How to make a pie ? Reproducible Research for Empirical Economics and Econometrics

Christophe Bontemps & Valérie Orozco
Toulouse School of Economics, INRA

Porto - December, 2019

HAVE YOU EVER HEARD (OR SAID YOURSELF) ?

"But where the hell do these numbers come from ?"

HAVE YOU EVER HEARD (OR SAID YOURSELF) ?

"But where the hell do these numbers come from ?"

"I can't remember how I constructed this variable..."

HAVE YOU EVER HEARD (OR SAID YOURSELF) ?

"But where the hell do these numbers come from ?"

"What does this program exactly ?"

"I can't remember how I constructed this variable..."

HAVE YOU EVER HEARD (OR SAID YOURSELF) ?

"But where the hell do these numbers come from ?"

"I don't understand why I get a different result here..."

"What does this program exactly ?"

"I can't remember how I constructed this variable..."

HAVE YOU EVER HEARD (OR SAID YOURSELF) ?

"But where the hell do these numbers come from ?"

"I'm bored of redoing that stuff again. I did it once, so it must be correct..."

"I don't understand why I get a different result here..."

"What does this program exactly ?"

"I can't remember how I constructed this variable..."

Have you ever heard (or said yourself) ?

"But where the hell do these numbers come from ?"

"I left this work 6 months ago and I'm not able to find anything now..."

"I'm bored of redoing that stuff again. I did it once, so it must be correct..."

"I don't understand why I get a different result here..."

"What does this program exactly ?"

"I can't remember how I constructed this variable..."

HAVE YOU EVER HEARD (OR SAID YOURSELF) ?

"But where the hell do these numbers come from ?"

"I left this work 6 months ago and I'm not able to find anything now..."

"I'm bored of redoing that stuff again. I did it once, so it must be correct..."

"I don't understand why I get a different result here..."

"Where does this result/table come from ?"

"What does this program exactly ?"

"I can't remember how I constructed this variable..."

Replication failures in Science

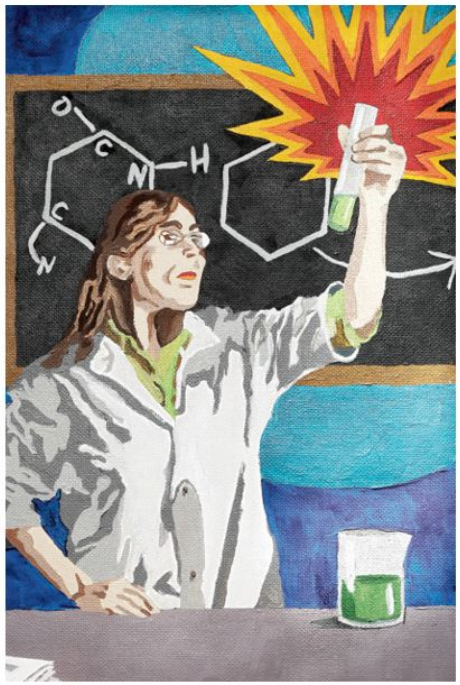
- In sciences in general : Nature (2013), (2017), (2018) *et al*...

Replication failures in Science

- In sciences in general : Nature (2013), (2017), (2018) *et al*...

Replication failures in Science

- In sciences in general : Nature (2013), (2017), (2018) *et al*...



RESEARCH POLICY

PUBLISHED ON JAN 15, 2014

BASIC RESEARCH

The Problem of Replication

It's disappointing when seemingly groundbreaking studied can't be repeated. But it's happening a lot.

IN 2012, WHEN THE JOURNAL *Science* **published a study** by a group of Cleveland researchers touting a cancer drug's success against Alzheimer's disease in mice, it looked like a potential breakthrough in a field long marked by setbacks and failures. As many as 5.4 million Americans have Alzheimer's, and by one estimate, there's a new case diagnosed every 69 seconds. While existing treatments can alleviate symptoms, there's nothing to stop the progression of the ultimately fatal disease.

But in this study, researchers from Case Western Reserve University School of Medicine described how a drug called bexarotene appeared to clear the buildup of amyloid-beta plaques—protein deposits in the brain that are hallmarks of Alzheimer's—in mice genetically engineered to have a condition similar to the human disease. The drug also seemed to improve the mice's memory, cognitive abilities and social behavior. "The plaque reduction was an astounding finding," says **Sangram Sisodia**, director of the Center for Molecular Neurobiology at the University of Chicago.

Replication failures in Science

- In sciences in general : Nature (2013), (2017), (2018) *et al*...

nature human behaviour

PERSPECTIVE

PUBLISHED: 10 JANUARY 2017 | VOLUME: 1 | ARTICLE NUMBER: 0021

OPEN

A manifesto for reproducible science

Marcus R. Munafò[1,2*], Brian A. Nosek[3,4], Dorothy V. M. Bishop[5], Katherine S. Button[6], Christopher D. Chambers[7], Nathalie Percie du Sert[8], Uri Simonsohn[9], Eric-Jan Wagenmakers[10], Jennifer J. Ware[11] and John P. A. Ioannidis[12,13,14]

Improving the reliability and efficiency of scientific research will increase the credibility of the published scientific literature and accelerate discovery. Here we argue for the adoption of measures to optimize key elements of the scientific process: methods, reporting and dissemination, reproducibility, evaluation and incentives. There is some evidence from both simulations and empirical studies supporting the likely effectiveness of these measures, but their broad adoption by researchers, institutions, funders and journals will require iterative evaluation and improvement. We discuss the goals of these measures, and how they can be implemented, in the hope that this will facilitate action toward improving the transparency, reproducibility and efficiency of scientific research.

Replication failures in Science

- In sciences in general : Nature (2013), (2017), (2018) *et al*...

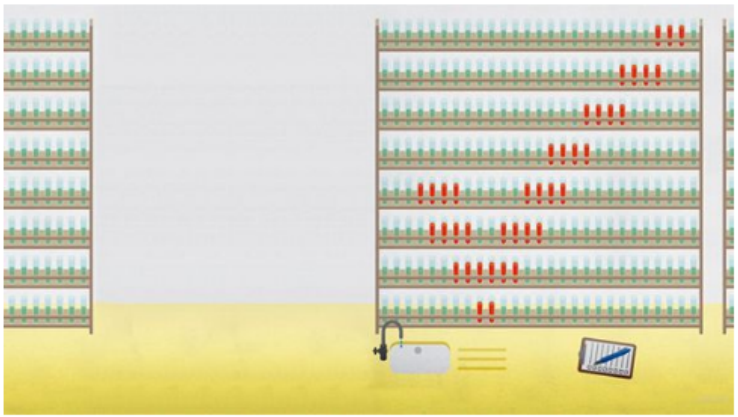
nature.com

Search Login

COLLECTION | 26 APRIL 2018

Scientific Rigour and Reproducibility

Science moves forward by corroboration – when researchers verify others' results. Progress is faster when people waste less time pursuing false leads. No research paper can ever be considered to be the final word, but there are too many that do not stand up to... show more



ERRORS AND DOUBTS

- Voluntarily errors (Stapel case of falsification of data, Psychology)

The New York Times | http://nyti.ms/unVj1J

RESEARCH

Fraud Case Seen as a Red Flag for Psychology Research

By BENEDICT CAREY NOV. 2, 2011

A well-known psychologist in the Netherlands whose work has been published widely in professional journals falsified data and made up entire experiments, an investigating committee has found. Experts say the case exposes deep flaws in the way science is done in a field, psychology, that has only recently earned a fragile respectability.

"In a system where there are few checks and balances, where people work alone, I took the wrong turn."
Dr Stapel (ex-phD)

ERRORS AND DOUBTS

- Voluntarily errors (Stapel case of falsification of data, Psychology)
  In Medecine

FIGURE – An analysis of Medline retractions 1988-2008 (`http://blogs.nature.com/news/2011/10/the_reasons_for_retraction.html`)

ERRORS MAY BE INVOLUNTARILY

Even on published papers !

ERRORS MAY BE INVOLUNTARILY

Even on published papers !

Excel Spreadsheet Error: Lessons from the Reinhart-Rogoff Controversy

How big can the errors get?

- In January 2010, academics Carmen Reinhart and Kenneth Rogoff published *"Growth in a Time of Debt"*.
- Their report was widely cited by politicians as theoretical and research based support for reducing public debt and public spending.
- Later analysis reveals errors with the underlying spreadsheet analysis; countries are excluded from the average because of a 'range error'.
- Great Brittan slashes spending by £10 billion, in response to the research and increase in debt following in the GFC.

	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
42	Germany	1946-2009	3.9	0.9	n.a.	n.a.	3.2
43	France	1949-2009	4.9	2.7	3.0	n.a.	5.2
44	Finland	1946-2009	3.8	2.4	2.5	n.a.	7.0
45	Denmark	1950-2009	3.5	2.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	3.1	n.a.	2.2
47	Belgium	1947-2009	n.a.	4.2	3.1	2.6	n.a.
48	Austria	1948-2009	5.2	3.3	2.8	n.a.	5.7
49	Australia	1951-2009	3.2	4.9	4.0	n.a.	5.9
50							
51			4.1	2.8	2.8	=AVERAGE(L30:L44)	

Source: Quartz website, http://qz.com/75119
Pasted from <http://qz.com/75119/how-to-avoid-making-an-excel-mistake-like-rogoff-and-reinhart/>

ERRORS MAY BE INVOLUNTARILY

Even on published papers !

- and even worse : not being able to reproduce results

ERRORS MAY BE INVOLUNTARILY

Even on published papers !

- and even worse : not being able to reproduce results
  - others results

ERRORS MAY BE INVOLUNTARILY

Even on published papers !

- and even worse : not being able to reproduce results
  - others results
  - or worse...

ERRORS MAY BE INVOLUNTARILY

Even on published papers !

- and even worse : not being able to reproduce results
  - others results
  - or worse...
  - own results !

WHAT ABOUT ECONOMICS AND ECONOMETRICS ?

- Dewald et al. (1988) tried to reproduce 54 articles from the "*Journal of Money, Credit and Banking*".

WHAT ABOUT ECONOMICS AND ECONOMETRICS ?

- Dewald et al. (1988) tried to reproduce 54 articles from the "*Journal of Money, Credit and Banking*".
- They succeeded in reproducing

WHAT ABOUT ECONOMICS AND ECONOMETRICS ?

- Dewald et al. (1988) tried to reproduce 54 articles from the "*Journal of Money, Credit and Banking*".
- They succeeded in reproducing

2 !

WHAT ABOUT ECONOMICS AND ECONOMETRICS ?

- Dewald et al. (1988) tried to reproduce 54 articles from the "*Journal of Money, Credit and Banking*".
- They succeeded in reproducing

  2 !
- Later McCullough and Vinod (2003),McCullough et al. (2006) and McCullough et al. (2008), tried...

WHAT ABOUT ECONOMICS AND ECONOMETRICS ?

- Dewald et al. (1988) tried to reproduce 54 articles from the "*Journal of Money, Credit and Banking*".
- They succeeded in reproducing

  2 !
- Later McCullough and Vinod (2003),McCullough et al. (2006) and McCullough et al. (2008), tried...
- McCullough deduced from those experiences :

WHAT ABOUT ECONOMICS AND ECONOMETRICS ?

- Dewald et al. (1988) tried to reproduce 54 articles from the "*Journal of Money, Credit and Banking*".
- They succeeded in reproducing

2 !

- Later McCullough and Vinod (2003),McCullough et al. (2006) and McCullough et al. (2008), tried...
- McCullough deduced from those experiences :

"ECONOMIC RESEARCH IS NOT REPLICABLE"

WHAT ABOUT ECONOMICS AND ECONOMETRICS ?

Chang and Li (2015) attempted to replicate **67** papers published in economics journals :

- Successfully replicate the key qualitative results of **33%** of the papers without contacting the authors.

WHAT ABOUT ECONOMICS AND ECONOMETRICS ?

Chang and Li (2015) attempted to replicate **67** papers published in economics journals :

- Successfully replicate the key qualitative results of **33%** of the papers without contacting the authors.
- Replicate **49%** of the papers with assistance from the authors

WHAT ABOUT ECONOMICS AND ECONOMETRICS ?

Chang and Li (2015) attempted to replicate **67** papers published in economics journals :

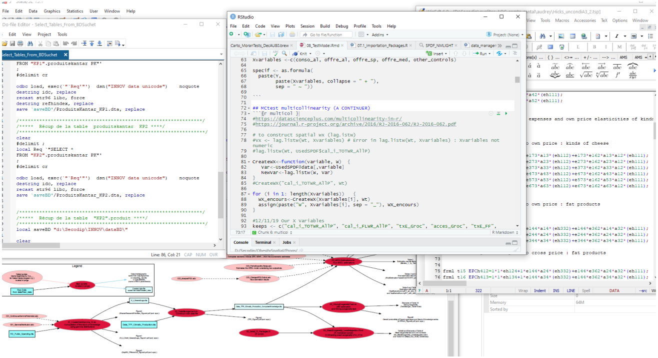
- Successfully replicate the key qualitative results of **33%** of the papers without contacting the authors.
- Replicate **49%** of the papers with assistance from the authors

"we assert that economics research is usually not replicable"

Because our job has changed !

Coding is increasing

- ► More and more data means more code !



BECAUSE OUR JOB HAS CHANGED !

Coding is increasing

- More and more data means more code !

↪ Every empirical researcher is now also a programmer !

BECAUSE OUR JOB HAS CHANGED !

Coding is increasing

- More and more data means more code !

↪ Every empirical researcher is now also a programmer !

↪ A lot of the time spent in writing, reading, and debugging code

BECAUSE OUR JOB HAS CHANGED !

Coding is increasing

- More and more data means more code !

↪ Every empirical researcher is now also a programmer !

↪ A lot of the time spent in writing, reading, and debugging code

- Decrease in theoretical articles in the 3-top economic journals (AER, QJE, JPE) from 51% in 1963 to 19% in 2011 Hamermesh (2013)

BECAUSE OUR JOB HAS CHANGED !

Coding is increasing

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

BECAUSE OUR JOB HAS CHANGED !

Coding is increasing

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

↪ But many researchers are unaware of the tools and best practices to write their code and learn-by-doing (Gentzkow and Shapiro (2014), Millman and Pérez (2014))

BECAUSE OUR JOB HAS CHANGED !

Coding is increasing

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

↪ But many researchers are unaware of the tools and best practices to write their code and learn-by-doing (Gentzkow and Shapiro (2014), Millman and Pérez (2014))

↪ ...leading to non-readability of the code, non reproducibility of the results and generation of errors

BECAUSE OUR JOB HAS CHANGED !

Collaboration is increasing

BECAUSE OUR JOB HAS CHANGED !

Collaboration is increasing

- Nb of authors/published paper in the top-5 economic journals has increased over time :

BECAUSE OUR JOB HAS CHANGED !

Collaboration is increasing

- Nb of authors/published paper in the top-5 economic journals has increased over time :
- On average : 1.3 authors/paper in 70's
- Now 2.2 authors/paper ! (Card and DellaVigna (2013))

BECAUSE OUR JOB HAS CHANGED !

Collaboration is increasing

- Nb of authors/published paper in the top-5 economic journals has increased over time :
- On average : 1.3 authors/paper in 70's
- Now 2.2 authors/paper ! (Card and DellaVigna (2013))

BECAUSE OUR JOB HAS CHANGED !

Collaboration is increasing

- Nb of authors/published paper in the top-5 economic journals has increased over time :
- On average : 1.3 authors/paper in 70's
- Now 2.2 authors/paper ! (Card and DellaVigna (2013))

↪ ... need of an improved work organisation, for sharing and exchanging files, emails and documentation writing.

BECAUSE OUR JOB HAS CHANGED !

Pressure is increasing

- Journals'policies include data and code sharing (AER, JAE,...) :

BECAUSE OUR JOB HAS CHANGED !

Pressure is increasing

- Journals'policies include data and code sharing (AER, JAE,...) :

Rank	Journal	Mandatory replication policy			
		2003	2009	2019	2019 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	–	–	YES	encourage (data + code) sharing
9	Rev Econ Studies	–	YES	YES	mandatory (data + code) sharing
10	J Econometrics	–	–	YES	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

BECAUSE OUR JOB HAS CHANGED !

Pressure is increasing

- Journals'policies include data and code sharing (AER, JAE,...) :

Rank	Journal	Mandatory replication policy			
		2003	**2009**	**2019**	2019 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	–	–	YES	encourage (data + code) sharing
9	Rev Econ Studies	–	YES	YES	mandatory (data + code) sharing
10	J Econometrics	–	–	YES	encourage (data + code) sharing

BECAUSE OUR JOB HAS CHANGED !

Pressure is increasing

- Journals'policies include data and code sharing (AER, JAE,...) :

Rank	Journal	Mandatory replication policy			
		2003	**2009**	**2019**	2019 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	–	–	YES	encourage (data + code) sharing
9	Rev Econ Studies	–	YES	YES	mandatory (data + code) sharing
10	J Econometrics	–	–	YES	encourage (data + code) sharing

Some even check/reproduce the results !

BECAUSE OUR JOB HAS CHANGED !

Pressure is increasing

- Publications of replications

BECAUSE OUR JOB HAS CHANGED !

Pressure is increasing

- Publications of replications
  - Some journals have a replication section

BECAUSE OUR JOB HAS CHANGED !

Pressure is increasing

- Publications of replications
  - Some journals have a replication section
  - Some journals are devoted to the publications of replications : ReScience (Rougier et al., 2017)

BECAUSE OUR JOB HAS CHANGED !

Pressure is increasing

- Publications of replications
  - Some journals have a replication section
  - Some journals are devoted to the publications of replications : ReScience (Rougier et al., 2017)
  - International Journal for Re-Views in Empirical Economics (IREE) founded in 2017

BECAUSE OUR JOB HAS CHANGED !

Pressure is increasing

- Publications of replications
  - Some journals have a replication section
  - Some journals are devoted to the publications of replications : ReScience (Rougier et al., 2017)
  - International Journal for Re-Views in Empirical Economics (IREE) founded in 2017
- A replication wiki in economics exists

BECAUSE OUR JOB HAS CHANGED !

Pressure is increasing

- Publications of replications
  - Some journals have a replication section
  - Some journals are devoted to the publications of replications : ReScience (Rougier et al., 2017)
  - International Journal for Re-Views in Empirical Economics (IREE) founded in 2017
- A replication wiki in economics exists

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	Keywords
Matthew Berg, Brian Hartley	Debt to GDP Ratios and Growth: Country Heterogeneity and Reverse Causation, the Case of Japan (Ultra Wonky)	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	Lowess regressions, distributed lag models	Macro data	Government debt and GDP ratios	Japan	Stata

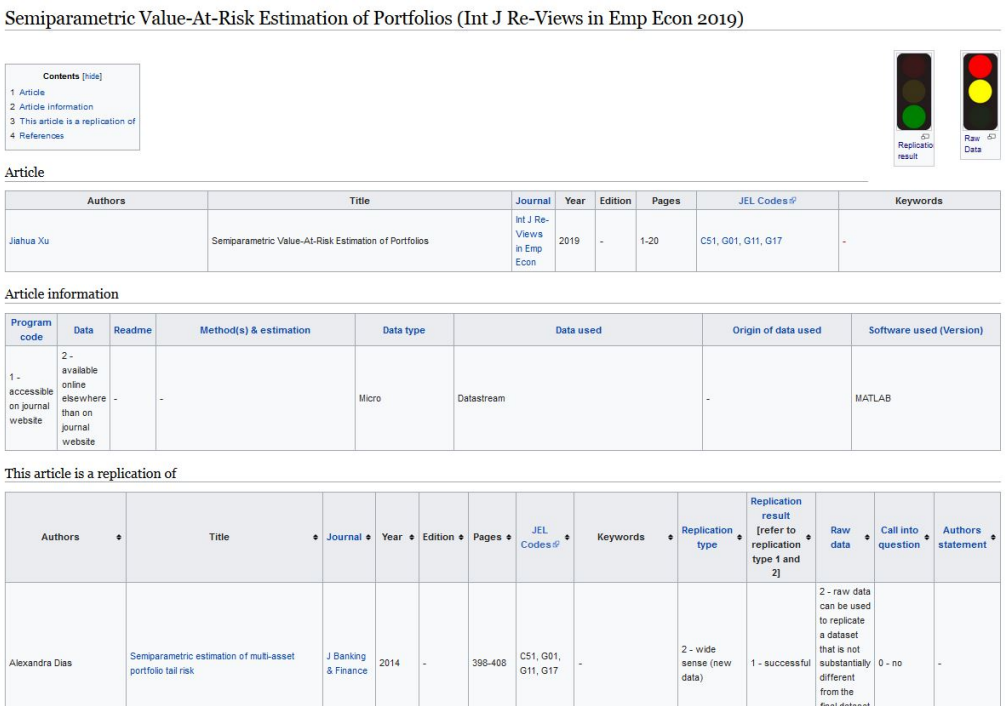
This article is a replication of

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	Authors statement
Carmen Reinhart, Kenneth Rogoff	Growth in a Time of Debt	AER	2010	1	-	E23, E31, H63, O11, O47	-	2 - wide sense (new data)	1 - successful	3 - final dataset can be exactly replicated	2 - partly	1 - rejected results

BECAUSE OUR JOB HAS CHANGED !

