Reproducible Dataviz Workflows with R

BristolR

Martin John Hadley
@martinjhnhadley

Download today's code from [xxx]

Martin John Hadley
@martinjhnhadley

UNIVERSITY OF OXFORD

Interactive Data Network

idn.it.ox.ac.uk

OxShef: dataviz

oxshef.io

VISIBLE DATA

visibledata.co.uk

Interactive Data Network

idn.it.ox.ac.uk

- Promote reproducible research workflows and good research data management
- Support researchers in using interactive dataviz to promote Open Data
- Host interactive dataviz built by researchers at Oxford (currently using shinyapps.io)

OxShef: dataviz

oxshef.io

- A very young collaboration between University of Oxford and University of Sheffield
- Open Source/Access resources for all researchers on reproducible data visualisation workflows
- OxShef: dataviz does not seek to reinvent the wheel or build new tools - we consolidate material and aim to answer questions researchers actually ask..

OxShef: dataviz aims to answer four common questions

Which chart should I use?

charts.oxshef.io provides advice on how to select the most appropriate charts for your data and how to avoid common mistakes in data visualisation. Subjects covered include:

- Charts (barcharts, piecharts, bubble charts)
- Maps (scattergeo, choropleth, meteorological charts)
- Time series (financial charts, time series)
- Networks

OxShef: dataviz aims to answer four common questions

Which chart should I use?

charts.oxshef.io provides advice on how to select the most appropriate charts for your data and how to avoid common mistakes in data visualisation. Subjects covered include:

- Charts (barcharts, piecharts, bubble charts)
- Maps (scattergeo, choropleth, meteorological charts)
- Time series (financial charts, time series)
- Networks

Which dataviz tool should I use?

tools.oxshef.io provides an overview of the most popular tools used in both academia and industry for building interactive dataviz.

There are dedicated sites for dataviz tools that support a *reproducible data visualisation workflow*, currently this includes; shiny.oxshef.io.

OxShef: dataviz aims to answer four common questions

Which chart should I use?

charts.oxshef.io provides advice on how to select the most appropriate charts for your data and how to avoid common mistakes in data visualisation. Subjects covered include:

- Charts (barcharts, piecharts, bubble charts)
- Maps (scattergeo, choropleth, meteorological charts)
- Time series (financial charts, time series)
- Networks

Where can I deposit datasets?

repositories.oxshef.io provides an overview of the differences between file-sharing platforms like Dropbox, GoogleDrive, OneDrive and DOI-issuing data repositories.

Visit repositories.oxshef.io for advice on the most appropriate data repository for your datasets and how to use them in a *reproducible data visualisation workflow*.

Which dataviz tool should I use?

tools.oxshef.io provides an overview of the most popular tools used in both academia and industry for building interactive dataviz.

There are dedicated sites for dataviz tools that support a *reproducible data visualisation workflow*, currently this includes; shiny.oxshef.io.

OxShef: dataviz aims to answer four common questions

Which chart should I use?

charts.oxshef.io provides advice on how to select the most appropriate charts for your data and how to avoid common mistakes in data visualisation. Subjects covered include:

- Charts (barcharts, piecharts, bubble charts)
- Maps (scattergeo, choropleth, meteorological charts)
- Time series (financial charts, time series)
- Networks

Which dataviz tool should I use?

tools.oxshef.io provides an overview of the most popular tools used in both academia and industry for building interactive dataviz.

There are dedicated sites for dataviz tools that support a *reproducible data visualisation workflow*, currently this includes; shiny.oxshef.io.

Where can I deposit datasets?

repositories.oxshef.io provides an overview of the differences between file-sharing platforms like Dropbox, GoogleDrive, OneDrive and DOI-issuing data repositories.

Visit repositories.oxshef.io for advice on the most appropriate data repository for your datasets and how to use them in a *reproducible data visualisation workflow*.

Do publishers care about dataviz?

Academic publishers are seeking to differentiate themselves technologically, and it's clear that embedding dataviz alongside journal articles is an active area of interest.

