

Open Data Science

<http://training.theodi.org/ODS>

David Tarrant · @davetaz



Course bought to you in collaboration with the
European Data Science Academy Project

Introductions

Your name & role?

Why is data science important to you?

What do you want to do differently after this course?



Course aim

Equip you with the knowledge and tools to help you upskill as modern data scientists.



Session 1

Discovering Open Data Science

Session 2

Machine learning and classification

Session 3

Visualisation and communication

Session 1

Discovering Open Data Science



Outcomes

Define open data science

Describe a number of key data science stories

Identify the characteristics of open data science projects



Prof. Dr. Frank Bensberg

**Osnabrück University of Applied Science
Faculty of Business Management &
Social Sciences
Capriviustrasse 30a
D-49076 Osnabrück
F.Bensberg@hs-osnabrueck.de**

List the 4 aspects of

Scientist

Analyst

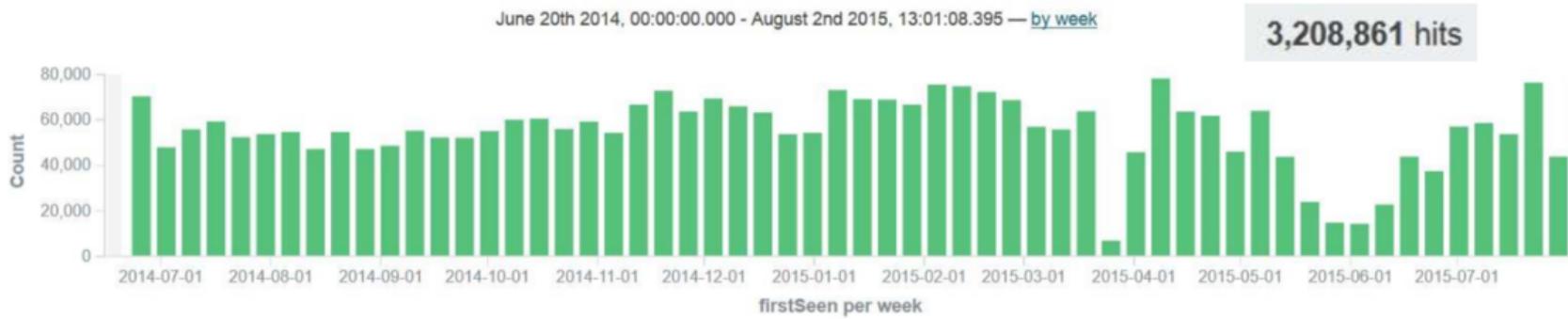
- Domains • Products
- Concepts • Soft skills

Engineer

Visualiser

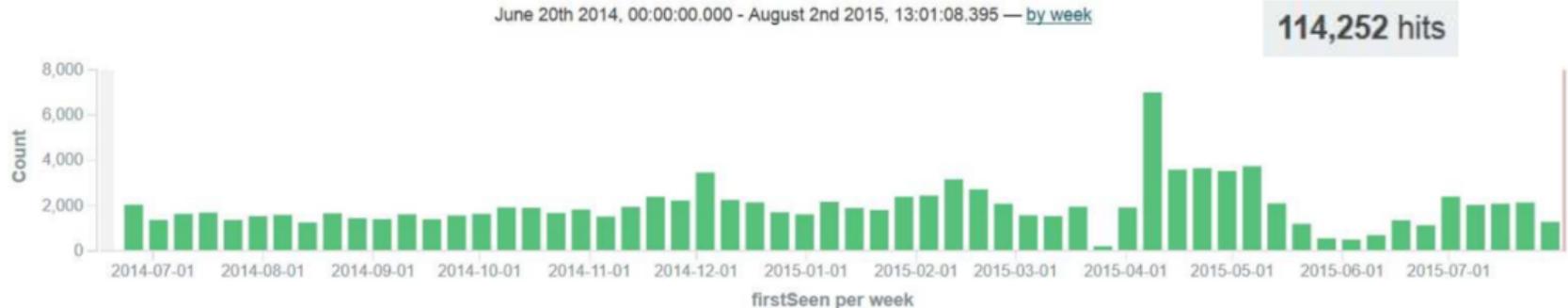
Database

Total Ads Collected



Big Data Ads

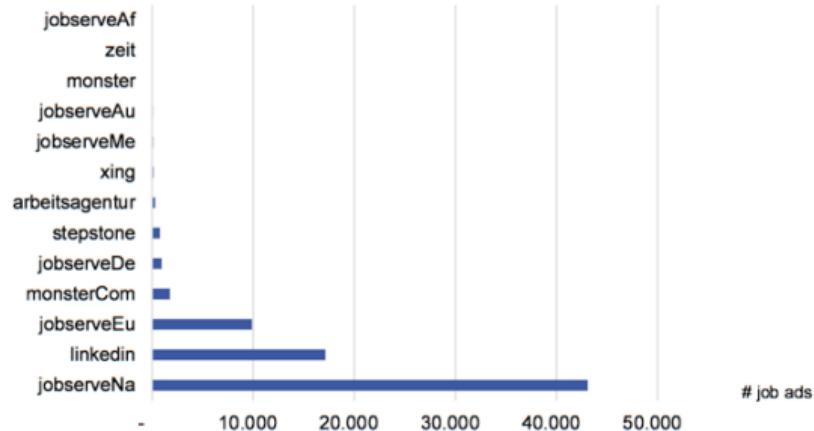
description:("big data" OR "data analytics" OR "data scien")



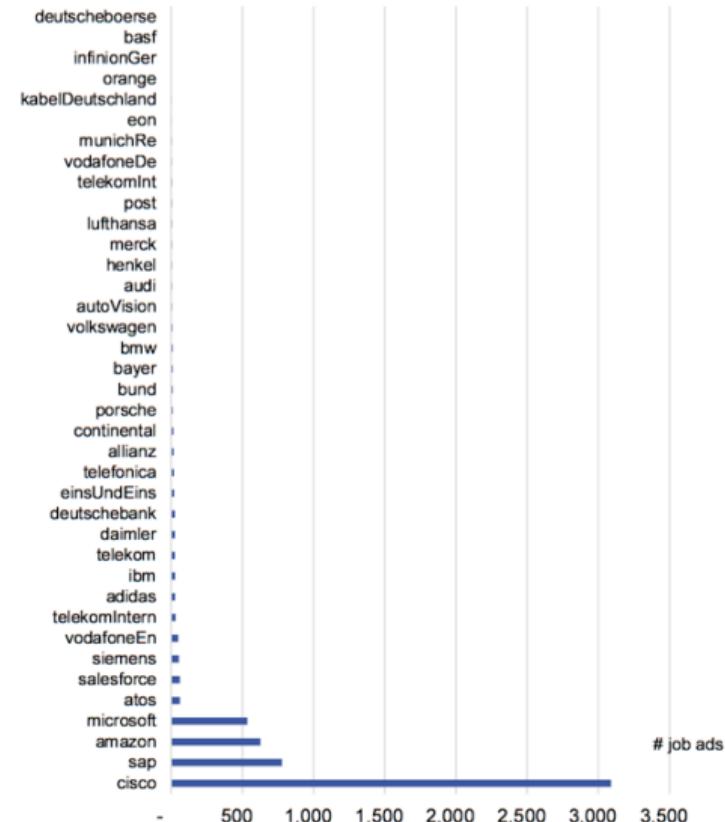
Data Set

- Basic data pool comprises about 3m of job ads for the ICT sector collected since June 2014
- Job ads are periodically collected from public job portals (jobserve, monster, ...) and from company-specific portals (Cisco, SAP, ...)
- Selection of job ads which contain at least one of the keywords *big data*, *data science*, *data scientist*, *smart data* or *fast data* (n=80.014)

Public Job Portals



Company-specific Portals



Data Scientist

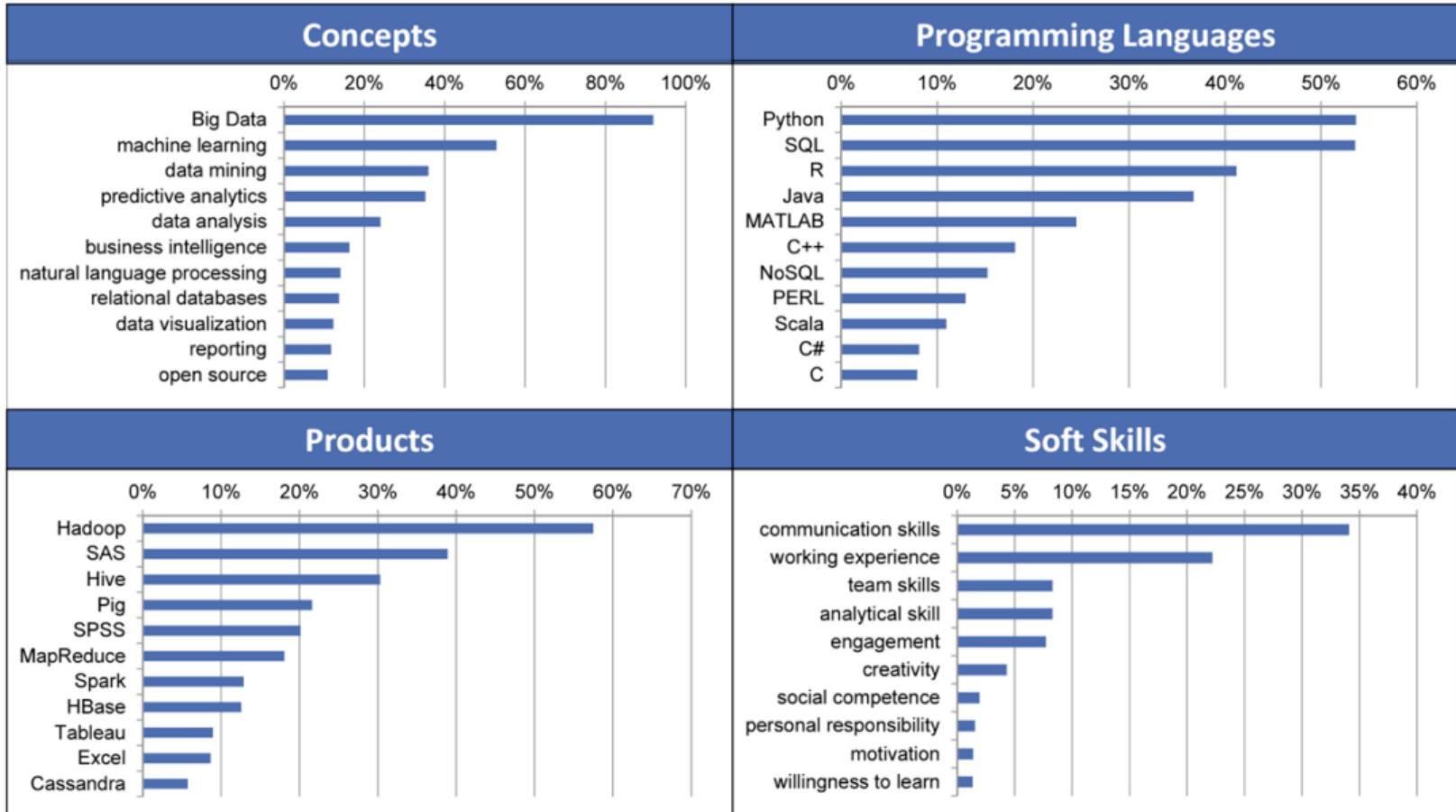
Data Set

# Job Ads	1.942	Sample Job Titles
Organizations (Top 10)	Liberty Personnel Services; Opus Recruitment Solutions Ltd; BigR.io; Career Brokers; Next Ventures Ltd; ITBM Consulting Inc;A3Logics US;Engage3;All In Analytics;Alivia technology	Data Scientist Big Data Scientist Senior Data Scientist Principal Data Scientist (R, Matlab, [...], Spark) Senior Data Scientist (Python, Matlab, R, SQL)

Professional Activities

Position	Freq	Object	Activities
1	39%	analytics	work, develop, use, include, provide, require, deliver, create, apply
2	39%	model	work, develop, provide, include, unite, use, deliver, look, analyze, create
3	37%	analysis	work, develop, use, provide, deliver, look, create, design, apply
4	35%	statistic	work, develop, use, provide, require, deliver, analyze, apply, solve
5	27%	algorithm	work, develop, use, provide, solve, apply, require, deliver, build
6	27%	machine learning	work, develop, use, apply, solve, build, design, deliver
7	24%	data set	work, develop, use, include, analyze, provide, require, design, deliver, create
8	11%	decision tree	develop, work, design, create, join, require, provide, utilize, deliver, grow
9	10%	data source	work, use, develop, require, provide, apply, help, prefer, understand
10	4%	data analysis	work, use, develop, provide, require, bring, build, drive, apply

