

R For Data Science



Hani Bounoua

4 Janvier 2020

R-Ladies Algiers meet-up

Sommaire

- Qu' est ce que la Data Science?
- Quels sont les outils de R qui peuvent répondre a un tel besoin?



Qu' est ce que la Data Science?

Qu'est ce que la Data Science?

“ Data Science is using **Data** to create as much **Impact** as possible for your company.”

Jonathan Ma



Qu' est ce que la Data Science?

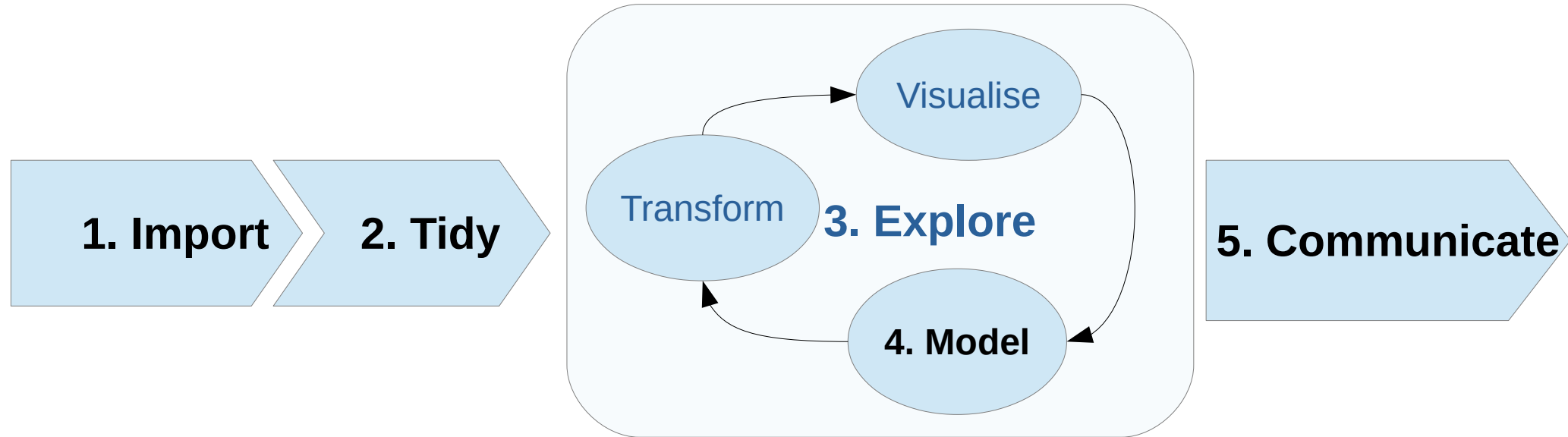
- **Data:**

- Tabulaire.
- Documents
- Images.
- Videos.
- Json.

- **Impacte:**

- Perspective de l'entreprise.
- Financiers
- Produits (Services).

Processus





Quels sont les outils de R qui peuvent répondre à un tel besoin?

Environnement de développement:

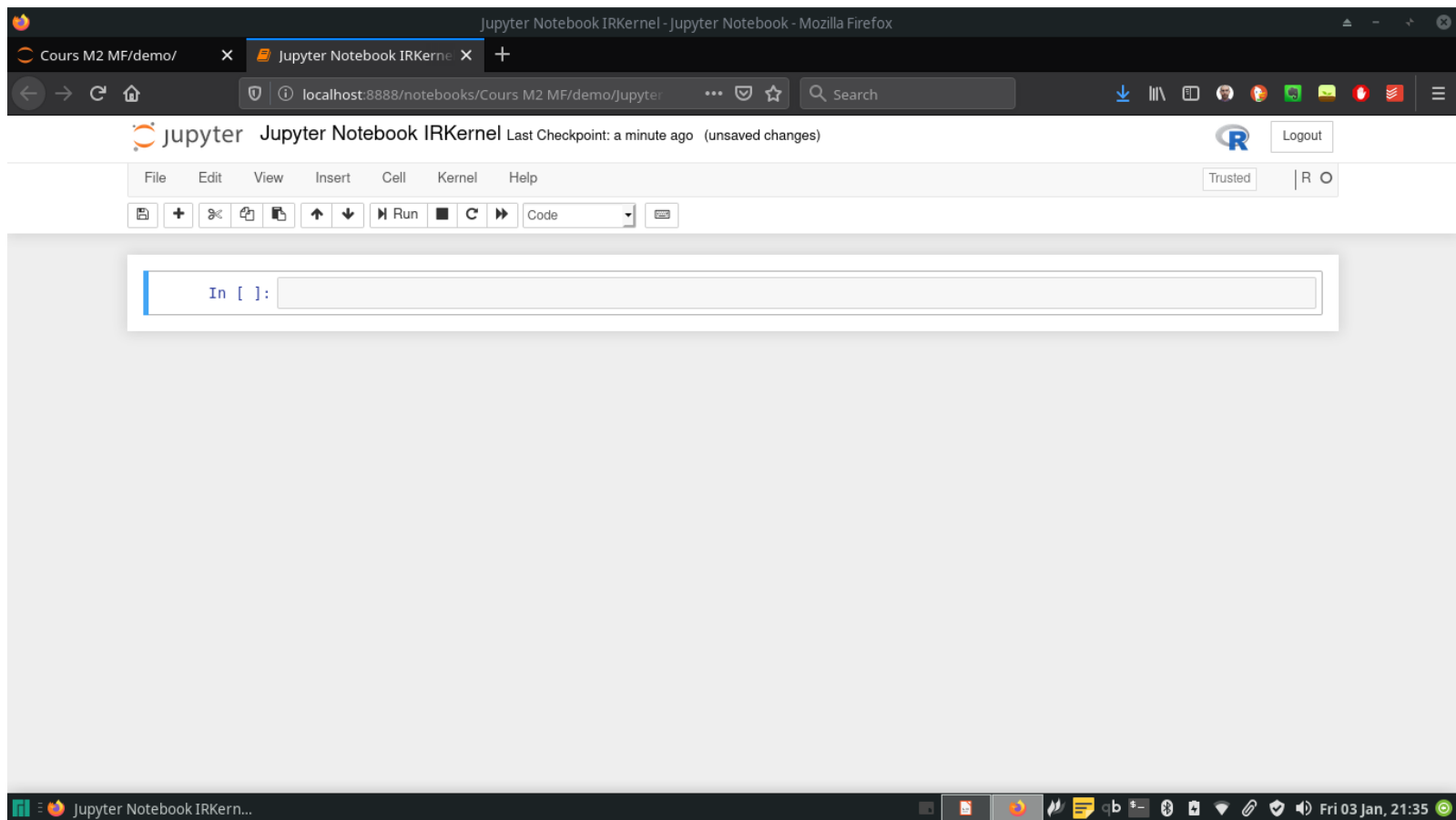
1. Jupyter Notebook.



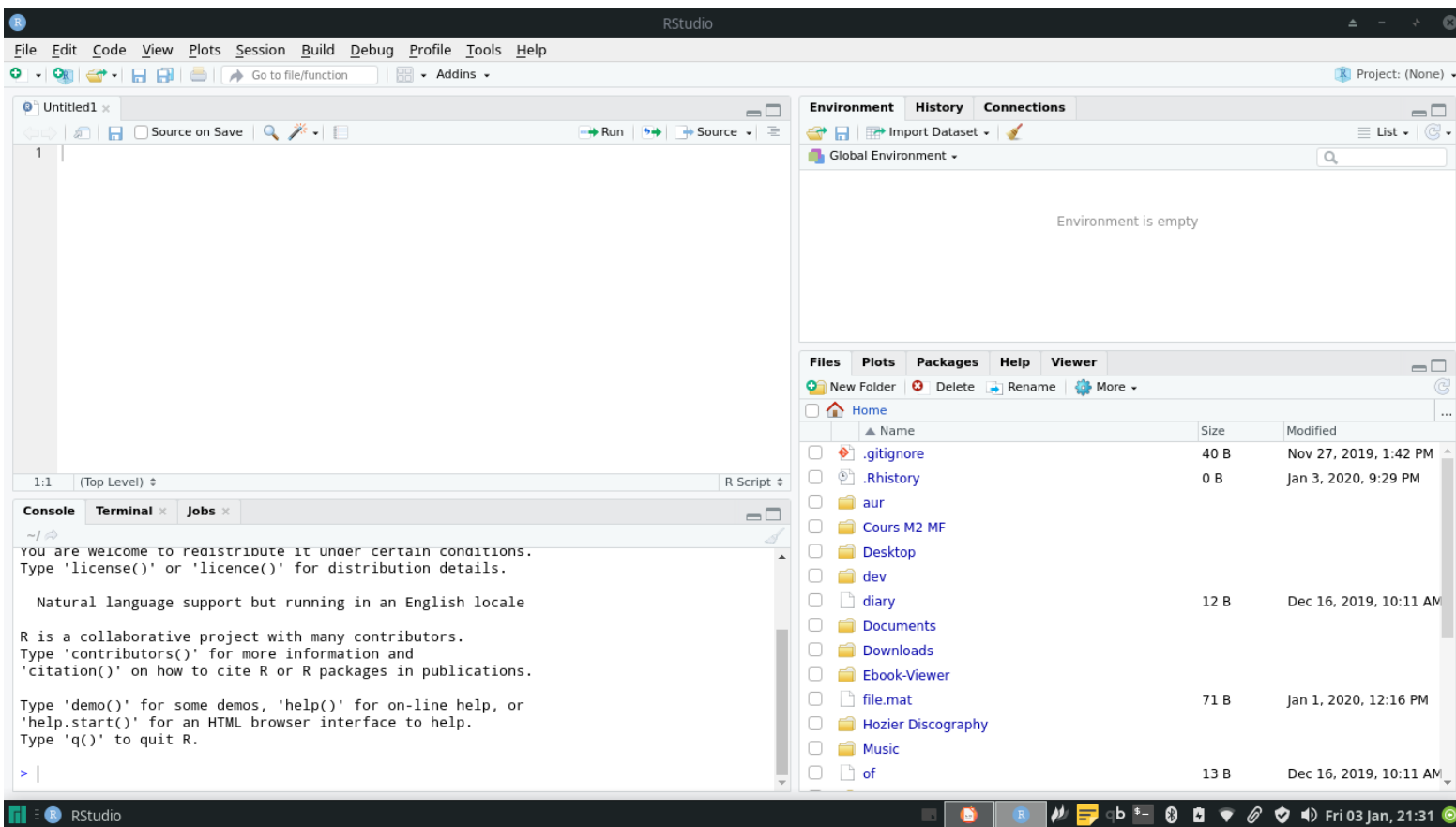
2. Rstudio.



