

# Emacs, Org-Mode, and Reproducible Research

Zheng Tian

# Outline

- 1 Introduction
- 2 Emacs
- 3 Org-Mode
- 4 Reproducible research with org-mode

# Have you ever experienced this?

## PROGRAMMING FOR NON-PROGRAMMERS



JORGE CHAM © 2014

WWW.PHDCOMICS.COM

# Reproducible research

*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.*  
– Buckheit and Donoho (1995)

# How can we do reproducible research?

The screenshot shows a GNU Emacs window with a presentation slide. The slide title is "Emacs, Org-Mode, and Reproducible Research". The slide content includes a welcome message to GNU Emacs, important help menu items, useful tasks, and GNU Emacs version information. The Emacs buffer shows the Org-Mode content of the slide, including the title, author (Zheng Tian), and a table of contents with items like "Introduction", "Emacs...", and "Org-Mode...".

```

Welcome to GNU Emacs, a part of the GNU operating system.
To follow a link, click Mouse-1 on it, or save to it and type RET.
To quit a partially entered command, type Control-g.

Important Help menu items:
Emacs Tutorial      Learn basic Emacs keystroke commands
Read the Emacs Manual  View the Emacs manual using Info
(Non)Warranty       GNU Emacs comes with ABSOLUTELY NO WARRANTY
Copying Conditions   Conditions for redistributing and changing Emacs
More Manuals / Ordering Manuals  How to order printed manuals from the FSF

Useful tasks:
Visit New File       Specify a new file's name, to edit the file
Open Home Directory Open your home directory, to operate on its files
Customize StartUp    Change initialization settings including this screen

GNU Emacs 25.1.1 (x86_64-apple-darwin13.4.0, NS appkit-1265.21 Version 10.9.5 (Build 13F1911))
of 2016-09-20
Copyright (C) 2016 Free Software Foundation, Inc.

If an Emacs session crashed recently, type M-x recover-session RET
to recover the files you were editing.

#*TITLE: Emacs, Org-Mode, and Reproducible Research
#*AUTHOR: Zheng Tian
#*EMAIL: zngtian@gmail.com
#*DATE:
#*PROPERTY: header-args:R :session #R-workshop :eval no

#*STARTUP: beamer
#*OPTIONS: toc:1 Hi:3
#*LATEX_CLASS: beamer
#*LATEX_CLASS_OPTIONS: [presentation]
#*BEAMER_THEME: AnnArbor
#*BEAMER_COLOR_THEME:
#*BEAMER_ENV: \documentclass[10pt]{beamer}
#*COLUMNS: %\setitemize[10pt]{beamer}
#*PROPERTY: BEAMER_col:ALL 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 0.0 :ETC

#*OPTIONS: reveal_centerit reveal_progressit reveal_history:nil reveal_controlit
#*OPTIONS: reveal_rolling_links:t reveal_keyboard:t reveal_overview:t num:nil
#*OPTIONS: reveal_width:1000 reveal_height:800
#*REVEAL_MARGIN: 0.2
#*REVEAL_MIN_SCALE: 0.5
#*REVEAL_MAX_SCALE: 2.5
#*REVEAL_TRANS: linear
#*REVEAL_THEME: moon
#*REVEAL_LEVEL: 2
#*REVEAL_PLUGINS: (highlight notes zoom)
#*REVEAL_EXTRA_JS: ( src "file:///Users/ztian/reveal.js/plugin/reveal.js-menu/menu.js"
#*REVEAL_MATHJAX_URL: file:///Users/ztian/gtdDownloads/MathJax/MathJax.js?config=TeX-
MML-As-Images

• TODO Introduction...

• Emacs...

• Org-Mode...

```

# What is Emacs? A text editor.

GNU Emacs is a free, portable, extensible text editor.

