

推荐书单

亲测可读的 10 本书

黄湘云

2019 年 1 月 21 日

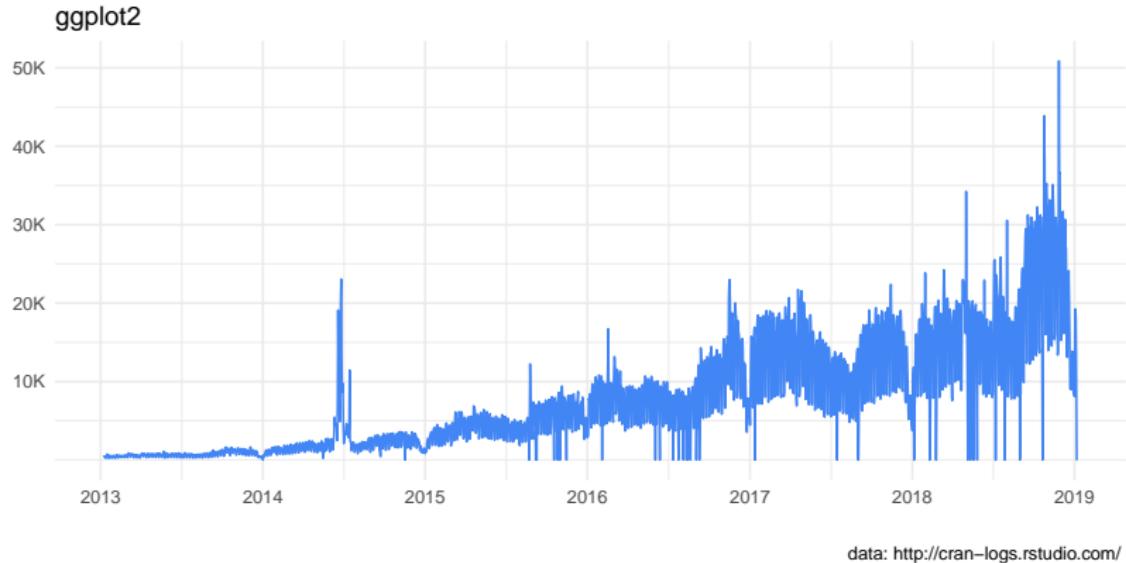


What is R? And why R?¹

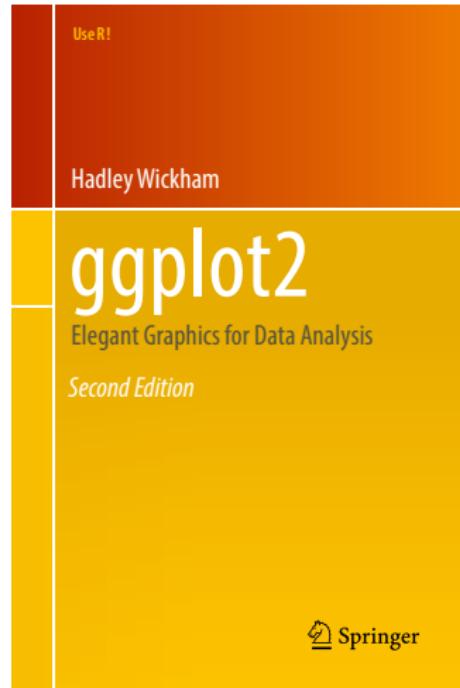


¹<https://www.r-project.org/>

ggplot2 每日的下载量

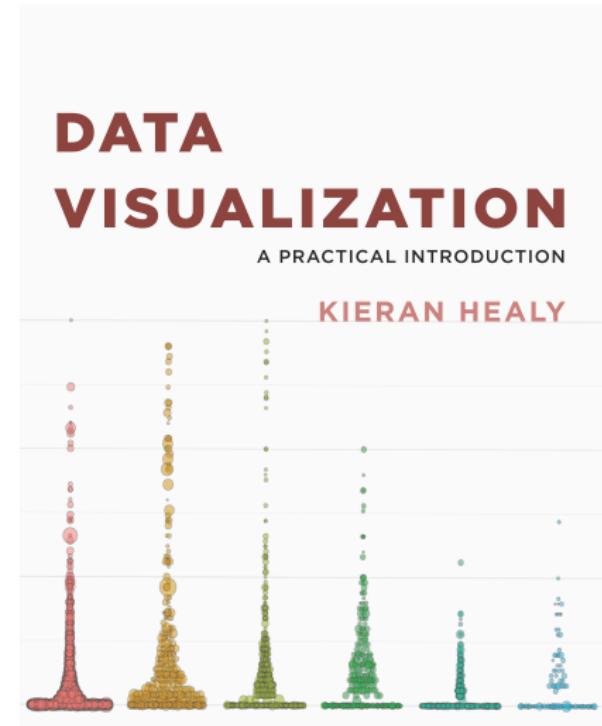


