Teaching Ethics for Biostatisticians: *ACTS, August 2014*

Kathryn Chaloner

Professor and Head Department of Biostatistics,
University of Iowa

August 2014 JSM

History

- Institutional requirement for all students to take *Responsible Conduct of Research* course offered centrally by Graduate College (not just required for NIH trainees).
- Dissatisfaction with Graduate College course led to the development of a specific course for Biostatisticians (and Statisticians). Jeff Dawson and I developed it.
- Jeff took over and gained approval from the Graduate College that our course would substitute for theirs.

Outline: Some Topics Covered at The University of Iowa

- Is your analysis reproducible?
- Why is genomics data analysis hard to reproduce?
- Are Mendel's results too good?
- Is your clinical trial ethical?

References for Responsible Research

- Jasny, Chin, Chong, Vignieri (2011). Introduction to Special Section. Science, 334, p.1225.
- Peng (2011). Reproducible research in computational science. Science 334:1226.
- Robert Gentleman and Duncan Temple Lang, "Statistical Analyses and Reproducible Research" (May 2004). Bioconductor Project Working Papers. Working Paper 2.

  http://www.bepress.com/bioconductor/paper2

- Yihui Xie (1012). *Dynamic Documents with R and knitr.* CRC Press.

Good Statistical Practice

1. Save and date your data.
2. Save and date your code.
3. Save and date your seed.
4. Document your analyses.
5. Think literate programming and dynamic documents.
6. Avoid GUIs.
7. File documents carefully and systemmatically.

From Gentleman and Temple Lang, 2007

- "Statistical methodology generally involves algorithmic concepts .... Expressing these concepts in a purely textual format ... is seldom entirely satisfactory ...."
- Data sets and code can be posted on the web or put in an Appendix. Many of us try to do this or at least make sure we carefully store a copy of our code and our data that goes with any publication. (PhD Theses have no page limit). But this is really not enough. Users and readers have difficulty with
  - To what variable does x, y etc correspond?
  - I want to apply this method to my own data, but the code does not let me do so easily.
- The Compendium Concept (Gentleman and Temple Lang) in many ways is similar to the practice in the basic and translational sciences for providing supplemental material to document reagents, primers, and assays as well as the details of the statistical sciences.

Literate Programming

- Donald Knuth. "Literate Programming (1984)" in Literate Programming. CSLI, 1992, pg. 99.
- *I believe that the time is ripe for significantly better documentation of programs, and that we can best achieve this by considering programs to be works of literature. Hence, my title: "Literate Programming."*

From Knuth, 1984

1. *Let us change our traditional attitude to the construction of programs: Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do.*
2. *The practitioner of literate programming can be regarded as an essayist, whose main concern is with exposition and excellence of style. Such an author, with thesaurus in hand, chooses the names of variables carefully and explains what each variable means. He or she strives for a program that is comprehensible because its concepts have been introduced in an order that is best for human understanding, using a mixture of formal and informal methods that reinforce each other.*
3. At a minimum this argues for writing comments in your code that explain what you are telling the computer to do.

Tools for Generating Reproducible Research

- Sweave
- Friedrich Leisch. Sweave: Dynamic generation of statistical reports using literate data analysis. In Wolfgang Härdle and Bernd Rönz, editors, Compstat 2002 - Proceedings in Computational Statistics, pages 575-580. Physica Verlag, Heidelberg, 2002. ISBN 3-7908-1517-9.
- Friedrich Leisch and Anthony J. Rossini. Reproducible statistical research. Chance, 16(2):46-50, 2003.
- knitr
- Yihui Xie (1012). *Dynamic Documents with R and knitr.* CRC Press.
- Sweave and knitr allow us to avoid the error prone cut and paste of getting tables, calculations and other results of R calculations into documents. Sweave allows results of calculations to be directly put into latex documents. knitr exends Sweave and also allows interfaces with other kinds of documents.

