Interactive Visualization with R

SAGE campus

Martin John Hadley
@martinjhnhadley

Slides and code available here:
https://goo.gl/Nxc3Ww

Interactive Visualization with R for Social Sciences

Martin John Hadley
@martinjhnhadley

Martin John Hadley
@martinjhnhadley

UNIVERSITY OF OXFORD

Interactive Data Network

idn.it.ox.ac.uk

Martin John Hadley
@martinjhnhadley

UNIVERSITY OF OXFORD

Interactive Data Network

idn.it.ox.ac.uk

Martin John Hadley
@martinjhnhadley

UNIVERSITY OF OXFORD

Interactive Data Network

idn.it.ox.ac.uk

OxShef: dataviz

oxshef.netlify.com

Martin John Hadley
@martinjhnhadley

UNIVERSITY OF OXFORD

Interactive Data Network

idn.it.ox.ac.uk

OxShef: dataviz

oxshef.netlify.com

Why visualise data?

What is a reproducible dataviz workflow?

How can we interactively visualise social sciences data with R?

Martin John Hadley
@martinjhnhadley

Why visualise data?

Research data is too often born and buried in a table

“Most [research] data is created in a form and organization that facilitates its generation rather than focusing on its eventual use.”

Table 2. Successes and failures for articles with non-zero metric scores, aggregated by journal, and only including journals for which there it is at least one success or failure.

Metric+	Mostly success	Mostly failure	Z	Equal	Journals
Tweets**	1097 (58%)	646 (34%)	10.8	148 (8%)	1891
**	1032 (59%)	586 (33%)	11.1	139 (8%)	1757
FbWalls**	414 (53%)	282 (36%)	5.0	86 (11%)	782
**	308 (55%)	188 (34%)	5.4	62 (11%)	558
RH	276 (51%)	221 (41%)	2.5	47 (9%)	544
	193 (51%)	157 (41%)	1.9	30 (8%)	380
Blogs**	190 (58%)	104 (32%)	5.0	32 (10%)	326
**	129 (57%)	70 (31%)	4.2	26 (12%)	225
Google+	61 (50%)	53 (44%)	0.7	7 (6%)	121
	25 (48%)	24 (46%)	0.1	3 (6%)	52
MSM	29 (56%)	17 (33%)	1.8	6 (12%)	52
	13 (52%)	9 (36%)	0.9	3 (12%)	25
Reddits	22 (51%)	17 (40%)	0.8	4 (9%)	43
	9 (47%)	7 (37%)	0.5	3 (16%)	19
Forums	5 (83%)	1 (17%)	1.6	0 (0%)	6
	3 (100%)	0 (0%)	1.7	0 (0%)	3
Q&A	4 (67%)	1 (17%)	1.3	1 (17%)	6
	2 (67%)	0 (0%)	1.4	1 (33%)	3
Pinners	2 (67%)	1 (33%)	0.6	0 (0%)	3
	0 (−%)	0 (−%)	-	0 (−%)	0
LinkedIn	0 (−%)	0 (−%)	-	0 (−%)	0
	0 (−%)	0 (−%)	-	0 (−%)	0

+ In each cell the upper figure is for all journals and the lower figure is for journals with at least 10 articles tested. * Ratio of successes to failures significantly different from 0.5 at $p = 0.05$, ** Significant at $p = 0.01$; both Bonferroni corrected for $n = 11$.
doi:10.1371/journal.pone.0064841.t002

Fox, P., & Hendler, J. (2011). Changing the Equation on Scientific Data Visualization. Science, 331(6018), 705–708. doi.org/10.1126/science.1197654

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. PloS One, 8(5), e64841. doi.org/10.1371/journal.pone.0064841

