An Introduction to `knitr` and `RMarkdown`

https://github.com/sahirbhatnagar/knitr-tutorial

Sahir Bhatnagar

June 26, 2018

McGill Univeristy

Acknowledgements

- Toby, Matthieu, Vaughn, Ary
- Maxime Turgeon (Windows)
- Kevin McGregor (Mac)
- Greg Voisin
- Don Knuth (TEX)
- Friedrich Leisch (Sweave)
- Yihui Xie (knitr)
- John Gruber (Markdown)
- John MacFarlane (Pandoc)
- You

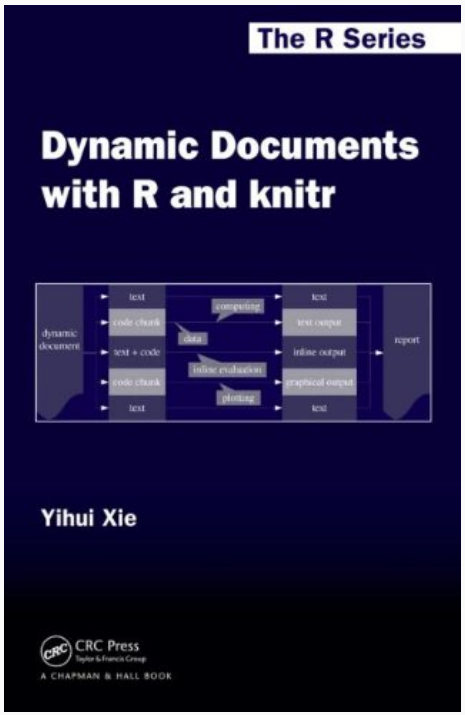
Disclaimer #1

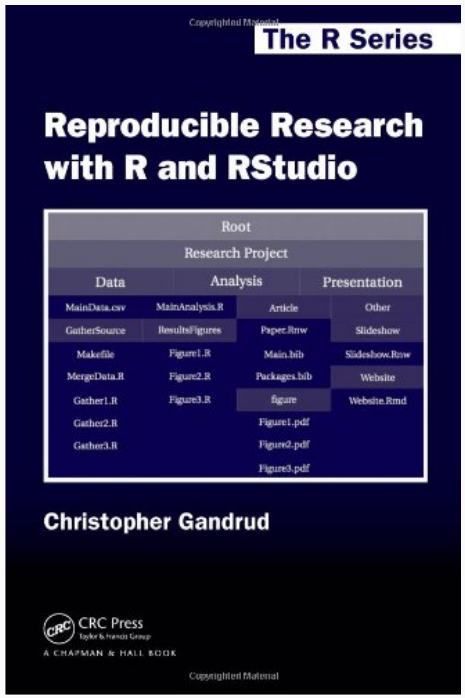
LaTeX

I don't work for, nor am I an author of any of these packages. I'm just a messenger.

Disclaimer #2

- Material for this tutorial comes from many sources. For a complete list see: https://github.com/sahirbhatnagar/knitr-tutorial
- Alot of the content in these slides are based on these two books





