

effective computing for research reproducibility

Laura Fortunato

University of Oxford

Santa Fe Institute

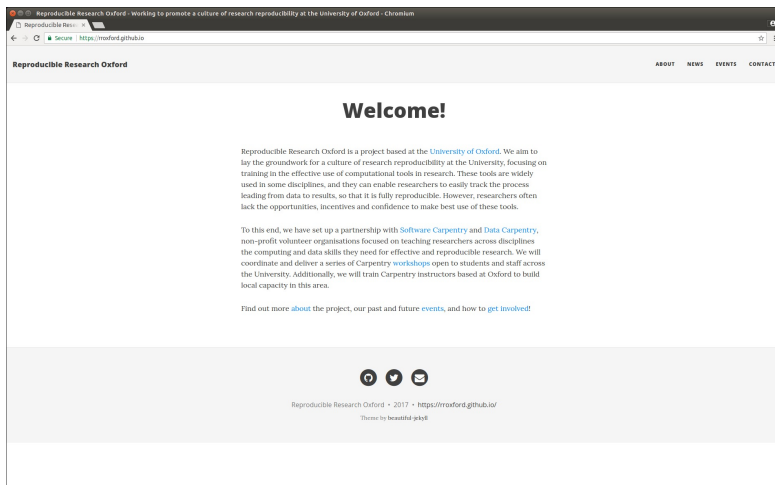
tweet @anthrolog

web <http://www.santafe.edu/~fortunato/>

email laura.fortunato@anthro.ox.ac.uk

Reproducible Research Oxford

<http://rroxford.github.io/>



The screenshot shows a web browser window with the title "Reproducible Research Oxford - Working to promote a culture of research reproducibility at the University of Oxford - Chromium". The address bar shows the URL "https://rroxford.github.io". The website has a navigation bar with the title "Reproducible Research Oxford" on the left and links "ABOUT", "NEWS", "EVENTS", and "CONTACT" on the right. The main content area features a large "Welcome!" heading, followed by two paragraphs of text and a line of links. The footer contains social media icons, the text "Reproducible Research Oxford • 2017 • <https://rroxford.github.io/>", and "Theme by beautiful-jekyll".

Reproducible Research Oxford - Working to promote a culture of research reproducibility at the University of Oxford - Chromium

Reproducible Res... x

Secure | <https://rroxford.github.io>

Reproducible Research Oxford




ABOUT NEWS EVENTS CONTACT

Welcome!

Reproducible Research Oxford is a project based at the [University of Oxford](#). We aim to lay the groundwork for a culture of research reproducibility at the University, focusing on training in the effective use of computational tools in research. These tools are widely used in some disciplines, and they can enable researchers to easily track the process leading from data to results, so that it is fully reproducible. However, researchers often lack the opportunities, incentives and confidence to make best use of these tools.

To this end, we have set up a partnership with [Software Carpentry](#) and [Data Carpentry](#), non-profit volunteer organisations focused on teaching researchers across disciplines the computing and data skills they need for effective and reproducible research. We will coordinate and deliver a series of Carpentry [workshops](#) open to students and staff across the University. Additionally, we will train Carpentry instructors based at Oxford to build local capacity in this area.

Find out more [about](#) the project, our past and future [events](#), and how to [get involved](#)!

Reproducible Research Oxford • 2017 • <https://rroxford.github.io/>

Theme by [beautiful-jekyll](#)

reproducibility crisis

the problem

a research result can only be trusted if it can be reproduced
but many published results are not reproducible!

e.g. data management malfunction

Brown et al. (2005)

Dance reveals symmetry especially in young men.

Nature, **438**:1148–1150



e.g. data management malfunction

Brown et al. (2005)

Dance reveals symmetry especially in young men.
Nature, **438**:1148–1150

Reich (2013)

Symmetry study deemed a fraud.
Nature, **497**:170–171



*“... Brown says that it is unclear
which data set is the original
because many versions exist.”*

reproducibility crisis

the problem

a research result can only be trusted if it can be reproduced
but many published results are not reproducible!