Data Visualization | 推荐指数 ****



Hadley Wickham

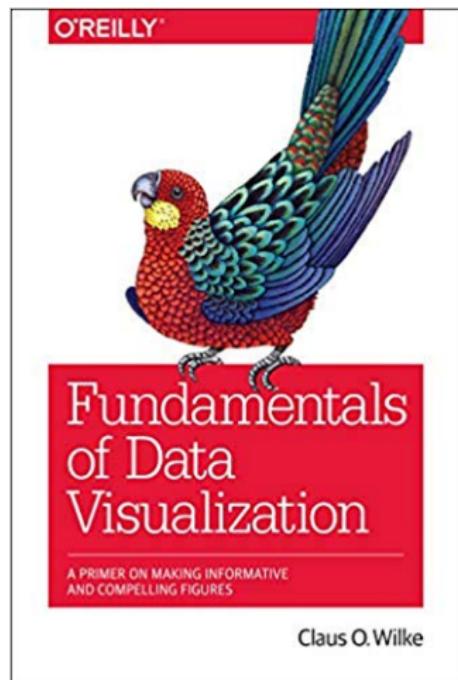
- <https://github.com/hadley/ggplot2>- book



Kieran Healy

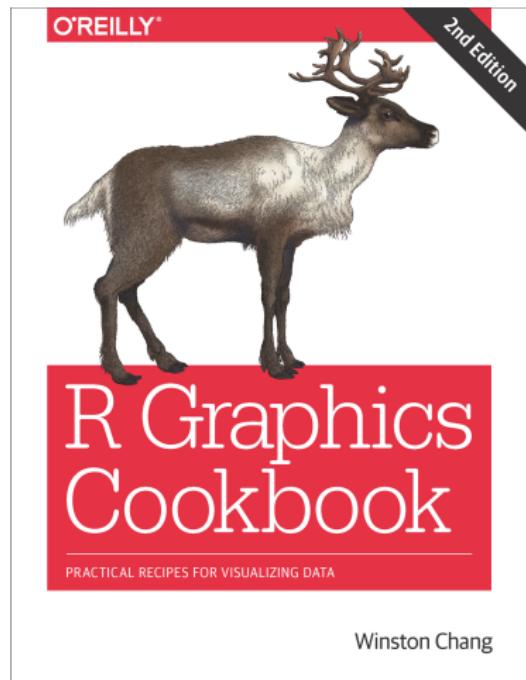
- <https://github.com/kjhealy/socviz>
- <https://socviz.co/>

