

Data Analysis and Visualization: R Workflow

Dr. Olga Scrivner, Research Scientist

January 16, 2018

Indiana University



CNS

Cyberinfrastructure for
Network Science Center

cns.iu.edu

Upcoming Events:

Monday, January 22, 2018 | 4:00 PM



Visual/Data Literacy Related to Maps

Theresa Quill

Indiana University Libraries

Monday, January 29, 2018 | 4:00 PM



Visualizing Science Using VOSviewer

Ludo Waltman

Centre for Science and Technology Studies (CWTS) at Leiden University

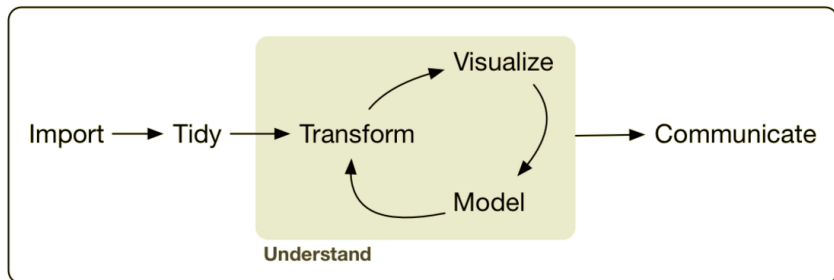
1. Understand RStudio
2. Understand the difference between R scripts and R projects
3. Learn how to plan and manage R project
4. Learn how to deploy and share R project

TBA:

- ⦿ R crash course
- ⦿ Shiny basics and Shiny advanced

Overview

1. Project set-up and planning
2. Reporting and documenting
3. Visualizing
4. Sharing

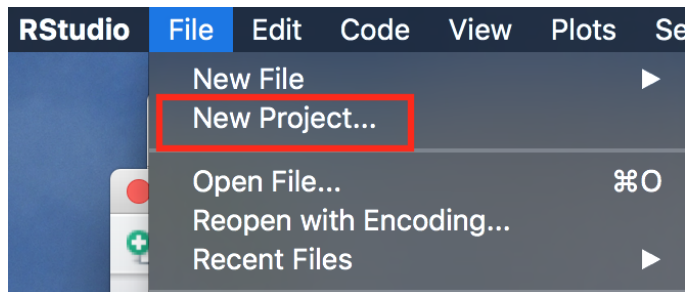


Project-oriented Workflow

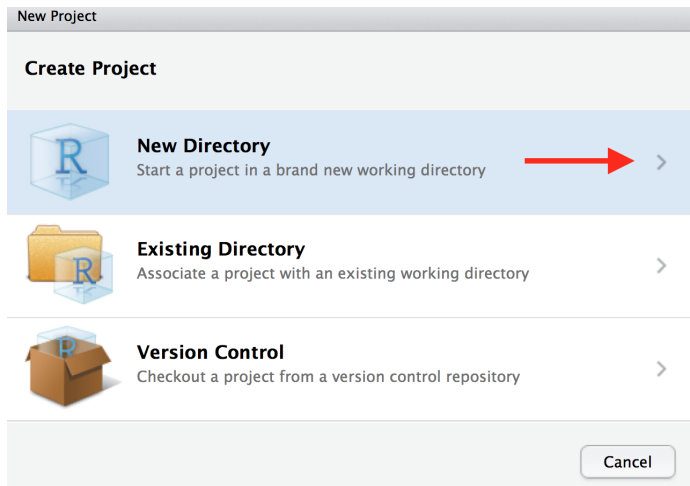


Organize each data into a **project**:

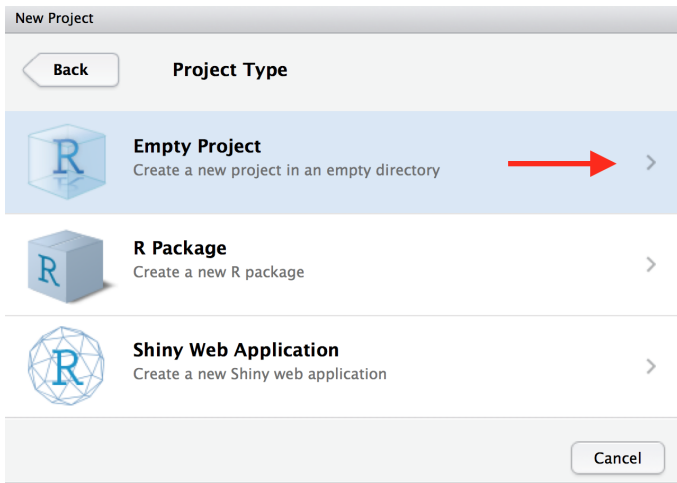
File → **New Project**



Create Project




Project Type



Project Directory

New Project

Back **Create New Project**



Directory name:

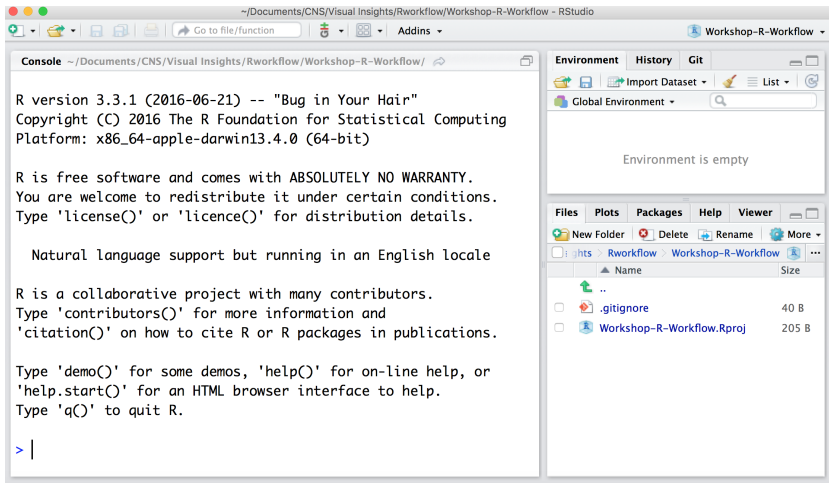
Create project as subdirectory of: **Browse...**

Create a git repository **Optional**

Use packrat with this project

Open in new session **Create Project** **Cancel**

New Project



~/Documents/CNS/Visual Insights/Rworkflow/Workshop-R-Workflow - RStudio

Go to file/function Addins Workshop-R-Workflow

Console ~/Documents/CNS/Visual Insights/Rworkflow/Workshop-R-Workflow/

```
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

Environment History Git

Import Dataset List

Global Environment

Environment is empty

Files Plots Packages Help Viewer

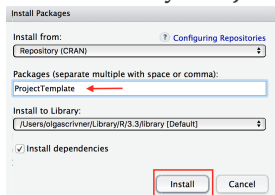
New Folder Delete Rename More

github Rworkflow Workshop-R-Workflow

| Name | Size |
|---------------------------|-------|
| .. | |
| .gitignore | 40 B |
| Workshop-R-Workflow.Rproj | 205 B |

Building a Directory Structure

1. Install library ProjectTemplate



2. Open a new R script: File → New File → R Script

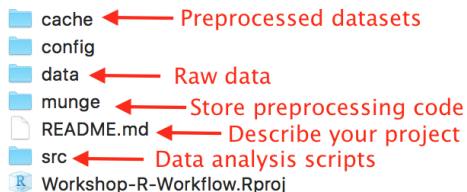
```
1 library('ProjectTemplate')
2
3 create.project('Project Name', minimal = TRUE,
4               merge.strategy = 'allow.non.conflict')
5
```

Use your project's name

3. TIP: Change your working directory - one level up from your project

Session → Set Working Directory → Choose Directory  CNS 10

Directory Structure



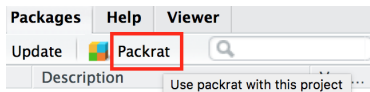
Extensions:

- ⊙ R script - **.R**
- ⊙ Readme file - **.md** (Markdown)
- ⊙ Project - **.Rproj**

Packrat - stores your package dependencies inside the project.

Advantages:

1. **Isolation:** Installing a new or updated package for one project will not break your other projects.
2. **Portability:** Easily transport your projects from one computer to another, even across different platforms.
3. **Reproducibility:** Packrat records the exact package versions you depend on.



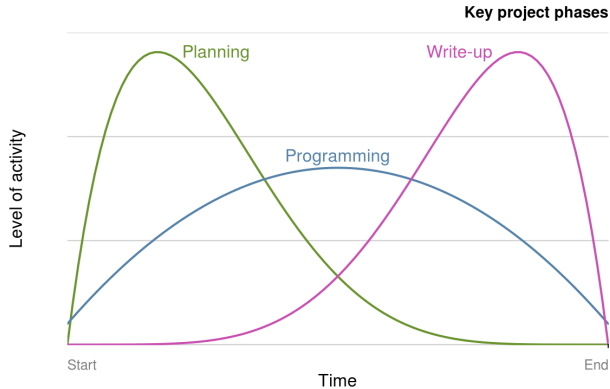
<http://rstudio.github.io/packrat/>

Project Planning



Project Planning

“smart preparation minimizes work” (Berkun, 2005)



<https://csgillespie.github.io/efficientR/workflow.html>

1. **Specific:** is the objective clearly defined and self-contained?
2. **Measurable:** is there a clear indication of its completion?
3. **Attainable:** can the target be achieved?
4. **Realistic:** have sufficient resources been allocated to the task?
5. **Time-bound:** is there an associated completion date or milestone?

<https://csgillespie.github.io/efficientR/workflow.html>

Gantt chart

