

# That's not [data] science!

Mara Averick ([@dataandme](https://twitter.com/dataandme))

Tidyverse Developer Advocate, RStudio

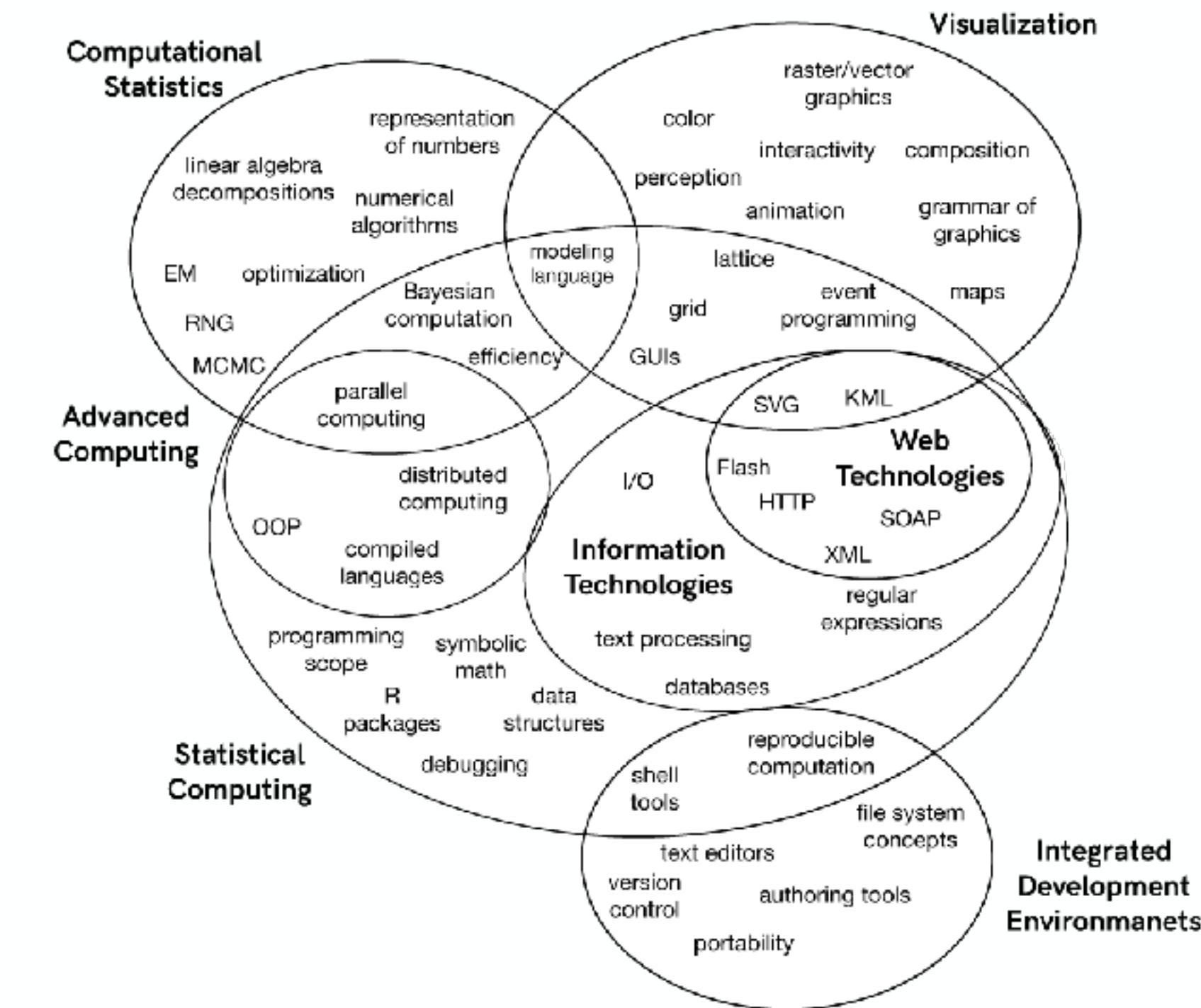
<http://bit.ly/not-science>





# Mara Averick

TIDYVERSE DEV ADVOCATE, RSTUDIO



Not a Real Data Scientist™



# Mara Averick

TIDYVERSE DEV ADVOCATE, RSTUDIO

Not a Real Data Scientist™



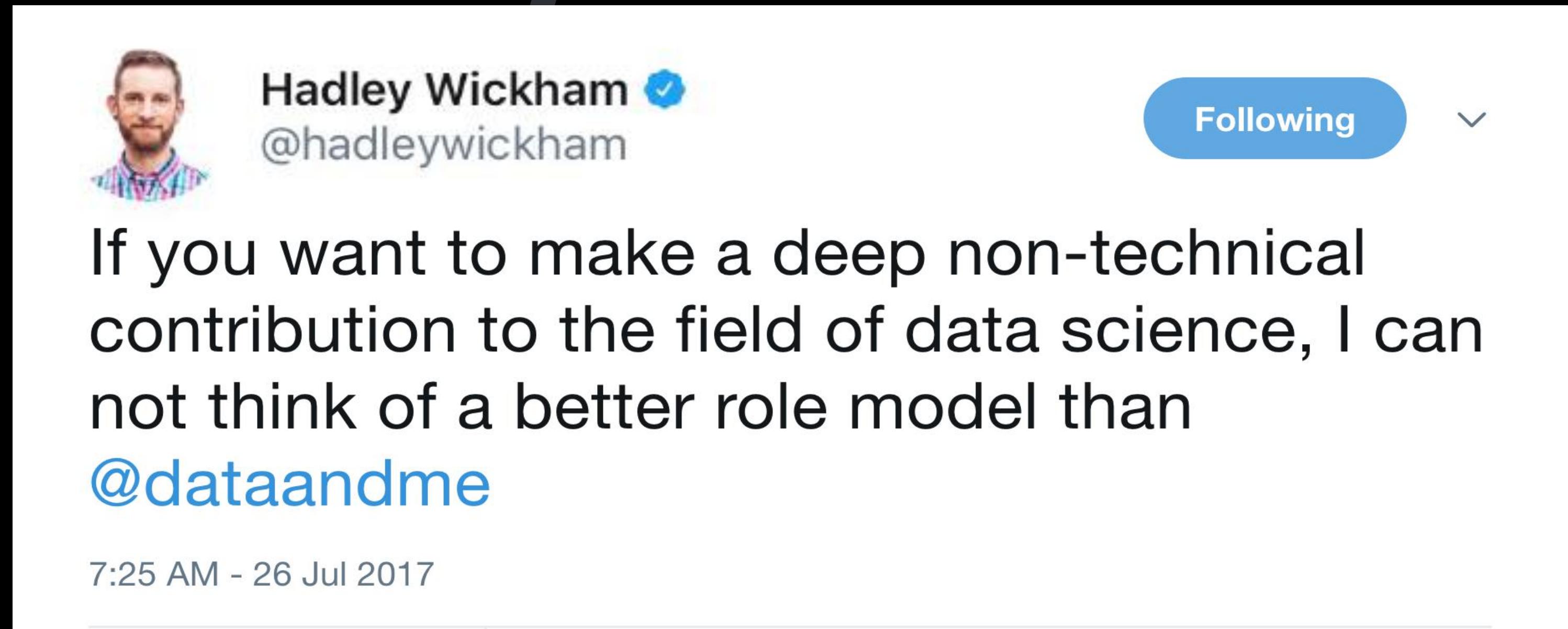


# Mara Averick

TIDYVERSE DEV ADVOCATE, RSTUDIO



# but...



Hadley Wickham   
@hadleywickham

Following ▾

If you want to make a deep non-technical contribution to the field of data science, I can not think of a better role model than  
[@dataandme](#)

7:25 AM - 26 Jul 2017

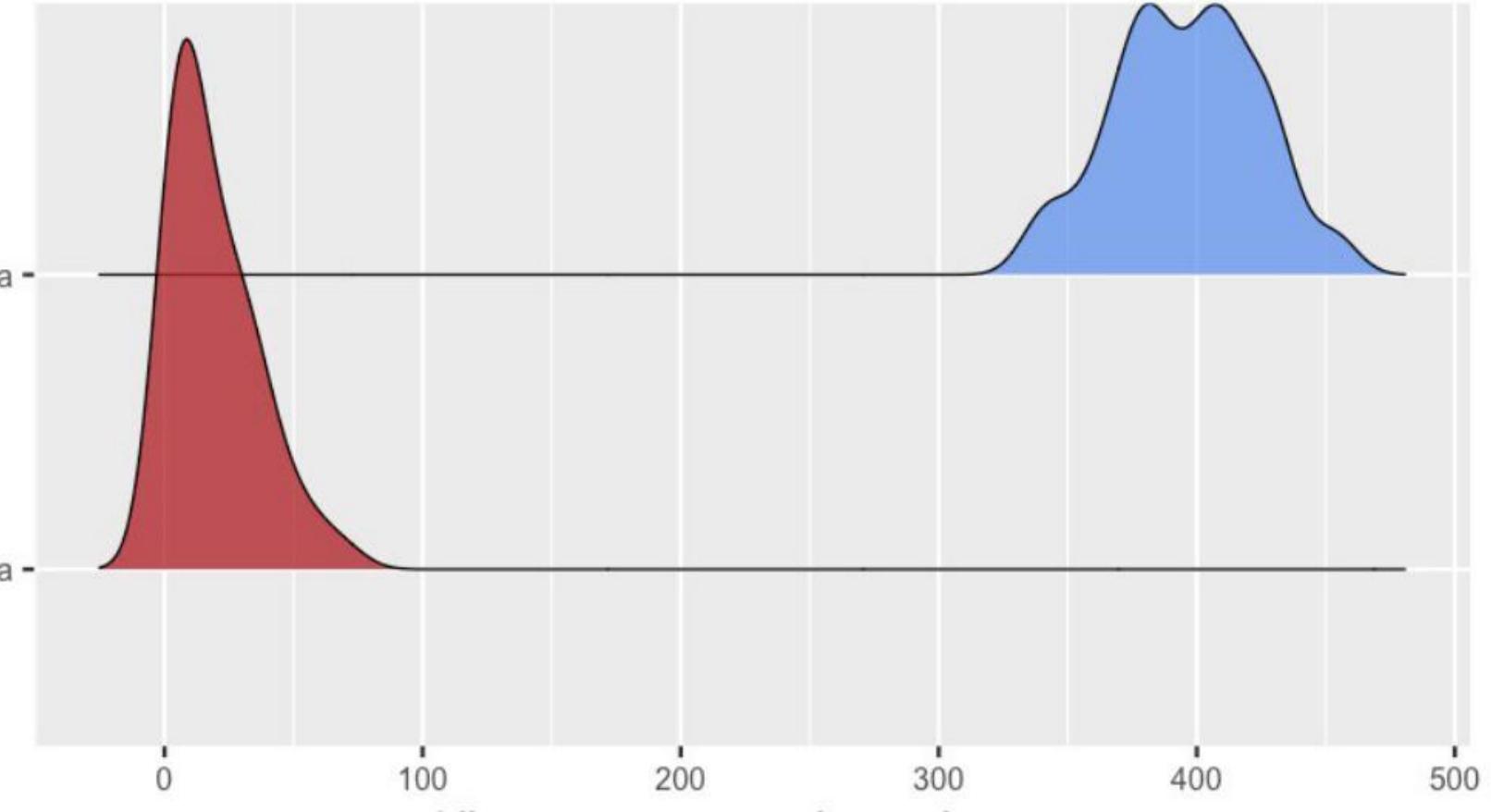
# but...

 **Edwin Thoen**  
@edwin\_thoen

Following ▾

After 9 months as an R blogger I discovered  
the key to writing a succesfull blog: please  
[@dataandme!](#)

What determines a succesful blog post?