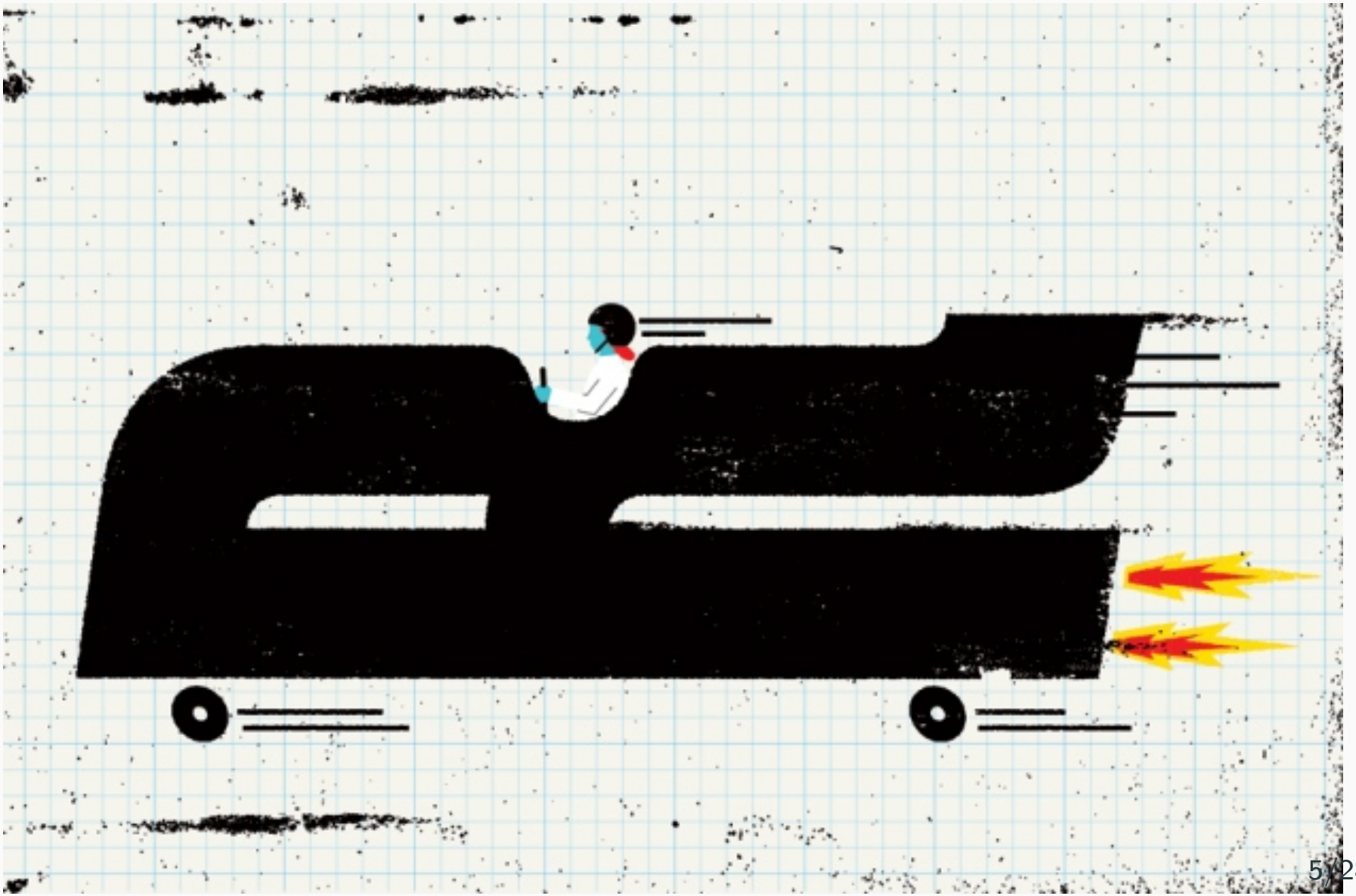
Objectives for today

- Create a reproducible document (`pdf`, `html`)

Objectives for today

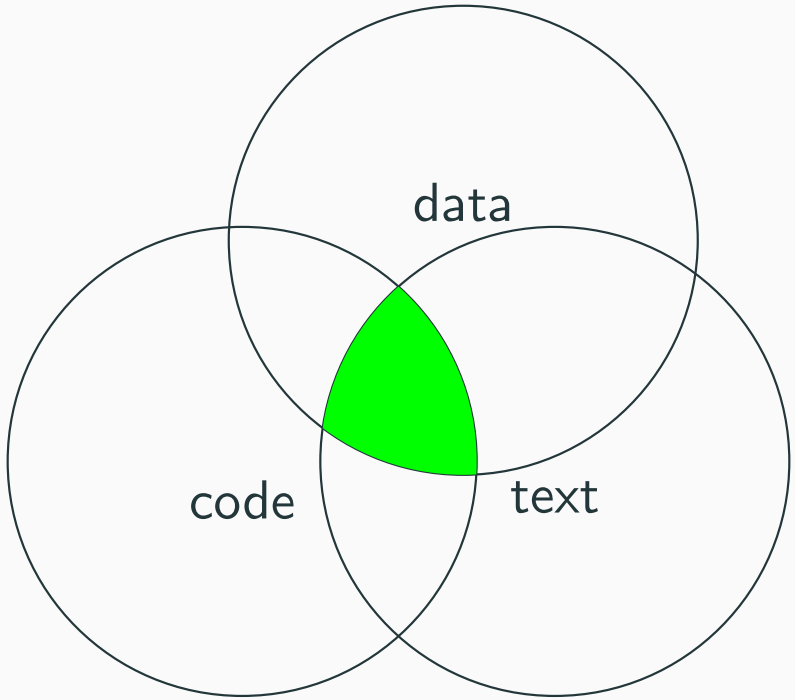
- Create a reproducible document (`pdf`, `html`)