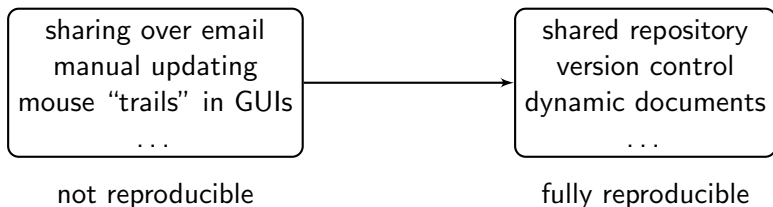
reproducibility crisis

the problem

a research result can only be trusted if it can be reproduced
but many published results are not reproducible!

(part of) the solution

change in approach to computing in research



a slogan

Buckheit and Donoho (1995) WaveLab and reproducible research.

https://doi.org/10.1007/978-1-4612-2544-7_5

*An article about computational science
in a scientific publication is **not** the scholarship itself,
it is merely **advertising** of the scholarship.*

*The actual scholarship is the complete
software development environment and the complete
set of instructions which generated the figures.*

but where to start...

books, articles, etc.

- ▶ Haddock S. H. D. & Dunn, C. W. (2011)
Practical computing for biologists. Sinauer Associates, Inc.
- ▶ Wilson G. et al. (2014)
Best practices for scientific computing.
PLoS Biology 12(1): e1001745.
<https://doi.org/10.1371/journal.pbio.1001745>
- ▶ Wilson G. et al. (2017)
Good enough practices in scientific computing.
PLoS Computational Biology 13(6): e1005510.
<https://doi.org/10.1371/journal.pcbi.1005510>

but where to start...

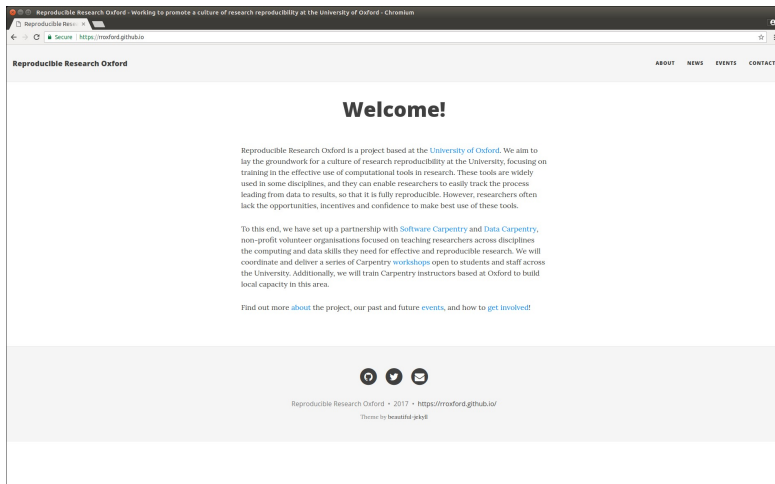
tutorials, workshops, etc

- ▶ various courses available through
<http://courses.it.ox.ac.uk/catalogue>
- ▶ Software Carpentry lessons, workshops
<https://software-carpentry.org/>
- ▶ Data Carpentry lessons, workshops
www.datacarpentry.org/



Reproducible Research Oxford

<http://rroxford.github.io/>



The screenshot shows a web browser window with the title "Reproducible Research Oxford - Working to promote a culture of research reproducibility at the University of Oxford - Chromium". The address bar shows the URL "https://rroxford.github.io". The website has a navigation bar with the title "Reproducible Research Oxford" on the left and links for "ABOUT", "NEWS", "EVENTS", and "CONTACT" on the right. The main content area features a large "Welcome!" heading, followed by two paragraphs of text and a line of links. The footer contains social media icons for GitHub, Twitter, and Email, the text "Reproducible Research Oxford • 2017 • <https://rroxford.github.io/>", and the phrase "Theme by beautiful-jekyll".

