

AVEZ-VOUS DÉJÀ ENTENDU (OU DIT) ?

**"J'avais écrit ces chiffres,
mais je ne sais plus d'où ils
viennent..."**

AVEZ-VOUS DÉJÀ ENTENDU (OU DIT) ?

**"J'avais écrit ces chiffres,
mais je ne sais plus d'où ils
viennent..."**

**"Je ne sais plus
comment cette
variable a été
construite..."**

AVEZ-VOUS DÉJÀ ENTENDU (OU DIT) ?

**"J'avais écrit ces chiffres,
mais je ne sais plus d'où ils
viennent..."**

**"Je ne comprends pas
pourquoi ça ne me donne
plus la même chose..."**

**"Que fait
exactement ce
programme?"**

**"Je ne sais plus
comment cette
variable a été
construite..."**

AVEZ-VOUS DÉJÀ ENTENDU (OU DIT) ?

**"J'avais écrit ces chiffres,
mais je ne sais plus d'où ils
viennent..."**

**"J'ai un peu la flemme
de tout refaire. Je
l'avais fait avant, donc
c'est juste..."**

**"Je ne comprends pas
pourquoi ça ne me donne
plus la même chose..."**

**"Que fait
exactement ce
programme?"**

**"Je ne sais plus
comment cette
variable a été
construite..."**

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ≡ ≡ ≡ ↺ 🔍 ↻

AVEZ-VOUS DÉJÀ ENTENDU (OU DIT) ?

"J'avais écrit ces chiffres, mais je ne sais plus d'où ils viennent..."

"Je n'y ai pas touché depuis 6 mois, je ne retrouve rien..."

"J'ai un peu la flemme de tout refaire. Je l'avais fait avant, donc c'est juste..."

"Je ne comprends pas pourquoi ça ne me donne plus la même chose..."

"Que fait exactement ce programme?"

"Je ne sais plus comment cette variable a été construite..."

"D'où sort ce tableau de résultats?"

AVEZ-VOUS DÉJÀ ENTENDU (OU DIT) ?

"J'avais écrit ces chiffres, mais je ne sais plus d'où ils viennent..."

"Je n'y ai pas touché depuis 6 mois, je ne retrouve rien..."

"J'ai un peu la flemme de tout refaire. Je l'avais fait avant, donc c'est juste..."

"Je ne comprends pas pourquoi ça ne me donne plus la même chose..."

"Que fait exactement ce programme?"

"Je ne sais plus comment cette variable a été construite..."

"D'où sort ce tableau de résultats?"

"Mince, mon étudiant vient de partir..."

CRISE DE LA REPRODUCTIBILITÉ

► Mondiale

CRISE DE LA REPRODUCTIBILITÉ

- ▶ Mondiale
- ▶ Toutes les disciplines sont concernées

CRISE DE LA REPRODUCTIBILITÉ

- ▶ Mondiale
- ▶ Toutes les disciplines sont concernées
- ▶ Nombreux résultats publiés ne sont pas reproductibles

CRISE DE LA REPRODUCTIBILITÉ

- En médecine : Ioannidis (2005), Horvath (2013), ... Flier (2022)

Why Most Published Research Findings Are False

John P.A. Ioannidis

CRISE DE LA REPRODUCTIBILITÉ EN ÉCONOMIE APPLIQUÉE ET ÉCONOMÉTRIE

- Les résultats ne sont pas (tous) reproductibles...

CRISE DE LA REPRODUCTIBILITÉ EN ÉCONOMIE APPLIQUÉE ET ÉCONOMÉTRIE

- ▶ Les résultats ne sont pas (tous) reproductibles...
- ▶ Dewald et al. (1988) ont essayé de reproduire **54** articles et ont réussi à reproduire les résultats de...

CRISE DE LA REPRODUCTIBILITÉ EN ÉCONOMIE APPLIQUÉE ET ÉCONOMÉTRIE

- ▶ Les résultats ne sont pas (tous) reproductibles...
- ▶ Dewald et al. (1988) ont essayé de reproduire **54** articles et ont réussi à reproduire les résultats de...

2!

(articles issus du "*Journal of Money, Credit and Banking*")

- ▶ McCullough en déduit en 2009 : "*ECONOMIC RESEARCH IS NOT REPLICABLE*"

CRISE DE LA REPRODUCTIBILITÉ EN ÉCONOMIE APPLIQUÉE ET ÉCONOMÉTRIE

Chang and Li (2015) ont essayé de répliquer 67 papiers. Ils ont pu répliquer :

- ▶ les principaux résultats (qualitatifs) de 22 des 67 papiers (33%) **sans contacter les auteurs**,
- ▶ 29 des 59 (49%) **avec l'aide des auteurs** *(en enlevant données confidentielles et logiciels non disponibles)*

"we assert that economics research is usually not replicable"

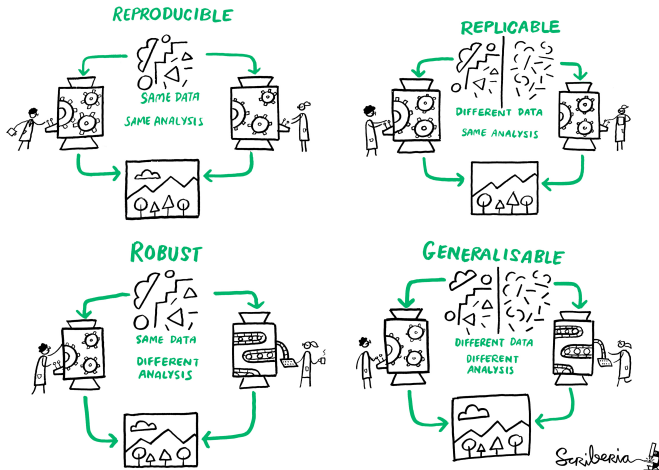
CRISE DE LA REPRODUCTIBILITÉ EN ÉCONOMIE APPLIQUÉE ET ÉCONOMÉTRIE

D'autres tentatives plus récentes :

Travaux de :	A partir de :	Reproduction
Gertler et al. (2018)	415 articles 9 journaux en économie	14%
Wood et al. (2018)	109 articles 10 journaux	25 %
Herbert et al. (2021)	303 articles American Economic Journal Applied Economics	38%

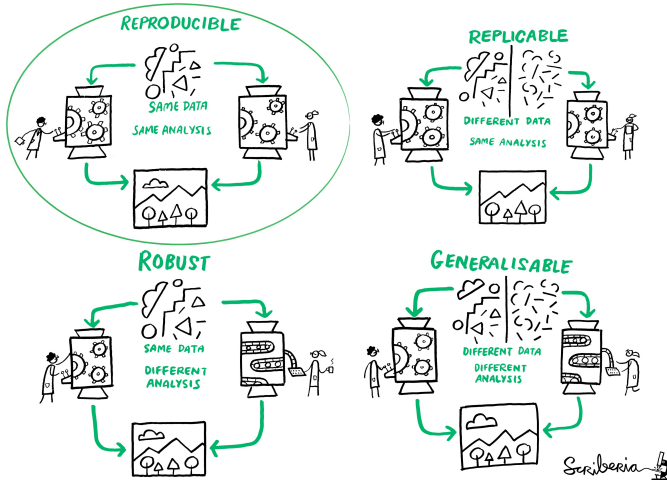
“RECHERCHE REPRODUCTIBLE”...

DE QUOI PARLE T-ON ?



MA PRÉSENTATION

► Reproductibilité



[CC-BY The Turing Way Community and Scriberia, 2020]

MA PRÉSENTATION

► Reproductibilité

↔ Au sens : même data + même code → mêmes résultats

MA PRÉSENTATION

- ▶ Reproductibilité
- ↔ Au sens : même data + même code → mêmes résultats
- ▶ Retour d'expériences ...

MA PRÉSENTATION

- ▶ Reproductibilité
- ↔ Au sens : même data + même code → mêmes résultats
- ▶ Retour d'expériences ...
 - ... en Economie appliquée / Econométrie

POURQUOI ON EN EST LÀ ?... NOTRE MÉTIER A CHANGÉ !

Plus de code

- Baisse des papiers théoriques : de **51%** en 1963 à **19%** en 2011 (dans le top 3 des revues en éco (AER, QJE, JPE), Hamermesh (2013))

POURQUOI ON EN EST LÀ ?... NOTRE MÉTIER A CHANGÉ !

Plus de code

- ▶ Baisse des papiers théoriques : de **51%** en 1963 à **19%** en 2011 (dans le top 3 des revues en éco (AER, QJE, JPE), Hamermesh (2013))
- ▶ Plus de données → plus d'analyses → plus de code !

POURQUOI ON EN EST LÀ?... NOTRE MÉTIER A CHANGÉ!

Plus de code

- ▶ Baisse des papiers théoriques : de **51%** en 1963 à **19%** en 2011 (dans le top 3 des revues en éco (AER, QJE, JPE), Hamermesh (2013))
- ▶ Plus de données → plus d'analyses → plus de code!
- ↪ Un chercheur/ingénieur qui fait de l'empirique est maintenant un programmeur!

POURQUOI ON EN EST LÀ ?... NOTRE MÉTIER A CHANGÉ !

Plus de code

- ▶ Baisse des papiers théoriques : de **51%** en 1963 à **19%** en 2011 (dans le top 3 des revues en éco (AER, QJE, JPE), Hamermesh (2013))
- ▶ Plus de données → plus d'analyses → plus de code!
- ↪ Un chercheur/ingénieur qui fait de l'empirique est maintenant un programmeur!
- ↪ Passe beaucoup de temps à écrire, à lire, et à débbuguer du code

POURQUOI ON EN EST LÀ ?... NOTRE MÉTIER A CHANGÉ !

Plus de code

“Constructing a computer program isn’t so different from constructing a formal proof” (LeVeque, 2009)

POURQUOI ON EN EST LÀ ?... NOTRE MÉTIER A CHANGÉ !

Plus de code

"Constructing a computer program isn't so different from constructing a formal proof" (LeVeque, 2009)

- ↪ Mais pas forcément les "bagages" (outils informatiques, pratiques) pour coder (Gentzkow and Shapiro (2014), Millman and Pérez (2014))
- ↪ Problèmes de :
 - ▶ lisibilité de code,
 - ▶ de non reproductibilité des résultats,
 - ▶ voire de génération d'erreurs

POURQUOI ON EN EST LÀ ?... NOTRE MÉTIER A CHANGÉ !

Plus de collaborations

Nombre d'auteurs par papier publié a augmenté :

- ▶ 1.3 dans les années 70's → 2.2 in 2012
(moyenne, top 5 des revues en économie, Card and DellaVigna (2013))
- ▶ Baisse des papiers rédigés à 1 seul auteur
50% en 1996 → 20% en 2014
- ▶ Forte augmentation des papiers à 3 voire + de 4 auteurs
(Kuld and O'Hagan (2017))

POURQUOI ON EN EST LÀ ?... NOTRE MÉTIER A CHANGÉ !

Plus de collaborations

Nombre d'auteurs par papier publié a augmenté :

- ▶ 1.3 dans les années 70's → 2.2 in 2012
(moyenne, top 5 des revues en économie, Card and DellaVigna (2013))
- ▶ Baisse des papiers rédigés à 1 seul auteur
50% en 1996 → 20% en 2014
- ▶ Forte augmentation des papiers à 3 voire + de 4 auteurs
(Kuld and O'Hagan (2017))

POURQUOI ON EN EST LÀ ?... NOTRE MÉTIER A CHANGÉ !

Plus de collaborations

Nombre d'auteurs par papier publié a augmenté :

- ▶ 1.3 dans les années 70's → 2.2 in 2012

(moyenne, top 5 des revues en économie, Card and DellaVigna (2013))

- ▶ Baisse des papiers rédigés à 1 seul auteur
50% en 1996 → 20% en 2014

- ▶ Forte augmentation des papiers à 3 voire + de 4 auteurs

(Kuld and O'Hagan (2017))

↪ ... besoin d'une bonne organisation pour partager,
échanger les fichiers, et échanger les informations (emails,
documentations)

UN CONTEXTE EN ÉVOLUTION...

Evolution du côté des revues

- Les revues ont développé leur politique en terme de données et code (Orozco et al., 2020)

UN CONTEXTE EN ÉVOLUTION...

Evolution du côté des revues

- Les revues ont développé leur politique en terme de données et code (Orozco et al., 2020)

Rank	Journal	Mandatory replication policy			
		2003	2009	2020	2020 policy* details
1	Am Econ Review	–	YES	YES	mandatory (data + code) sharing*
2	J Finance	–	–	YES	mandatory code, encourage data sharing*
3	Q J Economics	–	–	YES	mandatory (data + code) sharing*, AER
4	Econometrica	–	YES	YES	mandatory (data + code) sharing
5	J Financial Econ	–	–	–	encourage (data + code) sharing
6	J Political Econ	–	YES	YES	mandatory (data + code) sharing*, AER
7	Rev Financial Stud	–	–	–	–
8	J Econ Theory	–	–	–	encourage (data + code) sharing
9	Rev Econ Studies	–	YES	YES	mandatory (data + code) sharing
10	J Econometrics	–	–	–	encourage (data + code) sharing
11	J Econ Literature	–	–	YES	mandatory (data + code) sharing*
12	J Monetary Econ	–	–	–	encourage (data + code) sharing
13	J Econ Perspectives	–	YES	YES	mandatory (data + code) sharing*
14	Rev Econ & Stat	–	YES	YES	mandatory (data + code) sharing
15	Eur Econ Review	–	–	–	encourage (data + code) sharing
16	Int Econ Review	–	–	–	–
17	J Int Econ	–	–	–	mandatory data, encourage code sharing
18	Economic Journal	–	–	–	–
19	J Public Econ	–	–	–	encourage (data + code) sharing
20	Game Econ Behav	–	–	–	encourage (data + code) sharing
21	RAND J Economics	–	–	–	–
22	J Money Credit Bank	YES	YES	YES	mandatory (data + code) sharing*
23	Economic Theory	–	–	–	encourage data sharing
24	J Bus & Econ Stat	–	–	–	encourage (data + code) sharing
25	Economics Letters	–	–	–	mandatory data, encourage code sharing
...					
41	J Appl Econometrics	–	–	YES	mandatory data, encourage code sharing**
Specialized journals (data not available before 2019)					
	Eur Review of Agri Econ	n.a	–	YES	mandatory (data + code) sharing*
	Ecological Econ	n.a	–	–	encourage (data + code) sharing
	Am J of Agri Econ	n.a	–	YES	mandatory (data + code) sharing
	Food Policy	n.a	–	–	encourage (data + code) sharing
	Applied Econ	n.a	–	–	encourage data sharing
	Resource and Energy Econ	n.a	–	–	encourage (data + code) sharing

UN CONTEXTE EN ÉVOLUTION...

