CHAIRE DÉCISIONNEL
CONNAISSANCE CLIENT

Fondation UBS Université Bretagne Sud

CHAIRE D-CC "DÉCISIONNEL-CONNAISSANCE CLIENT" - IUT de Vannes - 8 rue Montaigne - 56017 Vannes Cx / 02 97 62 64 64 / chaire.d-cc@univ-ubs.fr

Faire des données une opportunité de croissance : la data-science et la relation-client

M. Gousseff-Fondation UBS

15 juin, 2017

Data-science et data-scientists
"La" data-science ?
Entreprise orientée données
Entreprise orientée données (2)
Entreprise orientée client
Une démarche itérative : le cadre CRISP-DM
La data-science comme carrefour de compétences
Le data-scientist full-stack, un mythe ?
Actions de la chaire D-CC pour la Data-Science en entreprise
La répartition des rôles dans un projet data
Par où commencer ?

Data-science et data-scientists

"La" data-science ?

Data-science et data-scientists
"La" data-science ?
Entreprise orientée données
Entreprise orientée données (2)
Entreprise orientée client
Une démarche itérative : le cadre CRISP-DM
La data-science comme carrefour de compétences
Le data-scientist full-stack, un mythe ?
Actions de la chaire D-CC pour la Data-Science en entreprise
La répartition des rôles dans un projet data
Par où commencer ?

Des sciences nées lorsque les données étaient rares... et propres !

Statistiques :
Inférer une population à partir d'un échantillon

- ✔ Proposer un modèle
- ✔ Poser des hypothèses
- ✔ produire les données
- ✔ Prédire / expliquer

Informatique :
Inventer le lien hardware/software

- ✔ Faire exécuter des algorithmes logiques à une machine
- ✔ Stocker et accéder à des données
- ✔ Interfacer l'homme et la machine

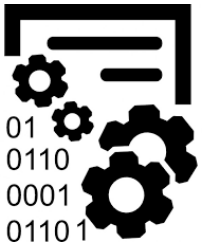
Entreprise orientée données

Data-science et data-scientists
"La" data-science ?
Entreprise orientée données
Entreprise orientée données (2)
Entreprise orientée client
Une démarche itérative : le cadre CRISP-DM
La data-science comme carrefour de compétences
Le data-scientist full-stack, un mythe ?
Actions de la chaire D-CC pour la Data-Science en entreprise
La répartition des rôles dans un projet data
Par où commencer ?

- ✔ De la donnée, plein de données !
  - ✘ Explosion des capacités de stockage
  - ✘ Sources de données multiples
    - ✓ Rdv, téléphone, internet
    - ✓ Smartphone, tablettes, IoT
  - ✘ Les données forcent l'évolution des sciences
    - ✓ Évolution des capacités de calcul
    - ✓ Statistique → data-mining, statistical learning
    - ✓ Informatique → Big Data

Entreprise orientée données (2)

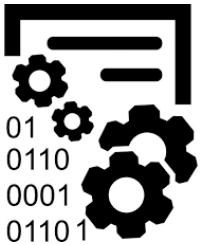
Data-science et data-scientists
"La" data-science ?
Entreprise orientée données
Entreprise orientée données (2)
Entreprise orientée client
Une démarche itérative : le cadre CRISP-DM
La data-science comme carrefour de compétences
Le data-scientist full-stack, un mythe ?
Actions de la chaire D-CC pour la Data-Science en entreprise
La répartition des rôles dans un projet data
Par où commencer ?



- ✔ De la convergence à l'émergence
  - ✘ Convergence de l'informatique, de l'algorithmique, des statistiques, de l'apprentissage
  - ✘ Émergence ou renouveau : deep Learning, intelligence artificielle...
- ✔ Appropriation par les entreprises :
  - ✘ Nées dans la données : GAFA...
  - ✘ Transformation numérique des entreprises traditionnelles

Entreprise orientée données (2)

Data-science et data-scientists
"La" data-science ?
Entreprise orientée données
Entreprise orientée données (2)
Entreprise orientée client
Une démarche itérative : le cadre CRISP-DM
La data-science comme carrefour de compétences
Le data-scientist full-stack, un mythe ?
Actions de la chaire D-CC pour la Data-Science en entreprise
La répartition des rôles dans un projet data
Par où commencer ?



- ✔ De la convergence à l'émergence
  - ✘ Convergence de l'informatique, de l'algorithmique, des statistiques, de l'apprentissage
  - ✘ Émergence ou renouveau : deep Learning, intelligence artificielle...
- ✔ Appropriation par les entreprises :
  - ✘ Nées dans la données : GAFA...
  - ✘ Transformation numérique des entreprises traditionnelles

→ Et le client, dans tout ça ?

Entreprise orientée client

Data-science et data-scientists
"La" data-science ?
Entreprise orientée données
Entreprise orientée données (2)
Entreprise orientée client
Une démarche itérative : le cadre CRISP-DM
La data-science comme carrefour de compétences
Le data-scientist full-stack, un mythe ?
Actions de la chaire D-CC pour la Data-Science en entreprise
La répartition des rôles dans un projet data
Par où commencer ?

- ✔ Du marketing transactionnel
  - ✘ Centré sur le produit
  - ✘ Échelle de temps : l'acte d'achat
- ✔ Au marketing relationnel
  - ✘ Centré sur le client, son expérience d'achat
  - ✘ Échelle de temps : le cycle de vie du client

Entreprise orientée client

Data-science et data-scientists
"La" data-science ?
Entreprise orientée données
Entreprise orientée données (2)
Entreprise orientée client
Une démarche itérative : le cadre CRISP-DM
La data-science comme carrefour de compétences
Le data-scientist full-stack, un mythe ?
Actions de la chaire D-CC pour la Data-Science en entreprise
La répartition des rôles dans un projet data
Par où commencer ?

- ✔ Du marketing transactionnel
  - ✘ Centré sur le produit
  - ✘ Échelle de temps : l'acte d'achat
- ✔ Au marketing relationnel
  - ✘ Centré sur le client, son expérience d'achat
  - ✘ Échelle de temps : le cycle de vie du client

→ Rien de nouveau ?

Entreprise orientée client

Data-science et data-scientists
"La" data-science ?
Entreprise orientée données
Entreprise orientée données (2)
Entreprise orientée client
Une démarche itérative : le cadre CRISP-DM
La data-science comme carrefour de compétences
Le data-scientist full-stack, un mythe ?
Actions de la chaire D-CC pour la Data-Science en entreprise
La répartition des rôles dans un projet data
Par où commencer ?

- ✔ Du marketing transactionnel
  - ✘ Centré sur le produit
  - ✘ Échelle de temps : l'acte d'achat

- ✔ Au marketing relationnel
  - ✘ Centré sur le client, son expérience d'achat
  - ✘ Échelle de temps : le cycle de vie du client

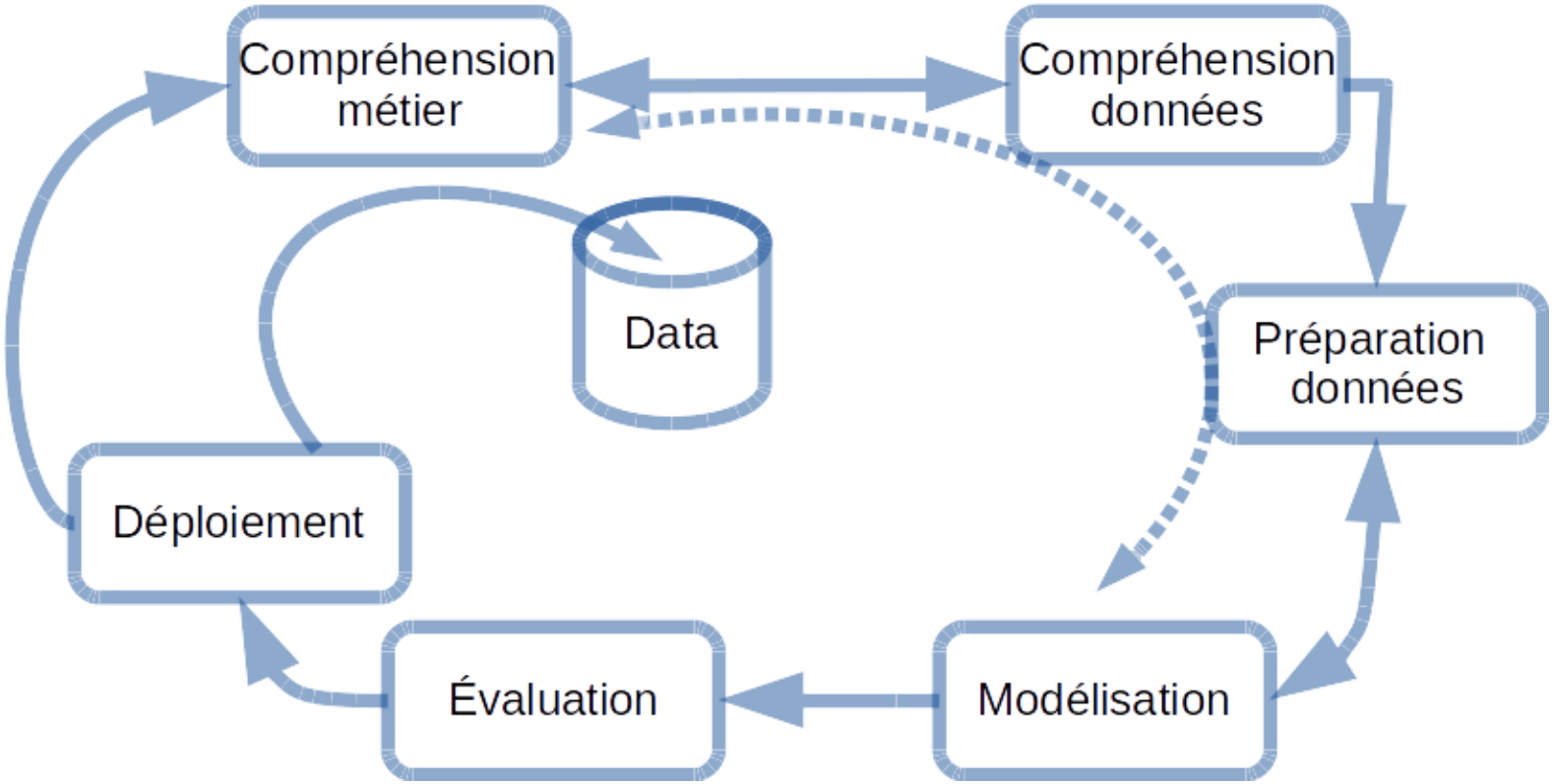
→ Rien de nouveau ?

