

Comment mesurer l'efficacité collective de l'équipe de France de rugby à XV ?



Le conférencier

La conférence est présentée par Avner BAR-HEN, professeur à l'Université Paris Descartes – UFR de mathématiques et informatique. Les sujets de prédilection d'Avner BAR-HEN sont les statistiques spatiales en écologie et l'analyse de données génomiques.

Avner BAR-HEN est très actif dans le domaine de la vulgarisation scientifique, notamment par ses articles sur le site « *Images des mathématiques* » du CNRS et par son blog pour le magazine « *Sciences et Avenir* ». Les professeurs du secondaire peuvent y trouver matière à illustrer leurs cours.

<http://images.math.cnrs.fr/Bar-Hen-Avner.html>

<http://statistique.blogs.sciencesetavenir.fr/>

Un compte rendu de la conférence d'après les notes de Mohammed Mesmoudi, professeur au collège Jacques-Yves Cousteau de Bussy-Saint-Georges

Le sport génère énormément d'argent (paris, publicités, diffusion dans les médias...). Des scientifiques de différents domaines travaillent maintenant pour le sport. Aux États-Unis on parle actuellement *Sport Scientists*. Le film « Le stratège », avec Brad Pitt, inspiré d'une histoire vraie, raconte ainsi comment un entraîneur, n'ayant pas les meilleurs joueurs, fait appel à la statistique pour élaborer une stratégie pour gagner.

Avec le développement technologique, le sport génère énormément de données numériques. On trouve par exemple des GPS sur des balles de tennis ou des ballons de rugby qui donnent en temps réel la vitesse de la balle, la force avec laquelle a été frappée, sa trajectoire... Les montres GPS renseignent sur le rythme cardiaque, la distance parcourue, le temps, la vitesse... Les personnes qui travaillent dans le domaine du sport cherchent maintenant de plus en plus à exploiter ces données pour en tirer un maximum d'informations.

Une section *Statistics in Sports* a ainsi été fondée dès 1992 au sein de l'*American Statistical Journal* : <http://community.amstat.org/sis/home>

L'INSEP est créé en 1975 sous le nom d'Institut national du sport et de l'éducation physique, il change de nom en 2009 pour devenir Institut national du sport, de l'expertise et de la performance. Sa mission est d'assurer aux sportifs de haut niveau toutes les conditions nécessaires à leur épanouissement. Il est situé près du bois de Vincennes (Paris 12^{ème}) et s'étale sur plusieurs hectares.

Le conférencier présente des chercheurs avec qui il collabore. Ils sont membres de l'IRMES (institut de recherche biomédicale et d'épidémiologie du sport) à l'INSEP et d'Agro Paris Tech – INRIA.

Après une revue historique du rugby, les différentes sources de données sont listées. Les données proviennent des progrès technologiques. Chaque joueur peut être équipé d'un GPS qui donne en temps réel sa position sur le terrain, sa vitesse, ses temps d'arrêts, les regroupements des joueurs. Le ballon du rugby est aussi équipé d'un GPS qui renseigne aussi sur sa vitesse, sa position sur le terrain. Ces données peuvent renseigner en temps réels sur la distribution des joueurs par rapport à la position du ballon. Elles peuvent donc servir pour faire des simulations, créer des jeux informatiques, et étudier le jeu de l'équipe adverse lors des préparations.

