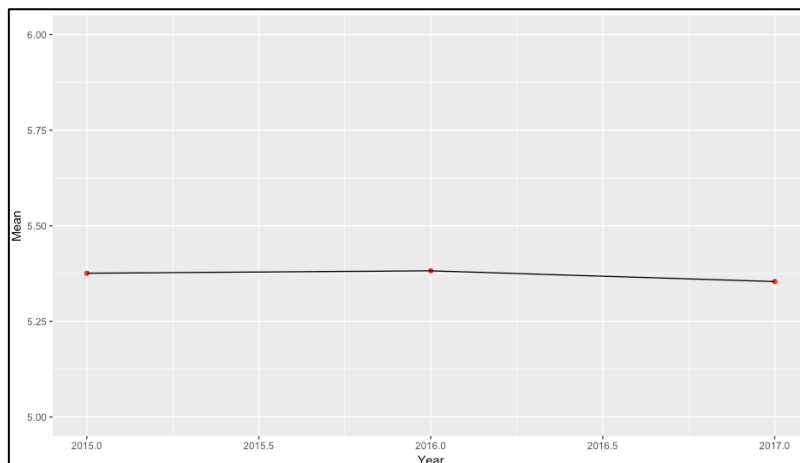


## Introduction :

Dans ce rapport, nous avons décidé de nous intéresser aux rapports sur le bonheur mondial des années 2015 à 2017. Le plus récent a été publié aux Nations Unies lors de la journée internationale du bonheur le 20 mars. Dans ces rapports, 155 pays sont classés selon leur niveau de bonheur. Afin de calculer les scores de chaque pays, les habitants interrogés doivent situer leur vie sur une échelle où 10 est la meilleure vie possible et 0 la pire. Les colonnes qui suivent la mesure du bonheur introduisent le contexte (économie, soutien social, espérance de vie...) dans lequel peut se développer ce taux de bonheur.

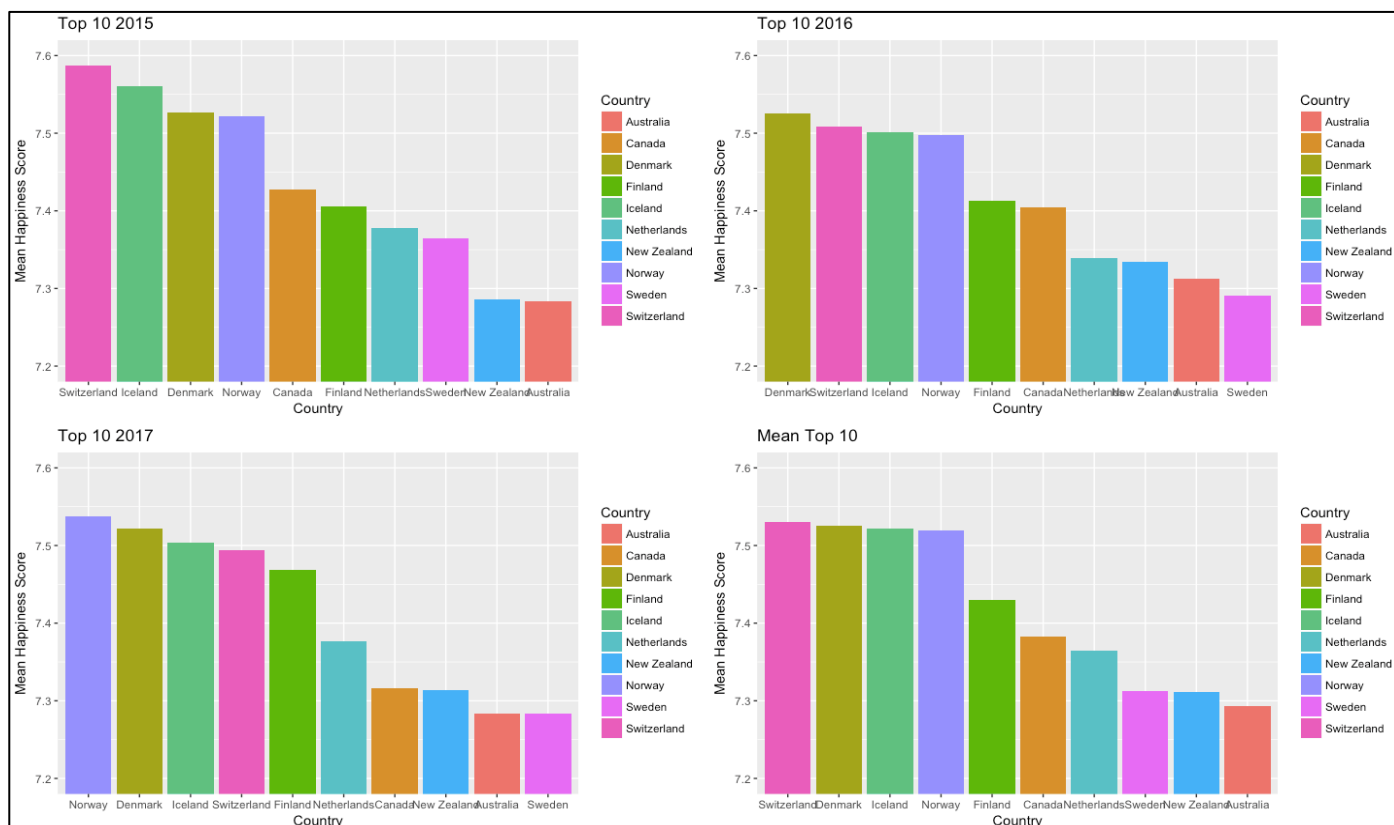
La base de données utilisée est disponible en open source sur le site Kaggle via l'adresse suivante : <https://www.kaggle.com/conniehe/happiness/data>.

## Analyse des données :



*Evolution du Score Moyen*

Le graphique ci-dessus décrit l'évolution du "Happiness Score" Moyen. On peut voir que ce bonheur moyen est stable. En effet, même si on perçoit une légère baisse entre 2016 et 2017, la différence est de l'ordre de  $10^{-2}$ . En conclusion, on peut affirmer que le bonheur moyen stagne au niveau mondial.



*10 pays avec les meilleurs Happiness Scores de 2015 à 2017 ainsi que la moyenne*

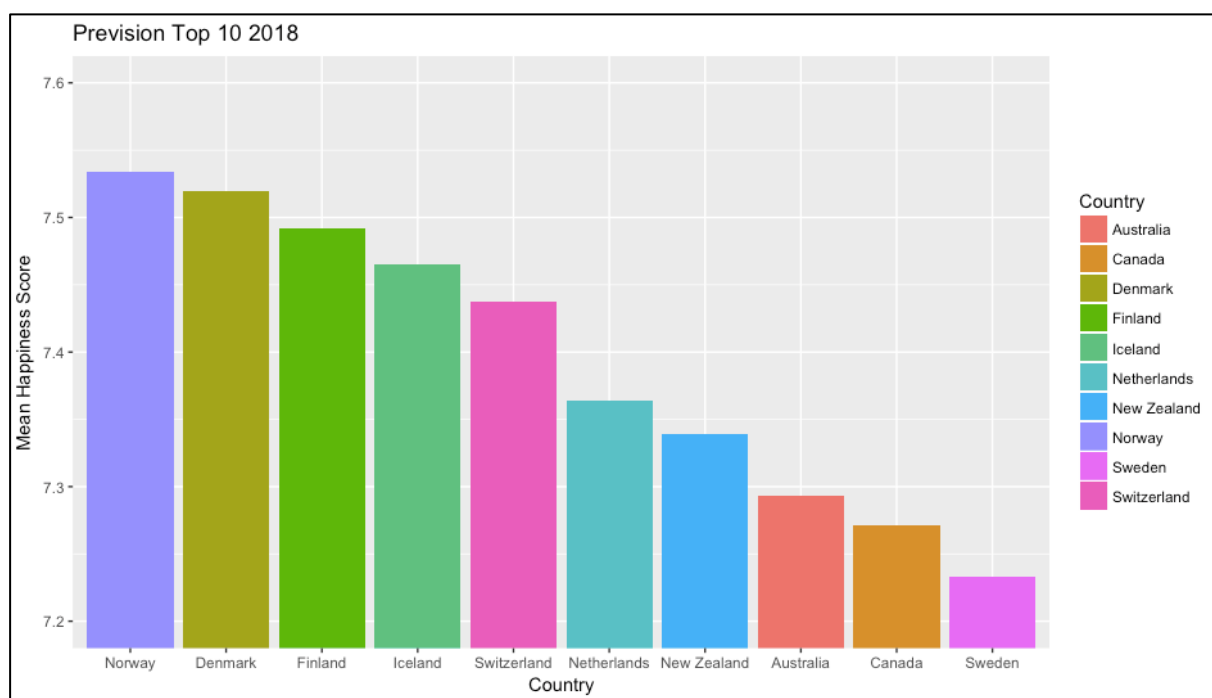
Les trois premiers graphiques ci-dessus montrent le classement des 10 pays ayant eu les meilleurs "Happiness Score". Le dernier, quant à lui, expose le classement des 10 pays ayant en moyenne les meilleurs "Happiness Score". On peut alors constater que la majorité des pays apparaissant dans ce classement sont des pays de l'Europe du Nord.

Une explication à ces résultats peut se retrouver dans une entraide de voisinage entre les citoyens et dans les programmes de soutien de l'État pour ceux qui en ont besoin. Les individus veulent se sentir en sécurité et bénéficier également d'une communauté sur laquelle ils peuvent compter - un environnement que les pays scandinaves réussissent mieux que les autres à créer. En effet, en plus de posséder un PIB par habitant élevé, les pays scandinaves apportent beaucoup d'importance au soutien social.