Evolution du côté des revues

- ▶ Les revues ont développé leur politique en terme de données et code (Orozco et al., 2020)
 - ▶ Certaines vérifient la reproductibilité des résultats! (par ex. *AJPS*, *AEA*)
 - ▶ Rôle de *data editor* (par ex. *Review of Economic Studies*, ...)

UN CONTEXTE EN ÉVOLUTION...OPEN SCIENCE

Evolution du côté des revues

De nouveaux supports de publications (revues ouvertes :
PLOS, arXiv, ...)

UN CONTEXTE EN ÉVOLUTION...OPEN SCIENCE

Evolution du côté des revues

De nouveaux supports de publications (revues ouvertes : *PLOS*, *arXiv*, ...)

Certaines exigent les matériels et la reproductibilité des résultats, par ex. *Computo* , journal de la SFDS, créé fin 2021

UN CONTEXTE EN ÉVOLUTION...OPEN SCIENCE

Valorisations de réplifications/reproductions

- ▶ Certaines revues publient des réplifications
 - ▶ par ex. *Journal of Applied Econometrics* (JAE) depuis 2003
 - ▶ par ex. *Spatial Economic Analysis* depuis 2022

UN CONTEXTE EN ÉVOLUTION...OPEN SCIENCE

Valorisations de réplifications/reproductions

- ▶ Des revues spécialisées dans la publication de réplifications :
 - ▶ *ReScience*, créé en 2015
 - ▶ *International Journal for Re-Views in Empirical Economics (IREE)*, créé en 2017

UN CONTEXTE EN ÉVOLUTION...OPEN SCIENCE

Valorisations de répliques/reproductions

► Un wiki autour de la réplication en économie existe

http://replication.uni-goettingen.de/wiki/index.php/Main_Page

Debt to GDP Ratios and Growth: Country Heterogeneity and Reverse Causation, the Case of Japan (Ultra Wonky) (NEP 2013)

Contents [hide]

- 1 Article
- 2 Article information
- 3 This article is a replication of
- 4 References



Article

Authors	Title	Journal	Year	Edition	Pages	JEL Codes ^a	Keywords
Matthew Berg, Brian Hartley	Debt to GDP Ratios and Growth: Country Heterogeneity and Reverse Causation, the Case of Japan (Ultra World) ^b	NEP	2013	-	573 - 578	-	-

Article information

Program code	Data	Readme	Method(s) & estimation	Data type	Data used	Origin of data used	Software used (Version)
0 - not available online	0 - not available online	0 - not available online	Least squares regressions, distributed lag models	Macro data	Government debt and GDP ratios	Japan	Stata

This article is a replication of

Authors	Title	Journal	Year	Edition	Pages	JEL Codes	Keywords	Replication type	Replication result [refer to replication type 1 and 2]	Raw data	Call into question	Author statement
Carmen Reinhart, Kenneth Rogoff	Growth in a Time of Debt	AER	2010	1	-	E23, E31, H53, O11, O47	-	2 - wide sense (new data)	1 - successful	3 - final dataset can be exactly replicated from raw	2 - partly	1 - rejected results

UN CONTEXTE EN ÉVOLUTION...OPEN SCIENCE

Valorisations de répliques/reproductions

- Un wiki autour de la réplication en économie existe

http://replication.uni-goettingen.de/wiki/index.php/Main_Page

Semiparametric Value-At-Risk Estimation of Portfolios (Int J Re-Views in Emp Econ 2019)

Contents [hide]

1 Article
2 Article information
3 This article is a replication of
4 References

Replicate
result

Peer
Data

Article

Authors	Title	Journal or Re- Views in Eng Exam	Year	Edition	Pages	JEL Codes if	Keywords
Jehua Xu	Semiparametric Value-At-Risk Estimation of Portfolios		2019	-	1-20	C51, G61, G11, G17	-

Article information

Program code	Data	Readme	Method(s) & estimation	Data type	Data used	Origin of data used	Software used (Version)
1 - accessible on journal website	2 - available online elsewhere on journal website	-	-	flrccr	Datastream	-	MATLAB

This article is a replication of

This article is a replication of

Authors	Title	Journal	Year	Edition	Pages	JEL Codes ^a	Keywords	Replication type	Replication result (refer to replication type 1 and 2)	Raw data	Call into question	Authors' statement
Alexandre Dries	Semiparametric estimation of multi-asset portfolio risk ratio	J Banking & Finance	2014	-	366-400	G31, G61, G11, G17	-	2 - wide sense (new data)	1 - successful	2 - raw data can be used to replicate a dataset that is not substantially different from the final dataset used for the publication	0 - no	-

UN CONTEXTE EN ÉVOLUTION...OPEN SCIENCE

Valorisations de répliques/reproductions

► Un wiki autour de la réplication en économie existe

http://replication.uni-goettingen.de/wiki/index.php/Main_Page

Association between economic growth and early childhood undernutrition: evidence from 121 Demographic and Health Surveys from 36 low-income and middle-income countries (Lancet Global Health 2014)

Contents [hide]

- 1 Article
- 2 Article information
- 3 Replication of this study
- 4 References for the replicated study



Article

Authors	Title	Journal	Year	Edition	Pages	JEL Codes#	Keywords
Selvanathan Velmer, Kenneth Harligen, Malakia A. Subramanyam, Jocelyn Prasad, Stephen Kissen, S V Subramanyam	Association between economic growth and early childhood undernutrition: evidence from 121 Demographic and Health Surveys from 36 low-income and middle-income countries#	Lancet Global Health	2014	4	225-34	-	-

Article information

Program code	Data	Readme	Method(s) & estimation	Data type	Data used	Origin of data used	Software used (Version)
0 - not available online	0 - not available online	0 - not available online	Clustering	-	Demographic and health Surveys	36 low-income and middle-income countries	-

Replication of this study

Authors	Title	Journal	Year	Edition	Pages	JEL Codes	Keywords	Replication type	Replication result (refer to replication type 1 and 2)	Raw data	Call into question	Authors' statements
Stephen A O'Connell, Caroline Smith	Economic growth and child undernutrition	Lancet Global Health	2016	12	961–2	-	-	-	2 - different results	-	1 - yes	1 - rejected results

UNE RIGUEUR EXIGÉE PAR NOTRE PROFESSION!

► Intégrité scientifique

**MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE
ET DE L'INNOVATION**

Décret n° 2021-1572 du 3 décembre 2021 relatif au respect des exigences de l'intégrité scientifique par les établissements publics contribuant au service public de la recherche et les fondations reconnues d'utilité publique ayant pour activité principale la recherche publique

...

Décrète :

Art. 1^{er}. – L'intégrité scientifique mentionnée à l'article L. 211-2 du code de la recherche se définit comme l'ensemble des règles et valeurs qui doivent régir les activités de recherche pour en garantir le caractère honnête et scientifiquement rigoureux.

Art. 2. – Les établissements publics et fondations reconnues d'utilité publique mentionnés au troisième alinéa de l'article L. 211-2 du code de la recherche :

1° Assurent la sensibilisation et la formation de leurs personnels au respect des exigences de l'intégrité scientifique, ainsi que de leurs étudiants dans le cadre de la formation à et par la recherche ;

2° Veillent à ce que l'organisation des travaux de recherche de leurs personnels soit menée dans le respect de ces exigences ;

3^e Promeuvent la diffusion des publications en accès ouvert et la mise à disposition des méthodes et protocoles, des données et des codes sources associés aux résultats de la recherche afin d'en garantir la traçabilité et la reproductibilité. Ils incitent à la publication des résultats de recherche dits négatifs ;

■■■

Art. 5. – Les résultats bruts des travaux scientifiques mentionnés au troisième alinéa de l'article L. 211-2 du code de la recherche sont constitués des données produites au cours du processus de recherche, ou à défaut des données traitées et scientifiquement validées, ainsi que des codes sources utilisés dans le traitement de ces données.

Extrait du décret MESRI relatif au respect des exigences de l'intégrité scientifique des établissements de recherche publique, décembre 2021.

UNE RIGUEUR EXIGÉE PAR NOTRE PROFESSION !

- ▶ Intégrité scientifique
- ▶ Mission intégrité scientifique depuis 2018 au CNRS
- ▶ Document de politique d'Intégrité scientifique INRAE, janvier 2021

UNE RIGUEUR EXIGÉE PAR NOTRE PROFESSION !

- ▶ Intégrité scientifique
- ▶ Mission intégrité scientifique depuis 2018 au CNRS
- ▶ Document de politique d'Intégrité scientifique INRAE, janvier 2021
- ▶ Charte de l'INED pour la science ouverte, 2022 :

"en application du **décret 2021-1572** relatif au respect de l'[intégrité scientifique](#) par les établissements de recherche... l'Ined s'engage à [promouvoir les méthodes et outils facilitant la répliquabilité, voire la reproductibilité](#), des recherches scientifiques ... L'enjeu est à la fois de rendre la recherche plus cumulative et d'accroître la [confiance](#) dans ses résultats."

UNE RIGUEUR EXIGÉE PAR NOTRE PROFESSION !

- ▶ Intégrité scientifique
- ▶ Chacun se doit de progresser (quel que soit son point de départ!)
- ▶ Et c'est possible!

OBJECTIFS :

- ▶ Proposer des principes et pratiques

OBJECTIFS :

- ▶ Proposer des principes et pratiques
- ▶ Montrer que des solutions simples existent

OBJECTIFS :

- ▶ Proposer des principes et pratiques
- ▶ Montrer que des solutions simples existent

Note that the software is not always the problem and will never be the solution. (Orozco et al., 2020)

HOW TO MAKE A PIE?



LE PROCESSUS DE RECHERCHE EST LONG !

- Le processus de recherche menant à un rapport/papier peut être long et complexe

LE PROCESSUS DE RECHERCHE EST LONG !

- ▶ Le processus de recherche menant à un rapport/papier peut être long et complexe
- ▶ Délais de publication longs :

17.70 mois (sd 7.52) en économie, 13.3 (sd 6.87) en mathématiques

(médecine 9.47, physique 1.93, arts et lettres 14.21, Björk and Solomon (2013))

LE PROCESSUS DE RECHERCHE EST LONG !

- ▶ Le processus de recherche menant à un rapport/papier peut être long et complexe
- ▶ Délais de publication longs :

17.70 mois (sd 7.52) en économie, 13.3 (sd 6.87) en mathématiques

(médecine 9.47, physique 1.93, arts et lettres 14.21, Björk and Solomon (2013))

- ▶ Publier prend du temps, le processus de recherche entier est encore plus long !

HOW TO MAKE A PIE ? BESOIN D'INGRÉDIENTS (DONNÉES)



HOW TO MAKE A PIE ? BESOIN DE LA RECETTE (CODE)

Phases techniques pour Mille-feuilles :



1 Pour réaliser cette recette de mille-feuilles, commencer par réaliser la pâte feuilletée inversée. Cette recette pourrait bien évidemment être réalisée avec une pâte feuilletée classique. Alors pourquoi j'ai choisi un feuilletage inversé ?



2 Cette pâte est un petit peu plus délicate, cependant tout à fait réalisable à partir du moment où l'on maîtrise la pâte feuilletée classique. Elle a l'avantage d'être plus rapide à préparer car elle comporte des tours doubles (expliqués dans la recette pas à pas). Elle gonfle plus à la cuisson et donne un feuilletage plus croustillant et fondant à la fois.



3 Une fois terminée, envelopper la pâte dans du papier film alimentaire et la laisser reposer au frais toute une nuit (c'est mieux si vous avez la possibilité de la préparer la veille).



4 Abaisser la pâte feuilletée inversée sur le plan de travail légèrement fleuré...



5 ...en une bande de 3 à 4 mm d'épaisseur. L'usage d'un laminoir peut faciliter la tâche autant dans la préparation de la pâte feuilletée en elle-même que dans son utilisation. Ici j'utilise un rouleau à pâtisserie en bois, ustensile plus couramment trouvé dans les cuisines.



6 Diviser la bande de pâte en 3 morceaux égaux.



7 Placer la première abaisse de pâte sur une plaque à pâtisserie recouverte d'une feuille de papier sulfurisé posée en deux par la moitié, puis réouverte.



8 Piquer la pâte avec un rouleau pique vite sur toute sa surface. Ceci évitera à la pâte de trop gonfler à la cuisson...



9 ...car il est important dans cette préparation d'obtenir des plaques de pâte feuilletée cuites pas trop épaisses, bien feuilletées pour garder le côté croustillant, et régulières.



10 Replier la feuille de papier sulfurisé, elle doit recouvrir la totalité de la pâte.



11 Poser sur le dessus une seconde plaque à pâtisserie, identique à la première. Elle fera pression sur la pâte et lui évitera ainsi de trop se développer à la cuisson.



12 Faire de même avec la seconde abaisse de pâte, en la plaçant sur une feuille de papier sulfurisé posée en deux, puis en la coquant entre 2 plaques. Idem avec la troisième abaisse de pâte. Nous nous retrouvons donc avec au total 4 plaques à pâtisserie emboîtées les unes sur les autres, entre lesquelles se trouvent 3 abaisses de pâte feuilletée piquées, elles-mêmes emprisonnées dans du papier sulfurisé.



13 Cuire ainsi à four chaud, 180°C (four air pulsé de préférence) pendant 20 minutes environ.



14 Au terme de la cuisson, les pâtes ont très peu levé...



15 ...retirer la plaque du dessus uniquement...



16 ...et séparer les 3 étapes...



17 ...de sorte que les 3 pâtes finissent de cuire individuellement une dizaine de minutes de plus. À ce stade, la pâte ne risque plus de gonfler, elle va juste finir de cuire. Notez que la pâte qui se trouvait sur le dessus sera cuite en premier.



