

# La main à la pâte : une analyse répliquable

Konrad Hinsen

CBM, CNRS Orléans et Synchrotron SOLEIL

`konrad.hinsen@cnrs.fr`

30 août 2018

# Outline

M3-S1 : Une analyse répliquable, c'est quoi ?

M3-S2 : Étude de cas : l'incidence de syndromes grippaux

M3-S3A : Importer les données (Jupyter)

M3-S3B : Importer les données (RStudio)

M3-S3C : Importer les données (OrgMode)

M3-S4A/B/C : Vérification et inspection

M3-S5A/B/C : Questions et réponses

### 3. La main à la pâte : une analyse répliquable

1. Une analyse répliquable, c'est quoi ?
2. Étude de cas : l'incidence de syndromes grippaux
3. Importer les données. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode
4. Vérification et inspection. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode
5. Questions et réponses. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode

# L'analyse de données traditionnelle

résumé  
méthodologique

résultats

discussion

## L'analyse de données répliquable

code

explication

résultats

discussion

# Pourquoi faire répliquable ?

- ▶ Facile à refaire si les données changent
- ▶ Facile à modifier
- ▶ Facile à inspecter et vérifier

### 3. La main à la pâte : une analyse répliquable

1. Une analyse répliquable, c'est quoi ?
2. Étude de cas : l'incidence de syndromes grippaux
3. Importer les données. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode
4. Vérification et inspection. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode
5. Questions et réponses. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode

## Les points clés à retenir

- ▶ Aucune modification des données "à la main".
- ▶ Du code pour tout !

### 3. La main à la pâte : une analyse répliquable

1. Une analyse répliquable, c'est quoi ?
2. Étude de cas : l'incidence de syndromes grippaux
3. **Importer les données.** Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode
4. Vérification et inspection. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode
5. Questions et réponses. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode

# Choix techniques

- ▶ Notebook Jupyter
- ▶ Langage Python 3
- ▶ Bibliothèques :
  - ▶ pandas
  - ▶ matplotlib
  - ▶ isoweek

## Les points clés à retenir

- ▶ Lecture des données directement de la source
- ▶ Gestion des données manquantes

### 3. La main à la pâte : une analyse répliquable

1. Une analyse répliquable, c'est quoi ?
2. Étude de cas : l'incidence de syndromes grippaux
3. **Importer les données**. Au choix :
  - ▶ Jupyter
  - ▶ **RStudio**
  - ▶ OrgMode
4. Vérification et inspection. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode
5. Questions et réponses. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode

# Choix techniques

- ▶ Environnement RStudio
- ▶ Langage R
- ▶ Bibliothèque : parsedate

## Les points clés à retenir

- ▶ Lecture des données directement de la source
- ▶ Attention aux données manquantes

### 3. La main à la pâte : une analyse répliquable

1. Une analyse répliquable, c'est quoi ?
2. Étude de cas : l'incidence de syndromes grippaux
3. **Importer les données**. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ **OrgMode**
4. Vérification et inspection. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode
5. Questions et réponses. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode

# Choix techniques

- ▶ Environnement Environnement Emacs + Org mode
- ▶ Langages :
  - ▶ Python 3 pour préparer les données
  - ▶ R pour l'analyse

## Les points clés à retenir

- ▶ Lecture des données directement de la source
- ▶ Gestion des données manquantes

### 3. La main à la pâte : une analyse répliquable

1. Une analyse répliquable, c'est quoi ?
2. Étude de cas : l'incidence de syndromes grippaux
3. Importer les données. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode
4. **Vérification et inspection.** Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode
5. Questions et réponses. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode

# Les points clés à retenir

- ▶ Pré-traitement des données
  - ▶ Adapter aux conventions des logiciels
  - ▶ Faciliter l'analyse
- ▶ Vérifier autant que possible
  - ▶ Inspection visuelle
  - ▶ Code de validation

### 3. La main à la pâte : une analyse répliquable

1. Une analyse répliquable, c'est quoi ?
2. Étude de cas : l'incidence de syndromes grippaux
3. Importer les données. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode
4. Vérification et inspection. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode
5. Questions et réponses. Au choix :
  - ▶ Jupyter
  - ▶ RStudio
  - ▶ OrgMode

# Questions

1. Quelles années ont connu les épidémies les plus fortes ?
2. Quelle est la fréquence d'épidémies faibles, moyennes, et fortes ?

## Les points clés à retenir

- ▶ Une analyse répliquable doit contenir **toutes les étapes** de traitement des données sous une forme **exécutable**.
- ▶ Il est important d'**expliquer** tous les choix qui peuvent influencer les résultats.
- ▶ Ceci nécessite d'exposer beaucoup de **détails techniques**, parce que c'est à ce niveau qu'on fait **le plus d'erreurs** !