Une baisse de bonheur se retrouve dans les pays possédant des cultures de recrutements et de licenciements faciles. Face à cette instabilité, les individus ont tendance à devenir mécontent sur leur lieu de travail. C'est le cas pour les États-Unis qui possèdent des programmes de chômage précaires, et c'est une des raisons pour lesquelles ce pays n'est pas mentionné dans les classement ci-dessus. En revanche, la plupart des pays scandinaves ont, quant à eux, une variété de services pour les chômeurs.

Enfin, certaines théories suggèrent que, historiquement, les communautés qui vivaient dans des conditions météorologiques plus difficiles ont été rassemblées par un plus grand soutien mutuel. Ainsi, le climat plus froid des pays d'Europe du Nord pourrait en fait rendre le soutien social plus facile.

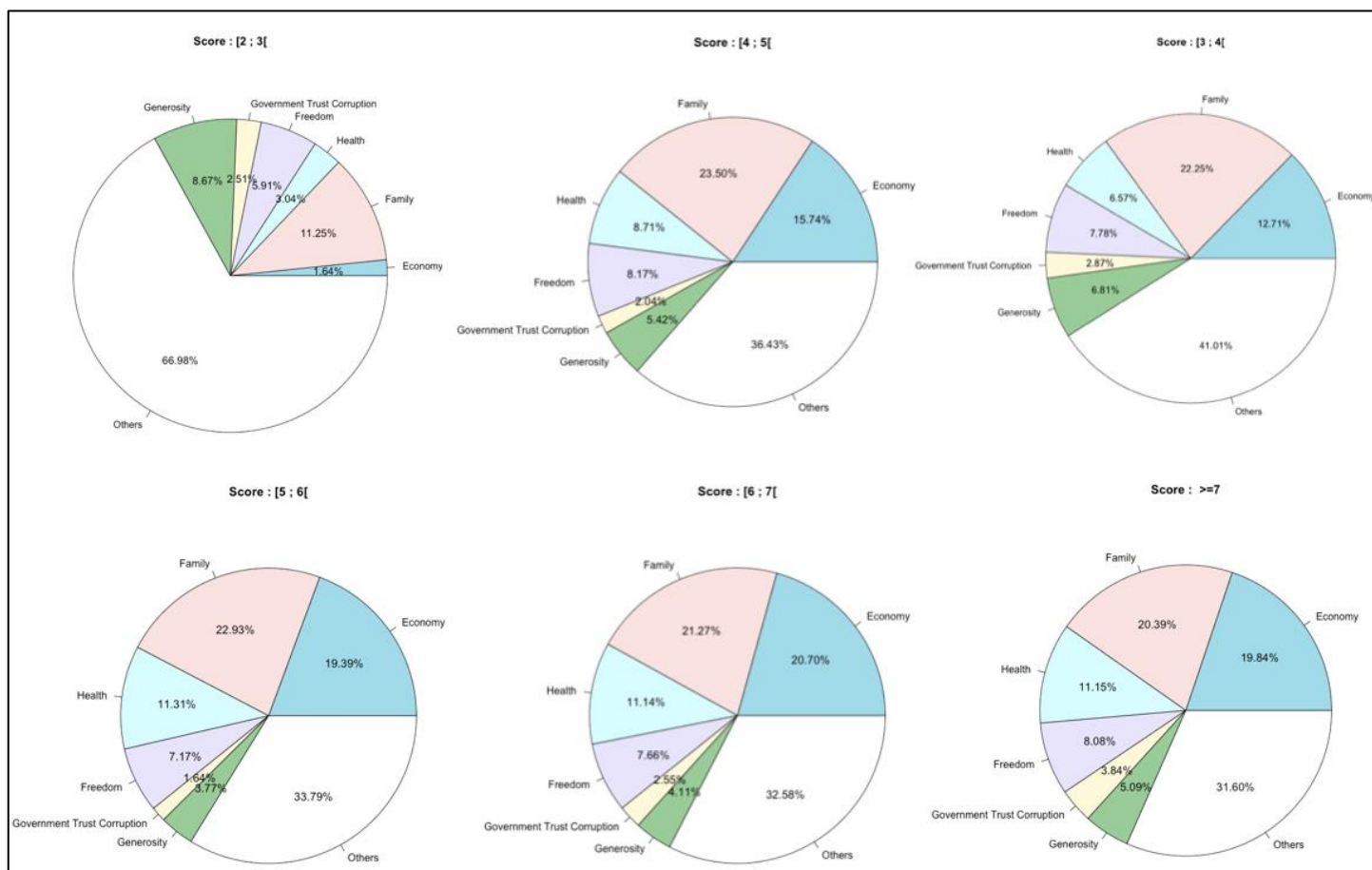
Nous avons ensuite effectué une régression linéaire afin de prédire quels seraient les scores de bonheur des pays pour l'année à venir. Voici ci-dessous ce que nous avons obtenu.



*Prévision des 10 pays avec les meilleurs Happiness Scores en 2018*

Une fois de plus, les pays qui seront les plus heureux en 2018 font majoritairement partie de l'Europe du Nord ou du moins des pays développés. C'est la Norvège qui semble se placer en tête pour l'année à venir. Ce pays jouit en effet d'un équilibre sain entre la prospérité, et le capital social. Cela se traduit par une forte cohésion sociale, des inégalités faibles mais aussi la confiance dans le gouvernement.

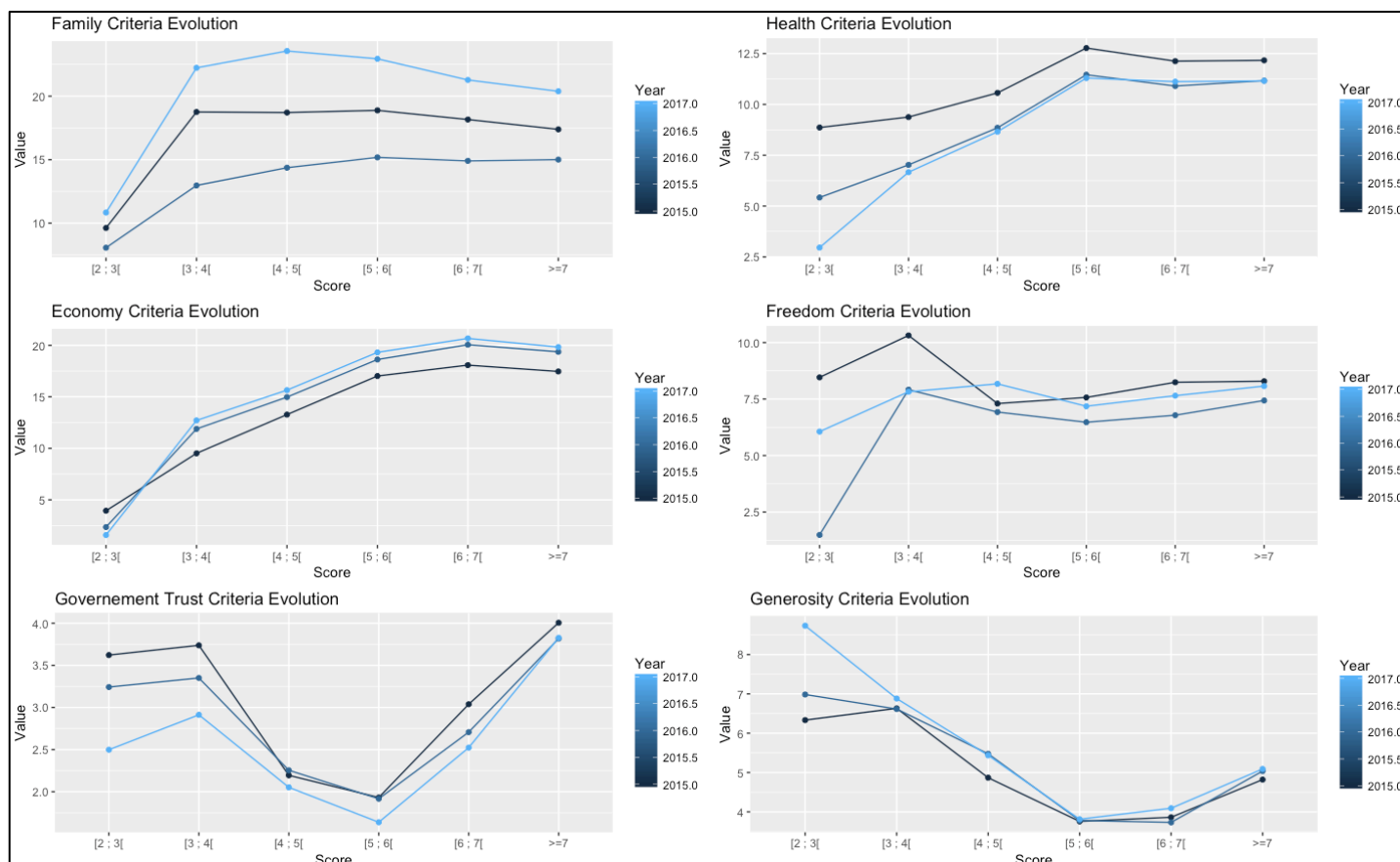
Ce dernier point reste important car son opposé influe grandement sur le mécontentement. C'est le cas de la France qui, malgré une espérance de vie forte et un PIB élevé, a perdu 3 places de 2015 à 2016, suite à la fragilisation de la confiance politique lors de la campagne présidentielle.