```
library('DiagrammeR')  
# Define the Gantt chart and plot the result  
mermaid("gantt  
    Section Initiation  
    Planning           :a1, 2018-01-16, 10d  
    Data processing    :after a1 , 30d")
```



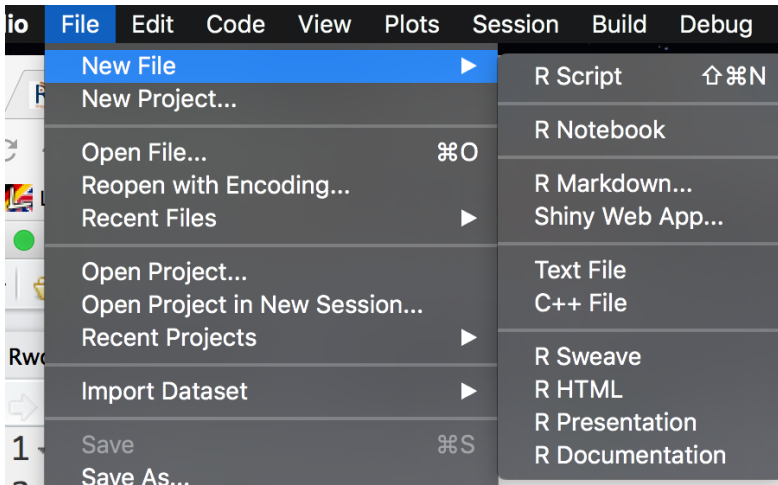
- ⊙ **Section** refers to the project's section (useful for large projects, with milestones)
- ⊙ **Line** refers to a task
- ⊙ Example: **Planning** begins on Jan 16 2018 and lasts for 10 days

Large projects: regular meetings, division of labour, tracking progress, issues and priorities (Gillespie and Lovelace, 2017, Chapter 4)

1. The interactive code sharing site **GitHub**
2. **ZenHub**, a browser plugin that is “the first and only project management suite that works natively within GitHub”
3. Web-based and easy-to-use tools such as **Trello**
4. Dedicated desktop project management software such as **ProjectLibre** and **GanttProject**
5. Fully featured, enterprise scale open source project management systems such as **OpenProject** and **redmine**

Documenting and Reporting

R Options



R Script

https:

The screenshot shows the RStudio environment with several annotations:

- Script:** Points to the source editor containing the following R code:

```
1 h <- "Hello"
2 yourname <- readline("what is your name?")
3 print(paste(h, yourname))
4
```
- Execute:** Points to the `Run` button in the toolbar.
- Source button:** Points to the `Source` button in the toolbar.
- Console results:** Points to the output in the console window:

```
> source("~/active-rstudio-document", echo=TRUE)
> h <- "Hello"
> yourname <- readline("what is your name?")
what is your name?Andrie
> print(paste(h, yourname))
[1] "Hello Andrie"
> |
```
- Workspace variables:** Points to the `Values` pane showing:

```
h      "Hello"
yourname "Andrie"
```
- Install libraries (packages):** Points to the `Packages` pane, which lists installed packages such as `braid`, `cleandata`, `conjoint`, `devtools`, `ggcluster`, `ggdendro`, `lats`, `lubridate`, and `mosaic`.

1. Copyright statement comment
2. Author comment (Use #)
3. File description comment, including purpose of program, inputs, and outputs
4. **source()** and **library()** statements
 - `source("file name")` - read R code from a file
 - `library(name)` - package name
5. Function definitions
6. Executed statements, if applicable (e.g., plot)

R Style Guide: Naming

File Names

| Good | Bad |
|-----------------------------------|--------------------|
| <code>predict_ad_revenue.R</code> | <code>foo.R</code> |

Variable Names

| Good | Bad |
|--|----------------------------|
| <code>variable.name</code> (preferred) | <code>variable_Name</code> |
| <code>variableName</code> (accepted) | |

Function Names (use action verbs)

| Good | Bad |
|---------------------------------|--|
| <code>CalculateAvgClicks</code> | <code>calculate_avg_clicks</code> <code>calculateAvgClicks</code> |

R Style Guide: Spacing and Assignment

- ⦿ Spaces around all binary operators (`=`, `+`, `-`, `<-`)
- ⦿ Use `<-`, not `=`, for assignment
- ⦿ Space after a comma

Incorrect: `total == sum(x[1,])`

R Style Guide: Spacing and Assignment

- ⦿ Spaces around all binary operators (=, +, -, <-)
- ⦿ Use <-, not =, for assignment
- ⦿ Space after a comma

Incorrect: `total == sum(x[1,])`

Correct: `total <- sum(x[1,])`

Learn more: <http://adv-r.had.co.nz/Style.html>

“R Markdown files are the ultimate R reporting tool”
(Grolemund, 2014)



R Markdown is a file format for making dynamic documents with R.

Markdown - an easy-to-write plain text format.

R Markdown files can be converted into HTML, PDF, and Word documents.

R Markdown

The screenshot displays the RStudio environment with two panes. The left pane shows the source R Markdown file, and the right pane shows the rendered HTML output.

Source R Markdown (Left Pane):

