

Design : Equipe UQAM 1

Composition de l'équipe

Notre équipe est composée de 7 personnes :

- 3 personnes de Polytechnique Montréal :
 - Justine Deloménie 1988908
 - Gautier Dervaux 1971606
 - Marina Menghetti 1971617
- 4 personnes de l'UQAM :
 - Mélissa Aubert
 - Jean-Baptiste Demouy
 - Hannah Raphaëlle Petitclerc-Vilandré
 - Théo Sardaigne

Rappel du sujet

Etude et visualisation de comment sont vus les différents journaux sur Twitter. Comparaison de la presse française et québécoise, ainsi que de la presse télévisée VS presse écrite et internet.

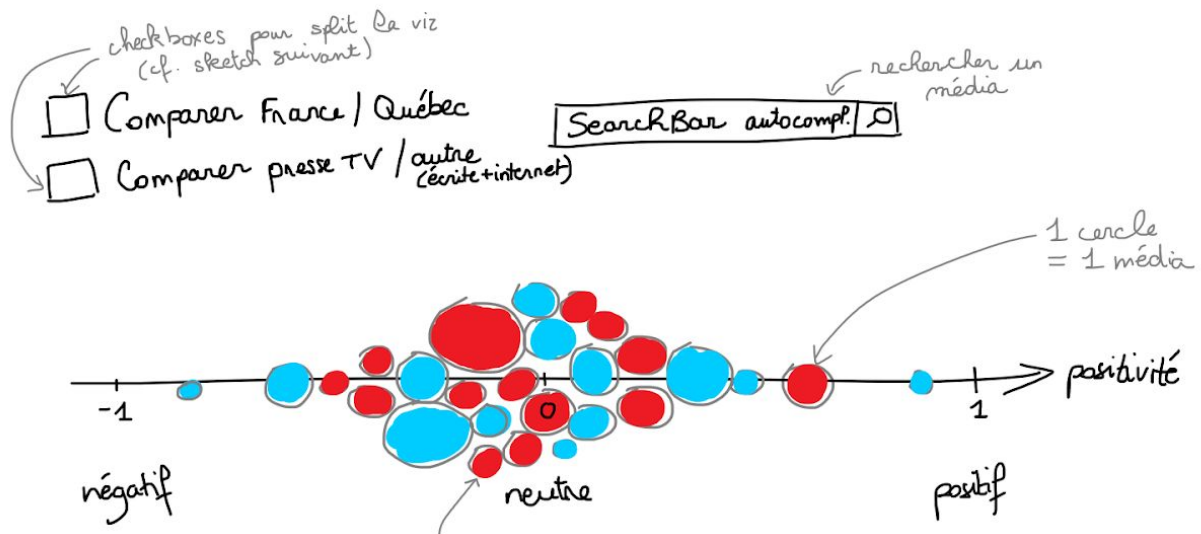
Récupération et processing des données

Les données sont récupérées grâce à l'API de Twitter, en cherchant différents mots-clés. Pour la visualisation la plus complète, nous recherchons des termes du type "@nom_du_compte_twitter_du_journal". Ceci nous permet de filtrer plus ou moins bien les tweets s'adressant directement au journal pour donner des critiques positives ou négatives.

Un algorithme permettant de récupérer un score de positivité/négativité de sentiments dans les tweets est utilisé. La librairie Vader semble donner les meilleurs résultats de façon assez rapide. Chacun des termes est associé à un score positif négatif ou neutre grâce à un entraînement déjà réalisé par la librairie, sur des données de réseau sociaux. La librairie prend également en compte l'utilisation de smileys, ce qui est très intéressant pour notre analyse. L'analyse d'un tweet donne sa fraction de positivité, sa fraction de neutralité, et sa fraction de négativité. Un coefficient récapitulatif entre -1 et 1 est ensuite calculé. Afin d'appliquer cette librairie sur des tweets écrits en français, nous devons traduire les tweets analysés (par exemple à l'aide de la librairie TextBlob ou autre). Nous arrivons à des résultats plutôt bons, avec quelques erreurs parfois. Cependant, dans certains cas, un tweet peut parler positivement d'un article sur un sujet très négatif. C'est dans ce genre de cas que les résultats peuvent ne pas bien correspondre. D'autres erreurs peuvent venir d'une mauvaise traduction vers l'anglais, ou bien de la non compréhension de certains mots écrits avec une orthographe plus ou moins approximative.