18 Au terme de la cuisson, sortir les plaques du four et laisser refroidir.



19 Crème mousseline : Porter à ébullition le lait avec la gousse de vanille fendue en deux sur la longueur.



20 Dans un cul de poule, battre l'œuf entier.



21 Ajouter le sucre en poudre...

Images et recettes issues de : <https://www.meilleurduchef.com>

HOW TO MAKE A PIE ? BESOIN DE LA RECETTE (CODE)

 22 ...et blanchir la préparation au fouet.	 23 Ajouter la poudre à crème...	 24 ...et bien mélanger...
 25 ...jusqu'à son incorporation complète.	 26 Lorsque le lait entre en ébullition...	 27 ...Le verser en une seule fois sur les œufs blanchis...
 28 ...en remuant à l'aide d'un fouet.	 29 Transvaser l'appareil obtenu dans la casserole de cuisson du lait...	 30 ...et cuire la préparation en la fouettant constamment de sorte que la crème n'attache pas au fond de la casserole.
 31 Elle va se mettre à épaissir au bout de quelques secondes. Poursuivre la cuisson à feu modéré pendant 2 à 3 minutes.	 32 Au terme de la cuisson débarrasser la crème dans un récipient froid.	 33 Ajouter la moitié du beurre coupé en morceaux dans la crème chaude.
 34 Bien mélanger avec le fouet.	 35 Une fois la crème refroidie (à température ambiante), ajouter le restant de beurre pommade...	 36 ...en l'incorporant au batteur électrique ou au fouet à main. Réserver jusqu'au moment de l'utilisation.
 37 Découper les 3 pâtes feuilletées en carrés de dimensions identiques. Commencer par la première plaque que j'ai découpée ici en un carré de 20 x 20 cm.	 38 Positionner le carré obtenu sur la deuxième plaque et découper tout autour pour obtenir un second morceau aux mêmes dimensions. Faire de même avec la troisième plaque. L'usage d'un couteau à dents de type couteau-scie à genouise est recommandé pour scier la pâte et non l'écraser.	 39 On obtient ainsi nos 3 plaques de pâte feuilletée cuite, aux mêmes dimensions.
 40 Remplir une poche à douille munie d'une douille unie Ø 1,5 cm de crème mousseline. Pocher la crème sur la première plaque de pâte feuilletée en dressant des lignes parallèles, collées les unes aux autres.	 41 Poser sur le dessus une seconde plaque de pâte feuilletée.	 42 Bien ajuster les bords, les uns au dessus des autres.
 43 Pocher des cordons de crème mousseline sur la seconde plaque, de la même façon que l'étage inférieur.	 44 Veiller à bien les accoler les uns contre les autres, et ne pas laisser des espaces sans crème.	 45 Déposer délicatement la troisième plaque de pâte feuilletée sur la crème et presser légèrement pour bien faire adhérer.

HOW TO MAKE A PIE ? BESOIN DE LA RECETTE (CODE)

	49 Faire chauffer au bain-marie... ○○○○○○○○○		50 ...en remuant fréquemment à la spatule jusqu'à ce que le fondant devienne liquide. Il est important de ne pas dépasser la température de 35 à 40°C pour qu'il reste blanc et brillant.		51 Ajouter 1 ou 2 cuillères à soupe d'eau chaude.
	52 Bien mélanger.		53 Contrôler la température de fonte avec un thermomètre à visée laser, pas plus de 40°C.		54 Prélever la valeur de 2 cuillères à soupe de fondant...
	55 ...et le colorer avec un peu d'extraît de café (Trablit).		56 Bien mélanger.		57 Le verser dans un cornet à décor.
	58 Bien replier les bords...		59 ...afin d'éviter au fondant de s'échapper du cornet durant son utilisation.		60 Réserver dans un endroit tiède. Il faut que le fondant reste liquide.
	61 Verser le fondant blanc sur le mille-feuilles.		62 Le répartir sur toute sa surface.		63 L'étaler rapidement avec une spatule métallique coudeuse en une épaisseur régulière et fine autant que possible (2 à 3 mm).
	64 Veiller à étaler le fondant jusqu'aux bords. Il faut que ce travail soit effectué rapidement, avant que le fondant ne se solidifie.		65 Avec le cornet à décor, tracer des lignes parallèles espacées de 2 cm environ.		66 Ce travail doit également se faire rapidement, tant que le fondant est encore liquide.
	67 Tirer les lignes avec la pointe d'un couteau, pour terminer le décor.		68 Tirer toutes les lignes dans un sens pour commencer...		69 ...puis dans l'autre sens. Il faut qu'il y ait une ligne dans un sens, puis la suivante dans l'autre sens et ainsi de suite.
	70 Faire adhérer sur tout le pourtour du mille-feuilles des amandes caramélisées préalablement broyées grossièrement au couteau.		71 Bien les faire adhérer pour que les côtés soient entièrement recouverts, qu'il ne reste pas de vide.		72 Faire ainsi sur tout le pourtour du gâteau.
	73 Avant de le trancher, il est conseillé de le laisser se raffirmer une heure minimum au frais. Pour le couper, je conseille d'utiliser un couteau scie à génoise.		74 Il faut le couper en donnant des coups de scie d'avant en arrière. Il ne faut en aucun cas couper en appuyant verticalement sous risque d'écraser le mille-feuilles, de voir la crème sortir sur les côtés, et donc de perdre tout l'esthétique du gâteau.		

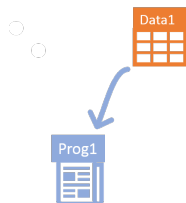
HOW TO MAKE A PIE ? EN ÉCO APPLIQUÉE/STAT...



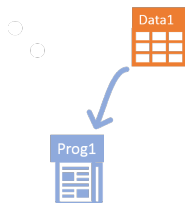
Data1		



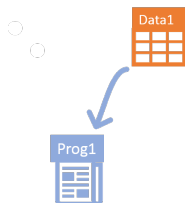
HOW TO MAKE A PIE ? EN ÉCO APPLIQUÉE/STAT...



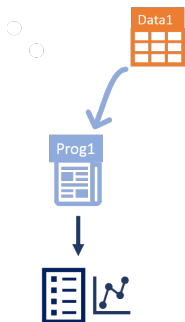
HOW TO MAKE A PIE ?



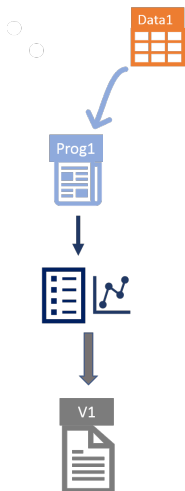
HOW TO MAKE A PIE ?



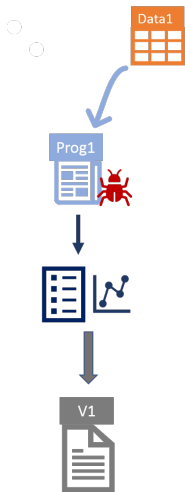
HOW TO MAKE A PIE ?



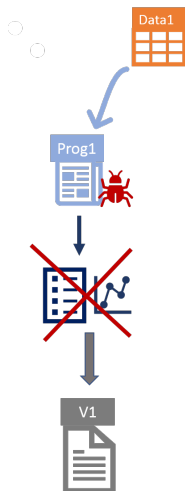
HOW TO MAKE A PIE ?



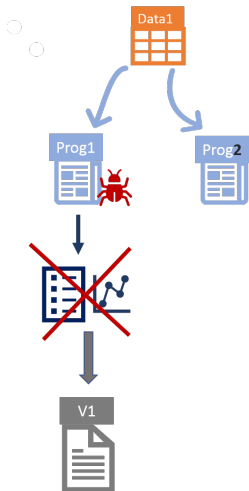
HOW TO MAKE A PIE ?



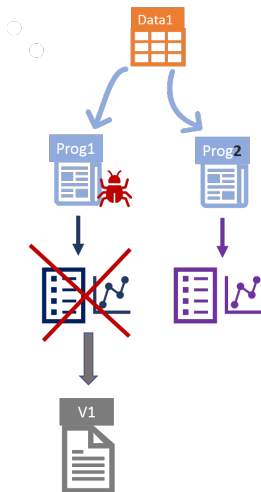
HOW TO MAKE A PIE ?



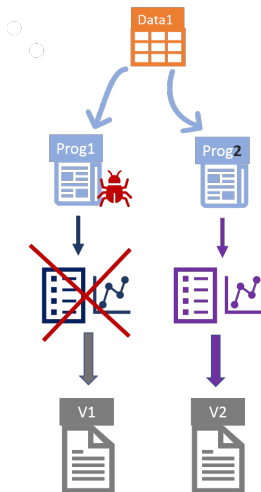
HOW TO MAKE A PIE ?



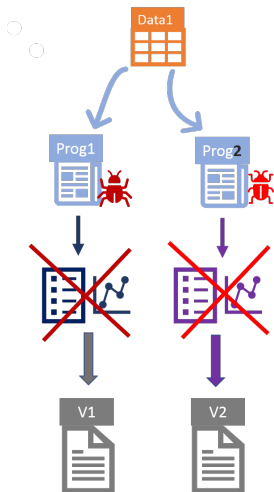
HOW TO MAKE A PIE ?



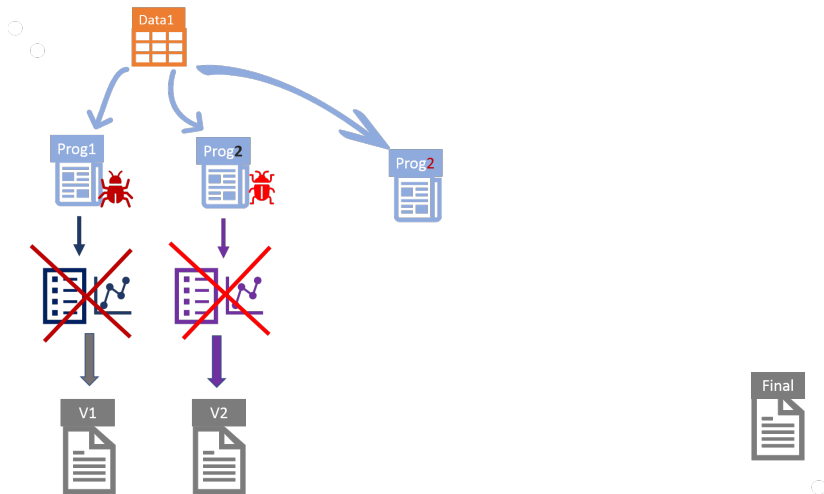
HOW TO MAKE A PIE ?



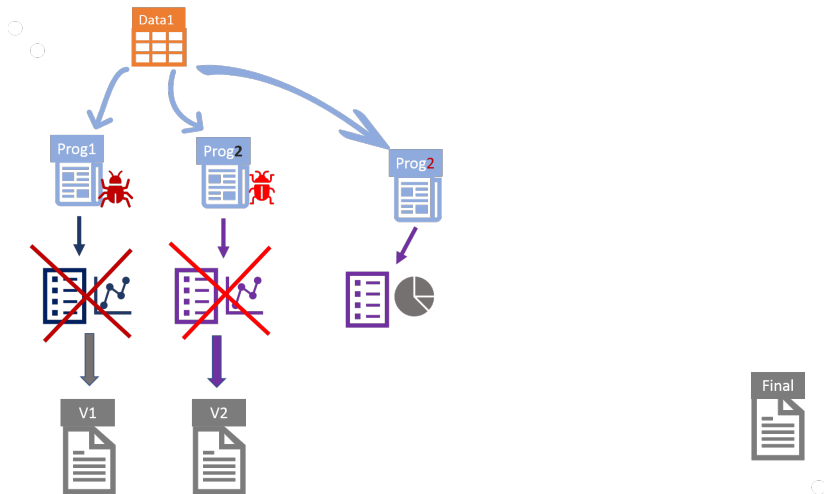
HOW TO MAKE A PIE ?



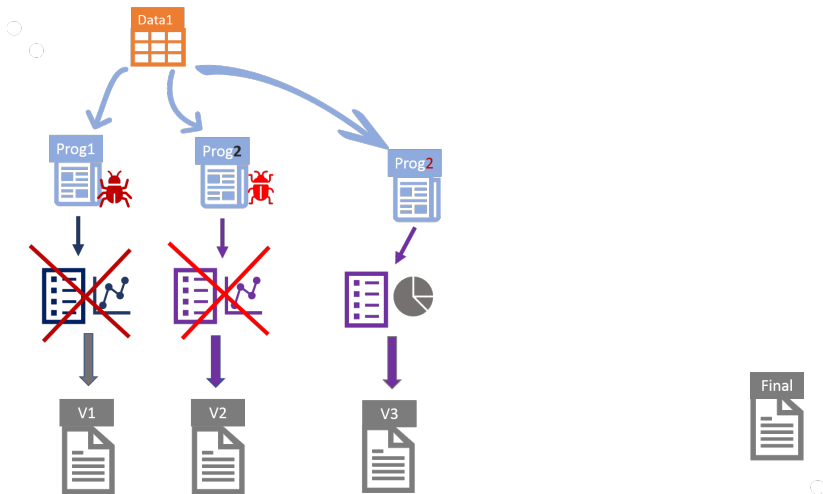
HOW TO MAKE A PIE ?



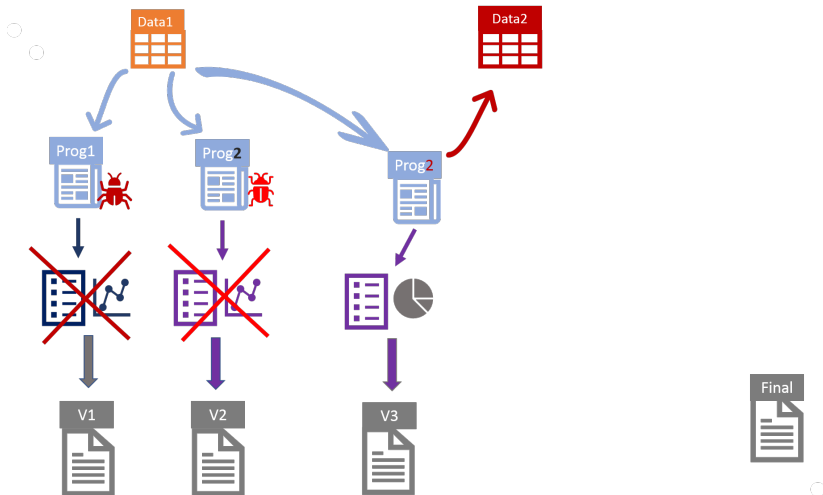
HOW TO MAKE A PIE?



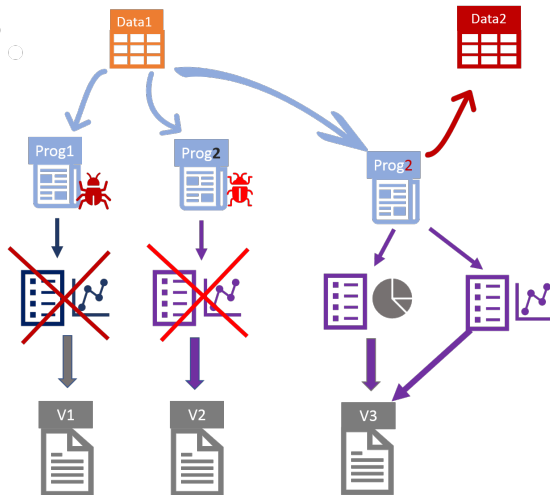
HOW TO MAKE A PIE?