*Importances des critères selon le Happiness Score (réalisé pour l'année 2017)*

Les 6 critères de la base de données additionnés à la colonne *residuals* qui correspond aux autres critères donnent la valeur du score pour chaque pays. Ainsi, ces critères représentent en fait l'importance de chacun d'eux dans le bonheur indiqué. Globalement, le score du bonheur est majoritairement dû à la famille quel que soit l'intervalle du score. On observe également que le bonheur semble être moins important dans les pays où l'économie vacille. Effectivement, le premier graphique représente des scores très bas (entre 2 et 3 sur 10), et l'économie n'a quasi pas d'influence sur ce score, c'est donc certainement que cette économie n'est pas très performante. Au contraire, dans les pays ayant un score plus élevé, l'économie du pays a une des influences les plus importantes. Le paramètre qui semble globalement avoir le moins d'influence sur le bonheur est celui de la confiance envers le gouvernement. Tous les autres sont donc jugés plus essentiels par les habitants des pays.

Ces résultats confirment les suppositions émises sur les raisons pour lesquelles les pays les mieux classés sont des pays d'Europe du Nord.



*Evolution de l'importances des critères selon le Happiness Score*

Sur le graphique ci-dessus, il est possible de voir non seulement la part de chaque critère dans le bonheur en fonction du score de ce bonheur mais également l'évolution de cette part dans le temps. On peut observer tout d'abord que pour les pays ayant un score de bonheur très bas, la générosité a plus d'importance que partout ailleurs et la famille, la santé et l'économie bien moins d'importance.

D'un point de vue évolution dans le temps, la famille est le critère qui a le plus gagné en influence en 2017. La santé, quant à elle, tend à prendre de moins en moins de place dans la mesure du bonheur. Il en va de même pour la confiance envers le gouvernement, quel que soit le pays et son score. A l'inverse, de manière assez mesurée, les critères de générosité et d'économie semblent devenir de plus en plus importants. La liberté dans les pays ayant un score très bas a bien moins d'importance aujourd'hui que ce qu'elle avait en 2015.

Pour l'année 2018 à venir, il semblerait que la tendance soit à l'augmentation de l'influence des critères famille, générosité et économie, la diminution de ceux de la santé et de la confiance envers le gouvernement et la stagnation pour la liberté (possible fluctuation difficilement prédictible pour les pays aux scores les plus bas).

## **Conclusion :**

Le bonheur est une donnée complexe et variable dans le temps. Les pays qui semblent être les plus "heureux" correspondent majoritairement à des pays développés et d'Europe du Nord. Les critères influant ce bonheur dépendent énormément des pays concernés mais sont principalement la famille et l'économie. La tendance des pays d'Europe du Nord en tête du classement semble se poursuivre pour l'année (et sûrement les années) à venir.

## Rapport - Analyse exploratoire et visualisation de données

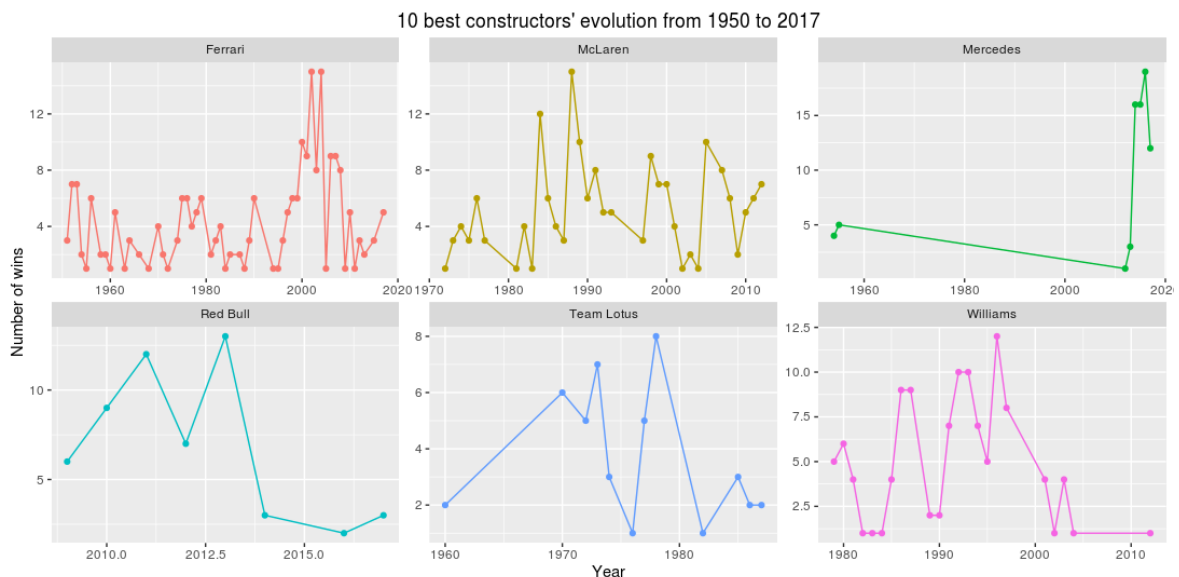
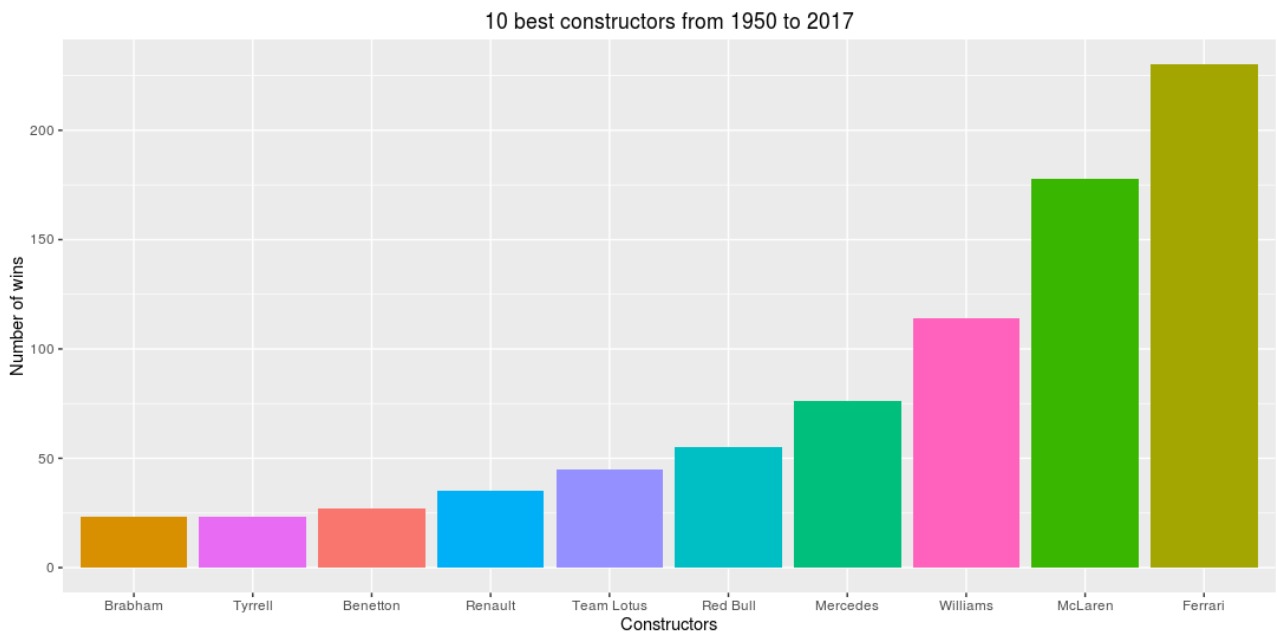
### Formule 1 : quels facteurs externes influencent les résultats des pilotes ?

#### Premier axe de recherches : la puissance des bolides.

Cela paraît évident : un véhicule plus rapide améliore grandement les performances d'un pilote. Nous avons été curieux ici sur deux points : quels constructeurs proposent les bolides les plus rapides ? et comment les véhicules se sont-ils améliorés au cours de temps ?

#### Les constructeurs

Pouvons-nous établir une corrélation entre la victoire d'un participant son écurie ? Il paraît évident que de tous les facteurs entrant en compte dans la performance d'un pilote, la qualité de son bolide n'est pas le moins important. En outre, la qualité de la prise en charge (anticipation des anomalies, diagnostic du véhicule, rapidité d'exécution) du pilote lors de ses arrêts aux stands va aussi grandement dépendre de ce paramètre. Intéressons-nous donc à l'évolution de ces écuries entre 1950 et 2017 :



Pour déterminer les 10 écuries les plus performantes, nous avons identifiés les vainqueurs de chaque course de formule 1 sans exception entre 1950 et 2017. Pour chaque année, nous avons fait la somme des victoires de chaque écurie, puis nous avons totalisé ces victoires pour chaque écurie de la base de donnée. Enfin nous avons gardé en mémoire les 10 écuries ayant totalisé le plus de victoires de 1950 à 2017. Comme l'on pouvait s'y attendre, Ferrari et McLaren sont de loin les écuries comptant le plus de victoire. Nous avons aussi pensé qu'il pouvait être intéressant de mesurer le développement de ces constructeurs au fil des ans : il ressort que Ferrari et McLaren sont des candidats assez sérieux, et ce depuis le début de notre analyse et que chacun d'eux a eu son « quart d'heure de gloire ». En revanche, leurs performances semblent s'amoinrir à l'approche de 2017 au profits de newcomers tels que Mercedes. On peut tirer différentes conclusions de cette première analyse : soit Ferrari et McLaren sont les écuries les plus performantes(/plus riches), soit Ferrari et McLaren s'arrangent pour recruter les coureurs les plus performants, soit il y a un peu des deux. On peut aussi envisager un futur dans lequel Mercedes (qui n'est peut-être pas moins perforant que Williams), opposera une concurrence féroce aux deux géants de la course automobile.

### L'amélioration de la technologie

Nous avons ensuite tenté d'analyser l'évolution de la puissance des bolides depuis 1950. Pour ce faire, nous avons compté pour chaque circuit le nombre d'années où il a été intégré au championnat. Nous avons ensuite gardé les six plus courus afin de tracer l'évolution du temps mis par le gagnant pour faire un tour.