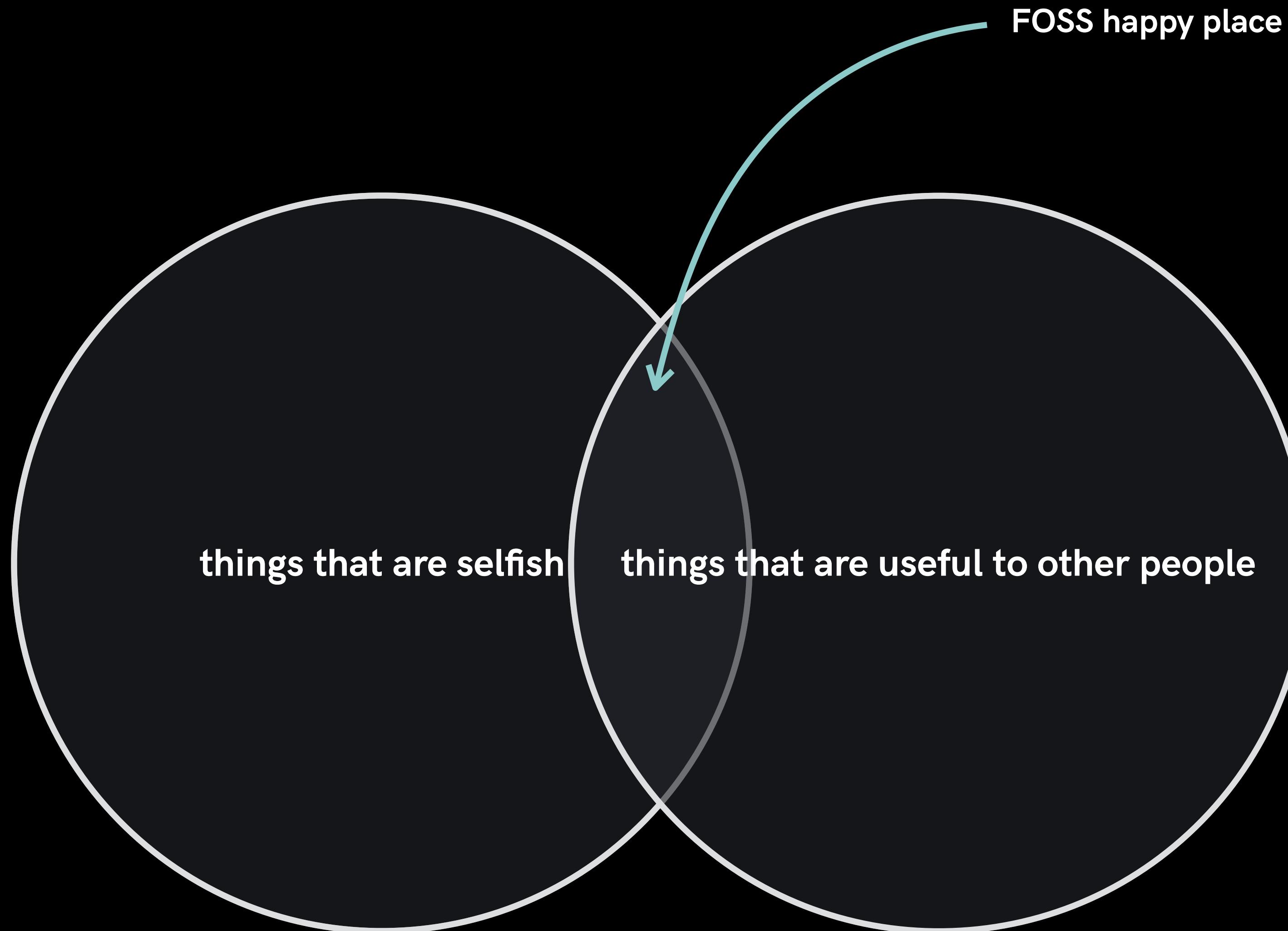
Tweeted about by Mara -

Not tweeted about by Mara -

Likes, retweets, pageviews, whatever

5:36 AM - 27 Aug 2017

# but...

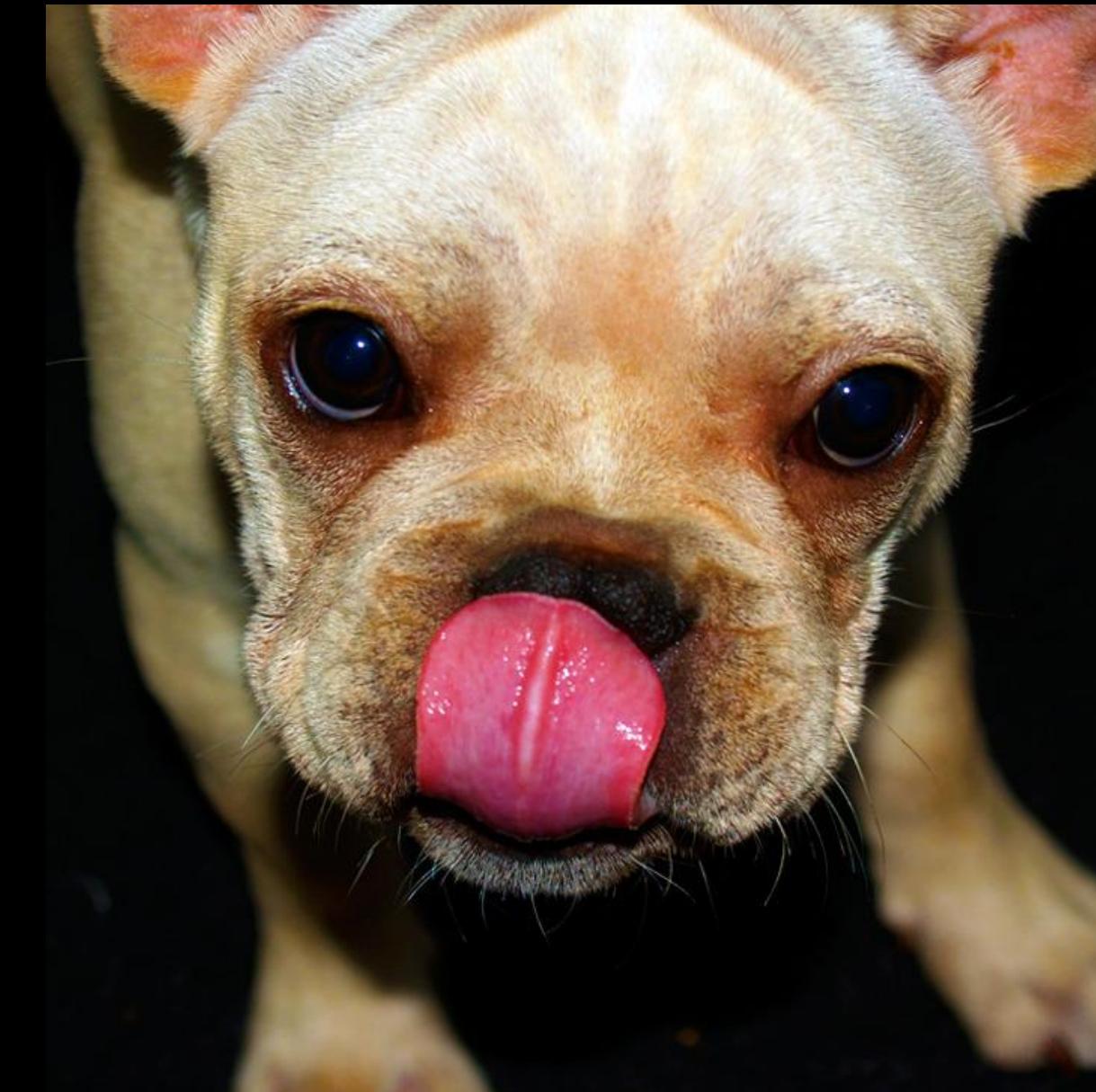


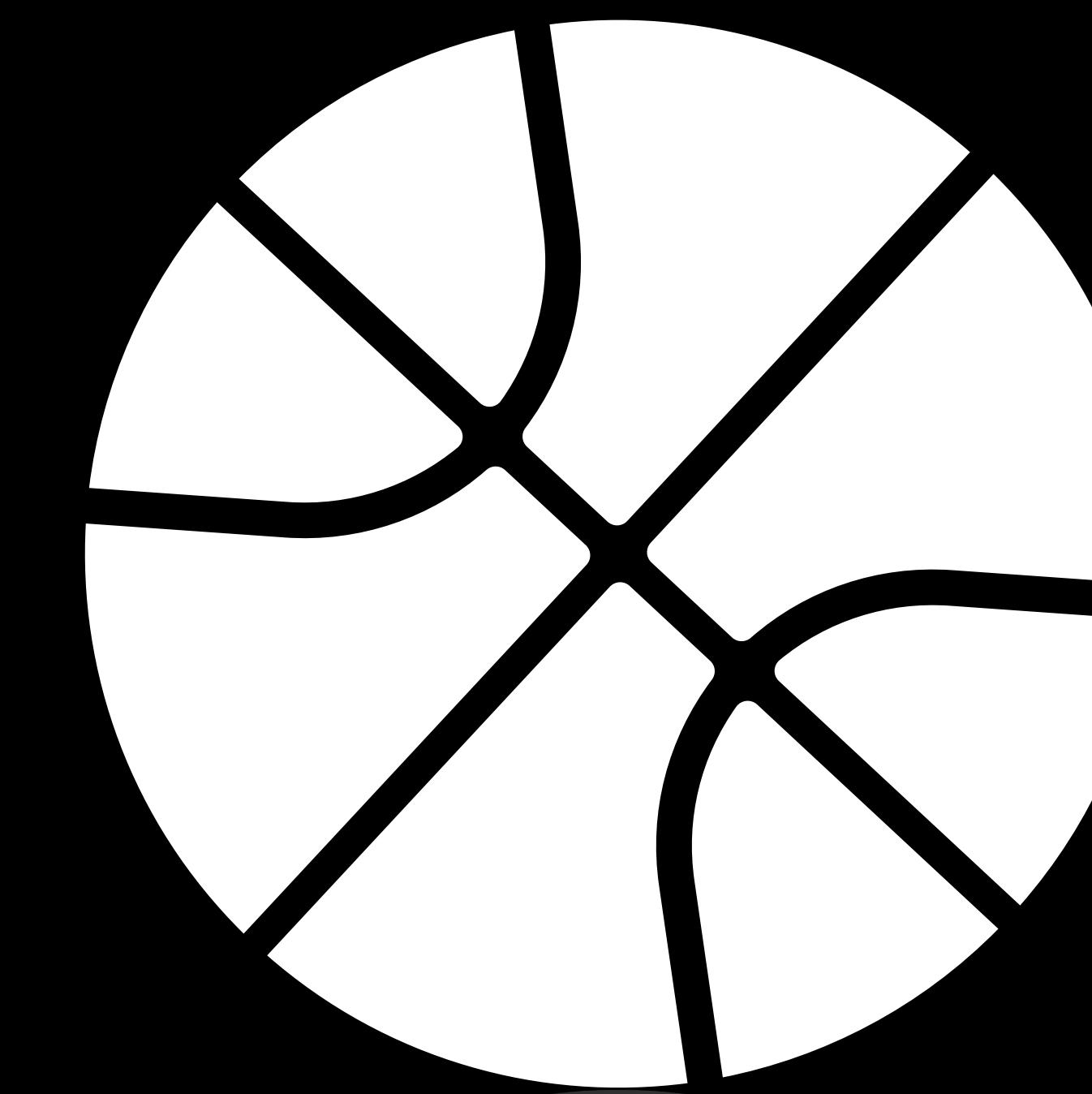
# origins...

✓ Basketball

✓ The Batpig

✓ @buffer









**Dean Oliver**

@DeanO\_Lytics

Following

To make the case to draft Draymond, though, you want to be more EXplicit in your discussions. Saying "the model says he's good" not enough.

[twitter.com/KaiserLindeman ...](https://twitter.com/KaiserLindeman ...)

This Tweet is unavailable.

11:16 AM - 30 Jun 2017

[Follow](#)

If u have great players, you're going 2 win games. If u have bad players, you're going 2 lose games. What does analytics have 2 do with it?

5:44 PM - 23 Feb 2016

real, verified sportsball reporter

The image shows a tweet card on a dark background. At the top left is a blacked-out profile picture with a blue checkmark. To its right is a 'Follow' button with a blue outline and the word 'Follow' in blue. Next to it is a downward-pointing arrow icon. A light blue curved arrow points from the text 'real, verified sportsball reporter' at the top to the blue checkmark on the profile picture. The main text of the tweet is a quote in white font: 'You put your best analytic team out there and he rolls out his best, one of you will lose. So at best, analytics works just 50% of the time.' Below the quote is the timestamp '6:44 PM - 24 Feb 2016'.

You put your best analytic team out there  
and he rolls out his best, one of you will lose.  
So at best, analytics works just 50% of the  
time.

6:44 PM - 24 Feb 2016



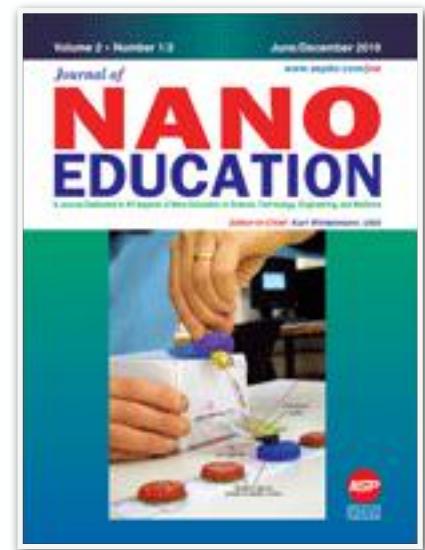
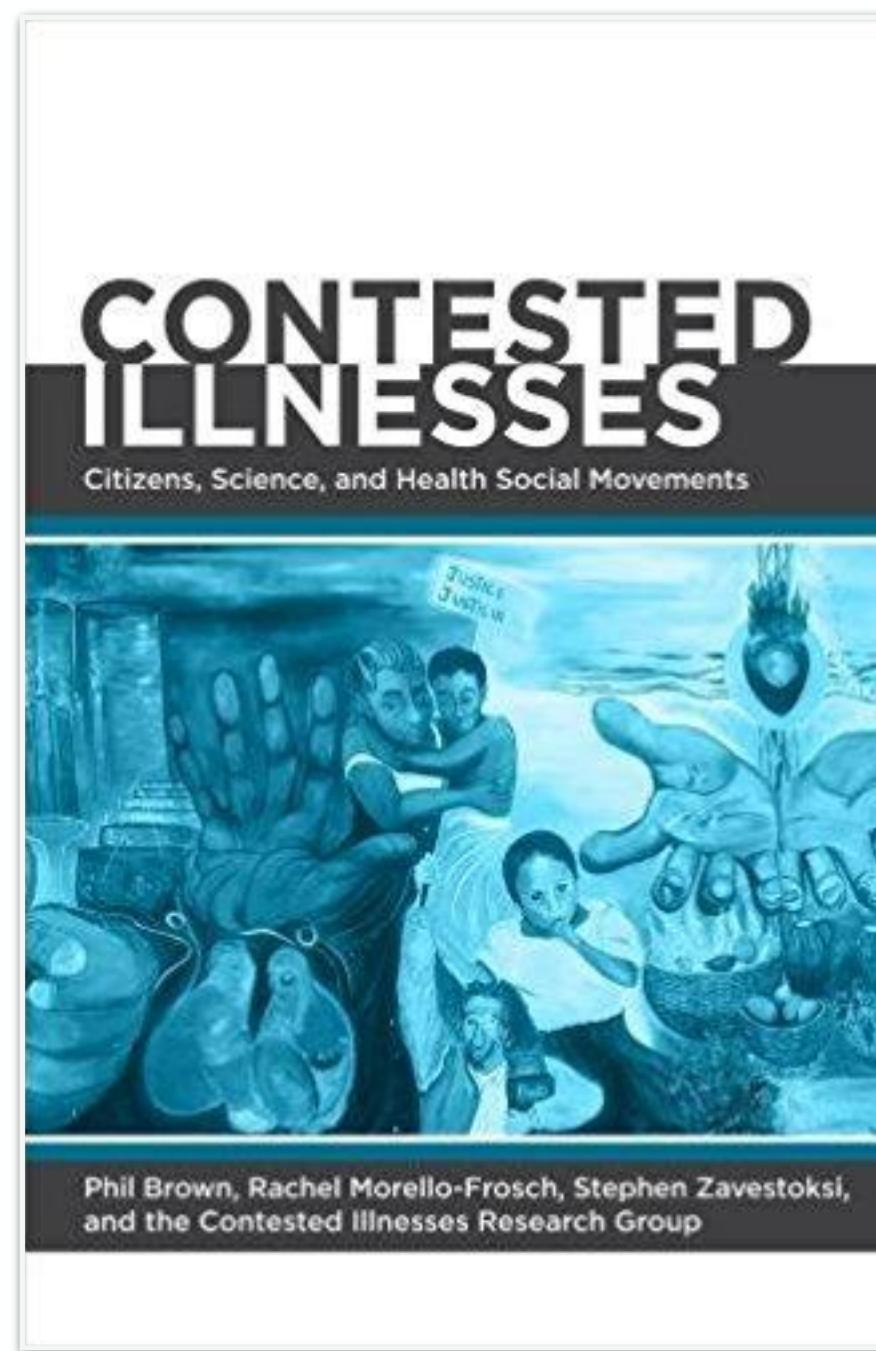
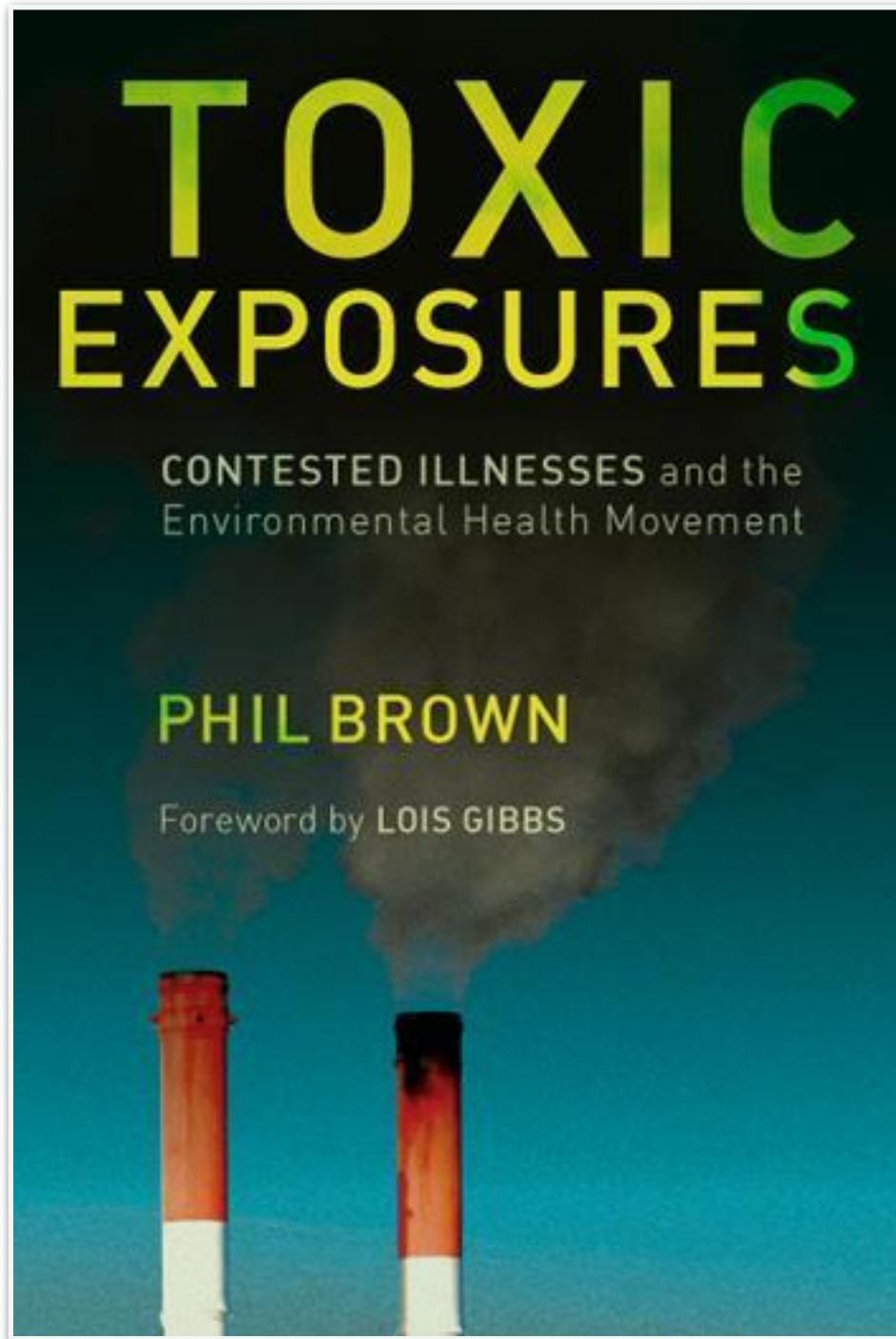
SCIENCE & SOCIETY

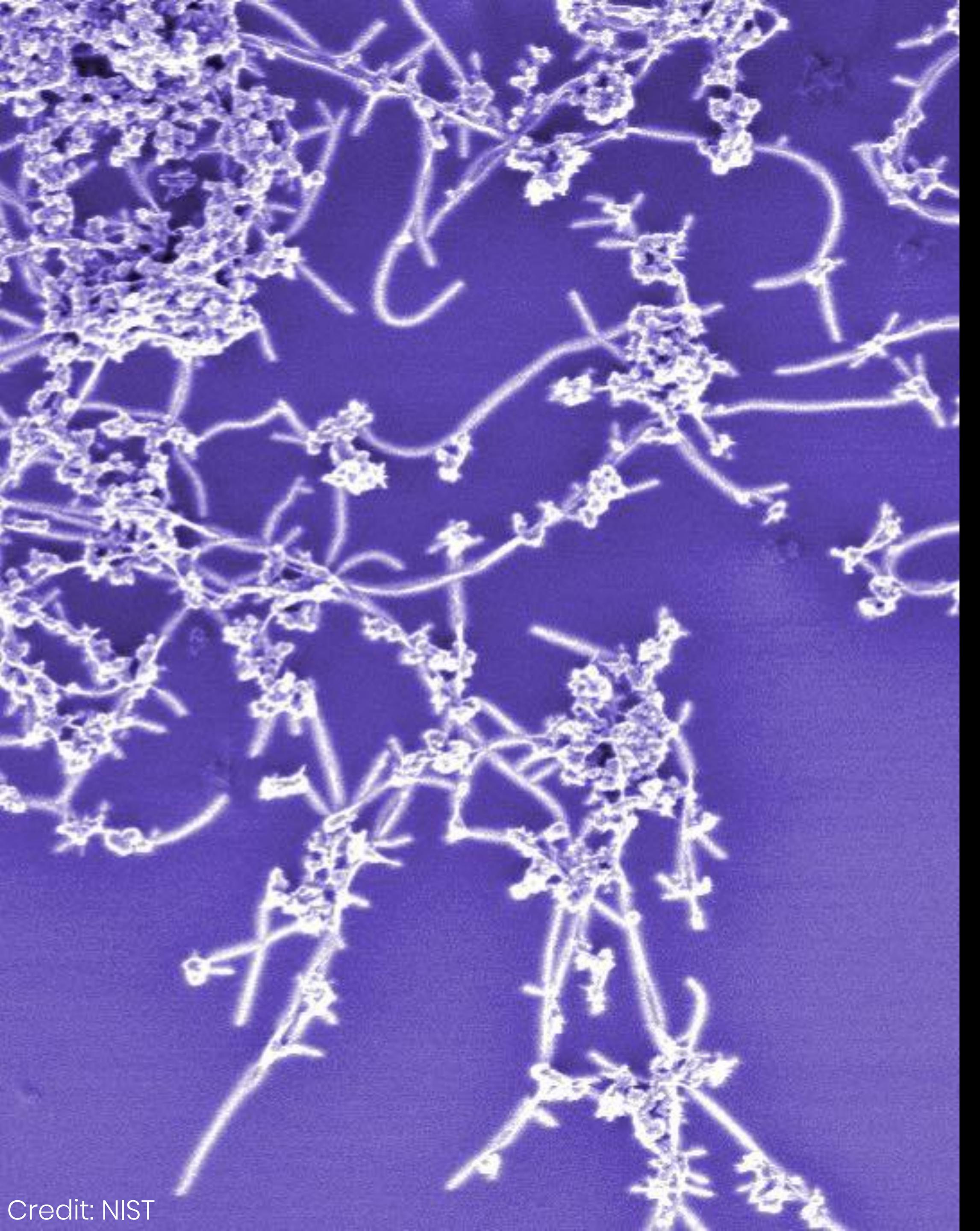
# Studying scientists...

“Like many social groups that do not reproduce themselves biologically, the experimental particle physics community renews itself by training novices.”

— Sharon Traweek, *Pilgrim's Progress: Male Tales Told During a Life in Physics*, 1988

# academic origins...





# Ethical, Legal & Social Implications of Nanotechnology

Nanotechnology Interdisciplinary Research Team



Hadley Wickham   
@hadleywickham

Follow

Replying to @juliasilge @arnicas @dataandme

I really didn't want to imply that Mara doesn't have awesome technical skills 😞 She does!