- Free: Open source, freely copyable and redistributable.
- Portable: Run on many machines under different operating systems.
- Extensible: Customizable for all aspect and many contributed packages,

Emacs is a  $\text{\LaTeX}$  editor via AucTex

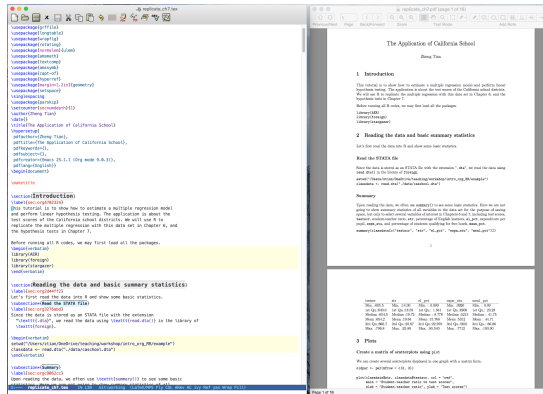


Figure: An illustration of the AucTex mode

# Emacs is IDE for programming languages

- Edit code with syntax highlighting
- Execute code within Emacs

```

library(eggplant2)
library(dplyr)

# Ch.2 - ml_data_frame(egg)

## 3.2.1 Exercises
summary(egg)
# str(egg)
# head(egg)
# tail(egg)
# names(egg)
# class(egg)
# class(egg)

data(package = "eggplant2")
# 1 mile = 1.6093 km, 1 gallon = 3.7854 liters,
# 1 min/gallon = 3.7854/1.6093/60 = 3.7854/96.5616

egg1km <- function(egg) return(egg * 3.7854/(1.6093*60))
egg1km(egg)

model <- egg %>%
  group_by(manufacturer) %>%
  summarise(nmodel = length(unique(model)))

with(nmodel, manufacturer[which.max(nmodel)])

ggplot(egg, aes(displ, hwy)) + # data: egg, aes(x, y, other settings)
  geom_point()

ggplot(egg, aes(displ, cty, color = class)) +
  geom_point()

ggplot(egg, aes(displ, hwy)) +
  geom_point() +
  facet_wrap(~class)

library(MASS)
ggplot(egg, aes(displ, hwy)) +
  geom_point() +
  geom_smooth(method = "lm")

library(MASS)
ggplot(egg, aes(displ, hwy)) +
  geom_point() +
  geom_smooth(method = "lm")

library(MASS)
ggplot(egg, aes(drv, hwy)) +
  geom_smooth(method = "lm")

# egg1km <- function(egg) return(egg * 3.7854/(1.6093*60))
# egg1km(egg)
# model <- egg %>%
#   group_by(manufacturer) %>%
#   summarise(nmodel = length(unique(model)))
# with(nmodel, manufacturer[which.max(nmodel)])
# ggplot(egg, aes(displ, hwy)) + # data: egg, aes(x, y, other settings)
#   geom_point()
# ggplot(egg, aes(displ, cty, color = class)) +
#   geom_point()
# ggplot(egg, aes(displ, hwy)) +
#   geom_point() +
#   facet_wrap(~class)
# library(MASS)
# ggplot(egg, aes(displ, hwy)) +
#   geom_point() +
#   geom_smooth(method = "lm")
# library(MASS)
# ggplot(egg, aes(displ, hwy)) +
#   geom_point() +
#   geom_smooth(method = "lm")
# library(MASS)
# ggplot(egg, aes(drv, hwy)) +
#   geom_smooth(method = "lm")

```

Figure: An illustration of the ESS mode



# Emacs is an operating system

- Emacs is an operating system, easily managing files and folders within a dired-directory interface.

