

# Reproductibilité de l'analyse

Pourquoi est-ce important et comment le faire ?

8 Juillet 2021

# Objectifs d'apprentissage

## À propos d'aujourd'hui:

- Comprendre ce que vous pouvez gagner de la reproductibilité de l'analyse.
- Connaître les principales exigences techniques à mettre en place pour qu'une analyse soit reproductible.
- Démonstration d'une manière pratique de faire un gâteau en utilisant les données d'une enquête auprès des ménages : analyse et interprétation et "data story" !

## Pas pour aujourd'hui:

- Formation d'initiation au langage R ! Pour cela, dirigez-vous vers [Formation UNHCR Learn & Connect-R](#)

## Une vision pour l'analyse des données

"Des équipes multifonctionnelles, dotées d'une maîtrise des données renforcée, mènent régulièrement des sessions conjointes d'interprétation de données documentées pour définir leurs orientations stratégiques sur la base de preuves statistiques"

# Théorie du changement: analyse de données

## Utilisateur approprié des données pour le plaidoyer et la prise de décision programmatique

- ↪ Des **standards** organisationnels existent pour définir comment encoder et traiter l'ensemble de données des enquêtes auprès des ménages
- ↪ Les experts en données de terrain sont formés sur la base de recettes précises et d'outils prédéfinis à chaque étape du **cycle de vie des données**
- ↪ Les données sont présentées, discutées et liées aux connaissances d'experts lors de sessions d'**interprétation** de données avec une équipe multifonctionnelle
- ↪ Toutes les interprétations valides potentielles, y compris les vues divergentes, sont systématiquement **enregistrées**
- ↪ **Persuasion** Des "data stories" et des notes d'analyse de politiques publiques sont générés

# La science des données, c'est comme la cuisine

Quand un chef commence avec un nouveau plat...

- Hypothèses de Dégustation - Poser les bonnes questions
- Ingrédients = source des Données
- Lavez votre nourriture = nettoyez vos données
- Ingénierie des saveurs = créer des variables calculées et dérivées
- Goûter et explorer = remodeler & visualiser les données
- Ajuster votre four = modélisation statistique
- Art de la mise en form = utiliser la marque de style
- Documentez votre recette = ajoutez des commentaires techniques

**Data**



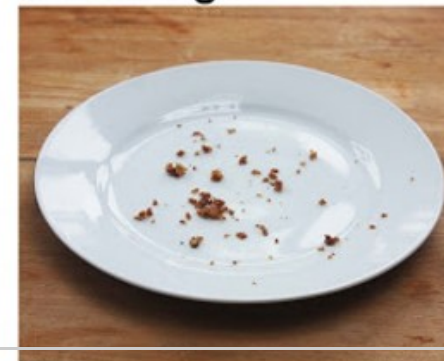
**Information**



**Presentation**



**Knowledge**



# Information Anxiété & Analyse paralysie

*Quand les gens ne veulent pas manger de gâteau...*

Source potentielle de réticence...

- Je ne sais pas comment le manger : je vois tous ces éléments dessus sans pouvoir comprendre pourquoi ils y ont été ajoutés et comment cela fonctionne...
- Je ne fais pas confiance à ce gâteau : Comment a-t-il été créé ? As-tu bien suivi la recette ? Les ingrédients étaient-ils frais ? Puis-je faire confiance à la façon dont vous vous êtes procuré les ingrédients?
- Ce n'est pas le gâteau dont j'ai besoin ! Il a l'air trop lourd & trop gros : je ne pourrai pas le digérer...
- Je n'ai pas faim et je ne sais même pas quel gâteau je veux...



