



Sharing your science: Three easy pieces

Denis Schluppeck

*ABDSA Limuru
Kenya, 2024*

1. **the big picture:** science communication
2. **the nitty-gritty:** some practical tips (formats, basic principles, version control)
3. **an example:** bringing a jupyter notebook to life



Engaging the media / public + presenting your work – a scientist's view

Denis Schluppeck

*ABDSA Limuru
Kenya, 2024*

Some thoughts about our wider “responsibilities” *as scientists*

- ... increase **public awareness** of (your) science / technology
- try to influence your institutions / local administrators (+ government) to support our work
- build communication competence skills (also of students and colleagues around you)

Challenges

— my view

Challenges

— my view

- get **news organisations** interested in covering research and science (sales? incentives?)

Challenges

— my view

- get **news organisations** interested in covering research and science (sales? incentives?)
- get **readers / consumers** to engage

Challenges

— my view

- get **news organisations** interested in covering research and science (sales? incentives?)
- get **readers / consumers** to engage
- get **scientists and researchers** to understand the importance of media / public engagement

Challenges

my view

- get **news organisations** interested in covering research and science (sales? incentives?)
- get **readers / consumers** to engage
- get **scientists and researchers** to understand the importance of media / public engagement



“The **BSA Media Fellowships** provide a unique opportunity for practising scientists, clinicians and engineers to spend 2 to 6 weeks working at the heart of a media outlet...”

–British Science Association website

Science for Africa Foundation

AAAS: <https://www.aaas.org/programs/mass-media-fellowship>
BSA website: <https://www.britishscienceassociation.org/>

4 weeks at the

FINANCIAL TIMES

Denis Schluppeck

Mentor: Clive Cookson

... my aims for this bit:

- How did I get into this & what did I do?
- What have I learnt?
- How can *you* get scientists and researchers around you to engage with “sci-comm”, print and broadcast journalists, online media, ...?



How did I get into this?

Fwd: Would you like to experience life as a science journalist? UoN BSA Media Fellowships - entries close midni...

McGraw Paul 

Inbox - Exchange 11 March 2016 at 16:05

Fwd: Would you like to experience life as a science journalist? UoN BSA Media Fellowships - entries close midnight 16 March 2016

To: LP-Academic

MP

Please see opportunity below. If you are interested then let me know.

Best wishes,

Paul

Dear Paul

This year The University of Nottingham is funding three places for academics to participate in the British Science Association (BSA) Media Fellowship scheme. It is a unique opportunity for practising scientists, clinicians, engineers and social sciences and arts academics to spend two to six weeks working at the heart of a media outlet such as the Guardian, the BBC or the Times. Fellows are mentored by professional journalists. They learn how the media operates and reports on science, how to communicate with the media and engage the wider public with science through the media. This year there are three funded places — for Arts, Social Sciences and Science.

This placement would involve support from their Head of School.

I wondered if it would be possible to flag this golden opportunity to all School of Psychology academics?

Kind regards
Lindsay

Lindsay Brooke
Media Relations Manager (Science)
Press Office
External Relations
The University of Nottingham

Fwd: Would you like to experience life as a science journalist? UoN BSA Media Fellowships - entries close midni...

McGraw Paul 

Inbox - Exchange 11 March 2016 at 16:05

Fwd: Would you like to experience life as a science journalist? UoN BSA Media Fellowships - entries close midnight 16 March 2016

To: LP-Academic

MP

Please see opportunity below. If you are interested then let me know.

Best wishes,

Paul

Dear Paul

This year The University of Nottingham is funding three places for academics to participate in the British Science Association (BSA) Media Fellowship scheme. It is a unique opportunity for practising scientists, clinicians, engineers and social sciences and arts academics to spend two to six weeks working at the heart of a media outlet such as the Guardian, the BBC or the Times. Fellows are mentored by professional journalists. They learn how the media operates and reports on science, how to communicate with the media and engage the wider public with science through the media. This year there are three funded places — for Arts, Social Sciences and Science.

This placement would involve support from their Head of School.

I wondered if it would be possible to flag this golden opportunity to all School of Psychology academics?

Kind regards
Lindsay

Lindsay Brooke
Media Relations Manager (Science)
Press Office
External Relations
The University of Nottingham



Lindsay
Brooke



Hosts (2016)

The Times

The Guardian

Financial Times

The I

The Daily Mirror

Daily Telegraph

Daily Mail

BBC Radio Science (+online)

Channel 4

BBC Breakfast

Open Democracy

Nature News

Next stop: *FT*



FINANCIAL TIMES

- daily readership: 2.2m
 - circulation >234k (in 2014)
 - daily, emphasis on business and economic news
 - founded 1888, in 2015 bought by Nikkei, Inc. (Tokyo)



source: via wikipedia

my mentor



Clive Cookson

started as journalist
straight out of Oxford (1974)

Times Higher ES (Science)
The Times (Technology)
BBC Radio Science

FT (1987)
Science Editor / FT (1991)

my mentor



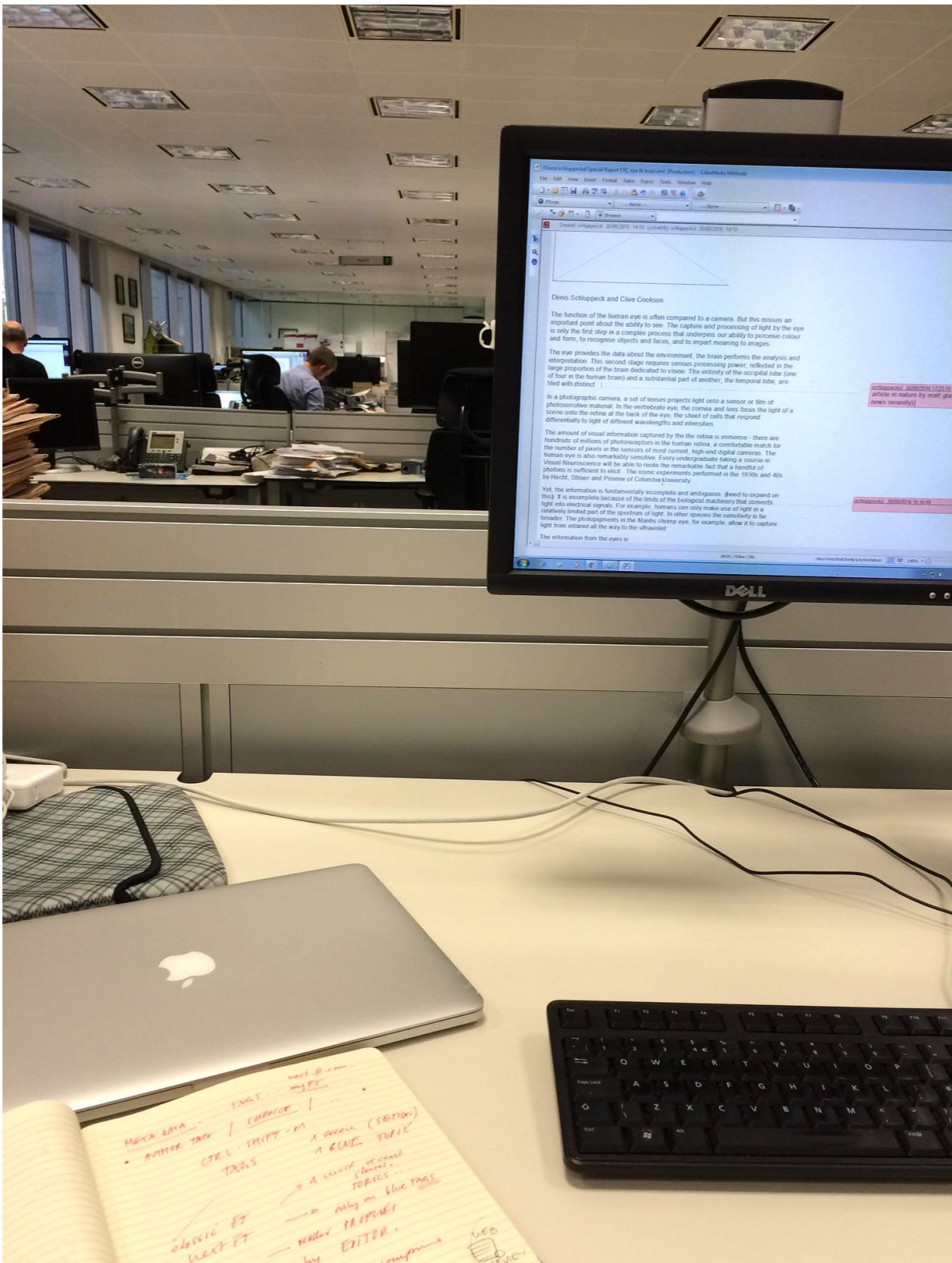
Clive Cookson

started as journalist
straight out of Oxford (1974)

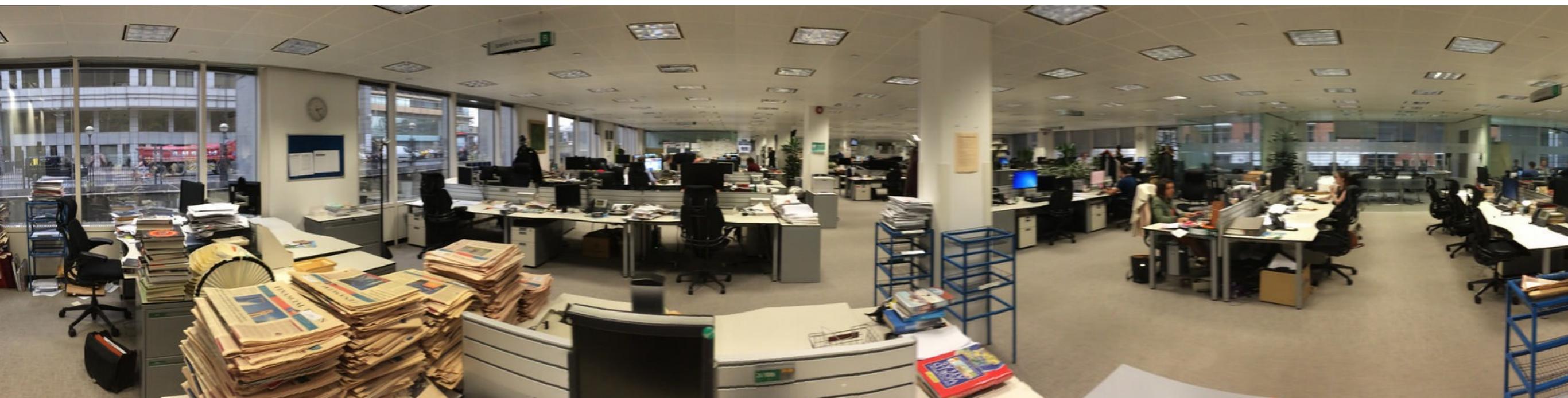
Times Higher ES (Science)
The Times (Technology)
BBC Radio Science

FT (1987)
Science Editor / FT (1991)

at the FT (Aug-Sept)



work !



a typical day

- 730am: wake up and start checking **EurekAlert** website
- 9am. arrive at desk (via Borough Market coffee shop)
- Discuss possible stories with Clive
- 10am Hop on tube to get to **press briefing**
Today: the Royal Astronomical Society; ESA Gaia mission
- 1245h. Back at FT. Check in with Clive. Start story (lunch *al desko*)
- 1500-1530h, **a coherent 450w piece** is ready. Grab an espresso.
- 1600h. **File story with the newsdesk.** Within 20-30 min, the editor assigned to the story comes back with any questions and an edited version.
- 1700h. The **picture desk gets in touch.** What photo to run with the story.

ated for 10 years. Don't believe

migration. Switzerland, by contrast, has

Over a decade-long period, Switzer-

santander UK and a former Eu

to reach the European single r
tinued access is se

A memo on Jap
posted on the mi
website, calls on
negotiate a post
guards almost
rights in the sing

The memo s
lured some Jap
ain on the basi
gateway to Eu
has a moral o
promises.

"We strong
consider this
in a respons
any harmfu
nesses," sai
Japan's posi
negotiators.

"Japanes
European
may decid
office fund
if EU laws
UK after
says.

It adds t
tions mig
ations fro
ments in
their rig
obtained i
market.

Almost i
ment inter
to the UK
main des
ment sto
of last ye

But Ja
difficult
closely
tionshi
voters
reform

Japa
access
the U
tariff
cedu

The
nise
betw

M
"cor
nati
min
me

M
G2
Ha
th
th
m
st
co

paper 'sting'

z to 'stand de' from ne affairs mittee

MP Keith Vaz is expected
airman of the Commons
select committee within
newspaper report about his

nister, whose role on the
given him a high profile
immigration and drugs
is speaking to his solicitor
that he was targeted in

ports in the Sunday
I paid for the services
Vaz said: "It is deeply
national newspaper
individuals to have

s currently carrying
itain's prostitution
t issued to the Mail
id he would inform
row of his inten
It was not clear if
y or permanent.
er solicitor and
was first elected
ast in 1987. He
in Tony Blair's
r Valerie is MP

Mr Vaz's positi
m and for the
Mr Vaz — who
tional executi
favour of Jer
getting on to
dership elec
cy party has
ontest.

rted that Mr
oppers, the
e two male
ed attempts

mer culture
r Mr Vaz to
n the areas
esponsible,
a sensible

Heads up Digital focus on crew of Mary Rose

Scientists have produced highly
detailed 3D reconstructions of skulls
found on the Mary Rose as they try to
unravel the mystery of the 460 crew
who died when Henry VIII's flagship
sank in 1545.

The crew's identities are largely
unknown, but researchers from
Swansea and Oxford universities hope
to glean hints about their age, health
and even occupation from the
interactive, photorealistic models of
each skull, and by sharing the results
online.

Capturing museum collections
digitally is now commonplace, but a
particular focus of this project is to
establish whether the quality of the 3D
reconstruction is high enough for
serious study of skeletal remains.

"This technology, and the appetite of
museums and researchers to open their

collections to larger global
communities, can have huge
implications for [the] speed that
science is done," said Dr Richard
Johnston of Swansea University.

For archaeology, digital imagery on a
much larger scale has already brought
a revolution. Google Earth has been
embraced by scientists as a useful tool
for archaeological digs and detailed
imagery has been used to reconstruct a
replica of an arch from Palmyra in
Trafalgar Square.

In other areas of science, from
astrophysics to zoology, the push
towards sharing data online has made
it easier for scientists to collaborate
across the globe. Successful projects
such as *Galaxy Zoo* have also allowed
researchers to involve the public in
their research.

The Mary Rose, which sank off

A 3D image of a skull believed to be that of a
carpenter on the Mary Rose, left, a facial
reconstruction of the carpenter, above, and a
painting of Henry VIII's flagship, below



A 3D image of a skull believed to be that of a carpenter on the Mary Rose, left, a facial reconstruction of the carpenter, above, and a painting of Henry VIII's flagship, below

Portsmouth, continues to capture
the imagination. The large number
of finds inside the hull have
provided an invaluable insight into
the Tudor period. Many thousands
of artefacts and the carefully
conserved hull of the ship are now
on display in Portsmouth's Mary
Rose Museum.

One of the skulls featured in the
project is thought to be that of a
carpenter. It was found immediately
below the Master Carpenter's cabin
near a number of woodworking
tools. Ship's carpenters were
particularly important during
battles and were stationed below
decks for repairing damage
below the waterline.

The skull reconstruction will be
available at www.virtualtudors.org.
Denis Schlupeck

myfirststory

≡ SEARCH

FINANCIAL TIMES

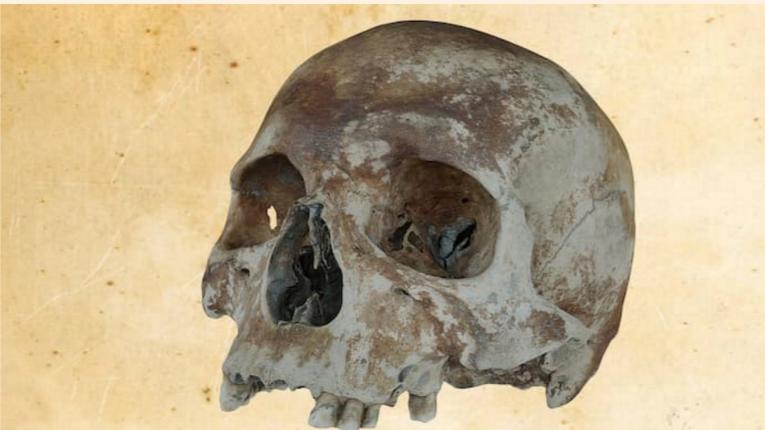
myFT

HOME WORLD US COMPANIES MARKETS OPINION WORK & CAREERS LIFE & ARTS Portfolio My Account

Medical science ✓ Added

Digital research aims to unveil Tudor mystery

Researchers assess skulls of Mary Rose crew from sinking of 1545



A 3D image of the skull of a carpenter found on the Mary Rose

September 5, 2016 by: Denis Schlupeck

Scientists have produced highly detailed 3D reconstructions of skulls found on the Mary Rose as they try to unravel the mystery of the 460 crew who died when Henry VIII's flagship sank in 1545.

Twitter Facebook LinkedIn Email 3 Save

Read latest:
If you want to learn, sleep on it
1 SECONDS AGO

4 weeks at the FT

10 articles in

≡ myFT

FINANCIAL TIMES

Science & Environment ✓ Added

Scientists question genetic tests for sporting ability

Trying to spot athletic talent in the lab raises ethical questions



Would genetic tests have spoilt Mo Farah's athletic ability? © PA

September 10, 2016 by: Denis Schlupeck

At commercial laboratories in the UK and abroad, children can now have genetic tests for sporting ability. Some of the tests even try to predict suitability for a particular sport. But should these tests be used to help spot athletic talent in children? "Don't waste your money," is the verdict of Mike McNamee, Professor of Applied Ethics at Swansea

Twitter Facebook LinkedIn Email 4 Save

FINANCIAL TIMES

Science + Add to myFT

Is prostate cancer as good as surgery or radiotherapy

The chance of survival after 10 years



Is PSA testing for prostate specific antigen © Jarun01/Dreamstime

September 10, 2016 by: Denis Schlupeck

A clinical trial of prostate cancer treatment has found that active monitoring offers the same chance of survival after 10 years as surgery or radiotherapy

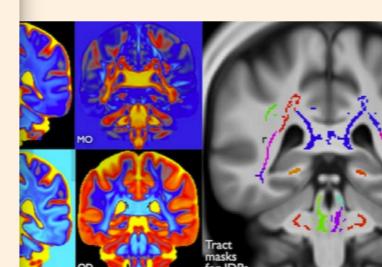
Twitter Facebook LinkedIn Email 8 Save

FINANCIAL TIMES

Science + Add to myFT

Brain scanning study provides 'window into the brain'

UK Biobank imaging project finds links between anatomy and cognitive function



The UK Biobank imaging project © UKBiobank

September 10, 2016 by: Denis Schlupeck

From the world's biggest body scanning project have revealed significant links between brain anatomy and cognitive function, allowing researchers to look for links between the two.

Twitter Facebook LinkedIn Email 2 Save

FINANCIAL TIMES

Science + Add to myFT

Learn to spot deadly food poisoning bacteria

Which cattle strains could spread to humans



September 10, 2016 by: Denis Schlupeck

Bacteria live in cattle and are harmless to them but some strains can spread to cause deadly disease. Until now, telling which ones are likely to be able to do this has been very difficult.

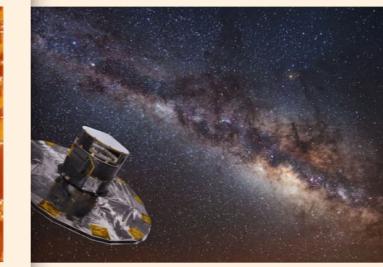
Twitter Facebook LinkedIn Email 0 Save

FINANCIAL TIMES

Science Agency + Add to myFT

Map the Milky Way

European Space Agency publish first data from Gaia space telescope



September 10, 2016 by: Denis Schlupeck

The telescope was launched 1,000 days ago with a mission to produce the most accurate map ever of the Milky Way. On Wednesday, the European Space Agency published first data from Gaia on star positions and movements in our galaxy. This will use the information to investigate how the Milky Way formed and evolved

Twitter Facebook LinkedIn Email 2 Save



20160905–maryRose/
20160906–bsf–oesophagealCancer/
20160906–supercomputingDrugs/
20160907–cloudyWithAChanceOfPain/
20160908–sleepLab/
20160909–geneticTestingInSports/
20160914–esa–mission/
20160914–prostateCancerTrial/
20160915–eyeVisionBrain/
20160919–biobank–brainImaging/
20160920–machineLearningEColi/
20160921–evolutionOfScience/
20160922–cornWorms/
20160923–hangoverFreeAlcohol/



What have I learnt?

For print journalists, press release is king!

The screenshot shows the homepage of eurekalert.org. At the top, there's a red header with the EurekAlert! logo and "The Global Source for Science News". Below it is a grey navigation bar with links for HOME, NEWS, MULTIMEDIA, MEETINGS, PORTALS, ABOUT, LOGIN, and REGISTER. A search bar with "SEARCH ARCHIVE" and "ADVANCED SEARCH" options is also present. The main content area is titled "TRENDING SCIENCE NEWS" and features a grid of six news items, each with a thumbnail image, a title, and a source. The news items are:

- Monitoring birds by drone** (American Ornithological Society Publications Office)
- Ancient jars found in Judea reveal earth's magnetic field is fluctuating, not diminishing** (American Friends of Tel Aviv University)
- Canadian glaciers now major contributor to sea level change, UCI study shows** (University of California - Irvine)
- NASA's OSIRIS-REx takes its first image of Jupiter** (NASA/Goddard Space Flight Center)
- Two from UW-Madison contribute to human gene editing report** (University of Wisconsin-Madison)
- Your brain's got rhythm** (Salk Institute)
- Researchers pinpoint watery past on Mars** (Trinity College Dublin)
- Fossil discovery rewrites understanding of reproductive evolution** (University of Queensland)
- To please your friends, tell them what they already know** (Association for Psychological Science)
- Ventura fault could cause stronger shaking, new research finds** (University of California - Riverside)
- Weight loss actually possible after menopause** (The North American Menopause Society (NAMS))

At the bottom left, the URL <https://www.eurekalert.org/> is visible.

snappy title

a summary that catches the eye

embargo!

nice imagery / photos
(not graphs, tables, ...)