Pressure is increasing

- Publications of replications
  - Some journals have a replication section
  - Some journals are devoted to the publications of replications : ReScience (Rougier et al., 2017)
  - International Journal for Re-Views in Empirical Economics (IREE) founded in 2017
- A replication wiki in economics exists



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

Replicatio result | Raw Data

Article

Authors	Title	Journal	Year	Edition	Pages	JEL Codes	Keywords
Jiahua Xu	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	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 than on journal website	-	-	Micro	Datastream	-	MATLAB

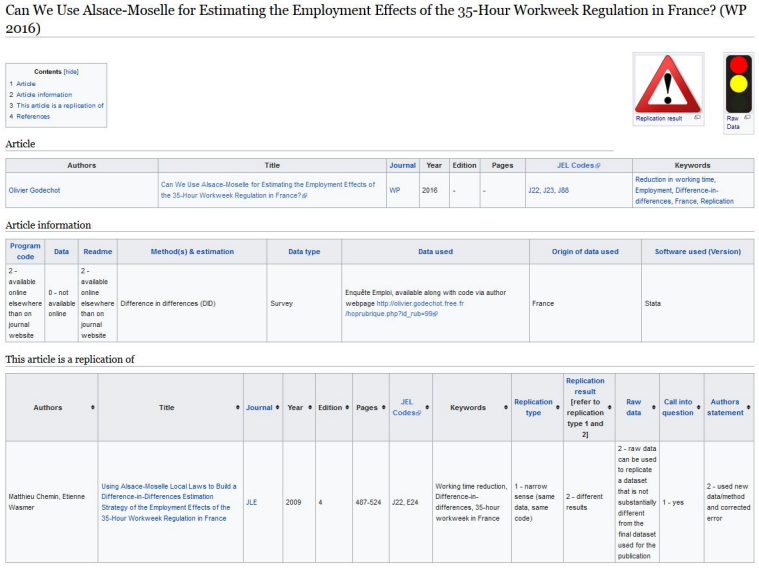
This article is a replication of

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	Authors statement
Alexandra Dias	Semiparametric estimation of multi-asset portfolio tail risk	J Banking & Finance	2014	-	398-408	C51, G01, 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	0 - no	-

BECAUSE OUR JOB HAS CHANGED !

Pressure is increasing

- Publications of replications
  - Some journals have a replication section
  - Some journals are devoted to the publications of replications : ReScience (Rougier et al., 2017)
  - International Journal for Re-Views in Empirical Economics (IREE) founded in 2017
- A replication wiki in economics exists



Can We Use Alsace-Moselle for Estimating the Employment Effects of the 35-Hour Workweek Regulation in France? (WP 2016)

Contents [hide]
1 Article
2 Article information
3 This article is a replication of
4 References

Replication result | Raw Data

Article

Authors	Title	Journal	Year	Edition	Pages	JEL Codes	Keywords
Olivier Godechot	Can We Use Alsace-Moselle for Estimating the Employment Effects of the 35-Hour Workweek Regulation in France?	WP	2016	-	-	J22, J23, J88	Reduction in working time, Employment, Difference-in-differences, France, Replication

Article information

Program code	Data	Readme	Method(s) & estimation	Data type	Data used	Origin of data used	Software used (Version)
2 - available online elsewhere than on journal website	0 - not available online	2 - available online elsewhere than on journal website	Difference in differences (DID)	Survey	Enquête Emploi, available along with code via author webpage http://olivier.godechot.free.fr/hoprubrique.php?id_rub=99	France	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	Authors statement
Matthieu Chemin, Etienne Wasmer	Using Alsace-Moselle Local Laws to Build a Difference-in-Differences Estimation Strategy of the Employment Effects of the 35-Hour Workweek Regulation in France	JLE	2009	4	487-524	J22, E24	Working time reduction, Difference-in-differences, 35-hour workweek in France	1 - narrow sense (same data, same code)	2 - different results	2 - raw data can be used to replicate a dataset that is not substantially different from the final dataset used for the publication	1 - yes	2 - used new data/method and corrected error

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

Contents [hide]
1 Article
2 Article information
3 This article is a replication of
4 References

Replication result | Raw Data

Article

Authors	Title	Journal	Year	Edition	Pages	JEL Codes	Keywords
William Easterly, Ross Levine, David Roodman	Aid, Policies, and Growth: Comment	AER	2004	3	774-780	-	-

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	Ordinary least squares (OLS)	Macro, Panel	Foreign aid, economic policies and growth of per capita GDP from World Bank	Sub-Saharan Africa and East Asian Countries	-

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	Authors statement
Craig Burnside, David Dollar	Aid, policies and growth	AER	2000	4	847-868	O19, F35, O23	Gender and development, Development economics & aid effectiveness, Economic theory & research, Environmental economics & policies, School health, Achieving shared growth, Governance indicators, Inequality	2 - wide sense (new data)	2 - different results	0 - not available at all	1 - yes	

Our Goal :

- Provide some principles and practices

Our Goal :

- Provide some principles and practices
- Show that simple solutions exist even in EEE

Our Goal :

- Provide some principles and practices
- Show that simple solutions exist even in EEE
- Prove that the costs are low, the benefit high

Our Goal :

- Provide some principles and practices
- Show that simple solutions exist even in EEE
- Prove that the costs are low, the benefit high

Our Goal :

- Provide some principles and practices
- Show that simple solutions exist even in EEE
- Prove that the costs are low, the benefit high

Note that the software is not always the problem and will never be the solution.

How to make a Pie ?

Research process is long !

- The process leading to a report or a research paper can be long and complex

Research process is long !

- The process leading to a report or a research paper can be long and complex
- Long publishing delay in economics and econometrics :

RESEARCH PROCESS IS LONG !

- The process leading to a report or a research paper can be long and complex
- Long publishing delay in economics and econometrics :

  According to Björk and Solomon (2013) economics journals have the longest publishing delay :

  17.70 months (sd 7.52)
  (bio-medicine 9.47, physics 1.93, mathematics 13.3, arts and letters 14.21)

RESEARCH PROCESS IS LONG !

- The process leading to a report or a research paper can be long and complex
- Long publishing delay in economics and econometrics :

  According to Björk and Solomon (2013) economics journals have the longest publishing delay :

  17.70 months (sd 7.52)
  (bio-medicine 9.47, physics 1.93, mathematics 13.3, arts and letters 14.21)

- Publishing takes time, but the whole research process is even longer !

How to make a Pie ? We need ingredients (data)

How to make a Pie ? We need the recipe (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.

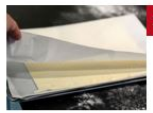
7 Placer la première abaisse de pâte sur une plaque à pâtisserie recouverte d'une feuille de papier sulfurisé pliée en deux par la moitié, puis réouverte.

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 pliée en deux, puis en la coinçant 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 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 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...

How to make a Pie ? We need the recipe (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.

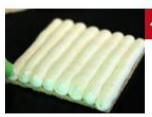

35 Une fois la crème refroidie (à température ambiante), ajouter le restant de beurre pommade...


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.

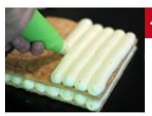

38 Positionner le carré obtenu sur la deuxième plaque et découper tout autour pour obtenir un second morceau aux mêmes dimensions. Faire de même avec la troisième plaque. L'usage d'un couteau à dents de type couteau-scie à génoise est recommandé pour scier la pâte et non l'écraser.


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 ? We need the recipe (code)

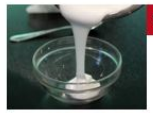

49 Faire chauffer au bain-marie...


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'extrait de café (Trablit).

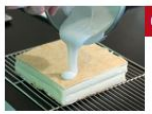

56 Bien mélanger.


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 coudée 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.

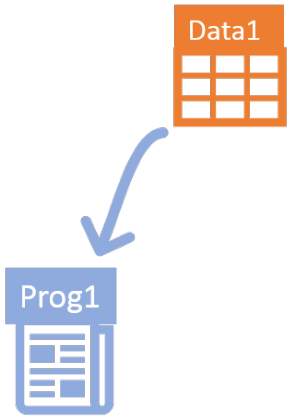

72 Faire ainsi sur tout le pourtour du gâteau.


73 Avant de le trancher, il est conseillé de le laisser se raffermir une heure minimum au frais. Pour le couper, je conseille d'utiliser un couteau scie à génoise.

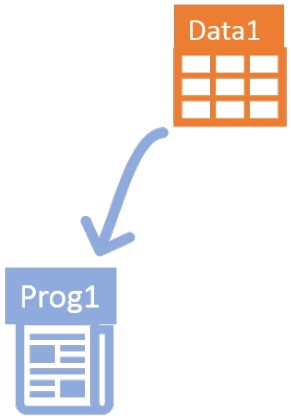

74 Il faut le couper en donnant des coups de scie d'avant en arrière. Il ne faut en aucun cas couper en appuyant verticalement sous risque d'écraser le mille-feuilles, de voir la crème sortir sur les côtés, et donc de perdre tout l'esthétique du gâteau.

How to make a pie ? In economics...

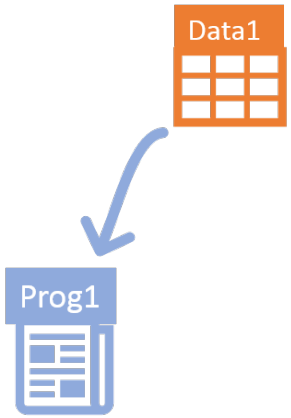
How to make a pie ? In economics...



How to make a pie ?

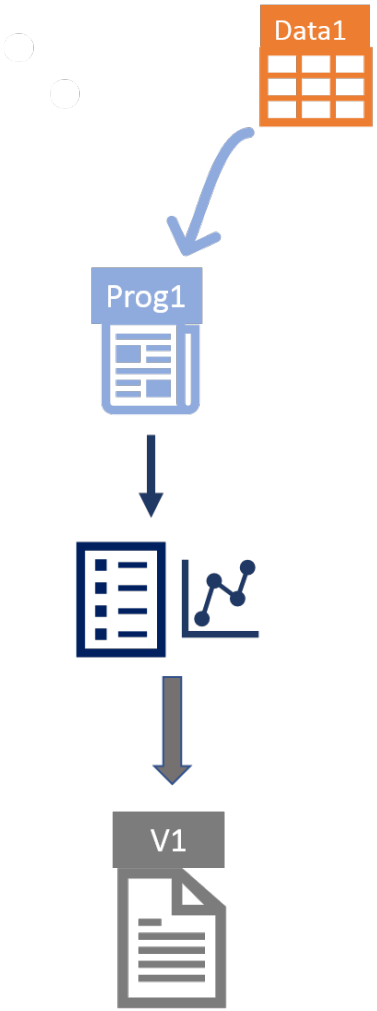


How to make a pie ?

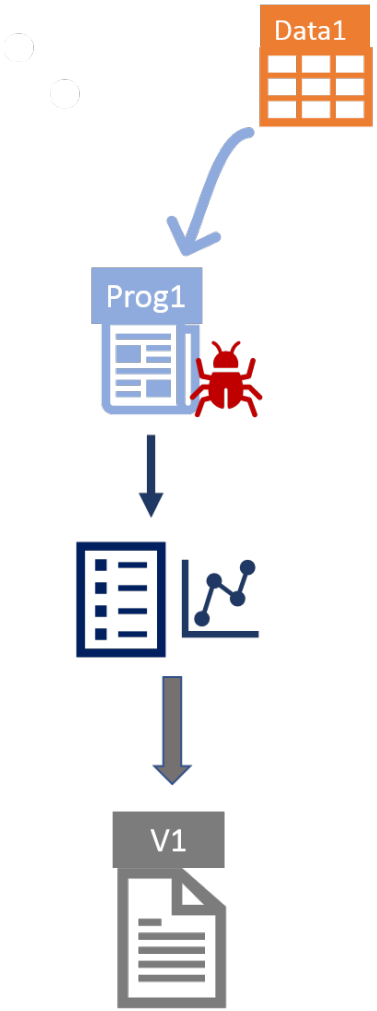


How to make a pie ?

How to make a pie ?

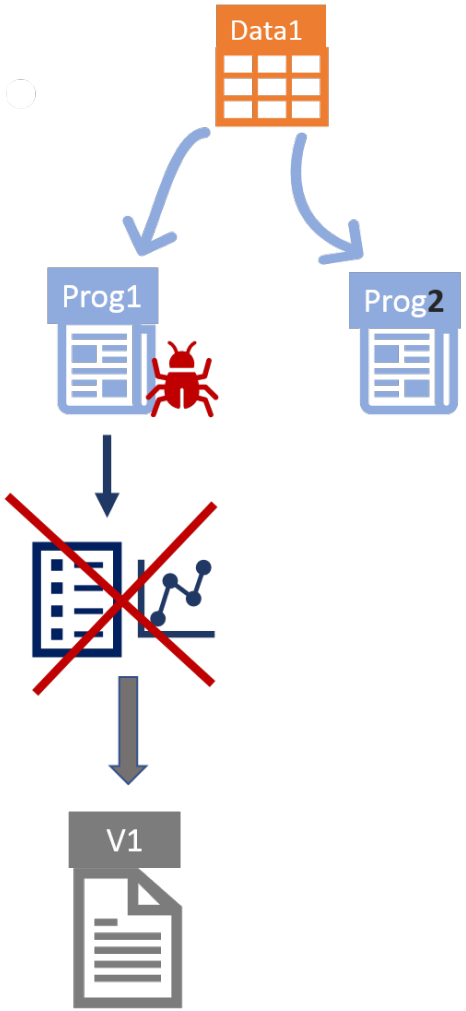


How to make a pie ?

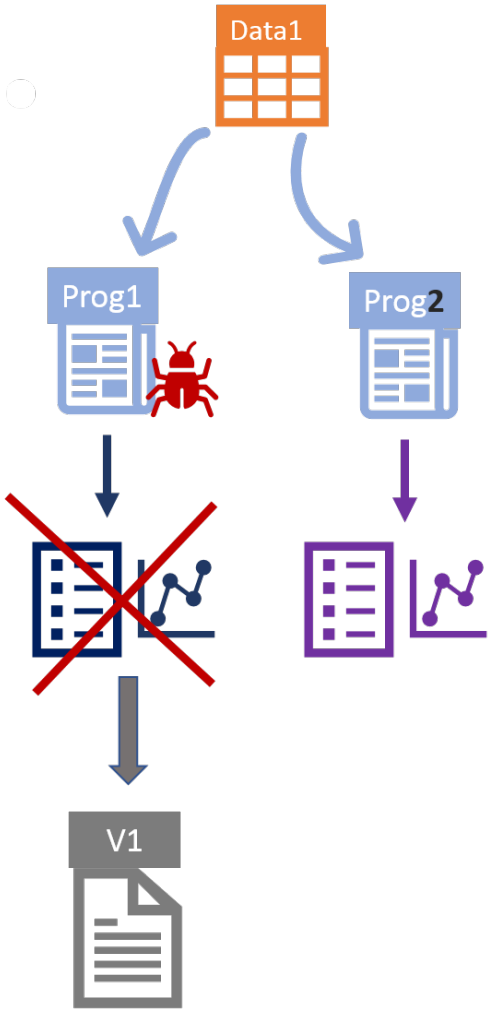


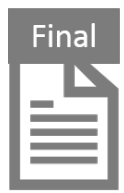
How to make a pie ?

How to make a pie ?

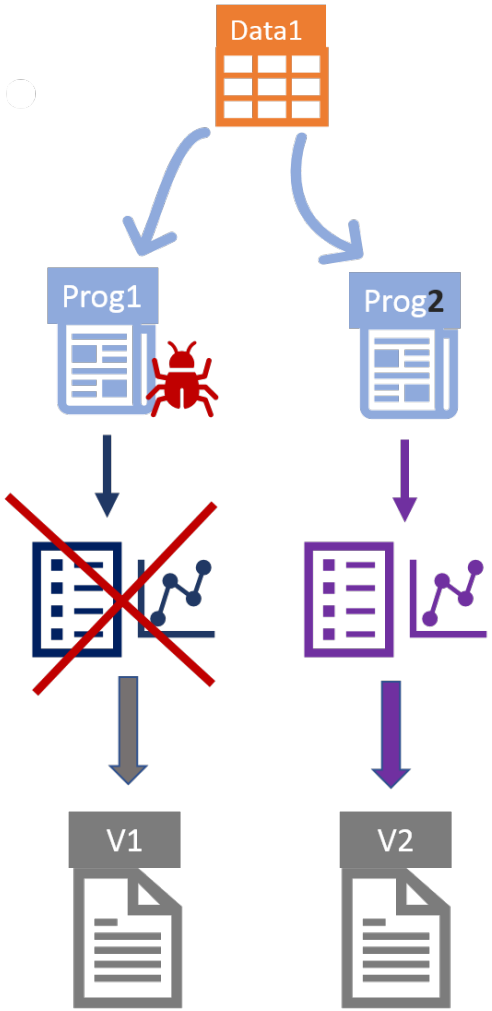


How to make a pie ?