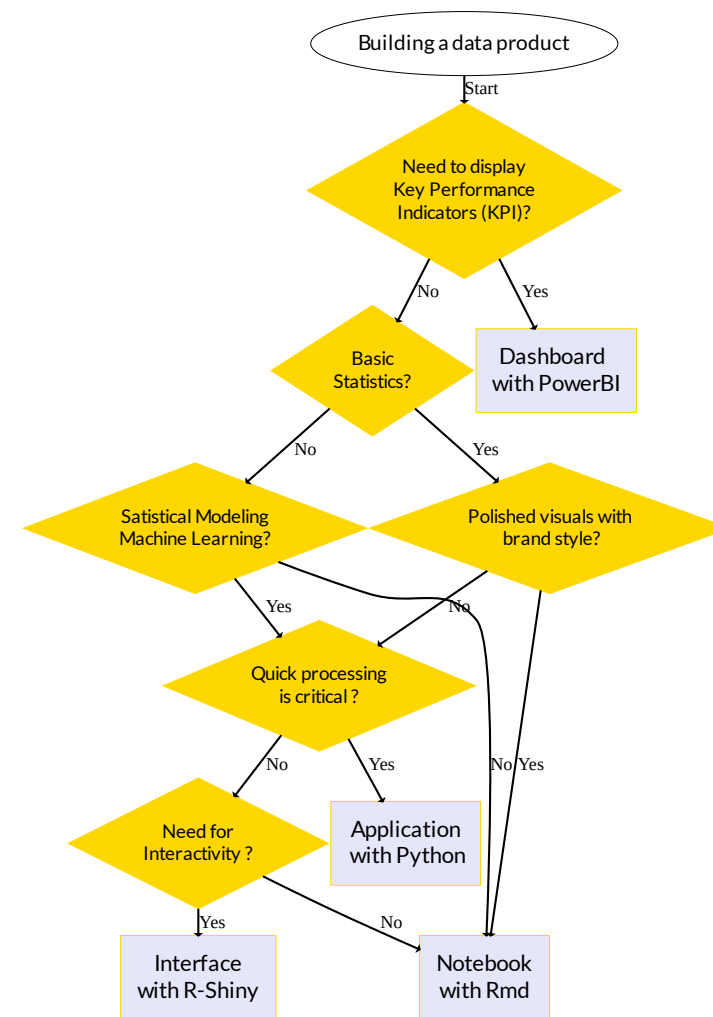
# Produits de données : quand quoi ?

Les tableaux de bord sont pertinents pour afficher les KPI ! (*comme quand vous conduisez votre voiture...*)

Les indicateurs clés de performance (KPI) sont des indicateurs spécifiquement conçus pour montrer les progrès vers un résultat attendu, c'est-à-dire un **cible** prédéfini

Créer une base analytique pour la **prise de décision**, i.e. Business Intelligence

Aidez à attirer l'attention de la direction Snr sur ce qui compte le plus - un bon tableau de bord doit être **concis**



# Pourquoi devons-nous travailler de manière reproductible ?

Éthique, Productivité, Apprentissage



# Éthique : la science, c'est « *montrez-moi* »! - pas « *faites-moi confiance* »...

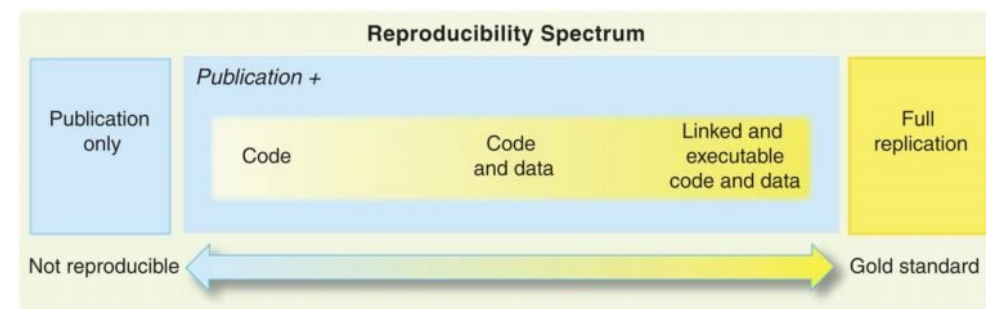
La reproductibilité permet l'**examen par les pairs**

L'examen par les pairs permet la **transparence**

La transparence permet **contrôle**

Le contrôle permet la **responsabilité**

*C'est bien de faire des erreurs, tant qu'on peut les détecter et qu'on peut en tirer des leçons...*

