

# Ingénierie mathématique et analyse statistique de données sportives

Sébastien Déjean  
Ingénieur de recherche statisticien

`math.univ-toulouse.fr/~sdejean`

8 avril 2022



# Ingénieur en mathématique ?

- 1993 - Baccalauréat série C, mathématiques et sciences physiques, mention AB, (merci l'option latin !)
- 1994-1997 - Institut Universitaire Professionnalisé, **Mathématiques Industrielles**, Calcul Scientifique et Statistique (DEUG, licence, maîtrise)
- 1998 - Diplôme d'Études Approfondies en mathématiques appliquées
- 2002 - Doctorat en mathématiques appliquées.
- 2019 - Habilitation à Diriger les Recherches. *Quinze ans de recherche appliquée en science des données.*

## BAP E « Informatique, Statistiques et Calcul scientifique » - Statistiques - IR

## E1D44 - Expert-e en information statistique

| Mission                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                                                   |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| Concevoir et piloter la réalisation de systèmes d'information statistique et animer le fonctionnement                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                   |
| Métiers si besoin                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                   |
| ▪ responsable d'une cellule statistique dans un rectorat, une université, un organisme de recherche                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | ▪ ingénieur responsable des aspects statistiques dans un laboratoire de recherche |
| Famille d'activité professionnelle                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Correspondance statutaire                                                         |
| Statistiques                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | Ingénieur de recherche                                                            |
| Famille d'activité professionnelle REME                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | Emploi-type de rattachement REME                                                  |
| Études et évaluation des politiques publiques                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Ingénieur, Chargé d'études et d'évaluation                                        |
| Activités principales                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                   |
| <ul style="list-style-type: none"> <li>Piloter des projets statistiques</li> <li>Concevoir des systèmes d'information à but décisionnel</li> <li>Coordonner une équipe autour de projets en statistique</li> <li>Définir un plan de recueil et de gestion des données ainsi que la chaîne de traitement associée</li> <li>Élaborer des enquêtes et des études</li> <li>Concevoir les méthodes pour traiter les données stockées</li> <li>Contrôler la qualité des résultats et de leur interprétation</li> <li>Apporter aux partenaires d'un projet des conseils de haut niveau sur les méthodes statistiques à employer et les outils pertinents disponibles</li> <li>Participer à des projets au plan national, voire international et aux publications associées</li> <li>Effectuer des présentations et des formations pour assurer un transfert de compétences</li> </ul> |                                                                                   |
| Conditions particulières d'exercice                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                   |
| <ul style="list-style-type: none"> <li>Obligation de respecter le secret statistique ou professionnel dans le cadre législatif existant</li> <li>Contraintes de calendrier en fonction de la nature du projet</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                   |

| Ancien code de l'emploi-type<br>REFERENS | Ancien intitulé de l'emploi-type<br>REFERENS        |
|------------------------------------------|-----------------------------------------------------|
| E1D24                                    | Chef de projet ou expert en information statistique |



[http://referens.esr.gouv.fr/pages/fiche\\_emploi\\_type\\_referens\\_iii\\_?refine.referens\\_id=E1D44](http://referens.esr.gouv.fr/pages/fiche_emploi_type_referens_iii_?refine.referens_id=E1D44)

| Compétences principales                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Connaissances                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <ul style="list-style-type: none"> <li>Statistiques (expertise)</li> <li>Techniques statistiques et informatiques de collecte et de traitement de données (expertise)</li> <li>Outils et méthodes de projection (expertise)</li> <li>Logiciels statistiques (maîtrise)</li> <li>Mathématiques (maîtrise)</li> <li>Langage de programmation (application)</li> <li>Méthodologie de conduite de projet</li> <li>Langue anglaise : B2 (cadre européen commun de référence pour les langues)</li> </ul> |
| Compétences opérationnelles                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| <ul style="list-style-type: none"> <li>Piloter un projet</li> <li>Encadrer / Animer une équipe</li> <li>Gérer la sécurité de l'information</li> <li>Assurer une veille</li> <li>Mettre au point ou adapter des techniques nouvelles</li> <li>Apporter des réponses à des besoins spécifiques</li> <li>Identifier et restituer de l'information</li> <li>Réaliser des synthèses</li> <li>Accompagner et conseiller</li> <li>Communiquer et faire preuve de pédagogie</li> </ul>                      |
| Compétences comportementales                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| <ul style="list-style-type: none"> <li>Capacité de prospective</li> <li>Capacité de conceptualisation</li> <li>Créativité / Sens de l'innovation</li> </ul>                                                                                                                                                                                                                                                                                                                                         |
| Diplôme réglementaire exigé - Formation professionnelle si souhaitable                                                                                                                                                                                                                                                                                                                                                                                                                              |
| <ul style="list-style-type: none"> <li>Doctorat, diplôme d'ingénieur</li> <li>Domaine de formation souhaité : statistiques et probabilités, économétrie, mathématiques appliquées</li> </ul>                                                                                                                                                                                                                                                                                                        |
| Tendances d'évolution                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Facteurs d'évolution à moyen terme                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| <ul style="list-style-type: none"> <li>Renforcement de l'utilisation des outils et techniques informatiques de gestion et de traitement de très grandes masses de données</li> <li>Développement d'outils d'aide à la décision analysant des données complexes en un temps réduit</li> </ul>                                                                                                                                                                                                        |
| Impacts sur l'emploi-type (qualitatif)                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| <ul style="list-style-type: none"> <li>Adaptation aux techniques et outils statistiques de traitement de grandes masses de données</li> </ul>                                                                                                                                                                                                                                                                                                                                                       |



# D'autres métiers

[data.enseignementsup-recherche.gouv.fr/pages/la\\_bap/?refine.referens\\_bap\\_id=E#EE](https://data.enseignementsup-recherche.gouv.fr/pages/la_bap/?refine.referens_bap_id=E#EE)

**BAP A : Sciences du vivant, de la terre et de l'environnement**  
> 5 familles professionnelles > 36 emplois-types

**BAP B : Sciences chimiques et Sciences des matériaux**  
> 4 familles professionnelles > 19 emplois-types

**BAP C : Sciences de l'Ingénieur et instrumentation scientifique**  
> 4 familles professionnelles > 37 emplois-types

**BAP D : Sciences Humaines et Sociales**  
> 4 familles professionnelles > 17 emplois-types

**BAP E : Informatique, Statistiques et Calcul scientifique**  
> 5 familles professionnelles > 18 emplois-types

**BAP F : Culture, Communication, Production et diffusion des savoirs**  
> 4 familles professionnelles > 44 emplois-types

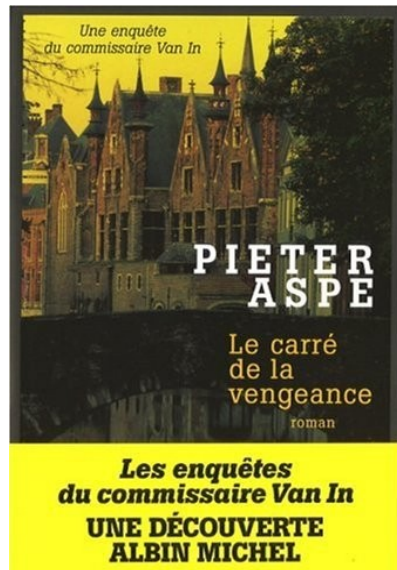
**BAP G : Patrimoine immobilier, Logistique, Restauration et Prévention**  
> 3 familles professionnelles > 38 emplois-types

**BAP J : Gestion et Pilotage**  
> 6 familles professionnelles > 33 emplois-types



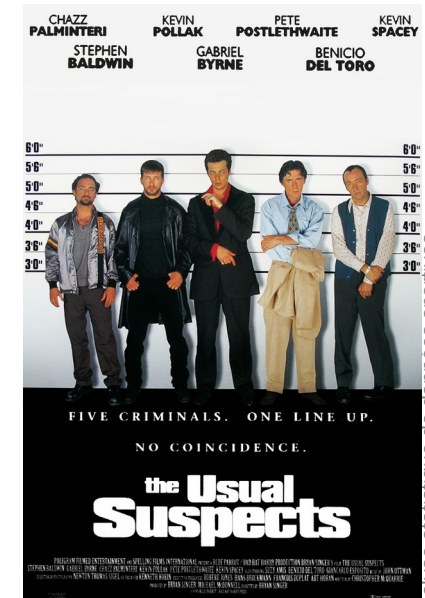
| Statistiques                                        | Calcul scientifique                                |
|-----------------------------------------------------|----------------------------------------------------|
| Expert-e en information statistique<br><b>E1D44</b> | Expert-e en calcul scientifique<br><b>E1E45</b>    |
| Ingénieur-e statisticien-ne<br><b>E2D46</b>         | Ingénieur-e en calcul scientifique<br><b>E2E47</b> |
| Assistant-e statisticien-ne<br><b>E3D44</b>         |                                                    |

# Un métier à risque...



*Selon le profil que j'ai établi en me basant sur les deux fax, le plus vieux des ravisseurs est très certainement **universitaire**. Il doit être **ingénieur** ou **mathématicien**. Je ne serais pas étonné si on me disait que c'est un **statisticien** ou un spécialiste du **calcul des probabilités**.*

Le carré de la vengeance, Pieter Aspe  
(traduction Emmanuèle Sandron)



# ... mais sympa quand même

[hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century](https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century)

Analytics And Data Science

## Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and DJ Patil

From the Magazine (October 2012)



Andrew J Buboltz, silk screen on a page from a high school yearbook, 8.5" x 12", 2011 Tamar Cohen



Cassie Kozyrkov

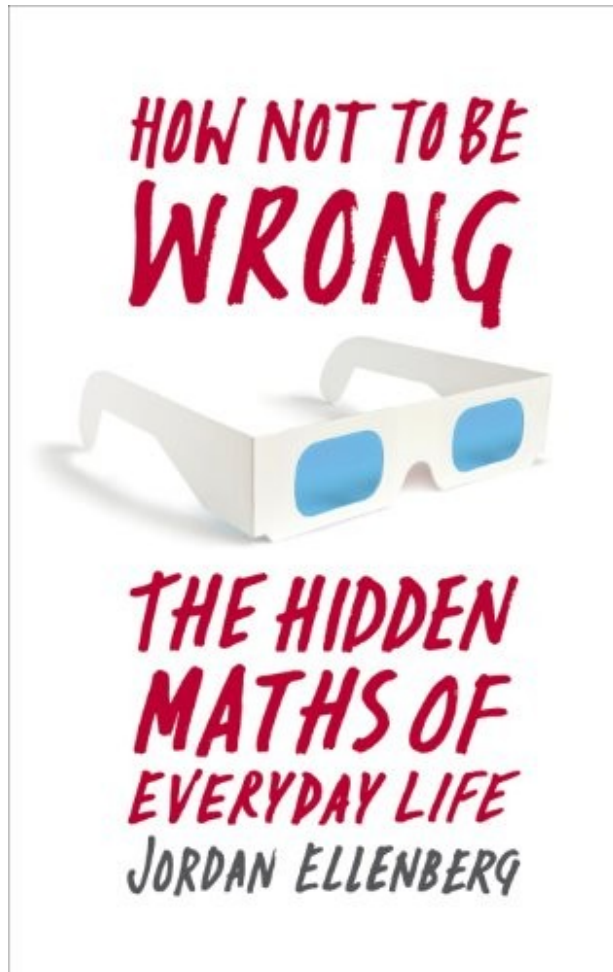
Mar 15, 2019 · 6 min read · [Listen](#)



## Data scientist: The sexiest job of the 22nd century

[towardsdatascience.com/the-sexiest-job-of-the-22nd-century-ffe753e1d155](https://towardsdatascience.com/the-sexiest-job-of-the-22nd-century-ffe753e1d155)

# Répondre à de vraies questions



Je n'aime pas les problèmes "issus de la vie courante". Ils donnent une image fausse du rapport entre mathématiques et réalité. "Bobby a 300 billes ; il en donne 30% à Jenny" [...]

