

# Reproducible Research Tools for R

January 2017

Boriana Pratt  
Princeton University



# Literate programming

## Literate programming (1984)

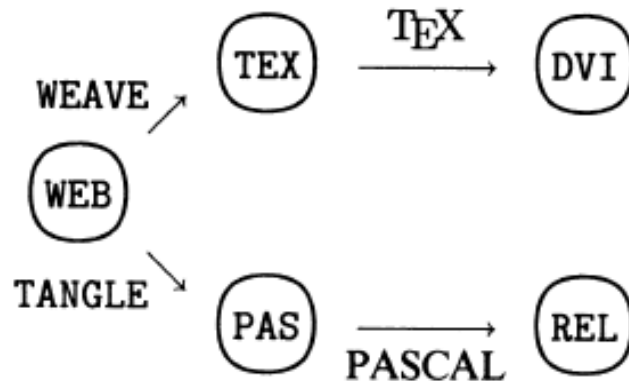
“ I believe that the time is right for significantly better documentation of programs, and that we can best achieve this by considering programs to be works of literature. Hence, my title: ‘Literate Programming’. “

Donald E. Knuth. Literate Programming. The Computer Journal, 27(2):97-111, May1984.

<http://comjnl.oxfordjournals.org/content/27/2/97.full.pdf+html>

## WEB

“... is a combination of two other languages: (1) a document formatting language and (2) a programming language.”



# terms

- Tangle  
extract the code parts (code chunks), then run them sequentially
- Weave  
extract the text part (documentation chunks) and weave back in the code and code output

# Noweb

Norman Ramsey (1989) – Noweb – simple literate programming tool...

<https://www.cs.tufts.edu/~nr/noweb/>

“*Literate programming* is the art of preparing programs for human readers. “

Noweb syntax includes two parts:

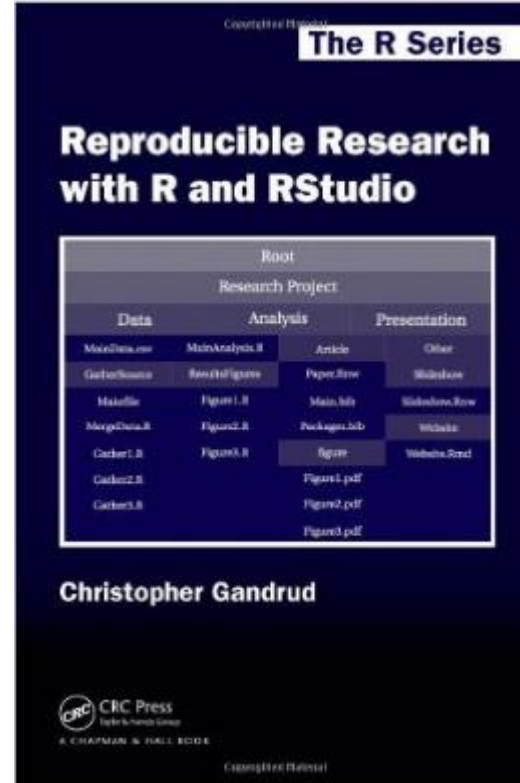
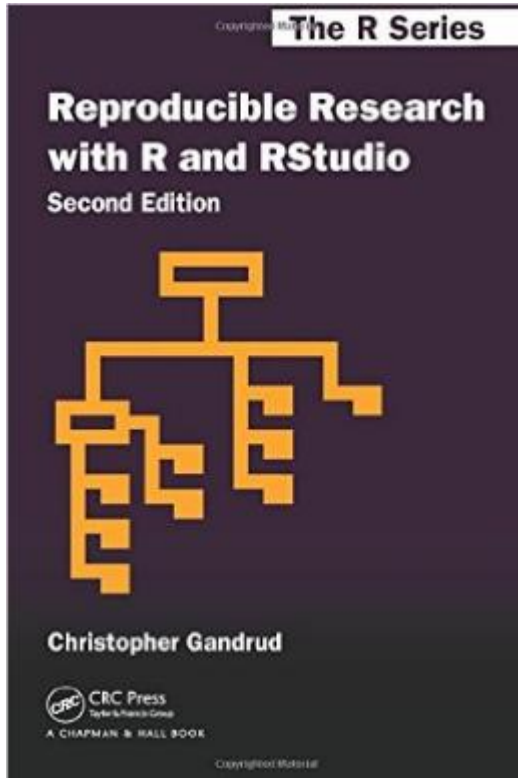
Code chunk:

<<chunk name>>=      - section that starts with <<*name*>>=

Documentation chunk:

@                      - line that starts with @ followed by a space; default for the first chunk

# Tools for R



# Tools for R

- Sweave

<https://stat.ethz.ch/R-manual/R-devel/library/utils/doc/Sweave.pdf>

## **What is Sweave?**

A tool that allows to embed the R code in LaTeX documents. The purpose is to create dynamic reports, which can be updated automatically if data or analysis change.