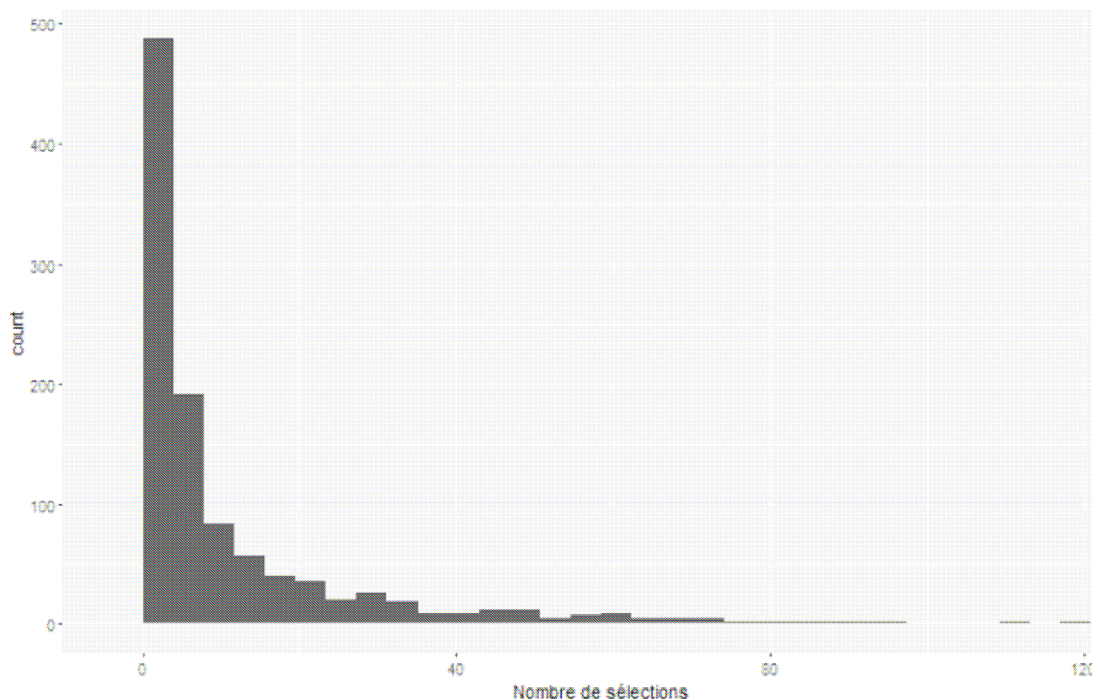
On a aussi des statistiques sur chaque joueur et sur chaque équipe.

Ces statistiques peuvent être trouvées sur le site de la fédération française du rugby (FFR).

Que peut-on ou doit-on faire avec ces données ?

On se pose, par exemple, la question de l'influence de l'expérience collective sur la performance d'un équipe. Pour ce faire, on étudie les données de plusieurs années.

Un histogramme donne le nombre des joueurs en fonction de leur nombre de sélections.



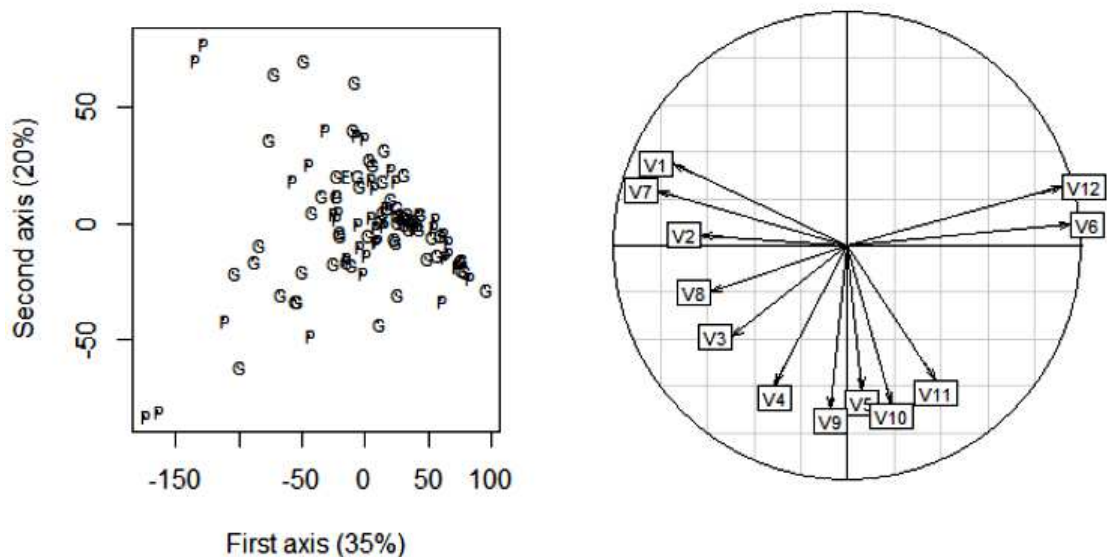
Les données peuvent servir à représenter chaque joueur comme un point dans un espace multidimensionnel. L'analyse en composantes principales (ACP) est alors une méthode permettant d'explorer des données en réduisant leur dimensionnalité, pour ne laisser que

l'essentiel. On projette dans un espace de plus petite dimension. On perd ainsi de l'informations. En faisant l'analyse on doit tenir compte de ce que l'on a perdu.