Les données et leur exploitation permettent d'améliorer la relation client de manière rationnelle, objectivable, scalable !

Une démarche itérative : le cadre CRISP-DM

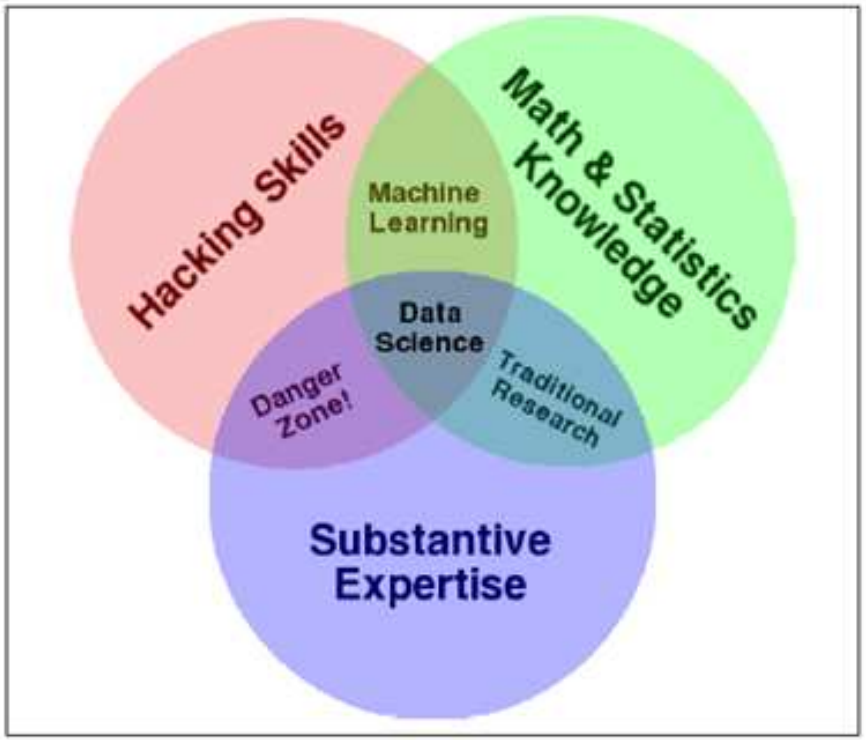
Data-science et data-scientists
"La" data-science ?
Entreprise orientée données
Entreprise orientée données (2)
Entreprise orientée client
Une démarche itérative : le cadre CRISP-DM
La data-science comme carrefour de compétences
Le data-scientist full-stack, un mythe ?
Actions de la chaire D-CC pour la Data-Science en entreprise
La répartition des rôles dans un projet data
Par où commencer ?

CRoss Industry Standard Process for Data-Mining : une démarche itérative et transversale



La data-science comme carrefour de compétences

Data-science et data-scientists
"La" data-science ?
Entreprise orientée données
Entreprise orientée données (2)
Entreprise orientée client
Une démarche itérative : le cadre CRISP-DM
La data-science comme carrefour de compétences
Le data-scientist full-stack, un mythe ?
Actions de la chaire D-CC pour la Data-Science en entreprise
La répartition des rôles dans un projet data
Par où commencer ?



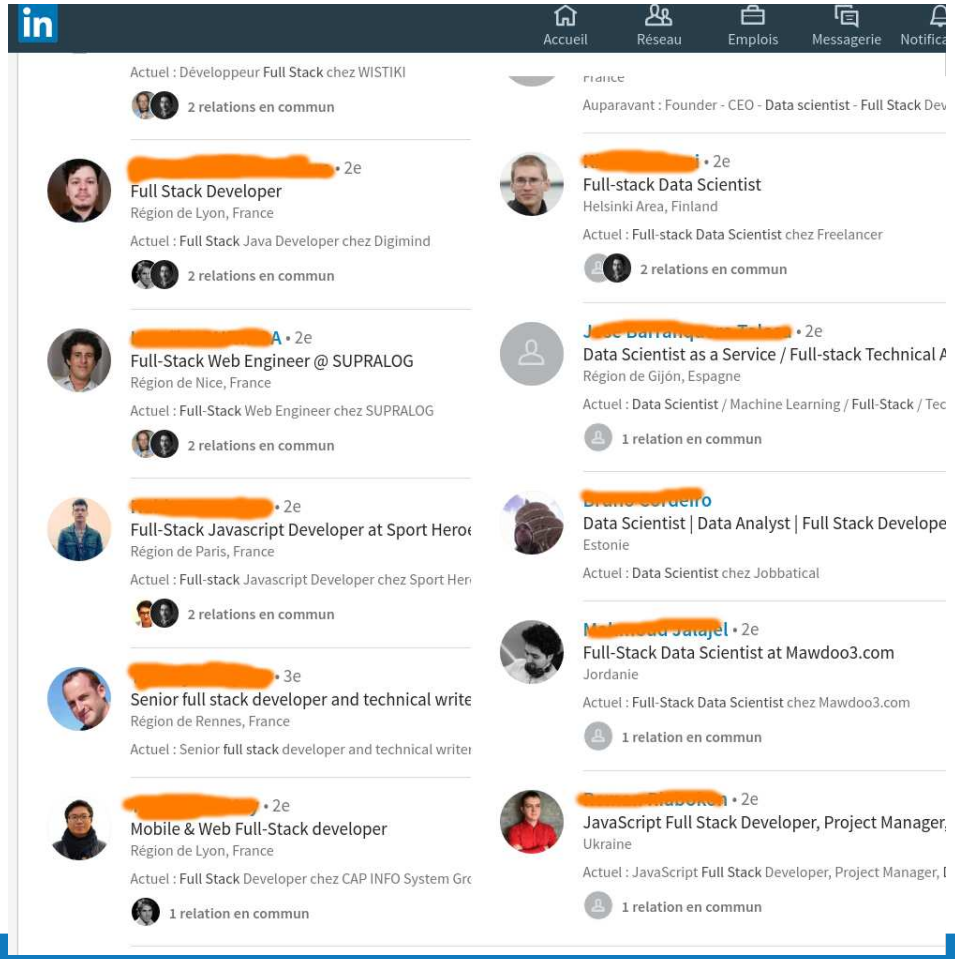
Data Scientist = Programmeur + Informaticien + Mathematicien + Story teller + Expert Métier

https://intellipaat.com/tutorial/data-science-tutorial/introduction-of-data-science

Le data-scientist full-stack, un mythe ?

Data-science et data-scientists
"La" data-science ?
Entreprise orientée données
Entreprise orientée données (2)
Entreprise orientée client
Une démarche itérative : le cadre CRISP-DM
La data-science comme carrefour de compétences
Le data-scientist full-stack, un mythe ?
Actions de la chaire D-CC pour la Data-Science en entreprise
La répartition des rôles dans un projet data
Par où commencer ?