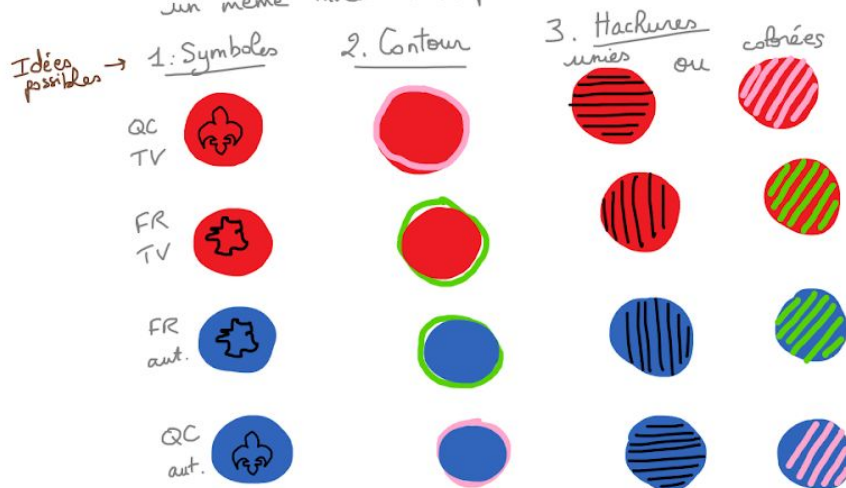
Idées de design

Nous partons sur une visualisation principale montrant la répartition des différents journaux écrits et télévisés, de la France et du Québec, sur un même graphique. La position des médias sur l'axe de positivité représente une moyenne de la positivité des tweets correspondants.