Mais les questions du monde réel ne ressemblent pas à ce type d'énoncés. **Un problème dans le monde réel**, c'est quelque chose comme : "La récession et ses suites ont-elles été particulièrement dures pour l'emploi des femmes, et dans ce cas, dans quelle mesure est-ce le résultat de l'administration Obama?" Votre calculette n'a pas de bouton pour ça.

Car pour donner une réponse vous devez connaître autre chose que de simples chiffres. [...] Ce n'est qu'après avoir formulé ces questions que vous pouvez prendre votre calculette. Mais à ce point, le véritable travail intellectuel est déjà terminé. **Diviser un nombre par un autre, ce n'est que du calcul ; savoir ce que vous devez diviser par quoi, ça, c'est des mathématiques.**

# Répondre à de vraies questions

*[...] the rocky road of real problems in preference to smooth road of unreal assumptions, arbitrary criteria, and abstract results without real attachments.*

John Tukey, 1962, The Future of Data Analysis



# Domaines d'application

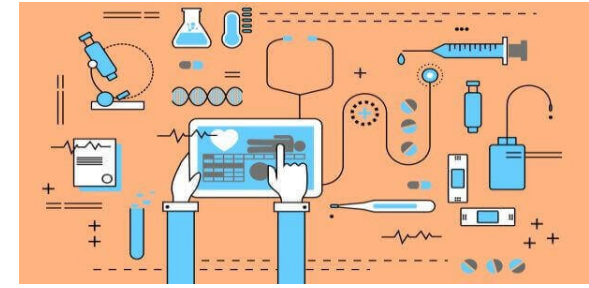


**Imaginez** une flotte d'avions qui transmettraient l'ensemble de leurs données de vol **sur une seule et même plate-forme numérique**, de la moindre vanne défectueuse à un décollage poussif, en passant par des vibrations anormales de la voilure. **Collectées, agrégées, retraitées, ces données formeraient un trésor. Il deviendrait possible** pour l'utilisation de chaque avion **de s'enrichir de l'exploitation des autres. À la clé, des millions d'euros d'économie.**

Changer cette pièce juste avant qu'elle ne tombe en panne, définir la meilleure trajectoire de vol pour réduire la consommation de carburant... **Une révolution pour** l'aéronautique, **un changement de modèle économique pour** les industriels, **un scénario de rêve** pour les compagnies aériennes.

[usinenouvelle.com/article/avec-skywise-airbus-lance-l-internet-des-avions.N618963](https://usinenouvelle.com/article/avec-skywise-airbus-lance-l-internet-des-avions.N618963)

*12 Examples of Big Data Analytics In Healthcare That Can Save People*  
Mona Lebed, Business Intelligence  
Jul 18th 2018



**Imaginez** un ensemble de patients qui transmettraient l'ensemble de leurs données de santé **sur une seule et même plate-forme numérique**, de la moindre prise de tension artérielle jusqu'aux examens génétiques ou d'imagerie lors d'un diagnostic de cancer, en passant par les résultats du bilan lors d'une lombalgie commune. **Collectées, agrégées, retraitées, ces données formeraient un trésor. Il deviendrait possible** pour le soin de chaque patient **de s'enrichir de l'exploitation des autres. À la clé, des vies sauvées et des millions d'euros d'économie.** Mettre en place des mesures de prévention et de dépistages personnalisés, définir le meilleur parcours de soin pour réduire l'impact des maladies et le handicap qui en découle... **Une révolution pour** la santé, **un changement de modèle économique** pour les hôpitaux, **un scénario de rêve** pour les patients et les cliniciens.

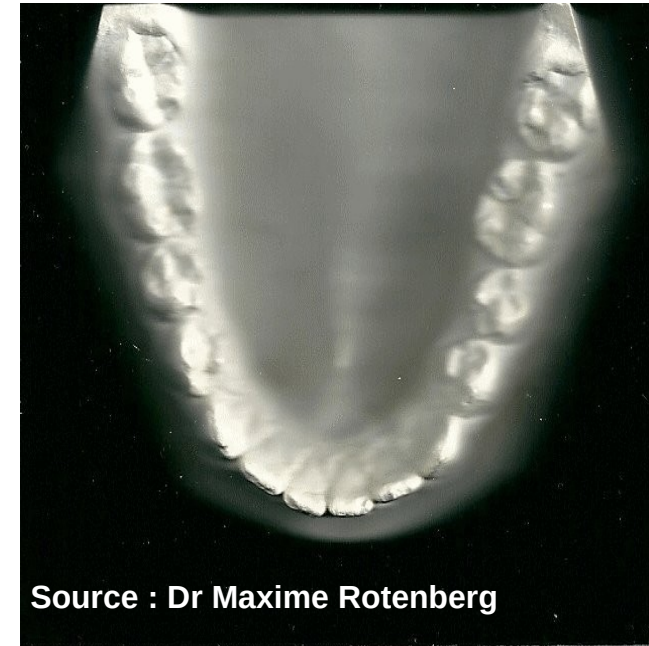
Adaptation réalisée avec l'aide de Pr X. de Boissezon (CHU Toulouse)

# Orthodontie

Pour un sourire éclatant...



... il faut en passer par ici...

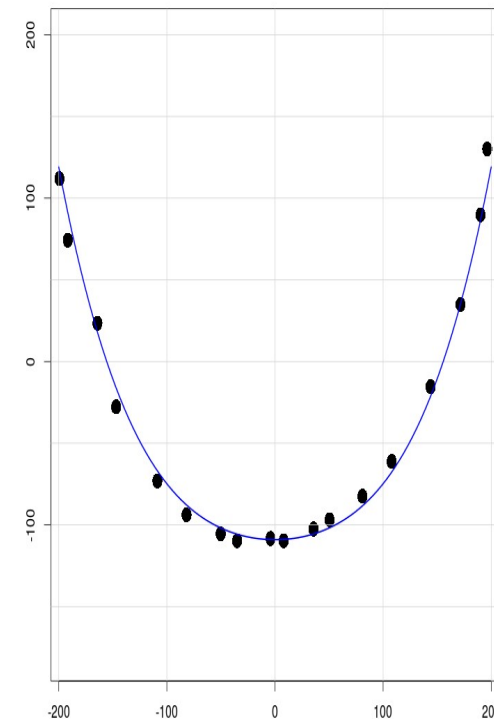


... et peut-être aussi par là !

Modélisation de l'arcade dentaire  
par un polynôme de degré 4 sans  
termes de degrés impairs  
(symétrie axiale)

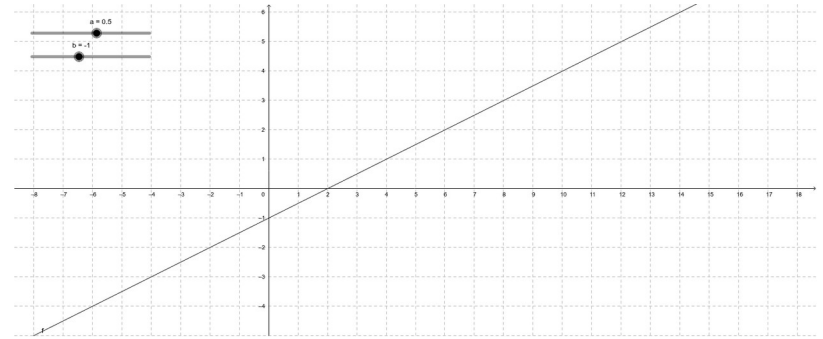
$$Y = b_0 + b_2 X^2 + b_4 X^4$$

- comparer les courbes avant et après traitement
- comparer une courbe à des références (charte)

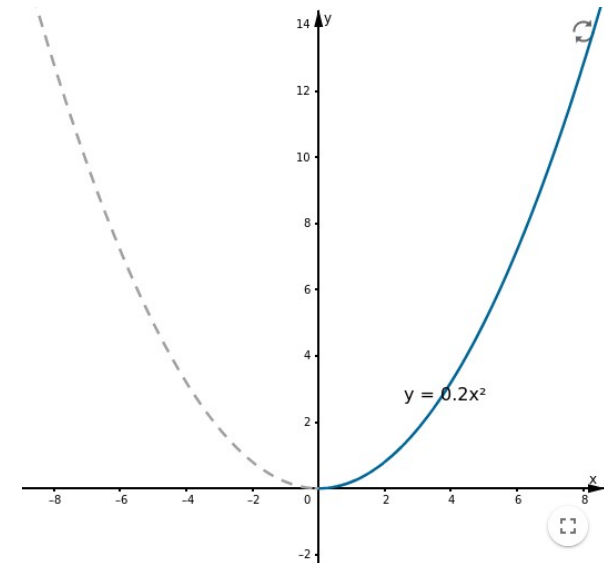


# Travaux théoriques

$$Y = aX + b$$



$$Y = aX^2 + bX + c$$



$$\sum_{l=1}^n \left( \frac{\cos(2\pi\theta_l) \left[ \cos\left(\frac{R_l}{m}t \cdot U_l\right) - 1\right] + \sin(2\pi\theta_l) \left[ \sin\left(\frac{R_l}{m}t \cdot U_l\right) \right]}{R_l^{h(t)+1/2}} \right)$$



