

Writing CHI Proceedings Papers With R Markdown

BEN TROVATO* and G.K.M. TOBIN*, Institute for Clarity in Documentation,
LARS THØRVÄLD, The Thørväld Group, Iceland
VALERIE BÉRANGER, Inria Paris-Rocquencourt, France
APARNA PATEL, Rajiv Gandhi University, India
HUIFEN CHAN, Tsinghua University, China
CHARLES PALMER, Palmer Research Laboratories,
JOHN SMITH, The Thørväld Group,
JULIUS P. KUMQUAT, The Kumquat Consortium,



Fig. 1. Seattle Mariners at Spring Training, 2010.

CCS Concepts: • **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

Additional Key Words and Phrases: datasets, neural networks, gaze detection, text tagging

ACM Reference Format:

Ben Trovato, G.K.M. Tobin, Lars Thørväld, Valerie Béranger, Aparna Patel, Huifen Chan, Charles Palmer, John Smith, and Julius P. Kumquat. 2018. Writing CHI Proceedings Papers With R Markdown. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2021)*, May 8–13, 2021, Yokohama, Japan. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/1122445.1122456>

*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

1 INTRODUCTION

Using a tool like [R Markdown](#) to write scientific papers makes your work more transparent and reproducible. It also reduces the risk of errors, because you can dynamically insert tables, figures, and summary statistics directly from the data they are generated from instead of transferring results manually from statistical software to manuscript.

This example illustrates how to use the ACM Master LaTeX template with R Markdown to write papers for the [CHI conference](#), in the CHI proceedings format. The content in this example is adapted and adjusted from content in the `sample-sigchi.tex` template included with the ACM template, to illustrate how to create the same content through the R Markdown workflow as well as to showcase additional features enabled by R Markdown.

2 PAPER META DATA

Set meta data (copyright, authors, keywords, title, keywords, optional teaser figure, etc.) in the YAML header of the .Rmd file in which you write the manuscript. This is done in the form of key: value pairs, e.g. `title: Writing CHI Proceedings Papers With R Markdown`. When compiling to a PDF (in RStudio, just click the 'Knit' button), the information in the YAML header is plugged into the CHI Extended Abstracts LaTeX template. (If you were to take a look at this template file inside of the `rticles` package, you would see e.g. `\title[$short-title][$title$]` where stuff between dollar signs is interpreted as a variable to be searched for in the YAML header and plugged into the template when generating a PDF).

Note the sole exception for adding paper meta data: The CCS Concepts are messy to insert from the YAML header, so you should manually insert this into the `ccsxml.tex` file from which it will be included into your manuscript.

3 THE BODY OF THE PAPER

Typically, the body of a paper has a hierarchical structure, with numbered or unnumbered headings for sections, subsections, sub-subsections, and paragraphs. Whereas in LaTeX you use the command `\section` for main sections, in R Markdown you simply use `#`, as in `# The Body of The Paper`. For subsections, or sub-subsections, use additional hashes, as in `## This Become a Subsection`, and `#### This Becomes a Paragraph Heading`.¹

If you want some section to be unnumbered in the output, add `{-}` after the section name, as in `# Unnumbered Section{-}`.

Indicate the start of a new paragraph with a blank line in your input file; that is why this sentence forms a separate paragraph. This line, however, does not form a separate paragraph.

3.1 Type Changes and *Special* Characters

Make words or phrases *italicized* by surrounding them with a single `*`; **embolden** them by surrounding them with `**two**`. Typewriter-style (for instance, for computer code) you create by surrounding text with ``backticks``.²

3.2 Citations

Citations to articles [1, 2, 4], conference proceedings [3] or maybe books [5, 6] listed in the Bibliography section of your article will occur throughout the text of your article. To insert a reference in the R Markdown syntax, type `@` followed by the citation key. The key is a short reference uniquely identifying each entry in the .bib file for your article, in which your references are listed in [BibTeX](#) format.

¹By the way, this is how to insert footnotes.

