Module 1 - Lesson 01

History & Importance of Reproducibility & Transparency

Melinda K. Higgins, PhD.

2017-09-02

Outline

- Timeline Reproducible Research & Transparency
- People
- Books
- Literate Programming > Dynamic Documentation > [R]Markdown
- The Big Picture

Timeline Reproducible Research & Transparency[1]

YEAR	Event
1992	Jon Claerbout coined the term "reproducible research" in his book "EARTH SOUNDINGS ANALYSIS: Processing versus Inversion (PVI)" [2]
1996	CONSORT statement introduced standards for reporting clinical trials [3]
2004	International Committee of Medical Journal Editors (ICMJE) stated they would not publish a clinical trial that had not been registered. [4]
2005	Ioannidis, J. P. A. Why most published research findings are false. PLoS Med. 2, e124 (2005) [5]

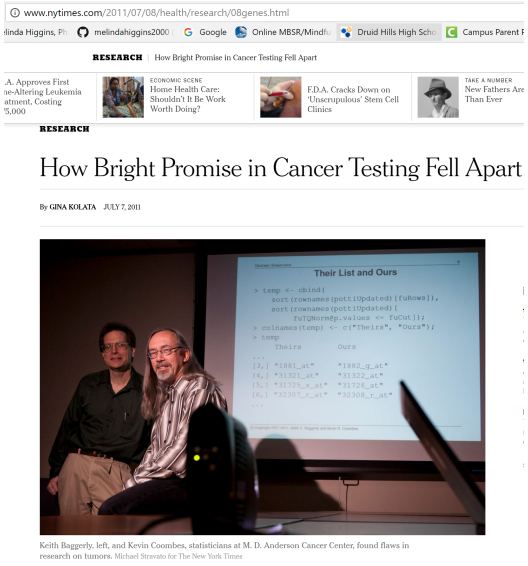
1. Timeline partially based on PLOS Blog December 2016 http://blogs.plos.org/absolutely-maybe/2016/12/05/reproducibility-crisis-timeline-milestones-in-tackling-research-reliability/
2. http://sepwww.stanford.edu/sep/jon/reproducible.html
3. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 276:637-639.
4. http://www.icmje.org/news-and-editorials/update_2005.html
5. https://doi.org/10.1371/journal.pmed.0020124

Timeline Reproducible Research & Transparency

YEAR	Event
2007	FDA Amendments Act (FDAAA) required more types of clinical trials to be registered (final rules took effect January 2017) [6]
2009	Journal of Biostatistics institutes policy to work with authors to publish articles that meet a standard of reproducibility.[7]
2011	Alsheikh-Ali, et.al. (2011), report the low percentage of researchers satisfying the policies regarding the availability and sharing of their data.[8]

6. https://clinicaltrials.gov/ct2/manage-recs/fdaaa
7. https://academic.oup.com/biostatistics/article/10/3/405/293660/Reproducible-research-and-Biostatistics & https://academic.oup.com/biostatistics/pages/General_Instructions
8. Alsheikh-Ali, A. A., Qureshi, W., Al-Mallah, M. H. & Ioannidis, J. P. Public availability of published research data in high-impact journals. PloS ONE 6, e24357, 2011; https://doi.org/10.1371/journal.pone.0024357

Cancer Testing Falls Apart



www.nytimes.com/2011/07/08/health/research/08genes.html

RESEARCH | How Bright Promise in Cancer Testing Fell Apart

...A. Approves First ...ne-Altering Leukemia ...atment, Costing ...5,000

ECONOMIC SCENE
Home Health Care: Shouldn't It Be Work Worth Doing?