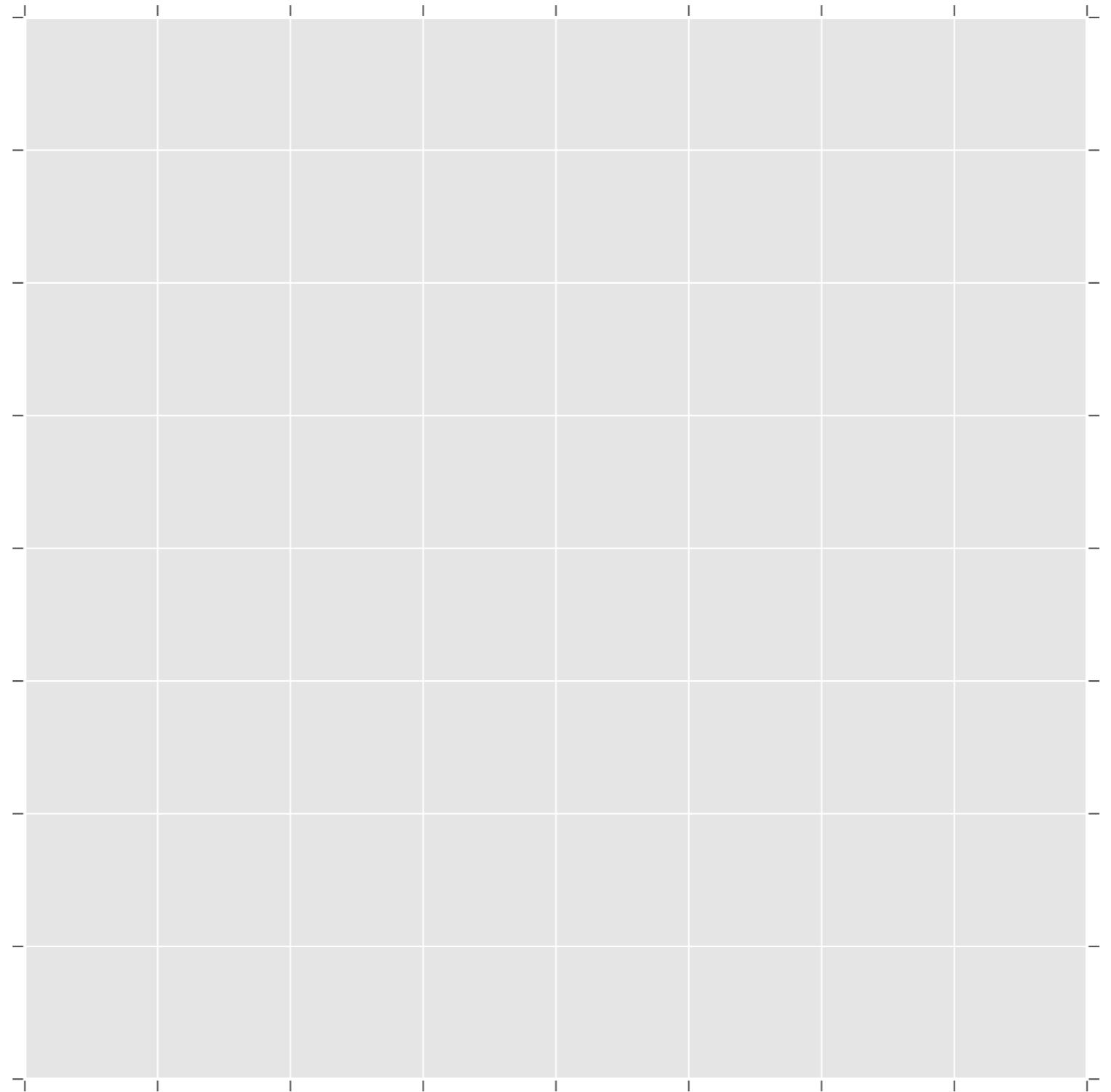
Key sites

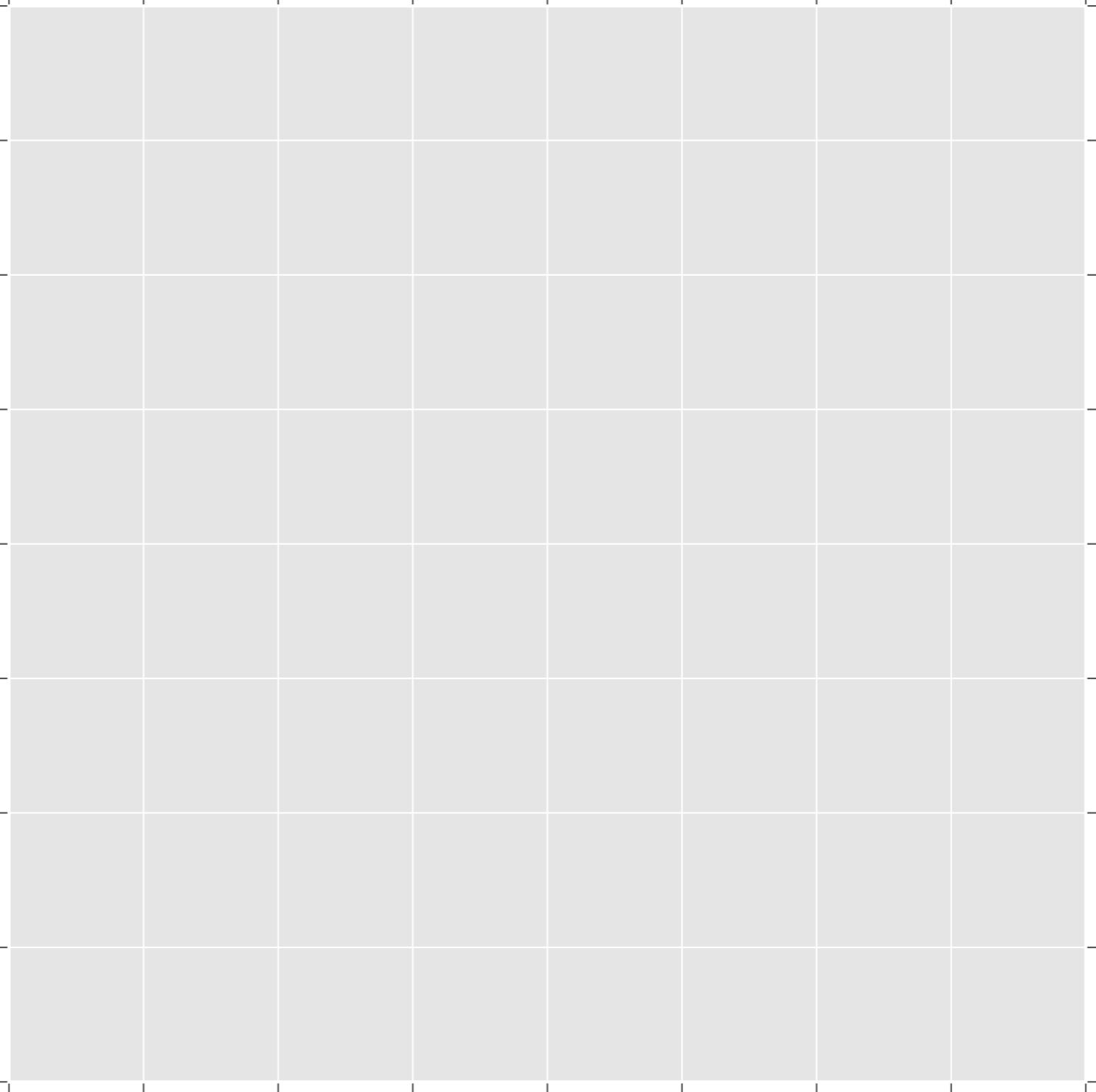
<https://www.eurekalert.org/>

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

*directly or via
journals*

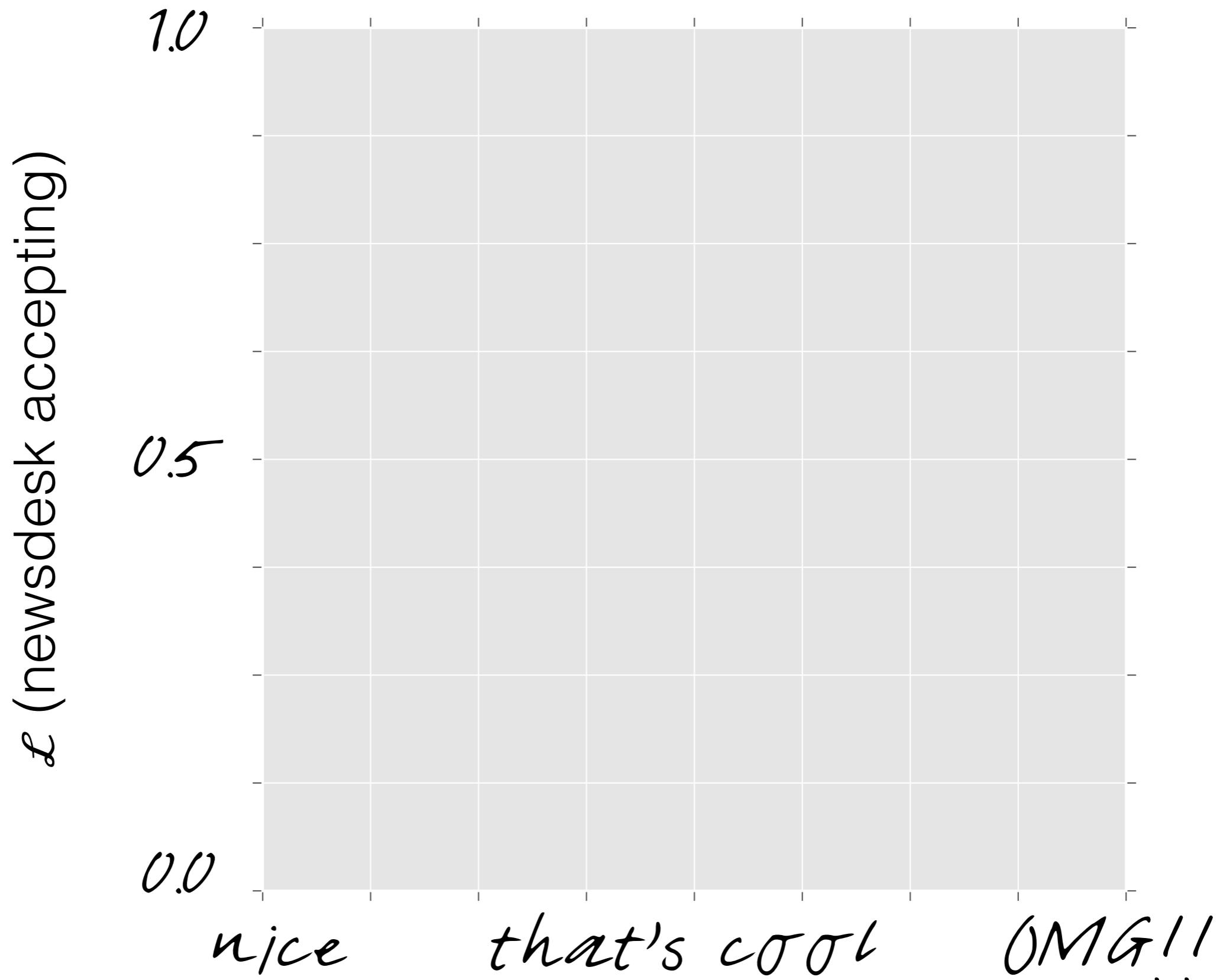
things I learnt about
myself



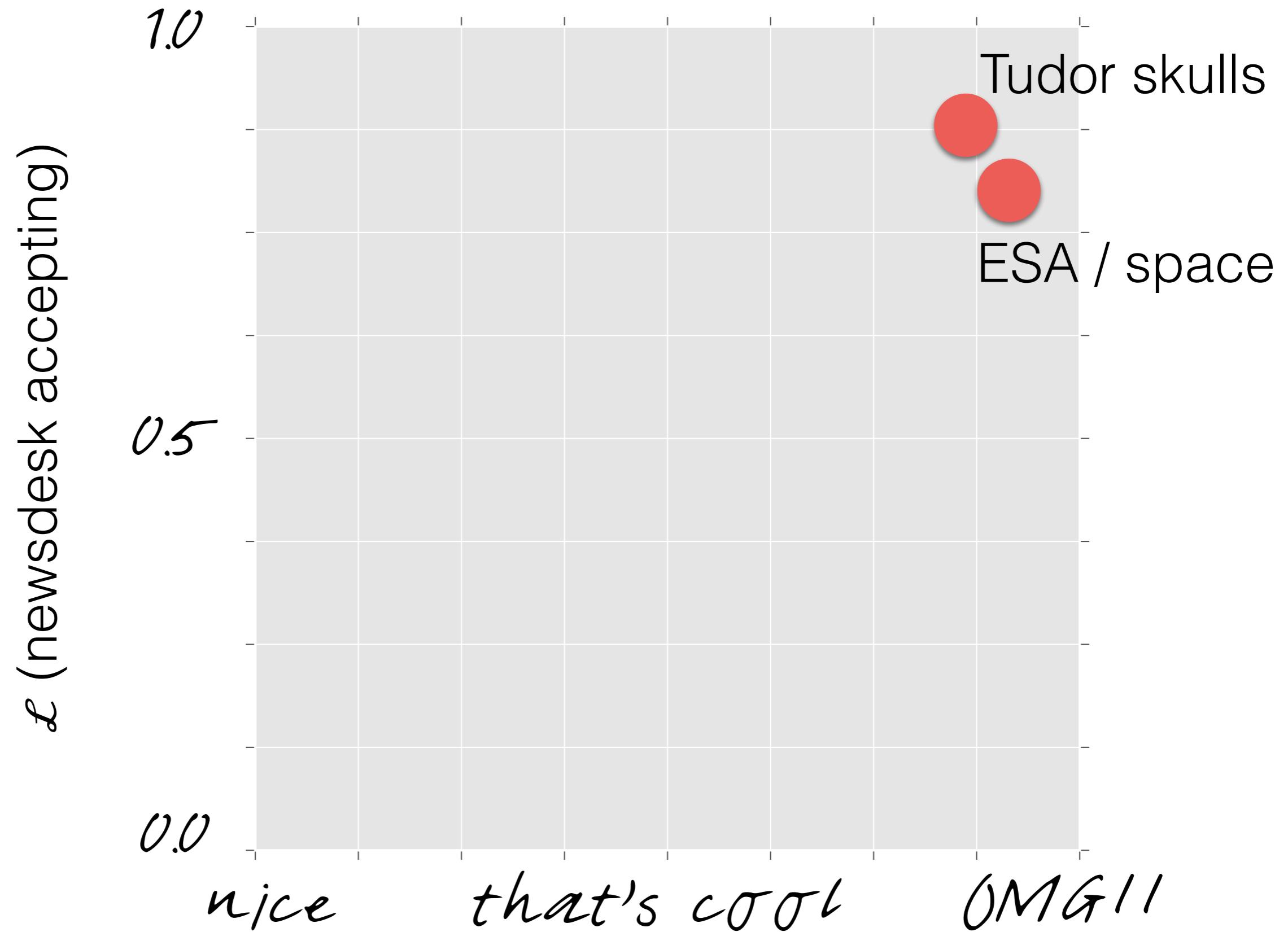


nice that's cool OMG!!

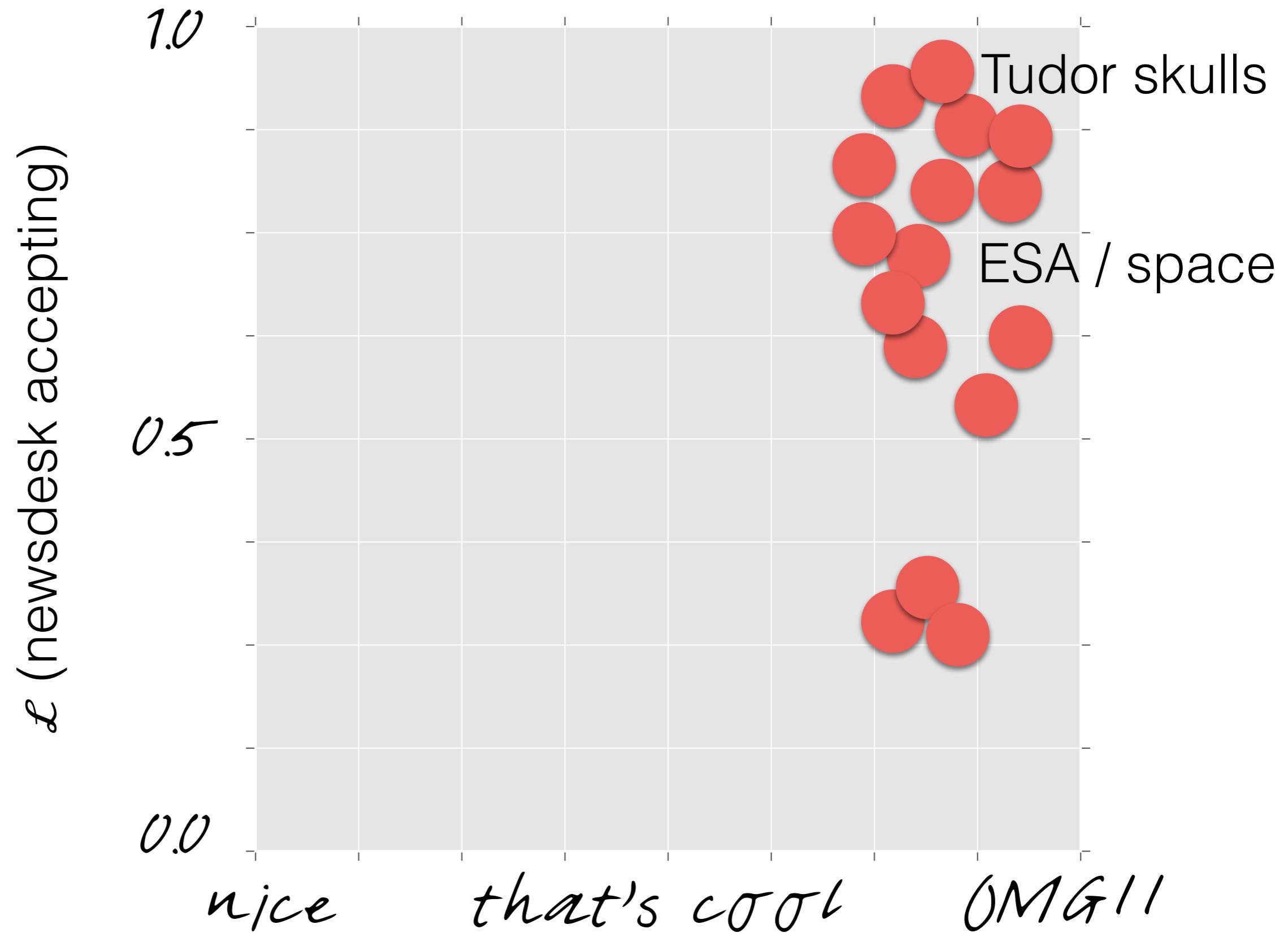
How fascinating do I find this story?



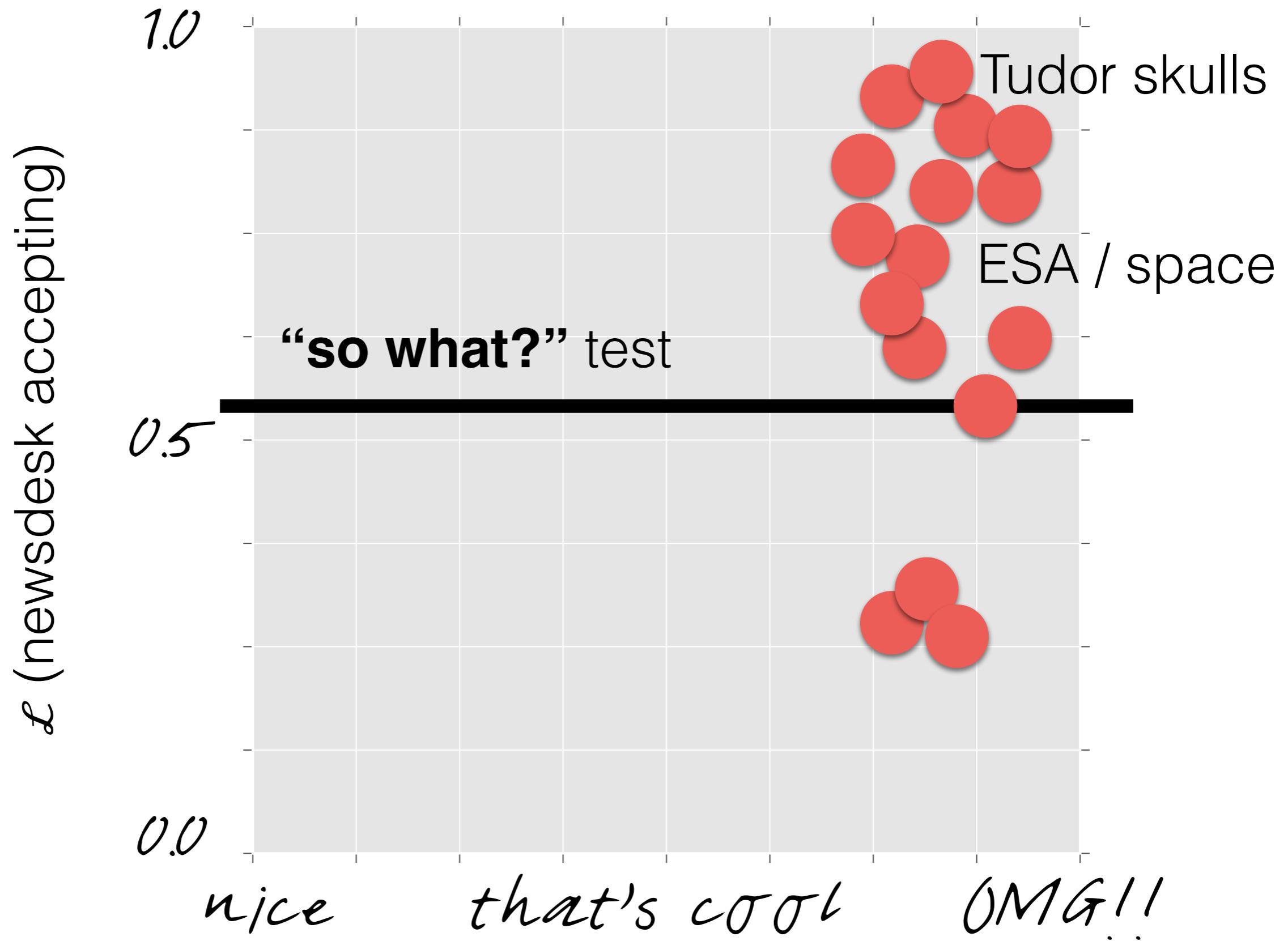
How fascinating do I find this story?