# Un programme informatique

Extrait de programme en langage C

```

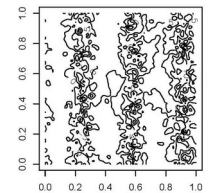
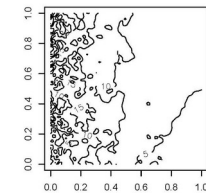
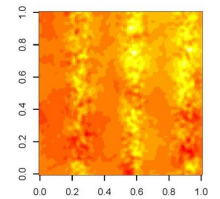
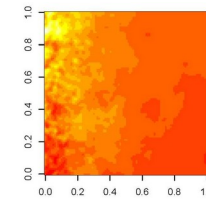
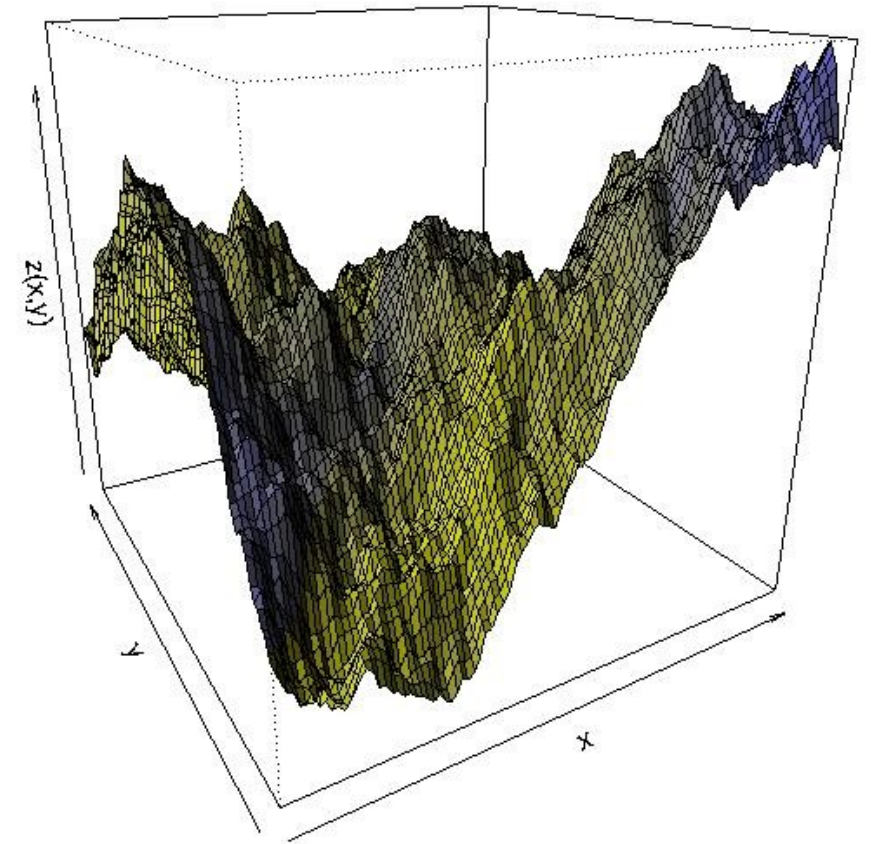
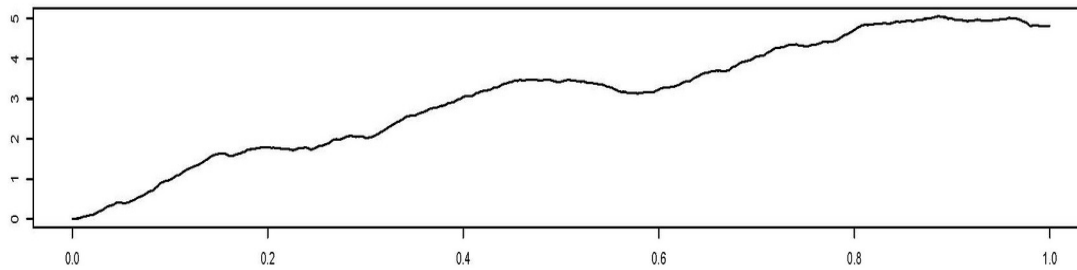
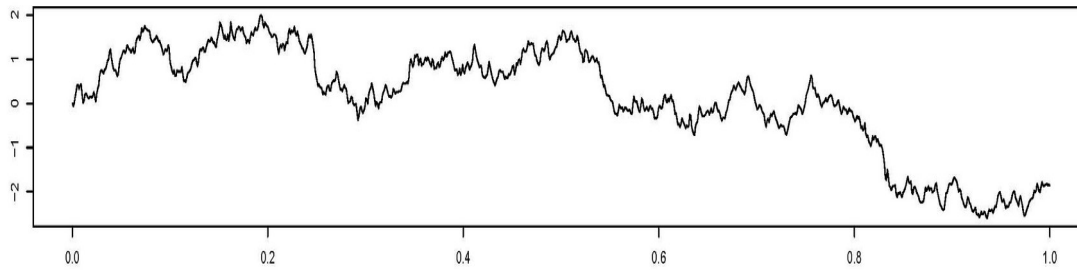
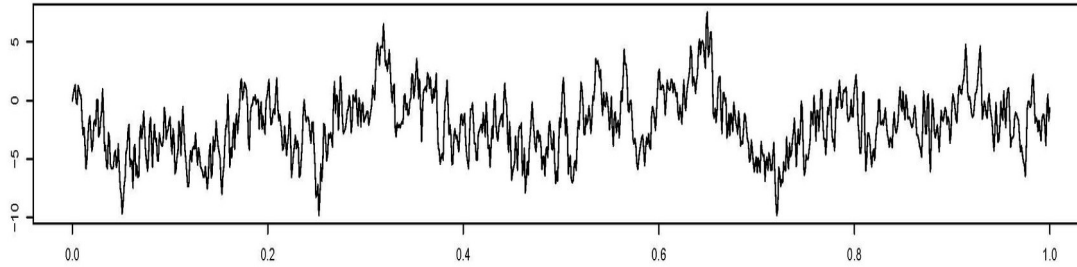
...
for (t=0;t<T;t++) nombre de valeurs de t, T termes
for (n=0;n<N;n++) calcul de la somme, N termes
{
tmp1 = cos(2*M PI*theta)*(cos(t*R*U) - 1);
tmp2 = sin(2*M PI*theta)*sin(t*R*U);
tmp3 = pow(R,0.5+H[t]);
tmp = (tmp1+tmp2)/tmp3;
res[t] = res[t] + tmp;
}
...

```

$$\sum_{l=1}^n \left( \frac{\cos(2\pi\theta_l) \left[ \cos\left(\frac{R_l}{m}t.U_l\right) - 1 \right] + \sin(2\pi\theta_l) \left[ \sin\left(\frac{R_l}{m}t.U_l\right) \right]}{R_l^{h(t)+1/2}} \right)$$



# Des dessins !



# Revue de presse

latribune.fr | 14/03/2013

L'an dernier, **le salaire moyen brut a progressé** de 2,1%, soit près d'un point de plus que l'inflation (+1,2%). Mais cette hausse ralentit.

**Petite éclaircie pour le pouvoir d'achat des Français.** L'Agence centrale des organismes de Sécurité sociale (Acos) a annoncé les chiffres sur l'évolution du salaire moyen brut en France. En 2012, il a augmenté de 2,1% pour s'établir à 2 410 euros brut [...]

# Un exemple

Supposons qu'une entreprise compte 12 personnes :

- 8 ouvriers : entre 970€ et 1 100 € chacun
- Un chef d'atelier : 2 000 €
- Un directeur technique : 4 000 €
- Un directeur des ressources humaines : 6 000 €
- Un directeur général : 10 000 €

970  
975  
980  
985  
990  
995  
1005  
1100  
2000  
4000  
6000  
10000

Moyenne : 2500

Médiane : 1000

**Salaire moyen : 2500 € - Salaire médian : 1000 €**

# Comment augmenter le salaire moyen dans cette société ?

Solution 1 : augmenter tous les salaires de x %

Solution 2 : le dirigeant augmente son salaire

Solution 3 : « remercier » quelques postes de bas salaire

|            |         |            |       |            |         |
|------------|---------|------------|-------|------------|---------|
|            | 1018.50 |            | 970   |            |         |
|            | 1023.75 |            | 975   |            |         |
|            | 1029    |            | 980   |            | 980     |
|            | 1034.25 |            | 985   |            | 985     |
|            | 1039.50 |            | 990   |            | 990     |
|            | 1044.75 |            | 995   |            | 995     |
|            | 1055.25 |            | 1005  |            | 1005    |
|            | 1155    |            | 1100  |            | 1100    |
|            | 2100    |            | 2000  |            | 2000    |
| Solution 1 | 4200    | Solution 2 | 4000  | Solution 3 | 4000    |
|            | 6300    |            | 6000  |            | 6000    |
|            | 10500   |            | 12000 |            | 10000   |
| Moyenne :  | 2625    | Moyenne :  | 2666  | Moyenne :  | 2805.50 |
| Médiane :  | 1050    | Médiane :  | 1000  | Médiane :  | 1052.50 |

# La suite de l'article

latribune.fr | 14/03/2013

L'an dernier, **le salaire moyen brut a progressé** de 2,1%, soit près d'un point de plus que l'inflation (+1,2%). Mais cette hausse ralentit.

**Petite éclaircie pour le pouvoir d'achat des Français.** L'Agence centrale des organismes de Sécurité sociale (Acos) a annoncé les chiffres sur l'évolution du salaire moyen brut en France. En 2012, il a augmenté de 2,1% pour s'établir à 2.410 euros brut, soit 0,9 point de plus que l'inflation qui s'est tenu à 1,2%.

## **Les effectifs salariés en légère baisse**

La progression du salaire moyen s'explique en grande partie par la hausse de la masse salariale (+1,7%) alors que **le nombre de salariés du secteur privé s'est légèrement contracté** [...] L'Acos note que la hausse de la masse salariale s'inscrit en net ralentissement puisqu'elle avait été de 3,5% l'année précédente. **La baisse des effectifs** est davantage marquée dans le secteur immobilier [...]



# Et d'autres surprises...

- **200** élèves : **100** filles, **100** garçons
- **2** matières au choix : physique et biologie

