

How to make a pie?

Comment améliorer nos pratiques pour aller vers une recherche (plus) reproductible?

Valérie Orozco

Toulouse School of Economics, INRAE (Département EcoSocio)

*en collaboration avec Christophe Bontemps, Elise Maigné, Virginie Piguet,
Annie Hofstetter, Anne Lacroix, Fabrice Levert, Jean-Marc Rousselle*



AG RIS - 16 mars 2021

AVEZ-VOUS DÉJÀ ENTENDU (OU DIT)?

**"J'avais écrit ces chiffres,
mais je ne sais pas de quel
programme ils sont issus..."**

AVEZ-VOUS DÉJÀ ENTENDU (OU DIT) ?

**"J'avais écrit ces chiffres,
mais je ne sais pas de quel
programme ils sont issus..."**

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

AVEZ-VOUS DÉJÀ ENTENDU (OU DIT)?

**"J'avais écrit ces chiffres,
mais je ne sais pas de quel
programme ils sont issus..."**

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

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

AVEZ-VOUS DÉJÀ ENTENDU (OU DIT) ?

**"J'avais écrit ces chiffres,
mais je ne sais pas de quel
programme ils sont issus..."**

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

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

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

AVEZ-VOUS DÉJÀ ENTENDU (OU DIT) ?

**"J'avais écrit ces chiffres,
mais je ne sais pas de quel
programme ils sont issus..."**

**"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 a été
construite cette
variable..."**

AVEZ-VOUS DÉJÀ ENTENDU (OU DIT) ?

"J'avais écrit ces chiffres, mais je ne sais pas de quel programme ils sont issus..."

"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 a été construite cette variable..."

AVEZ-VOUS DÉJÀ ENTENDU (OU DIT) ?

"J'avais écrit ces chiffres, mais je ne sais pas de quel programme ils sont issus..."

"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 a été construite cette variable..."

"D'où sort ce tableau de résultats ?"

AVEZ-VOUS DÉJÀ ENTENDU (OU DIT) ?

"J'avais écrit ces chiffres, mais je ne sais pas de quel programme ils sont issus..."

"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 a été construite cette variable..."

"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), ...

Why Most Published Research Findings Are False

John P. A. Ioannidis

- ▶ En Sciences en général : Nature (2013), (2017), (2018) *et al...*

ANNOUNCEMENT

Reducing our irreproducibility

CRISE DE LA REPRODUCTIBILITÉ

- ▶ Des erreurs (souvent involontaires)... même sur des papiers publiés ! Par exemple AER 2010

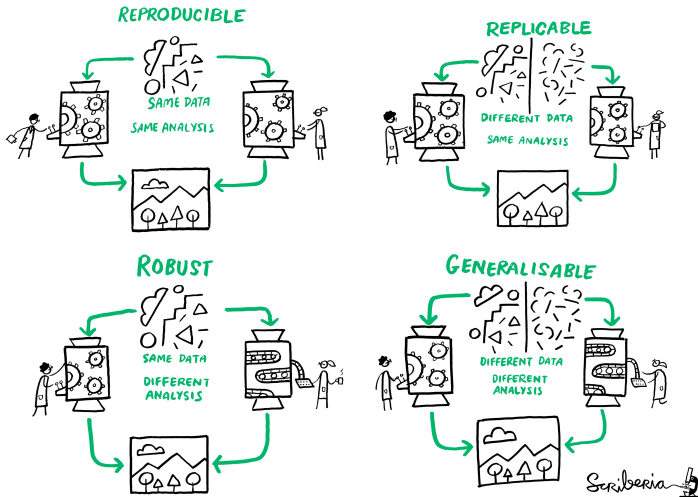
Comment Thomas Herndon, 28 ans, a mis à jour les failles de l'étude Rogoff/Reinhart

Par *Karl de Meyer* | 20/04 | 08:47 | mis à jour le 22/04 | 15:21 | [8](#) commentaires

En démontrant que deux stars de l'économie, Kenneth Rogoff et Carmen Reinhart, avaient commis des erreurs dans les feuilles de calcul Excel qui sous-tendent une de leurs études de 2010, il a déclenché une tempête médiatique.

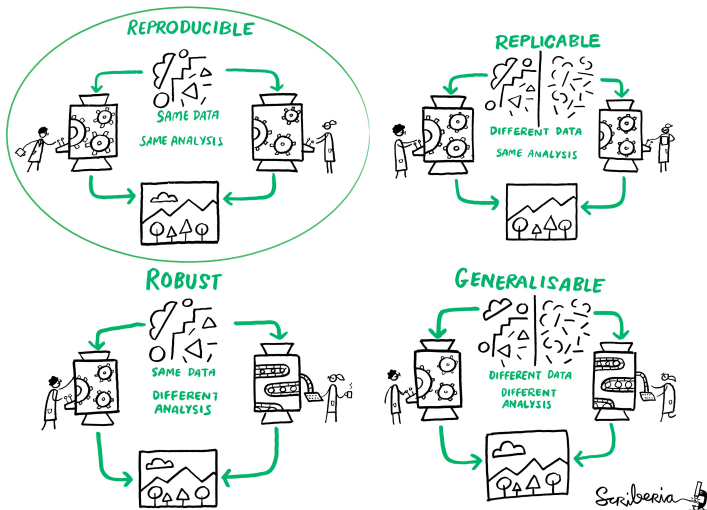
	B	C	I	J	K	L	M
2			Real GDP growth				
3			Debt/GDP				
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	30 or less
26			3.7	3.0	3.5	1.7	5.5
27	Minimum		1.6	0.3	1.3	-1.8	0.8
28	Maximum		5.4	4.9	10.2	3.6	13.3
29							
30	US	1946-2009	n.a.	3.4	3.3	-2.0	n.a.
31	UK	1946-2009	n.a.	2.4	2.5	2.4	n.a.
32	Sweden	1946-2009	3.6	2.9	2.7	n.a.	6.3
33	Spain	1946-2009	1.5	3.4	4.2	n.a.	9.9
34	Portugal	1952-2009	4.8	2.5	0.3	n.a.	7.9
35	New Zealand	1948-2009	2.5	2.9	3.9	-7.9	2.6
36	Netherlands	1956-2009	4.1	2.7	1.1	n.a.	6.4
37	Norway	1947-2009	3.4	5.1	n.a.	n.a.	5.4
38	Japan	1946-2009	7.0	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1948-2009	4.4	4.5	4.0	2.4	2.9
41	Greece	1970-2009	4.0	0.3	2.7	2.9	13.3

DE QUOI PARLE T-ON ?



MA PRÉSENTATION

► Reproductibilité



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

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))

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 de code!

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 de code !
- ↪ Un chercheur/ingénieur qui fait de l'empirique est maintenant un programmeur !

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 de code !
- ↪ Un chercheur/ingénieur qui fait de l'empirique est maintenant un programmeur !
- ↪ Beaucoup de temps passé à écrire, à lire, et à débbuguer du code

NOTRE MÉTIER A CHANGÉ!

Plus de code

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

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

NOTRE MÉTIER A CHANGÉ

Plus de collaborations

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))

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

NOTRE MÉTIER A CHANGÉ!

Une pression croissante

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

NOTRE MÉTIER A CHANGÉ !

Une pression croissante

- ▶ 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

NOTRE MÉTIER A CHANGÉ!

Une pression croissante

- ▶ 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*, ...)

NOTRE MÉTIER A CHANGÉ!

Une pression croissante

- ▶ Publication de répliquions/reproductions

NOTRE MÉTIER A CHANGÉ!

Une pression croissante

- ▶ Publication de réplifications/reproductions
 - ▶ Certaines revues publient des réplifications (par ex. JAE)

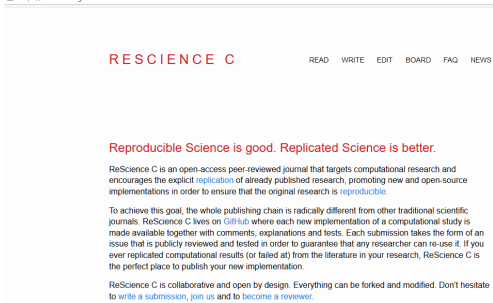
NOTRE MÉTIER A CHANGÉ !