Data Scientist

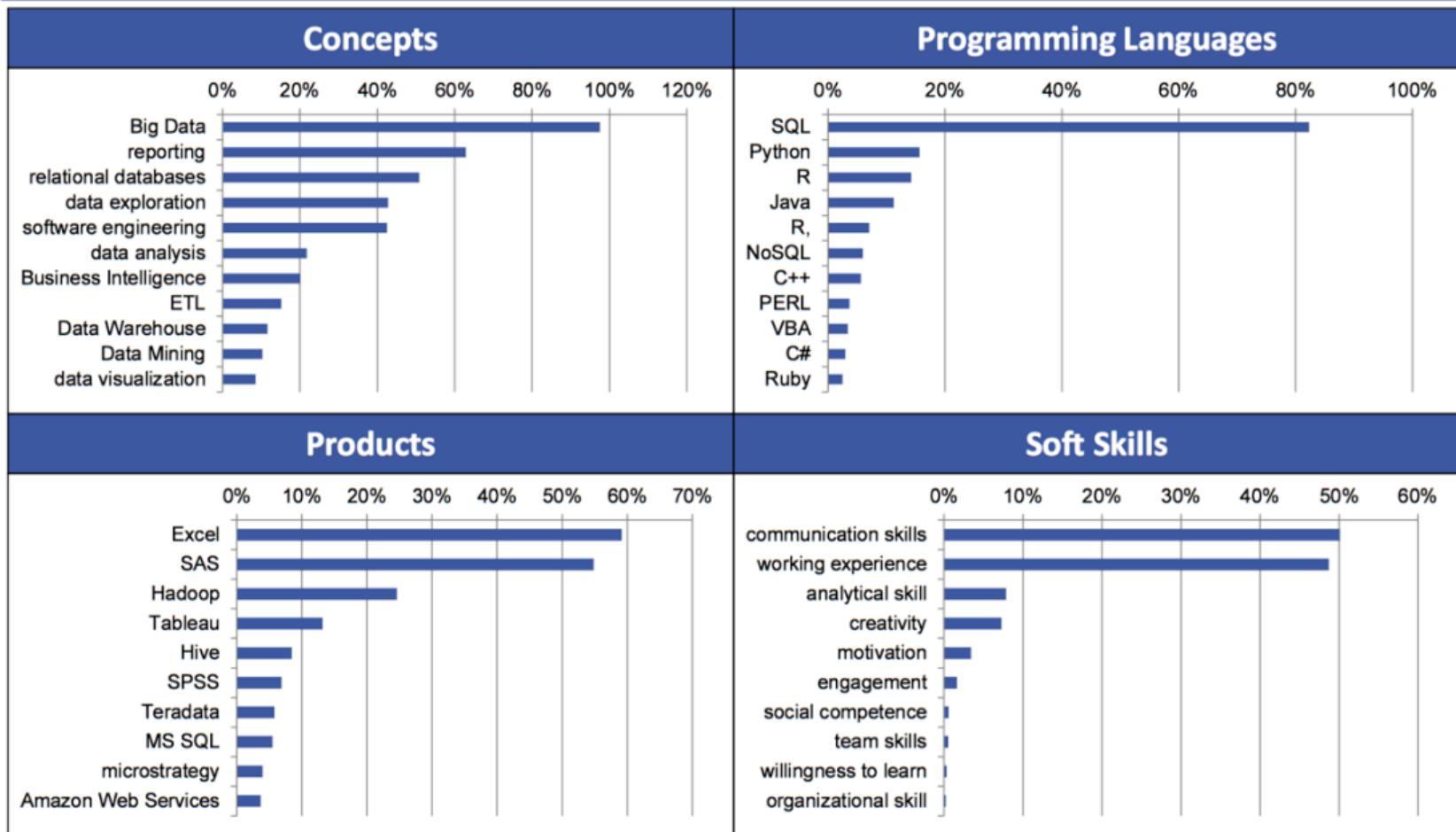


Data Analyst

Data Set			
# Job Ads	1.690	Sample Job Titles	
Organizations (Top 10)	MedeAnalytics, Liberty Personnel Services, FILD, TalentRISE, R Square, Libertyjobs.com, LS Direct Marketing, The Judge Group, UnitedHealth Group Cognius	Business Data Analyst - Health Plan Solutions Data Analyst Senior Data Analyst Big Data Analyst Data Analyst (BIG DATA - Finance Analytics)	
Professional Activities			
Position	Freq	Object	Activities
1	59%	analytics	work, develop, use, provide, require, support, deliver, solve, look, make, apply
2	55%	model	use, work, provide, develop, analyze, support, make, deliver, apply, require
3	53%	statistic	work, use, provide, develop, require, look, deliver, solve, analyze, make
4	48%	decision	make, analyze, develop, deliver, apply, require, solve, leverage, gather
5	46%	dataset	work, provide, develop, use, deliver, support, amke, apply, require, look, solve
6	44%	workflow	use, provide, analyze, develop, deliver, make, apply, gather, look, leverage
7	43%	risk	analyze, identify, reduce, look, plan, improve, change, demonstrate
8	42%	data exploration	use, develop, help, apply, provide, solve, offer, demonstrate, support, make

Annotation: data set shows a strong sectoral bias (Healthcare).

Data Analyst



Data Engineer

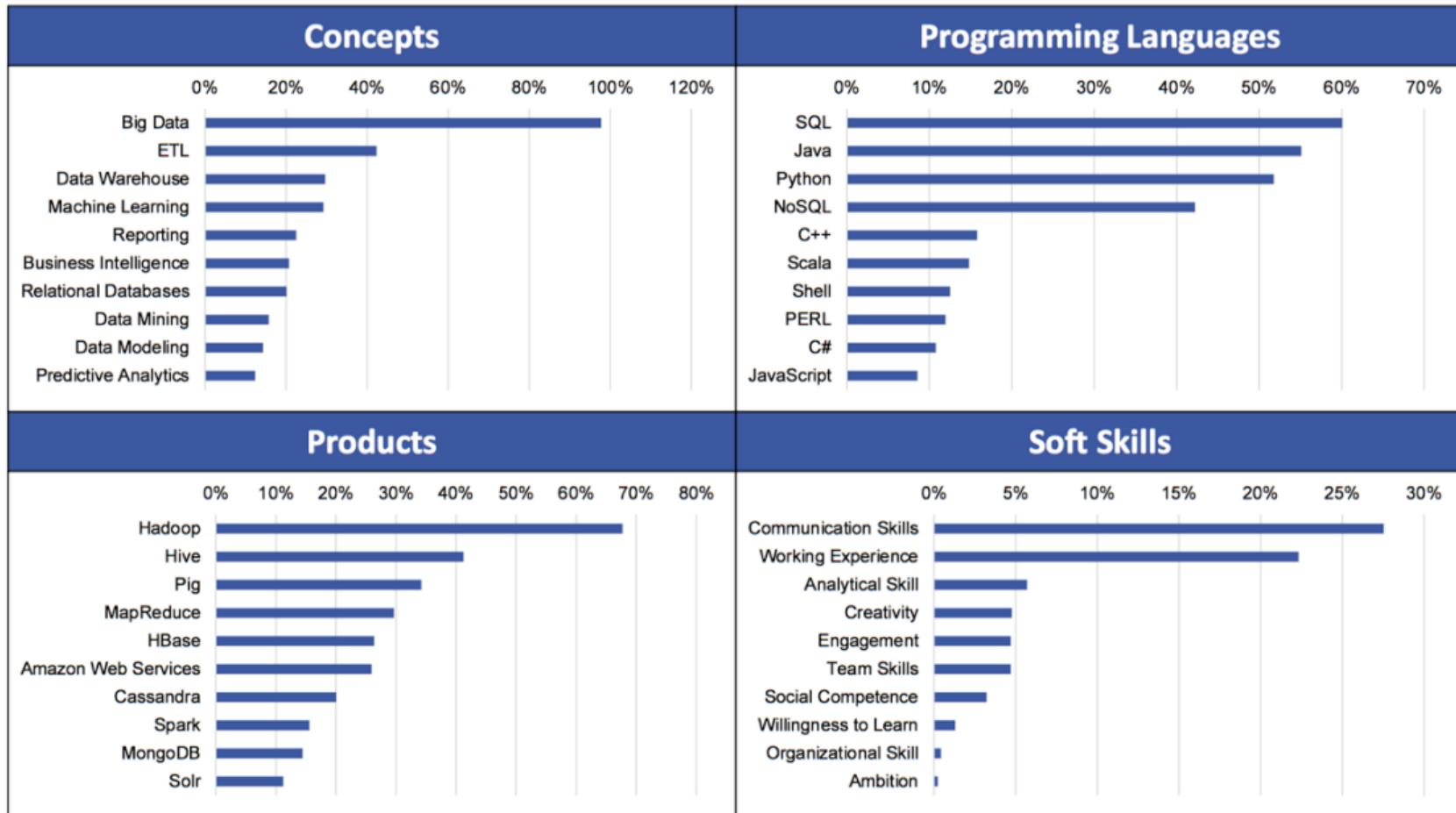
Data Set

# Job Ads	1.613*	Sample Job Titles
Organizations (Top 10)	FILD, Prosum, Liberty Personnel Services, Praedicat, Libertyjobs, Emprise Technologies, Amazon, Teradata, ColdLight, ARC IT Recruitment	Big Data Engineer Data Engineer Senior Data Engineer Senior Big Data Engineer Director of Data Engineering

Professional Activities

Position	Freq	Object	Activities
1	29%	Analytics	work, develop, include, use, design, build, provide, implement, grow
2	28%	ETL	work, develop, design, unite, provide, include, build, implement, create
3	21%	Machine Learning	work, unite, use, develop, design, build, include, grow, create, implement, provide
4	20%	Algorithm	work, use, develop, build, unite, include, create, grow, design, analyze, implement
5	16%	Data Set	work, use, develop, design, unite, include, build, process, create, grow, make, implement, analyse, provide
6	15%	Large Data	work, use, develop, unite, process, create, design, grow, build
7	13%	Unstructured Data	work, develop, include, design, use, unite, build, create, analyze, provide
8	11%	Data Platform	develop, design, work, unite, support, analyze, build, implement, grow, provide
9	11%	Predictive Analytics	work, build, develop, create, use, unite, grow

Data Engineer



Visualization Engineer

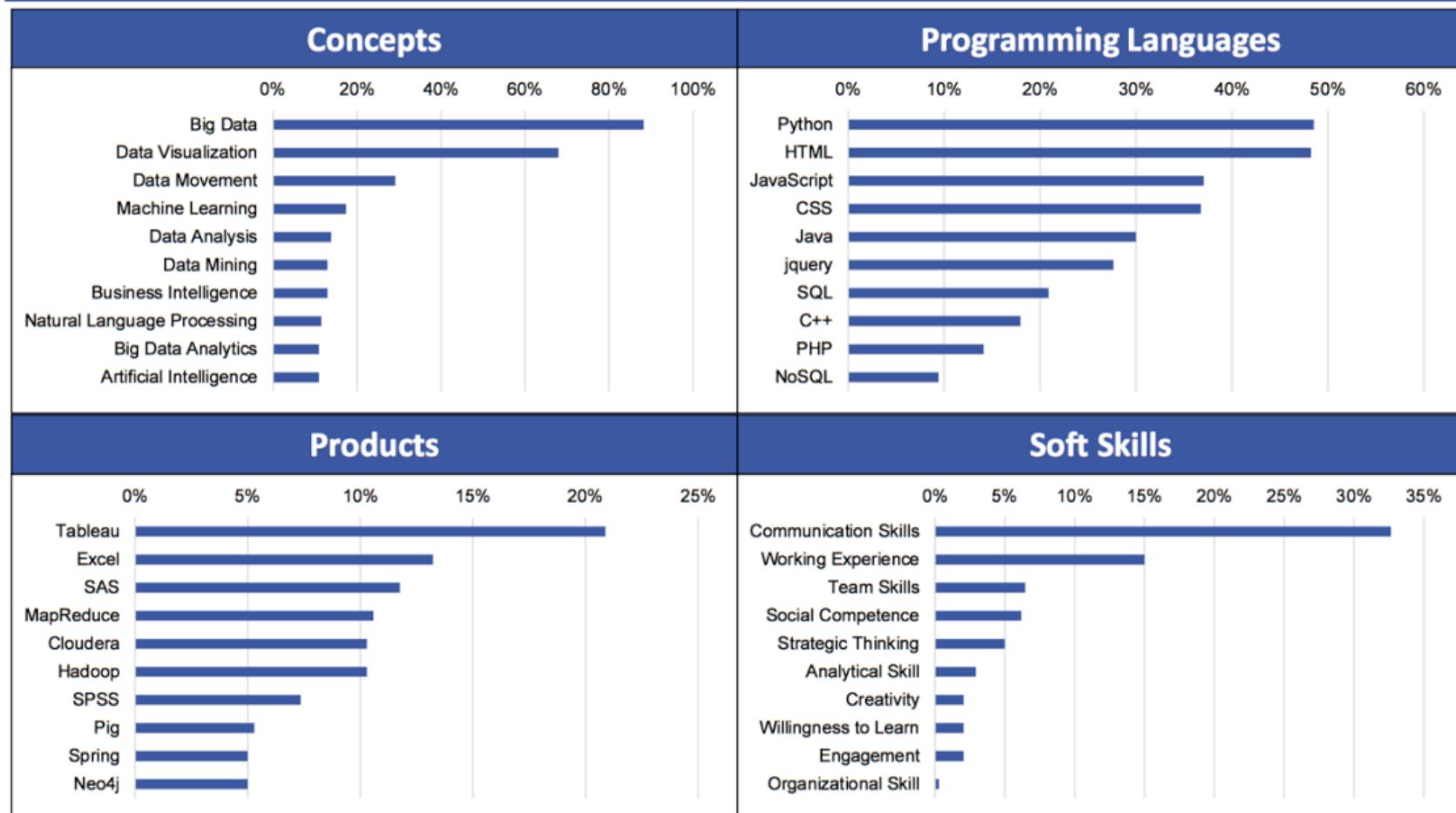
Data Set

# Job Ads	340	Sample Job Titles
Organizations (Top 10)	BigR.io, Infinite Resources, MGRS Group, FIELD, Analyze Corporation, Opus Recruitment, TASC, Splunk, N-Tier Solutions, Globys, AT&T	Data Visualization Engineer Software Engineer - Data Visualization/Graphic Design Java Visualization Software Eng. JavaScript Engineer - Python - Data Visualization - Big Data Challenge Visual & UI Designer - Adobe Creative, HTML, CSS

