

# R 能给我们生活带来什么？

---

王敏杰

2020 年 9 月 19 日

四川师范大学

# R 能给我们生活带来什么？

这个问题，好比人生三大终极问题：

- R 是什么？
- R 能干什么？
- 为什么是 R ？

# R 是什么

---

# R 那些事

- 1992 年，新西兰奥克兰大学统计学教授 Ross Ihaka 和 Robert Gentleman，为了方便地给学生教授统计学课程，他们设计开发了 R 语言（他们名字的首字母都是 R）。



**Ross Ihaka**



**Robert Gentleman**

# R 是什么

官网定义：<https://www.r-project.org/>

R 语言是用于统计分析，图形表示和报告的编程语言：

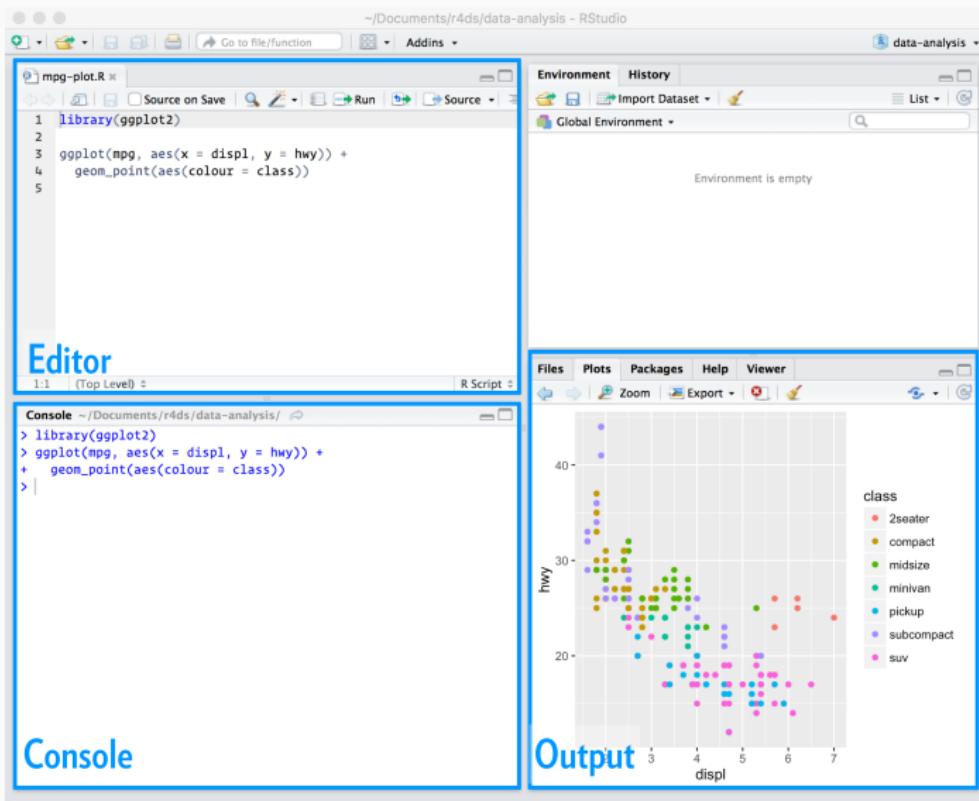
- R 是一个**统计编程**语言 (statistical programming)
- R 可运行于多种平台之上，包括 Windows、UNIX 和 Mac OS X
- R 拥有顶尖水准的**制图**功能
- R 是免费的
- R 应用广泛，拥有丰富的**库包**
- 活跃的**社区**

# R 语言发展趋势

Jul 2020	Jul 2019	Change	Programming Language	Ratings	Change
1	2	▲	C	16.45%	+2.24%
2	1	▼	Java	15.10%	+0.04%
3	3		Python	9.09%	-0.17%
4	4		C++	6.21%	-0.49%
5	5		C#	5.25%	+0.88%
6	6		Visual Basic	5.23%	+1.03%
7	7		JavaScript	2.48%	+0.18%
8	20	▲	R	2.41%	+1.57%
9	8	▼	PHP	1.90%	-0.27%
10	13	▲	Swift	1.43%	+0.31%
11	9	▼	SQL	1.40%	-0.58%
12	16	▲	Go	1.21%	+0.19%
13	12	▼	Assembly language	0.94%	-0.45%
14	19	▲	Perl	0.87%	-0.04%

