

Parcoursup - DataESR

Julien Gossa

Dans son effort de transparence, le gouvernement a publié des données nationales sur les deux premières campagnes Parcoursup¹, ainsi que la dernière campagne APB. Ces données présentent à la fois un immense intérêt et de sérieuses limites.

Description des données

Ces données sont orientées filières de formation. Chaque ligne est une formation présentée sur la plateforme Parcoursup et proposées aux candidats à l'entrée dans le supérieur. On y trouve des indications comme l'établissement qui donne la formation, l'intitulé précisé de la formation, le nombre de candidats et de candidates, leurs profils, et le profil des admis et admises.

Ce faisant, ces jeux de données ne permettent pas de comprendre les campagnes d'affectation dans le supérieur du point de vue des candidat.e.s. Notamment, il est impossible de connaître les temps d'attente avant proposition, ou le degré de satisfaction. Tou.tes les candidat.e.s pourraient aussi bien être affecté.e.s dans leur dernier choix, que ça ne se verrait pas. Il serait opportun qu'en plus des données actuellement disponibles, le ministère de l'enseignement supérieur et de la recherche, conjointement avec d'autres ministères qui habilitent aussi des formations, fasse une enquête sur le degré de satisfaction après deux ans par rapport aux vœux des étudiants sur Parcoursup.

De plus il est difficile de repérer les établissements qui relèvent des autres ministères (culture, agriculture, etc) que celui de l'enseignement supérieur et de la recherche. Il est donc très difficile de savoir si tous les établissements publics et privés relevant des différents ministères sont toutes répertoriées dans Parcoursup.

Ces jeux de données sont centrés sur Parcoursup. Les admissions hors de cette plateforme échappent donc à cette analyse. Dans cette limite, ils permettent de percevoir les caractéristiques des formations et les flux de néo-bacheliers admis dans le supérieur.

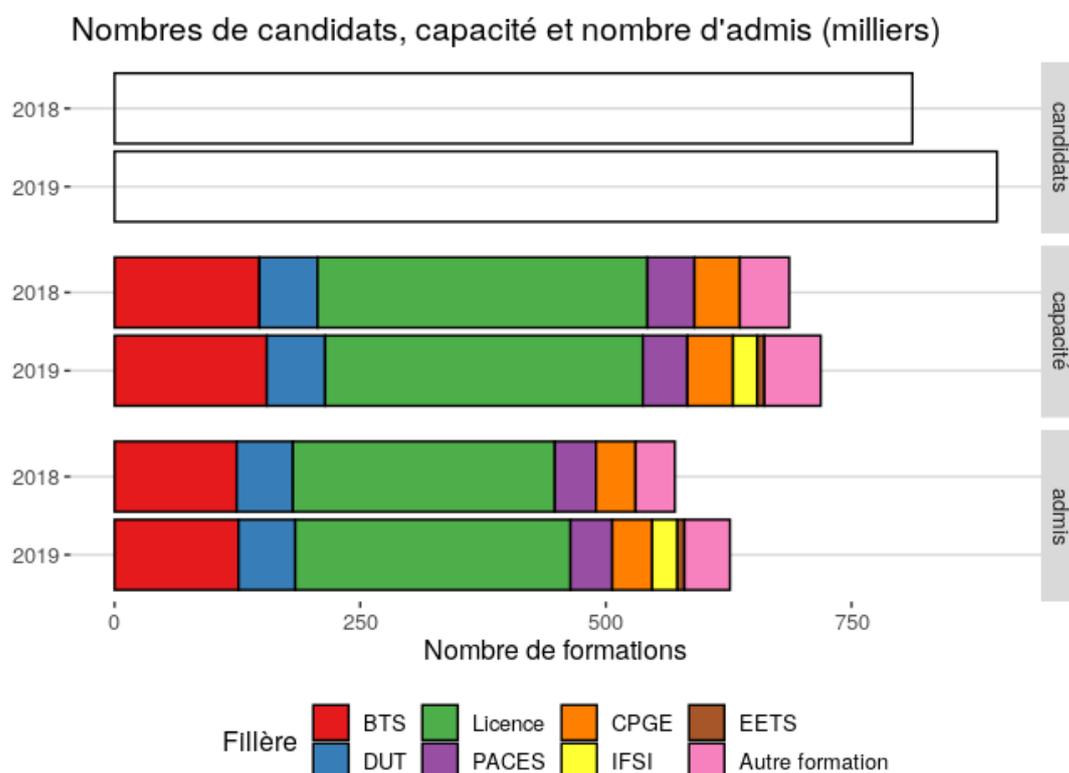
NB : un changement de méthodologie² a été annoncé qui pourrait empêcher les comparaisons entre les jeux de données 2018 et 2019. Le lecteur doit en rester conscient.