C'est parti

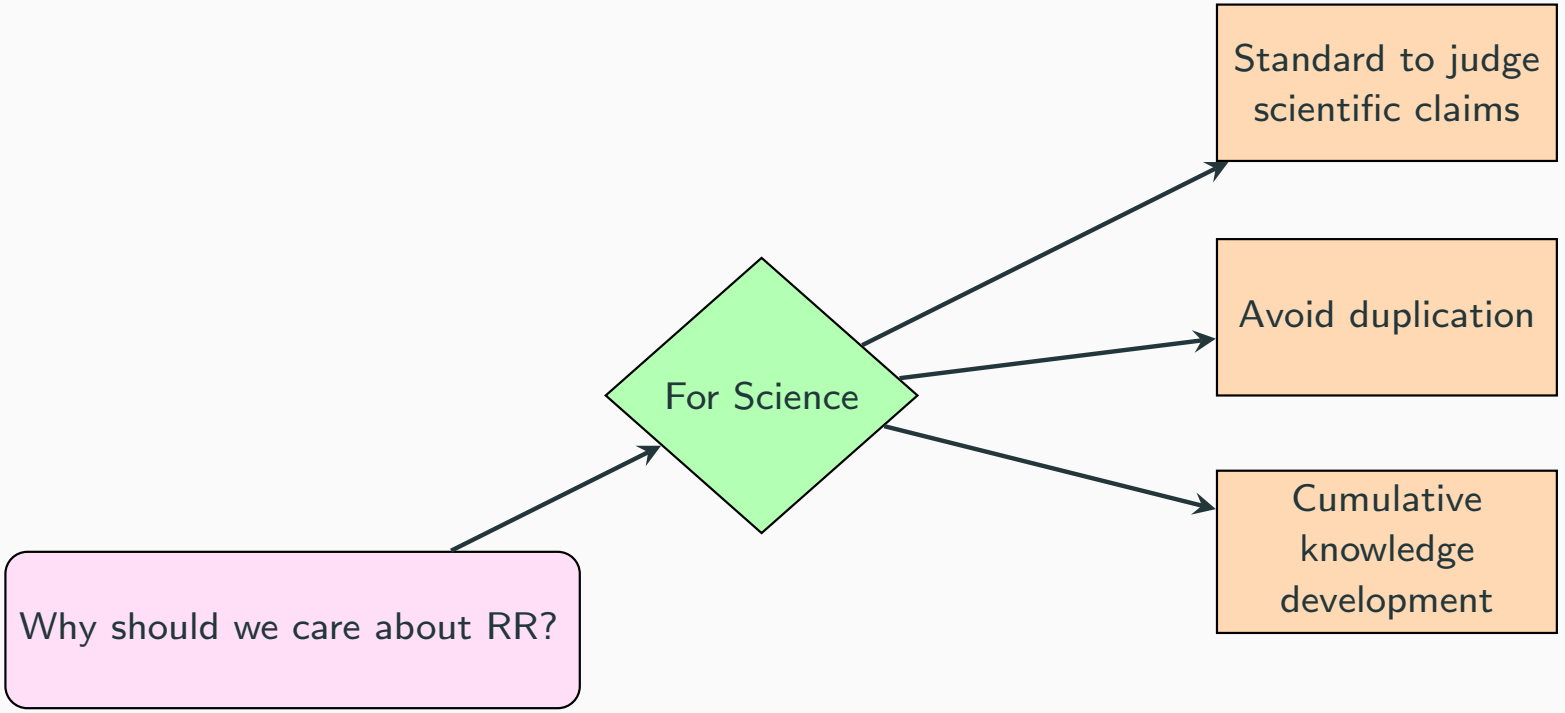


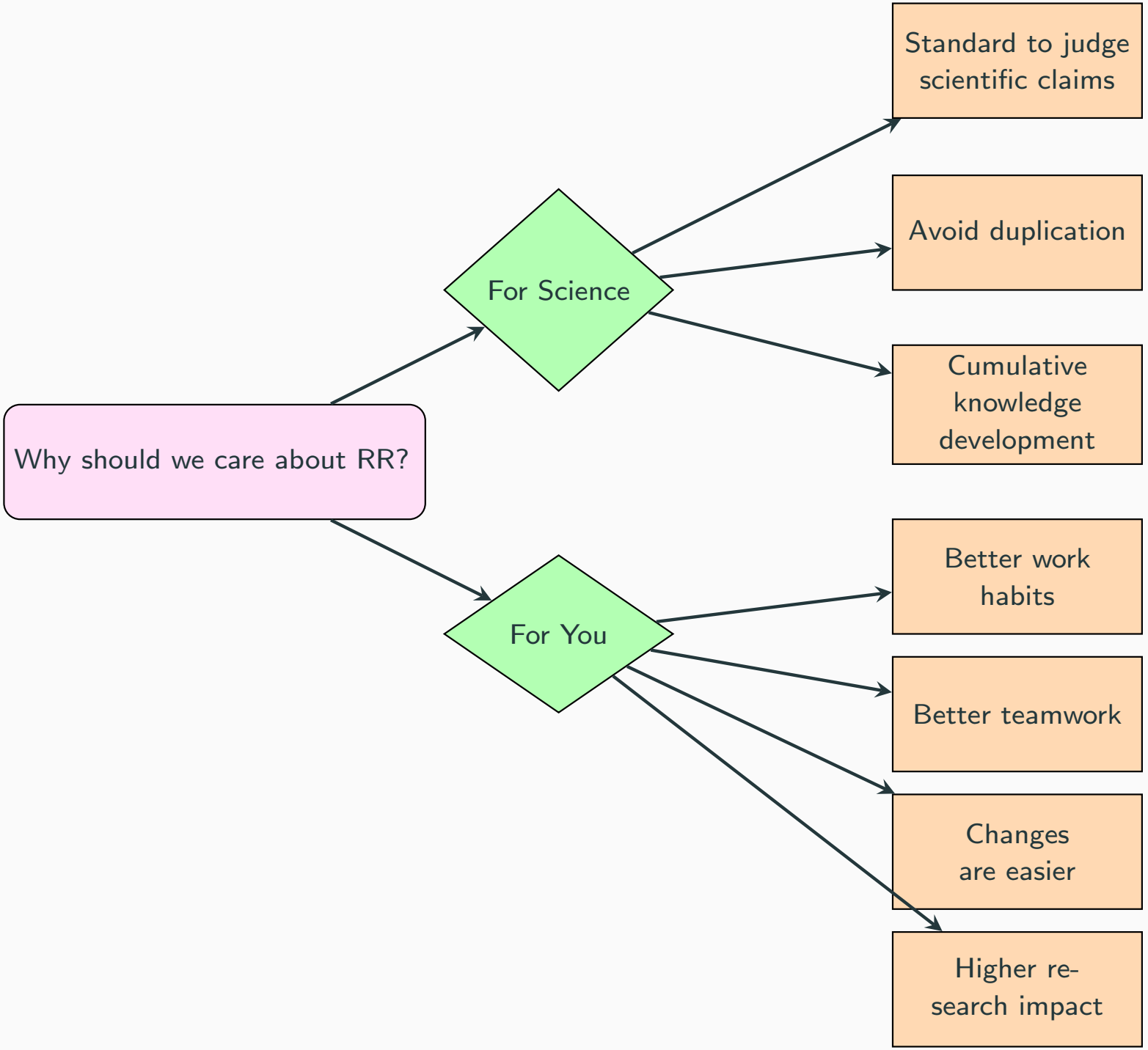
What?

What is needed for Reproducible research?



Why?


Why should we care about RR?
For Science
Standard to judge scientific claims
Avoid duplication
Cumulative knowledge development


Why should we care about RR?
For Science
For You
Standard to judge scientific claims
Avoid duplication
Cumulative knowledge development
Better work habits
Better teamwork
Changes are easier
Higher research impact

001-motivating-example

A Motivating Example

Demonstrate: 001-motivating-example

How?

