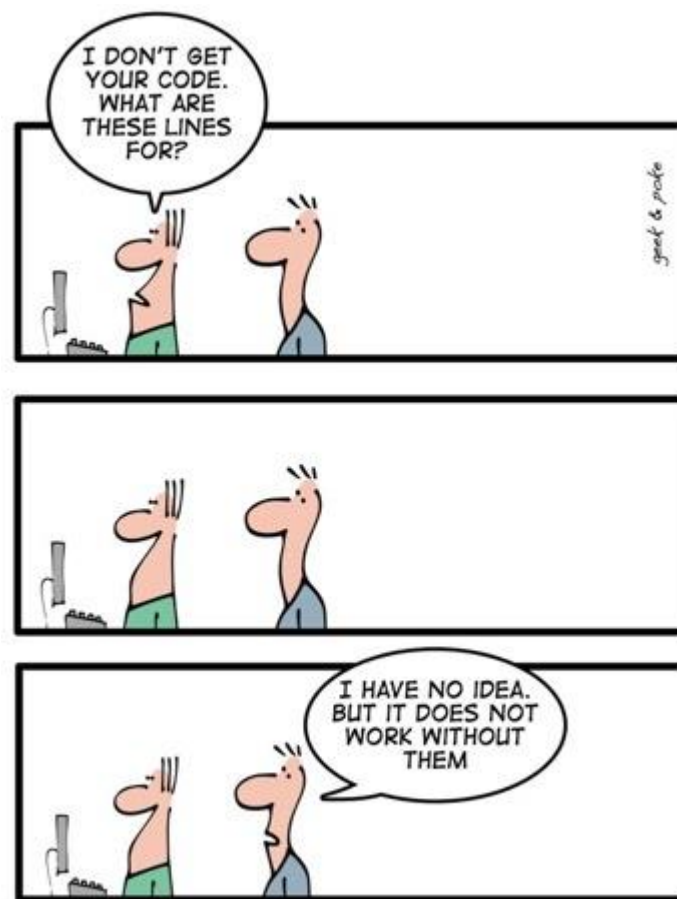


大數據分析方法

R 101

曾意儒 助理教授

長庚大學 資訊管理學系



THE ART OF PROGRAMMING - PART 2: KISS

Outlines

- Why R?
- What is Git?
- What is GitHub?
- How to Use DataCamp?











Why R?

Free & Open!



R

5th Place in IEEE Language Rankings









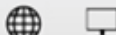

Language Rank Types Spectrum Ranking

1. Java		100.0	2015
2. C		99.9	
3. C++		99.4	
4. Python			
5. C#			
6. R			
7. PHP			
8. JavaScript			
9. Ruby			
10. Matlab			

Language Rank Types Spectrum Ranking

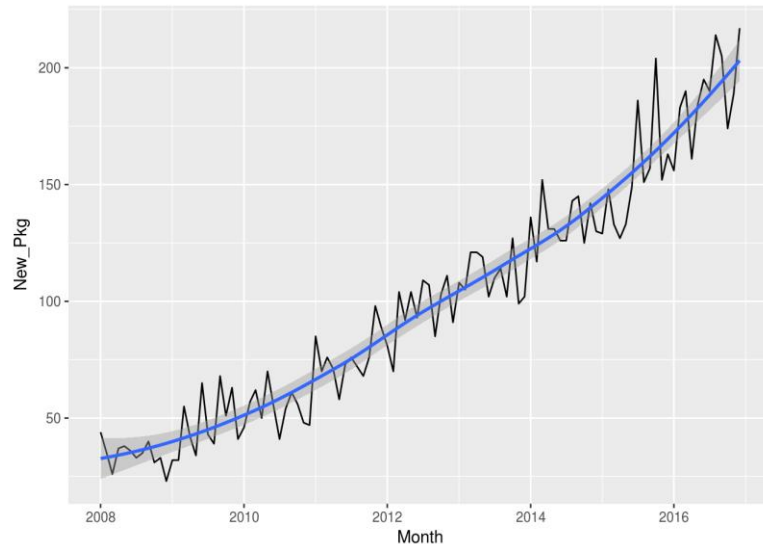
1. C		100.0	2016
2. Java		98.1	
3. Python			
4. C++			
5. R			
6. C#			
7. PHP			
8. JavaScript			
9. Ruby			
10. Go			

