

# Alternative to $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ for High Quality Reports

## HTML AND CSS FOR PAGINATED DOCUMENTS

uRos2018

Romain Lesur

Deputy Head of the Statistical Service



---

Retrouvez-nous sur  
[justice.gouv.fr](https://justice.gouv.fr)

# Benefits of R Markdown

A reproducible workflow	Documents are built from source.
Multiple output formats	Interactive document <code>.html</code>
	Static portable document <code>.pdf</code> with $\text{\LaTeX}$
	Ms Word <code>.docx</code>
	LibreOffice <code>.odt</code>
	...

# Use of R Markdown in an organization

Take an organization where documents have to comply with a corporate design...



...a Grid Model...

...template done with publishing softwares:

InDesign®

MS Publisher

Scribus...

# Statistical publications

## The usual workflow

Statisticians send `.docx` or `.odt` files & spreadsheets with data (for plots) to publishing assistants.

## A non-reproducible workflow!

## Solution

### Develop a PDF document template

Build a `ggplot2` theme

# Developping a PDF document template

With  $\text{\LaTeX}$ ?

$\text{\LaTeX}$  documents are perfect!

$\text{\LaTeX}$  has a painful learning curve.

$\text{\LaTeX}$  templates are hard to maintain.

Is  $\text{\LaTeX}$  legacy?

[LaTeX is Dead \(long live LaTeX\)](#) by Deyan Ginez

[In HTML and the Web I Trust](#) by Yihui Xie

# Another approach for publishing

## Evolution in the publishing industry

Major publishers had re-built their print-publishing toolchains using **HTML** and **CSS for Paged Media** .

see [Streamlining CSS Print Design with Sass](#) for O'Reilly Media and [Beyond XML: Making Books with HTML](#) for Hachette Book

## Main idea

Thanks to RMarkdown, R users can easily produce a HTML document.

Many R users learned HTML and CSS in order to customize their HTML documents or their Shiny apps.

**Make a PDF template with CSS rules**

# What is CSS for Paged Media?

## The CSS for Paged Media Standard

The *CSS for Paged Media* standard is a subset of the [W3C CSS specifications](#):

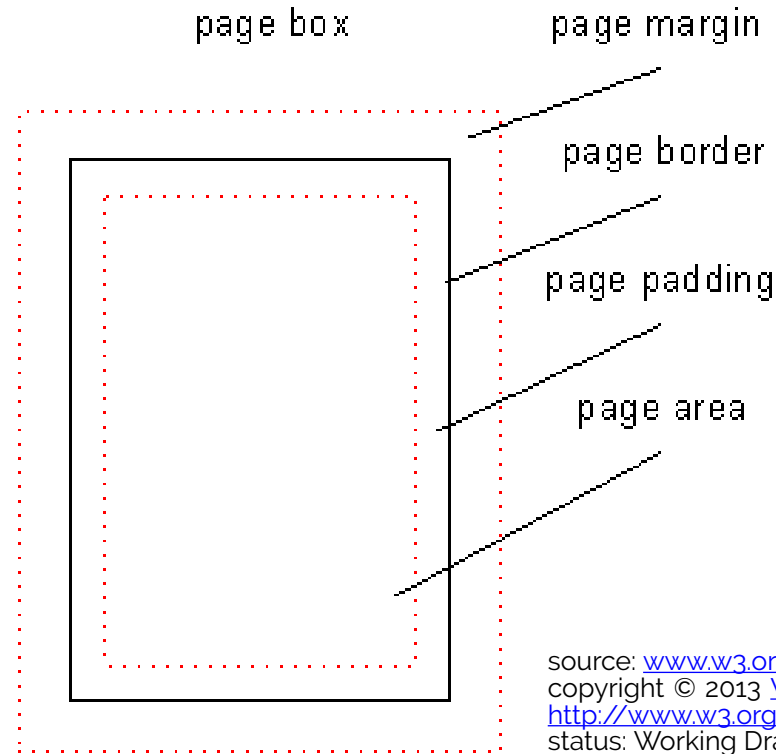
[CSS Paged Media Module Level 3](#)

[CSS Generated Content for Paged Media Module](#)

[CSS Page Floats](#)

[CSS Fragmentation Module Level 3](#)

# The page model



source: [www.w3.org/TR/css3-page](http://www.w3.org/TR/css3-page)  
copyright © 2013 [World Wide Web Consortium](http://www.w3.org/Consortium/), (MIT, ERCIM, Keio, Beihang).  
<http://www.w3.org/Consortium/Legal/2015/doc-license>  
status: Working Draft



# The margin boxes

top-left-corner	top-left	top	top-right	top-right-corner
left-top				right-top
left-middle				right-middle
left-bottom				right-bottom
bottom-left-corner	bottom-left	bottom	bottom-right	bottom-right-corner

# New CSS at-rules

Size, margins and page numbering

```
@page {  
    size: a4;  
    margin: 15mm;  
    @bottom-right-corner {  
        content: "p. " counter(page);  
    }  
}
```

## Pseudo classes

```
@page :right {  
  @bottom-right-corner {  
    content: counter(page) " of " counter(pages);  
  }  
}  
  
@page :left {  
  @bottom-left-corner {  
    content: counter(page) " of " counter(pages);  
  }  
}
```

# Control page breaks

Page break before a new section

```
h1.level1 {  
    page-break-before: always;  
}
```

Avoid page break inside tables

```
table {  
    page-break-inside: avoid;  
}
```

# How to generate a PDF with CSS for Paged Media?

## Browsers support

Top browsers in use do not support *CSS for Paged Media*.  
You cannot rely on (headless) browsers to generate a PDF file from HTML.

