

## What We Do

THE DATA LAB ENABLES NEW COLLABORATIONS BETWEEN INDUSTRY, PUBLIC SECTOR AND UNIVERSITIES DRIVEN BY COMMON INTERESTS IN THE EXPLOITATION OF DATA SCIENCE. IT PROVIDES RESOURCES AND FUNDING TO KICK-START PROJECTS, DELIVER SKILLS AND TRAINING, AND HELP TO DEVELOP THE LOCAL ECOSYSTEM BY BUILDING A COHESIVE DATA SCIENCE COMMUNITY.

### Collaborative Innovation

The main focus of The Data Lab is to strengthen Scotland's local industry and transfer world-leading research in informatics and computer science into the global marketplace.

[READ MORE](#)

### Skills & Training

The skills needed to exploit the data opportunity are in great demand and short supply. The Data Lab will work with industry and seek to create a pipeline of talented data scientists, equipped with the relevant skills.

[READ MORE](#)

### Community Building

The Data Lab connects world-leading researchers and data scientists with local industry and public sector organisations, giving them access to experts who can help collaborate on solutions to key problems.

[READ MORE](#)

# API's and the open source data science ecosystem

Roman Popat - Data Scientist  
[roman.popat@thedatalab.com](mailto:roman.popat@thedatalab.com)  
twitter: @rmnppt  
web: [rmnppt.com](http://rmnppt.com)  
Github: [rmnppt](https://github.com/rmnppt)

# What is an API?

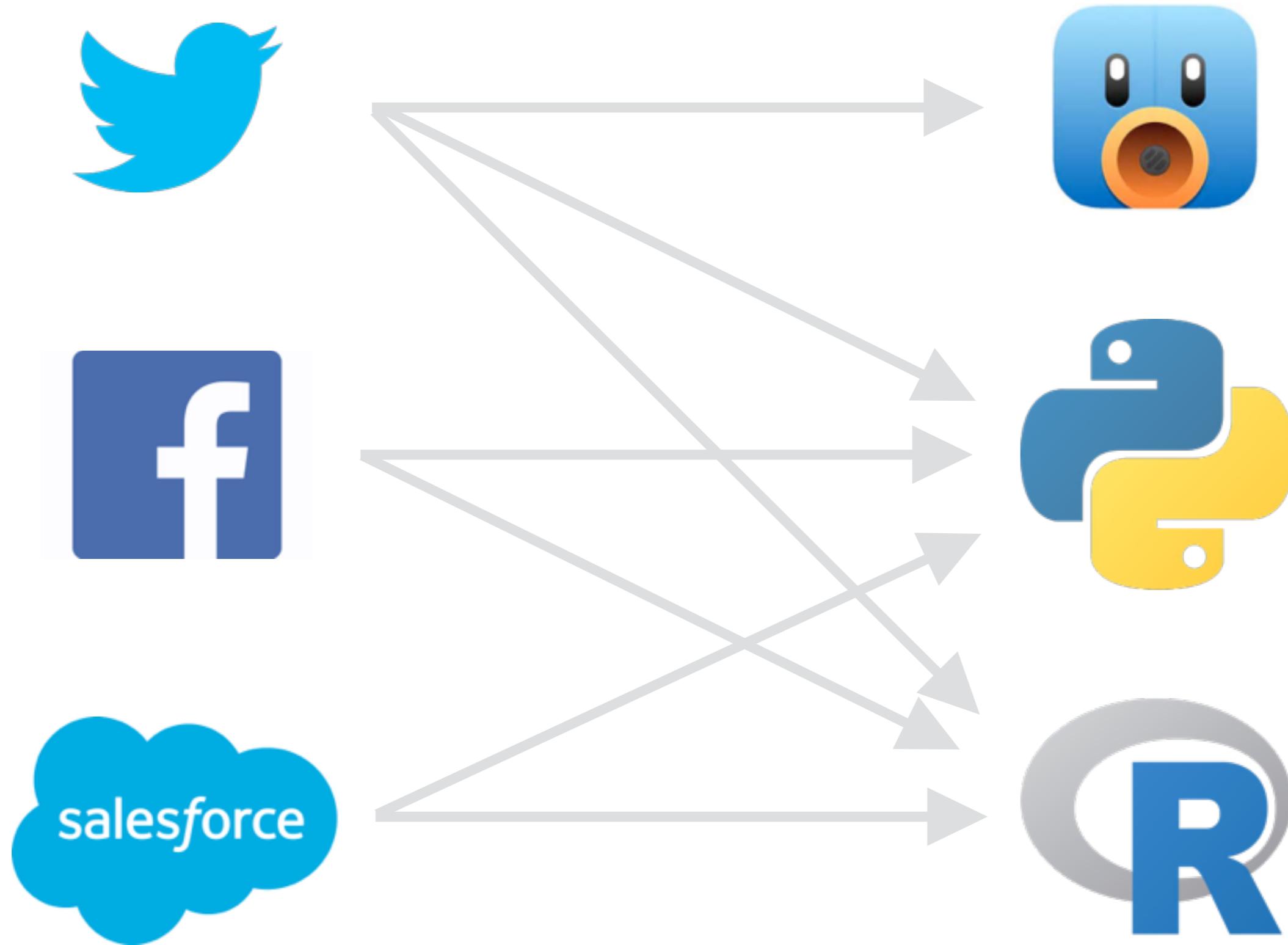
## Web APIs [ edit ]