For a list of publishers who support embedding dataviz please visit publishers.oxshef.io. This site also provides advice on best practices for linking to dataviz that are not hosted by the publisher.

Reproducible dataviz workflow requirements

- ✓ All data shown in the dataviz is public and available via a DOI
- ✓ All code for the dataviz is public and available via a DOI
- ✓ All authors of the code and data are attributed by an ORCID

Digital Object Identifiers (DOI)

Digital Object Identifiers (DOI)

Digital Object Identifiers (DOI) are the alphanumeric codes often seen in citations for academic papers, eg DOI:10.1126/science.1102896. (slide 1/4)

Digital Object Identifiers (DOI)

DOI are incredibly important (and useful) because they provide a permanent link to a resource (usually a PDF) which should never break. (slide 2/4)

Digital Object Identifiers (DOI)

DOI are issued by academic publishers and data repositories. Read more at repositories.oxshef.io. (slide 3/4)

Digital Object Identifiers (DOI)

Commercial useRs have sadly not embraced the DOI.

If you care about your stuff being available freely, easily and permanently you need a DOI.

GitHub is not good enough. (slide 4/4)

Digital Object Identifiers (DOI)

Commercial useRs have sadly not embraced the DOI.

If you care about your stuff being available freely, easily and permanently you need a DOI.

GitHub is not good enough. (slide 4/4)

Deposit data, code and publications separately

Data (and code) are fundamental components of the majority of research in the sciences and increasingly often in the humanities too. However, they are often relegated to “supplementary materials”.

Publications, data and code should all be considered 1st rate research outputs.

- ✓ rfigshare is an **excellent** package for creating/editing/publishing content to figshare.com
- ✓ zenodo provides basic functionality for depositing content to zenodo.org

ORCID

ORCID

Uniquely identifying humans by their names, handles and usernames is **hard**.

ORCID is a ready made solution from academia - it automatically builds a publication history for you, e.g. 0000-0002-3039-6849

Register for FREE and use it everywhere.

Martin John Hadley
@martinjhnhadley

UNIVERSITY OF OXFORD

Interactive Data Network

idn.it.ox.ac.uk

OxShef: dataviz

oxshef.io

VISIBLE DATA

visibledata.co.uk

DATAVIZ FIRST

Why visualise data?

~~Researeh~~ 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

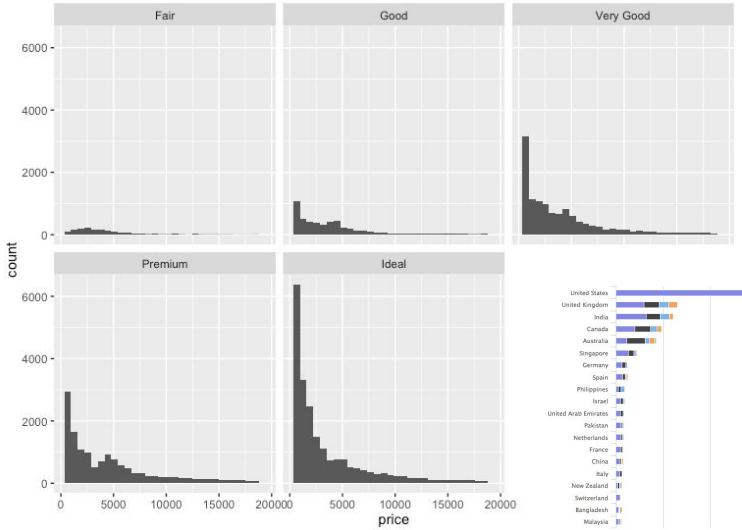
Why visualise data?