Une pression croissante

- ▶ Publication de réplifications/reproductions
 - ▶ Certaines revues publient des réplifications (par ex. JAE)
 - ▶ Des revues spécialisées dans la publication de réplifications :

ReScience

<https://rescience-c.github.io>



The screenshot shows the homepage of ReScience C. At the top, the text "RESCIENCE C" is displayed in red, with navigation links "READ", "WRITE", "EDIT", "BOARD", "FAQ", and "NEWS" to its right. Below this, a red heading reads "Reproducible Science is good. Replicated Science is better." The main text describes the journal as an open-access peer-reviewed journal for computational research, emphasizing the explicit replication of published research. It mentions that the journal uses GitHub for submissions and reviews, and encourages collaborative and open design.

NOTRE MÉTIER A CHANGÉ !

Une pression croissante

- ▶ Publication de réplifications/reproductions
 - ▶ Certaines revues publient des réplifications (par ex. JAE)
 - ▶ Des revues spécialisées dans la publication de réplifications : ReScience
 - ▶ International Journal for Re-Views in Empirical Economics (IREE), créé en 2017

The screenshot shows the mission statement of the International Journal for Re-Views in Empirical Economics (IREE). The page title is "International Journal for Re-Views in Empirical Economics". Below the title, it states: "WE PUBLISH REPLICATION STUDIES IN ECONOMICS - INDEPENDENT OF THEIR RESULT. PEER-REVIEWED. OPEN ACCESS. NO AUTHOR FEES." There is a search bar and a "SEARCH IREE" button. The "Our Mission" section contains the following text: "Replications are pivotal for the credibility of empirical economics. Only replicable and robust results can be generalized. And only generalizable findings can serve as evidence-based advice to economic policy. And yet, replication studies are rarely published in economic journals. The International Journal for Re-Views in Empirical Economics (IREE) is the first journal dedicated to the publication of replication studies in empirical economics." Below this, it says: "Until recently, it was difficult for authors of replication studies to gain recognition for their work and to be accepted to publish in stable publication outlets. Since researchers in economics have had little to no".

NOTRE MÉTIER A CHANGÉ !

Une pression croissante

- ▶ Publication de réplifications/reproductions
 - ▶ Certaines revues publient des réplifications (par ex. JAE)
 - ▶ Des revues spécialisées dans la publication de réplifications : ReScience
 - ▶ International Journal for Re-Views in Empirical Economics (IREE), créé en 2017
- ▶ Un wiki autour de la réplification en économie existe



<https://replication.uni-goettingen.de/wiki/index.php>

NOTRE MÉTIER A CHANGÉ !

Une pression croissante

- ▶ Publication de réplifications/reproductions
 - ▶ Certaines revues publient des réplifications (par ex. JAE)
 - ▶ Des revues spécialisées dans la publication de réplifications : ReScience
 - ▶ International Journal for Re-Views in Empirical Economics (IREE), créé en 2017
- ▶ Un wiki autour de la réplification en économie existe

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

Contents (7/6)	
1	Article
2	Article information
3	This article is a replication of
4	References



Article

Authors	Title	Journal	Year	Edition	Pages	JEL Codes ¹	Keywords
Matthew Berg, Brian Hankley	Debt to GDP Ratios and Growth: Country Heterogeneity and Reverse Causation, the Case of Japan (Ultra Wonky) ²	NEP	2013	-	E73-E76	-	-

Article information

Program code	Data	Readme	Method(s) & estimation	Data type	Data used	Origin of data used	Software used (if/when)
0 - not available online	0 - not available online	0 - not available online	Linear 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 ¹	Keywords	Replication type	Replication result (refer to replication type 1 and 2)	Raw data	Call into question	Authors statement
Carmen Reinhart, Kenneth Rogoff	Growth in a Time of Debt	AER	2010	1	-	E23, E31, H63, O11	-	2 - wide sense (new)	1 - successful	3 - final dataset can be easily	2 - partly	1 - unrelated results

NOTRE MÉTIER A CHANGÉ !


Une pression croissante

- ▶ Publication de réplifications/reproductions
 - ▶ Certaines revues publient des réplifications (par ex. JAE)
 - ▶ Des revues spécialisées dans la publication de réplifications : ReScience
 - ▶ International Journal for Re-Views in Empirical Economics (IREE), créé en 2017
- ▶ Un wiki autour de la réplification en économie existe

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

Contents (1/6)

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

Article

Authors	Title	Journal	Year	Edition	Pages	JEL Codes ^{1/2}	Keywords
Jahjouk Yu.	Semiparametric Value-At-Risk Estimation of Portfolios	Int J Re-Views in Emp Econ	2019	-	1-20	C51, G01, G11, G17	-

Article information

Program code	Date	Revised	Methods ^{1/2} & estimators	Data type	Data used	Origin of data used	Software used (Version)
2 - available online elsewhere 1 - accessible on journal website	-	-	-	Micro	Dataset	-	MATLAB

This article is a replication of

Authors	Title	Journal	Year	Edition	Pages	JEL Codes ^{1/2}	Keywords	Replication type	Replication result ^{1/2} (refer to replication type 1 and 2)	Raw data	Call into question	Authors statement
Alexandra Diaz	Semiparametric estimation of multi-asset portfolio VaR risk	J Banking & Finance	2014	-	368-408	C51, G01, G11, G17	-	2 - with source (raw data)	1 - successful	2 - raw data can be used to reproduce a dataset that is not substantially different from the	0 - no	-

NOTRE MÉTIER A CHANGÉ !

Une pression croissante

- ▶ Publication de réplifications/reproductions
 - ▶ Certaines revues publient des réplifications (par ex. JAE)
 - ▶ Des revues spécialisées dans la publication de réplifications : ReScience
 - ▶ International Journal for Re-Views in Empirical Economics (IREE), créé en 2017
- ▶ Un wiki autour de la réplification en économie existe

Aid, Policies, and Growth: Comment (AER 2004)

Article		Title		Journal	Year	Volume	Pages	JEL Codes?	Keywords			
Authors: Robert E. Lucas, David Stockman Vol. Notes, and Growth Comment?				AER	2004	3	176-180					
Article Information												
Original author	Title	Year	Journal	Volume	Pages	JEL Codes	Keywords	Origin of data used	Software used (see text)			
Robert E. Lucas, David Stockman	Aid, policies and growth: Comment	2004	AER	3	176-180			Sub-Subsidiary Africa and East Asia-Subsidiary				
This article is a replication of												
Authors	Title	Journal	Year	Volume	Pages	JEL Codes	Keywords	Replication type	Replication year	Year data	Cell data updates	Authors addresses
David Stockman, Robert E. Lucas	Aid, policies and growth	AER	2004	3	181-189	E15, E24, E25	Growth and development; Development economics; Aid effectiveness; Economic Policy & Law; Environmental Economics & Policy; International Development; Income and Wealth; Growth; Government Expenditure; Inequality	2 - with source data	2 - official results	2 - not available	1 - yes	

UNE NÉCESSITÉ

Une rigueur exigée par notre profession !

► Intégrité scientifique

Typologie indicative des manquements à l'intégrité scientifique

Les manquements à l'intégrité reconnus comme les plus graves et relevant de la fraude sont :

- > La fabrication de données, ...
- > La falsification : ...
- > Le plagiat : ...

Outre ces trois types de manquements de nature clairement frauduleuse, on reconnaît une « zone grise » de pratiques non conformes au respect de l'intégrité scientifique, notamment :

- > Des pratiques concernant les données, le matériel et les résultats comme par exemple :
 - L'usage de méthodes non reproductibles pour la production ou l'analyse des données ;
 - Le mauvais usage des méthodes statistiques, à dessein (alors proche de la falsification) ou par incompetence ;

....

....

Extrait du document de politique d'Intégrité scientifique
INRAE, janvier 2021.

UNE NÉCESSITÉ

Une rigueur exigée par notre profession !

► Intégrité scientifique

Référent intégrité scientifique

L'intégrité scientifique se réfère à l'ensemble des règles et valeurs qui régissent l'activité scientifique et en garantissent le caractère **honnête**. Elle est indispensable à la **crédibilité** de la science et à la **confiance** que lui accorde la société. On désigne, en général les **méconduites scientifiques** par la **trilogie des fraudes FFP** (Fabrication de données, Falsification de données, Plagiat), complétée d'une large « zone grise », qui comprend, entre autres, des **comportements critiquables** relatifs aux publications ou encore des expertises faites en cachant des liens d'intérêt.

Le Référent Intégrité Scientifique (RIS) anime la Mission à l'Intégrité Scientifique (MIS), qui intervient d'une part dans le traitement des allégations de méconduites et d'autre part, de façon concertée avec le référent déontologue, dans la formation et la sensibilisation des personnels aux questions d'intégrité scientifique et de déontologie.