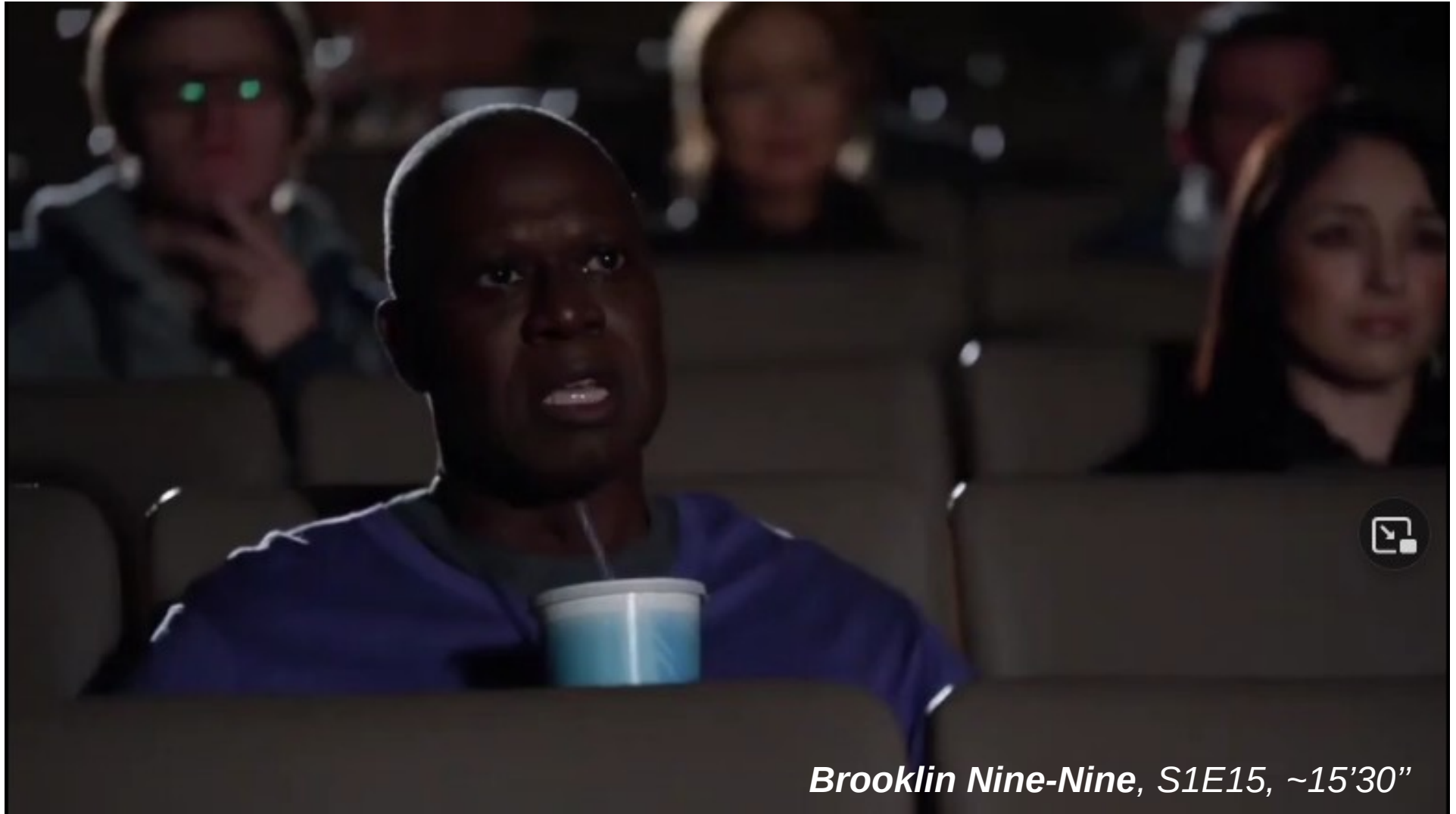
Exemple fictif tiré de Pour la Science, Février 2018 ; J.-P. Delahaye citant *Statistiques Méfiez-vous !* de N. Gauvrit

|          | Physique |    | Biologie |    |
|----------|----------|----|----------|----|
|          | G        | F  | G        | F  |
| Réussite | 80       | 10 | 4        | 50 |
| Echec    | 10       | 0  | 6        | 40 |
| Total    | 90       | 10 | 10       | 90 |

Qui réussit le mieux ? Garçons ou filles ?

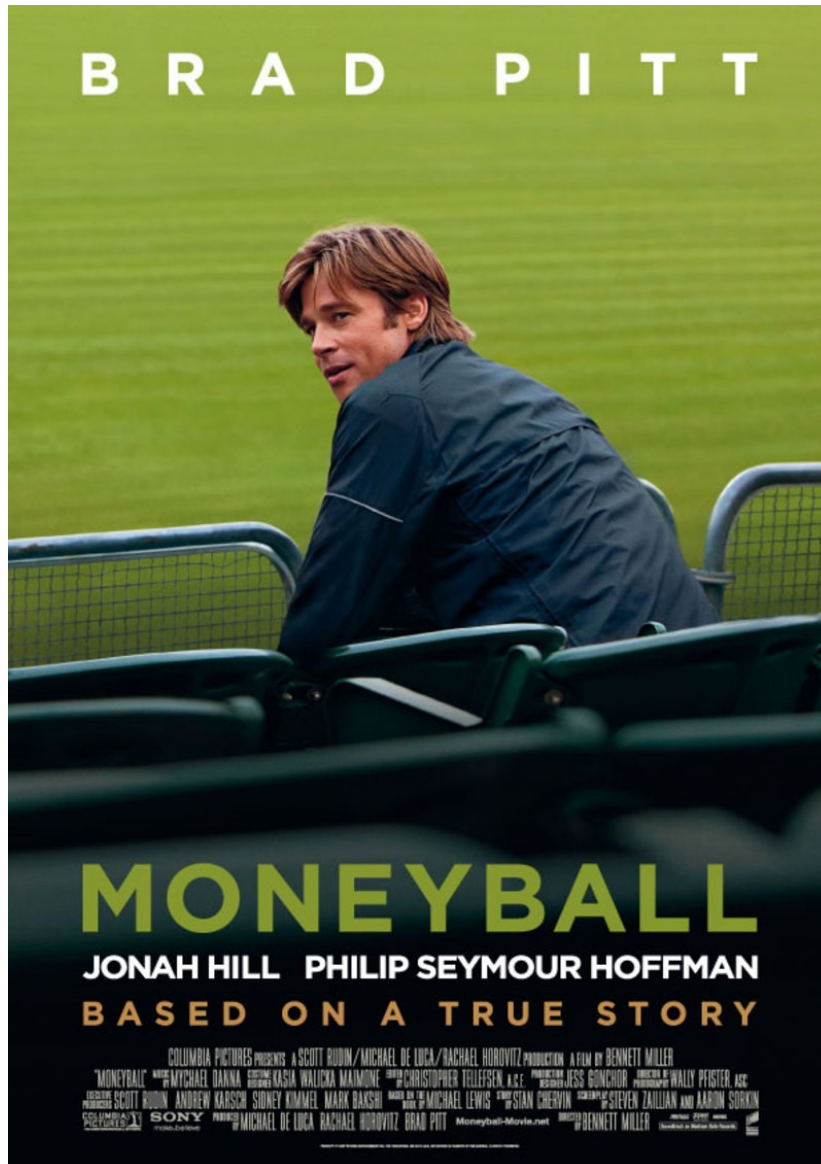
# Analyse statistique de données sportives

*This statistical analysis is so beautiful*



*Brooklyn Nine-Nine, S1E15, ~15'30"*

# Moneyball



## Synopsis

[www.allocine.fr/film/fichefilm\\_gen\\_cfilm=140005.html](http://www.allocine.fr/film/fichefilm_gen_cfilm=140005.html)

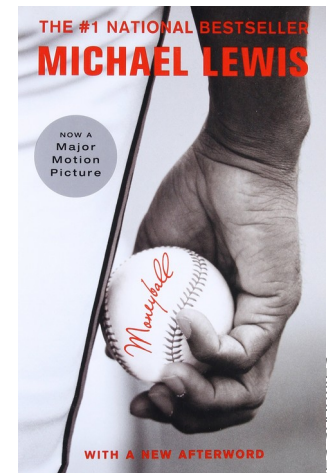
Voici l'histoire vraie de Billy Beane, un ancien joueur de baseball prometteur qui, à défaut d'avoir réussi sur le terrain, décida de tenter sa chance en dirigeant une équipe comme personne ne l'avait fait auparavant...

**Il va s'appuyer sur des théories statistiques [...] contre tous les principes, ils reconsidèrent la valeur de chaque joueur sur la base des statistiques [...]**

La **sabermétrie** est une **approche statistique du baseball**. Le mot tire son origine de l'acronyme SABR (pour Society for American Baseball Research)[...].

Cette approche a été popularisée dans la dernière décennie par le **livre Moneyball** de Michael Lewis [...] et par le film du même nom [...]

[fr.wikipedia.org/wiki/Saberm%C3%A9trie](http://fr.wikipedia.org/wiki/Saberm%C3%A9trie)



[www.honus.fr/moneyball-interview-dalain-usereau-membre-de-la-sabr](http://www.honus.fr/moneyball-interview-dalain-usereau-membre-de-la-sabr)

# Analyse statistique de données sportives



[www.lebigdata.fr/sport-et-big-data](http://www.lebigdata.fr/sport-et-big-data)

## Sport et Big Data – Quand la science des données donne l'avantage sur le terrain

INSOLITE. Quand les datas des joueurs du RAF se retrouvent dans des exercices de maths

LADEPECHE.fr



Les performances de Florian David et ses coéquipiers servent de bases pour des exercices de maths / DDM - HAUTE-GARONNE

<https://www.ladepeche.fr/2022/03/27/insolite-quand-les-datas-des-joueurs-du-raf-se-retrouvent-dans-des-exercices-de-maths-10198130.php>

ACTUALITE

## Coupe du monde: Comment le Big Data coach l'équipe d'Allemagne

Par Rachid Karavan  
Publié le 26/06/2014 à 15:00



La Mannschaft s'entraîne depuis quelques mois avec l'appui d'un logiciel analysant les données biométriques et les déplacements des joueurs, ainsi que l'historique des matches joués par

DESER COPIAIE | Réservez tous vos articles en illimité pour 1€ seulement - Le m'abonnement

## TFC : Avec Damien Comolli, la révolution du Big Data va déferler sur Toulouse

**FOOTBALL** Très probable futur président du TFC, Damien Comolli est un pionnier de l'usage des statistiques dans le football, notamment dans le recrutement



Nicolas Stival | Publié le 14/06/20 à 10h15 — Mis à jour le 14/06/20 à 10h15

### Le déclic « Moneyball »

L'autre déclic chez Comolli est d'origine familiale : « Son frère installé aux Etats-Unis lui a envoyé le livre *Moneyball* de Michael Lewis qui a été une révélation pour lui : on pouvait donc rationaliser le sport plutôt que se fier à des avis purement subjectifs », décrit l'ancien journaliste. Cet ouvrage de 2003, adapté au cinéma sous le titre *Le Stratège* en VF retrace

[www.20minutes.fr/sport/2798987-20200614-tfc-damien-comolli-revolution-big-data-va-deferler-toulouse](http://www.20minutes.fr/sport/2798987-20200614-tfc-damien-comolli-revolution-big-data-va-deferler-toulouse)



# Exemple avec les coupes du monde de foot

- Données disponibles sur le site **STATSBOMB**.com
- Données des finales homme 2018 (France – Croatie) et femme 2019 (USA – Pays-Bas)
- **1 685** passes réalisées pendant ces 2 matchs
- Informations recueillies pour chaque passe : temps écoulé, nom du joueur ou de la joueuse, son poste, **longueur de la passe**, angle donnant la direction de la passe, joueur ou joueuse ayant reçu la passe, hauteur de la passe (sol, basse, haute), partie du corps (pied droit, pied gauche, tête, main...)