Example of knitr and Latex from Xie's Book

```
\documentclass{article}
\begin{document}
\title{example.Rnw}
\author{Yihui Xie}
\maketitle
We examine the relationship between speed and stopping distance
 using a linear regression
 model:  $y=\beta_{0} + \beta_{1} x \epsilon$.
<<model, fig.width=4, fig.height=3, fig.align='center'>>=
par(mar=c(4,4,1,1),mgp=c(2,1,0), cex=0.8)
plot(cars, pch=20, col='darkgray')
fit <- lm(dist ~ speed, data=cars)
abline(fit, lwd =2)
@
The slope of a simple linear regression is
\Sexpr{coef(fit)[2]}.
\end{document}
```

In R

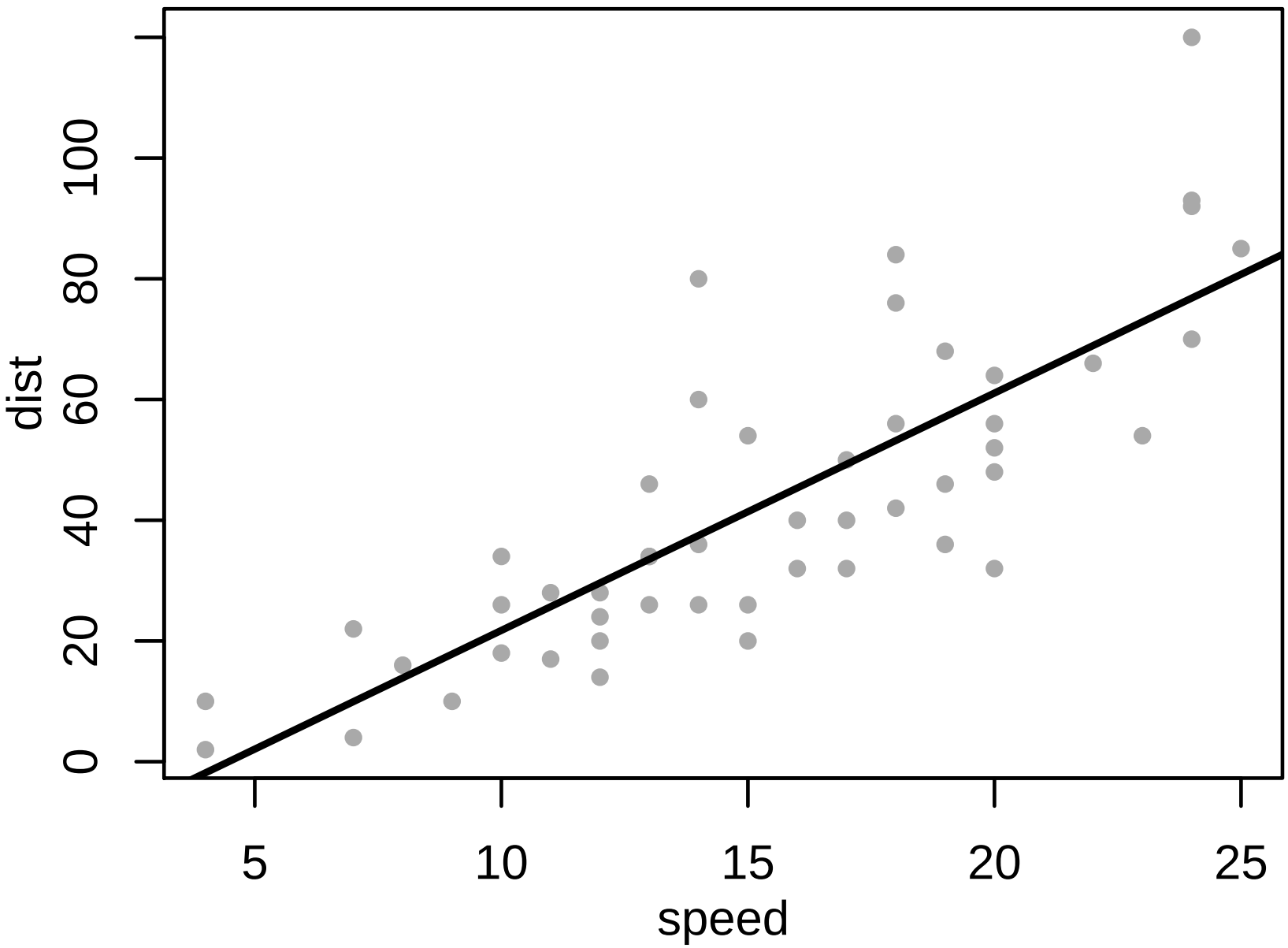
- Load and install package knitr
- `> knit(example.Rnw)`
- This command creates an example.tex in your working directory. It also creates a directory with