Statistiques générales sur les formations

En premier lieu, il convient de comparer le nombre de candidats (données ajoutées, non comprises dans les jeux de données originaux), le nombre de places et le nombre d'admis.

1 Source : <https://data.enseignementsup-recherche.gouv.fr/explore/dataset/fr-esr-parcoursup/information/>

2 https://data.enseignementsup-recherche.gouv.fr/api/datasets/1.0/fr-esr-parcoursup/attachments/methodo_opendata_2019_pdf/



Le décalage entre le nombre de candidats et le nombre d'admis s'explique entre autres par l'exclusion des candidats en reprise d'étude de ce jeu de données, alors qu'ils sont présents dans le nombre de candidats.

L'augmentation de 2018 à 2019 s'explique d'ailleurs essentiellement par une hausse de ce nombre de candidats en reprise d'étude³, et l'intégration des EETS (Educateur technique spécialisé) et IFSI (Institut de formation en soins infirmiers) qui ont attirés énormément de candidats pour très peu de places.

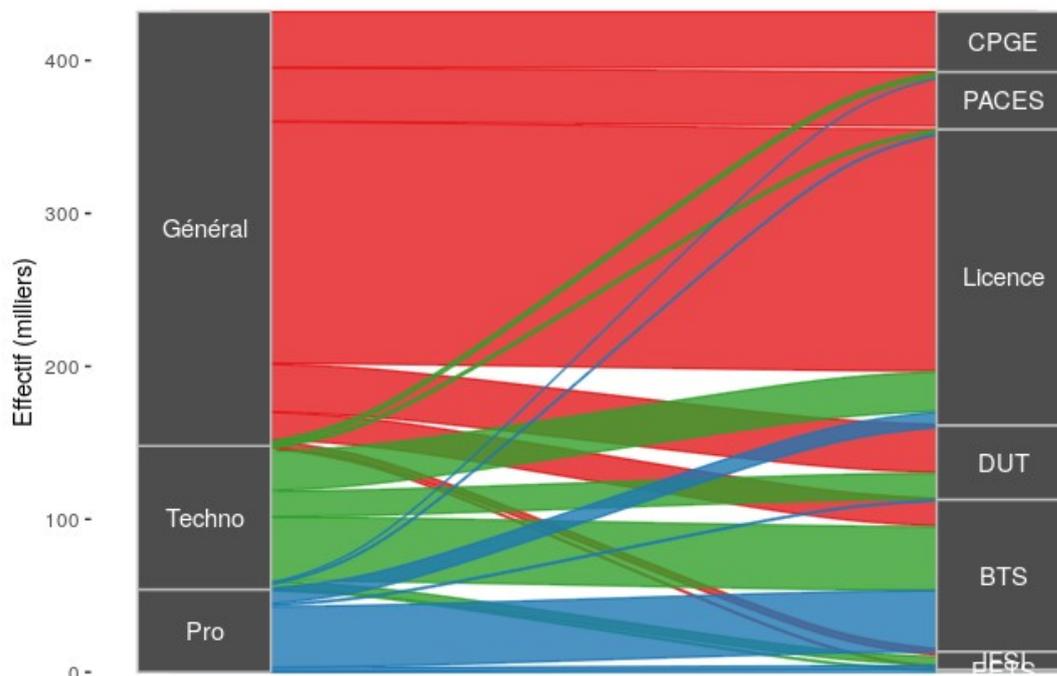
Une différence importante entre 2018 et 2019 est la baisse des capacités d'accueil en Licence, et la hausse d'admis dans cette filière. Il est fort probable que la baisse viennent d'un réajustement après la première année, où certaines capacités d'accueil ont pu être fixées sans grande rigueur.

Flux de néo-bacheliers entrant dans le supérieur

Les flux de néo-bacheliers admis dans le supérieur public peuvent se modéliser par un diagramme de Sankey, comportant à gauche le bac (général, technologique ou professionnel) et à droite les formations supérieures.