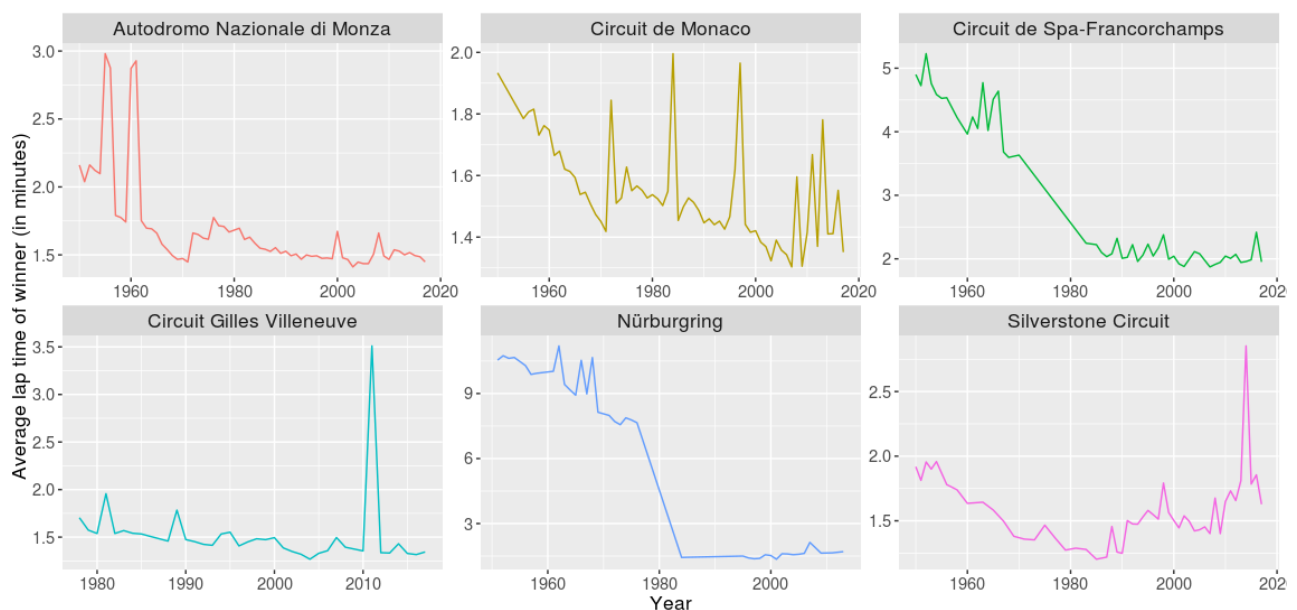


Illustration 1: Temps moyen par tour et par circuit depuis 1950

Dans un premier temps, nous pouvons effectivement observer une tendance des circuits à être exécutés de plus en plus rapidement. Cela est visible sur le circuit de **Monza** par exemple, pour lequel le temps moyen par tour du gagnant est passé progressivement de 1,7 minutes environ en 1980 à 1,5 minutes en 2010.

Cependant, ces courbes contiennent aussi d'autres informations que l'on peut recouper avec le contenu des pages Wikipédia du fichier *circuits.csv* (chaque course de chaque année a sa page sur Wikipédia). Nous obtenons ainsi plusieurs autres facteurs intéressants :

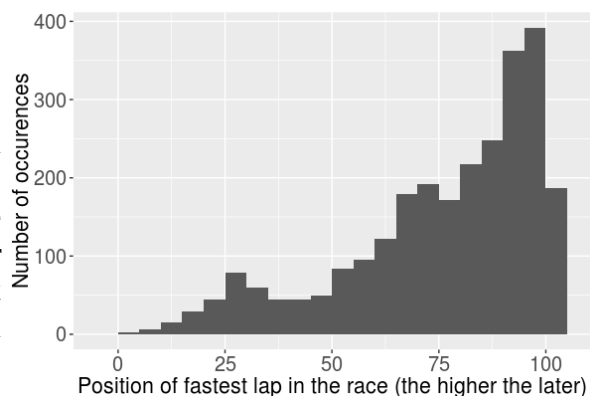
- On observe que les courbes ont tendance à "coller" par le haut à leur tendance, avec des anomalies ponctuelles sous forme de fort pic local. Cela est particulièrement visible sur le **Circuit Gilles Villeneuve**, où le meilleur temps est doublé en 2011 par rapport aux années d'avant. Wikipédia indique qu'il y a eu de très fortes pluies pour ce grand prix là. Ces courbes peuvent donc nous donner des indices sur les conditions météo.

- Ensuite, certaines courbes ont des cassures, comme par exemple **Nürburgring** entre 1975 et 1985, ou **Monaco** en 1971. Ceux-ci sont liés à des restructurations des circuits.

## Deuxième axe de recherches : la psychologie des pilotes.

Il y a dans *results.csv* une colonne digne d'une analyse : *fastestLap*, indiquant à quel moment de la course le pilote a fait son meilleur tour. Nous avons voulu voir si les pilotes avaient une tendance à être plus performants en début ou en fin de course.

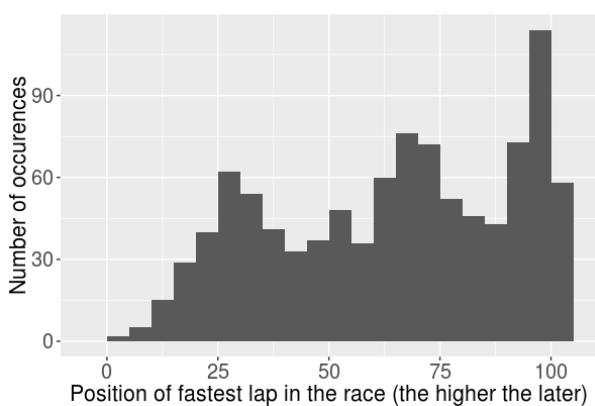
Nous avons donc gardé pour chaque course uniquement les pilotes ayant fait la totalité des tours, puis divisé leur *fastestLap* par le nombre total de tours, donnant un pourcentage indiquant à quel moment les instants de gloire ont eu lieu. Nous avons obtenu l'histogramme ci-contre.



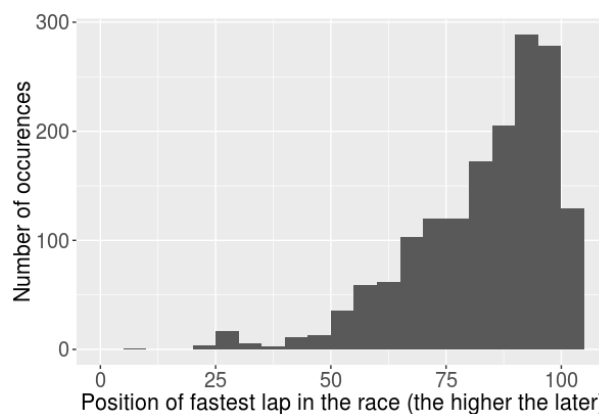
*Illustration 2: Tours les plus rapides depuis 1950*

Cependant, en recherchant des explications pour le petit pic à 25 %, nous avons trouvé (toujours sur Wikipedia), que les règles des arrêts aux stands ont changées en 2010. En effet, alors que le remplissage du réservoir d'essence était monnaie courante auparavant, il a été interdit à partir de cette année là, forçant les pilotes à courir avec un seul plein pour toute la course.

Sachant que le carburant représente un poids non négligeable influençant la vitesse du bolide, nous avons donc séparé les données en deux pour regarder les résultats avant et après 2010. Nous obtenons les deux graphiques suivants :



*Illustration 3: Tours les plus rapides entre 1950 et 2010*



*Illustration 4: Tours les plus rapides entre 2010 et 2017*

On voit ainsi que l'état du réservoir est crucial pour la vitesse du véhicule : les données depuis 2010 montrent clairement que les tours les plus rapides se situent vers la fin de la course.

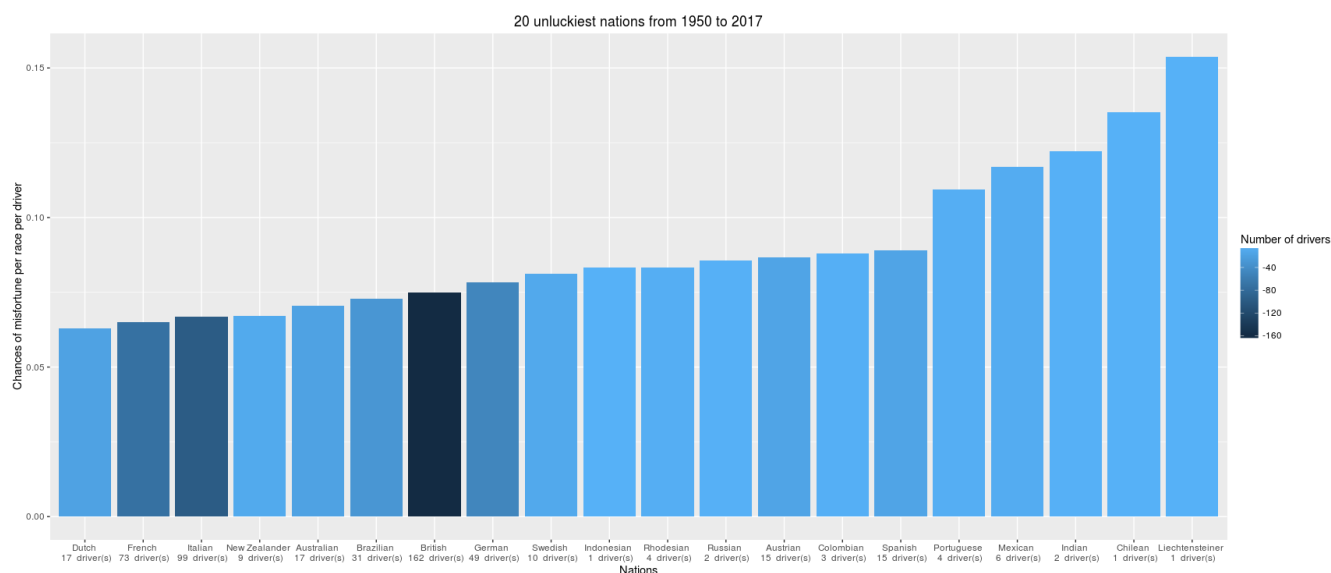
Sur le graphique d'avant 2010, il est assez simple d'expliquer les pics à 25 % et juste avant 75 %. En effet, avant l'interdiction de faire le plein, les pilotes faisaient la plupart du temps deux arrêts aux stands, avec un véhicule plus lourd en sortie. Les chances de faire un tour plus rapide augmentant plus le dernier plein était ancien, on peut en déduire qu'en général, les pilotes s'arrêtaient entre 30 et 40 % de la course une première fois et entre 60 et 80 % pour la deuxième fois. Le léger pic à 50 % doit provenir de ceux qui optent pour trois arrêts aux stands.