Full-stack : capable de mener un projet de bout en bout.
Mais quel périmètre pour le projet ?



Actions de la chaire D-CC pour la Data-Science en entreprise

Data-science et data-scientists
"La" data-science ?
Entreprise orientée données
Entreprise orientée données (2)
Entreprise orientée client
Une démarche itérative : le cadre CRISP-DM
La data-science comme carrefour de compétences
Le data-scientist full-stack, un mythe ?
Actions de la chaire D-CC pour la Data-Science en entreprise
La répartition des rôles dans un projet data
Par où commencer ?

- ✔ Interaction Experts métiers / Data-scientists
- ✔ Une proposition de démarche générale d'analyse des bases de données clients
- ✔ Livrables : de la méthodologie au service open source

La répartition des rôles dans un projet data
Contre exemple classique
Interactions Métier-Datascience
Freins aux interactions
Tous sur le même bateau !
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source

La répartition des rôles dans un projet data

Contre exemple classique

La répartition des rôles dans un projet data

Contre exemple classique

Interactions Métier-Datascience

Freins aux interactions

Tous sur le même bateau !

Par où commencer ?

La chaire D-CC : de la méthodologie au service open source

Contre exemple classique

La répartition des rôles dans un projet data
Contre exemple classique
Interactions Métier-Datascience
Freins aux interactions
Tous sur le même bateau !
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source

"Expert métier : Tu sais monter un cluster Hadoop ?

Contre exemple classique

La répartition des rôles dans un projet data
Contre exemple classique
Interactions Métier-Datascience
Freins aux interactions
Tous sur le même bateau !
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source

"Expert métier : Tu sais monter un cluster Hadoop ?
Data-scientist : Éventuellement, Mais pour quoi faire ?

Contre exemple classique

La répartition des rôles dans un projet data
Contre exemple classique
Interactions Métier-Datascience
Freins aux interactions
Tous sur le même bateau !
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source

"Expert métier : Tu sais monter un cluster Hadoop ?
Data-scientist : Éventuellement, Mais pour quoi faire ?
- Pour faire du Map Reduce.

Contre exemple classique

La répartition des rôles dans un projet data
Contre exemple classique
Interactions Métier-Datascience
Freins aux interactions
Tous sur le même bateau !
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source

"Expert métier : Tu sais monter un cluster Hadoop ?
Data-scientist : Éventuellement, Mais pour quoi faire ?
- Pour faire du Map Reduce.
- Oui, bien-sûr, mais . . . dans quel but ?

Contre exemple classique

La répartition des rôles dans un projet data
Contre exemple classique
Interactions Métier-Datascience
Freins aux interactions
Tous sur le même bateau !
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source

"Expert métier : Tu sais monter un cluster Hadoop ?
Data-scientist : Éventuellement, Mais pour quoi faire ?
- Pour faire du Map Reduce.
- Oui, bien-sûr, mais . . . dans quel but ?
- Mais parce qu'on est à l'ère du Big Data !

Contre exemple classique

La répartition des rôles dans un projet data
Contre exemple classique
Interactions Métier-Datascience
Freins aux interactions
Tous sur le même bateau !
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source

"Expert métier : Tu sais monter un cluster Hadoop ?
Data-scientist : Éventuellement, Mais pour quoi faire ?
- Pour faire du Map Reduce.
- Oui, bien-sûr, mais . . . dans quel but ?
- Mais parce qu'on est à l'ère du Big Data !
- Mais pour quelles données ? On a de grandes tables, mais elles entrent dans la mémoire du serveur, alors...

Contre exemple classique

La répartition des rôles dans un projet data
Contre exemple classique
Interactions Métier-Datascience
Freins aux interactions
Tous sur le même bateau !
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source

"Expert métier : Tu sais monter un cluster Hadoop ?
Data-scientist : Éventuellement, Mais pour quoi faire ?
- Pour faire du Map Reduce.
- Oui, bien-sûr, mais . . . dans quel but ?
- Mais parce qu'on est à l'ère du Big Data !
- Mais pour quelles données ? On a de grandes tables, mais elles entrent dans la mémoire du serveur, alors...
- Pourquoi ne veux pas faire du Hadoop ? Tu ne sais pas faire du Hadoop ?"

Contre exemple classique

La répartition des rôles dans un projet data
Contre exemple classique
Interactions Métier-Datascience
Freins aux interactions
Tous sur le même bateau !
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source

"Expert métier : Tu sais monter un cluster Hadoop ?
Data-scientist : Éventuellement, Mais pour quoi faire ?
- Pour faire du Map Reduce.
- Oui, bien-sûr, mais . . . dans quel but ?
- Mais parce qu'on est à l'ère du Big Data !
- Mais pour quelles données ? On a de grandes tables, mais elles entrent dans la mémoire du serveur, alors...
- Pourquoi ne veux pas faire du Hadoop ? Tu ne sais pas faire du Hadoop ?"

Idem avec, réseaux de neurones, forêts aléatoires, réseaux bayésiens...

Interactions Métier-Datascience

La répartition des rôles dans un projet data
Contre exemple classique
Interactions Métier-Datascience
Freins aux interactions
Tous sur le même bateau !
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source

Expert métier

- ✔ Poser la bonne question
- ✔ Connaître les variables et indicateurs
- ✔ Assurer l'appropriation et le bon usage des modèles

Data Scientist

- ✔ De la question au modèle
- ✔ Raffiner le modèle : compromis
  précision appropriation
- ✔ Former et transmettre

Freins aux interactions

La répartition des rôles dans un projet data
Contre exemple classique
Interactions Métier-Datascience
Freins aux interactions
Tous sur le même bateau !
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source

Côté métier

- ✔ Je connais mieux mes clients qu'un algorithme
- ✔ Rien ne remplace l'expérience terrain
- ✔ On peut faire dire n'importe quoi aux stats
- ✔ . . .

Côté DS

- ✔ Mon modèle trouve les corrélations tout seul
- ✔ Les variables ne sont pas bonnes
- ✔ Voici le meilleur modèle
- ✔ . . .

Tous sur le même bateau !

La répartition des rôles dans un projet data
Contre exemple classique
Interactions Métier-Datascience
Freins aux interactions
Tous sur le même bateau !
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source

- ✔ Les navigateurs : les experts métiers
- ✔ Le moteur : les algorithmes
- ✔ Le carburant : les données :
  - ✘ Techniquement correctes
  - ✘ Statistiquement pertinentes

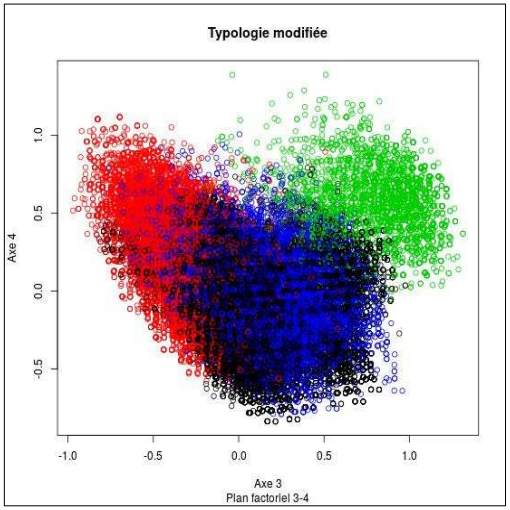
La répartition des rôles dans un projet data
Par où commencer ?
Une démarche générale possible
Un exemple concret : proposer un produit à un client
À dire d'expert : à vous !
Score statistique : démarche
Score statistique
De la bonne question...
... À l'appropriation du modèle
La chaire D-CC : de la méthodologie au service open source

Par où commencer ?