<https://www.cnrs.fr/fr/ethique-deontologie-integrite-scientifique-et-lancement-dalerte>,
7 février 2020

OBJECTIF :

- ▶ Proposer des principes et pratiques

OBJECTIF :

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

OBJECTIF :

- ▶ 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.

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 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 laminair 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 plié en deux par la moitié, puis redouverte.



8 Riguer la pâte avec un rouleau plaque 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 plié en deux, puis en la plaçant entre 2 plaques. Idem avec la troisième abaisse de pâte. Nous nous retrouvons donc avec au total 4 plaques à pâtisserie empilées les unes sur les autres, entre lesquelles se trouvent 3 abaisses de pâte feuilletée, 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 étages...



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)

 <p>22 ...et blanchir la préparation au fouet.</p>	 <p>23 Ajouter la poudre à crème...</p>	 <p>24 ...et bien mélanger...</p>
 <p>25 ...Jusqu'à son incorporation complète.</p>	 <p>26 Lorsque le lait entre en ébullition...</p>	 <p>27 ...Le verser en une seule fois sur les œufs blanchis...</p>
 <p>28 ...en remuant à l'aide d'un fouet.</p>	 <p>29 Transvaser l'appareil obtenu dans la casserole de cuisson du lait...</p>	 <p>30 ...et cuire la préparation en la fouettant constamment de sorte que la crème n'attache pas au fond de la casserole.</p>
 <p>31 Elle va se mettre à épaissir au bout de quelques secondes. Poursuivre la cuisson à feu modéré pendant 2 à 3 minutes.</p>	 <p>32 Au terme de la cuisson débarrasser la crème dans un récipient frais.</p>	 <p>33 Ajouter le moitié du beurre coupé en morceaux dans la crème chaude.</p>
 <p>34 Bien mélanger avec le fouet.</p>	 <p>35 Une fois la crème refroidie (à température ambiante), ajouter le restant de beurre pommadé...</p>	 <p>36 ...en l'incorporant au batteur électrique ou au fouet à main. Réserver jusqu'au moment de l'utilisation.</p>
 <p>37 Découper les 3 pâtes feuilletées en cercle de dimensions identiques. Commencer par la première plaque que l'on découpe ici en un carré de 20 x 20 cm.</p>	 <p>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 genoise est recommandé pour scier la pâte et non l'écraser.</p>	 <p>39 On obtient ainsi nos 3 plaques de pâte feuilletée cuites, aux mêmes dimensions.</p>
 <p>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.</p>	 <p>41 Poser sur le dessus une seconde plaque de pâte feuilletée.</p>	 <p>42 Bien ajuster les bords, les uns au dessus des autres.</p>
 <p>43 Pocher des cordons de crème mousseline sur la seconde plaque, de la même façon que l'étape inférieure.</p>	 <p>44 Veiller à bien les accoler les uns contre les autres, et ne pas laisser des espaces sans crème.</p>	 <p>45 Déposer délicatement la troisième plaque de pâte feuilletée sur la crème et presser légèrement pour bien faire adhérer.</p>

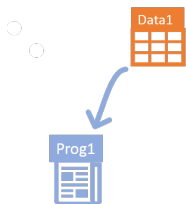
HOW TO MAKE A PIE ? BESOIN DE LA RECETTE (CODE)

 <p>49 Faire chauffer au bain-marie...</p>	 <p>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.</p>	 <p>51 Ajouter 1 ou 2 cuillères à soupe d'eau chaude.</p>
 <p>52 Bien mélanger.</p>	 <p>53 Contrôler la température de fonte avec un thermomètre à vitre laser, pas plus de 40°C.</p>	 <p>54 Prélever la valeur de 2 cuillères à soupe de fondant...</p>
 <p>55 ...et le colorer avec un peu d'extrait de café (Trabù).</p>	 <p>56 Bien mélanger.</p>	 <p>57 Le verser dans un cornet à décor.</p>
 <p>58 Bien replier les bords...</p>	 <p>59 ...afin d'éviter au fondant de s'échapper du cornet durant son utilisation.</p>	 <p>60 Réserver dans un endroit tiède. Il faut que le fondant reste liquide.</p>
 <p>61 Verser le fondant blanc sur le mille-feuilles.</p>	 <p>62 Le répartir sur toute sa surface.</p>	 <p>63 L'étaler rapidement avec une spatule métallique coudée en une épaisseur régulière et fine autant que possible (2 à 3 mm).</p>
 <p>64 Veiller à étaler le fondant jusqu'aux bords. Il faut que ce travail soit effectué rapidement, avant que le fondant ne se solidifie.</p>	 <p>65 Avec le cornet à décor, tracer des lignes parallèles espacées de 2 cm environ.</p>	 <p>66 Ce travail doit également se faire rapidement, tant que le fondant est encore liquide.</p>
 <p>67 Tirer les lignes avec la pointe d'un couteau, pour terminer le décor.</p>	 <p>68 Tirer toutes les lignes dans un sens pour commencer...</p>	 <p>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.</p>
 <p>70 Faire adhérer sur tout le pourtour du mille-feuilles des amandes caramélisées préalablement broyées grossièrement au couteau.</p>	 <p>71 Bien les faire adhérer pour que les côtés soient entièrement recouverts, qu'il ne reste pas de vide.</p>	 <p>72 Faire ainsi sur tout le pourtour du gâteau.</p>
 <p>73 Avant de le trancher, il est conseillé de le laisser se raffermir une heure minimum au frais. Pour le couper, je conseille d'utiliser un couteau scie à génoise.</p>	 <p>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.</p>	

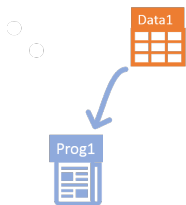
HOW TO MAKE A PIE ? EN ÉCO APPLIQUÉE / STAT...



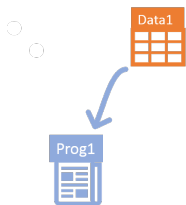
HOW TO MAKE A PIE ? EN ÉCO APPLIQUÉE / STAT...



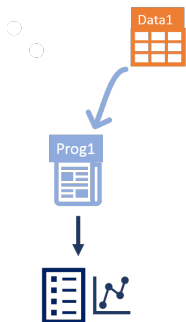
HOW TO MAKE A PIE ?



HOW TO MAKE A PIE ?



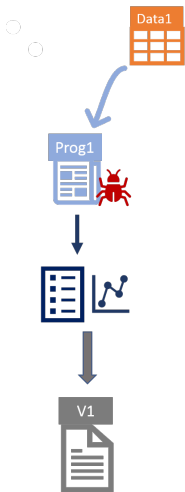
HOW TO MAKE A PIE ?



HOW TO MAKE A PIE ?



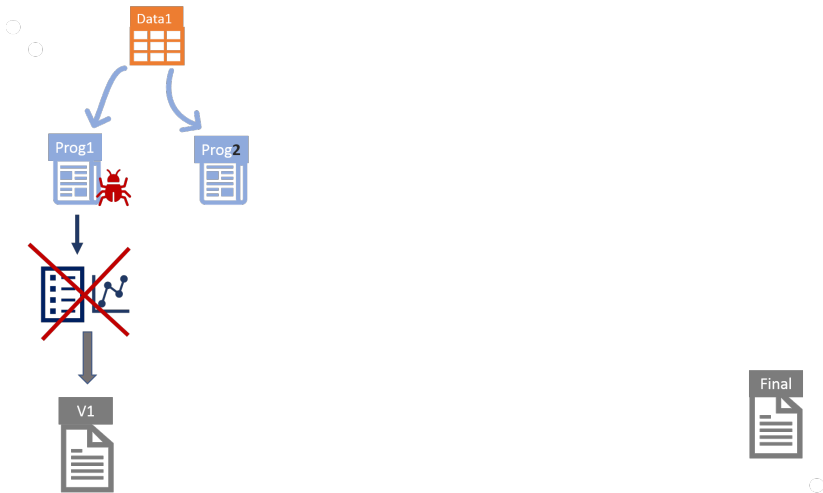
HOW TO MAKE A PIE ?