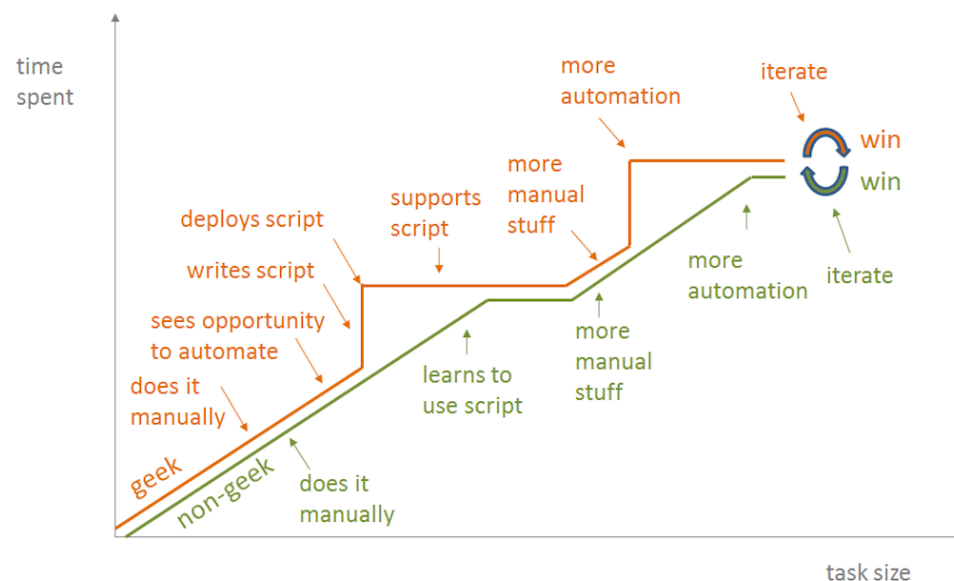


# Productivité : faire avancer les choses rapidement et en toute sécurité !

L'automatisation via des fonctions et des scripts peut aider à sauter des **tâches répétitives**

Les tâches qui impliquent des **manipulations de données** récurrentes sont entreprises par des équipes... .. mais tout le monde dans l'équipe n'a pas besoin d'être un **geek/coder** !

Lorsqu'un investissement suffisant peut être fait, une **Interface utilisateur graphique (GUI)** peut être développée pour des fonctions spécifiques afin de faciliter la courbe d'apprentissage des nouveaux utilisateurs pendant qu'ils sont encore en train de développer leurs compétences personnelles en R.



# Une Communauté R tournée vers l'apprentissage

Quelle approche est l'exercice le plus attrayant parmi les 2 proposés à part ?!!

Commencez à partir d'un produit final et faire **reverse engineering** !

Mangez le gâteau d'abord! (puis jouez avec et changez les ingrédients...)

(a)

- Declare the following variables
- Then, determine the class of each variable

```
# Declare variables
x ← 8
y ← "monkey"
z ← FALSE

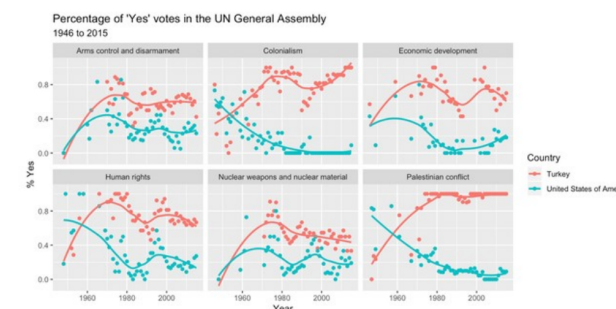
# Check class of x
class(x)
#> [1] "numeric"

# Check class of y
class(y)
#> [1] "character"

# Check class of z
class(z)
#> [1] "logical"
```

(b)

- Open today's demo project
- Knit the document and discuss the results with your neighbor
- Then, change **Turkey** to a different country, and plot again



