

# R

## Visualisation de données avec R

Présenté par François Pelletier

Atelier du 15 octobre 2016

Nous explorerons trois outils pour visualiser des données:

- leaflet qui permet de mettre des données sur une carte géographique interactive
- plotly qui permet de créer des graphiques interactifs pour le web
- ggplot2 qui permet de créer des graphiques personnalisés pour les publications

Ces outils sont utilisés par les scientifiques, les professionnels de l'intelligence d'affaires et les journalistes.

Vous aurez les bases nécessaires pour créer des graphiques de qualité similaires à ceux du New York Times (voir <https://twitter.com/nytgraphics>) suite à cet atelier.

Matériel requis:

- Un ordinateur portable avec Linux, MacOS ou Windows et au moins 2go de RAM
- Le logiciel GNU R
- L'éditeur RStudio <https://www.rstudio.com/> ou Emacs avec le plugin ESS
- Les librairies tidyverse, ggplot2, leaflet et plotly

Installation de R:

- disponible sur <https://cran.rstudio.com/> ou MRAN <https://mran.microsoft.com>

Installation de RStudio (facultatif):

- Visiter <https://rstudio.com>

Installation de Emacs ESS (facultatif):

- Sur Windows et Mac:
  - Visiter <http://vgoulet.act.ulaval.ca/emacs/windows/>
- Sur Linux:

```
sudo apt-get install ess
```

Installation des packages nécessaires pour l'atelier (dans R)

```
install.packages(c("ggplot2", "leaflet", "plotly", "tidyverse"))
```

Le contenu source de la présentation est disponible sur ce dépôt GitLab:

<https://gitlab.com/franc00018/r-dataviz-formation>

En installant l'ensemble des bibliothèques nécessaires et en utilisant l'IDE RStudio, vous pourrez répliquer exactement ma présentation.

Le contenu compilé en HTML (certaines fonctionnalités non disponibles sans Shiny) est disponible dans cette archive: <https://francoispelletier.org/owncloud/index.php/s/OKc9aYPkn3ZWdNg>

# Le logiciel statistique R pour l'exploration de données ouvertes

## But de cette page

Cette page est un résumé de l'atelier du 31 octobre 2015 sur le logiciel statistique R présenté par François Pelletier ([francois@francoispelletier.org](mailto:francois@francoispelletier.org))

## Installations des logiciels

- MRO (Microsoft R Open) est un interpréteur du langage R. C'est la version libre de Microsoft R Enterprise.
  - Téléchargement et installation avec Qapt
- MKL (Math Kernel Library, ATTENTION logiciel non-libre)
  - Installation facultative
- Rstudio qui est l'environnement de développement intégré le plus populaire avec R
  - Téléchargement et installation avec Qapt

## Présentation des logiciels

- Logiciel libre, implémente le langage de programmation S, multi-paradigme (objet, impératif et fonctionnel)
- Environnement : fonctions de base, extension avec des paquets (+3000), souvent écrites en C, C++ et FORTRAN.
- `library("MASS")` (Modern Applied Statistics with S+ : excellent livre pour les statistiques)
- Développer en R (outils les plus populaires): Rstudio, Emacs + ESS, Eclipse + extension StatET
- Les opérateurs de base
- Les objets en R
- Informatif : interface graphique Shiny.
- Installer les paquets : `install.packages("actuar")` du professeur Vincent Goulet de l'Université Laval.
- Voir CRAN pour la liste de tous les packages.
- `library("devtools")` permet d'installer des paquets à partir de GitHub, BitBucket,
- Langage Yacas pour symbolique (algèbre) via `library("ryacas")`

## Exploration

Démarrage et survol rapide de Rstudio.

## Ressources

- [Projet R](#) pour les statistiques (voir section Manuals)