Language Rank Types Spectrum Ranking

1. Python		100.0	2017
2. C		99.7	
3. Java		99.5	
4. C++		97.1	
5. C#		87.7	
6. R		87.7	
7. JavaScript		85.6	
8. PHP		81.2	
9. Go		75.1	
10. Swift		73.7	

More Than 10,000 Packages!

- Standing on the shoulders of giants - Hal R. Varian, Google
- To load data
- To manipulate data
- To visualize data
- To model data
- To report results
- For spatial data
- For time series and financial data
- To write high performance R code
- To write your own R packages

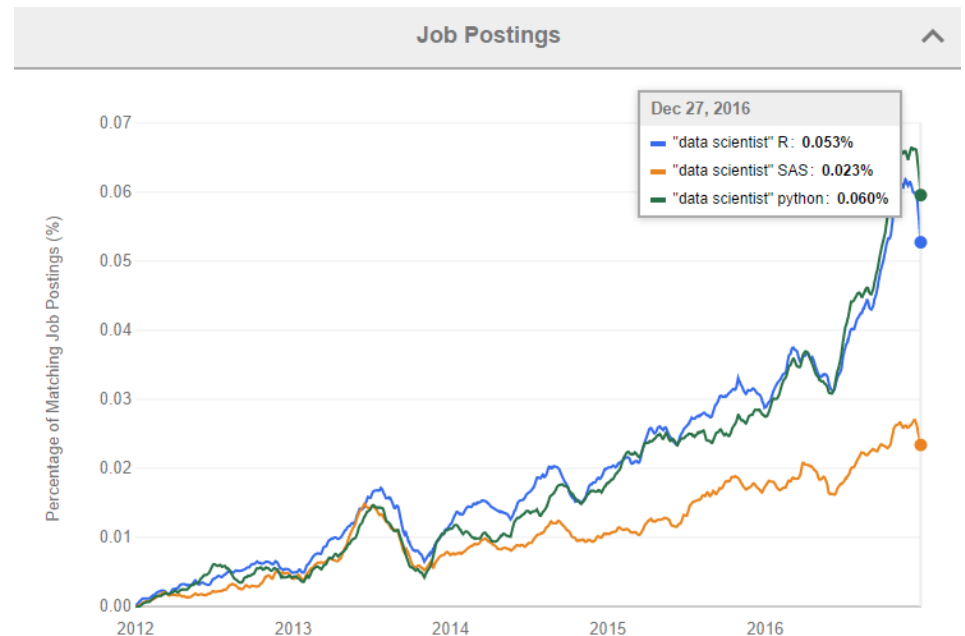
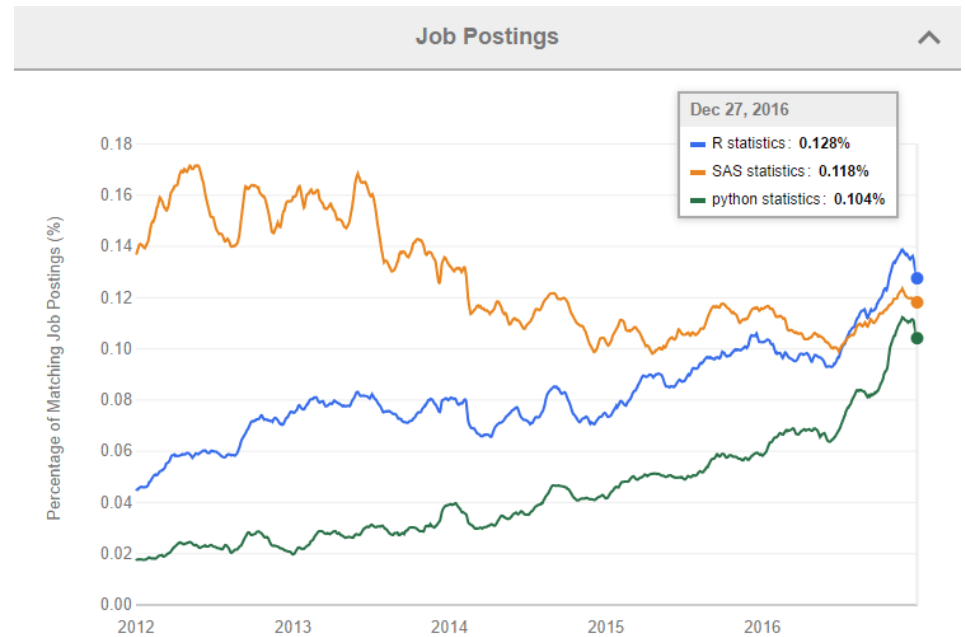