  `figure/model.pdf`

  a file with the plot in.
- put this through LaTex to produce

  `example.pdf`

model.pdf



example.pdf

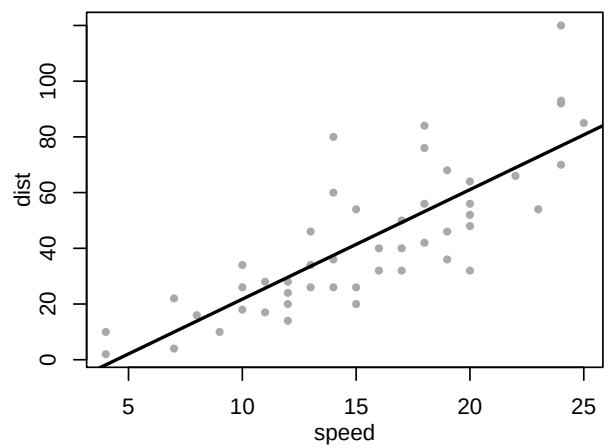
A Minimal Example

Yihui Xie

July 30, 2014

We examine the relationship between speed and stopping distance using a linear regression model: $y = \beta_0 + \beta_1 x \epsilon$.

```
par(mar = c(4, 4, 1, 1), mgp = c(2, 1, 0), cex = 0.8)
plot(cars, pch = 20, col = "darkgray")
fit <- lm(dist ~ speed, data = cars)
abline(fit, lwd = 2)
```



The slope of a simple linear regression is 3.9324

- If your data changes, for example if a single error is corrected, or if your data is updated for a DSMB interim report for a clinical trial, it is straightforward to update your analyses.
- Of course, if the discussion and interpretation of your data analysis changes, that must be updated!

Extreme Example of Research that was not Reproducible:

- Statisticians at MD Anderson were asked by clinical investigators to look at really interesting, high impact, 2006 publications by Duke researchers (Nevins and Potti) so that Baggerly and Coombs could analyse similar data from MD Anderson.
- The publications were on genetic markers for selecting specific therapies that would be very effective for particular cancer patients ("individualized treatments", or "targeted therapies").
- Big impact publication. Corresponding publicity.
- Clinical trials at Duke recruited subjects.
- Nevins (senior, professor) and Potti (junior, associate professor) had not made their data or R code available on the web so Baggerly and Coombs emailed them to request them.
- Baggerly and Coombs *eventually* got the data and tried to reproduce the result. The data was poorly documented.
- They identified multiple errors in the data analysis. The journal that published the original paper of Potti and Nevins rejected a paper with their concerns ($\simeq$2006). Most of their concerns were published in a statistical journal (2009, *Annals of Applied Statistics*).
- Baggerly and Coombs were very persistent.

Continued....