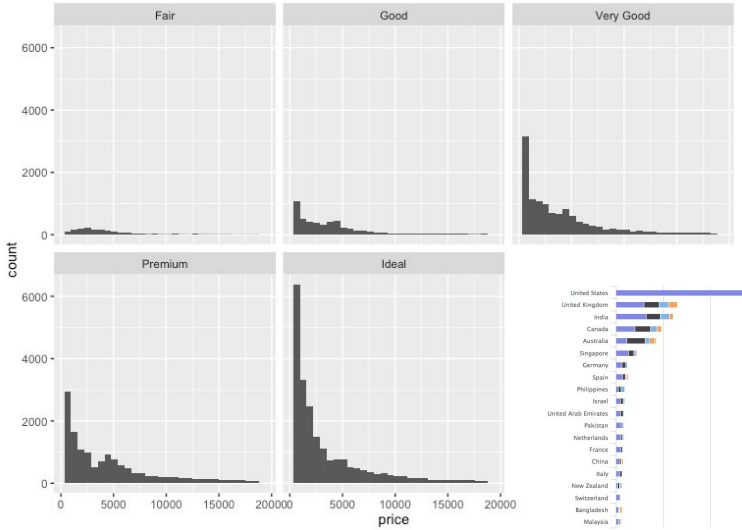
Why visualise data?

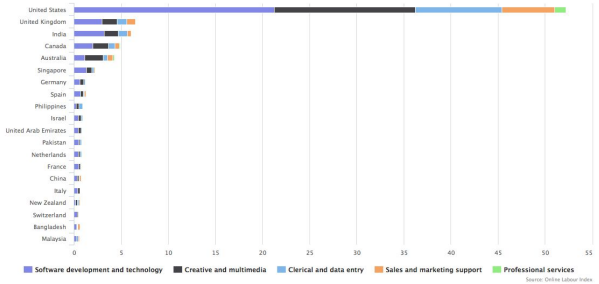
Exploratory data analysis

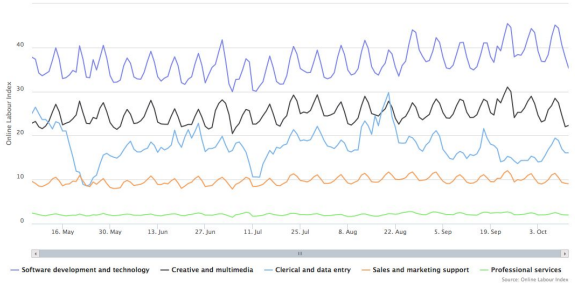
Summarise trends in an easily consumable manner

Physically demonstrate comparisons between groups of data

Present connections otherwise difficult to communicate







Moving beyond dead trees

“... interactivity is the new colour chart…”

(cc) BY

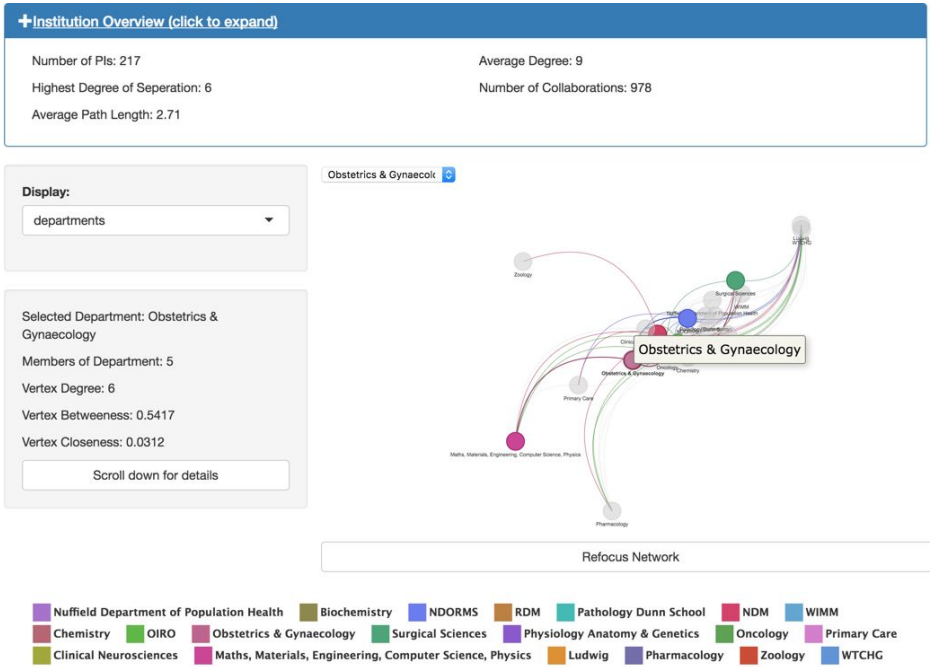
@photosteve101 (flickr.com/photos/42931449@N07/5263540555/in/album-72157625505293849/)

What does interactivity provide?