²Another footnote here. Let's make this a rather long one to see how it looks.

For example, to cite the article “Deciding equivalences among conjunctive aggregate queries” from our .bib file, write `@Cohen07`. If you drop the `[]`'s, you get author names, as well as the citation: Cohen et al. [4]. See [this short guide for more](#).

4 DYNAMIC REPORTING

One of the most important benefits of writing in R Markdown (aside from being able to compile to other formats than PDF, such as HTML or even Microsoft Word), is the ability to insert results dynamically into your manuscript using code chunks or inline code. This means that you can do analyses **directly** in your manuscript or, probably better, read file(s) with data, summaries, or results directly into your manuscript and refer to them dynamically.

This is important for two (related) reasons: 1. You avoid initial manual transfer of results from statistical software to manuscript, which reduces the risk of error. 2. If at a later stage you update the analysis files, the results reported in your manuscript are automatically also updated - this again reduces the risk of mistakes, because you don't need to manually update figures and tables.

In R Markdown syntax, **code chunks** have the following form (cf. *R Markdown: The Definitive Guide*):

```
```{coding_language chunk-label, chunk_options}
your code goes here
```
```

Inline code has the form ``coding_language #code here``.

4.1 Setup chunks and figure descriptions

The first chunk in an R Markdown document is usually used to load packages and set default chunk options, for example like so (we normally add the chunk option `include=FALSE` to not include output from this chunk in the manuscript; here we just add `message=FALSE` to suppress the message that the tidyverse package has been loaded):

```
library(tidyverse)
knitr::opts_chunk$set(echo = FALSE,
  message = FALSE, warning = FALSE)
# these options will exclude code output,
# messages, or warnings in knitted manuscript
```

In addition, the ACM Master template adds the ability to provide descriptions of figures via the latex command `\Description{my description}`. To be able to add these descriptions easily, as well as an option to position chunks vertically, include this code in your initial setup chunk:

```
# create additional chunk options
hook_chunk = knit_hooks$get('chunk')
knit_hooks$set(chunk = function(x, options) {
  txt = hook_chunk(x, options)
  # add chunk option 'vspaceout' which positions
  # chunks vertically with \vspace
  if (!is.null(options$vspaceout)) {
```

Table 1. Frequency of Special Characters

| Non-English or Math | Frequency | Comments |
|---------------------|-------------|-------------------|
| Ø | 1 in 1,000 | For Swedish names |
| π | 1 in 5 | Common in math |
| \$ | 4 in 5 | Used in business |
| Ψ_1^2 | 1 in 40,000 | Unexplained usage |

```

latex_vspace <- paste0("\\1\\\\vspace\\{",
  options$vspaceout, "\\}")
txt <- sub('(\\\\begin[^\n]+\n)',
  latex_vspace, txt)
}
# add chunk option 'description' which adds
# \Description{...} to figures
if (!is.null(options$description)) {
  latex_include <- paste0("\\1\\\\Description\\{",
    options$description, "\\}")
  gsub('(\\\\includegraphics[^\n]+\n)',
    latex_include, txt)
} else {
  return(txt) # pass to default hook
}
})

```

You can then add descriptions to your figures by setting `description="my description"` as a chunk option to images and plots as you will see below.

4.2 Inline results

You might read in a made-up data set of goals scored by basketball players like so:

```
data <- read_csv("data/fakeBasketData.csv")
```

We can use inline code to dynamically report properties of this data set. For example, “there are a total of 270 observations of goals scored. The mean number of goals made by any player in a given game is: 17.2555556”.

4.3 Tables

For tables, you could use LaTeX syntax directly. This might be useful if your table itself contains LaTeX syntax, as in Table 1.

However, the power of writing in R Markdown is that you can read in data and automatically create corresponding LaTeX tables. The easiest way is probably to use the [kable function](#). For example, Table 2 shows the first 5 rows in our basket data set.