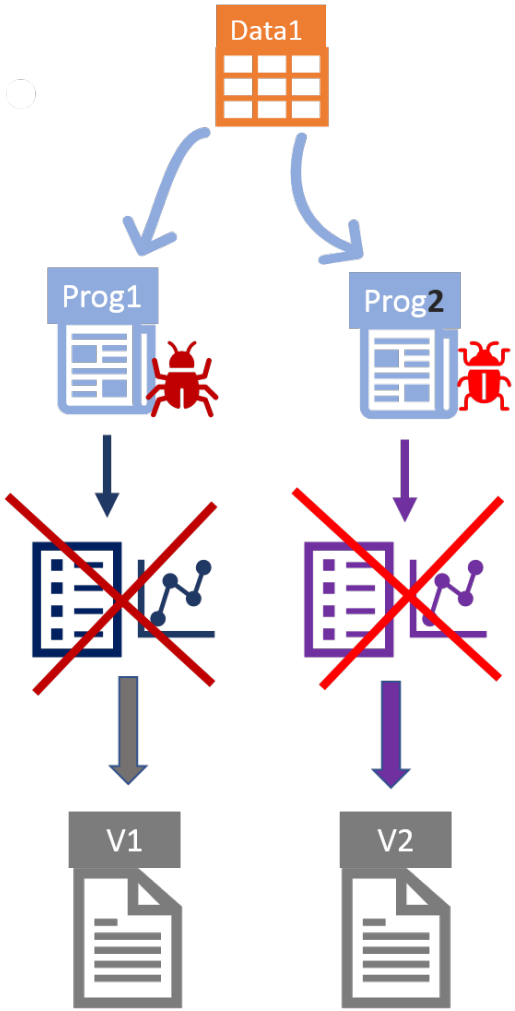




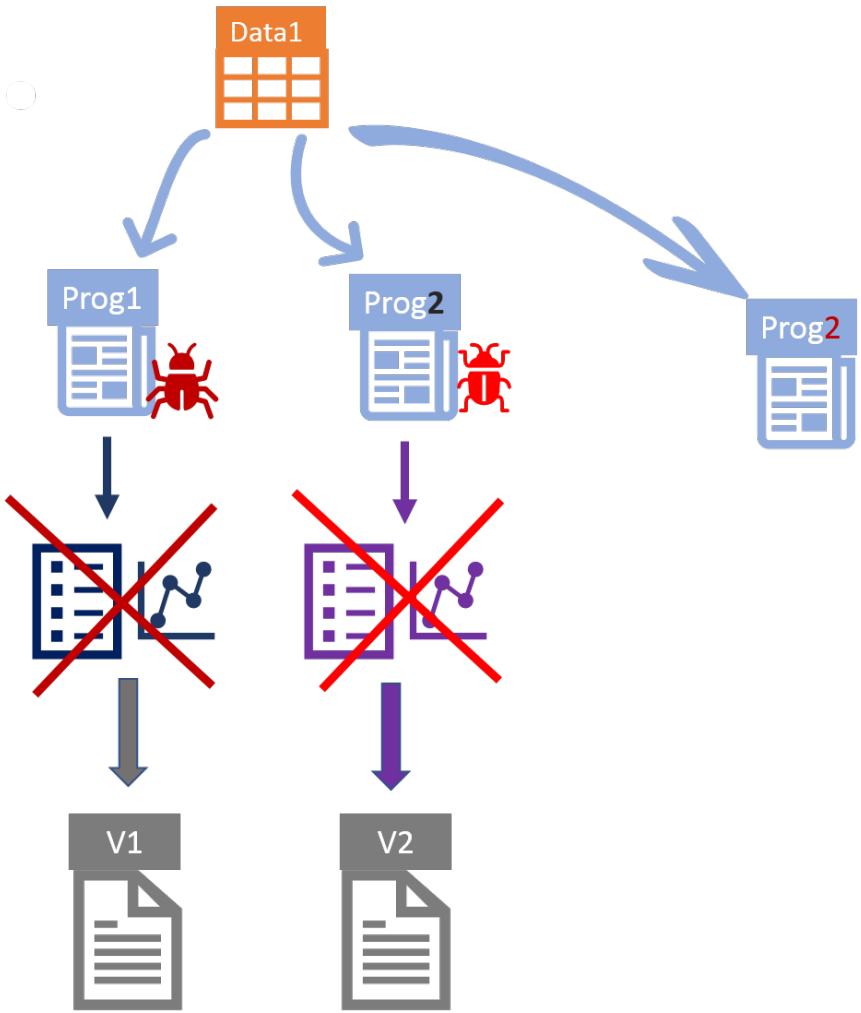
How to make a pie ?

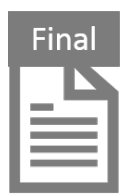


How to make a pie ?

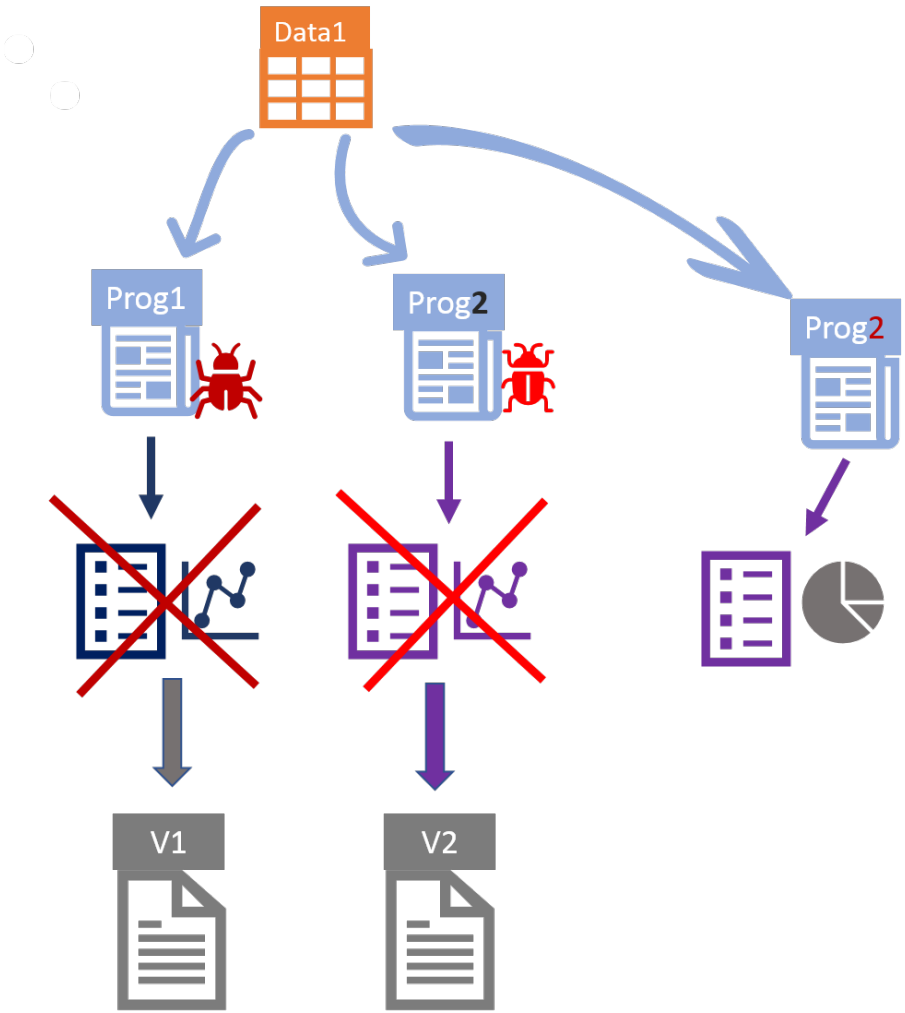


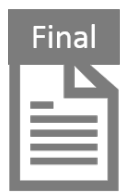
How to make a pie ?



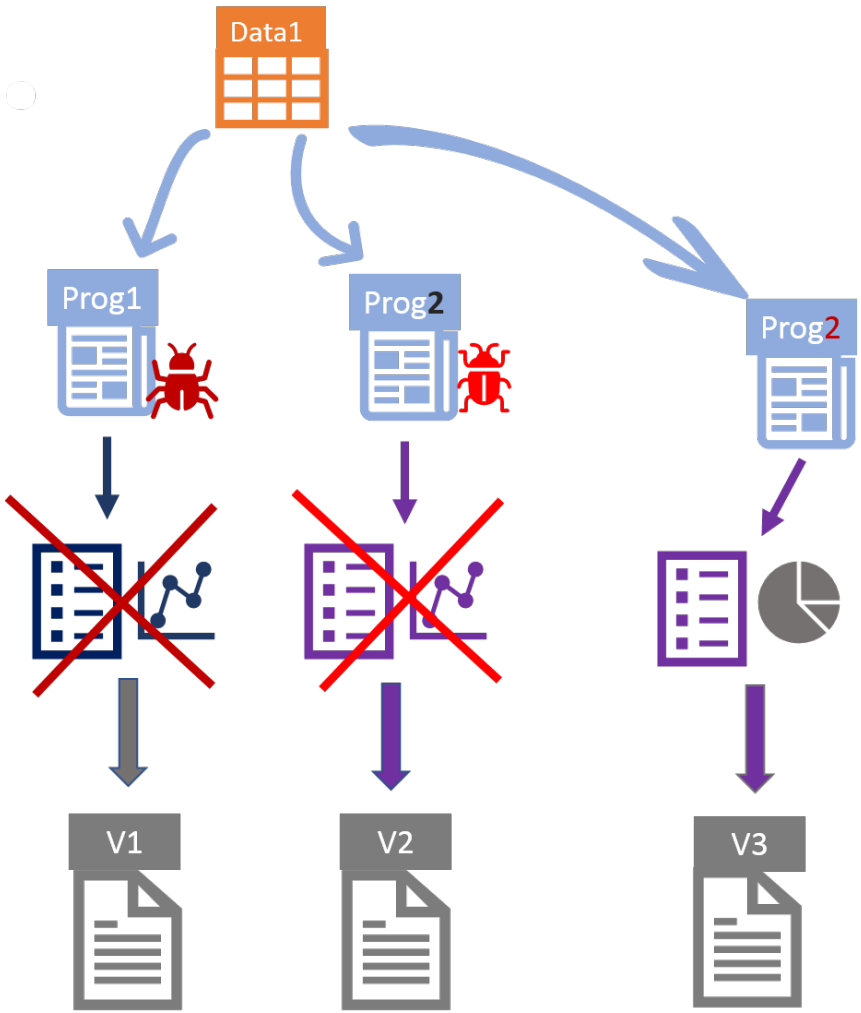


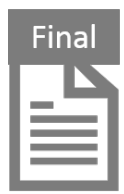
How to make a pie ?



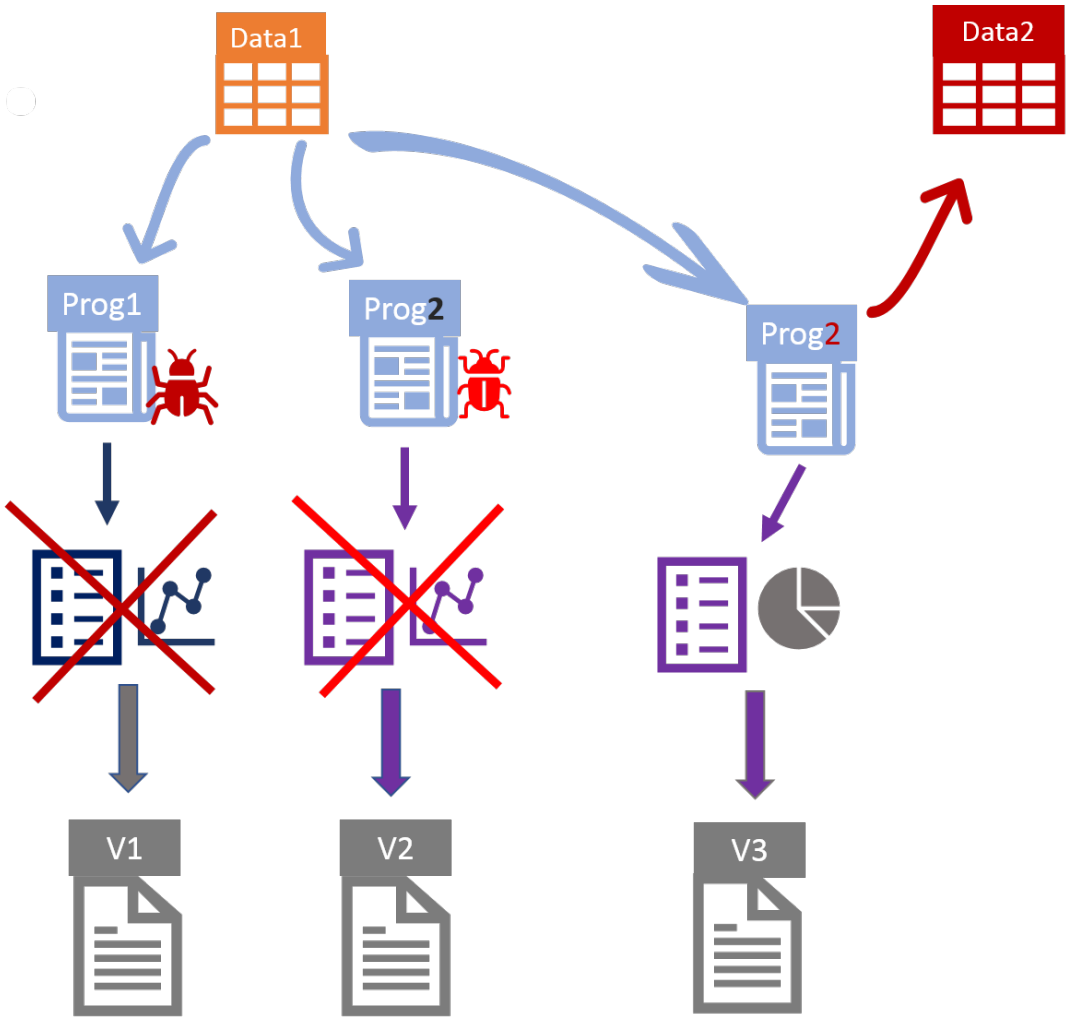


How to make a pie ?

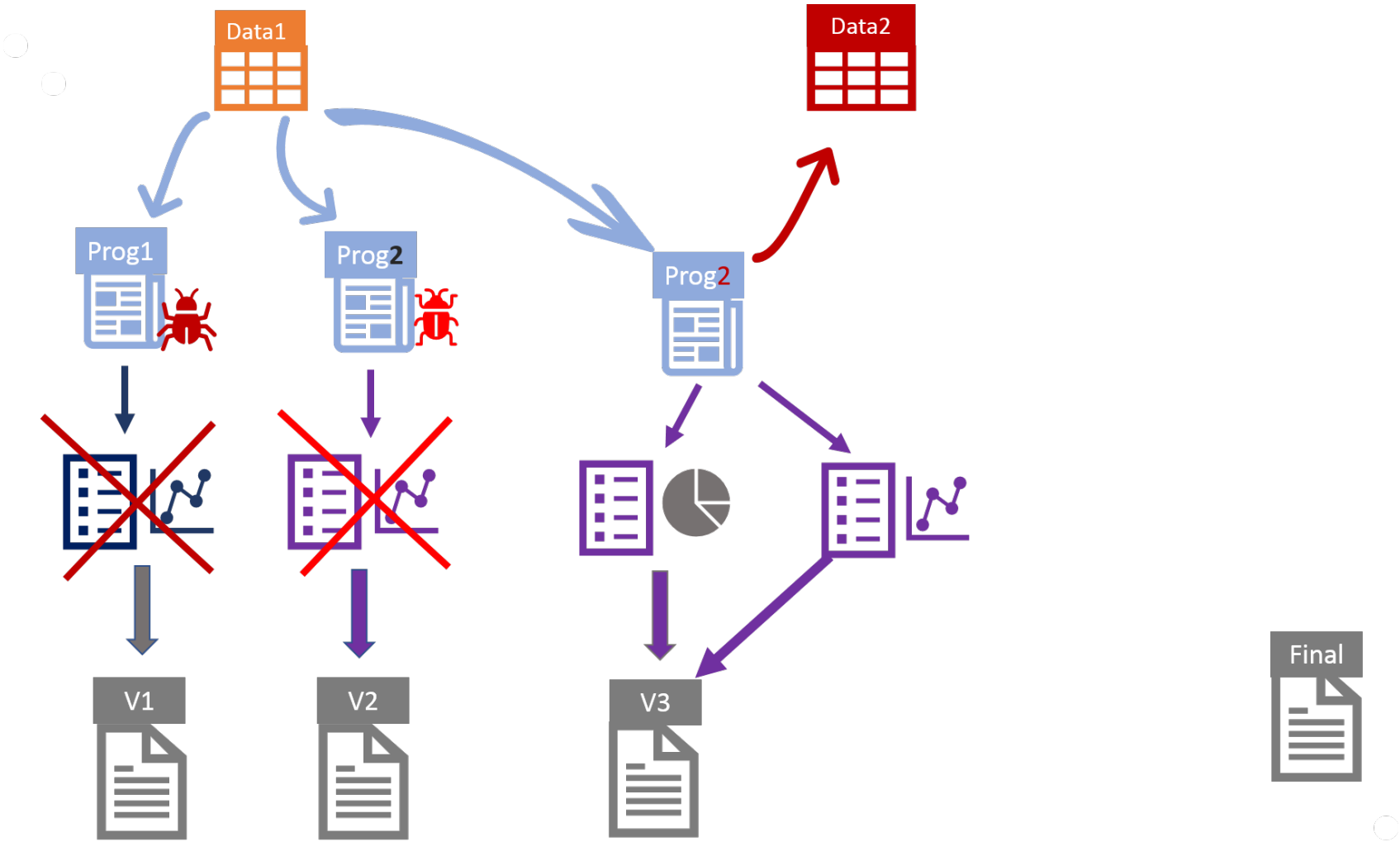




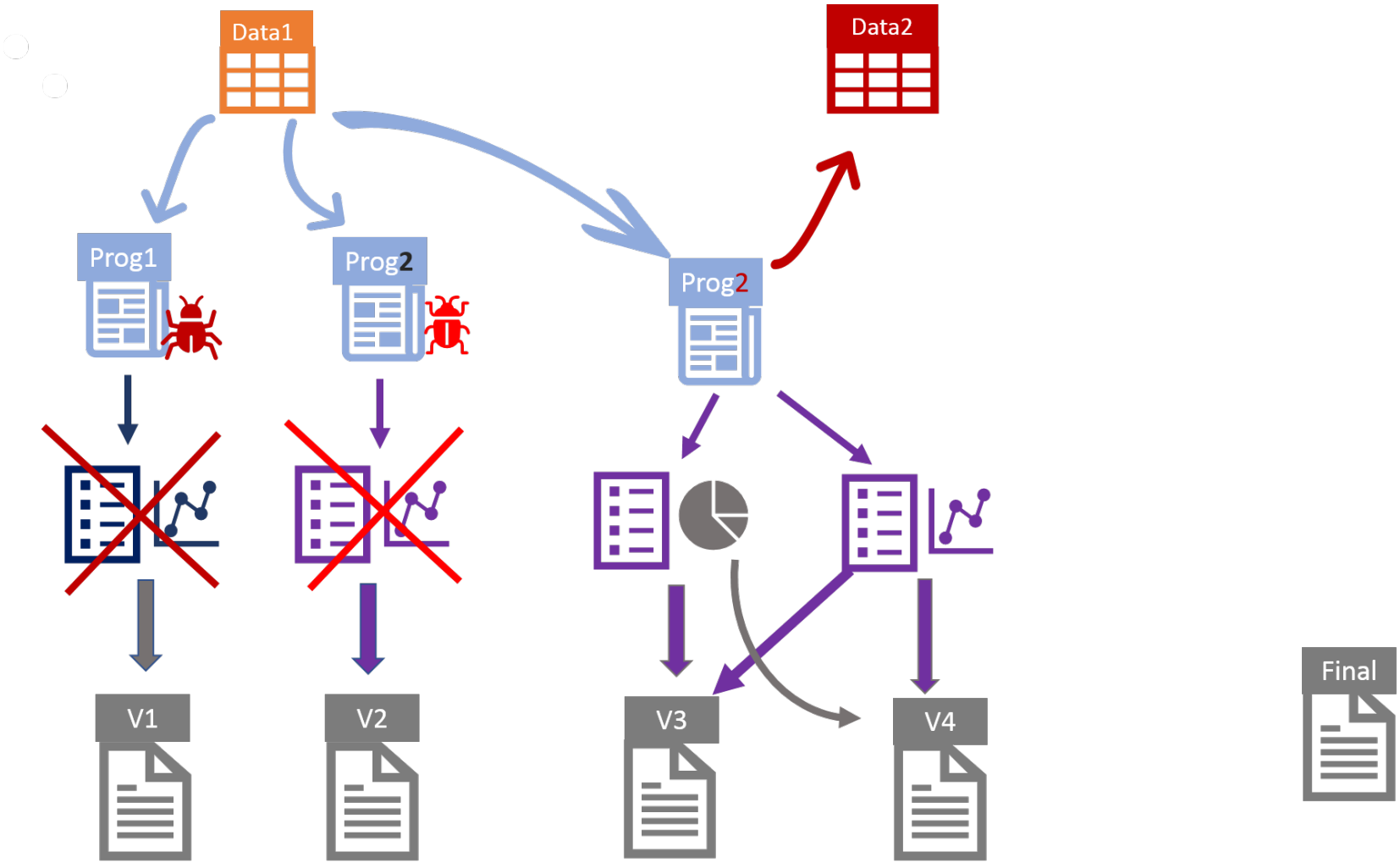
How to make a pie ?