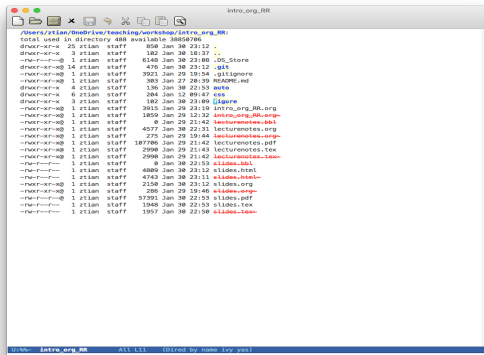


Figure: Emacs as an operating system with the dired mode

# Emacs is a Git tool via Magit

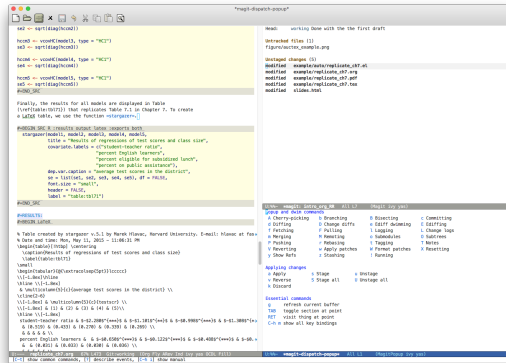


Figure: An illustration of the Magit mode

# Emacs can do many other things

- Emacs can do spell checking, reading news, checking and sending emails, etc., through plenty of contributed packages.
- Most importantly, Emacs enable researchers to manage research project, take notes, and write dynamic documentation.

# Installation

- Homepage of GNU Emacs:  
<https://www.gnu.org/software/emacs/>
- Vincent Goulet's binary files:  
<http://vgoulet.act.ulaval.ca/en/emacs/>

I personally prefer the second option because it has already included some of the mostly used packages.

# Configuration

Emacs is customizable and all customized configuration can be done with either a `.emacs` file or `init.el` under the directory `~/.emacs.d.` With some settings, we can use an `org` file to organize and apply your customization.

# My settings

All my settings have been uploaded to Github from where you can download or clone git clone <https://github.com/zngtian/.emacs.d.git>.

## A peek of my settings

[init.el](#) <https://github.com/zngtian/.emacs.d/blob/master/init.el>

[myconfig.org](#) <https://github.com/zngtian/.emacs.d/blob/master/myconfig.org>

# Notation

In Emacs documentation, we often see the following notations

**C-x** Press Control key and x

**M-x** Press Alt key and x

**RET** Press the return key

**SPC** Press the space bar

**ESC** Press the escape key

**S-<TAB>** Press shift and tab keys

# Buffer and windows

## The basic user interface of Emacs in buffers and windows

The screenshot illustrates the Emacs editor interface with several components highlighted:

- Window 1:** The main editing area showing R code for reading data from a file.
- Buffer 2:** A buffer containing R code, likely the source file being edited.
- Buffer 3:** A buffer containing R code, likely the source file being edited.
- Window 2:** A window displaying the output of an R command, showing a summary of a matrix.
- Buffer 1:** A buffer containing R code, likely the source file being edited.
- This is a mini buffer:** A small buffer at the bottom left showing the command 'switch to buffers'.

The Emacs interface also includes a sidebar on the left with the Emacs menu, a status bar at the bottom right showing the current file and line, and a mini-buffer at the bottom left showing the command 'switch to buffers'.



# Motion within a buffer

C-f and M-f move forward by one character and by one word

C-b and M-b move backward by one letter and by one word

C-n and C-p move downward and upward

C-v and M-v scroll down and up

M-< and M-> move to the start and to the end of a buffer

# Switch buffers and windows

`C-x 2` open a new buffer

`C-x 0` close the current buffer

`C-x b` switch to a buffer

`C-x o` switch between two opened buffers

`C-x 4 b` switch to a buffer and open it as a new one

`C-x 5 2` and `C-x 5-0` open and close a new window

# Open, save, quit, and help

`C-x C-f` open a new file

`C-x C-s` save the current buffer

`C-x s` save all files

`C-g` cancel the currently invoked command. VERY IMPORTANT!

`C-x C-c` exit Emacs

`C-h ?/m/a` get help

# Copy, yank, and delete

**C-SPC** set a mark and move the cursor around to select a region

**C-w** kill (cut)

**M-w** copy

**C-y** yank (paste)

**DEL** and **C-d** delete a character backward and forward

**M-DEL** and **M-d** delete a word backward and forward

**C-k** kill a line.

**C-x u** undo the previous editing.

# Tutorial and cheat sheet

`C-h t` open the complete tutorial

A guided tour <https://www.gnu.org/software/emacs/tour/>

Cheat sheet <https://www.gnu.org/software/emacs/refcards/pdf/refcard.pdf>

# What is org-mode

Org mode is one of the most popular contributed packages in Emacs. It can accomplish a variety of work including, but not limited to,

- taking notes with structured documentation,
- assigning tasks and scheduling them,
- editing tables and doing calculation,
- exporting to pdf, html, odt files,
- working with source code.