Pour pouvoir visualiser les points, on crée un premier axe de « composante principale » en minimisant ses distances avec les points. Cet axe suit la direction la plus allongée du nuage de points. Puis un deuxième axe de composante principale perpendiculaire au premier, il y a moins de variabilité. Puis un troisième, de plus petite variabilité pour former un repère orthogonal dans un espace tridimensionnel. Pour illustrer cette représentation, le conférencier donne l'exemple des dimensions d'un poisson : le premier axe est dans sa longueur, le deuxième dans sa largeur et le troisième dans son épaisseur.

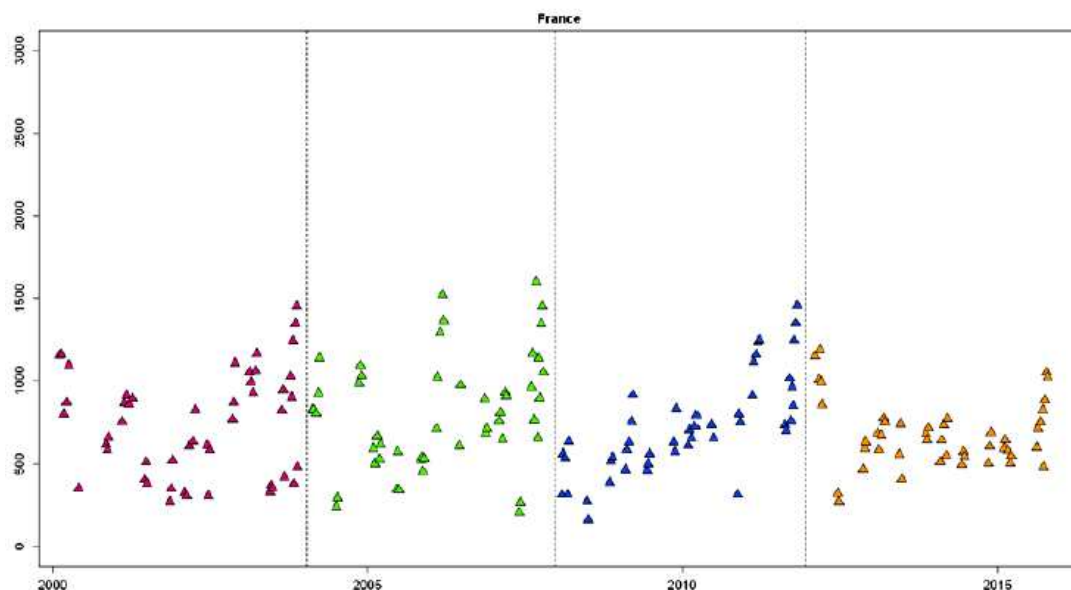
Une autre représentation possible consiste à effectuer des régressions linéaires. Deux exemples sont fournis : représentation des résultats vs les expériences individuelles des joueurs représentation des résultats vs les expériences collectives des joueurs de l'équipe (les résultats négatifs représentent les scores de pertes). Dans les deux cas, les points sont trop dispersés pour conclure.

Un exemple d'ACP d'expérience des joueurs par paires :