Jupyter Notebook



RStudio

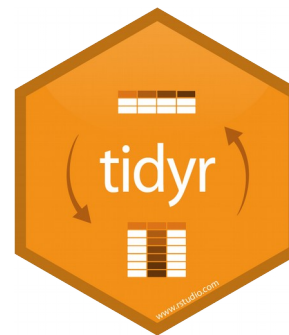


1. Importer les Données:

- **readr**: csv, tsv...
- **readxl**: xls, xlsx.
- **haven**: SPSS, SAS, Stata.
- **jsonlite**: Json.
- **xml2**: XML
- ...etc.

2. Tidy Data:

Package: tidyr.



country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	666	20593360
Brazil	1999	3737	172006362
Brazil	2000	488	174504898
China	1999	21258	127201272
China	2000	1766	128012583

variables

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	666	20593360
Brazil	1999	3737	172006362
Brazil	2000	488	174504898
China	1999	21258	127201272
China	2000	1766	128012583

observations

country	year	cases	population
Afghanistan	99	745	19987071
Afghanistan	00	666	20593360
Brazil	99	3737	172006362
Brazil	00	488	174504898
China	99	21258	127201272
China	00	1766	128012583

values

3. Exploration:

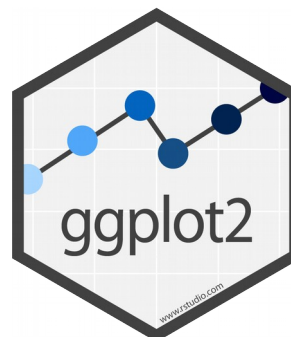
Transformation:

- **Package:** dplyr.



Visualisation:

- **Packages:** ggplot2, plotly.



4. Modelisation:

Package: caret

- Classification.
- Régression.

Autres packages:

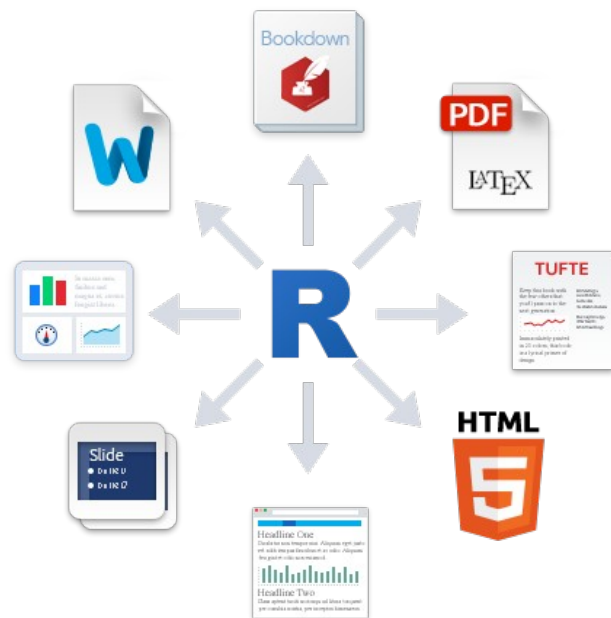
- a) **rpart**: Decision Tree.
- b) **e1071**: Support Vector Machine.
- c) **randomForest**: Random Forest.
- d) **FactoMineR**: PCA, HCA.
- e) **h2o**: Machine learning.
- f) **Keras**: Deep Learning.
- ...etc

5. Communication:

Package: shiny



Package: rmarkdown.



Références:

- R For Data Science: <https://r4ds.had.co.nz/index.html>
- Rstudio: <https://rstudio.com/>
- Rmakdown: <https://rmarkdown.rstudio.com/>
- Jonathan Ma: <https://twitter.com/jomaoppa>
<https://www.linkedin.com/in/jomaa>
<https://medium.com/@joma>



Merci pour votre attention.