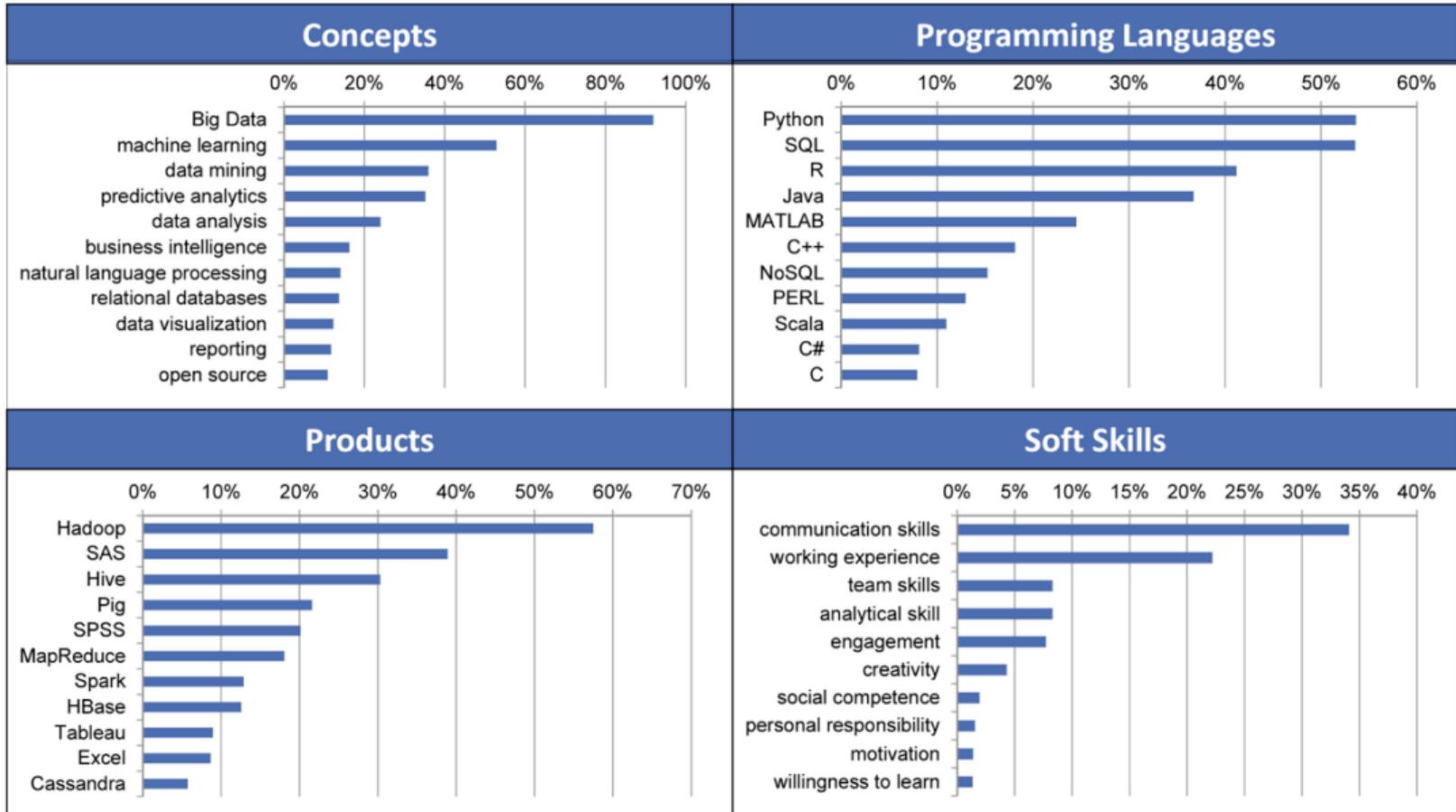
Professional Activities

Position	Freq	Object	Activities
1	44%	analytics	work, use, help, unite, develop, deliver, provide, focus, bring
2	42%	data visualization	work, use, develop, build, deliver, provide
3	31%	interface	provide, deliver, focus, use, bring, poise
4	26%	data model	use, deliver, unite, look, help, work, provide
5	22%	complex data	work, unite, deliver, provide, use
6	22%	chart	use, deliver, provide, build, bring, focus
7	21%	performance data	pinpoint, assure, dig, detail
8	15%	machine learning	develop, work, use
9	14%	interaction	work, develop, design, provide
10	9%	artificial intelligence	program, develop, design, conceptualize, implement

Visualization Engineer



Data Scientist

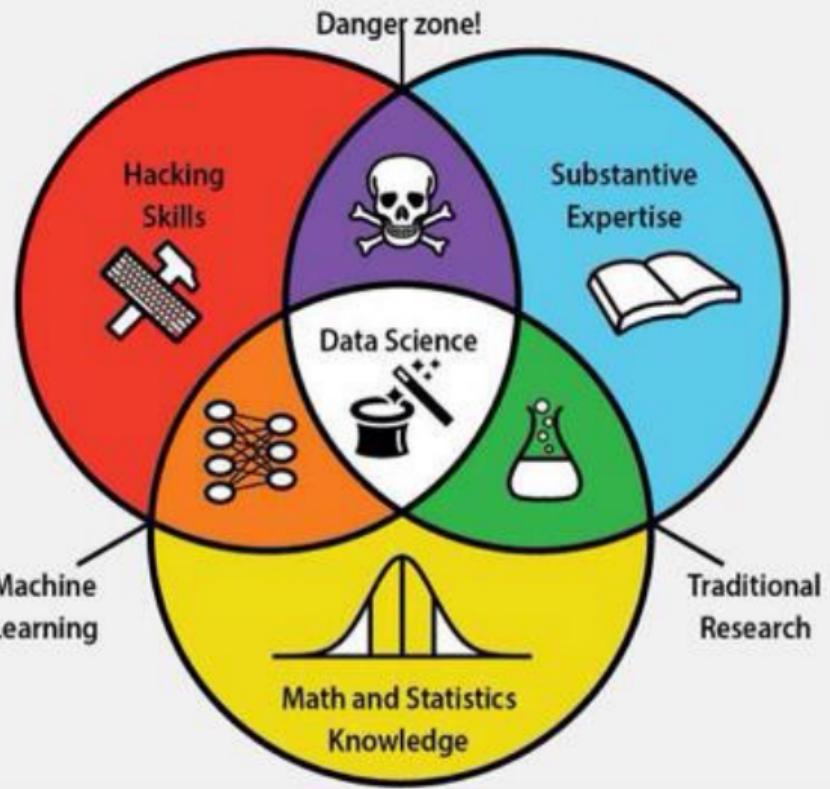


Key areas

- 1) Big Data
- 2) Machine Learning and Prediction
- 3) Data Collection and Analysis
- 4) Maths and Statistics
- 5) Interpretation and Visualisation
- 6) Advanced Computing and Programming
- 7) Business Intelligence and Domain Expertise
- 8) Open Source Tools and Concepts



DATA SCIENCE SKILLSET



Data science, due to its interdisciplinary nature, requires an intersection of abilities: **hacking skills, math and statistics knowledge**, and **substantive expertise** in a field of science.



Hacking skills are necessary for working with massive amounts of electronic data that must be acquired, cleaned, and manipulated.



Math and statistics knowledge allows a data scientist to choose appropriate methods and tools in order to extract insight from data.



Substantive expertise in a scientific field is crucial for generating motivating questions and hypotheses and interpreting results.



Traditional research lies at the intersection of knowledge of math and statistics with substantive expertise in a scientific field.

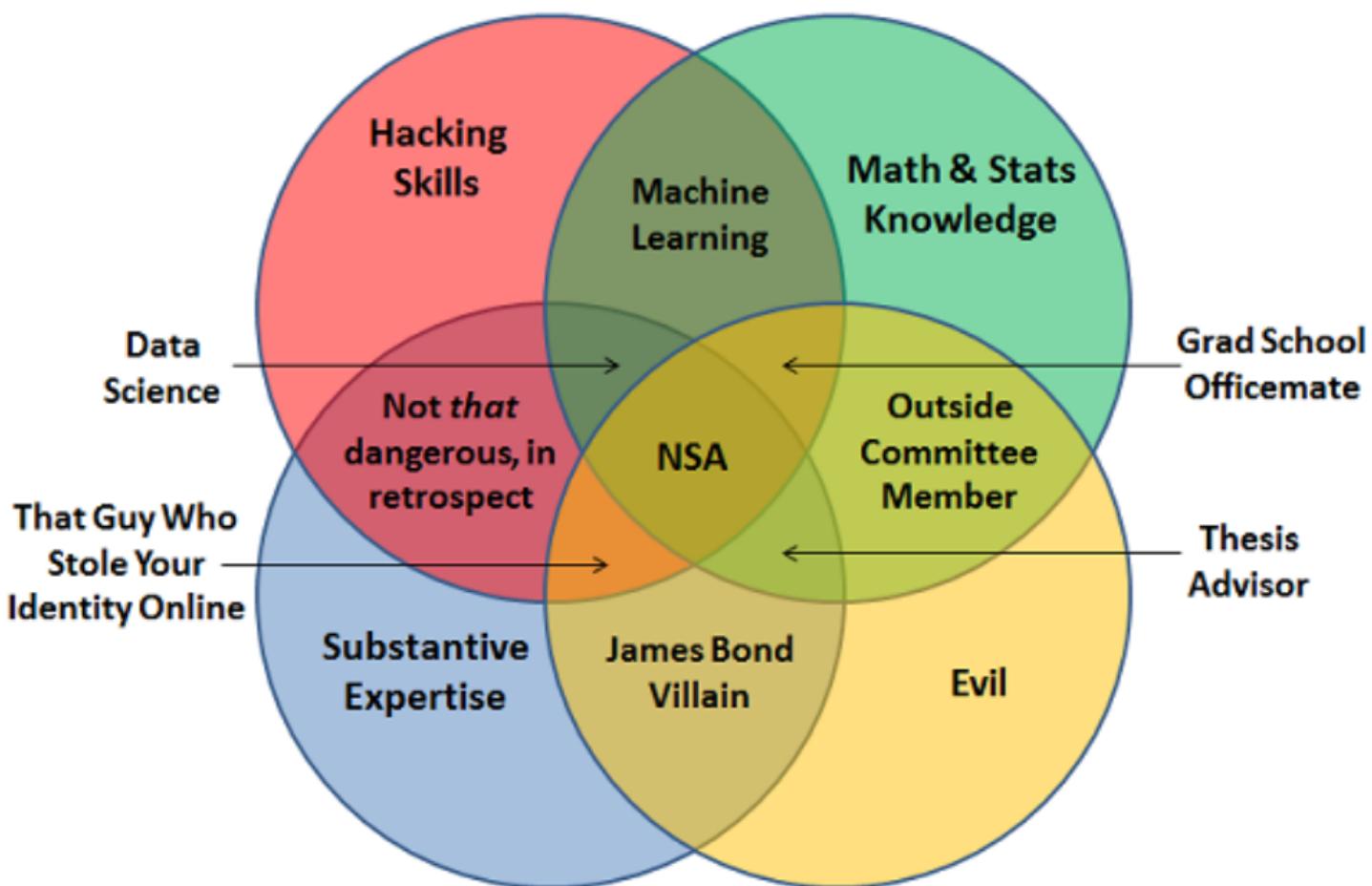


Machine learning stems from combining hacking skills with math and statistics knowledge, but does not require scientific motivation.



Danger zone! Hacking skills combined with substantive scientific expertise without rigorous methods can beget incorrect analyses.





