Sane's STATS Academy of Statistics

Research Presentation

Sane's STATS Academy of Statistics

Introduction

Sane's STATS Academy of Statistics

Important Terms

- First let us understand few terms:
  - Dynamic Document
  - Reproducible Research
  - Markdown
  - Literate Programming

Dynamic Documents

- A living document or dynamic document is a document that is continually edited and updated.
- Unlike static documents dynamic documents can get easily updated

Reproducible Research

- It is the research intended to facilitate the replication of scientific findings by including data and software code
- This allows readers to verify the findings for themselves

Markdown

- Markdown is a lightweight mark-up language with plain text formatting syntax designed
- It can be converted to HTML and many other formats using a tool by the same name
- Markdown is often used
  - to format readme files,
  - for writing messages in online discussion forums,
  - to create rich text using a plain text editor.

Sane's STATS Academy of Statistics

Literate Programming

- Literate programming is an approach to programming introduced by Donald Knuth in which a program is given as an explanation of the program logic in a natural language, such as English
- Literate Programming has following properties:
  - Article being a mix of text and code
  - Analysis code divided into text and code
  - Presentation code formats the results (tables, graphs etc.)

Resemblance of Markdown with HTML

text using Markdown syntax	the corresponding HTML produced by a Markdown processor	the text viewed in a browser

```
Heading
=======

Sub-heading
-----------

### Another deeper heading

Paragraphs are separated
by a blank line.

Leave 2 spaces at the end of a line to do a
line break

Text attributes *italic*, **bold**,
`monospace`, ~~strikethrough~~ .

Shopping list:

  * apples
  * oranges
  * pears

Numbered list:

  1. apples
  2. oranges
  3. pears
```

```
<h1>Heading</h1>

<h2>Sub-heading</h2>

<h3>Another deeper heading</h3>

<p>Paragraphs are separated
by a blank line.</p>

<p>Leave 2 spaces at the end of a line to
do a<br />
line break</p>

<p>Text attributes <em>italic</em>,
<strong>bold</strong>,
<code>monospace</code>,
<s>strikethrough</s>.</p>

<p>Shopping list:</p>

<ul>
<li>apples</li>
<li>oranges</li>
<li>pears</li>
</ul>

<p>Numbered list:</p>
```

Heading

Sub-heading

Another deeper heading

Paragraphs are separated by a blank line.

Leave 2 spaces at the end of a line to do a
line break

Text attributes *italic*, **bold**, `monospace`, ~~strikethrough~~.

Shopping list:

- apples
- oranges
- pears

Numbered list:

1. apples
2. oranges
3. pears

The rain—not the reign—in Spain.

A link.

Sane's STATS Academy of Statistics

Markdown

Short Tutorial

Text Effects

Syntax Applied	Effect
Business Analytics	*Business Analytics* (italic)
Business Analytics	**Business Analytics** (bold)
# Header 1	Header 1
## Header 2	Header 2
### Header 3	Header 3
Superscript^2^	Superscript^2
- R - Python - SAS	• R • Python • SAS
1. R 2. Python 3. SAS	1. R 2. Python 3. SAS

Text Effects

Syntax Applied	Effect
[Sane's Academy](www.saneacademy.com)	Sane's Academy
I spend so much time on [Twitter][1] and [Facebook][2] [1]: http://www.twitter.com "Twitter" [2]: http://www.facebook.com "Facebook"	I spend so much time on Twitter and Facebook
Newlines require a double space after the end of the line	

Sane's STATS Academy of Statistics

R Markdown

What is R Markdown?

- R Markdown is an authoring format that enables easy creation of ***dynamic documents***, presentations, and reports from R
- It combines the core syntax of markdown with embedded R code chunks
- We can easily edit the markdown documents in RStudio

Package rmarkdown

- Markdown functionality is provided by package *rmarkdown* which needs to be installed
- Markdown can be converted to HTML document with the help of package *rmarkdown*

Package knitr

- *knitr* is a package that enables integration of R code into LaTeX, LyX, HTML, Markdown, AsciiDoc, and reStructuredText documents
- The purpose of *knitr* is to allow reproducible research in R through the means of **Literate Programming**.