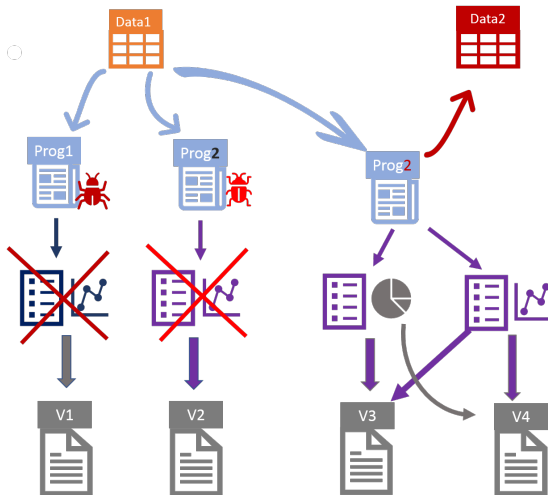
HOW TO MAKE A PIE?



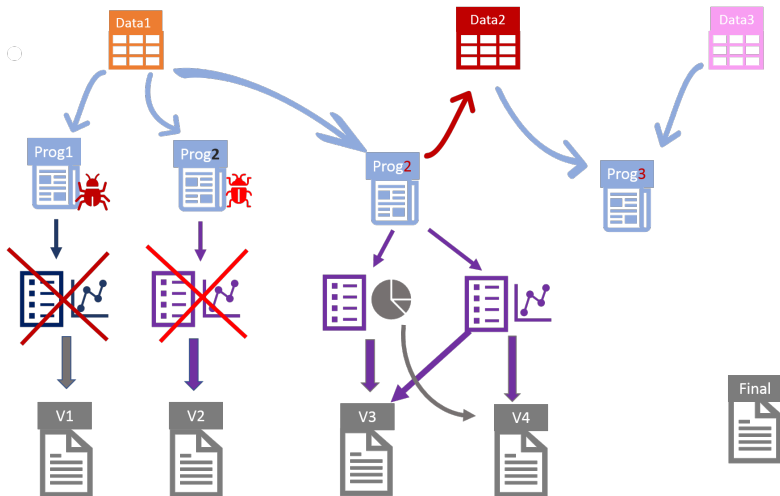
HOW TO MAKE A PIE ?



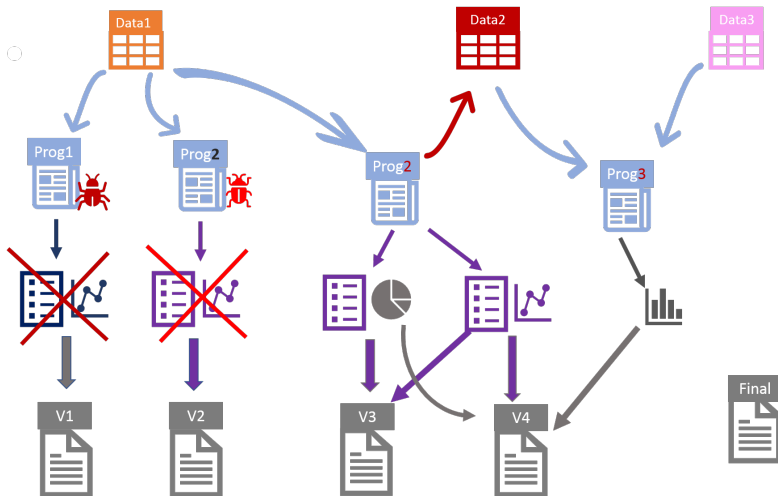
HOW TO MAKE A PIE?



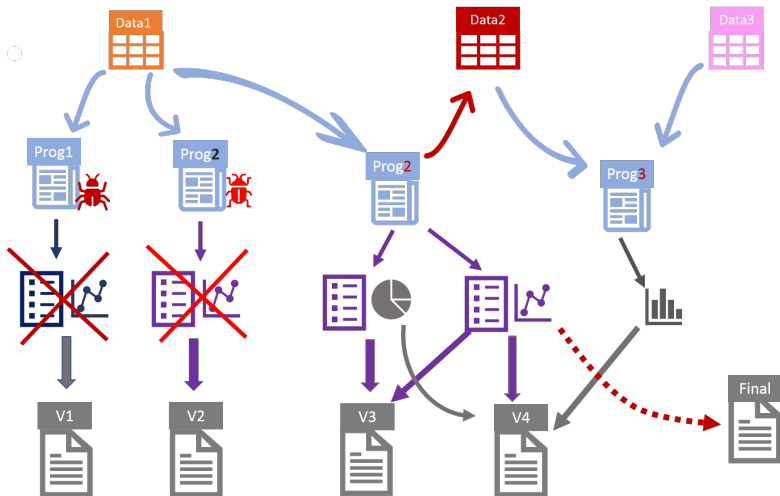
HOW TO MAKE A PIE?



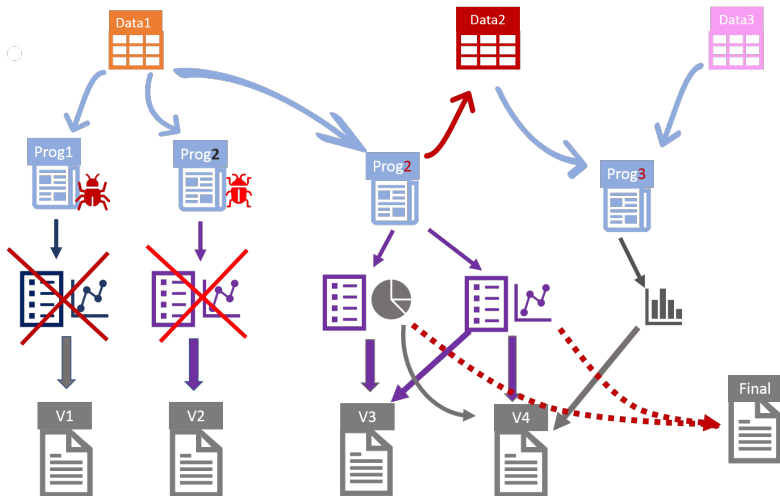
HOW TO MAKE A PIE?



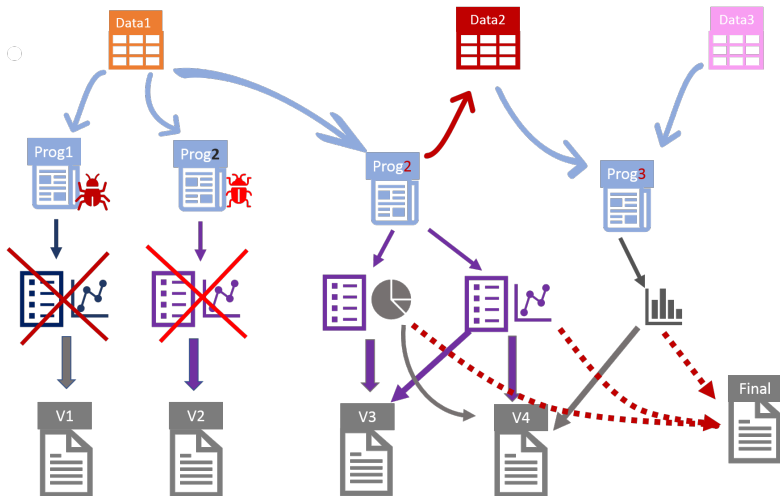
HOW TO MAKE A PIE?



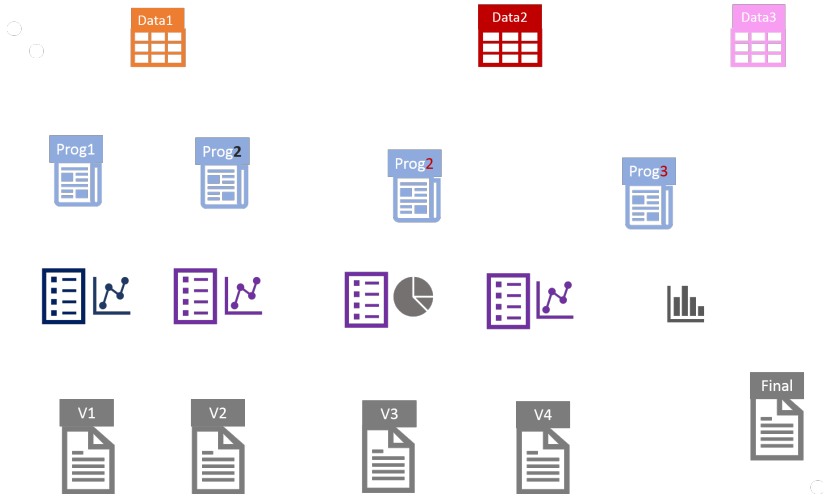
HOW TO MAKE A PIE?



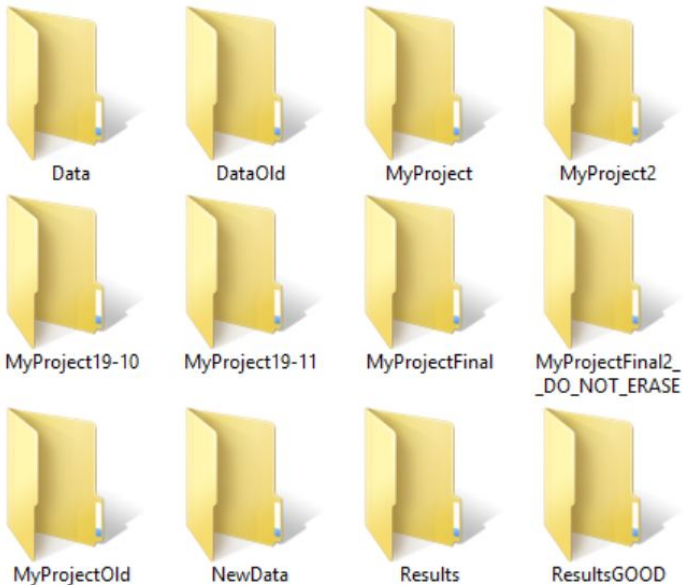
HOW TO MAKE A PIE?



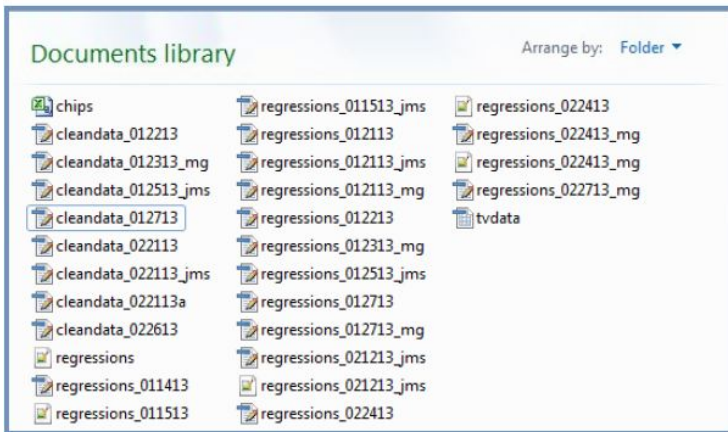
2 MOIS (6 ? 12 ?) PLUS TARD ...



2 MOIS (6 ? 12 ?) PLUS TARD ...



2 MOIS (6 ? 12 ?) PLUS TARD ...

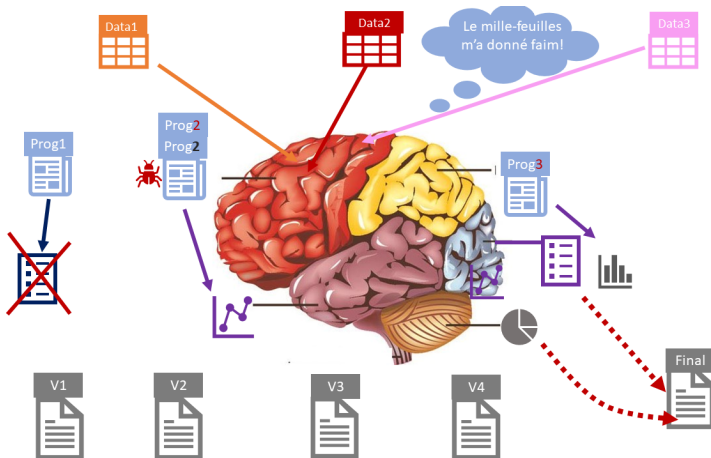


De notre propre expérience et Gentzkow and Shapiro (2014)

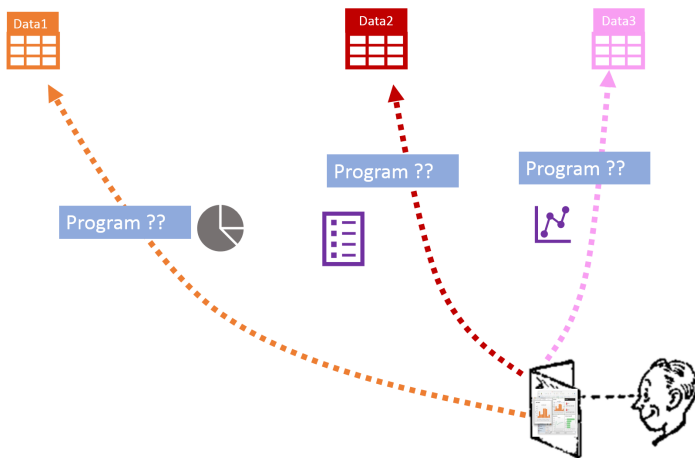
2 MOIS (6 ? 12 ?) PLUS TARD ...



CERTAINS DIRONT POUVOIR COMPTER SUR LEUR MÉMOIRE...



MAIS POUR LE LECTEUR : COMMENT DÉCODER ?



LE LECTEUR A SEULEMENT LE GÂTEAU (PAPIER) ET QUELQUES INGRÉDIENTS !



“RECHERCHE REPRODUCTIBLE”

- ▶ Mouvement né à Stanford dans les années 80 impulsé par *Jon Claerbout* qui a souligné :
 - ▶ Pénibilité de la reconstruction de travaux de recherche après un break,
 - ▶ Difficulté d'évaluer des travaux de recherche uniquement sur la base de la publication,
 - ▶ Importance de la réplication/ré-utilisation par d'autres
- ▶ Idée générale : les données, outils et programmes sont importants dans le processus de recherche

3 PRINCIPES :

1. Organiser son travail
2. Coder pour les autres (y compris pour soi)
3. Automatiser le plus possible

Appliqués selon le contexte (plusieurs co-auteurs, contraintes techniques,...)

PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitrisez toutes les étapes

- Garder/enregistrer chaque étape

PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitrisez toutes les étapes

- ▶ Garder/enregistrer chaque étape
- ▶ On évite les copiers/collers des résultats

PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitrisez toutes les étapes

- ▶ Garder/enregistrer chaque étape
- ▶ On évite les copiers/collers des résultats
- ▶ On évite les utilisations “clic bouton” de certains logiciels (la plupart permettent d’exporter du code!)

PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitrisez toutes les étapes

- ▶ Garder/enregistrer chaque étape
- ▶ On évite les copiers/collers des résultats
- ▶ On évite les utilisations “clic bouton” de certains logiciels (la plupart permettent d’exporter du code!)
- ▶ **Code!**

PRINCIPE 1 : ORGANISER SON TRAVAIL

Avoir une organisation claire des fichiers

- ▶ Séparer les fichiers : données (brutes, transformées), programmes, résultats et documentation

PRINCIPE 1 : ORGANISER SON TRAVAIL

Avoir une organisation claire des fichiers

- Séparer les fichiers : données (brutes, transformées), programmes, résultats et documentation

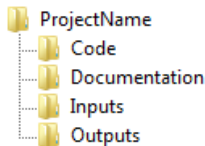


FIGURE – Exemple d’une arborescence bien organisée.

PRINCIPE 1 : ORGANISER SON TRAVAIL

Avoir une organisation claire des fichiers

- ▶ Séparer les fichiers : données (brutes, transformées), programmes, résultats et documentation
- ▶ Rendre les chemins portables (chemins relatifs)

PRINCIPE 1 : ORGANISER SON TRAVAIL

Avoir une organisation claire des fichiers

- ▶ Séparer les fichiers : données (brutes, transformées), programmes, résultats et documentation
- ▶ Rendre les chemins portables (chemins relatifs)

```
/**** Stata EXAMPLE ****/  
/**** Definition of the useful path ***/  
local CodeFolder "c:/ApplePie/Progs"  
  
/**** Positioning ***/  
cd "'CodeFolder'"  
  
/**** Using data that is in another folder ***/  
use ../Raw_Data/Sugar.dta, replace  
append using ../Raw_Data/Apple.dta  
  
save ../Final_Data/ApplePie.dta, replace  
qui log close
```

```
#### R EXAMPLE ####  
# Definition of the useful path  
CodeFolder <- "c:/ApplePie/Progs"  
GraphFolder <- "../Graphs/"  
  
# Positioning  
setwd(CodeFolder)  
  
# Saving the graph to another folder  
file <- paste(GraphFolder, "MySuperPie.png", sep="")  
png(filename = file)  
pie(rep(1,8), col=1:8)  
dev.off()
```

voir aussi package *here()*

PRINCIPE 1 : ORGANISER SON TRAVAIL

Utiliser des conventions de nommage (des fichiers)

- Pour le code : nom explicite, indication du workflow, (version)

PRINCIPE 1 : ORGANISER SON TRAVAIL

Utiliser des conventions de nommage (des fichiers)

- Pour le code : nom explicite, indication du workflow, (version)

Usual

```
prog1.do  
prog2.do  
test.R  
final.R
```

PRINCIPE 1 : ORGANISER SON TRAVAIL

Utiliser des conventions de nommage (des fichiers)

- Pour le code : nom explicite, indication du workflow, (version)

Usual

```
prog1.do  
prog2.do  
test.R  
final.R
```

Better

```
01_preparing_data.do  
02_stat_desc.do  
03_model1.R  
03_model2.R
```

PRINCIPE 1 : ORGANISER SON TRAVAIL

Utiliser des conventions de nommage (des fichiers)

- ▶ Pour le code : nom explicite, indication du workflow, (version)
- ▶ Pour les sorties (données créées, logs, tables, graphiques)

PRINCIPE 1 : ORGANISER SON TRAVAIL

Utiliser des conventions de nommage (des fichiers)

- ▶ Pour le code : nom explicite, indication du workflow, (version)
- ▶ Pour les sorties (données créées, logs, tables, graphiques)

Better

```
02_stats_desc_table.tex  
03_model1_table.tex  
03_model2_fig.JPG
```

PRINCIPE 1 : ORGANISER SON TRAVAIL

Utiliser des conventions de nommage (des fichiers)

- ▶ Pour le code : nom explicite, indication du workflow, (version)
- ▶ Pour les sorties (données créées, logs, tables, graphiques)
- ▶ Pour tous types de fichiers y compris les articles de bibliographie

PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitrisez toutes les étapes

- Garder/enregistrer chaque étape (pas de copier/coller, pas de clic bouton). Code!

PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitrisez toutes les étapes

- Garder/enregistrer chaque étape (pas de copier/coller, pas de clic bouton). Code!
- Maitriser le workflow

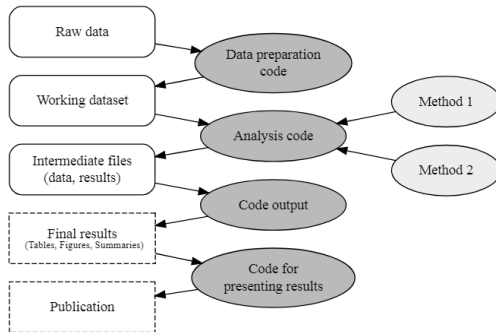


FIGURE – Un exemple simple de workflow

PRINCIPE 1 : ORGANISER SON TRAVAIL

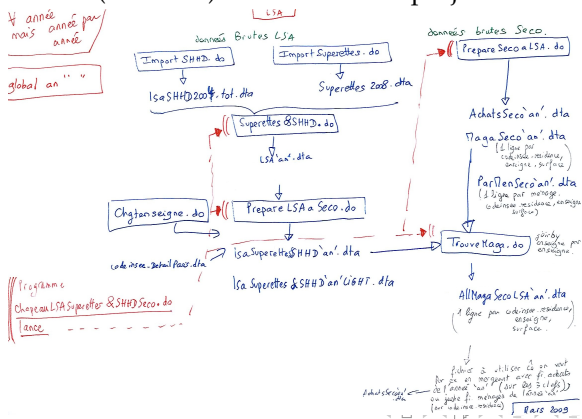
Maitrisez toutes les étapes

- ▶ Garder/enregistrer chaque étape (pas de copier/coller, pas de clic bouton). Code!
- ▶ Maitriser le workflow
 - ▶ Documenter (Dessiner) le workflow du projet

PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitrisez toutes les étapes

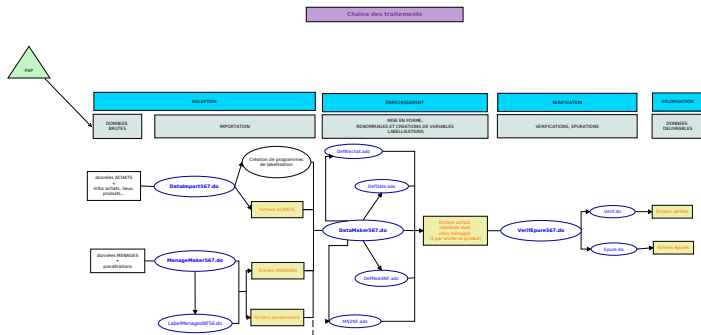
- ▶ Garder/enregistrer chaque étape (pas de copier/coller, pas de clic bouton). Code!
- ▶ Maîtriser le workflow
 - ▶ Documenter (Dessiner) le workflow du projet *A la main*



PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitrisez toutes les étapes

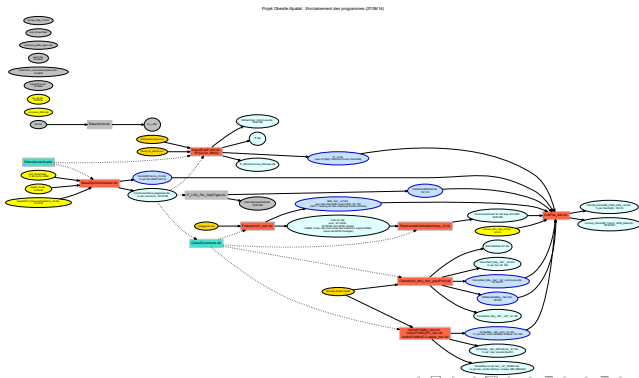
- ▶ Garder/enregistrer chaque étape (pas de copier/coller, pas de clic bouton). Code !
- ▶ Maitriser le workflow
 - ▶ Documenter (Dessiner) le workflow du projet Avec un logiciel (dessin) (Dia)



PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitrisez toutes les étapes

- Garder/enregistrer chaque étape (pas de copier/coller, pas de clic bouton). Code!
- Maitriser le workflow
 - Documenter (Dessiner) le workflow du projet Avec *GraphViz* (voir aussi package *RGraphViz*)



PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitriser les versions des codes

- Utiliser un système de gestion de versions

PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitriser les versions des codes

- ▶ Utiliser un système de gestion de versions
- ▶ Par exemple Git via les forges (*GitHub*), GitLab, SourceSup (Renater)

PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitriser les versions des codes

- ▶ Utiliser un système de gestion de versions
- ▶ Par exemple Git via les forges (*GitHub*), GitLab, SourceSup (Renater)
- ▶ “*Version control is like an undo command for everything*”

(Gentzkow and Shapiro, 2014)

The screenshot shows a Git web interface. On the left, a commit history list shows various commits by 'XtopheB' over the last year, including 'Loop to create Mfx tables with CI', 'Complete file with Bootstrapped MFX', 'With heterov specified', 'Document de compilation des résultats', 'Calcul de la fenetre CV', 'Sorties triées dans l'ordre croissant des fenêtres...', 'Sorties LaTeX included...', 'Sortie MFX', 'Avec MFX', 'Version complète des programmes', and 'VersionDebut Aout'. The 'With heterov specified' commit is selected. On the right, a diff view shows the changes in 'LoopSpecial.do'. The diff highlights several lines of Stata code, including a line that sets 'gen Special = -(prix_2008 - mprix_2008)' and another that sets 'set linesize 88'. The diff also shows changes to the 'quietly' command and the 'trim' function.

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

“Let us concentrate rather on explaining to humans what we want the computer to do”, (Knuth, 1984)

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

"Let us concentrate rather on explaining to humans what we want the computer to do", (Knuth, 1984)

- Utiliser des conventions de **layout** (commentaires, indentations, longueur de ligne...)

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

“Let us concentrate rather on explaining to humans what we want the computer to do”, (Knuth, 1984)

- ▶ Utiliser des conventions de **layout** (commentaires, indentations, longueur de ligne...)
 - ▶ *pylint* (PEP8 Python), packages R *check*, *formatR*

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

- Utiliser des conventions de **nommage** (des variables)

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

- Utiliser des conventions de **nommage** (des variables)

Usual

```
gen sex = 1 if gender == 101  
rep sex = 2 if gender == 102  
gdp <- gdp /1.02  
size <- 180
```

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

- Utiliser des conventions de **nommage** (des variables)

Usual

```
gen sex = 1 if gender == 101  
rep sex = 2 if gender == 102  
gdp <- gdp /1.02  
size <- 180
```

Better

```
gen I_Men = (gender == 101)  
  
GDP_New <- gdp /1.02  
Height <- 180
```


PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

- Bien **documenter** : environnement computationnel, données, code, toutes les hypothèses

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

- Bien **documenter** : environnement computationnel, données, code, toutes les hypothèses

En-tête des programmes

Program for pie cooking technology

Goal: Generate the Chocolate Foam estimations

Date: 2017/01/05

Author: Jamie Oliver

Running under R version 3.2.2 (2015-08-14)

Platform: x86_64-w64-mingw32/x64 (64-bit)

Input files: chocolate.csv, eggs.txt

Output: ChocolateFoam.R, ChocolateFoam.tex

Version 4 of the program: + function fct_coef_variation

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

- **Bien documenter** : environnement computationnel, données, code, toutes les hypothèses

(Si on n'utilise pas de système de versioning)

```
/* ===== */
/* ----- PROGRAMME DATAMAKERG ----- */
/* ===== */
/* Ce programme NE TRAITE PAS LES FICHIERS MENAGES (voir MenagesMaker567.do) */
/* DataMakerNF56 créé à partir de DataMakerNF345.do */
/* Version 1.1 29/01/08 sa3 = range.appellation et sa2 = range (pour cohérence avec les fichiers ancienne formule) via un rename */
/* Version 1.2 04/02/08 Initialisation des locales (NbAnPasObs AnSansObs) pour test sur les années manquantes (r(N)>0) */
/* Version 1.3 13/02/2008 Modification du fichier ProduitNF56.dta si produit sans obs. */
/* Version 1.4 7/03/2008 Initialisation locale varlistannee_sans_prem */
/* Version 1.41 26/03/2008 drop de an */
/* Version 1.42 22/04/2008 FusionMarques56 intégrée et MN2NF corrigée (moda de Leader Price) */
/* Version 2.0 17/12/2008 Nouvel envoi 2005 et 2006 */
/* Version 2.1 15/06/2009 2005 2006 et 2007 */
/* Version 2.2 8/06/2010 Fichier des caractéristiques produits (Product_Desc_1a0.txt) modifié pour 2005 */
/* (avant c'était le fichier 1er envoi 2005, now celui 2ème envoi 2003-2006) */
/* Version 2.3 25/08/2010 Pour nouvelles données 2006n, 2007n et 2008 */
/* correction labellisation fichier */
/* Version 3.0 27/07/11 ajout condition pr liste des produits : besoin que l'info panel soit renseignée */
/* ex: ds Produits678.dta , pdt 538 n'a pas d'info panel donc prog plante (je ne sais pas pourquoi il ne plantait pas avant...) (V) */
/* Version 3.1 30/08/11 : Changement de pu en Pu et qu en QU (conforme à notre règle typographique) */
/* Version 4.0 1/09/11 : version générique; chgt de boucle : 1 produit par année */
/* : liste des produits à partir du fichier ProduitsNFXXX.do DANS la boucle année */
/* Version 5.0 31/07/13 : adaptation aux fichiers 2011 créés par DataImportG11.do */
/* Version 5.1 07/08/13 : Correction d'un bug apparu en 2011 lors de la concaténation des périodes pour chaque produit */
/* Version 5.2 12/08/13 : des "capture" rajoutés pour éviter des pbs lorsqu'un produit n'a pas la variable sa2 (par ex produit 152) */
/* Version 5.3 27/08/13 : correction des modalités de tuva pour conversion de Qu (V) ; déplacement du lancement de "LabelVarProdCom6.do" une fois "LabelSPA'annee'.do" lancé */
/* (pour que Spel aient ses moda labellisées car ensuite Spel est renommé Encode) */
/* Version 5.4 31/03/15 : Variables sa2 et sa4 maintenant créées dans DataImportG11.do */
/* Version 5.5 2/04/15 : Locales rajoutées pour rendre le programme générique avant/après 2011 */
/* Version 5.6 8/04/15 : maj de la note sur Qu et Pu */
/* 7/08/15 : On lance sur 2013 */
/* Version 5.7 6/01/16 : rajout labellisation sa4 (à partir de 2011 son programme de labellisation est créé dans DataImportG11.do) (V.0) */
/* ===== */
local version "5.7"
```

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

- Bien **documenter** : environnement computationnel, données, code, toutes les hypothèses

Métadonnées associées à des fichiers (ici *Stata*)

```
. use J:\Secodip\Data2012\Produits\0005\p0005NF_E.dta
(5_PRODUITS SUCRANTS (0005) EPURE, annees (2012) ( 8 Apr 2015). PRIX EN EUROS)

. note

_dta:
1. Créé avec la version 2.3 de DataImportG11.do
2. Qu= qorig*gawa*pweigh ; ptwa=ptwa*gawa*pweight (donc Pu=ptwa_orig/quorig)
3. Programme de labelisation des ménages : version M3.1
4. Créé avec la version 3.2 de MenageMakerG.do
5. Version 2.1 de LabelVarProdComG.do
6. Version 3.0 de MN2NF.adoc
7. ATTENTION : 3 unités (tuwa) coexistent sur ce produit !!!!!
8. Créé avec la version 5.5 de DataMakerG.do
9. Créé avec la version 4.0 de VerifG.do
10. Créé avec la version 1.1 de EpureG.do
```

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

- **Bien documenter** : environnement computationnel, données, code, toutes les hypothèses

Fichier *"Read Me"*

Christophe Bontemps, Michel Simioni, and Yves Surry, "Semiparametric Hedonic Price Models: Assessing the Effects of Agricultural Nonpoint Source Pollution", *Journal of Applied Econometrics*, Vol. 23, No. 6, 2008, pp. 825-842.