## **How do I cite Sweave?**

To cite Sweave please use the paper describing the first version:

Friedrich Leisch. Sweave: Dynamic generation of statistical reports using literate data analysis. In W. Härdle and B. Rönz, editors, Compstat 2002 – Proceedings in Computational Statistics, pages 575-580. Physica Verlag, Heidelberg, 2002.

# Tools for R

- **How to use Sweave in R?**

1) Create .Rnw file ("example.Rnw")

Important syntax :

ϕ Documentation chunk (the default):

    @ as first character of a line, followed by a space or newline

ϕ Code chunk:

    <<name>>= at the beginning of a line

ϕ Everything should be between

    \begin{document}

    and

    \end{document}

For Latex syntax see: <http://wch.github.io/latexsheet/latexsheet.pdf>

2) Run `Sweave("example.Rnw")`

3) Then run `tools::texi2pdf("example.tex")`

# Sweave Example

From 'example.Rnw':

```
\documentclass{article}
\usepackage{amsmath}
\usepackage{graphicx}

\begin{document}

\title{Sweave R Example}
\author{B. Pratt}
\date{\today}
\maketitle

\section{Example}
This is an example of a linear regression (also known as OLS). The
arithmetic is:
\begin{equation} \label{eq:linear-predictor}
Y_i = \alpha + \beta X_i
\end{equation}

\subsection{The data}
First we get the data of measures of muscle contraction in rat hearts:
<<data>>=
library(MASS)
data(muscle)
@
and look to see what variables are there:
<<vars>>=
names(muscle)
@
Then we use regression model to model the relationship between length and
concentration:
<<lm>>=
mfit <- lm( Length ~ Conc, data=muscle)
summary(mfit)
@
```



# Sweave Example

## Sweave R Example

B. Pratt

November 16, 2016

To 'example.pdf':

### 1 Example

This is an example of a linear regression (also known as OLS). The arithmetic is:

$$Y_i = \alpha + \beta X_i \quad (1)$$

#### 1.1 The data

First we get the data of measures of muscle contraction in rat hearts:

```
> library(MASS)
> data(muscle)
```

and look to see what variables are there:

```
> names(muscle)

[1] "Strip" "Conc"  "Length"
```

Then we use regression model to model the relationship between length and concentration:

```
> mfit <- lm( Length ~ Conc, data=muscle)
> summary(mfit)
```

Call:

```
lm(formula = Length ~ Conc, data = muscle)
```

Residuals:

| Min     | 1Q     | Median | 3Q    | Max    |
|---------|--------|--------|-------|--------|
| -12.884 | -4.097 | 1.060  | 4.487 | 10.064 |

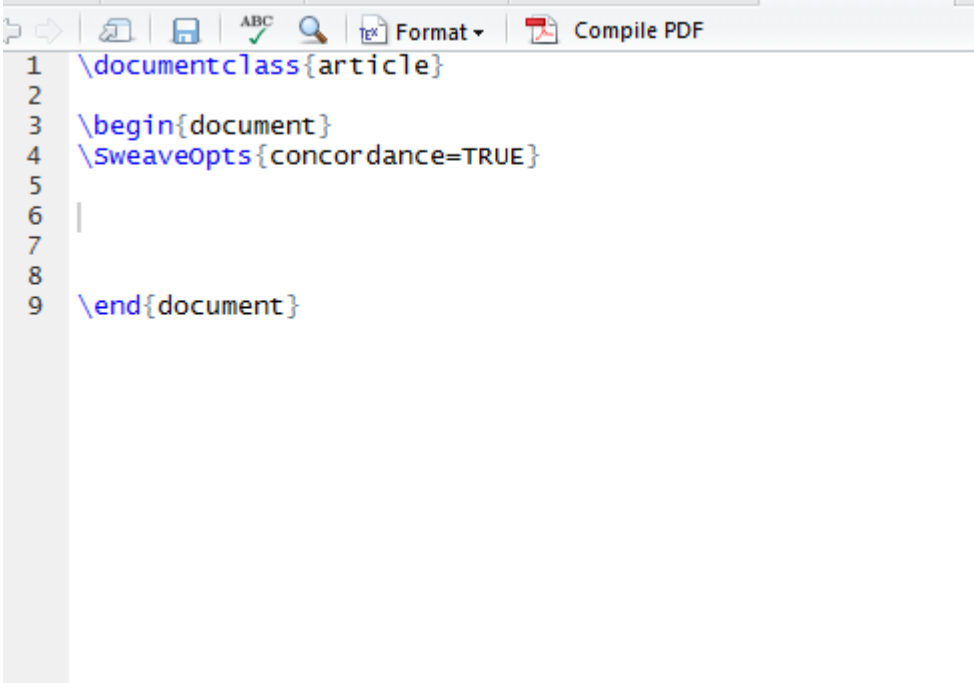
Coefficients:

|             | Estimate | Std. Error | t value | Pr(> t )     |
|-------------|----------|------------|---------|--------------|
| (Intercept) | 13.5330  | 1.4229     | 9.511   | 1.93e-13 *** |
| Conc        | 5.4030   | 0.7653     | 7.060   | 2.32e-09 *** |

---

# Sweave Example

- in Rstudio: File > New File > R Sweave

A screenshot of the RStudio interface showing a new R Sweave file. The top toolbar includes icons for navigation, saving, and compiling. The main editor area contains the following LaTeX code:

```
1 \documentclass{article}
2
3 \begin{document}
4 \SweaveOpts{concordance=TRUE}
5
6 |
7
8
9 \end{document}
```

try it...