# Conditions de reproductibilité.

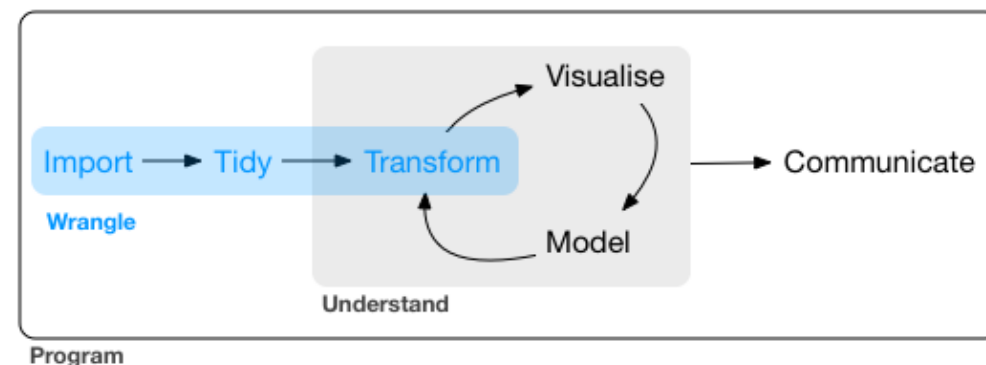
Données d'approvisionnement, analyse de la documentation et  
sortie d'emballage

# Préparation des données

le "Data Wrangling" prend généralement plus de 80% du temps de tout projet de données...

Imaginez si vous avez besoin de revenir en arrière dans votre analyse...

Corrigez à toutes les étapes du processus et réexécutez tout.



# Documenter l'analyse

DO

*Cutting corners to meet arbitrary management deadlines*



*Essential*

Copying and Pasting  
from Stack Overflow

*How to actually learn any new programming concept*



*Essential*

Changing Stuff and  
Seeing What Happens

*The internet will make those bad words go away*



*Essential*

Googling the  
Error Message

*Software can be chaotic, but we make it work*



*Expert*

Trying Stuff  
Until it Works

*Where's the fun in just knowing what the code is supposed to do?*



*Essential*

Excuses for Not  
Writing Documentation

*Does it run? Just leave it alone.*



Writing Code that  
Nobody Else Can Read

*git commit -m "changes"*



*Writing*

Useless Git  
Commit Messages

*Who are you kidding?*



"Temporary"  
Workarounds

DON'T



**UNHCR**  
The UN Refugee Agency

# De la "Fonction" jusqu'au "Package"

## Automatisation progressive

- niveau 1 : écrire une commande
- niveau 2: organiser plusieurs commandes ensemble en fonction réutilisable
- Niveau 3 : organiser plusieurs fonctions ensemble dans un package
- Niveau 4: comprend les données de test et la documentation
- Niveau 5 : Tests unitaires, alias revue de code
- Niveau 6 : Interface utilisateur graphique (GUI)