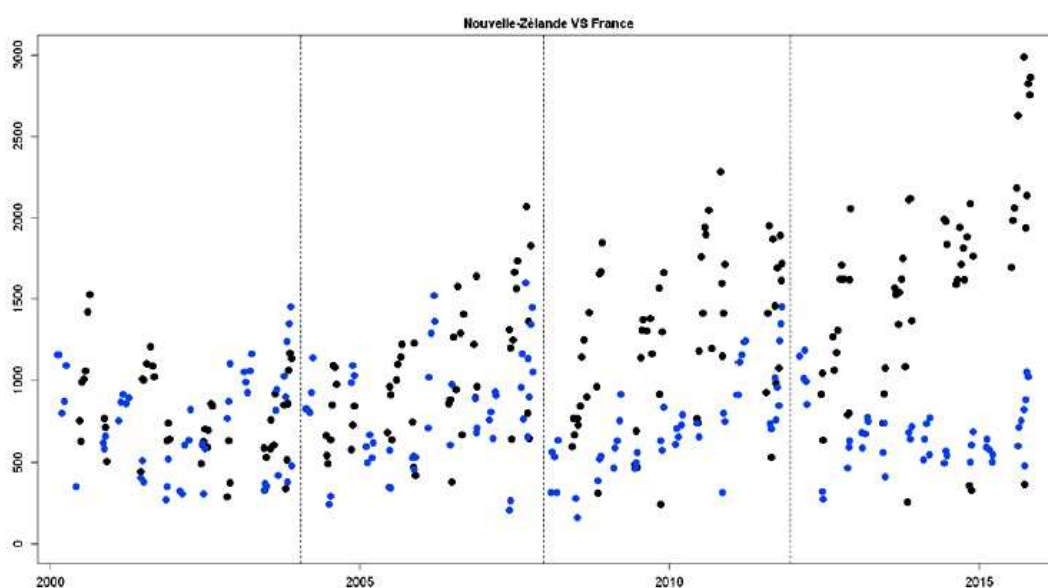


On voit que le nombre de joueurs avec plus de cinq sélections et le nombre de paires de joueurs ayant plus de cinq sélections en commun sont les deux variables qui se détachent et s'opposent aux autres variables. De plus, les matchs avec ces deux variables ont tendance à être plutôt gagnés.

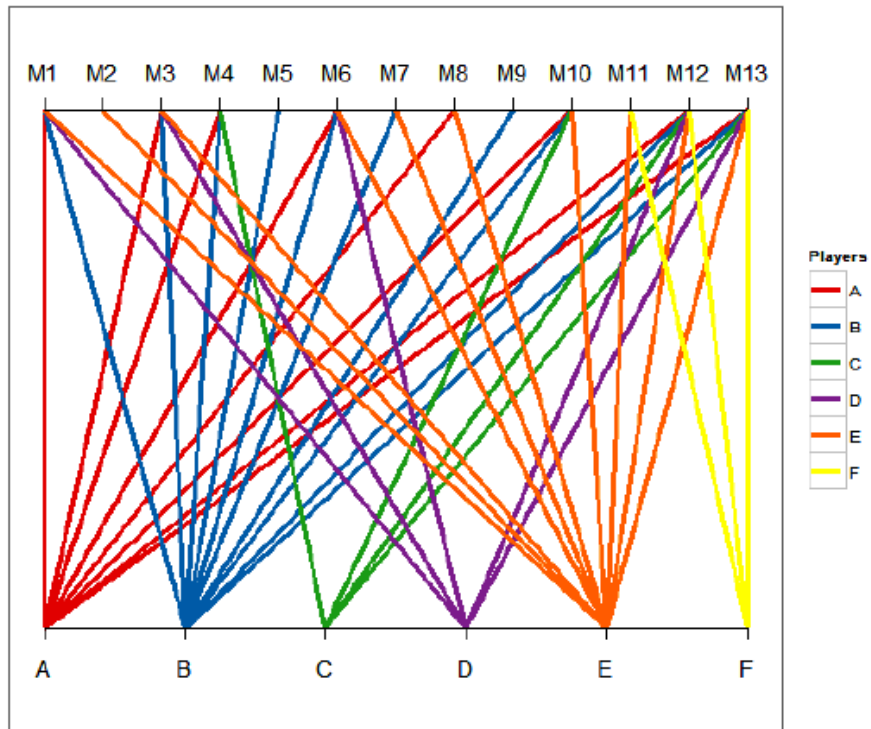
Le graphique suivant représente le nombre de sélections en fonction du temps pour l'équipe nationale (chaque point représente un match). On voit un collectif plus ou moins important se former à la fin de chaque période.