Data Visualization II 推荐指数 *****



Claus O. Wilke

- <https://github.com/clauswilke/>
- <https://serialmentor.com/dataviz/>



Winston Chang

- <https://github.com/wch>
- <http://www.cookbook-r.com/>

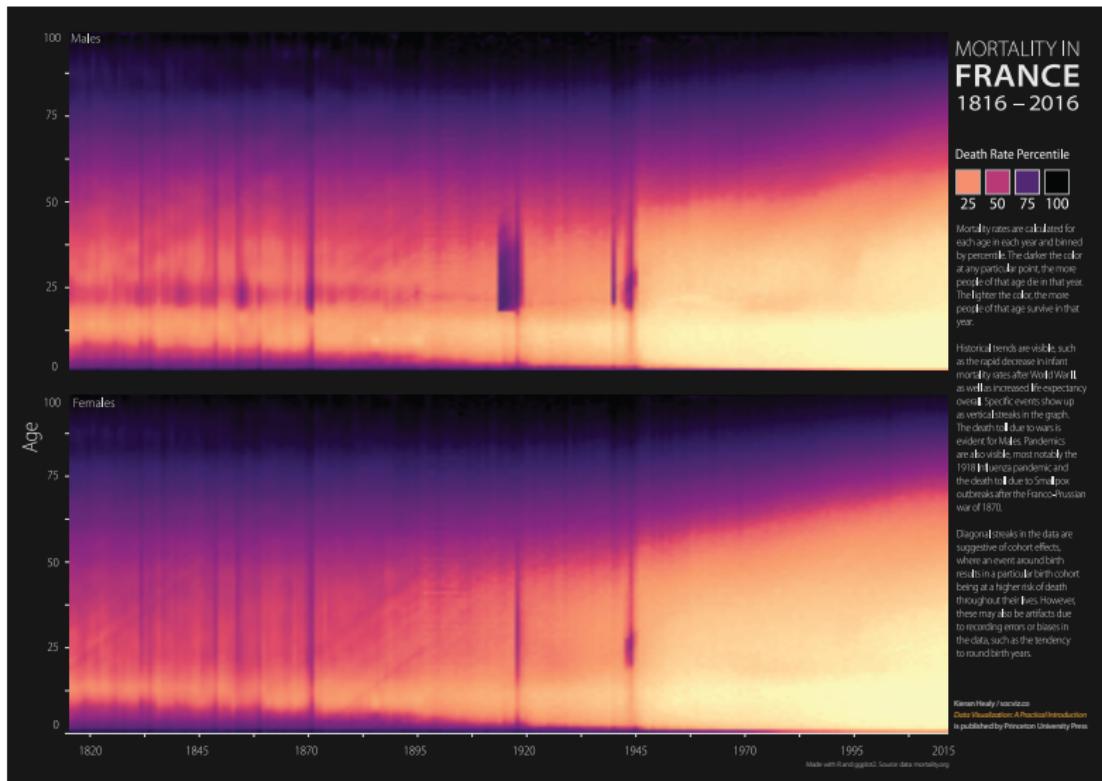
Using R packages and education to scale Data Science at Airbnb²



- Github: https://github.com/ricardo-bion/medium_visualization

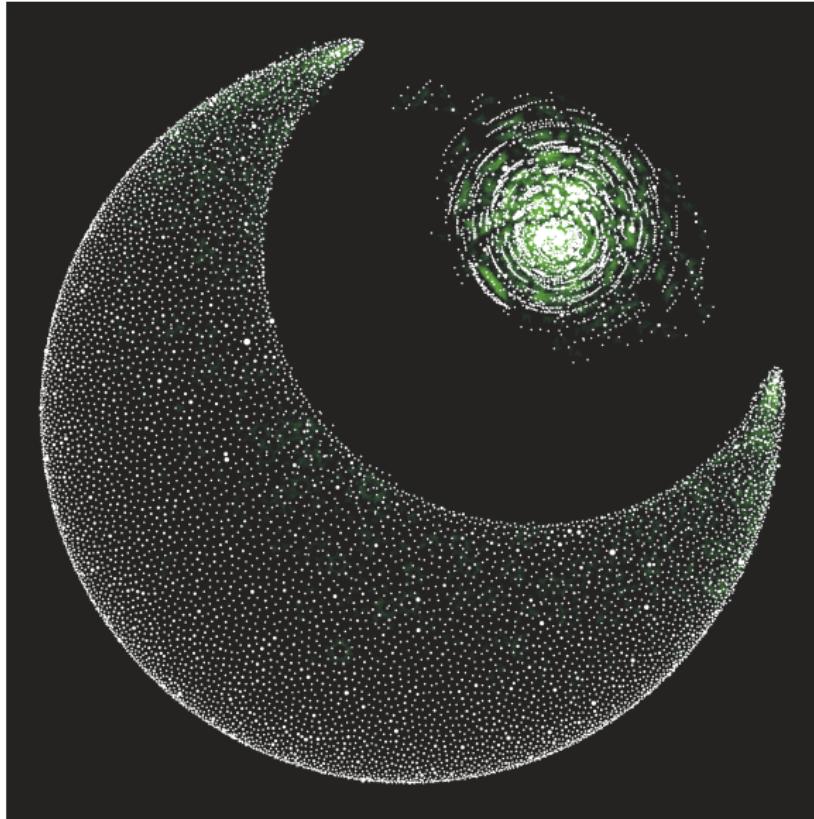
²<https://medium.com/airbnb-engineering/using-r-packages-and-education-to-scale-data-science-at-airbnb-906faa58e12d>