# Aperçu des données

| competition    | competiti | period | minute | second | timestamp    | player.name          | position.name             | pass.length | pass.angle | pass.recipient.name      | pass.height.name | pass.body_part.name |
|----------------|-----------|--------|--------|--------|--------------|----------------------|---------------------------|-------------|------------|--------------------------|------------------|---------------------|
| FIFA World Cup | 1         | 0      | 0      | 0      | 00:00:00.400 | Mario Mandžukić      | Center Forward            | 11.0        |            | 3.1 Marcelo Brozović     | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 0      | 2      | 0      | 00:00:02.347 | Marcelo Brozović     | Center Defensive Midfield | 11.2        |            | 2.0 Luka Modrić          | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 0      | 3      | 0      | 00:00:03.947 | Luka Modrić          | Right Center Midfield     | 23.0        |            | 1.6 Šime Vrsaljko        | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 0      | 6      | 0      | 00:00:06.240 | Šime Vrsaljko        | Right Back                | 40.3        |            | -2.3 Danijel Subašić     | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 0      | 9      | 0      | 00:00:09.400 | Danijel Subašić      | Goalkeeper                | 64.1        |            | -0.1 Mario Mandžukić     | High Pass        | Right Foot          |
| FIFA World Cup | 1         | 0      | 13     | 0      | 00:00:13.627 | Raphaël Varane       | Right Center Back         | 10          |            | 1.6 NA                   | High Pass        | NA                  |
| FIFA World Cup | 1         | 0      | 16     | 0      | 00:00:16.680 | Ivan Perišić         | Left Wing                 | 12.6        |            | 2.8 Ivan Rakitić         | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 0      | 19     | 0      | 00:00:19.360 | Ivan Rakitić         | Left Center Midfield      | 17.9        |            | -1.1 Ivan Strinić        | Ground Pass      | Left Foot           |
| FIFA World Cup | 1         | 0      | 20     | 0      | 00:00:20.120 | Ivan Strinić         | Left Back                 | 25.3        |            | 1.7 Marcelo Brozović     | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 0      | 23     | 0      | 00:00:23.547 | Marcelo Brozović     | Center Defensive Midfield | 3.6         |            | -2.6 Ivan Perišić        | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 0      | 25     | 0      | 00:00:25.107 | Ivan Perišić         | Left Wing                 | 12.4        |            | 2.9 Domagoj Vida         | Ground Pass      | Left Foot           |
| FIFA World Cup | 1         | 0      | 27     | 0      | 00:00:27.267 | Domagoj Vida         | Left Center Back          | 7.6         |            | 0.4 Luka Modrić          | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 0      | 27     | 0      | 00:00:27.920 | Luka Modrić          | Right Center Midfield     | 10.4        |            | -2.9 Domagoj Vida        | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 0      | 29     | 0      | 00:00:29.427 | Domagoj Vida         | Left Center Back          | 11.2        |            | -1.1 Ivan Rakitić        | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 0      | 31     | 0      | 00:00:31.747 | Ivan Rakitić         | Left Center Midfield      | 12.8        |            | 2.5 Domagoj Vida         | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 0      | 34     | 0      | 00:00:34.560 | Domagoj Vida         | Left Center Back          | 19.1        |            | 1.7 Dejan Lovren         | Ground Pass      | Left Foot           |
| FIFA World Cup | 1         | 0      | 38     | 0      | 00:00:38.627 | Dejan Lovren         | Right Center Back         | 37.6        |            | -1.1 Ivan Rakitić        | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 0      | 41     | 0      | 00:00:41.387 | Ivan Rakitić         | Left Center Midfield      | 16.1        |            | 0.5 Ivan Perišić         | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 0      | 54     | 0      | 00:00:54.747 | Benjamin Pavard      | Right Back                | 19.2        |            | -1.7 N'Golo Kanté        | Low Pass         | NA                  |
| FIFA World Cup | 1         | 0      | 56     | 0      | 00:00:56.507 | N'Golo Kanté         | Left Defensive Midfield   | 5.1         |            | 1.4 Paul Pogba           | High Pass        | Right Foot          |
| FIFA World Cup | 1         | 0      | 59     | 0      | 00:00:59.187 | Marcelo Brozović     | Center Defensive Midfield | 32.6        |            | 2.7 Ivan Perišić         | High Pass        | Head                |
| FIFA World Cup | 1         | 1      | 0      | 0      | 01:00:01.107 | Ivan Perišić         | Left Wing                 | 6.4         |            | 2.5 Ivan Rakitić         | Low Pass         | Right Foot          |
| FIFA World Cup | 1         | 1      | 1      | 0      | 01:01:01.947 | Antoine Griezmann    | Right Center Forward      | 6.3         |            | 1.2 Kylian Mbappé Lottin | Ground Pass      | Left Foot           |
| FIFA World Cup | 1         | 1      | 10     | 0      | 01:10.467    | Benjamin Pavard      | Right Back                | 33.5        |            | -2.8 Raphaël Varane      | Low Pass         | NA                  |
| FIFA World Cup | 1         | 1      | 14     | 0      | 01:14.667    | Raphaël Varane       | Right Center Back         | 31.6        |            | -2.2 Hugo Lloris         | Ground Pass      | Left Foot           |
| FIFA World Cup | 1         | 1      | 17     | 0      | 01:17.227    | Hugo Lloris          | Goalkeeper                | 17.8        |            | -0.9 Samuel Yves Umtiti  | Ground Pass      | Left Foot           |
| FIFA World Cup | 1         | 1      | 36     | 0      | 01:36.867    | Samuel Yves Umtiti   | Left Center Back          | 15.1        |            | -0.1 Blaise Matuidi      | Ground Pass      | Left Foot           |
| FIFA World Cup | 1         | 1      | 39     | 0      | 01:39.027    | Blaise Matuidi       | Left Midfield             | 16.5        |            | -2.9 Samuel Yves Umtiti  | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 1      | 41     | 0      | 01:41.787    | Samuel Yves Umtiti   | Left Center Back          | 27.0        |            | 0.0 Olivier Giroud       | Ground Pass      | Left Foot           |
| FIFA World Cup | 1         | 1      | 42     | 0      | 01:42.880    | Olivier Giroud       | Left Center Forward       | 4           |            | 3.1 Blaise Matuidi       | Low Pass         | Left Foot           |
| FIFA World Cup | 1         | 1      | 43     | 0      | 01:43.440    | Blaise Matuidi       | Left Midfield             | 6.4         |            | 2.2 Olivier Giroud       | Low Pass         | Head                |
| FIFA World Cup | 1         | 1      | 45     | 0      | 01:45.440    | Olivier Giroud       | Left Center Forward       | 23.3        |            | 2.3 Raphaël Varane       | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 1      | 47     | 0      | 01:47.107    | Raphaël Varane       | Right Center Back         | 20.6        |            | 1.1 Paul Pogba           | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 1      | 51     | 0      | 01:51.427    | Paul Pogba           | Right Defensive Midfield  | 42.2        |            | -0.2 Antoine Griezmann   | High Pass        | Right Foot          |
| FIFA World Cup | 1         | 1      | 53     | 0      | 01:53.867    | Domagoj Vida         | Left Center Back          | 13.6        |            | -0.9 Ivan Strinić        | Low Pass         | Left Foot           |
| FIFA World Cup | 1         | 1      | 55     | 0      | 01:55.587    | Ivan Strinić         | Left Back                 | 2.2         |            | 1.1 NA                   | High Pass        | Left Foot           |
| FIFA World Cup | 1         | 2      | 9      | 0      | 02:09.907    | Marcelo Brozović     | Center Defensive Midfield | 5.8         |            | -1.0 Ivan Strinić        | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 2      | 11     | 0      | 02:11.507    | Ivan Strinić         | Left Back                 | 24.1        |            | 0.1 Mario Mandžukić      | High Pass        | Left Foot           |
| FIFA World Cup | 1         | 2      | 13     | 0      | 02:13.067    | Mario Mandžukić      | Center Forward            | 24          |            | 1.6 Ante Rebić           | High Pass        | Head                |
| FIFA World Cup | 1         | 2      | 42     | 0      | 02:42.747    | Kylian Mbappé Lottin | Right Midfield            | 12.6        |            | -2.8 Benjamin Pavard     | High Pass        | NA                  |
| FIFA World Cup | 1         | 2      | 53     | 0      | 02:53.507    | Samuel Yves Umtiti   | Left Center Back          | 18.7        |            | 1.3 NA                   | Ground Pass      | Right Foot          |
| FIFA World Cup | 1         | 3      | 15     | 0      | 03:15.680    | Ivan Strinić         | Left Back                 | 23.2        |            | 1.4 Mario Mandžukić      | High Pass        | NA                  |
| FIFA World Cup | 1         | 3      | 17     | 0      | 03:17.267    | Mario Mandžukić      | Center Forward            | 8.5         |            | -2.4 Ivan Rakitić        | High Pass        | Head                |
| FIFA World Cup | 1         | 3      | 19     | 0      | 03:19.787    | Kylian Mbappé Lottin | Right Midfield            | 32.2        |            | 0.5 NA                   | High Pass        | Right Foot          |

## Exemple : La première ligne du fichier

**Competition :** FIFA World Cup  
**Period :** 1  
**Minute :** 0  
**Second :** 0  
**Timestamp :** 00:00:00.400  
**Player.name :** Mario Mandžukić  
**Position.name :** Center Forward  
**Pass.length :** 11.0  
**Pass.angle :** 3.1  
**Pass.recipient.name :** Marcelo Brozović  
**Pass.height.name :** Ground Pass  
**Pass.body\_part.name :** Right Foot



# Quelques indicateurs statistiques

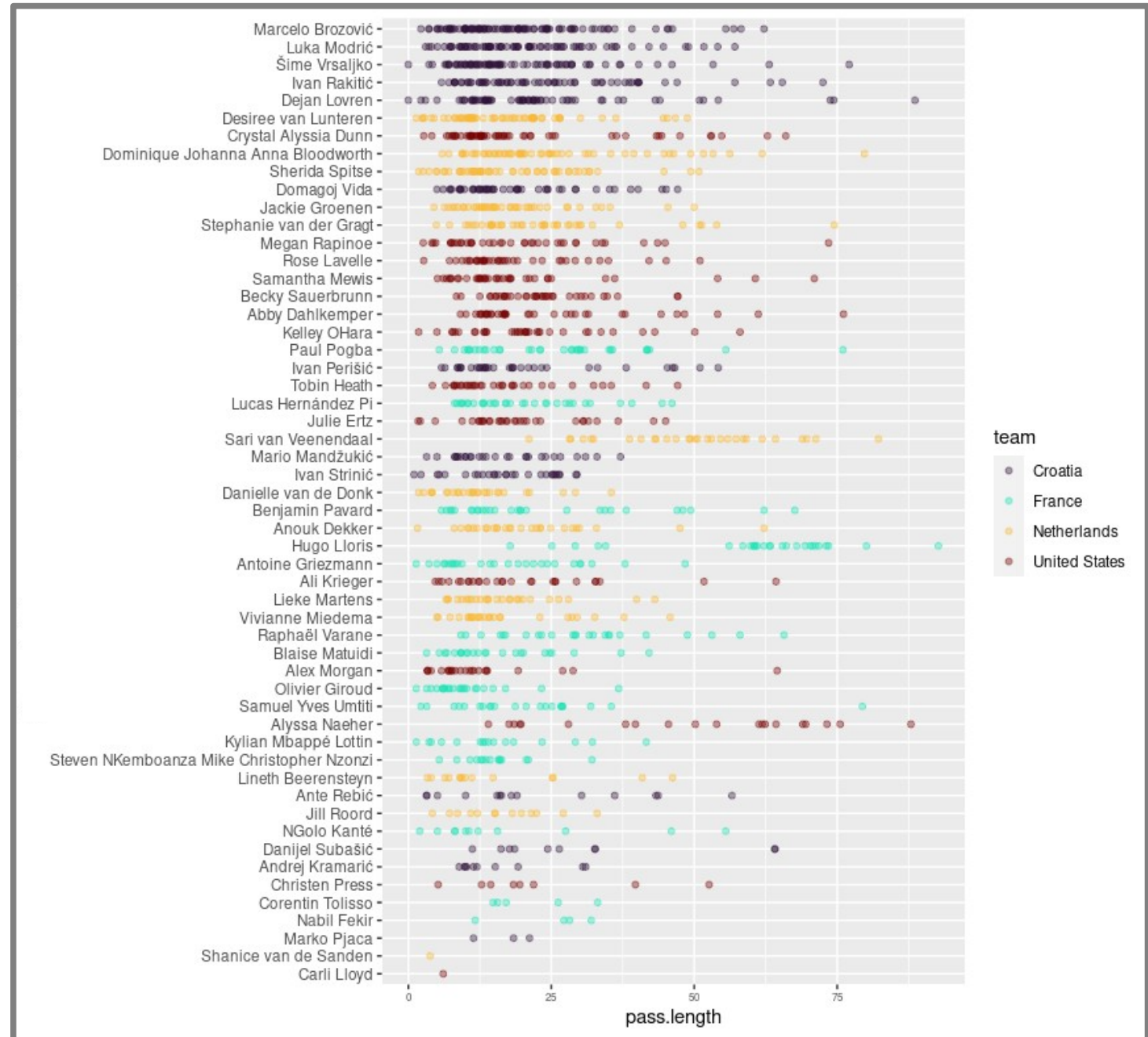
- Nombre de passes, moyenne des longueurs de passe, médiane des longueurs de passe, étendue des longueurs de passe, différence moyenne-médiane, équipe, poste
- Joueurs et joueuses triés dans l'ordre décroissant de la moyenne des longueurs de passes