```
1 ---
2 title: "R markdown sample"
3 author: "Olga Scrivner"
4 date: "January 16, 2018"
5 output: html_document
6 ---
7
8 ## R Markdown
9
10 Click the **Knit** to generate the output of any embedded
11 R code chunks within the document.
12 You can embed an R code chunk like this:
13 ```{r cars}
14 summary(cars)
15 ```
```

Rendered HTML Output (Right Pane):

R markdown sample

Olga Scrivner
January 16, 2018

R Markdown

Click the **Knit** to generate the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

| ## | speed | dist |
|----|--------------|----------------|
| ## | Min. : 4.0 | Min. : 2.00 |
| ## | 1st Qu.:12.0 | 1st Qu.: 26.00 |
| ## | Median :15.0 | Median : 36.00 |
| ## | Mean :15.4 | Mean : 42.98 |
| ## | 3rd Qu.:19.0 | 3rd Qu.: 56.00 |
| ## | Max. :25.0 | Max. :120.00 |

Did You Know?

Markdown is used:

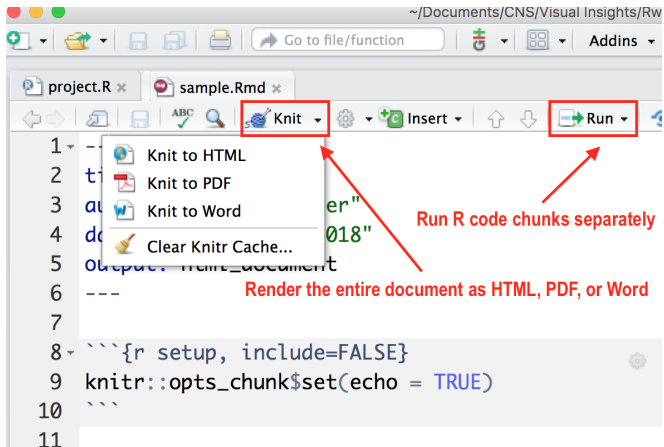
- ⦿ Github
- ⦿ StackOverflow
- ⦿ Reddit

R Markdown

| Text using Markdown syntax | Corresponding HTML produced by a Markdown processor | Text viewed in a browser |
|--|---|---|
| Heading ===== ## Sub-heading Paragraphs are separated by a blank line. Two spaces at the end of a line leave a line break. Text attributes <i>_italic_</i> , **bold** , <code>`monospace`</code> . Horizontal rule: --- Bullet list: * apples * oranges * pears Numbered list: 1. wash 2. rinse 3. repeat | <pre><h1>Heading</h1> <h2>Sub-heading</h2> <p>Paragraphs are separated by a blank line.</p> <p>Two spaces at the end of a line leave a line break.</p> <p>Text attributes italic, bold, <code>monospace</code>. </p> <p>Horizontal rule:</p> <hr /> <p>Bullet list:</p> apples oranges pears <p>Numbered list:</p> </pre> | <h2>Heading</h2> <hr/> <h3>Sub-heading</h3> <p>Paragraphs are separated by a blank line.</p> <p>Two spaces at the end of a line leave a line break.</p> <p>Text attributes <i>italic</i>, bold, monospace .</p> <hr/> <p>Bullet list:</p> <ul style="list-style-type: none">• apples• oranges• pears <p>Numbered list:</p> <ol style="list-style-type: none">1. wash2. rinse3. repeat <p>A link.</p> |

<https://en.wikipedia.org/wiki/Markdown>

Rendering



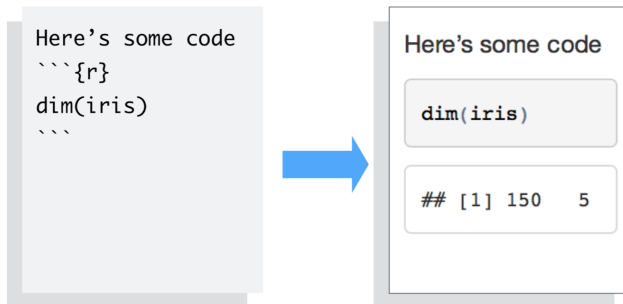
The screenshot shows the RStudio interface with the Knit menu open. The menu options are: Knit to HTML, Knit to PDF, Knit to Word, and Clear Knitr Cache... The Run button is also visible in the toolbar. Red arrows point from the text annotations to the Knit menu and the Run button.

Run R code chunks separately

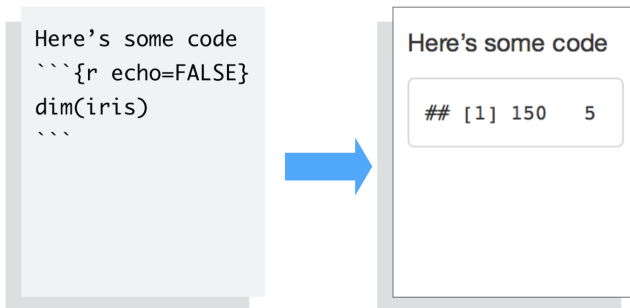
Render the entire document as HTML, PDF, or Word

```
1 -
2 t
3 al
4 d
5 O
6 ---
7
8 - ````{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ````
11
```

<https://en.wikipedia.org/wiki/Markdown>



<http://shiny.rstudio.com/articles/rmarkdown.html>

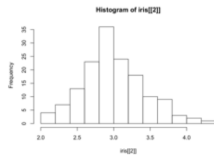


<http://shiny.rstudio.com/articles/rmarkdown.html>

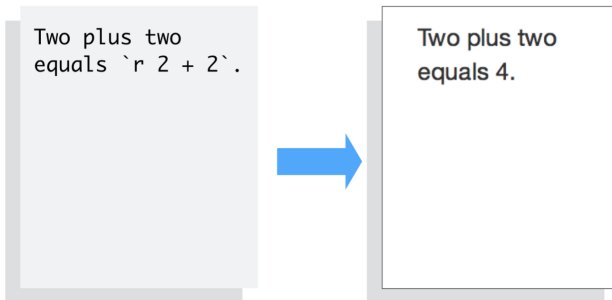

