

Préambule

Introduction

Dans un projet Data Science, on peut énumérer les différentes étapes suivantes:

- Définition de la problématique
- Collecte des données
- Stockage des données
- Nettoyage et analyse
- Visualisation des données
- Prise de décision

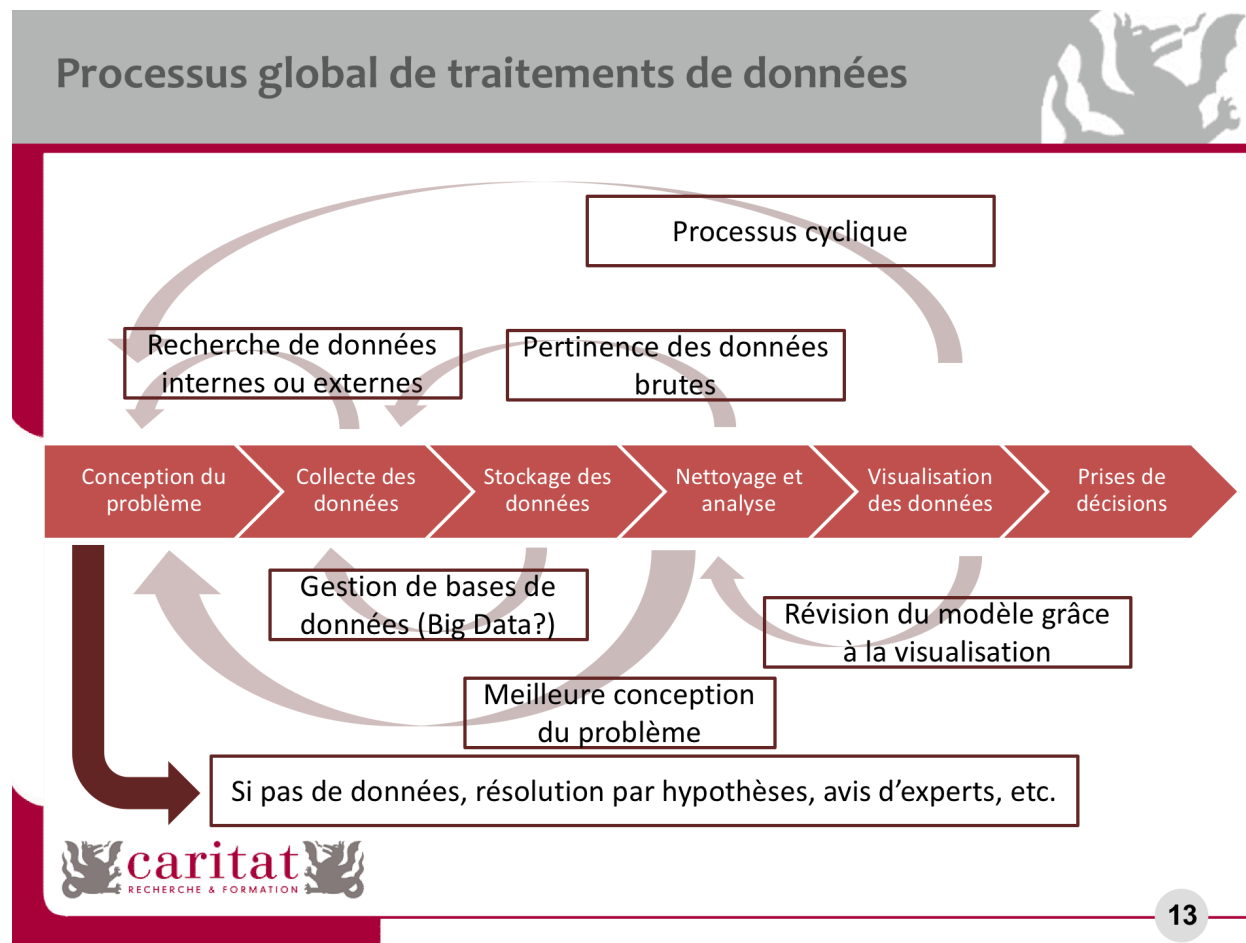


Figure 1:

La création de graphiques peut intervenir à plusieurs étapes de ces traitements:

- Lors de l'exploration des données, des graphiques rapides et simples peuvent aider à mieux comprendre la structure des données
- La visualisation des résultats finaux permet de construire un tableau de bord

- La visualisation des algorithmes permet de mieux comprendre les principes mathématiques