---

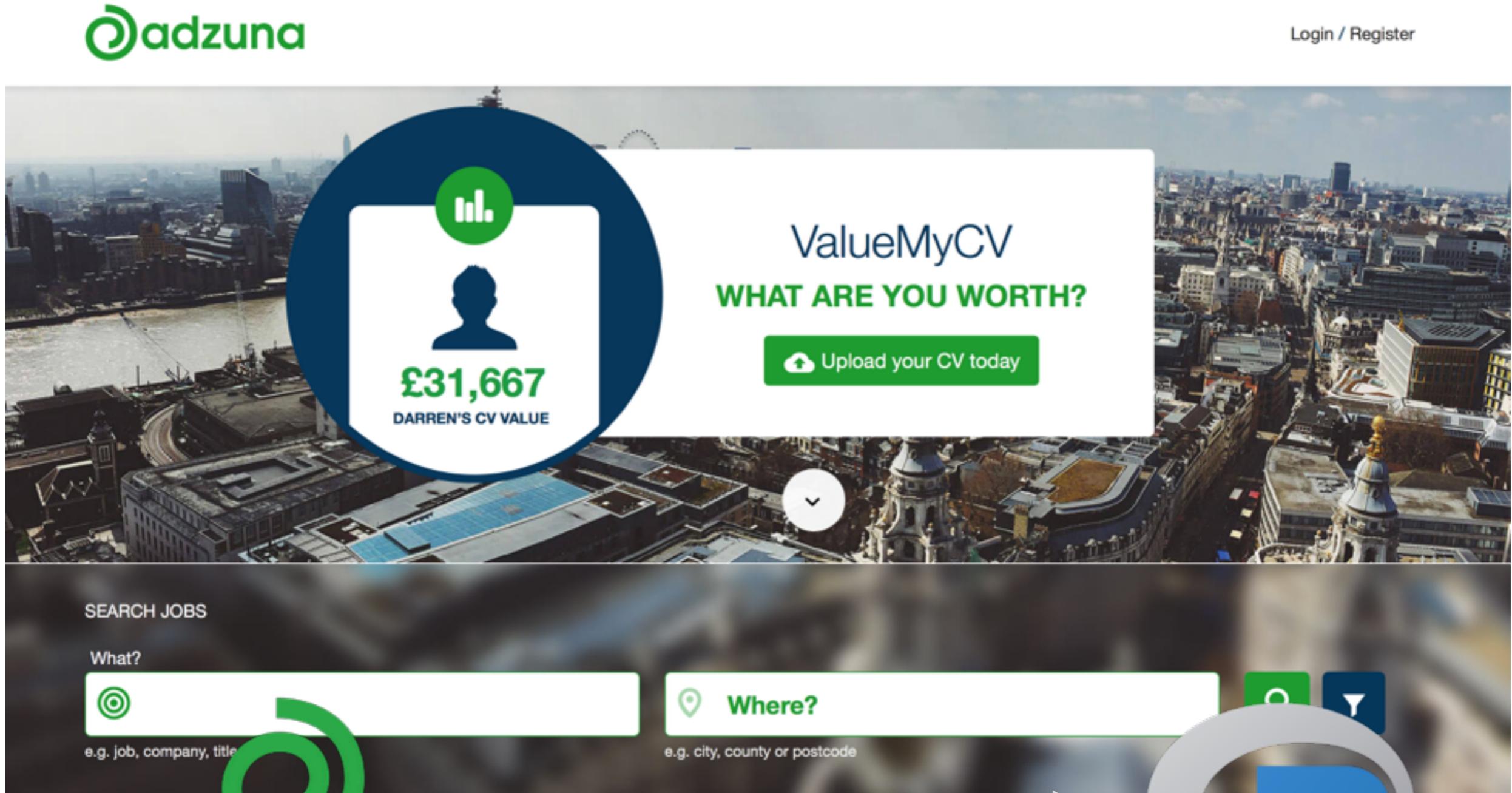
*Main article: [Web API](#)*

Web APIs are the defined interfaces through which interactions happen between an enterprise and applications that use its assets. An API approach is an architectural approach that revolves around providing programmable interfaces to a set of services to different applications serving different types of consumers.<sup>[10]</sup> When used in the context of [web development](#), an API is typically defined as a set of [Hypertext Transfer Protocol](#) (HTTP) request messages, along with a definition of the structure of response messages, which is usually in an Extensible Markup Language ([XML](#)) or JavaScript Object Notation ([JSON](#)) format. While "web API" historically has been virtually synonymous for [web service](#), the recent trend (so-called [Web 2.0](#)) has been moving away from Simple Object Access Protocol ([SOAP](#)) based web services and [service-oriented architecture](#) (SOA) towards more direct [representational state transfer](#) (REST) style [web resources](#) and [resource-oriented architecture](#) (ROA).<sup>[11]</sup> Part of this trend is related to the [Semantic Web](#) movement toward [Resource Description Framework](#) (RDF), a concept to promote web-based [ontology engineering](#) technologies. Web APIs allow the combination of multiple APIs into new applications known as [mashups](#).<sup>[12]</sup>

# Some examples



# A case study



The screenshot shows the Adzuna homepage. At the top left is the Adzuna logo. At the top right are 'Login / Register' links. A large banner in the center features a blue circular badge with a green icon and the text '£31,667 DARREN'S CV VALUE'. To the right of the badge, the text 'ValueMyCV WHAT ARE YOU WORTH?' is displayed, along with a green button that says 'Upload your CV today'. Below the banner is a search bar with fields for 'What?' and 'Where?'. The 'What?' field contains a placeholder 'e.g. job, company, title' and has a magnifying glass icon. The 'Where?' field contains a placeholder 'e.g. city, county or postcode' and has a location pin icon. A large grey arrow points from the search bar area towards the bottom right, where there is a large blue 'QR' logo.

adzuna



[Home]

Download

CRAN

R Project

About R

Logo

Contributors

What's New?

Reporting Bugs

Development Site

Conferences

Search

R Foundation

Foundation

Board

# The R Project for Statistical Computing

## Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

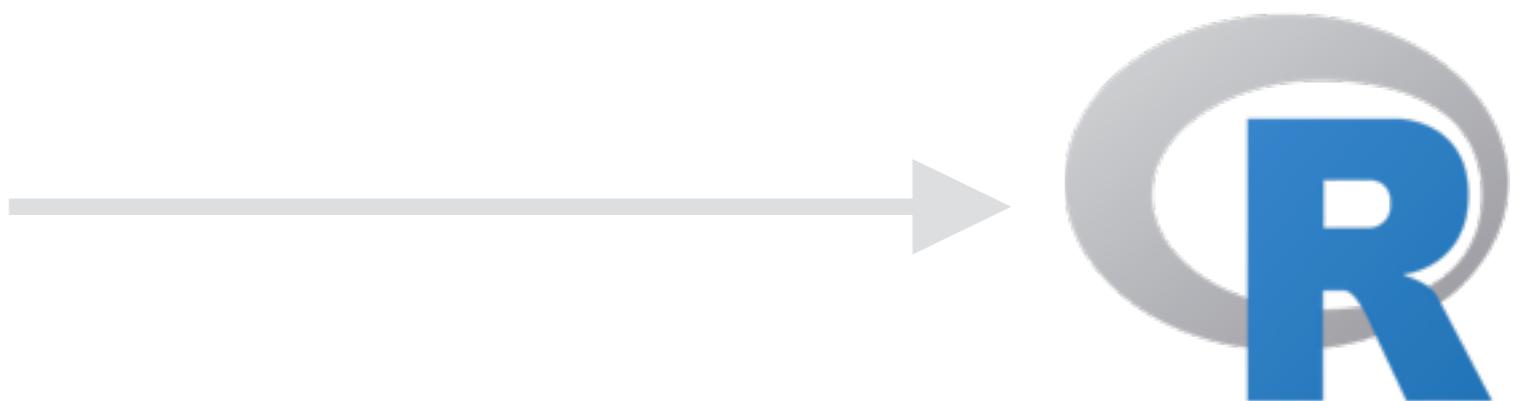
## News

- [The R Journal Volume 8/1](#) is available.
- The useR! 2017 conference will take place in Brussels, July 4 - 7, 2017, and details will appear here in due course.
- [R version 3.3.1 \(Bug in Your Hair\)](#) has been released on Tuesday 2016-06-21.
- [R version 3.2.5 \(Very, Very Secure Dishes\)](#) has been released on 2016-04-14. This is a rebadging of the quick-fix release 3.2.4-revised.



# What does an API call look like?

`api.adzuna.com:80/v1/api/jobs/gb/search/  
1?  
app_id=134859f2&app_key=1d215eccbc34d374  
30ce8693d575c51b&what=data%2Bscience`



# Data returned from API



# JSON



A screenshot of a web browser window titled "Adzuna API". The address bar shows the URL "api.adzuna.com/v1/api/jobs/gb/search/1?app\_id=134859f2&amp;app\_key=1d215eccbc34d37430ce8693d575c51b&amp;what=data%2Bscier". The page content displays a JSON response structure.