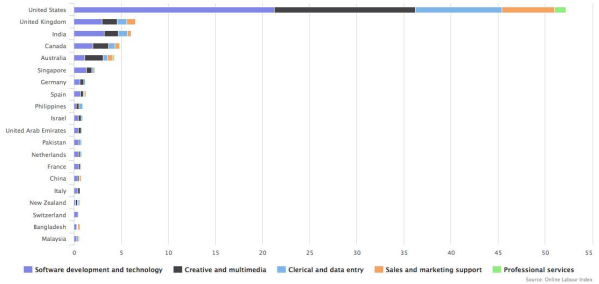
Exploratory data analysis

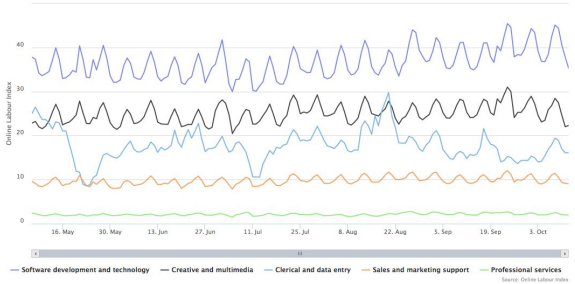
Summarise trends in an easily consumable manner

Physically demonstrate comparisons between groups of data

Present connections otherwise difficult to communicate







Moving beyond dead trees

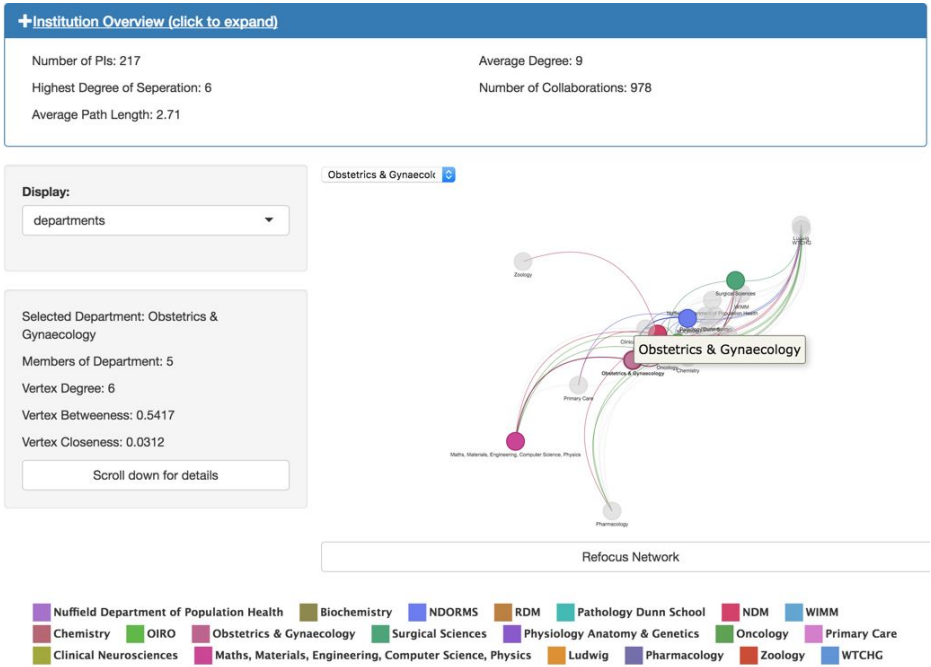
“... interactivity is the new colour chart…”

What does interactivity provide?

Provide alternative methods to access data

Allow users to slice through datasets

Combine summary and detailed information



CHARTABLE

A blog by Datawrapper

datawrapper.de is excellent for two things:

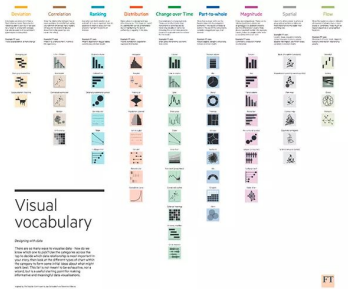
- Point and click chart builder
- Excellent articles on dataviz design

OxShef: Charts

charts.oxshef.io

eagereyes

eagereyes.org is an incredible resource for data-driven advice on dataviz design and ideas.



The FT Visual Vocabulary toolkit is an extremely easy to use tool for choosing the most appropriate dataviz for your data.

#rstats interactive dataviz

R is the programming (or scripting) language we're using to make interactive viz!

If you need a quick intro:
datacamp.com/courses/intro-to-statistics-with-r-introduction

R is the programming (or scripting) language we're using to make interactive viz!

