3.3 PDF document

To create a PDF document from R Markdown, you specify the `pdf_document` output format in the YAML metadata:

```
---
title: "Habits"
author: John Doe
date: March 22, 2005
output: pdf_document
---
```

Within R Markdown documents that generate PDF output, you can use raw LaTeX, and even define LaTeX macros. See Pandoc's documentation on the raw_tex extension for details.

Note that PDF output (including Beamer slides) requires an installation of LaTeX (see Chapter 1).

3.3.1 Table of contents

You can add a table of contents using the `toc` option and specify the depth of headers that it applies to using the `toc_depth` option. For example:

```
---
title: "Habits"
output:
  pdf_document:
    toc: true
    toc_depth: 2
---
```

If the TOC depth is not explicitly specified, it defaults to 2 (meaning that all level 1 and 2 headers will be included in the TOC), while it defaults to 3 in `html_document` .

You can add section numbering to headers using the `number_sections` option:

```
---
title: "Habits"
output:
  pdf_document:
    toc: true
    number_sections: true
---
```

If you are familiar with LaTeX, `number_sections: true` means `\section{}`, and `number_sections: false` means `\section*{}` for sections in LaTeX (it also applies to other levels of "sections" such as `\chapter{}`, and `\subsection{}`).

3.3.2 Figure options

There are a number of options that affect the output of figures within PDF documents:

- `fig_width` and `fig_height` can be used to control the default figure width and height (6.5x4.5 is used by default).
- `fig_crop` controls whether the `pdfcrop` utility, if available in your system, is automatically applied to PDF figures (this is `true` by default).
    - If you are using TinyTeX as your LaTeX distribution, we recommend that you run `tinytex::tlmgr_install("pdfcrop")` to install the LaTeX package `pdfcrop`. You also have to make sure the system package `ghostscript` is available in your system for `pdfcrop` to work. For macOS users who have installed Homebrew, `ghostscript` can be installed via `brew install ghostscript`.
    - If your graphics device is `postscript`, we recommend that you disable this feature (see more info in the **knitr** issue #1365).
- `fig_caption` controls whether figures are rendered with captions (this is `true` by default).
- `dev` controls the graphics device used to render figures (defaults to `pdf`).

For example:

```yaml
---
title: "Habits"
output:
  pdf_document:
    fig_width: 7
    fig_height: 6
    fig_caption: true
---
```

3.3.3 Data frame printing

You can enhance the default display of data frames via the `df_print` option. Valid values are presented in Table 3.3.

TABLE 3.3: The possible values of the `df_print` option for the `pdf_document` format.

Option	Description
default	Call the `print.data.frame` generic method
kable	Use the `knitr::kable()` function
tibble	Use the `tibble::print.tbl_df()` function
A custom function	Use the function to create the table. See 3.1.6.2

For example:

```yaml
---
title: "Habits"
output:
  pdf_document:
    df_print: kable
---
```

3.3.4 Syntax highlighting

The `highlight` option specifies the syntax highlighting style. Its usage in `pdf_document` is the same as `html_document` (Section 3.1.4). For example:

```
---
title: "Habits"
output:
  pdf_document:
    highlight: tango
---
```

3.3.5 LaTeX options

Many aspects of the LaTeX template used to create PDF documents can be customized using *top-level* YAML metadata (note that these options do not appear underneath the `output` section, but rather appear at the top level along with `title`, `author`, and so on). For example:

```
---
title: "Crop Analysis Q3 2013"
output: pdf_document
fontsize: 11pt
geometry: margin=1in
---
```

A few available metadata variables are displayed in Table 3.4 (consult the Pandoc manual for the full list):

TABLE 3.4: Available top-level YAML metadata variables for LaTeX output.

Variable	Description
lang	Document language code
fontsize	Font size (e.g., `10pt`, `11pt`, or `12pt`)
documentclass	LaTeX document class (e.g., `article`)
classoption	Options for documentclass (e.g., `oneside`)
geometry	Options for geometry class (e.g., `margin=1in`)
mainfont, sansfont, monofont, mathfont	Document fonts (works only with `xelatex` and `lualatex`)
linkcolor, urlcolor, citecolor	Color for internal, external, and citation links

3.3.6 LaTeX packages for citations

By default, citations are processed through `pandoc-citeproc`, which works for all output formats. For PDF output, sometimes it is better to use LaTeX packages to process citations, such as `natbib` or `biblatex`. To use one of these packages, just set the option `citation_package` to be `natbib` or `biblatex`, e.g.

```
---
output:
  pdf_document:
    citation_package: natbib
---
```

3.3.7 Advanced customization

3.3.7.1 LaTeX engine

By default, PDF documents are rendered using `pdflatex`. You can specify an alternate engine using the `latex_engine` option. Available engines are `pdflatex`, `xelatex`, and `lualatex`. For example:

```
---
title: "Habits"
output:
  pdf_document:
    latex_engine: xelatex
---
```

The main reasons you may want to use `xelatex` or `lualatex` are: (1) They support Unicode better; (2) It is easier to make use of system fonts. See some posts on Stack Overflow for more detailed explanations, e.g., https://tex.stackexchange.com/q/3393/9128 and https://tex.stackexchange.com/q/36/9128.

3.3.7.2 Keeping intermediate TeX

R Markdown documents are converted to PDF by first converting to a TeX file and then calling the LaTeX engine to convert to PDF. By default, this TeX file is removed, however if you want to keep it (e.g., for an article submission), you can specify the `keep_tex` option. For example:

```
---
title: "Habits"
output:
  pdf_document:
    keep_tex: true
---
```

3.3.7.3 Includes

You can do more advanced customization of PDF output by including additional LaTeX directives and/or content or by replacing the core Pandoc template entirely. To include content in the document header or before/after the document body, you use the `includes` option as follows:

```yaml
---
title: "Habits"
output:
  pdf_document:
    includes:
      in_header: preamble.tex
      before_body: doc-prefix.tex
      after_body: doc-suffix.tex
---
```

3.3.7.4 Custom templates

You can also replace the underlying Pandoc template using the `template` option:

```yaml
---
title: "Habits"
output:
  pdf_document:
    template: quarterly-report.tex
---
```

Consult the documentation on Pandoc templates for additional details on templates. You can also study the default LaTeX template as an example.

3.3.8 Other features

Similar to HTML documents, you can enable or disable certain Markdown extensions for generating PDF documents. See Section 3.1.10.4 for details. You can also pass more custom Pandoc arguments through the `pandoc_args` option (Section 3.1.10.5), and define shared options in `_output.yml` (Section 3.1.11).