```
Here's a plot  
```\r echo=FALSE}  
hist(iris[[2]])
```\
```



Here's a plot



R Chunks - Inline




<http://shiny.rstudio.com/articles/rmarkdown.html>

Gallery: Get Inspiration

Websites


R Markdown makes it easy to build webpages straight from .Rmd files.



The screenshot shows the R Markdown website. It features a central diagram with the R logo in the middle, surrounded by icons for various services and tools like GitHub, Docker, and RStudio. Below the diagram, there are sections for 'Getting Started' and 'Learning More'.

R Markdown


The R [Markdown website](#) is itself built with R Markdown. [Example Code.](#)



The screenshot shows the flexDashboard website. It displays a grid of various dashboard examples, including maps, charts, and data visualizations.

flexDashboard

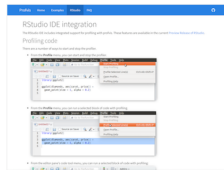
flexdashboard extends R Markdown to make administrative dashboards. Its website is also built from R Markdown. [Example Code.](#)



The screenshot shows the bookdown website. It features a list of books created with bookdown, each with a cover image and a brief description.

bookdown

Bookdown extends R Markdown to make books. Its website is built with R Markdown and CSS styling. [Example Code.](#)



The screenshot shows the profvis website. It displays several screenshots of the RStudio IDE with profiling tools overlaid on the code editor.

profvis

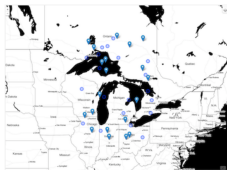
profvis provides profiling tools for R code, as well as a website made with R Markdown. [Example Code.](#)

<http://rmarkdown.rstudio.com/gallery.html>

Gallery: Get Inspiration

Interactive Documents

Combine R Markdown with htmlwidgets or the shiny package to make interactive documents.



HTML Widgets

Add interactive graphics with htmlwidgets, such as the leaflet map widget.

UNCCC Data Report
March 2016

- 1 Disclaimer
- 2 Data Overview

1 Disclaimer

This report is prepared for the purpose of providing information and does not constitute an offer of investment or any other financial product. The information contained herein is for informational purposes only and should not be used as a basis for investment decisions. The user assumes full responsibility for any loss or damage resulting from the use of the information contained herein.

2 Data Overview

| Year | Value |
|------|-------|
| 2010 | 100 |
| 2011 | 105 |
| 2012 | 110 |
| 2013 | 115 |
| 2014 | 120 |
| 2015 | 125 |

HTML Widgets

Embed htmlwidgets such as dygraphs and datatables directly into your reports.

Shiny leaflet example

Version 0 - Use observers