8:32 AM - 26 Jul 2017

8:32 AM - 26 Jul 2017



Lucy   
@LucyStats

Follow

Replying to @hadleywickham @juliasilge and 2 others

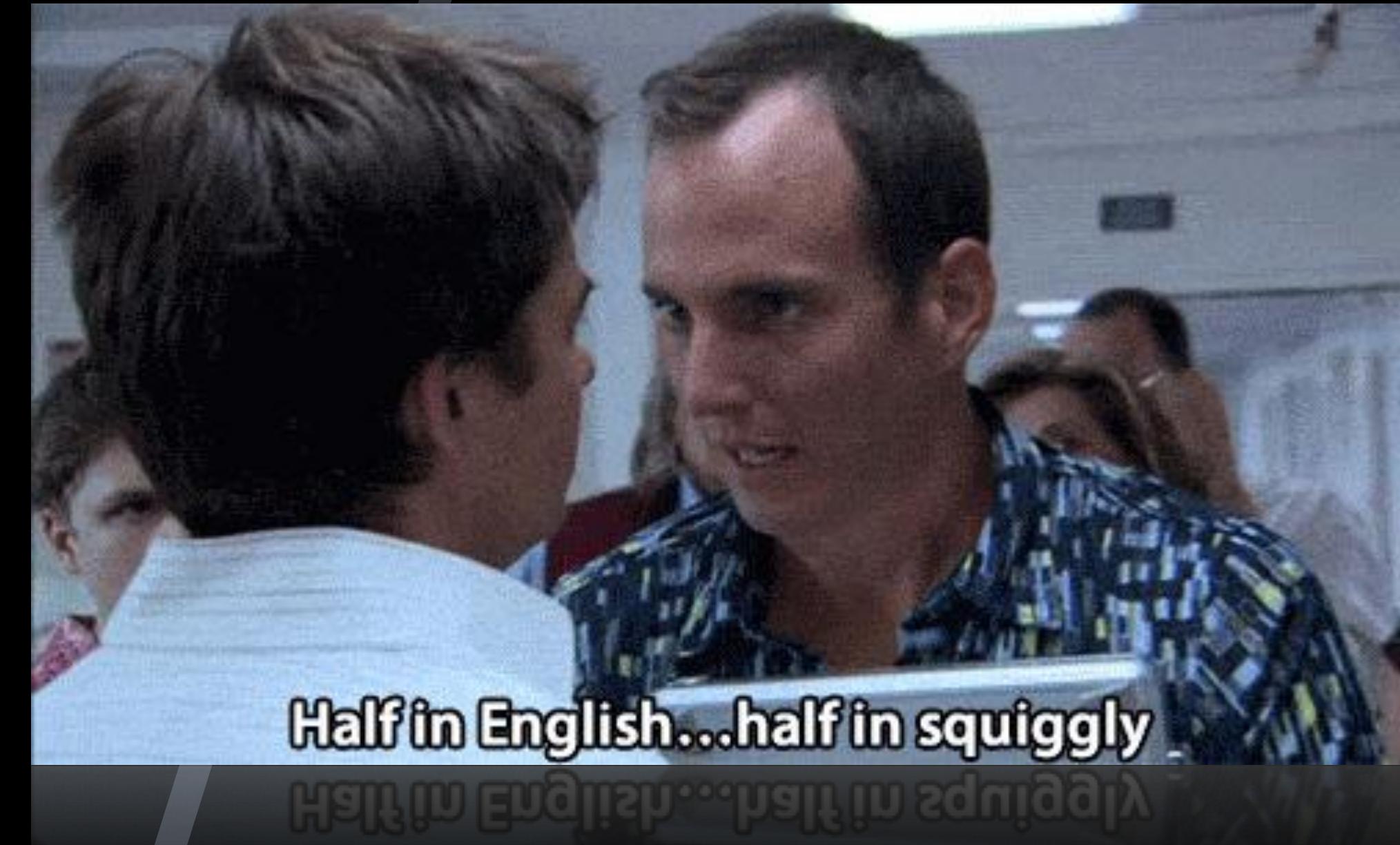
I read as @dataandme has the ability to translate quite technical material to a non-tech audience, so if you're non-tech she's a 🔥 go to! 🚀🏆

8:37 AM - 26 Jul 2017

# ex•o•ter•ic

adj. *understandable by outsiders*

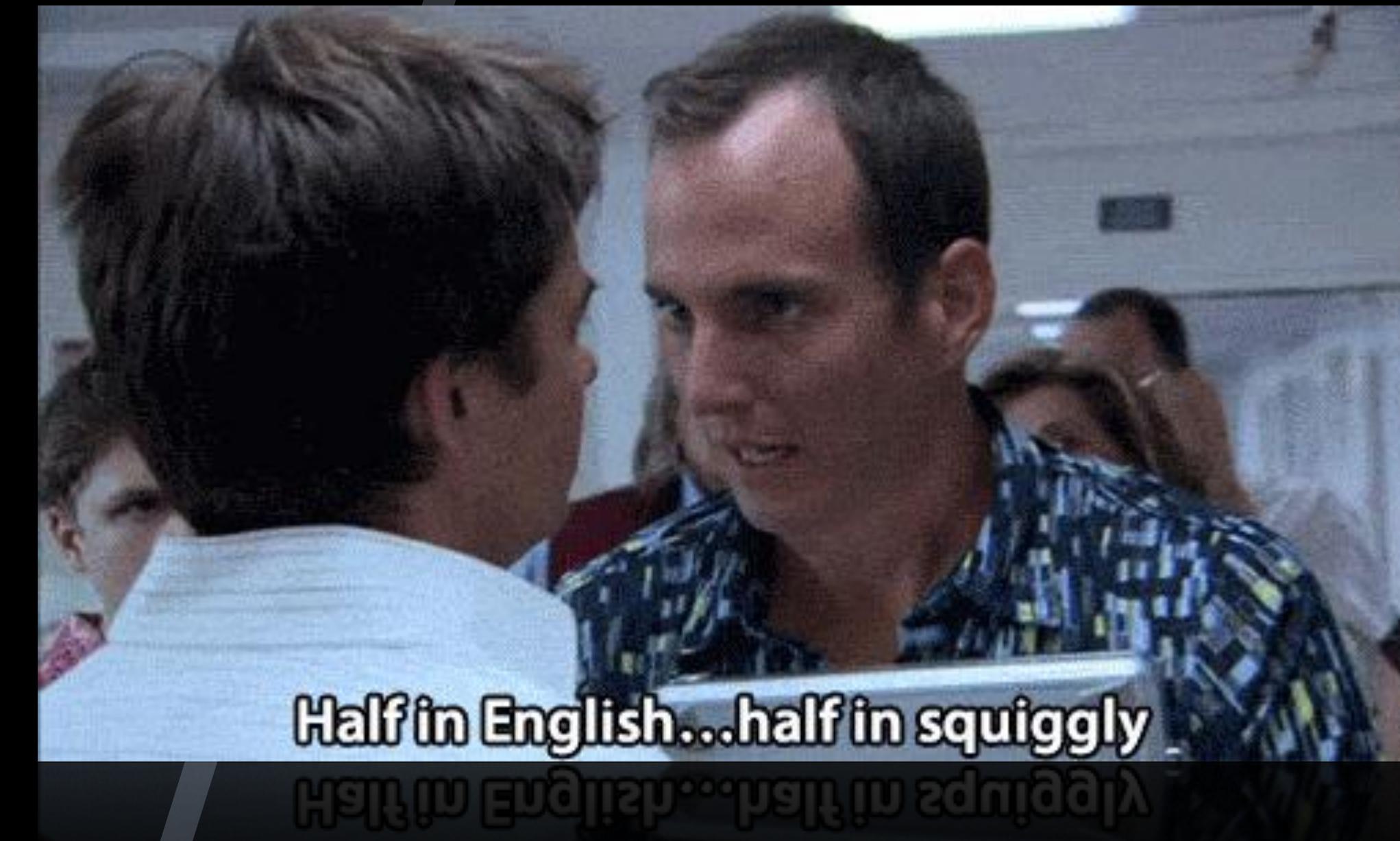
*or the general public*



Half in English...half in squiggly

Half in English...half in squiggly

## Communicating across the English-squiggly divide



Half in English...half in squiggly  
Half in English...half in squiggly

# Who cares about boundaries?

“Nonstop categorization is every bit as indispensable to our survival in the world as is the nonstop beating of our hearts.”

- Hofstadter and Sanders 2013, p.15

# Who cares about boundaries?

"Similitude is still an indispensable border of knowledge. For no equality or relation of order can be established between two things unless their resemblance has at least occasioned their comparison."

- Foucault 1973, p.67

# Who cares about boundaries?

“Teaching in these new [data science] programs has significant overlap in curricular subject matter with traditional statistics courses; yet many academic statisticians perceive the new programs as ‘*cultural appropriation*.’”

– Donoho 2017, p.745

science



natural  
philosophy  
*science*

natural  
philosophy

science





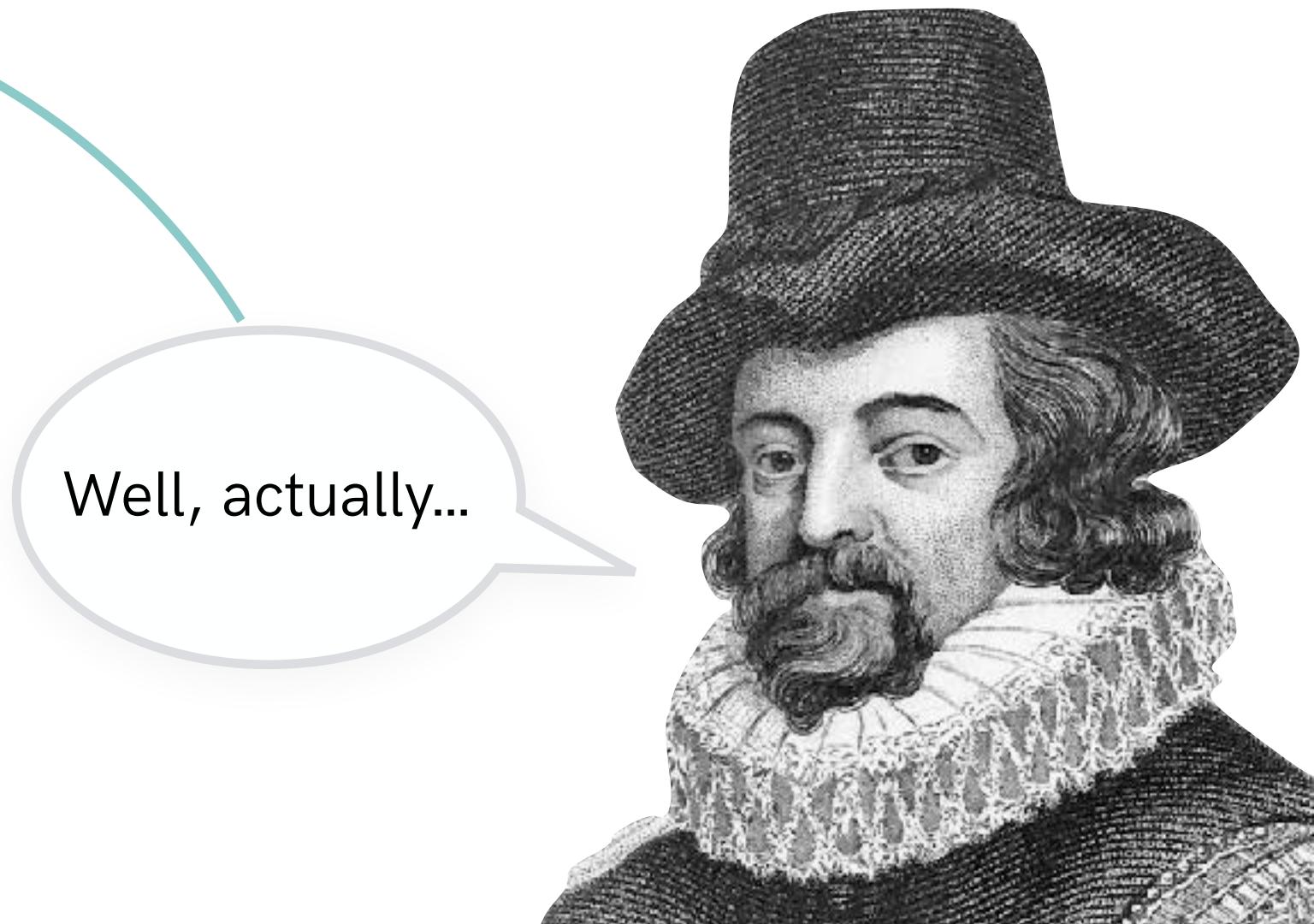
**ORGANON**



# NOVUM ORGANON



Well, actually...



# MAP of the SYSTEM of HUMAN KNOWLEDGE

## UNDERSTANDING

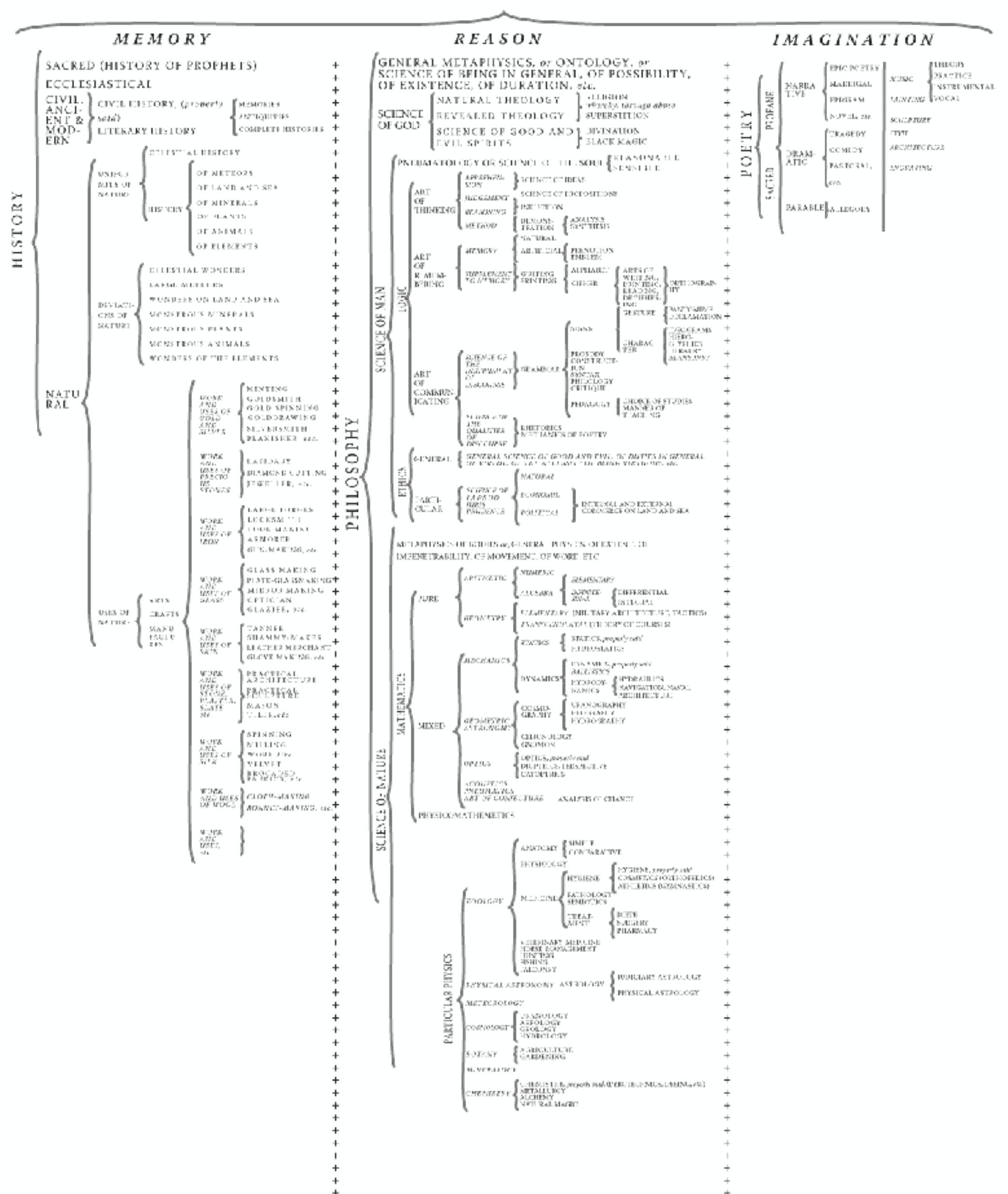


Diagram translated by Benjamin Heller of the University of Michigan and designed by Marketa Kubacakova.(6 Nov. 2013.). The Encyclopedia of Diderot & d'Alembert Collaborative Translation Project. *Quod.lib.umich.edu*. Retrieved from <https://quod.lib.umich.edu/d/did/tree.html>

