

## Projet 6 – Exploration et visualisation des données

Dans ce projet, vous produirez des visualisations pour trois ensembles de données.

**Données:** *GlobalCitiesPBI.csv*; *2016collisionsfinal.csv*; *HR\_2016\_Census\_simple.csv*

### Tâches:

Pour chaque ensembles de données:

1. Établissez une liste de variables (7 au maximum) que vous jugez essentielles à une bonne compréhension de l'ensemble de données. Justifiez vos choix.
2. Créez 2 visualisations ridicules et atroces; en mettant en évidence les informations inutiles et/ou les "idées" trompeuses qui en découlent.
3. Créez (au moins) 4 visualisations multivariées pour l'ensemble de données en utilisant (certaines) des variables énumérées à l'étape 1.
4. Produisez (au moins) 2 visualisations "définitives" pour l'ensemble de données. Utilisez les principes discutés en classe (y compris la documentation, les présentations, les annotations, etc.). Vous devez mettre l'accent sur le contenu ET sur la présentation.

### Conseils et suggestions:

- Utilisez ggplot2/seaborn et des outils pour éditer les résultats au besoin.
- Envisagez la création d'un nombre raisonnablement élevé de graphiques en utilisant une sélection aléatoire de variables afin de minimiser les chances de manquer des informations utiles ou des renseignements importants. C'est une approche particulièrement importante quand on fait affaire à des ensembles comptant un nombre élevé d'attributs, et elle devrait vous éclairer dans vos choix à l'étape 1, ci-dessus.