Mortality in France 1816–2016³



³<https://kieranhealy.org/blog/archives/2018/12/27/french-mortality-poster/>

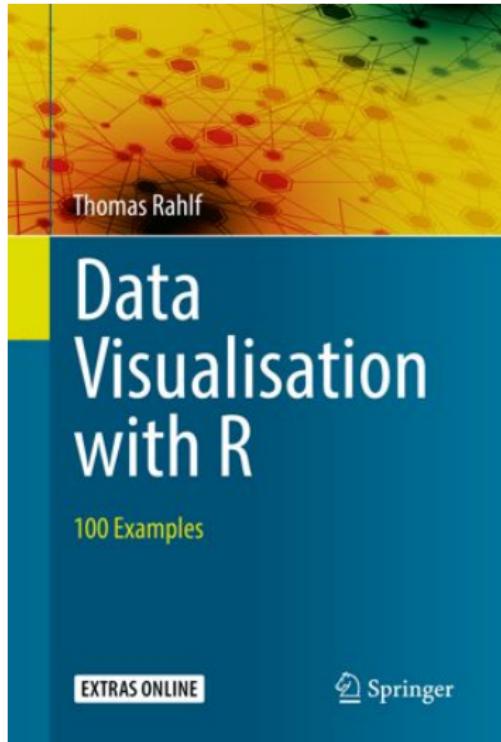
CRAN contributors



网络规模

- R 包 11800+
- 开发者 6000+
- 数据日期: 2017-07-15

Data Visualisation with R⁴ 推荐指数 *****



Prof. Qiwei Yao

I really like this book: those 100 interesting real data examples with the detailed R-scripts for producing beautiful graphical visualisation are particularly helpful for learning R graphics. I will use it as a main textbook for data visualization for my statistics course.

Paul Murrell

The images look REALLY nice!

Uwe Ligges

The graphics look VERY professional, with some I would not have thought that they come from R.

⁴Thomas Rahlf <http://www.datavisualisation-r.com/>

highlights

336

10.3.3 Choropleth Map of Europe at Country-Level



About the figure: Choropleth maps do not necessarily have to show small-scale data. Characterisation of individual countries with variable attributes in different colours can also be useful. In this example, we colour European countries according to a value concerning atheism. Again, we restrict ourselves to one colour and select appropriate gradations with a Brewer palette. In this instance, an additional second colour is used to indicate "missing values". This is useful if we want to point out that certain countries were included in the analysis, but no concrete data for the depicted variable are available. In the chosen layout, not only the legend, but also the title are written within the map. That way, the figure itself can be bigger (see Sect. 9.2.2). Data derive from the European Values Study and were kindly provided by Mira Hassan. They are based on results by Siegers, Pascal (2012): Alternative Spiritualität: neue Formen des Glaubens in Europa: eine empirische Analyse. Akteure und Strukturen. Frankfurt/New York: Campus. See annex A, NUTS, for map data.

```
pdf_file<-pdf("maps_europe_choropleth_countries.pdf")
cairo_pdf(bg="grey98", pdf_file, width=13, height=11)
par(omi=c(0,0,0,0), mai=c(0,0,0,0), family="Lato Light", las=1)
```