F.D.A. Cracks Down on 'Unscrupulous' Stem Cell Clinics

TAKE A NUMBER
New Fathers Are ... Than Ever

RESEARCH

How Bright Promise in Cancer Testing Fell Apart

By GINA KOLATA JULY 7, 2011

Keith Baggerly, left, and Kevin Coombes, statisticians at M. D. Anderson Cancer Center, found flaws in research on tumors. Michael Stravato for The New York Times

The Duke saga began when a prestigious journal, Nature Medicine, published a paper on Nov. 6, 2006, by Dr. Anil Potti, a cancer researcher at Duke University Medical Center; Joseph R. Nevins, a senior scientist there; and their colleagues. They wrote about genomic tests they developed that looked at the molecular traits of a cancerous tumor and figured out which chemotherapy would work best.

First, though, he asked two statisticians at M. D. Anderson, Keith Baggerly and Kevin Coombes, to check the work. Several other doctors approached them with the same request.

Dr. Baggerly and Dr. Coombes found errors almost immediately. Some seemed careless — moving a row or a column over by one in a giant spreadsheet — while others seemed inexplicable. The Duke team shrugged them off as "clerical errors."

And the Duke researchers continued to publish papers on their genomic signatures in prestigious journals. Meanwhile, they started three trials using the work to decide which drugs to give patients.

Dr. Baggerly and Dr. Coombes tried to sound an alarm. They got the attention of the National Cancer Institute, whose own investigators wanted to use the Duke system in a clinical trial but were dissuaded by the criticisms. Finally, they published their analysis in The Annals of Applied Statistics, a journal that medical scientists rarely read.

http://www.nytimes.com/2011/07/08/health/research/08genes.html

2010 Video Presentation by Keith A. Baggerly
http://videolectures.net/cancerbioinformatics2010_baggerly_irrh/

The Excel-Error Heard Around the World


https://newrepublic.com/article/112951/rogoff-reinhart-and-world-excel-error-research

melindahiggins2000 Google Online MBSR/Mindfu Druid Hills High Scho Campus Parent Portal A Series of R Worksh U14 Boys

NEW REPUBLIC

The Weird and Very Real World of Excel-Error Research

The Rogoff-Reinhart blunder is a prominent example of a very common problem

BY **ROBERT LONG** | April 18, 2013

They're calling it the "Excel Error Heard Round the World": Kenneth Rogoff and Carmen Reinhart's widely cited paper about the relationship between public debt and economic growth was revealed Monday to have grossly misstated economic growth for high-debt countries, all because of a forehead-smackingly simple error in an Excel spreadsheet. ("It is sobering that such an error slipped into one of our papers despite our best efforts to be consistently careful," the paper's authors said on Wednesday.)

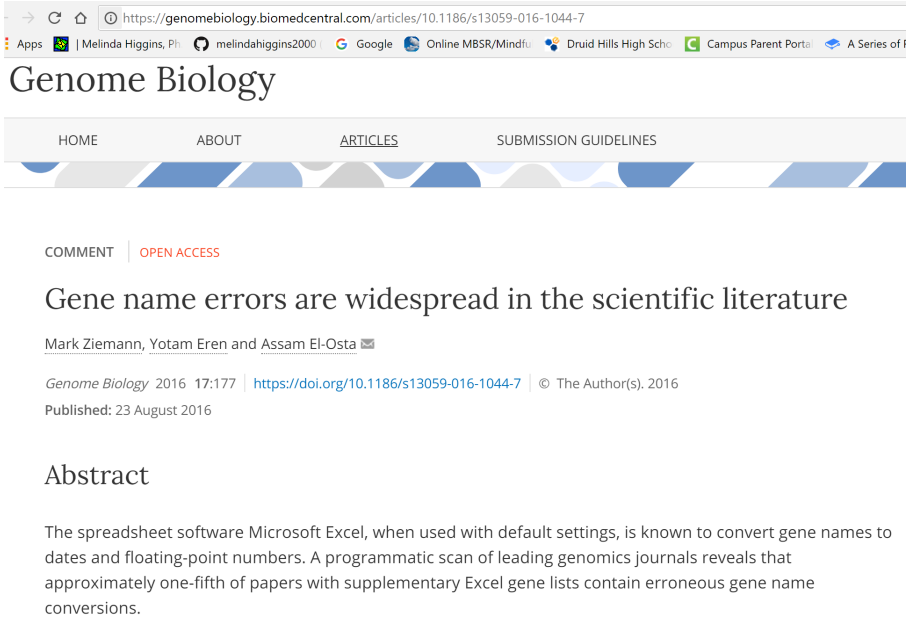
https://newrepublic.com/article/112951/rogoff-reinhart-and-world-excel-error-research