# Tools for R

- Knitr

<http://yihui.name/knitr/demo/manual/>

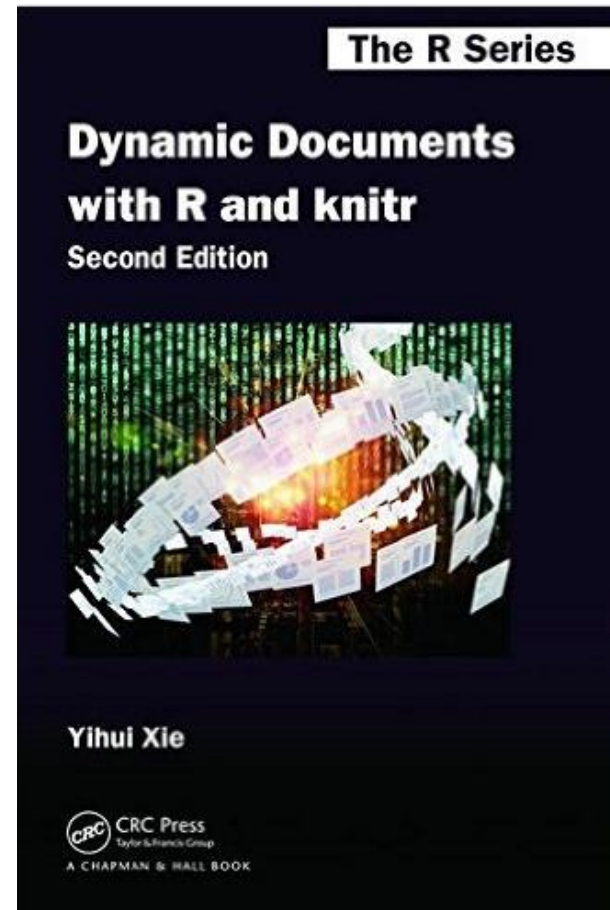
<http://yihui.name/knitr/demo/graphics/>

Knitr is an R package that enables integration of R code into LaTeX, LyX, HTML and other documents.

You write the R code in ‘code chunks’.

Code chunks are between ````` and `````.

Knitr package on GitHub: <https://github.com/yihui/knitr>



# knitr

Setting global option for the all code chunks:

```
```{r global_options, include = FALSE}  
knitr::opts_chunk$set(fig.width = 12, fig.height = 8, echo = FALSE, |  
                      warning = FALSE, message = FALSE)  
```
```

Options to use for code chunks:

## R Code Chunk Arguments

| option     | default  | effect  |
|------------|----------|---|
| eval       | TRUE     | Whether to evaluate the code and include its results      |
| echo       | TRUE     | Whether to display code along with its results            |
| warning    | TRUE     | Whether to display warnings                               |
| error      | FALSE    | Whether to display errors                                 |
| message    | TRUE     | Whether to display messages                               |
| tidy       | FALSE    | Whether to reformat code in a tidy way when displaying it |
| results    | "markup" | "markup", "asis", "hold", or "hide"                       |
| cache      | FALSE    | Whether to cache results for future renders               |
| comment    | "##"     | Comment character to preface results with                 |
| fig.width  | 7        | Width in inches for plots created in chunk                |
| fig.height | 7        | Height in inches for plots created in chunk               |

# Tools for R

## Rmarkdown ( Yihui Xie, 2014)

From [http://rmarkdown.rstudio.com/authoring\\_quick\\_tour.html](http://rmarkdown.rstudio.com/authoring_quick_tour.html) :

“ Markdown is a simple formatting language designed to make authoring content easy for everyone. Rather than write in complex markup code (e.g. HTML or LaTeX), you write in plain text with formatting cues. Pandoc uses these cues to turn your document into attractive output. “

How it works:

1. Create an .Rmd file that contains a combination of markdown and R code chunks.
2. The .Rmd file is fed to [knitr](#), which executes all of the R code chunks and creates a new markdown (.md) document which includes the R code and it's output.
3. The .md file is then processed by [pandoc](#), which creates a finished web page, PDF, MS Word document, slide show, handout, book, dashboard, package vignette or other format.