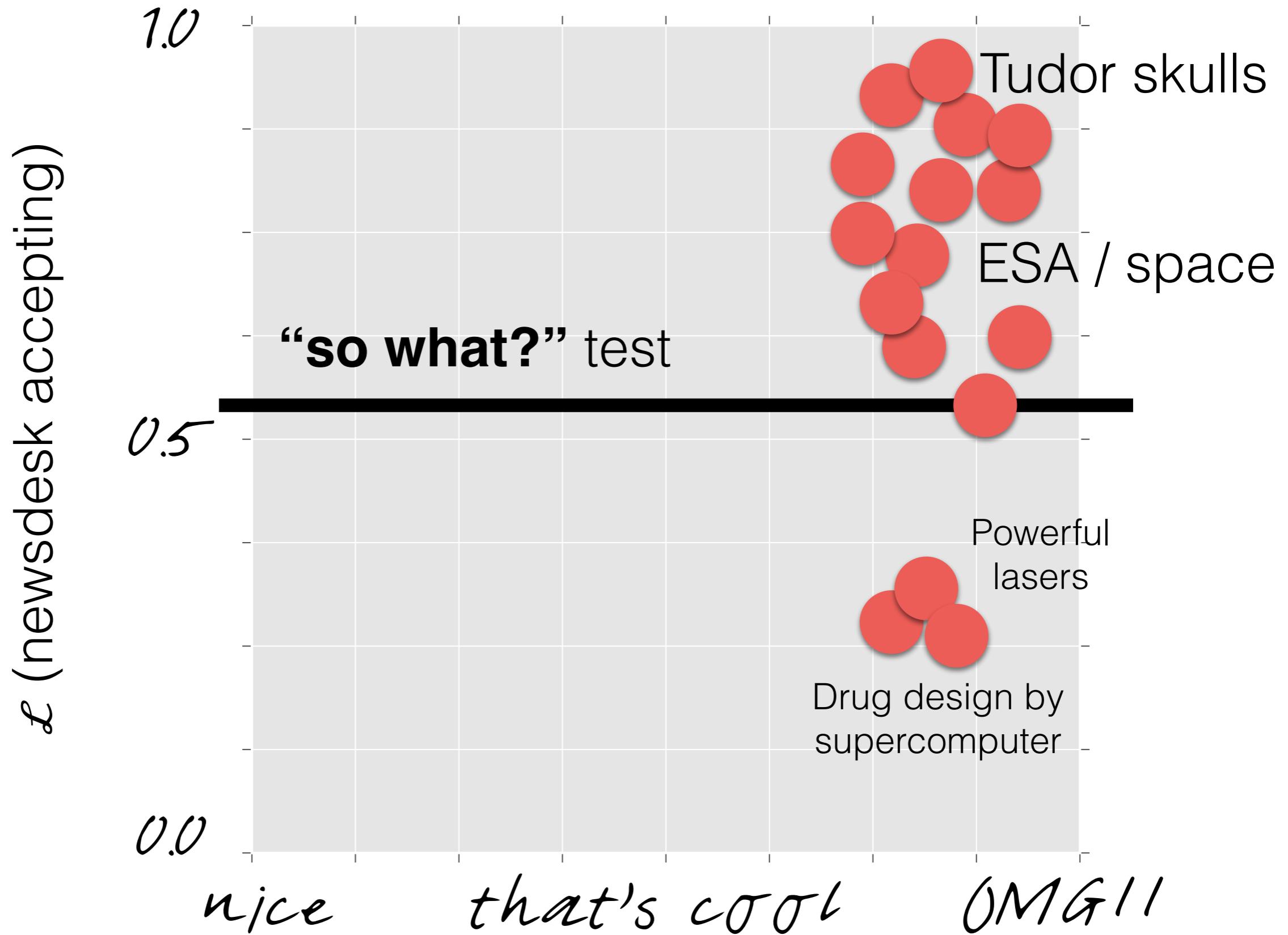
How fascinating do I find this story?



How fascinating do I find this story?



How fascinating do I find this story?



How fascinating do I find this story?

- **Clarity of writing** and **relevance to the reader** are the most important things.

- **Clarity of writing** and **relevance to the reader** are the most important things.
- If you want to improve, **look for real feedback**, not only what you want to hear

- **Clarity of writing** and **relevance to the reader** are the most important things.
- If you want to improve, **look for real feedback**, not only what you want to hear
- **journalists are busy** and the pace is pretty crazy

“If a good scientific paper is a 5-course menu, a well-written news piece is the perfect serving of sushi.”

–Denis Schluppeck



What can you (and
researchers) do to increase
engagement?

Build relationships

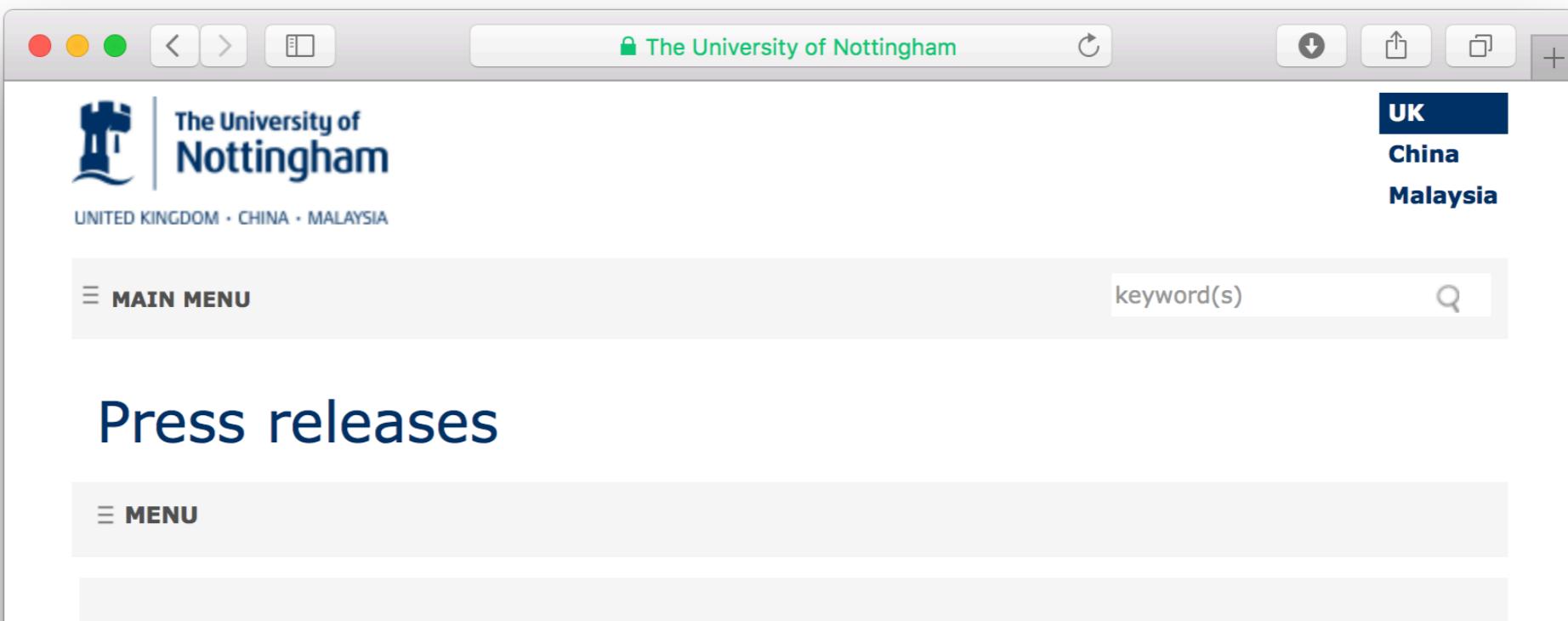
- try to meet science journalists - in person.
 - different outlets (broadsheets – to magazines)
 - local radio (for some stories)
 - ... may lead to national coverage
 - ... may lead to the next story getting picked up

at ~~Nottingham~~...
your institution

U o Nairobi: <https://www.uonbi.ac.ke/news>

U o Rwanda: [https://ur.ac.rw/spip.php?
page=news-and-events](https://ur.ac.rw/spip.php?page=news-and-events)

... look up your own media / press office



Timescales

- we / scientists need to learn to plan ahead: talk to media people about work that might be relevant
- media deadlines are much tighter than academic ones – we need to put ourselves in their shoes + be ready
- may require culture shift +/- help from experienced professionals

Different work – different channels

- if a story doesn't end up getting in, don't take it personally [many reasons why things get "bumped"]
- not everything is front page news
- some work might be more suitable for a blog
- things get re-tweeted, re-discovered, shared on Facebook, go viral...

theconversation.com

Edition: United Kingdom | Donate | Events | Get newsletter | Facebook | Twitter | RSS | Become an author | Sign up as a reader | Sign in

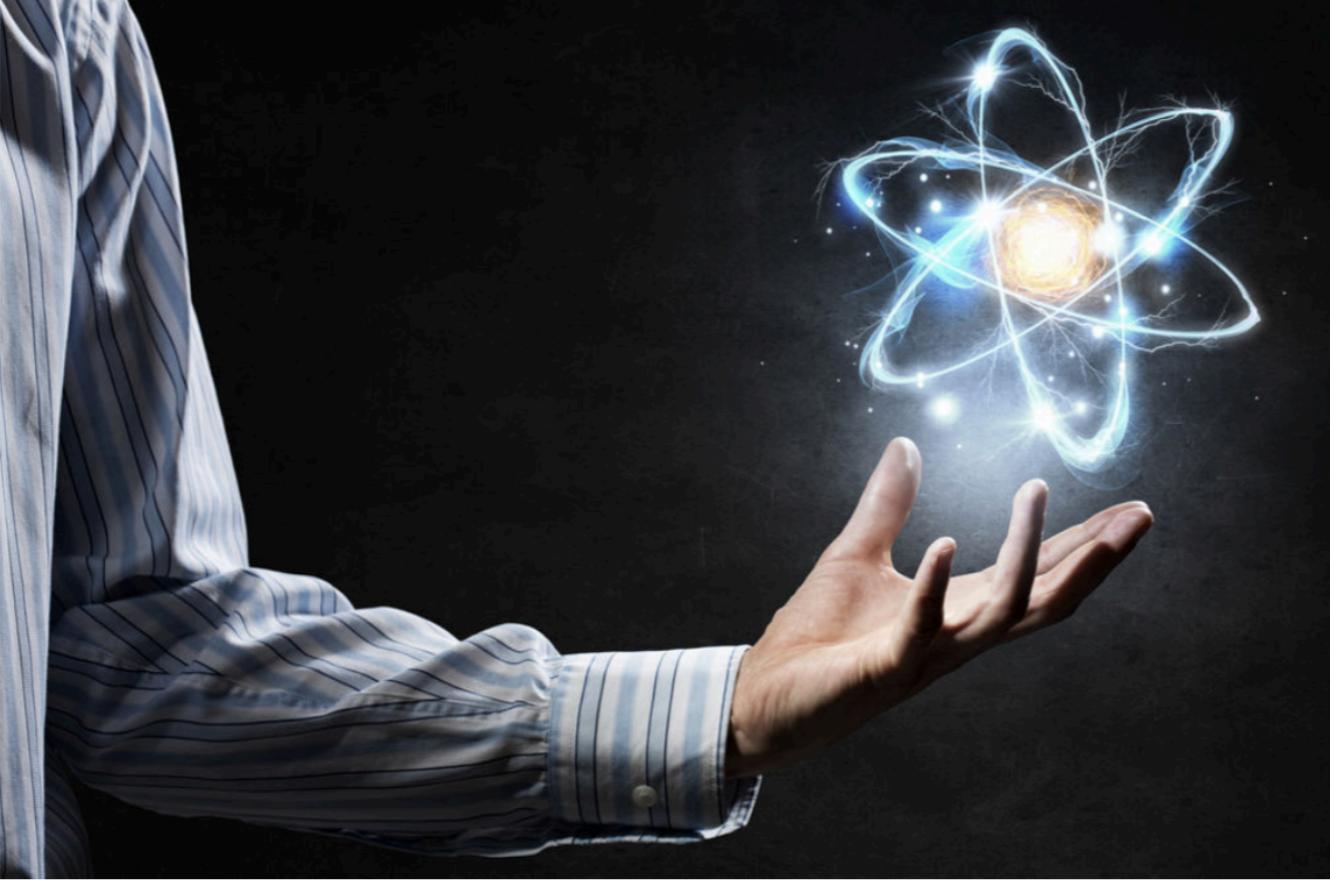
THE CONVERSATION

Academic rigour, journalistic flair

Search analysis, research, academics...

Arts + Culture Business + Economy Education Environment + Energy Health + Medicine Politics + Society Science + Technology Brexit

Follow Topics Cities Climate change Space Syria Migration Brexit Health



Why can't we see the spaces? Shutterstock

If atoms are mostly empty space, why do objects look and feel solid?

Podcast



Selected updates

Curated Tweets by @ConversationUK

 **John Jewell** @jjohnjewell 
#BritishComedyclassnationandidentity My piece on the success of Private Eye... theconversation.com/private-eye-ci... ... via @ConversationUK



become an author!

The screenshot shows a web browser window for the URL <https://theconversation.com/become-an-author>. The page header includes the site's logo, navigation links for 'Edition: Africa', 'Get newsletter', 'Become an author', 'Sign up as a reader', and 'Sign in'. A search bar is also present. The main content features a large heading 'Can you write for The Conversation?'. Below it, a text block specifies the requirements: 'To be published by The Conversation you must be currently employed as a researcher or academic with a university or research institution. PhD candidates under supervision by an academic can write for us, but we don't currently publish articles from Masters students.' Three steps are outlined at the bottom: '1. Verify Institution', '2. Education History', and '3. Account Password'.

Become an author - The Conv X +

← → ⌛ https://theconversation.com/become-an-author

Home Edition: Africa Get newsletter Become an author Sign up as a reader Sign in

THE CONVERSATION

Search analysis, research, academics...

Can you write for The Conversation?

To be published by The Conversation you must be currently employed as a researcher or academic with a university or research institution. PhD candidates under supervision by an academic can write for us, but we don't currently publish articles from Masters students.

1. Verify Institution
Please identify your current institution.

2. Education History
Tell us a bit about your formal qualifications.

3. Account Password
Set your password, agree the terms and write!

1. ~~the big picture: science communication~~
2. **the nitty-gritty:** some practical tips (formats, basic principles, version control)
3. **an example:** bringing a jupyter notebook to life

- **tidy data** principles
- **version control** with git/github
- **markdown** is queen / king

Tidy data

Rethinking your plotting (+ data sharing) habits

Aim

Make us think about how to handle data to be

- reproducible (+ easy to understand later)
- more flexible for exploration
 - in other kinds of plots, visualisations
 - with other tools / programming languages
 - by other people (colleagues, people downloading when you publish)

inspiration for this



Journal of Statistical Software
MMMMMM YYYY, Volume VV, Issue II. <http://www.jstatsoft.org/>

Tidy Data

Hadley Wickham
RStudio

Abstract

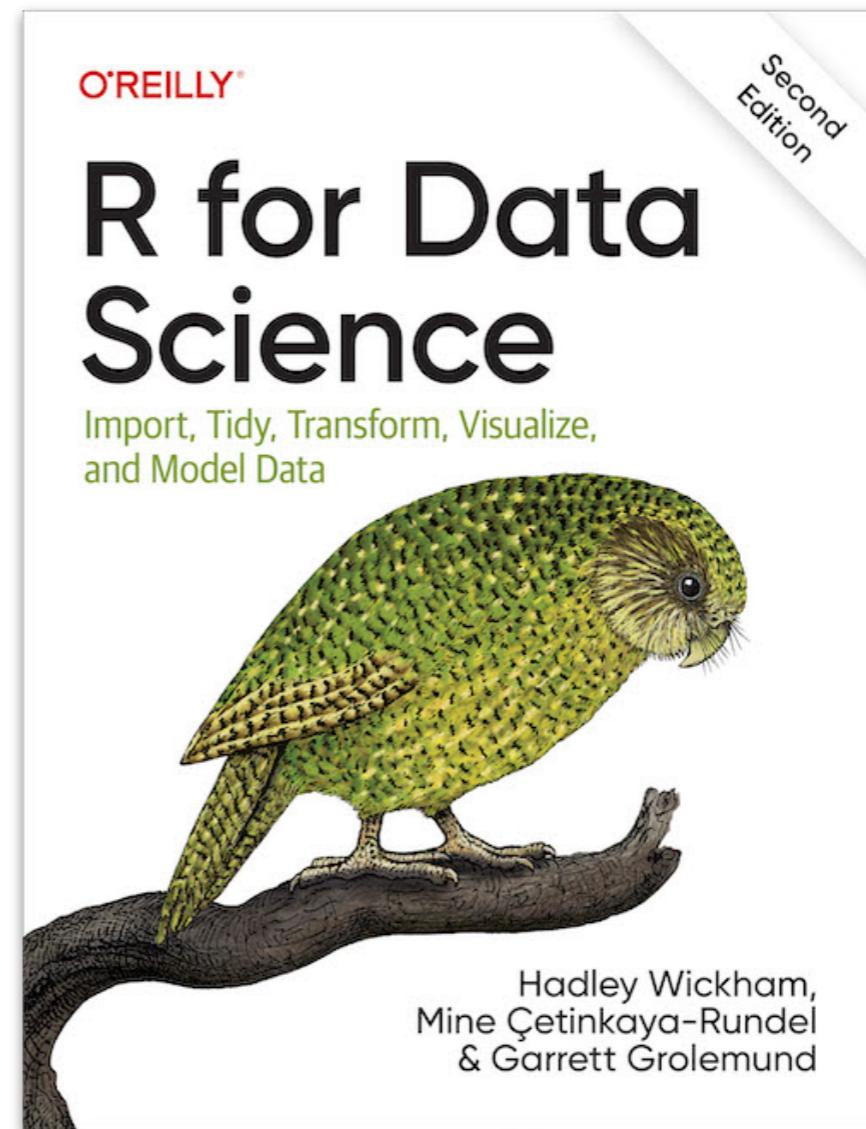
A huge amount of effort is spent cleaning data to get it ready for analysis, but there has been little research on how to make data cleaning as easy and effective as possible. This paper tackles a small, but important, component of data cleaning: data tidying. Tidy datasets are easy to manipulate, model and visualise, and have a specific structure: each variable is a column, each observation is a row, and each type of observational unit is a table. This framework makes it easy to tidy messy datasets because only a small set of tools are needed to deal with a wide range of un-tidy datasets. This structure also makes it easier to develop tidy tools for data analysis, tools that both input and output tidy datasets. The advantages of a consistent data structure and matching tools are demonstrated with a case study free from mundane data manipulation chores.

Keywords: data cleaning, data tidying, relational databases, R.

1. Introduction

It is often said that 80% of data analysis is spent on the process of cleaning and preparing the data (Dasu and Johnson 2003). Data preparation is not just a first step, but must be repeated many over the course of analysis as new problems come to light or new data is collected. Despite the amount of time it takes, there has been surprisingly little research on how to clean data well. Part of the challenge is the breadth of activities it encompasses: from outlier checking, to date parsing, to missing value imputation. To get a handle on the problem, this paper focusses on a small, but important, aspect of data cleaning that I call data **tidying**: structuring datasets to facilitate analysis.

The principles of tidy data provide a standard way to organise data values within a dataset. A standard makes initial data cleaning easier because you don't need to start from scratch and reinvent the wheel every time. The tidy data standard has been designed to facilitate initial exploration and analysis of the data, and to simplify the development of data analysis tools that work well together. Current tools often require translation. You have to spend time



links on **references** slide

inspiration for this

The image shows the cover of a journal article titled "Tidy Data" by Hadley Wickham. The cover features a small illustration of a bird on the left. The title "Journal of Statistical Software" is at the top, followed by the volume and issue information "MMMMMM YYYY, Volume VV, Issue II". The URL "http://www.jstatsoft.org/" is also present. Below the title, the author's name "Hadley Wickham" and affiliation "RStudio" are listed. A section titled "Abstract" follows, containing a detailed description of the paper's content. At the bottom, the keywords "data cleaning, data tidying, relational databases, R" are listed.

Tidy Data

Hadley Wickham
RStudio

Abstract

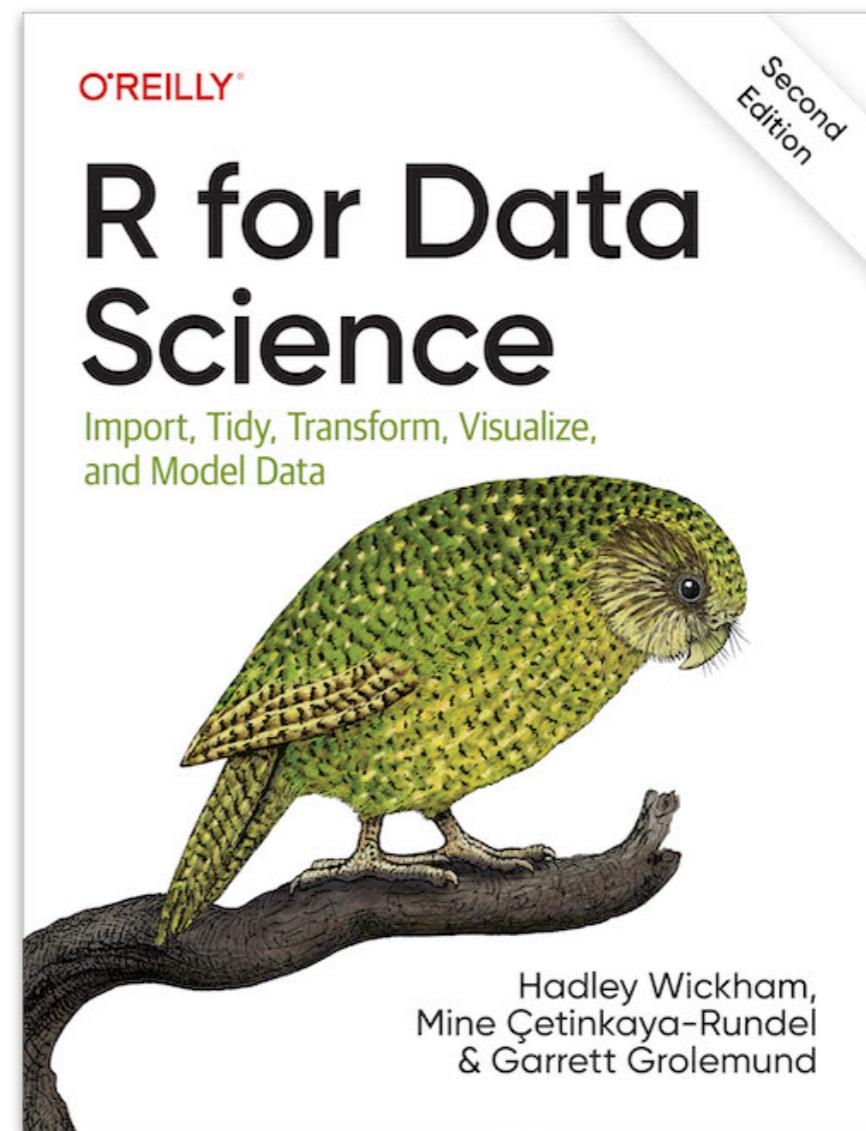
A huge amount of effort is spent cleaning data to get it ready for analysis, but there has been little research on how to make data cleaning as easy and effective as possible. This paper tackles a small, but important, component of data cleaning: data tidying. Tidy datasets are easy to manipulate, model and visualise, and have a specific structure: each variable is a column, each observation is a row, and each type of observational unit is a table. This framework makes it easy to tidy messy datasets because only a small set of tools are needed to deal with a wide range of un-tidy datasets. This structure also makes it easier to develop tidy tools for data analysis, tools that both input and output tidy datasets. The advantages of a consistent data structure and matching tools are demonstrated with a case study free from mundane data manipulation chores.

Keywords: data cleaning, data tidying, relational databases, R.

