

Exoplanets dataviz

Thomas Mauran



Figure 1: thumbnail

Les données

Mon jeu de donnée provient du site kaggle il est composé de 4575 lignes. Il contient des informations sur des exoplanètes, comme leur nom, la méthode utilisée pour les découvrir, l'année de leur découverte, etc. On y trouve également des détails sur les étoiles elles-mêmes, comme leur température, leur taille, leur masse, etc.

J'ai choisi de prendre ce dataset car l'espace est un sujet qui me passionne et qu'il semblait avoir assez de données pour être pertinent.

Load des données

Combien de planètes ont été découvertes chaque année et quelles méthodes ont été utilisées pour les découvrir ?

Population: L'ensemble des exoplanètes de l'univers

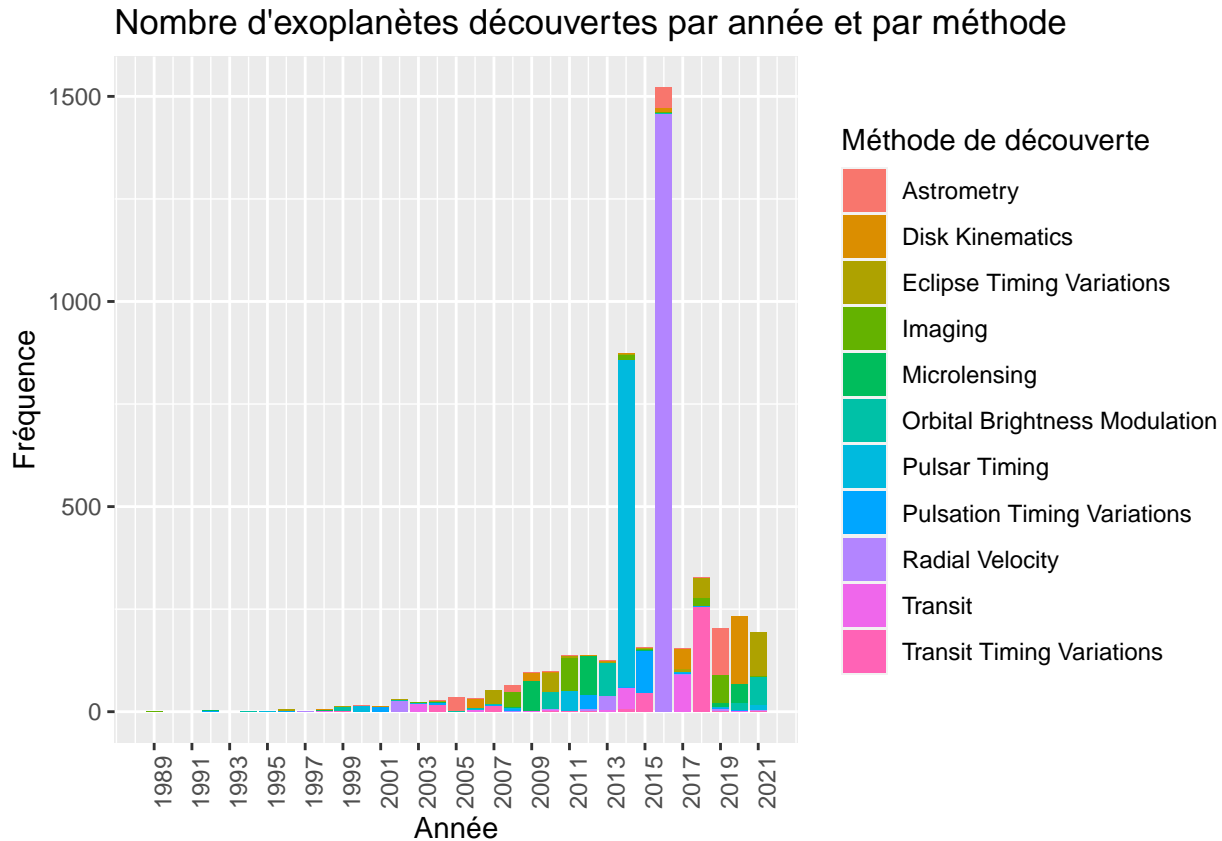
Echantillon: Les 4575 exoplanètes presentes dans le dataset