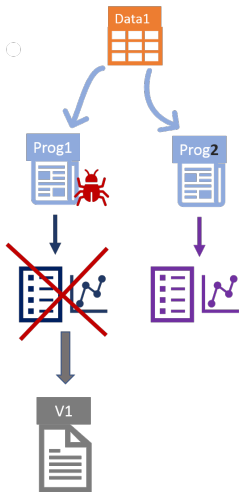
HOW TO MAKE A PIE ?



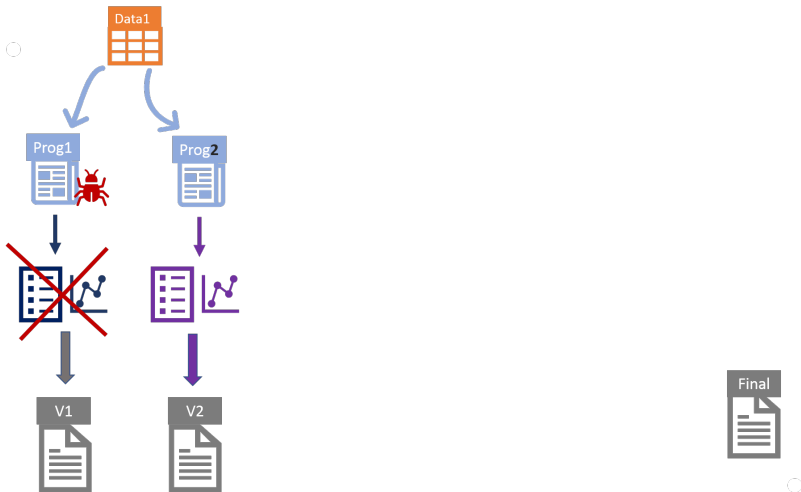
HOW TO MAKE A PIE ?



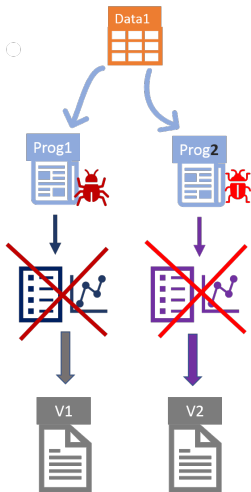
HOW TO MAKE A PIE ?



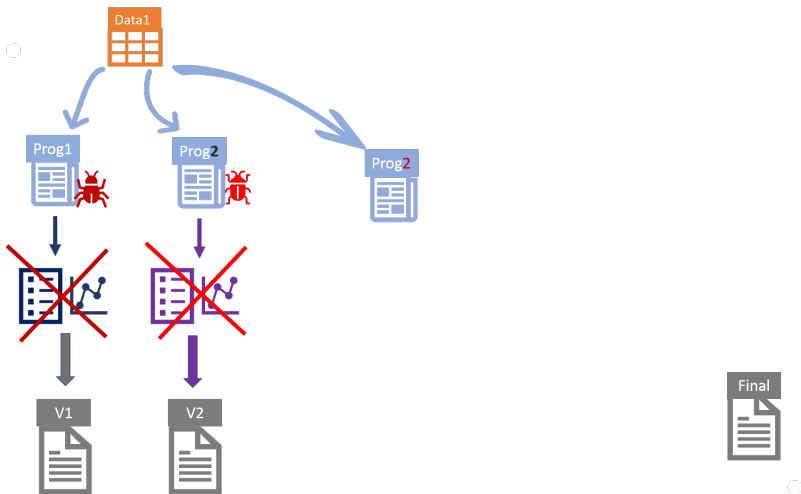
HOW TO MAKE A PIE?



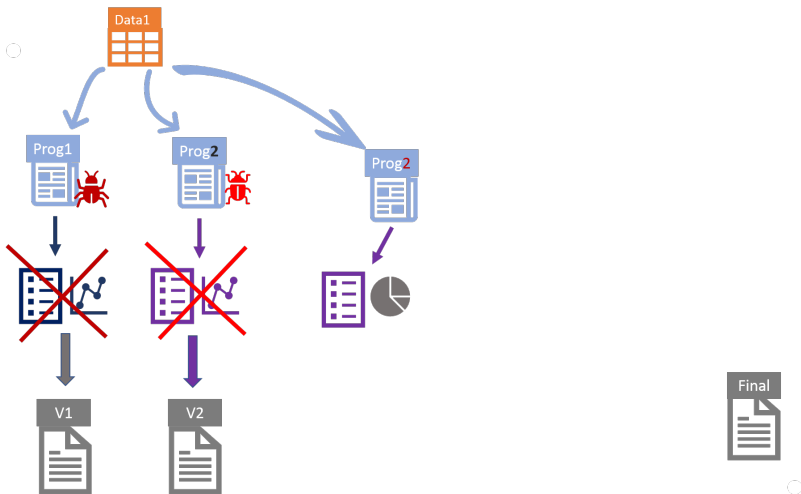
HOW TO MAKE A PIE ?



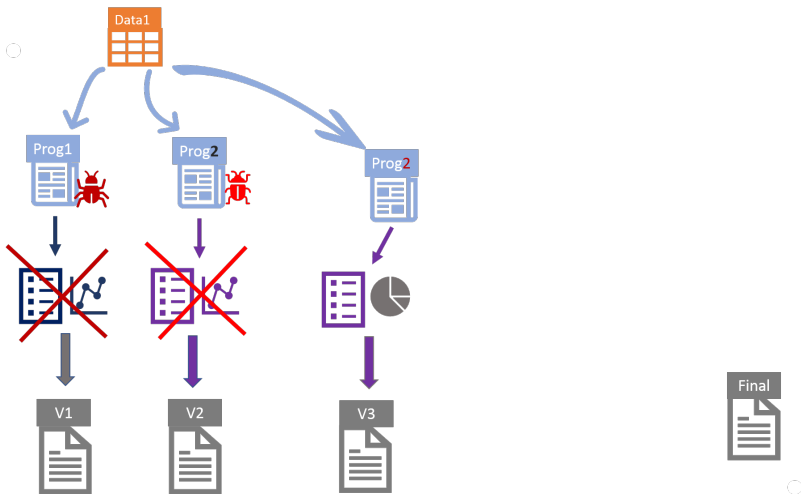
HOW TO MAKE A PIE ?



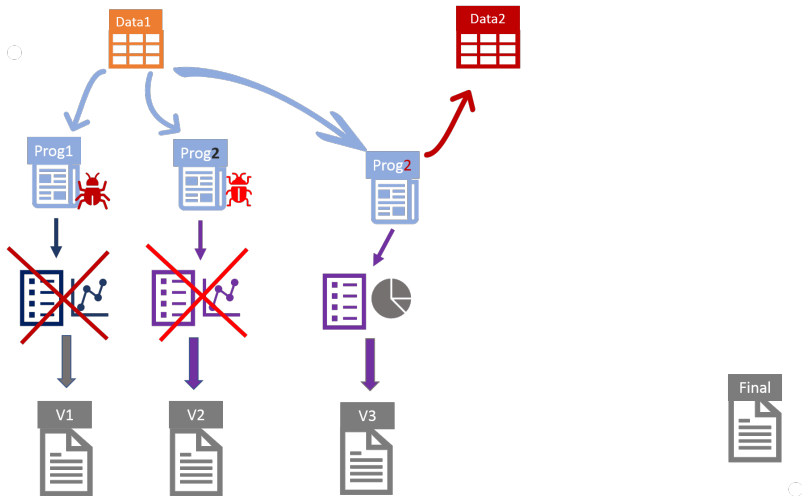
HOW TO MAKE A PIE?



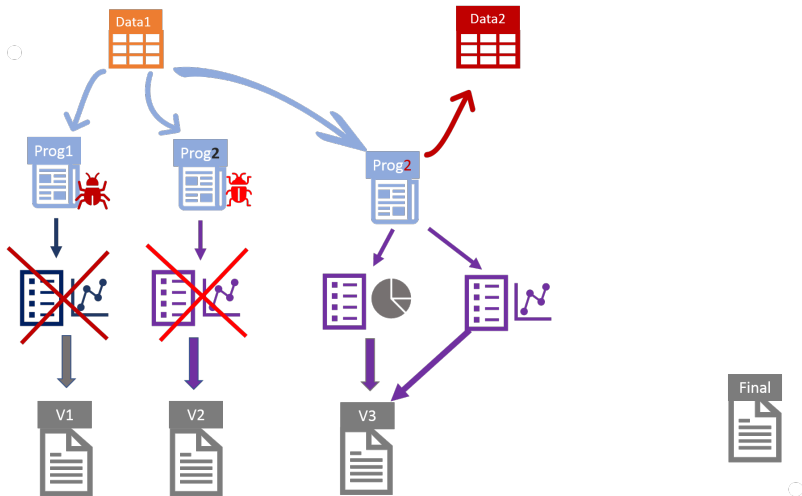
HOW TO MAKE A PIE?