³ <https://start.lesechos.fr/apprendre/universites-ecoles/parcoursup-le-nombre-de-candidats-en-reprise-d-etudes-explose-15056.php>

Flux de néo-bacheliers admis dans le supérieur

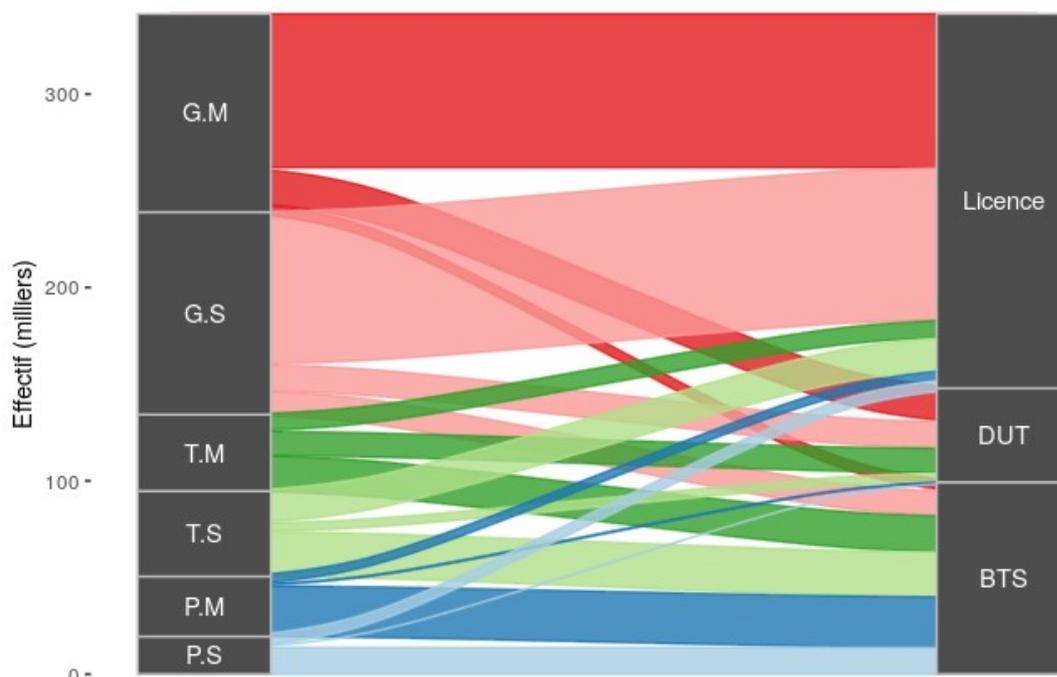


NB : Les deux dernières formations en bas à droite sont les IFSI et EETS

Ce diagramme permet de constater les inégalités de poursuite d'étude en fonction du bac : CPGE et PACES étant quasiment réservées au bac généraux, quand les autres formations panachent plus ou moins.

Pour plus de précision, on peut modéliser les flux avec (M) et sans mention (S) pour les trois bac (G, T et P) :

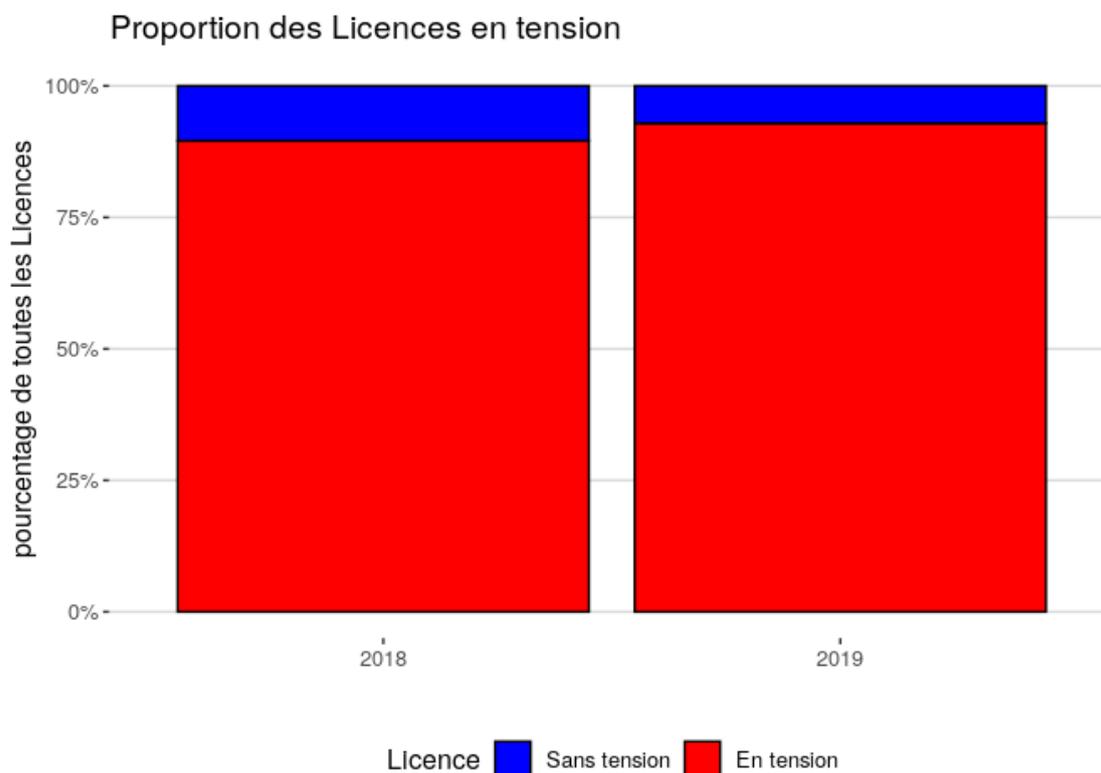
Flux de néo-bacheliers admis dans le supérieur