Variable:

- Le nombre d'exoplanètes découvertes
- Leurs année de découvertes
- Leurs méthode de découvertes

Modalités

- Le nombre de planète découvertes => nombre entre 0 et environ 1600
- année de découverte => Nombre entre 1989 et 2021
- Méthode => variable qualitative, méthode de découverte

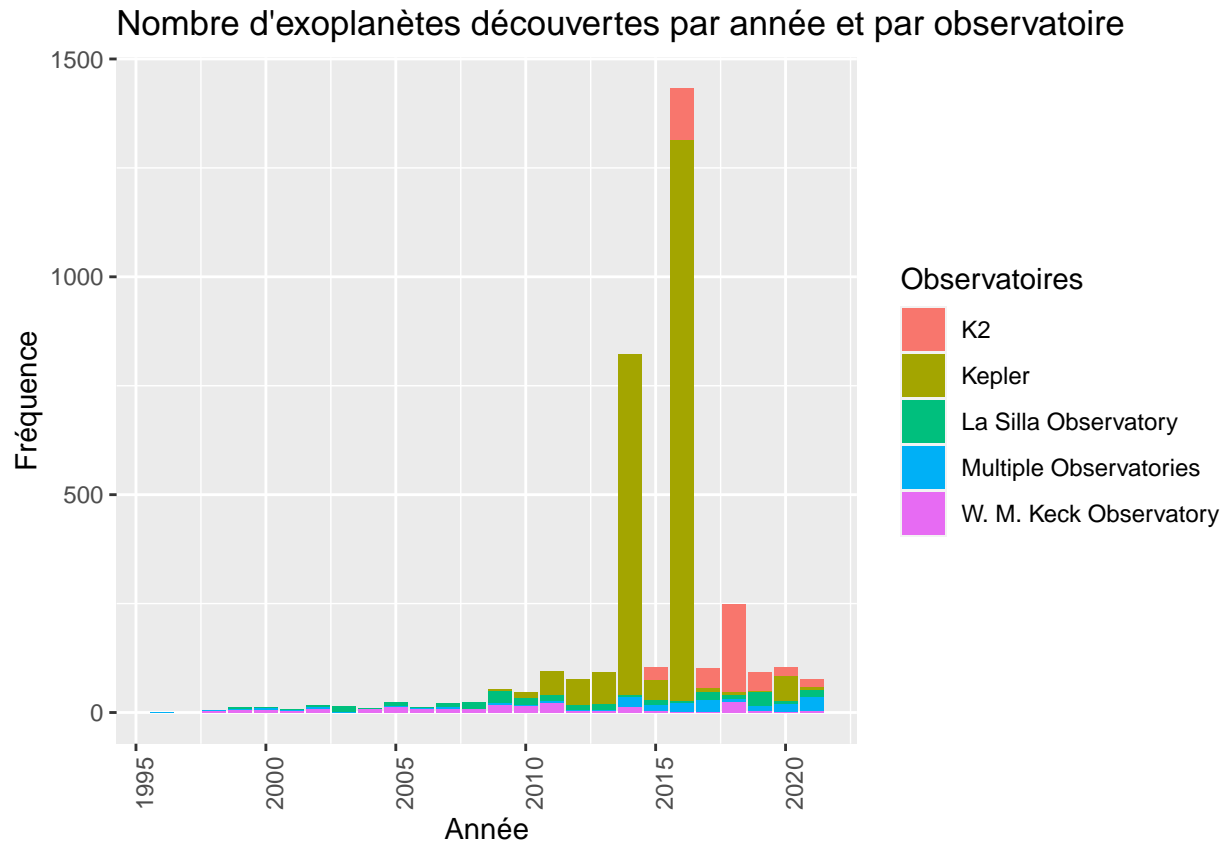


J'ai choisi d'utiliser un diagramme à barre empilée pour répondre à cette question. Ce type de graphique permettant une interprétation efficace et rapide d'une variable quantitative (le nombre d'exoplanètes découvertes) sur une période de temps (chaque année). J'ai choisi de colorer chaque bloc avec la méthode de découverte afin d'avoir rapidement un aperçu de l'efficacité des méthodes et de leurs popularités au fil du temps de l'efficacité des méthodes et de leurs popularités au fil du temps.