Reproducible Research Oxford - Working to promote a culture of research reproducibility at the University of Oxford - Chromium

Reproducible Res... x

Secure | <https://rroxford.github.io>

Reproducible Research Oxford

ABOUT NEWS EVENTS CONTACT

Welcome!

Reproducible Research Oxford is a project based at the [University of Oxford](#). We aim to lay the groundwork for a culture of research reproducibility at the University, focusing on training in the effective use of computational tools in research. These tools are widely used in some disciplines, and they can enable researchers to easily track the process leading from data to results, so that it is fully reproducible. However, researchers often lack the opportunities, incentives and confidence to make best use of these tools.

To this end, we have set up a partnership with [Software Carpentry](#) and [Data Carpentry](#), non-profit volunteer organisations focused on teaching researchers across disciplines the computing and data skills they need for effective and reproducible research. We will coordinate and deliver a series of Carpentry [workshops](#) open to students and staff across the University. Additionally, we will train Carpentry instructors based at Oxford to build local capacity in this area.

Find out more [about](#) the project, our past and future [events](#), and how to [get involved](#)!

Reproducible Research Oxford • 2017 • <https://rroxford.github.io/>

Theme by [beautiful-jekyll](#)

the project

overall aim

seed a process of change in approach to research computing
by teaching researchers the skills they need for reproducibility

the project

overall aim

seed a process of change in approach to research computing by teaching researchers the skills they need for reproducibility

partnership with Software/Data Carpentry

funded by the IT Innovation seed fund in June 2016

- ▶ delivery of 4 workshops/year
- ▶ training of local instructors



get involved!

contact

tweet @RR_Oxford
#RROxford

web <https://RROxford.github.io/>

mailist <https://web.maillist.ox.ac.uk/ox/info/rroxford>

email ReproducibleResearchOxford@gmail.com

get involved!

contact

tweet @RR_Oxford
#RROxford

web <https://RROxford.github.io/>

mailist <https://web.maillist.ox.ac.uk/ox/info/rroxford>

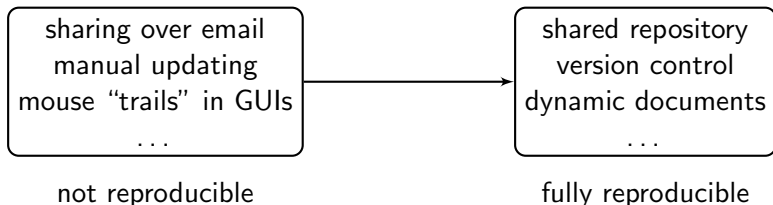
email ReproducibleResearchOxford@gmail.com

Software/Data Carpentry workshops

- ▶ attend as a learner
- ▶ volunteer as a host, helper, and/or instructor
- ▶ next workshop: **October 12–13**, Software Carpentry (Unix shell, version control with git, programming in R)

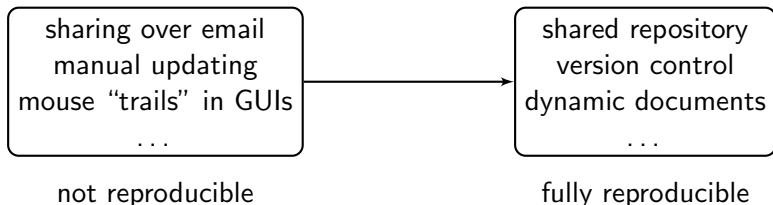
building mental models

two approaches



building mental models

two approaches

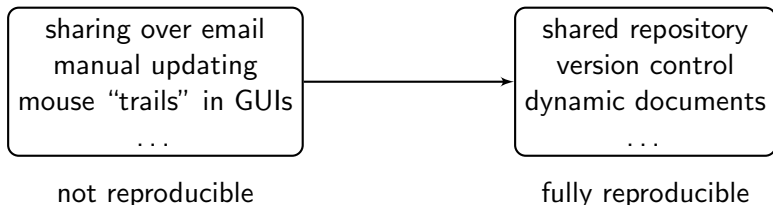


a simple task

given a set of text files (e.g. machine output)
find the file with the smallest no. of lines