<https://www.rstudio.com/rviews/2017/01/06/10000-cran-packages/>

<https://support.rstudio.com/hc/en-us/articles/201057987-Quick-list-of-useful-R-packages>

R statistics,
SAS statistics, and
python statistics Job Trends
@indeed.com

SAS ↓
R & python ↑



軟體安裝

- R
 - <https://www.r-project.org/>
- RStudio
 - IDE for R
 - <https://www.rstudio.com/>
- 詳細安裝步驟
 - <http://yijutseng.github.io/DataScienceRBook/install.html>

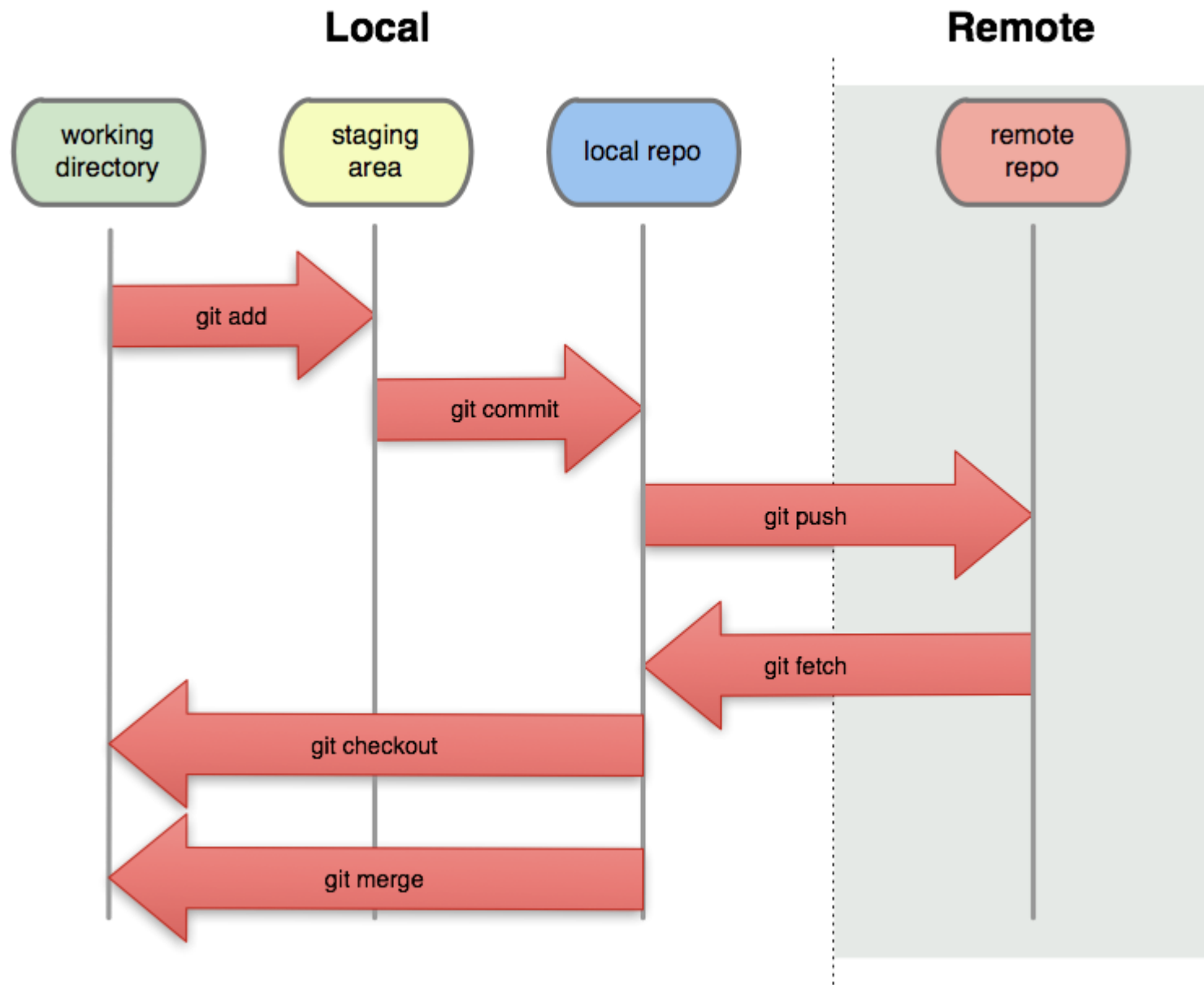
R的起源

- S Language
 - John Chambers @ Bell's Lab, 1976
- R Language
 - Ross Ihaka, Robert Gentleman @ University of Auckland, 1993
- 2000 R 1.0.0

What Is Git?

What Is Git?

- Version control (like TFS)
- Created by the same people who developed Linux
- Free
- Open source
- Command line
- GIT版本控制教學
<https://kingofamani.gitbooks.io/git-teach/content/index.html>