# Headline

\* Top level headline

\*\* Second level

\*\*\* 3rd level

some text

\*\*\* 3rd level

more text

\* Another top level headline

<TAB> hit at a headline to see and hide the content under it

S-<TAB> global cycling.

M-left and M-right promote and demote a heading

# Lists

- Unordered list
  - + Item 1
  - + Item 2
- Ordered list
  1. first thing
  2. second thing
  3. third thing
- Description
  - Tom :: a cat
  - Jerry :: a mouse
- List with check box
  - [X] Do this
  - [ ] Do that



# Links

- The basic syntax for a link:  
`[[link][description]]` or `[[link]]`
- Internal link: 24  
`[[Lists]]`
- External link: slides.tex  
`[[file:slides.tex]]`
- URL: `http://rri.wvu.edu/`  
`[[http://rri.wvu.edu/]]`

# Blocks

- Blocks are defined by `#+BEGIN_...` and `#+END_...`
- The CENTER block

This sentence will be centered in the exported file

```
#+BEGIN_CENTER
```

This sentence will be centered in the exported file

```
#+END_CENTER
```

- The QUOTE block

*Everything should be made as simple as possible, but not  
any simpler – Albert Einstein*

```
#+BEGIN_QUOTE
```

Everything should be made as simple as possible,  
but not any simpler -- Albert Einstein

```
#+END_QUOTE
```

# Mathematics

Org mode can contain  $\text{\LaTeX}$  math fragments that don't need any special marking. Just do as in  $\text{\LaTeX}$ .

```
\begin{equation}
  x=\sqrt{b}
\end{equation}
```

If  $a^2=b$  and  $(b=2)$ , then the solution must be either  $a=+\sqrt{2}$  or  $a=-\sqrt{2}$

$$x = \sqrt{b} \tag{1}$$

If  $a^2 = b$  and  $b = 2$ , then the solution must be either

$$a = +\sqrt{2} \text{ or } a = -\sqrt{2}$$

# Table

- '|' as the first non-whitespace character starts a table. The following texts yield a table in HTML export

```
| Name | Age | Score |
|-----+-----+-----|
| Peter | 17 | 1234 |
| Anna | 25 | 4321 |
```

Name	Age	Score
Peter	17	1234
Anna	25	4321

# Calculation in a table

- We can define formula for a field, a row, or a column by starting a field with "="

```
| Name | Age | Score |
|-----+-----+-----|
| Peter | 17 | 1234 |
| Anna | 25 | 4321 |
|      | 21 | 5555 |
```

```
#+TBLFM: @4$2=vmean(@2..@3)::@4$3=vsum(@2..@3)
```

Name	Age	Score
Peter	17	1234
Anna	25	4321
	21	5555

# Exporting

- An org file can be exported to a variety of formats, including latex, beamer, html, odt, etc.
- `C-c C-e`: start the export dispatcher.
- Then you can select from several options. Try `C-c C-e 1 0` to generate the beamer file.
- This presentation is exported with the `og-reveal` package.
- Export settings can be set using some keywords, such as `#+TITLE`, `#+AUTHOR`, `#+OPTIONS`, `#+LATEX_HEADER`, `#+HTML_HEADER`, etc.

# A sneak peek of my agenda

Org mode is not just a text editor that can include a rich variety of elements but also a handy tool to plan daily life and manage research projects.

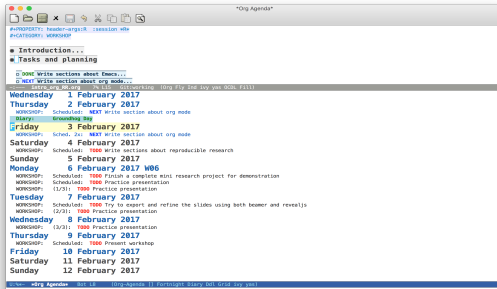


Figure: An illustration of agenda view

# To-do items

TODO items in org mode are headlines defined by TODO keywords after asterisks.