1. Introduction

It is often said that 80% of data analysis is spent on the process of cleaning and preparing the data (Dasu and Johnson 2003). Data preparation is not just a first step, but must be repeated many over the course of analysis as new problems come to light or new data is collected. Despite the amount of time it takes, there has been surprisingly little research on how to clean data well. Part of the challenge is the breadth of activities it encompasses: from outlier checking, to date parsing, to missing value imputation. To get a handle on the problem, this paper focusses on a small, but important, aspect of data cleaning that I call data **tidying**: structuring datasets to facilitate analysis.

The principles of tidy data provide a standard way to organise data values within a dataset. A standard makes initial data cleaning easier because you don't need to start from scratch and reinvent the wheel every time. The tidy data standard has been designed to facilitate initial exploration and analysis of the data, and to simplify the development of data analysis tools that work well together. Current tools often require translation. You have to spend time



links on **references** slide

Kinds of data?

3d, 4d arrays (+ higher)

imaging data

imaging data + some metadata
+ TXT (e.g. bvecs)

NIfTI

GIFTI,
freesurfer, VTK
...

Kinds of data?

3d, 4d arrays (+ higher)

imaging data

imaging data + some metadata
+ TXT (e.g. bvecs)



other stuff (for summary / figures, etc)

tables

key:value

hierarchical

Kinds of data?

3d, 4d arrays (+ higher)

imaging data

imaging data + some metadata
+ TXT (e.g. bvecs)

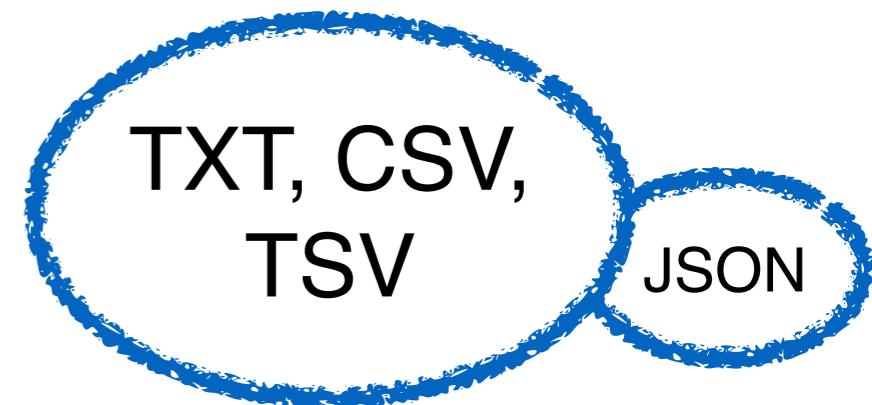


other stuff (for summary / figures, etc)

tables

key:value

hierarchical



idea: for images

www.nature.com/scientificdata

SCIENTIFIC DATA



OPEN

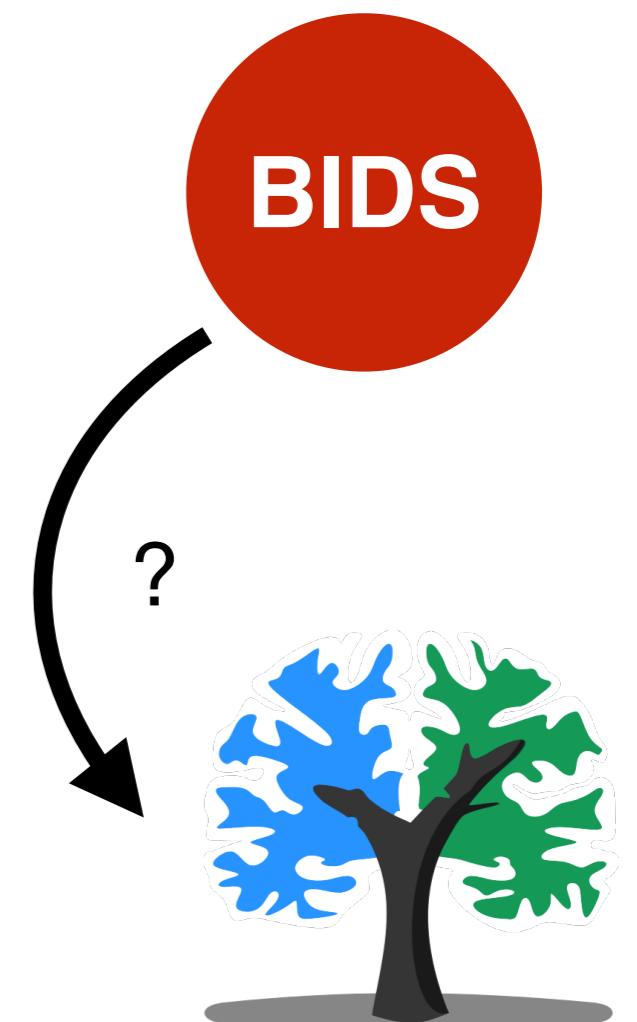
SUBJECT CATEGORIES
» Data publication and archiving
» Research data

Received: 18 December 2015
Accepted: 19 May 2016
Published: 21 June 2016

The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments

Krzysztof J. Gorgolewski¹, Tibor Auer², Vince D. Calhoun^{3,4}, R. Cameron Craddock^{5,6}, Samir Das⁷, Eugene P. Duff⁸, Guillaume Flandin⁹, Satrajit S. Ghosh^{10,11}, Tristan Glatard^{7,12}, Yaroslav O. Halchenko¹³, Daniel A. Handwerker¹⁴, Michael Hanke^{15,16}, David Keator¹⁷, Xiangrui Li¹⁸, Zachary Michael¹⁹, Camille Maumet²⁰, B. Nolan Nichols^{21,22}, Thomas E. Nichols^{20,23}, John Pellman⁶, Jean-Baptiste Poline²⁴, Ariel Rokem²⁵, Gunnar Schaefer^{1,26}, Vanessa Sochat²⁷, William Triplett¹, Jessica A. Turner^{3,28}, Gaël Varoquaux²⁹ & Russell A. Poldrack¹

The development of magnetic resonance imaging (MRI) techniques has defined modern neuroimaging. Since its inception, tens of thousands of studies using techniques such as functional MRI and diffusion weighted imaging have allowed for the non-invasive study of the brain. Despite the fact that MRI is routinely used to obtain data for neuroscience research, there has been no widely adopted standard for organizing and describing the data collected in an imaging experiment. This renders sharing and reusing data (within or between labs) difficult if not impossible and unnecessarily complicates the application of automatic pipelines and quality assurance protocols. To solve this problem, we have developed the Brain Imaging Data Structure (BIDS), a standard for organizing and describing MRI datasets. The BIDS standard uses file formats compatible with existing software, unifies the majority of practices already common in the field, and captures the metadata necessary for most common data processing operations.



fMRI_report()
QA toolbox
MRIQC

idea: for tabular data



Journal of Statistical Software

MMMMMM Volume II.

<http://www.jstatsoft.org/>

Tidy Data

Hadley Wickham

Abstract

A huge amount of effort is spent cleaning data to get it ready for analysis, but there has been little research on how to make data cleaning as easy and effective as possible. This paper tackles a small, but important, component of data cleaning: data tidying. Tidy datasets are easy to manipulate, model and visualise, and have a specific structure:

“Happy families are all alike; every unhappy family
is unhappy in its own way.” — Leo Tolstoy

“Happy families are all alike; every unhappy family
is unhappy in its own way.” — Leo Tolstoy

“Tidy datasets are all alike, but every messy dataset
is messy in its own way.” — Hadley Wickham

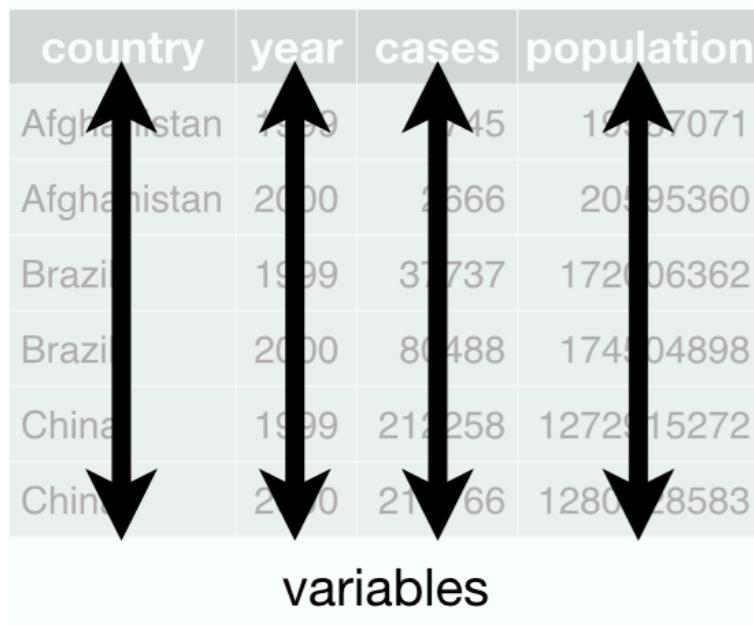
Principles

1. Each variable must have its own column.
2. Each observation must have its own row.
3. Each value must have its own cell.

Principles

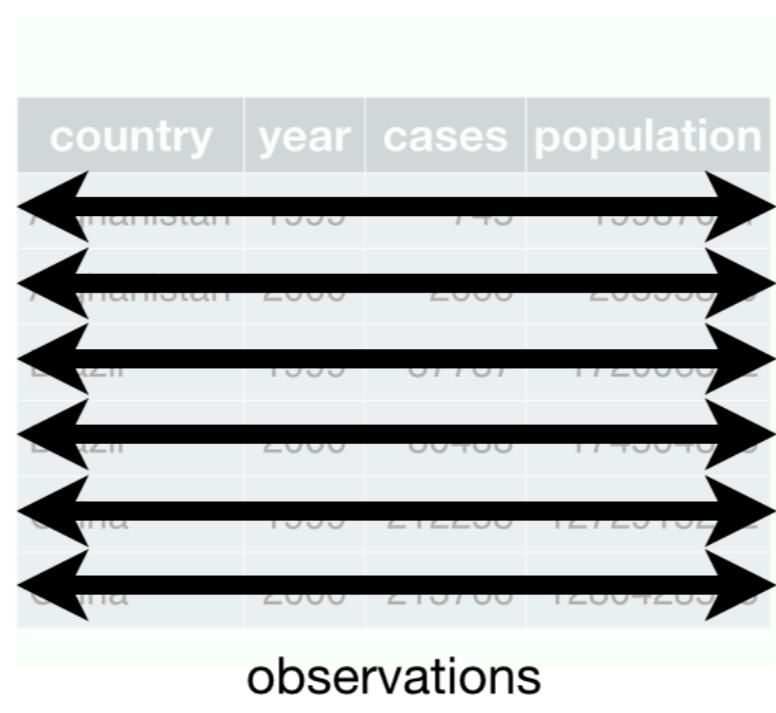
country	year	cases	population
Afghanistan	1999	745	19357071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	128042583

variables



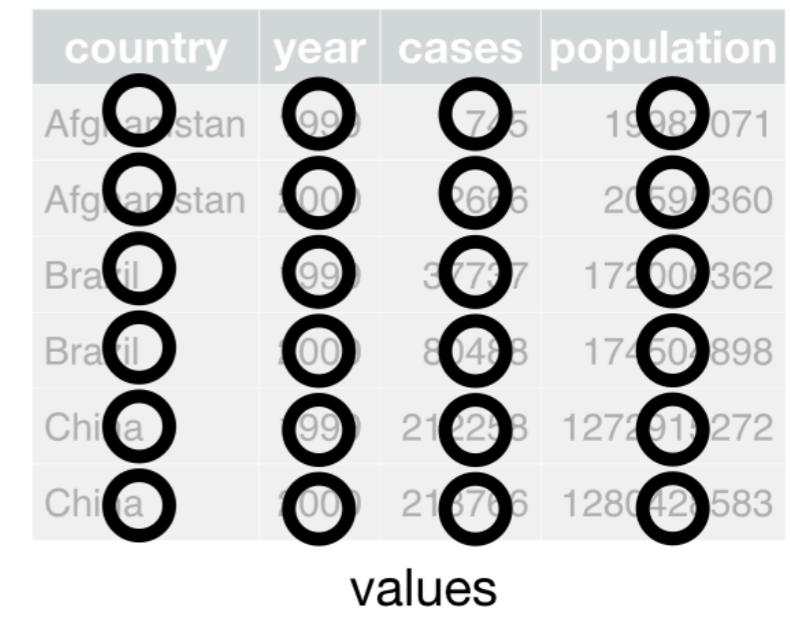
country	year	cases	population
Afghanistan	1999	745	19357071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	128042583

observations



country	year	cases	population
Afghanistan	1999	745	19357071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	128042583

values



+ even if you have megabytes
of data, consider

Principles

country	year	cases	population
Afghanistan	1999	745	19357071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	128042583

variables

This diagram illustrates the concept of variables. It shows a data frame with four columns: country, year, cases, and population. Four vertical arrows point upwards from the column labels to the first row of data, and four vertical arrows point downwards from the last row of data back to the column labels, indicating that each column represents a specific variable.

country	year	cases	population
Afghanistan	1999	745	19357071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	128042583

observations

This diagram illustrates the concept of observations. It shows a data frame with six rows, each representing an observation of the variables country, year, cases, and population for different countries and years. Six horizontal arrows point to the right from the first column of each row, indicating that each row represents a single observation.

country	year	cases	population
Afghanistan	999	745	19357071
Afghanistan	000	2666	20595360
Brasil	999	37737	172006362
Brasil	000	80488	174504898
China	999	212258	1272915272
China	000	213766	128042583

values

This diagram illustrates the concept of values. It shows a data frame where each cell contains a black circle. The circles are arranged in a grid corresponding to the data in the table above. This visualizes how each cell in the table represents a single data value.

+ even if you have megabytes
of data, consider



If you stick to this...

... much goodness will result!

If you stick to this...

... much goodness will result!

- predictable (in *your* head + in software)
- many tools handle these kind of data
- secondary analysis + plotting + visualising becomes a breeze

Visualisation tools

- R / ggplot2 (<https://ggplot2.tidyverse.org/>)
- Matlab / gramm (<https://github.com/piermorel/gramm>)
- python / pandas (and friends)
 - plotly
 - seaborn
 - plotnine (<https://plotnine.readthedocs.io/en/stable/>)

“verbs” for dealing with tables

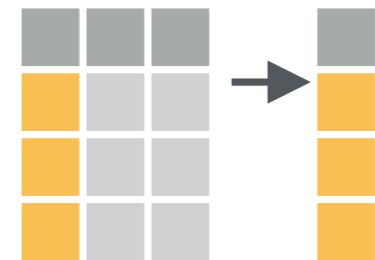
`select()`



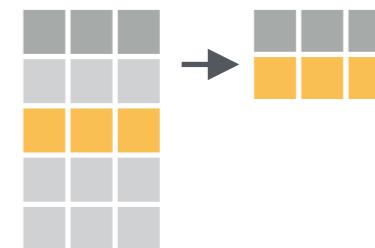
varyiations across APIs, but some version of them usually exists

“verbs” for dealing with tables

`select()`



`filter()`



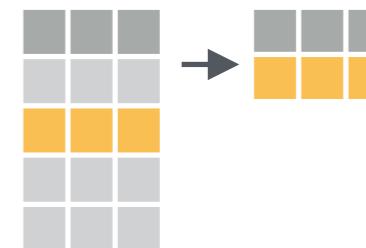
varyiations across APIs, but some version of them usually exists

“verbs” for dealing with tables

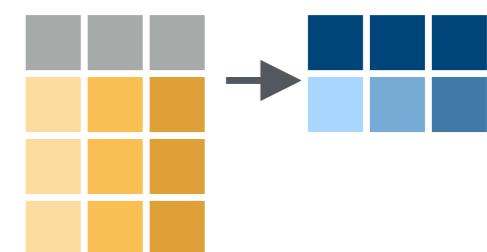
`select()`



`filter()`



`mutate()`



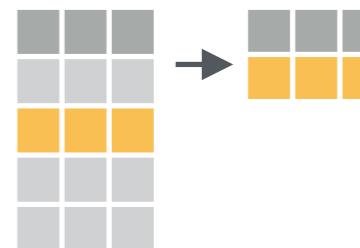
varyiations across APIs, but some version of them usually exists

“verbs” for dealing with tables

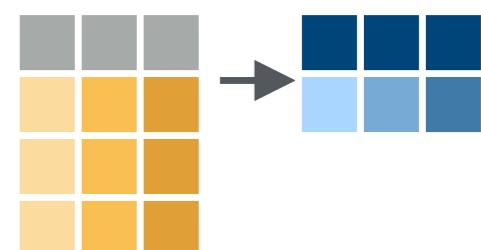
`select()`



`filter()`



`mutate()`



`groupby()/summarize()`

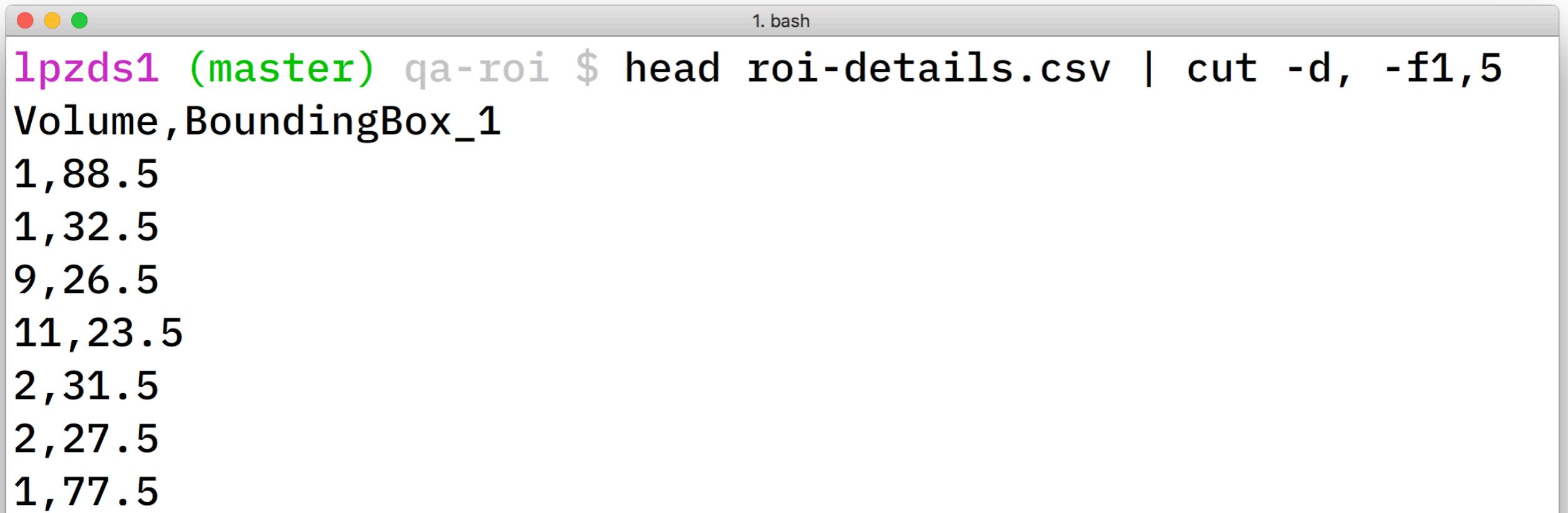
`group_by()/agg()`

varying across APIs, but some version of them usually exists

unix magic

head roi-details.csv

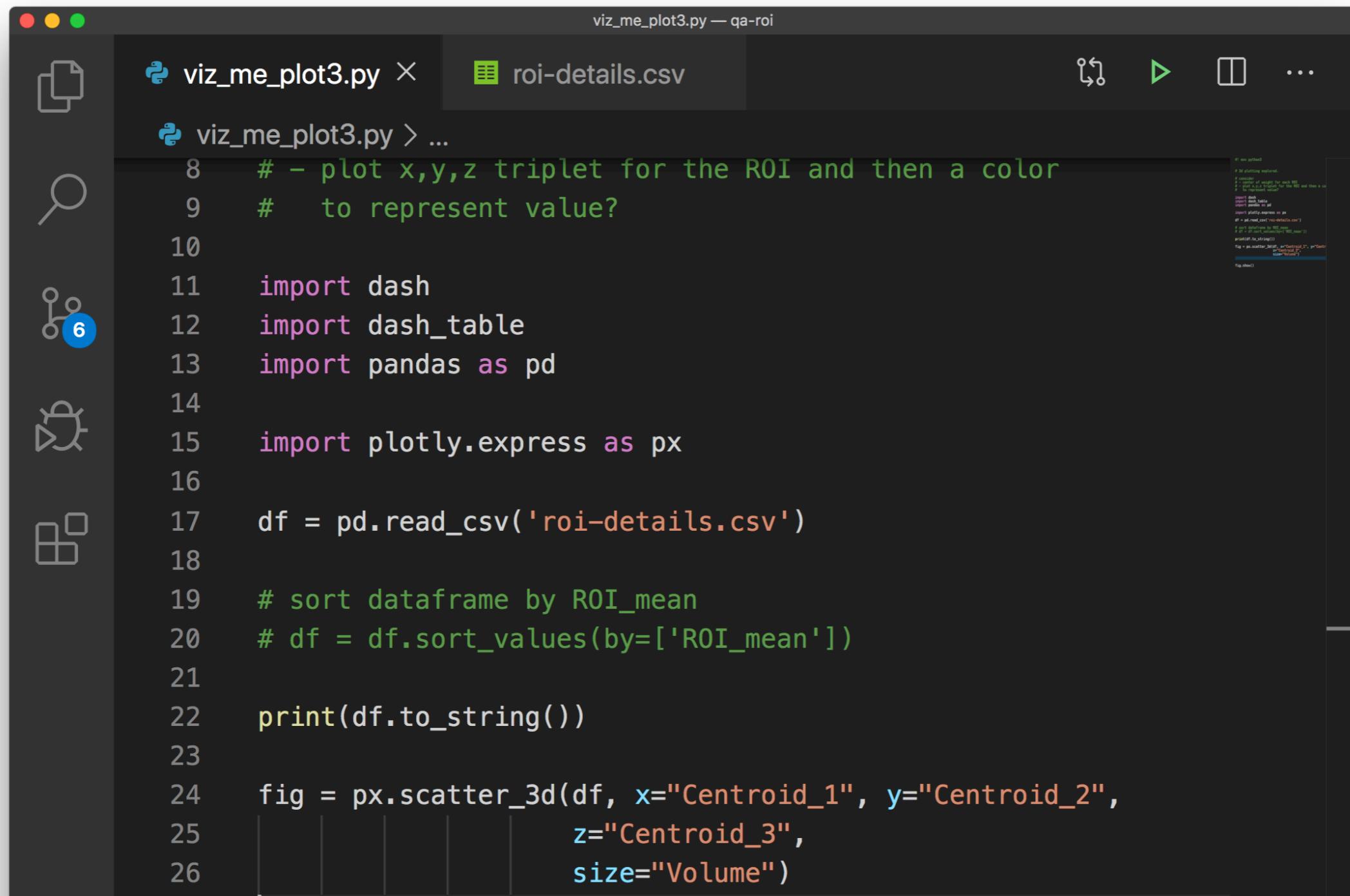
head roi-details.csv | cut -d, -f1,5



A screenshot of a Mac OS X terminal window titled "1.bash". The window shows a command being run: "head roi-details.csv | cut -d, -f1,5". The output of the command is displayed below the command line, showing the first five columns of the CSV file "roi-details.csv". The output is as follows:

```
lpzds1 (master) qa-roi $ head roi-details.csv | cut -d, -f1,5
Volume,BoundingBox_1
1,88.5
1,32.5
9,26.5
11,23.5
2,31.5
2,27.5
1,77.5
```