Timeline Reproducible Research & Transparency

YEAR	Event
2012	Begley and Ellis reviewed 53 "landmark" studies and only 6 (11%) had the scientific findings confirmed.[9]
2013	Center for Open Science launches & by 2014 the Open Science Framework has 7000 users with more than 45,000+ and over 15 institutions by 2017[10]
2014	NIH publishes their guidelines for addressing reproducibility[11]
2015	The Open Science Collaboration reports that they were only able to replicate between 1/3 to 1/2 of the results from 100 studies[12]

9. http://www.nature.com/nature/journal/v483/n7391/full/483531a.html
10. https://cos.io/about/brief-history-cos-2013-2017/ & https://osf.io/
11. https://www.nih.gov/research-training/rigor-reproducibility
12. Science, 28 Aug 2015: Vol. 349, Issue 6251, aac4716; DOI: 10.1126/science.aac4716; http://science.sciencemag.org/content/349/6251/aac4716

Wide-Spread Gene Name Errors



https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-1044-7

Genome Biology

HOME ABOUT ARTICLES SUBMISSION GUIDELINES

COMMENT | OPEN ACCESS

Gene name errors are widespread in the scientific literature

Mark Ziemann, Yotam Eren and Assam El-Osta

Genome Biology 2016 **17**:177 | https://doi.org/10.1186/s13059-016-1044-7 | © The Author(s). 2016

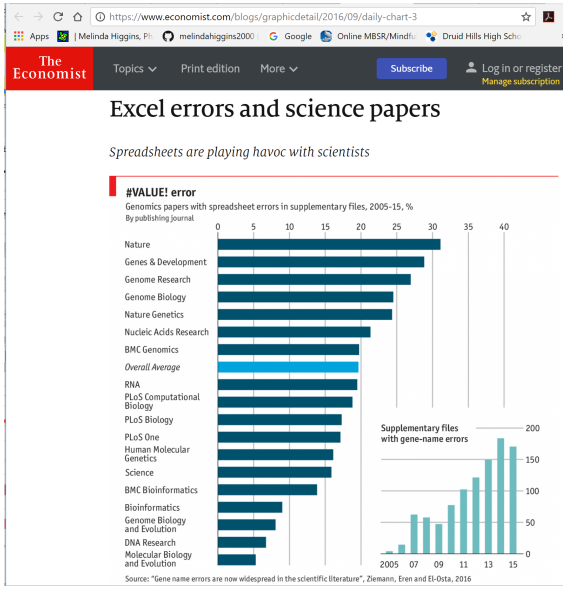
Published: 23 August 2016

Abstract

The spreadsheet software Microsoft Excel, when used with default settings, is known to convert gene names to dates and floating-point numbers. A programmatic scan of leading genomics journals reveals that approximately one-fifth of papers with supplementary Excel gene lists contain erroneous gene name conversions.

https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-1044-7

Wide-Spread Gene Name Errors



https://www.economist.com/blogs/graphicdetail/2016/09/daily-chart-3

People

- Victoria Stodden https://ischool.illinois.edu/people/faculty/vcs
  - presentation on History of the Reproducibility Movement https://web.stanford.edu/~vcs/talks/ICERM-Dec102012STODDEN.pdf
  - co-author "Implementing Reproducible Research" book https://www.crcpress.com/Implementing-Reproducible-Research/Stodden-Leisch-Peng/p/book/9781466561595
- Roger Peng http://www.biostat.jhsph.edu/~rpeng/index.html
  - Associate Editor for Reproducible Research - Biostatistics Journal https://academic.oup.com/biostatistics/pages/Editorial_Board
  - co-author "Implementing Reproducible Research" book https://www.crcpress.com/Implementing-Reproducible-Research/Stodden-Leisch-Peng/p/book/9781466561595