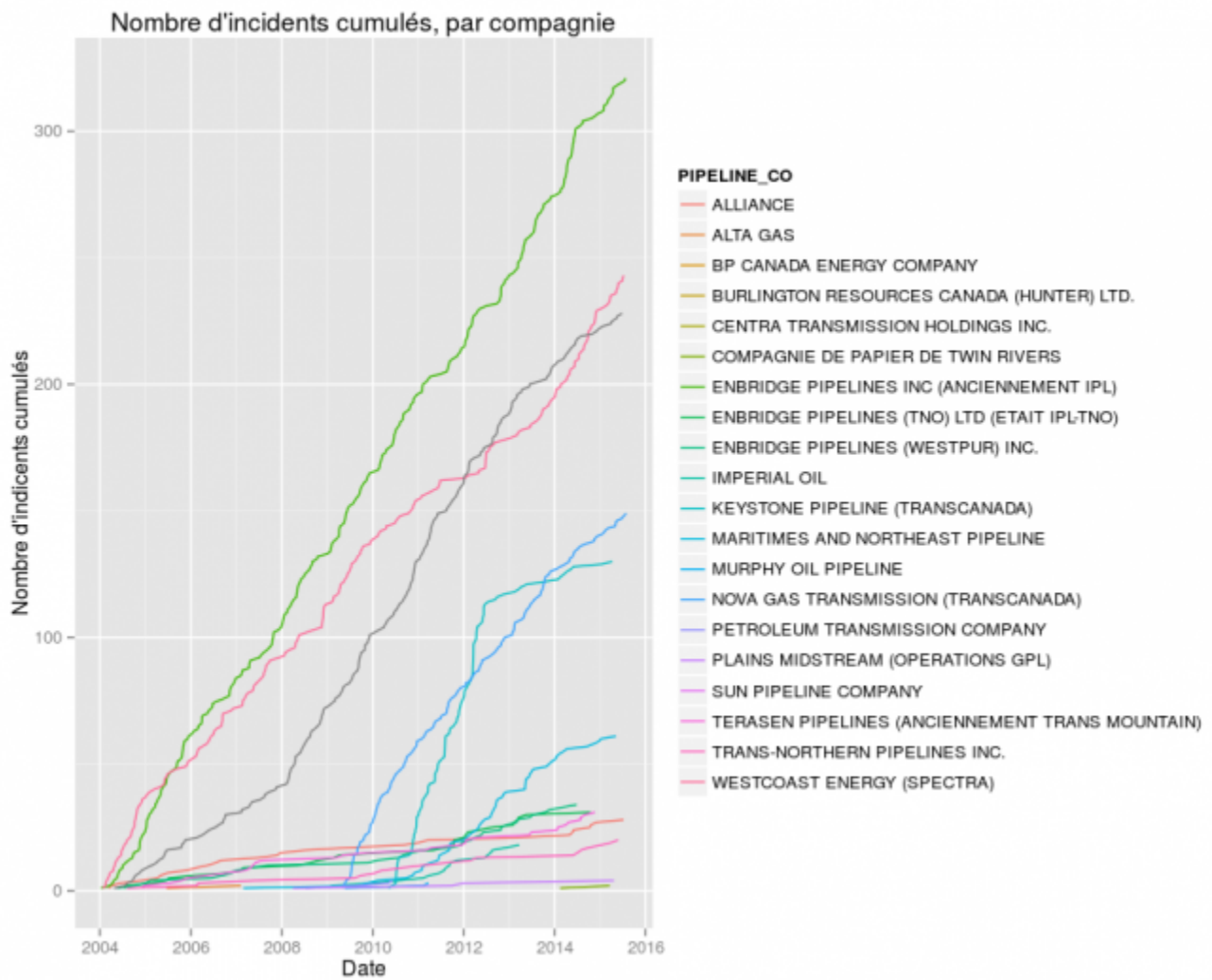
- [Le journal R](#) (voir section Books)
- [CRAN](#) (Comprehensive R Architecture Network)
- [Wiki](#) de la programmation en R
- [Blog](#) du langage R (nouvelles et tutoriels)

## Présentation d'un cas

- François nous a présenté un travail qu'il a effectué en utilisant le langage R.
- Analyse d'un jeu de données ouvertes provenant du site du gouvernement ouvert canadien : [Données sur les événements de pipeline de Janvier 2004 au présent](#)
- [Branche créée sur GitHub : PipelineViz](#) sur Git Analyse (jeu de données ouvertes).
- Explication des étapes suivies.
- Présentation de la carte des incidents avec déversements produite à partir des données :

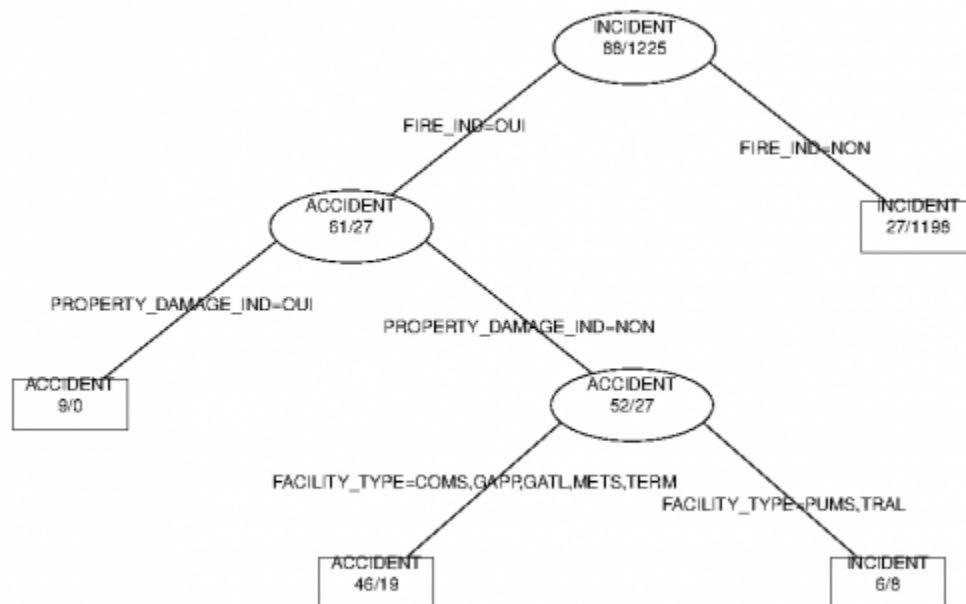


- Présentation d'un schéma du nombre d'incidents cumulés, par compagnie :



- Présentation d'un arbre de décision :

## Arbre de classification pour Type d'événement



- Présentation de diapositives sur son GitLab
- dplyr
- Une personne très active avec le langage R : Hadley Wickam hadley.co

2015-10-31 : Première présentation par François Pelletier (francois@francoispelletier.org), rédaction par Jean Leblond (jeanleblondpublic@gmail.com)

2016-10-15 : Deuxième présentation avec mise à jour par François Pelletier (francois@francoispelletier.org)

From:  
<https://linuq.org/> - **LinuQ: Logiciels libres à Québec**

Permanent link:  
<https://linuq.org/logiciels/r>

Last update: **2018/06/25 07:47**