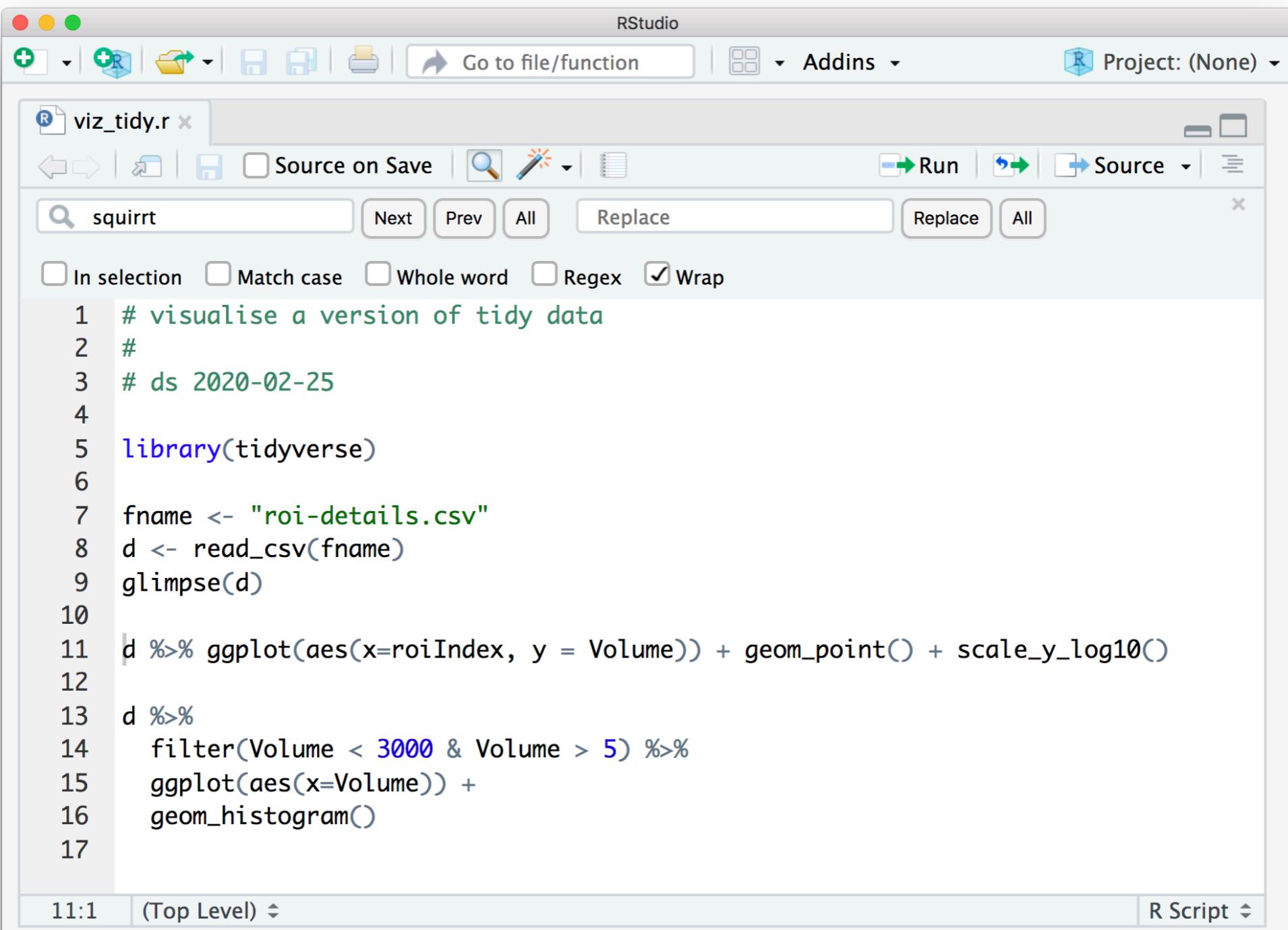
python / pandas magic



The screenshot shows a Jupyter Notebook interface with a dark theme. The top bar displays the file name "viz_me_plot3.py — qa-roi". The left sidebar contains icons for file operations, search, dependencies (with a '6' notification), and cell execution. The main area shows the following Python code:

```
# see methods
# We plotting explored.
# a centroid
# calculate weight for each ROI
# plot A, Z, & V (plot for the ROI and then a color
# to represent value?
import dash
import dash_table
import pandas as pd
import plotly.express as px
df = pd.read_csv('roi-details.csv')
# sort datafame by ROI_mean
# df = df.sort_values(by=['ROI_mean'])
print(df.to_string())
fig = px.scatter_3d(df, x="Centroid_1", y="Centroid_2",
                     z="Centroid_3",
                     size="Volume")
```

R / dplyr magic



The screenshot shows the RStudio interface with an R script file open. The file is titled "viz_tidy.r". The code in the script demonstrates the use of the dplyr package for data manipulation. It starts by visualizing a version of tidy data, then reads a CSV file named "roi-details.csv", and filters the data to show volumes between 5 and 3000. The code uses the ggplot2 package to create a scatter plot of Volume vs. ROI index.

```
# visualise a version of tidy data
#
# ds 2020-02-25

library(tidyverse)

fname <- "roi-details.csv"
d <- read_csv(fname)
glimpse(d)

d %>% ggplot(aes(x=roiIndex, y = Volume)) + geom_point() + scale_y_log10()

d %>%
  filter(Volume < 3000 & Volume > 5) %>%
  ggplot(aes(x=Volume)) +
  geom_histogram()
```

References

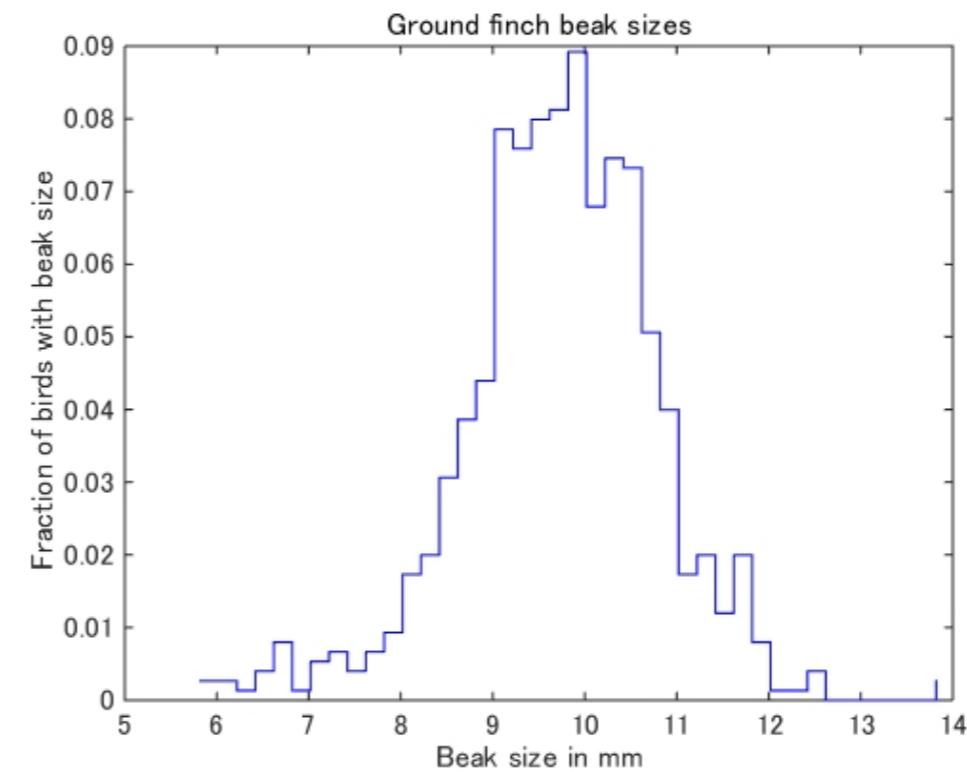
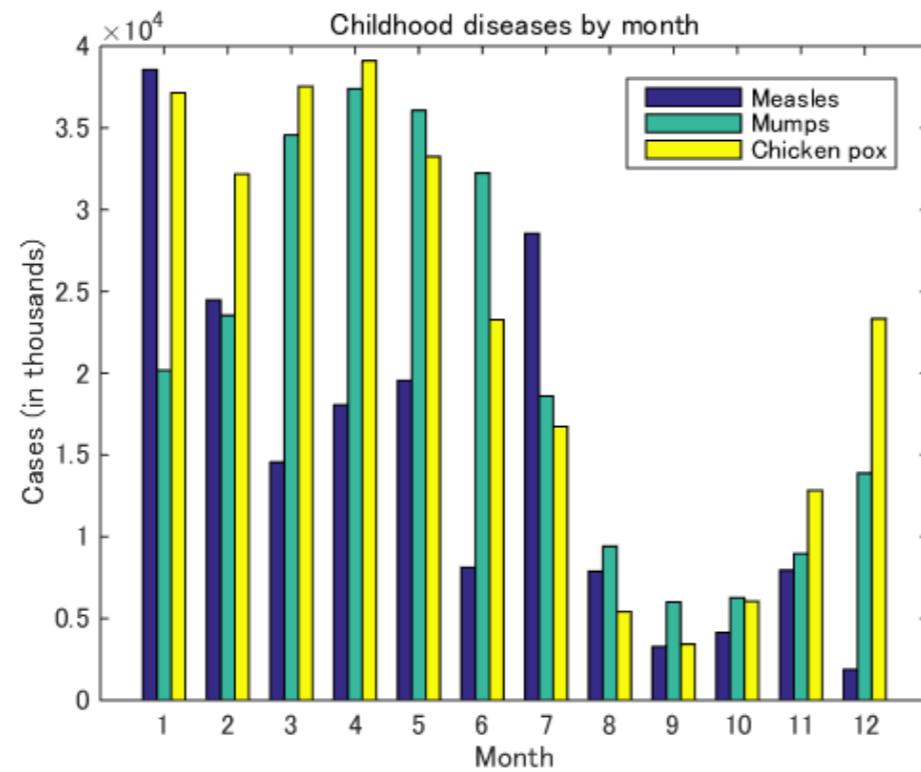
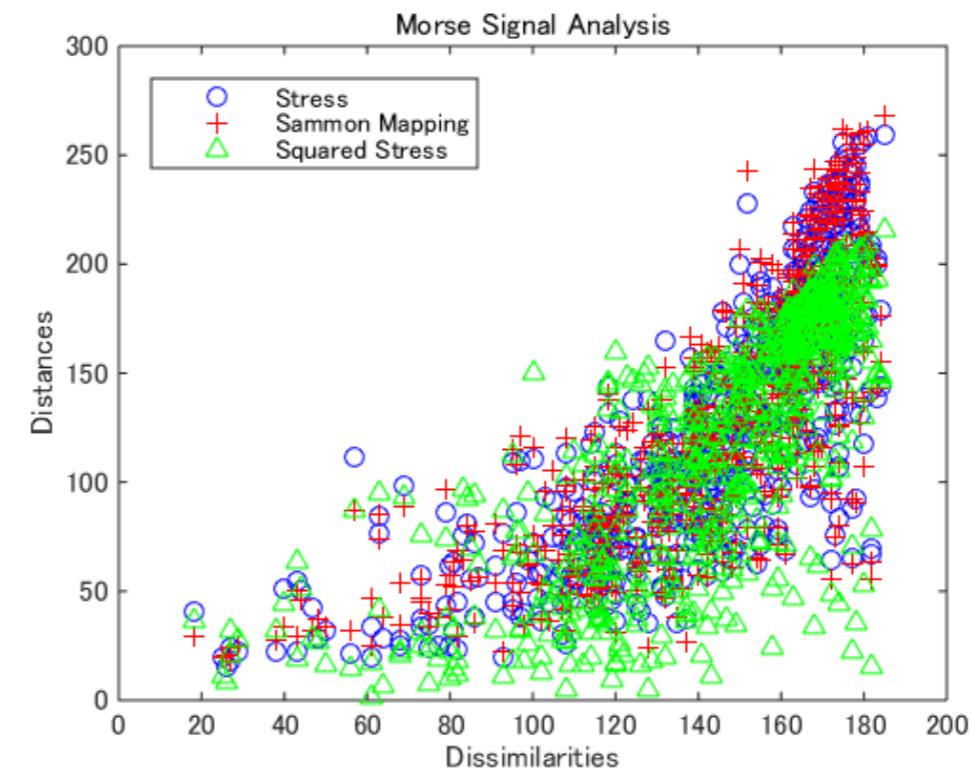
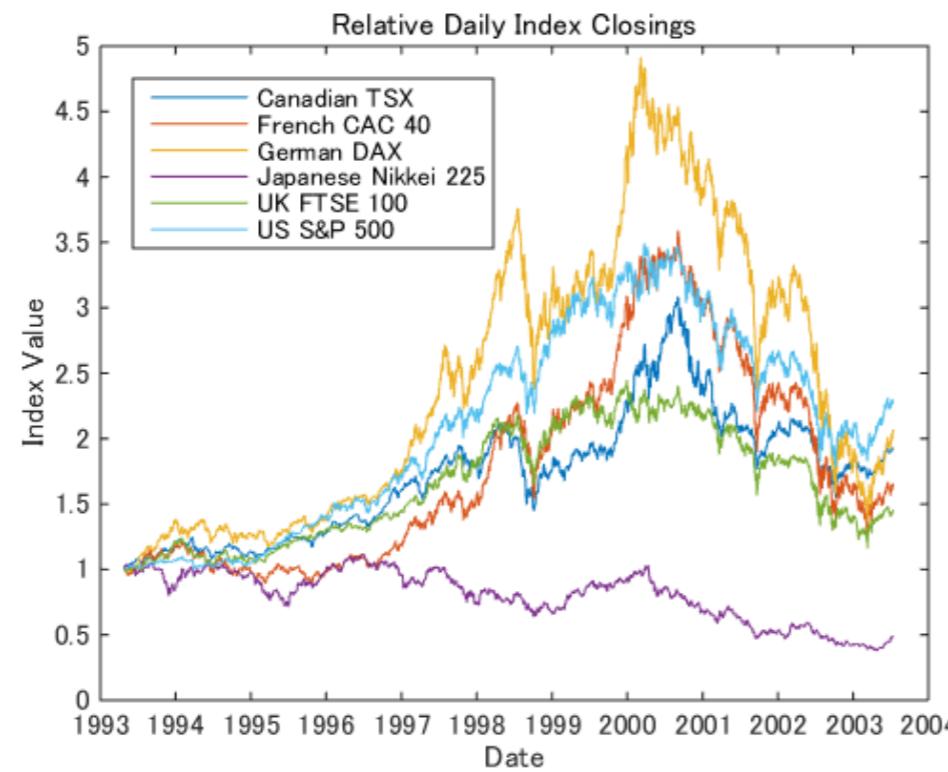
<https://r4ds.hadley.nz/>

<https://vita.had.co.nz/papers/tidy-data.pdf>

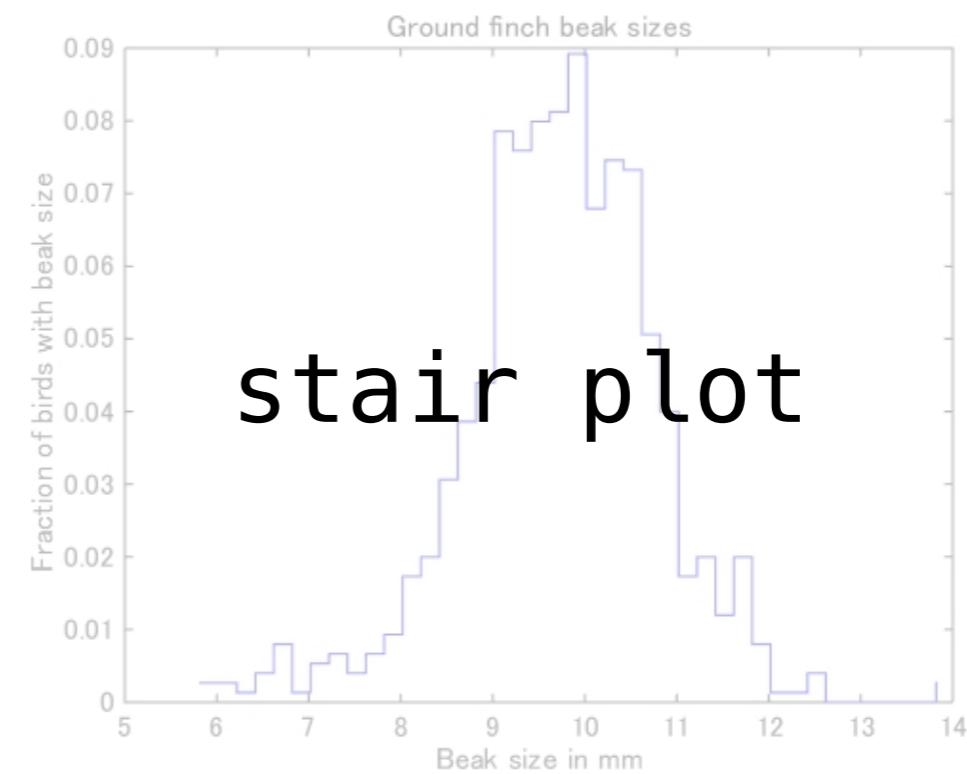
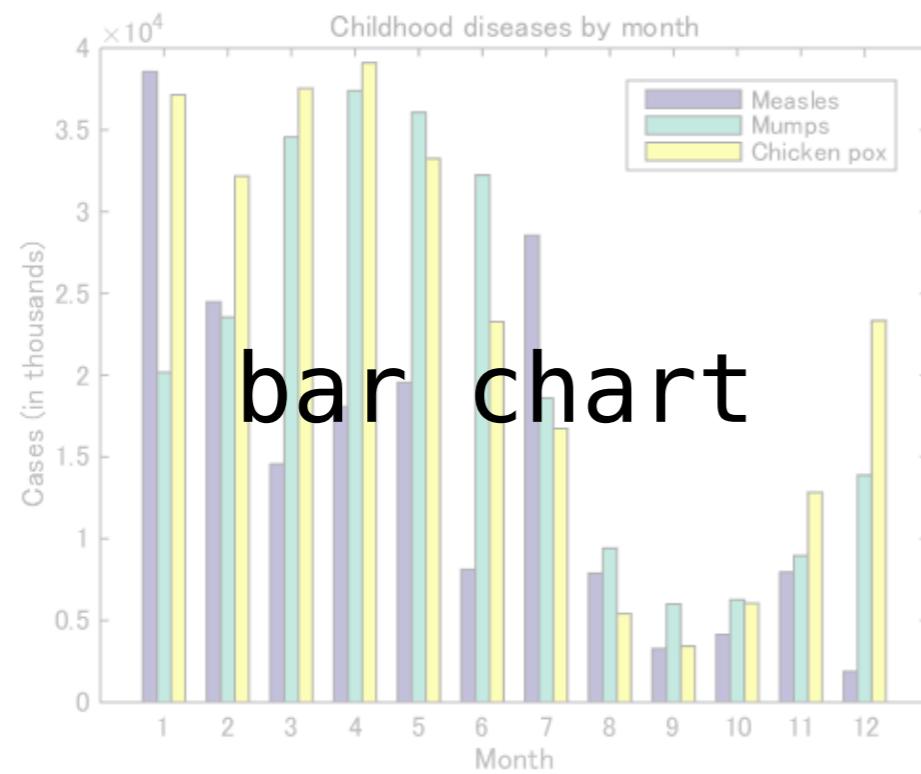
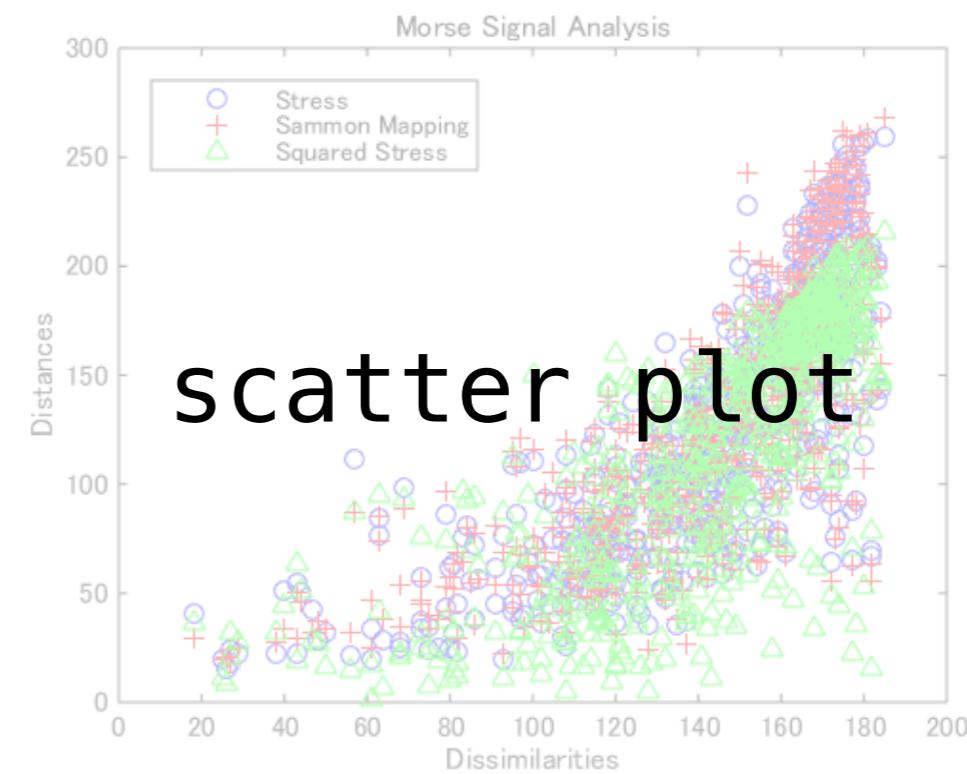
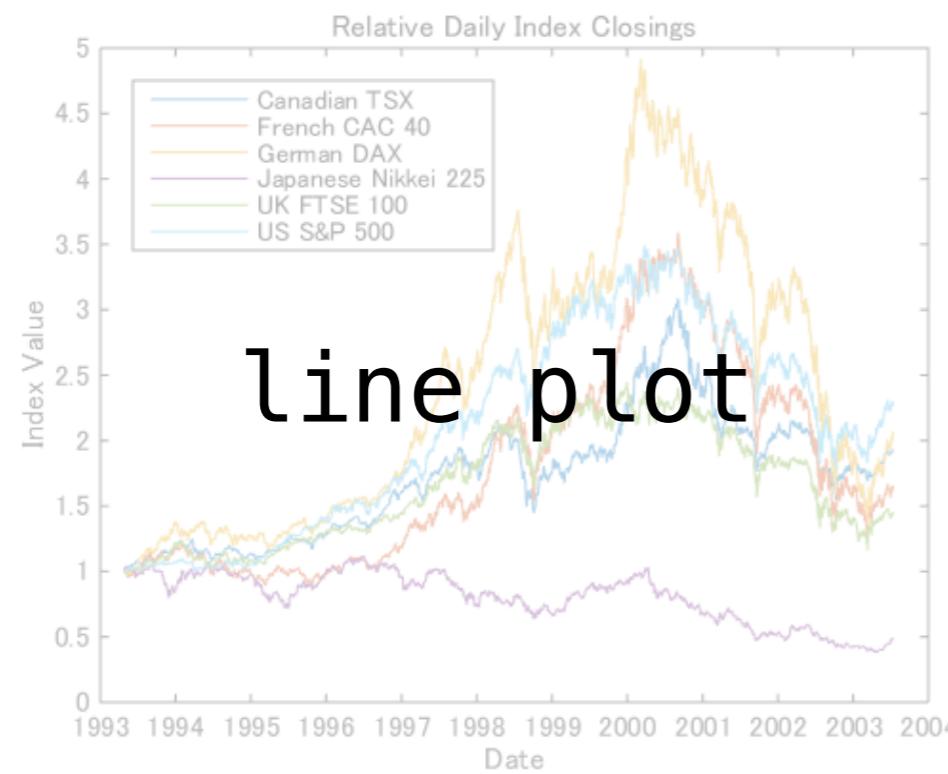
<https://pandas.pydata.org/docs/>

/visualise
/visualize

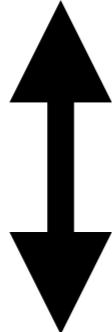
imperative



imperative



imperative



decide on plot type,
build plot step by step, ...