Table 2. The first 5 rows of some made-up basket data.

| Player | goals |
|-----------------|-------|
| Carmelo Anthony | 4 |
| Carmelo Anthony | 2 |
| Carmelo Anthony | 10 |
| Carmelo Anthony | 3 |
| Carmelo Anthony | 1 |

Table 3. Summary statistics of goals scored by top players in made-up basketball season.

| Player | Total goals scored |
|-----------------|--------------------|
| Blake Griffin | 406 |
| Brook Lopez | 776 |
| Carmelo Anthony | 166 |
| Damian Lillard | 808 |
| David Lee | 362 |
| David West | 492 |
| Demar Derozan | 972 |
| Deron Williams | 365 |
| Dwyane Wade | 312 |

Table 4. Bigger display of more summary statistics of goals scored by top players in made-up basketball season.

| Player | Total goals scored | Goals per game |
|-----------------|--------------------|----------------|
| Blake Griffin | 406 | 13.533333 |
| Brook Lopez | 776 | 25.866667 |
| Carmelo Anthony | 166 | 5.533333 |
| Damian Lillard | 808 | 26.933333 |
| David Lee | 362 | 12.066667 |
| David West | 492 | 16.400000 |
| Demar Derozan | 972 | 32.400000 |
| Deron Williams | 365 | 12.166667 |
| Dwyane Wade | 312 | 10.400000 |

You can reference Table 2 with `\@ref(tab:basket-data)`.

You can also do arbitrary transformations and analyses of the data before creating a table, as in Table 3.

When using a two-column format you can create a wider table, which takes up the whole width of the page's live area, by adding the parameter `table.env = 'table*'` to the `kable` function, like in Table 4. In the LaTeX output, this puts the table in a `\table*` environment.

4.4 Figures

4.4.1 Static figures. Figures are similarly included via code chunks. You can include arbitrary image files, as in Figure 2.



Fig. 2. Here's a little pretty fly.



Fig. 3. A sample black and white graphic that has been resized with the `out.height` and `out.width` chunk options.

If you don't give it a caption in the chunk options (with something like `fig.cap="My caption"`), the figure does not float:



You can resize the figures with the chunk options `out.height` and `out.width`, as in Figure 3. If you only care about LaTeX output, you can resize e.g. in inches or relative to the column width (`out.height = '1in'` or `out.height = '0.50\columnwidth'`), but if you want to get maximum value out of R Markdown and be able to output also to html formats, set it with a percentage (`out.height = '50%'` - when outputting to PDF via LaTeX, this will be translated into `out.height = '.5\linewidth'`, see the [bookdown reference](#)).

If you need to style text in a caption, or include references in the caption, you have two options (see [bookdown on 'text references'](#)):

1. set the caption with the chunk option `fig.cap` and use LaTeX rather than markdown syntax. As the figure caption is a string, you must escape the LaTeX syntax's `\` with another `\`. The caption for Figure 3 would then have been written like this: `fig.cap="A sample black and white graphic that has been resized with the \texttt{out.height} and \texttt{out.width} chunk options."`.
2. write the caption in the body text with the syntax `(ref:chunk_label)` My caption here. and then refer to it in the chunk options with `fig.cap='(ref:chunk_label)` as we did for the resized fly caption.

4.4.2 Dynamic figures. Again, the power of R Markdown is that you can include e.g. plots that are dynamically generated from the underlying data. For example, Figure 4 is a simple visualisation of the basket data.

As with tables, you may want a figure to span two columns. To do this, set the environment to `figure*` with the chunk option `fig.env = 'figure*'`. You can fiddle around with the size and aspect ratio of the generated plot with the chunk options `fig.height` and `fig.width`. If your image is very large, you may want to restrict its width with `out.width`.