Data Science

“part analyst, part artist”

-- Anjul Bhambhani (VP for Big Data, IBM)



Outcomes

Define open data science

Describe a number of key data science stories

Identify the characteristics of open data science projects

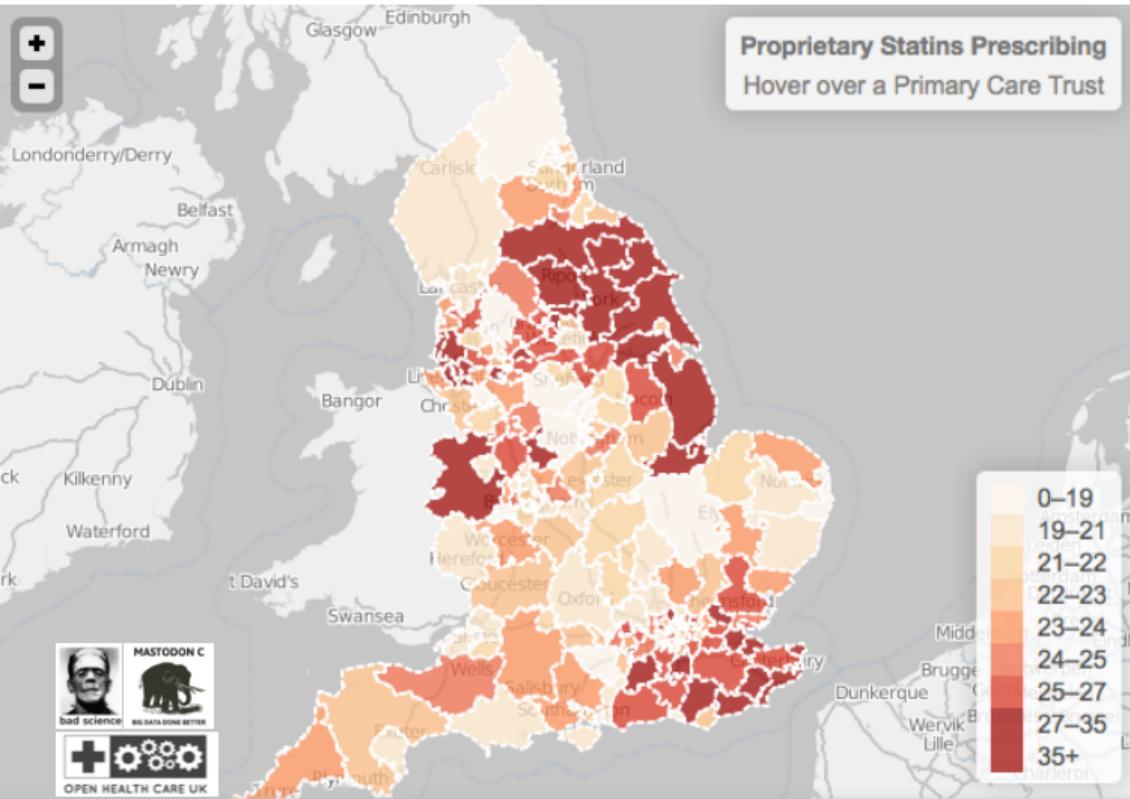


Stories that combine areas?

- 1) Big Data
- 2) Machine Learning and Prediction
- 3) Data Collection and Analysis
- 4) Maths and Statistics
- 5) Interpretation and Visualisation
- 6) Advanced Computing and Programming
- 7) Business Intelligence and Domain Expertise
- 8) Open Source Tools and Concepts



£200m potential saving identified in 6 weeks



Convened domain-experts

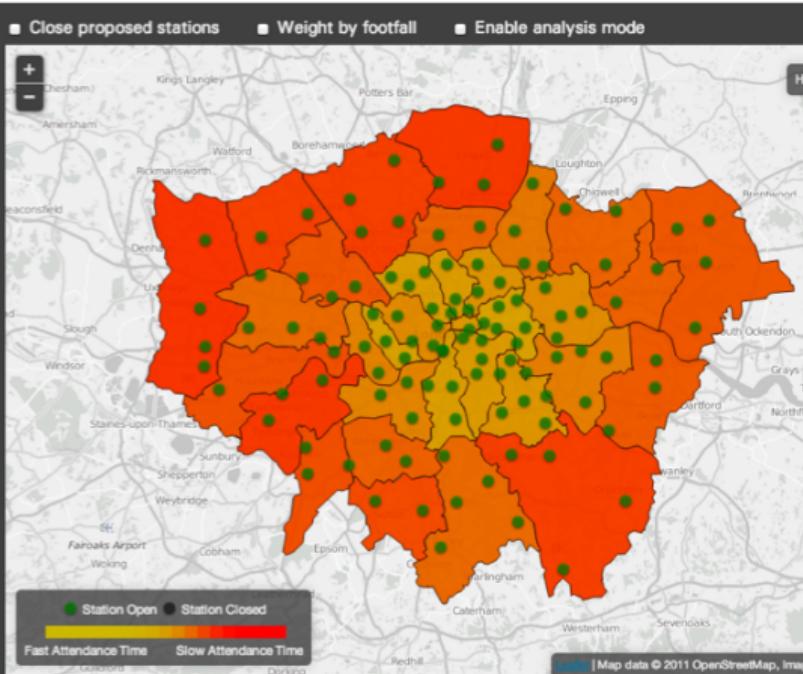
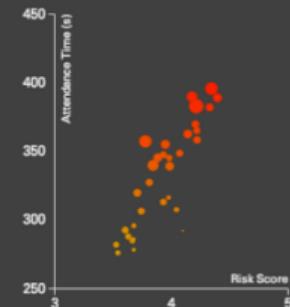
Analysed 35m records

National & international reach

Scalable to £1bn [Bn]

Tools that aid policy makers and are open to the public

London
Average attendance time:
5m34s



<http://theodi.org/stories>

Convened domain experts

Real-time big data processing

1 expert analysis tool

MARYLAND STATESTAT WAR ROOM | 14:00



Unlocking data on the web

Formats and Structures



What's the difference between a data structure and a data format?

Exercise

Exploring data formats

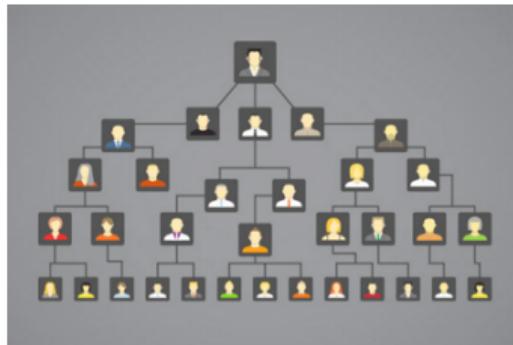


<http://training.theodi.org/UnlockingData/>

Structures



Tabular



Hierarchical



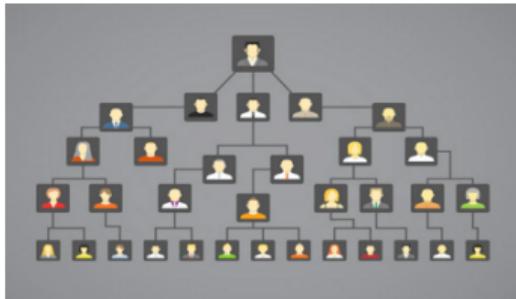
Network

Structures and formats



Tabular
csv, xls, json

Hierarchical
json, rdf*, xls**



Network
rdf*, json, html

* rdf is not a format, it is a schema that can be serialised into many formats.

** xls can be abused to create hierarchical databases using multiple worksheets

Choose a structure and format:

1. A list of events
2. A shopping catalog
3. Alternative names for companies

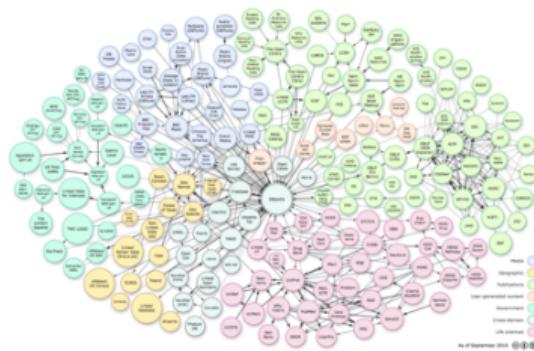


Approaches to publishing data

ON the web



IN the web

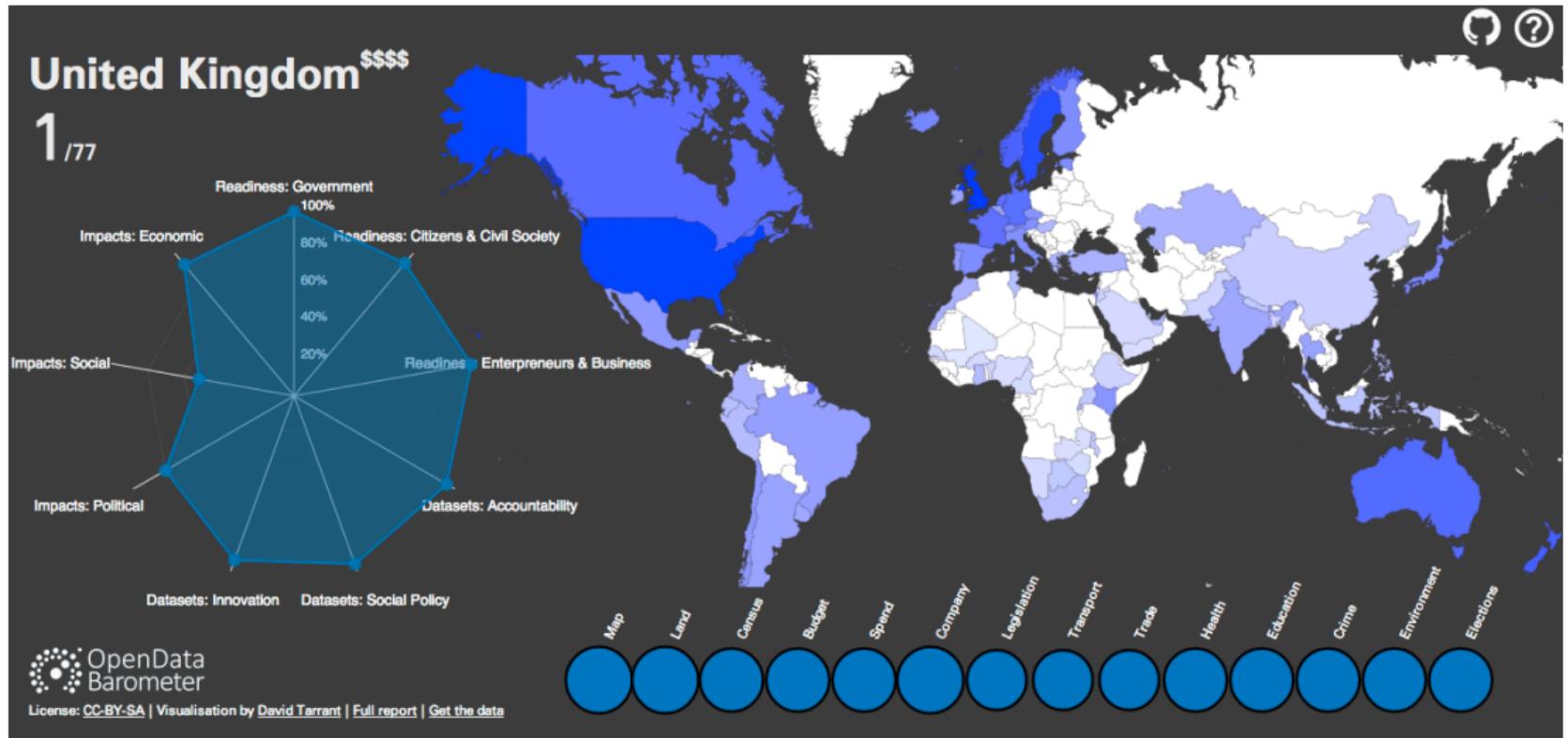


Finding data on the web (of documents)

- Government data
- Private sector data
- Google advanced
- Aggregators and portals
- Scraping



Government data



data.gov.XX

The screenshot shows the DATA.GOV.UK homepage with a search bar for datasets. A specific result for "Live traffic information from the Highways Agency" is highlighted, showing a map of the US-Mexico border and a snippet of the dataset description.

DATA.GOV.UK
Opening up Government

Home Data Apps Interact

Datasets Map Search Data Requests Publishers Public Roles & Salaries Spend Reports Site Analytics Reports

/ Datasets

Search for data...

19266 Results

Sort by

Show only...
Published datasets (15180)

Live traffic information from the Highways Agency
Highways Agency
Live traffic information data showing traffic information on the strategic road

OpenData Burkina Faso

Accueil Comprendre l'Open Data Thèmes Producteurs Jeux de données Applications Partenaires A propos

THEMES

- Agriculture
- Éducation
- Diplomatie
- Infrastructures
- Eau et Environnement
- TIC
- Santé
- Collectivités territoriales
- Sécurité Publique
- Tourisme Culture

Tous les thèmes

Open Data Burkina Faso statistiques

- 78 jeux de données
- 28 organisations
- 11 groupes

CC BY SA

The screenshot shows the DATOS.GOB.MX BETA homepage with a search bar and a map of Mexico and Central America. It displays a list of 127 datasets found, with a section titled "RATING SOCIAL PROGRAMS" and a news feed.

DATOS.GOB.MX BETA
OPEN TO PARTICIPATE

Data Stories Advances

datasets

Buscar conjuntos de datos...