Package xtable

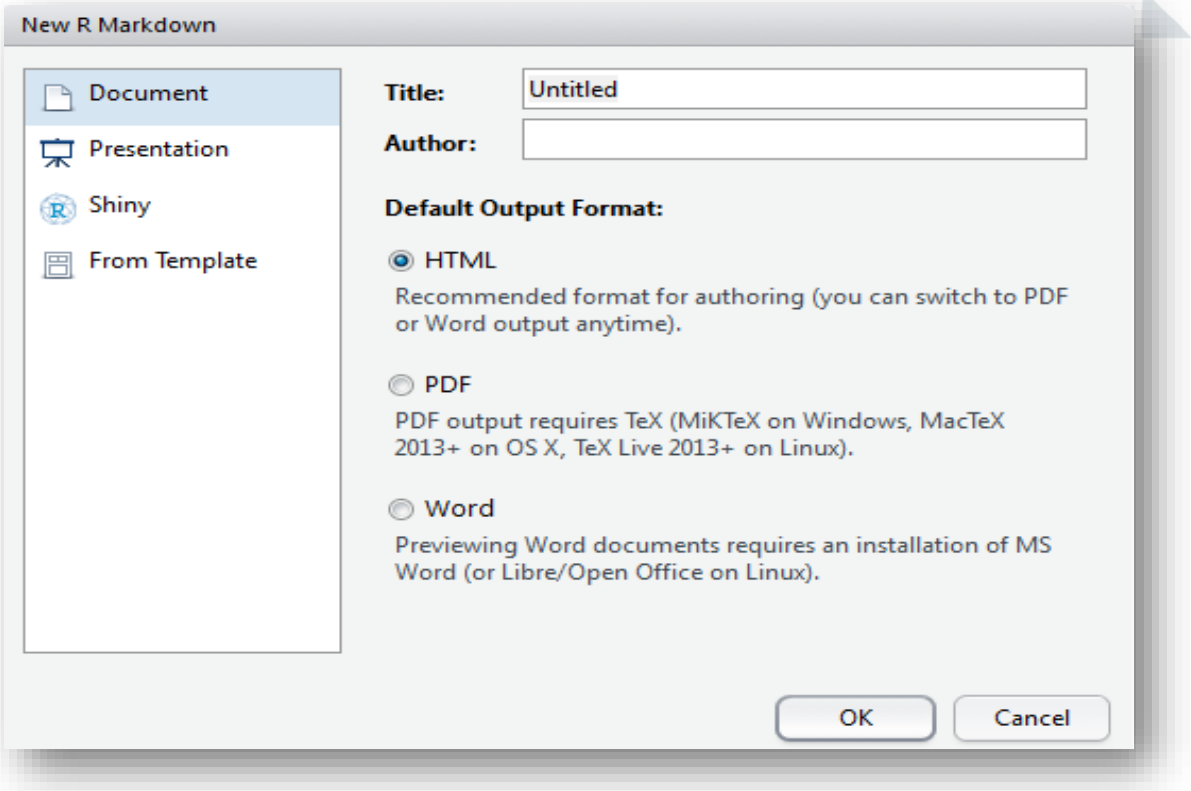
- Package *xtable* includes function *xtable()* which attractively formats data frames and matrices for inclusion in reports
- *xtable()* can also format objects produced by *lm(), glm(), aov(), table(), ts()* etc.