Tools for Reproducible Research

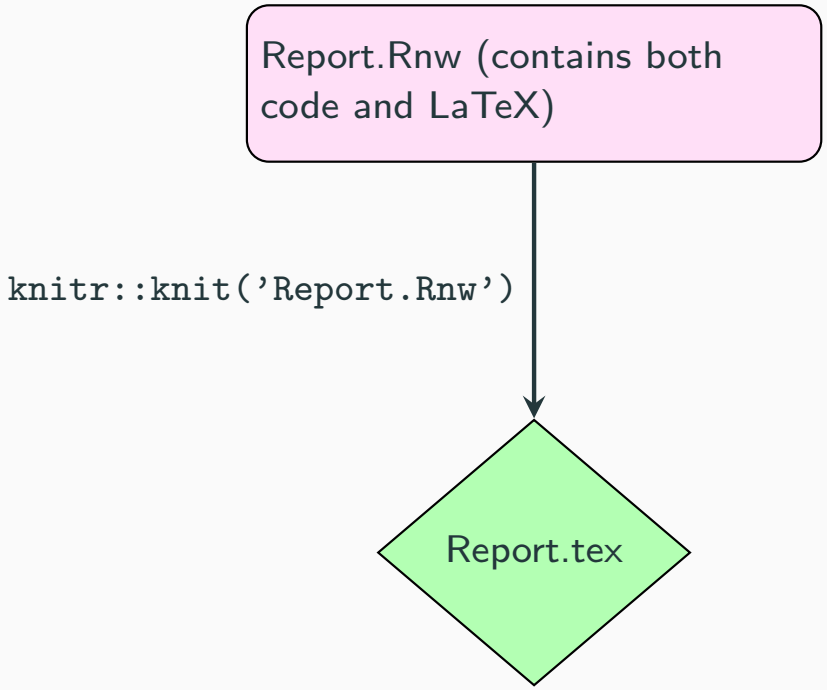
Free and Open Source Software

- `RStudio`: Creating, managing, compiling documents
- LaTeX: Markup language for typesetting a pdf
- Markdown: Markup language for typesetting an html
- `R`: Statistical analysis language
- `knitr`: Integrate LaTeXand `R` code. Based on Prof. Friedrich Leisch's `Sweave`

knitr

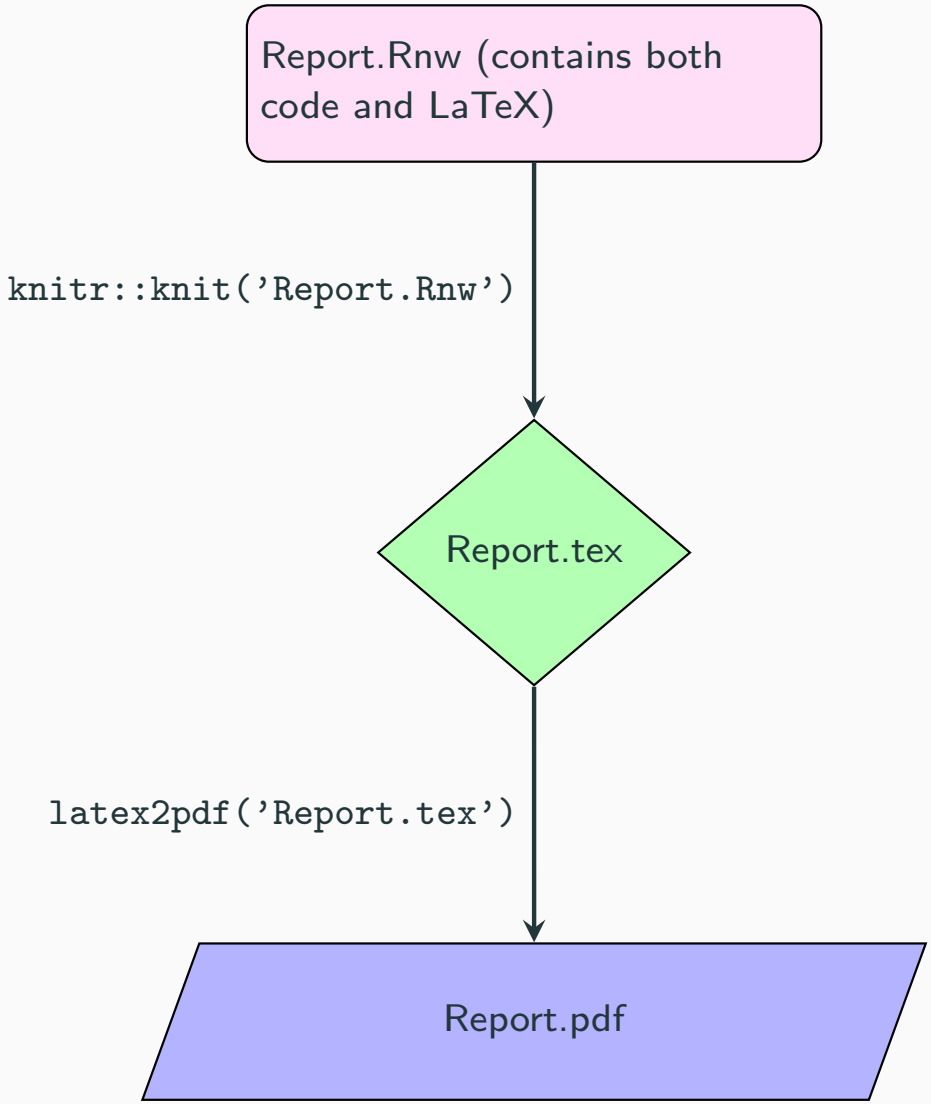
What `knitr` does

LaTeX example:



What `knitr` does

LaTeX example:

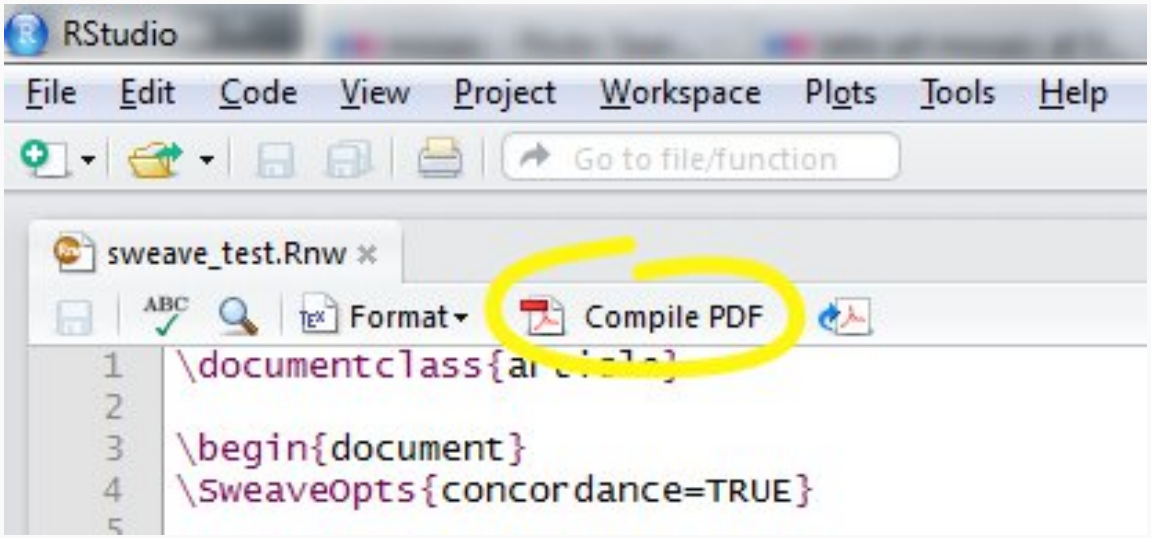


Compiling a `.Rnw` document

The two steps on previous slide can be executed in one command:

```
knitr::knit2pdf()
```

or in RStudio:



Incorporating R code

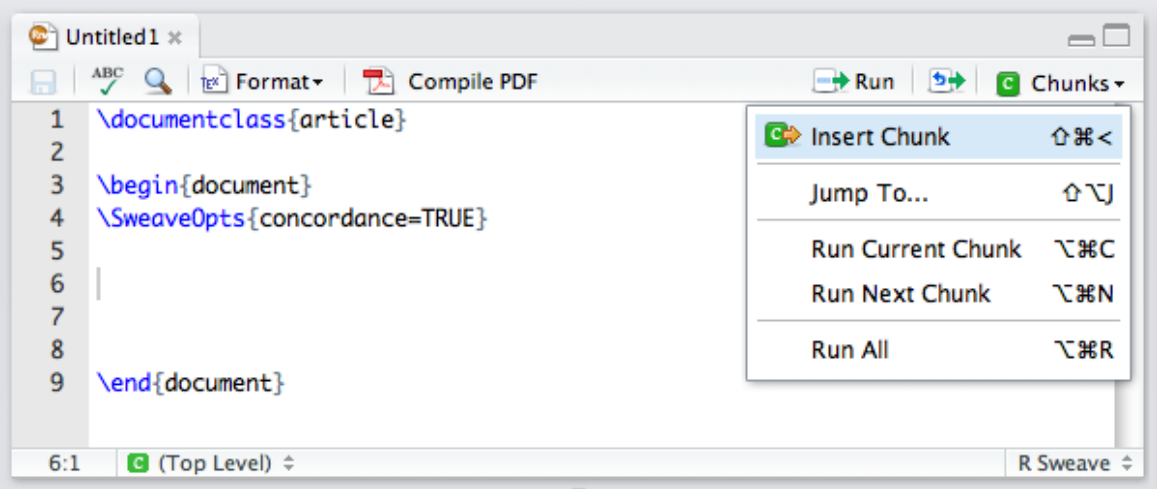
- Insert R code in a **Code Chunk** starting with

  << >>=

  and ending with

  @

In RStudio:



Example 1: Show code and results

```
<<example-code-chunk-name, echo=TRUE>>=
x <- rnorm(50)
mean(x)
@
```

produces

```
x <- rnorm(50)
mean(x)

## [1] 0.12
```

Example 2: Tidy code

```
<<example-code-chunk-name2, echo=TRUE, tidy=TRUE>>=
for(i in 1:5){ print(i+3)}
@
```

produces

```
for (i in 1:5) {
    print(i + 3)
}


## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
```

Example 2.2: don't show code

```
<<example-code-chunk-name3, echo=FALSE>>=
for(i in 1:5){ print(i+3)}
@
```

produces

```
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
```

Example 2.3: don't evaluate and don't show the code

```
<<example-code-chunk-name4, echo=FALSE, eval=FALSE>>=
for(i in 1:5){ print(i+3)}
@
```

produces

R output within the text

- Include R output within the text
- We can do that with "S-expressions" using the command `\Sexpr{...}`

Example:

The iris dataset has `\Sexpr{nrow(iris)}` rows and `\Sexpr{ncol(iris)}` columns

produces

The iris dataset has 150 rows and 5 columns

Include a Figure

```
<<fig.ex, fig.cap='Linear Regression',fig.height=3,fig.width=3>>=
plot(mtcars[ , c('disp','mpg')])
fit <- lm(mpg ~ disp , data = mtcars)
abline(fit,lwd=2)
@
```

