**Projet Tidyverse / shiny**

*Partie 1 : Extraction de données et Tidyverse*

**Objectif final du projet** : concevoir une application de Shiny permettant de suivre l’évolution des campagnes de financement participatif du site Ulule

**Jeu de données**: *ulule\_2019.csv*

**Objectif de cette première étape**: préparer un script R permettant d’extraire les données sous le périmètre/ format voulu, et réaliser les principales visualisations qui seront nécessaires ultérieurement à l’application.

**Echéance**: scriptà préparer avant la séance du 8 Novembre

***Pas de rendu à prévoir, l’objectif est d’éviter une surcharge de travail au moment où commencera la seconde partie du projet sur Shiny***

* **Périmètre :**
* Extraire le jeu de données
* Exclure du périmètre les campagnes annulées
* Se restreindre au périmètre des 8 pays ayant le plus de campagnes au total
* Exclure le mois le plus récent où les données seraient éventuellement incomplètes
  + Exemple : dernière campagne enregistrée le 12/01/19 -> supprimer les données du mois de Janvier 2019
* **Manipulations de données à prévoir :**
* Ne conserver que les variables utiles aux traitements et à la visualisation
  + - Exclure d’entrée les variables n’ayant aucune utilité pour les visualisations / les traitements
* Convertir les montants de toutes les devises en euro
* Créer pour cela une table à part de taux de conversion, ou bien une instruction de type if\_else
* Recalculer le nombre de jours des campagnes
* Différence en jours entre date de début et date de fin
* Créer une nouvelle variable de catégorie au format plus « macro » regroupant les 6 principales catégories
* Les autres étant regroupées dans un « autre », ou bien avec l’une des six principales si le regroupement semble pertinent (ex : regrouper « jeux » et « BD »)
* **Visualisations  à préparer:**
* Nombre de campagnes par catégorie (sur 6 catégories / regroupements de catégories)
* Nombre de campagnes par pays (sur 8 pays / regroupement de pays)
* Distribution de la durée des campagnes
* Evolution du nombre de campagnes crées au fil des mois, *sur les deux dernières années*
* Evolution de la proportion de campagnes financées au fil des mois, *sur les deux dernières années*
* Evolution des montants moyens des campagnes financées au fil des mois / années, *sur les deux dernières années*

**Consignes générales :**

* **Toutes les manipulations de données doivent être réalisées dans l’esprit du « Tidyverse** », en utilisant les différents packages qui le composent.
  + Viser le minimum de $, de sélections de dataframe via les crochets [], de fonctions comme table ou subset…
  + Enchaîner les pipe (%>%) plutôt que multiplier les affectations
* Le **code doit être aussi clair que possible**, avec notamment une distinction des différentes parties du script (importation, librairies, exploration initiale, pré-traitements, visualisations…)
  + Un code clair facilitera son implémentation future dans shiny
* **Soigner les visualisations**, avec un type de visualisation approprié à l’objectif et un soin accordé aux détails de finition
  + Les graphiques doivent être aussi clairs que possibles, compréhensibles du premier coup d’œil
  + Penser aux détails : libellés des axes, titre, choix raisonné des couleurs, affichage des libellés, thème personnalisé...
  + Graphiques à construire avec ggplot, ou éventuellement plotly

**Notes sur le jeu de données :**

* Une ligne = une campagne de financement participatif
* Données réelles basées sur le site de financement participatif Ulule
* Une campagne financée est une campagne dont le montant récolté a atteint ou dépassé son objectif