127 datasets found

Sort by: Relevance

RATING SOCIAL PROGRAMS

Database containing the evaluation of the results of social programs from the federal government subject to the annual assessment.

data.gov.my

Portal Rasmi Open Data Kerajaan Malaysia
The Government of Malaysia's Open Data Official Portal

HOME ABOUT US CATALOGUE INFOGRAPHICS MOBILE APPS CONTACT

DATASETS INFO
117 Datasets
11 Ministries
10 Sectors

RESOURCES
Malaysia Directory
Malaysian Open Source Centre
Open Government Data

MOBILE APPS
myHealth | myJalani | KPONIK SPAD | SELAWAT | myMAHTAS
5 Android Mobile Apps
4 iOS Mobile Apps

QUICK LINKS

Malaysia Informativ Data Centre
Malaysian Population Quick Info
Tourism Malaysia - Facts & Figures

Latest Datasets Top Publishers Feedback

Applikasi Pemetaan Belia Malaysia
1 Info belia berdasarkan domain
by : Ministry of Youth and Sports

Last Updated 2014-06-26 11:30:19

Applikasi Pemetaan Belia Malaysia
2 Statistik Sikap Belia Malaysia mengikut
by : Ministry of Youth and Sports

Last Updated 2014-06-26 11:27:30

Applikasi Pemetaan Belia Malaysia
3 Statistik Pemetaan Belia Malaysia dalam bentuk aplikasi
by : Ministry of Youth and Sports

Last Updated 2014-06-26 11:24:52

Latest News and Events

Currently, there are no latest event.

Suppliers



X OPEN DATA

<http://manufacturingmap.nikeinc.com/#>

You agree not to change or delete any ownership notices from materials downloaded or printed from the Platform. You agree not to modify, copy, translate, broadcast, perform, display, distribute, frame, reproduce, republish, download, display, post, transmit or sell any Intellectual Property or Content appearing on the Platform



Google advanced

Google site:gov filetype:xls

Web Images Maps Shopping More Search tools

About 4,150,000 results (0.22 seconds)

[XL] [Code List or Concept \(Acronym\)](#) ●
www.acquisition.gov/short_codeListsTS.xls Share
File Format: Microsoft Excel - View as HTML
A, B, C, D, F, G, H, 1, Code List or Concept (Acronym), Definition, Authoritative or Maintenance Agency, Current Version, Update Frequency and/or ...

[XL] [Approps - Foreign Assistance.gov](#) ●
www.foreignassistance.gov/Full_ForeignAssistanceData.xls
File Format: Microsoft Excel
A, B, C, D, E, F, G, H, 1, Planning Data, 2, 3, 4, Fiscal Year, Fiscal Year Type, Account Name, Agency Name, Operating Unit, Category, Sector, Amount, ...

[XL] [TSB Monthly Cash Flow Projection](#) ●
www.dca.vt.gov/units/bsa/cashflow.xls

site: Get results only from certain sites or domains

link: Find pages that link to a certain page

related: Find sites similar to one you already know

filetype: Find certain file types only

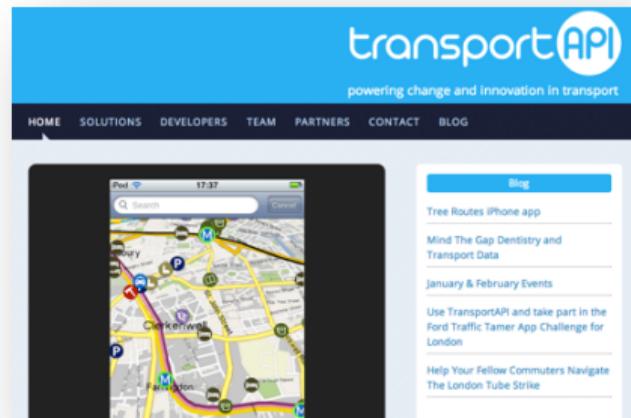


Aggregators and portals

Collect together data from across the web into one place.



enigma.io



transportAPI



Scraping

If you can't obtain usable data (csv, xls) then you may have to resort to scraping.



“excellent, so excited beyond description”
George Ofosu, Doctoral Student, UCLA

pdftables.com

A screenshot of the import.io web interface. At the top, there's a search bar with the placeholder "Enter a URL, for a list page" and a pink "Extract Data" button. Below the search bar, there are several examples of websites being scraped, each with a preview image and the name of the website below it: "Reseller Ratings", "Zoopla", "500px", "Growth Hackers", "Udemy", and "Stack Exchange".

magic.import.io

5-Stars



<http://5stardata.info/>

ON THE WEB



DL RE



DL



DL RE OF



IN THE WEB



http://data.../



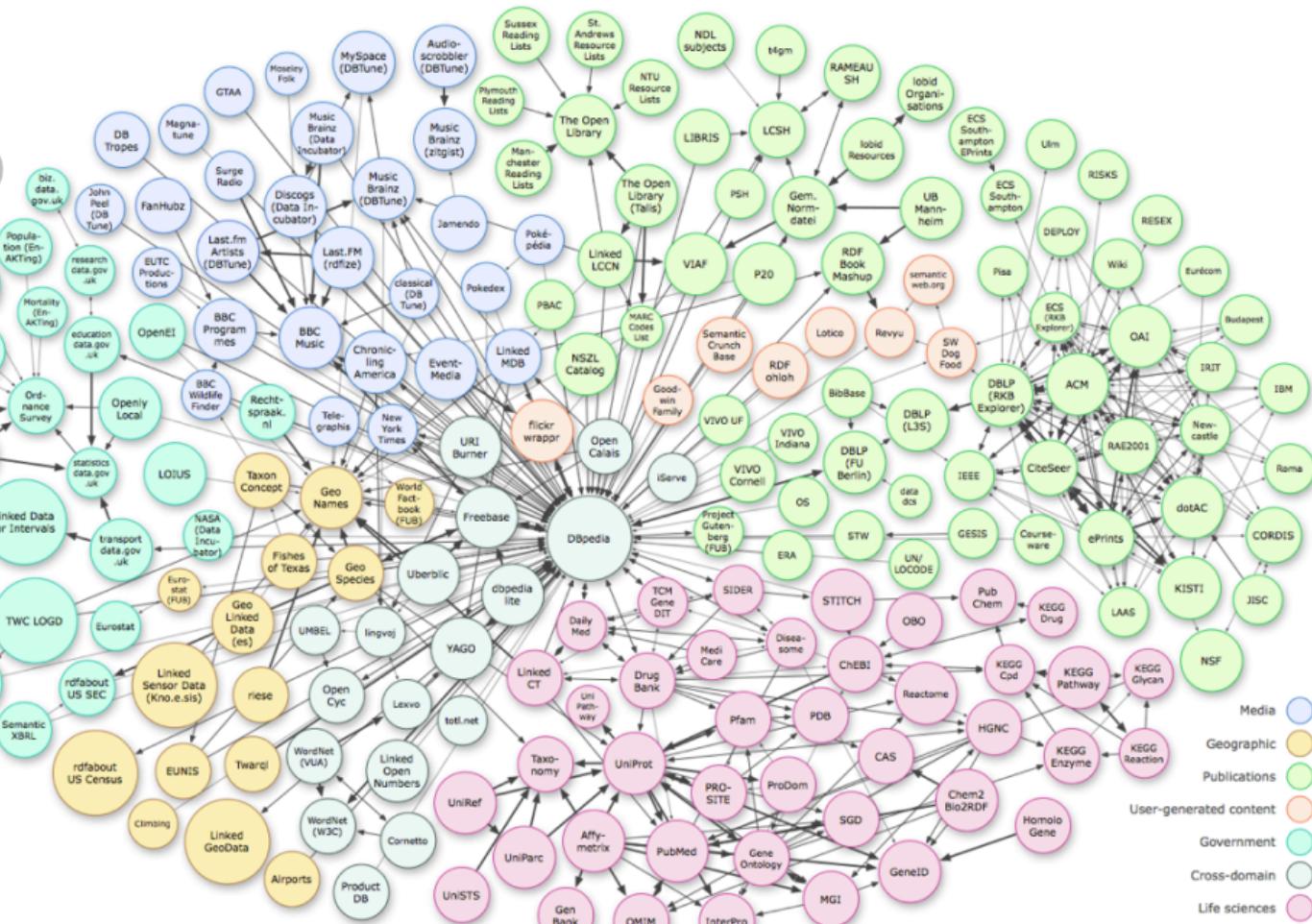
DL RE OF URI LD



OPEN DATA



Data IN the web



Finding data on the web (of data)

1. Add random extensions (.xml, .json, .csv etc)
2. Look for alternative links (rss feeds etc)
3. Look for embedded data **IN THE WEB**
4. Do some content negotiation
5. Spot the API
6. Scrape (or search google again)

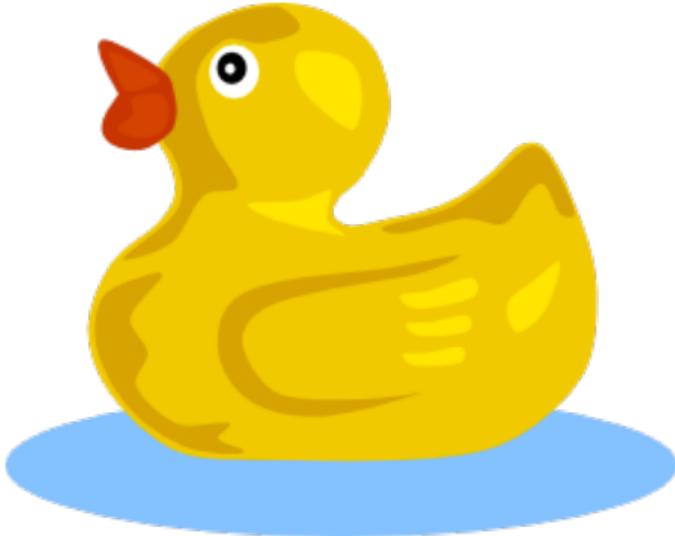


How the web should work,
but people forgot that Tim
put this in when he
invented it!

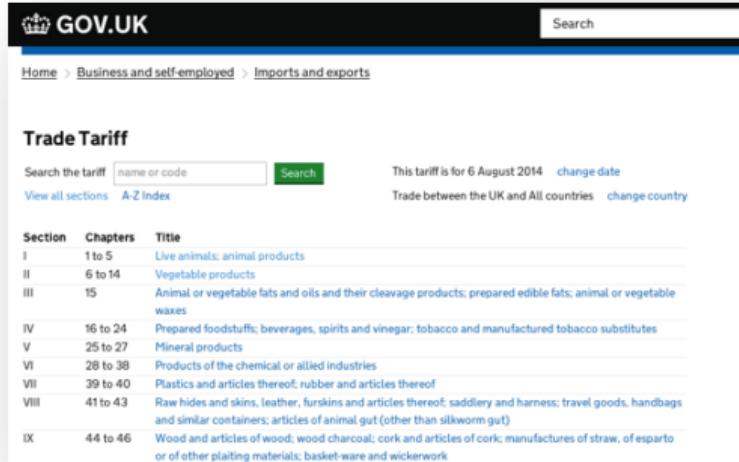
Duck typed data

If it looks like a duck
and quacks like a duck,
then it's probably a duck.

Basically, keep an eye out for tables,
lists and other stuff that looks like data.



1. Adding random extensions



The screenshot shows the UK Trade Tariff page on the GOV.UK website. At the top, there's a search bar and a navigation menu with links to Home, Business and self-employed, and Imports and exports. Below this, the title "Trade Tariff" is displayed. A search bar allows users to search by name or code, with a "Search" button. To the right, a note says "This tariff is for 6 August 2014" and provides options to change date and country. A sidebar lists "View all sections" and "A-Z Index". The main content area is organized into sections (I-IX) and chapters (1 to 46), each with a title and a brief description of the products included.

UK Trade Tariff



The screenshot shows the BBC Music and Programmes website for the TV show Doctor Who. The header features the BBC One logo and the show's title. Below the header, a navigation menu includes Home, Episodes, Clips, Galleries, Latest News, Characters, Monsters, Fun and Games, and More. The main content area features a thumbnail for an episode with two characters. To the right, a section titled "On iPlayer" says "This programme will be available shortly after broadcast". Another section titled "On TV" shows a thumbnail for "The Day of the Doctor" and provides broadcast details: SATURDAY 19:00 BBC THREE, All upcoming (0 NEW AND 1 REPEAT).

BBC Music and Programmes

Try using the following: .csv .json .xml .rss .rdf



2. Look for alternative links



Business Insight - NEWSASIA

NEWS TV WATCH LIVE

ASIA PACIFIC SINGAPORE WORLD BUSINESS SPORT ENTERTAINMENT TECHNOLOGY HEALTH LIFESTYLE VIDEOS WEATHER MORE ▾

CHANGINGLIVES LUMINARY AWARDS START-UP

APORE STORIES

Raise of up to 12% for Home Team officers, with sign-on bonuses of up to S\$30,000

National Day Award 2014

SP (2) Ng (SP)

MEDIACORP

Officers from the Home Team, both past and present were recognised at the Home Team National Day Observance Ceremony on Wednesday.

9 hours ago

Pay rise, special bonus for about 23,000 nurses

10 hours ago

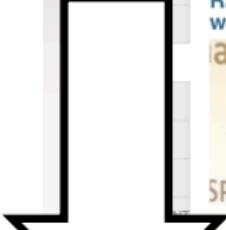
50,000 openings on Jobs Bank for Singaporeans, PRs

1 hour ago

NUS University Town identified as a high-risk dengue cluster

10 hours ago

Scroll down!