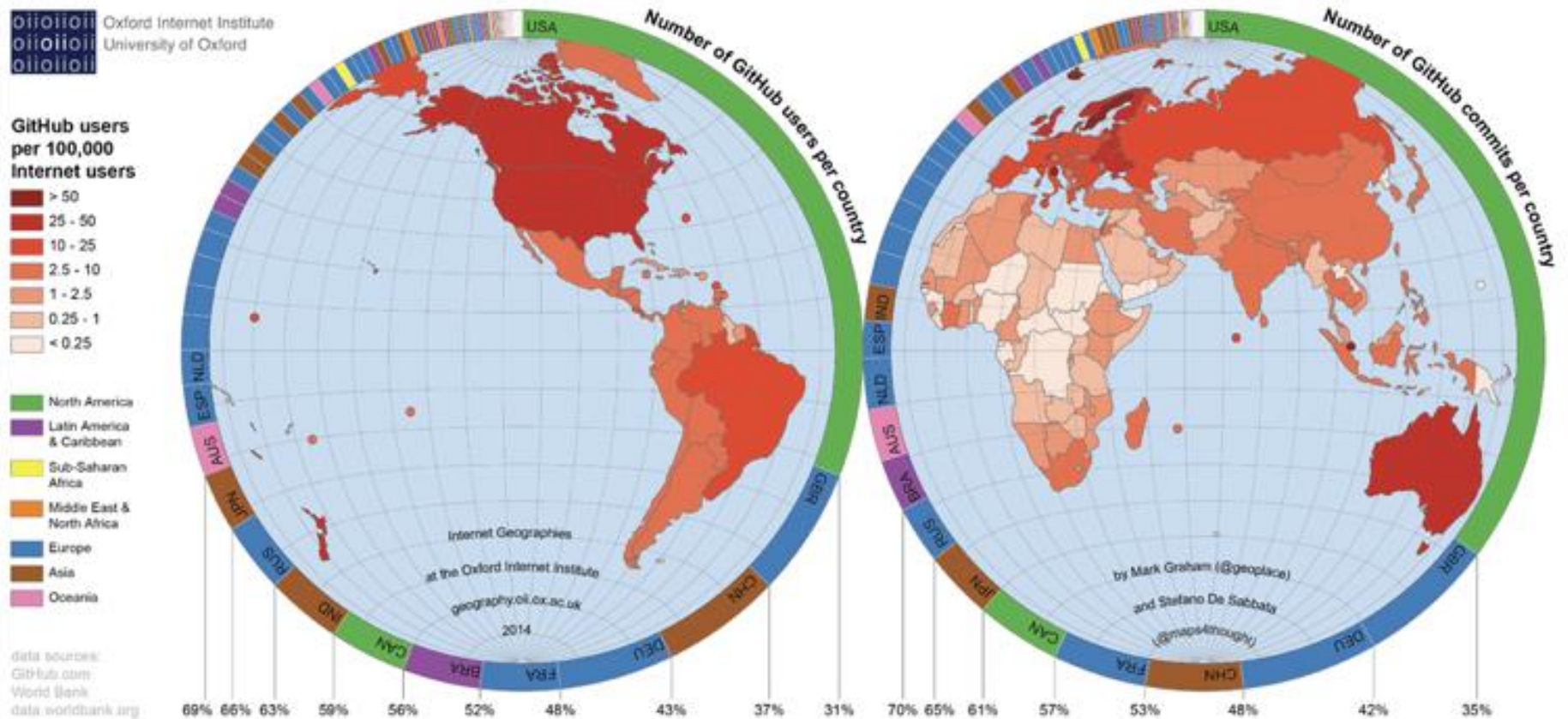
What Is GitHub?

What Is GitHub?

- <https://github.com/>
- April 2008
- Web-based Git repository hosting service
- Free public account

- Git != GitHub
- Local vs. Remote

As of February 2016, GitHub reports having over 12 million users and over 31 million repositories



GitHub | Mapping collaborative software



Hadley Wickham
hadley

RStudio
 Houston, TX
 h.wickham@gmail.com
 <http://had.co.nz>
 Joined on Apr 1, 2008

5.6k Followers
111 Starred
7 Following

Organizations

Contributions

Repositories

Public activity

Unfollow



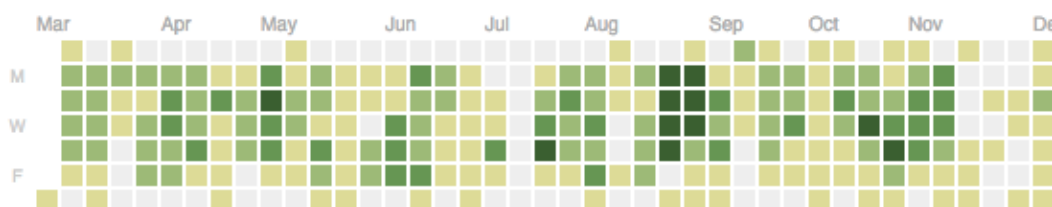
Popular repositories

ggplot2	An implementation of the Grammar of Graphic...	1,699 ★
devtools	Tools to make an R developer's life easier	1,204 ★
dplyr	Plyr specialised for data frames: faster & with ...	1,092 ★
adv-r	Advanced R programming: a book	701 ★
rvest	Simple web scraping for R	503 ★