If you need a quick intro:
datacamp.com/courses/intro-to-statistics-with-r-introduction

RStudio is *the* IDE for data science (and package development) for R.

The company behind it - RStudio - build tools which work seamlessly together to make your lives easier (and to make R more powerful).

htmlwidgets is a framework for building bindings between JavaScript libraries and R.

htmlwidgets allows develops to build R packages that provide other useRs with the ability to create interactive charts

htmlwidgets.org

htmlwidgets for R

Home | Showcase | Develop | Flexdashboard | Crosstalk | Gallery | GitHub

HTML widgets work just like R plots except they produce interactive web visualizations. A line or two of R code is all it takes to produce a D3 graphic or Leaflet map. HTML widgets can be used at the R console as well as embedded in R Markdown reports and Shiny web applications. In addition to the widgets featured below you may also want to check out the htmlwidgets gallery.

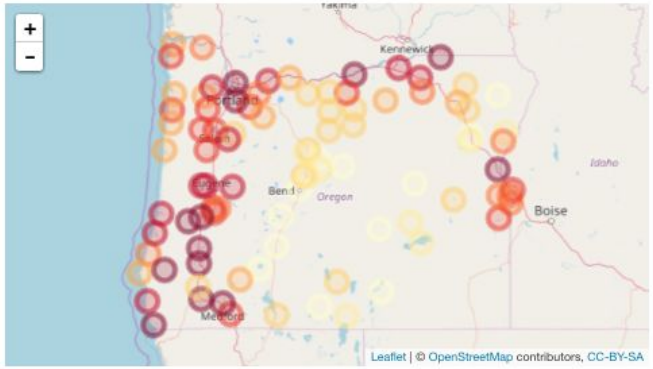
- **Leaflet** — Geo-spatial mapping
- dygraphs — Time series charting
- Plotly — Interactive graphics with D3
- rbokeh — R interface to Bokeh
- Highcharter — R interface to Highcharts
- visNetwork — Graph data visualization with vis.js
- networkD3 — Graph data visualization with D3
- d3heatmap — Interactive heatmaps with D3
- DataTables — Tabular data display
- threejs

Leaflet

http://rstudio.github.io/leaflet/

Leaflet is a JavaScript library for creating dynamic maps that support panning and zooming along with various annotations like markers, polygons, and popups.

```
library(leaflet)
pal <- colorQuantile("YlOrRd", NULL, n = 8)
leaflet(orstationc) %>%
  addTiles() %>%
  addCircleMarkers(color = ~pal(tann))
```



- Scatter geo plot with leaflet
- Choropleth with mapedit and leaflet
- Network visualisation with visNetwork
- Stacked barcharts with plotly and highcharter

linkedin.com/learning has a 5h+ course produced by Martin John Hadley covering the most popular htmlwidget libraries.

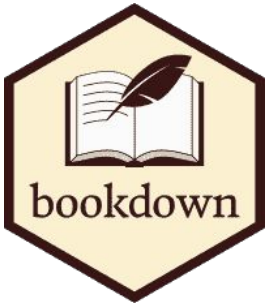
Note that this is a subscription learning service from which Martin receives royalties.

RMarkdown is a simple way to build reports and presentations that include R code.

RMarkdown can output html, PDF and .docx documents.

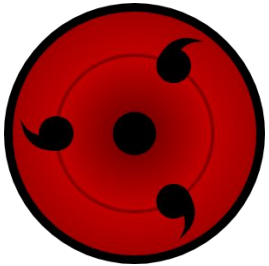
blogdown is a package for creating websites (in particular, blogs) with RMarkdown.

Read more are bookdown.org/yihui/blogdown/



bookdown is a package for creating long form documents (HTML and PDF) with RMarkdown.