## Package

- DESCRIPTION
- R/
- tests/
- man/
- vignettes/
- data/
- NAMESPACE

SETUP

WRITE CODE

TEST

DOCUMENT

TEACH

ADD DATA

ORGANIZE

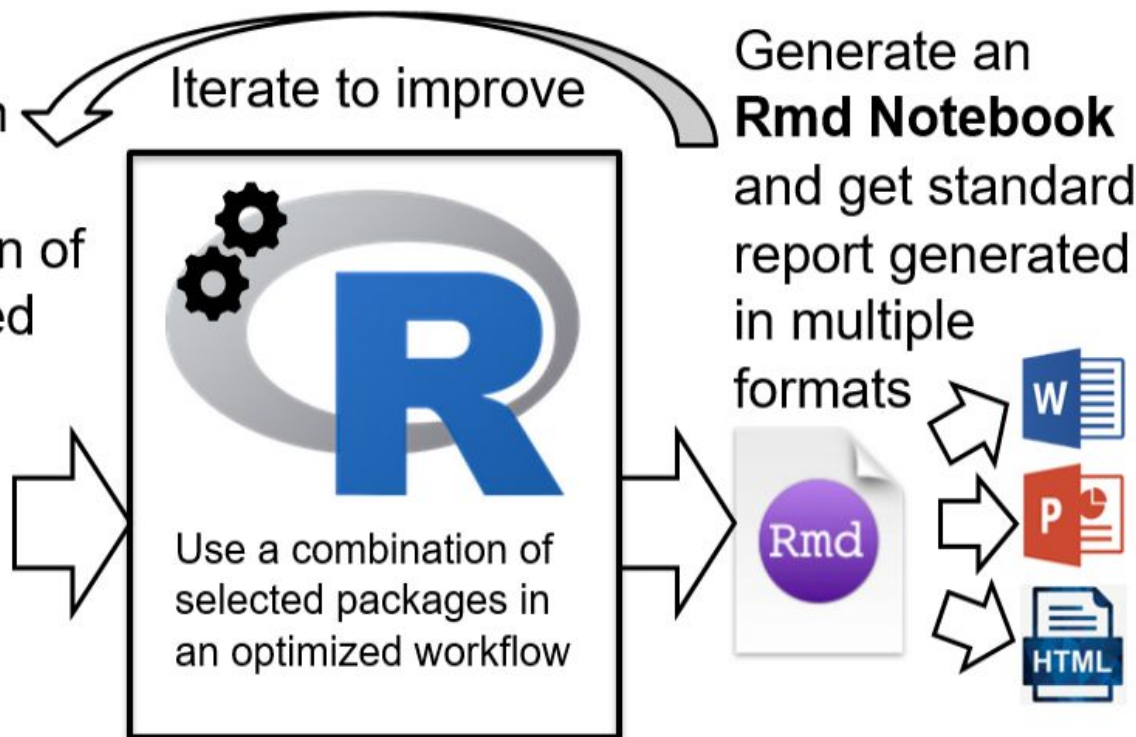
# Exercice pratique : données d'enquête auprès des ménages

Décryptage, interprétation et diffusion



# Étape 1 - Notebook pour l'exploration automatique des données, alias "crunching"

**Configure the analysis plan** in Excel within an extended version of the XlsForm used for the survey



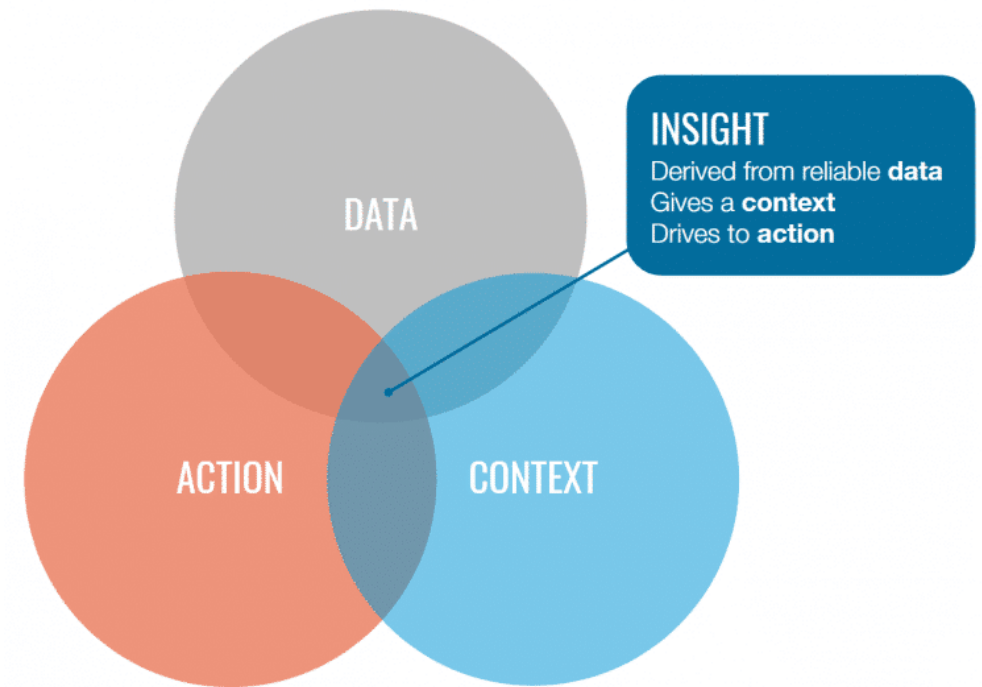
# Étape 2 - Notebook de Documentation pour les éclairage via les données : Analysis Repo