Ici on remarque deux années importantes, 2014 et 2016, visiblement ces deux années ont été très riches en découvertes d'exoplanètes. Ces pics sont liés aux recherches du satellite Kepler lancées en 2009 et qui a eu pour mission jusqu'en 2018 de chercher des exoplanètes et notamment des planètes plus ou moins habitable. Après des recherches les pics en 2014 et en 2016 s'expliquent par des conditions d'observations très favorables. au cours de ces 2 années Kepler a eu la chance de pouvoir faire des observations à long terme sur des zones ciblées de notre univers.

La représentation est plutôt intéressante, néanmoins on peut noter qu'il est difficile de lire précisément le nombre de planètes découvertes dans les petites barres à cause des 2 années aberrantes. On peut aussi noter que le graphique ne commence pas en 1989 mais plutôt en 1995, après analyse des données une exoplanète a bien été découverte en 1989, néanmoins ce nombre étant trop petit par rapport aux autres années on ne peut pas voir de barre sur le graphique

Afin de mettre en exergue l'efficacité de Kepler depuis sa mise en service j'ai choisi de faire ce second petit graphique afin que le nombre d'exoplanètes découvertes par année et par satellites (en prenant seulement les 5 premiers par ordre de découverte car sinon la légende avait trop de valeurs).



Avec cette représentation on se rend bien compte de l'efficacité de Kepler sur ses années de mise en service, on remarque aussi K2 visiblement efficace à partir de 2015.

Quels sont les observatoires qui découvrent le plus d'exoplanètes ?

Population: L'ensemble des exoplanètes de l'univers

Echantillon: Les 4575 exoplanètes présentes dans le dataset

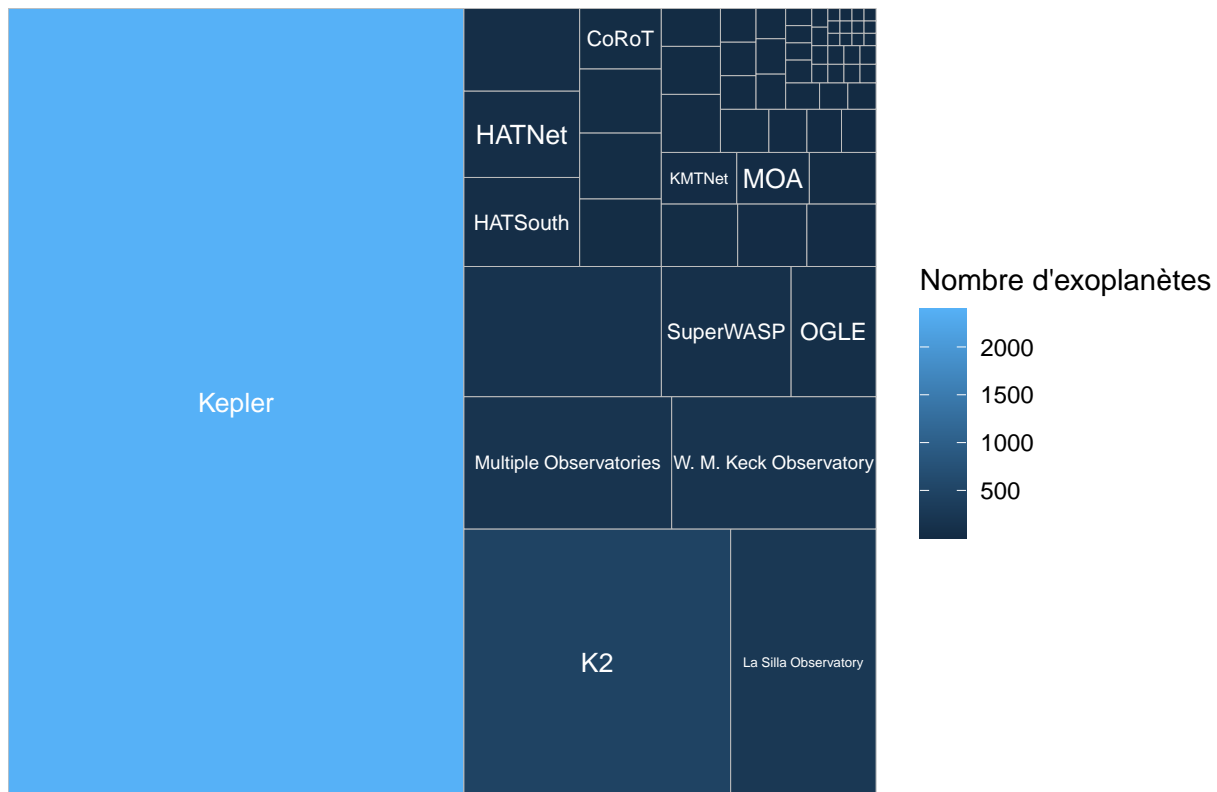
Variable:

- Le nombre d'exoplanètes découvertes
- L'observatoire / satellite qui les a découvertes

Modalités

- Le nombre de planète découvertes => variable quantitative nombre entre 0 et environ 1600
- Observatoire => variable qualitative, nom de l'observatoire

Nombre d'exoplanètes découvertes par observatoire



Pour cette question je voulais avoir une représentation visuelle efficace des observatoires les plus efficaces. J'aurais pu utiliser un diagramme en camembert mais je souhaitais apprendre à manier un diagramme en "tree map" afin de voir si le rendu était visuellement intéressant. Ici on utilise la taille des rectangles et les couleurs pour connaître le nombre d'exoplanètes découvertes.

On remarque tout de suite en lisant le graphique que c'est bien Kepler qui a découvert le plus d'exoplanètes, visiblement environ 2000.

Cette représentation est pratique pour rapidement savoir quel observatoire ont découvert le plus d'exoplanètes mais il est difficile d'avoir précisément une idée du nombre exact. Dans ce type de graphique on peut aussi s'apercevoir que la quantité est représentée à la fois par la taille des boîtes et leurs couleurs, cette information doublée n'est pas très pertinente. Il est aussi important de noter que ce type de graphique tronque les noms trop long ce qui n'est pas très optimal pour la lecture du graphique.

Y a-t-il un lien entre la masse stellaire et le rayon des exoplanètes?

Population: L'ensemble des exoplanètes de l'univers

Echantillon: Les 888 exoplanètes présentes dans le dataset avec les catégories mass radius et equilibrium temperature