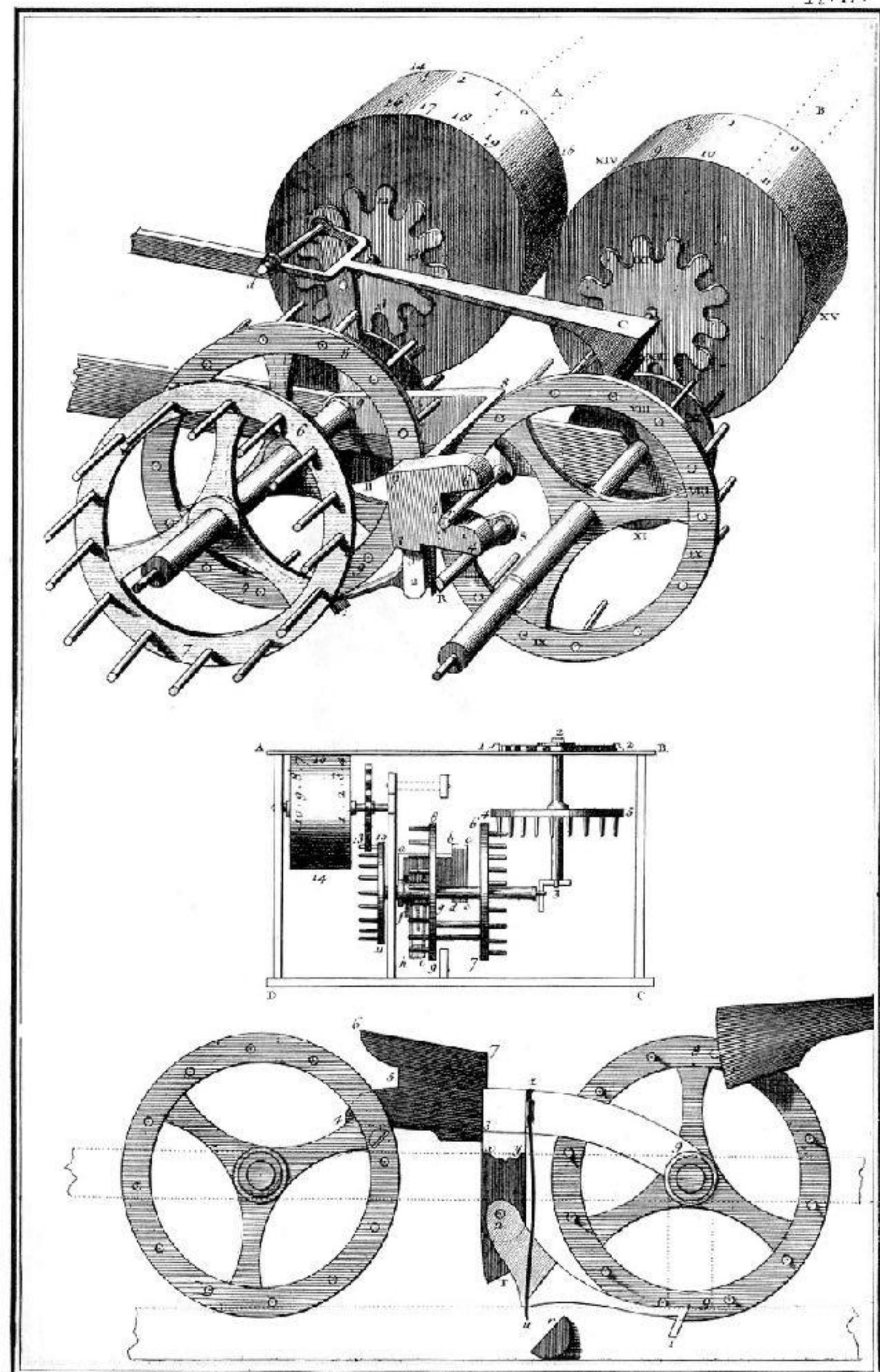
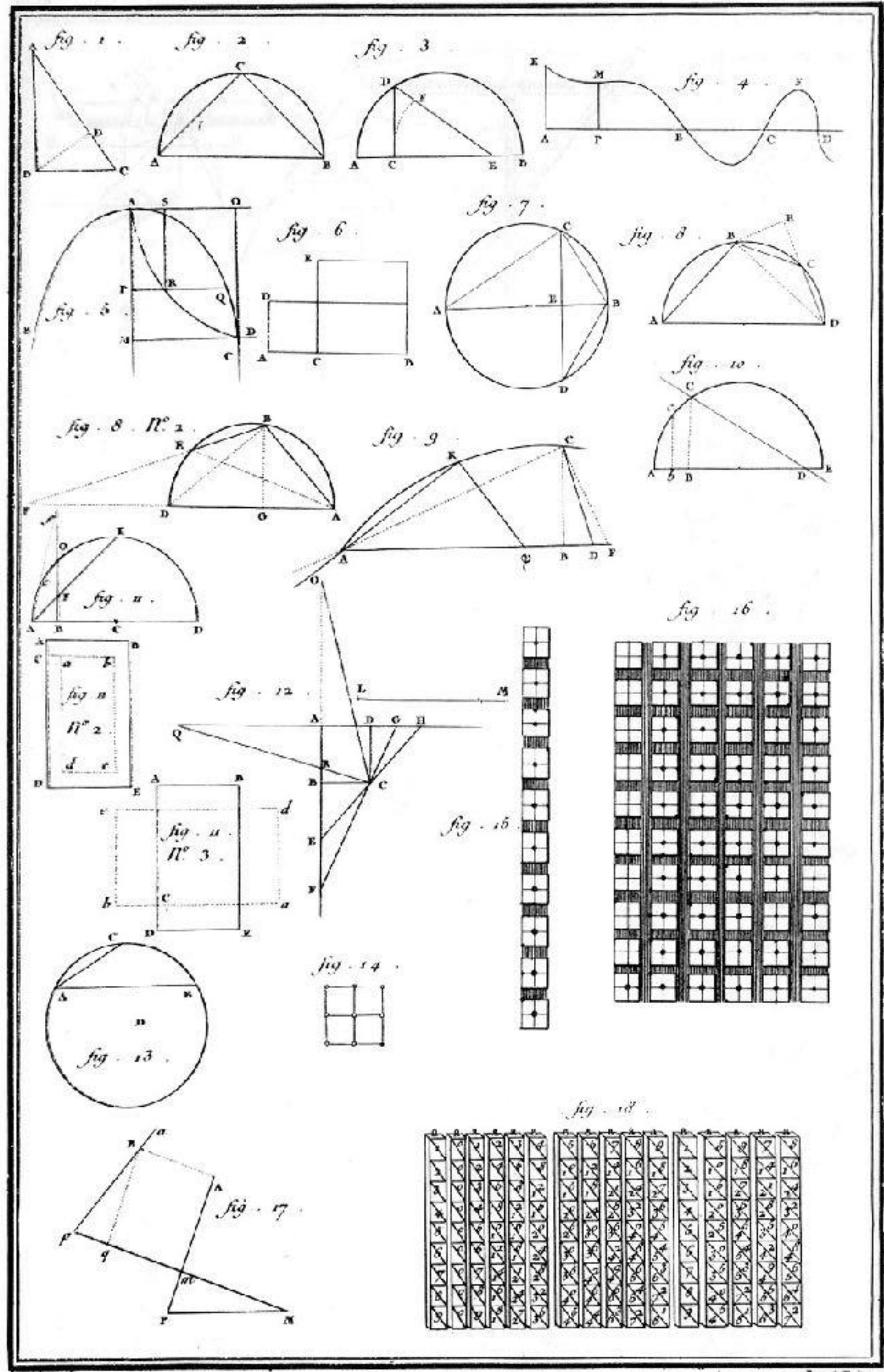
# denis diderot

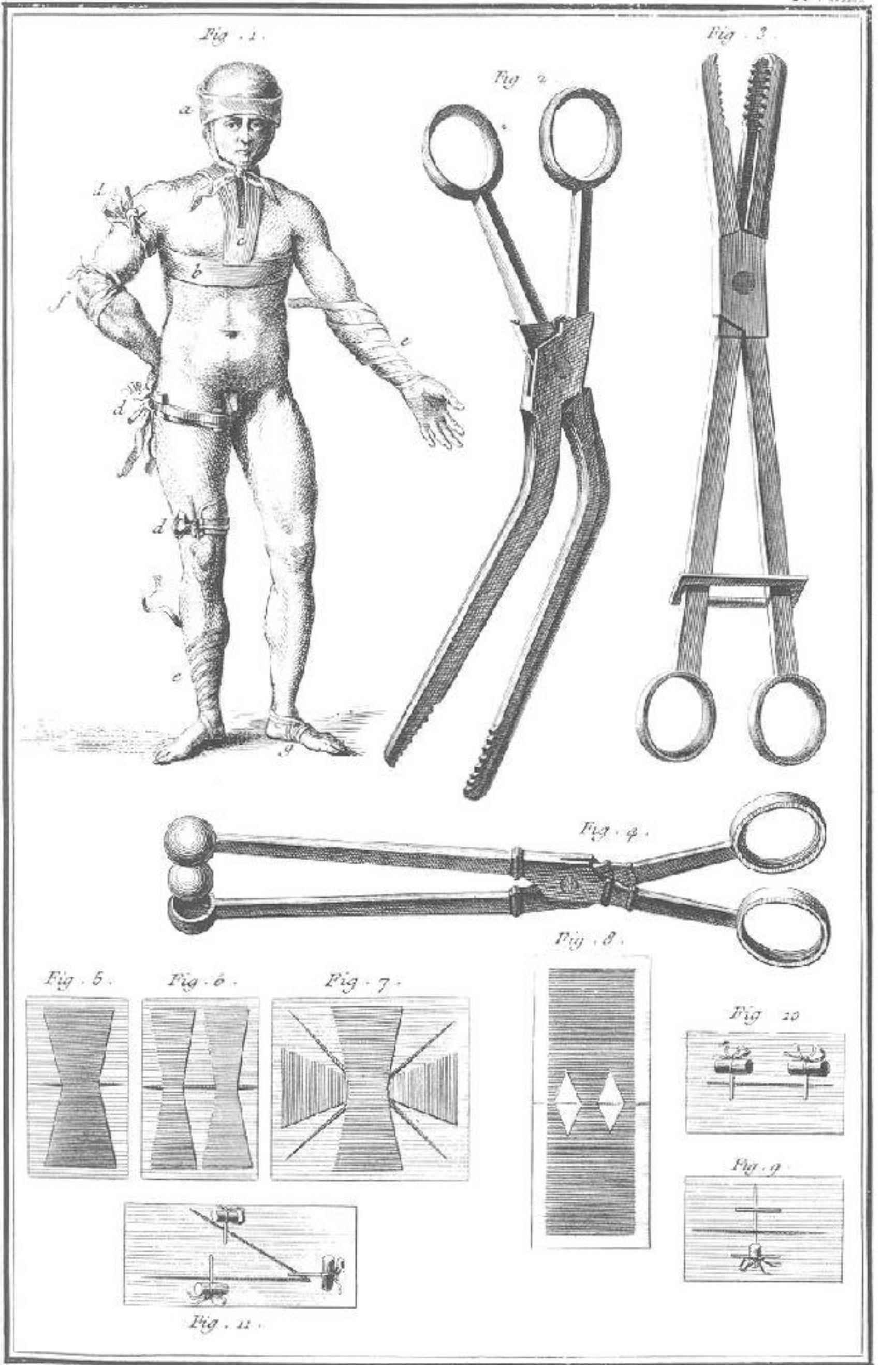
“This is a work that cannot be completed except by a society of men of letters and skilled workmen, each working separately on his own part, but all bound together solely by their zeal for the best interests of the human race and a feeling of mutual good will.”



(1713-1784)

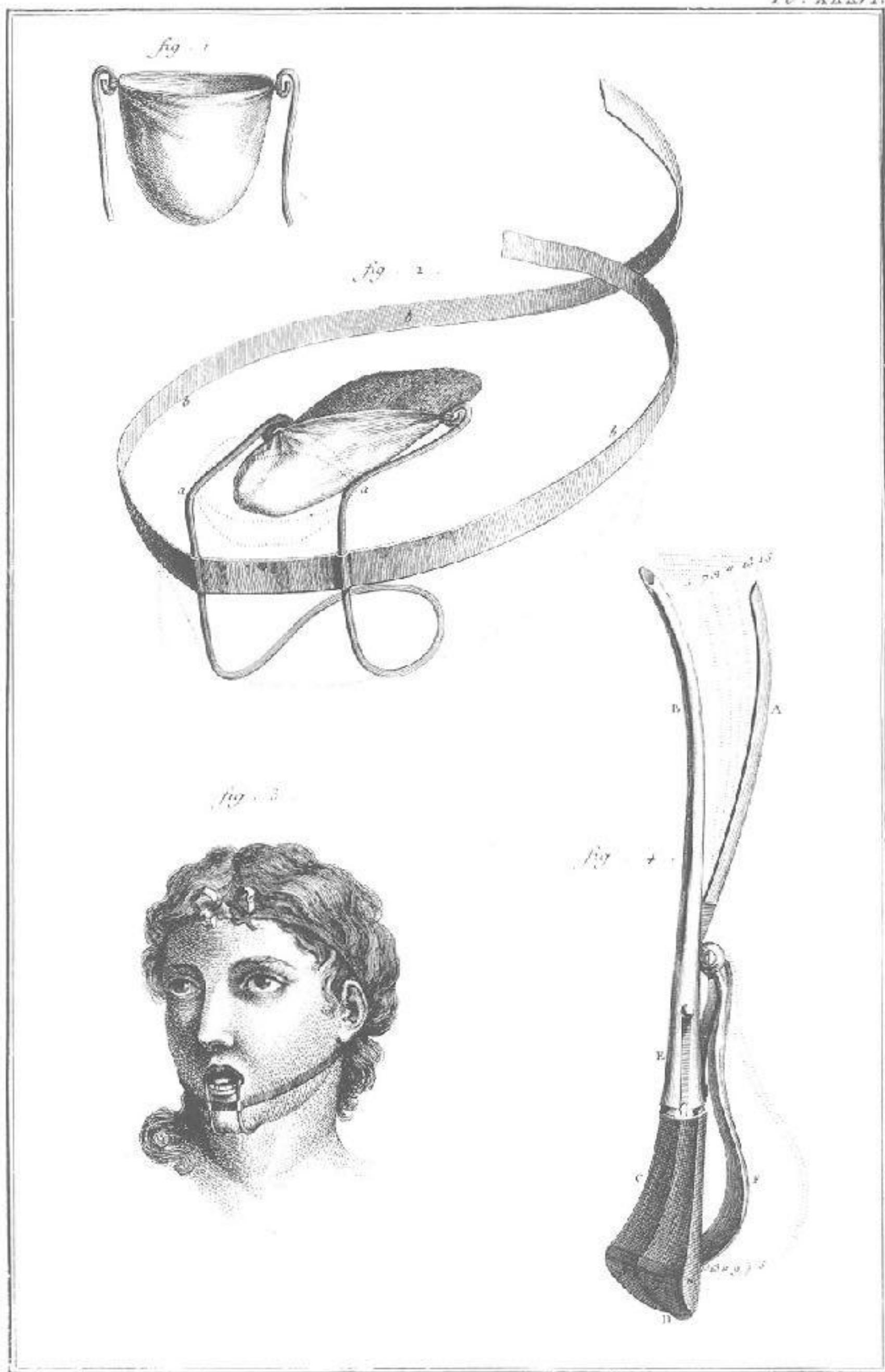
*On L'Encyclopédie, as quoted on The Encyclopedia of Diderot & d'Alembert Collaborative Translation Project. Quod.lib.umich.edu.*





Gouverneur del

## *Chirurgie.*



Cahier des

## *Chirurgie.*

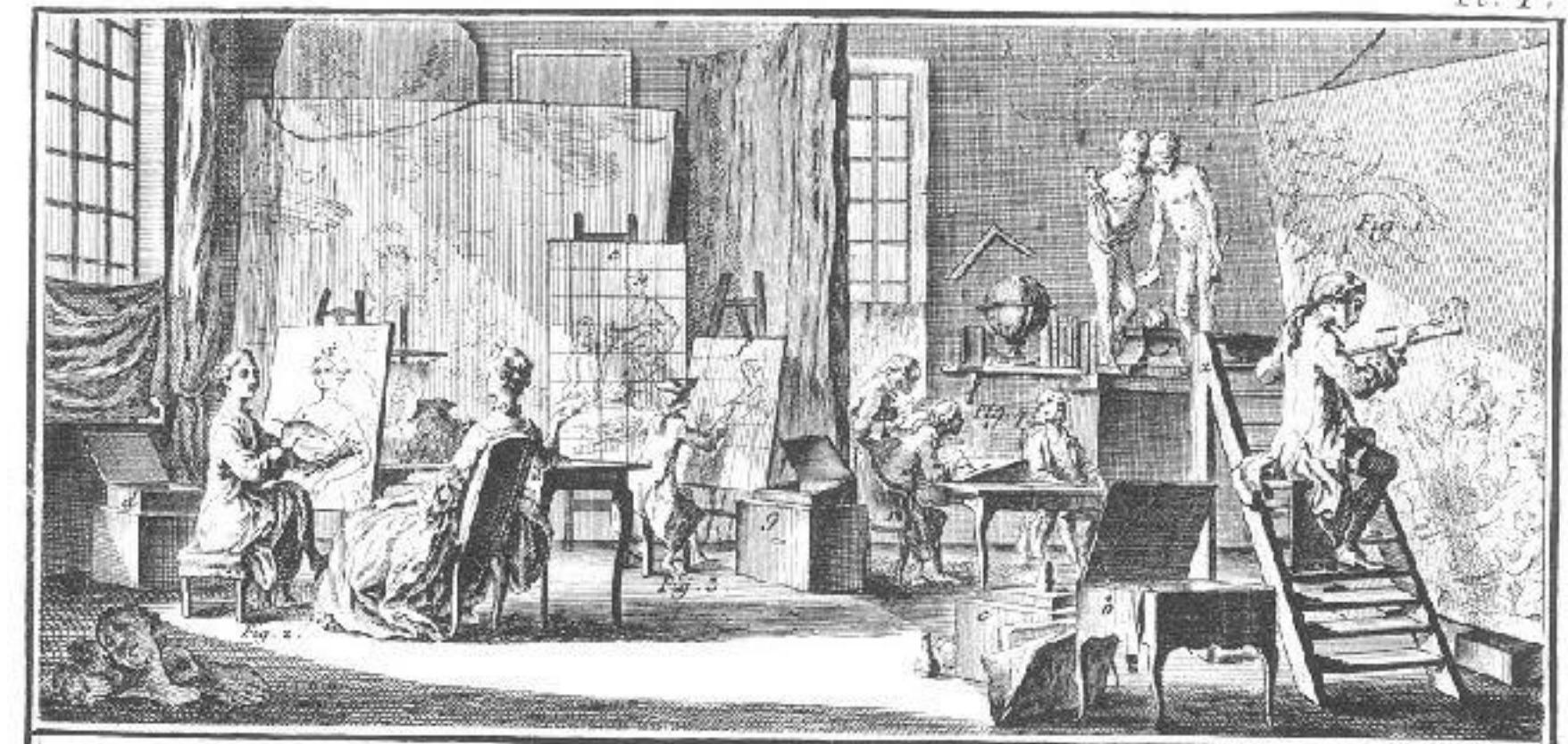


Fig. 1.

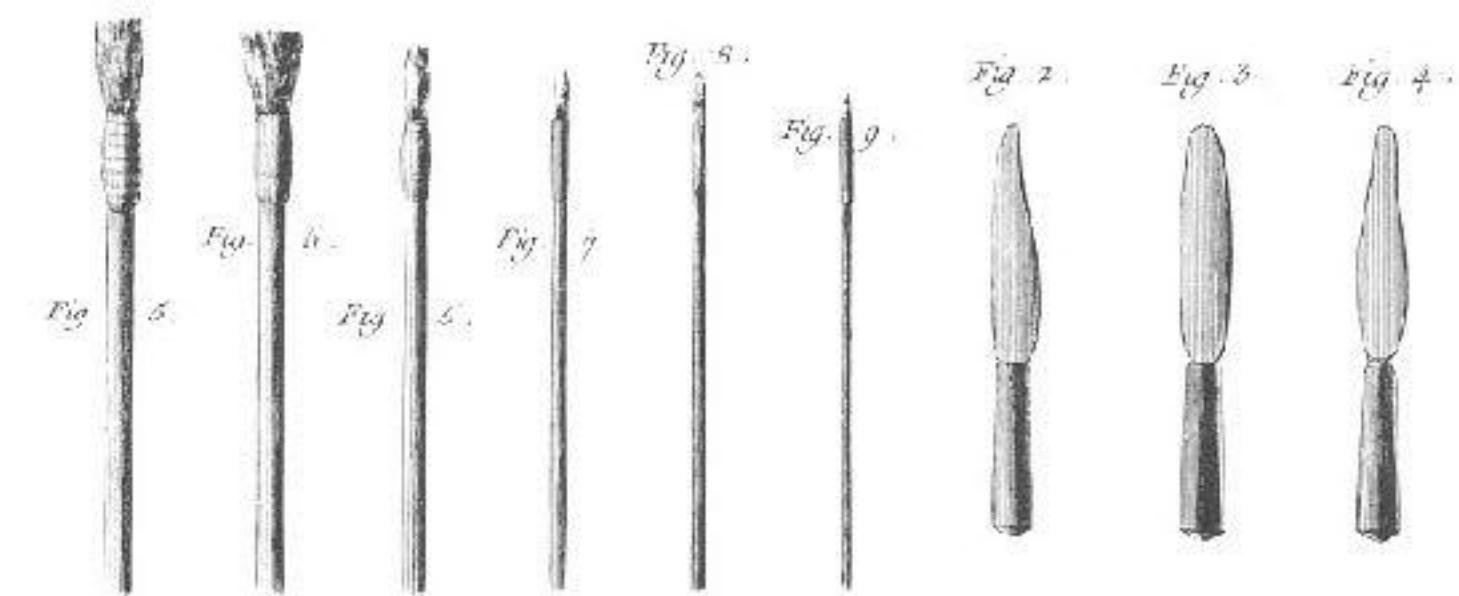


Fig. 11.