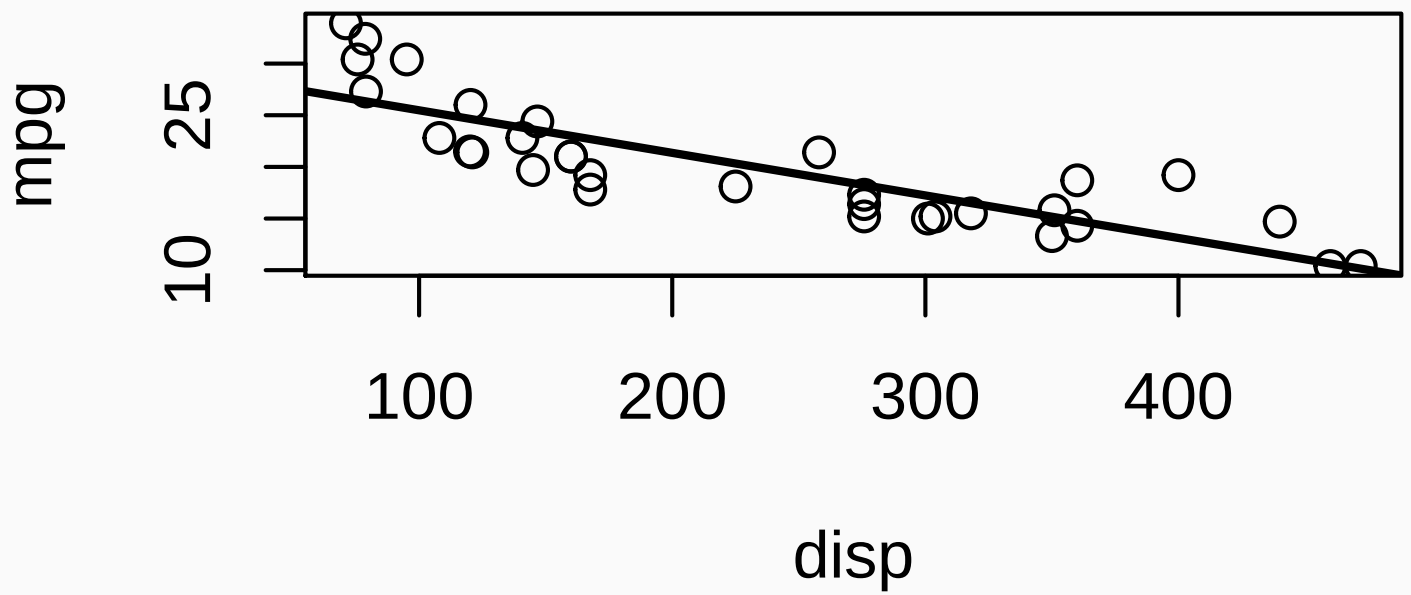


Figure 1: Linear regression

Include a Table

```
<<table.ex, results='asis'>>=
library(xtable)
tab <- xtable(iris[1:5,1:5],caption='Sample of Iris data')
print(tab, include.rownames=FALSE)
@
```

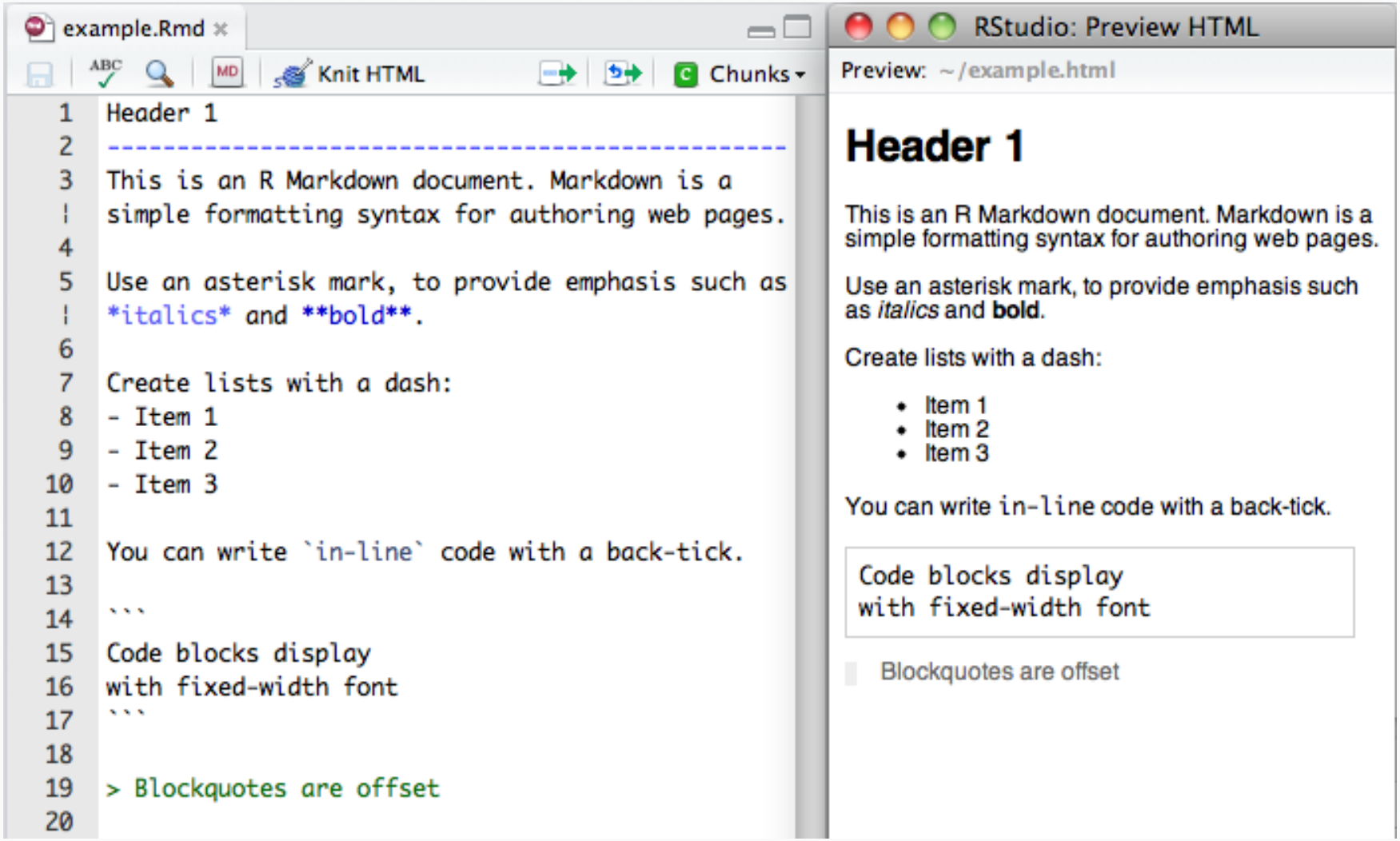
```
library(xtable)
tab <- xtable(iris[1:5,1:5], caption = 'Sample of Iris data')
print(tab, include.rownames = F)
```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.10	3.50	1.40	0.20	setosa
4.90	3.00	1.40	0.20	setosa
4.70	3.20	1.30	0.20	setosa
4.60	3.10	1.50	0.20	setosa
5.00	3.60	1.40	0.20	setosa