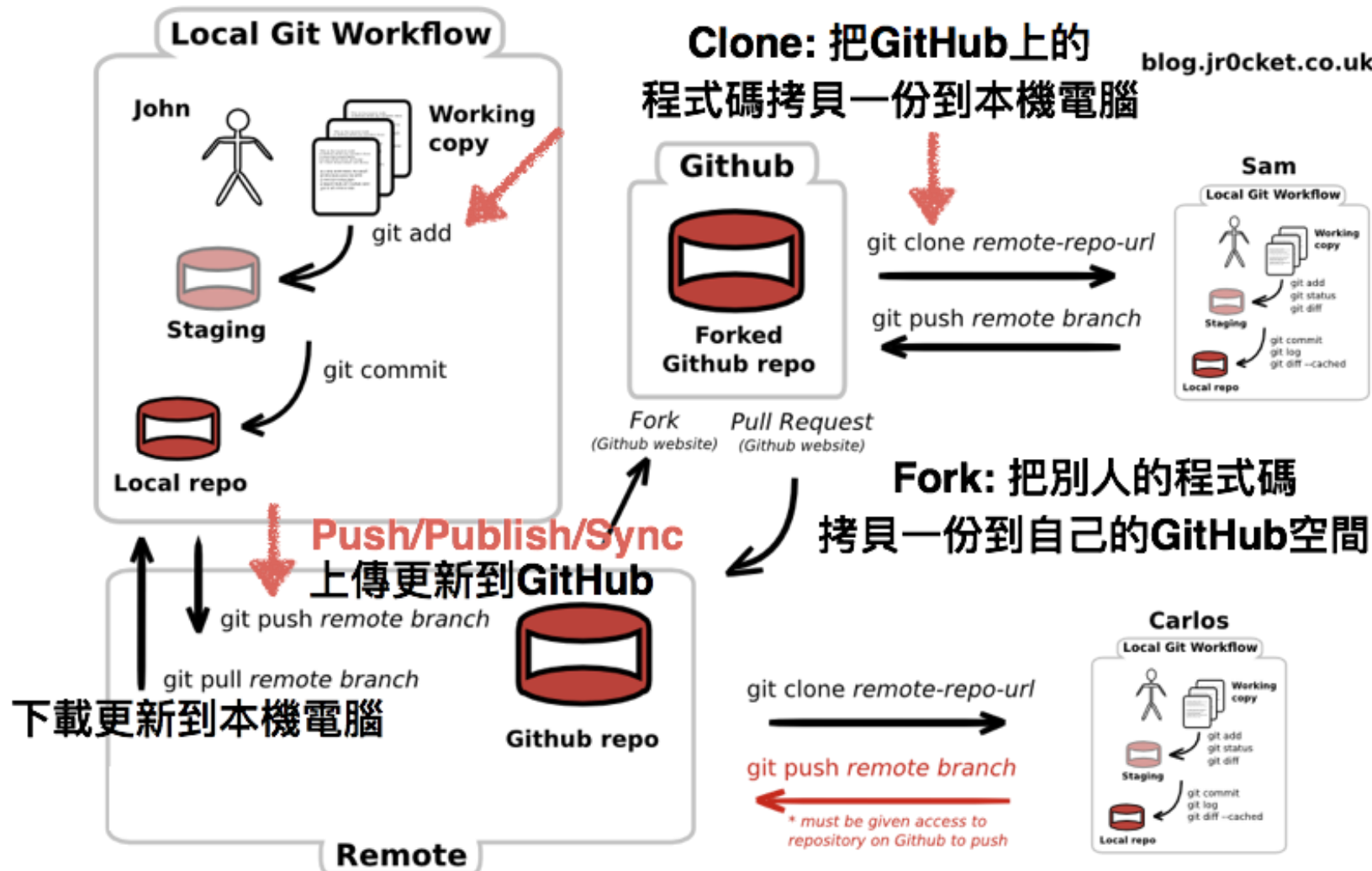
Repositories contributed to

krmlr/tibble	Data frames and table sources in "dplyr" style	12 ★
klutometis/roxygen		166 ★
davidgohel/gdtools	R utilities for graphical rendering	6 ★
jeroenooms/curl	An R connection interface to libcurl.	36 ★
garrettgman/ggsubplot	Embed subplots in ggplot2 g	