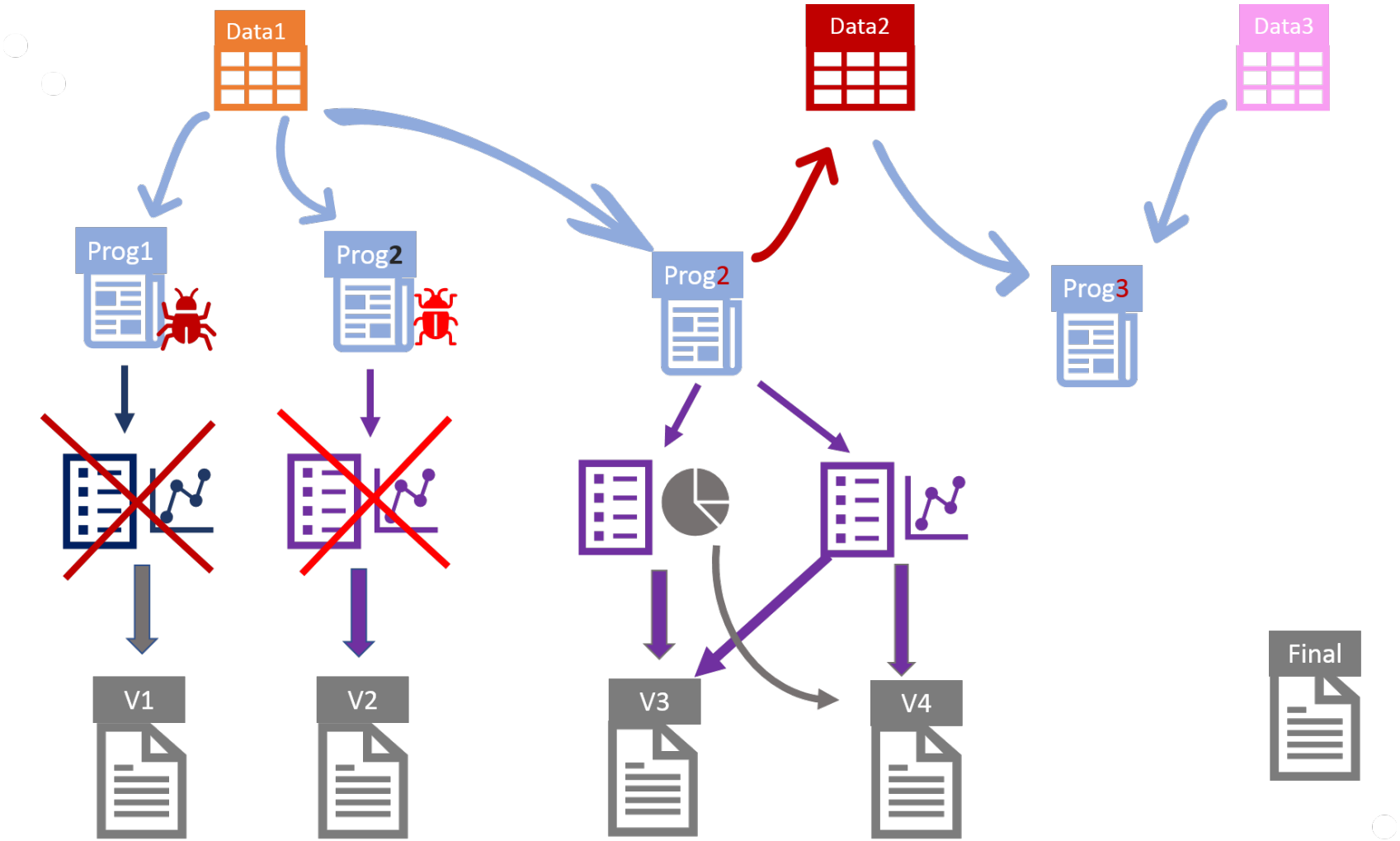
How to make a pie ?



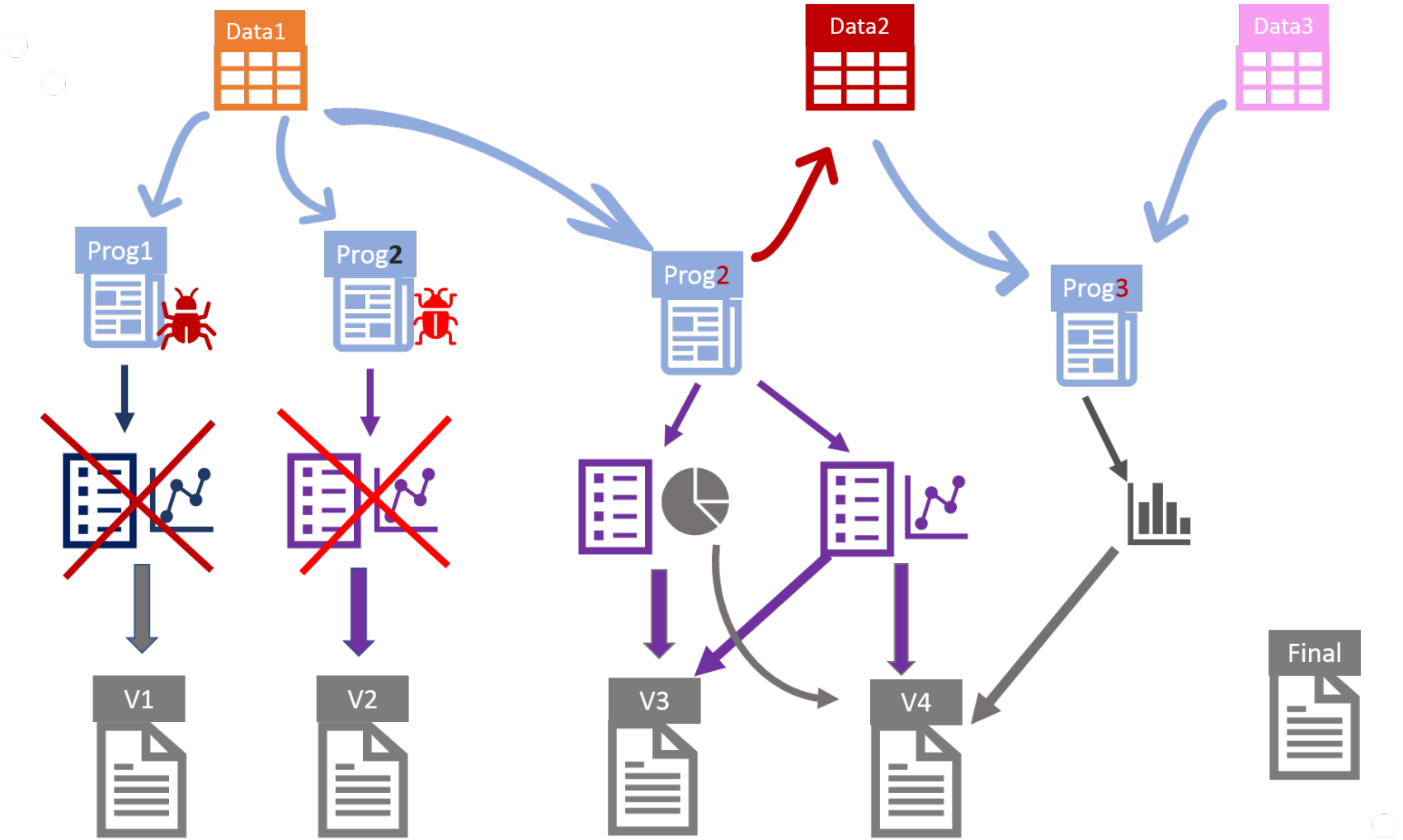
How to make a pie ?



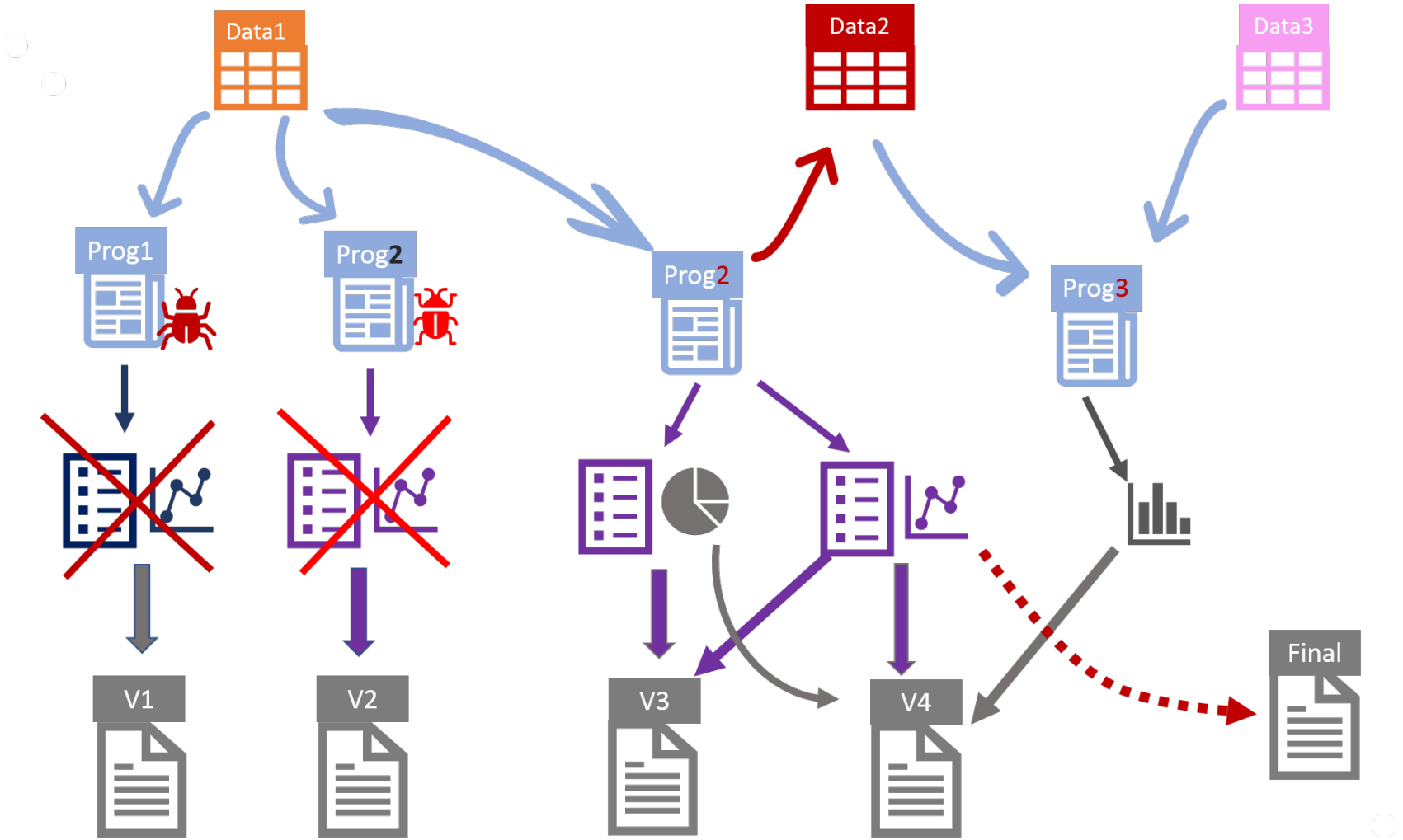
How to make a pie ?



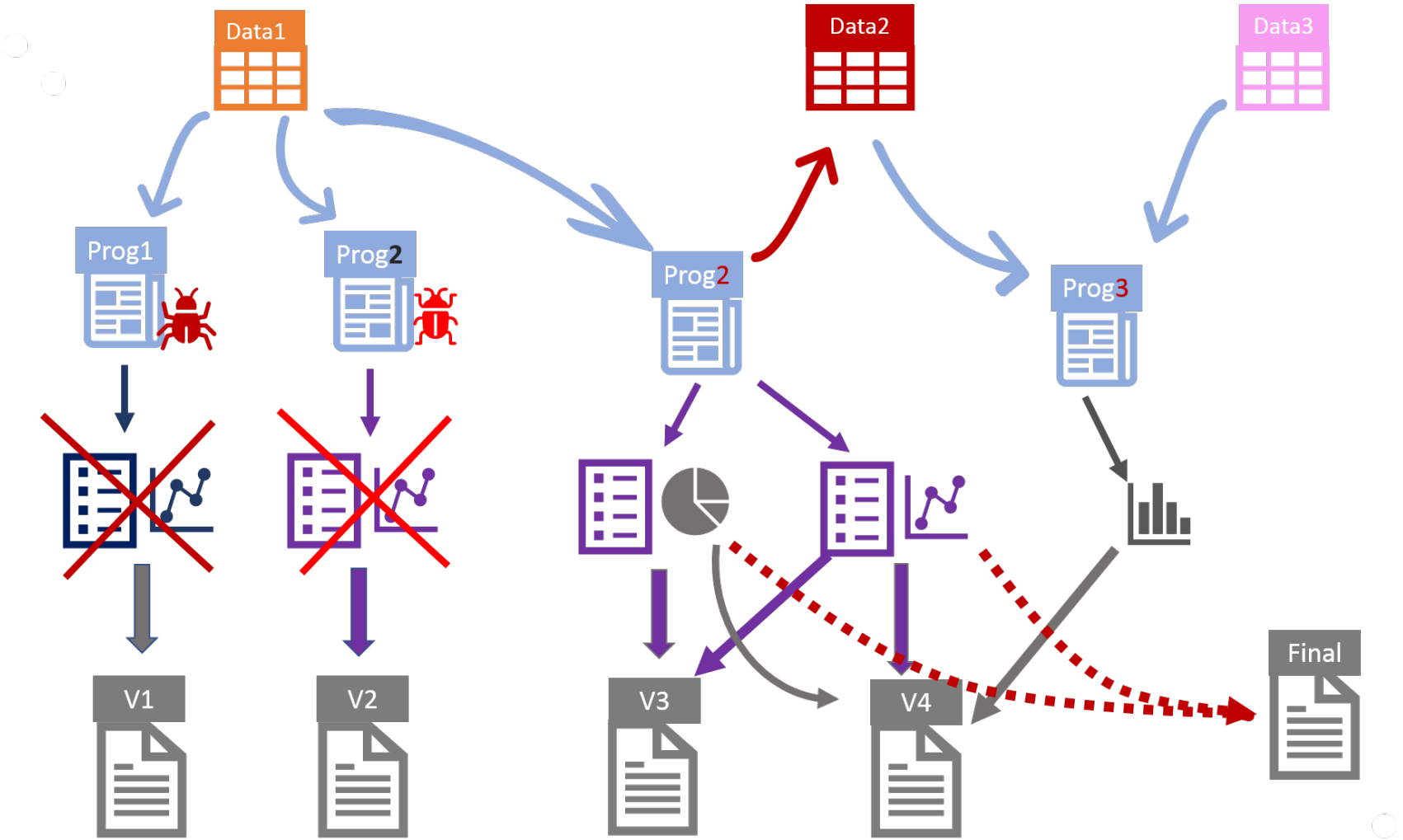
How to make a pie ?



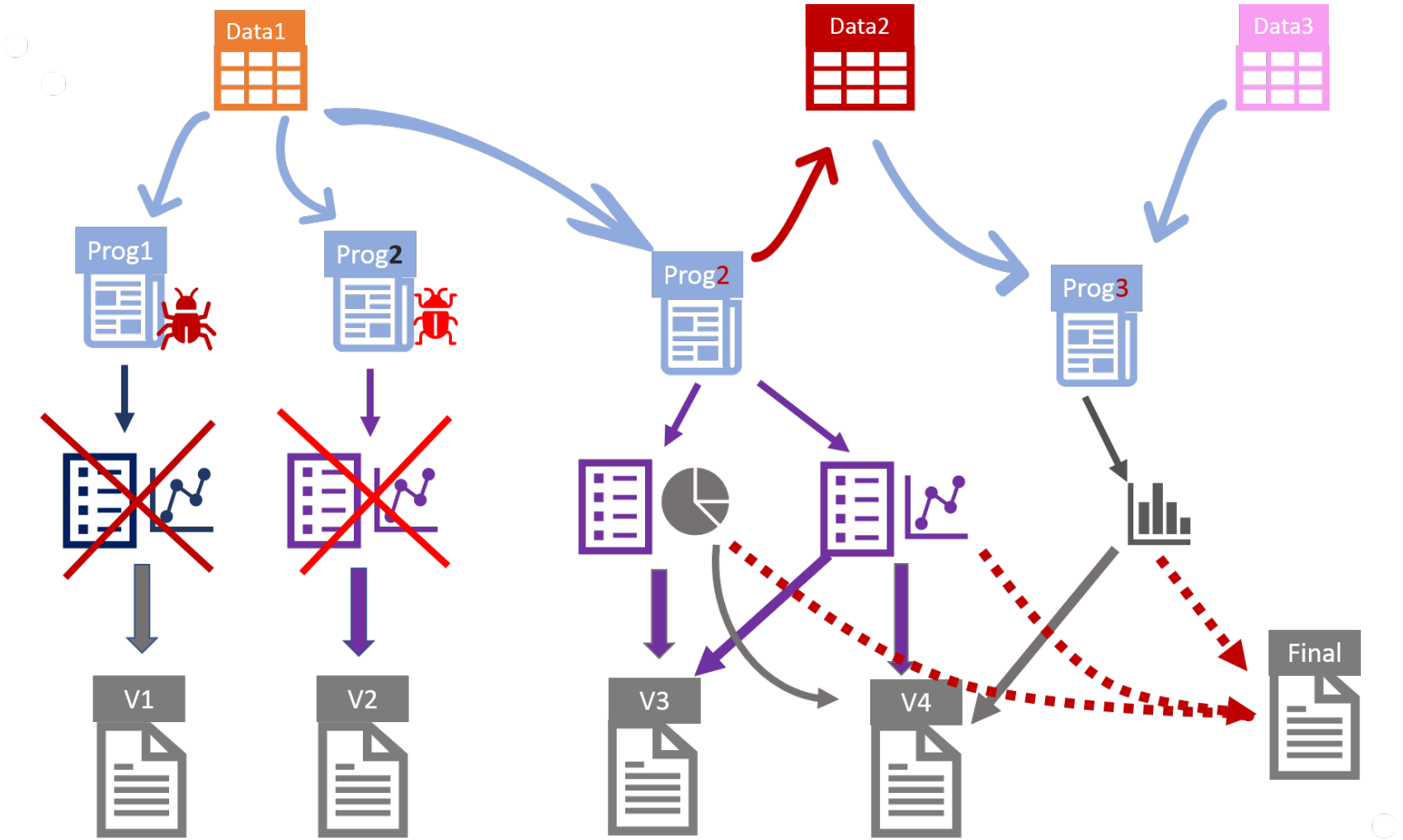
How to make a pie ?



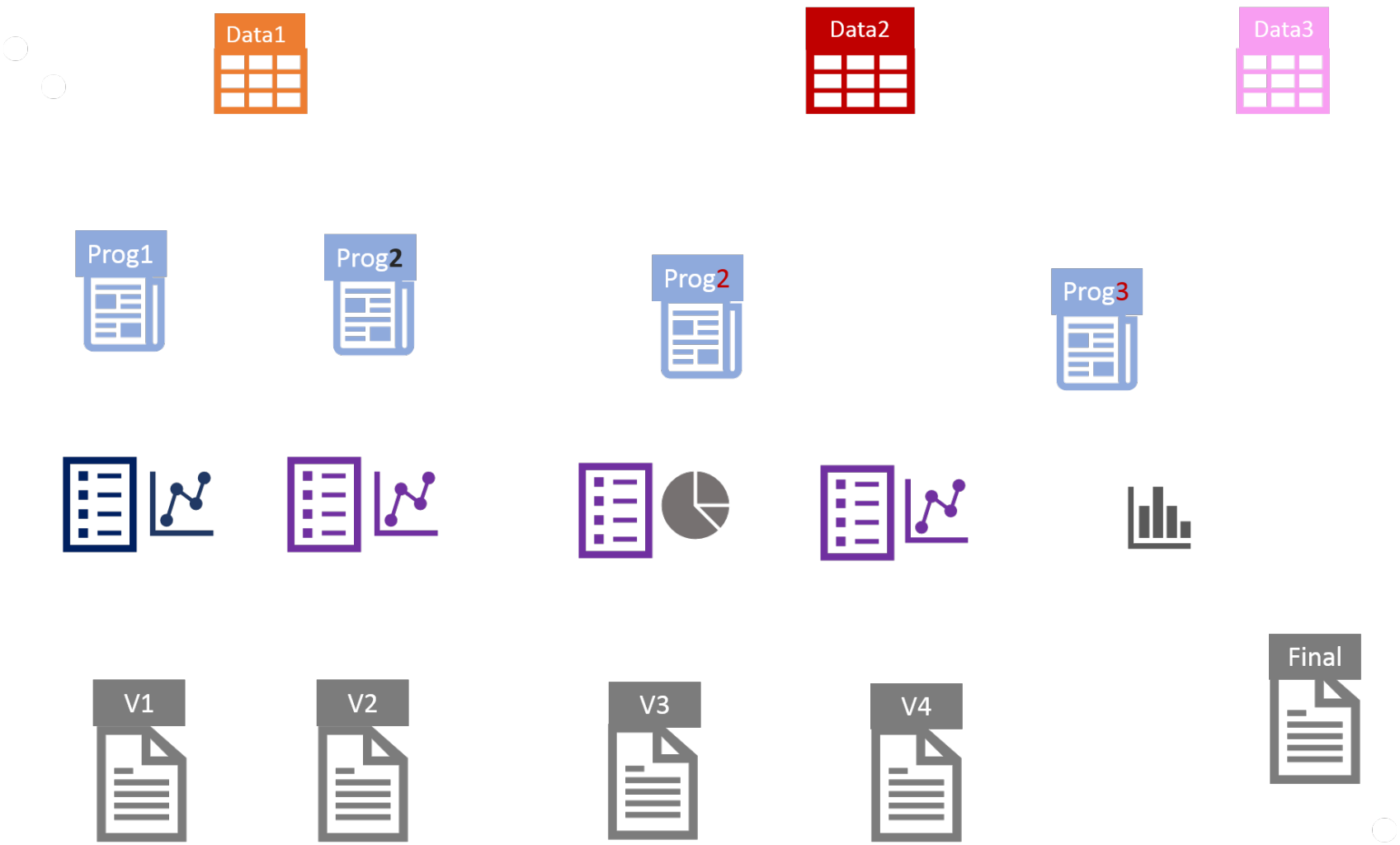
How to make a pie ?