building mental models

two approaches



a simple task

given a set of text files (e.g. machine output)
find the file with the smallest no. of lines

demo

manual vs. Unix shell scripting

basic operations

<http://swcarpentry.github.io/shell-novice/>

```
1 # change directory
2 cd molecules
3
4 # list content of directory
5 ls
6
7 # print content of file cubane.pdb
8 more cubane.pdb
9
10 # print first 5 lines of file
11 head -n 5 cubane.pdb
```

word count

<http://swcarpentry.github.io/shell-novice/>

```
1 # print first 5 lines of file
2 head -n 5 cubane.pdb
3
4 # print no. of lines, words, chars
5 wc cubane.pdb
6
7 # as above,
8 # for any file ending in .pdb
9 wc *.pdb
10
11 # as above,
12 # but print lines only
13 wc -l *.pdb
```

redirection

<http://swcarpentry.github.io/shell-novice/>

```
1 # print no. of lines for any file ending in .pdb
2 wc -l *.pdb
3
4 # as above,
5 # with output redirected to file lengths.txt
6 wc -l *.pdb > lengths.txt
7
8 # print content of file
9 more lengths.txt
```

redirection

<http://swcarpentry.github.io/shell-novice/>

```
1 # print content of file
2 more lengths.txt
3
4 # sort (numeric) content of file
5 sort -n lengths.txt
6
7 # as above,
8 # with output redirected to file lengths-sorted.txt
9 sort -n lengths.txt > lengths-sorted.txt
10
11 # print first line of file
12 head -n 1 lengths-sorted.txt
```

pipes

<http://swcarpentry.github.io/shell-novice/>

recap

```
1 # print no. of lines for any file ending in .pdb,  
2 # with output redirected to file lengths.txt  
3 wc -l *.pdb > lengths.txt  
4  
5 # sort (numeric) content of file lengths.txt,  
6 # with output redirected to file lengths-sorted.txt  
7 sort -n lengths.txt > lengths-sorted.txt  
8  
9 # print first line of file lengths-sorted.txt  
10 head -n 1 lengths-sorted.txt
```

pipes

<http://swcarpentery.github.io/shell-novice/>

alternatively...

```
1 # print no. of lines for any file ending in .pdb
2 wc -l *.pdb
3
4 # as above,
5 # with output piped through sort
6 wc -l *.pdb | sort -n
7
8 # as above,
9 # with output piped through head
10 wc -l *.pdb | sort -n | head -n 1
```


scripts

<http://swcarpentry.github.io/shell-novice/>

typing commands at the prompt...

```
1 # print no. of lines for shortest file
2 wc -l *.pdb | sort -n | head -n 1
```

scripts

<http://swcarpentry.github.io/shell-novice/>

typing commands at the prompt...

```
1 # print no. of lines for shortest file
2 wc -l *.pdb | sort -n | head -n 1
```

...vs. saved as script find-shortest.sh

```
# Show filename with smallest no. of lines
# Usage: bash find-shortest.sh filename(s)

wc -l "$@" | sort -n | head -n 1
```

scripts

<http://swcarpentry.github.io/shell-novice/>

typing commands at the prompt...

```
1 # print no. of lines for shortest file
2 wc -l *.pdb | sort -n | head -n 1
```

...vs. saved as script find-shortest.sh

```
1 # print content of file find-shortest.sh
2 more find-shortest.sh
3
4 # run script on files ending in .pdb
5 bash find-shortest.sh *.pdb
```

version control

<http://swcarpentry.github.io/hg-novice/>

<http://swcarpentry.github.io/git-novice/>

who did what, when?

```
1 # show revision history
2 hg log
```