Public contributions






Less More




Fork

 This repository

[Pull requests](#) [Issues](#) [Gist](#)

 **Fossj117 / NBAdata**


[Watch](#) 15 [Star](#) 34 [Fork](#) 31

[Code](#) [Issues 0](#) [Pull requests 0](#) [Wiki](#) [Pulse](#) [Graphs](#)

Contains scraper in R for grabbing NBA Sport Tracking Data

[6 commits](#) [1 branch](#) [0 releases](#) [1 contributor](#)

[Branch: master](#) [New pull request](#) [New file](#) [Upload files](#) [Find file](#) [HTTPS](#) <https://github.com/Fossj117/NBAdata> [Download ZIP](#)

 **Fossj117** Added new README Latest commit c813894 on Jul 15, 2014

README.md	Added new README	2 years ago
nba.csv	Added final .csv	2 years ago
nba_data_5-10.csv	Added end of season data pull	2 years ago
nbascraper.R	Revised readIt function using textConnection()	2 years ago
position_data.csv	Added script for pulling position data	2 years ago
pullPositions.py	Added script for pulling position data	2 years ago

GitHub Classroom

GitHub Classroom

Your course assignments on GitHub



細節等要交作業時再說~

Learn to ship software like a pro



Stop emailing code

The days of sharing source code by email and thumb drives are long gone. Using GitHub, it's a piece of cake to keep your team up to date with the latest changes to your code. Thanks to Git's powerful merge features, you don't even have to worry about overwriting someone else's changes.



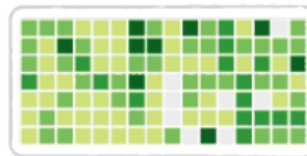
Never lose your work

No need to worry about your dog eating your homework or your hard drive exploding. Your latest and greatest code is always available to you and your [Collaborators](#) on GitHub. Search through your history and easily revert to a known good state when you need to.



Stay organized

A GitHub [Repository](#) is a single place to go for everything related to a project. Keep tabs on new feature ideas and document known bugs using [Issues](#). Discuss proposed changes inline with the code using [Pull Requests](#). Organize your team and goals using [Milestones](#).



Build a portfolio

Using public Repositories, you can build a portfolio of projects and show them off to potential employers. Who needs a resume when you can show what you've built?

How to Use DataCamp?

<https://www.datacamp.com/>

My Groups

[Create New Group](#)

DHLab

FREE ADMIN

[Go to the Group >](#)



Big Data Analytical M...

PREMIUM ADMIN

[Go to the Group >](#)



專題

FREE ADMIN

[Go to the Group >](#)