For pandoc file formats, see here: <http://pandoc.org/>

# Rmarkdown Example

R code chunk:

```
```${r}  
data(muscle)  
names(muscle)  
```
```

See Rmarkdown cheat sheet here:

<https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf>

In Rstudio:

.Rmd  $\xrightarrow{\text{knitr}}$  .md  $\xrightarrow{\text{pandoc}}$  .pdf

# Rmarkdown in R

To run Rmarkdown in R (requires pandoc to be installed) :

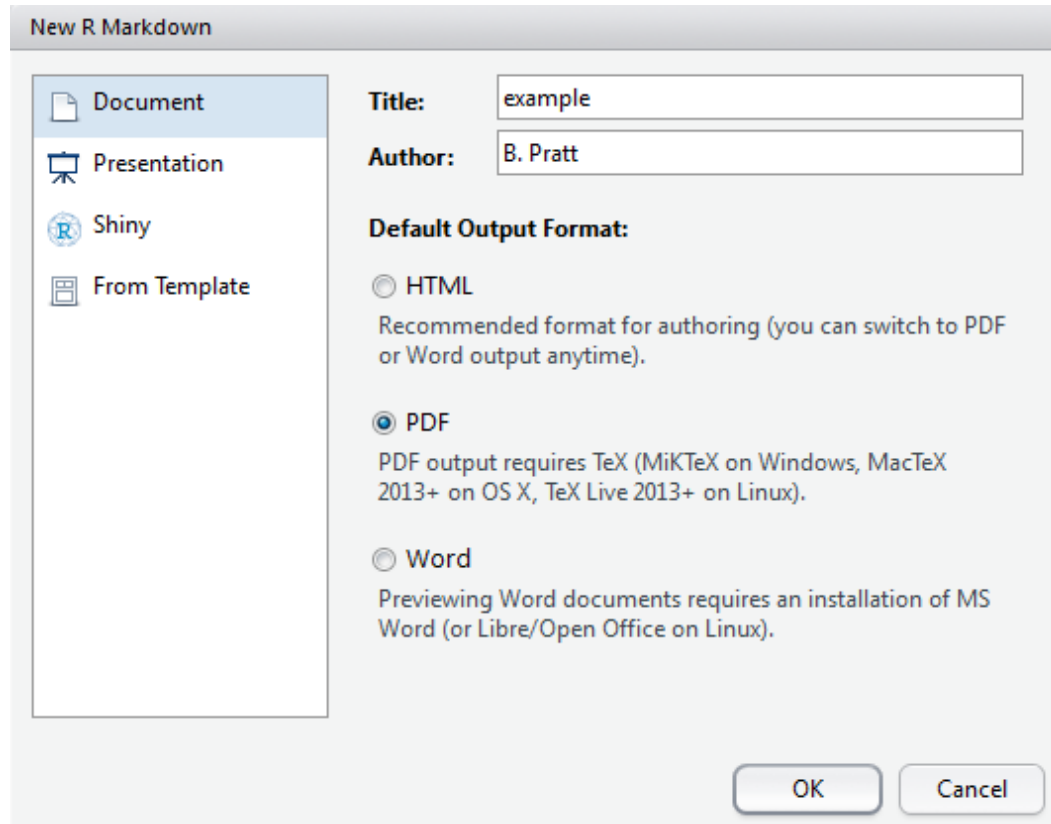
```
install.packages("rmarkdown")  
library(rmarkdown)  
render("example.Rmd")
```

Default output is .pdf, could specify html:

```
render("example.Rmd", html_document())
```

# Rmarkdown in Rstudio

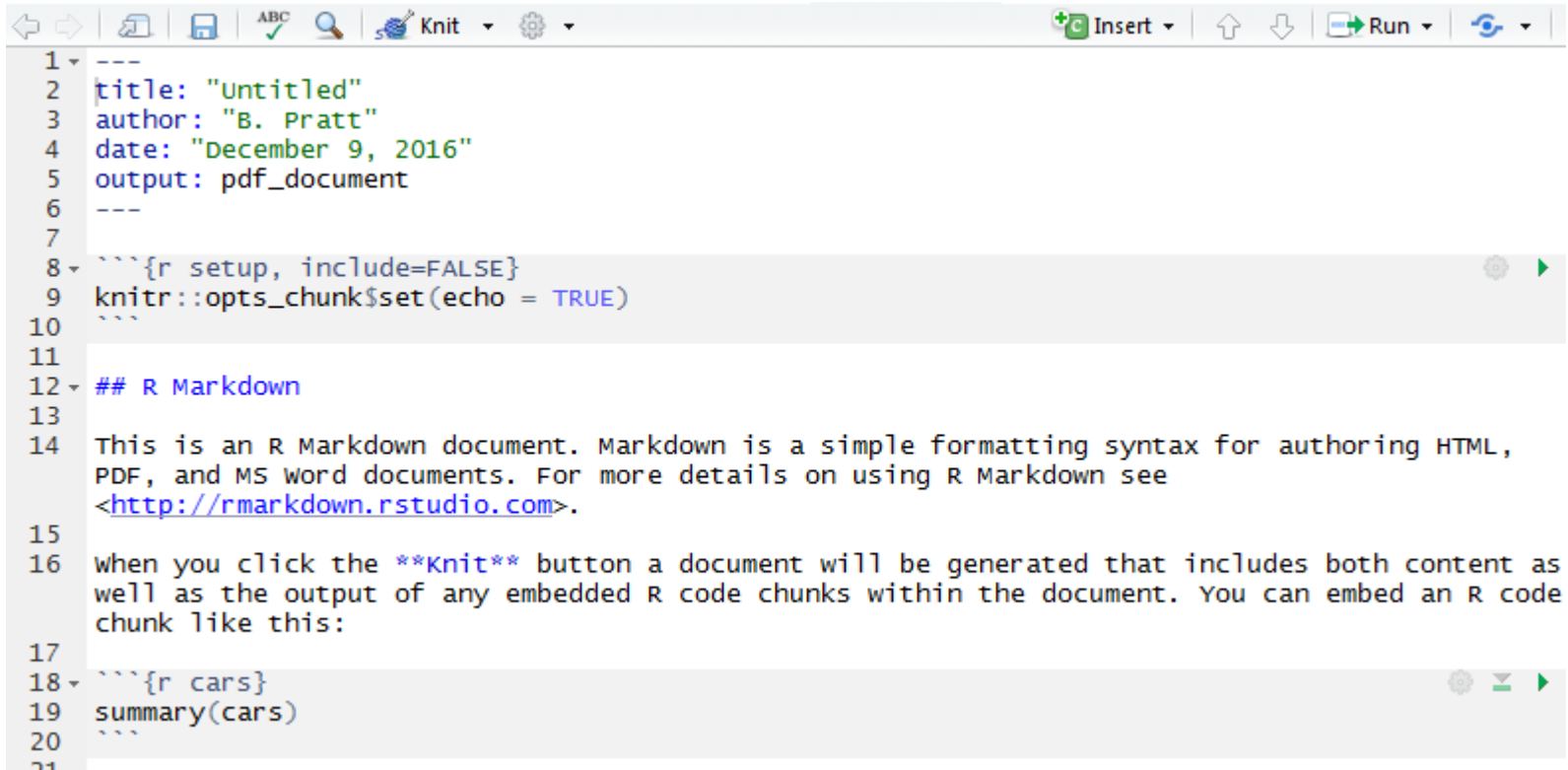
In Rstudio: File > New File > R Markdown