## Converters

Need to use a converter that supports the *CSS for Paged Media* standard  
Comprehensive list of converters: [print-css.rocks/tools.html](http://print-css.rocks/tools.html)  
This slideshow was made with Prince XML: [www.princexml.com](http://www.princexml.com)

**Note:** a promising project of the [Paged Media initiative](#) named [Paged.js](#)  
(polyfill for Paged Media)  
This could be a game changer.

# How to use CSS for Paged Media with R?

weasydoc

[github.com/RLesur/weasydoc](https://github.com/RLesur/weasydoc)

R package dedicated to CSS for Paged Media (WIP)

```
devtools::install_github('RLesur/weasydoc')
```

For now, weasydoc provides some basic RMarkdown output formats:

`hpdf_document()` and `hpdf_document_base()`

`hpdf_book()` and `hpdf_document2()` for bookdown

# System requirements

## pandoc

version 2.1.3 min., 2.2.1 recommended

[pandoc.org](https://pandoc.org)

Compatible pandoc version in the latest RStudio preview

## Supported converters

### WeasyPrint

[weasyprint.org](https://weasyprint.org) open source software (BSD-3 License)

### Prince XML

[Prince](https://prince.xml) is a commercial software.

Free license for non-commercial purposes

# Don't want to install anything?

## Docker image

[rlesur/weasydoc](https://rlesur.github.io/weasydoc/) built from rocker/verse:3.5.0 (not for production!)

```
docker run --rm -dp 8787:8787 rlesur/weasydoc:latest
```

## MyBinder environment

[frama.link/weasydoc](https://frama.link/weasydoc)

Access to RStudio with WeasyPrint and Prince available on system

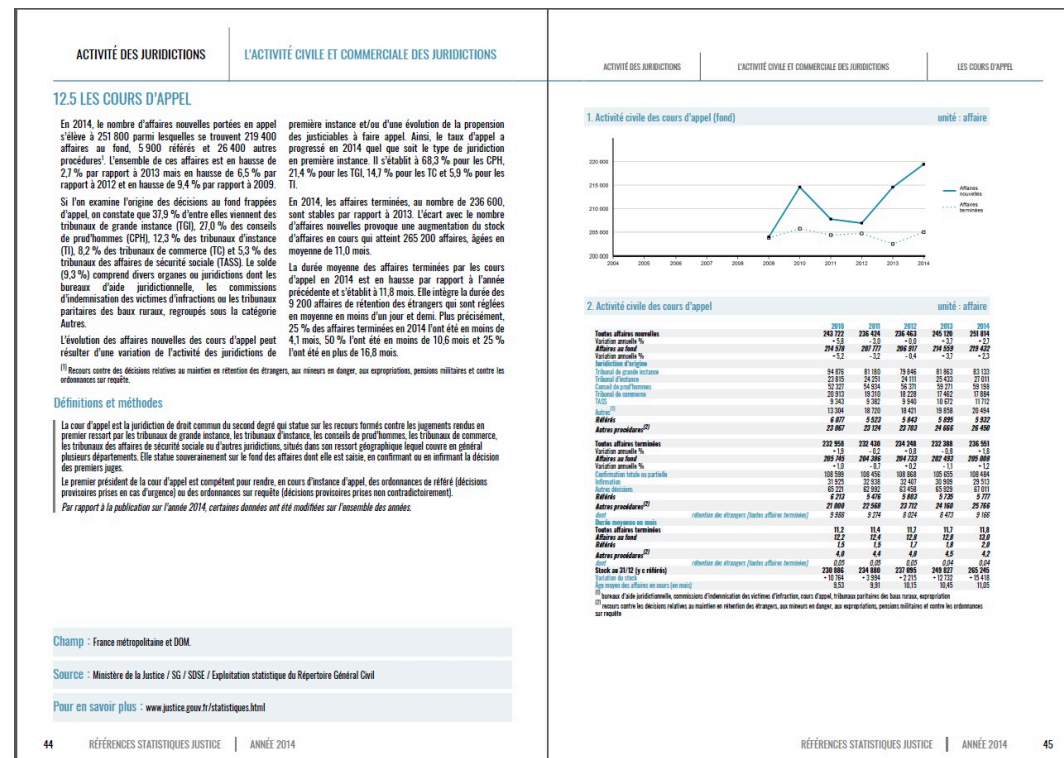


# What can be done with CSS for Paged Media?

Chapters with  
RMarkdown

Book with bookdown

ggplot2 theme



# Document quality

Use a professional  
converter if you need

**CMYK** colors (4 inks)

**Accessibility** (PDF/UA) not so easy with  $\LaTeX$ ...

**Graphics exchange** with publishing firms (PDF/X)

# Going further

There are a lot of great ressources on the web to learn *CSS for Paged Media*:

[A Guide To The State Of Print Stylesheets In 2018](#) by Rachel Andrew

[print-css.rocks](#) website by Andreas Jung

[Introduction to CSS for Paged Media](#) by Tony Graham, Antenna House - XML  
Prague 2018 Conference.

[Prince User Guide](#)

# Going further

## O'Reilly Media tutorials on Youtube:

[Part 1: Introduction to HTML and CSS](#)

[Part 2: Basic Layout and Text Formatting](#)

[Part 3: Paged Media Basics](#)

[Part 4: Generated Content - Counters & Strings](#)