## CLASS

Adzuna::API::Response::JobSearchResults

### count

1452

### mean

56323.02

### results

#### 0. CLASS

Adzuna::API::Response::Job

##### adref

eyJhbGciOiJIUzI1NiJ9.eyJpjoIiDE4NTI5MDg4IiwicyI6IjBGU0F2QkRIVGhxTXFoaWFVZ1B0RmcifQ.gX0vUPgbRUd\_WX9mxpw\_HbXHD8Bc\_iXx2FVfmoOiycw

##### category

#### CLASS

Adzuna::API::Response::Category

##### label

Accounting & Finance Jobs

##### tag

accounting-finance-jobs

### company

#### CLASS

Adzuna::API::Response::Company

##### canonical\_name

Endsleigh Insurance

##### display\_name

Endsleigh Insurance

### contract\_type

permanent

### created

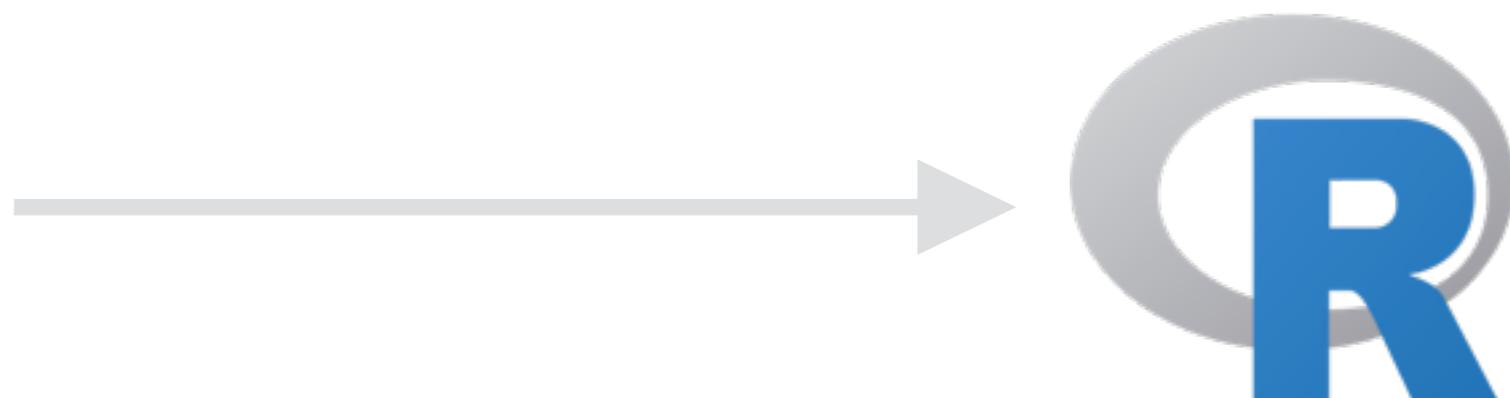
# The R function

```
get_country_page <- function(  
  keyword,  
  country,  
  app_id,  
  app_key,  
  page  
) {  
  
  this_url <- paste0(  
    "http://api.adzuna.com:80/v1/api/jobs/",  
    country,  
    "/search/",  
    page, "?",  
    "app_id=", app_id,  
    "&app_key=", app_key,  
    "&results_per_page=50",  
    "&what=", keyword  
)  
  
  dat <- fromJSON(this_url)  
  
  return(dat)  
}
```



# Using the function

```
get_country_page(  
    keyword = "data science",  
    country = "gb",  
    app_id = id,  
    app_key = key,  
    page = 1  
)
```