Creating R Markdown Document

- To incorporate R output details in a document, create a text document which contains:
  - Report text
  - Markdown syntax
  - R code chunks (R code surrounded by delimiters)
- The text file for R Markdown is saved with extension .Rmd

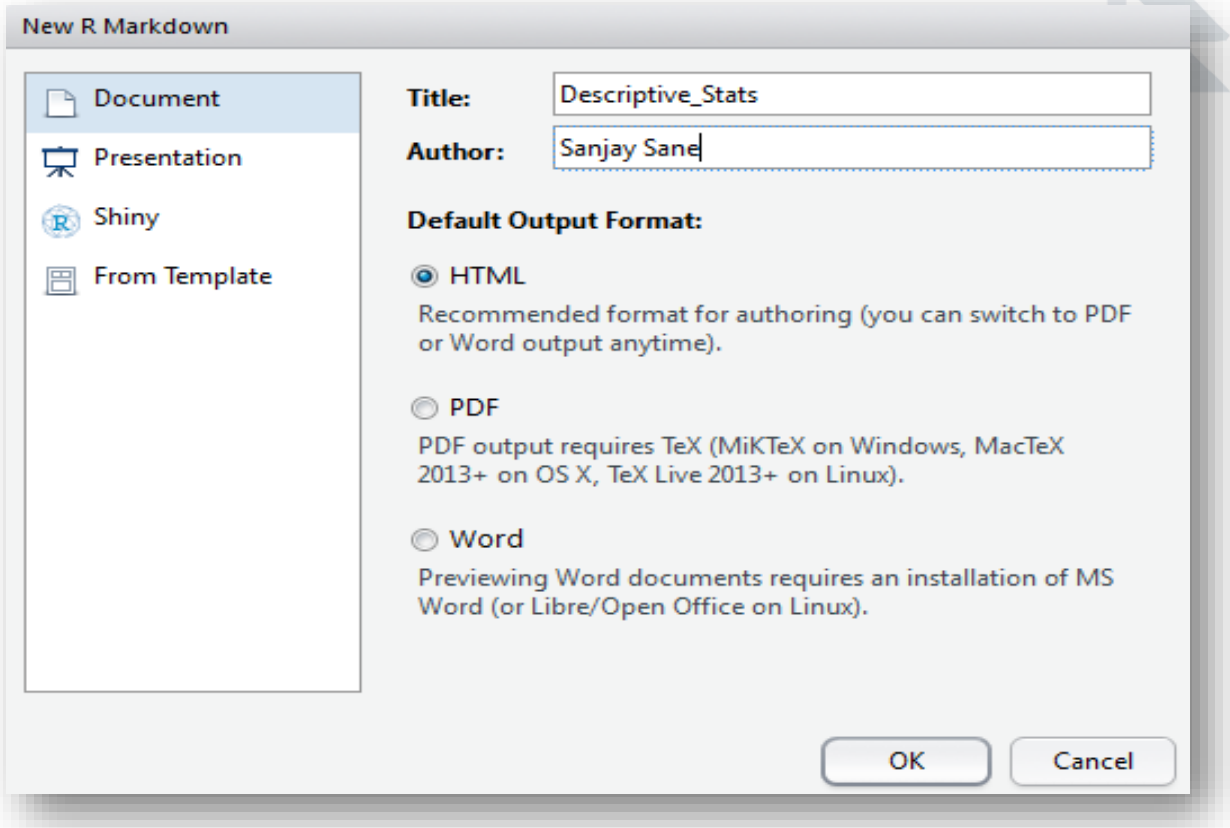
Using R Studio for Markdown

- Choose File > New File > R Markdown from the R Studio Menu
- You will see the following dialog box:



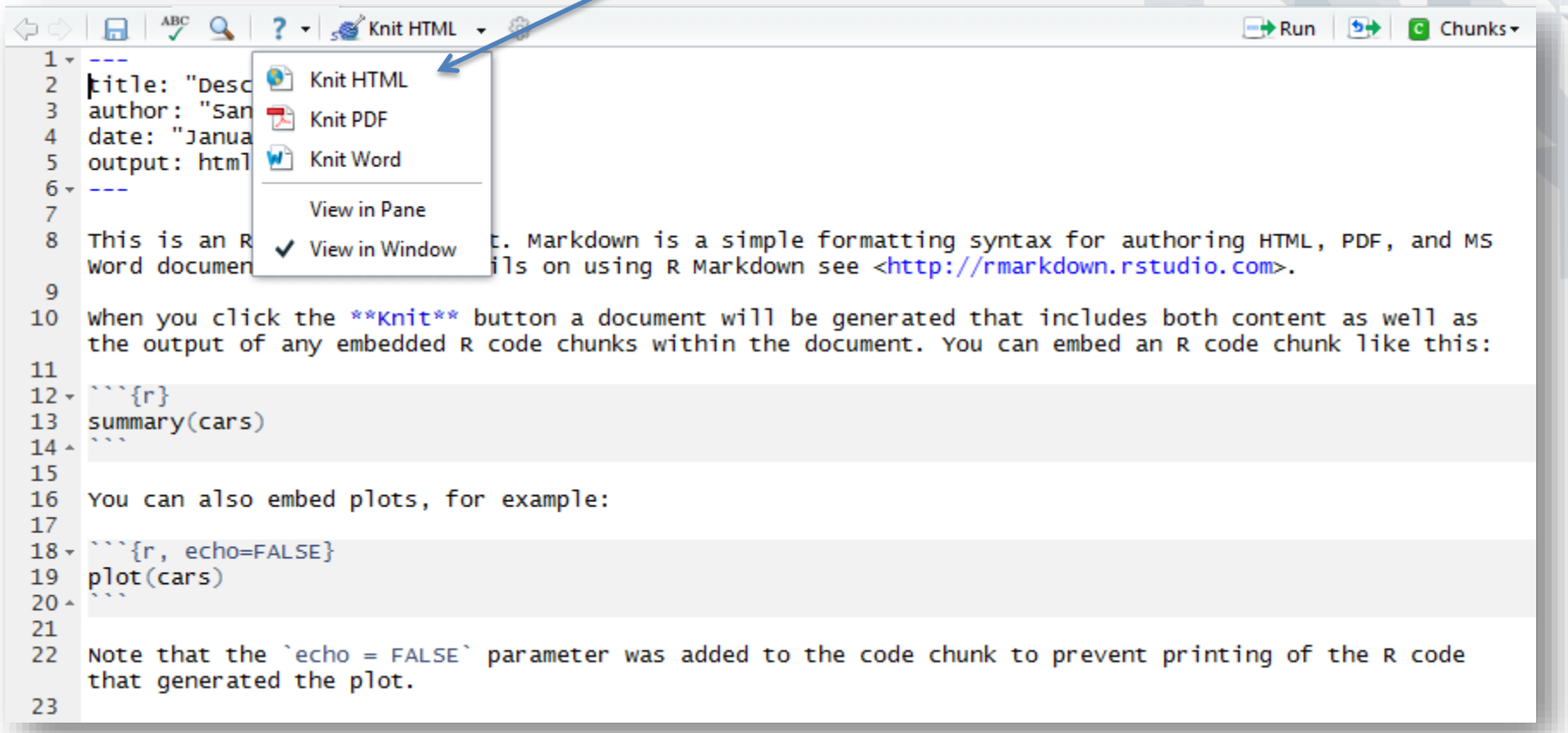
Using R Studio for Markdown

- Choose the type of the report you want to generate and R Studio will create an outline of the file for you.



Using R Studio for Markdown

- You can choose the rendering option from the Knit dropdown list as shown:



R Markdown Example: Desc_Stats_Grp.Rmd

R Markdown Code:

````
###Descriptive Measures
####Mean
The function used is mean(). It is also necessary to specify whether there are NA values in the input variable with the option *na.rm=TRUE*. By default **na.rm** is FALSE which will give NA as the output even if the input variable contains just one NA value

```{r}
Sales <- c(23000,50000,20000,NA,34000,78000,83000)
mean(Sales) # Not recommended

mean(Sales,na.rm = TRUE) # Recommended

```
````

R Markdown Output:

Descriptive Measures

Mean

The function used is mean(). It is also necessary to specify whether there are NA values in the input variable with the option *na.rm=TRUE*. By default **na.rm** is FALSE which will give NA as the output even if the input variable contains just one NA value

```
Sales <- c(23000,50000,20000,NA,34000,78000,83000)
mean(Sales) # Not recommended
```

```
## [1] NA
```

```
mean(Sales,na.rm = TRUE) # Recommended
```

```
## [1] 48000
```