-ve: leads to repetitive
code / work

declarative

declarative

/dɪ'klærətɪv/

adjective

1. of the nature of or making a declaration.
"declarative statements"

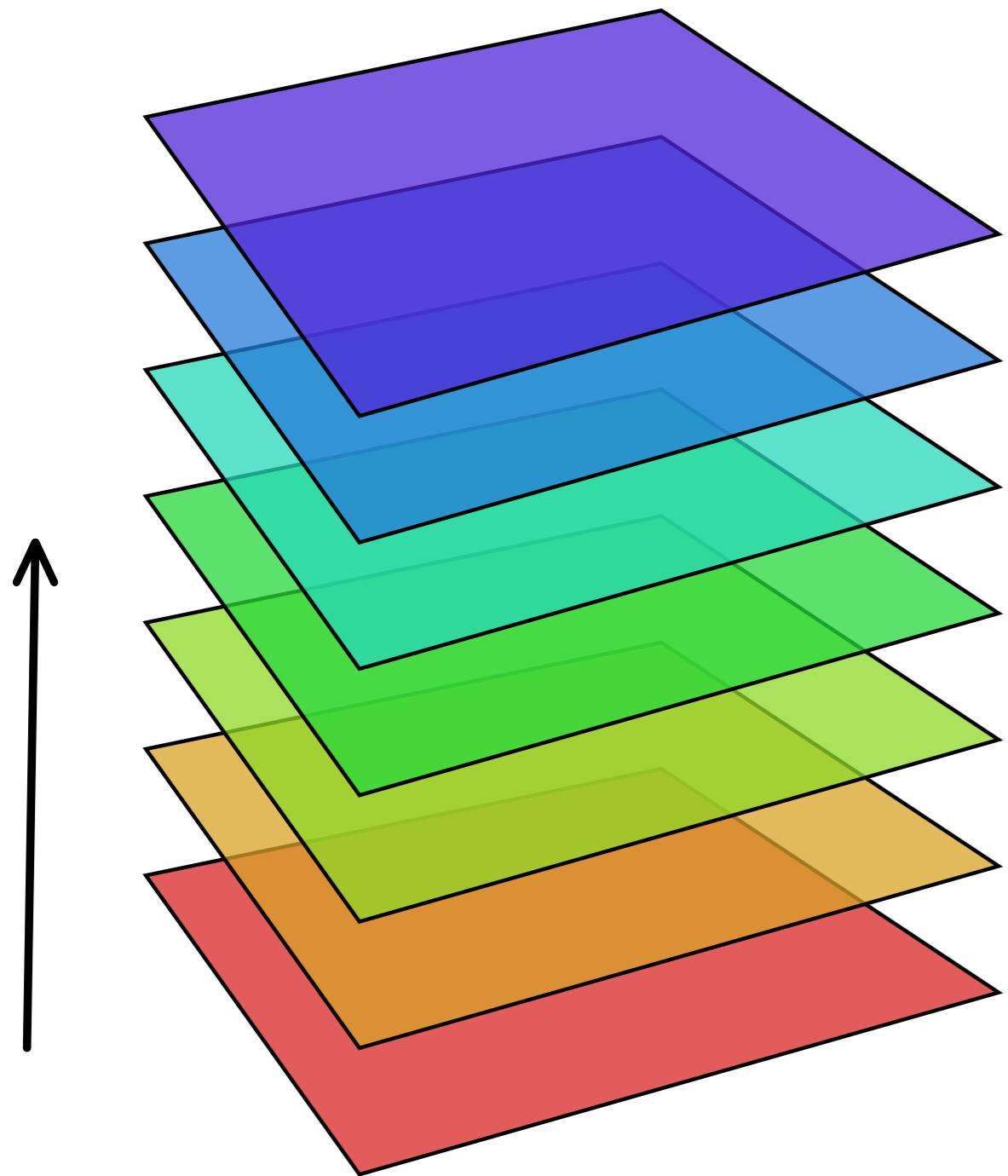
2. COMPUTING

denoting high-level programming languages which can be used to solve problems without requiring the programmer to specify an exact procedure to be followed.

noun

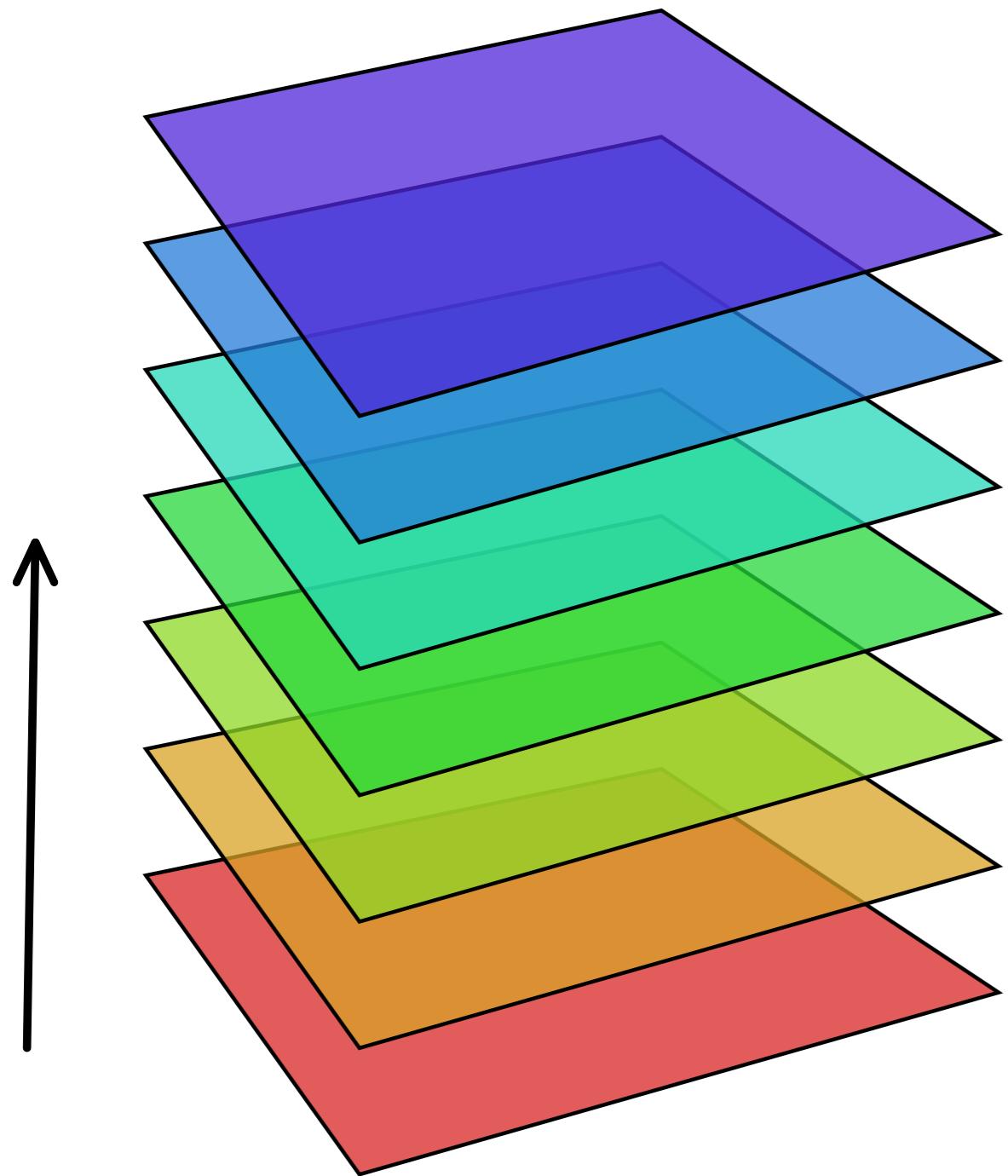
1. a statement in the form of a declaration.

instead: common framework



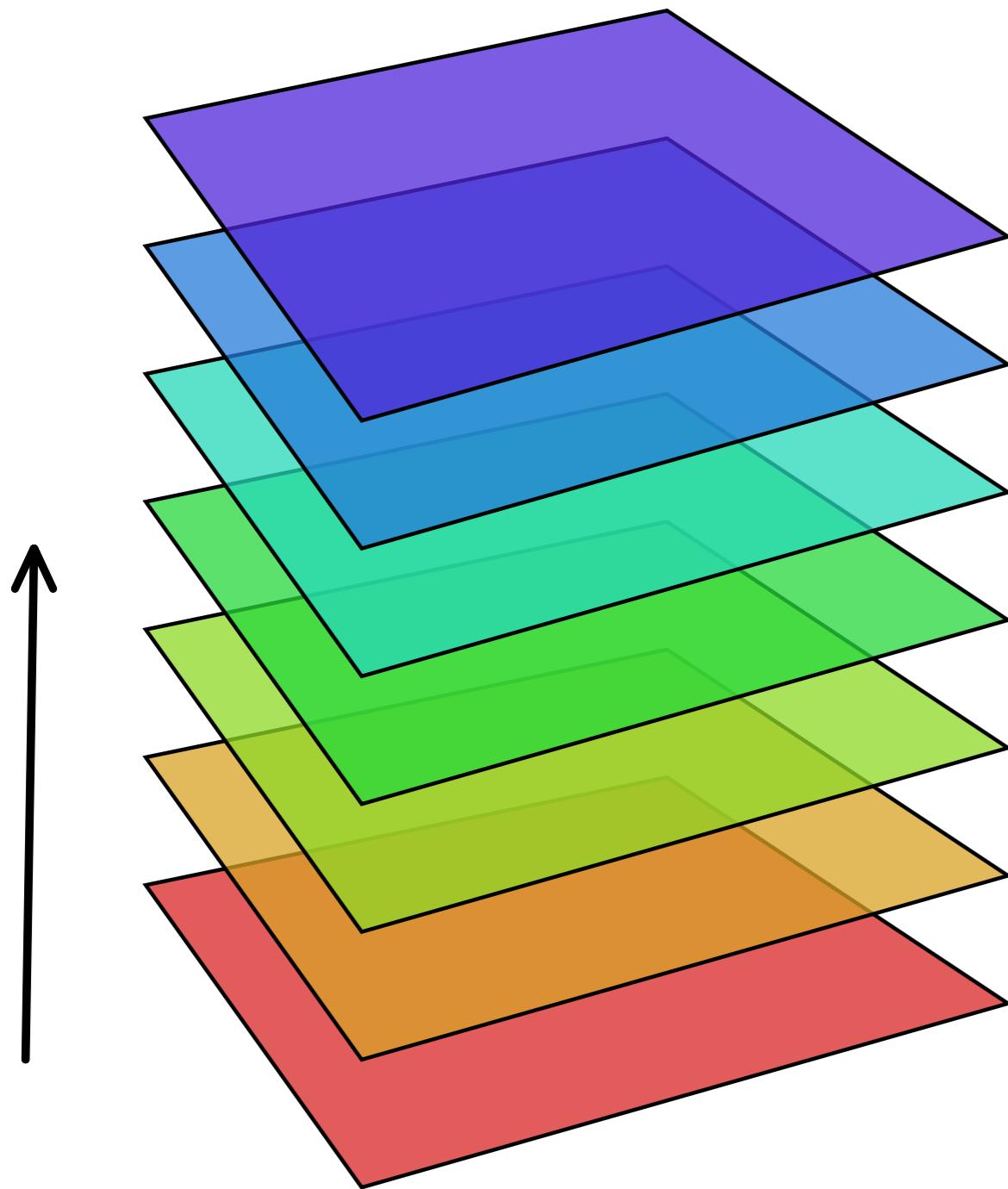
geometry
aesthetics
data

instead: common framework



coordinates
stats
facets
scales
geometry
aesthetics
data

a grammar instead: ~~common framework~~



coordinates
stats
facets
scales
geometry
aesthetics
data

an example

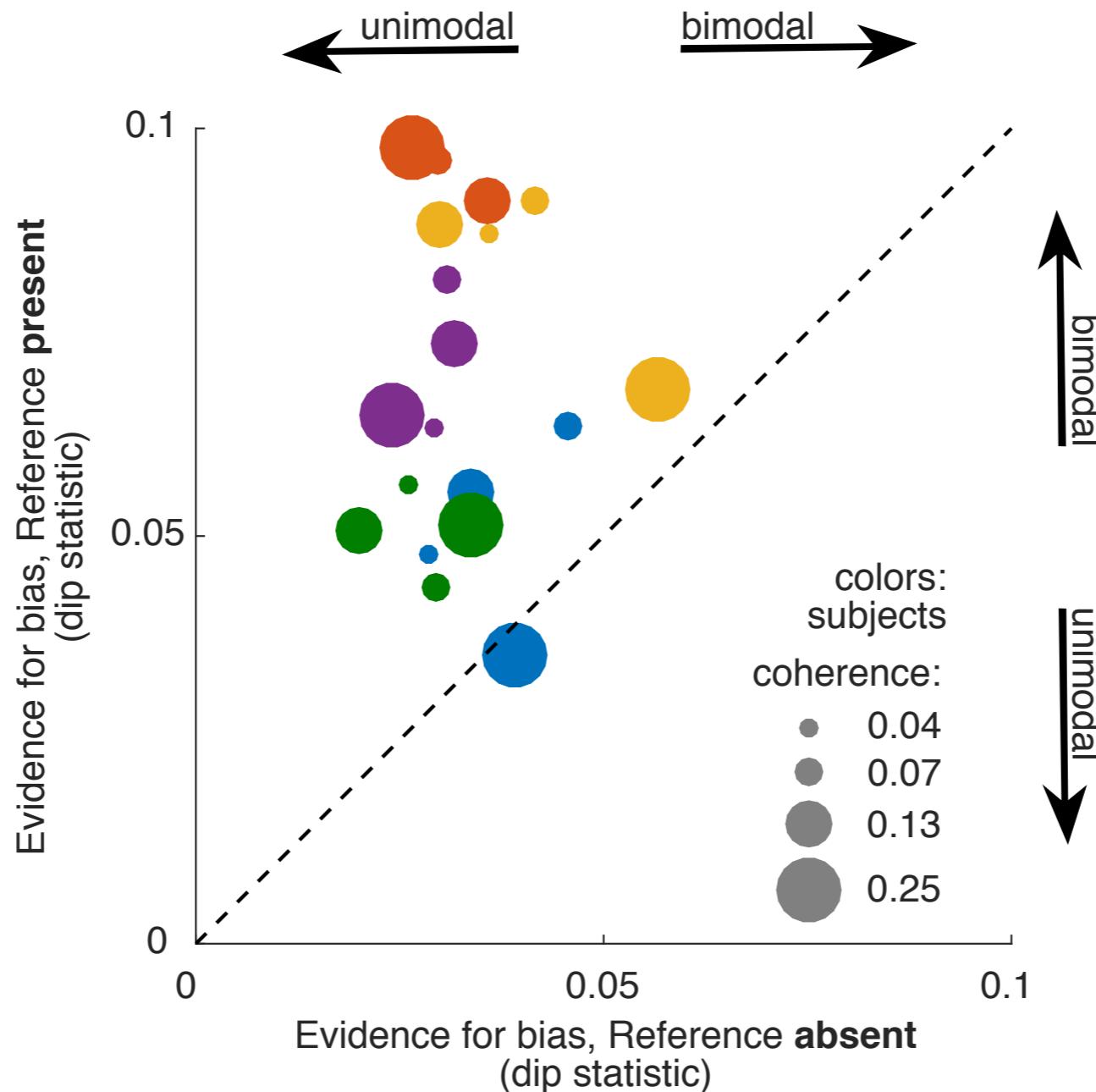


Figure 3b

Zamboni et al (2016) Proc Biol Sci.

recreating this graphic
in R/ggplot2

code:

[https://gist.github.com/schluppeck/
9a54b9b7a37793d8959779629b4cd2fc](https://gist.github.com/schluppeck/9a54b9b7a37793d8959779629b4cd2fc)

data, d

head(d)

	X	subject	coherence	absent	present
1	1	1		4 0.035996	0.087079
2	2	2		7 0.026042	0.056272
3	3	3		13 0.028604	0.047791
4	4	4		25 0.029221	0.063149
5	5	5		4 0.025157	0.096832
6	6	1		7 0.041558	0.091160

4 variables that
we want to **map**
into a plot

aesthetics

- x, y (position)
- alpha, color, fill
- size
- shape
- linetype

data

head(d)

	X	subject	coherence	absent	present
1	1	1		4 0.035996	0.087079
2	2	2		7 0.026042	0.056272
3	3	3		13 0.028604	0.047791
4	4	4		25 0.029221	0.063149
5	5	5		4 0.025157	0.096832
6	6	1		7 0.041558	0.091160

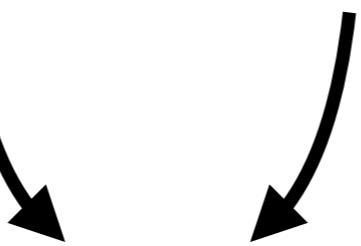
4 variables that
we want to **map**
into a plot

data

`head(d)`

	X	subject	coherence	absent	present
1	1	1	4	0.035996	0.087079
2	2	2	7	0.026042	0.056272
3	3	3	13	0.028604	0.047791
4	4	4	25	0.029221	0.063149
5	5	5	4	0.025157	0.096832
6	6	1	7	0.041558	0.091160

4 variables that
we want to **map**
into a plot



x, y (position)

data

`head(d)`

	X	subject	coherence	absent	present
1	1	1	4	0.035996	0.087079
2	2	2	7	0.026042	0.056272
3	3	3	13	0.028604	0.047791
4	4	4	25	0.029221	0.063149
5	5	5	4	0.025157	0.096832
6	6	1	7	0.041558	0.091160

4 variables that
we want to **map**
into a plot

`size`

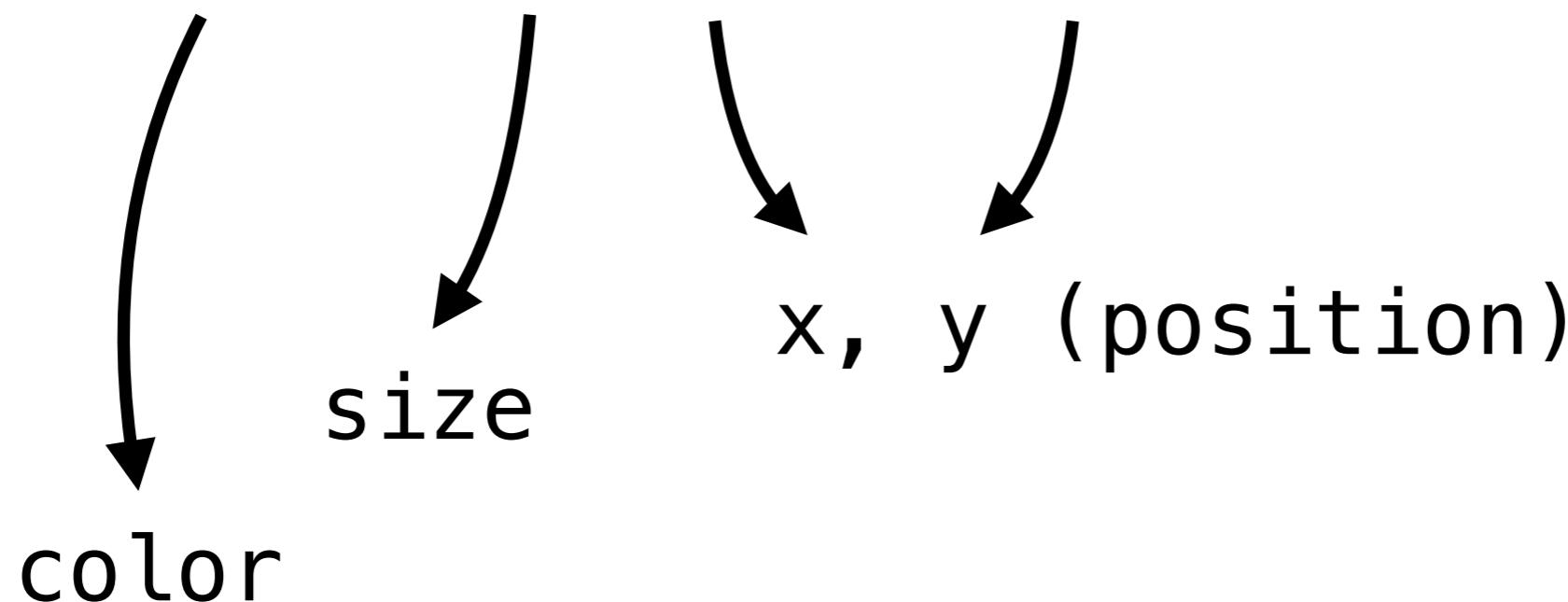
`x, y (position)`

data

`head(d)`

	X	subject	coherence	absent	present
1	1	1	4	0.035996	0.087079
2	2	2	7	0.026042	0.056272
3	3	3	13	0.028604	0.047791
4	4	4	25	0.029221	0.063149
5	5	5	4	0.025157	0.096832
6	6	1	7	0.041558	0.091160