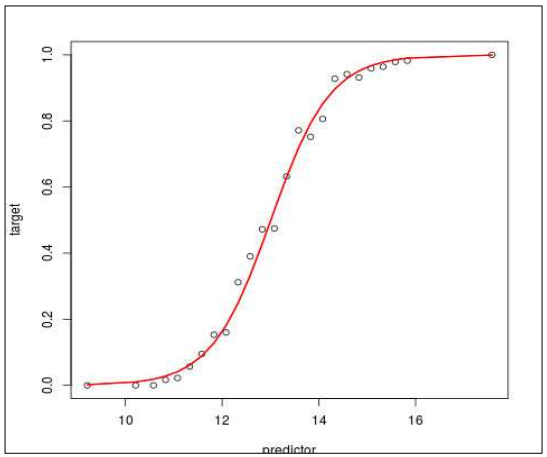
La répartition des rôles dans un projet data
Par où commencer ?
Une démarche générale possible
Un exemple concret : proposer un produit à un client
À dire d'expert : à vous !
Score statistique : démarche
Score statistique
De la bonne question...
... À l'appropriation du modèle
La chaire D-CC : de la méthodologie au service open source

✔ Segmentation → Ciblage/Scores → Trajectoires

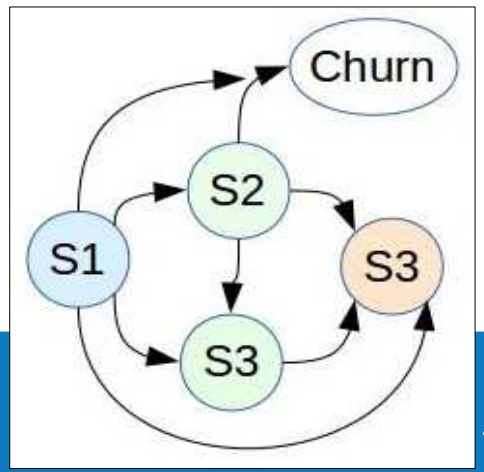
Explorer la diversité de ses clients = segmenter



Mesurer l'intérêt d'un client pour un produit ou service = scorer



Prendre en compte la dynamique client



Un exemple concret : proposer un produit à un client

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

- ✔ Pourquoi cibler des clients ?
  - ✘ Temps des commerciaux limité
  - ✘ Coût des envois
  - ✘ Pression marketing

- ✔ Comment cibler des clients ?
  - ✘ À dire d'experts
  - ✘ À l'aide d'un score statistique

Un exemple concret : proposer un produit à un client

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

- ✔ Pourquoi cibler des clients ?
  - ✘ Temps des commerciaux limité
  - ✘ Coût des envois
  - ✘ Pression marketing

- ✔ Comment cibler des clients ?
  - ✘ À dire d'experts
  - ✘ À l'aide d'un score statistique

⇒ Travailler avec les experts ou avec les data-scientists ?

À dire d'expert : à vous !

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

Différentes cibles

- ✔ Quels clients cibler pour du crédit à la consommation ?
- ✔ Quels clients cibler pour de l'assurance vie ?

À dire d'expert : à vous !

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

Différentes cibles

- ✔ Quels clients cibler pour du crédit à la consommation ?
- ✔ Quels clients cibler pour de l'assurance vie ?

De nombreuses questions

- ✔ Quelles variables pour chaque cible ?
- ✔ Quelles bornes pour chaque variable ?
- ✔ Quel ordre de priorité pour les clients retenus ?
- ✔ Quel taux de transformation attendre ?

Nombre de variables et de clients en pleine croissance !

Score statistique : démarche

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

Apprentissage

Score statistique : démarche

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

Apprentissage

- ✔ On propose le produit à un échantillon de clients/prospect

Score statistique : démarche

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

Apprentissage

- ✔ On propose le produit à un échantillon de clients/prospect
- ✔ Certains achètent d'autres non

Score statistique : démarche

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

Apprentissage

- ✔ On propose le produit à un échantillon de clients/prospect
- ✔ Certains achètent d'autres non
- ✔ On mesure d'autres variables = prédictrices

Score statistique : démarche

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

Apprentissage

- ✔ On propose le produit à un échantillon de clients/prospect
- ✔ Certains achètent d'autres non
- ✔ On mesure d'autres variables = prédictrices

Score statistique : démarche

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

Apprentissage

- ✔ On propose le produit à un échantillon de clients/prospect
- ✔ Certains achètent d'autres non
- ✔ On mesure d'autres variables = prédictrices
  - ✘ Variables socio professionnelles, comportement antérieur, autres secteurs d'achat, etc.

Score statistique : démarche

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

Apprentissage

- ✔ On propose le produit à un échantillon de clients/prospect
- ✔ Certains achètent d'autres non
- ✔ On mesure d'autres variables = prédictrices
  - ✘ Variables socio professionnelles, comportement antérieur, autres secteurs d'achat, etc.
- ✔ On construit une équation de prédiction

$$P(achat) = Fonction(predictrices)$$

Score statistique : démarche

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

Apprentissage

- ✔ On propose le produit à un échantillon de clients/prospect
- ✔ Certains achètent d'autres non
- ✔ On mesure d'autres variables = prédictrices
  - ✘ Variables socio professionnelles, comportement antérieur, autres secteurs d'achat, etc.
- ✔ On construit une équation de prédiction

$$P(achat) = Fonction(predictrices)$$

Déploiement sur la base client

- ✔ On calcule le score de chaque client

Score statistique : démarche

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

Apprentissage

- ✔ On propose le produit à un échantillon de clients/prospect
- ✔ Certains achètent d'autres non
- ✔ On mesure d'autres variables = prédictrices
  - ✘ Variables socio professionnelles, comportement antérieur, autres secteurs d'achat, etc.
- ✔ On construit une équation de prédiction

$$P(achat) = Fonction(predictrices)$$

Déploiement sur la base client

- ✔ On calcule le score de chaque client
- ✔ On ordonne la population par score décroissant

Score statistique : démarche

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

Apprentissage

- ✔ On propose le produit à un échantillon de clients/prospect
- ✔ Certains achètent d'autres non
- ✔ On mesure d'autres variables = prédictrices
  - ✘ Variables socio professionnelles, comportement antérieur, autres secteurs d'achat, etc.
- ✔ On construit une équation de prédiction

$$P(achat) = Fonction(predictrices)$$

Déploiement sur la base client

- ✔ On calcule le score de chaque client
- ✔ On ordonne la population par score décroissant
- ✔ On cible les scores élevés

Score statistique : démarche

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

Apprentissage

- ✔ On propose le produit à un échantillon de clients/prospect
- ✔ Certains achètent d'autres non
- ✔ On mesure d'autres variables = prédictrices
  - ✘ Variables socio professionnelles, comportement antérieur, autres secteurs d'achat, etc.
- ✔ On construit une équation de prédiction

$$P(achat) = Fonction(predictrices)$$

Déploiement sur la base client

- ✔ On calcule le score de chaque client
- ✔ On ordonne la population par score décroissant
- ✔ On cible les scores élevés n'ayant pas encore acheté le produit

Score statistique

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

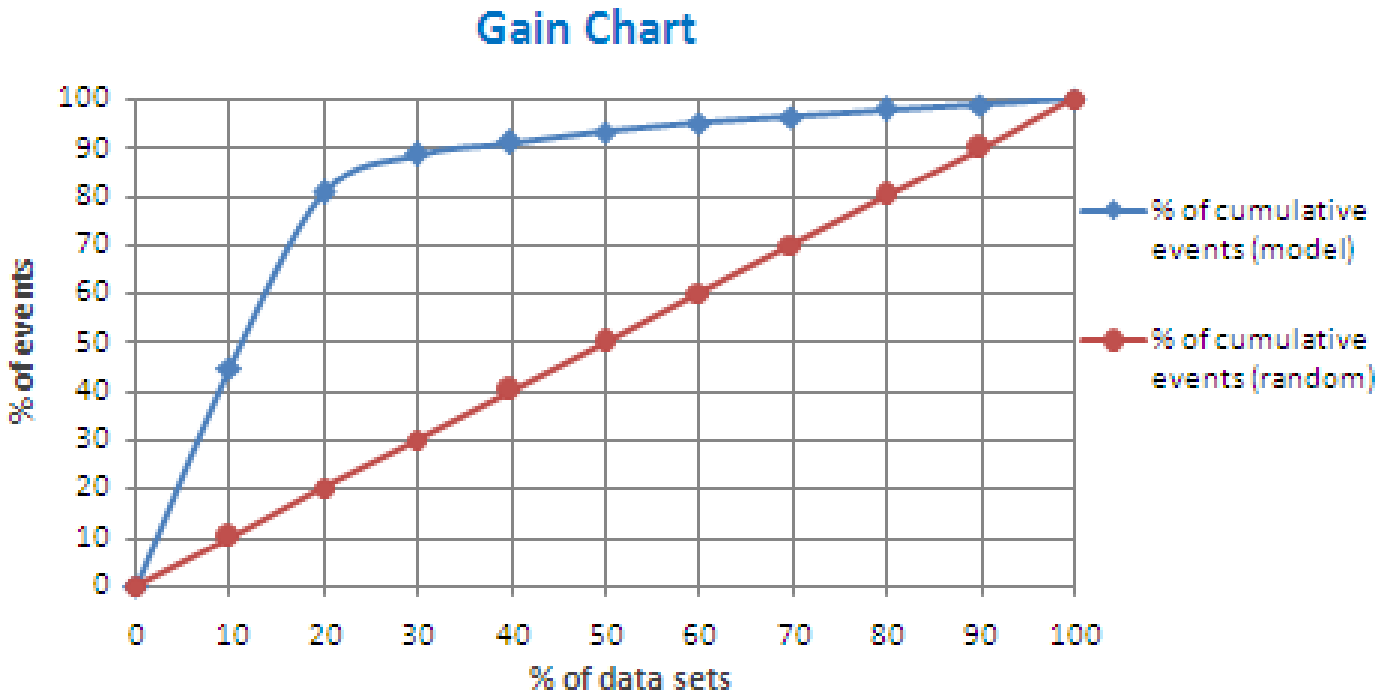
Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

- ✔ Explicatif : Identification des variables les plus pertinentes
- ✔ Prédictif : Estimation du taux de retour probable, du levier



Plus besoin des experts ?