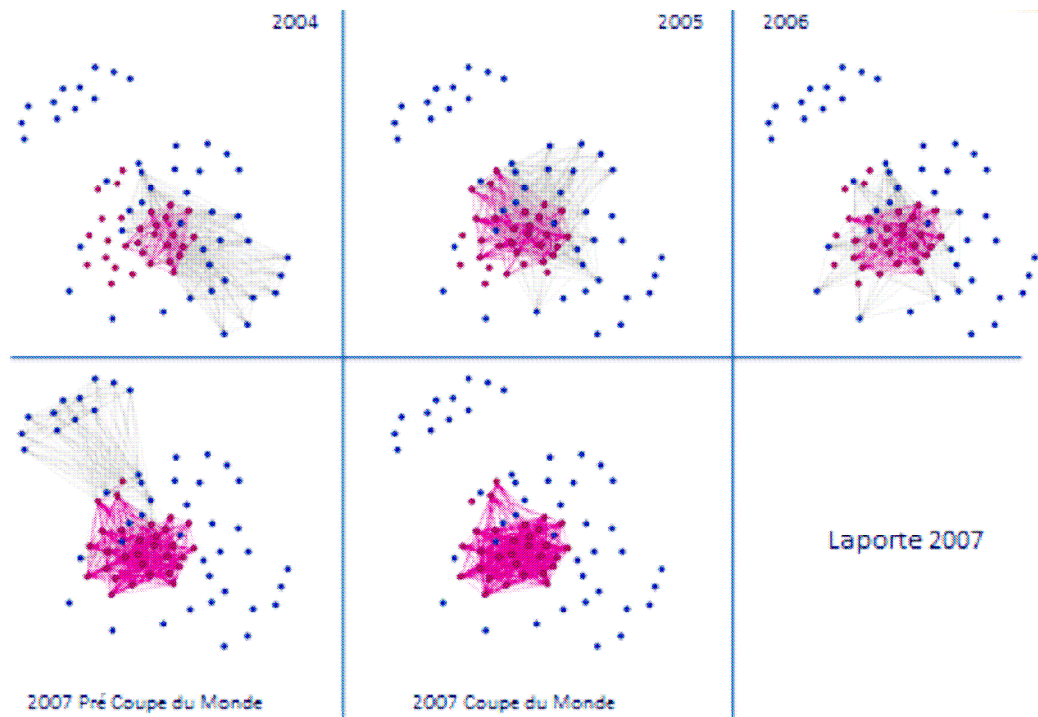
Le nombre de sélections communes est un facteur important pour le collectif. En 2015, l'écart est net : on voit le collectif se former plus tôt pour l'équipe de Nouvelle Zélande (en noir) que pour l'équipe de France (en bleu).



Une autre représentation graphique est possible sous forme de graphe : les matchs en haut et les joueurs en bas. Une arête relie chaque joueur aux matchs qu'il a joués. Cette représentation devient cependant vite illisible.



On peut recourir à un graphe de type « réseau ». Deux joueurs sont connectés s'ils ont déjà joué ensemble.



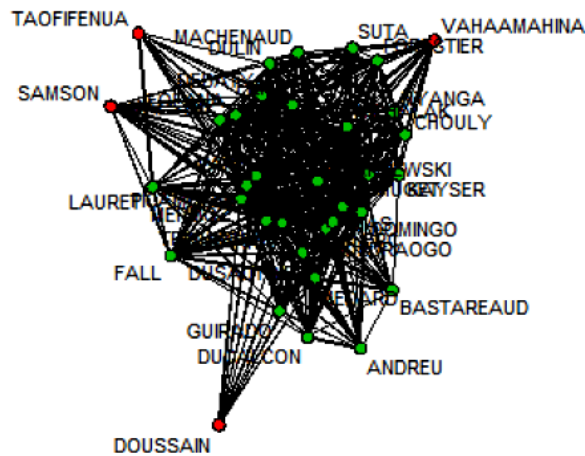
L'analyse d'un tel graphe, peut notamment se faire par sa densité et sa transversalité. La densité est le rapport du nombre d'arêtes sur le nombre d'arêtes possible. Et la transversalité se définit par : « l'ami de mon ami est un ami ». C'est-à-dire que trois sommets définissent un triangle (avec 3 côtés matérialisés) dans le graphe. Le pourcentage de la transversalité est défini par le rapport des cas réels aux cas possibles.

Le degré d'un sommet est le nombre d'arêtes issues de ce sommet.

La centralité est une notion qui mesure dans un graphe l'importance des sommets.

L'exemple du graphe de l'équipe de France 2013 est plus particulièrement examiné.