Cependant, pour ce graphique d'avant 2010, nous n'arrivons pas à expliquer le maximum qui se trouve vers 90 %. Peut-être qu'effectivement, les pilotes sont plus à cran à la fin de la courses et roulent donc plus vite...

Ainsi, nous n'avons pas pu dégager d'informations sûres concernant la psychologie des pilotes, mais nous avons trouvé un autre facteur d'une importance certaine : le réservoir.

### Troisième axe de recherches : la nationalité, la dangerosité, et la poisse.

Le monde de la formule 1 est malheureusement impitoyable et peut se montrer dangereux à certaines occasions, d'autant plus que les pilotes peuvent atteindre des vitesses vertigineuses et qu'il devient alors difficile de prendre des décisions sur la volée. Un moteur qui s'enflamme, un virage mal négocié ou une anomalie non détectée par les mécaniciens et un drame est vite arrivé. Nous pensons que l'aptitude du pilote à maîtriser son véhicule et à réagir dans l'immédiat fait partie intégrante de son talent. Il peut donc être intéressant de porter un regard plus attentif à la « sûreté » d'un candidat. Puis nous avons pensé qu'il pouvait être encore plus intéressant de déterminer s'il existe un pays dont les ressortissants sont les plus dangereux :



Pour établir ce graphe, nous avons calculé pour chaque pilote, le nombre total d'accidents (blessures, collisions, exclusions) entre 1950 et 2017, que nous avons divisé par le nombre total de courses courues. Nous obtenons ainsi un pourcentage proposant une probabilité pour l'individu de de rencontrer une mésaventure.

Ensuite, nous avons sommé ces pourcentages pour chaque nationalité et divisé par le nombre total de pilotes du pays, pour obtenir une répartition de ce pourcentage « par pilote, par pays ». Évidemment, il existera des pilotes plus ou moins dangereux et ce pourcentage sera d'autant moins pertinent que le nombre de pilotes du pays sera faible, mais nous pensons que cela peut donner un ordre d'idée. Par ailleurs, nous souhaitons préciser que nous gardons en tête que l'origine et l'ethnie sont à dissocier du talent. Si les nationalités pouvaient influencer les performances de leurs pilotes, ce serait avant tout un problème de fonds, d'organisation, d'expérience ou d'importance accordée au sport (entre autres).

Nous observons ici que des pays tels que le Liechtenstein et le Chili disposent chacun d'un pilote particulièrement fougueux avec plus de 10 % de chances d'accidents par course. Les Mexicains, ayant 6 pilotes, ont moins d'excuses avec leurs 10 % d'accidents par course par pilote. Parmi les pays comptant le plus de coureurs, l'Allemagne et l'Angleterre sont les plus dangereux (même s'ils sont en 13<sup>e</sup> et 14<sup>e</sup> place du podium derrière d'autres pays disposant de moins de coureurs). Nous bons français sommes en bonne position en 19<sup>e</sup> place avec environ 6 % d'accidents par course, par pilote pour 73 membres. Des pays tels que les États-Unis comptabilisent si peu de chances d'accidents qu'il n'apparaissent même pas parmi les 20 premiers.



---

## Table des matières

Introduction . . . . .	1
1 Le candidat type pour un speed dating . . . . .	2
1.1 Âge et origine . . . . .	2
1.2 Études réalisées . . . . .	3
2 Qu'est-ce que les candidats attendent de cet événement ? . . . . .	4
2.1 Qu'est ce que les participants attendent de la relation ? . . . . .	4
2.2 Qu'est ce que les participants recherchent chez leur partenaire ? . . . . .	4
3 Y'a-t-il des corrélations entre ces variables ? . . . . .	6
3.1 Différence d'objectifs en fonction des revenus . . . . .	6
3.2 Age et matchs . . . . .	6
Conclusion . . . . .	8

## Introduction

Trouver le partenaire idéal, celui qui nous correspond peut parfois s'avérer compliqué et peut être chronophage pour un résultat qui n'était pas celui escompté... C'est pourquoi de nombreuses personnes se sont tournées vers le speed dating. C'est l'objet de notre analyse, nous nous demandons qui participe à des speed dating, que recherche les participants ou encore si l'on peut observer une corrélation entre certaines variables et les matchs entre deux personnes ? Notre analyse vise à fournir une compréhension basique du comportement humain dans ce genre de situation. Le dataset utilisé retrace plusieurs vagues de personnes qui ont été interrogées. Une vague est déterminée par la date de réalisation de l'étude d'un speed dating. L'étude contient un grand nombre de variables, allant des plus basiques telles que l'âge, l'origine, les études ou la profession, à des variables plus complexes telles que les attentes des personnes, leur objectifs, ce qu'ils pensent des attentes des autres candidats et bien plus encore.

## 1 Le candidat type pour un speed dating

En premier lieu on peut se demander quel est le profil des personnes qui se présentent à des événements de speed dating, en étudiant la répartition des participants en fonction de leur ethnie et de leur âge.

### 1.1 Âge et origine

On remarque que les hommes participant au speed datings sont légèrement plus âgés que les femmes. Cependant chez les deux sexes, le plus gros des participants se situe entre 20 et 30 ans. En terme d'ethnie, la répartition homme/femme est relativement respectée. On trouve dans cette étude des participants avec des origines variées.

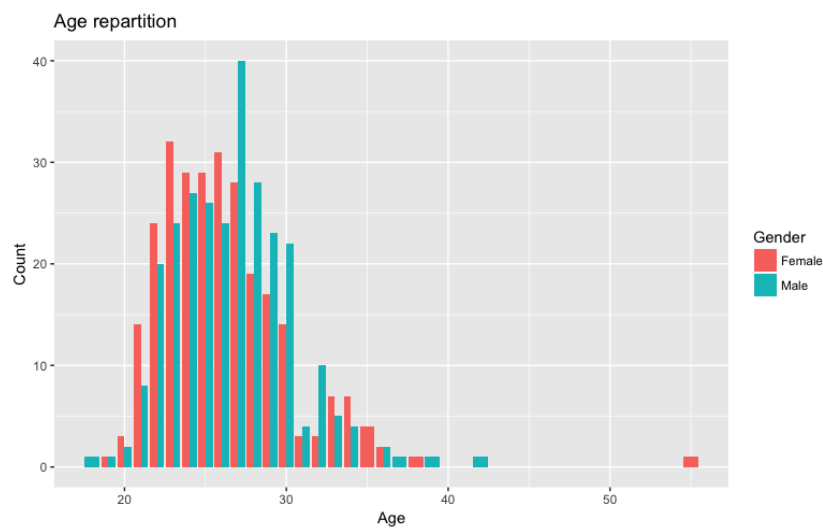


Fig. 1: Répartition des participants selon leur age

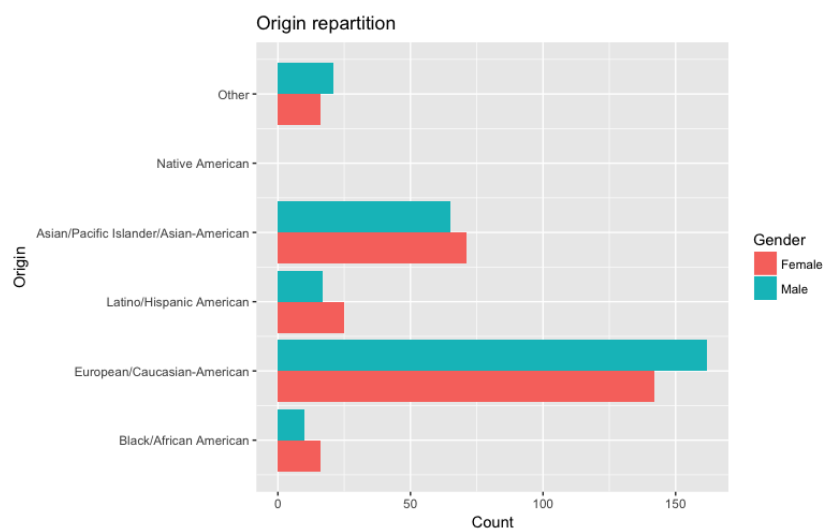


Fig. 2: Répartition des participants selon leur origine

## 1.2 Études réalisées

Sur l'ensemble des vagues étudiées nous intéressons maintenant aux études effectuées par les participants. On remarque sur la figure 3 que pour les hommes une large partie des candidats sont issus d'études en économie, business et finance. Pour les femmes, la provenance scolaire est assez bien répartie. Les études prédominantes de ces dernières se trouvent être les sciences sociales et la psychologie suivies de près à nouveau par les études d'économie et de finance.