| player.name                       | number | mean | median | range | mean-median | team                  | position                 |
|-----------------------------------|--------|------|--------|-------|-------------|-----------------------|--------------------------|
| Hugo Lloris                       | 27     | 59,9 | 63,2   | 74,9  | -3,3        | France                | Goalkeeper               |
| Sari van Veenendaal               | 29     | 49,6 | 50,2   | 61,1  | -0,6        | Netherlands Women's   | Goalkeeper               |
| Alyssa Naeher                     | 20     | 48,5 | 52     | 74    | -3,5        | United States Women's | Goalkeeper               |
| Danijel Subašić                   | 10     | 30,8 | 25,4   | 52,9  | 5,4         | Croatia               | Goalkeeper               |
| Raphaël Varane                    | 24     | 30,6 | 29,2   | 56,5  | 1,4         | France                | Right Center Back        |
| Paul Pogba                        | 34     | 25,9 | 25,1   | 70,6  | 0,8         | France                | Right Defensive Midfield |
| Dominique Johanna Anna Bloodworth | 55     | 25,7 | 20,2   | 73,9  | 5,5         | Netherlands Women's   | Left Center Back         |
| Abby Dahlkemper                   | 36     | 25,3 | 17     | 67    | 8,3         | United States Women's | Right Center Back        |
| Ivan Rakitić                      | 65     | 24,8 | 20,6   | 66,6  | 4,2         | Croatia               | Left Center Midfield     |
| Nabil Fekir                       | 4      | 24,8 | 27,7   | 20,4  | -2,9        | France                | Right Midfield           |
| Stephanie van der Gragt           | 39     | 24,5 | 21,8   | 69,6  | 2,7         | Netherlands Women's   | Center Back              |
| Benjamin Pavard                   | 28     | 24,2 | 18,6   | 61,8  | 5,6         | France                | Right Back               |
| Becky Sauerbrunn                  | 37     | 23,3 | 22,4   | 38,7  | 0,9         | United States Women's | Left Center Back         |
| Dejan Lovren                      | 64     | 23,3 | 20,1   | 88,6  | 3,2         | Croatia               | Right Center Back        |

|                       |    |      |      |      |     |                       |                     |
|-----------------------|----|------|------|------|-----|-----------------------|---------------------|
| Alex Morgan           | 24 | 12,9 | 9,6  | 61,1 | 3,3 | United States Women's | Center Forward      |
| Danielle van de Donk  | 28 | 12,8 | 11,3 | 33,7 | 1,5 | Netherlands Women's   | Left Wing           |
| Olivier Giroud        | 23 | 10,3 | 8,9  | 35,4 | 1,4 | France                | Left Center Forward |
| Carli Lloyd           | 1  | 6,1  | 6,1  | 0    | 0,0 | United States Women's | Right Wing          |
| Shanice van de Sanden | 1  | 3,8  | 3,8  | 0    | 0,0 | Netherlands Women's   | Right Wing          |

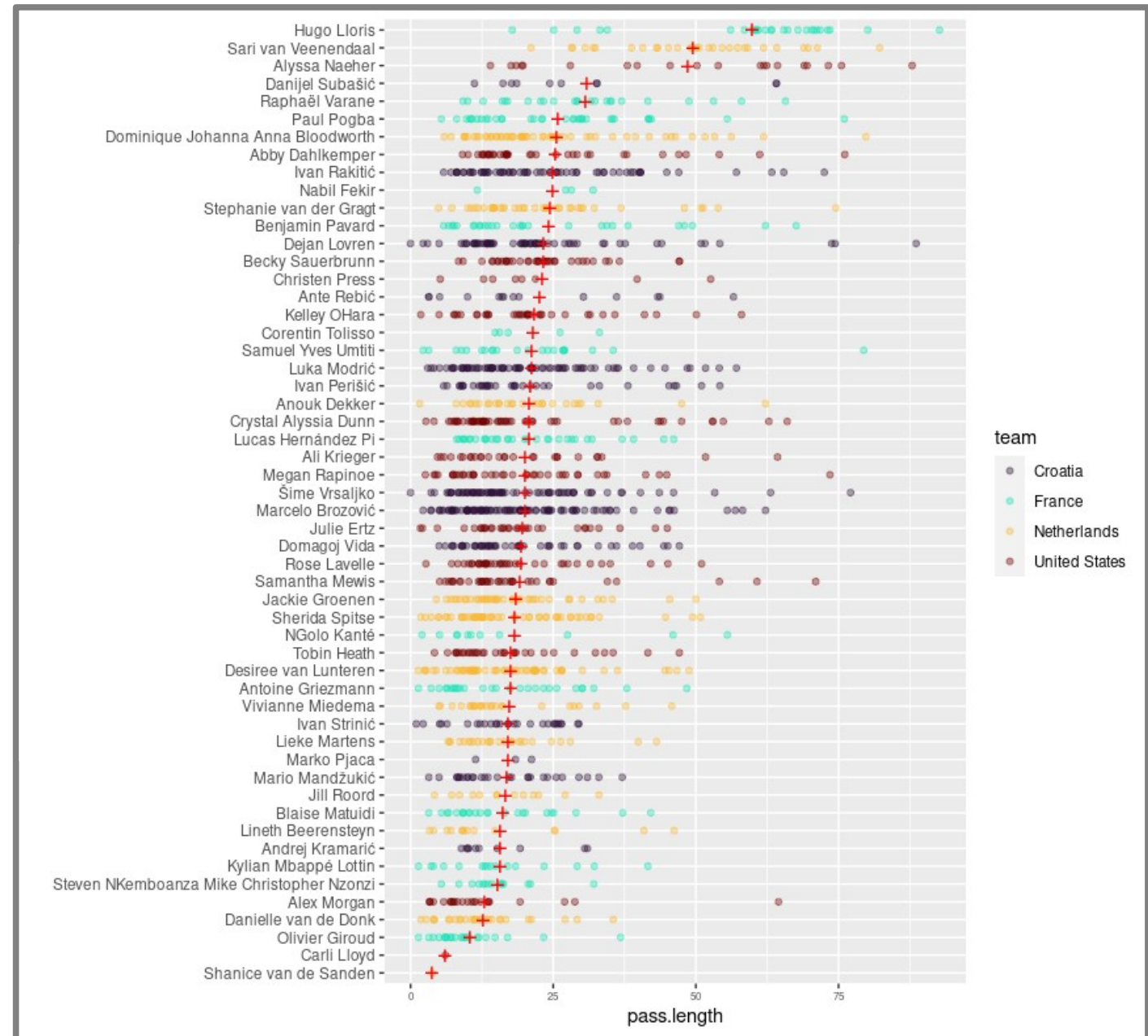
# Représentations graphiques

- longueur des passes de tous les joueurs et joueuses des finales
- de haut en bas, dans l'ordre décroissant du nombre de passes réalisées
- la couleur indique l'équipe



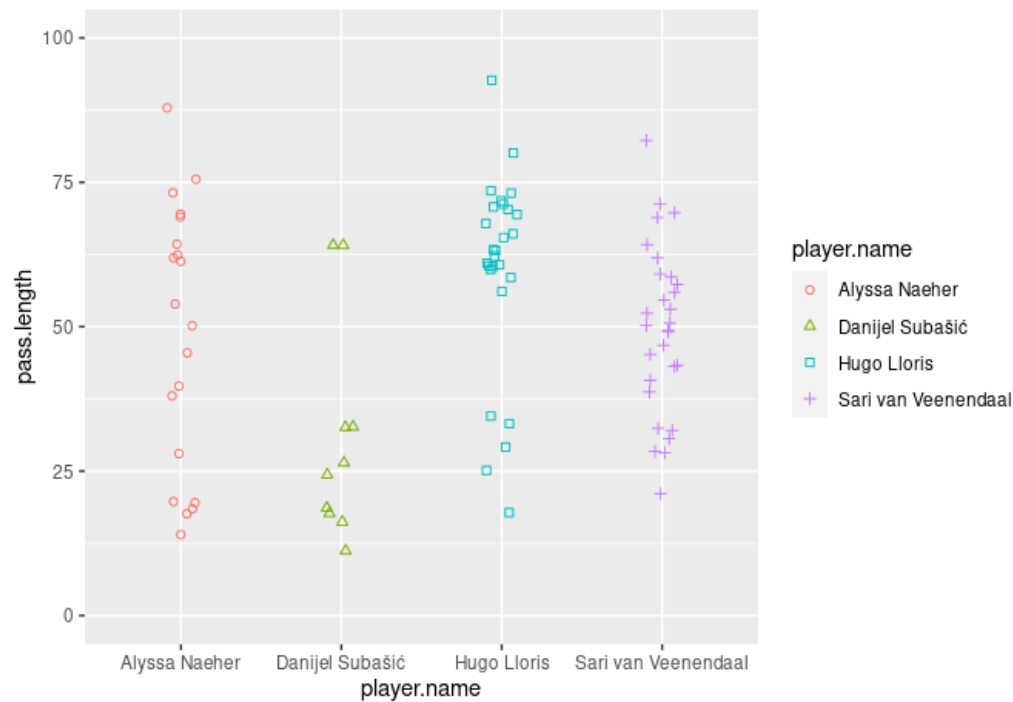
# Représentations graphiques

- longueur des passes de tous les joueurs et joueuses des finales
- de haut en bas, **dans l'ordre décroissant de la longueur moyenne des passes**
- la couleur indique l'équipe
- une croix rouge indique la moyenne de chaque joueur ou joueuse