2. Look for alternative links



 CHANNEL NEWSASIA MediaCorp News Group. © 2014 MediaCorp Pte Ltd. All Rights Reserved. Terms and Conditions Privacy Policy About MediaCorp Pte Ltd	NEWS Asia Pacific Singapore World Business Sport Entertainment Technology Health Lifestyle Videos Photos Special Reports Archives	TV Live TV TV Videos TV Schedule SERVICES Weather ADVERTISE WITH US Online Advertising Mobile Advertising TV Advertising Contact Sales	ABOUT US About Channel NewsAsia Our Logo Our Coverage Our Tagline Presenters and Correspondents Contact Us GET OUR NEWS 
---	---	---	---



RSS



3. Look for embedded data

ODI Experiment

Hidden data extractor

open
data
institute

Hidden data extractor

Enter the URL of any webpage to see what JSON data is hidden within it.

Submit

Try these

[Products from Marks and Spencer UK](#)

[Products from ASOS](#)

CC BY SA

Finding data on the web (of data)

1. Add random extensions (.xml, .json, .csv etc)
2. Look for alternative links (rss feeds etc)
3. Look for embedded data **IN THE WEB**
4. Do some content negotiation
5. Spot the API
6. Scrape (or search google again)



How the web should work,
but people forgot that Tim
put this in when he
invented it!

Outcomes

Define open data science

Describe a number of key data science stories

Identify the characteristics of open data science projects



Session 1

Discovering Open Data Science

Session 2

Machine learning and classification

Session 3

Visualisation and communication

Session 2

Machine learning and classification



Outcomes

Explain the core aspects of Machine Learning

Apply classification to a set of data to create a decision tree

Perform a number of statistical experiments on data



Machine Learning

“construction and study of systems that can learn from data”

Can be seen as building blocks to make computers learn to
behave more intelligently

There are various *techniques* with various *implementations*.



Terminology

Features

The number of features or distinct traits that can be used to describe each item in a quantitative manner.

Samples

A sample is an item to process (e.g. classify). It can be a document, a picture, a sound, a video, a row in database or CSV file, or whatever you can describe with a fixed set of quantitative traits.

Feature vector

is an n-dimensional vector of numerical features that represent some object.

Feature extraction

Preparation of feature vector

Transforms the data in the high-dimensional space to a space of fewer dimensions.

Training/Evolution set

Set of data to discover potentially predictive relationships.



Learning (Training)



Features:

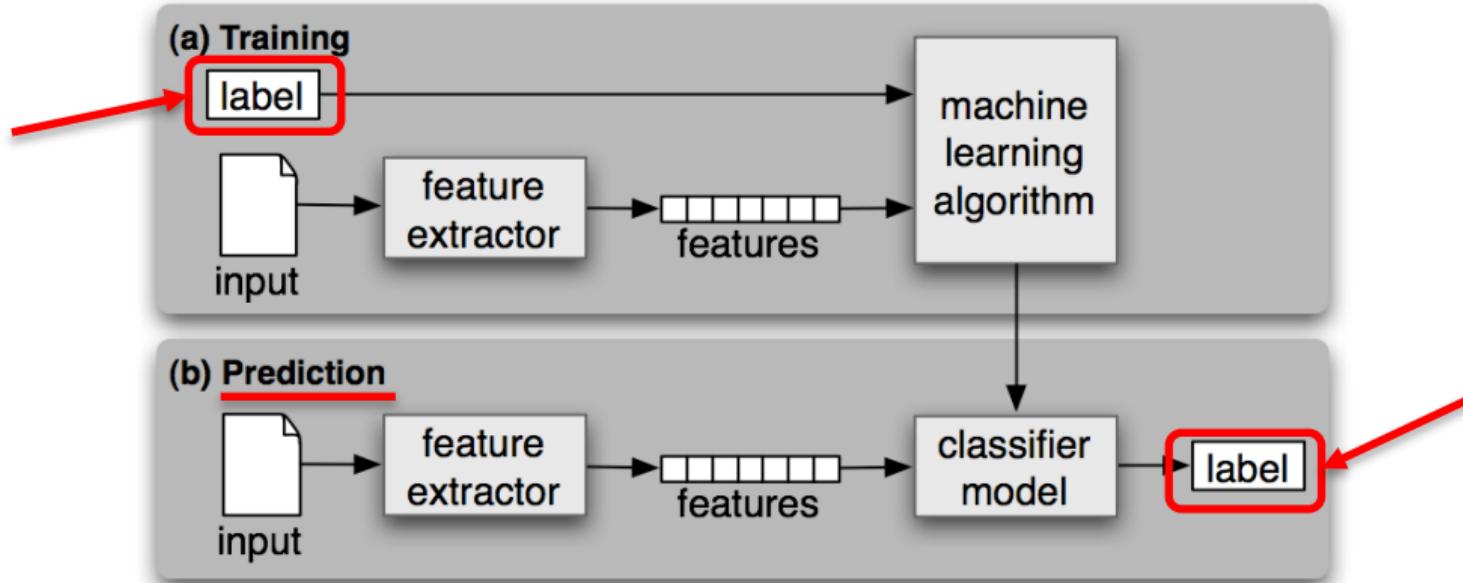
1. Color: **Radish/Red**
2. Type : **Fruit**
3. Shape
- etc...



Features:

1. **Yellow**
2. **Fruit**
3. Shape
- etc...

Workflow



Property locations?

Each table has a set of “Top Trump” training-set cards.

Build a decision tree to sort them into “New York” and “San Francisco”.

You cannot use the target to sort them.



Categories

Supervised Learning

Unsupervised Learning

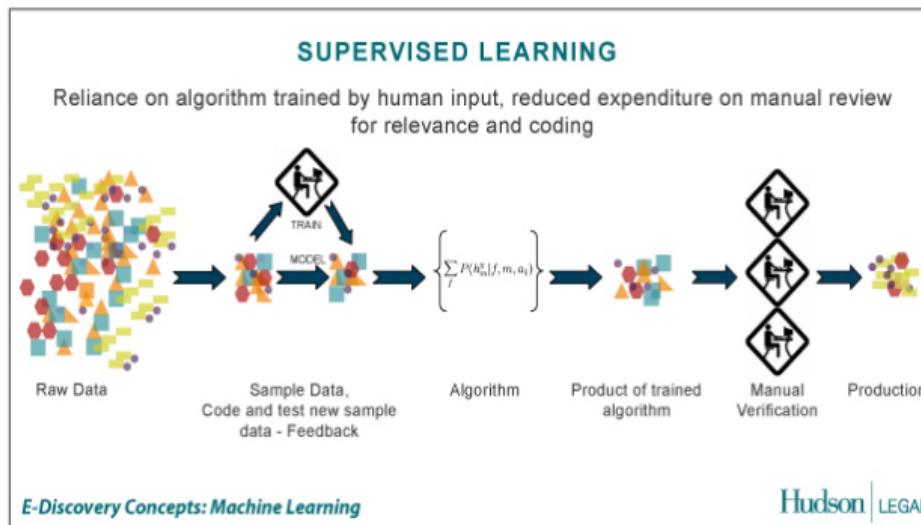
Semi-Supervised Learning

Reinforcement Learning



Supervised Learning

The correct labels of the training data are known



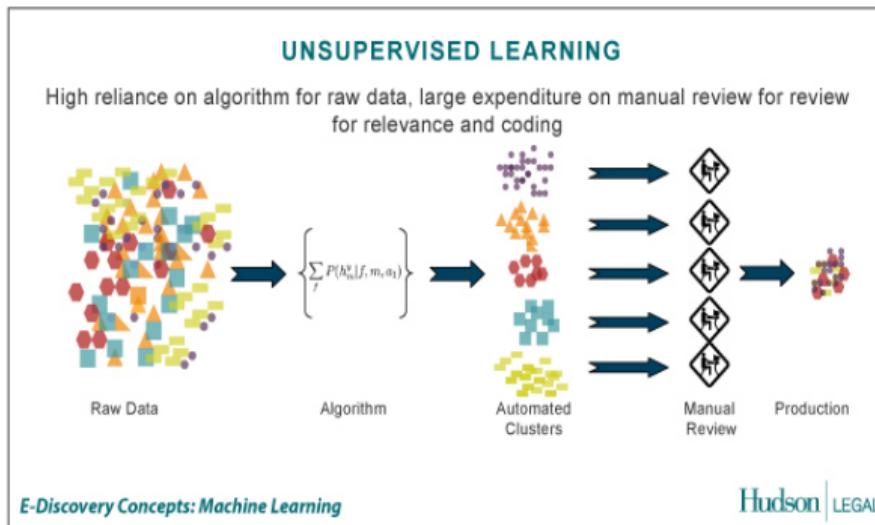
Credit: <http://us.hudson.com/legal/blog/postid/513/predictive-analytics-artificial-intelligence-science-fiction-e-discovery-truth>



Slide ideas thanks to @rahuldausa

Unsupervised Learning

The correct labels of the training data are not known.
Cluster discovery and pattern generation.



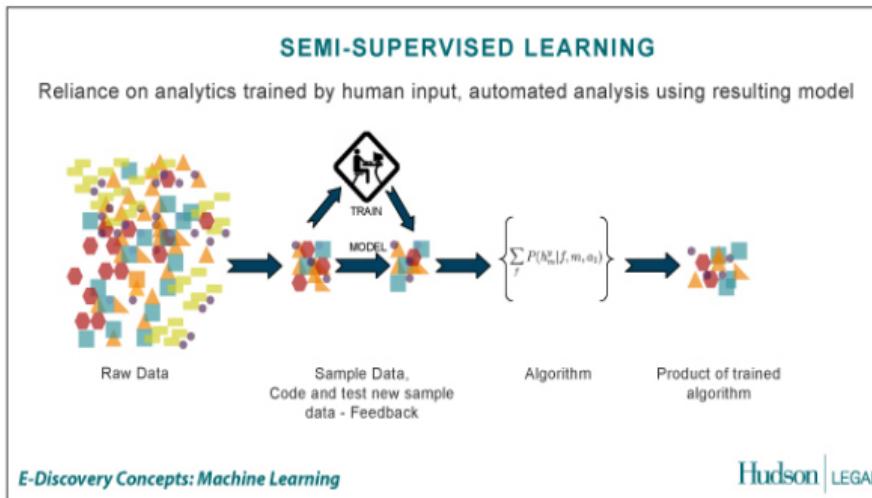
Credit: <http://us.hudson.com/legal/blog/postid/513/predictive-analytics-artificial-intelligence-science-fiction-e-discovery-truth>

Slide ideas thanks to @rahuldausa



Semi-Supervised Learning

A Mix of Supervised and Unsupervised learning.
Some labels are known, but majority are not.

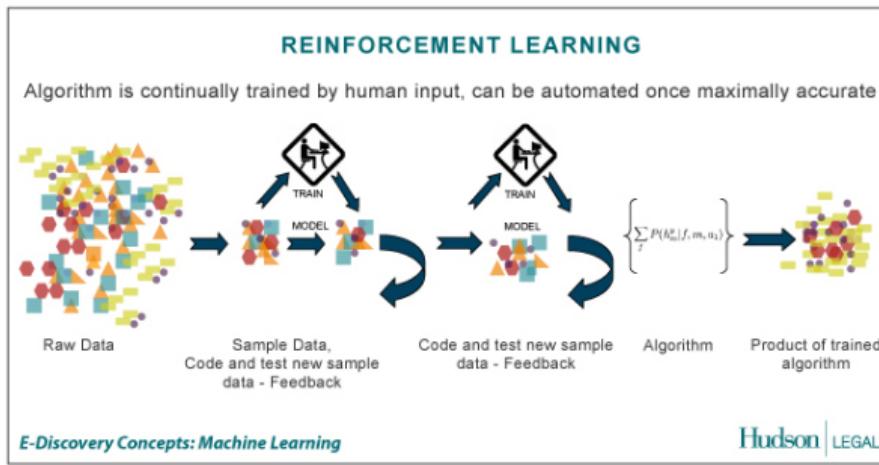


Credit: <http://us.hudson.com/legal/blog/postid/513/predictive-analytics-artificial-intelligence-e-science-fiction-e-discovery-truth>



Reinforcement Learning

Software agent learns based on feedback and/or reward.



Techniques

classification

predict class from observations

clustering

group observations into “meaningful” groups

regression (prediction):

predict value from observations



Question?

What instances of machine learning have you come across?

What types are they?



Use-Cases

- Spam Email Detection
- Machine Translation (Language Translation)
- Image Search (Similarity)
- Clustering (KMeans) : Amazon Recommendations
- Classification : Google News



Use-Cases (contd.)