TIOBE index

# 界面很友好



# R 路上的大神

2019 年 8 月，国际统计学年会将考普斯总统奖（被誉为统计学的诺贝尔奖）奖颁给 tidyverse 的作者



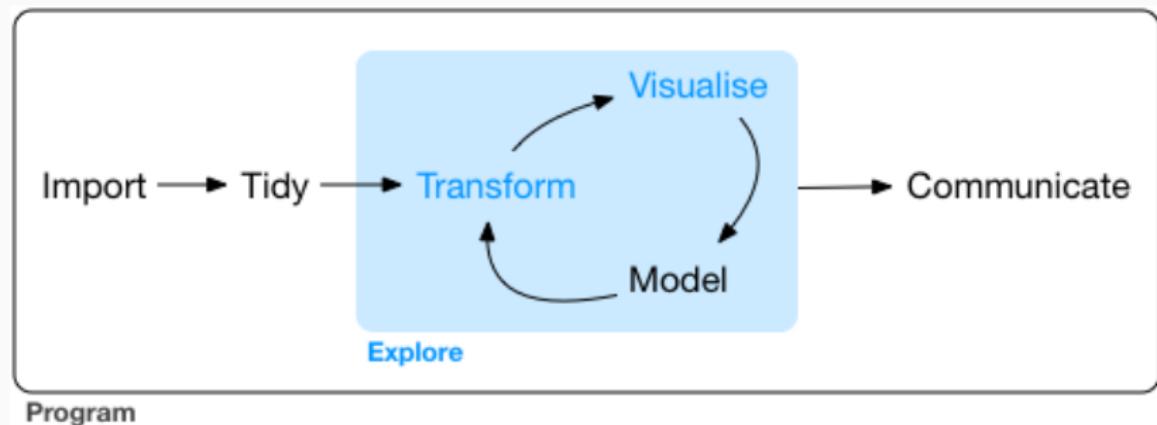
- Hadley Wickham
- R 路上的大神
- 一个改变了 R 语言的人

# R 能干什么

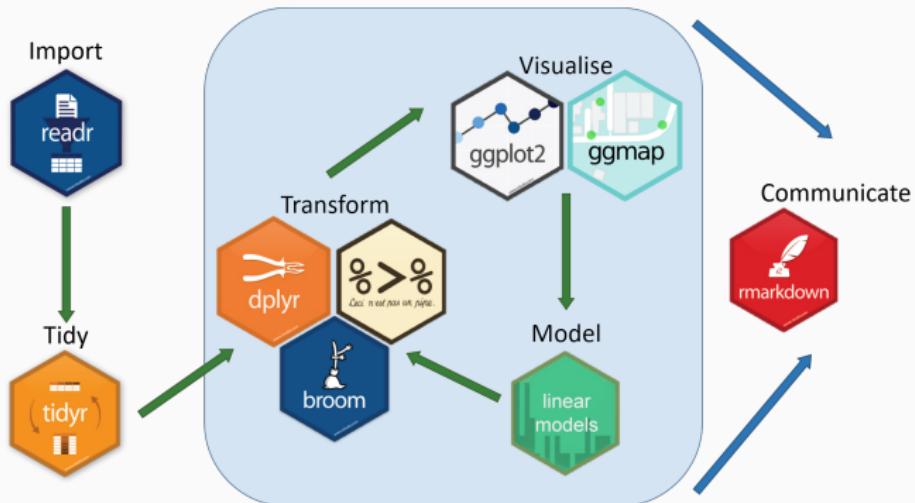
---

# 数据科学的流程

Hadley Wickham 定义了数据科学的工作流程



# tidyverse 套餐



<https://www.tidyverse.org/>

# R & tidyverse

序号	内容	代码演示
1	统计	1_stats.R
2	可视化	2_visual.R
3	探索性分析	3_eda.R
4	可重复性报告	4_reproducible.R