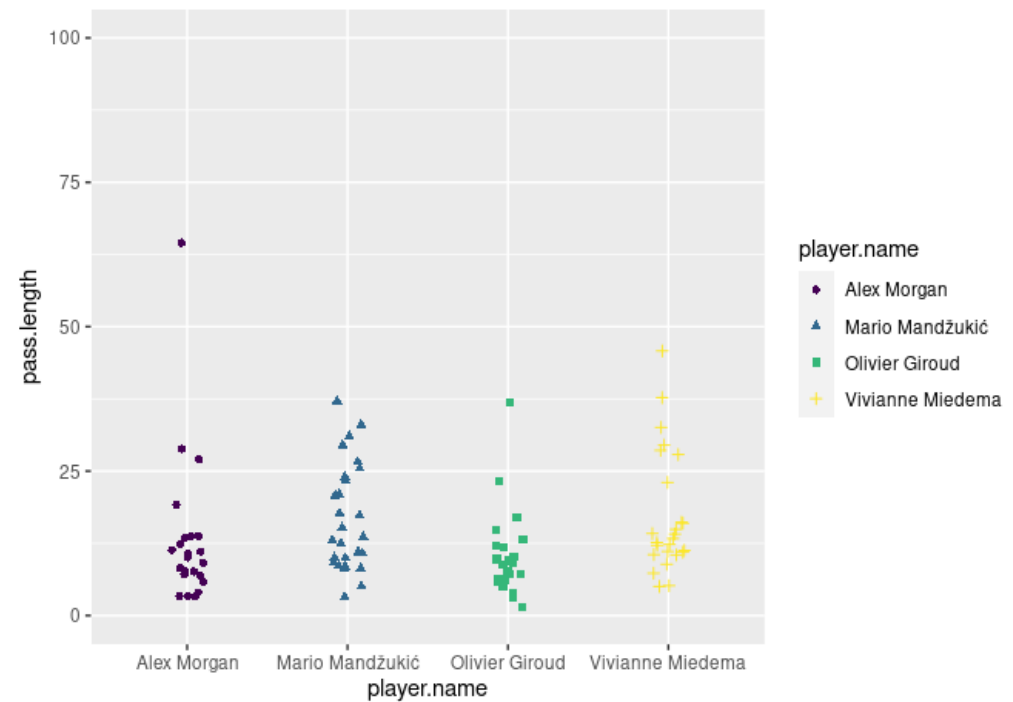
# Deux postes spécifiques

## Gardien.nes de but



|                     | Min. | Median | Mean | Max. |
|---------------------|------|--------|------|------|
| Alyssa Naeher       | 14.0 | 52.1   | 48.5 | 87.9 |
| Danijel Subašić     | 11.2 | 25.4   | 30.8 | 64.1 |
| Hugo Lloris         | 17.8 | 63.2   | 70.5 | 92.7 |
| Sari van Veenendaal | 21.1 | 50.2   | 49.6 | 82.2 |

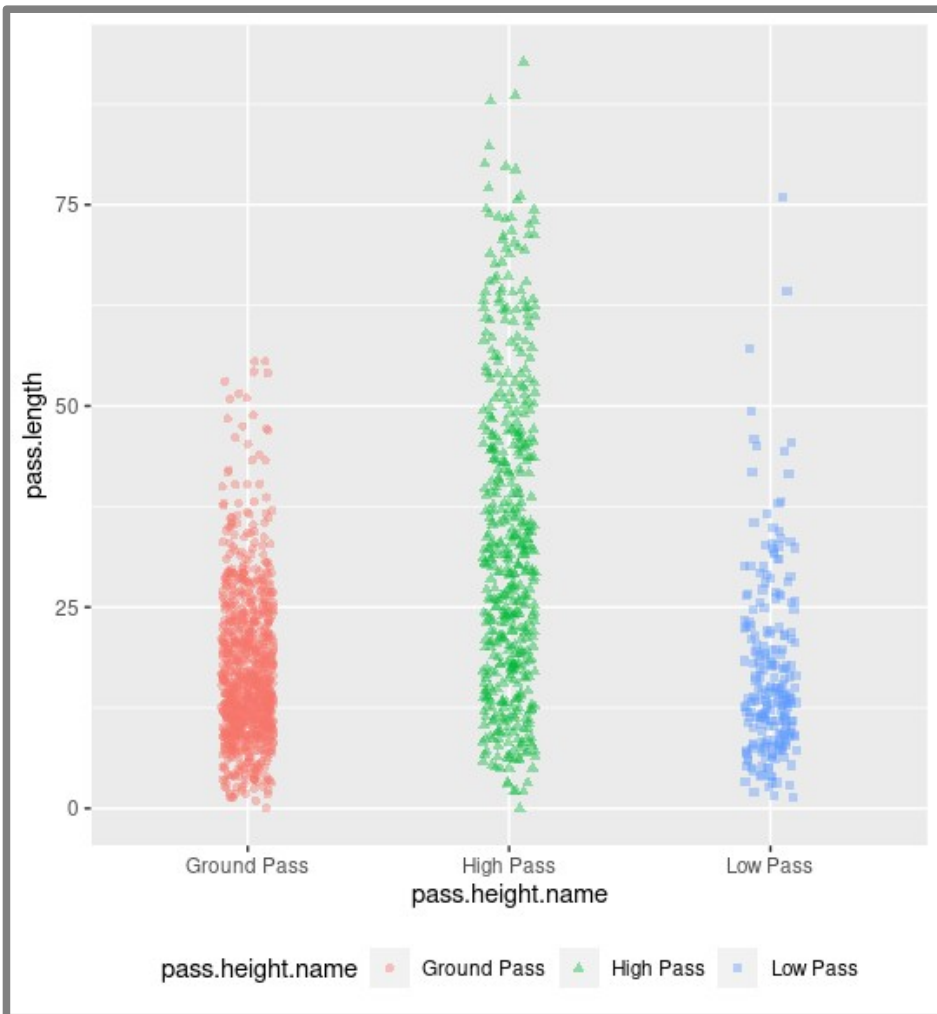
## Attaquant.es



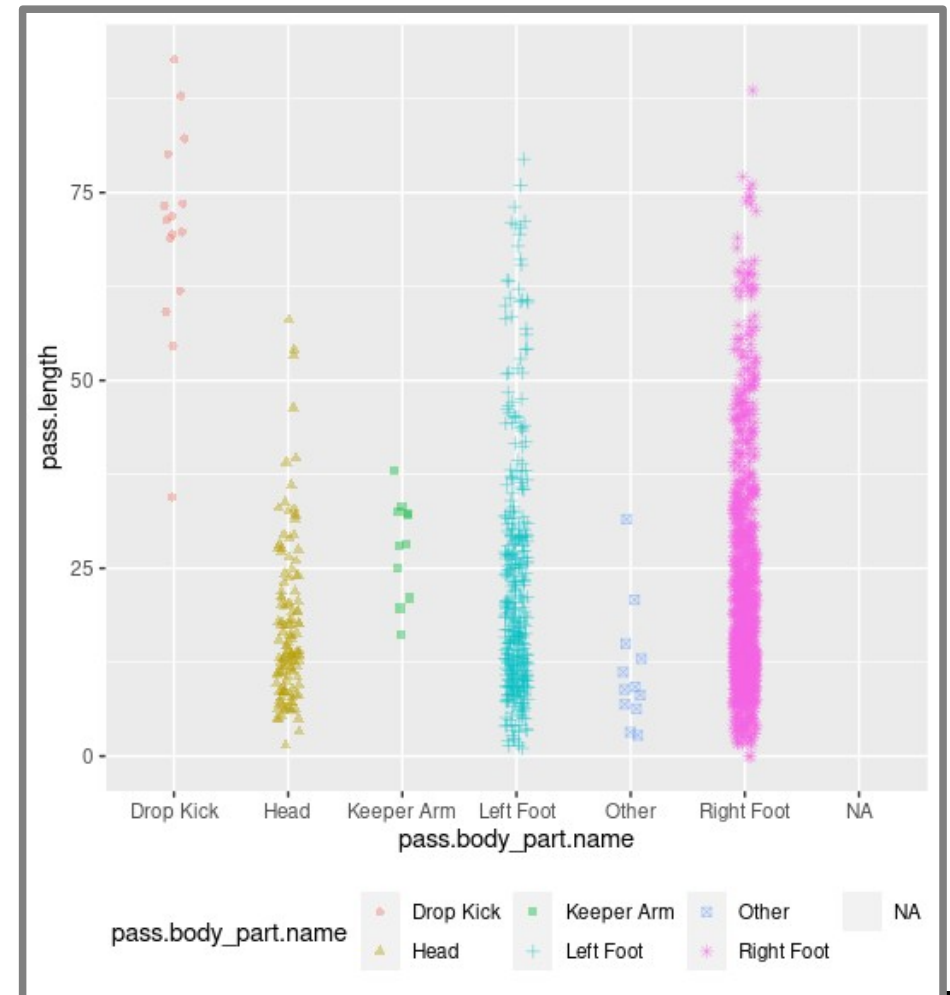
|                  | Min. | Median | Mean | Max. |
|------------------|------|--------|------|------|
| Alex Morgan      | 3.3  | 9.6    | 12.9 | 64.5 |
| Mario Mandžukić  | 3.2  | 14.4   | 16.9 | 37.1 |
| Olivier Giroud   | 1.4  | 8.9    | 10.3 | 36.8 |
| Vivianne Miedema | 5.0  | 13.3   | 17.3 | 45.8 |