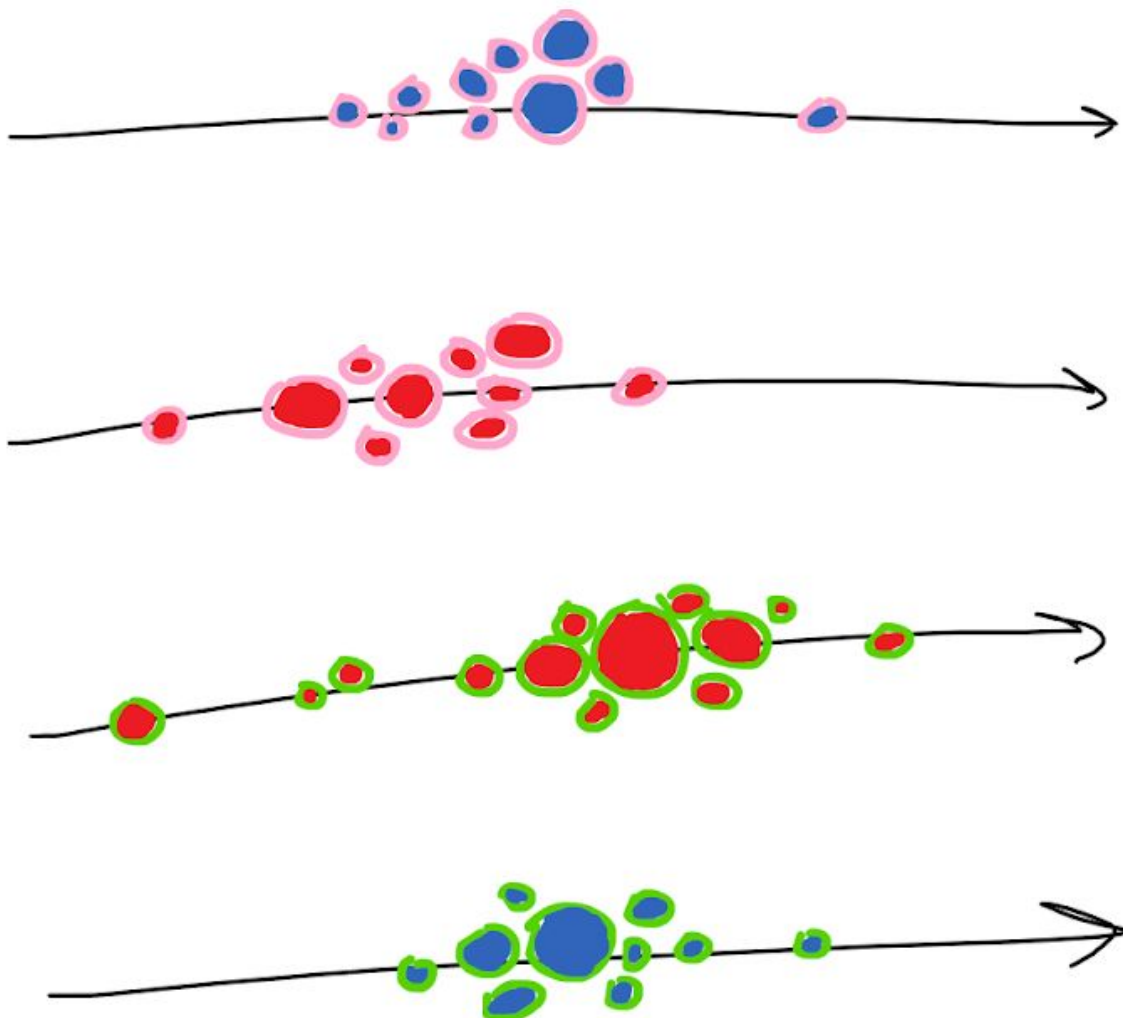


ENCODAGE DES MÉDIAS

- * taille : À quel pt un média est connu ?
 → nombre de followers / abonnés sur twitter
 Prendre en compte la différence du nombre d'habitants en France et au Québec ?
 → Diviser par le nombre d'habitants ⇒ nb de followers / habitant
- * couleur : Catégorie : France ou Québec ? TV ou écrite/internet ?
 Quelle comparaison favoriser ? Peut-on trouver un encodage pour les mettre sur un même niveau d'importance ?