Insight : La capacité d'acquérir une compréhension précise et profonde de quelqu'un ou de quelque chose

Tous les graphiques n'émuleront pas le besoin d'interprétation - l'analyste de données doit générer celui qui peut créer des **débats**.

Les graphiques doivent être **fabriqués** - par exemple, utilisez le titre du graphique encadré comme "question d'ouverture"...

Seule une équipe multifonctionnelle est capable d'expliquer des **motifs inattendus**, de remettre en cause ou de réviser des **hypothèses existantes**, ou d'identifier des preuves justifiant un **Appel à l'action**.



# Étape 3- Notebook pour communiquer avec les données : Microsite

## D'une hypothèses vers une déclaration fondée sur des preuves

Les données doivent soutenir Narrative - et non l'inverse !

Tirez parti de la narration avec des données pour :

- Expliquer,
- Éclairer,
- Engager



# Conclusion

# R en contexte humanitaire

## Vous n'êtes pas seul

Plus de 450 utilisateurs de plusieurs organisations dans le [humanitarian-useR-group](#)

Environ déjà ≈20 R champions au sein de l'UNHCR contre plus de 420 utilisateurs de PowerBI Pro

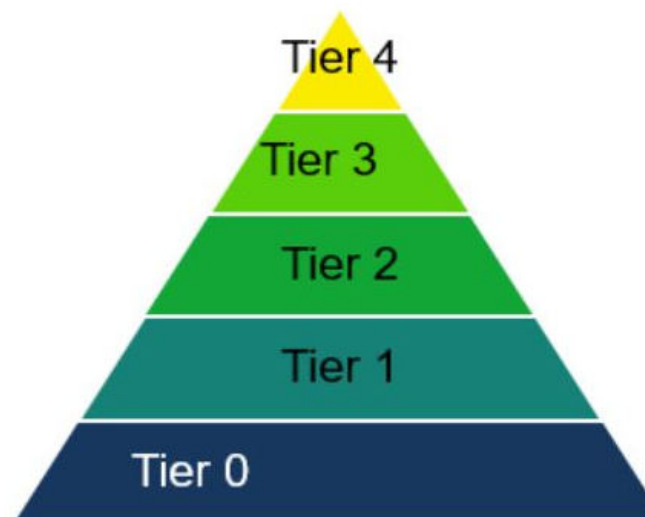
Essayez de commencer en utilisant les packages UNHCR existants et partez d'un projet que vous pouvez reproduire



# Un appel à l'institutionnalisation

Utilisation du modèle de support informatique standard à plusieurs niveaux...

- Niveau 4 : Examen du code et assurance qualité / Entreprise sous contrat avec un accord-cadre mondial
- Niveau 3 : Développement de packages internes / Équipe de champions R internes (coût : une réunion Rdev annuelle pour inciter le personnel contributeur)
- Niveau 2 : Induction des utilisateurs & Support utilisateur avancé / Global Data Service/DIMA (Data Science Team)
- Niveau 1 : Dépannage de base de l'utilisateur/Centre de services mondial (WIPRO selon le scénario documenté)
- Niveau 0 : Auto-support / Documentation du package (maintenue et améliorée en continu)



Votre opinion compte

Veillez **remplir ce sondage**  
pour partager votre opinion et  
vos réflexions sur le sujet  
présenté ici

Slides are available in [English](#) & [French](#) - slides note can be displayed by pressing keyboard shortcut p - and in PDF ([English](#) & [French](#)) within this [github repo](#)