10 Maps

M3 Choropleth Maps

```
library(mapprojtools)
library(RColorBrewer)
library(sp)
library(rgdal) # for spTransform
library(gdata)

# Import data and prepare chart
myData<-read.csv("myData/prop.table EVS cntr.csv",sep=";",dec=",",
                  na="")
myData[is.na(myData$ATHE),"ATHE"]<-99
data(wrlid_simpl)
wrlid_simpl[(wrlid_simpl$data[["ISO2"]] %in% myData$Country),]
wrlid_simpl[!(wrlid_simpl$data[["NAME"]] == "Antarctica")]
wrlid_simpl$NUTS_ID<-c("00000000000000000000000000000000")
wrlid_simpl$NAME<-c("Antarctica")
wrlid_simpl$COUNTRY<-c("Antarctica")
wspTransform(wrlid_simpl,CRS=CRS("+proj=merc"))

# Create chart
plot(m,xlim=c(-20000000,5000000),ylim=c(40000000,10000000),col=rgb(
  160,160,160,.1,100,maxColorValue=255),border=F)
x<-readShapeSpatial("myData/NUTS-2010/NUTS_RG_50M_2010.shp",>
  proj4string=CRS("+proj=longlat"))
yc<-(x$NUTS_ID %in% myData$Country,]
wspTransform(y,CRS=CRS("+proj=merc"))

myClasses<-c(<-10,0,10,20,30,50)
myColourPalette<-c("cornflowerblue",brewer.pal(4,"Reds"))

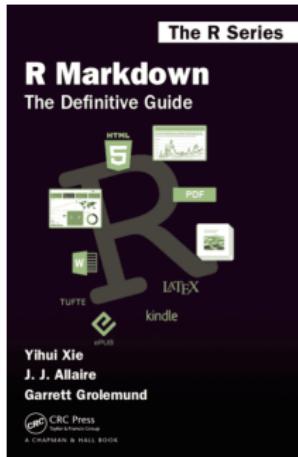
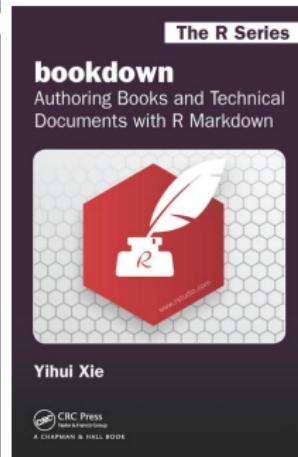
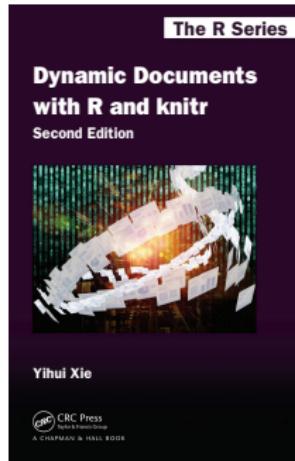
id<-m$NUTS_ID
n<-length(id)
position<-vector()
for (i in 1:n){
  position[i]<-match(m$NUTS_ID[i], myData$Country)
}
myColour_no<-cut(myData$ATHE[position],myClasses)
levels(myColour_no)<-c("missing","0 to 10","10 to 20","20 to 30",
  "30 to 50")
plot(m,col=myColourPalette[myColour_no],border="white",add=T)

legend("bottomleft",levels(myColour_no),cex=1.45,border=F,bty="n",
  fill=myColourPalette,text.col="black")

# Titling
mtext("Atheism in Europe",at=-1300000,cex=2,adj=0,line=-3)
mtext("",at=-1300000,cex=2,adj=0,line=-4.8)
mtext("Source: European Values Study",1,at=3200000,cex=1.7,adj=0,
  line=-2.3,font=3)
dev.off()
```

We start the script by defining a window that is 13 inches wide and 11 inches high. The inner and outer margins are set to 0, and the background to white. We require four libraries: maptools, RColorBrewer, sp, and rgdal for the spTransform()

Dynamic Documents with R and knitr⁸ 推荐指数 ***



- bookdown: Authoring Books and Technical Documents with R Markdown (2016)⁵
- blogdown: Creating Websites with R Markdown (2017)⁶
- R Markdown: The Definitive Guide (2018)⁷
- 作者主页: <https://yihui.name/>