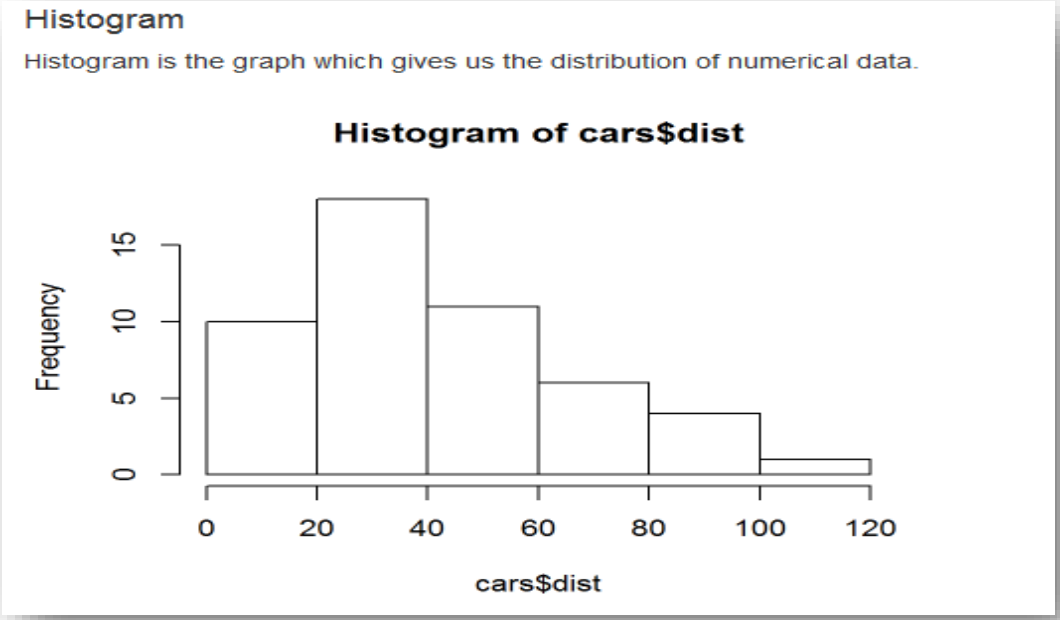
R Markdown Example: Desc_Stats_Grp.Rmd

R Markdown Code:

```
####Histogram
Histogram is the graph which gives us the distribution of numerical data.
```{r, echo=FALSE,  fig.width=5, fig.height=4}
hist(cars$dist)
```
```

echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot. fig.width and fig.height indicate dimensions of output

R Markdown Output:



Example:

R Markdown Code:

```
Regression output of Insurance data with `r n` records is given by:
```{r results='asis'}
library(xtable)
options(xtable.comment=FALSE)
print(xtable(sfit), type="html",html.table.attributes="border=1")
```
It's R^2^ = `r sfit$r.squared`.
```

Function xtable() is used to format the table output. Also note the use of R inline code such as `r n`