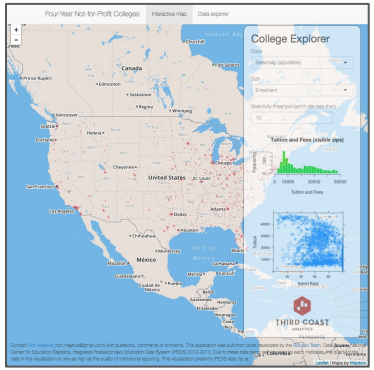
Provide alternative methods to access data

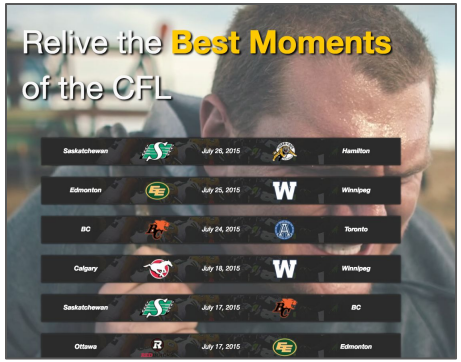
Allow users to slice through datasets

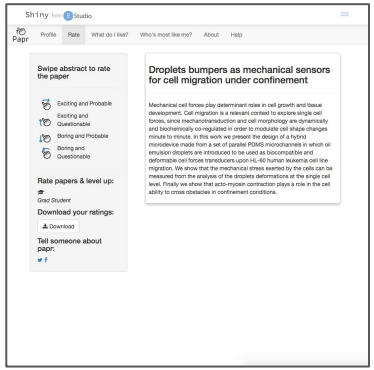
Combine summary and detailed information



Shiny allows interactive web applications to be built entirely with R (no HTML, CSS or JavaScript)







Shiny on your local machine

```
> install.packages("shiny")
```

Shiny on a server (for others to use)

shinyapps.io

Fully hosted solution for Shiny apps

(Includes a free tier!)