Score statistique

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

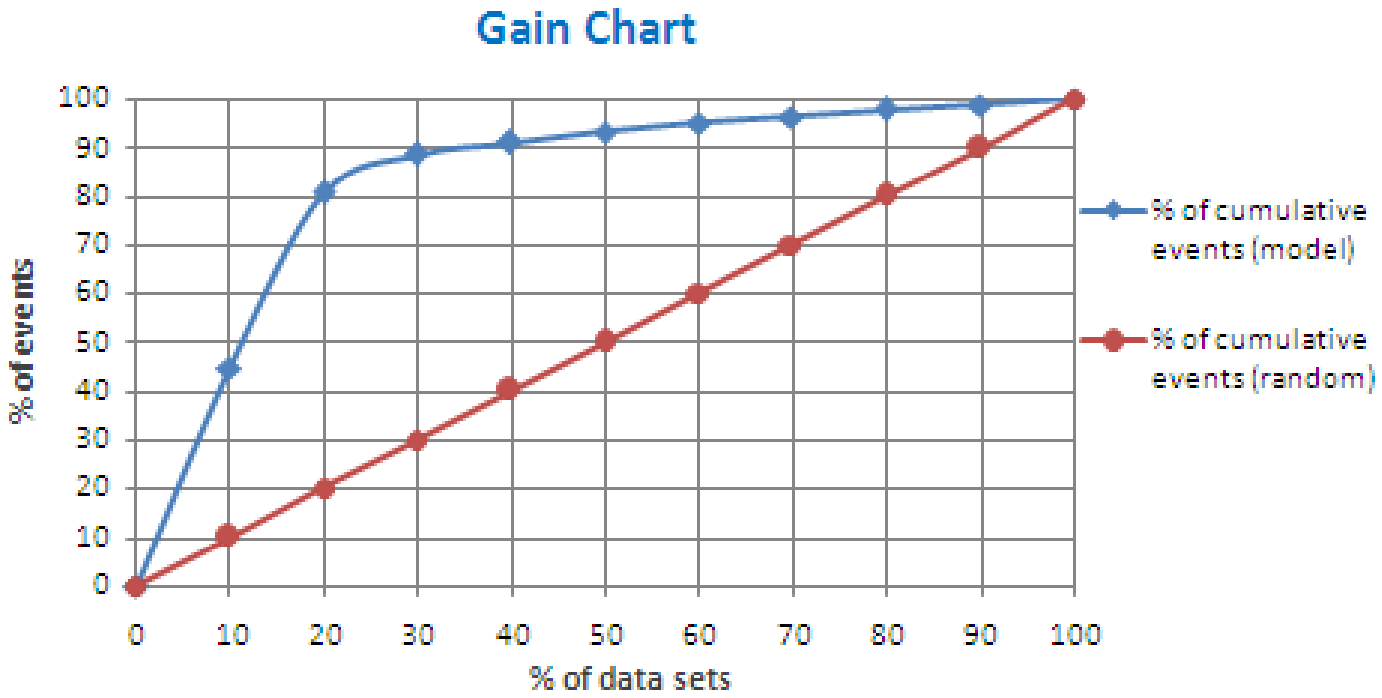
Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

- ✔ Explicatif : Identification des variables les plus pertinentes
- ✔ Prédictif : Estimation du taux de retour probable, du levier



Plus besoin des experts ? Au contraire !

De la bonne question...

La répartition des rôles dans un projet data
Par où commencer ?
Une démarche générale possible
Un exemple concret : proposer un produit à un client
À dire d'expert : à vous !
Score statistique : démarche
Score statistique
De la bonne question...
... À l'appropriation du modèle
La chaire D-CC : de la méthodologie au service open source

Il n'y a pas de vent favorable à celui qui ne sait pas où il va.

Sénèque

Exemple de questions :

- ✔ Quels clients sont les plus susceptibles de souscrire à notre nouvelle offre de crédit?
- ✔ Quel DVD proposer à un client qui a acheté Un nouvel espoir puis L'empire contre attaque ?
- ✔ Faut-il contacter ce client par mail, téléphone, ou lui proposer un rdv en agence pour lui proposer d'acheter un produit d'épargne ?

Les data-scientists n'ont pas de légitimité dans la formulation des questions, ils n'ont que des outils pour y répondre.

... À l'appropriation du modèle

La répartition des rôles dans un projet data

Par où commencer ?

Une démarche générale possible

Un exemple concret : proposer un produit à un client

À dire d'expert : à vous !

Score statistique : démarche

Score statistique

De la bonne question...

... À l'appropriation du modèle

La chaire D-CC : de la méthodologie au service open source

- ✔ La question est posée par l'expert et comprise par le DS
- ✔ Le modèle est proposé par le DS et discuté avec l'expert
- ✔ Les variables retenues sont un compromis entre des critères DS et métier
- ✔ Le modèle est rendu intelligible et fait l'objet de formation auprès de ceux qui l'utiliseront
  - ✘ Le fonctionnement général du modèle est compris
  - ✘ Les variables retenues sont connues
  - ✘ Le sens et l'ampleur générale de l'influence des variables est accessible
  - ✘ L'incertitude éventuelle autour du modèle est connue

La répartition des rôles dans un projet data
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source
Une chaire partenariale
typoChooseR
Situation de départ
ACM et distance du Chi 2
Les deux fonctions de l'ACM
Il faudrait plus de six minutes !
Comment convaincre ?
Comparer les partitions
Premier onglet : choisir les variables actives
Deuxième onglet : explorer les

La chaire D-CC : de la méthodologie au service open source

Une chaire partenariale

La répartition des rôles dans un projet data
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source
Une chaire partenariale
typoChooseR
Situation de départ
ACM et distance du Chi 2
Les deux fonctions de l'ACM
Il faudrait plus de six minutes !
Comment convaincre ?
Comparer les partitions
Premier onglet : choisir les variables actives
Deuxième onglet : explorer les

- ✔ Des entreprises mécènes
- ✔ Des laboratoires de recherche
- ✔ Des problèmes business aux questions de recherche
- ✔ Pour l'intérêt général

typoChooseR

La répartition des rôles dans un projet data
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source
Une chaire partenariale
typoChooseR
Situation de départ
ACM et distance du Chi 2
Les deux fonctions de l'ACM
Il faudrait plus de six minutes !
Comment convaincre ?
Comparer les partitions
Premier onglet : choisir les variables actives
Deuxième onglet : explorer les

Comment :

- ✔ partitionner une population sur laquelle on mesure des variables qualitatives ?
- ✔ communiquer auprès de data-miners en entreprise sur des questions statistiques/algébriques non triviales ?
- ✔ faire changer les pratiques en entreprise à partir des résultats de recherche ?
- ✔ aboutir à un service utilisable avec des technologies "libres" ?

Situation de départ

La répartition des rôles dans un projet data
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source
Une chaire partenariale
typoChooseR
Situation de départ
ACM et distance du Chi 2
Les deux fonctions de l'ACM
Il faudrait plus de six minutes !
Comment convaincre ?
Comparer les partitions
Premier onglet : choisir les variables actives
Deuxième onglet : explorer les

- ✔ Population à segmenter...
- ✔ Sur laquelle on mesure des variables qualitatives...
- ✔ Nativement ou après discrétisation...

Calculer des distances pour utiliser des méthodes géométriques :

- ✔ K-means
- ✔ CAH
- ✔ Mix des deux

ACM et distance du Chi 2

La répartition des rôles dans un projet data

Par où commencer ?

La chaire D-CC : de la méthodologie au service open source

Une chaire partenariale

typoChooseR

Situation de départ

ACM et distance du Chi 2

Les deux fonctions de l'ACM

Il faudrait plus de six minutes !

Comment convaincre ?

Comparer les partitions

Premier onglet : choisir les variables actives

Deuxième onglet : explorer les

La distance du Chi2 étant peu répandue, la pratique est de

- ✔ Faire une ACM...
- ✔ Pour récupérer *un certains nombre* d'axes quantitatifs...
- ✔ Calculer une distance euclidienne classique.

Si on garde tous les axes : distance du Chi2. Mais...