Table 1: Sample of Iris data

RMarkdown

Markdown: HTML without knowing HTML



R + Markdown = RMarkdown

chunks.Rmd

Knit HTML | Chunks

```
1  R Code Chunks
2  ==========
3
4  With R Markdown, you can insert R code
   chunks including plots:
5
6  ```{r qplot, fig.width=4, fig.height=3,
   message=FALSE}
7  # quick summary and plot
8  library(ggplot2)
9  summary(cars)
10 qplot(speed, dist, data=cars) +
11       geom_smooth()
12 ```
13
```

RStudio: Preview HTML

Preview: ~/chunks.html | Save As | Publish

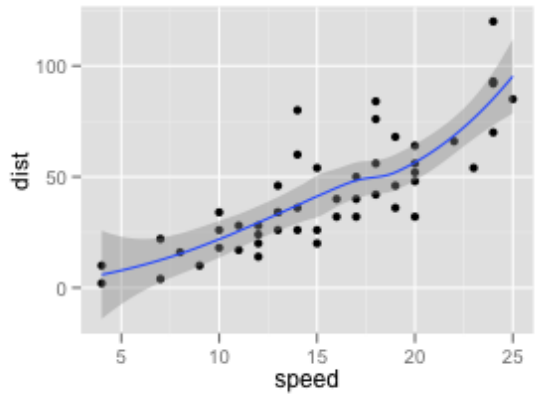
R Code Chunks

With R Markdown, you can insert R code chunks including plots:

```
# quick summary and plot
library(ggplot2)
summary(cars)
```

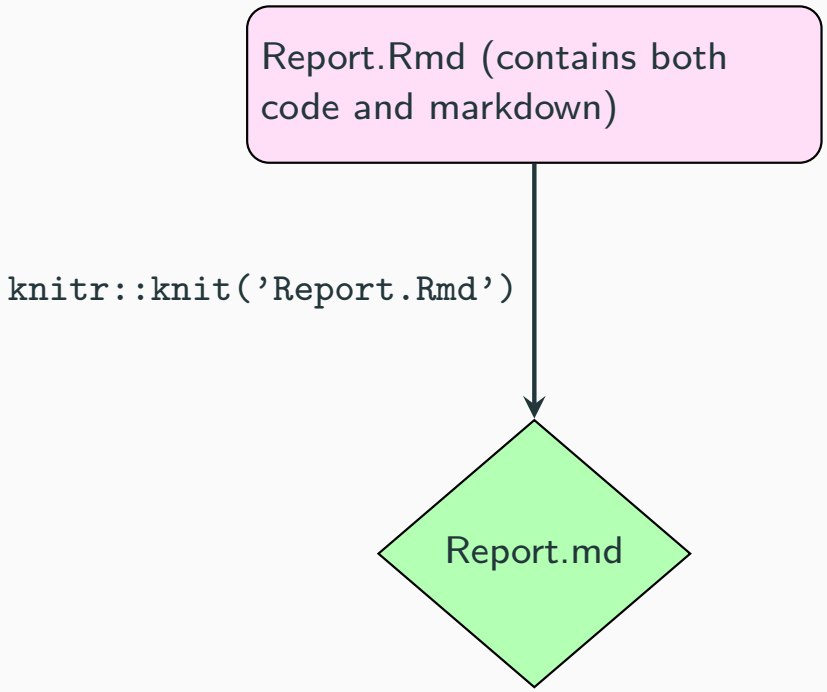
```
##      speed           dist
##  Min.   : 4.0   Min.   :  2
##  1st Qu.:12.0   1st Qu.: 26
##  Median :15.0   Median : 36
##  Mean   :15.4   Mean   : 43
##  3rd Qu.:19.0   3rd Qu.: 56
##  Max.   :25.0   Max.   :120
```

```
qplot(speed, dist, data = cars) + geom_smooth()
```