⁵<https://bookdown.org/yihui/bookdown/>

⁶<https://bookdown.org/yihui/blogdown/>

⁷<https://bookdown.org/yihui/rmarkdown/>

⁸<https://yihui.name/knitr/>

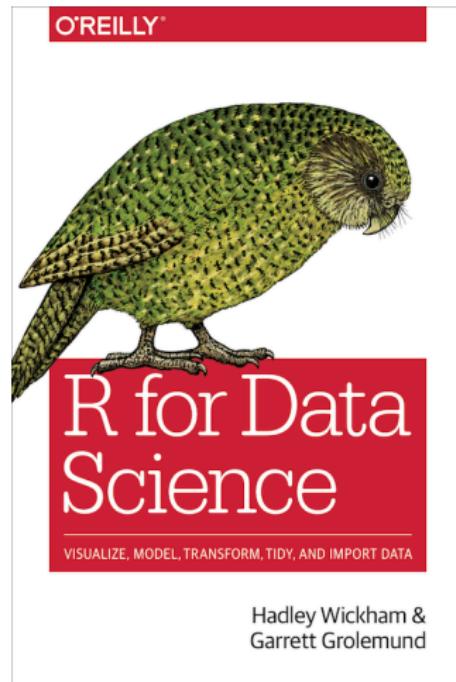
男神 — Hadley Wickham⁹



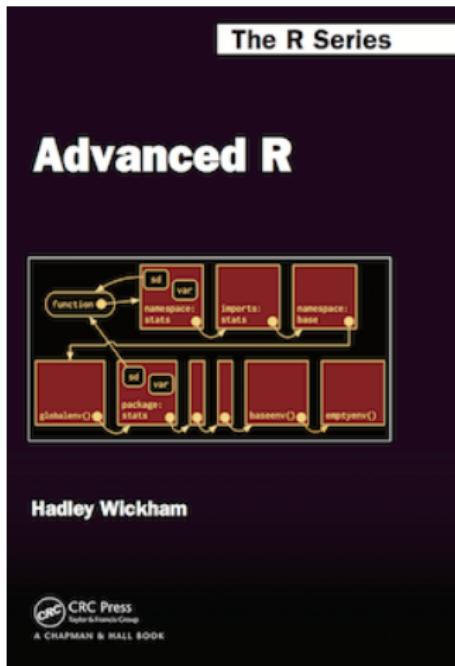
- 改变做图的姿势 — ggplot2
- 改变开发的姿势 — devtools
- 改变数据的姿势 — dplyr

⁹图片来自 <http://hadley.nz/>

R for Data Science 和 Advanced R 推荐指数 **



- <https://github.com/hadley/r4ds>
- <https://r4ds.had.co.nz/>



- <https://github.com/hadley/adv-r>
- <https://adv-r.hadley.nz/>

awesome bloggers

ILYA KASHNITSKY

BLOG ARCHIVE ME PUB CONF DO TSA CONTACTS



demographer's notes



December 03, 2018

Compare population age structures of states, 2 regions and the US country using ternary color-coding

On 28 November 2018 I presented a poster at DataDemograph Day in Utrecht. Here it is:



September 23, 2018

Update: all SciMago Journal & Country Rank data, ready for R



July 25, 2018

Regional population structures at a glance



June 01, 2018

Deep Catalan roots: playing in strength

Probable



(a) Ilya Kashnitsky
<https://ikashnitsky.github.io/>

Shirin's playground

Dr. Shirin Glander

Biologie Journal für Numerische Kernal Data Science

Home Archives Categories Tags

Code for case study - Customer Churn with Keras/TensorFlow and H2O

This is code that accompanies a book chapter on customer churn that I have written for the German Springer Verlag. The book is in German and will probably appear in February:
<https://www.dipity.de/buch/erstes-patrick-schuster-data-science.html>. The code you find below can be used to recreate all figures and analyses from this book chapter. Because the content is exclusively for the book, my descriptions around the code had to be minimal. If I'm sure, you can get the pdf, even without the book.

Contributing

Trust in ML models. Slides from TWIML & AI EMEA Meetup + IX Articles

Update: There is now a recording of the meetup up on YouTube. Here you find my slides from the TWIML & AI EMEA Meetup about Trust in ML models, where I presented the Anscombe paper by Carlos Grajeda et al.. I have also just written two articles for the German IT magazine iX about the same topic of Explaining Black-Box Machine Learning Models: A short article in the IX 12/18 and a longer one in the IX 01/19.

(c) Shirin Glander
<https://shirinsplayground.netlify.com/>

黄湘云

MAX WOOLF

@minimaxir

ABOUT

CODE PORTFOLIO

DATA PORTFOLIO PATREON



Max Woolf (@minimaxir) is a Data Scientist at BuzzFeed in San Francisco. He is also an ex-Apple employee and Carnegie Mellon University graduate.

In his spare time, Max uses Python to gather data from public APIs and ggplot2 to plot plenty of pretty charts from that data. On special occasions, he uses Keras for fancy deep learning projects.

You can learn more about Max here, view his data analysis portfolio here, or view his coding portfolio here.



(b) Max Woolf <https://minimaxir.com/>

Thoughts + Data + R

HOME ABOUT ME CONTACT

Laminar flow with ggplot2 and gganimate

Watermain Breaks in the City of Toronto

Regular Expression & Treemaps to Visualize Emergency Department Visits

(d) Eugene Joh <https://incidental-ideas.org/>

推荐书单

2019年1月21日

14 / 15

Thanks!

Q & A

- Github: <https://github.com/XiangyunHuang>
- homepage: <https://www.xiangyunhuang.com.cn/>
- slides:
<https://xiangyunhuang.github.io/awesome-readings/awesome-readings.pdf>