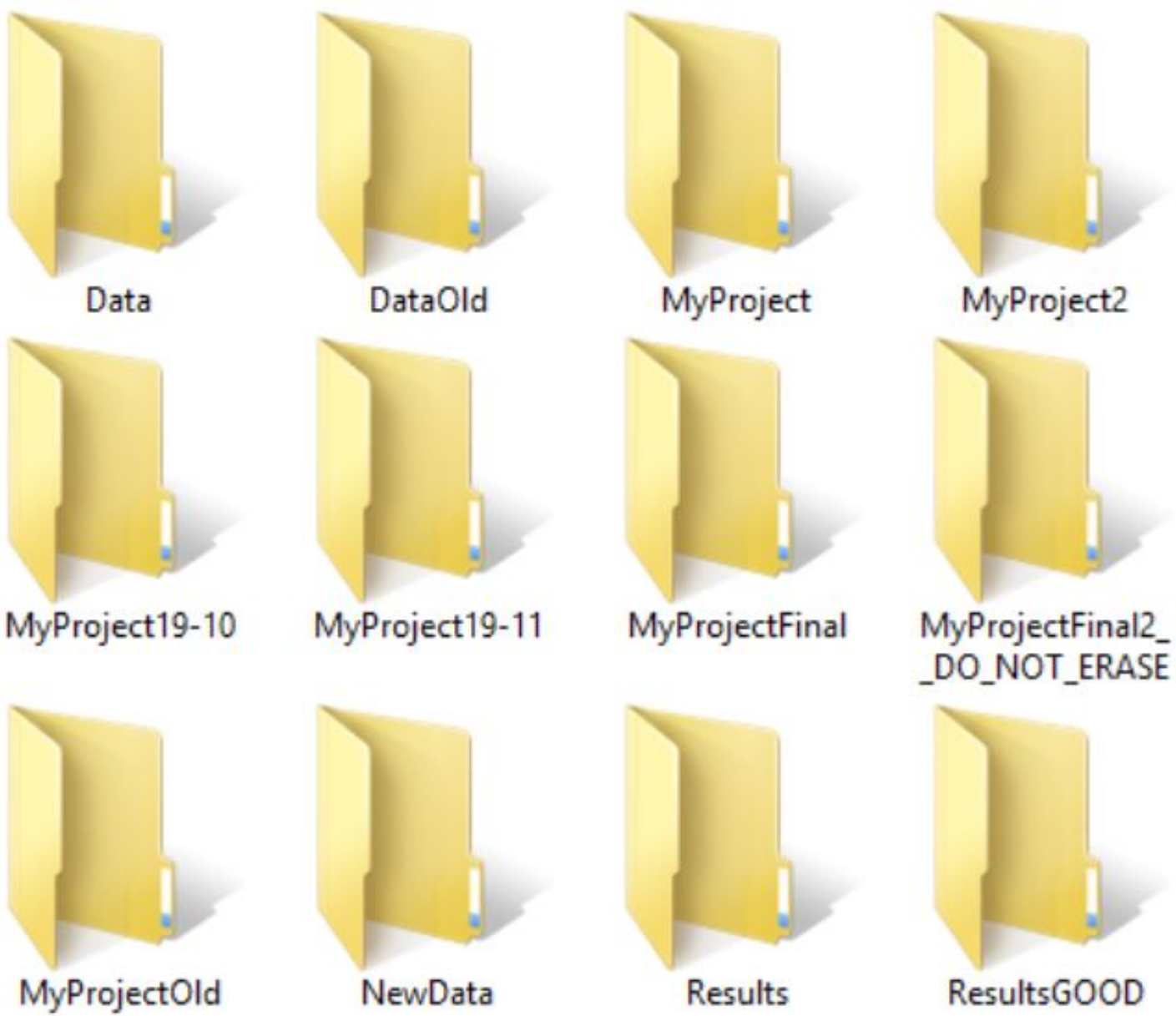
How to make a pie ?



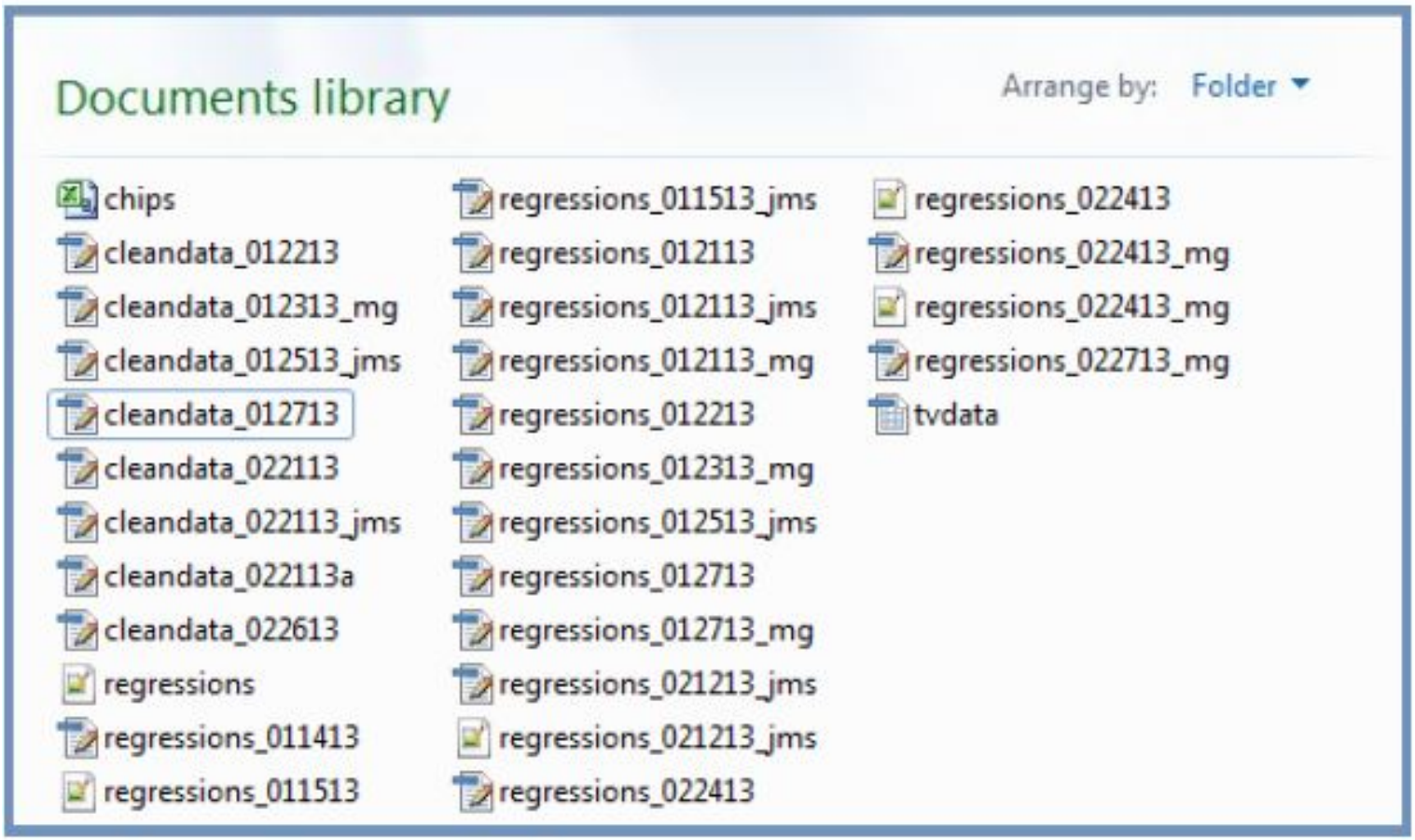
PAPER PROCESSING : 2 MONTHS LATER



PAPER PROCESSING : 2 MONTHS LATER

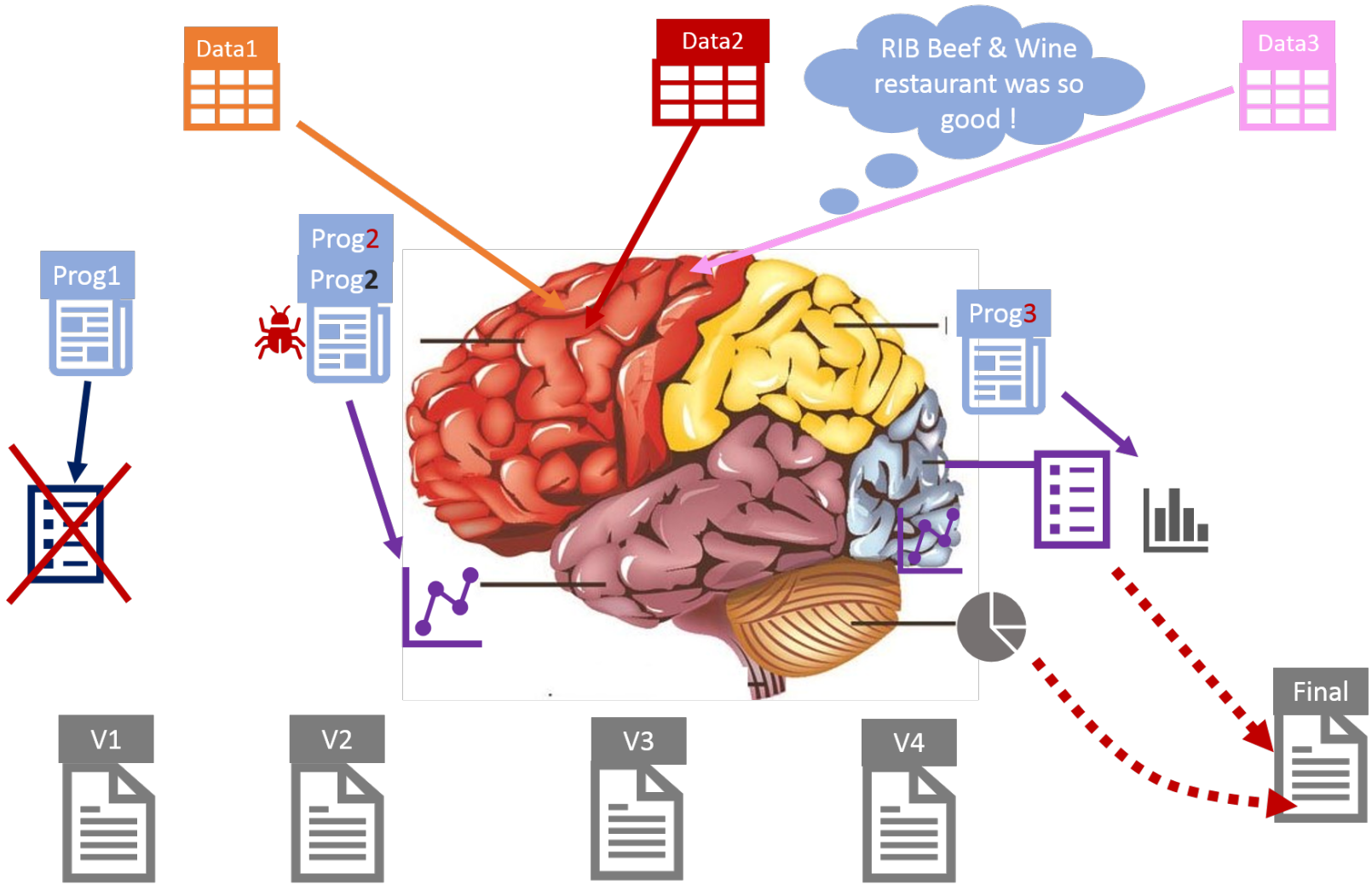


PAPER PROCESSING : 2 MONTHS LATER

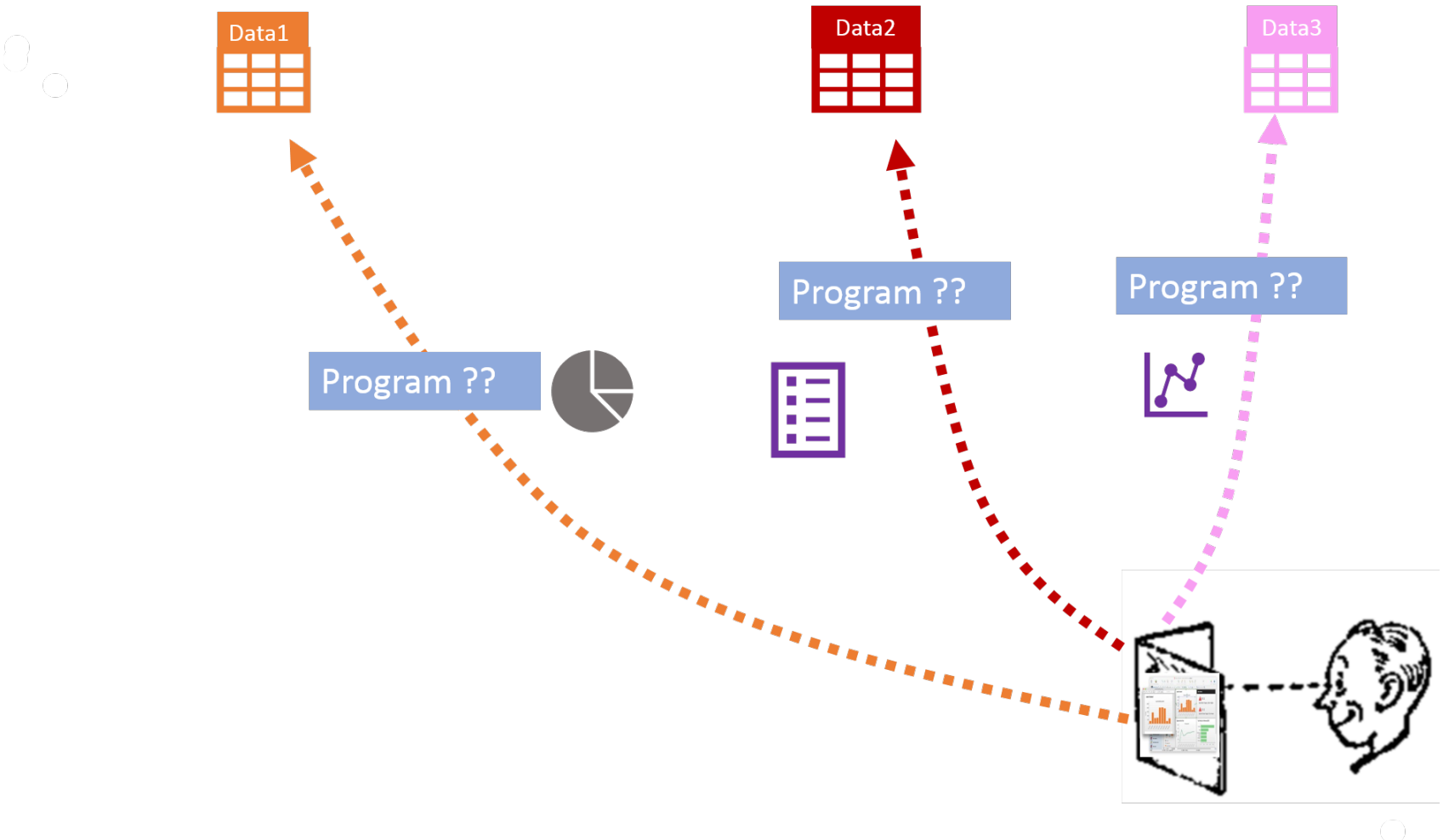


From our own experience and Gentzkow and Shapiro (2014)

Author's view : It's all in your Brain



READER'S VIEW : HOW TO DECODE



THE READER HAS ONLY THE PIE AND SOME INGREDIENTS !

Reproducible Research

- Movement born in Stanford in the 80s under the impulsion of *Jon Claerbout* who stated that :

Reproducible Research

- Movement born in Stanford in the 80s under the impulsion of *Jon Claerbout* who stated that :
  - It is quite painful to return to a research program after a stop

Reproducible Research

- Movement born in Stanford in the 80s under the impulsion of *Jon Claerbout* who stated that :
  - It is quite painful to return to a research program after a stop
  - It is quite difficult to evaluate research based on publication only,

Reproducible Research

- Movement born in Stanford in the 80s under the impulsion of *Jon Claerbout* who stated that :
  - It is quite painful to return to a research program after a stop
  - It is quite difficult to evaluate research based on publication only,
  - It is impossible to climb on "*giants shoulders*" ... if there is no ladder !

Reproducible Research

- Movement born in Stanford in the 80s under the impulsion of *Jon Claerbout* who stated that :
  - It is quite painful to return to a research program after a stop
  - It is quite difficult to evaluate research based on publication only,
  - It is impossible to climb on "*giants shoulders*" ... if there is no ladder !
- The general idea is that data, programs, and tools are **an important part** of any research program.