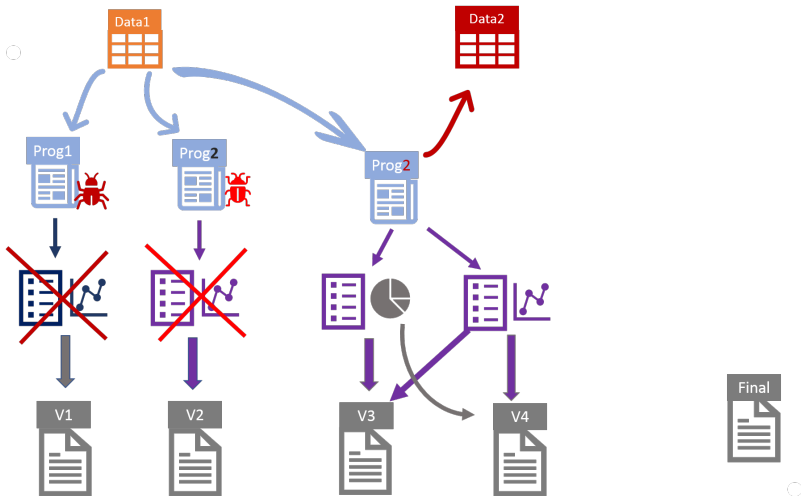
HOW TO MAKE A PIE?



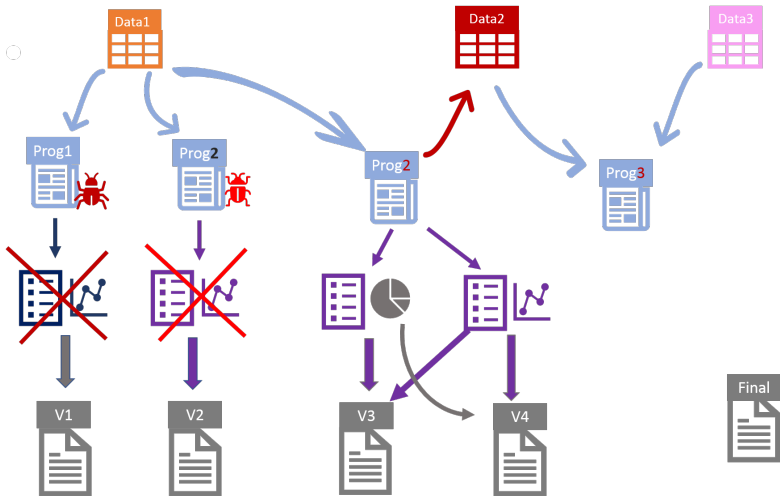
HOW TO MAKE A PIE?



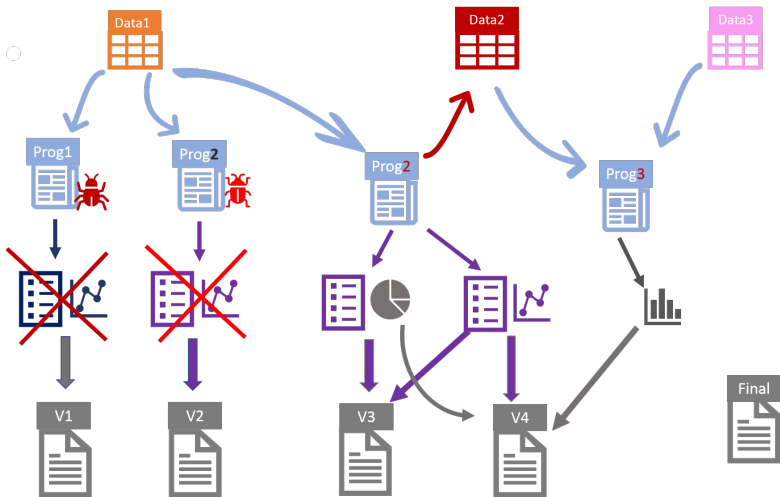
HOW TO MAKE A PIE?



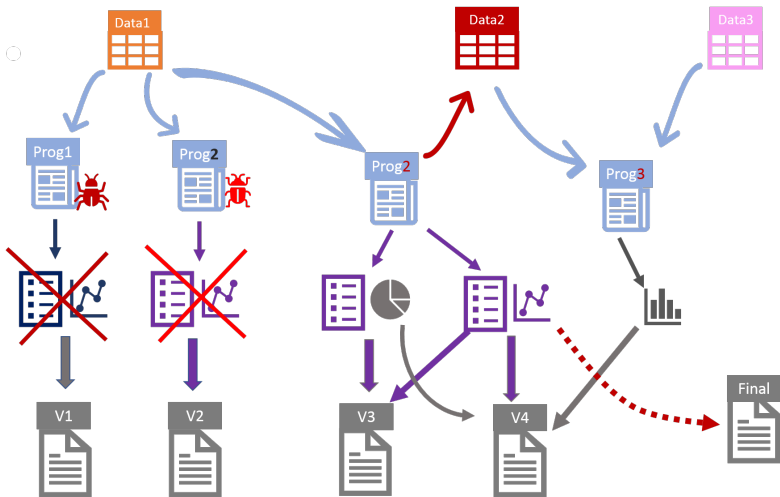
HOW TO MAKE A PIE?



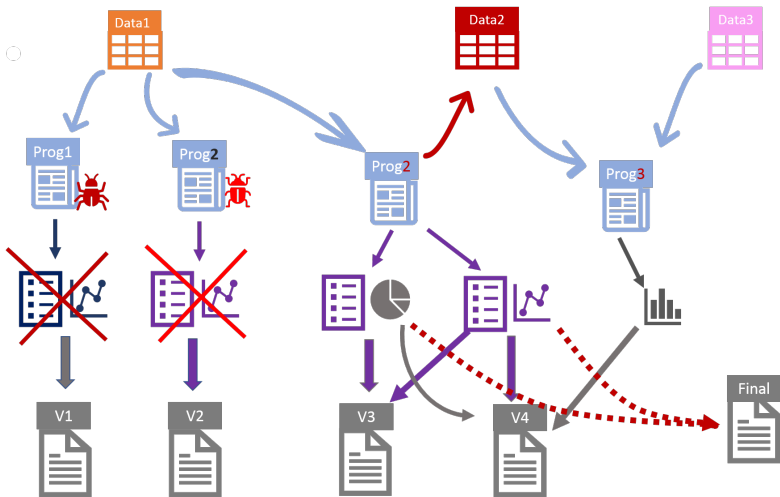
HOW TO MAKE A PIE?



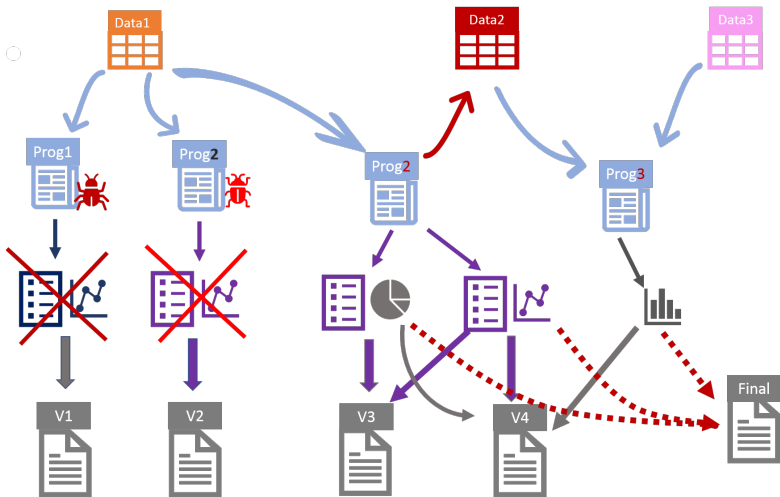
HOW TO MAKE A PIE ?