On constate alors que le flux de bacheliers technologiques et professionnels sans mention en Licence est très conséquent, ce qui pourrait confirmer le rôle de “filet de sécurité” de cette formation.

La sélection dans le supérieur

Au premier abord, le taux de « filière en tension » en Licence comme étant officiellement « lorsque le nombre de candidatures excède les capacités d’accueil d’une formation » :



On constate que la quasi totalité des Licences sont virtuellement en tension, et que la proportion augmente entre 2018 et 2019.

Il convient donc d’avoir une meilleure métrique pour définir la sélectivité d’une formation que la tension pour la sélection :

On peut définir le taux de sélection comme le rapport entre le nombre d’appelés (admis ou non, grâce au rang du dernier appelé), et le nombre de candidats en phase principale :

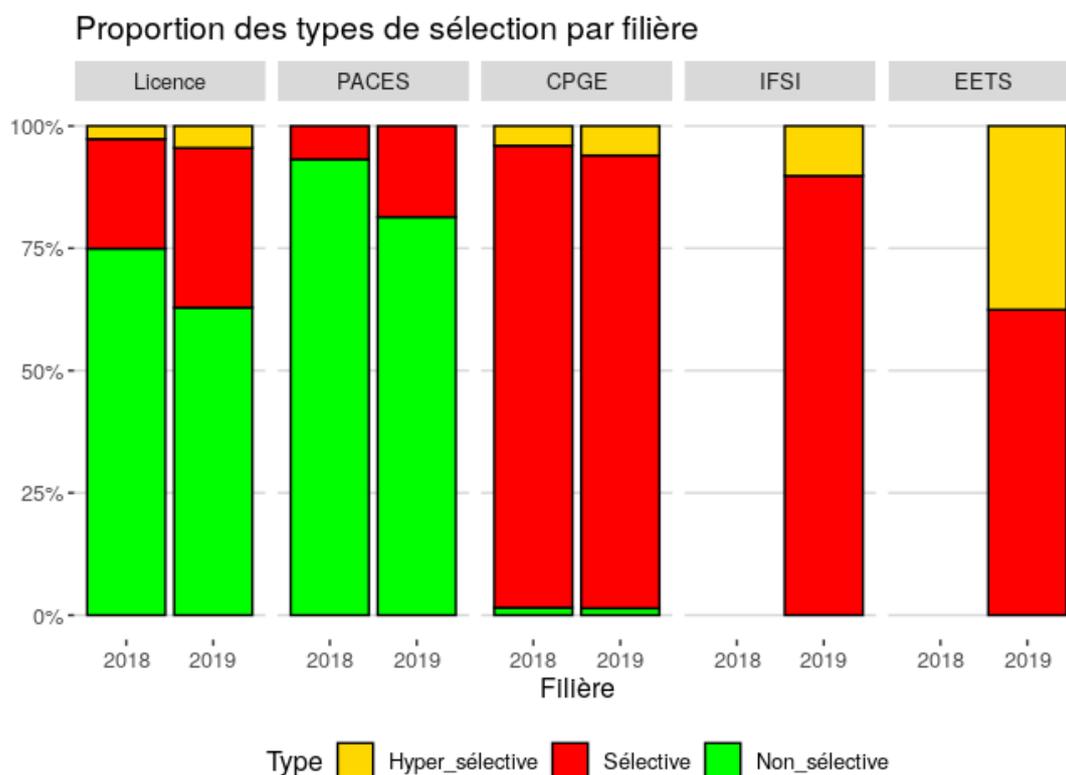
$$\textit{sélection} = \frac{\textit{rang.du.dernier.appelé}}{\textit{candidats}}$$

Interprétation : un taux de sélection de 100% signifie que tous les candidats ont été appelés. Plus le taux est bas, plus la formation est sélective.