难吗？

感觉很难吗？  
如果是，那说明你认真听了

看了这些代码，可能第一眼感觉是这样的



图 1：图片来自电影《降临》

但我更希望这门课结束后

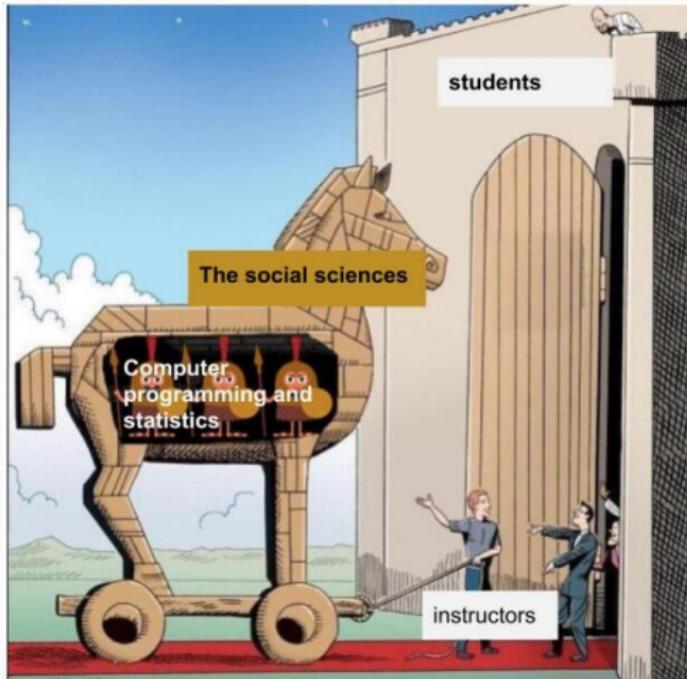


图 2：图片来自美剧《权利的游戏》

# 为什么是 R

---

# 社会科学需要统计



我们不是学统计的，但需要统计

# 社会科学需要可视化



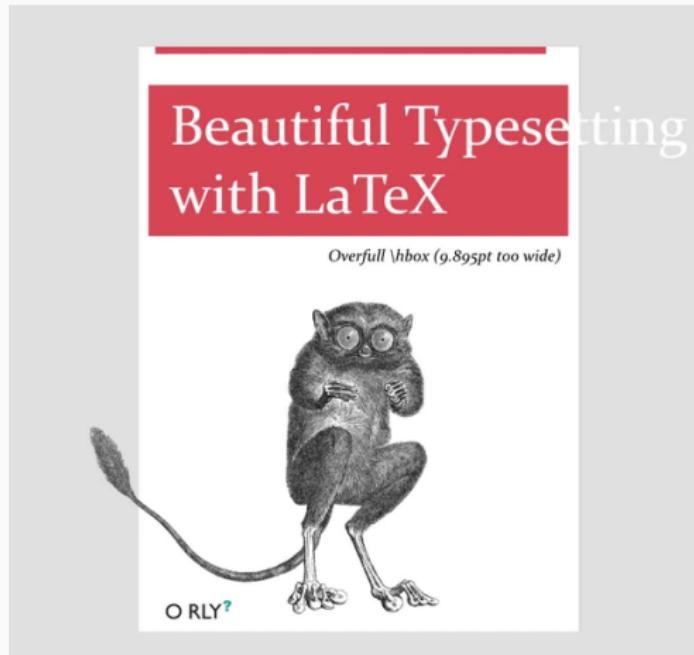
我们不是学美术的，但要可视化

# 社会科学需要编程



我们不是学计算机的，但需要编程

# 你的论文需要排版



我们不是学设计的，但要操心**交叉引用**的事

# 挖掘机技术到底哪家强？

你有需求，而



很专业

序号	内容	特性	评价
1	统计分析	看家本领	好用
2	ggplot2 画图	颜值担当	好看
3	tidyverse 语法	简单易懂	好学
4	可重复性报告	方便快捷	好玩

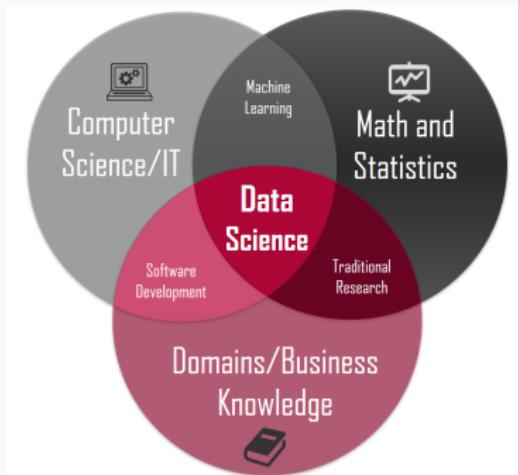
一见钟情，还是相见恨晚？



# 关于学习

---

# 我们的课程不会枯燥



- 数据科学是为社会科学服务的，我们会有很多案例
- 编程是工具，统计是灵魂，专业是核心

# 关于学习

我很少使用

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

更多的是

```
library(tidyverse)
summary_monthly_temp <- weather %>%
  group_by(month) %>%
  summarize(mean = mean(temp),
            std_dev = sd(temp))
```

# 关于学习

## 课程目标

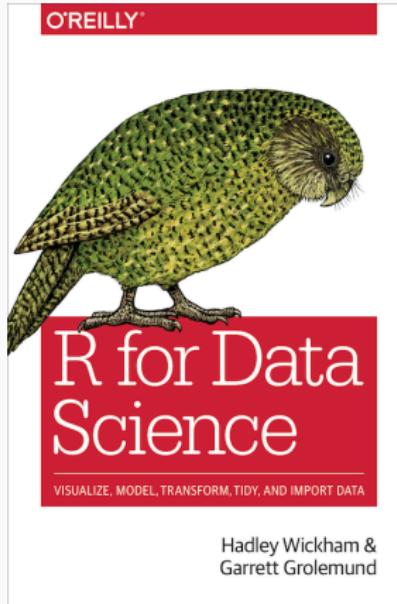
- 训练数据思维，提升编程技能，培养创新能力

## 学习方法

- 问题驱动型学习
- 多实践（光看李小龙的电影，是学不会功夫的）
- 不是 learning R，而是 learning with R
- 把 R 看做是知识学习的脚手架



# 参考书目



- R for Data Science
- <https://bookdown.org/wangminjie/R4DS/>