

R マークダウンと Pandoc で楽々レポート 作成

@kohske

2014/3/1

目 次

1	はじめに	1
1.1	役に立つ資料	2
2	メタ情報の記述	2
2.1	簡易記法	2
2.2	YAML 記法	2
3	レポート生成コマンド (おなじない)	3
3.1	R 上で pandoc を使う	4
4	例: あやめの解析 (またかよ ... orz)	4
4.1	データの雰囲気	4
4.2	データの解析	5
4.3	データの可視化	5
5	最後に	5

1 はじめに

R マークダウンでドキュメントとコード書いて `knit()` `pandoc` (html
| pdf | docx) します。

1.1 役に立つ資料

- @teramonagi さんの資料
 - Tokyo.R@36 ~ knitr+pandoc ではじめる ~ 『R Markdown で Reproducible Research』 <http://www.slideshare.net/teramonagi/tokyo-r36-20140222>
 - Tokyo.R@36 ~ knitr パッケージではじめる ~ 『R Markdown で Reproducible Research』の基礎編のコード http://rpubs.com/teramonagi/TokyoR36_Basic
 - Tokyo.R@36 ~ knitr パッケージではじめる ~ 『R Markdown で Reproducible Research』の応用編のコード http://rpubs.com/teramonagi/TokyoR36_Advanced
- Pandoc ユーザーズガイド 日本語版 <http://sky-y.github.io/site-pandoc-jp/users-guide/>
- TeX Wiki <http://oku.edu.mie-u.ac.jp/~okumura/texwiki/>
- マークダウン用 github.css <https://gist.github.com/andyferra/2554919>

も参考にしてくださいね～

2 メタ情報の記述

マークダウンファイルにはメタ情報を含めることができます。

2.1 簡易記法

ファイル先頭を

1	% タイトル
2	% 著者
3	% 日付

で始めることができます。

2.2 YAML 記法

ファイルの先頭に YAML でメタ情報を入れることができます。次の例を参考にしてください。

```

1 ---
2 title: RマークダウンとPandocで楽々レポート作成
3 author: "@ohske"
4 tags: [R, pandoc, Dynamic Documentation]
5 abstract: Rマークダウンでドキュメントとコード書いて knit() pandoc (
        html | pdf | docx)します。
6 ---

```

3 レポート生成コマンド (おなじない)

まずは、

```

1 library(knitr)
2 knit("pandoc-md.Rmd")

```

として R マークダウンファイル (.Rmd) からマークダウンファイルを作成します。続いて、マークダウンファイルを Pandoc によって様々な形式に変換します。

- HTML ファイルの作成

```

1 $ pandoc -s --toc -c github.css --mathjax pandoc-md.md -o
    pandoc-md.html

```

github.css というファイルを同じフォルダに入れときます。

- LaTeX ファイルの作成

```

1 $ pandoc -s --toc --number-sections -V documentclass=ltjarticle
    pandoc-md.md -o pandoc-md.tex

```

- PDF ファイルの作成

```

1 $ pandoc -s --toc --number-sections --listings -V documentclass=
    ltjarticle --latex-engine=lualatex pandoc-md.md -o pandoc-md.
    pdf

```

- DOCX ファイルの作成

```

1 $ pandoc -s pandoc-md.md -o pandoc-md.docx

```

3.1 R上でpandocを使う

knitr パッケージには `pandoc()` という関数があるんですが、オプション渡すのが面倒なので `system()` で `pandoc` を実行します。

```
1 system("pandoc -s --toc -c_github.css --mathjax_pandoc-md.md -o_pandoc-md.html")
2 system("pandoc -s --toc --number-sections V documentclass=ltjarticle_pandoc-md.md -o_pandoc-md.tex")
3 system("pandoc -s --toc --number-sections --listings V documentclass=ltjarticle --latex-engine=lualatex_pandoc-md.md -o_pandoc-md.pdf")
4 system("pandoc -s_pandoc-md.md -o_pandoc-md.docx")
```

4 例：あやめの解析 (またかよ … orz)

あやめとは、さかな植物の名前です。おそらく、世界中でも最も多く解析にさらされた植物でしょう。

学名は *Iris sanguinea* といいます。イリスではなくて、アイリスです。

大きい声では言えませんが今でも「イリス」と呼んでます。

4.1 データの雰囲気

```
1 pandoc::pandoc(head(iris), caption="あやめのデータ (1-6行)", split.tables = 100)
```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa

表 1: あやめのデータ (1-6 行)

4.2 データの解析

```
1 cor(iris[, -5])

1 ## Sepal.Length Sepal.Width Petal.Length Petal.Width
2 ## Sepal.Length 1.0000 -0.1176 0.8718 0.8179
3 ## Sepal.Width -0.1176 1.0000 -0.4284 -0.3661
4 ## Petal.Length 0.8718 -0.4284 1.0000 0.9629
5 ## Petal.Width 0.8179 -0.3661 0.9629 1.0000
```

等幅フォントにできるかな

4.3 データの可視化

ヒストグラムを作って、正規分布 ($\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$) と比べてみます。

```
1 par(mar=c(2.5, 2.5, 1.5, 1))
2 hist(scale(iris[, 1]), probability = TRUE, ylim=c(0, 0.5))
3 curve(dnorm(x), add=TRUE)
```

5 最後に

Enjoy!!

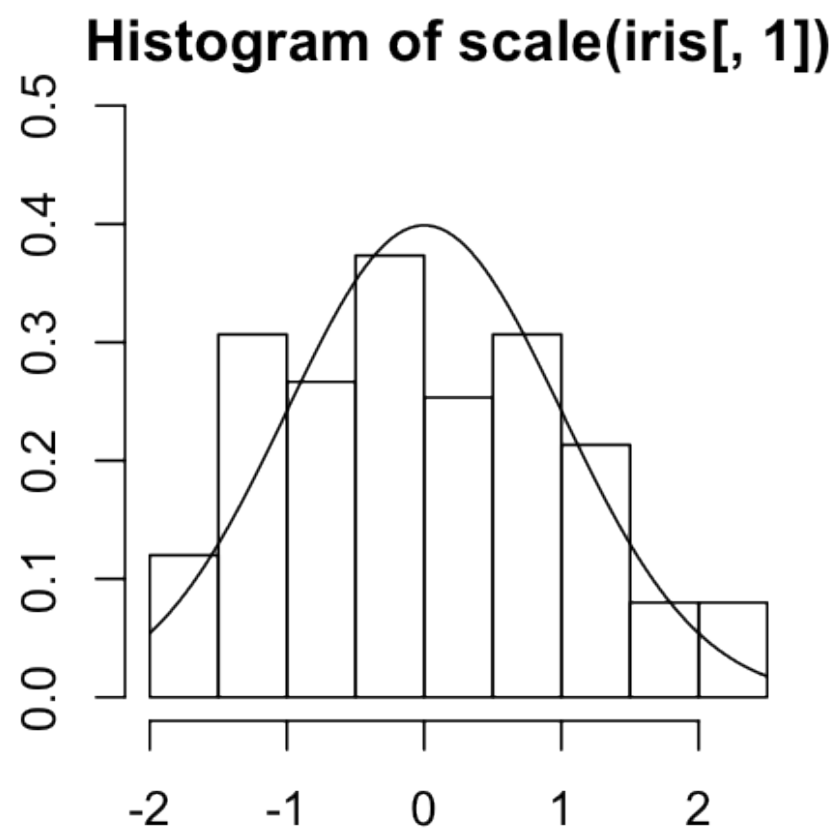


図 1: ヒストグラム