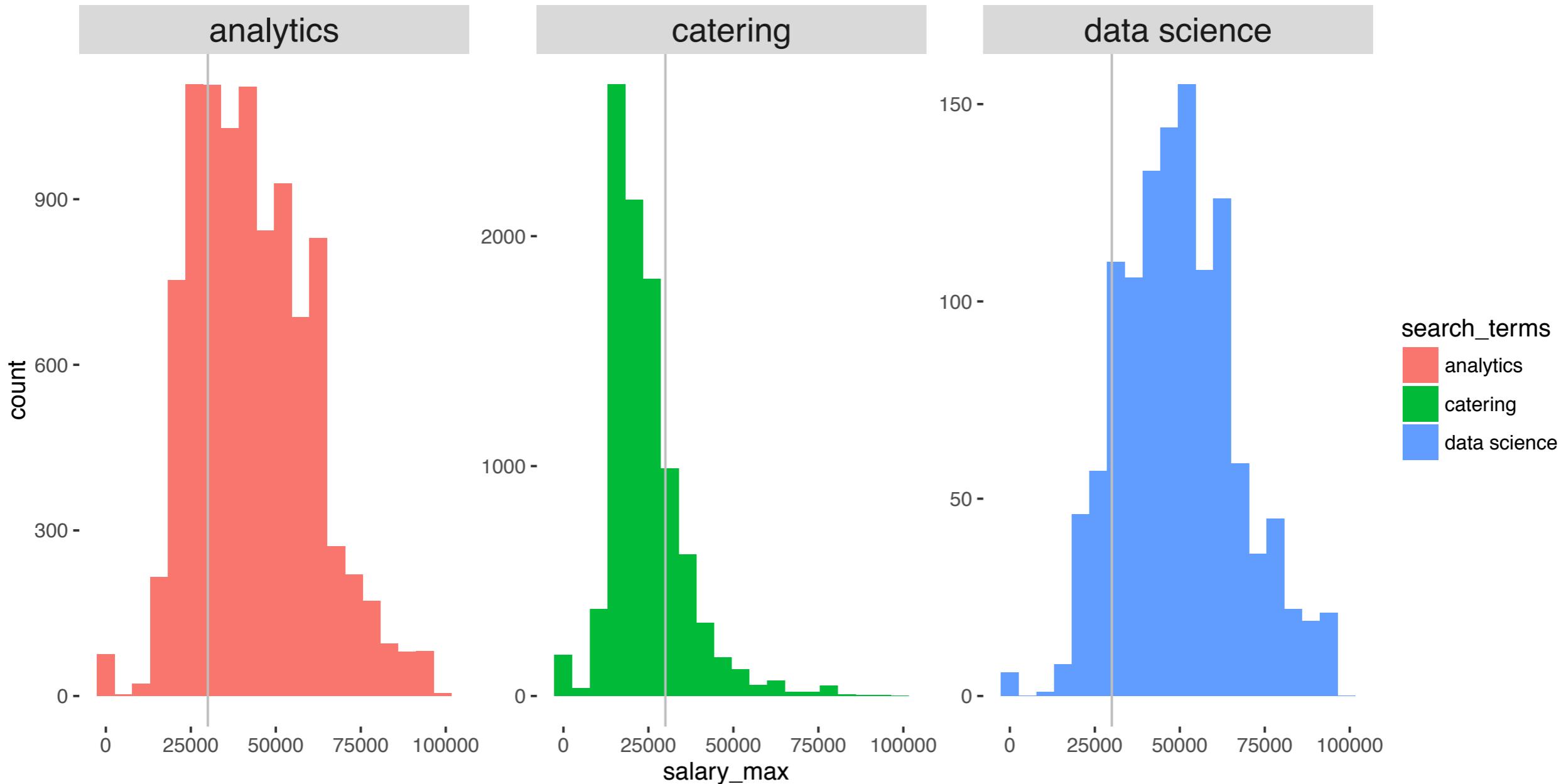
# Why abstract this? -Reproducibility



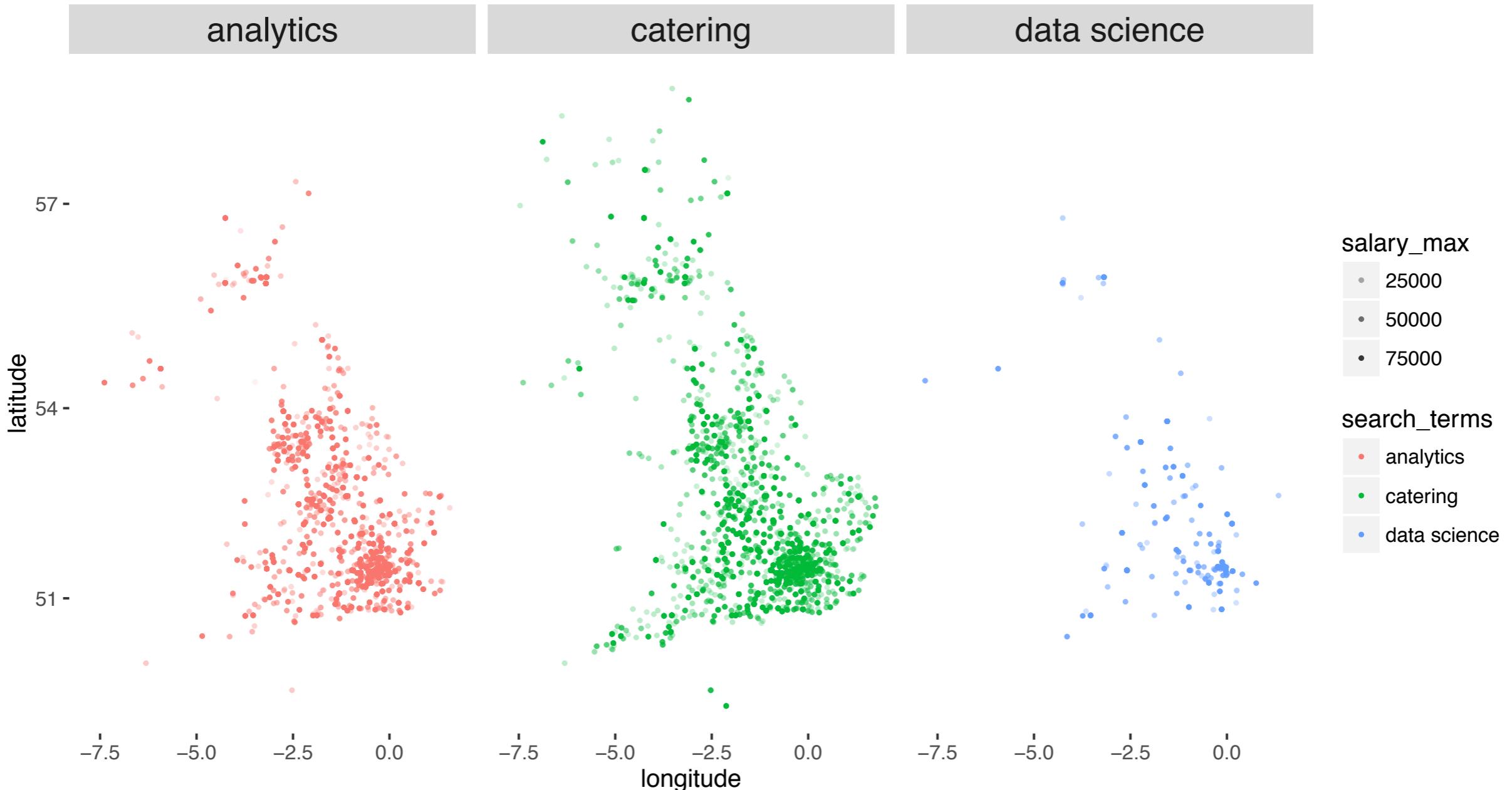
Make several queries at one time

Make repeated queries through time