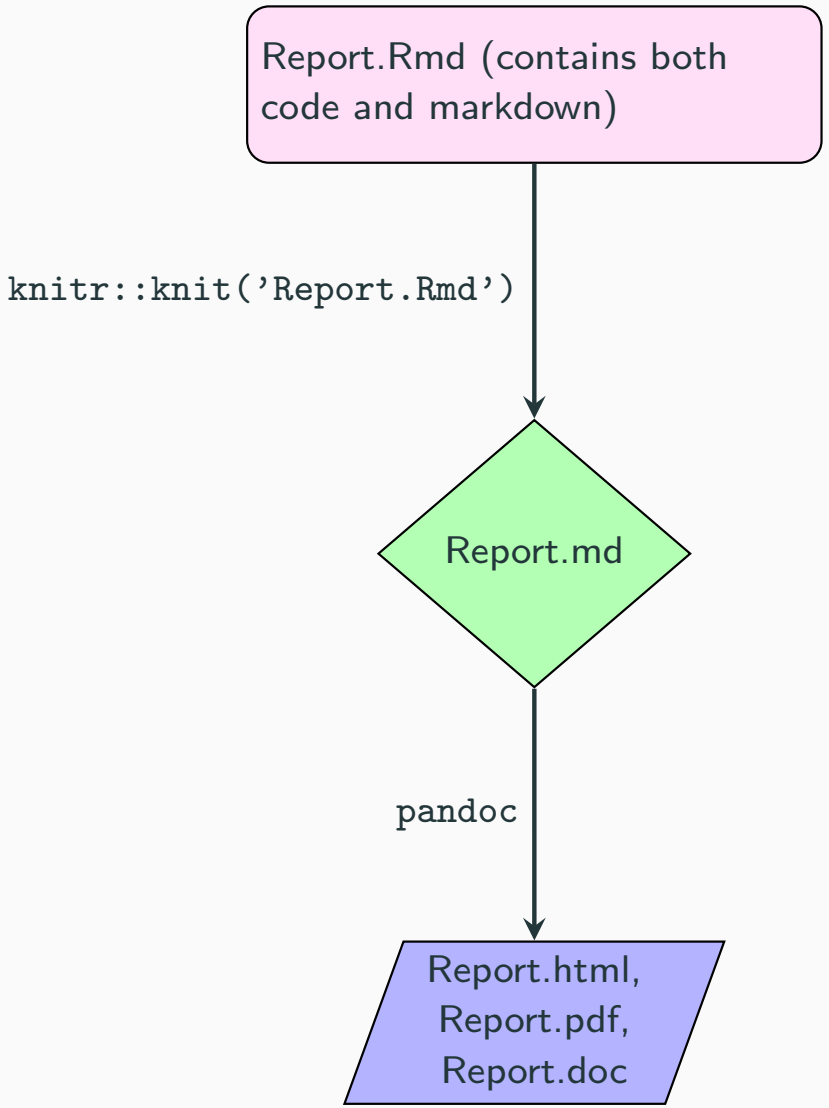
What `rmarkdown` does

`RMarkdown` example:



What `rmarkdown` does

`RMarkdown` example:

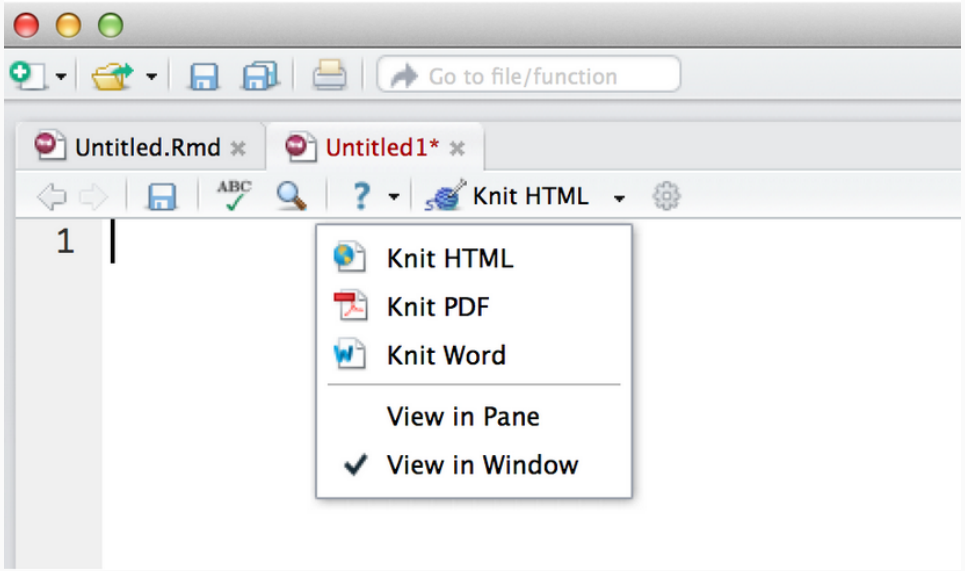


Compiling a `.Rmd` document

The two steps on previous slide can be executed in one command:

```
rmarkdown::render()
```

or in RStudio:



Final Remarks

How to choose between LaTeX and Markdown ?

LaTeX

- math/stat symbols
- beamer presentations
- customized documents
- publish to journals, arXiv

Markdown

- quick and easy reports
- use javascript libraries
- interactive plots
- publish to websites

Opinion: Reproducible research can still be wrong: Adopting a prevention approach

Jeffrey T. Leek[a,1] **and Roger D. Peng**[b]

[a]*Associate Professor of Biostatistics and Oncology and* [b]*Associate Professor of Biostatistics, Johns Hopkins University, Baltimore, MD*

computational tools such as knitr, iPython notebook, LONI, and Galaxy (8) have simplified the process of distributing reproducible data analyses.

Always Remember ...

$$\text{Reproducibility} \propto \frac{1}{\text{copy paste}}$$

Is the juice worth the squeeze?

Session Info

- R version 3.4.4 (2018-03-15), `x86_64-pc-linux-gnu`
- Running under: `Ubuntu 16.04.4 LTS`
- Matrix products: default
- BLAS: `/usr/lib/openblas-base/libblas.so.3`
- LAPACK: `/usr/lib/libopenblasp-r0.2.18.so`
- Base packages: base, datasets, graphics, grDevices, methods, stats, utils
- Other packages: data.table 1.10.4-3, dplyr 0.7.5, ggplot2 2.2.1.9000, knitr 1.20, xtable 1.8-2
- Loaded via a namespace (and not attached): assertthat 0.2.0, bindr 0.1.1, bindrcpp 0.2.2, colorspace 1.3-2, compiler 3.4.4, evaluate 0.10.1, formatR 1.5, glue 1.2.0, grid 3.4.4, gtable 0.2.0, highr 0.7, lazyeval 0.2.1, magrittr 1.5, munsell 0.4.3, pillar 1.2.3, pkgconfig 2.0.1, plyr 1.8.4, purrr 0.2.5, R6 2.2.2, Rcpp 0.12.17, rlang 0.2.0.9001, scales 0.5.0.9000, stringi 1.2.3, stringr 1.3.1, tibble 1.4.2, tidyselect 0.2.4, tools 3.4.4, withr 2.1.2