

Big Data

Joseph Saint Pierre

Université de Toulouse

rencontre Capitoul 18 février 2015

Plan

Introduction.

Le Big Data, une mode ?

Un peu d'histoire.

Le calcul et les outils techniques.

La place des statistiques et des probabilités.

L'utilisation en sciences sociales.

Conclusion.

Introduction

Emplois...

Enseignement.

Recherche.

Exposés, journées, séminaires.

Articles, émissions, télévision, radio.

Mode

Quelques critiques.

Grandes sociétés américaines : Google, Amazon, Facebook.

Méthodes magiques.

Détection des maladies. Recommandations infaillibles.

Trop fort...

Big Data, Big Brother.

Histoire

Big Data et Data Mining

Analyse de données

Traitement de données.

Causalité et corrélation. ACP

Histoire du calcul, des statistiques, des probabilités.

Calcul et outils

Abaques, cailloux, bouliers.

Chiffres romains, chiffres indiens

Papier, imprimerie.

Machines pour calculer.

Données en astronomie.

Lunettes, télescopes.

Statistiques

Histoire de l'astronomie.

Galile

Huygens

Gauss

Quetelet

Recensements, élections.

Hollerith. IBM.

Nate Silver.

Confusion

Citation

There are a lot of small data problems that occur in big data, they don't disappear because you've got lots of the stuff. They get worse.

Confusion de variables.

Un tableau université Toulouse 3 Égalité Femme-Homme
Nombres de femmes et d'hommes parmi les professeurs et maîtres de conférence suivant les disciplines.

Trois variables, sexe, niveau, discipline.

Seule deux disciplines, physique et médecine sont retenues.

Voici le tableau

Niveau	Disicipline	Homme	Femme
Professeur	Physique	28	2
Professeur	Medecine	150	20
MCF	Physique	29	6
MCF	Medecine	37	53

La physique est la discipline avec la plus faible proportion de femmes pour les professeurs et les maîtres de conférence dans l'université Toulouse 3. Dans le document sur l'égalité il y a tableau pour les professeurs et un tableau pour les maîtres de conférence.

La juxtaposition des deux tableaux permet de voir un autre aspect des données, l'écart important entre la proportion de femmes professeurs de médecine et la proportion de femmes maîtres de conférence, dans cette discipline.

Dans cette étude il y a d'autres variables.

Sciences sociales

Internet.

Forums.

Blogs.

Réseaux sociaux. Facebook, tweeter

Objets connectés. Sigfox.

Données ouvertes.

Journalisme de données.

Modifications des enquêtes.

Changements sociaux ?

Conclusion

Pour un assez long moment il y a aura un nombre croissant de données.

Il y aura besoin de moyens techniques et scientifiques pour les traiter.

Ce ne sera pas la panacée.

Il est toujours pertinent de regarder le passé pour appréhender le futur. C'est essentiel en statistiques.

Merci.