La qualité des données impose une certaine souplesse sur cet indicateur, et c’est pourquoi sont qualifiés de :

- “Non-sélective” : les formation dont le taux de sélection est supérieur ou égal à 95% (Pour des questions de lisibilité, leur taux de sélection est forcé à 100%) ;
- “Sélective” : les formation dont le taux de sélection est inférieur à 95% ;
- “Hyper-Sélective” : les formation dont le taux de sélection est inférieur à 20%.

On peut ainsi regarder la proportion de ces trois types pour chaque filière :



N.B. : DUT et BTS ont été filtrés en raison de trop nombreuses absences du rang du dernier appelé pour ces formations dans le jeu de donnée.

D'abord, il faut remarquer que les IFSI et EETS sont des filières remarquablement hyper-sélectives, et ensuite que les taux de sélection augmentent dans toutes les filières.

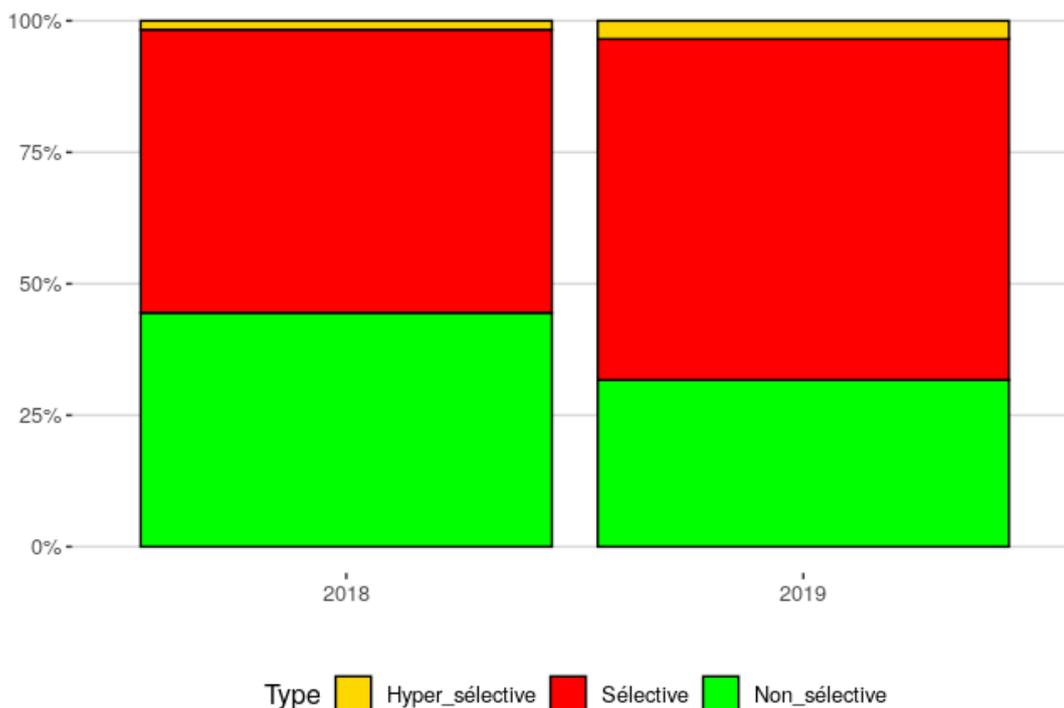
Ensuite, on constate que des CPGE, même en très faible nombre, sont également non sélectives. D'un autre côté, un tiers des Licences sont sélectives, voire hyper-sélectives. Cela démontre que l'appellation administrative « filière non sélective » n'a plus lieu d'être.

Il convient donc d'enlever toute trace de l'expression « filière non sélective » de Parcoursup et des textes de loi.

La sélection vue depuis la génération de bacheliers

Il est possible de mesurer quelle part des bacheliers a été admise dans une filière effectivement non sélective, et quelles parts ont été admises dans les filières effectivement sélectives ou hyper-sélectives :

Distribution des admissions par type de sélection



On constate sur ce graphique qu'un peu moins de la moitié des bacheliers ont été admis dans une filière effectivement non sélective en 2018, mais que cette proportion est tombée à un tiers en 2019. Ce taux sera amené à bouger au fil des années, selon les politiques nationales et locales d'organisation du premier cycle : la création de places apaisera la tension, quand la création de petites sections locales très sélectives l'accentuera.

Il convient donc que la proportion de bacheliers intégrant une formation non sélective fasse partie des indicateurs nationaux des campagnes Parcoursup.