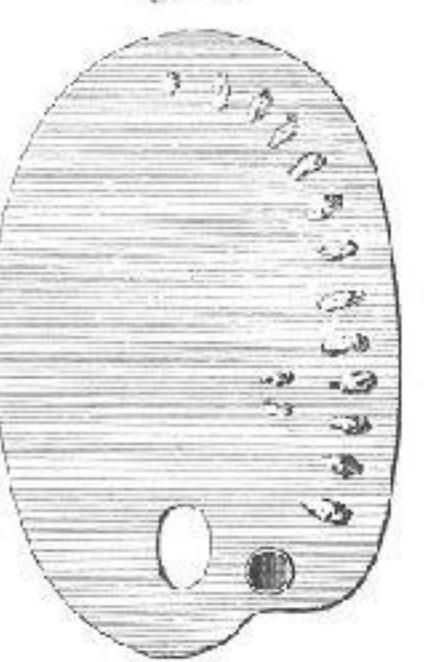
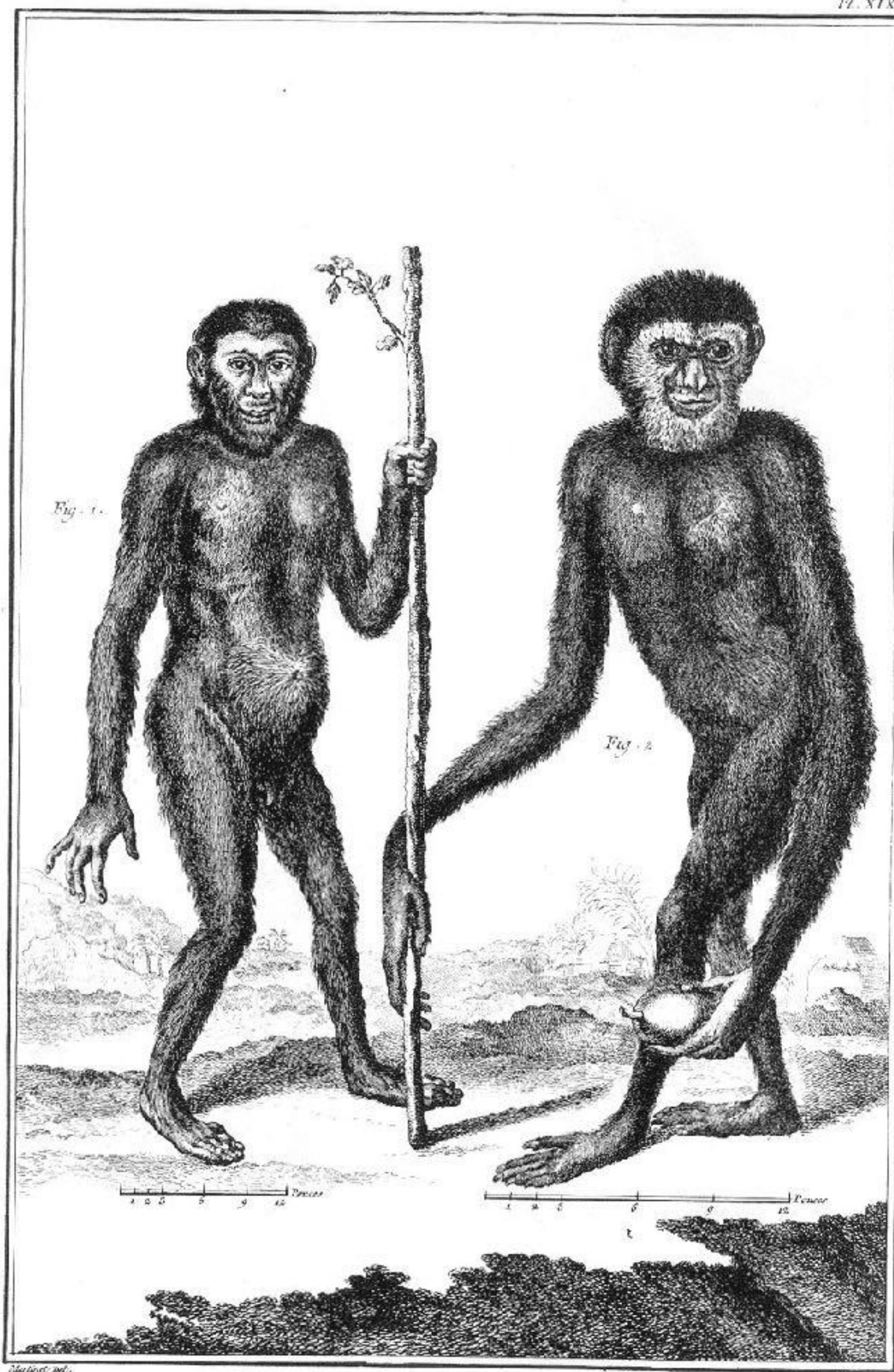


Fig. 12.



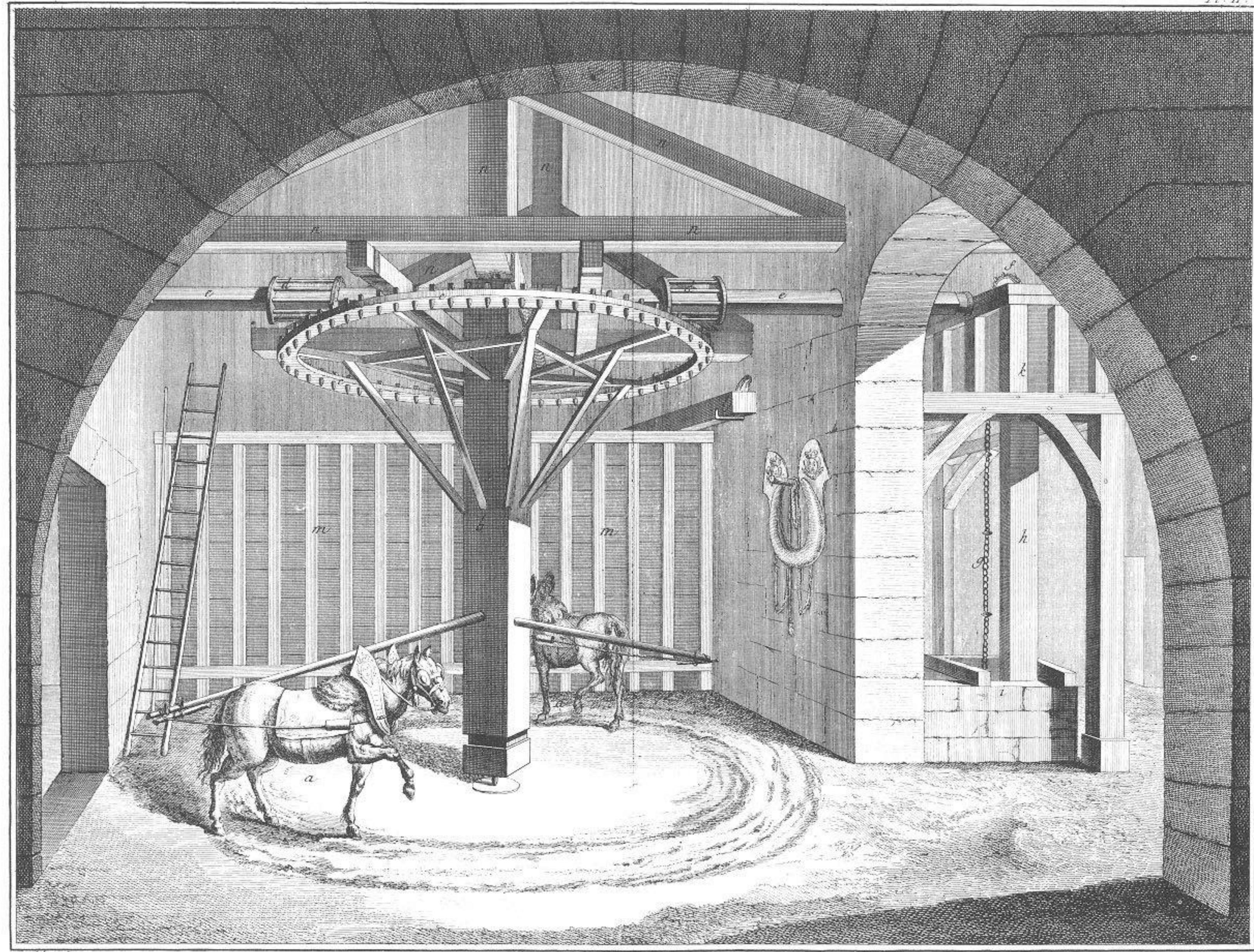
Fig. 13.



Pl. XIX

*Histoire Naturelle,*  
Fig. 1. LE JOCKO. Fig. 2. LE GIBBON.

*Peinture,*  
Atelier, Palettes et Pinceaux.



Rudot del.

Bernard Picot.

Teinture des Gobelins, vue de la machine qui sert à tirer l'eau de la citerne.

*ENCYCLOPEDIE,*  
OU  
*DICTIONNAIRE RAISONNÉ  
DES SCIENCES,  
DES ARTS ET DES MÉTIERS,*  
*PAR UNE SOCIÉTÉ DE GENS DE LETTRES.*

Mis en ordre & publié par M. DIDEROT, de l'Académie Royale des Sciences & des Belles-Lettres de Prusse; & quant à la PARTIE MATHEMATIQUE, par M. D'ALEMBERT, de l'Académie Royale des Sciences de Paris, de celle de Prusse, & de la Société Royale de Londres.

*Tanum series junturaque pollet,  
Tanum de modis sumptis accedit honoris?* HORAT.

TOME PREMIER.



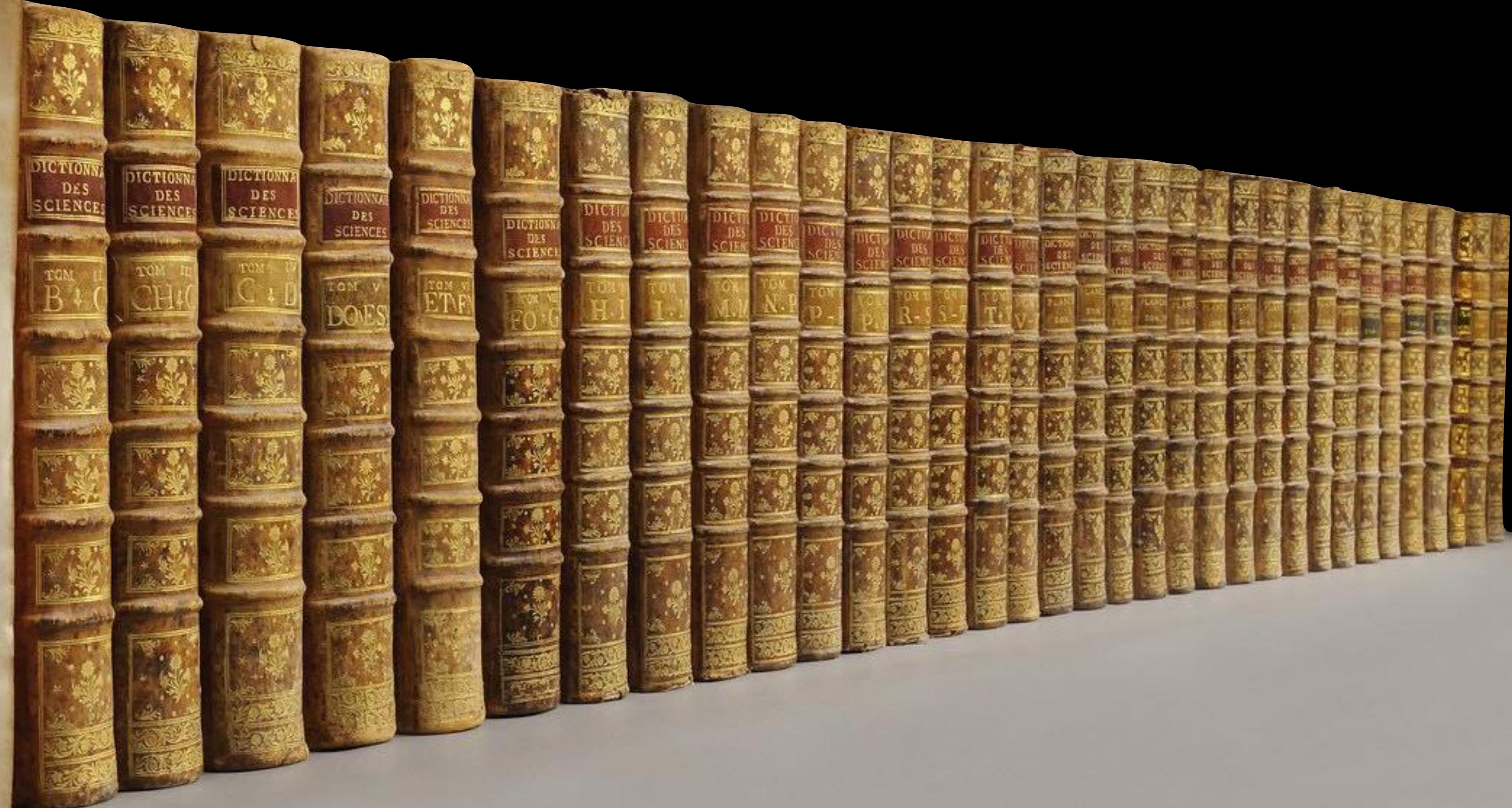
A P A R I S

A PARIS,

chez BREASSON, rue Saint-Jacques, à la Scierie.  
DAVID l'aîné, rue Saint-Jacques, à la Plume-Verte.  
LE BRETON, imprimeur ordinaire du Roi, rue de la Harpe.  
DURAND, rue Saint-Jacques, à Saint-Landry, le au Griffon.

M.DCC.LL

M. DCC. L<sup>E</sup>  
AVEC APPROBATION ET PRIVILEGE DU ROY.



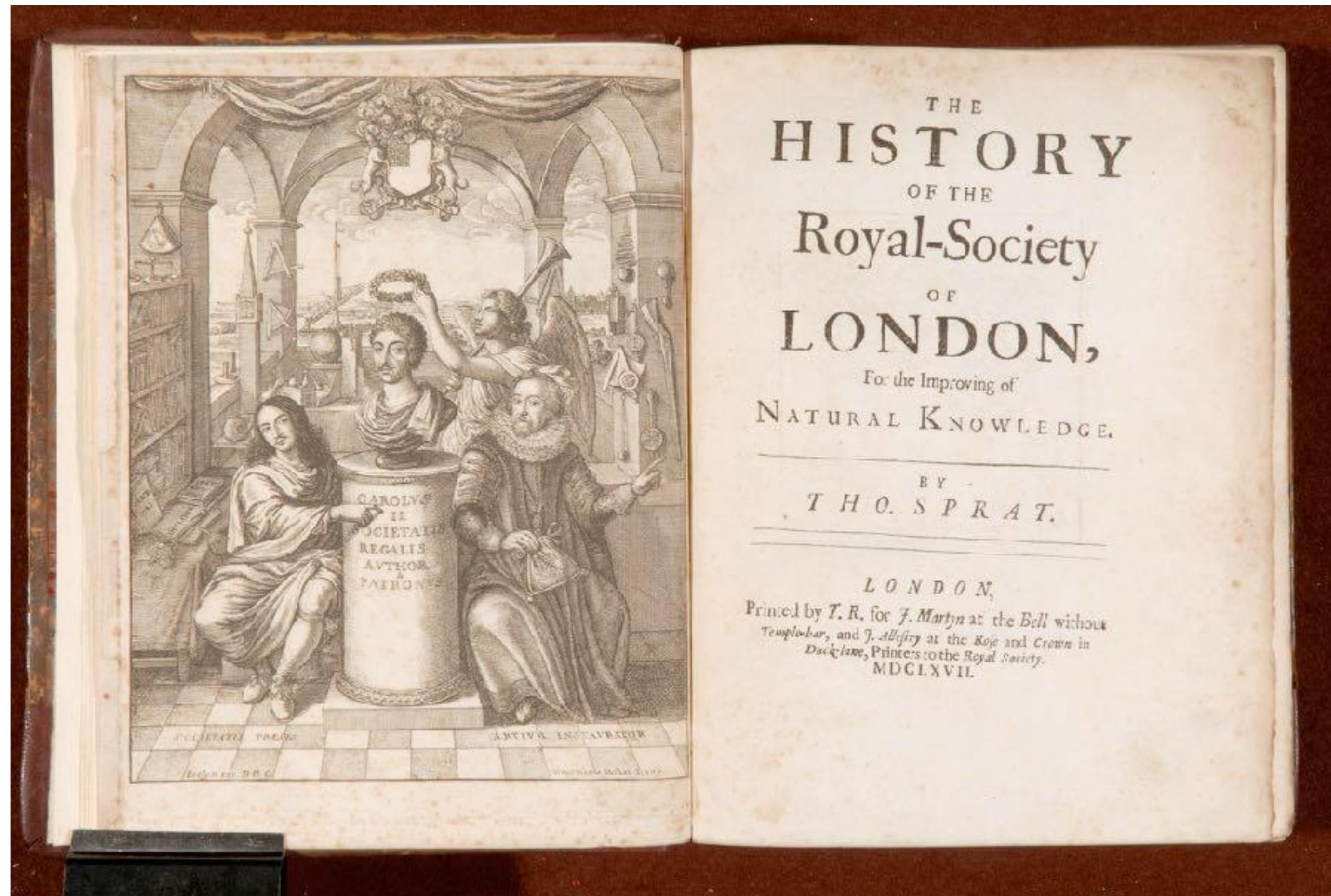
# Institutions of "science"