Reproducible Research

Definition :

		Data	
		Same	Different
Analysis	Same	**Reproducible**	**Replicable**
	Different	**Robust**	**Reusable**

Reproducible Research

Definition :

		Data	
		Same	Different
Analysis	Same	**Reproducible**	**Replicable**
	Different	**Robust**	**Reusable**

- Alternative and conflicting terminologies :

Reproducible Research

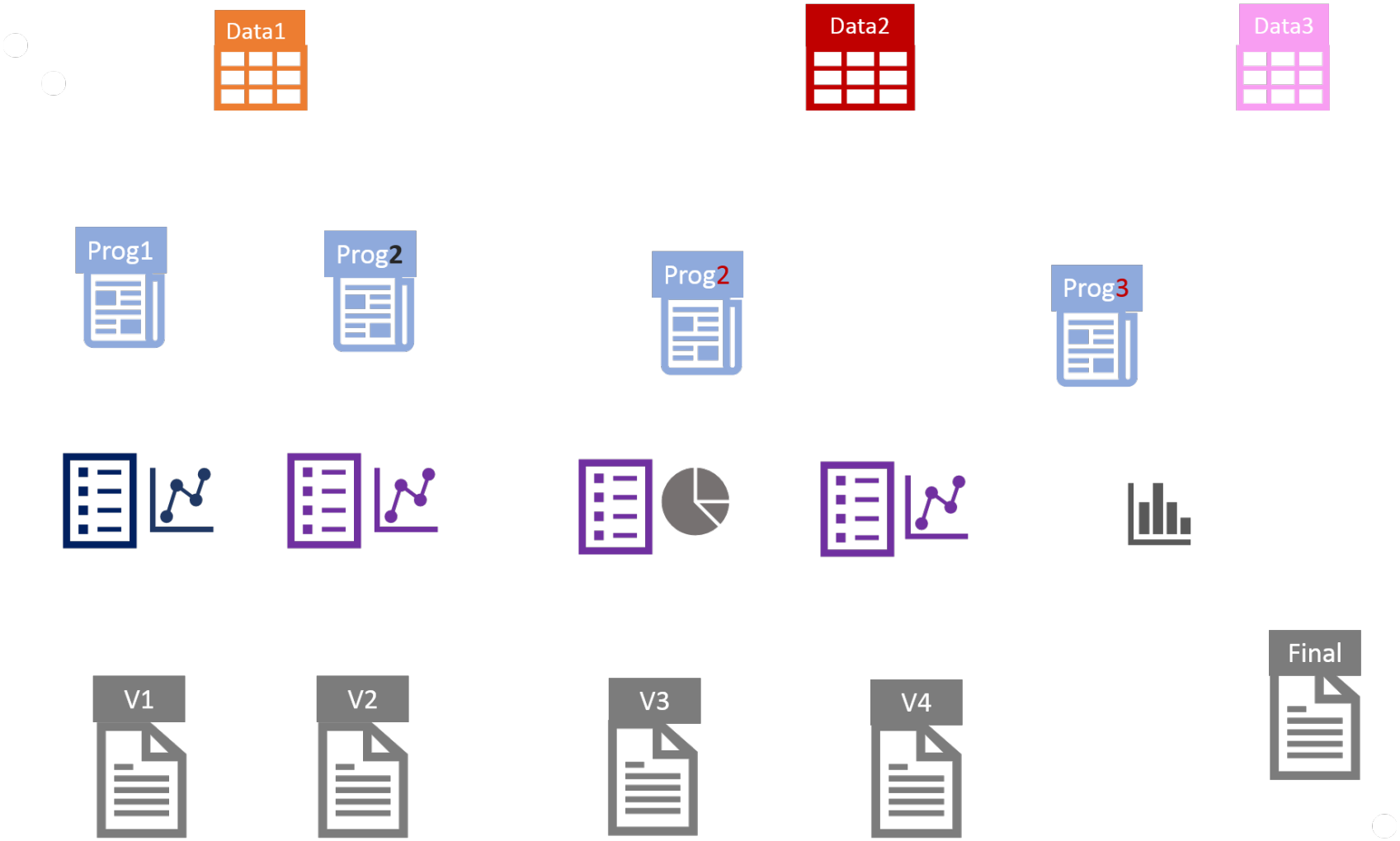
Definition :

		Data	
		Same	Different
Analysis	Same	**Reproducible**	**Replicable**
	Different	**Robust**	**Reusable**

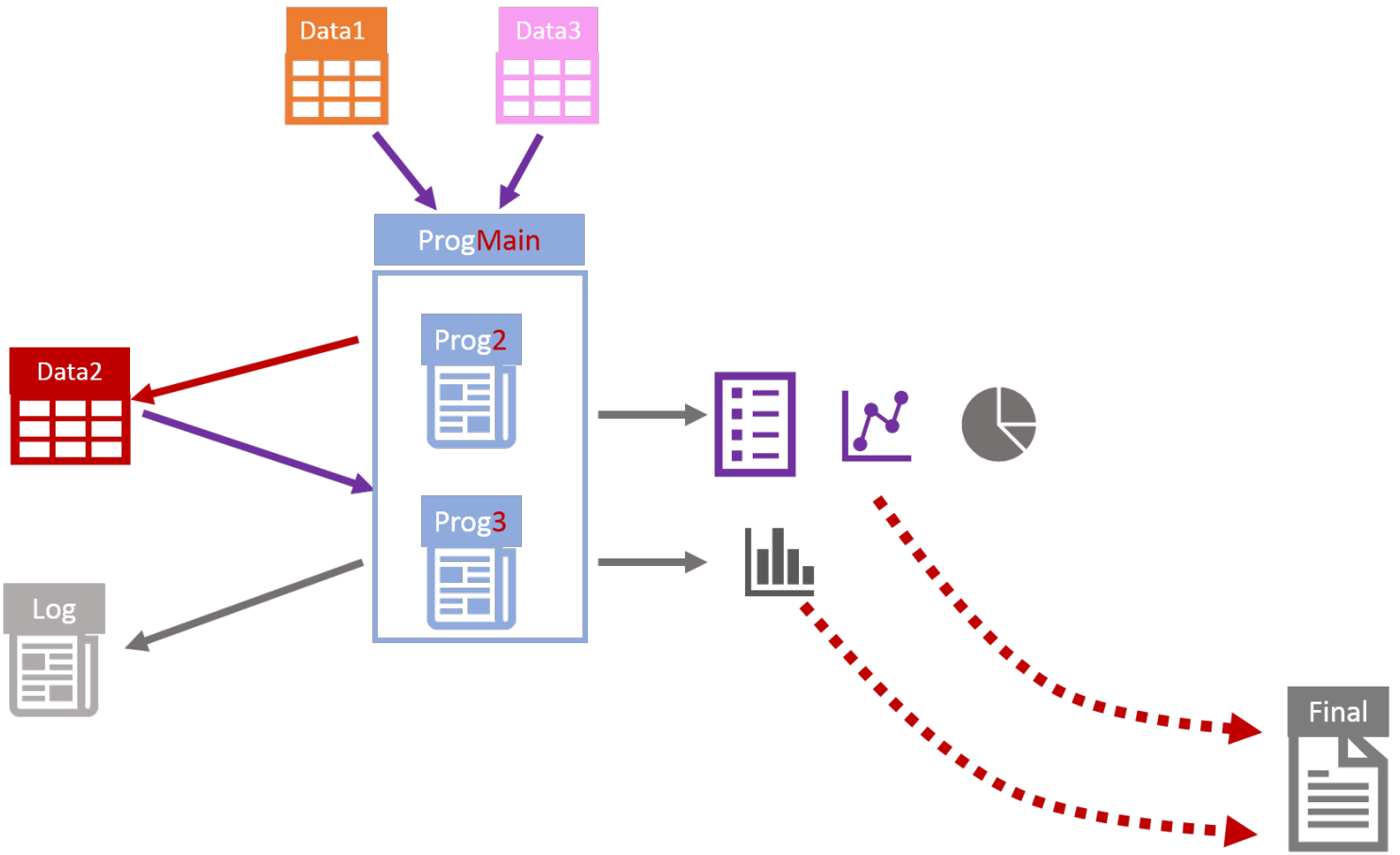
- Alternative and conflicting terminologies :

  Reviewable Research, Confirmable Research, Auditable Research, Reanalysis... (Stodden et al., 2013; Barba, 2018; Clemens, 2017)

PRINCIPLES ILLUSTRATED : FROM...



PRINCIPLES ILLUSTRATED : TO...



3 MAIN PRINCIPLES :

1. Organize your work

3 MAIN PRINCIPLES :

1. Organize your work
2. Code for others (including your future self)

3 MAIN PRINCIPLES :

1. Organize your work
2. Code for others (including your future self)
3. Automate as much as you can

3 MAIN PRINCIPLES :

1. Organize your work
2. Code for others (including your future self)
3. Automate as much as you can

3 MAIN PRINCIPLES :

1. Organize your work
2. Code for others (including your future self)
3. Automate as much as you can

Apply this in context (co-authors, software constraints,...)

Principle 1 : Organize your work

Have a clear directory structure

- Separate files into data (raw, transformed), programs, results and documentation

PRINCIPLE 1 : ORGANIZE YOUR WORK

Have a clear directory structure

- Separate files into data (raw, transformed), programs, results and documentation

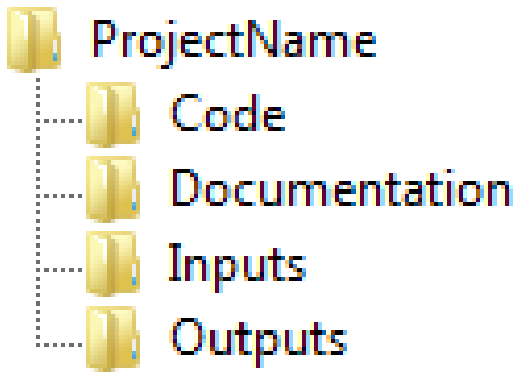


FIGURE – Example of a well-organized directory structure.

Principle 1 : Organize your work

Have a clear directory structure

- Separate files into data (raw, transformed), programs, results and documentation
- Make directories portable (relative path)

PRINCIPLE 1 : ORGANIZE YOUR WORK

Have a clear directory structure

- Separate files into data (raw, transformed), programs, results and documentation
- Make directories portable (relative path)

```
/**** 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)

  # Example of use in a path used to save a graph
  file <- paste(GraphFolder, "MySuperPie.png", sep="")
  png(filename = file)
  pie(rep(1,8), col=1:8)
  dev.off()
```

Principle 1 : Organize your work

Use naming conventions

- For your code (explicit name, indication of workflow, version)

PRINCIPLE 1 : ORGANIZE YOUR WORK

Use naming conventions

- For your code (explicit name, indication of workflow, version)

Usual

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

Principle 1 : Organize your work

Use naming conventions

- For your code (explicit name, indication of 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
```

Principle 1 : Organize your work

Use naming conventions

- For your code (explicit name, indication of workflow, version)
- For your outputs (created data, logs, tables, graphics)

Principle 1 : Organize your work

Use naming conventions

- For your code (explicit name, indication of workflow, version)
- For your outputs (created data, logs, tables, graphics)

Better

```
Stats_desc_table.tex
regress_model1_table.tex
regress_model1_fig.JPG
```

Principle 1 : Organize your work

- Keep/record every step (no cut and paste, no point and click software). Code !

Principle 1 : Organize your work

- Keep/record every step (no cut and paste, no point and click software). Code !
- Manage the workflow

Principle 1 : Organize your work

- Keep/record every step (no cut and paste, no point and click software). Code !
- Manage the workflow
  - Draw the workflow of your project

PRINCIPLE 1 : ORGANIZE YOUR WORK

- Keep/record every step (no cut and paste, no point and click software). Code !
- Manage the workflow
  - Draw the workflow of your project

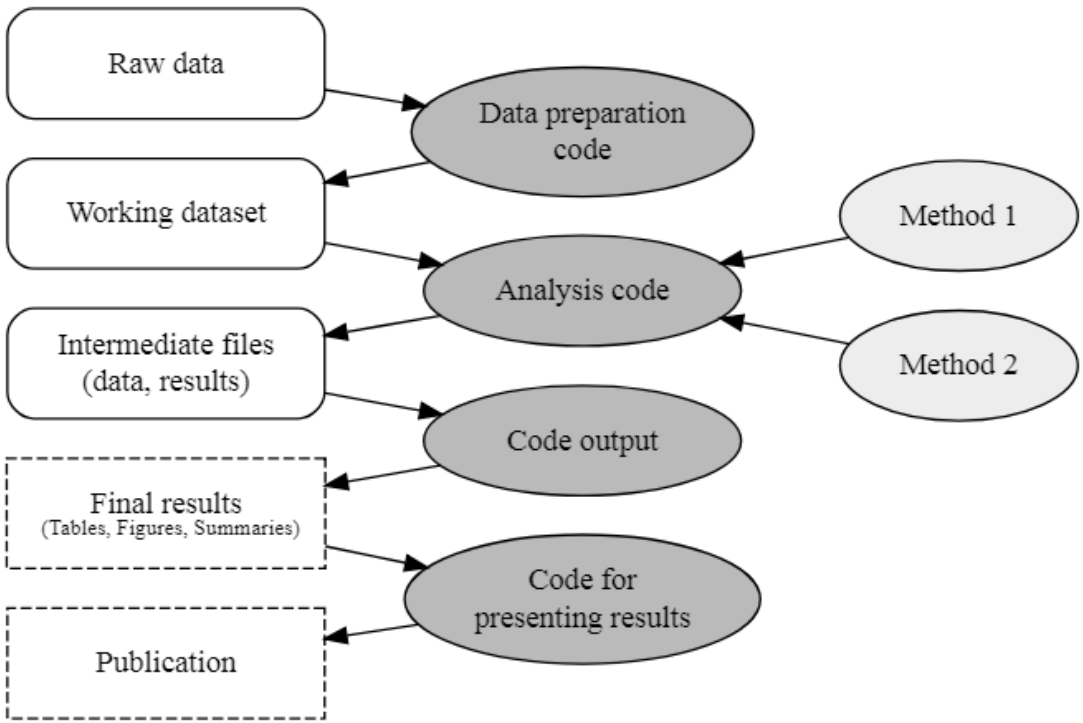
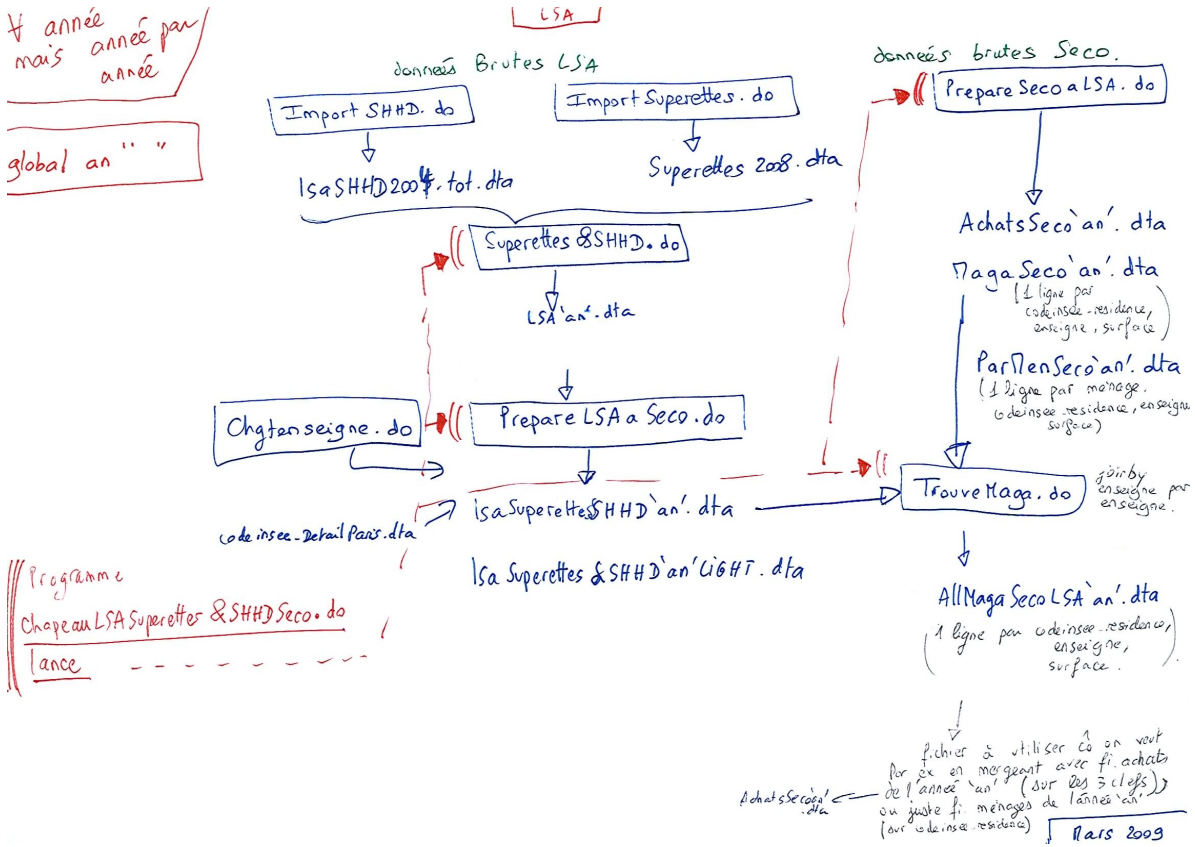


FIGURE – A simple example of a workflow

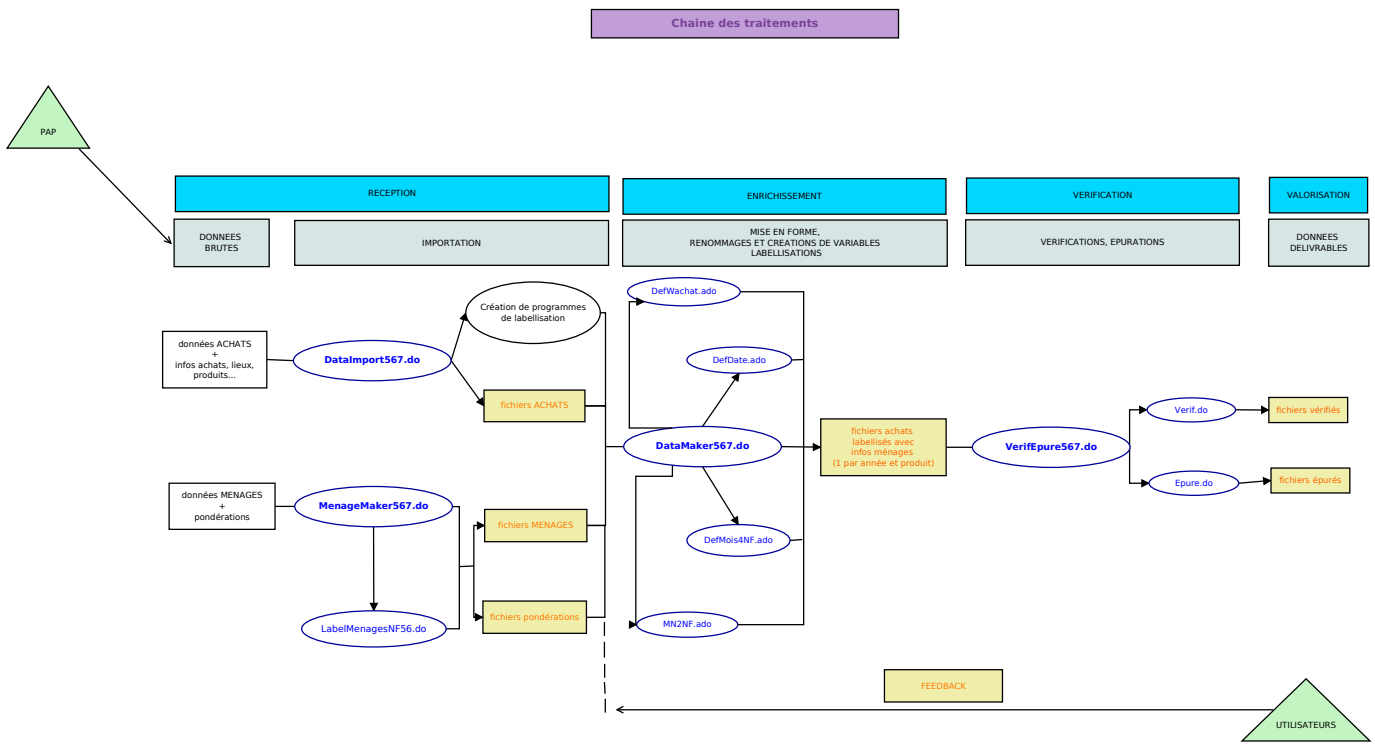
Principle 1 : Organize your work

- Keep/record every step (no cut and paste, no point and click software). Code !
- Manage the workflow
  - Draw the workflow of your project *By hand*



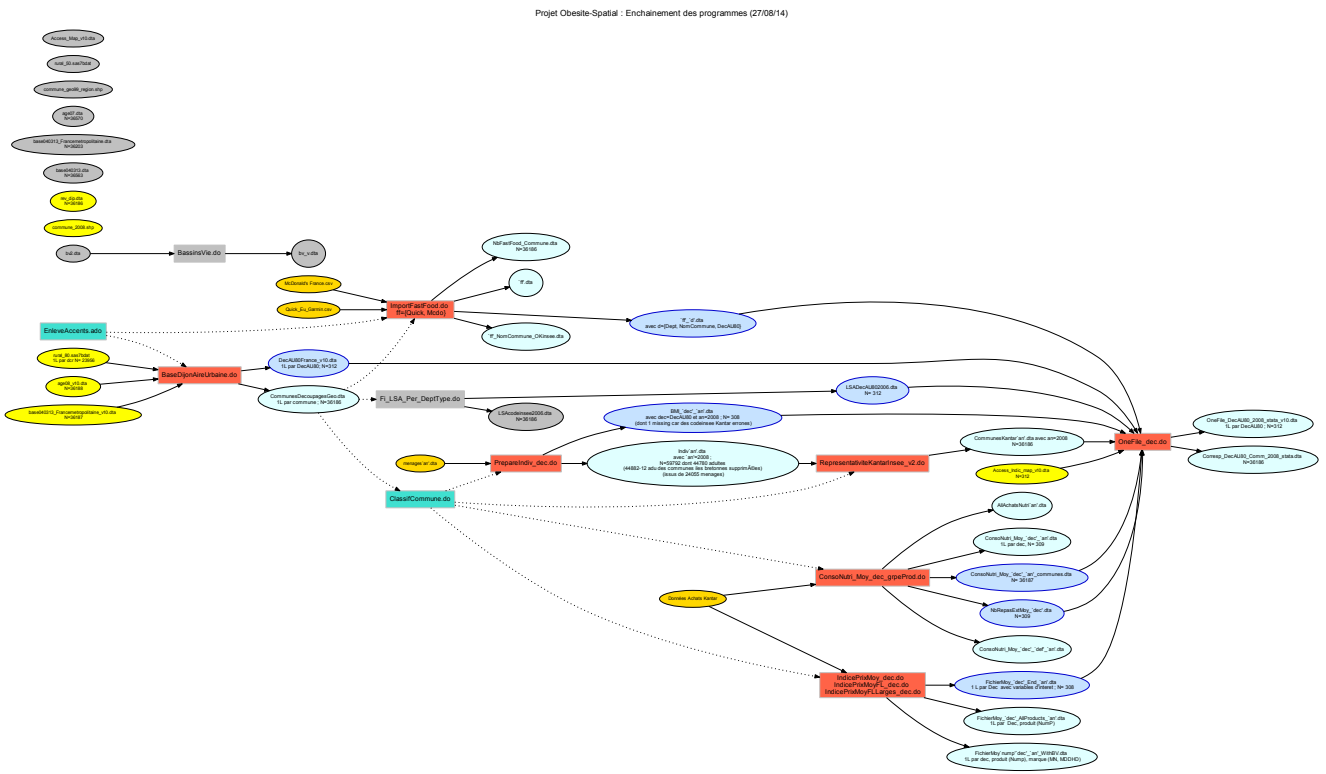
Principle 1 : Organize your work

- Keep/record every step (no cut and paste, no point and click software). Code !
- Manage the workflow
  - Draw the workflow of your project With a tool (Dia)



Principle 1 : Organize your work

- Keep/record every step (no cut and paste, no point and click software). Code !
- Manage the workflow
  - Draw the workflow of your project With *GraphViz*



Principle 1 : Organize your work

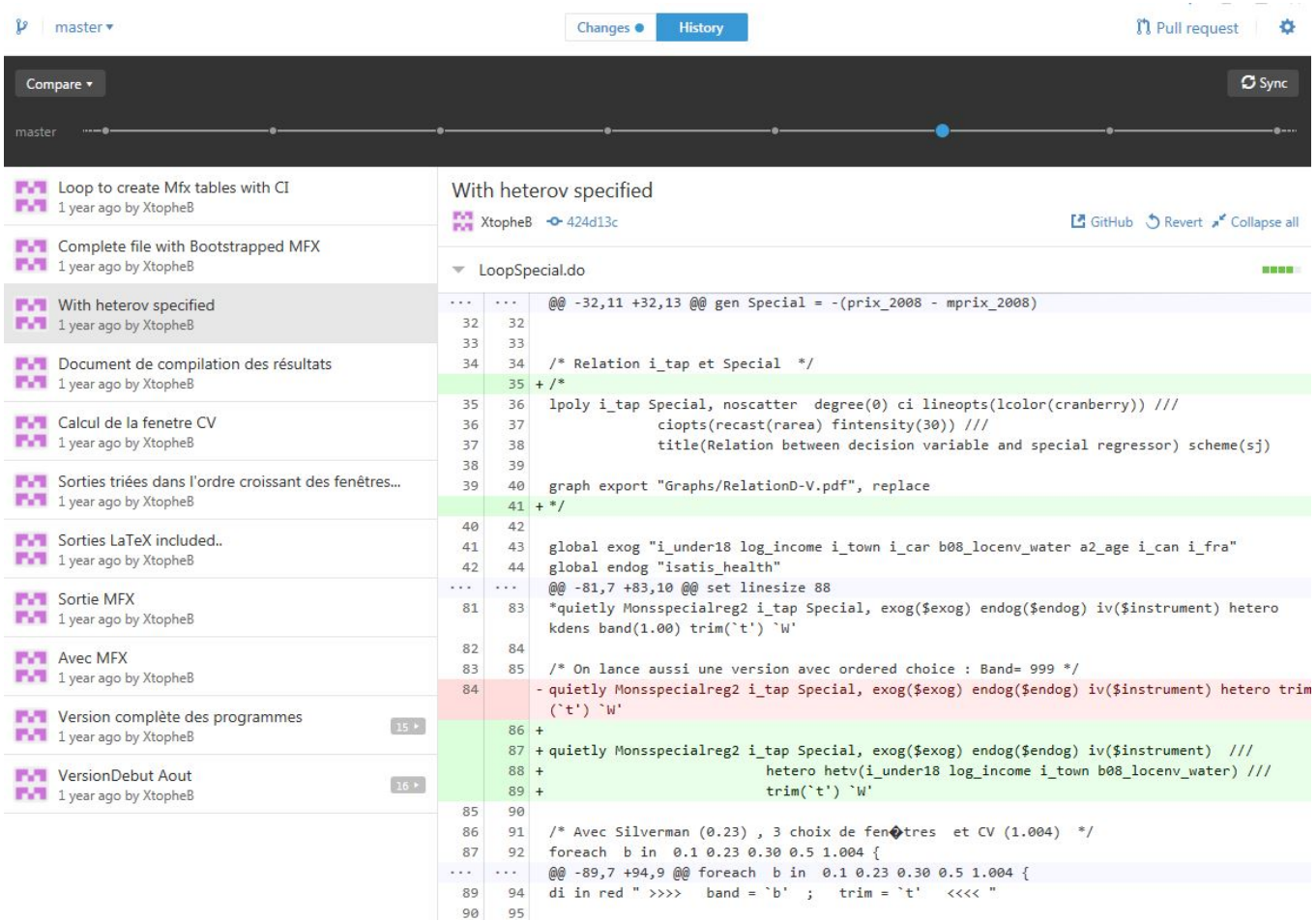
Use a version control system (example GitHub)

- *"Version control is like an undo command for everything"*, Gentzkow and Shapiro (2014)

PRINCIPLE 1 : ORGANIZE YOUR WORK

Use a version control system (example GitHub)

- *"Version control is like an undo command for everything"*, Gentzkow and Shapiro (2014)



Principle 2 : Code for others (including your future self)

Program with style :

- The code should be easy to read (Literate programming)

Principle 2 : Code for others (including your future self)

Program with style :

- The code should be easy to read (Literate programming)

Principle 2 : Code for others (including your future self)

Program with style :

- The code should be easy to read (Literate programming)

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

Principle 2 : Code for others (including your future self)

Program with style :

- The code should be easy to read (Literate programming)

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

- Use conventions on layout (Comments, indentation,...)

Principle 2 : Code for others (including your future self)

Program with style :

- Use conventions on naming

Principle 2 : Code for others (including your future self)

Program with style :

- Use conventions on naming

Usual

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

Principle 2 : Code for others (including your future self)

Program with style :

- Use conventions on naming

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"
```