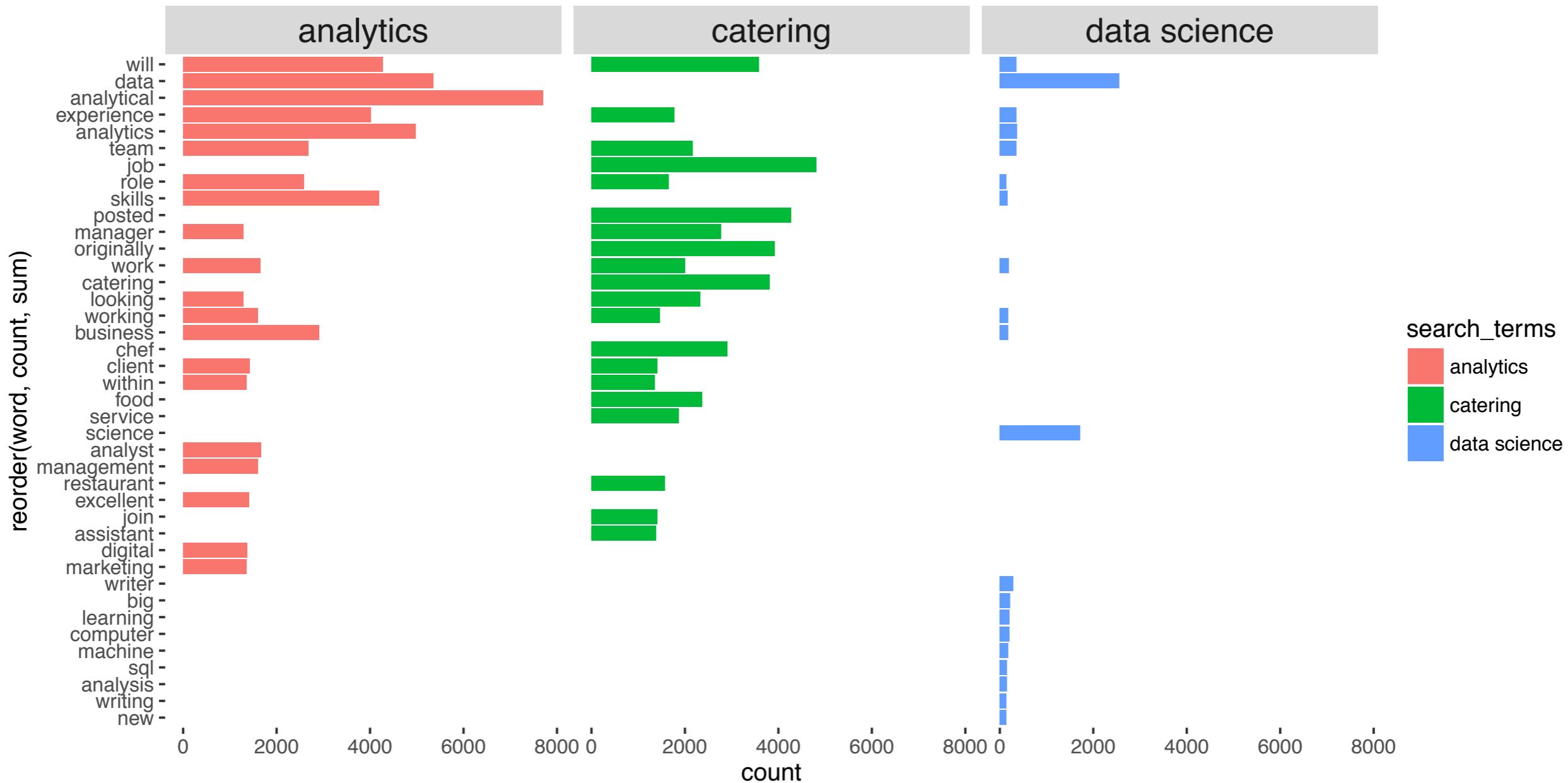
# Salaries



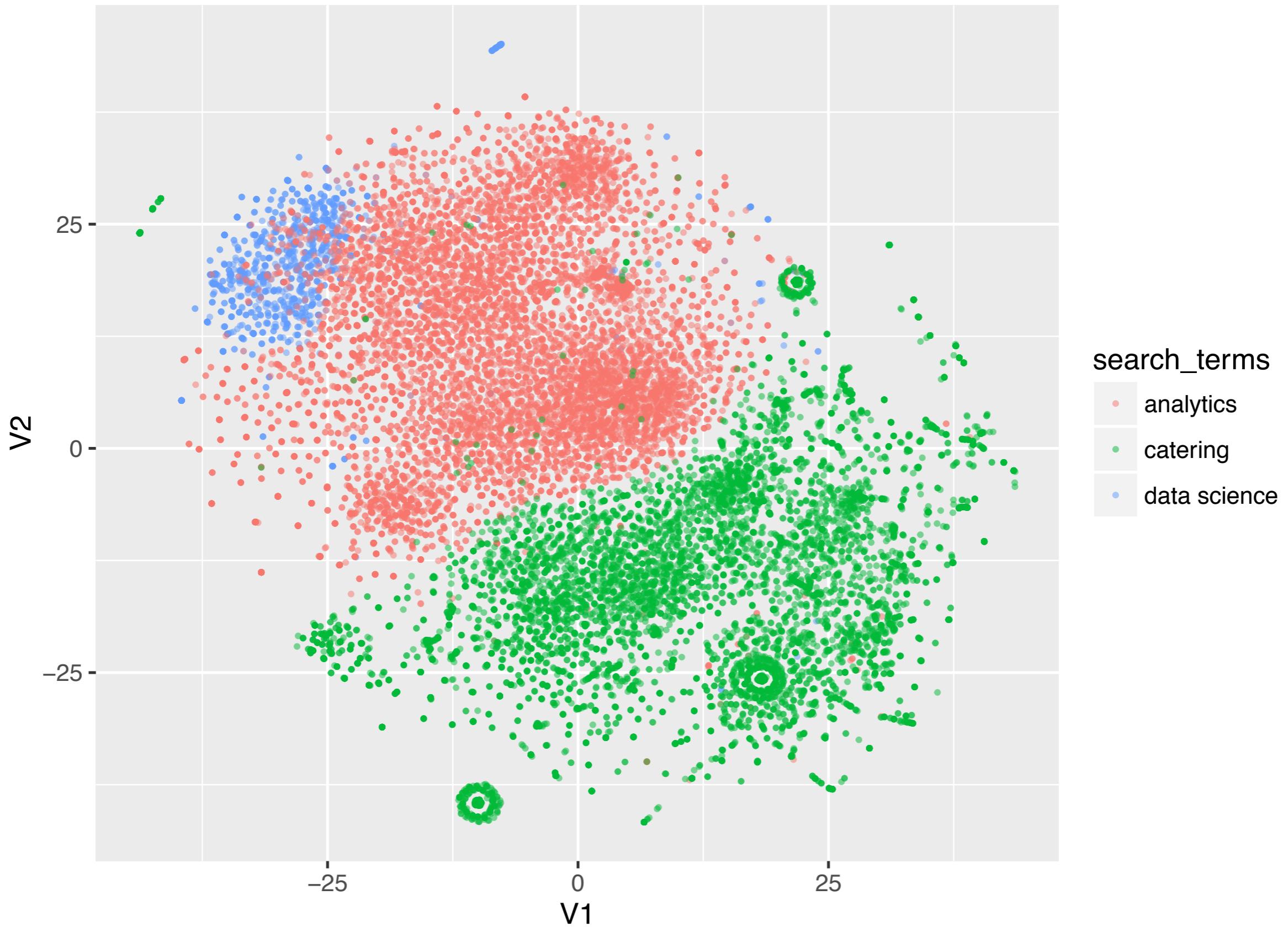
# Geographies



# Word frequencies



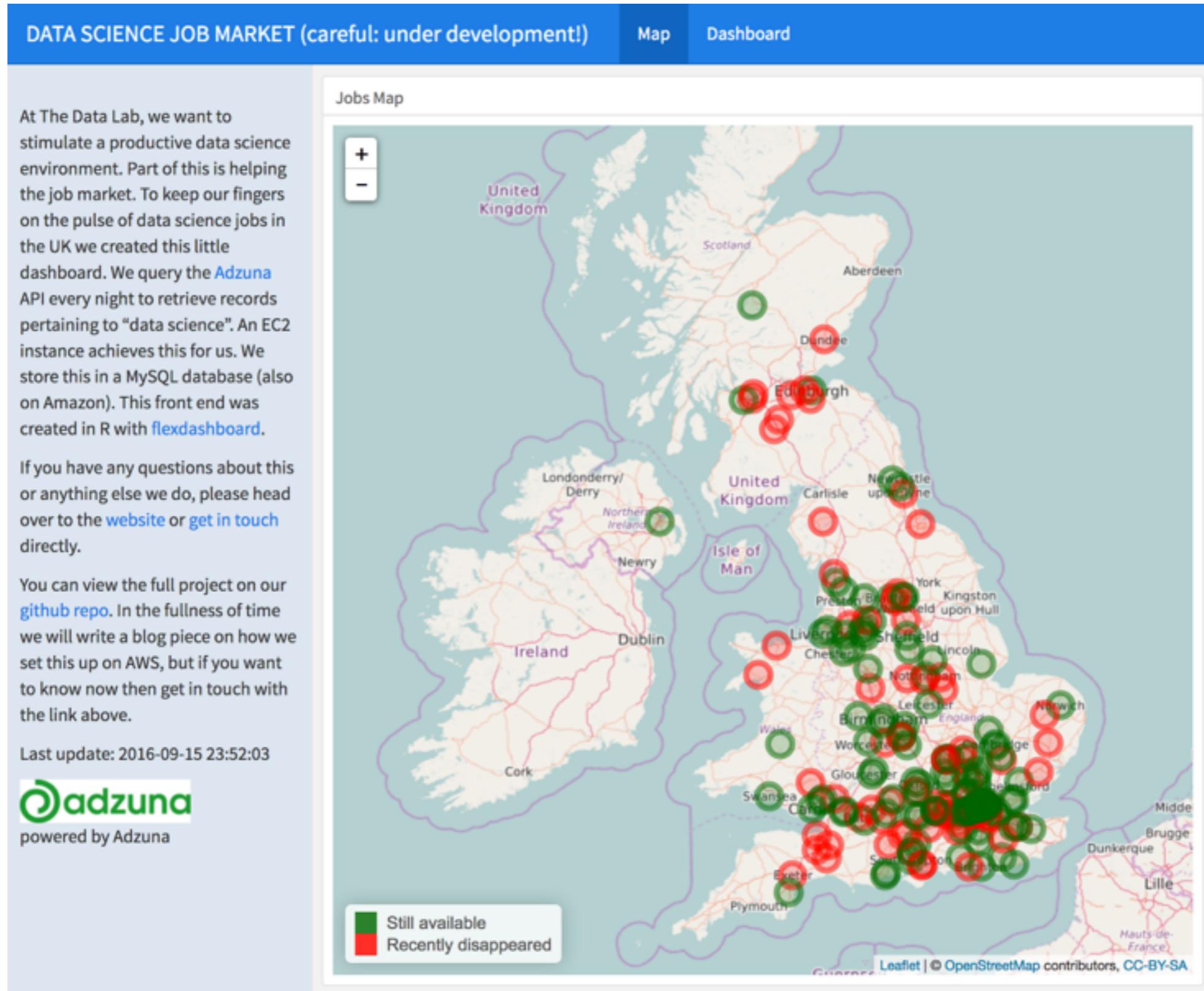
# Language (tSNE)