- \* [#A] TODO Do this first.
- \* DONE This task has been done

M-S RET quick enter a TODO item

S-right/left cycle through TODO status

S-up/down cycle through priorities.



# Schedule and deadline

We can set schedule and deadline to TODO items.

`C-c C-s` set a day and time to begin doing this item

`C-c C-d` set a deadline

\* [#A] TODO Do this first.

SCHEDULED: <2017-02-03 Fri>

\* DONE This task has been done

DEADLINE: <2017-02-03 Fri>

- Time stamps are generated using the calendar minor mode.

# Agenda view

All TODO items, schedules, and deadlines can be viewed in the Agenda view in org mode.

`C-c a a` start the agenda view

`C-c a t` see all TODO items

`C-c a m` filter TODO items by tags

Within the agenda view, you can filter by tag, change the status, and go to the headline of a TODO item.

# What is reproducible research?

*The data and code used to make a finding are available and they are sufficient for an independent researcher to recreate the finding. – Gandrud (2015)*

# Why should we do reproducible research?

## For readers

- Easy for reviewers to test and validate your findings.
- Easy for readers to reuse your code in their research.
- Make your paper a reliable citation.

## For ourselves

- Easy for us to tract and retrospect what we have done.
- Helpful to have good research habits and workflow.
- Facilitating team work.

# What is a workflow of reproducible research?

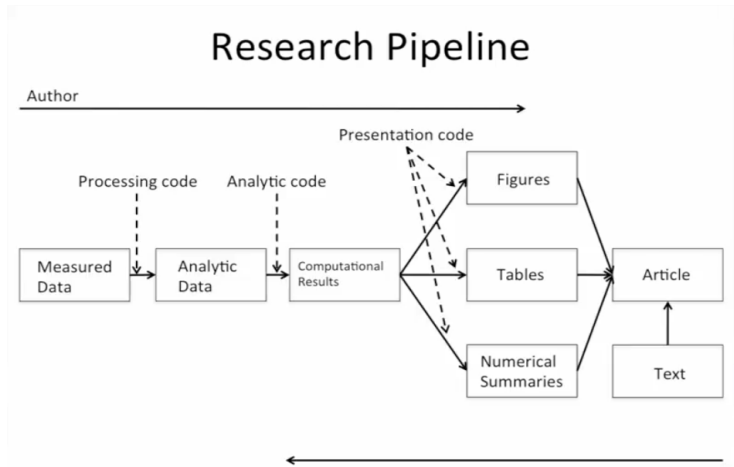


Figure: A workflow of reproducible research (Source: Peng (2015))

# What are necessary elements of reproducible research?

Roger Peng (2015) summarizes four essential elements to make results reproducible:

- Analytical data
- Analytical code
- Documentation
- Distribution

# What is literate programming?

Literate programming (Donald Knuth, 1992) is the central part of reproducible research.

Typically, literate programming involves the following three steps (Xie, 2015):

- ➊ parse the source document and separate the code from narratives;
- ➋ execute the source code and return results;
- ➌ mix results from the source code with the original narratives.

# Available tools for literate programming

- WEB (Knuth, 1983)
- Noweb (Ramsey, 1994)
- roxygen2 (Wickham et al., 2015)
- knitr (Xie, 2015b)
- Jupyter(IPython) Notebook
- Emacs org mode



# Source code block

The basic structure of code blocks is as follows

```
#+NAME: <name>
#+BEGIN_SRC <language> <switches> <header arguments>
  <body>
#+END_SRC
```

The structure of an inline code block is

```
src_<language> [<header arguments>] {<body>}
```

# Basic settings

```
#+BEGIN_SRC emacs-lisp :eval no
(org-babel-do-load-languages
 'org-babel-load-languages
 '((R . t)
   (python . t)
   (emacs-lisp . t)
   (calc . t)
   (latex . t)
   (org . t)
   (sh . t)))

(setq org-confirm-babel-evaluate nil)
#+END_SRC
```

# Header arguments

Header arguments fine-tune the behaviors of a source block.