All data are in the file `bss-data.txt`, which is an ASCII file in DOS format. It is zipped in the file `bss-data.zip`. Unix users should use `"unzip -a"`.

Data come from various sources described in section 4:

- A real estate database, known as MIN, maintained by the association of French notaries
- The French census, collected by the INSEE
- The regional branch (Brittany) of the French Ministry of Agriculture, Fisheries and Forestry.

The list of variables follows. A detailed explanation of the nature and sources of each of them is given in section 4 of the paper.

Variable	Description	Units
PRICE	Market Price (in log)	French Francs(x10000)
AGE	Age	Year
REPAIR	State of repair	=1 if good
ROOMS	Number of rooms	#
LOT	Lot size	m ² (x1000)
COUNTY	County location	=1 if "ille et Vilaine"
VACANT	Vacant Housing	Percentage
POP	County population	# (x1000)
AVINC	Average income	French Francs(x1000)
THEAD	Temporary meadows	Percentage
NITRO	Nitrogen concentration	kg/ha

NB: All the computations done in the paper used prices and average income expressed in French Francs. Results are expressed in Euros using the formula 1 Euro = 6.55957 French Francs (see for instance some summary statistics in Table 1).

The total data sample is of 2092 observations distributed over 516 districts with less than 5000 inhabitants (see Table 1 for summary statistics). The

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

- ▶ Bien **documenter** : environnement computationnel, données, code, toutes les hypothèses
- ▶ Utilisation de **cahiers de laboratoire numériques (ELN)** comme *Evernote*, *OneNote*, *eLabFTW*

cf rapport Comité pour la science ouverte (Mathieu et al., 2021)

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

Usual

```
coeff_variation_Sugar_Qty <- 2.1201803 # sd / mean = 4234 / 1997  
coeff_variation_Chocolate_Qty <- 4 # sd / mean = 4/1
```


PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

Usual

```
coeff_variation_Sugar_Qty <- 2.1201803 # sd / mean = 4234 / 1997  
coeff_variation_Chocolate_Qty <- 4 # sd / mean = 4/1
```

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

Usual

```
coeff_variation_Sugar_Qty <- 2.1201803 # sd / mean = 4234 / 1997
```

Better

```
standard_deviation_Sugar_Qty <- sd(Sugar_Qty)  
mean_Sugar_Qty <- mean(Sugar_Qty)  
  
coeff_variation_Sugar_Qty <- standard_deviation_Sugar_Qty / mean_Sugar_Qty
```

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

Usual

```
coeff_variation_Sugar_Qty <- 2.1201803 # sd / mean = 4234 / 1997
```

Better

```
standard_deviation_Sugar_Qty <- sd(Sugar_Qty)  
mean_Sugar_Qty <- mean(Sugar_Qty)  
  
coeff_variation_Sugar_Qty <- standard_deviation_Sugar_Qty / mean_Sugar_Qty
```

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

Usual

```
coeff_variation_Sugar_Qty <- 2.1201803 # sd / mean = 4234 / 1997
```

Even better

```
coeff_variation_Sugar_Qty <- sd(Sugar_Qty) / mean(Sugar_Qty)
```

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

- Le code le plus “*self-documenting*” possible

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

- ▶ Le code le plus “*self-documenting*” possible
- ▶ Le code le plus **générique** possible (principe DRY -*Don't Repeat Yourself* - (Wilson et al., 2014))

```
fct_coef_variation <- function(numvector)
{
  if( is.numeric(numvector) == F | is.vector(numvector) == F )
  {
    stop( "The data should be a numeric vector" )
  }
  standard_deviation_data <- sd(numvector)
  mean_data <- mean(numvector)
  coef_variation_data <- standard_deviation_data / mean_data
  return(coef_variation_data)
}
# Call the function for Sugar
fct_coef_variation(Sugar_Qty)
# Call for Chocolate
fct_coef_variation(Chocolate_Qty)
```

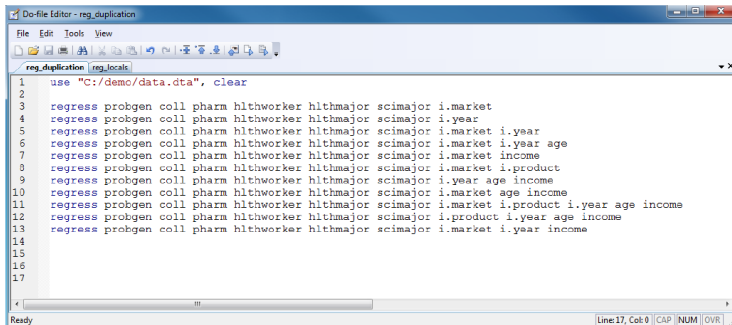
FIGURE – Exemple de code générique (fonction R)

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

- ▶ Le code le plus “*self-documenting*” possible
- ▶ Le code le plus **générique** possible (principe DRY -*Don't Repeat Yourself* - (Wilson et al., 2014))



```
Do-file Editor - reg_duplication
File Edit Tools View
reg_duplication reg_locals
1 use "C:/demo/data.dta", clear
2
3 regress probgen coll pharm hlthworker hlthmajor scimajor i.market
4 regress probgen coll pharm hlthworker hlthmajor scimajor i.year
5 regress probgen coll pharm hlthworker hlthmajor scimajor i.market i.year
6 regress probgen coll pharm hlthworker hlthmajor scimajor i.market i.year age
7 regress probgen coll pharm hlthworker hlthmajor scimajor i.market income
8 regress probgen coll pharm hlthworker hlthmajor scimajor i.market i.product
9 regress probgen coll pharm hlthworker hlthmajor scimajor i.year age income
10 regress probgen coll pharm hlthworker hlthmajor scimajor i.market age income
11 regress probgen coll pharm hlthworker hlthmajor scimajor i.market i.product i.year age income
12 regress probgen coll pharm hlthworker hlthmajor scimajor i.product i.year age income
13 regress probgen coll pharm hlthworker hlthmajor scimajor i.market i.year income
14
15
16
17
Ready
Line 17, Col 0 CAP NUM OVR
```

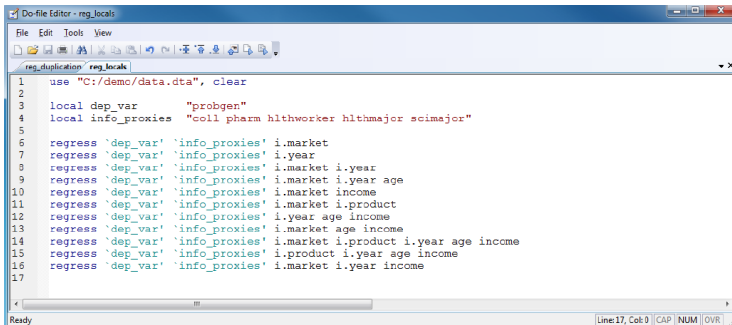
FIGURE – Exemple (Stata) issu de Gentzkow and Shapiro (2013)

PRINCIPE 2 :

CODER POUR LES AUTRES (Y COMPRIS POUR SOI-MÊME (FUTUR))

Faciliter la lisibilité et compréhension du code :

- ▶ Le code le plus “*self-documenting*” possible
- ▶ Le code le plus **générique** possible (principe DRY -*Don't Repeat Yourself* - (Wilson et al., 2014))



```
1 use "C:/demo/data.dta", clear
2
3 local dep_var      "probgén"
4 local info_proxies "coll pharm hlthworker hlthmajor scimajor"
5
6 regress `dep_var' `info_proxies' i.market
7 regress `dep_var' `info_proxies' i.year
8 regress `dep_var' `info_proxies' i.market i.year
9 regress `dep_var' `info_proxies' i.market i.year age
10 regress `dep_var' `info_proxies' i.market income
11 regress `dep_var' `info_proxies' i.market i.product
12 regress `dep_var' `info_proxies' i.year age income
13 regress `dep_var' `info_proxies' i.market age income
14 regress `dep_var' `info_proxies' i.market i.product i.year age income
15 regress `dep_var' `info_proxies' i.product i.year age income
16 regress `dep_var' `info_proxies' i.market i.year income
17
```

FIGURE – Exemple (Stata) issu de Gentzkow and Shapiro (2013)

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

- Tout **coder**! (pas de manipulation à la main, pas de copier-coller, pas de clic-bouton,...)

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

- ▶ Tout **coder**! (pas de manipulation à la main, pas de copier-coller, pas de clic-bouton,...)
- ▶ **Automatiser** le workflow !

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

- ▶ Tout **coder**! (pas de manipulation à la main, pas de copier-coller, pas de clic-bouton,...)
- ▶ **Automatiser** le workflow !
 - ▶ Ecrire un programme “*Maître*”

EXAMPLE 1 (Stata code): global.do

```
// Definition of the project's directory
local CodeFolder "c:/ApplePie/Progs"
cd "'CodeFolder'"
```

```
// Preparation of the data
do DataPreparation.do
```

```
// Some analysis code
do AnalysisCode.do
```

```
// Production of figures
do OutputCode.do
```

```
// Production of the paper
do MakingPaper.do
```

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

- ▶ Tout **coder**! (pas de manipulation à la main, pas de copier-coller, pas de clic-bouton,...)
- ▶ **Automatiser** le workflow !
 - ▶ Utiliser un fichier "*shell*" reliant les programmes (*.bat*)

EXAMPLE 2 (Batch file): global.bat

```
REM Definition of the project's directory
set CodeFolder="C:\ApplePie\Progs"
cd %CodeFolder%
```

```
REM Preparation of the data
R CMD BATCH DataPreparation.R
```

```
REM Some analysis code
stata-se /e do AnalysisCode.do
```

```
REM Production of figures
stata-se /e do OutputCode.do
```

```
REM Production of the paper
R CMD BATCH MakingPaper.R
```

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

- ▶ Tout **coder**! (pas de manipulation à la main, pas de copier-coller, pas de clic-bouton,...)
- ▶ **Automatiser** le workflow !
 - ▶ Utiliser un “*MakeFile*” (*Snakemake* pour Python)

```
# 1. Preparation of the data:
WorkingDataset.dta: RawData.csv DataPreparation.do
stata-se -b do "DataPreparation.do"

# 2. Some analysis code:
StatisticalTable.tex: WorkingDataset.dta AnalysisCode.do
stata-se -b do "AnalysisCode.do"

# 3. Production of two figures. The '%' character can be used as a shortcut:
Figure%.pdf: WorkingDataset.dta OutputCode.R
Rscript "OutputCode.R"

# 4. Production of the paper (from figures, table and bibliography):
Paper.pdf: Paper.tex biblio.bib Figure1.pdf Figure2.pdf StatisticalTable.tex
pdflatex "Paper.tex"

# 5. Production of a zip file
zip MyZipFile.zip Paper.pdf Paper.tex RawData.csv /
DataPreparation.do AnalysisCode.do OutputCode.R
```

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

Coder toutes les sorties de résultats

- ▶ Exporter les résultats et insertion dans un document

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

Coder toutes les sorties de résultats

► Exporter les résultats et insertion dans un document

	OLS		2SLS	
Price	−0.001***	(0.000)	−0.001	(0.001)
Cooker level	0.161***	(0.006)	0.161***	(0.006)
Number of different ingredients	0.030***	(0.007)	0.040	(0.036)
Number of servers	−0.042	(0.038)	−0.044	(0.039)
French recipe dummy	0.016*	(0.009)	0.016*	(0.009)
Michelin rating rank	0.050***	(0.008)	0.049***	(0.009)
Constant	−0.051	(0.113)	−0.098	(0.201)
Observations	428		428	
R^2	0.736		0.734	
Sargan statistic			0.923	
Sargan p			0.630	

Standard errors are in parentheses.

IV are input prices: sugar, flour and egg prices.

The Sargan test is an overidentification test of all instruments.

This is a fictive example (no real interpretation).