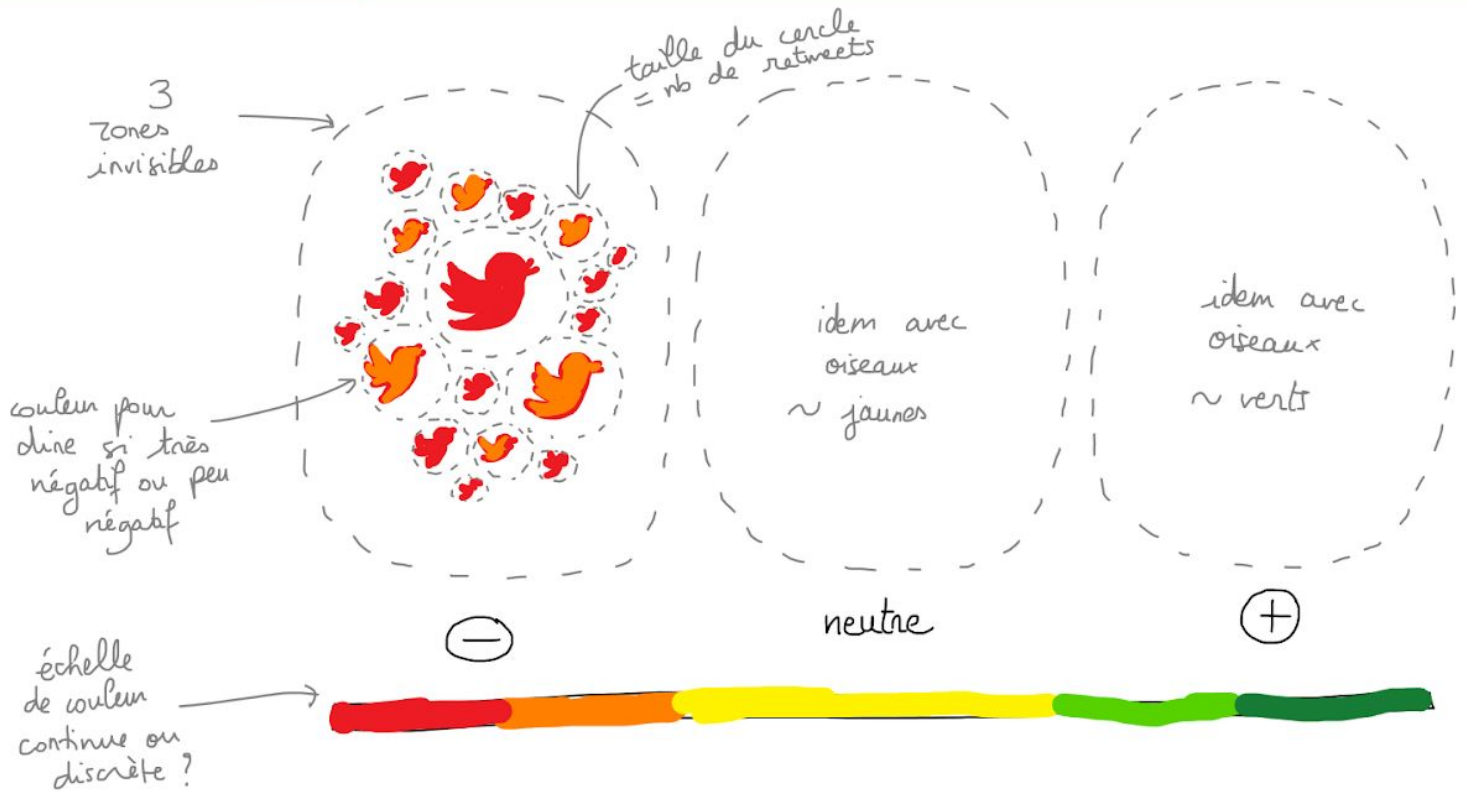
En cochant les checkbox correspondantes dans la première visualisation, il est possible de séparer la visualisation en 2, puis en 4 si les 2 checkboxs sont cochées. Ceci permettrait de mieux visualiser la répartition par catégorie. Une transition permettrait de voir les cercles se déplacer d'un graphique à l'autre. Le sketch ci-dessous montre le graphique séparé en 4 par catégories avec l'encodage par contours colorés pour l'exemple. Il est également possible que le graphique ne soit séparé qu'en 2 si une seule case est cochée. Une légende sera bien sûr présente.



Sur chacun de ces graphiques, il est possible à l'aide d'une barre verticale ou autre d'indiquer où se situe la moyenne/médiane, pondérée par la taille des cercles par exemple.

On notera également la présence d'un tip au survol (non représenté dans les schémas), qui décrirait : le nom du média (éventuellement un logo), sa catégorie TV/autre et pays/province, son nombre de tweets récupérés pris en compte pour la visualisation, nombre de followers Twitter (ou nombre de followers par habitant en pourcentage par exemple).

En sélectionnant un journal grâce à la barre de recherche ou en cliquant sur un cercle, il est ensuite possible de visualiser plus précisément les tweets récupérés utilisant le “@nom_journal_cliqué” .



En survolant un tweet, plusieurs informations sur le tweet seront montrées : le texte du tweet, le nombre de retweets, la positivité du tweet par exemple. La taille des tweets serait en fonction du nombre de retweets+1 (car le tweet de base compte aussi). Les zones peuvent être visibles ou délimitées par une forme autour de la zone.