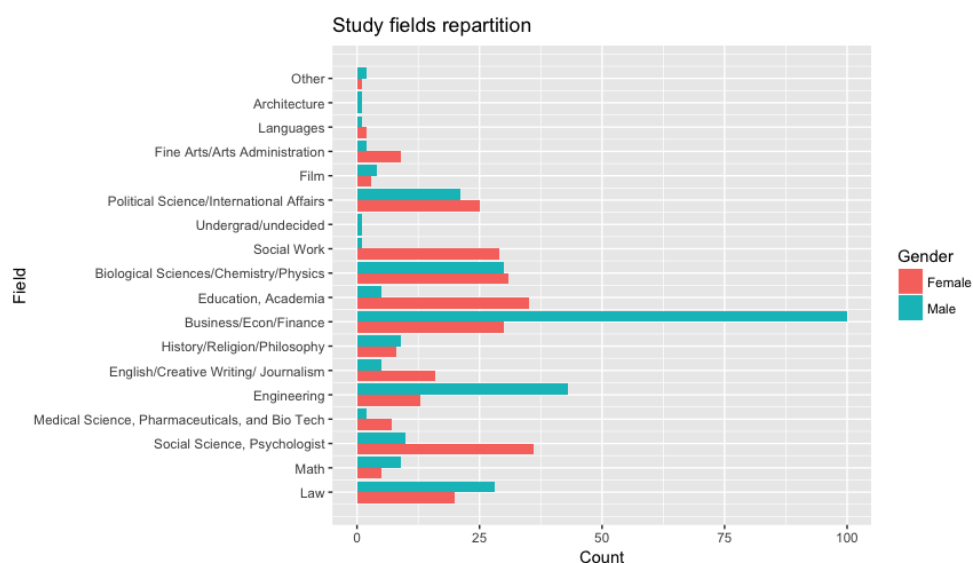


Fig. 3: Répartition des participants selon leur domaine d'études

## 2 Qu'est-ce que les candidats attendent de cet événement ?

On s'est ensuite attardé sur les attentes des personnes lors d'un speed dating. Notamment ce que recherchent les participants, leur objectifs, et même ce qu'ils pensent que le sexe opposé recherche chez eux.

### 2.1 Qu'est ce que les participants attendent de la relation ?

Les participants possèdent différentes attentes de ce genre d'événements. C'est ce qu'essaie de représenter la figure 4. On en déduit que pour une grande majorité ils sont simplement là pour s'amuser ou rencontrer du monde que ce soit pour les hommes ou pour les femmes. Certains sont venus pour décrocher un rendez-vous ou s'engager dans une relation, mais ils sont très minoritaires. En revanche, si l'on oppose cette analyse à la moyenne des matches prévus

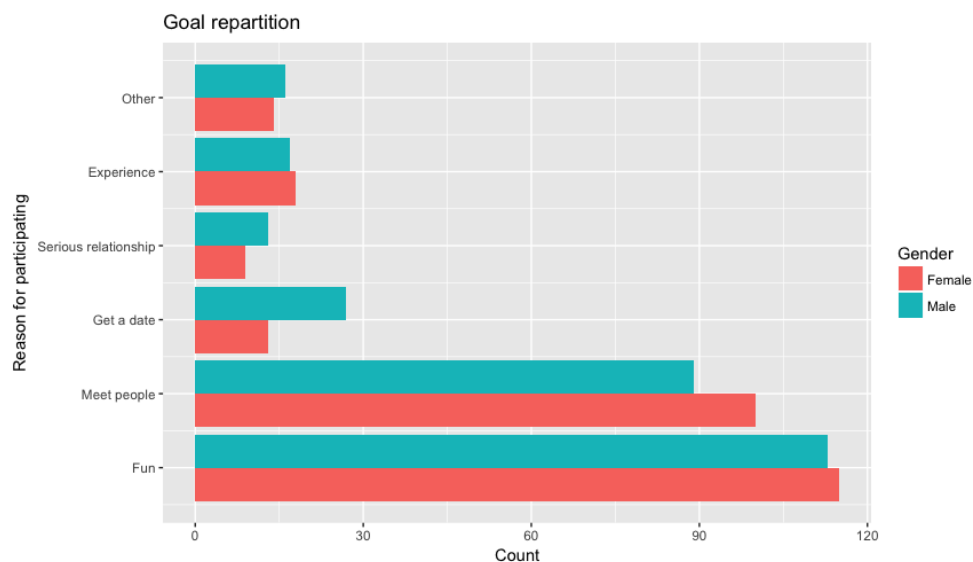


Fig. 4: Ce que les participants recherchent de la relation

par chacun on se rend compte que les participants n'ont pas forcément dévoilé leurs réelles intentions. En effet cette moyenne est de **3** pour les femmes et **3,4** pour les hommes.

### 2.2 Qu'est ce que les participants recherchent chez leur partenaire ?

La figure 5 montre ce que les participants recherchent chez leur éventuel partenaire, en différenciant les hommes et les femmes. Les participants devaient répartir 100 points entre divers traits de caractères :

- L'attractivité physique
- La sincérité
- L'intelligence
- L'humour
- L'ambition
- Le fait d'avoir des intérêts/hobbys en commun

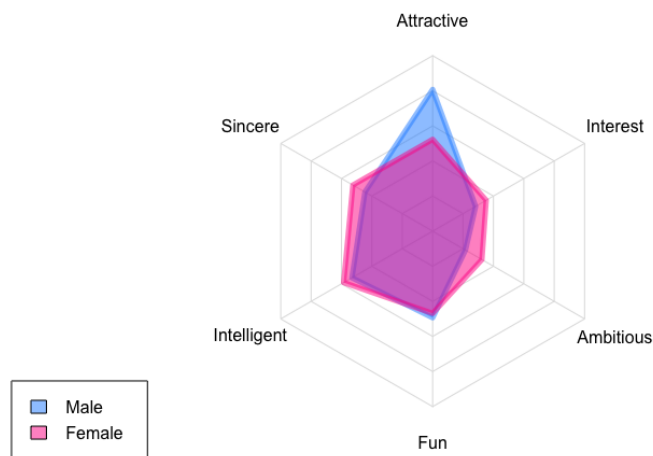


Fig. 5: Ce que les participants recherchent chez leur éventuel match

L'analyse du graphe est intéressante puisqu'elle montre une différence très nette au niveau de l'attractivité. Elle représente un gros pic chez les hommes, suivi de l'humour. Tandis que chez les femmes, l'attractivité se situe au même niveau que la sincérité et l'intelligence.

### 3 Y'a-t-il des corrélations entre ces variables ?

Dans cette partie, nous allons étudier certaines variables en fonction d'autres pour essayer de montrer quelques corrélations.

#### 3.1 Différence d'objectifs en fonction des revenus

On cherche ici à montrer que les objectifs d'un speed dating peuvent varier plus ou moins suivant le revenu annuel brut des personnes. En effet, on remarque que l'espérance pour laquelle les revenus sont les plus élevés, environ 60k\$ bruts annuels, concerne la catégorie de personnes qui recherche des relations sérieuses. Au contraire, l'espérance pour laquelle les revenus sont les plus bas, environ 35k\$ bruts annuels, concerne plutôt la catégorie de personnes qui sont simplement là pour l'expérience. Ainsi, les objectifs recherchés et les revenus de chacun semblent être des variables plutôt en lien.

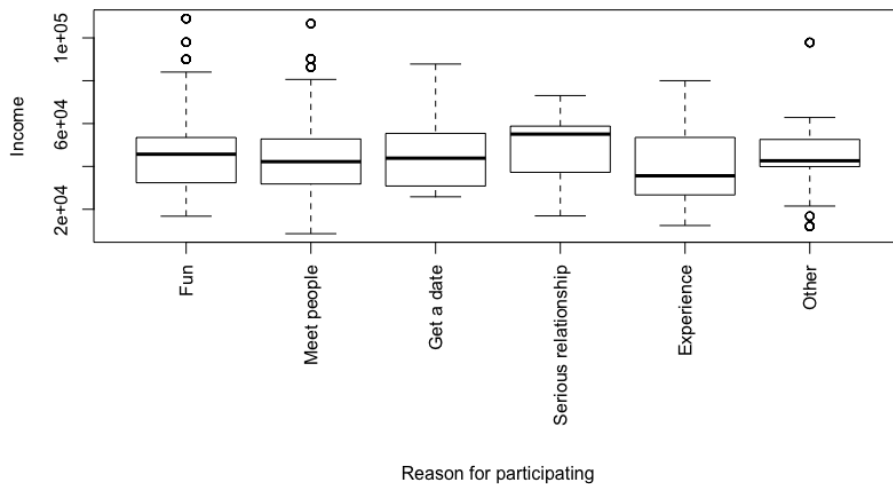


Fig. 6: But recherché en fonction des revenus

#### 3.2 Age et matches

La figure 7 nous montre une corrélation assez importante entre le nombre de matches et la différence d'âge entre les participants. En effet, plus la différence d'âge est importante, moins il y a de matches entre deux personnes. La corrélation entre ces deux variables est de  $-0.9$ .