# Analysing language

1. word count
  - `library(tm)`
2. determine similarity
  - `library(ecodist)`
3. compress for visualisation
  - `library(Rtsne)`
4. plot
  - `library(ggplot2)`

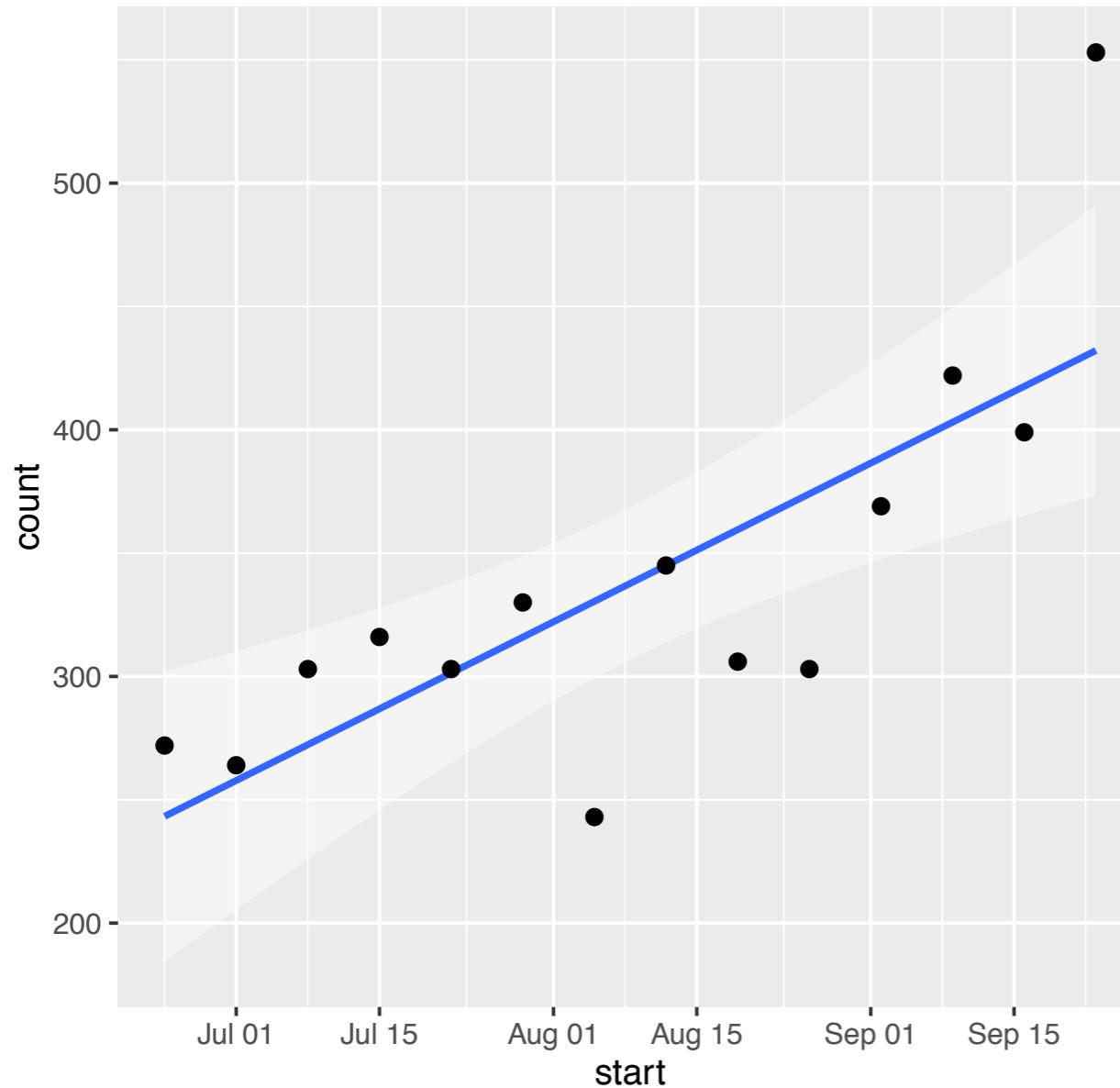
# We're stashing some data



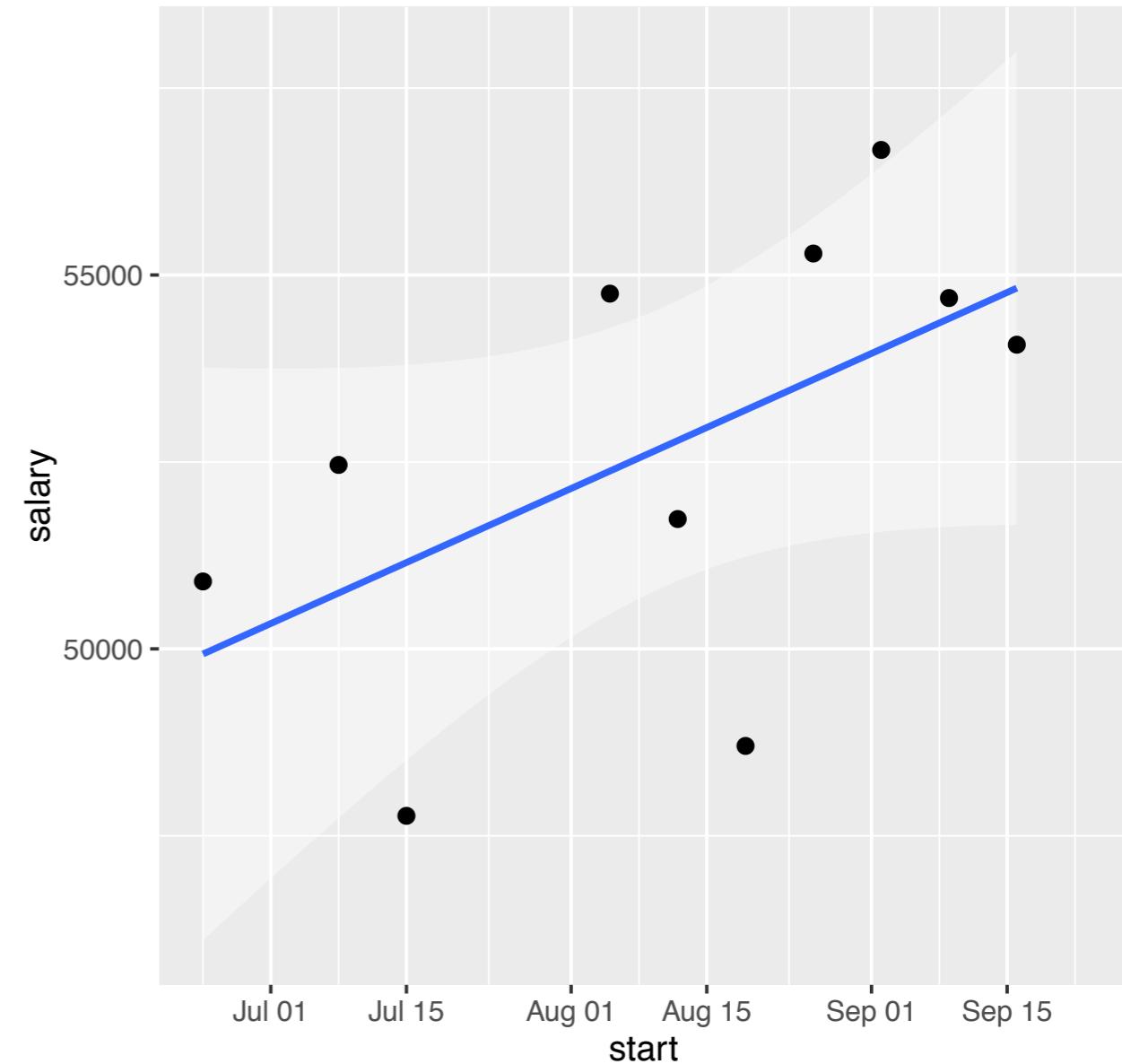
<https://s3-eu-west-1.amazonaws.com/adzunadata/dashboard/AdzunaFlexDashboard.html>

