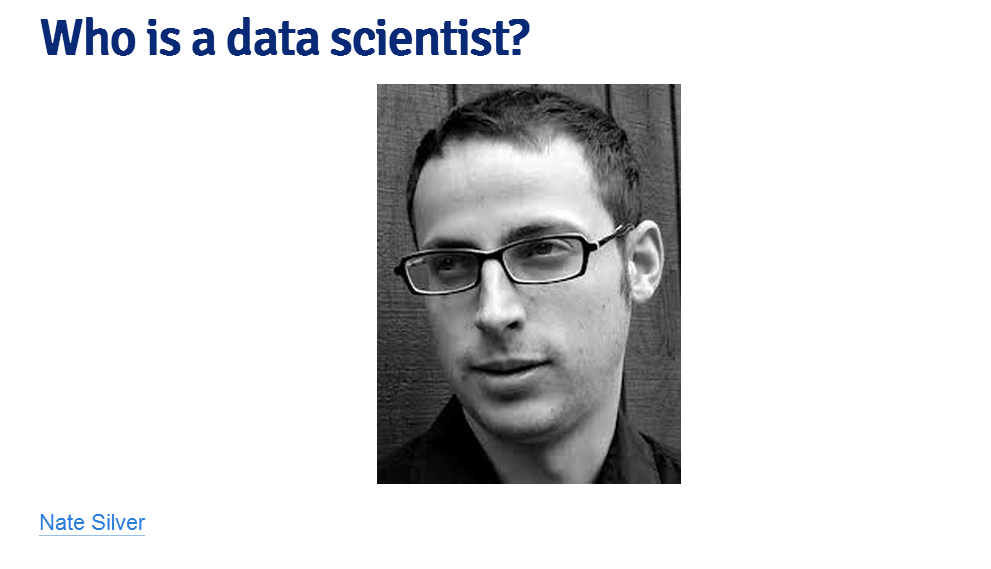
# 1. The Data Scientist’s Toolbox

## 1.1 Series Motivation

The key word in data science is not data; it is science



<http://fivethirtyeight.blogs.nytimes.com/>

noise and signal

## 1.2一般的数据分析的步骤

* Define the question
* Define the ideal data set
* Determine what data you can access
* Obtain the data
* Clean the data
* Exploratory data analysis
* Statistical prediction/modeling
* Interpret results
* Challenge results
* Synthesize/write up results
* Create reproducible code
* Distribute results to other people

## 1.3 一些帮助函数

**Access help file**

?rnorm

**Search help files**

help.search("rnorm")

**Get arguments**

args("rnorm")

这些都可以直接在窗口右下侧的help里直接搜索函数

Short-refcard是介绍R的一些常用的命令函数，我这里有个中文版的，可以从群里下载

## 2.3Introduction to Git

（1）这里有个Git教程

<http://www.liaoxuefeng.com/wiki/0013739516305929606dd18361248578c67b8067c8c017b000>

windows版git下载 <http://msysgit.github.io/>

（2）local Git 与 Github连接

$ mkdir ~/test-repo

$ cd ~/test-repo

$ git init

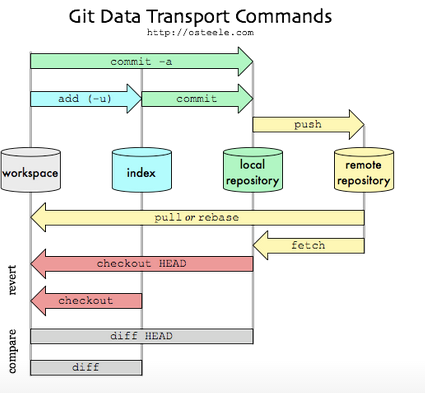
$ git remote add origin <https://github.com/yourUserNameHere/test-repo.git>

git remote add origin https://github.com/tjzzz/ExData\_Plotting1.git

Clone the Repo将github上的glone到本地

git clone https://github.com/tjzzz/RepData\_PeerAssessment1.git

## 2.6 Basic Git commands



* 第一步， git add ，注意，可反复多次使用，添加多个文件；

git add .  adds all new files

git add -u  updates tracking for files that changed names or were deleted

git add -A  does both of the previous

* 第二步， git commit

git commit -m "message"

* 第三步， git push

Git add .

Git commit –m “”

Git push

在本地新建一个rep

Cd rep

Git init

#关联上

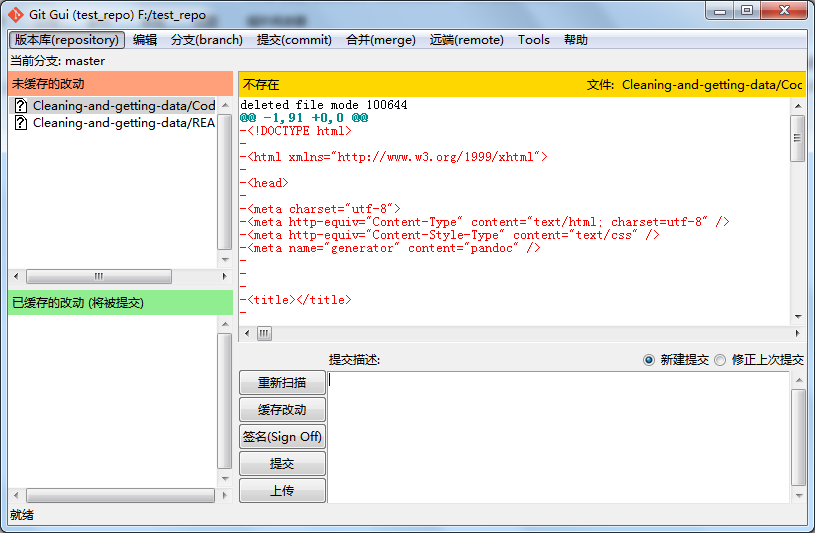
git remote add origin https://github.com/tjzzz/Stackedit.git

Git add .

Git commit –m “”

Git push

GIT GUI界面操作更简洁



**Git鼓励大量使用分支：**

查看分支：git branch

创建分支：git branch *name*

切换分支：git checkout *name*

创建+切换分支：git checkout -b *name*

合并某分支到当前分支：git merge *name*

删除分支：git branch -d *name*

其他介绍可参见 [廖雪峰的博客](http://www.liaoxuefeng.com/wiki/0013739516305929606dd18361248578c67b8067c8c017b000)

## 2.7 Basic markdown commands

Markdown Quick Reference 可以在Rstuido中查到。详细的可以参照 [Using Markdown with RStudio](http://www.rstudio.org/links/using_markdown) to learn more.

#### *Emphasis*

\*italic\* \*\*bold\*\*

\_italic\_ \_\_bold\_\_

#### *Headers*

Header 1

=========================

Header 2

-------------------------

### Header 3

#### Header 4

##### Unordered List

\* Item 1

\* Item 2

\* Item 2a

\* Item 2b

##### Ordered List

1. Item 1

2. Item 2

3. Item 3

\* Item 3a

\* Item 3b

## 3.1 types of data science questions

* Descriptive

描述统计，主要是一些图和表

* Exploratory

探索性数据分析

* Inferential

从样本去推断总体的性质

* Predictive
* Causal 因果分析

设置随机对照组，一般在临床医学生物统计方便比较多。经济方面会有格兰杰因果检验等

* Mechanistic