R Markdown Output:

```
library(xtable)
```

```
## Warning: package 'xtable' was built under R version 3.2.3
```

```
options(xtable.comment=FALSE)
print(xtable(sfit), type="html",html.table.attributes="border=1")
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-10084.2131	3504.1812	-2.88	0.0237
Home	167.3267	5.8955	28.38	0.0000
Automobile	54.1053	3.6559	14.80	0.0000

It's R^2 = 0.9967184.

Reference for R Markdown

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

Apps Sanjay

rmarkdown-cheatsheet 1 / 2

R Markdown Cheat Sheet

learn more at rmarkdown.rstudio.com

rmarkdown 0.2.50 Updated: 8/14

RStudio

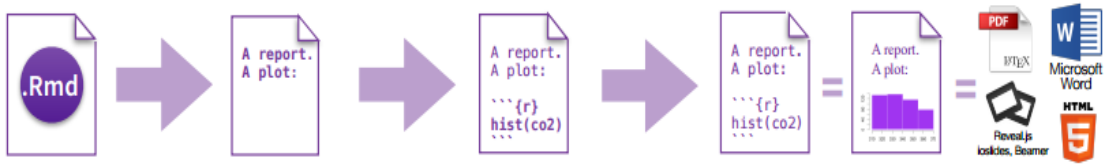
1. Workflow R Markdown is a format for writing reproducible, dynamic reports with R. Use it to embed R code and results into slideshows, pdfs, html documents, Word files and more. To make a report:

i. Open - Open a file that uses the .Rmd extension.

ii. Write - Write content with the easy to use R Markdown syntax

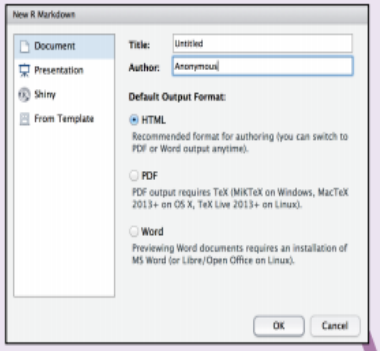
iii. Embed - Embed R code that creates output to include in the report

iv. Render - Replace R code with its output and transform the report into a slideshow, pdf, html or ms Word file.



2. Open File Start by saving a text file with the extension .Rmd, or open an RStudio Rmd template

- In the menu bar, click **File ▸ New File ▸ R Markdown...**
- A window will open. Select the class of output you would like to make with your .Rmd file
- Select the specific type of output to make with the radio buttons (you can change this later)
- Click OK



3. Markdown Next, write your report in plain text. Use markdown syntax to describe how to format text in the final report.

syntax

```
Plain text
End a line with two spaces to start a new paragraph.
*italics* and _italics_
**bold** and __bold__
superscript^2^
~~strikethrough~~
[link](www.rstudio.com)