- http://www.nature.com/nm/journal/v13/n11/full/nm1107-1276b.html A short letter to the editor and a defence from the authors.
- To cut a long story short ...
- http://www.nature.com/nm/journal/v12/n11/pdf/nm1491.pdf
- Original 2006 Nature Medicine paper was retracted (in 2011, 5 years after publication!).
- Trials at Duke were suspended in 2009, restarted in 2010, suspended then terminated later in 2010. In *CBS 60 minutes* the senior investigator, Nevins, blamed the junior investigator, Potti. Lawsuits and malpractice claims resulted.
- Articles in *Economist, NY Times, Cancer Letter*.
- Multiple papers were retracted.
- Lawsuites are in progress against Duke, Potti, and Nevins on behalf of patients who were enrolled as subjects in trials designed using the results.

Questions for Discussion in Class

- Who supported the effort of Baggerly and Coombes?
- How can this kind of thing be prevented?
- Was it sloppy statistics or was it deliberate?
- Nevins in presentation

  > *I did not recognize that a critical flaw in the research effort was one of data corruption, an apparent manipulation of validation data*

- What blame should be attached to the senior investigator? Do you know of other cases where the senior author was treated much more generously than the junior author? (eg: 1. UMN, Professor lost tenure but kept clinical faculty position, administrator went to jail. 2. *The Patchwork Mouse*, JR Hixson, 1976, Anchor Press)

References: Reproducible Research in Genomics Data

- Data sets are large.
- Ioannidis JPA, Khoury MJ (2011). Improving Validation in "Omics" Research. *Science*, 334, p. 1230-3.
- Ioannidis et al (2009). Repeatability of published microarray gene expression analyses. *Nature Genetics* 41, p. 149-155.
- Bell at al (2009). *Nature Methods* 6, p.423-430.

- Ioannidis and colleagues in 2009, Nature Genetics report the results of an empirical test.
  - 18 articles on microarray-based gene-expresion profiling, published in *Nature Genetics* in 2005-6. One table or figure was independently evaluated by two independent teams of analysts.
  - Results – two analyses were reproduced, six were partially reproduced with some discrepancies, ten could not be reproduced.
  - The primary reason for not reproducing was unavailability of data. Discrepancies primarily due to incomplete data annotation or specification of data processing and analysis.

- *Nature Genetics* requires public data availablility, and compliance with MIAME.
- "MIAME describes the Minimum Information About a Microarray Experiment that is needed to enable the interpretation of the results of the experiment unambiguously and potentially to reproduce the experiment. [Brazma et al., 2001, *Nature Genetics*]

Conclusions of Ionnidis and Khury

- Proliferation of platforms and technology – which are more reproducible?
- Empirical replication (repeat the experiment) is increassingly necessary.
- Funding incentives, reproducibility rewards, and targeted repeatability checks, may enhance science.

From Bell et al, 2009, *Nature Methods*

- Do different labs measure the same thing? Can results be replicated.
- Consider liquid chromatography-mass spectrometry based proteomics.
  - test sample with 20 proteins to 27 laboratories.
  - only 7 labs reported all 20 proteins correctly.

Are Mendel's results too good?

- Fisher RA, Has Mendel's work been rediscovered? *Annals of Science*, 1936, I:115–137.
- Edwards AWF, 1986. Are Mendel's Results Really too Close? *Biol. Rev.* 1986, 61:295–312.
- Pires AM, Branco JA, A Statistical Model to Explain the Mendel-Fisher Controversy, *Statistical Science*, 2010:25:545–565.

Edwards Quotes Fisher's Talk

It is interesting that Mendel's original results all fall within the limits of probable error; if his experiments were repeated the odds against getting such good results is about 16 to one. It may have been just luck; or it may be that the worthy German abbot, in his ignorance of probable error, unconsciously placed doubtful plants on the side which favoured his hypothesis (Norton & Pearson, 1976; Bennett, 1983).

Quote in Edwards paper of Fisher's talk in 1911.