# Historical trends

14.53 additional ads per week



407.81 £ per week



# Benefits of abstraction

api.adzuna.com:443/v1/api/jobs/gb/  
search/1?  
app\_id=134859f2&app\_key=1d215eccbc34d374  
30ce8693d575c51b&what=data%2Bscience

# Sharing Code

R session

```
create("adzunar")
```



Terminal

```
git init  
git add  
git commit  
git push
```



# Sharing Code

```
devtools::install_github("rmnppt/adzunar")
```

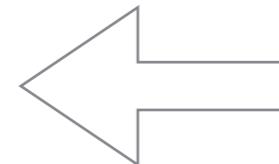
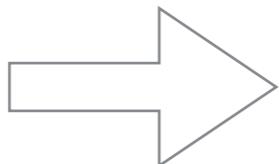
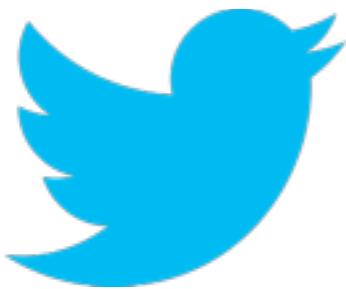
```
library(adzunar)
```

```
get_country_page()
```



# Why do we need bridges?

Assets



Techniques

Xgboost

randomForest

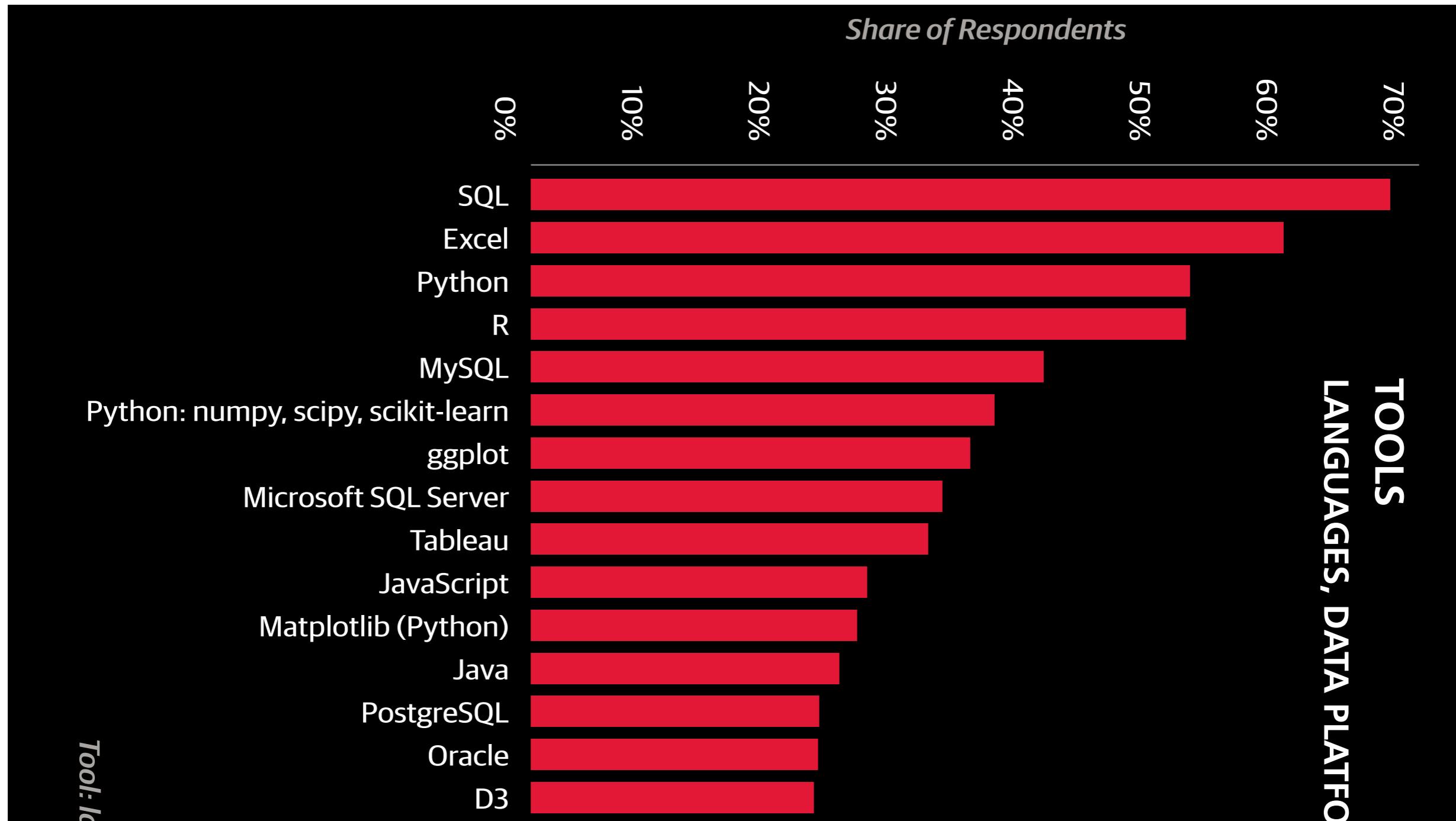
tSNE

DBScan

svm

convNet

# Insert title



<https://www.oreilly.com/ideas/2015-data-science-salary-survey>

# Why is OS so popular?



Free



Trustworthy