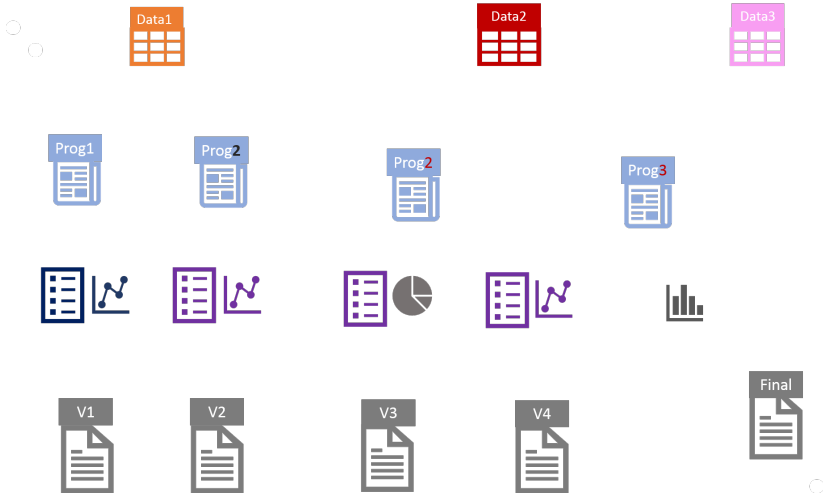
HOW TO MAKE A PIE ?



HOW TO MAKE A PIE ?



2 MOIS PLUS TARD ...



2 MOIS PLUS TARD ...



Data



DataOld



MyProject



MyProject2



MyProject19-10



MyProject19-11



MyProjectFinal



MyProjectFinal2_
_DO_NOT_ERASE



MyProjectOld



NewData

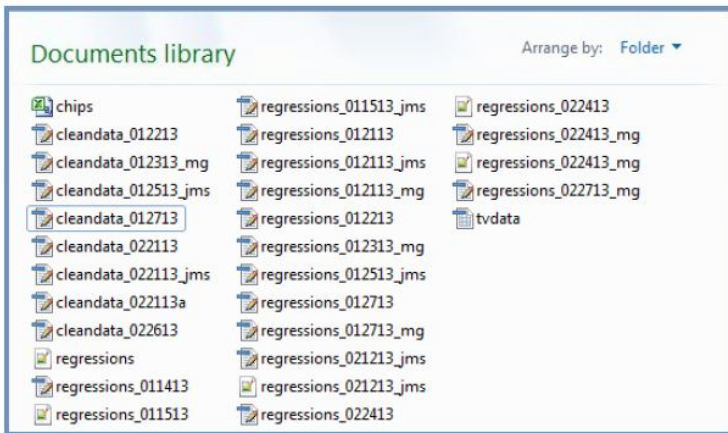


Results



ResultsGOOD

2 MOIS PLUS TARD ...

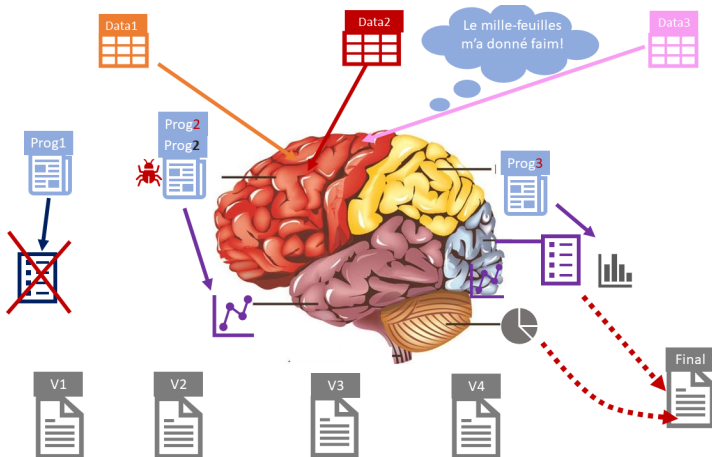


De notre propre expérience et Gentzkow and Shapiro (2014)

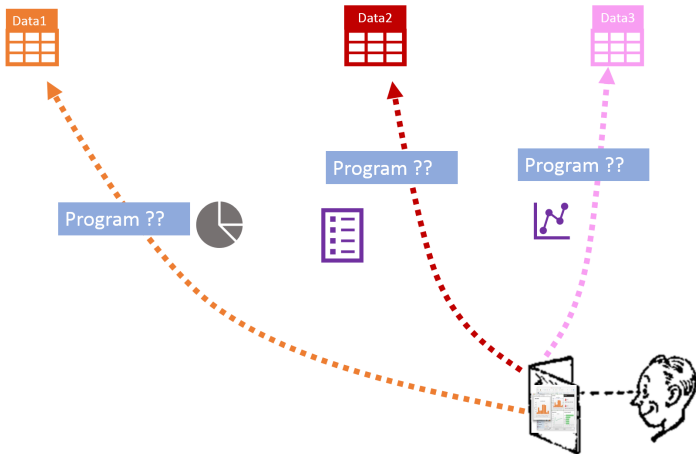
2 MOIS PLUS TARD ...



CERTAINS DIRONT POUVOIR COMPTER SUR LEUR MÉMOIRE...



DU POINT DE VUE DU 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,
 - ▶ Idée de 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

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

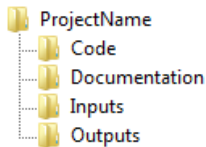


Figure – Exemple d'une arborescence bien organisée.

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()
```

PRINCIPE 1 : ORGANISER SON TRAVAIL

Utiliser des conventions de nommage

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

PRINCIPE 1 : ORGANISER SON TRAVAIL

Utiliser des conventions de nommage

- 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

- 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

- ▶ 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

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

Better

```
02_stats_desc_table.tex  
03_regress_modell1_table.tex  
03_regress_modell1_fig.JPG
```

PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitrisez toutes les étapes

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

PRINCIPE 1 : ORGANISER SON TRAVAIL

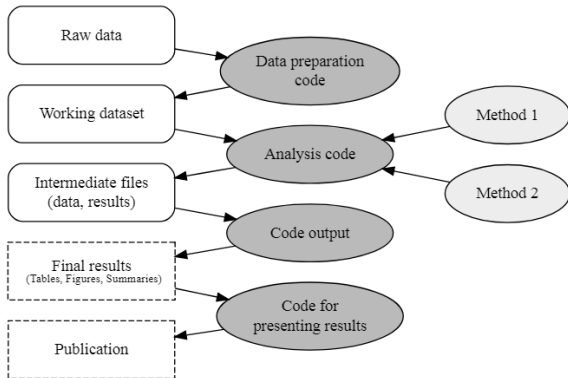
Maitrisez toutes les étapes

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

PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitrisez toutes les étapes

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



PRINCIPE 1 : ORGANISER SON TRAVAIL

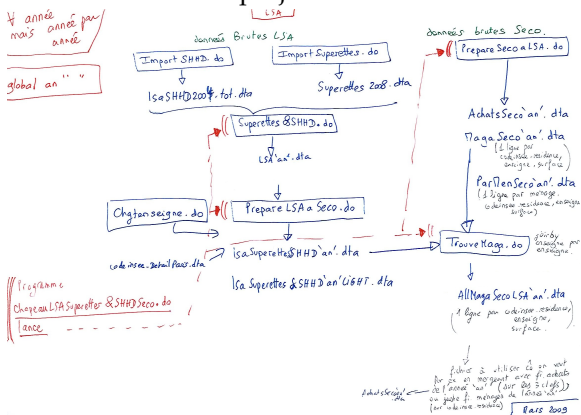
Maitrisez toutes les étapes

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

PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitrisez toutes les étapes

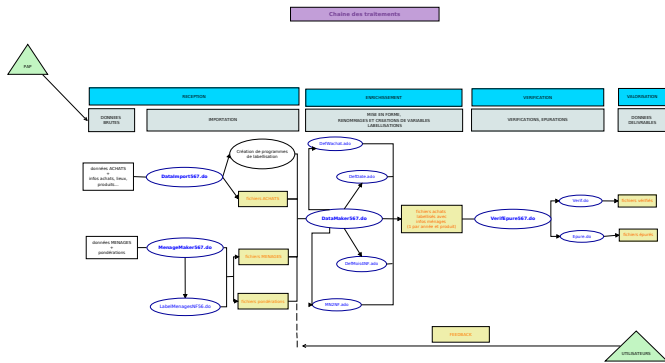
- ▶ Garder/enregistrer chaque étape (pas de copier/coller, pas de logiciel clic bouton). Code!
- ▶ Gérer/Maitriser le workflow
- ▶ Dessiner le workflow du projet *A la main*



PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitrisez toutes les étapes

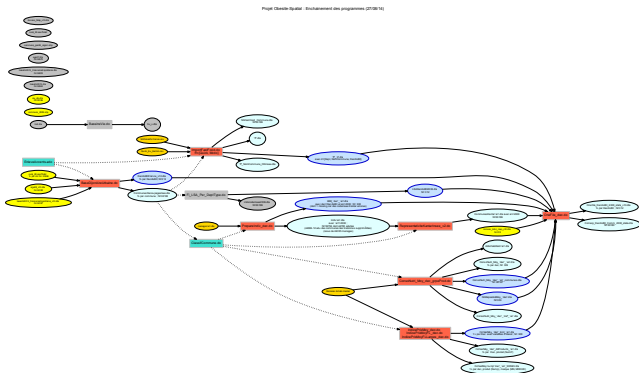
- ▶ Garder/enregistrer chaque étape (pas de copier/coller, pas de logiciel clic bouton). Code!
- ▶ Gérer/Maitriser le workflow
 - ▶ Dessiner le workflow du projet *Avec un logiciel (Dia)*



PRINCIPE 1 : ORGANISER SON TRAVAIL

Maitrisez toutes les étapes

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



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 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 GitHub, GitLab, SourceSup (Renater)
- ▶ “Version control is like an undo command for everything”,

(Gentzkow and Shapiro, 2014)

Compare • master • Changes • History Pull request Sync

Loop to create Mfx tables with CI 1 year ago by XtopheB

Complete file with Bootstrapped MFX 1 year ago by XtopheB

With hetero specified 1 year ago by XtopheB

Document de compilation des résultats 1 year ago by XtopheB

Calcul de la fenetre CV 1 year ago by XtopheB

Sorties triées dans l'ordre croissant des fenêtres... 1 year ago by XtopheB

Sorties LaTeX included. 1 year ago by XtopheB

Sortie MFX 1 year ago by XtopheB

Avec MFX 1 year ago by XtopheB

Version complète des programmes 1 year ago by XtopheB

VersionDebut Aout 1 year ago by XtopheB

With hetero specified

XtopheB 424d13c