Shiny Server

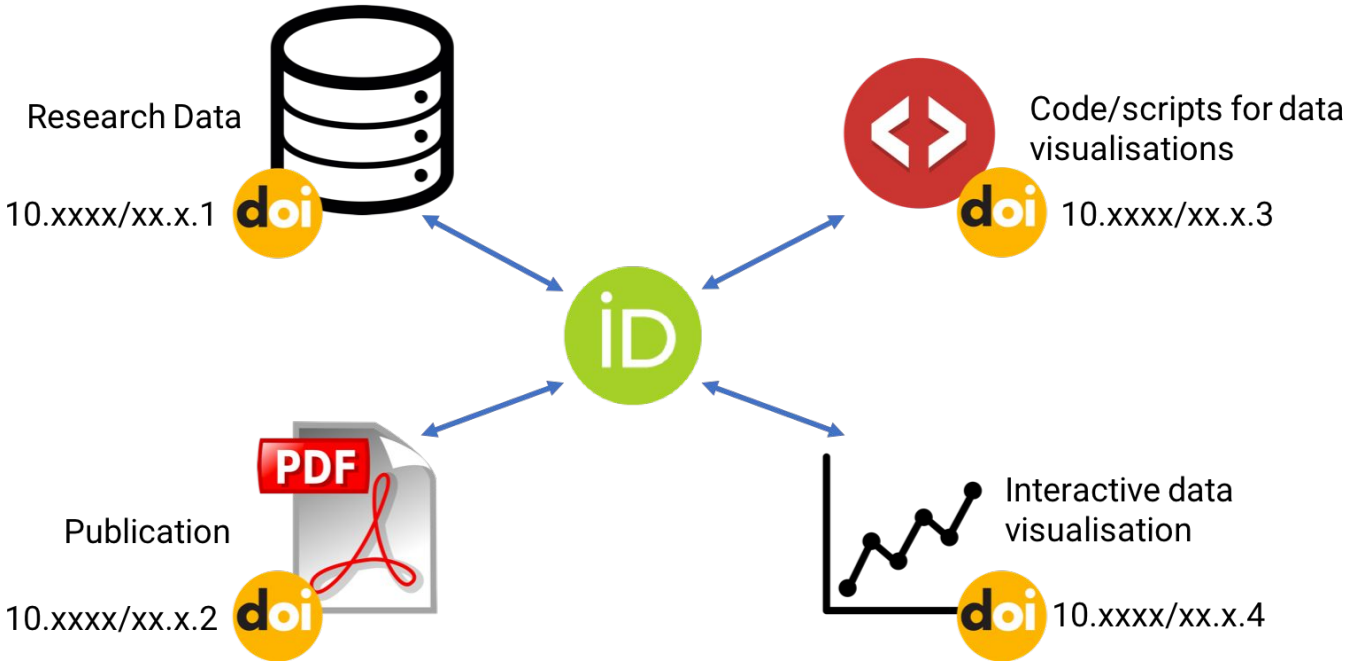
Shiny Server
Open Source

Shiny Server Pro

(Commercial License)

What is a reproducible dataviz workflow?