Les deux fonctions de l'ACM

La répartition des rôles dans un projet data
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source
Une chaire partenariale
typoChooseR
Situation de départ
ACM et distance du Chi 2
Les deux fonctions de l'ACM
Il faudrait plus de six minutes ! Comment convaincre ?
Comparer les partitions
Premier onglet : choisir les variables actives
Deuxième onglet : explorer les

- ✔ Fonction de représentation :
  - ✘ On ne voit qu'en 2D → "meilleurs plans"
- ✔ Fonction de réduction de la dimensionnalité
  - ✘ "Filtrer" une information non pertinente → "conjurer" l'influence de variables trop liées, éviter du bruit

La répartition des rôles dans un projet data

Par où commencer ?

La chaire D-CC : de la méthodologie au service open source

Une chaire partenariale

typoChooseR

Situation de départ

ACM et distance du Chi 2

Les deux fonctions de l'ACM

Il faudrait plus de six minutes !

Comment convaincre ?

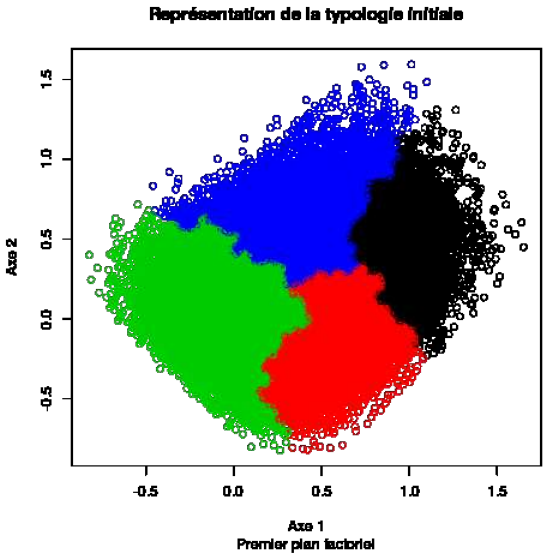
Comparer les partitions

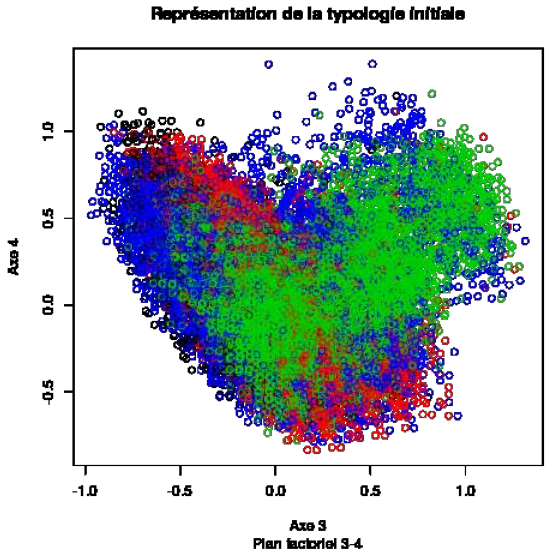
Premier onglet : choisir les variables actives

Deuxième onglet : explorer les

Pour expliquer pourquoi

- ✔ L'inertie en ACM est un critère discutable
- ✔ Par construction l'inertie expliquée par les premiers axes est faible
- ✔ Les transformations de valeurs propres sont dangereuses

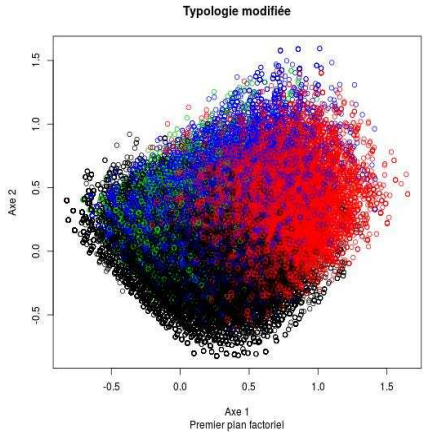


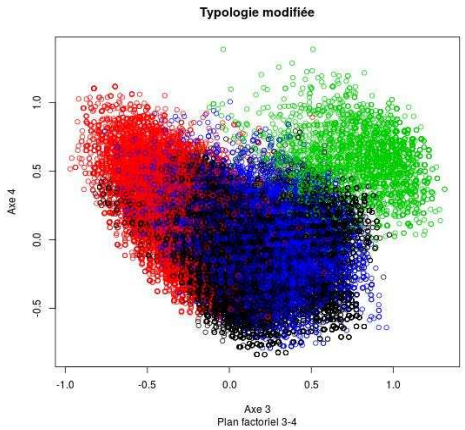


Comment convaincre ?

La répartition des rôles dans un projet data
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source
Une chaire partenariale
typoChooseR
Situation de départ
ACM et distance du Chi 2
Les deux fonctions de l'ACM
Il faudrait plus de six minutes !
Comment convaincre ?
Comparer les partitions
Premier onglet : choisir les variables actives
Deuxième onglet : explorer les

Qu'une typologie qui semble moins jolie sur le premier plan a en fait plus de sens si on regarde les dimensions supérieures ?





- ✔ Un rapport de recherche → "Mmm" poli
- ✔ Un Guide de bonnes pratiques → "Ah d'accord"

Mais il faut un outil pour faire changer les pratiques : typoChooser

Comparer les partitions

La répartition des rôles dans un projet data

Par où commencer ?

La chaire D-CC : de la méthodologie au service open source

Une chaire partenariale

typoChooseR

Situation de départ

ACM et distance du Chi 2

Les deux fonctions de l'ACM

Il faudrait plus de six minutes !

Comment convaincre ?

Comparer les partitions

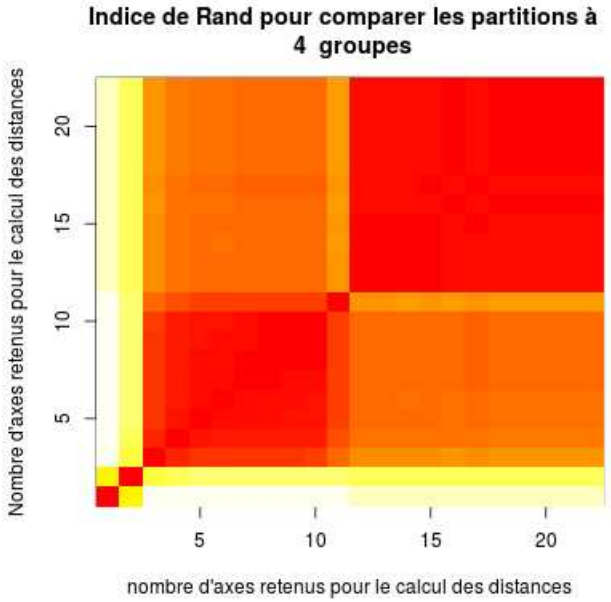
Premier onglet : choisir les variables actives

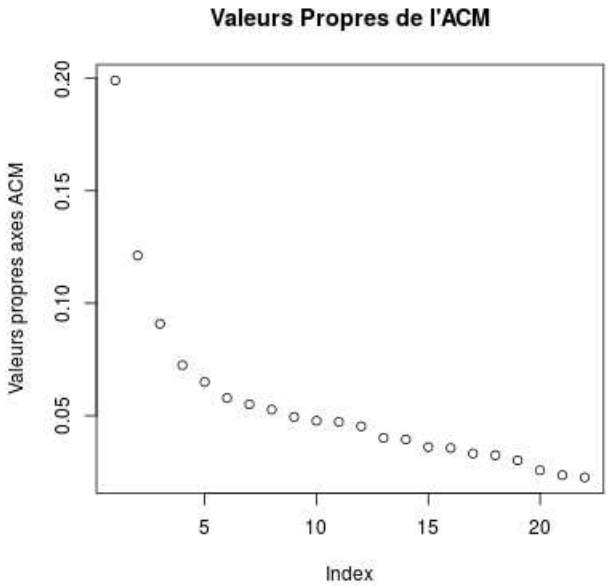
Deuxième onglet : explorer les

Indice de Rand :

- est-ce que deux partitions classent toutes les deux des individus ensemble, ou toutes les deux séparément ?
- est ce qu'au contraire les deux partitions sont en désaccord ?

On représente graphiquement cette proximité entre partitions construites à partir d'une, deux, trois... p composantes de l'ACM.