```
... LoopSpecial.do
... @@ -32,11 +32,13 @@ gen Special = -(prix_2008 - mprix_2008)
32 32
33 33
34 34 /* Relation i_tap et Special */
35 35 +/*
36 36 lpoly i_tap Special, noscatter degree(0) ci lineopts(icolor(cranberry)) ///
37 37 clopts(recast(rare) fintensity(30)) ///
38 38 title(Relation between decision variable and special regressor) scheme(sj)
39 39 graph export "Graphs/RelationD-V.pdf", replace
40 40 +/*
41 41
42 42
43 43 global exog "i_under10 log_income i_town i_car b08_locenv_water a2_age i_can i_fra"
44 44 global endog "isatis_health"
... @@ -81,7 +83,10 @@ set linesize 88
81 83 "quietly Monsspecialreg2 i_tap Special, exog($exog) endog($endog) iv($instrument) hetero
kdens band(1.00) trim('t') `w'
82 84
83 85 /* On lance aussi une version avec ordered choice : Band= 999 */
84 85 -quietly Monsspecialreg2 i_tap Special, exog($exog) endog($endog) iv($instrument) hetero trim
('t') `w'
86 86 +
87 87 +quietly Monsspecialreg2 i_tap Special, exog($exog) endog($endog) iv($instrument) ///
88 88 hetero hetv(i_under10 log_income i_town b08_locenv_water) ///
89 89 + trim('t') `w'
90 90
91 91 /* Avec Silverman (0.23) , 3 choix de fenetres et CV (1.004) */
```

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 :

- ▶ Utiliser des conventions de nommage

PRINCIPE 2 :

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

Faciliter la lisibilité et compréhension du code :

- ▶ Utiliser des conventions de nommage

Usual

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


PRINCIPE 2 :

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

Faciliter la lisibilité et compréhension du code :

- Utiliser des conventions de nommage

Usual

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

Better

```
gen I_Men = (gender == 101)

GDP_New <- gdp /1.02
Height <- "180cm"
```

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(NI)>0) */
/* Version 1.3 13/02/2008 Modification du fichier ProductNF56.dta si produit sans obs. */
/* Version 1.4 7/03/2008 Initialisation locale varlistanee_sans_prem */
/* Version 1.41 26/03/2008 drop de an */
/* Version 1.42 22/04/2008 Fusion Marques56 intégrée et MW2NF 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.dta 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 pbts 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 Gencode)*/
/*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 crée 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 labélisation 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.ado
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(x1000)
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)
TMEAD	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).

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)
```

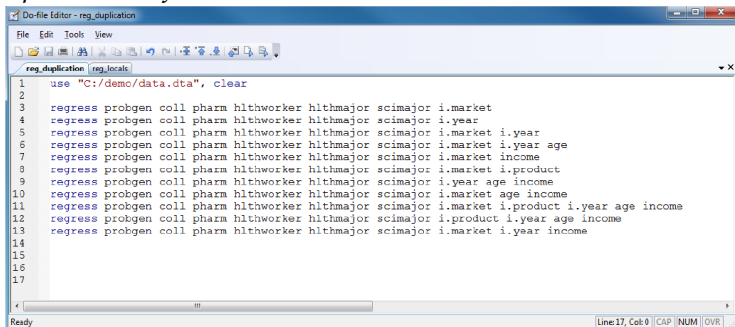
Figure – Exemple de code générique (fonction R)

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)

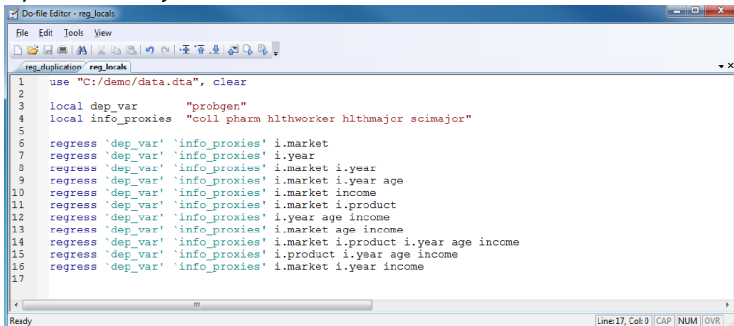


```
Do-file Editor - reg_duplication
File Edit Tools View
reg_duplication reg_locals
1 use "C:/demo/data.dta", clear
2
3 regress probgen coll pharm hlthworker hltbmajor scimajor i.market
4 regress probgen coll pharm hlthworker hltbmajor scimajor i.year
5 regress probgen coll pharm hlthworker hltbmajor scimajor i.market i.year
6 regress probgen coll pharm hlthworker hltbmajor scimajor i.market i.year age
7 regress probgen coll pharm hlthworker hltbmajor scimajor i.market income
8 regress probgen coll pharm hlthworker hltbmajor scimajor i.market i.product
9 regress probgen coll pharm hlthworker hltbmajor scimajor i.year age income
10 regress probgen coll pharm hlthworker hltbmajor scimajor i.market age income
11 regress probgen coll pharm hlthworker hltbmajor scimajor i.market i.product i.year age income
12 regress probgen coll pharm hlthworker hltbmajor scimajor i.product i.year age income
13 regress probgen coll pharm hlthworker hltbmajor scimajor i.market i.year income
14
15
16
17
Ready
Line17, Col 0 | CAP NUM OVR
```

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_locals
File Edit Tools View
reg_duplication reg_locals
1 use "C:/demo/data.dta", clear
2
3 local dep_var "probgem"
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
Ready
Line:17, Col:0 CAP NUM OVR
```

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 logiciel clic-bouton,...)

PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

- ▶ Tout coder! (pas de manipulation à la main, pas de copier-coller, pas de logiciel 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 logiciel 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 logiciel 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 logiciel 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
(exogenous ones "Xexo", endogenous 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 savefprefix(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
```


PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

Coder toutes les sorties de résultats

► Exporter les résultats et insertion RR dans un document

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

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)

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(200,2,10,40)
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    10   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")
```

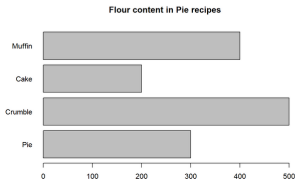


Figure – Exemple d'un document reproductible.

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

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


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")
...

