Carl Boettiger[a,*], Duncan Temple Lang[b], Peter C. Wainwright[a]

[a]*Center for Population Biology, University of California, Davis, California 95616*

[b]*Department of Statistics, University of California, Davis, California 95616*

knitcitations

An automated way to generate citations by dynamic lookup using Crossref DOIs or by bibtex files.

Markdown is becoming an increasingly popular platform for lightweight and online publishing. While traditional publishing tools like LaTeX and word processors have long had integrated bibliographic management, few tools handle citations for lightweight publishing. I am finding myself more and more drawn to markdown rather then tex/Rnw as my standard format (not least of which is the ease of displaying the files on github, particularly now that we have automatic image uploading. I've taken a little whack at generating in-text citations using knitr and other R tools.

Installation

I've put some simple functions in a `knitcitations` package. The functions use the crossref API to grab citation information given a doi, so I don't have to generate a bibtex file for papers I'm reading, (inspired by the kcite package for Wordpress). One can currently grab the package from Github

```r
library(devtools)
install_github("knitcitations", "cboettig")
```

Let's load the package into R and go for a spin.

```r
require(knitcitations)
```

---

*Corresponding author, cboettig@ucdavis.edu
Email address: `cboettig@ucdavis.edu` (Peter C. Wainwright)

20 *Generating citations*

21 *DOI Approach*

22 Then we can generate a citation given a DOI with the `ref` function:

```
r <- ref("10.1111/j.1461-0248.2005.00827.x")
print(r, "html")
```

23 Halpern BS, Regan HM, Possingham HP and Mccarthy MA (2006). "Accounting For Uncer-

24 tainty in Marine Reserve Design." Ecology Letters, 9. ISSN 1461-023X, http://dx.doi.org/10.1111/j.1461-

25 0248.2005.00827.x.

26 Note that R allows bibitems to print in a variety of formats, including html.

27 We can generate inline citations in the short name/date format with the `citet` function:

```
citet("10.1111/j.1461-0248.2005.00827.x")
```

28 [1] "Halpern *et. al.* (2006)"

29 Similarly we can generate parenthetical citations with the `citep` function,

```
citep(c("10.1111/j.1461-0248.2005.00827.x", "10.1890/11-0011.1"))
```

30 [1] "(Halpern *et. al.* 2006; Abrams *et. al.* 2012)"

31 Which can take a list of DOIs to cite parenthetically. The `citet` and `citep` functions are

32 automatically retrieving the available metadata via the Crossref API, and R is storing the

33 information to generate the final bibliography.

34 *Using and creating citation keys*

35 When specifying a DOI for a citation, we can also give the citation a simple key name so we can

36 use it later without having to remember the DOI, for instance, we can make the first citation of

37 a particular example as

```r
citep(c(Michaels = "10.1111/j.1755-263X.2012.00241.x"))
```

38 [1] "(Michaels & Tyre, 2012)"

39 and then later use

```r
citep("Michaels")
```

40 [1] "(Michaels & Tyre, 2012)"

41 If we do not pass a key for the DOI we create, knitcitations will automatically generate a key

42 of it's own using the last name of the first author and the year. For instance, based on one of

43 the DOI-citations we have already created, we can do

```r
citet("Halpern2006")
```

44 [1] "Halpern *et. al.* (2006)"

45 and knitcitations recognizes the key. The function will try to avoid collisions in the key: if it is

46 given or creates a key matching one that is already in use, it will see if the titles of the articles

47 match. If the are the same, the same key is used to avoid a duplicate entry. This makes it safe

48 to call

```r
citet(c(Halpern2006 = "10.1111/j.1461-0248.2005.00827.x"))
```

49 [1] "Halpern *et. al.* (2006)"

50 even if we have earlier or later cited by the doi alone. Collision checking also avoids duplicating

51 keys that correspond to different papers. If the titles are unique, knitcitations appends an

52 underscore at the end of the automatically generated key. For instance, here we call a DOI that

53 corresponds to a different citation with the same first author and year:

```r
citet("10.1111/j.1523-1739.2005.00258.x")
```

54 Warning message: Automatic key generation found a copy of this key, using Halpern2006_ ins

55 [1] "Halpern *et. al.* (2006)"

56 A warning (not printed by knitr when this is used inline, but included in the log file) alerts us

57 to the fact that this citation has been given an alternate key,

```
citet("Halpern2006_")
```

58 [1] "Halpern *et. al.* (2006)"

59 Of course if managing different keys sounds annoying, we can just call the DOI directly each

60 time.