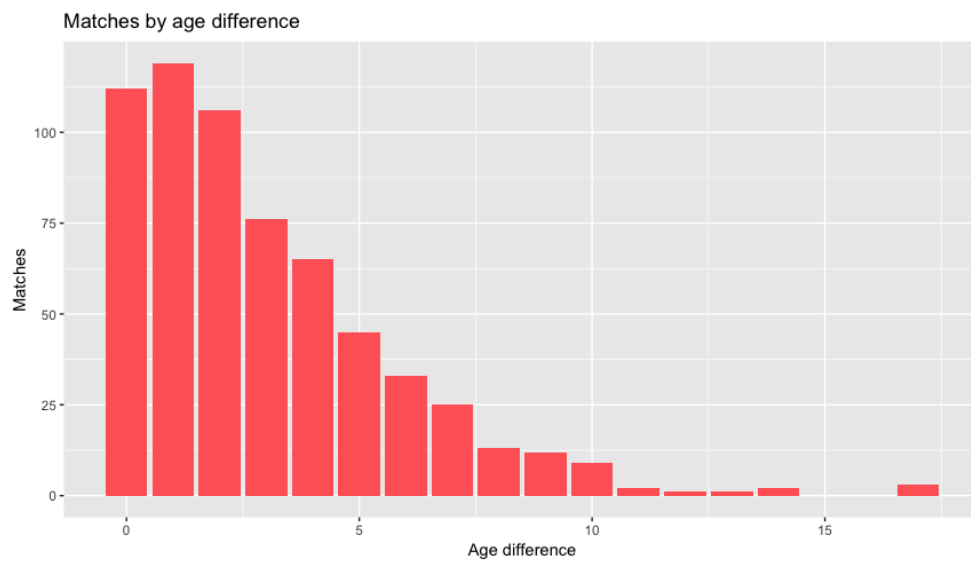


Fig. 7: Nombre de match en fonction de la différence d'âge entre les deux participants



## Conclusion

En conclusion, on a pu se faire une idée du type de personnes intéressées par les événements de speed dating et en particulier ce qu'ils attendent des rencontres. Il était également intéressant d'observer les corrélations entre les matchs et certaines variables, comme la différence d'âge. Il reste cependant de nombreuses analyses à faire sur ce dataset très riche. On pourrait par exemple faire une étude plus approfondie de ce que chacun pense des participants du sexe opposé, ou encore de l'évolution de leurs attentes avant et après les rencontres, ou bien des différences de perception de soi par l'autre en fonction du sexe ou de l'origine, si les personnes se surévaluent ou sous-évaluent, et bien plus encore.

## Portrait du Don Juan moderne.

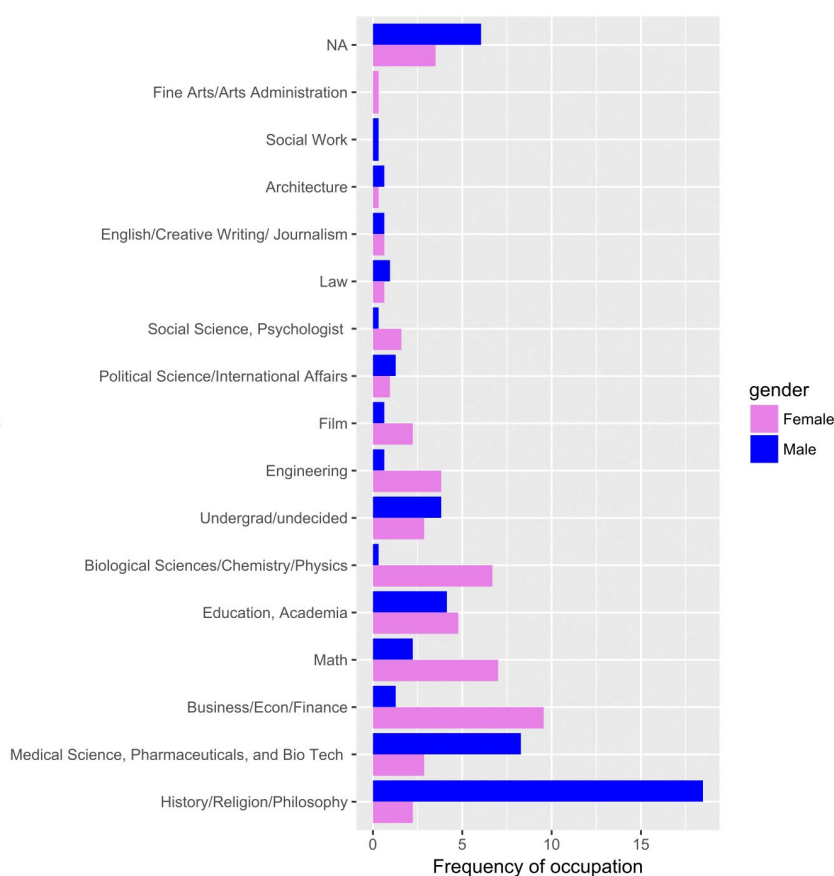
Nous essayons dans ce rapport de caractériser les individus les plus attractifs pour le sexe opposé à travers l'étude du jeu de données

“Speed dating experiment” [<https://www.kaggle.com/annavictoria/speed-dating-experiment>]

Ce jeu de données très riche rend compte pour chacune de ses lignes d'une rencontre ayant eu lieu dans le cadre d'une vague de speed-dating. Cependant il a fallu effectuer un tri dans ces données pour éviter de produire des résultats contradictoires. Par exemple, les participants ont pu noter leurs attentes pour ces rencontres soit sur une échelle de 1 sur 10 pour 5 critères soit en distribuant 100 points sur chacun de ces critères. Le choix du type de notation s'est effectué suivant le numéro des vagues. Nous avons considéré que ces deux méthodes étaient incompatibles et décidé de ne garder que les données des vagues avec distribution de points. Nous avons aussi exclu les vagues dans lesquelles les participants étaient limités en nombre de match. Nous avons comptabilisé un nombre total de participants de 314.

Par la suite, nous essayons de catégoriser les participants suivant leurs motivations et leur condition sociale.

### 1) Quels sont les individus qui composent le jeu de donnée

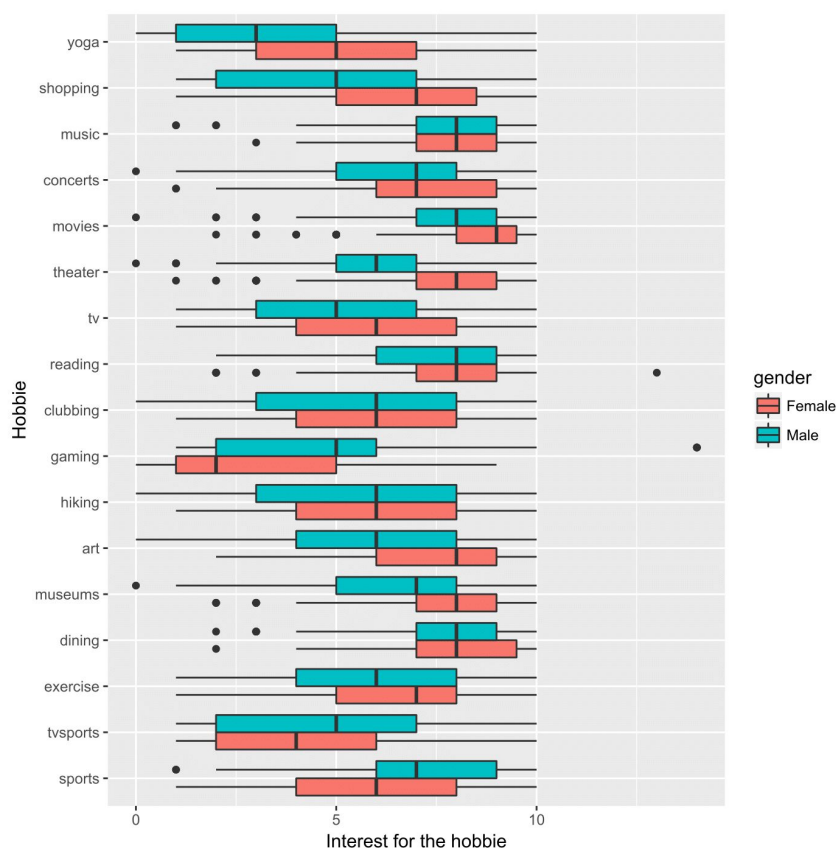
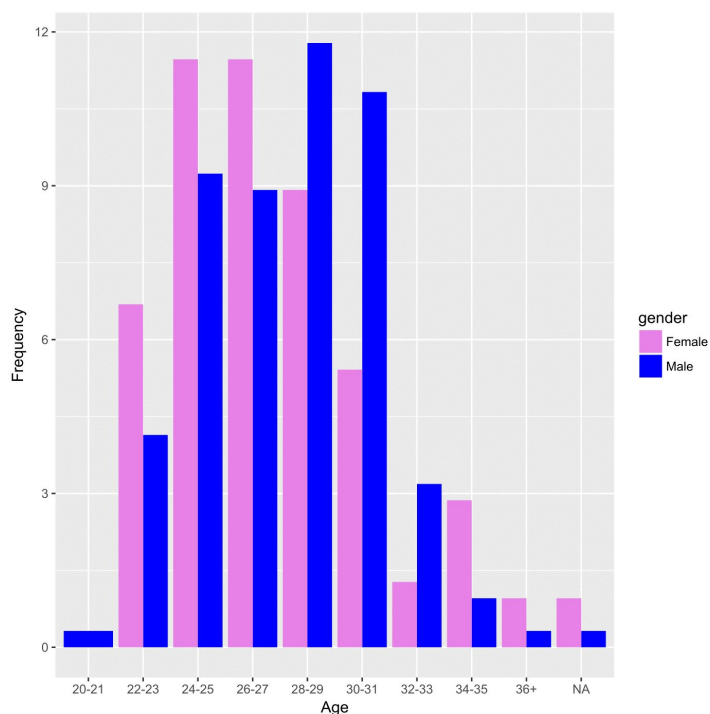


Les individus participant à ces séances de speed-dating sont des étudiants d'occupations diverses. Cependant, l'étude du graphe suivant révèle qu'une grande majorité des hommes présent étudient l'Histoire, la Religion, ou la Philosophie; du côté des femmes, ce sont plutôt les sciences dures qui sont représentées, avec une grande partie étudiant la Finance, les Maths, ou la Chimie.