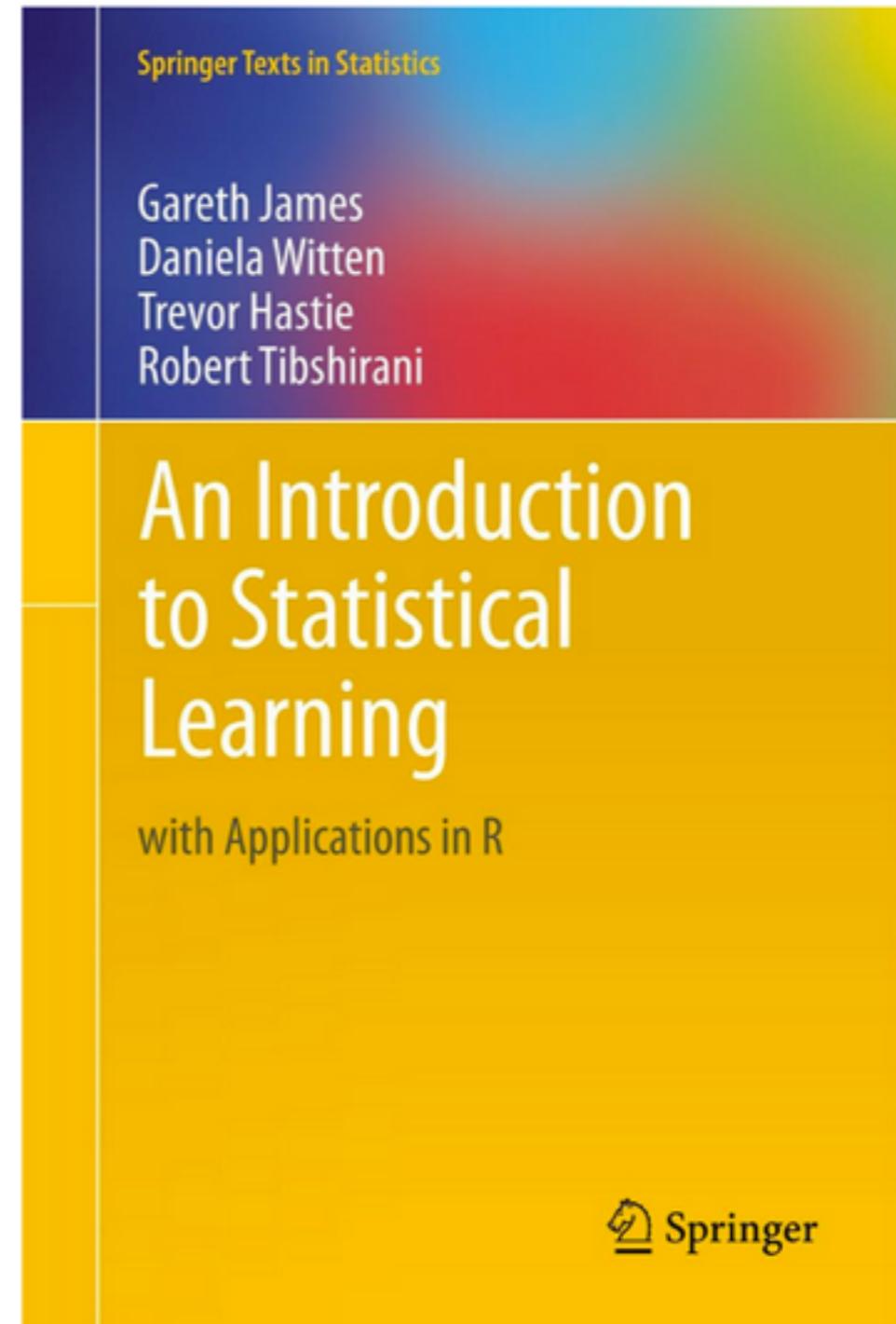
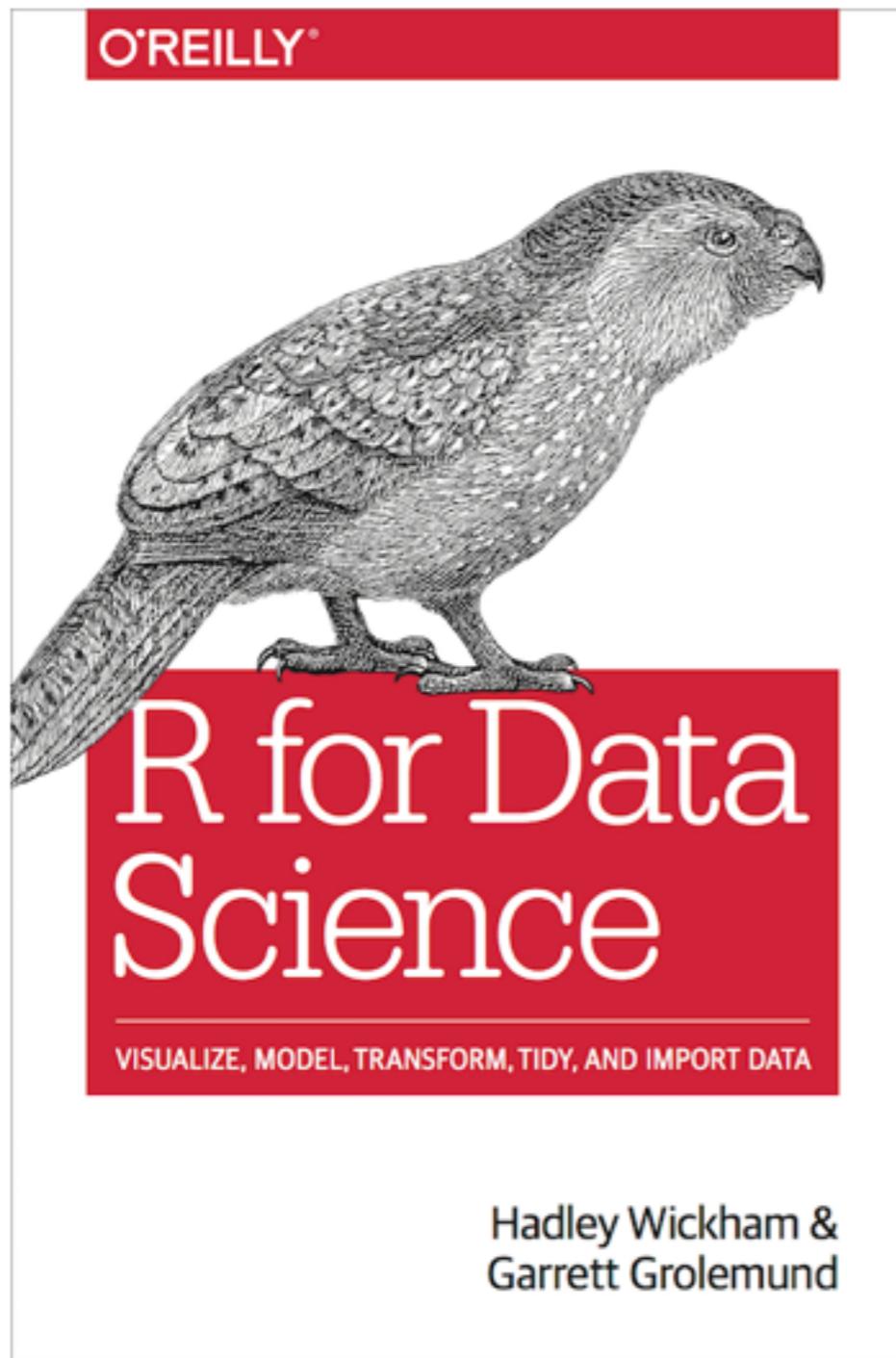


Reproducible



Extensible

# Starting Out



# Links etc

<https://developer.adzuna.com/>

<https://www.r-project.org/>

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

<http://r4ds.had.co.nz/>

<http://www-bcf.usc.edu/~gareth/ISL/>

<https://en.wikipedia.org/wiki/T->

[distributed\\_stochastic\\_neighbor\\_embedding](#)

[https://s3-eu-west-1.amazonaws.com/adzunadata/  
dashboard/AdzunaFlexDashboard.html](https://s3-eu-west-1.amazonaws.com/adzunadata/dashboard/AdzunaFlexDashboard.html)

# API's and building bridges in Data Science

Roman Popat - Data Scientist  
[roman.popat@thedatalab.com](mailto:roman.popat@thedatalab.com)  
twitter: @rmnppt  
web: [rmnppt.com](http://rmnppt.com)  
Github: [rmnppt](https://github.com/rmnppt)

# Pros and Cons

Microsoft R  
R user group

# Insert title