61 *Bibtex Approach*

62 If we have a bibtex file, we can use this for the citations as well. If you don't have a bibtex file

63 handy, you can make one containing the citation information for some of R's packages:

```
write.bibtex(c("bibtex", "knitr", "knitcitations"), file = "example.bib")
```

64 We could also have used the `ref` function write a bibtex file for our list of DOIs

```
refs <- lapply(c("10.1111/j.1461-0248.2005.00827.x", "10.1890/11-0011.1"),
    ref)
write.bibtex(refs, file = "refs.bib")
```

65 Once we have a bibtex file available, we must read it into R.

```
bib <- read.bibtex("example.bib")
```

66 We can now create citations from `bib` using the bibtex key,

```
bib[["knitr"]]
```

67 Xie Y (2012). *knitr: A general-purpose package for dynamic report generation in R*. R package
68 version 0.5.4, .

69 The inline citation tools can also now use this `bib` instead of a DOI to generate a citation,

```
citet(bib[["knitr"]])
```

70 [1] "Xie, (2012)"

```
citep(bib[c("knitr", "bibtex")])
```

71 [1] "(Xie, 2012; Francois, 2011)"

72 Like the case of the DOI, after we have used a citation once, we can cite by the bibkey name
73 directly, rather than having to reference the bibentry object (*e.g.* from the `bib` list):

```
citet("knitr")
```

74 [1] "Xie, (2012)"

75 *Using the inline citations*

76 The inline citation calls are designed to be used with knitr's inline code blocks. In markdown,
77 these are enclosed in ' r '. The output format will use the plain-text rendering rather than the
78 code markup. Thus we can use the line `citep("10.1111/j.1461-0248.2005.00827.x")` to
79 generate a parenthetical citation, (Halpern *et. al.* 2006). We can generate the in-text citations
80 with `citet`, such as Xie, (2012).

81 *Generating the final bibliography*

82 As we go along adding inline citations, R stores the list of citation info. Then at the end of
83 the document, use this command to print the bibliography generated by the use of our inline
84 citations.

```
print(bibliography(), "html")
```

Halpern BS, Regan HM, Possingham HP and Mccarthy MA (2006). "Accounting For Uncertainty in Marine Reserve Design." Ecology Letters, 9. ISSN 1461-023X, http://dx.doi.org/10.1111/j.1461-0248.2005.00827.x.

Abrams PA, Ruokolainen L, Shuter BJ and Mccann KS (2012). "Harvesting Creates Ecological Traps: Consequences of Invisible Mortality Risks in Predator–Prey Metacommunities." Ecology, 93. ISSN 0012-9658, http://dx.doi.org/10.1890/11-0011.1.

Michaels S and Tyre AJ (2012). "How Indeterminism Shapes Ecologists' Contributions to Managing Socio-Ecological Systems." Conservation Letters. http://dx.doi.org/10.1111/j.1755-263X.2012.00241.x.

Halpern BS, Pyke CR, Fox HE, Haney JC, Schlaepfer MA and Zaradic P (2006). "Gaps And Mismatches Between Global Conservation Priorities And Spending." Conservation Biology, 20. http://dx.doi.org/10.1111/j.1523-1739.2005.00258.x.

Xie Y (2012). knitr: A general-purpose package for dynamic report generation in R. R package version 0.5.4, http://yihui.name/knitr/.

Francois R (2011). bibtex: bibtex parser. R package version 0.3-0, http://CRAN.R-project.org/package=bibt

Reflections

We could have justed printed bibliography in plain text format using with `bibliography()`. Note that it contains only the citations created with the inline citation commands `citet` and `citep`, in the order cited. All of these citations are stored in a hidden options variable in R when the inline functions are called. To reset the citation list (*i.e.* empty the contents of "bibliography()") we can use the `cleanbib()` command, or set the option `bibliography(erase=TRUE)`. Typically we could hide the bibliography chunk using a inline knitr call or the chunk option "echo=FALSE".

I hope to add markup to format this a bit more nicely later. For instance, we want the links to appear as real links. Additionally, we may want to add markup around the citations, such as the reason for the citation into the link using the Citation Typing Ontology. Ideally I need a method to support different citation styles, even though it is silly in today's world that the

citation format is still a choice of the *publisher* and not a choice of the *reader*. This will probably require citeproc integration and a major upgrade. Please report any bugs, feature requests or citations on the Github issues tracker!

Comparing lightweight citation alternatives

Several citation alternatives are available for lightweight publishing outside of this option, each with its own advantages and limitations. John MacFarlane's Pandoc is probably the most widely used citation manager for markdown files, working with a bibtex source file but formatting the citation lists using citeproc. It has the advantage of a more concise citation syntax, consistent with the source-readable goals of markdown and citation formating. Of course it is a markdown extension and will not be read by other markdown interpreters. This would be less of an issue of Pandoc could run markdown -> markdown without garbling syntax of some other markdown interpreter, like Github-flavored markdown.

The only other tool I know of that provides dynamic citations by DOI look-up is Phil Lord's excellent Wordpress plugin, kcite. It now uses citeproc for formatting, automatically links the in-text citations to the bibliography, supports PubMed and ArXiv ids as well as DOIs and even web URLs (though not bibtex files). The major limitation for me is that it is limited to Wordpress with Wordpress specific shortcode. (Of course I introduce R-specific code here, but with the assumption of a knitr-based audience who probably uses github). A variety of other platform-specific plugins are available to convert bibtex files into citations for different blogging platforms, including Wordpress and Jekyll.

So why knitcitations? My goal is primarily to bring this functionality to knitr users who rely on the markdown format rather than the latex format and are interested in dynamic citations and web-based publishing. I hope it finds its niche.