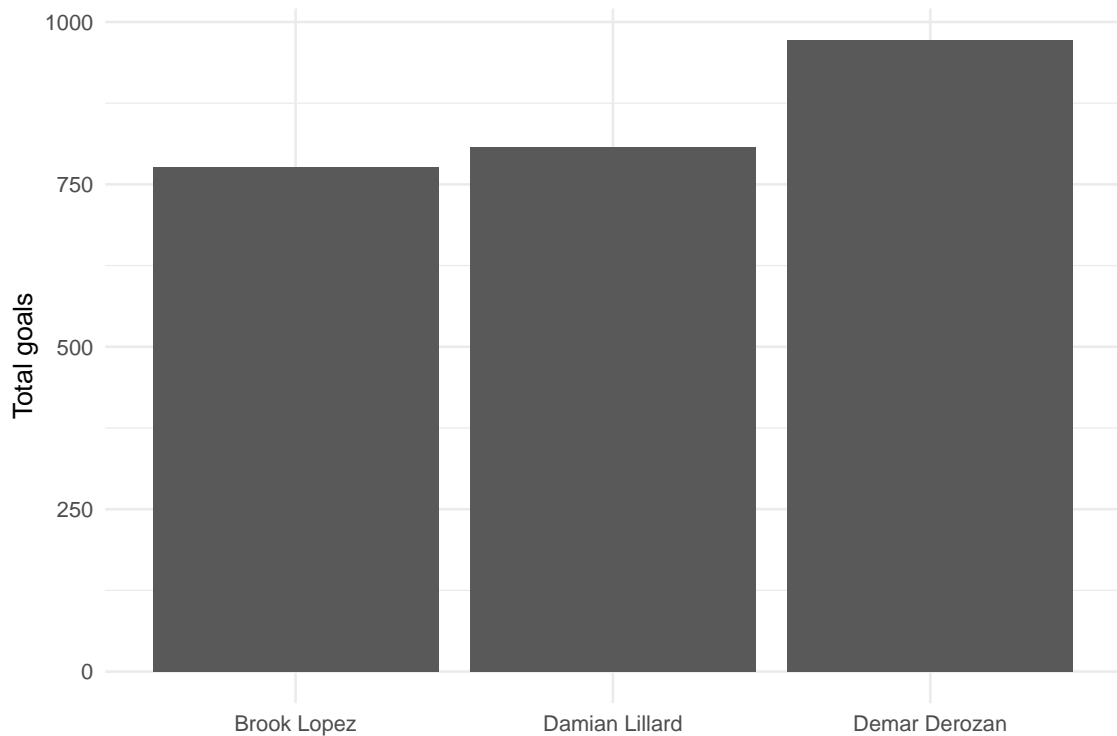


Fig. 4. Total number of goals by the top 3 players in made-up basketball season

4.5 Math Equations

You may want to display math equations in three distinct styles: inline, numbered or non-numbered display. Each of the three are discussed in the next sections. You can use usual LaTeX syntax directly, or [R Markdown](#).

4.5.1 Inline (In-text) Equations. A formula that appears in the running text is called an inline or in-text formula. In LaTeX it is produced by the `math` environment, which can be invoked by surrounding text with dollar signs: `$`. You can use any of the symbols and structures, from α to ω , available in LaTeX. For example, here's a nice equation inline: $\lim_{n \rightarrow \infty} x = 0$. If you're writing in RStudio, you can even hover over it to see the rendered output displayed!

4.5.2 Display Equations. A numbered display equation—one set off by vertical space from the text and centered horizontally—is produced by using LaTeX syntax directly to put the content in an equation environment³. So here's that nice equation from above:

$$\lim_{n \rightarrow \infty} x = 0 \quad (1)$$

They can be assigned labels with the syntax `(\#eq: label)`. Refer to the equation with `\@ref{eq:display-equation}`, e.g. see Equation (1).

To make an unnumbered display equation, surround the expression with two dollar signs:

$$\lim_{n \rightarrow \infty} x = 0$$

³In fact, you can use any arbitrary LaTeX syntax directly in your .Rmd document.