Principle 2 : Code for others (including your future self)

- Document everything you can : computing environment, data, code, all hypothesis

Principle 2 : Code for others (including your future self)

- Document everything you can : computing environment, data, code, all hypothesis

```
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
```

Principle 2 : Code for others (including your future self)

- Document everything you can : computing environment, data, code, all hypothesis

```
. 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.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
```

Principle 2 : Code for others (including your future self)

- Document everything you can : computing environment, data, code, all hypothesis

```
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)
```

Principle 2 : Code for others (including your future self)

- Document everything you can : computing environment, data, code, all hypothesis

```
/*==================================================================================================*/
/* ============================== 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_1aN.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 pbs lorsqu'un produit n'a pas la variable sa2 (par ex produit 152)*/
/*Version 5.3 27/08/13 : correction des modalités de tuwa pour conversion de Qu  (V) ; déplacement du lancement de "LabelVarProdComG.do" une fois "LabelSPA`annee'.do" lancé */
/*                      (pour que Spe1 aient ses moda labellisées car ensuite Spe1 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.O)*/
/*==================================================================================================*/
local version "5.7"
```

Principle 2 : Code for others (including your future self)

Program with style :

Principle 2 : Code for others (including your future self)

Program with style :

Usual

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

coeff_variation_Chocolate_Qty <- 4 # sd / mean = 4/1
```

Principle 2 : Code for others (including your future self)

Program with style :

Usual

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

coeff_variation_Chocolate_Qty <- 4 # sd / mean = 4/1
```

Principle 2 : Code for others (including your future self)

Program with style :

Usual

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

coeff_variation_Chocolate_Qty <- 4 # sd / mean = 4/1
```

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
```

Principle 2 : Code for others (including your future self)

Program with style :

Usual

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

coeff_variation_Chocolate_Qty <- 4 # sd / mean = 4/1
```

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
```

Principle 2 : Code for others (including your future self)

Program with style :

Usual

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

coeff_variation_Chocolate_Qty <- 4 # sd / mean = 4/1
```

Even better

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

Principle 2 : Code for others (including your future self)

Program with style :

- The code should be as modular as possible (keep it DRY ! Wilson et al. (2014))

Principle 2 :
Code for others (including your future self)

Program with style :

- The code should be as modular as possible (keep it DRY ! Wilson et al. (2014))

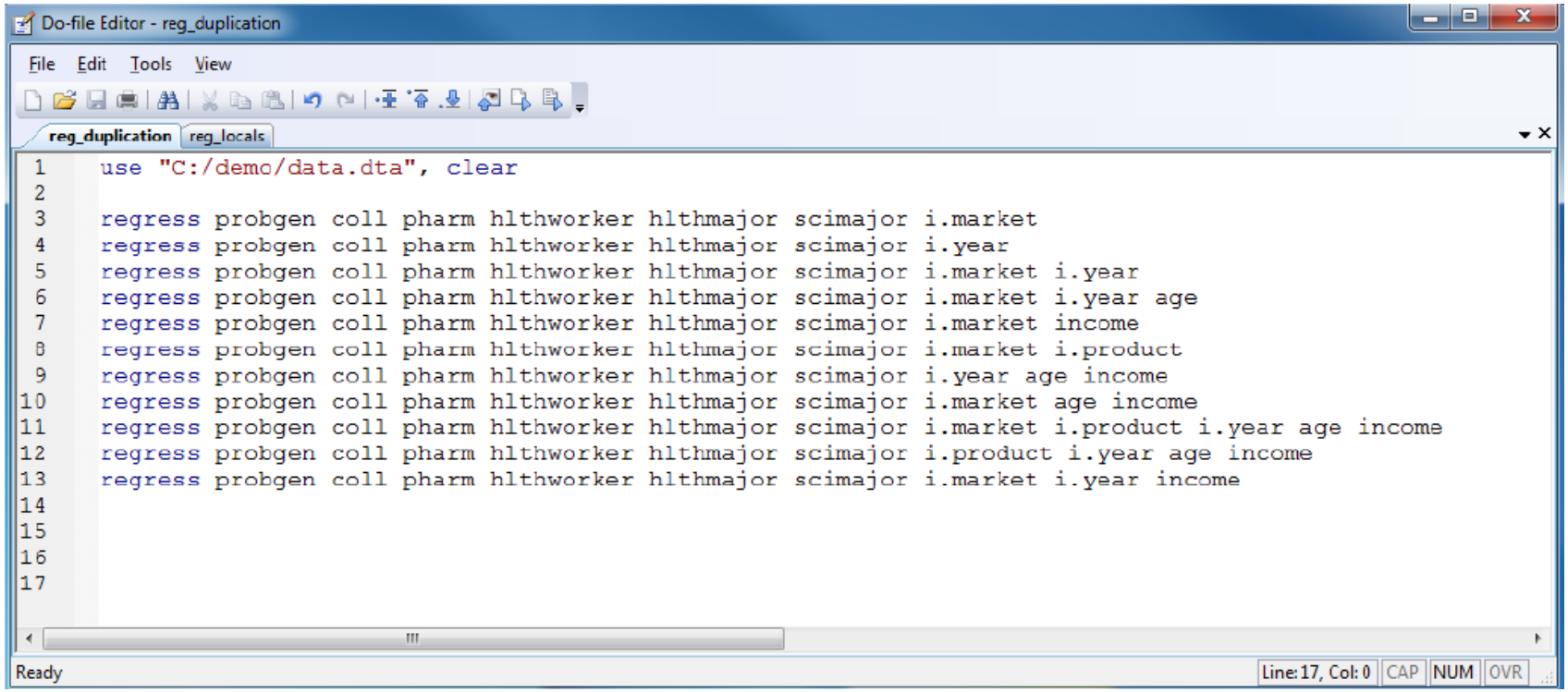


FIGURE – Example from Gentzkow and Shapiro (2013)

Principle 2 :
Code for others (including your future self)

Program with style :

- The code should be as modular as possible (keep it DRY ! Wilson et al. (2014))

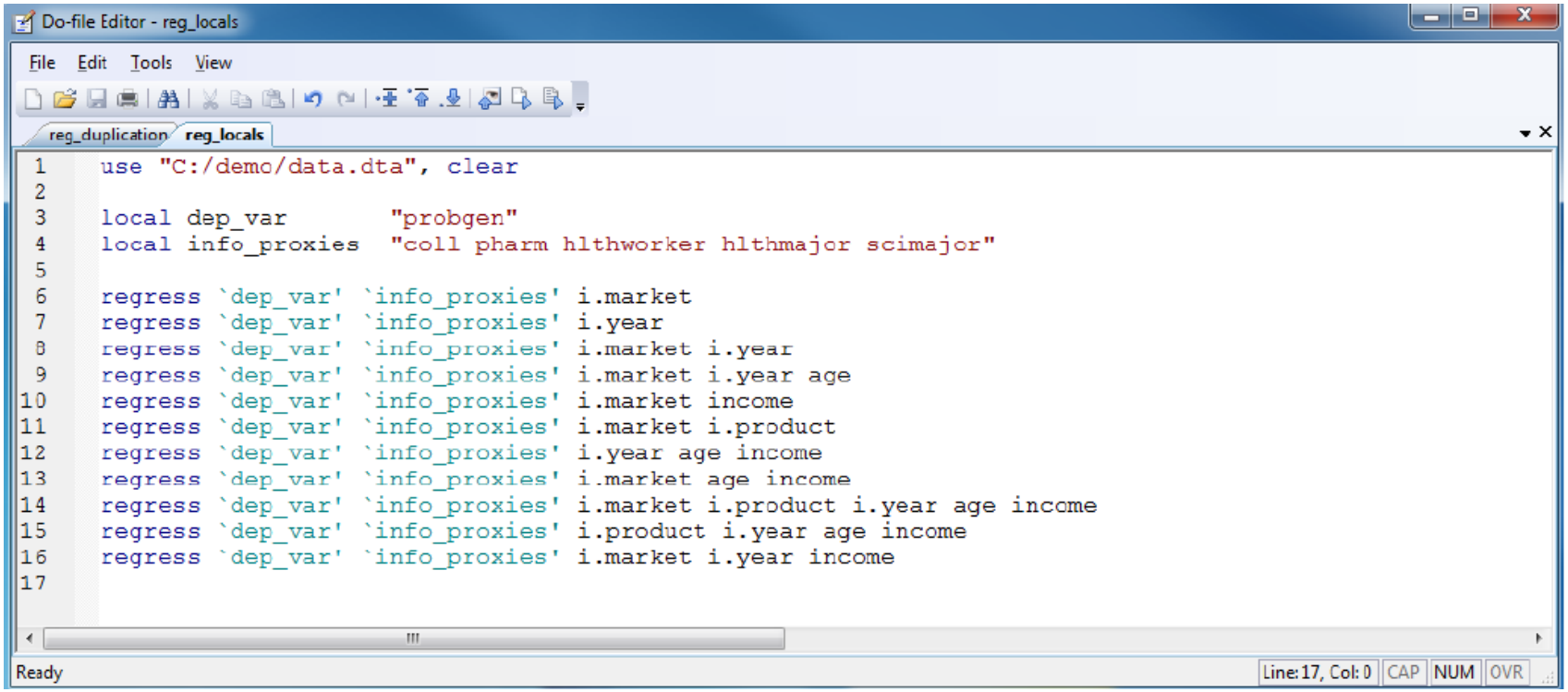


Figure – Example from Gentzkow and Shapiro (2013)

Principle 3 : Automate as much as you can

- Code everything (no cut and paste, no point and click software)

Principle 3 : Automate as much as you can

- Code everything (no cut and paste, no point and click software)
- Automate the workflow !

PRINCIPLE 3 : AUTOMATE AS MUCH AS YOU CAN

- Code everything (no cut and paste, no point and click software)
- Automate the workflow !
  - Write a *"Master"* program embedding all programs

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
```

PRINCIPLE 3 : AUTOMATE AS MUCH AS YOU CAN

- Code everything (no cut and paste, no point and click software)
- Automate the workflow !
  - Use a *"shell"* file linking your programs (*.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
```

PRINCIPLE 3 : AUTOMATE AS MUCH AS YOU CAN

- Code everything (no cut and paste, no point and click software)
- Automate the workflow !
  - Use a MakeFile

```
# 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
```

Principle 3 : Automate as much as you can

Create reproducible documents

- Automatically insert results into the final document

PRINCIPLE 3 : AUTOMATE AS MUCH AS YOU CAN

Create reproducible documents

- Automatically insert results into the final 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.

PRINCIPLE 3 : AUTOMATE AS MUCH AS YOU CAN

Create reproducible documents

- Automatically insert results into the final 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 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
```

PRINCIPLE 3 : AUTOMATE AS MUCH AS YOU CAN

Create reproducible documents

- Automatically insert results into the final 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)*, gdxxrw, xlexport, xldump *(.csv, .xls)*	gnuplot, gnuplotxyz *(.png)*
Mathematica	Export[ ] *(.xls)*, CloudExport[, "pdf"] *(.pdf)*	Export[ ] *(.gif, .jpg)*

Principle 3 : Automate as much as you can

Create reproducible documents

- Automatically insert results into the final document
- Link the article with the code (and results)

ILLUSTRATION :

How long does it take to redo a research paper ?

ILLUSTRATION :

How long does it take to redo a research paper ?

For example this one

ILLUSTRATION :

How long does it take to redo a research paper ?

For example this one

1 minute !

ILLUSTRATION :

How long does it take to redo a research paper ?

For example this one

1 minute !

- Let's see the R code

ILLUSTRATION :

Also possible in Stata

ILLUSTRATION :

Also possible in Stata

For example One html page

ILLUSTRATION :

Also possible in Stata

For example One html page

Let's try !