```

## 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

$$Sweight = \frac{(a+c) \cdot (b+d)}{2}$$

And draw the graphic:


```
{r, results="hide"}
barplot(Pie.Table$Flour, col = "gray",
  names.arg = rownames(Pie.Table),
  horiz = TRUE, las = 0,
  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

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

And draw the graphic:

```

barplot(Pie.Table$Flour, col = "gray",
 names.arg = rownames(Pie.Table),
 horiz = TRUE, las = 0,
 main = "Flour content in Pie recipes")

```

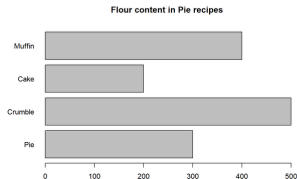


Figure – Exemple d'un document reproductible.

# EGALEMENT POSSIBLE DANS D'AUTRES LOGICIELS (PAR EX. STATA)

```

title: Reproducible document with Stata and Markdown
author: C. Bonhomme, V. Desautel
date: December 2019
bibliography: RR16.bib

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 better file for references is easy! (see [McCullough2009]).

Data

***/
local prod "0212"
qui use ".../data/p/prod/men.dta", clear
qui verify lib*
global lib "`*' (lib)"
qui sum annee
local p "`*' (mean)"
local hobs "`*' (H)"
...

We used data from the THS-Worldpanel (2008) on "`*' lib_`*'=" consumption for the year "`*'". The data contains "`*' hobs`'" observations.
***/
qui bysort CODE_DEPT : keep if _n==1
qui keep CODE_DEPT qtdsep lib
qui merge 1:1 CODE_DEPT using ".../data/PopDep99.dta", keepusing(pop99)
qui drop if _merge==1
qui drop _merge
qui gen qtdsep# = qtdsep / pop99
qui sum qtdsep# qtdsep
...
qui merge 1:1 CODE_DEPT using ".../data/fcsh.dta", keepusing(id X_CHF_LIQU Y_CHF_LIQU)
qui drop if _merge==1
qui drop _merge
...

Map

***/
qui cd ../data
qui sgmap qtdsep# using fccodest if id<97, id(id) fcolor(purple) title("Purchased Quantity/Inhabitant lib#") ///
subnote(("France 2008 per dept (ISSEE)")) ///
legend(on) legend(1) legend(ring(1) position(3)) legend(title(Quantity in Kg on L)) ///
scalewidth(units(800) scale(1/1000) aspect(80) label(8)lines(4))
qui cd ../prog
qui graph export map.png, width(400) replace
...
! Nice Map! map.png

Estimations

***/
qui use ".../data/p/prod/men.dta", clear
qui gen sum(hab==0)

Our main argument is that the BMI depends on the quantity of "`*' lib_`*'=" we eat. The models we are going to estimate are the following:
***/
nl_1=beta_0 + \beta_1/cdot X_1 + \nu_1residual_1
***/

qui estato clear
qui estato : reg BMI1 qtdsep#len zu mifi
qui estato : reg BMI1 qtdsep#len zu mifi ageind1
...
esttab , non s2 ar(2) label nonuser title(Estimations) xtitles("mod 1" "mod 2") addnote(("Source : lib data") b(4) s(1) replace

```

## Reproducible document with Stata and Markdown

C. Bonhomme, V. Desautel  
December 2019

### Introduction

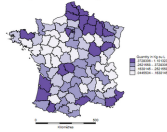
This small paper aims at providing some trends (geographical, time) about consumption. <br/>
 First section will introduce the data and the product we study. Then we will see a map in order to analyse if some regional behaviours exist. <br/>
 Section 3 will present some estimations. Let's see that using a better file for references is easy! (see [McCullough 2009]).

### Data

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

### Map

Purchased Quantity/Inhab. CHAMPAGNE  
(France 2008 per dept (ISSEE))



Nice 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:

$$BMI = \beta_0 + \beta_1 \cdot X_1 + \dots + \nu$$

estimation	mod 1	mod 2
probden	0.300***	0.300
IV	0.181**	0.181**
IV - 95%	0.111**	0.111**
IV - 90%	0.084**	0.084**
Observ	24,287***	21,287***
Observes	0.322	0.313
Observes - standard	0.152	0.149
Observations	1381	1381
Source : CHAMPAGNE_0212		
> prob.05. >> prob.01. >> prob.001		

That's all for now!

### Bibliography

McCullough, B. D. 2009. "Open Access Economic Journals and the Market for Reproducible Economic Research." Economic Inquiry Policy 39 (1): 117-28. <https://doi.org/10.1002/eic.1117>. URL

## 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...*)



# PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

## Créer un document reproductible (papier exécutable)

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

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**

- ▶ Encapsuler l'environnement de travail (conteneurs, par ex. *Docker*; systèmes de gestion de paquets, par ex. *Guix*)

## PRINCIPE 3 : AUTOMATISER LE PLUS POSSIBLE

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

- ▶ Encapsuler l'environnement de travail (conteneurs, par ex. *Docker*; systèmes de gestion de paquets, par ex. *Guix*)
- ▶ Partager des notebooks (*Binder*, *Guix-Jupyter*)

# POUR CORSER LE TOUT : TRAVAILLER AVEC MES CO-AUTEURS

## Un peu + difficile!

- ▶ Définir ensemble des conventions (Conventions de nommage, Arborescence, ...)  $\leftrightarrow$  *pair programming*
- ▶ Centraliser les informations
  - ▶ Données  $\leftrightarrow$  (Dropbox/Framadrive), serveurs sécurisés, plateformes d'archivage (*Web repository* : FigShare, Zenodo,...)
  - ▶ Programmes  $\leftrightarrow$  (Dropbox), (Sharepoint), GitLab/SourceSup
  - ▶ Environment computationnel  $\leftrightarrow$  Docker, Binder
  - ▶ Bibliographie  $\leftrightarrow$  Endnote, Zotero, BibTeX
  - ▶ Document de travail/Article/Rapport  $\leftrightarrow$  Overleaf/PLMlateX(L<sup>A</sup>T<sub>E</sub>X), (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 : ÉCHANGER AVEC MES CO-AUTEURS

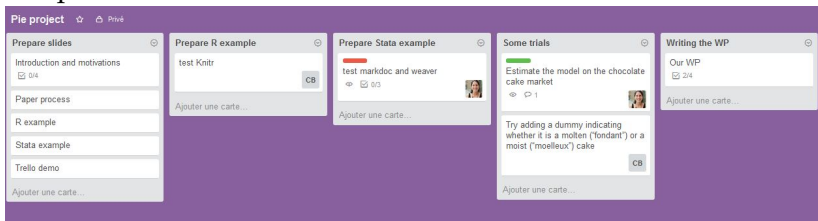
## Echange de messages / Gestion des tâches

- ▶ 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 (*Online management tasks system*) comme *Trello*, *Asana*, ...mieux que des post-it!





# 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, Exec&Share* (CNRS et ANR sont partenaires, (Hurlin et al., 2014)), *Binder, Code Ocean, WholeTale*

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

# A VENIR...

## 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

# A VENIR...

## 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, ...) devraient 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 Reproductible, piloté par l'INRIA), workshops, ...

# A VENIR...

## 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, ...) devraient 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 Reproductible, piloté par l'INRIA), workshops, ...
- ▶ Nous, chercheurs, ingénieurs, doctorants, enseignants, encadrants, referees doivent ouvrir la voie
- ▶ Vers une meilleure reproductibilité...

# A VENIR...

## 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, ...) devraient 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 Reproductible, piloté par l'INRIA), workshops, ...
- ▶ Nous, chercheurs, ingénieurs, doctorants, enseignants, encadrants, referees doivent ouvrir la voie
- ▶ Vers une meilleure reproductibilité...
- ▶ ...mais aussi plus de transparence (et 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.
- 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.
- Gentzkow, M. and Shapiro, J. M. (2014). Code and data for the social sciences : a practitioner's guide. University of Chicago mimeo.
- Hamermesh, D. S. (2013). Six decades of top economics publishing : Who and how ? *Journal of Economic Literature*, 51(1) :162–172.
- Horvath, J. (2013). The replication myth : Shedding light on one of science's dirty little secrets. *Scientific American*, December.
- Hurlin, C., Pérignon, C., and Stodden, V. (2014). Runmycode.org : a novel dissemination and collaboration platform for executing published computational results. *Open Science Framework*.

## REFERENCES II

- 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.
- 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).
- 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.
- 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.

## REFERENCES III

Pérignon, C., Gadouche, K., Hurlin, C., Silberman, R., and Debonnel, E. (2019). Certify reproducibility with confidential data. *Science*, 365(6449) :127–128.

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.