Ainsi, si on isole le problème de visualisation des données, en supposant que les données sont bien stockées, et que des analyses simples sont réalisées, on peut avoir les différentes étapes suivantes:

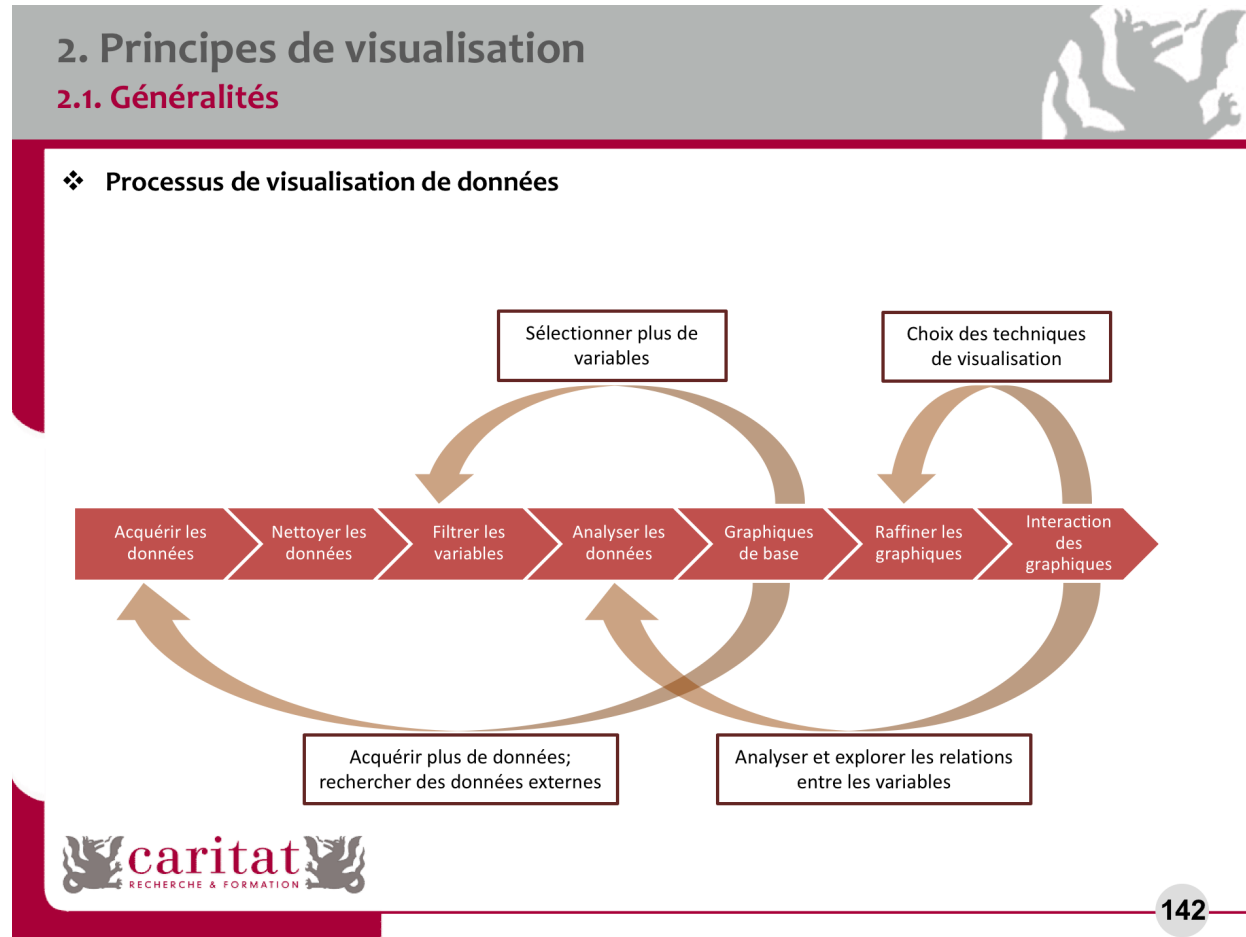


Figure 2:

Outils

Outil de programmation principal : Rstudio Packages R :

- Manipulation des données: data.table, dplyr, tidyr, stringr, lubridate
- Graphiques : plotly, ggplot2, ggvis, googleVis
- Cartographie : Leaflet, sp, maptools, maps, ggmap
- Séries temporelles : zoo, xts, quantmod, dygraphs
- Widget html: htmlwidgets
- Application web : Shiny
- Webscrapping : XML, rselenium, httr

De nombreux packages sont à découvrir par vous-mêmes.