*Académie des Sciences* 



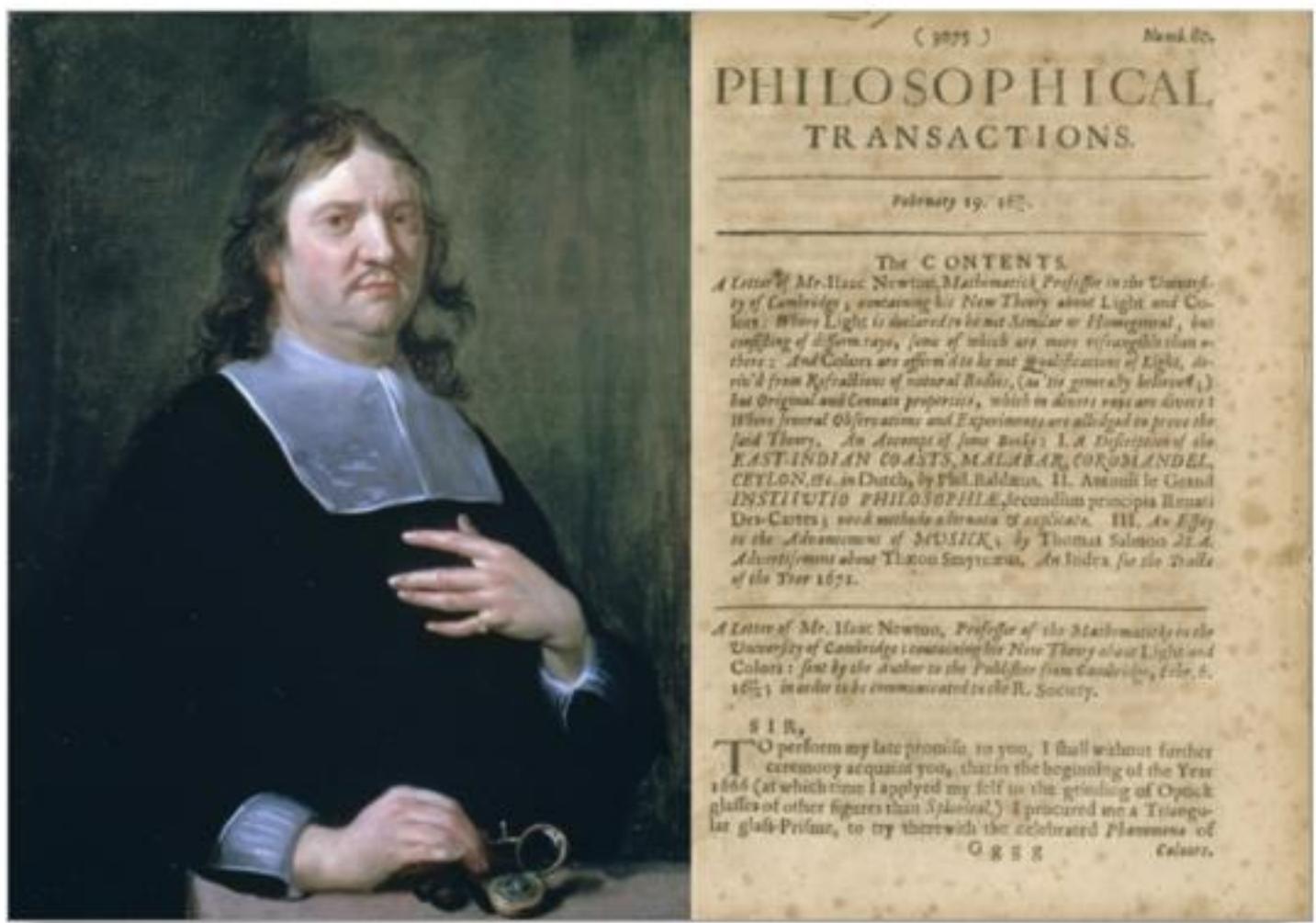
# Institutions of "science"

*Royal Society of London for the Improving of Natural Knowledge* 



# Institutions of "science"

*Royal Society of London for the Improving of Natural Knowledge* 



© The Royal Society

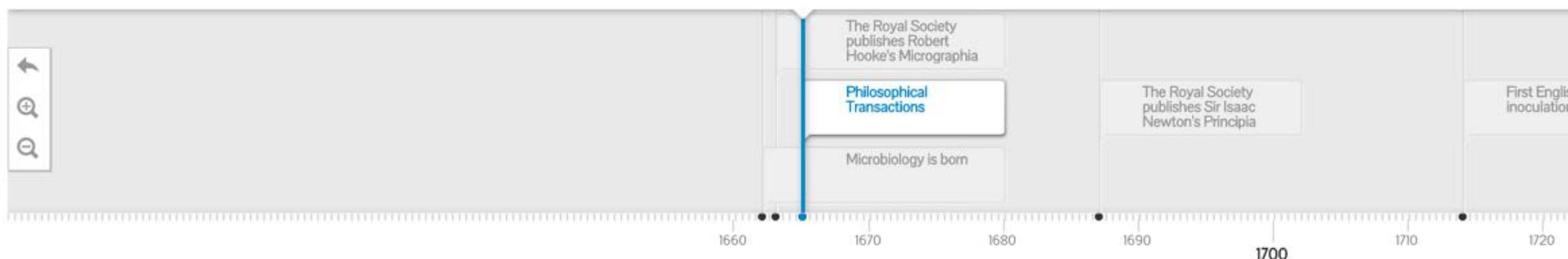
1663 Beginning of modern archaeology

1665 Microbiology is born

**Philosophical Transactions**

Philosophical Transactions of the Royal Society begins publication under the editorial guidance of Henry Oldenburg, Secretary of the Royal Society. This journal is now the oldest scientific journal in continuous publication in the world and established the concepts of scientific priority and peer review.

Portrait of Henry Oldenburg (left) by Jan van Cleve, 1668; and contents page of Philosophical Transactions of the Royal Society, Volume 6 (right).



# Royal Society comms: squad goals

- ...separate *Knowledge of Nature* from:
  - *Colours of Rhetoric*
  - *Devices of Fancy*
  - *Deceits of Fables*

— Thomas Sprat, 1667. *History of the Royal Society of London*

# Royal Society comms: squad goals

- ...separate *Knowledge of Nature* from:
  - *Colours of Rhetoric*
  - *Devices of Fancy*
  - *Deceits of Fables*
- Ditch the language of “*Wits and Scholars*”
- Return to...*primitive purity, and shortness*

— Thomas Sprat, 1667. *History of the Royal Society of London*

# Royal Society comms: squad goals

- ...separate *Knowledge of Nature* from:
  - *Colours of Rhetoric*
  - *Devices of Fancy*
  - *Deceits of Fables*
- Ditch the language of “*Wits and Scholars*”
- Return to...*primitive purity, and shortness*
- Get as near the *Mathematical plainness* as possible

— Thomas Sprat, 1667. *History of the Royal Society of London*

scientist

# Scientist : The Story of a Word

“The appellation ***scientist*** is considered a title of honour, hotly contended for by economists, engineers, physicians, psychologists, and others.”

— Sydney Ross, 1962. *Annals of Science*

“The appellation *scientist* is considered a title of honour, hotly contended for by economists, engineers, physicians, psychologists, and others.”

*data*

"The appellation **'scientist'** is considered a **sexy** title ~~of honour~~, hotly contended for by ~~economists, engineers, physicians, psychologists, and others.~~" *i dunno, but it makes me mad!*

# william whewell

"...by analogy with *artist*, they might form ***scientist***...there could be no scruple in making free with this termination when we have such words as ***sciolist***, *economist*, and *atheist*"



(1794-1866)

W. Whewell. (anonymously) 1834. *The Quarterly Review*, 51, 58-61.

# sci.o.list

*n* • a superficial pretender to knowledge

# Scientist : The Story of a Word

"When an appellation is accepted or rejected as the designation of a group of people...the motives, which are not usually admitted consciously, are dictated by...the image that the word provokes."

— Ross 1962, continued

# Scientist

"The designation *scientist*... was not in accord with the persona... To them the word scientist implied making a business of science; it degraded their **labours of love** to a **drudgery for profits** or salary."

— Ross 1962, p.66

# Data Scientist

“The now-contemplated field of *data science*...is motivated by **commercial** rather than **intellectual developments**. Choosing in this way is likely to miss out on the really important intellectual event of the next 50 years.”

— Donoho 2017, p.745

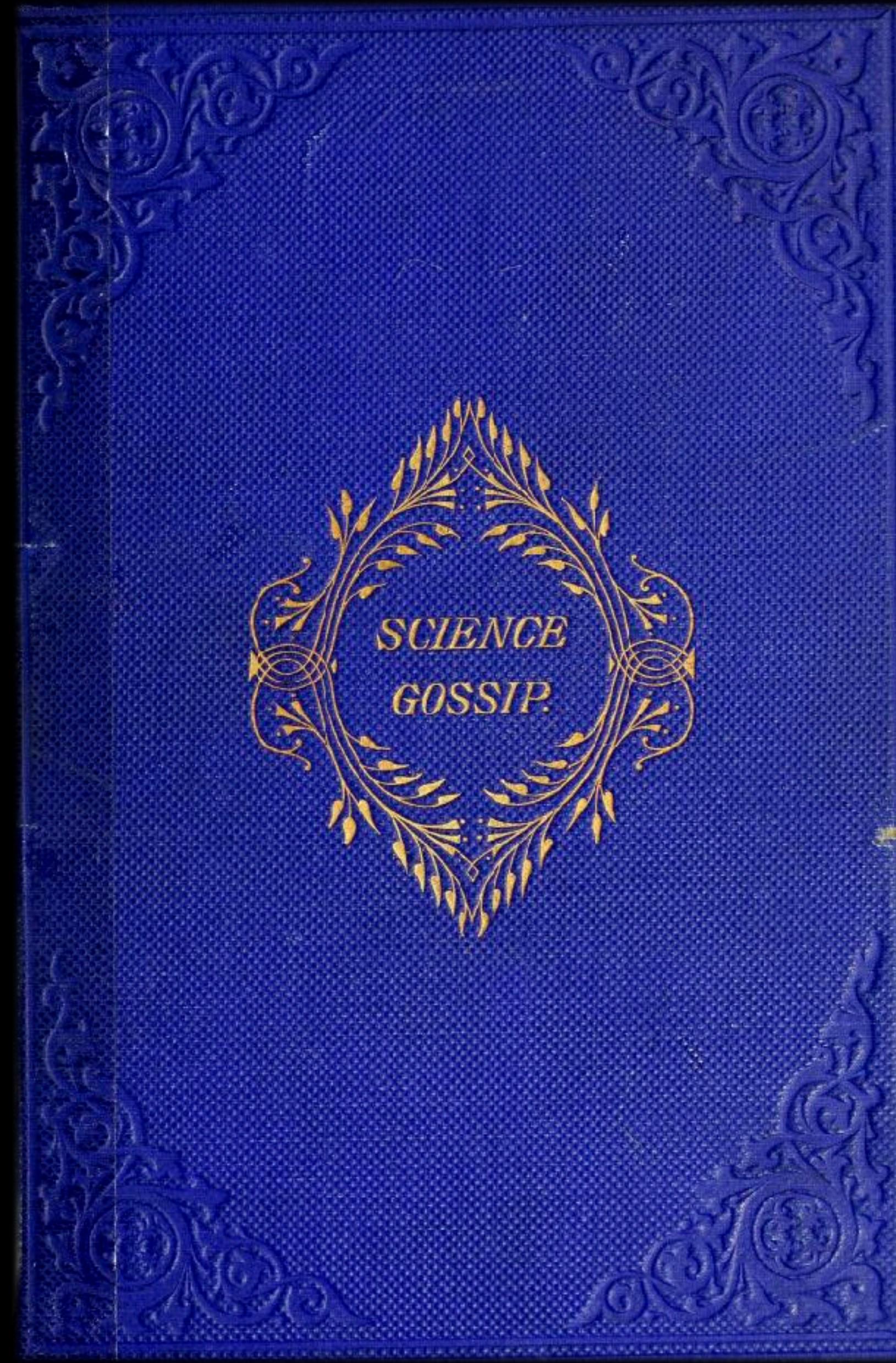
# thomas h. huxley

“To any one who respects the English language, I think '**Scientist**' must be about as pleasing a word as 'Electrocution.' I sincerely trust you will not allow the pages of **Science-Gossip** to be defiled by it.”



(1825-1895)

*Thos. H. Huxley 1894. Letter to J.T. Carrington, editor of *Science-Gossip*, in Ross 1962.*

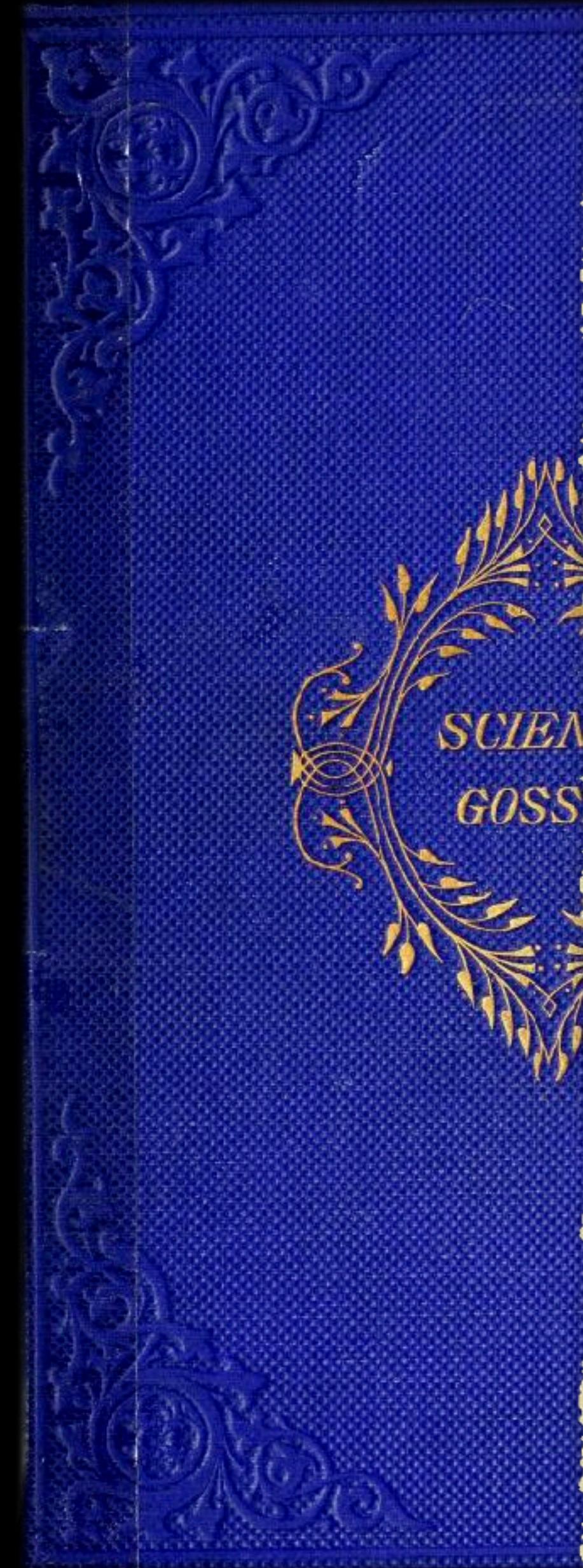


HARDWICKE'S  
**Science-Gossip:**  
AN ILLUSTRATED MEDIUM OF INTERCHANGE AND GOSSIP  
FOR STUDENTS AND  
**LOVERS OF NATURE.**

EDITED BY M. C. COOKE,  
AUTHOR OF "A PLAIN AND EASY ACCOUNT OF THE BRITISH FUNGI," "MICROSCOPIC FUNGI,"  
"A MANUAL OF BOTANICAL TERMS," AND OF "STRUCTURAL BOTANY,"  
THE "BRITISH REPTILES," ETC. ETC.



LONDON:  
ROBERT HARDWICKE, 192, PICCADILLY.  
1863.



## "SCIENTISTS,"—A PROTEST.

IT is idle to ignore the fact that the use of a most objectionable word, "scientist," is increasing, and it is doubtful if any number of protests will have effect on its discontinuance.

Its application is not satisfactory, and is usually the offspring of a paucity of erudition and expression which comes of the modern system of cramming with text-books and general hurry in education. Why not speak of nomenclators as "nameists," of a sempstress as a "sewist," or a conchologist as a "shellist"? All these words may come into use among "progressivists," but are equally abominable and offensive with "scientists."

The word was apparently first coined by Whewell, in 1840, in "Philosophy of the Inductive Sciences," and for some time past has frequently appeared in paragraphs in various newspapers, but it is only lately that one has found it in books of any pretentious character. Why cannot our friends confine themselves to elegance in words? There is abundance of expression in our vocabulary. Recently, when reviewing a book which was the result of years of patient labour, I found the opening paragraph and many subsequent pages disfigured by this abomination. Why the author, who really knows better, should have marred his otherwise excellent work with such a "penny-a-line" word one cannot imagine. Let me appeal to our readers to avoid this word, which is only suited to the pages of a slang dictionary. It is just as easy to select words of elegance and correctness as such mongrels as "scient-ists."

JOHN T. CARRINGTON.

HARDWICKE'S

## Science-Gossip:

THE MEDIUM OF INTERCHANGE AND GOSSIP

FOR STUDENTS AND

MEN OF NATURE.

EDITED BY M. C. COOKE,

AN EASY ACCOUNT OF THE BRITISH FUNGI," "MICROSCOPIC FUNGI,"  
AL OF BOTANICAL TERMS," AND OF "STRUCTURAL BOTANY,"  
THE "BRITISH REPTILES," ETC. ETC.



LONDON:

ERT HARDWICKE, 192, PICCADILLY.

1863.

## "SCIENTISTS,"—A PROTEST.

IT is idle to ignore the fact that the use of a most objectionable word, "scientist," is increasing, and it is doubtful if any number of protests will have effect on its discontinuance.

Its application is not satisfactory, and is usually the offspring of a paucity of erudition and expression which comes of the modern system of cramming with text-books and general hurry in education. Why not speak of nomenclators as "nameists," of a sempstress as a "sewist," or a conchologist as a "shellist"? All these words may come into use among "progressivists," but are equally abominable and offensive with "scientists."

The word was apparently first coined by Whewell, in 1840, in "Philosophy of the Inductive Sciences," and for some time past has frequently appeared in paragraphs in various newspapers, but it is only lately that one has found it in books of any pretentious character. Why cannot our friends confine themselves to elegance in words? There is abundance of expression in our vocabulary. Recently, when reviewing a book which was the result of years of patient labour, I found the opening paragraph and many subsequent pages disfigured by this abomination. Why the author, who really knows better, should have marred his otherwise excellent work with such a "penny-a-line" word one cannot imagine. Let me appeal to our readers to avoid this word, which is only suited to the pages of a slang dictionary. It is just as easy to select words of elegance and correctness as such mongrels as "scient-ists."

JOHN T. CARRINGTON.

opening paragraph and many subsequent pages disfigured by this abomination. Why the author, who really knows better, should have marred his otherwise excellent work with such a "penny-a-line" word one cannot imagine. Let me appeal to our readers to avoid this word, which is only suited to the pages of a slang dictionary. It is just as easy to select words of elegance and correctness as such mongrels as "scient-ists."

JOHN T. CARRINGTON.

His Grace the Duke of Argyle, K.G., K.T., F.R.S.

Gosford, Longniddry, N.B.;

*December 8th, 1894.*

Sir,—In reply to your question, I can only answer for myself, that I never use the word "Scientist" in any serious literary work, and that I regard it with great dislike.

Yours obediently,

ARGYLE.