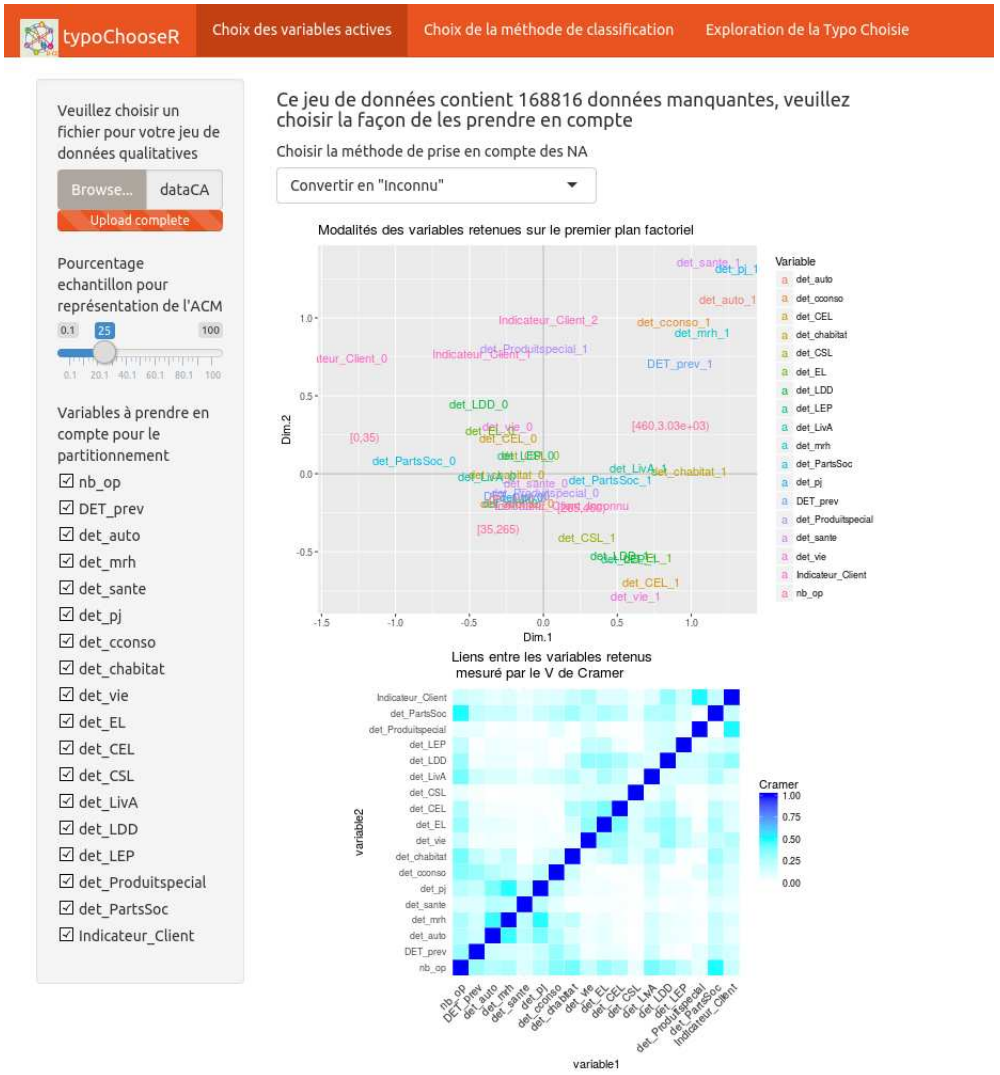




La répartition des rôles dans un projet data
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source
Une chaire partenariale
typoChooseR
Situation de départ
ACM et distance du Chi 2
Les deux fonctions de l'ACM
Il faudrait plus de six minutes !
Comment convaincre ?
Comparer les partitions
Premier onglet : choisir les variables actives
Deuxième onglet : explorer les

Premier onglet : choisir les variables actives

- ✔ Charger le fichier
- ✔ Choisir les variables actives



Deuxième onglet : explorer les méthodes

La répartition des rôles dans un projet data

Par où commencer ?

La chaire D-CC : de la méthodologie au service open source

Une chaire partenariale

typoChooseR

Situation de départ

ACM et distance du Chi 2

Les deux fonctions de l'ACM

Il faudrait plus de six minutes !

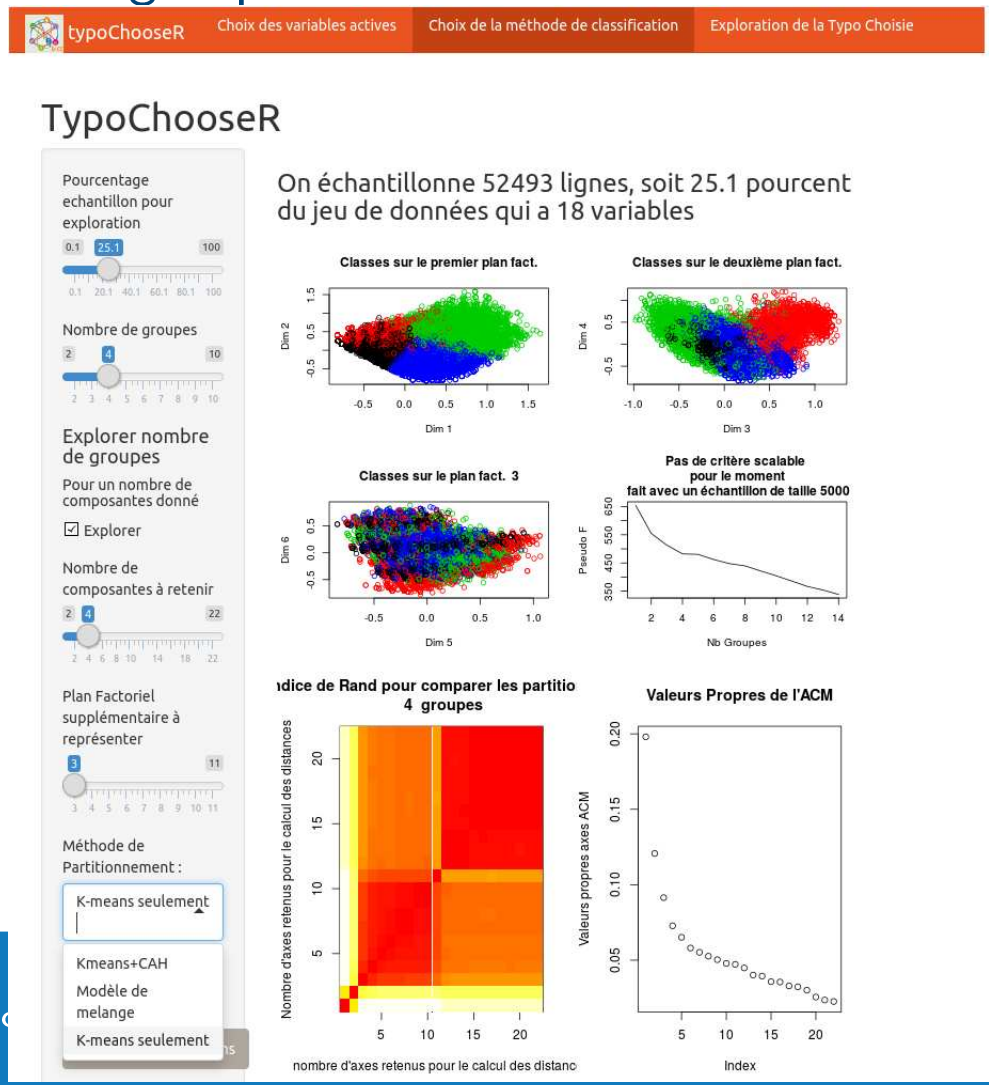
Comment convaincre ?

Comparer les partitions

Premier onglet : choisir les variables actives

Deuxième onglet : explorer les méthodes

- ✔ La méthode : géométrique ou modèle de mélange
- ✔ Le nombre de composantes à retenir (si géométrique)
- ✔ Le nombre de groupes à retenir



Troisième Onglet : explorer les groupes

La répartition des rôles dans un projet data

Par où commencer ?

La chaire D-CC : de la méthodologie au service open source

Une chaire partenariale

typoChooseR

Situation de départ

ACM et distance du Chi 2

Les deux fonctions de l'ACM

Il faudrait plus de six minutes !