People

- John P.A. Ioannidis https://profiles.stanford.edu/john-ioannidis?tab=publications
  - Professor of Medicine and of Health Research and Policy at Stanford University School of Medicine and a Professor of Statistics at Stanford University School of Humanities and Sciences
- Christopher Gandrud https://www.iq.harvard.edu/people/christopher-gandrud
  - research fellow at IQSS (Institute for Quantitative Social Science)
  - Book Author "Reproducible Research with R and RStudio" https://www.crcpress.com/Reproducible-Research-with-R-and-R-Studio/Gandrud/p/book/9781466572843

People

- Yihui Xie
  - software engineer for RStudiohttps://www.rstudio.com/about/
  - author of "Dynamic Documents with R and knitr"https://www.crcpress.com/Dynamic-Documents-with-R-and-knitr/Xie/p/book/9781482203530
  - author of "Bookdown: Authoring Books and Technical Documents with R Markdown" book https://www.crcpress.com/bookdown-Authoring-Books-and-Technical-Documents-with-R-Markdown/Xie/p/book/9781138700109 and `bookdown` R package https://cran.r-project.org/web/packages/bookdown/index.html
  - author of `blogdown` R package https://cran.r-project.org/web/packages/blogdown/index.html

People

- Friedrich Leisch
  - Professor of Applied Statistics at the University of Natural Resources and Life Sciences, Vienna
  - developer of Sweave for creating dynamic reports https://leisch.userweb.mwn.de/Sweave/
  - co-author "Implementing Reproducible Research" book https://www.crcpress.com/Implementing-Reproducible-Research/Stodden-Leisch-Peng/p/book/9781466561595

Books on Reproducibility and Tools of the Trade

Image	Book
	Implementing Reproducible Research by Victoria Stodden, Friedrich Leisch, Roger D. Peng https://www.crcpress.com/Implementing-Reproducible-Research/Stodden-Leisch-Peng/p/book/9781466561595
	Dynamic Documents with R and knitr (Chapman & Hall/CRC The R Series) 1st Edition by Yihui Xie https://www.crcpress.com/Dynamic-Documents-with-R-and-knitr/Xie/p/book/9781482203530
	bookdown: Authoring Books and Technical Documents with R Markdown by Yihui Xie https://www.crcpress.com/bookdown-Authoring-Books-and-Technical-Documents-with-R-Markdown/Xie/p/book/9781138700109 & read online https://bookdown.org/yihui/bookdown/

more books

Image	Book
	Happy Git and GitHub for the useR by Jenny Bryan; read online http://happygitwithr.com/
	Using R and RStudio for Data Management, Statistical Analysis, and Graphics, Second Edition by Nicholas J. Horton & Ken Kleinman https://www.crcpress.com/Using-R-and-RStudio-for-Data-Management-Statistical-Analysis-and-Graphics/Horton-Kleinman/p/book/9781482237368; also see Project MOSAIC, http://mosaic-web.org/
	ModernDive: An Introduction to Statistical and Data Sciences via R by Chester Ismay and Albert Y. Kim; read online https://ismayc.github.io/moderndiver-book/ & Getting used to R, RStudio, and R Markdown by Chester Ismay https://ismayc.github.io/rbasics-book/
.	… and lots more … see https://bookdown.org/