Statistical Science
2010, Vol. 25, No. 4, 545–565
DOI: 10.1214/10-STS342

A Statistical Model to Explain the Mendel–Fisher Controversy

Ana M. Pires and João A. Branco

Abstract. In 1866 Gregor Mendel published a seminal paper containing the foundations of modern genetics. In 1936 Ronald Fisher published a statistical analysis of Mendel's data concluding that "*the data of most, if not all, of the experiments have been falsified so as to agree closely with Mendel's expectations*." The accusation gave rise to a controversy which has reached the present time. There are reasonable grounds to assume that a certain unconscious bias was systematically introduced in Mendel's experimentation. Based on this assumption, a probability model that fits Mendel's data and does not offend Fisher's analysis is given. This reconciliation model may well be the end of the Mendel–Fisher controversy.

Is your clinical trial ethical?

- Clinical trials have a requirement for "equipoise' which is "uncertainty as to the outcome".
  1. theoretical equipoise (Zelen, Gehan and Friereich)
  2. clinical equipoise (Friedman)
- Why not elicit beliefs from clinicians to document equipoise?
- See Chaloner and Rhame (2001) Quantifying beliefs in clinical trials. *Statistics in Medicine* 2001; 20:581-600.
- Beliefs of 50+ clinicians elicited for two trials of HIV therapy.

Thank You!