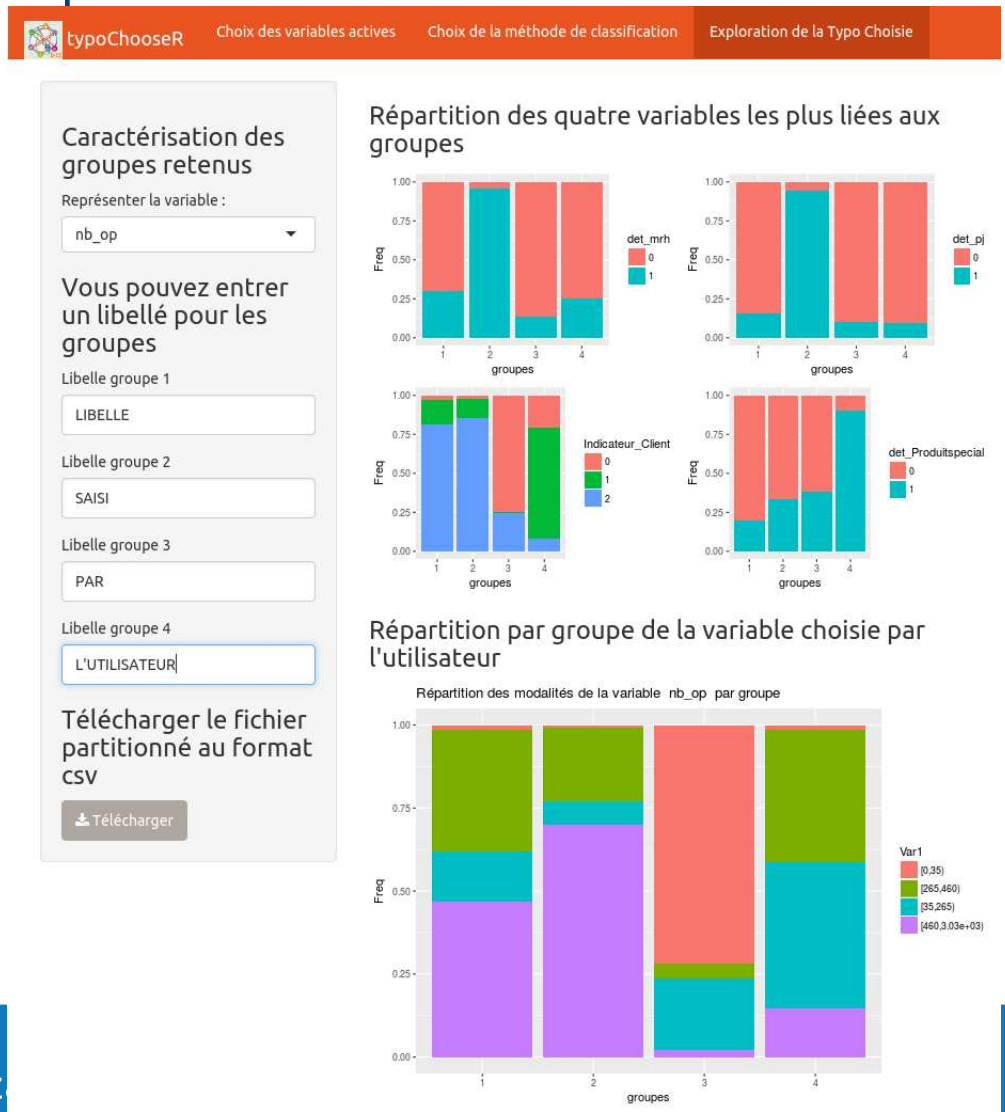
Comment convaincre ?

Comparer les partitions

Premier onglet : choisir les variables actives

Deuxième onglet : explorer les

- ✔ Nommer les Groupes
- ✔ Itérer ou exporter



D'un démonstrateur à une application

La répartition des rôles dans un projet data
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source
Une chaire partenariale
typoChooseR
Situation de départ
ACM et distance du Chi 2
Les deux fonctions de l'ACM
Il faudrait plus de six minutes !
Comment convaincre ?
Comparer les partitions
Premier onglet : choisir les variables actives
Deuxième onglet : explorer les

- ✔ Démonstrateur R+Shiny
  - ✘ Jouer avec des jeux de données réels
  - ✘ Voir l'impact sur les groupes retenus "en direct"
  - ✘ On n'insiste plus sur les failles ou les erreurs méthodologiques, mais sur les gains possibles avec d'autres méthodes
  - ✘ $\Rightarrow$ Volonté de changer
- ✔ Passage en production
  - ✘ Optimiser le code R
  - ✘ Installer un serveur Shiny ou encapsuler l'application
  - ✘ Implémenter la solution retenue dans technologie présente dans l'entreprise (SAS, Python...)

Une chaire + R + Shiny : Vecteur d'innovation

La répartition des rôles dans un projet data
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source
Une chaire partenariale
typoChooseR
Situation de départ
ACM et distance du Chi 2
Les deux fonctions de l'ACM
Il faudrait plus de six minutes !
Comment convaincre ?
Comparer les partitions
Premier onglet : choisir les variables actives
Deuxième onglet : explorer les

chaire.d-cc@univ-ubs.fr

decisionnelclient.org

https://chairedecisionnelconnaissanceclient.shinyapps.io/typoChooseR/

La répartition des rôles dans un projet data

Par où commencer ?

La chaire D-CC : de la méthodologie au service open source

Une chaire partenariale

typoChooseR

Situation de départ

ACM et distance du Chi 2

Les deux fonctions de l'ACM

Il faudrait plus de six minutes !

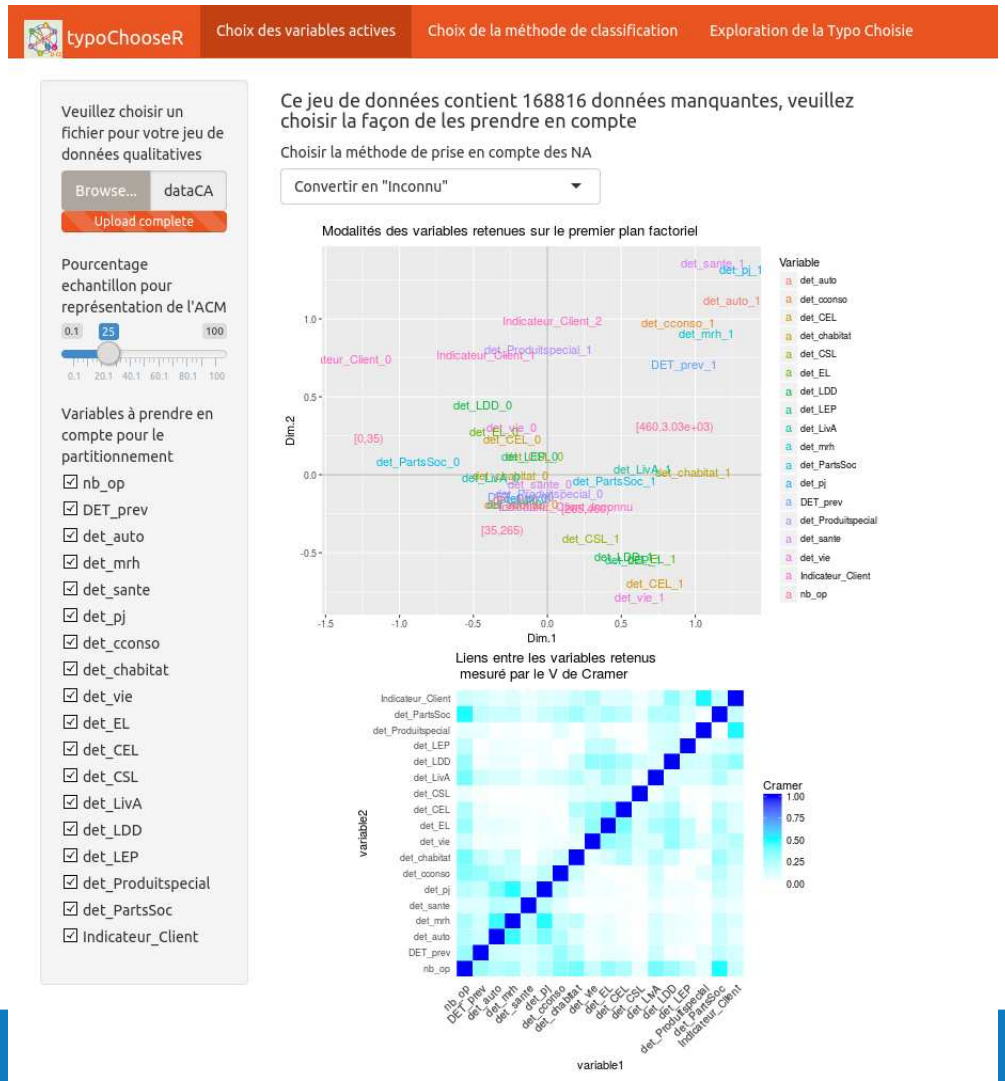
Comment convaincre ?

Comparer les partitions

Premier onglet : choisir les variables actives

Deuxième onglet : explorer les

Démonstration, à un rythme plus compatible avec un petit-déjeuner !



Questions, échanges, viennoiseries

La répartition des rôles dans un projet data
Par où commencer ?
La chaire D-CC : de la méthodologie au service open source
Une chaire partenariale
typoChooseR
Situation de départ
ACM et distance du Chi 2
Les deux fonctions de l'ACM
Il faudrait plus de six minutes !
Comment convaincre ?
Comparer les partitions
Premier onglet : choisir les variables actives
Deuxième onglet : explorer les

Merci de votre attention.

chaire.d-cc@univ-ubs.fr
matthieu.gousseff@univ-ubs.fr
veronique.bosc@univ-ubs.fr