還沒有收到DataCamp邀請信件的同學，請聯絡助教



Big Data Analytical Methods, Spring 2017

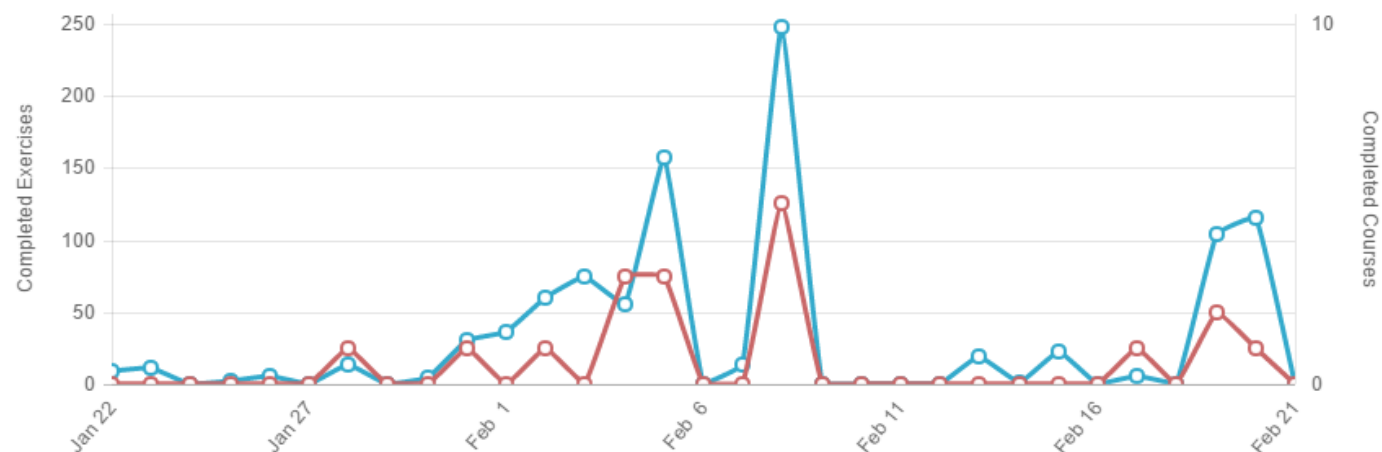
PREMIUM

[Buy More Seats](#)

Member Activity

○ Course Completions ○ Exercise Completions

30 Days | [90 Days](#) | [Last Year](#)



Manage Your Group

[Members](#)[Assignments](#)[Courses](#)[Settings](#)[Export Group Data](#)

Approved

Congratulations! Your DataCamp Classroom Account was approved. All invited students have full access until **20 Jul 2017**.

Active Members (8)

Pending Members (1)

[Invite Members](#)

Name ▾

Assignments ▾

Search



Anny Yen

19,670 XP

2

0

0



bingbyebye

0 XP

0

0

0



Big Data Analytical Methods, Spring 2017

PREMIUM

Buy More Seats

Assignments

Create Assignment

Active Assignments (2)

Name ▾

Assigned At

Due By



資料框整理技巧

Feb 20, 2017

Apr 17, 2017

3

-

-



R 語言導論

Feb 20, 2017

Apr 10, 2017

3

-

-



Manage Your Group

Members

Assignments

Courses

Settings

Export Group Data

Approved

Congratulations! Your DataCamp Classroom Account was approved. All invited students have full access until **20 Jul 2017**.



Welcome Yi-Ju



This is your personalized learning dashboard. Here you can check progress, decide what courses to take next, and catch up on DataCamp news.

IN PROGRESS

COMPLETED

COURSE LIBRARY

Data Visualization with ggplot2 (Part 3)

This course covers some advanced topics including strategies for handling large data sets and specialty plots.



RICK SCAVETTA

Data Scientist and cofounder of Science Craft



Working with the RStudio IDE (Part 1)

Learn the basics of the important features of the RStudio IDE.



GARRETT GROLEMUND

Data Scientist at RStudio



Working with the RStudio IDE (Part 2)

Further your knowledge of RStudio and learn how to integrate Git, LaTeX, and Shiny



GARRETT GROLEMUND








Data Scientist at RStudio





MY GROUPS

ASSIGNMENTS

NAME	ASSIGNED AT	DUE BY	STATUS
 R 語言導論	Dec 29, 2016	Jan 30, 2017	MISSED
 資料框整理技巧	Dec 29, 2016	Jan 30, 2017	MISSED
 Data Manipulation in R with dplyr	Dec 29, 2016	Feb 13, 2017	COMPLETE
 Data Analysis in R, the data.table Way	Dec 29, 2016	Feb 13, 2017	MISSED
 Exploratory Data Analysis in R: Case Study	Dec 29, 2016	Feb 20, 2017	MISSED
 Cleaning Data in R	Jan 20, 2017	Feb 6, 2017	MISSED
 Data Visualization with ggplot2 (Part 1)	Jan 20, 2017	Feb 6, 2017	MISSED

推薦課程