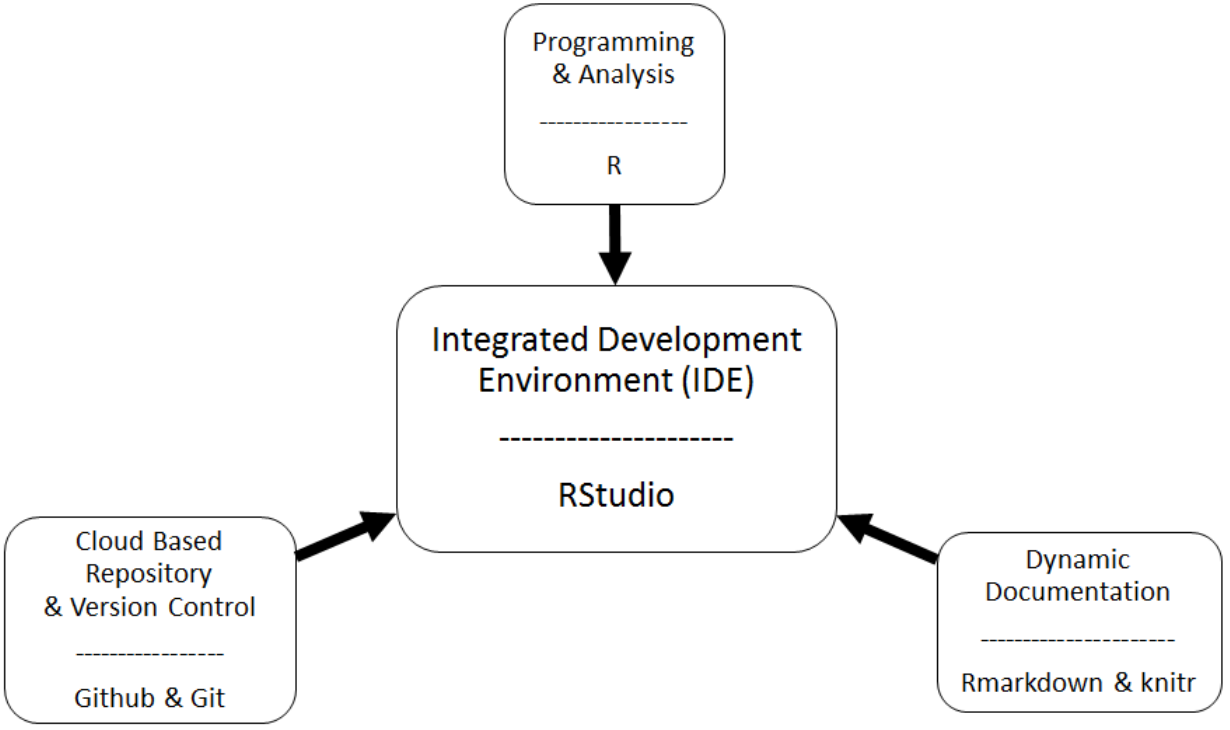
Literate Programming > Dynamic Documentation > [R]Markdown

YEAR	Event
1992	"Literate Programming" is introduced by Donald Knuth as "that (which) combines a programming language with a documentation language, thereby making programs more robust, more portable, more easily maintained, and arguably more fun to write than programs that are written only in a high-level language. **The main idea is to treat a program as a piece of literature, addressed to human beings rather than to a computer.**" http://www-cs-faculty.stanford.edu/~knuth/lp.html
2002	Friedrich Leisch introduces `SWEAVE` a program for "Dynamic generation of statistical reports using literate data analysis" https://leisch.userweb.mwn.de/Sweave/