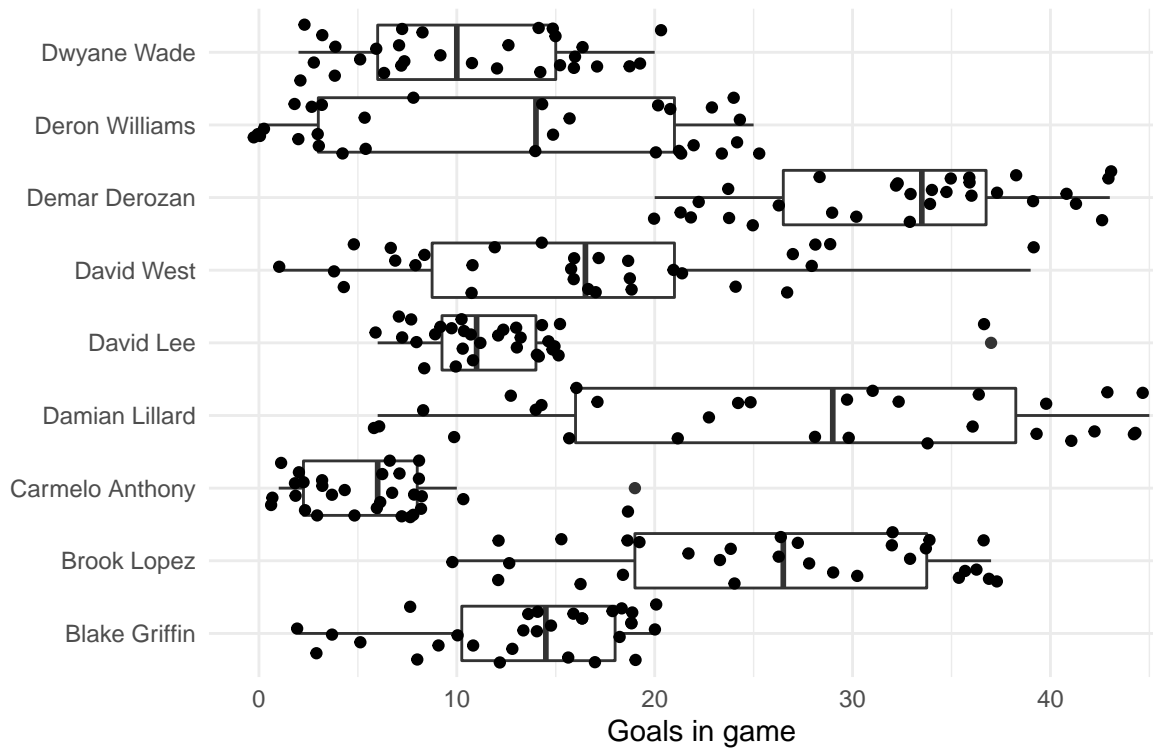


Fig. 5. Distribution of goals scored by game for players in made-up basketball season

4.6 Theorem-like Constructs

To create theorems, use this syntax (echo=TRUE is only necessary if you've set echo=FALSE as the default):

```
```{theorem, echo=TRUE}
Here is my theorem.
```
```

For example:

THEOREM 4.1. *Let f be continuous on $[a, b]$. If G is an antiderivative for f on $[a, b]$, then*

$$\int_a^b f(t) dt = G(b) - G(a).$$

Similarly for definitions, use the syntax

```
```{definition}
Here is my theorem.
```
```

For example:

Definition 4.1. If z is irrational, then by e^z we mean the unique number that has logarithm z :

$$\log e^z = z.$$

In the ACM LaTeX template, pre-defined theorem-like constructs are **theorem**, **conjecture**, **proposition**, **lemma** and **corollary**. The pre-defined definition-like constructs are **example** and **definition**.