Header arguments	Example
:exports	:exports results or :exports none
:results	:results value table or :results silent
:eval	:eval no
:cache	:cache yes
:file	:file ./img/figure1.png

# Results in raw format

```
#+BEGIN_SRC R :exports both :results output
library(ggplot2)
head(mpg[1:5])
#+END_SRC
```

#+RESULTS:

	manufacturer	model	displ	year	cyl
: 1	audi	a4	1.8	1999	4
: 2	audi	a4	1.8	1999	4
: 3	audi	a4	2.0	2008	4
: 4	audi	a4	2.0	2008	4
: 5	audi	a4	2.8	1999	6
: 6	audi	a4	2.8	1999	6

# Results in org tables

```
#+BEGIN_SRC R :exports results :results value table :colnames y
head(mpg[1:5])
#+END_SRC
```

```
#+RESULTS[f45a5d1174dd12cdb343701a0868203eda23a5bc]:
```

manufacturer	model	displ	year	cyl
audi	a4	1.8	1999	4
audi	a4	1.8	1999	4
audi	a4	2	2008	4
audi	a4	2	2008	4
audi	a4	2.8	1999	6
audi	a4	2.8	1999	6

# Results in figures

```
#+BEGIN_SRC R :exports both :results output graphics :file mpg.  
  ggplot(mpg, aes(displ, cty, colour = class)) +  
    geom_point()  
#+END_SRC  
  
#+ATTR_HTML: :width 600 :height 500  
#+ATTR_LATEX: :width 0.6\textwidth :height 0.6\textheight  
#+RESULTS:  
[[file:figure/mpg.png]]
```

# The figure generated

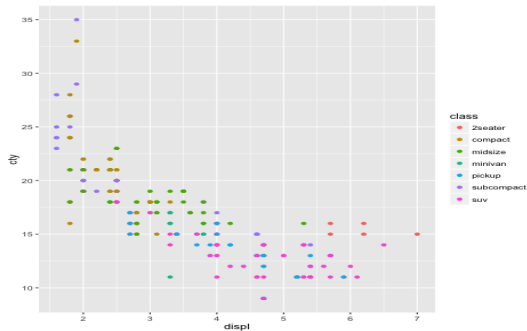


Figure: The Scatterplot Between the Engine Displacement and City MPG

# Results in latex

```
#+BEGIN_SRC R :exports both :results output latex
library(stargazer)
stargazer(mpg, header = FALSE)
#+END_SRC
```

```
#+RESULTS:
#+BEGIN_EXPORT latex
```

```
% Table created by stargazer v.5.2 by Marek Hlavac, Harvard Uni
% Date and time: Mon, Feb 06, 2017 - 09:45:31
```

```
\begin{table}[!htbp] \centering
  \caption{}
  \label{}
```

```
\begin{tabular}{@{\extracolsep{5pt}}lccccc}
\\[-1.8ex]\hline
\hline \\[-1.8ex]
```



# The $\text{\LaTeX}$ table generated

**Table:** Summary Statistics of the `=mpg=` dataset

Statistic	N	Mean	St. Dev.	Min	Max
displ	234	3.472	1.292	1.600	7.000
year	234	2,003.500	4.510	1,999	2,008
cyl	234	5.889	1.612	4	8
cty	234	16.859	4.256	9	35
hwy	234	23.440	5.955	12	44

# An mini example of literate programming

The following file is an example of reproducible research, which I used in teaching Econometrics.

`example/replicate_ch7.org`

# Tutorial sources

John Kitchin's blog <http://kitchingroup.cheme.cmu.edu/blog/2014/08/08/What-we-are-using-org-mode-for/>

John Kitchin's presentation on YouTube [https://www.youtube.com/watch?v=1-dUkyn\\_fZA&t=126s](https://www.youtube.com/watch?v=1-dUkyn_fZA&t=126s)

Vikas Rawal's tutorial <https://github.com/vikasrawal/orgpaper/blob/master/orgpapers.org>