Reproducible dataviz workflow



Reproducible dataviz workflow

RStudio

Projects

Reproducible examples

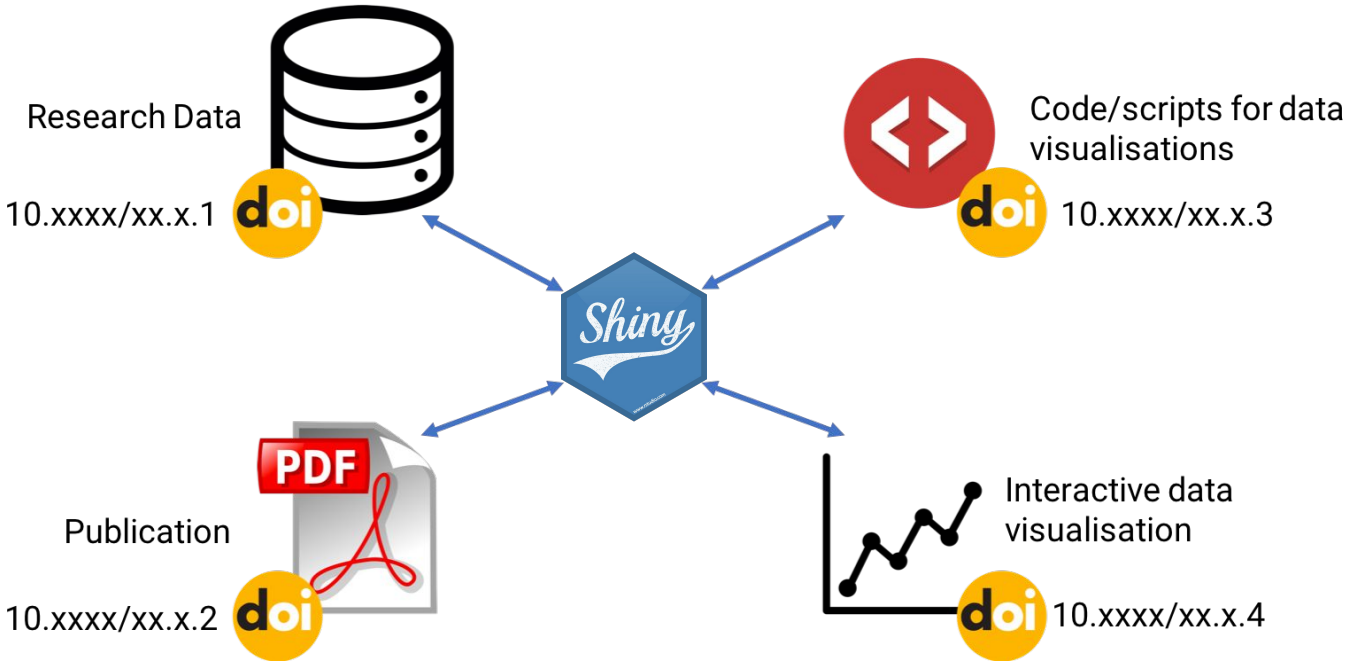
Reproducible dataviz workflow

RStudio

Projects

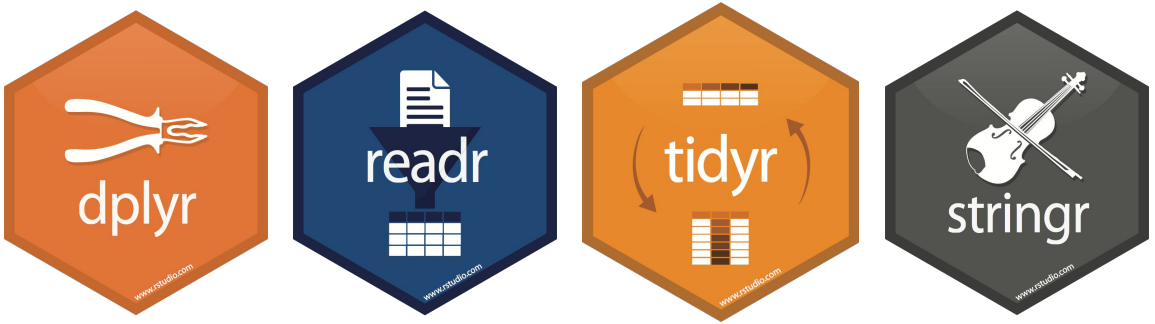
Reproducible examples

Reproducible dataviz workflow

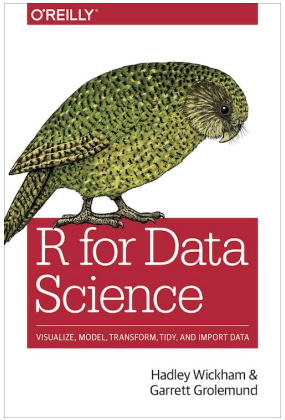


How can we interactively visualise social sciences data with R?

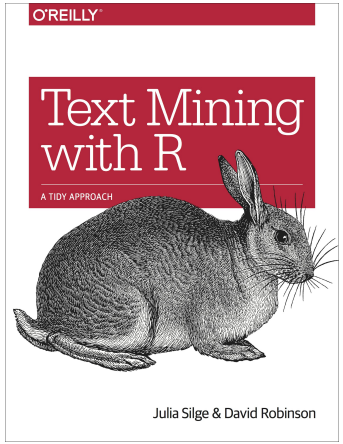
The tidyverse is a collection of R packages that make data analysis, modelling, visulalisation and comunication **as smooth as possible**



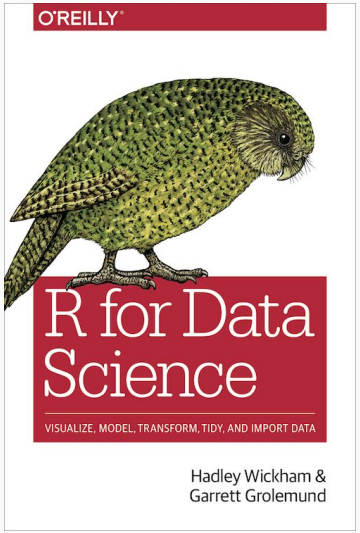
The R for Data Science book introduces the philosophy of ***tidy data*** and a consistent approach to working with data.



Specialised packages like *tidytext* provide a workflow that fits neatly into the tidyverse.



Where can I learn more?



#rstats

R Studio Community

Martin John Hadley
@martinjhnhadley

Slides and code available here:
https://goo.gl/Nxc3Ww