# Header 1

## Header 2

### Header 3

#### Header 4
```

becomes

Plain text
End a line with two spaces to start a new paragraph.
italics and *italics*
bold and **bold**
superscript^2
~~strikethrough~~
link

Header 1
Header 2
Header 3
Header 4

Sane's STATS Academy of Statistics

RPubs

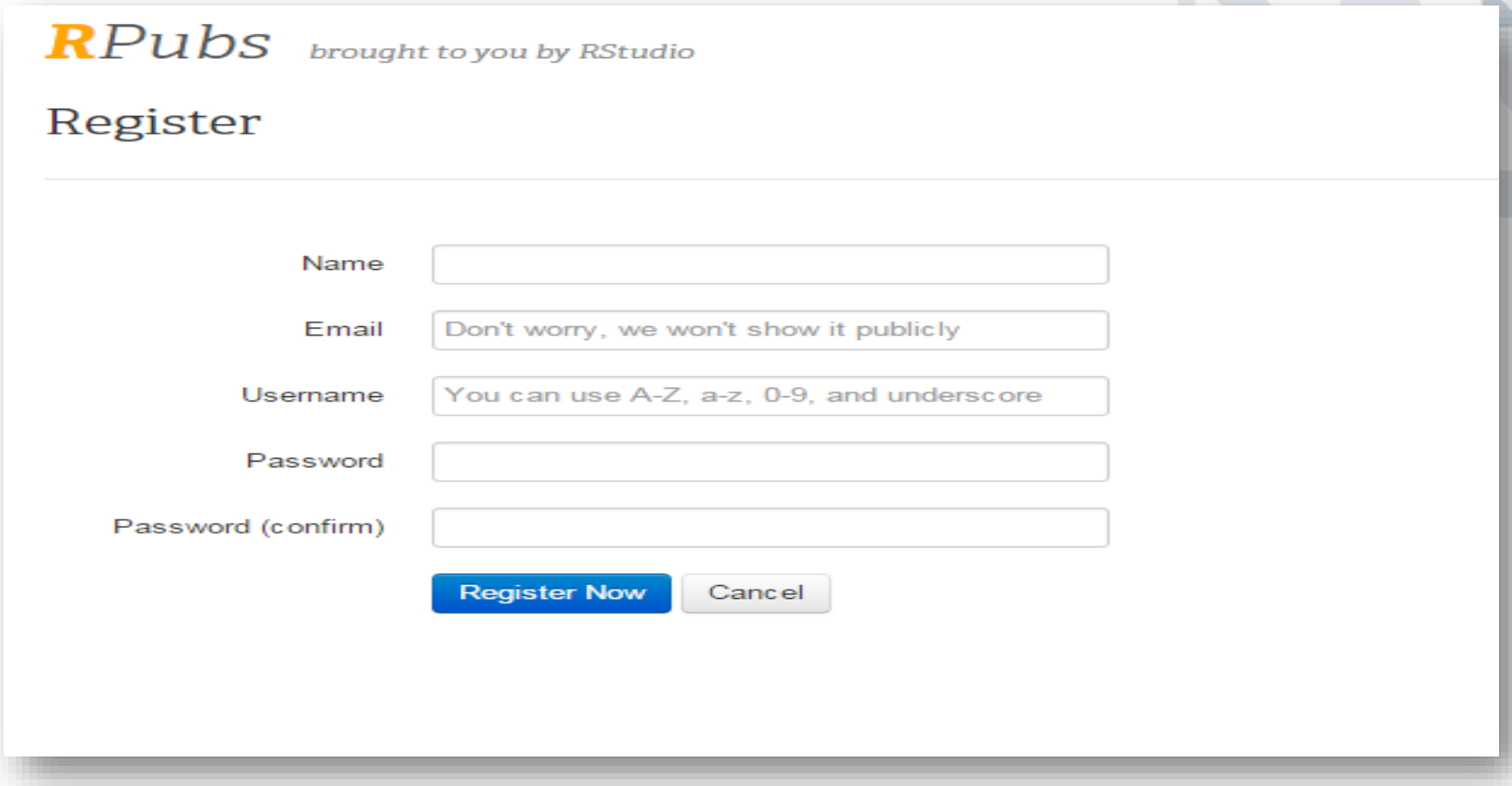
Sane's STATS Academy of Statistics

RPubs

- RPubs is a portal where you can share the R Markdown documents
- For sharing it requires you to be registered user

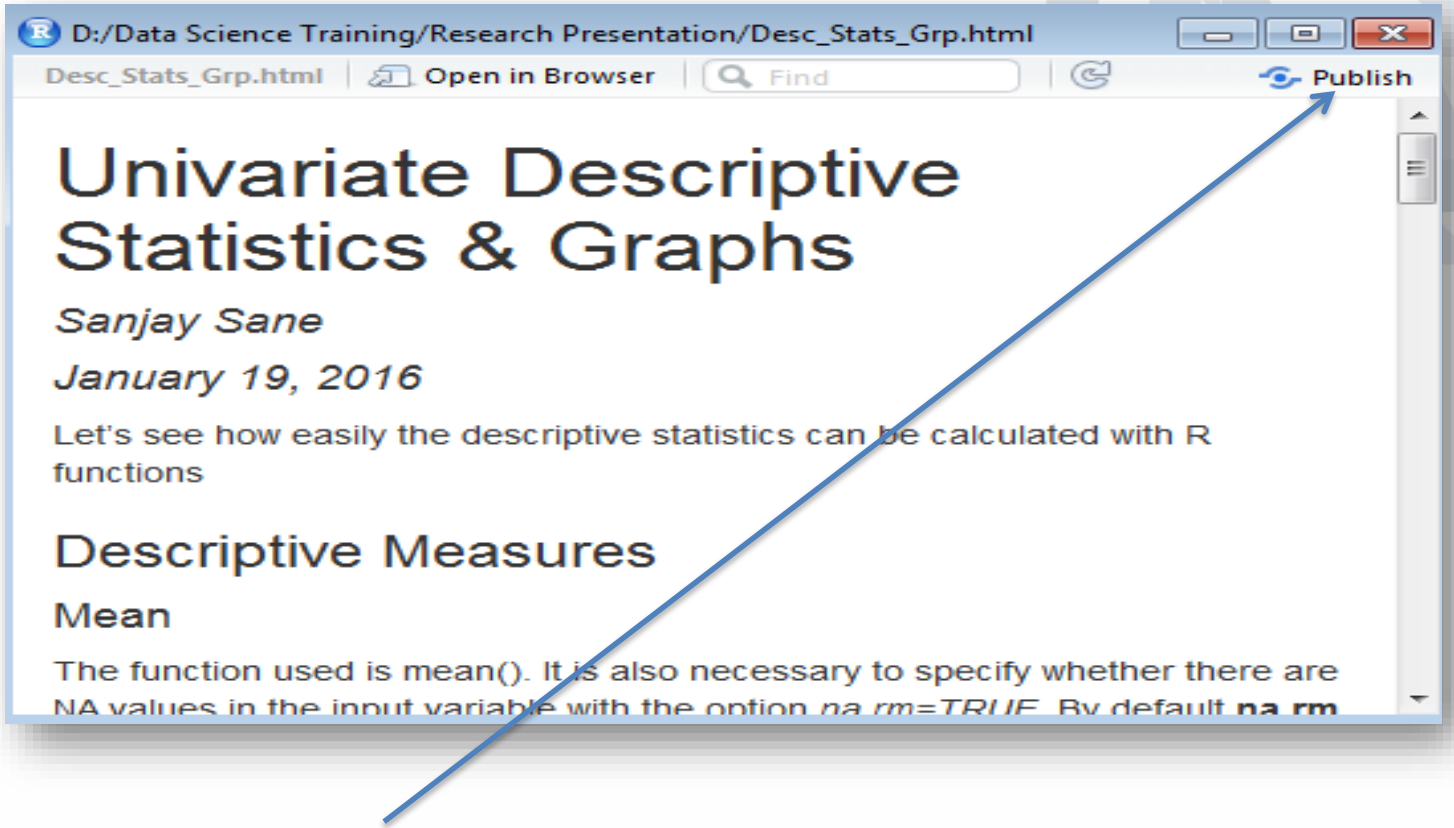
Steps to Publish a Document on RPubs