Unsurprisingly, for proofs use

```
```{proof}
Here is my theorem.
```
```

For example:

PROOF. Suppose on the contrary there exists a real number L such that

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = L.$$

Then

$$l = \lim_{x \rightarrow c} f(x) = \lim_{x \rightarrow c} \left[gx \cdot \frac{f(x)}{g(x)} \right] = \lim_{x \rightarrow c} g(x) \cdot \lim_{x \rightarrow c} \frac{f(x)}{g(x)} = 0 \cdot L = 0,$$

which contradicts our assumption that $l \neq 0$. □

5 CONCLUSIONS

This paragraph ends the body of this sample document. Remember that you might still have Acknowledgments or Appendices; brief samples of these follow. There is still the Bibliography to deal with; and we will make a disclaimer about that here: with the exception of the reference to the LaTeX book, the citations in this paper are to articles which have nothing to do with the present subject and are used as examples only.

6 MORE HELP FOR THE HARDY

For acknowledgements, you may want to use the LaTeX syntax for this from the ACM template example, in which case you'll put acknowledgement text in between `\begin{acks}` and `\end{acks}`. Alternatively, just start an unnumbered heading `# Acknowledgements{-}` and write your text, like this:

ACKNOWLEDGEMENTS

There's a lot of people to thank here.

A HEADINGS IN APPENDICES

The rules about hierarchical headings discussed above for the body of the article are different in the appendices. You begin the **appendix** section with the special header `# (APPENDIX) Appendix {-}`. Then, any subsequent top level headers (`#`) indicates the start of each Appendix, with alphabetic order designation (i.e., the first is A, the second B, etc.). So, if you need hierarchical structure *within* an Appendix, start with **subsection** (`##`) as the highest level.

A.1 Research Methods

A.1.1 Part One. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi malesuada, quam in pulvinar varius, metus nunc fermentum urna, id sollicitudin purus odio sit amet enim. Aliquam ullamcorper eu ipsum vel mollis. Curabitur quis dictum nisl. Phasellus vel semper risus, et lacinia dolor. Integer ultricies commodo sem nec semper.

A.1.2 Part Two. Etiam commodo feugiat nisl pulvinar pellentesque. Etiam auctor sodales ligula, non varius nibh pulvinar semper. Suspendisse nec lectus non ipsum convallis congue hendrerit vitae sapien. Donec at laoreet eros. Vivamus non purus placerat, scelerisque diam eu, cursus ante. Etiam aliquam tortor auctor efficitur mattis.

A.2 Online Resources

Nam id fermentum dui. Suspendisse sagittis tortor a nulla mollis, in pulvinar ex pretium. Sed interdum orci quis metus euismod, et sagittis enim maximus. Vestibulum gravida massa ut felis suscipit congue. Quisque mattis elit a risus ultrices commodo venenatis eget dui. Etiam sagittis eleifend elementum.

Nam interdum magna at lectus dignissim, ac dignissim lorem rhoncus. Maecenas eu arcu ac neque placerat aliquam. Nunc pulvinar massa et mattis lacinia.

REFERENCES

- [1] Mic Bowman, Saumya K. Debray, and Larry L. Peterson. 1993. Reasoning About Naming Systems. *ACM Trans. Program. Lang. Syst.* 15, 5 (November 1993), 795–825. <https://doi.org/10.1145/161468.161471>
- [2] Johannes Braams. 1991. Babel, a Multilingual Style-Option System for Use with LaTeX's Standard Document Styles. *TUGboat* 12, 2 (June 1991), 291–301.
- [3] Malcolm Clark. 1991. Post Congress Tristesse. In *TeX90 Conference Proceedings*. TeX Users Group, 84–89.
- [4] Sarah Cohen, Werner Nutt, and Yehoshua Sagie. 2007. Deciding equivalences among conjunctive aggregate queries. *J. ACM* 54, 2, Article 5 (April 2007), 50 pages. <https://doi.org/10.1145/1219092.1219093>
- [5] Leslie Lamport. 1986. *TeX: A Document Preparation System*. Addison-Wesley, Reading, MA.
- [6] S.L. Salas and Einar Hille. 1978. *Calculus: One and Several Variable*. John Wiley and Sons, New York.