The Right Hon. Sir John Lubbock, Bart., M.P., F.R.S., wrote:

High Elms, Farnborough, R.S.O., Kent;

*7th December, 1894.*

Sir,—I quite concur with you as to the word "Scientist," and have never used it myself. Why not retain the old word "Philosopher."

I am, your Obedient Servant,

JOHN LUBBOCK.

The Right Hon. Lord Rayleigh, F.R.S., writes:

Terling Place, Witham, Essex;

*10th December, 1894.*

Dear Sir,—I dislike the word "Scientist," and have never used myself; but I foresee a difficulty in avoiding it unless a substitute can be provided. Lord Kelvin's suggestion of reverting to the wider meaning of "Naturalist" might afford a solution.

Yours faithfully,

RAYLEIGH.

Dr. Albert Günther, F.R.S., the learned head of the Zoological Department of the British Museum, writes:

British Museum, Cromwell Road,

London, W.; *December 13th, 1894.*

Dear Sir,—The illegitimacy of formation of the word "Scientist" has been sufficiently exposed in the daily press of a week or so ago. I believe it has been shown to be an American importation. However, as within the last quarter of a century a crowd of writers has sprung up who dabble in science, and especially in the great scientific questions of our time, the word "Scientist" might be retained as an appropriate term for this class.

Yours truly, A. GÜNTHER.

# Ernest Rutherford

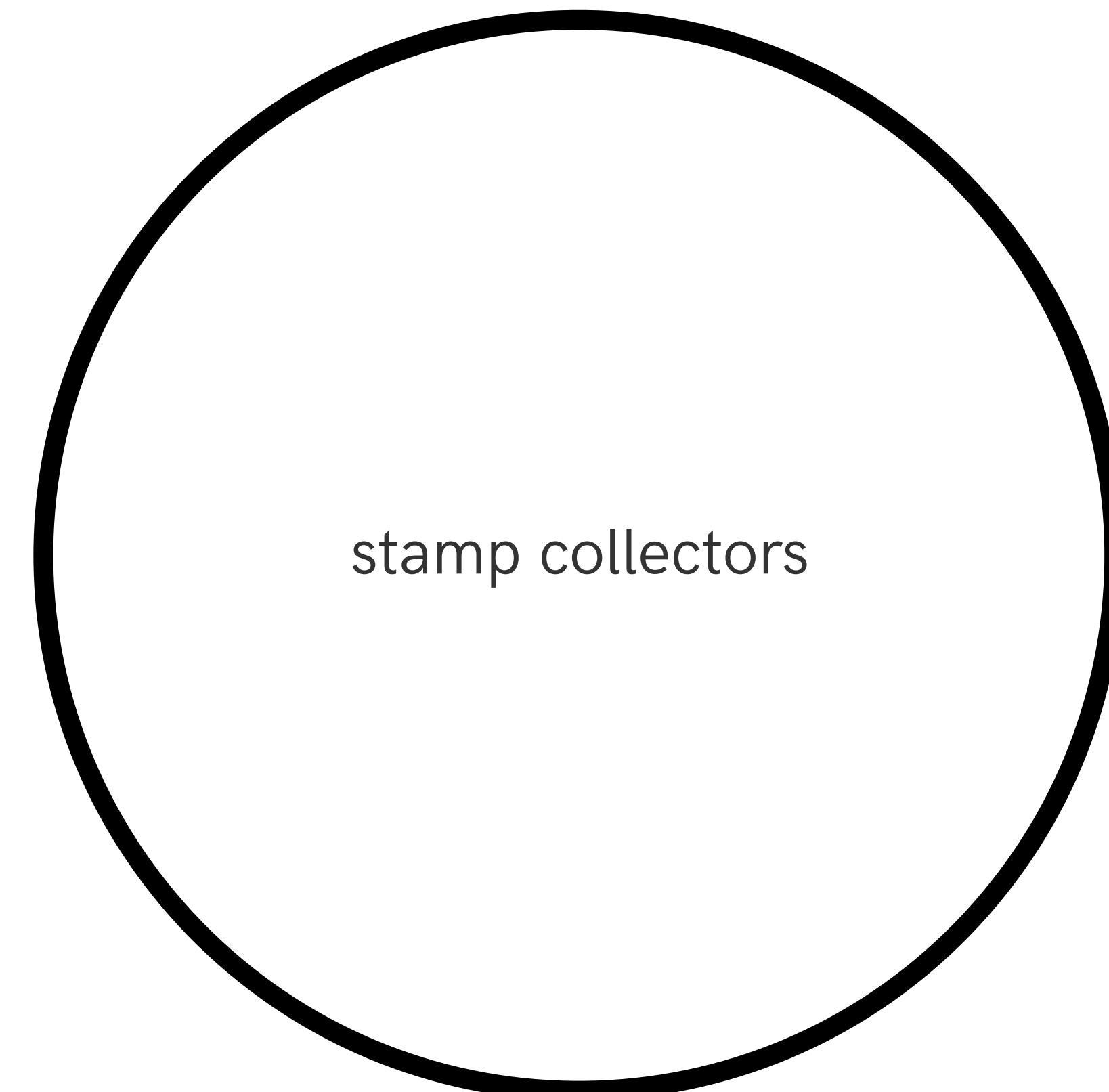
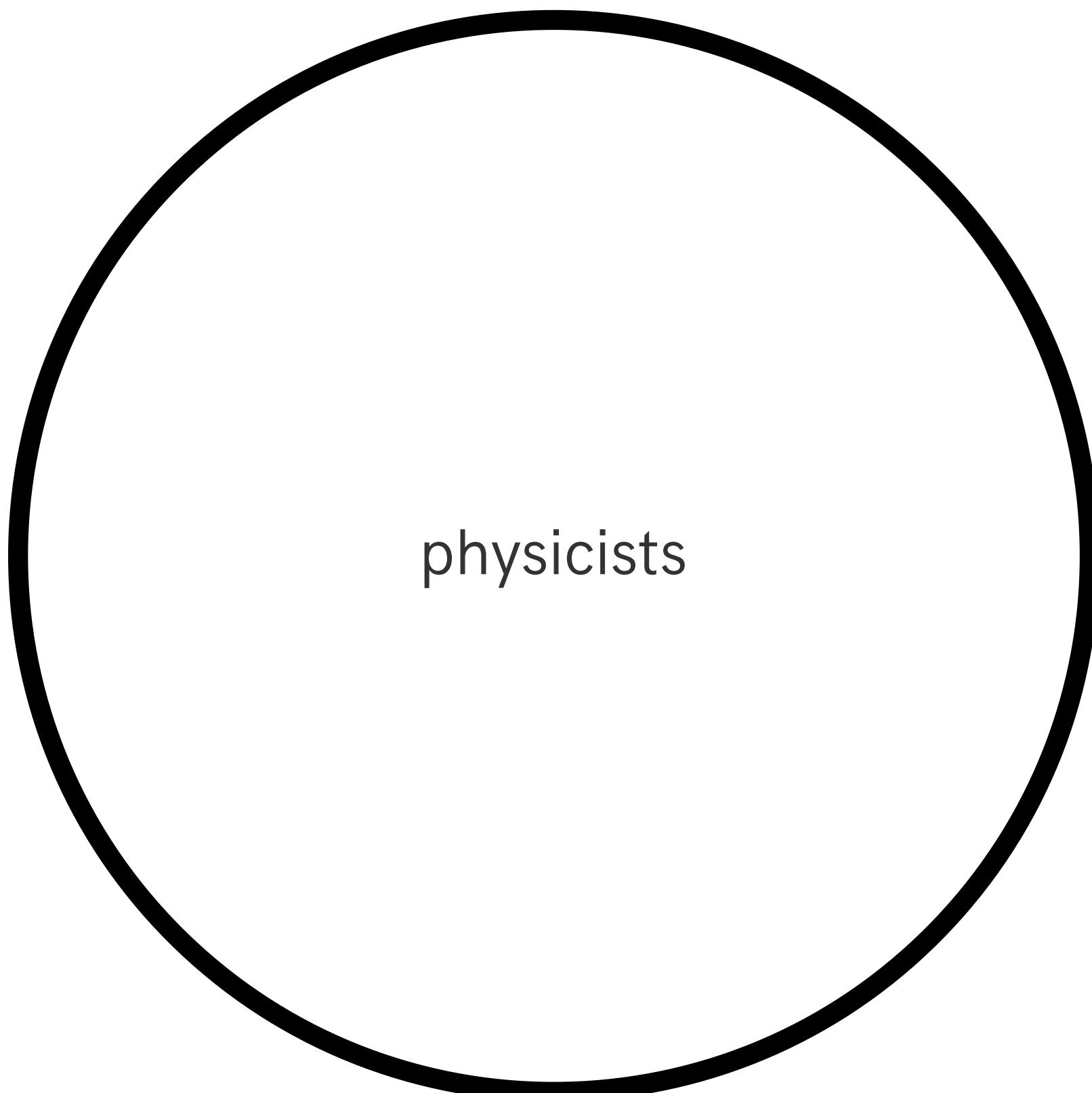
*“...said that scientists were divided into two categories—physicists and stamp collectors”*

*– Daniel Lang, The New Yorker, 1963.*

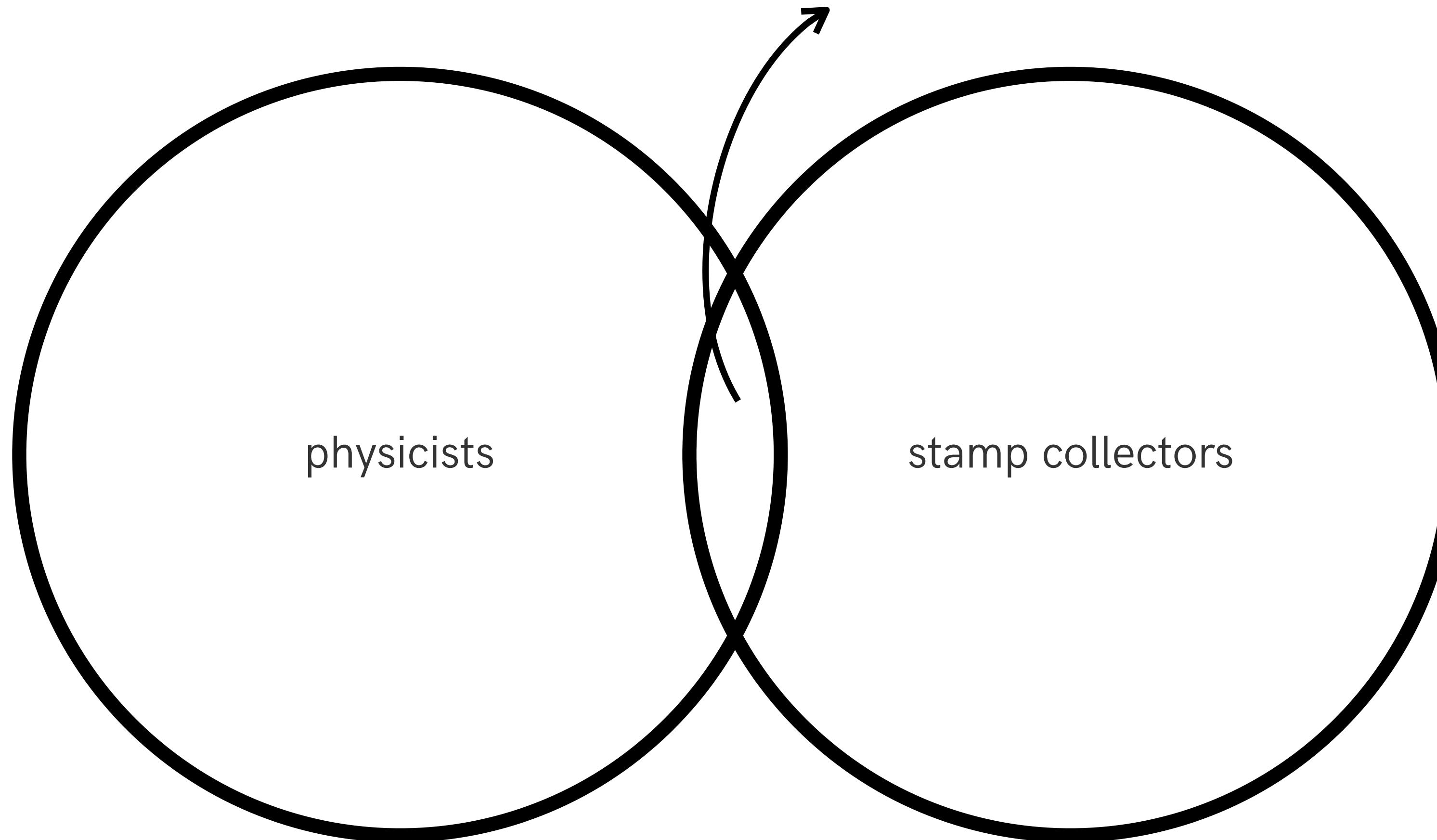


Ernest Rutherford At The McGill University In 1905

# all scientists



# physilatelists



# pan.chres.ton

*n* · an explanation or theory which can fit all cases, being used in such a variety of ways as to become meaningless

# “Mertonian” norms of science

1. Universalism
2. Communism (“Communalism”)
3. Disinterestedness
4. Organized skepticism

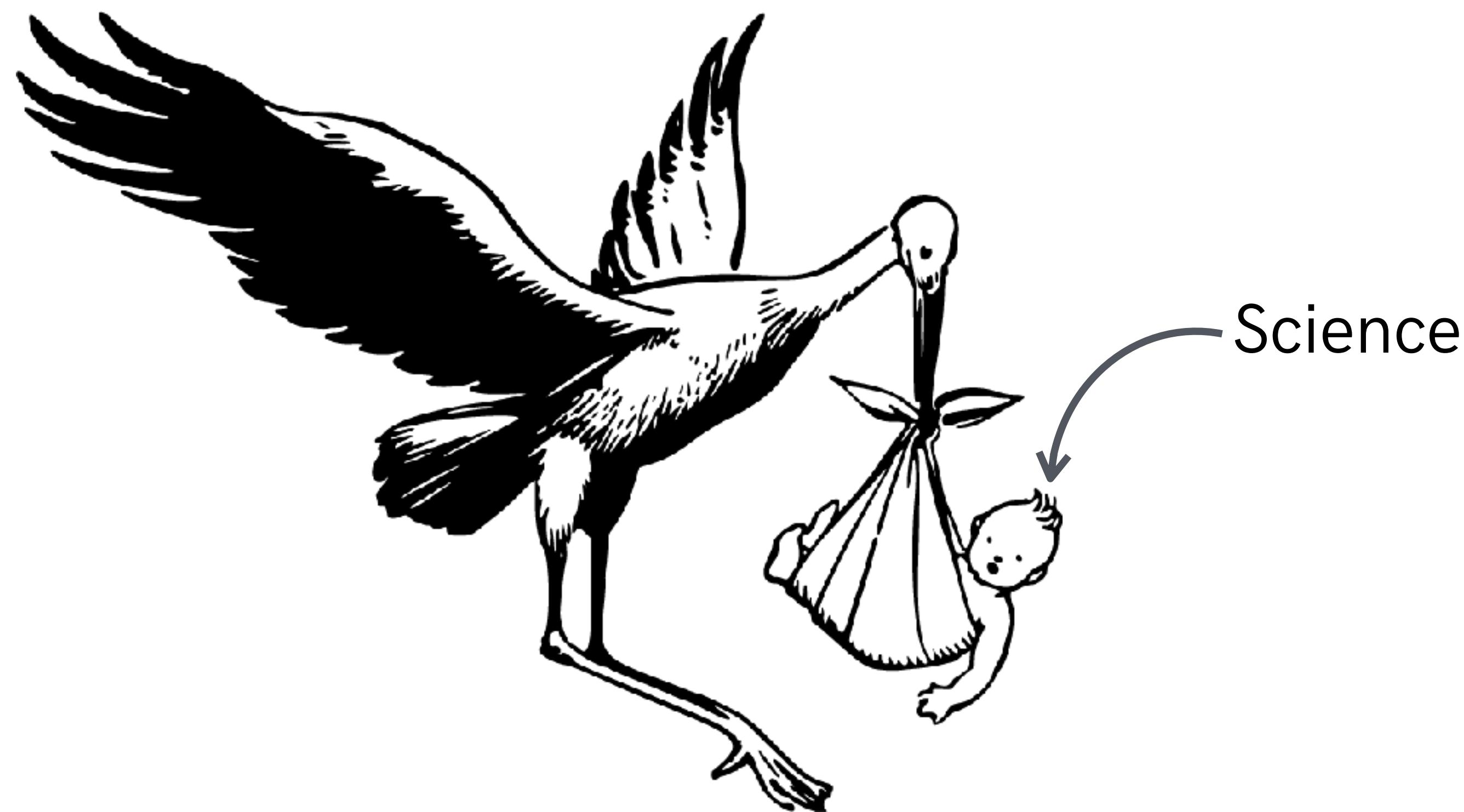


# data science-ing

```
library(tidyverse)  
  
data <- read_csv(data.csv)  
  
model <- fancy_algo(data)  
  
model
```



# Mom, where do data come from?



**“Is not science a  
dispassionate recording of  
facts, uncontaminated by  
value judgements?”**

(Ross 1962)

**science  
in context**



CONTEXT

CONTEXT



What does  
doing science  
look like?



What does  
doing science  
look like?



What does  
doing science  
look like?

What does  
doing science  
look like?