Read more are at bookdown.org



xaringan is a package for creating beautiful HTML presentations with RMarkdown

Read more are at slides.yihui.name/xaringan

R Markdown

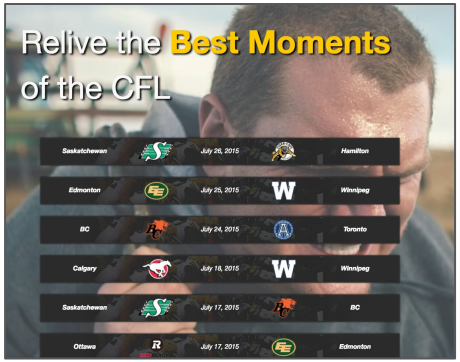
from R Studio

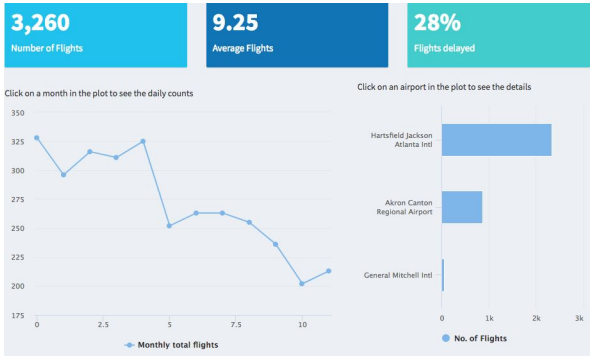
rmarkdown.rstudio.com provides tutorials, templates and examples of many different RMarkdown output types

DataCamp

datacamp.com has a number of excellent (not free) interactive tutorials on creating content with RMarkdown.

Shiny is a framework for creating interactive web applications using R.





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

RStudio Connect

shinyapps.io

Fully hosted solution for Shiny apps

(Includes a free tier!)

Shiny Server

Shiny Server Open Source

(Non-commercial use)

Shiny Server Pro

(Commercial use)

RStudio Connect

shinyapps.io

Fully hosted solution for Shiny apps

(Includes a free tier!)

Shiny Server

Shiny Server Open Source

(Non-commercial use)

Shiny Server Pro

(Commercial use)

RStudio Connect

RStudio Connect

(Commercial use, 45 day evaluation license available)

RStudio Connect is a content management system for htmlwidgets, RMarkdown documents and Shiny apps.

RStudio Connect allows parameterised RMarkdown documents to be scheduled.

#rstats interactive dataviz

VISIBLE DATA

visibledata.co.uk

- Oxford and Bristol based R consultancy focused on reproducible research, dataviz and data science
- Bespoke in-person training provider

jbi training

lynda.com

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

SAGE Publishing

ERICSSON

Perspectum Diagnostics

UBISOFT
FGOL

TTC GROUP
educating for safer roads

figshare

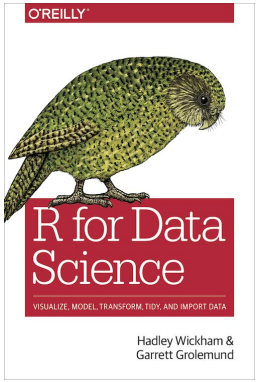
LinkedIn

UNIVERSITY OF OXFORD

Get started with R

datacamp.com/courses/free-introduction-to-r

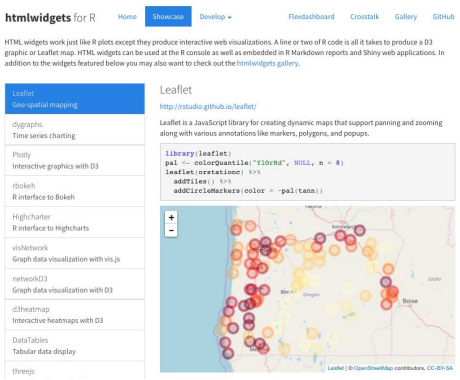
r4ds.had.co.nz/



htmlwidgets

bit.ly/htmlwidgets-course

htmlwidgets.org



shiny

shiny.rstudio.com

Shiny from RStudio

bit.ly/shiny-course

datacamp.com

Feel part of the R Community

#rstats

community.rstudio.com

Reproducible Dataviz Workflows with R

Oxford RUG

Martin John Hadley
@martinjhnhadley

Download today's code from
http://doi.org/10.6084/m9.figshare.7039997