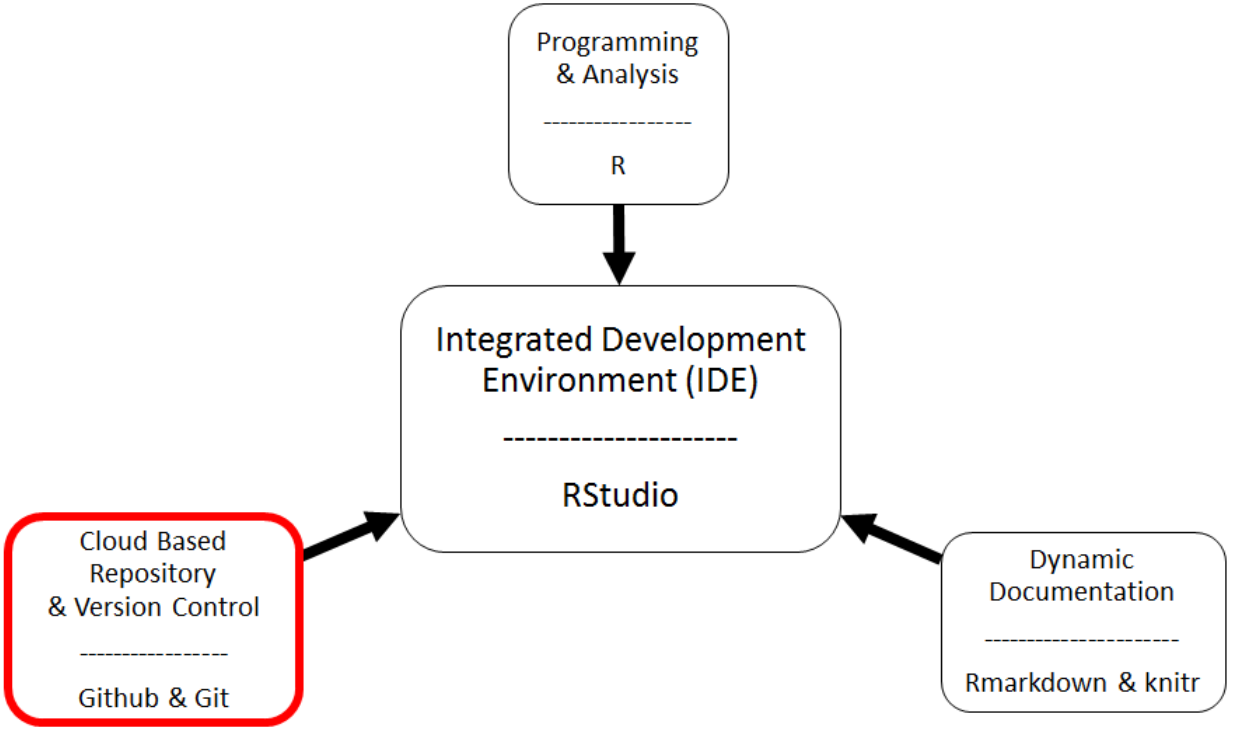
Literate Programming > Dynamic Documentation > [R]Markdown

YEAR	Event
2004	John Gruber created the `Markdown` language in 2004 in collaboration with Aaron Swartz - their goal was to "write using an easy-to-read, easy-to-write plain text format, and optionally convert it to structurally valid XHTML (or HTML)" https://daringfireball.net/projects/markdown/
2012	Yihui Xie releases `knitr` R package released - `knitr` was inspired by `SWEAVE`
2014	`rmarkdown` R package released - extends `Markdown` to work with R/RStudio environment