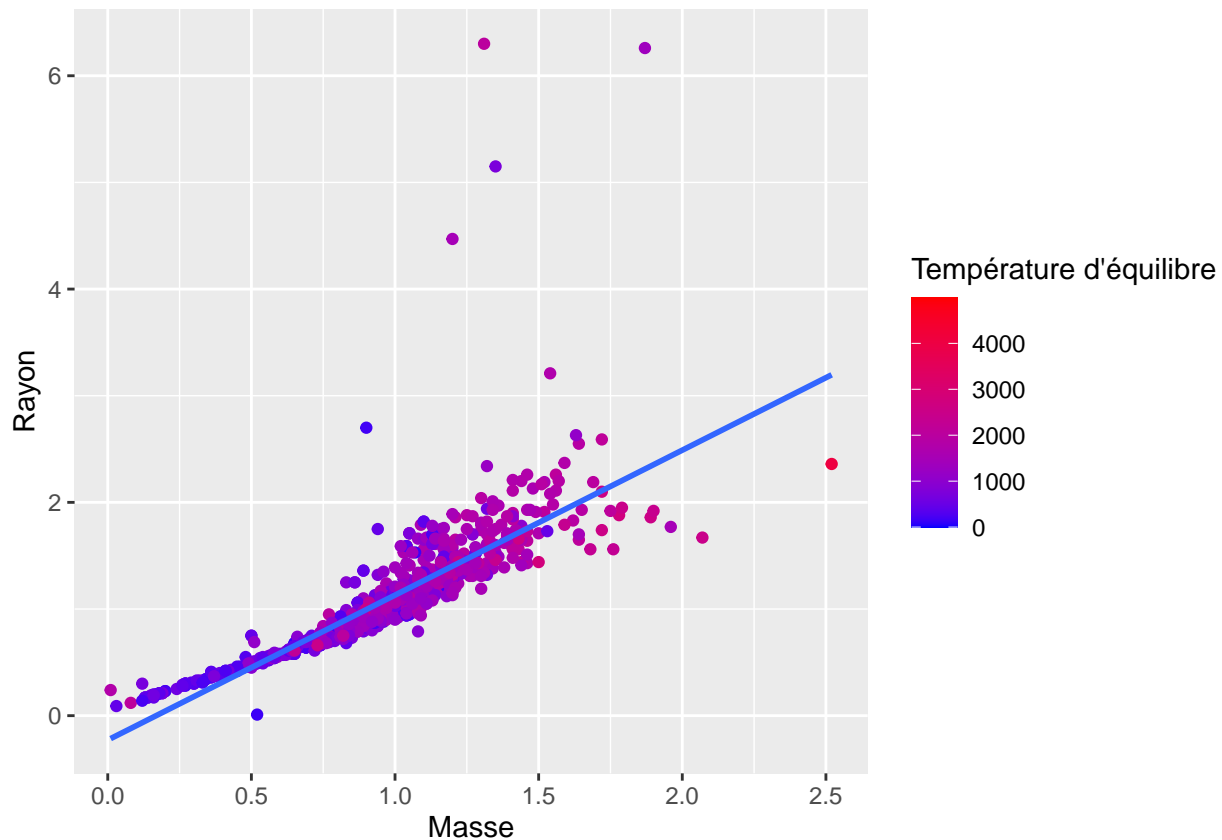
Variable:

- La masse stellaire des exoplanètes
- Le rayon des étoiles
- leur température d'équilibre

Modalités

- La masse stellaire des exoplanètes => quantitative nombre
- Le rayon des étoiles => quantitative nombre entre 0 et 7
- leurs température d'équilibre => quantitative nombre entre 0 et 5000

```
## 'geom_smooth()' using formula = 'y ~ x'
```



Ce graphique permet de voir le lien entre la masse des exoplanètes et leurs rayons. Le but était de pouvoir déterminer s'il y a un bien un lien entre ces deux variables. Pour essayer de déterminer un lien entre deux variables il me semblait pertinent de faire un graphique en nuage de points afin de mieux pouvoir observer la répartition de ceux-ci. Ce type de graphique permet aussi d'afficher une régression linéaire pour mettre encore plus en évidence un potentiel lien entre les données. Il me semblait aussi pertinent d'essayer de voir si ces deux variables pouvaient aussi avoir un lien avec la température d'équilibre des exoplanètes, c'est pour cela que j'ai décidé de mapper la température à la couleur des points.

Ici le graphique semble pertinent, en effet on peut bien constater un lien entre la masse et le rayon des exoplanètes. En règle générale il est visible que plus la masse des exoplanètes est importante plus leur rayon est important, la régression linéaire nous permet de confirmer cette affirmation, celle-ci étant croissante. On peut aussi noter qu'en règle générale plus ces deux variables augmentent plus la température d'équilibre augmente elle aussi.