Les participants du speed dating sont pour la plupart jeune, la moyenne d'âge se situant autour de 30 ans. Les femmes sont également en moyenne plus jeune que les hommes. La participation tend à s'abaisser rapidement au-dessus des 31 ans voir 30 ans pour les femmes.

On peut donc supposer que cette population jeune sera plus frivole et plus encline au match. Elle ne cherche pas à se poser en couple.

Dans le cas des intérêts des participants, on peut observer les différents boîtes à moustache des stéréotypes qui se vérifient. Ainsi, les hommes tendent beaucoup plus à aimer les jeux vidéos, le sport et les rediffusions télévisées que les femmes. Inversement ces dernières restent plus attirées que les hommes pour le shopping et le yoga. Ces écarts sont néanmoins relatifs, étant donné que les moyennes du yoga pour les femmes et des jeux vidéos, matchs sportifs pour les hommes sont relativement basses (5) par rapport à d'autres activités.



Cependant les deux sexes peuvent s'appuyer sur les valeurs sûres telles que la musique, les concerts, les dîners, les films. Ce sont des points qui pourront sûrement être abordés aux premiers échanges et être accueillis avec plaisir. Les écarts-types y sont faibles et les moyennes élevées.

Malgré un intérêt des hommes comme des femmes pour les arts, les femmes présentent bien plus d'intérêt pour les musées, le théâtre et les arts en général.

Sans statistiques représentant la population de manière générale, il est difficile de dire si l'on peut dresser un portrait d'une personne pratiquant le speed-dating.

## II)A la recherche du Don Juan moderne

Qu'est-ce qui caractérise un Don Juan? Est-ce la fierté et l'orgueil? Le mépris des règles sociales, morales et religieuses? Ou bien le pourcentage de matchs (accord réciproque de revoyure) qu'il ou elle parvient à obtenir lors d'une série de rencontres supervisées?

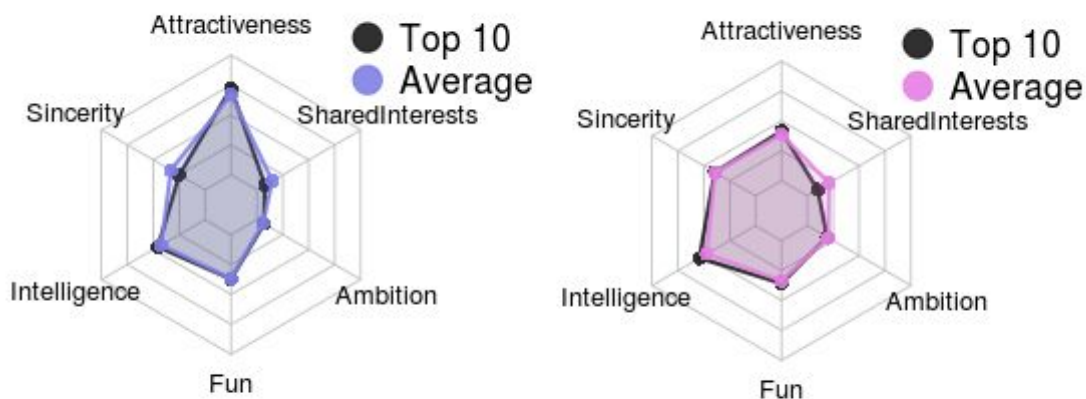
Cette dernière approche nous convient bien puisqu'elle nous permet d'utiliser notre jeu de données pour essayer d'en dégager des tendances.

Pour commencer, notons que le Don Juan moderne est une femme. Mais notons surtout qu'il ne semble pas y avoir à première vue de critère évident permettant de déterminer quels individus ont un taux de match élevé, si ce n'est que les hommes blancs (race == 2) semblent être bien positionnés.

	iid	gender	race	age	match_nb	match_ratio	liked_ratio
1	524	0	4	25	14	0.6363636	0.6363636
2	107	1	2	21	11	0.6111111	0.6666667
3	489	1	2	28	8	0.5333333	0.7333333
4	366	1	2	27	10	0.5000000	0.9000000
5	404	1	2	23	9	0.5000000	0.6111111
6	448	1	2	22	5	0.5000000	0.6000000
7	268	0	3	24	10	0.4761905	0.9523810
8	467	0	2	24	7	0.4666667	0.6666667
9	492	1	6	30	7	0.4666667	0.6666667
10	86	0	2	22	8	0.4444444	0.7777778

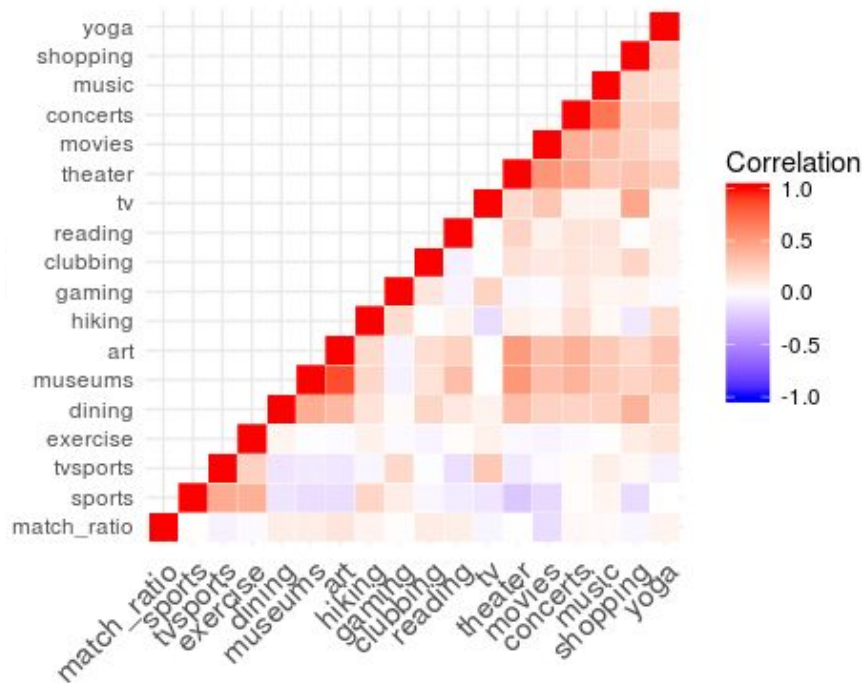
Nous essayons de rentrer dans la tête de nos Don Juan. Se pourrait-il qu'ils aient une idée différente du genre de personne qu'ils cherchent à rencontrer? Les graphes ci-contre représentent l'idée que les participants avaient du type de personne qu'ils recherchaient avant de commencer le speed-dating. Pour changer, les hommes sont représentés en bleu, les femmes en rose, et les 10 meilleurs "matcheurs" de chaque sexe en noir.

On remarque sur ce graphe que s'il n'y a pas de différence majeure entre les Don Juans et la moyenne, les Don Juans femmes et hommes s'en écartent tout deux vers plus d'intelligence et de beauté, mais moins d'intérêts en commun.



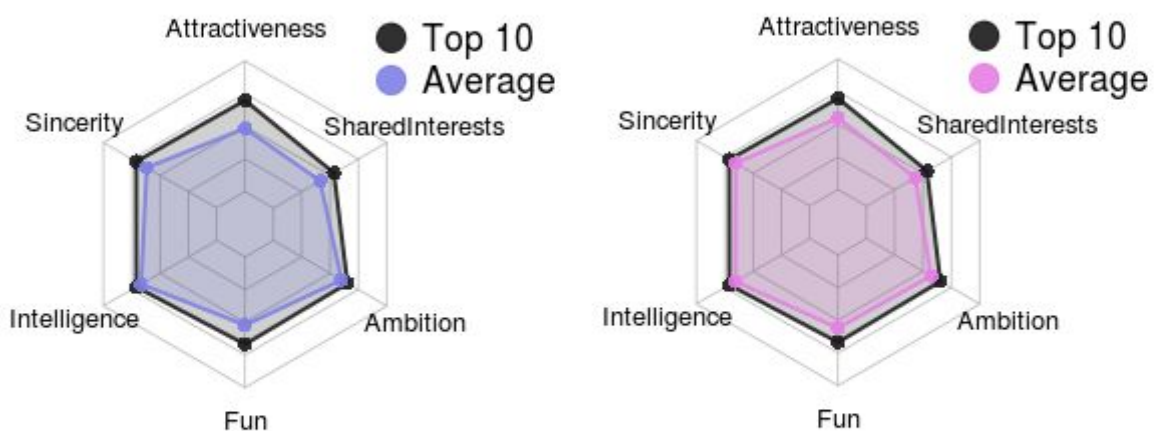
Voyons ensuite si les Don Juans se caractérisent par leurs centre d'intérêts.

La carte de chaleur qui suit représente la corrélation entre les centre d'intérêts des participants et leurs ratio de match. Elle permet aussi d'observer les corrélations entre activités. On observe de très faibles corrélations de match\_ratio avec les différentes activités. Ce qui indique que les loisirs influent peu sur l'habileté à être un Don Juan.



Une autre piste d'exploration est l'idée que la confiance en soi transcende un individu lambda en bourreau des coeurs. On établit donc un score de confiance en soi correspondant à la somme des points qu'il attribue à chacune de ses qualités. Mais encore une fois, la corrélation avec le ratio de match est trop faible pour être significative.

Enfin s'ils ne diffèrent pas par ces caractéristiques là, sont il perçu différemment par leurs partenaires? En effet, les graphes ci-dessous, qui correspondent à l'évaluation par le sexe opposé de caractéristiques du participant après leur rencontre. On remarque que nos Don Juans sont considérés comme plus beaux et sympas que la moyenne.



### Conclusion

Nous avons pu, au travers de ce rapport, caractériser les individus faisant partie de ce jeu de données, et tenté d'isoler ce qui fait de l'un d'eux un Don Juan.