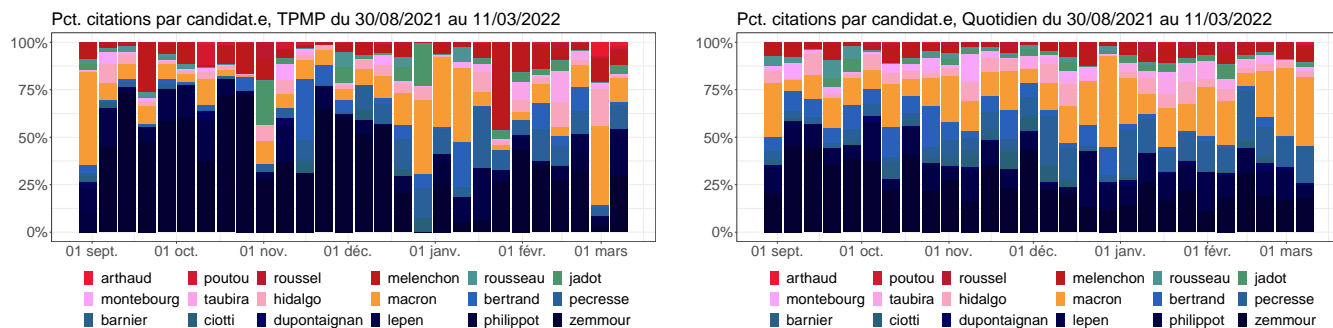


FIGURE 39 – Proportion moyenne de citations par candidat.e pour l'infotainment