Ioannidis *Science* 2011

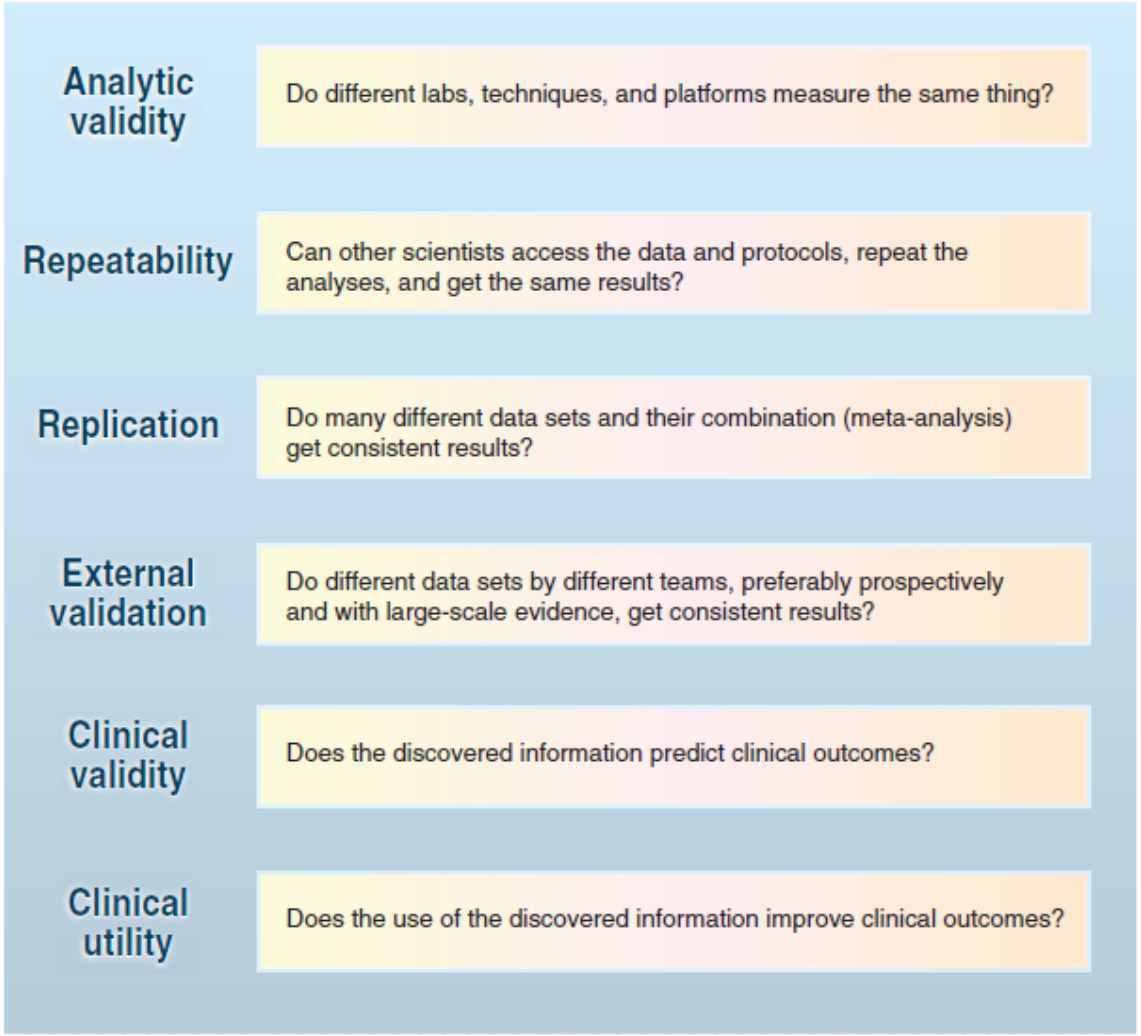


Fig. 1. The validation of omics research for use in medicine and public health requires fulfilling multiple steps. [Adapted from (7)]