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

Coder toutes les sorties de résultats

- Exporter les résultats et insertion dans un document

```
/*-----*/
/* Define dependent ("Y") and independent variables
(exogeneous ones "Xexo", endogeneous one "Xendo")
and instrumental variables ("IV") */
/*-----*/
local Y      "Taste"
local Xexo   "Cooker_Level NbIngredients NbServers French Michelin"
local Xendo   "price"
local IV      "Eggs_Price Flour_Price Sugar_Price"

/*-----*/
/* Estimations : OLS and 2SLS */
/*-----*/
eststo OLS : reg `Y' `Xendo' `Xexo'
eststo IV : ivreg2 `Y' `Xexo' (`Xendo' = `IV'), endog(`Xendo') first ///
savefirst saveprefix(First_Stage)

/*-----*/
/* Export a nice table (LaTeX format) with both OLS and 2SLS estimation
results */
/*-----*/
esttab OLS IV using RegressionTable.tex, ///
scalar("N Observations" "r2 R$-2$" "sargan Sargan statistic" ///
"sarganp Sargan p") b(3) not nonumber mtitle compress replace se ///
star(* 0.10 ** 0.05 *** 0.01) label ///
title(Regression table created using Stata \textit{esttab} command. ///
\label{ExampleNiceReg}) ///
addnote("Standard errors are in parentheses." ///
"IV are input prices: sugar, flour and eggs prices." ///
"Sargan test is an overidentification test of all instruments." ///
"This is a fictive example (no real interpretation).") ///
"\sym{*} p < 0.10, \sym{**} p < 0.05, \sym{***} p < 0.01." ///
mtitle("OLS" "2SLS") wide
```

FIGURE – Exemple de code Stata

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

Coder toutes les sorties de résultats

► Exporter les résultats et insertion dans un document

	Analysis output (descriptive statistics, estimation results)	Graph
R	xtable, texreg (<i>.tex</i> , <i>.html</i> , <i>.doc</i>), stargazer, tables	png(), jpeg(), pdf(), tiff() (<i>.png</i> , <i>.jpg</i> , <i>.pdf</i> , <i>.tiff</i>)
Stata	table, collect, etable (<i>.docx</i> , <i>.html</i> , <i>.tex</i> , <i>.pdf</i> , <i>.xlsx</i> , ...), esttab (<i>.tex</i> , <i>.rtf</i>)	graph export (<i>.eps</i> , <i>.pdf</i> , <i>.wmf</i> , <i>.png</i>)
SAS	ods rtf (<i>.doc</i>), ods html (<i>.html</i> , <i>.xls</i>), ods pdf (<i>.pdf</i>), ods tagsets.latex (<i>.tex</i>)	ods graphics (<i>.png</i> , <i>.tiff</i> , <i>.jpg</i> , <i>.ps</i>)
MATLAB	writetable (<i>.xls</i> , <i>.csv</i>), xlswrite (<i>.xls</i>)	saveas (<i>.png</i> , <i>.eps</i> , <i>.pdf</i>)
Gams	gams2tbl (<i>.txt</i> , <i>.tex</i> , <i>.prn</i> , <i>.html</i>),.gdxrw, xl-export, xldump (<i>.csv</i> , <i>.xls</i>)	gnuplot, gnuplotxyz (<i>.png</i>)
Mathematica	Export[] (<i>.xls</i>), CloudExport[, "pdf"] (<i>.pdf</i>)	Export[] (<i>.gif</i> , <i>.jpg</i>)

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

Coder toutes les sorties de résultats

- ▶ Exporter les résultats et insertion dans un document
- ▶ Lier l'article avec le code (et les résultats) (*literate programming*, écriture de document reproductible)

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

Coder toutes les sorties de résultats

ILLUSTRATION

How to Make a Pie?

The authors

June, 2018

What are the ingredients for *Reproducible Research*?

The basic structure of reproducible research document follows a logic of sequences of commands in some programming language code chunks embedded in the text (or *text chunks* embedded in the code). For example:

```
Apple = c(300,3,200,200)
Crumble = c(500,4,200,10)
Cheese = c(200,2,30,40)
Muffin = c( 400,3,100,120)

Pie.Table <- rbind(Apple,Crumble, Cheese, Muffin)
Pie.Table <- as.data.frame(Pie.Table)

colnames(Pie.Table) = list("Flour", "Eggs", "Sugar", "Milk")
rownames(Pie.Table) = list("Pie", "Crumble", "Cake", "Muffin")
```

```
##      Flour Eggs Sugar Milk
## Pie      300   3   200  200
## Crumble  500   4   200   10
## Cake     200   2    30   40
## Muffin   400   3   100  120
```

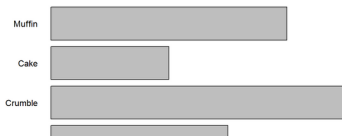
We can compute whatever we want in LaTeX

$$Weight = \frac{(a + c) * (b + d)}{2}$$

And draw the graphic:

```
barplot(Pie.Table$Flour, col = "gray",
        names.arg = rownames(Pie.Table),
        horiz = TRUE, las = 1,
        main = "Flour content in Pie recipes")
```

Flour content in Pie recipes



ILLUSTRATION

```
---
title: 'How to Make a Pie?'
author: "The authors"
date: "June, 2018"
output:
  html_document: default
  html_notebook:
    code_folding: hide
    fig_caption: yes
    theme: journal
---

### What are the ingredients for *Reproducible Research*?
The basic structure of reproducible research document follows a
logic of sequences of commands in some programming language *code
chunks* embedded in the text (or *text chunks* embedded in the
code).
For example:

```{r, results="hide"}
Apple = c(300,3,200,200)
Crumble = c(500,4,200,10)
Cheese = c(200,2,30,40)
Muffin = c(400,3,100,120)

Pie.Table <- rbind(Apple,Crumble, Cheese, Muffin)
Pie.Table <- as.data.frame(Pie.Table)

colnames(Pie.Table) = list("Flour", "Eggs", "Sugar", "Milk")
rownames(Pie.Table) = list("Pie", "Crumble", "Cake", "Muffin")
...

```{r, echo=FALSE}
#show table
Pie.Table
...

We can compute whatever we want in LaTeX

$$S_S \text{weight} = \frac{(a+c) \cdot (b+d)}{2}$$


And draw the graphic:

```{r, results="hide"}
barplot(Pie.Table$Flour, col = "gray",
 names.arg = rownames(Pie.Table),
 horiz = TRUE, las =1,
 main = "Flour content in Pie recipes")
...

```

# ILLUSTRATION

```

title: 'How to Make a Pie?'
author: "The authors"
date: "June, 2018"
output:
 html_document: default
 html_notebook:
 code_folding: hide
 fig_caption: yes
 theme: journal

What are the ingredients for "Reproducible Research"?
The basic structure of reproducible research document follows a
logic of sequences of commands in some programming language "code
chunks" embedded in the text (or "text chunks" embedded in the
code).
For example:

```{r, results="hide"}
Apple = c(300,3,200,200)
Crumble = c(500,4,200,10)
Cheese = c(200,2,30,40)
Muffin = c(400,3,100,120)

Pie.Table <- rbind(Apple,Crumble, Cheese, Muffin)
Pie.Table <- as.data.frame(Pie.Table)

colnames(Pie.Table) = list("Flour", "Eggs", "Sugar", "Milk")
rownames(Pie.Table) = list("Pie", "Crumble", "Cake", "Muffin")
...

```{r, echo=FALSE}
#show table
Pie.Table
...

We can compute whatever we want in LaTeX

$$Weight = \frac{(a+c) * (b+d)}{2}$$

And draw the graphic:

```{r, results="hide"}
barplot(Pie.Table[,Flour, col = "gray",
  names.arg = rownames(Pie.Table),
  horiz = TRUE, las =1,
  main = "Flour content in Pie recipes")
...

```

How to Make a Pie?

The authors
June, 2018

What are the ingredients for *Reproducible Research*?

The basic structure of reproducible research document follows a logic of sequences of commands in some programming language code chunks embedded in the text (or *text chunks* embedded in the code). For example:

```

Apple = c(300,3,200,200)
Crumble = c(500,4,200,10)
Cheese = c(200,2,30,40)
Muffin = c(400,3,100,120)

Pie.Table <- rbind(Apple,Crumble, Cheese, Muffin)
Pie.Table <- as.data.frame(Pie.Table)

colnames(Pie.Table) = list("Flour", "Eggs", "Sugar", "Milk")
rownames(Pie.Table) = list("Pie", "Crumble", "Cake", "Muffin")

```

```

##      Flour Eggs Sugar Milk
## Pie      300   3   200  200
## Crumble  500   4   200   10
## Cake     200   2    30   40
## Muffin   400   3   100  120

```

We can compute whatever we want in LaTeX

$$Weight = \frac{(a+c) * (b+d)}{2}$$

And draw the graphic:

```

barplot(Pie.Table[,Flour, col = "gray",
  names.arg = rownames(Pie.Table),
  horiz = TRUE, las =1,
  main = "Flour content in Pie recipes")

```

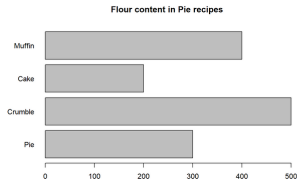


FIGURE — Exemple d'un document reproductible (*R Markdown*).

EGALEMENT POSSIBLE DANS D'AUTRES LOGICIELS

(PAR EX. STATA)

```

---
title: Reproducible document with Stata and Markdown
author: C. Rostenga, V. Ozorio
date: December 2019
bibliography: Bib.bib
---

Introduction
=====
This small paper aims at providing some trends (geographical, time) about consumption. <br/>
First section will introduce the data and the product we study. Then we will see a map in order to analyse if some regional behaviours exist. <br/>
Section 3 will present some estimations. It's a see that using a bib file for references is easy! (see [McCullough2009]).

Data
=====
***/
local prod "B111"
qui use "../data/p/prod/11an.dta", clear
qui wallist lib*
global lib `*'z(lib)`*
qui sum annee
local y "`z(mean)'"
local Nobs "`z(N)'"
***

We used data from the TNS-Worldpanel (2008) on "`s' lib_n==1" consumption for the year "`s' y". The data contains "`s' Nobs" observations.

***/
qui bysort CODE_DEPT : keep if _n==1
qui keep CODE_DEPT qtdep lib
qui merge 1:1 CODE_DEPT using "../data/PopDep93.dta", keepusing(pop93)
qui drop if _merge==2
qui drop _merge
qui gen qtdepq = qtdep / pop93
qui sum qtdepq qtdep

qui merge 1:1 CODE_DEPT using "../data/frdb.dta", keepusing(lid X_CHF_LIEU V_CHF_LIEU)
qui drop if _merge==2
qui drop _merge

***

Map
=====
***/
qui cd ../data
qui spmap qtdepq using frcoast if id!=97, id(id) fcolor(Purple) title("Purchased Quantity/Inhabitant $lib") ///
    addtitle("France 2008 per dep93[DEPT]") ///
    legend(on) legend(1) legend(ring(1) position(3)) legend(Quantity in Kg on L) ///
    scalebar(units(800) scale(1/1000) mpos(-80) label(Kilometres))
qui cd ../progs
qui graph export map.png, width(400) replace

!Rice Map! map.png

Estimations
=====
***/
qui use "../data/p/prod/11an.dta", clear
qui gen sum(hablibm)

Our main argument is that the BMI depends on the quantity of "`s' lib_n==1" we eat. The models we are going to estimate are the following:

***/
BMI_i = beta_0 + \beta_{s,i} \cdot X_i + \epsilon_i
***/
qui estato clear
qui estato : reg BMI1 qtdepqgen ru mifi Ageind1
esttab , not z2 ar(2) label nonumber title(Estimations) mtitles("Mod 1" "Mod 2") addnote("Source : $lib data") b(49.5f) replace

```

Reproducible document with Stata and Markdown

C. Rostenga, V. Ozorio
December 2019

Introduction

This small paper aims at providing some trends (geographical, time) about consumption.

 First section will introduce the data and the product we study. Then we will see a map in order to analyse if some regional behaviours exist.

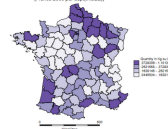
 Section 3 will present some estimations. Let's see that using a bib file for references is easy! (see [McCullough 2009]).

Data

We used data from the TNS-Worldpanel (2008) on CHAMPAGNE consumption for the year 2008. The data contains 3743 observations.

Map

Purchased Quantity/inhab. CHAMPAGNE
(France 2008 per dep93[DEPT])



See Map

Estimations

Our main argument is that the BMI depends on the quantity of CHAMPAGNE we eat. The models we are going to estimate are the following:

$$BMI_i = \beta_0 + \beta_1 \cdot X_i + \epsilon_i$$

estimations	mod 1	mod 2
qtdepqgen	0.000***	0.000
ru	0.045*	0.049*
Age as PC	0.164***	0.049
Age ind1	0.000***	0.000***
Constant	24.284***	21.280***
N-observ	3743	3743
Adjusted R-squared	0.02	0.00
Observations	3743	3743

Source : CHAMPAGNE DATA
* p<0.10, ** p<0.05, *** p<0.01

That's all for now!

Bibliography

McCullough, B. O. 2009. "Open Access Economics Journals and the Market for Reproducible Economic Research." *Economic Analysis*

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

Créer un *document reproductible (papier exécutable)*

- ▶ Lier l'article avec code (et résultats)

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

Créer un *document reproductible (papier exécutable)*

- ▶ Lier l'article avec code (et résultats)
- ▶ Notebooks (*R Notebook, Jupyter Notebook, Quarto...*)

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

Créer un document reproductible (papier exécutable)

- Lier l'article avec code (et résultats)
- Notebooks (*R Notebook*, *Jupyter Notebook*, *Quarto*...)

Language used for:				
Code	Text	Tool name	Output format	References
<i>Sweave-like tools</i>				
R	LaTeX	Sweave	TeX, Beamer, PDF	Leisch (2002), Meredith and Racine (2009)
R, Python, SAS, SQL, ...	Markdown	R Markdown	HTML, PDF, MS Word, Beamer, ...	Xie (2015), Gandrud (2015), Xie <i>et al.</i> (2018)
R, SAS	LaTeX, noweb	SASWeave	TeX, PDF	Lenth and Højsgaard (2007), Morrisson and Karafa (2012)
R, SAS, MATLAB, Stata, ...	LaTeX, OpenOffice	StatWeave*	TeX, ODT	Lenth and Højsgaard (2011), Lenth (2012)
Stata	Markdown	Markstat	TeX, PDF, HTML	Rodriguez (2017)
SAS	LaTeX	StatRep	TeX, PDF	Arnold and Kuhfeld (2012, 2015), Morrisson and Karafa (2012)
MATLAB	plain text markup	Publish	TeX, MS Word, HTML, PDF	Hunt <i>et al.</i> (2014), McCarthy (2018)
R, Stata, MATLAB, Python, ...	plain text markup	Org-mode	TeX, PDF, HTML, ODT, ...	Schulte and Davison (2011), Schulte <i>et al.</i> (2012), Dominik (2019)
<i>Notebooks</i>				
Python, R, SAS, Stata, ** MATLAB, Julia, ...	Markdown	Jupyter Notebook	HTML, rST, PDF	LeVeque (2009), Kluyver <i>et al.</i> (2016), de Kok (2016)
Mathematica	Wolfram language	Mathematica Notebook	HTML, PDF, TeX, ...	Wolfram Research, Inc. (2008)
R, Python, SAS, SQL, ...	Markdown	R Notebook	HTML, PDF, MS Word, Beamer, ...	Gandrud (2015)
MATLAB	Formatted text	Live Scripts	HTML, PDF	Hunt <i>et al.</i> (2014), McCarthy (2018)

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

Pour aller plus loin : Maîtriser son environnement pour s'assurer de l'entière reproductibilité sur une autre machine

- Figurer/partager ses versions de packages pour un projet, gestion dépendances (par ex. package *R renv()*)

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

Pour aller plus loin : Maîtriser son environnement pour s'assurer de l'entière reproductibilité sur une autre machine

- ▶ Figurer/partager ses versions de packages pour un projet, gestion dépendances (par ex. package *R renv()*)
- ▶ Encapsuler l'environnement de travail : conteneurs (par ex. *Docker, Singularity*) ; systèmes de gestion de paquet (par ex. *Guix, Nix*)

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

Pour aller plus loin : Maîtriser son environnement pour s'assurer de l'entière reproductibilité sur une autre machine

- ▶ Figurer/partager ses versions de packages pour un projet, gestion dépendances (par ex. package *R renv()*)
- ▶ Encapsuler l'environnement de travail : conteneurs (par ex. *Docker, Singularity*) ; systèmes de gestion de paquet (par ex. *Guix, Nix*)
- ▶ Partager des notebooks (*Binder, Guix-Jupyter*)

POUR CORSER LE TOUT : TRAVAILLER AVEC MES CO-AUTEURS

Un peu + difficile!