try it...



# Rmarkdown in Rstudio



```
1 ---
2 title: "Untitled"
3 author: "B. Pratt"
4 date: "December 9, 2016"
5 output: pdf_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ```
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML,
15 PDF, and MS word documents. For more details on using R Markdown see
16 <http://rmarkdown.rstudio.com>.
17
18 when you click the Knit button a document will be generated that includes both content as
19 well as the output of any embedded R code chunks within the document. You can embed an R code
20 chunk like this:
21
22 ```{r cars}
23 summary(cars)
24 ```
```

Top part is a YAML header. YAML is data serialization language.

For more on YAML, see: <http://yaml.org/spec/1.0/>

# Rmarkdown in Rstudio

Let's try to:

1. add a sentence
2. add a header
3. add a code chunk

Example code chunk:

```
```{r}
a <- 1:10
a
```
```

Some code chunk options:

```
```{r chunk_name, echo=FALSE, include=TRUE} – do not display code in output document
```{r scatter, fig.width=8, fig.height=6} – set size for figure
```{r, include=FALSE} – do not include output from this chunk
```

# Rmarkdown in Rstudio

## Inline Expressions:

The sum of three, five and eight is ``r 3+5+8``.

## Result:

The sum of three, five and eight is 16.

## Regression table:

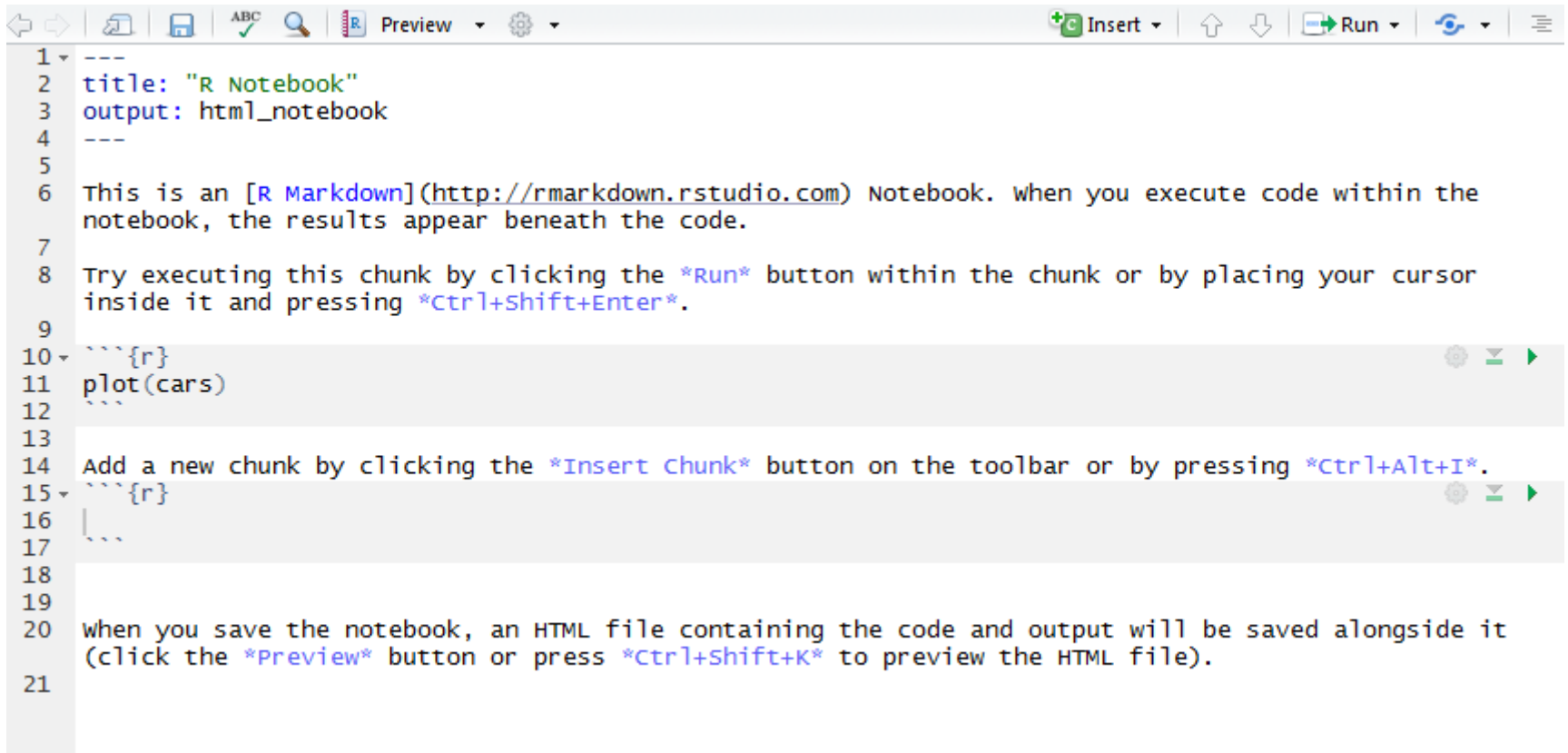