```
observe({
  # Update the map with the selected location
  map.setView(selectedLocation$lat, selectedLocation$lon, 15)
})
```

Shiny

Add interactive analysis with shiny, which lets your user rerun the actual analysis within your report.



Shiny

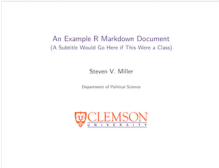
Shiny components and htmlwidgets will work in any HTML based output, such as a file, slide show or dashboard.

<http://rmarkdown.rstudio.com/gallery.html>

Gallery: Get Inspiration


Presentations

R Markdown supports several presentation (slide show) formats.




An Example R Markdown Document
(A Subtitle Would Go Here if This Were a Class)


Steven V. Miller
Department of Political Science




Beamer slideshow
Create pdf slides with Beamer. Example Code.



Yelp.com Revenue



Slidy slideshow
Create HTML-based slides with Slidy.



Shiny

Debugging with Shiny

Jonathan McPherson

ioslides slideshow
Create HTML-slides with ioslides. Example Code



Principles of Reactivity
Joe Cheng <@joe@rstudio.com>
#ShinyDevConf - January 30, 2016

reveal.js slideshow
Create HTML-based slides with reveal.js. Example Code

<http://rmarkdown.rstudio.com/gallery.html>

Gallery: Get Inspiration

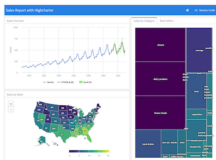
Dashboards

Combine R Markdown with the flexdashboard package to quickly assemble R components into administrative dashboards. Each example below contains a link to the source code within the dashboard.



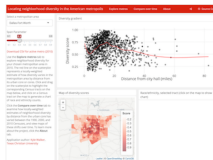
Dashboard with gauges and value boxes

Use flexdashboard to create dashboards with gauges and value boxes.



Dashboard with htmlwidgets

Add interactive graphics to a dashboard with htmlwidgets.



Dashboard with Shiny

Add interactive analysis to a dashboard with Shiny.



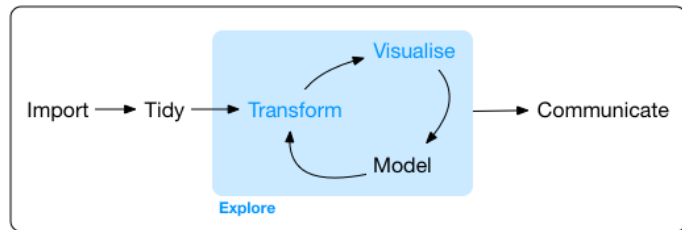
Dashboard with storyboard

Organize dashboards around a storyboard.

<http://rmarkdown.rstudio.com/gallery.html>

Data Analysis and Visualization Work- flow

Tidyverse Workflow



Program

1. **Import** data into R: `read_csv()`, `read_line()`, `read_delim()`
2. **Tidy** data - variables per column, observation per row
3. **Transform** with `dplyr`
4. **Visualize** with `ggplot` and `plotly`

(Wickham and Grolemund, 2017)

- ⊙ Pick observations by their values - **filter()**
- ⊙ Reorder the rows - **arrange()**
- ⊙ Pick variables by their names -**select()**
- ⊙ Create new variables with functions of existing variables
mutate()
- ⊙ Collapse many values down to a single summary
summarise()

(Wickham and Grolemund, 2017, Chapter 5)

Practice: <http://r4ds.had.co.nz/transform.html>

Recommended Reading -

<http://vita.had.co.nz/papers/layered-grammar.pdf>

Visualization Template

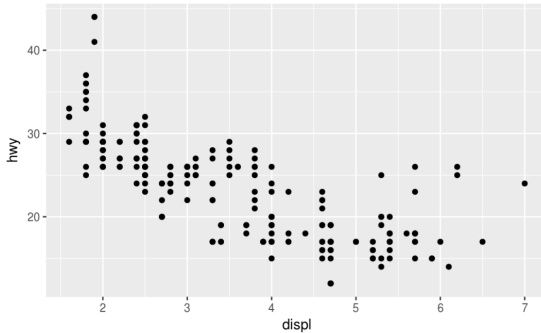
```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

Visualization Template

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy))
```

geom_function

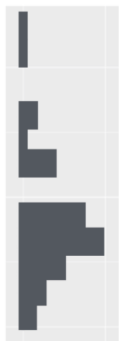
mapping



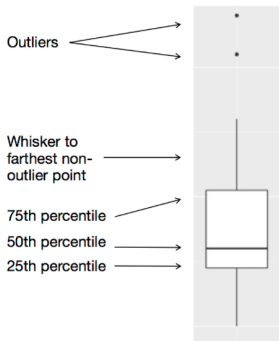
Exploratory Data Visualization



geom_point()



geom_bar()



geom_boxplot()

<http://r4ds.had.co.nz/exploratory-data-analysis.html>

Interactive Visualization - Plotly