- Before publishing you need to register yourself on RPubs Website https://**rpubs**.com.



Steps to Publish a Document on RPubs

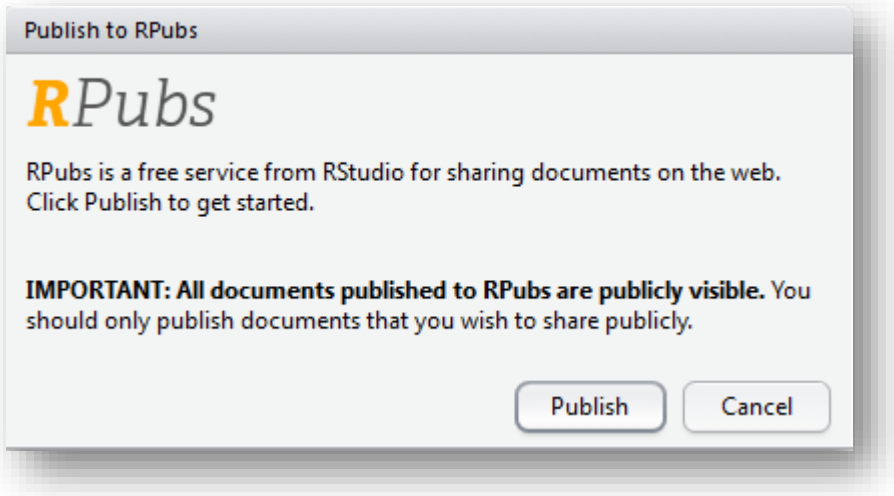
- Open the R Markdown document in output mode as shown



- Click on the *Publish* button on the top right corner

Steps to Publish a Document on RPubs

- You may be prompted to install some more packages if they aren't installed on your machine. Press Yes button to install them
- After installation of required packages we get the following window:



- Click on the *Publish* button on the dialog box