```
```{r, results='asis'}
library(knitr)
n <- 100
x <- rnorm(n)
y <- 2*x + rnorm(n)
out <- lm(y ~ x)
kable(summary(out)$coef, digits=2)
```
```

## Try to:

- . add to your .Rmd file a plot, then re-size it using the chunk options
- . add to your .Rmd file an inline expression
- . add a regression table (could use package `xtable`)

# R Notebook

In Rstudio (the newest version) go to File -> New File -> R Notebook to create a R Notebook file, which is a Rmarkdown file, accompanied by a .nb (notebook) file which could be previewed in a browser.



```
1 ---
2 title: "R Notebook"
3 output: html_notebook
4 ---
5
6 This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you execute code within the
7 notebook, the results appear beneath the code.
8
9 Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor
10 inside it and pressing *Ctrl+Shift+Enter*.
11
12 ```{r}
13 plot(cars)
14 ```
15
16 Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.
17
18 ```{r}
19
20 ```
21
22 When you save the notebook, an HTML file containing the code and output will be saved alongside it
23 (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).
```

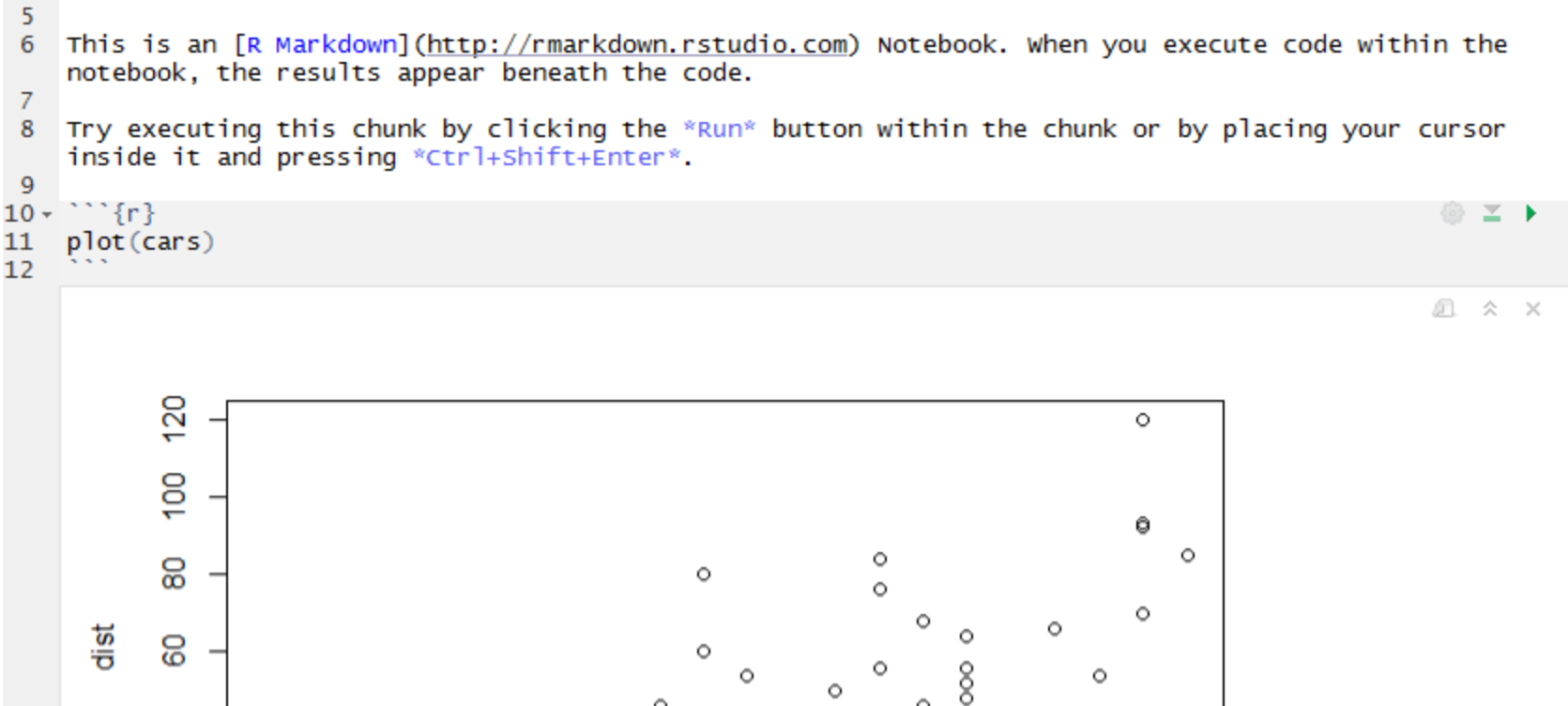
# R Notebook

You have the option to run some chunks and the output shows right under (in Rstudio click on the little green arrow next to the chunk).