```
install.packages("plotly")  
library(plotly)
```



plotly

Developer Support

PLOTCON

Consulting

Help

API Libraries

Ggplot2

Quick Start

Getting Started

User Guide

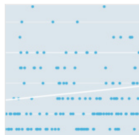
Examples

Basic

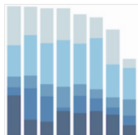
Statistical

Animations

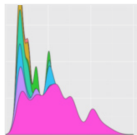
Layout Options



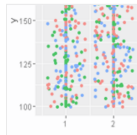
geom_abline



geom_bar



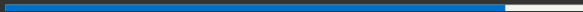
geom_density



geom_jitter

<https://plot.ly/ggplot2/>

Publishing



Publishing

The screenshot displays the RStudio interface with two panes. The left pane shows the source code for 'sample.Rmd':

```
1 ---
2 title: "R markdown sample"
3 author: "Olga Scrivner"
4 date: "January 16, 2018"
5 output: html_document
6 ---
7
```

The right pane shows the rendered HTML output. The title is 'R markdown', the author is 'Olga Scrivner', and the date is 'January 16, 2018'. Below this is a section titled 'R Markdown' with a paragraph of text: 'Click the Knit to generate the output of any embedded R code chunks within the document. You can embed an R code chunk like this:'. A context menu is open over the rendered output, showing 'Publish Document...' and 'Manage Accounts...' options.

Publishing

The image displays two sequential screenshots of the RStudio publishing interface.

Left Screenshot: Publish To

- Publish To**
- RPubs**
RPubs is a free service from RStudio for sharing documents on the web.
- RStudio Connect**
RStudio Connect is a server product from RStudio for secure sharing of applications, reports, and plots.
- Cancel** button at the bottom.

Right Screenshot: Publish to RPubS

- Back** button at the top left.
- Publish to RPubS** header.
- RPubs logo and text: "RPubs is a free service from RStudio for sharing documents on the web. Click Publish to get started."
- IMPORTANT: All documents published to RPubS are publicly visible. You should only publish documents you wish to share publicly.**
- Publish** and **Cancel** buttons at the bottom right.

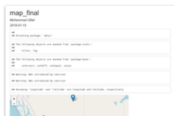
Easy web publishing from R

Write R Markdown documents in RStudio.

Share them here on RPubs. (It's free, and couldn't be simpler!)

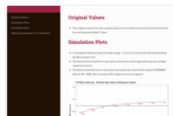
Get Started

Recently Published



A screenshot of an R Markdown document titled "서브윌 인더벤션 데이터 분석결과". The document displays a summary table with columns for "Year", "Age", "Gender", "Marital Status", "Education", "Income", "Health", and "Smoking". The table contains data for the years 2008, 2009, and 2010.

| Year | Age | Gender | Marital Status | Education | Income | Health | Smoking |
|------|-----|--------|----------------|-------------|--------|--------|---------|
| 2008 | 25 | Male | Married | High School | 30000 | Good | Yes |
| 2009 | 26 | Female | Single | College | 40000 | Good | No |
| 2010 | 27 | Male | Married | High School | 35000 | Good | Yes |



General Information

E583 | Z637 | Information Visualization MOOC 2018

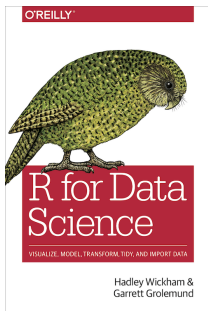
This graduate level course provides an overview of the state of the art in information visualization. The course teaches visualization theory and the process of producing effective and actionable visualizations that take the needs of users into account. Students apply the visualization knowledge and skills that they gain in the course by working in teams on real-world client projects.

CONCURRENT

SELF-PACED

Data Visualization Literacy

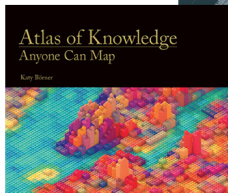
References



Börner, Katy and Ted Polley. 2014.
Visual Insights. The MIT Press.
<http://cns.iu.edu/ivmooocbook14.html>



Börner, Katy. 2015. *Atlas of Knowledge. Anyone Can Map*. The MIT Press.
<http://scimaps.org/atlas2>



THE
END

@katycns

@obscrivn

#IVMOOC