# Avec d'autres paramètres

Longueur de la passe en fonction de la **hauteur**

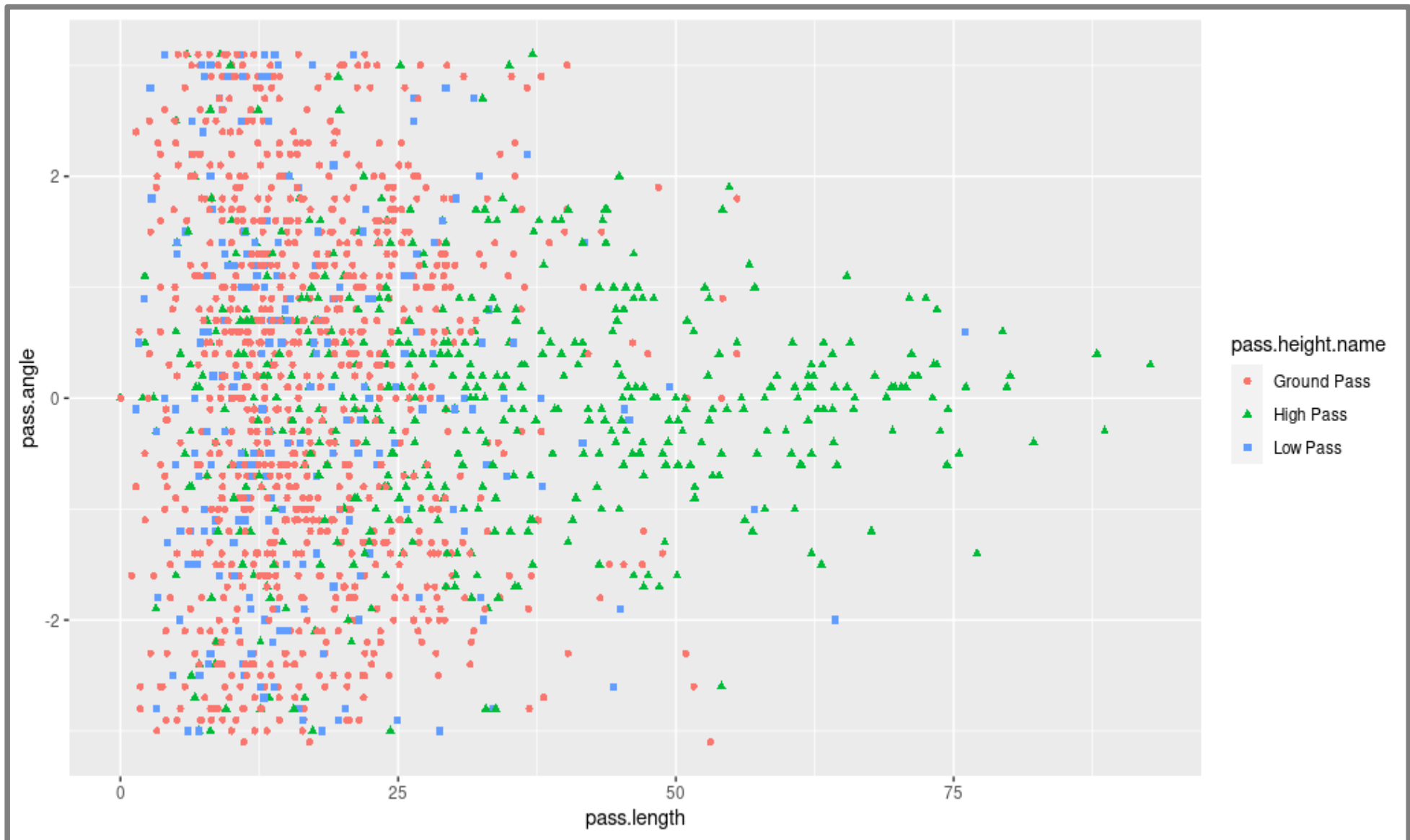


Longueur de la passe en fonction de la **partie du corps**



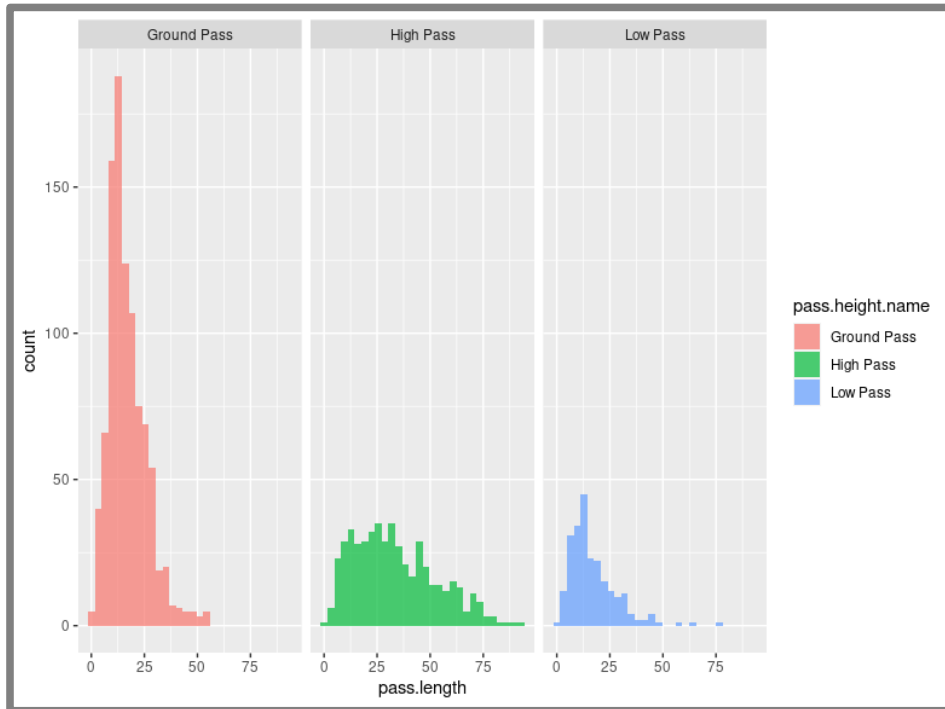
# Avec d'autres paramètres

**Angle** de la passe (axe vertical) en fonction de la **longueur** (axe horizontal) et de la **hauteur** (couleur)

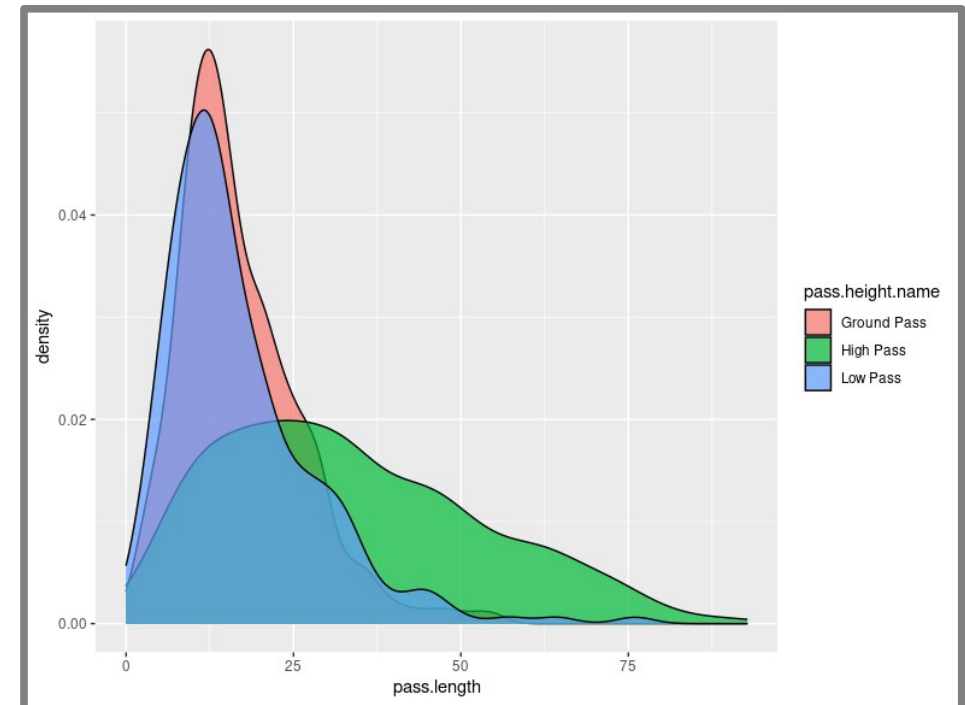
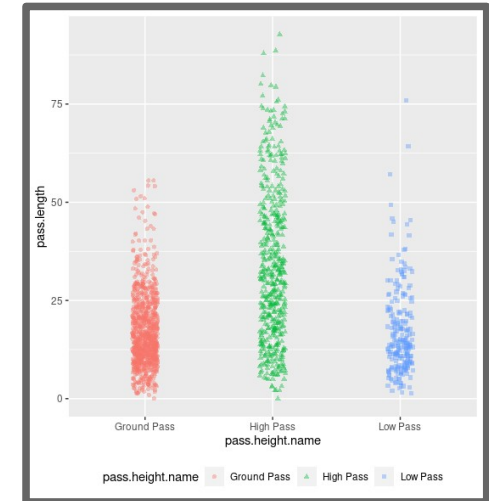


# Et d'autres représentations graphiques

Longueur de la passe en fonction de la **hauteur**



Histogrammes



Densités

# Un modèle *expected pass*

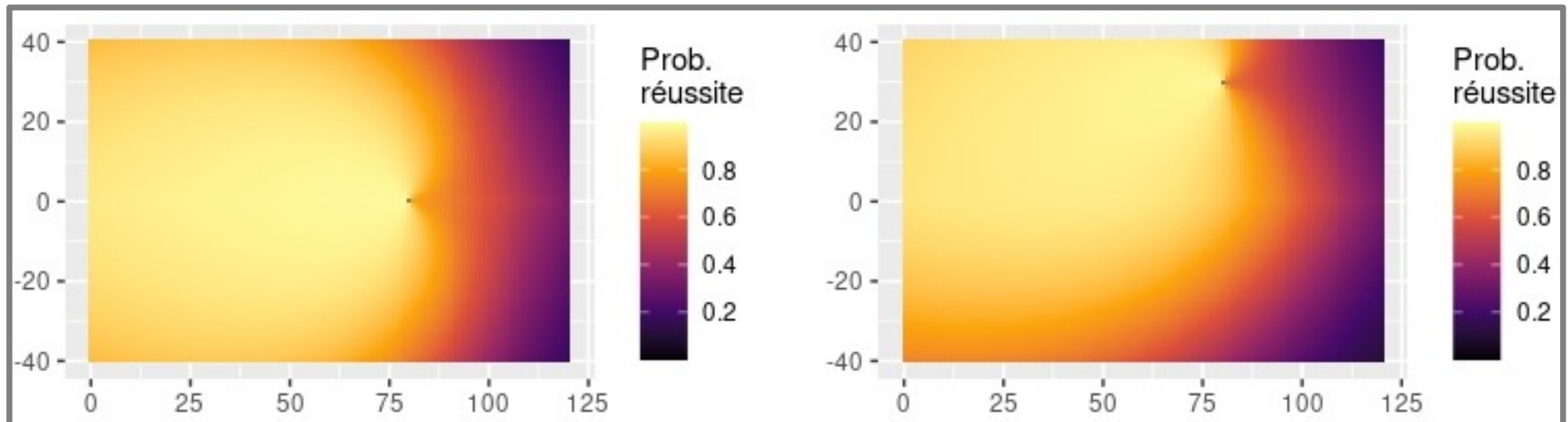
## Un modèle statistique

### Une base de données

- **64** matchs de la coupe du monde 2018
- **227 886** passes dont 215 172 réussies

$$\begin{aligned}
 \text{logit}(\pi(x_i)) &= \ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) \\
 &= \theta_0 + \theta_1 \text{length}_i + \theta_2 \text{angle}_i + \sum_{j=2}^6 \alpha_j \mathbb{1}_{\text{zone}_i=j} \\
 &\quad + \sum_{j=2}^6 \beta_j \mathbb{1}_{\text{receivingzone}_i=j} + \theta_3 \mathbb{1}_{\text{pressure}_i=\text{true}} + \theta_4 \mathbb{1}_{\text{bodypart}_i=\text{foot}} \\
 &\quad + \theta_5 \mathbb{1}_{\text{type}_i=\text{normal}} + \sum_{j=2}^{25} \gamma_j \mathbb{1}_{\text{position}_i=j}
 \end{aligned}$$

**Résultat** : prédiction de la probabilité de réussite d'une passe en fonction de la position du joueur



# Conclusion

Analytics And Data Science

## Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and DJ Patil

From the Magazine (October 2012)



Andrew J. Buboltz, silk screen on a page from a high school yearbook, 8.5" x 12", 2011. Tamar Cohen