```
1 ---
2 title: "R Notebook"
3 output: html_notebook
4 ---
5
6 This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you execute code within the
7 notebook, the results appear beneath the code.
8 Try executing this chunk by clicking the Run button within the chunk or by placing your cursor
9 inside it and pressing Ctrl+Shift+Enter.
10
11 plot(cars)
12
13
14 Add a new chunk by clicking the Insert Chunk button on the toolbar or by pressing Ctrl+Alt+I.
15
```

# R Notebook

You have the option to run some chunks and the output shows right under:



# R Notebook

When you open your Rnotebook in a browser, you can hide code chunk code:

## R Notebook

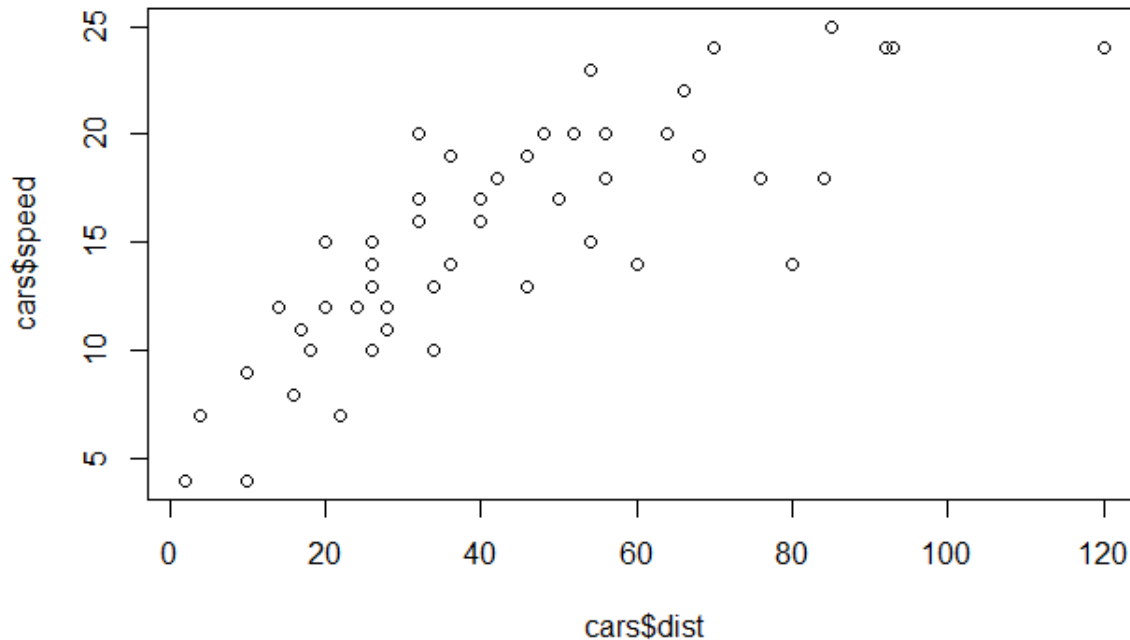
Code ▾

This is an [R Markdown](#) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