- Text Summarization - Google News
- Rating a Review/Comment: Yelp
- Fraud detection : Credit card Providers
- Decision Making : e.g. Bank/Insurance sector
- Sentiment Analysis
- Speech Understanding – iPhone with Siri
- Face Detection – Facebook's Photo tagging



Not Spam

Delete all spam messages now (messages that have been in Spam more than 30 days will be automatically deleted)

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	HDFC Bank	LOAN upto Rs 25 lac - Disbursal in 2 days - open in fresh tab -- If you do not want to receive any more newsletters, please click here	9:40 pm
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	iEntry	Welcome iEntry Member - Ultimate Guide To Assessing	9:23 pm
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	New-Zealand-Jobs.067L	Come to New Zealand to find a great job and settle here (Search for all Jobs from differ... - Search for all Jobs from different kinds of industries Find a Job in Enchanting	8:18 pm
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	CarSizzler	Assured Free Luxurious Ride worth Rs.300 with Uber Cabs - Home Home Buy New Car Buy New Car Sell Car Sell Car Tech Tics Tip & Tale Facebook 41727 others	6:05 pm
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Supermarket Promotion	Enjoy Rs.1700 voucher valid at any supermarket! - If you are unable to view this mailer Click here HOW TO CONTACT US? BY EMAIL: support@savethedeals.in	4:51 pm
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Entireweb Newsletter	Hire an SEO the Right Way - 6 Tips You Must Remember for Life - Unsubscribe me View web version Become a fan on Facebook Follow us on Twitter September 5th, 21	1:24 pm
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Max Bupa	A policy that understands your family's medical need - open in fresh tab -- If you do not want to receive any more newsletters, please click here	11:08 am
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Scoop.it	Your Scoop.it Daily Summary - How to Maximize Your LinkedIn Publishing Exposure SME a... - Scoop.it Facebook Twitter G+ H	9:30 am
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	standard charterer Bank	Instant approval on your Credit Card	7:27 am
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	CAR TRADE	Sell your car at no cost at all - If you are having trouble viewing this email,view web version View this message in your mobile	Sep 4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Uday	VPS Web Hosting Services Provider - Dear Sir, I am Uday Sharma, Business development executive. We are providing quality VPS hosting for	Sep 4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Mark Regan, SPN	How to Find Your Most Valuable Keywords [Free Guide] - This is a SiteProNews/ExactSeek Webmaster Exclusive Mailing! To drop your subscription, use the link	Sep 4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	HDFC Bank	LOAN upto Rs 25 lac - Disbursal in 2 days - open in fresh tab You have received this mailer from Shop@Best on behalf of HDFC Bank because you	Sep 4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	CAR TRADE	Sell your car at no cost at all - open in fresh tab -- If you do not want to receive any further newsletters, please click here	Sep 4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ICICI Bank	Home Loan Interest Rate starting from 10.15%. Get Instant Approval! - open in fresh tab -- If you do not want to receive any more newsletters, please Click Here	Sep 4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	calculateyourwealth	It's good when your bank helps you manage your wealth and fulfill your ambitions - Calculate Now Dreams you wish to realize in your lifetime require enough wealth. C	Sep 4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Angel Broking	Get Low Brokerage - Free Demat & Trading Account - open in fresh tab -- If you do not want to receive any further newsletters, please click here	Sep 4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Bankbazaar	7 Minute Instant Online Approval for your PERSONAL LOAN - Now get instant online Personal Loan approval in 7 minutes by BankBazaar.com from leading Banks in	Sep 4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Jayde	Welcome To The Jayde Newsletter! Welcome To WebProNews! Welcome To The Jayde Newsletter! Before we begin, make sure to add	Sep 4
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ineedhits noreply	[ineedhits] Your ineedhits Account and Password - ACCOUNT CREATION Account ID : A1588368 Dear Rah, Welcome to ineedhits. Y	Sep 4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Rekha	Mobility Apps for Your Business - Mobility has been a part of our lives for last 20 years, we could hardly imagine life without it. Life in	Sep 4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	SlideShare Newsletter	Top Tips From the World Champions of PowerPoint - View online version Remember to display images Meet the PowerPoint World Champs Top Tips From the	Sep 4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Dilshad Pathan	Feeling Hesitate to Discuss personal Health Queries - My Life Care Follow Us on facebook twitter linkedin Google+ Feeling Hesitate to Discuss personal	Sep 4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Vaishu	TAKE YOUR PICK. Register in SimplyMarry - TAKE YOUR PICK. Register in SimplyMarry – Regards Vaishu	Sep 4

Not a Spam

Not a Spam



NER (Named Entity Recognition)

Stanford Named Entity Tagger

Classifier: english.muc.7class.distsim.crf.ser.gz ▾

Output Format: highlighted ▾

Preserve Spacing: yes ▾

Please enter your text here:

When Mike Brannigan was 18 months old, he was diagnosed with autism. At the time, his doctors said he would likely need a special school and a group home. His mom, Edie, admits she thought he'd "never be able to function in the world." Fast-forward several years. Brannigan is now 17, and is a senior at Northport High School, a public school in Long Island, New York. He's doing well academically, he has friends -- and he also happens to be one of the best young athletes in the country. Continue Reading...

Submit

Clear

When **Mike Brannigan** was 18 months old, he was diagnosed with autism. At the time, his doctors said he would likely need a special school and a group home. His mom, **Edie**, admits she thought he'd "never be able to function in the world." Fast-forward several years. Brannigan is now 17, and is a senior at **Northport High School**, a public school in **Long Island, New York**. He's doing well academically, he has friends -- and he also happens to be one of the best young athletes in the country. Continue Reading...

Potential tags:

LOCATION

TIME

PERSON

ORGANIZATION

MONEY

PERCENT

DATE



Similar/Duplicate Images

About 81 results (0.70 seconds)



Image size:
250 × 321

No other sizes of this image found.

Best guess for this image: [taj mahal](#)

Visually similar images



[Report images](#)

Remember

Features ?

(Feature Extraction)

Can be:

- Width
- Height
- Contrast
- Brightness
- Position
- Hue
- Colors

Check this:

LIRE (Lucene Image REtrieval) library -
<https://code.google.com/p/lire/>

Credit: <https://www.google.co.in/>



Recommendations

More Items to Consider

You looked at

JavaScript: The Good Parts Paperback by Douglas Crockford
\$29.99 \$19.79

> Find similar items

You might also consider

JavaScript: The Definitive Guide Paperback by David Flanagan
\$49.99 \$31.49

CSS: The Missing Manual Paperback by David McFarland
\$34.99 \$23.09

Learning jQuery 1.2 Paperback by Chaffer, Kar
\$39.99 \$35.9

Related to Items You've Viewed

You looked at

Forms that Work: Designing Web Forms... Paperback by Steve Krug
\$34.99 \$24.99

You might also consider

DON'T MAKE ME THINK A Common Sense Approach to Web Usability by Steve Krug

Letting Go of the Words: Writing Web... Paperback by Steve Krug

Designing Web Interfaces: Principles and Patterns Paperback by Steve Krug

Today's Recommendations For You

Here's a daily sample of items recommended for you. Click here to [see all recommendations](#)

Even Faster Web Sites: Performance Secrets from the World's Fastest Websites (Paperback) by Steve Souders
★★★★★ (7) \$23.10

Fix this recommendation

SIMPLY JAVASCRIPT (Paperback) by Kevin Yank
★★★★★ (19) \$26.37

Fix this recommendation

THE ART & SCIENCE OF JAVA (Paperback) by Brian Goetz
★★★★★ (1)

Fix this recommendation

[Any Category](#) [Algorithms](#) [Boxed Sets](#) [Business & Culture](#) [Java](#)
[Graphic Design](#) [Microsoft](#) [Networking](#) [Networks, Protocols & APIs](#) [New](#)
[SQL](#)



Frameworks/Tools



R
Weka
Carrot2

Gate
OpenNLP

LingPipe

Stanford NLP

Mallet – Topic Modelling

Gensim – Topic Modelling (Python)

Apache Mahout

MLib – Apache Spark

scikit-learn - Python

LIBSVM : Support Vector Machines
and many more...

Exploring R

Improve your classification tree using R



Exploring in R

```
attach(houses)
str(houses)
summary(houses)
plot(houses)
Hist(elevation)
plot(x=price_per_sqft,y=elevation,col=target+1)
boxplot(formula = beds ~ target, data=houses)
Library(rpart)
fit <- rpart(formula = target ~ elevation+ price_per_sqft, data=houses,
method="class")
summary(fit)
plot(fit, uniform=TRUE, main="Classification Tree for Houses")
text(fit, use.n=TRUE, all=TRUE, cex=.8)
```