Quelles sont les méthodes qui permettent de trouver les exoplanètes les plus éloignées

Population: L'ensemble des exoplanètes de l'univers

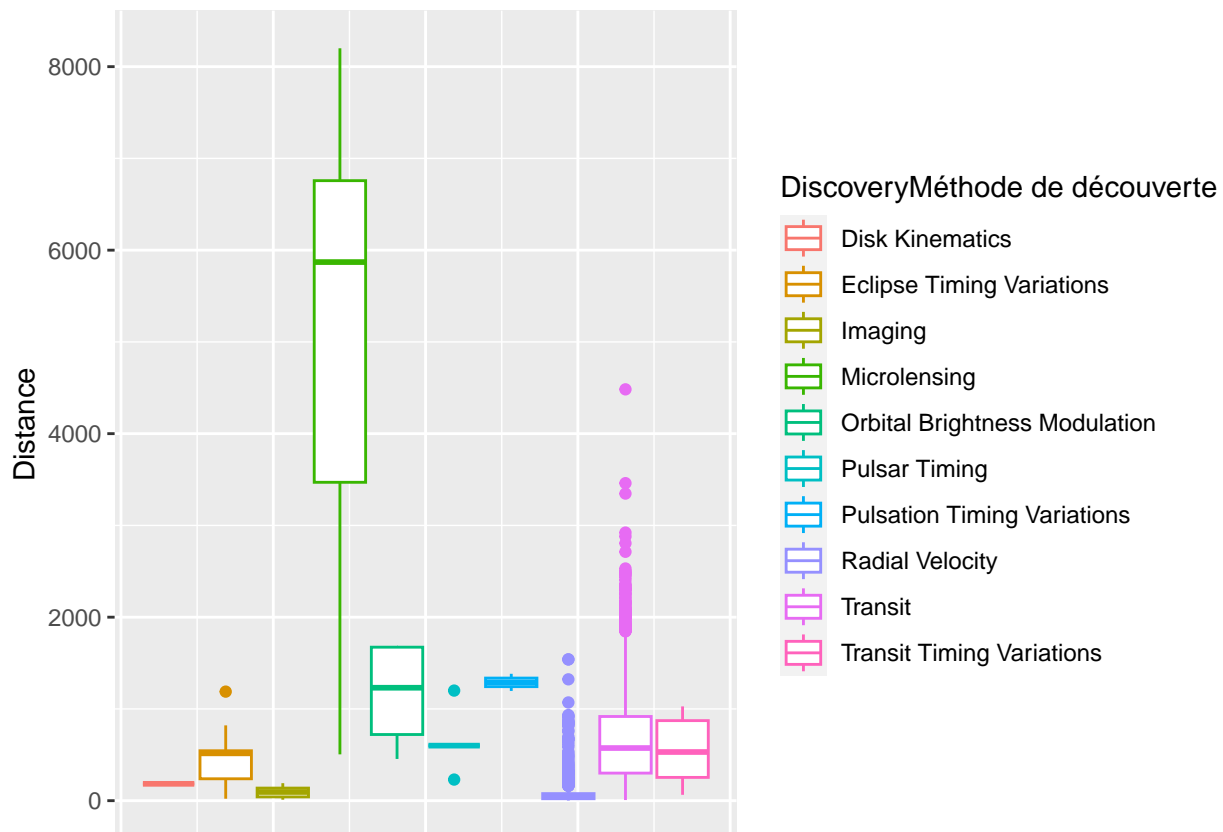
Echantillon: Les 4,471 exoplanètes presentes dans le dataset avec les catégories year, distance et method

Variable:

- L'année de découverte
- La distance depuis la terre
- La méthode de découverte

Modalités

- année de découverte => Nombre entre 1989 et 2021
- distance => variable quantitative nombre
- Méthode => variable qualitative, méthode de découverte



Ici le but était de voir si une méthode permettait de découvrir des exoplanètes plus éloignées que d'autres. Dans ce cas prendre un graphique en moustache permet de comparer les moyennes les quartiles et les valeurs extremes, cette représentation permet une lecture efficace pour ce type de question

Ce graphique nous permet de nous rendre que c'est la microlésion qui permet de trouver les exoplanètes les plus lointaines, on peut aussi noter que la méthode transit contient certaines valeurs aberrantes qui ont permis de découvrir des exoplanètes très lointaines mais beaucoup plus rarement que les méthodes de microlensing