4 variables that
we want to **map**
into a plot



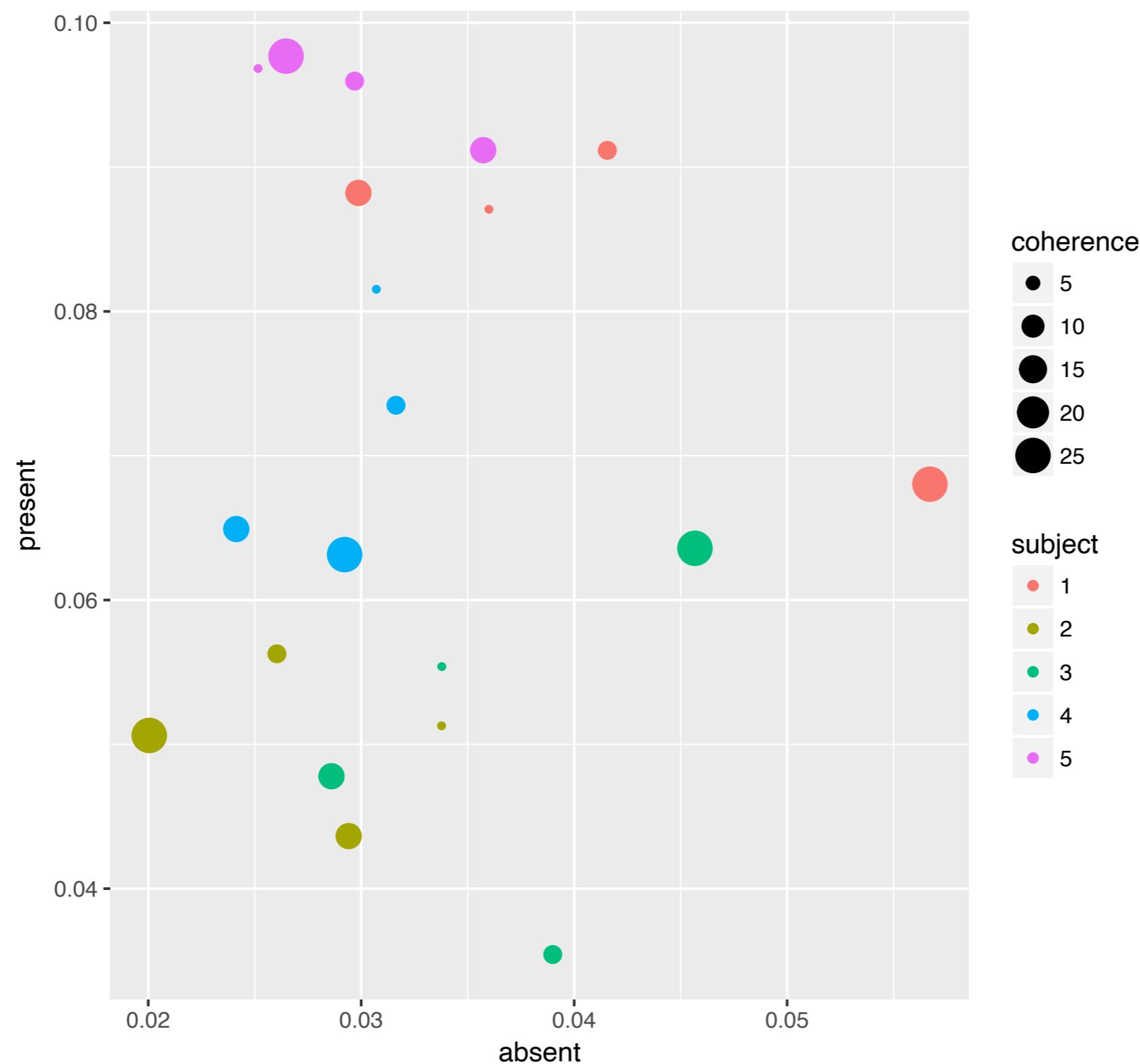
geometry

- 0d: points, text
- 1d: lines, paths
- 2d: polygons, intervals

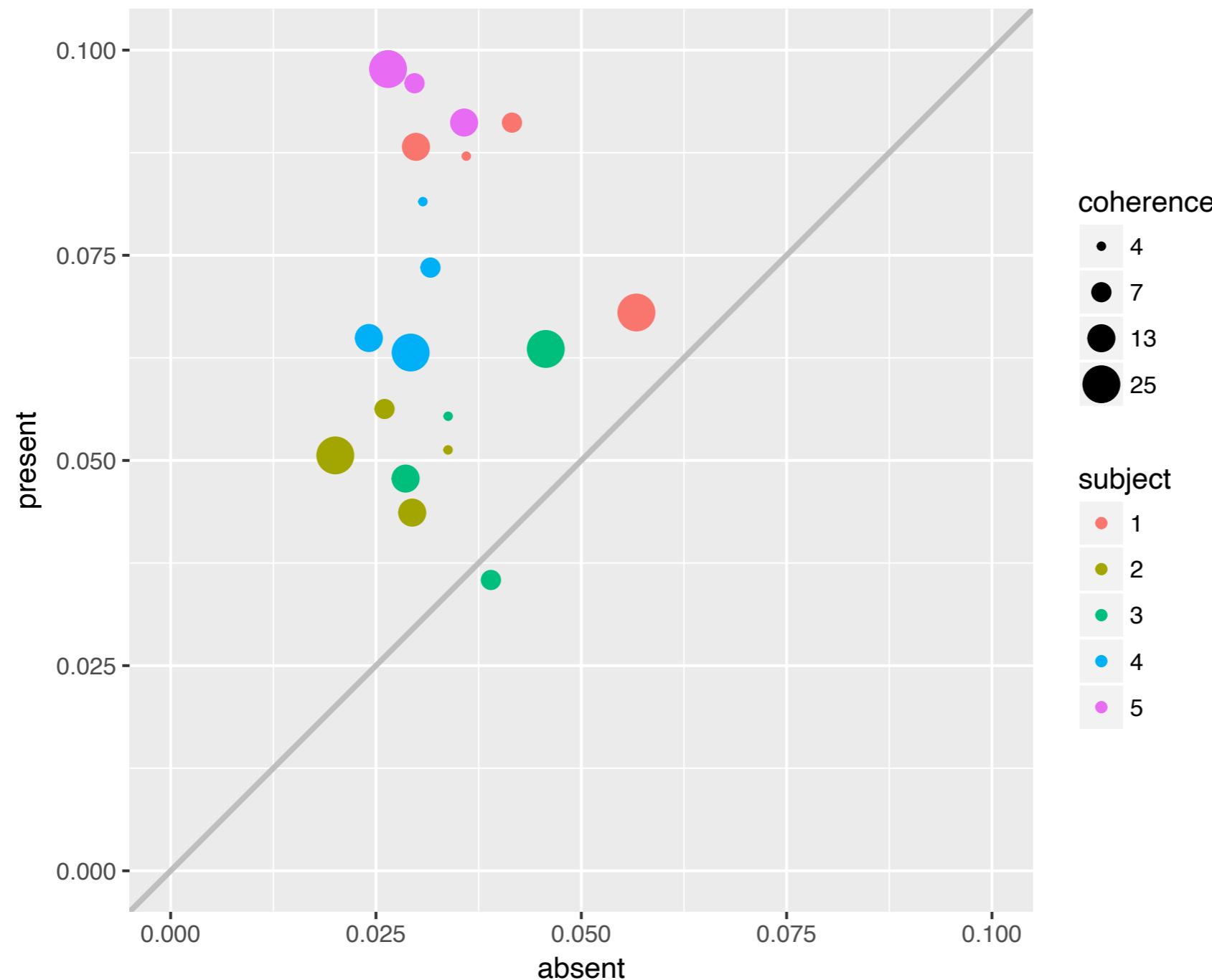
geometry

- 0d: points, text
- 1d: lines, paths
- 2d: polygons, intervals

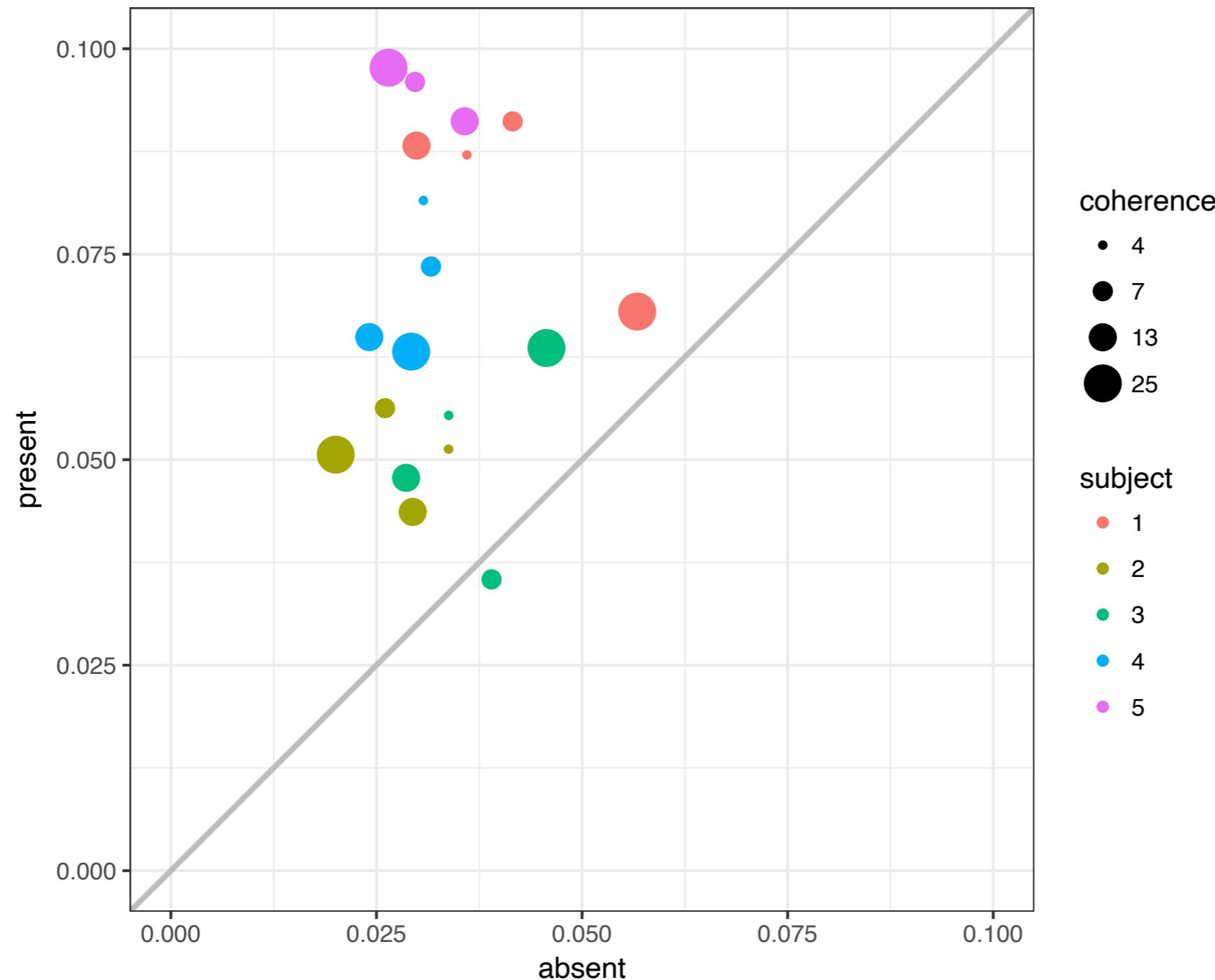




with default settings

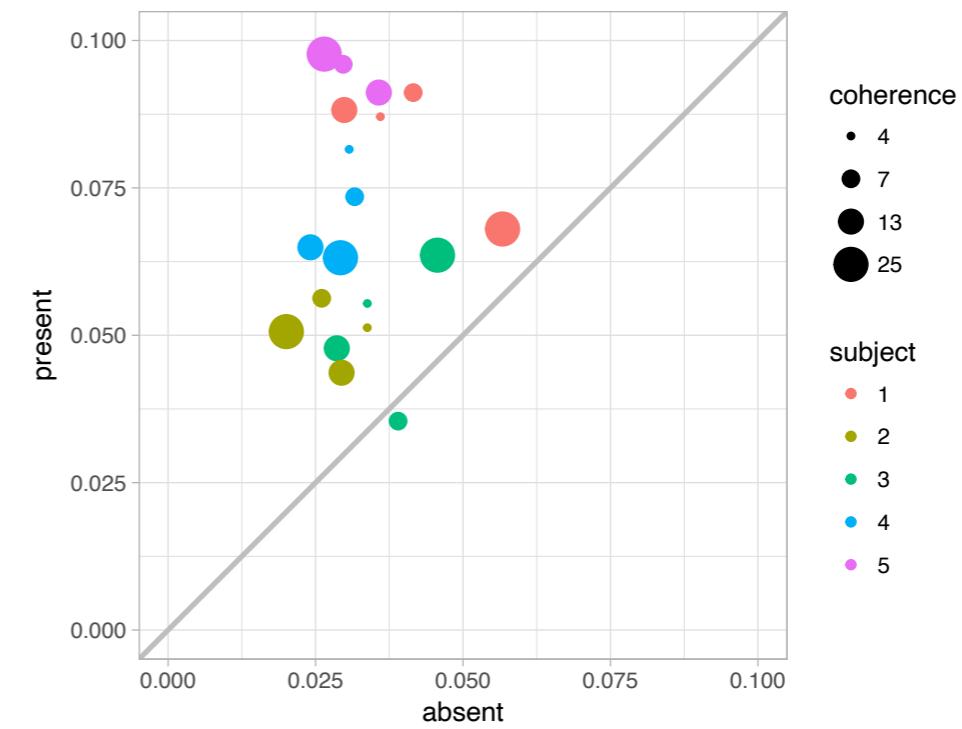
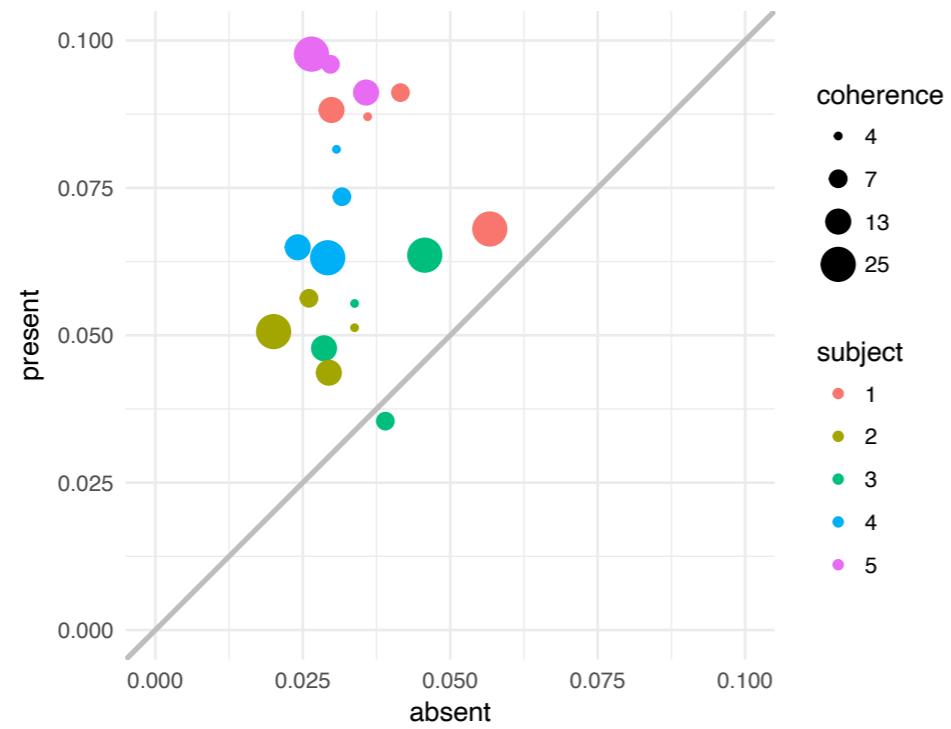
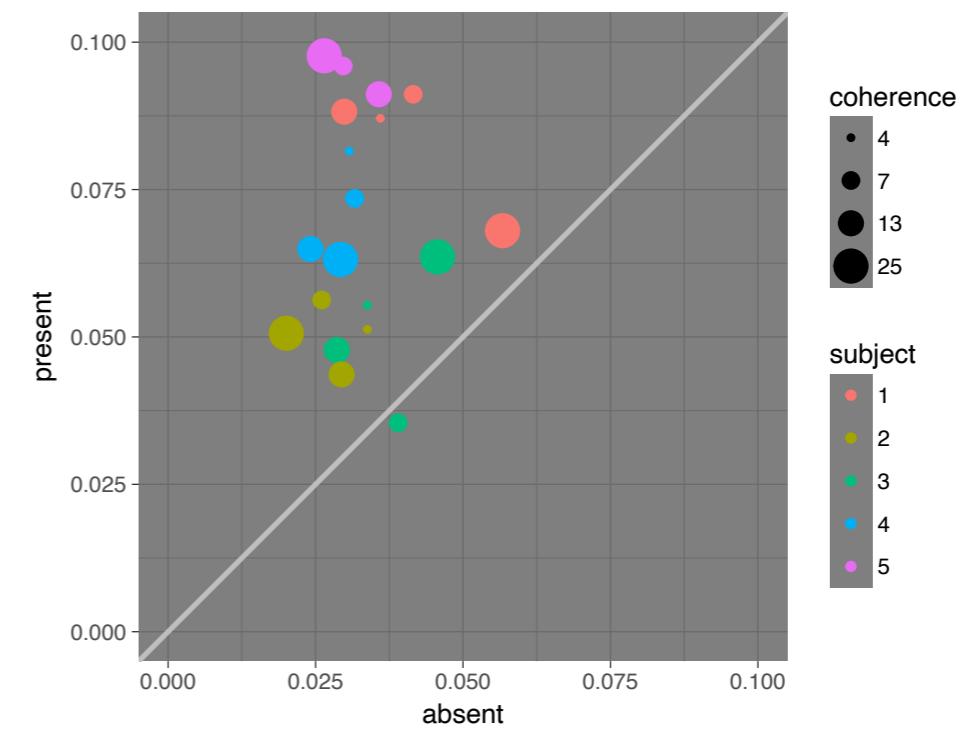
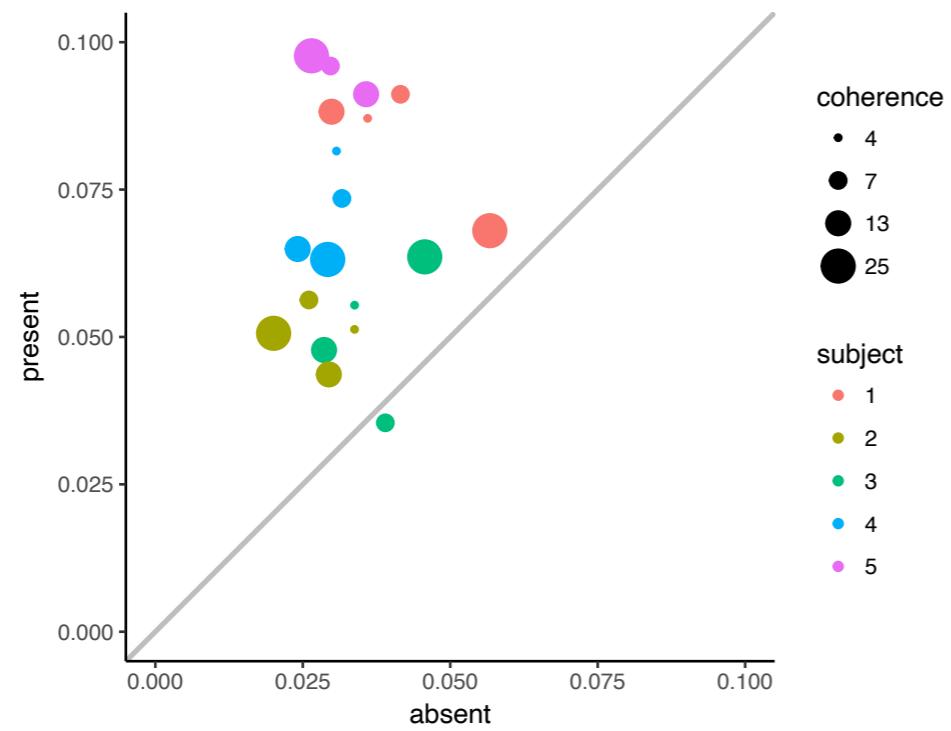


+ scale, coord (aspect ratio), unity line



+ scale, coord (aspect ratio), unity line

themes / look of plot



worth the hassle?