Appendix: The Compendium Concept of Gentleman and Temple Lang

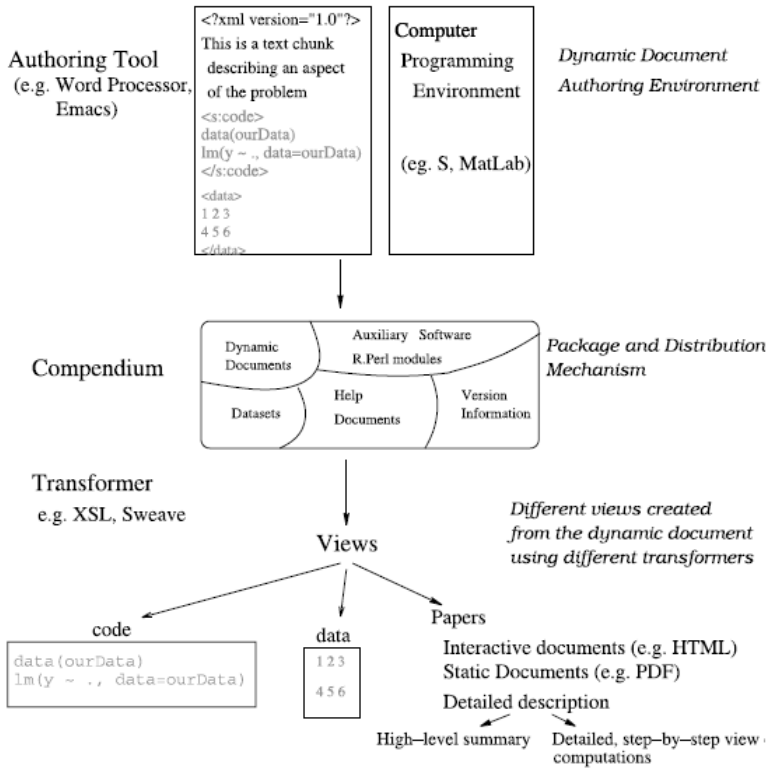


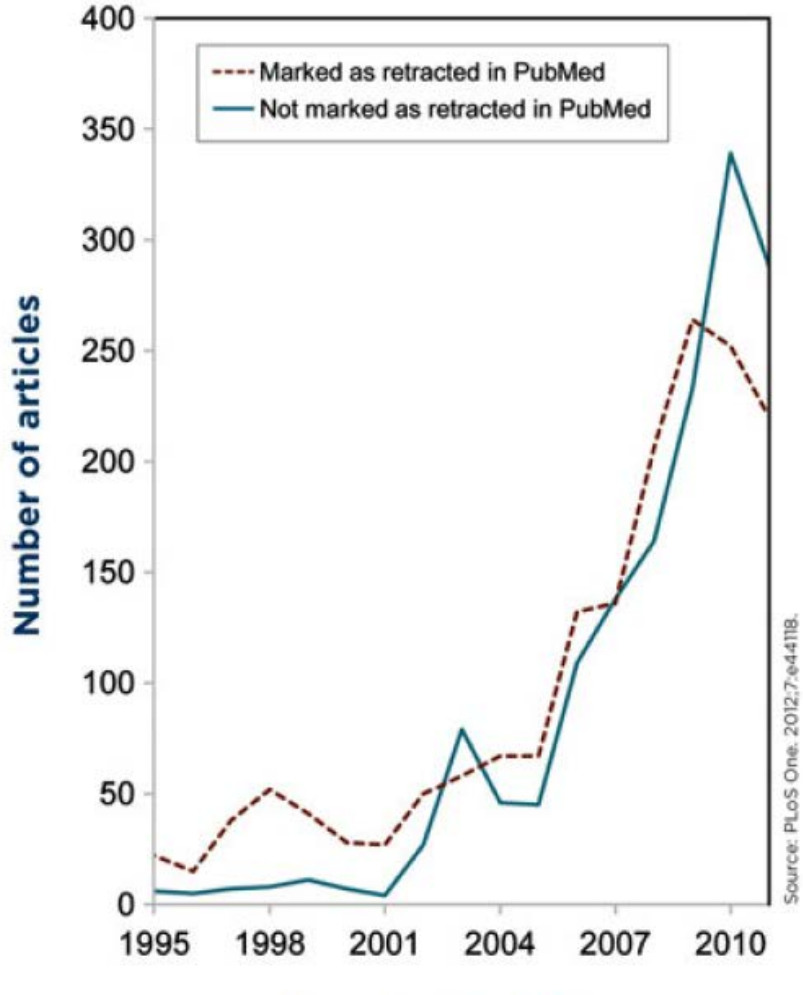
Figure 1. An overview of the components of the compendium concept. A dynamic document is a collection, or graph, of segments or nodes. Some of these segments are text, others are code, and others are (templates for) figures or tables or content that are generated when the computations are evaluated. A view is the result of a particular evaluation of all the code in the document, generating the specific results and content (i.e., figures, tables, and text). The result is that our original "document" is used to dynamically generate different views from the original "dynamic document" and the views can be static or interactive. A compendium acts as an archive or container for dynamic documents and associated datasets and software.

From: Gentleman and Temple Lang, 2007, JCGS.

Appendix: MIAME

- From www.mged.org
- The six most critical elements contributing towards MIAME are:
  1. The raw data for each hybridisation (e.g., CEL or GPR files)
  2. The final processed (normalised) data for the set of hybridisations in the experiment (study) (e.g., the gene expression data matrix used to draw the conclusions from the study)
  3. The essential sample annotation including experimental factors and their values (e.g., compound and dose in a dose response experiment)
  4. The experimental design including sample data relationships (e.g., which raw data file relates to which sample, which hybridisations are technical, which are biological replicates)
  5. Sufficient annotation of the array (e.g., gene identifiers, genomic coordinates, probe oligonucleotide sequences or reference commercial array catalog number)
  6. The essential laboratory and data processing protocols (e.g., what normalisation method has been used to obtain the final processed data)
- For more details, see MIAME 2.0.

Appendix: ML Grieneisen, M Zhang, 2012. PlosOne, 2012:7:e44118



- 42 databases examined.
- 4,449 scholarly publications retracted from 1928 to 2011.

  Most retracted articles do not contain flawed data; and the authors of most retracted articles have not been accused of research misconduct.