Outcomes

Explain the core aspects of Machine Learning

Apply classification to a set of data to create a decision tree

Perform a number of statistical experiments on data



Session 1

Discovering Open Data Science

Session 2

Machine learning and classification

Session 3

Visualisation and communication

Session 1

Visualisation and communication

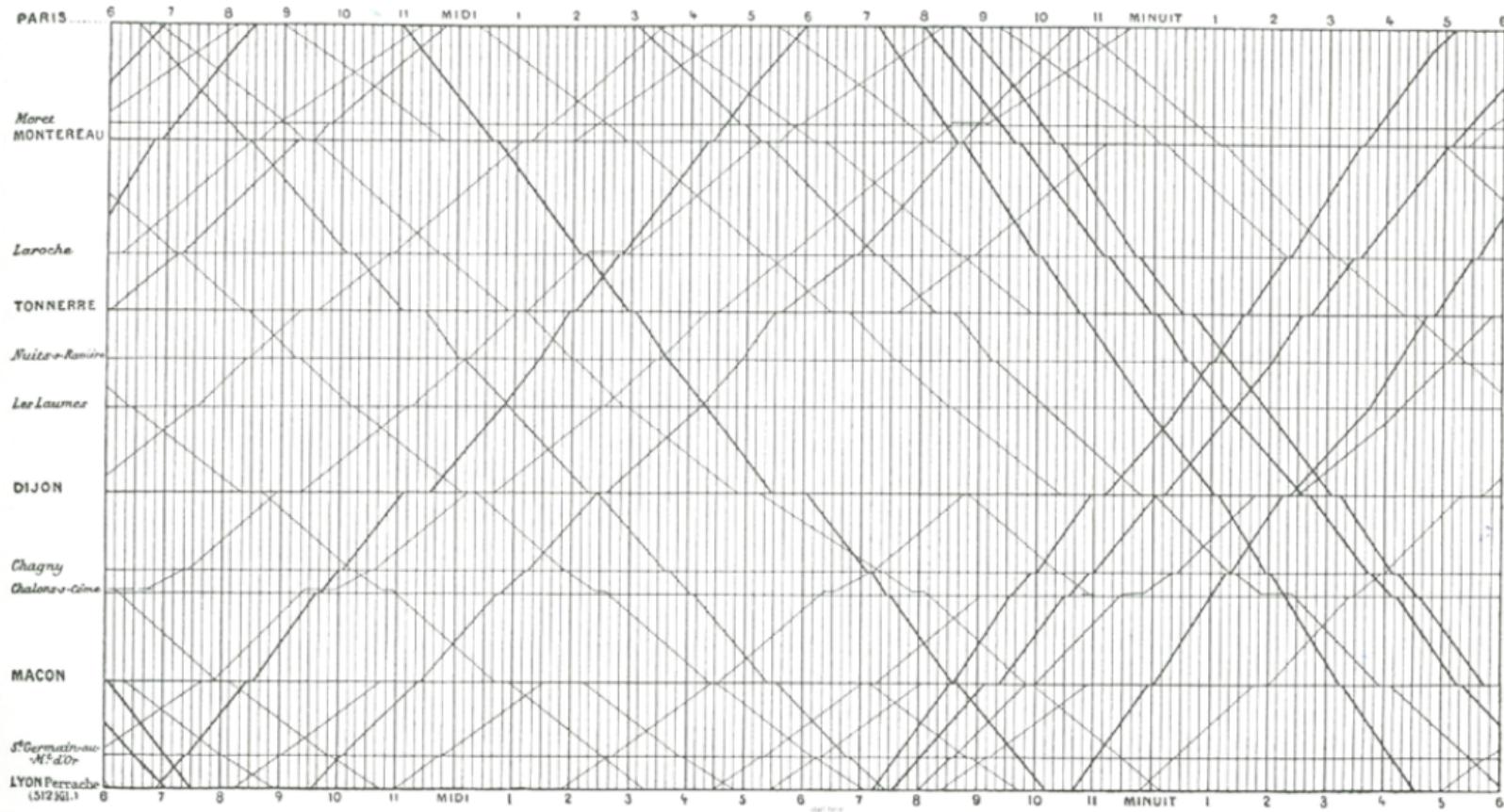


Outcomes

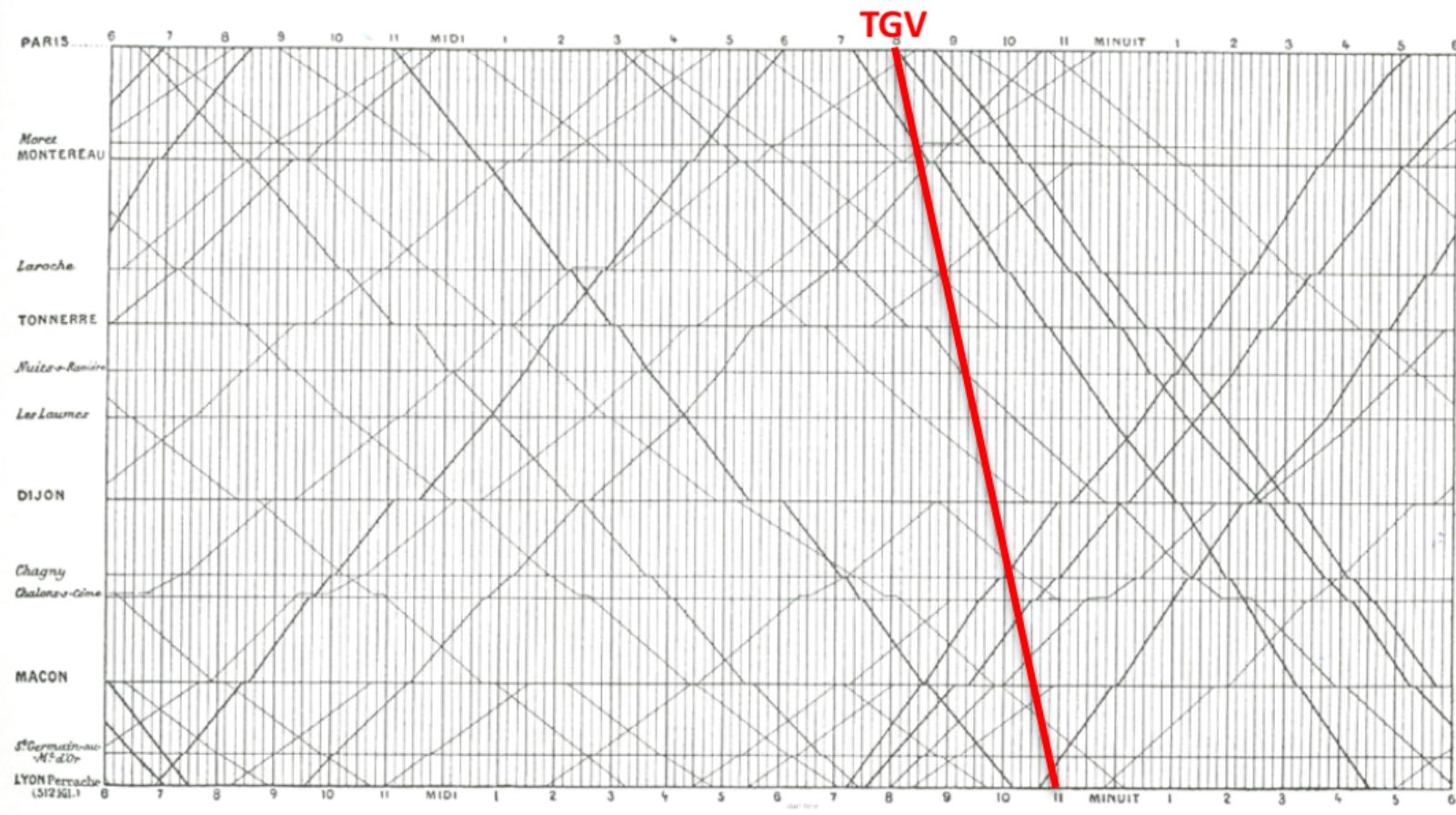
Differentiate between visualising for communication and understanding.
Explain the challenges of visualising complex data.
Create a visualisation to aid understanding of complex data.



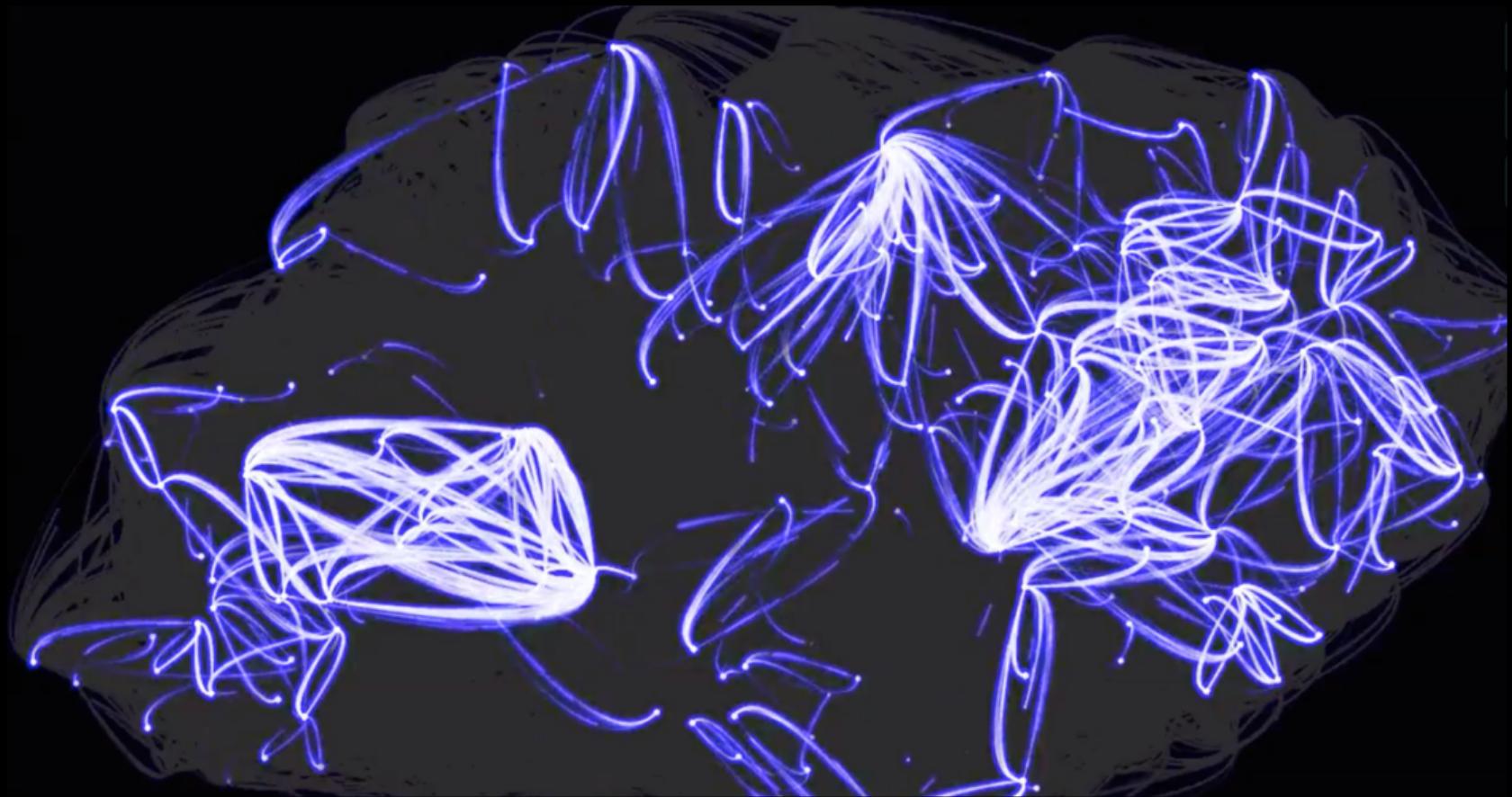




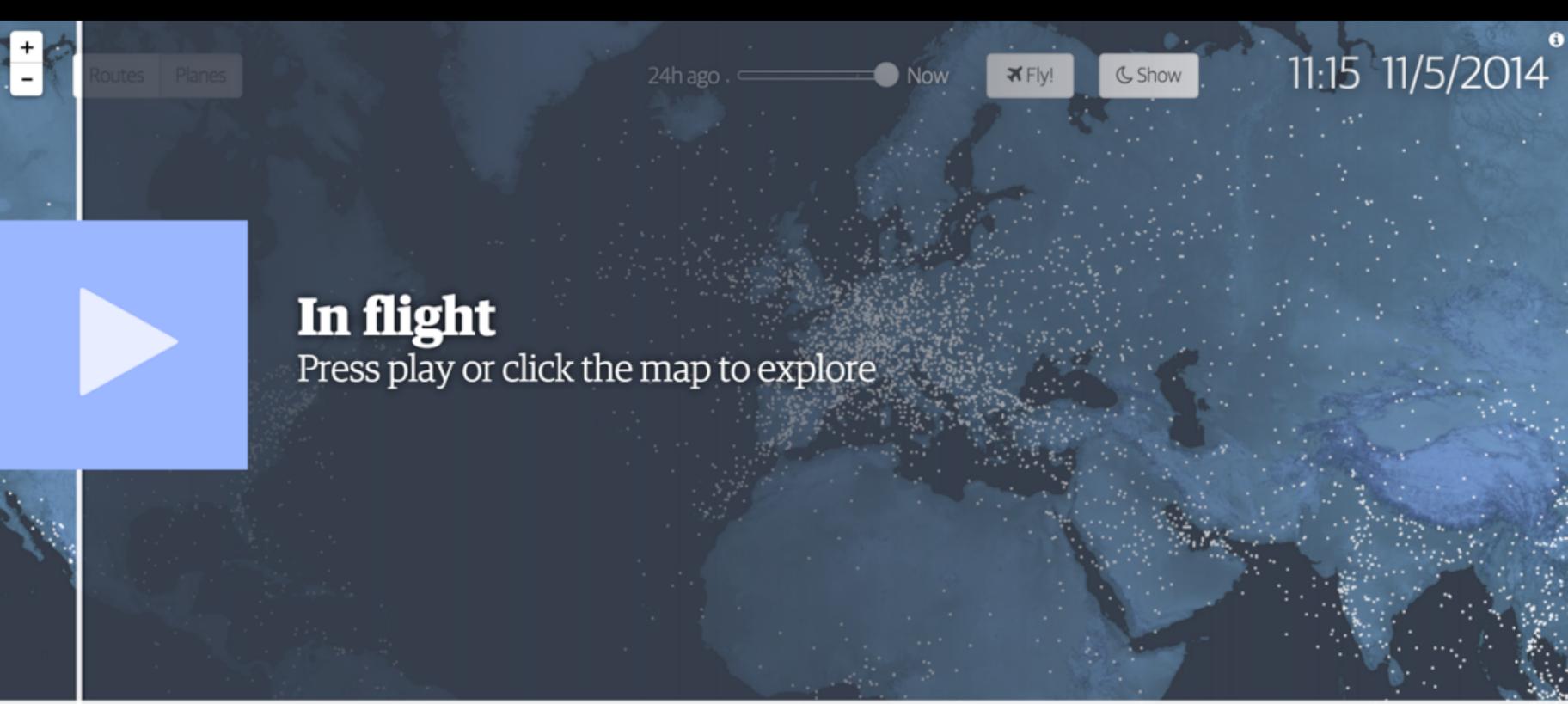
E. J. Marey, *La méthode graphique* (Paris, 1885), 20. The method is attributed to the French engineer, Ibry.



E. J. Marey, *La méthode graphique* (Paris, 1885), 20. The method is attributed to the French engineer, Iibry.



<https://vimeo.com/33712288>



g

1 Mapping the skies

2 Birth of an industry

3 A century of growth

4 Hitting the limits?



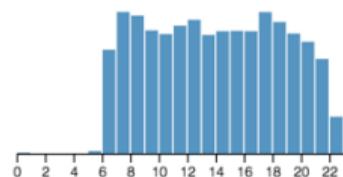
Share

Tweet

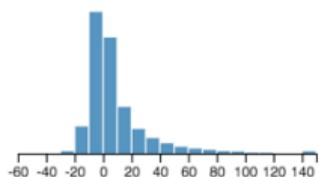


<http://www.theguardian.com/world/ng-interactive/2014/aviation-100-years>

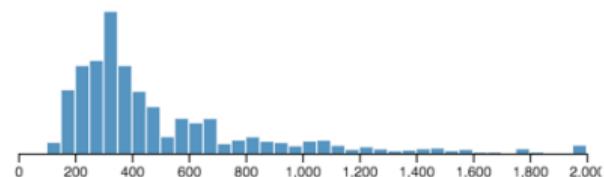
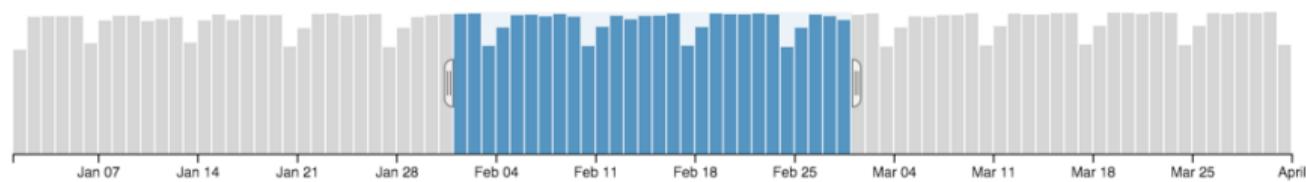
Time of Day



Arrival Delay (min.)



Distance (mi.)

Date [reset](#)

February 28, 2001

71,818 of 231,083 flights selected.

11:59 PM	LAS	LAX	236 mi.	+139 min.
11:58 PM	PHX	SAN	304 mi.	+83 min.
11:49 PM	SJC	PDX	569 mi.	+172 min.
11:42 PM	PHX	OAK	646 mi.	+97 min.
11:41 PM	PHX	LAX	370 mi.	+73 min.
11:40 PM	PHX	ONT	325 mi.	+92 min.
11:35 PM	PHX	ONT	325 mi.	+16 min.
11:25 PM	ONT	OAK	361 mi.	+75 min.



Exercise

Building a crossfilter for our houses



Gary Flake:

Is Pivot a turning point for web exploration?

TED2010 · 6:25 · Filmed Feb 2010

 24 subtitle languages 

 View interactive transcript





Thank-you