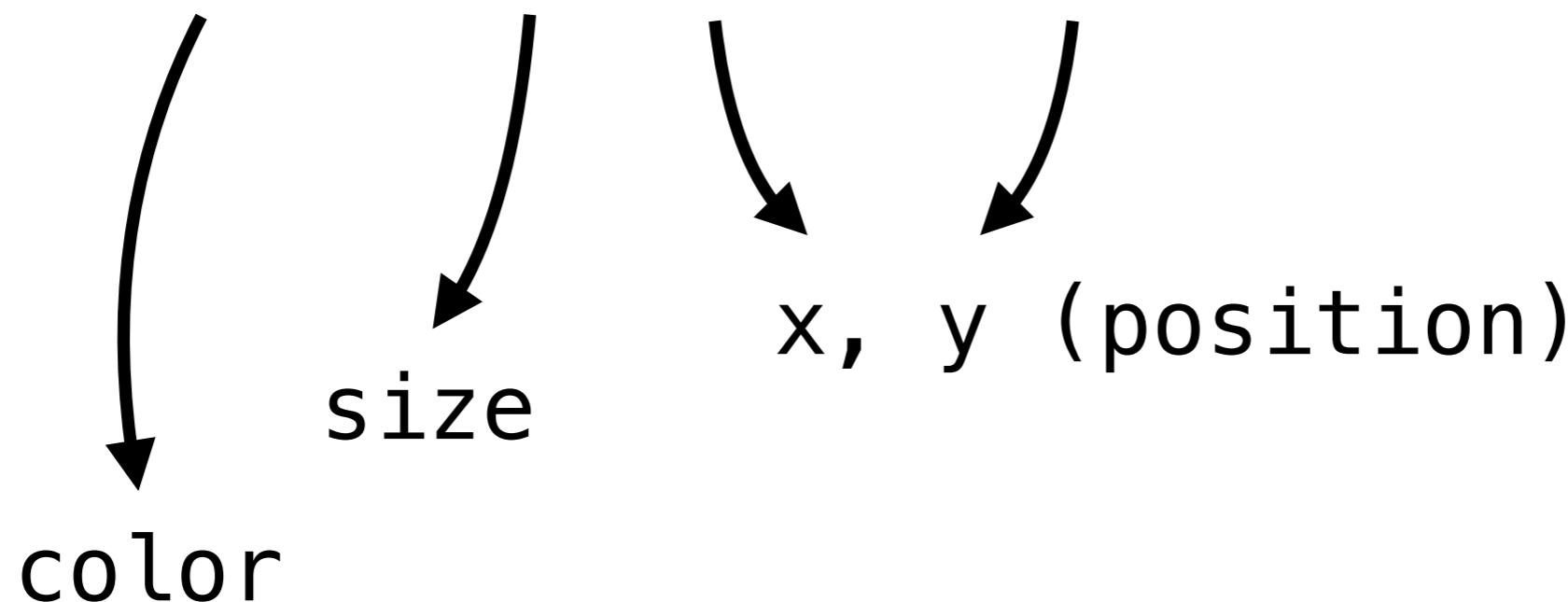
- I think yes: already for basic plotting
- for data exploration we often slice across different dimensions:
 - subjects, regions of interest, ...
 - measures: RT, % correct, fMRI response amplitude, ...

data

`head(d)`

	X	subject	coherence	absent	present
1	1	1	4	0.035996	0.087079
2	2	2	7	0.026042	0.056272
3	3	3	13	0.028604	0.047791
4	4	4	25	0.029221	0.063149
5	5	5	4	0.025157	0.096832
6	6	1	7	0.041558	0.091160

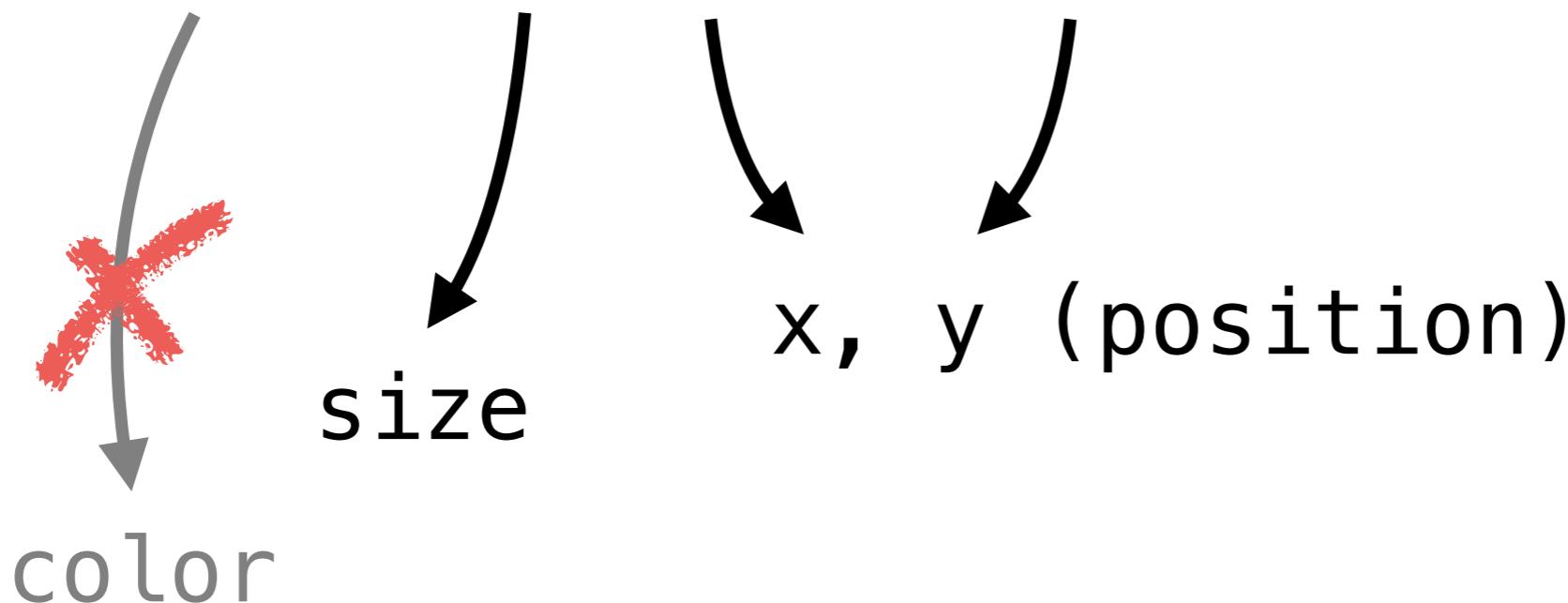
4 variables that
we want to **map**
into a plot



data

`head(d)`

	X	subject	coherence	absent	present
1	1	1	4	0.035996	0.087079
2	2	2	7	0.026042	0.056272
3	3	3	13	0.028604	0.047791
4	4	4	25	0.029221	0.063149
5	5	5	4	0.025157	0.096832
6	6	1	7	0.041558	0.091160

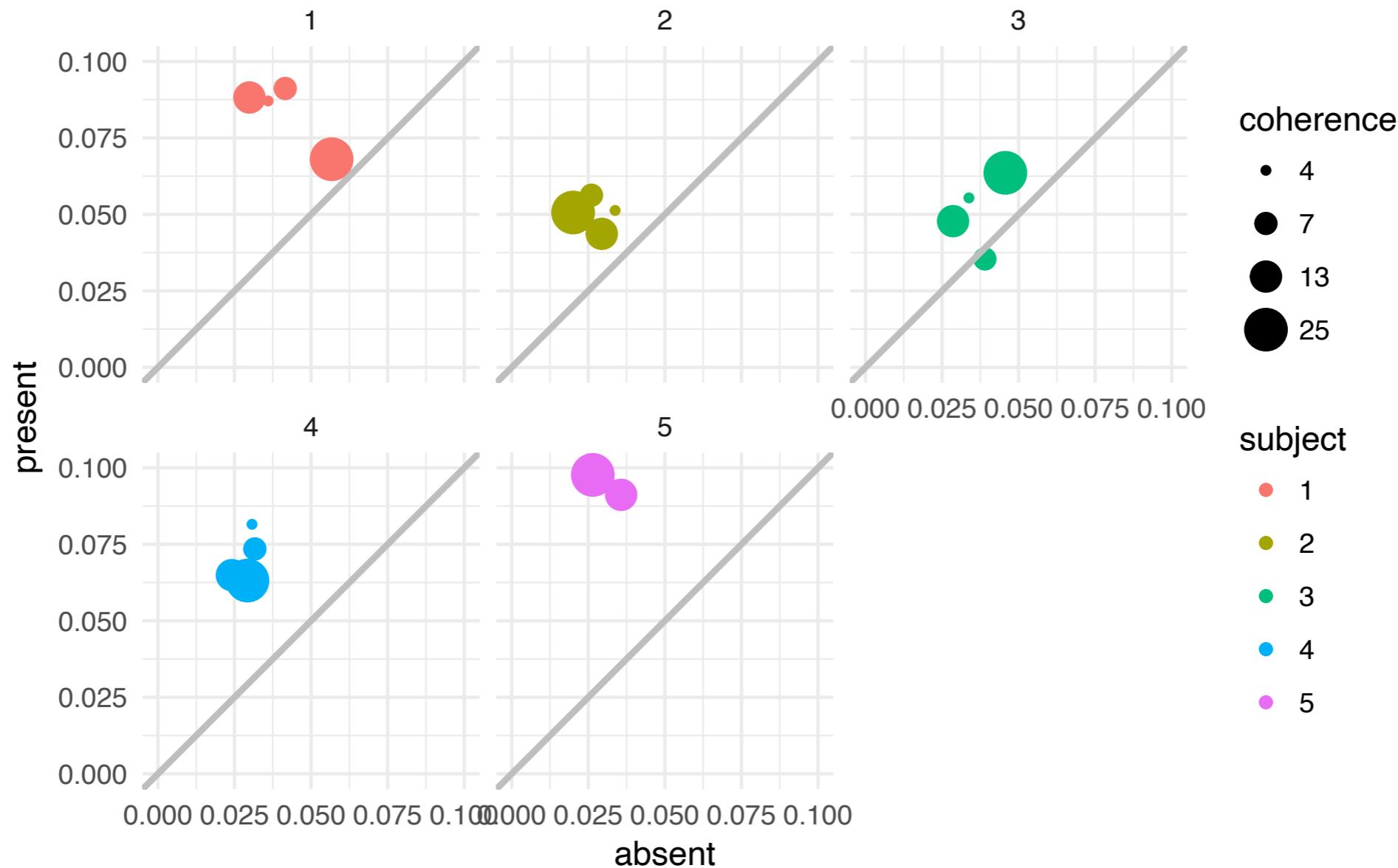


3

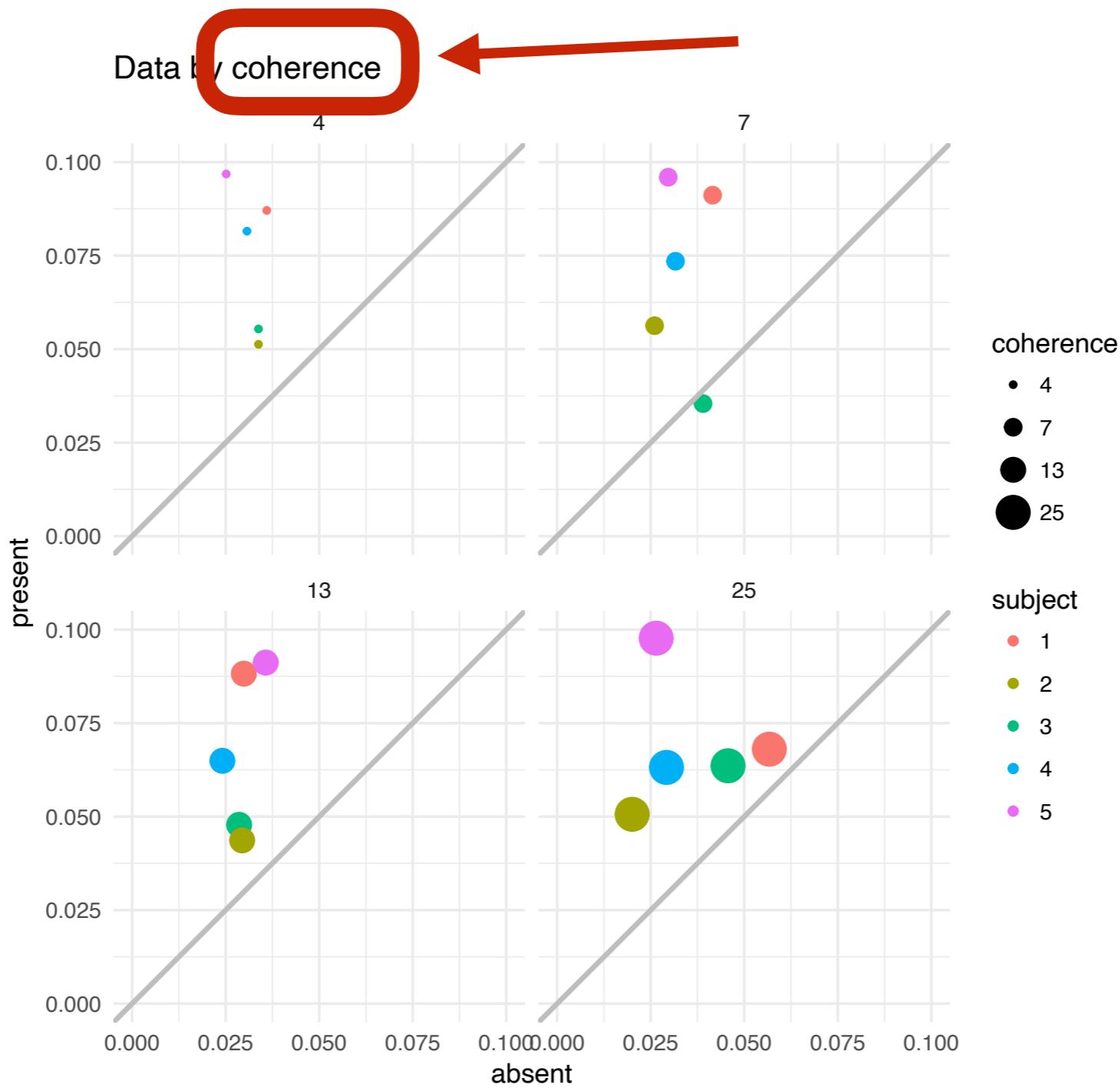
~~+variables that we want to **map** into a plot~~

facet (lattice)

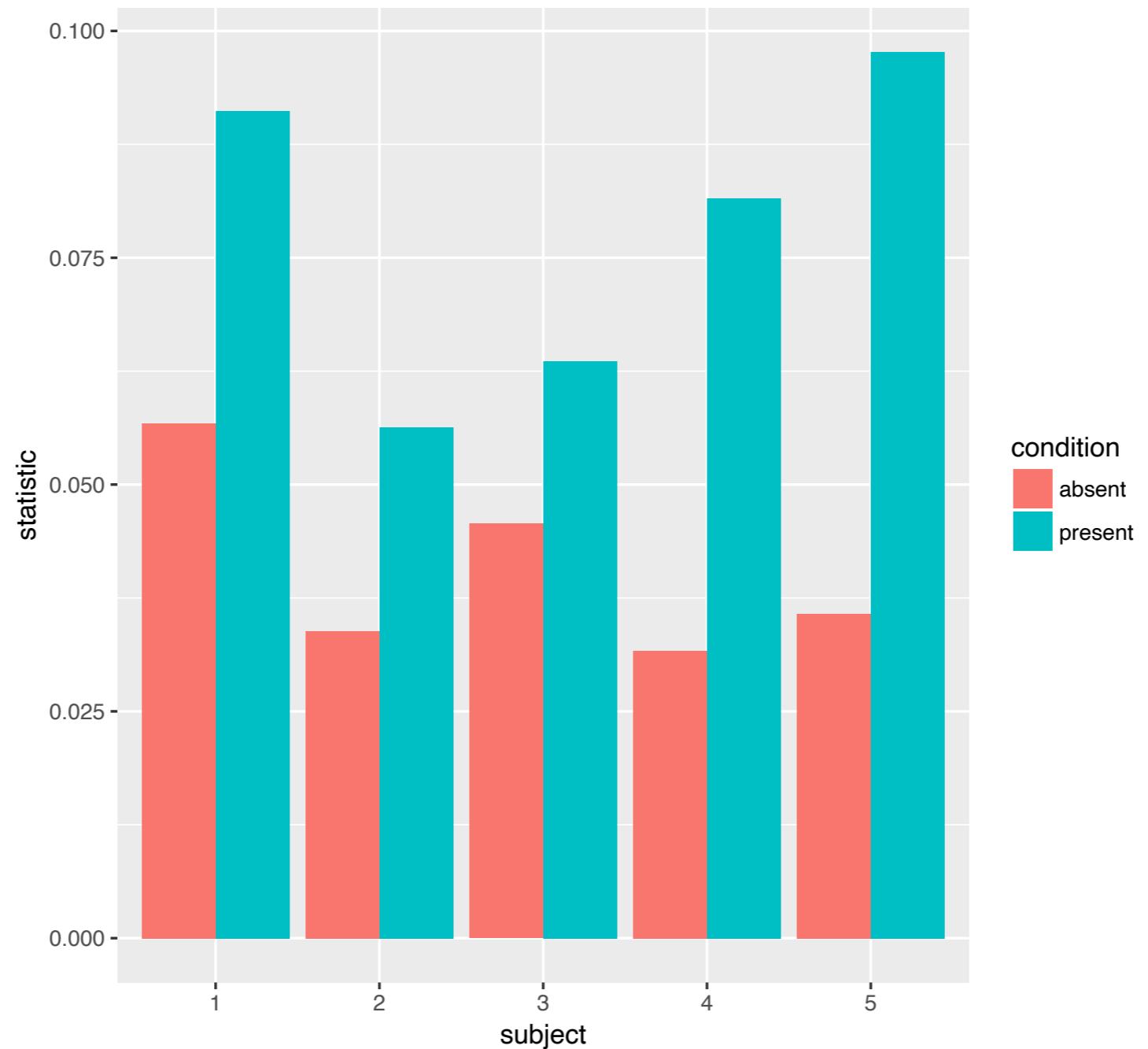
Data for different subjects



facet (lattice)



rearrange

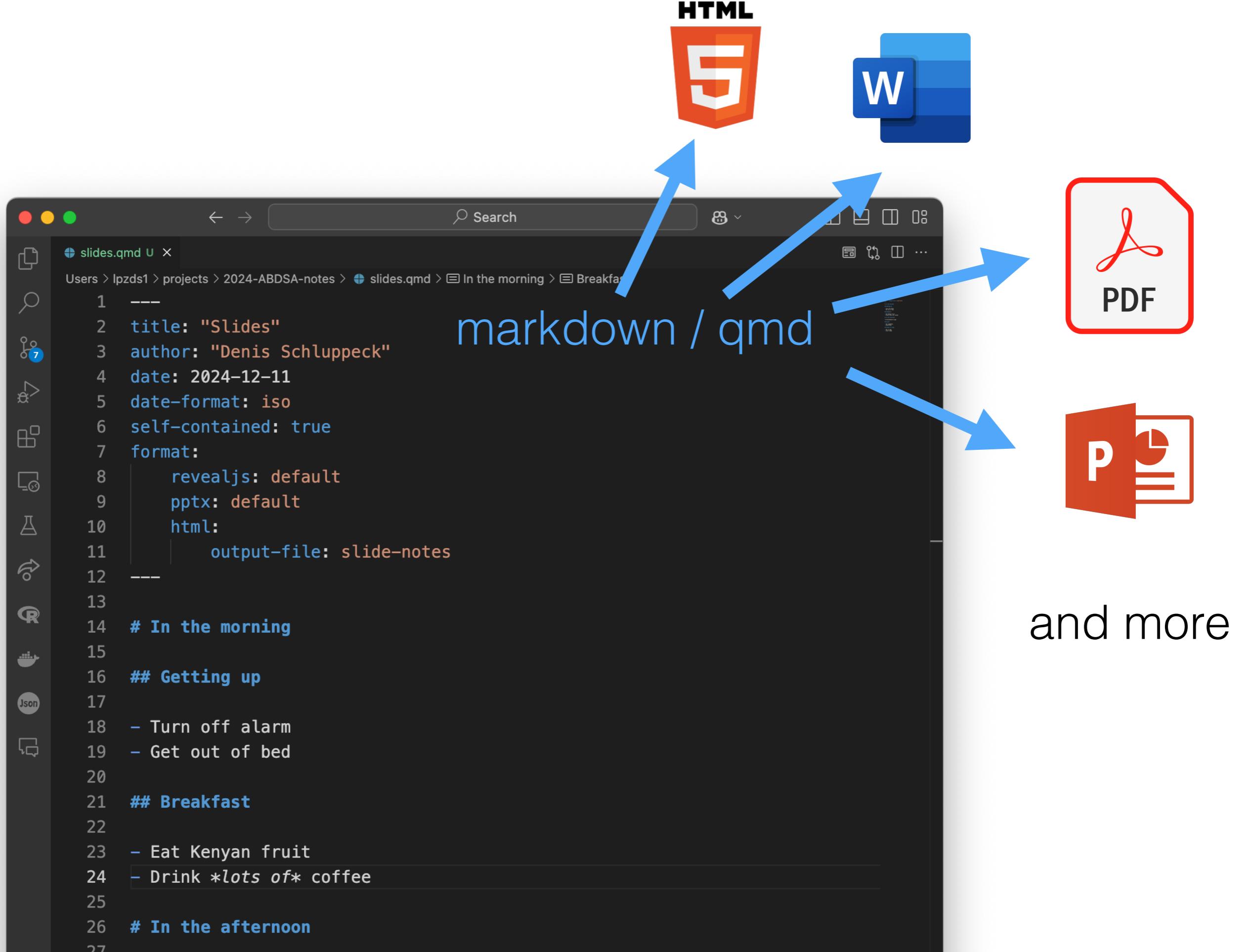


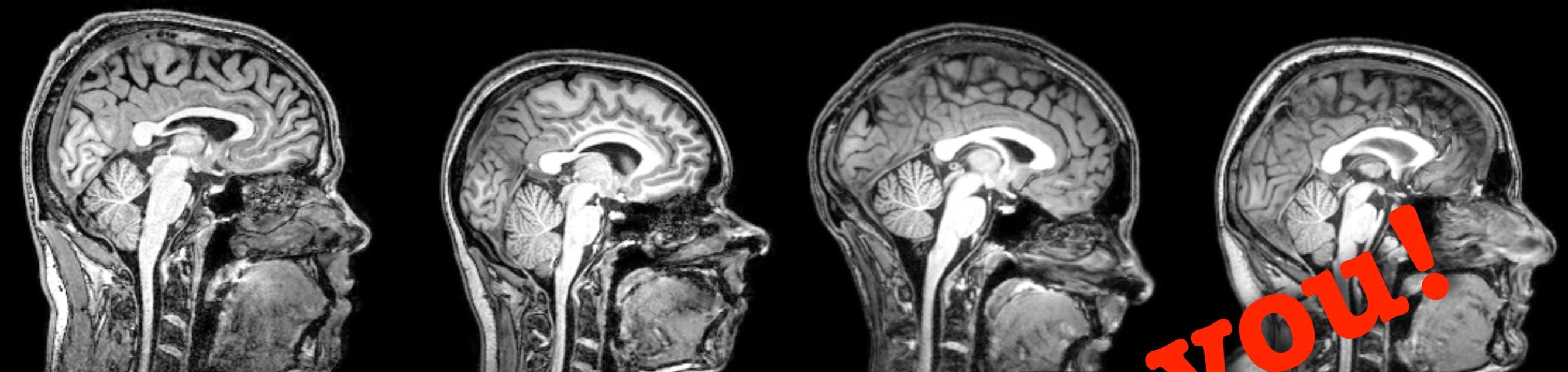
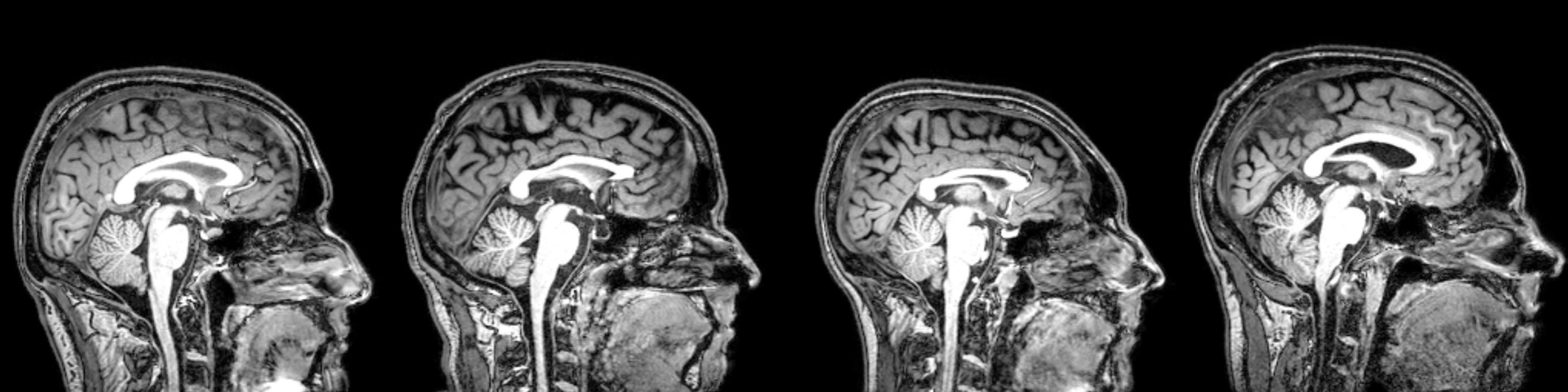
Examples

[https://github.com/
schluppeck/2024-
ABDSA-notes](https://github.com/schluppeck/2024-ABDSA-notes)

demo: editing on github website

<https://quarto.org/>





Thank you!