```
plot(cars$dist, cars$speed)
```

Hide



# R Notebook

Or you can show the code chunk part (if it was hidden):

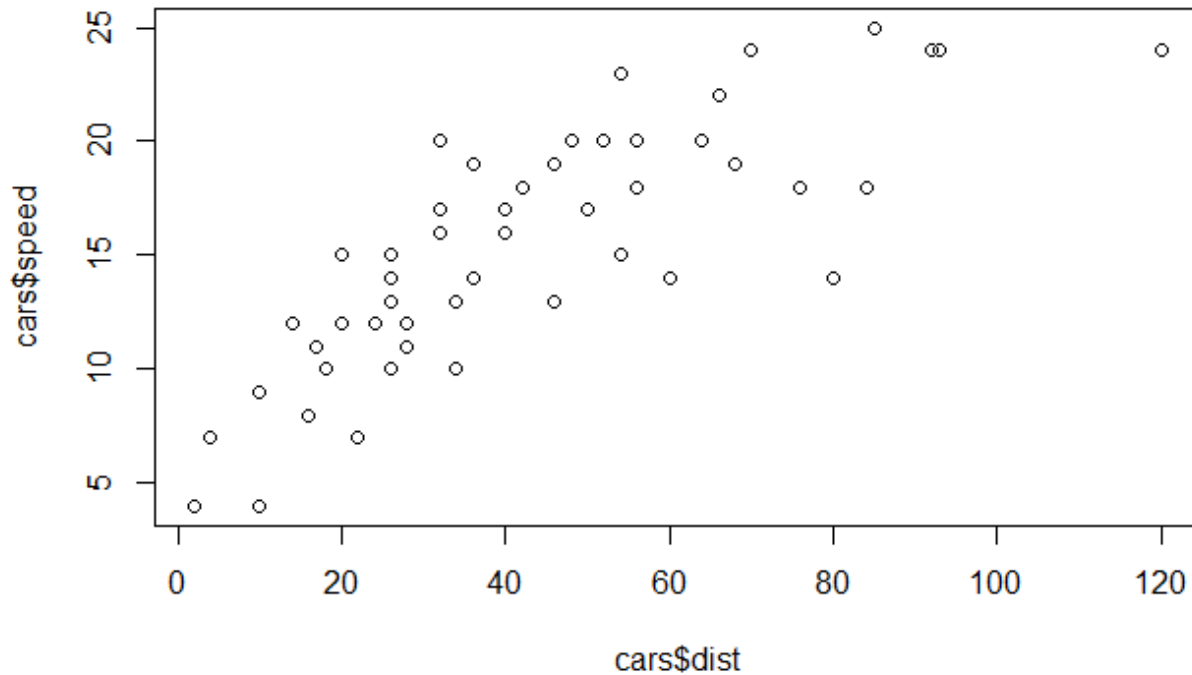
## R Notebook

Code ▾

This is an [R Markdown](#) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

Code





# R Notebook

You can hide all code chunks and only leave the output visible:

## R Notebook

This is an [R Markdown](#) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Sh*

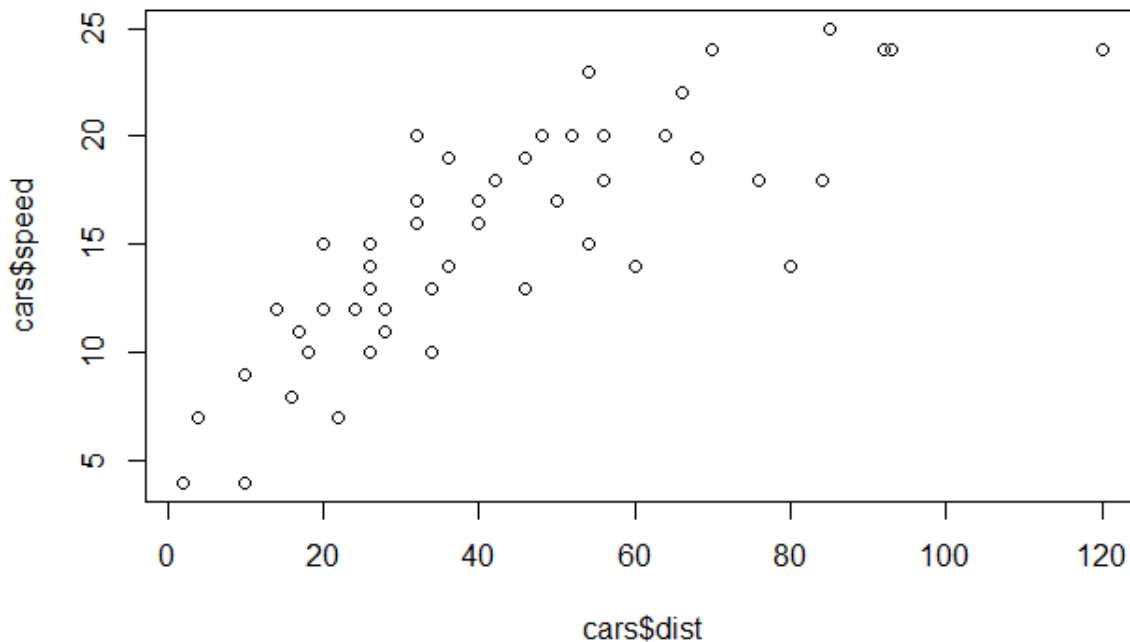
```
plot(cars$dist, cars$speed)
```

Code ▾

Show All Code

Hide All Code

Download Rmd



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

Thank you.