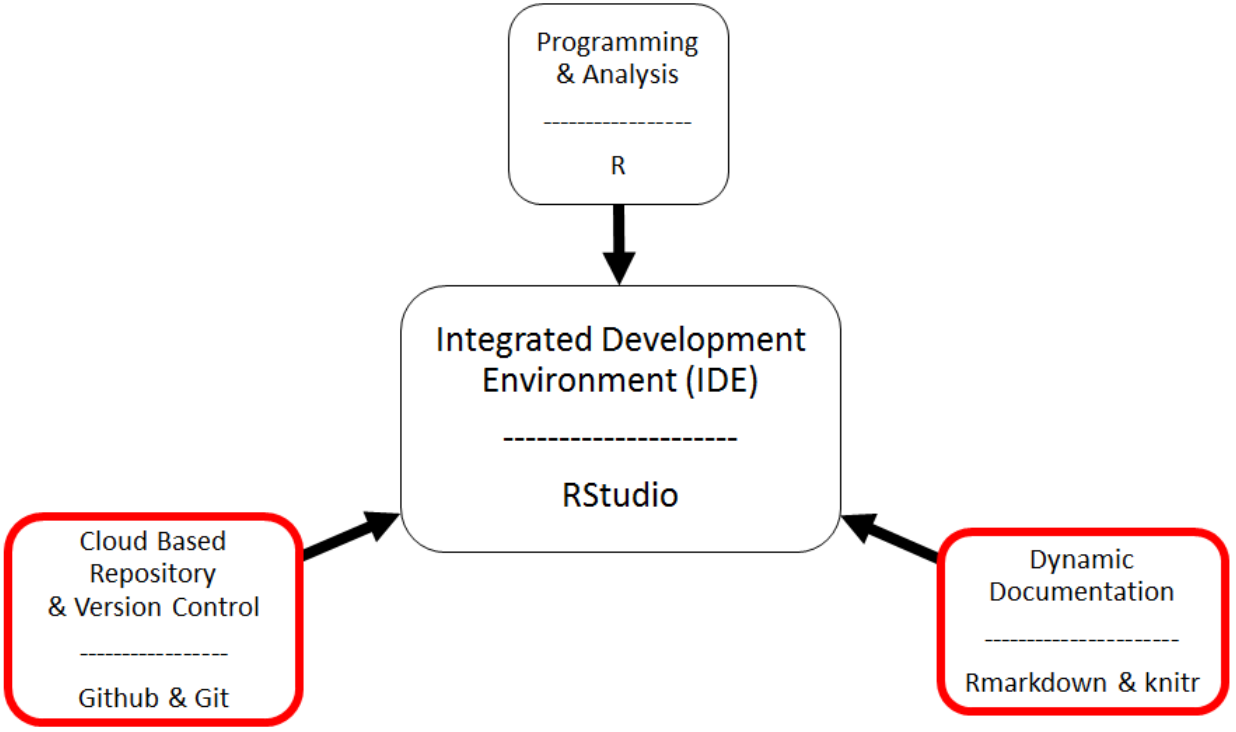
The Big Picture



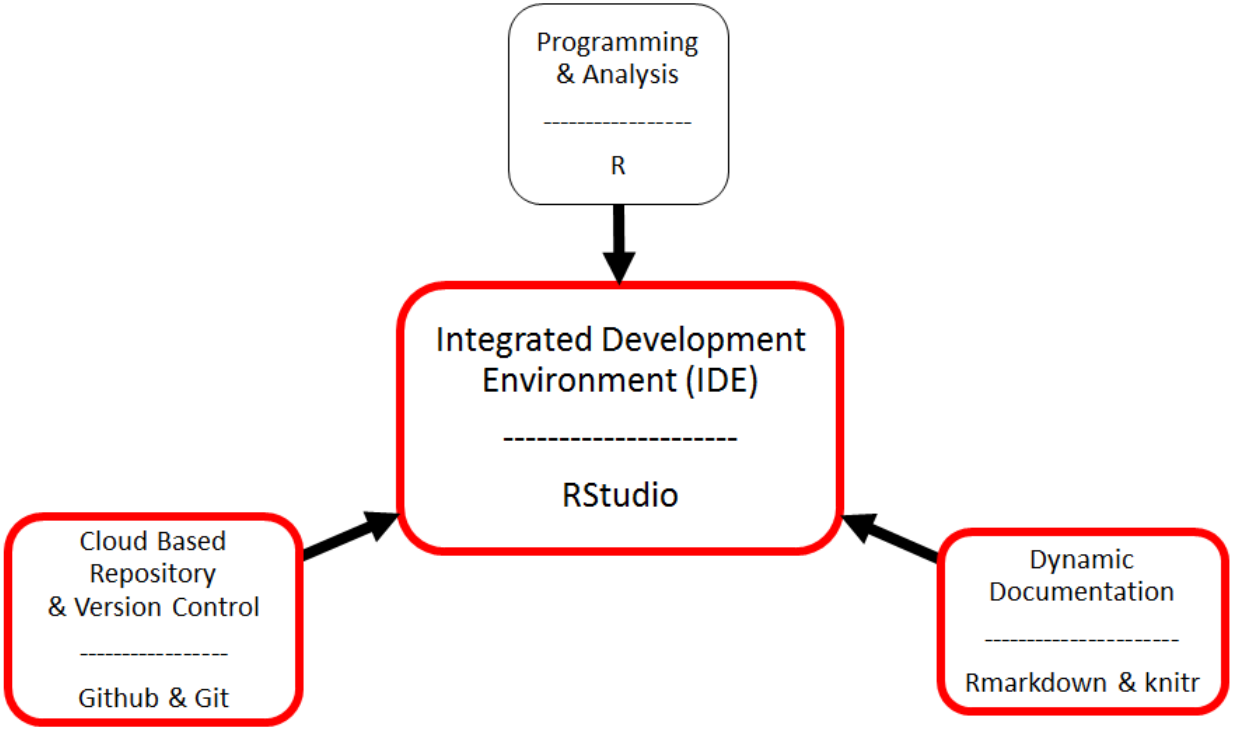
The Big Picture



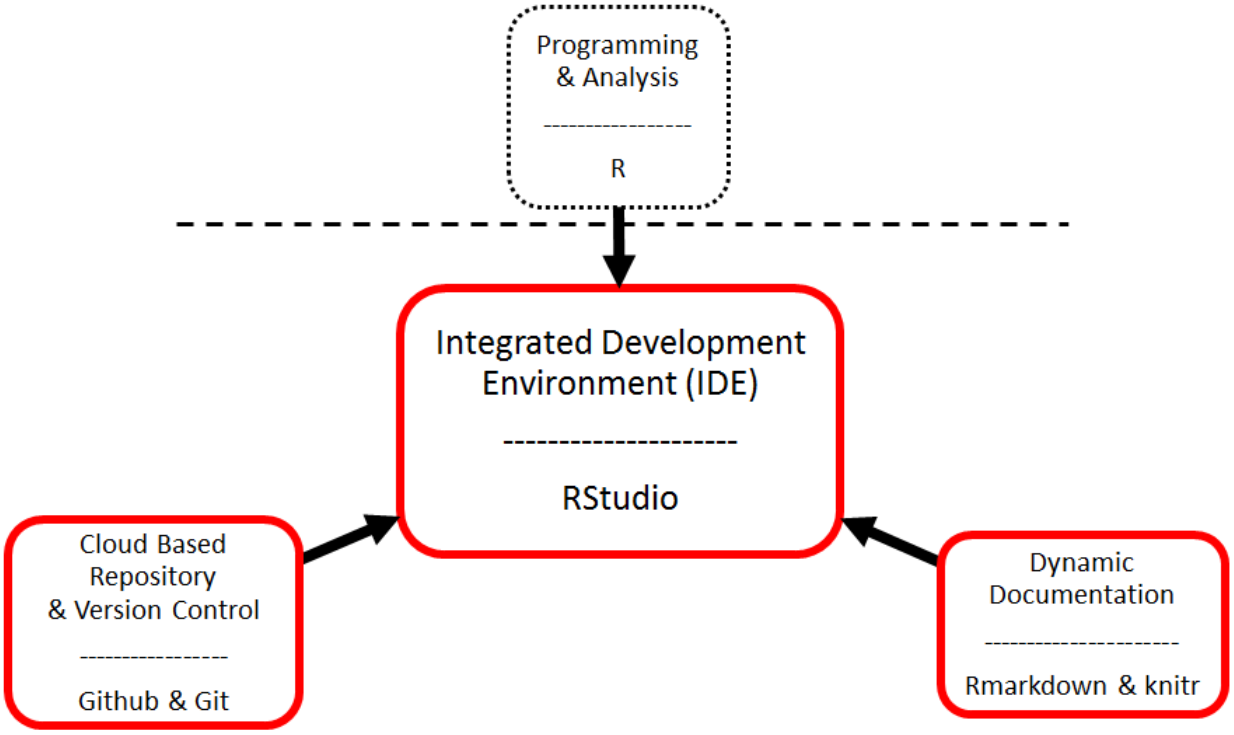
The Big Picture



The Big Picture



The Big Picture



Next in Lesson 02 …

Literate Programming

&

Dynamic Documentation