- ▶ Définir ensemble des **conventions** (nommage, arborescence, ...) $\hookrightarrow$ *pair programming*
- ▶ **Centraliser** les informations
 - ▶ Données $\hookrightarrow$ (*Dropbox/Framadrive*), serveurs sécurisés, plateformes d'archivage (*Web repository* : FigShare, Zenodo,...)
 - ▶ Programmes $\hookrightarrow$ (*Dropbox*), (*Sharepoint*), GitLab/SourceSup, archivage des codes sur le long terme : Software Heritage
 - ▶ Environment computationnel $\hookrightarrow$ Docker, Binder
 - ▶ Bibliographie $\hookrightarrow$ Endnote, Zotero, BibTeX
 - ▶ Document de travail/Article/Rapport $\hookrightarrow$ Overleaf/PLMlateX(L^AT_EX), (*Google Docs*), Sharepoint, OneDrive, Teams

POUR CORSER LE TOUT : ECHANGER AVEC MES CO-AUTEURS

Echange de messages / Gestion des tâches

POUR CORSER LE TOUT : ECHANGER AVEC MES CO-AUTEURS

Echange de messages / Gestion des tâches

De : Valerie Orozco

Envoyé : mercredi 16 mars 2016 15:41

À : Christophe Bontemps

Objet : Our pie project

As we decided two weeks ago, I estimate the model on the chocolate cake market and still found our strong significant effect of the butter content !

Could you please send me the test you did adding a dummy indicating whether it is a molten ("fondant") or a moist ("moelleux") cake? I'm writing the section in our WP.

De : Christophe Bontemps

Envoyé : jeudi 17 mars 2016 18:52

À : Valerie Orozco

Objet : RE : Our pie project

Great for the chocolate cake market.

For the dummy, I was thinking you did the job? Maybe Céline did?

POUR CORSER LE TOUT : ECHANGER AVEC MES CO-AUTEURS

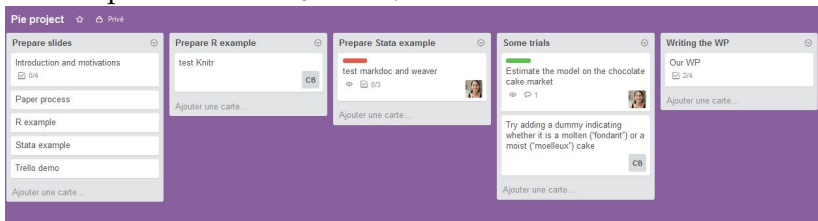
Echange de messages / Gestion des tâches

- ▶ Système de gestion de version : messages avec les "commits"

POUR CORSER LE TOUT : ECHANGER AVEC MES CO-AUTEURS

Echange de messages / Gestion des tâches

- Système de gestion de version : messages avec les "commits"
- Utiliser un système de gestion de tâches...mieux que des mails / post-it! (*Online management tasks system*) comme Trello, Asana,



A RETENIR

Notre expérience

- L'élément le plus important est le **workflow**

A RETENIR

Notre expérience

- ▶ L'élément le plus important est le **workflow**
- ▶ Chacun peut **progresser** dans la démarche RR
 - ▶ De nombreux outils/pratiques pour améliorer la qualité (reproductibilité), rigueur de nos recherches
 - ▶ S'améliorer plutôt que d'attendre de devoir subir la pression

VALORISATION DE LA RR

Des initiatives valorisant la RR ont émergé

2009 : Création des *badges* (BioStatistics, AJPS)

2011 : Papiers Exécutables sur ScienceDirect

2013 : *RunMyCode*, *Binder*, *Code Ocean*, *WholeTale*

2019 : *CodeCheck* : initiative d'exécution et test du code

VALORISATION DE LA RR

Des initiatives valorisant la RR ont émergé

2009 : Création des *badges* (BioStatistics, AJPS)

2011 : Papiers Exécutables sur ScienceDirect

2013 : *RunMyCode, Binder, Code Ocean, WholeTale*

2019 : *CodeCheck* : initiative d'exécution et test du code

CODECHECK certificate 2022-001

<https://doi.org/10.5281/zenodo.6040066>



Item	Value
Title	Geographically Weighted Regressions for prioritizing educational planning, policies, and interventions
Authors	Germán Vargas Mesa, Amélie A. Gagnon 
Reference	IIEP technical note (2021) http://www.iiep.unesco.org/en/publication/geographically-weighted-regressions-prioritizing-educational-planning-policies-and
Codechecker	Stephen J. Eglen 
Date of check	2022-01-19 10:00:00
Summary	R code for this project was reproducible.
Repository	https://github.com/codecheckers/GWR-in-educational-planning

Table 1: CODECHECK summary

<https://codecheck.org.uk>

VALORISATION DE LA RR

Des initiatives valorisant la RR ont émergé

2009 : Création des *badges* (BioStatistics, AJPS)

2011 : Papiers Exécutables sur ScienceDirect

2013 : *RunMyCode*, *Binder*, *Code Ocean*, *WholeTale*

2019 : *CodeCheck* : initiative d'exécution et test du code

2019 : Agence de certification CASCAD (*Certification Agency for Scientific Code And Data*) (CNRS, éco et management) (Pérignon et al., 2019)



February 10, 2022 - cascad#222

Execution Report

Title: **Equilibrium Bitcoin Pricing**

Author: **Bruno Biais, Christophe Bisière, Matthieu Bouvard,
Catherine Casamatta & Albert J. Menkveld**

Full references: Biais Bruno, Bisière Christophe, Bouvard Matthieu, Casamatta Catherine & Menkveld Albert J. "Equilibrium Bitcoin Pricing" Working paper, February 3, 2022.

The structure and contents of this execution report provided by **cascad** for the certification are similar to those recommended by the [ALA Data Editor](#).

1. DATA DESCRIPTION

This paper relies on data on returns, transactional benefits and costs from the MtGox bitcoin marketplace, collected from July 17, 2010 to December 31, 2018. For a thorough description of the data, please refer to Section 4 of the paper.

2. CODE DESCRIPTION

For the purpose of this certification, we aimed to check the results displayed in Table 1 and in Figures 9-12. The code for the other Figures were not provided to us (note Figures 1 and 3 contain no results).

The replication materials are divided into four self-explanatory subfolders: "code", "input", "temp", and "output". The "input" subfolder stores one data file, "crypto-pricing-series-daily-2009-2018.xlsx". A Python script, "equilibrium-bitcoin-pricing-model-calibration.py" loads the Excel file, implements the calibration, creates the Table and Figures and stores them in the "temp" folder, before writing down in a pdf file that is exported to the "output" folder.

AVANCER POUR UNE RECHERCHE + REPRODUCTIBLE...LE RÔLE DE TOUS !

Tous les acteurs de la recherche ont un rôle à jouer

- ▶ Les **journaux** devraient demander les données et les codes lors de la soumission...
- ▶ ... et tester la **reproductibilité** des résultats avant la publication

AVANCER POUR UNE RECHERCHE + REPRODUCTIBLE...LE RÔLE DE TOUS !

Tous les acteurs de la recherche ont un rôle à jouer

- ▶ Les **journaux** devraient demander les données et les codes lors de la soumission...
- ▶ ... et tester la **reproductibilité** des résultats avant la publication
- ▶ Les **producteurs de données** ont aussi un rôle à jouer
- ▶ Les **institutions de recherche (ERC, ANR, ...)** vont probablement monter en exigences dans les projets financés (cf. Rapport Commission Européenne sur RR, décembre 2020)
- ▶ **Enseignement** à la RR doit augmenter : MOOCs (par ex. Recherche Reproducible, piloté par l'INRIA), workshops, ...

AVANCER POUR UNE RECHERCHE + REPRODUCTIBLE...LE RÔLE DE TOUS !

Tous les acteurs de la recherche ont un rôle à jouer

- ▶ **Nous** (chercheurs, ingénieurs, doctorants, enseignants, encadrants, referees, ...) doivent ouvrir la voie
- ▶ Vers une meilleure reproductibilité...

Réseau français sur la recherche reproductible. <https://www.recherche-reproductible.fr/>

AVANCER POUR UNE RECHERCHE + REPRODUCTIBLE...LE RÔLE DE TOUS !

Tous les acteurs de la recherche ont un rôle à jouer

- ▶ **Nous** (chercheurs, ingénieurs, doctorants, enseignants, encadrants, referees, ...) doivent ouvrir la voie
- ▶ Vers une meilleure reproductibilité...
Réseau français sur la recherche reproductible. <https://www.recherche-reproductible.fr/>
- ▶ ...mais aussi vers plus de **transparence** (et de **partage**! → **Open Science**!)

POUR FINIR

*“Only results that can be replicated are **truly scientific results**. If there is no chance to replicate research results, they can be regarded as **no more than personal views**”*

(Huschka, 2013)

MERCI!

REFERENCES I

- Björk, B.-C. and Solomon, D. (2013). The publishing delay in scholarly peer-reviewed journals. *Journal of Informetrics*, 7(4) :914–923.
- Card, D. and DellaVigna, S. (2013). Nine facts about top journals in economics. *Journal of Economic Literature*, 51(1) :144–161.
- Chang, A. C. and Li, P. (2015). Is economics research replicable? sixty published papers from thirteen journals say ‘usually not’. Finance and Economics Discussion Series 2015-83, Board of Governors of the Federal Reserve System (U.S.).
- Dewald, W. G., Thursby, J. G., and Anderson, R. G. (1988). Replication in empirical economics : The journal of money, credit and banking project : Reply. *American Economic Review*, 78(5) :1162–1163.
- Flier, J. S. (2022). The problem of irreproducible bioscience research. *Perspectives in Biology and Medicine*.
- Gentzkow, M. and Shapiro, J. (2013). Nuts and bolts : Computing with large data. In *Summer Institute 2013 Econometric Methods for High-Dimensional Data*. Summer Institute 2013 Econometric Methods for High-Dimensional Data.

REFERENCES II

- Gentzkow, M. and Shapiro, J. M. (2014). Code and data for the social sciences : a practitioner's guide. University of Chicago mimeo.
- Gertler, P., Galiani, S., and Romero, M. (2018). How to make replication the norm. *Nature*, 554 :417–419.
- Hamermesh, D. S. (2013). Six decades of top economics publishing : Who and how ? *Journal of Economic Literature*, 51(1) :162–172.
- Herbert, S., Kingi, H., Stanchi, F., and Vilhuber, L. (2021). The reproducibility of economics research : A case study. Technical report.
- Horvath, J. (2013). The replication myth : Shedding light on one of science's dirty little secrets. *Scientific American*, December.
- Huschka, D. (2013). Why should we share our data, how can it be organized, and what are the challenges ahead ? Ratswd german data forum.
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med*, 2(8) :e124.
- Knuth, D. E. (1984). Literate programming. *The Computer Journal*, 27 :97–111.

REFERENCES III

- Kuld, L. and O'Hagan, J. (2017). The trend of increasing co-authorship in economics : New evidence. Technical report, Center for Economic Policy Research (CEPR).
- LeVeque, R. J. (2009). Python tools for reproducible research on hyperbolic problems. In *Special issue on Reproducible Research*, pages 19–27. Computing in Science and Engineering (CiSE).
- Mathieu, G., Pigeon, D., Rabemanantsoa, T., Chipeaux, C., Duvillard, S., Gruson-Daniel, C., Herbet, M. E., Legrand, A., Leon, N., Libri, D., Lily, J.-B., Pinet, A., Sabot, F., Seydi, M., and Theisen, V. (2021). Rapport du groupe de travail sur les cahiers de laboratoires électroniques. Research report, Comité pour la science ouverte.
- Millman, K. J. and Pérez, F. (2014). Developing open source scientific practice. In *Implementing Reproducible Research*, chapter 6, pages 149–183. CRC Press, Stodden, Victoria and Leisch, Friedrich and Peng, Roger D (ed.).
- Nature, E. (2013). Reducing our irreproducibility. *Nature*, 496 :398.

REFERENCES IV

- Orozco, V., Bontemps, C., Maigne, E., Piguet, V., Hofstetter, A., Lacroix, A., Levert, F., and Rousselle, J.-M. (2020). How to make a pie : Reproducible research for empirical economics & econometrics. *Journal of Economic Surveys*, 34(5) :1134–1169.
- Pérignon, C., Gadouche, K., Hurlin, C., Silberman, R., and Debonnel, E. (2019). Certify reproducibility with confidential data. *Science*, 365(6449) :127–128.
- Perkel, J. M. (2022). Cut the tyranny of copy-and-paste with these coding tools. *Nature*, 603(7899) :191–192.
- Wilson, G., Aruliah, D. A., Brown, C. T., Chue Hong, N. P., Davis, M., Guy, R. T., Haddock, S. H. D., Huff, K. D., Mitchell, I. M., Plumbley, M. D., Waugh, B., White, E. P., and Wilson, P. (2014). Best practices for scientific computing. *PLOS Biology* 12(1), e1001745.
- Wood, B. D., Müller, R., and Brown, A. N. (2018). Push button replication : Is impact evaluation evidence for international development verifiable? *PloS one*, 13(12) :e0209416.