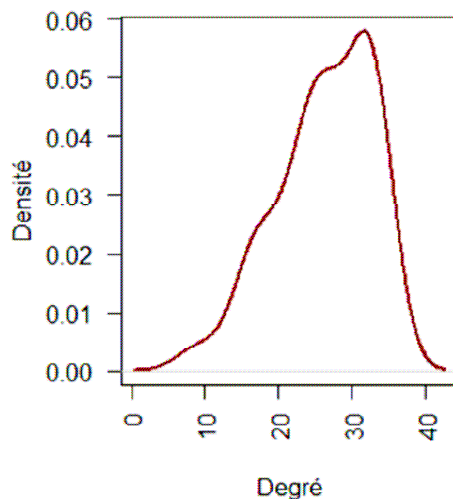
Graphe 2013



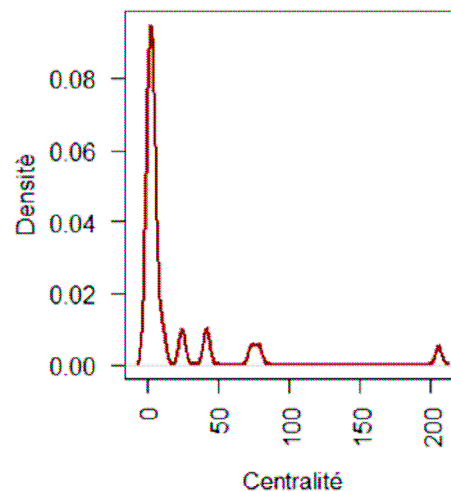
Densité = 0.7731092

Transitivité = 0.8492629

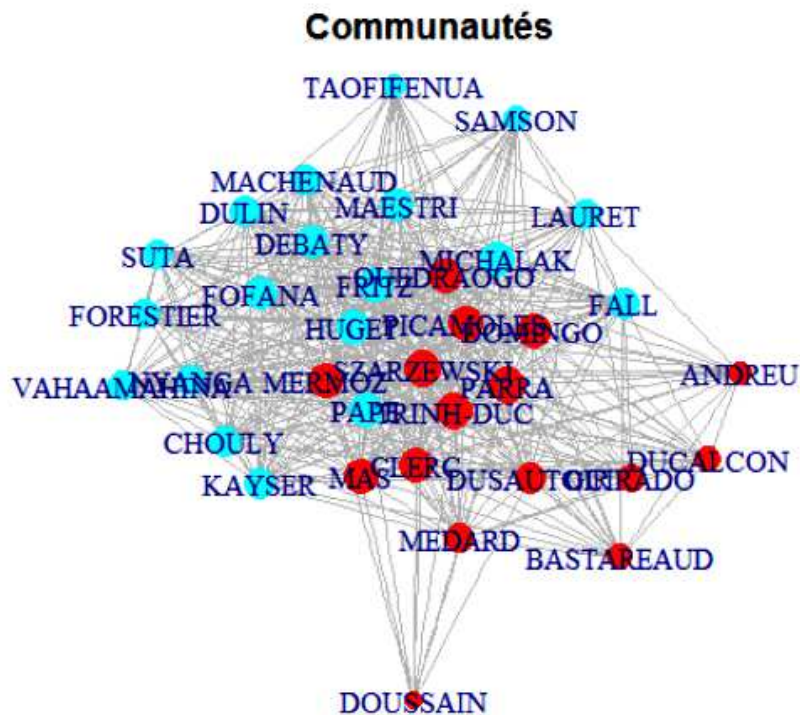
Distribution des degrés



distribution de la centralité



On peut effectuer une partition d'un graphe en « communautés » en regroupant les parties les plus denses et qui ont un faible lien les unes avec les autres. Voici la séparation du graphe précédent en deux communautés. Le degré moyen de la communauté rouge est de 32 et celui de la communauté bleue est de 26.



Et pour finir... du football !

On examine l'évolution du nombre de sélections communes partagées (SCP) en fonction du temps en ligue 1. On constate que plus le SCP est élevé moins il y a des blessures. Cela signifie que les joueurs les plus expérimentés et qui ont joué ensemble un nombre important de fois ont moins d'accidents pendant le jeu.

Conclusion :

Avec les progrès actuels, les mathématiques commencent à intéresser les médias : le prix Abel fait par exemple la une des journaux actuellement.

On parle de Big Data. À l'intérieur il y a énormément de données mais le problème est que faire de toutes ces données. C'est un chantier énorme qui s'ouvre pour chercher comment utiliser toutes ces données dont le nombre augmente de plus en plus.

Avec le développement de l'informatique, les graphiques sont de plus en plus beaux mais ne répondent pas toujours à la demande d'information. C'est un point qui reste à développer.

L'homme a créé la machine, la question est comment répondre au problème d'améliorer l'homme par la machine. La statistique est aujourd'hui ce que l'invention de la perspective a représenté au 15^{ème} siècle.