ILLUSTRATION :

```
---
title: Reproducible document with Stata and Markdown
author: C. Bontemps, V. Orozco
date: December 2019
bibliography: RR16.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. Let's see that using a bibtex file for references is esay! [see @McCullough2009].

Data
------------------------
```s
    local prod "0212"
    qui use "../data/p`prod'Men.dta", clear
    qui vallist lib*
    global lib "`r(list)'"
    qui sum annee
    local y "`r(mean)'"
    local Nobs "`r(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.
```s
    qui bysort CODE_DEPT : keep if _n==1
    qui keep CODE_DEPT qtedep lib
    qui merge 1:1 CODE_DEPT using "../data/PopDep99.dta", keepusing(pop99)
    qui drop if _merge==2
    qui drop _merge
    qui gen qtedepW = qtedep / pop99
    qui sum qtedepW qtedep

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

Map

```s
    qui cd ../data
          qui spmap qtedepW using frcoord if id<97,          id(id) fcolor(Purples)  title("Purchased Quantity/Inhabitant $lib") ///
          subtitle("(France 2008 per dept(INSEE))") ///
          legenda(on) legstyle(2) legend(ring(1) position(3)) legtitle(Quantity in Kg ou L) ///
          scalebar(units(500) scale(1/1000) xpos(-50)  label(Kilomètres))
    qui cd ../progs
    qui graph export map.png, width(400)  replace
```
![Nice Map](map.png)

Estimations

```s
    qui use "../data/p`prod'Men.dta", clear
    qui gen ru=(habi==0)
```
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:

$$
    BMI_i=\beta_0 + \beta_{1}\cdot X_1 ..... + \varepsilon_k
$$

```s
    qui eststo clear
    qui eststo : reg BMI1  qtedepMen  ru    mifi
    qui eststo : reg BMI1  qtedepMen  ru    mifi  ageind1
    esttab  ,   not r2  ar(2)  label   nonumber  title(Estimations)   mtitles("mod 1" "mod 2")   addnote("Source : $lib data")   b(%9.3f)  replac
```
```

Reproducible document with Stata and Markdown

C. Bontemps, V. Orozco

December 2019

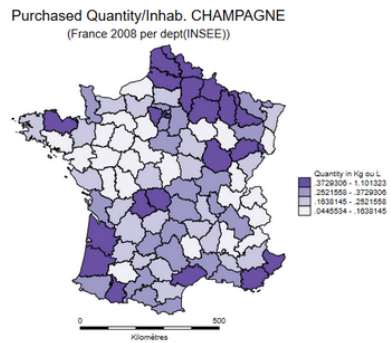
Introduction

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 bibtex file for references is esay! (see McCullough 2009).

Data

We used data from the TNS-Worldpanel (2008) on **CHAMPAGNE** consumption for the year 2008. The data contains 3740 observations.

Map



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:

$BMI_i = \beta_0 + \beta_1 \cdot X_1 ..... + \varepsilon_k$

Estimations

	mod 1	mod 2
qtedepMen	0.000***	0.000
ru	0.465*	0.480*
Nb de PC	0.154***	0.048
Age indiv 1		0.005***
Constant	24.284***	21.290***
R-squared	0.012	0.053
Adjusted R-squared	0.01	0.05
Observations	3395	3395

Source : CHAMPAGNE data
* p<0.05, ** p<0.01, *** p<0.001

That's all for now!

Bibliography

McCullough, B. D. 2009. "Open Access Economics Journals and the Market for Reproducible Economic Research." *Economic Analysis and Policy* 39 (1): 117–26. http://ideas.repec.org/a/eap/articl/v39y2009i1p117-126.html.

FIGURE – Stata code (left), and its resulting report (right)

Principle 3 : Automate as much as you can

Create reproducible documents

- Link the article and results with the code (use *Sweave, Knitr, R Markdown, Mardoc, Statweave, Markdown* ...)

PRINCIPLE 3 : AUTOMATE AS MUCH AS YOU CAN

Create reproducible documents

- Link the article and results with the code (use *Sweave, Knitr, R Markdown, Mardoc, Statweave, Markdown* ...)

Language used for: Code	Language used for: Text	Tool name	Output format	References
Sweave-like tools				
R	LaTeX	Sweave	TeX, Beamer, PDF	Leisch (2002), Meredith & Racine (2009)
R, Python, SAS, SQL, ...	Markdown	R Markdown	HTML, PDF, MS Word, Beamer, ...	Xie (2015), Gandrud (2015), Xie *et al.* (2018)
R, SAS	LaTeX, noweb	SASWeave	TeX, PDF	Lenth & Højsgaard (2007), Morrisson & Karafa (2012)
R, SAS, MATLAB, Stata, ...	LaTeX, OpenOffice	StatWeave*	TeX, ODT	Lenth & Højsgaard (2011), Lenth (2012)
Stata	Markdown	Markstat	TeX, PDF, HTML	Rodriguez (2017)
SAS	LaTeX	StatRep	TeX, PDF	Arnold & Kuhfeld (2012, 2015), Morrisson & Karafa (2012)
MATLAB	plain text markup	Publish	TeX, MS Word, HTML, PDF	Hunt *et al.* (2014) McCarthy (2018)
R, Stata, MATLAB, Python, ...	plain text markup	Org-mode	TeX, PDF, HTML, ODT, ...	Dominik (2010), Schulte & Davison (2011), Schulte *et al.* (2012)
Notebooks				
Python, R, SAS, Stata**, MATLAB, Julia, ...	Markdown	Jupyter Notebook	HTML, rST, PDF	LeVeque (2009), Kluyver *et al.* (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 *et al.* (2014) McCarthy (2018)

In addition :
My co-authors and I share everything

Always difficult !

- Naming conventions $\hookrightarrow$ pair programming

In addition :
My co-authors and I share everything

Always difficult !

- Naming conventions $\hookrightarrow$ pair programming
- Directory structure $\hookrightarrow$ relative paths in programs

In addition :
My co-authors and I share everything

Always difficult !

- Naming conventions $\hookrightarrow$ pair programming
- Directory structure $\hookrightarrow$ relative paths in programs
- Data* $\hookrightarrow$ Dropbox, Web repository, Zenodo,..

In addition :
My co-authors and I share everything

Always difficult !

- Naming conventions $\hookrightarrow$ pair programming
- Directory structure $\hookrightarrow$ relative paths in programs
- Data* $\hookrightarrow$ Dropbox, Web repository, Zenodo,..
- Programs $\hookrightarrow$ Dropbox, GitHub, Agora,

In addition :
My co-authors and I share everything

Always difficult !

- Naming conventions $\hookrightarrow$ pair programming
- Directory structure $\hookrightarrow$ relative paths in programs
- Data* $\hookrightarrow$ Dropbox, Web repository, Zenodo,..
- Programs $\hookrightarrow$ Dropbox, GitHub, Agora,
- Computing environment $\hookrightarrow$ Docker

In addition :
My co-authors and I share everything

Always difficult !

- Naming conventions $\hookrightarrow$ pair programming
- Directory structure $\hookrightarrow$ relative paths in programs
- Data* $\hookrightarrow$ Dropbox, Web repository, Zenodo,..
- Programs $\hookrightarrow$ Dropbox, GitHub, Agora,
- Computing environment $\hookrightarrow$ Docker
- Bibliography $\hookrightarrow$ Zotero

In addition :
My co-authors and I share everything

Always difficult !

- Naming conventions $\hookrightarrow$ pair programming
- Directory structure $\hookrightarrow$ relative paths in programs
- Data* $\hookrightarrow$ Dropbox, Web repository, Zenodo,..
- Programs $\hookrightarrow$ Dropbox, GitHub, Agora,
- Computing environment $\hookrightarrow$ Docker
- Bibliography $\hookrightarrow$ Zotero
- Current paper $\hookrightarrow$ Overleaf(LaTeX), DataJoy, Google Docs

In addition :
My co-authors and I share everything

Exchange messages/Manage your task

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.

In addition :
My co-authors and I share everything

Exchange messages/Manage your task

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?

In addition :
My co-authors and I share everything

Exchange messages/Manage your task

- If version control, the messages are attached to "commits"

In addition :
My co-authors and I share everything

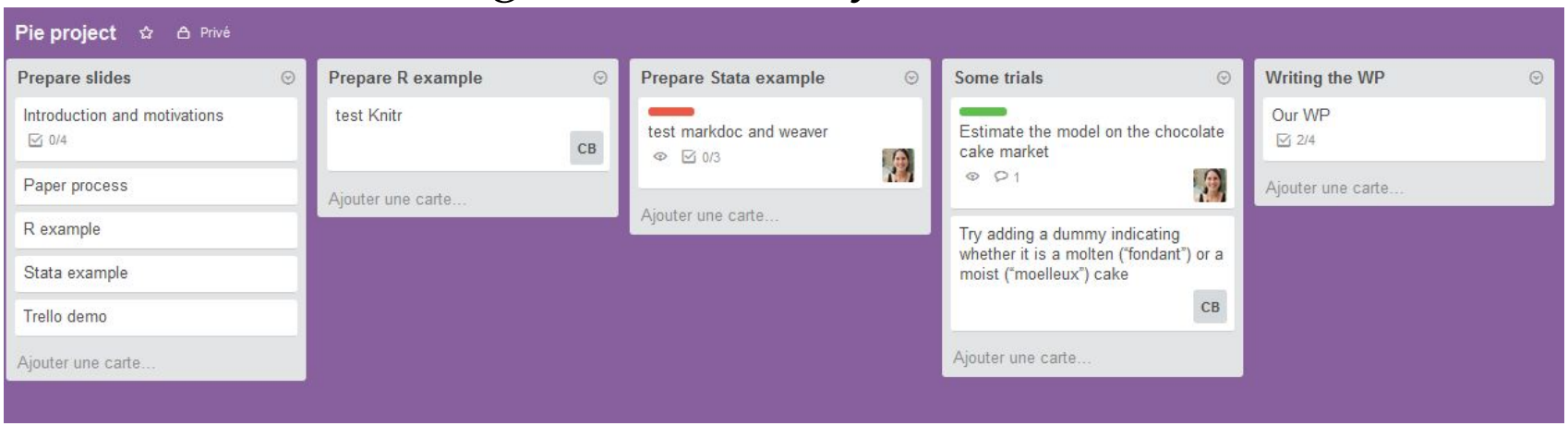
Exchange messages/Manage your task

- If version control, the messages are attached to "commits"
- Use an online management tasks system (Trello, Asana, ..)

In addition : My co-authors and I share everything

Exchange messages/Manage your task

- If version control, the messages are attached to "commits"
- Use an online management tasks system (Trello, Asana, ..)



Take away message

Benefits are greater that costs

- There are easy steps everybody can do

Take away message

Benefits are greater that costs

- There are easy steps everybody can do
- Automatic procedures reduce memory burden

Take away message

Benefits are greater that costs

- There are easy steps everybody can do
- Automatic procedures reduce memory burden
- Reusable code improves productivity

Take away message

Benefits are greater that costs

- There are easy steps everybody can do
- Automatic procedures reduce memory burden
- Reusable code improves productivity
- Sharing code improves productivity of others (network effect)

Take away message

Benefits are greater that costs

- There are easy steps everybody can do
- Automatic procedures reduce memory burden
- Reusable code improves productivity
- Sharing code improves productivity of others (network effect)
- Reproducible papers may be more cited (Gleditsch and Metelits, 2003)

Take away message

Benefits are greater that costs

- There are easy steps everybody can do
- Automatic procedures reduce memory burden
- Reusable code improves productivity
- Sharing code improves productivity of others (network effect)
- Reproducible papers may be more cited (Gleditsch and Metelits, 2003)
- Increase in the credibility of the research / trust in science (Leek and Peng, 2015)

Take away message

There are tools to help : Our experience

- The key element is the workflow

Take away message

There are tools to help : Our experience

- The key element is the workflow
- R is a nice software in term of reproducible features

TAKE AWAY MESSAGE

There are tools to help : Our experience

- The key element is the workflow
- R is a nice software in term of reproducible features
- Jupyter notebooks are nice for Python, Julia and others (including Stata)

TAKE AWAY MESSAGE

There are tools to help : Our experience

- The key element is the workflow
- R is a nice software in term of reproducible features
- Jupyter notebooks are nice for Python, Julia and others (including Stata)
- Reproducible documents and notebooks

TAKE AWAY MESSAGE

There are tools to help : Our experience

- The key element is the workflow
- R is a nice software in term of reproducible features
- Jupyter notebooks are nice for Python, Julia and others (including Stata)
- Reproducible documents and notebooks
  - are simple and attractive

Take away message

There are tools to help : Our experience

- The key element is the workflow
- R is a nice software in term of reproducible features
- Jupyter notebooks are nice for Python, Julia and others (including Stata)
- Reproducible documents and notebooks
  - are simple and attractive
  - can be complicated to have **only** one document

TAKE AWAY MESSAGE

There are tools to help : Our experience

- The key element is the workflow
- R is a nice software in term of reproducible features
- Jupyter notebooks are nice for Python, Julia and others (including Stata)
- Reproducible documents and notebooks
  - are simple and attractive
  - can be complicated to have **only** one document
  - are useful as companion papers

WHAT'S NEXT ?

Research actors have a role to play

- Journals should *ask* for data & code with submitted paper

WHAT'S NEXT ?

Research actors have a role to play

- Journals should *ask* for data & code with submitted paper
- Journals should **reproduce** results before publication

WHAT'S NEXT ?

Research actors have a role to play

- Journals should *ask* for data & code with submitted paper
- Journals should **reproduce** results before publication
- Research institutions (ERC, NFS, ANR, Universities,...) should **condition** their grants to RR projects

WHAT'S NEXT ?

Research actors have a role to play

- Journals should *ask* for data & code with submitted paper
- Journals should **reproduce** results before publication
- Research institutions (ERC, NFS, ANR, Universities,...) should **condition** their grants to RR projects
- Initiatives emerge :

WHAT'S NEXT ?

Research actors have a role to play

- Journals should *ask* for data & code with submitted paper
- Journals should **reproduce** results before publication
- Research institutions (ERC, NFS, ANR, Universities,...) should **condition** their grants to RR projects
- Initiatives emerge :

2011 : Executable papers on ScienceDirect,

WHAT'S NEXT ?

Research actors have a role to play

- Journals should *ask* for data & code with submitted paper
- Journals should **reproduce** results before publication
- Research institutions (ERC, NFS, ANR, Universities,...) should **condition** their grants to RR projects
- Initiatives emerge :

2011 : Executable papers on ScienceDirect,
2013 : *RunMyCode* and *ExecAndShare* (CNRS and ANR are partners) Hurlin et al. (2014)

WHAT'S NEXT ?

Research actors have a role to play

- Journals should *ask* for data & code with submitted paper
- Journals should **reproduce** results before publication
- Research institutions (ERC, NFS, ANR, Universities,...) should **condition** their grants to RR projects
- Initiatives emerge :

2011 : Executable papers on ScienceDirect,
2013 : *RunMyCode* and *ExecAndShare* (CNRS and ANR are partners) Hurlin et al. (2014)
2019 : Certification Agency for Scientific Code And Data, CASCAD (Pérignon et al., 2019)

WHAT'S NEXT ?

Research actors have a role to play

- Journals should *ask* for data & code with submitted paper
- Journals should **reproduce** results before publication
- Research institutions (ERC, NFS, ANR, Universities,...) should **condition** their grants to RR projects
- Initiatives emerge :

2011 : Executable papers on ScienceDirect,
2013 : *RunMyCode* and *ExecAndShare* (CNRS and ANR are partners) Hurlin et al. (2014)
2019 : Certification Agency for Scientific Code And Data, CASCAD (Pérignon et al., 2019)

- Education to RR should increase (MOOCs, workshops, ...)

WHAT'S NEXT ?

Research actors have a role to play

- Journals should *ask* for data & code with submitted paper
- Journals should **reproduce** results before publication
- Research institutions (ERC, NFS, ANR, Universities,...) should **condition** their grants to RR projects
- Initiatives emerge :

2011 : Executable papers on ScienceDirect,
2013 : *RunMyCode* and *ExecAndShare* (CNRS and ANR are partners) Hurlin et al. (2014)
2019 : Certification Agency for Scientific Code And Data, CASCAD (Pérignon et al., 2019)

- Education to RR should increase (MOOCs, workshops, ...)
- We, as researchers, teachers, referees, Deans can lead the way

WHAT'S NEXT ?

Research actors have a role to play

- Journals should *ask* for data & code with submitted paper
- Journals should **reproduce** results before publication
- Research institutions (ERC, NFS, ANR, Universities,...) should **condition** their grants to RR projects
- Initiatives emerge :

2011 : Executable papers on ScienceDirect,
2013 : *RunMyCode* and *ExecAndShare* (CNRS and ANR are partners) Hurlin et al. (2014)
2019 : Certification Agency for Scientific Code And Data, CASCAD (Pérignon et al., 2019)

- Education to RR should increase (MOOCs, workshops, ...)
- We, as researchers, teachers, referees, Deans can lead the way
- Data providers may help

REMEMBER

*"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)

HAVE YOU EVER HEARD (OR SAID YOURSELF) :

"I do not remember I did this..."

HAVE YOU EVER HEARD (OR SAID YOURSELF) :

"I do not remember I did this..."

"I don't understand anything at this code written by my previous student in intership"

HAVE YOU EVER HEARD (OR SAID YOURSELF) :

"I do not remember I did this..."

"I don't understand anything at this code written by my previous student in intership"

"It is surprising, I have different numbers now..."

HAVE YOU EVER HEARD (OR SAID YOURSELF) :

"I do not remember I did this..."

"I don't understand anything at this code written by my previous student in intership"

"Maybe I was using a different file with different individuals ? What have I done at that time ?"

"It is surprising, I have different numbers now..."

HAVE YOU EVER HEARD (OR SAID YOURSELF) :

"I do not remember I did this..."

"How did I compute this average value ? Arg... Maybe I removed some atypical observations..."

"I don't understand anything at this code written by my previous student in intership"

"Maybe I was using a different file with different individuals ? What have I done at that time ?"

"It is surprising, I have different numbers now..."

HAVE YOU EVER HEARD (OR SAID YOURSELF) :

"I do not remember I did this..."

"How did I compute this average value ? Arg... Maybe I removed some atypical observations..."

"I don't understand anything at this code written by my previous student in intership"

"Where the F...k is that damned file ?"

"Maybe I was using a different file with different individuals ? What have I done at that time ?"

"It is surprising, I have different numbers now..."

REFERENCES I

Barba, L. A. (2018). Terminologies for reproducible research. https://arxiv.org/abs/1802.03311.

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.).

Clemens, M. A. (2017). The meaning of failed replications : A review and proposal. *Journal of Economic Surveys*, 31(1) :326–342.

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.

REFERENCES II

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.

Gleditsch, N. P. and Metelits, C. (2003). The replication debate. *International Studies Perspectives*, 4(1) :72–79.

Hamermesh, D. S. (2013). Six decades of top economics publishing : Who and how ? *Journal of Economic Literature*, 51(1) :162–172.

Hurlin, C., Pérignon, C., and Stodden, V. (2014). Runmycode.org : a novel dissemination and collaboration platform for executing published computational results. *Open Science Framework*.

Huschka, D. (2013). Why should we share our data, how can it be organized, and what are the challenges ahead ? Ratswd german data forum.

Knuth, D. E. (1984). Literate programming. *The Computer Journal*, 27 :97–111.

REFERENCES III

Leek, J. T. and Peng, R. D. (2015). Opinion : Reproducible research can still be wrong : Adopting a prevention approach. *Proceedings of the National Academy of Sciences*, 112(6) :1645–1646.

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).

McCullough, B. D., McGeary, K. A., and Harrison, T. D. (2006). Lessons from the jmcb archive. *Journal of Money, Credit and Banking*, 38(4) :1093–1107.

McCullough, B. D., McGeary, K. A., and Harrison, T. D. (2008). Do economics journal archives promote replicable research ? *Canadian Journal of Economics*, 41(4) :pp. 1406–1420.

McCullough, B. D. and Vinod, H. D. (2003). Verifying the solution from a nonlinear solver : A case study. *American Economic Review*, 93(3) :873–892.

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

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

Rougier, N. P., Hinsen, K., Alexandre, F., Arildsen, T., Barba, L. A., Benureau, F. C., Brown, C. T., de Buyl, P., Caglayan, O., Davison, A. P., Delsuc, M.-A., Detorakis, G., Diem, A. K., Drix, D., Enel, P., Girard, B., Guest, O., Hall, M. G., Henriques, R. N., Hinaut, X., Jaron, K. S., Khamassi, M., Klein, A., Manninen, T., Marchesi, P., McGlinn, D., Metzner, C., Petchey, O., Plesser, H. E., Poisot, T., Ram, K., Ram, Y., Roesch, E., Rossant, C., Rostami, V., Shifman, A., Stachelek, J., Stimberg, M., Stollmeier, F., Vaggi, F., Viejo, G., Vitay, J., Vostinar, A. E., Yurchak, R., and Zito, T. (2017). Sustainable computational science : the rescience initiative. *PeerJ Computer Science*, 3 :e142.

Stodden, V., Bailey, D., Borwein, J., LeVeque, R., Rider, W., and Stein, W. (2013). Setting the default to reproducible : Reproducibility in computational and experimental mathematics.

References V

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.