OH, THAT'S WHAT YOU mean BY context

# **socio-technical systems**

# **socio-technical systems**

# **complex socio-technical systems**

(Norman & Stappers 2015)



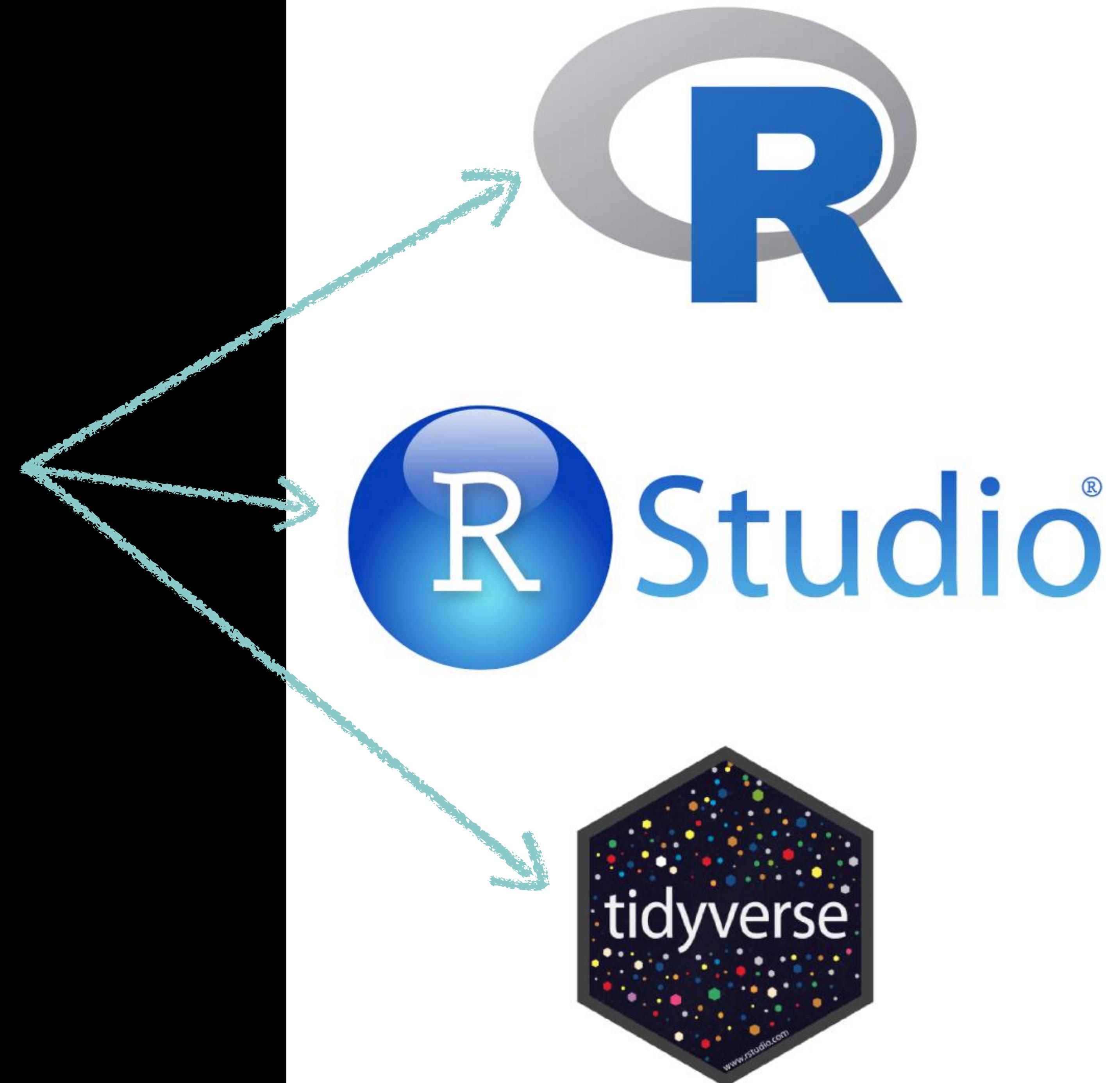
# What About R?

*Oh, I'm super into it!*

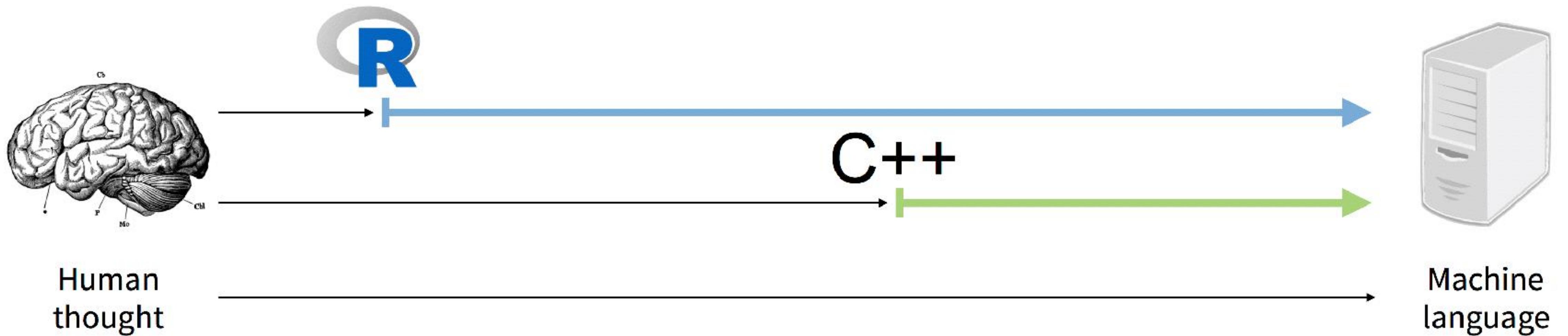
# KRIEGER'S MOST DISTINCTIVE TERMS



my favourite  
tools



# R - a computer language for scientists



# data wrangling



We gotta find a way to make this fit into the hole for this

*Rectangling*

# data wrangling



*Idea*

*Rectangling*

# ~~data wrangling~~



**We gotta find a way to make this fit into the hole for this**





**FIELD  
NOTES®**

48-Page Memo Book  
Durable Materials / Made in the U.S.

FIELD NOTES  
[fieldnotesbrand.com](http://fieldnotesbrand.com)

PROUDLY PRINTED AND MANUFACTURED IN THE U.S.A. — "A DOG CP JOINT"

MABA →

MABA →

with visualization we want to be sure that we communicate the right thing, and that we don't inadvertently communicate the wrong thing. In order to do so, we need to take into account the audience(s) and contexts in which these visualizations are used — we also need to account for the application/use of the visualization, esp. when being used as a "decision-support" mechanism. for example, what our adverse outcome is might inform the way in which we depict uncertainty. we have to ask ourselves what information is being removed in our rendering — are those conversations being eliminated? information fidelity is crucial, and/but we use visualization to help maximize our ability to communicate quite a lot of information. the user/viewer inherently brings his or her own experience to bear on the interpretation of the knowledge being conveyed (if it is in fact knowledge at that point in time). so, what is to be done? do we attempt to measure efficacy, and, if so, how? visualization, in fact, can also be a medium through which we consider new information; we're able to visualize information about our data before we actually know its contents sometimes this is a matter of good, efficient procedures. so, what tools do we even trust to help us w/ such an endeavor. how do we bring more of the tools in a way that helps us augment our strengths as humans?

## ## comment your life

what do you do when you have an idea but don't, at that moment, know how to "make" that idea happen?

a thing I actually have done/will likely continue to do in order to avoid things taking forever in life always  
and also b/c it allows a different kind of thinking

note: below I started to write as though someone else would actually read that (note the spacing)

at times we don't calibrate correctly for the audience, ourselves included

or this are "off" I for n later found by their

author (me) could not be properly parsed/interpreted. However, in trying to get back to what it was that I meant to convey, I am able to make use of clues/signals which, if asked to do so, I would likely be unable to "formalize" or make explicit. even when I do "get" it, I, as reader, bring things to the table that weren't there to begin with

with visualization we want to be sure that we communicate the right thing, and that we don't inadvertently communicate the wrong thing. In order to do so, we need to take into account the audience(s) and contexts in which these visualizations are used — we also need to account for the application/use of the visualization, esp. when being used as a "decision-support" mechanism. for example, what our adverse outcome is might inform the way in which we depict uncertainty. we have to ask ourselves what information is being removed in our rendering — are those conversations being eliminated?

information fidelity is crucial, and/but we use visualization to help maximize our ability to communicate quite a lot of information. the user/viewer inherently brings his or her own experience to bear on the interpretation of the knowledge being conveyed (if it is in fact knowledge at that point in time). so, what is to be done? do we attempt to measure efficacy, and, if so, how? visualization, in fact, can also be a medium through which we consider new information; we're able to visualize information about our data before we actually know its contents sometimes this is a matter of good, efficient procedures. so, what tools do we even trust to help us w/ such an endeavor. how do we bring more of the tools in a way that helps us augment our strengths as humans?

## ## comment your life

what do you do when you have an idea but don't, at that moment, know how to "make"

you consider your audience. if it's you, you capture the idea in a way that's different from how you would do so for a second-party or an audience

a thing I actually have done/will likely continue to do in order to avoid things taking forever in life always

and also b/c it allows a different kind of thinking

note: below I started to write as though someone else would actually read that (note the spacing)

at times we don't calibrate correctly for the audience, ourselves included

sometimes, your mental model/calculations for this are "off." I, for one have left myself a bunch of notes which, when later found by their author (me) could not be properly parsed/interpreted. However, in trying to get back to what it was that I meant to convey, I am able to make use of clues/signals which, if asked to do so, I would likely be unable to "formalize" or make explicit. even when I do "get" it, I, as reader, bring things to the table that weren't there at the time of their writing. to find the common ground

What is the inverse-opposite version  
of cooperative federalism?  
- moral federalism?

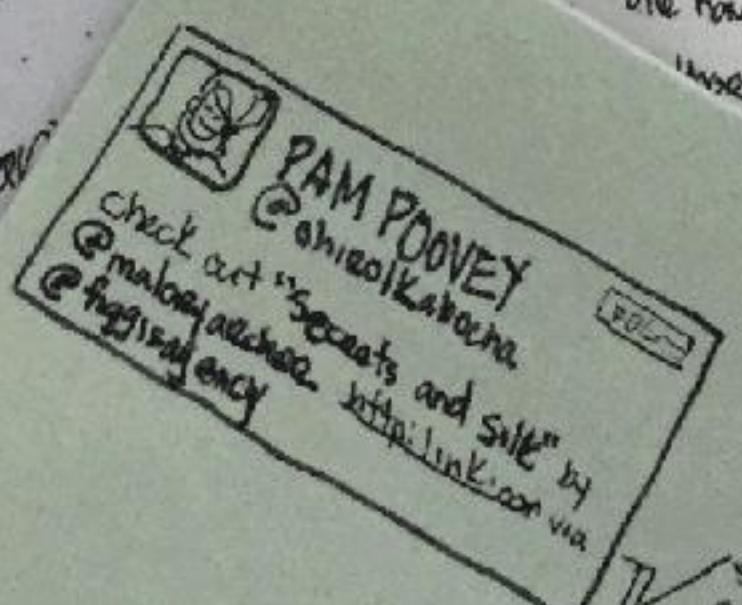
with visualization we want to be sure that we communicate the right thing, and that we don't inadvertently communicate the wrong thing. In order to do so, we need to take into account the audience(s).

high-level constructs such these visualizations are used - we application/use of the visualization, e.g. support mechanism. for example, inform the way in which we depict what information is being removed from the way in which we depict information being eliminated?

DOM  
Math  
Transition

"incremental elaboration" on  
mental models

Visualization favors certain mental models  
- Yui Chang @openic



PAM DOVEY  
@shirokabocha  
Check out "Secrets and Signs" by  
@makeachecklist, [@fogisfactory](http://link.com/via)

two pieces of info  
and combine  
accessed "matrix"  
note

And this ties into the  
"aspirational" reading  
we tend to do  
while scrolling a  
feed

it don't want to leave  
our "self" atm

"... scale forces us

generated using **Roxygen2**

## comment your life

what do you do when you have an idea but don't, at that moment, know how to "make"

you consider your audience.  
if it's you, you capture the idea in a way that's different from how you would do so for a second-party or an audience

a thing I actually have done/call likely continue to do in order to avoid things taking forever in life always

and also b/c it allows a different kind of thinking

: roxygen2 blocks

note: below I started to write as though someone else would actually read that (note the spacing)

at times we don't collaborate correctly for two audience, audience included

comments to .R → our document() → convert

to .Rd files

rect directive for funs

tag, and combines

importFrom in one

PACE file

: roxygen2 tags

comment to .R → our

document() → convert

to .Rd files

of foreknowledge that makes new  
very - if the goal is true "abstraction" it  
fundamentally destined to fail; humans  
are "doggish and metaphorical"; in order to understand  
similitude" (forecasts), and categorization

works

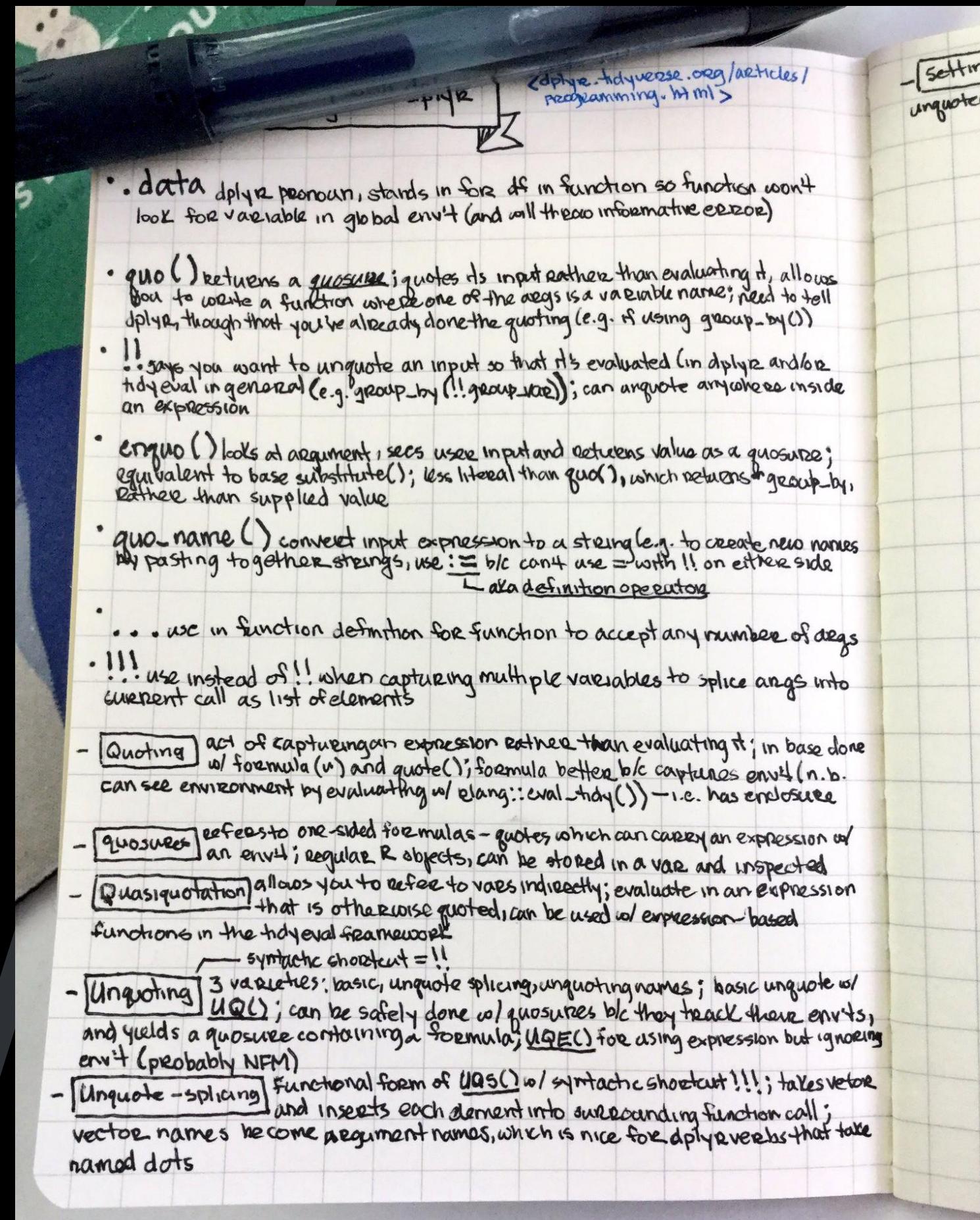
# Not Very Tidy



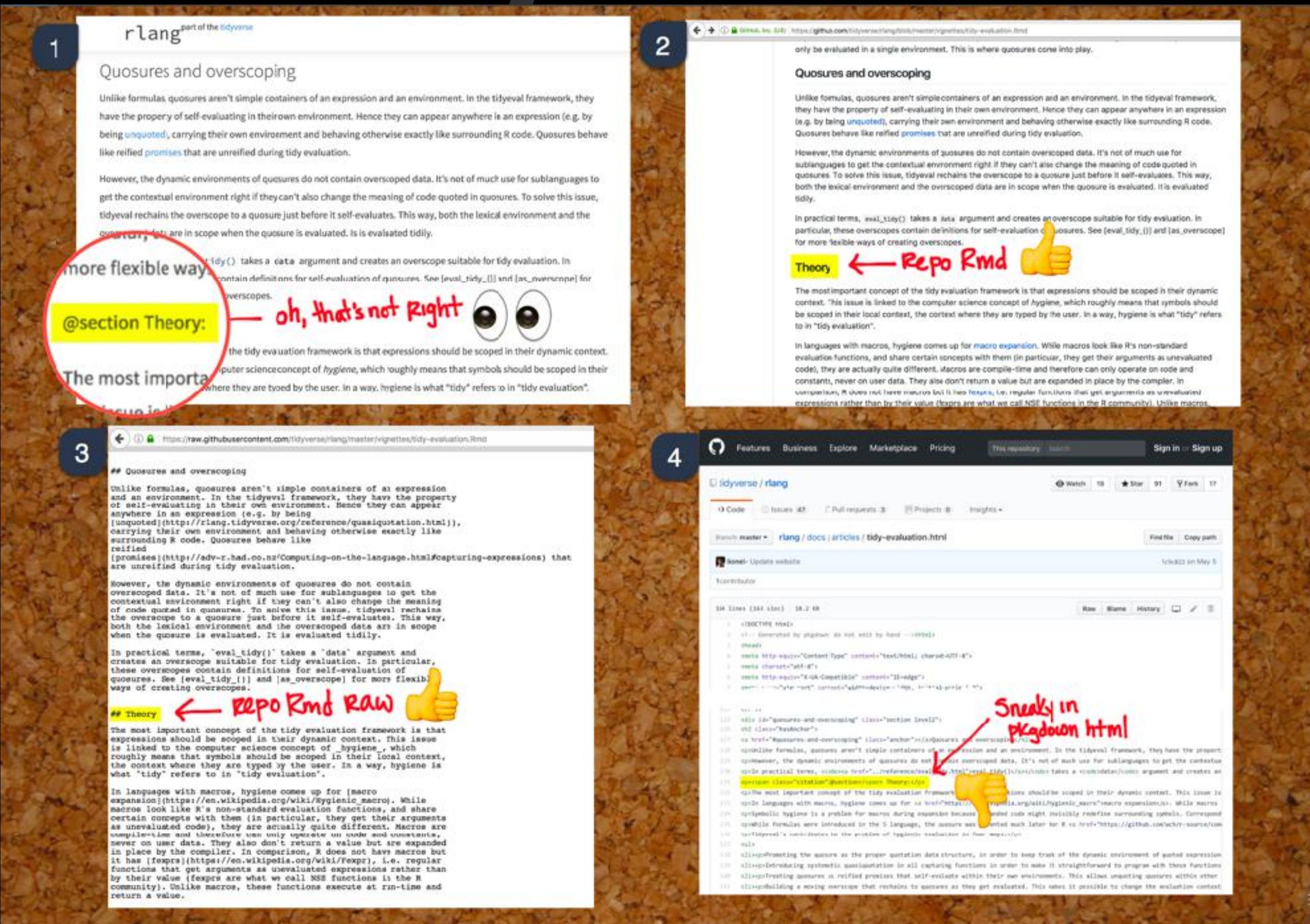
# Not Very Tidy



# Sharing untidy efforts



# Sharing untidy efforts

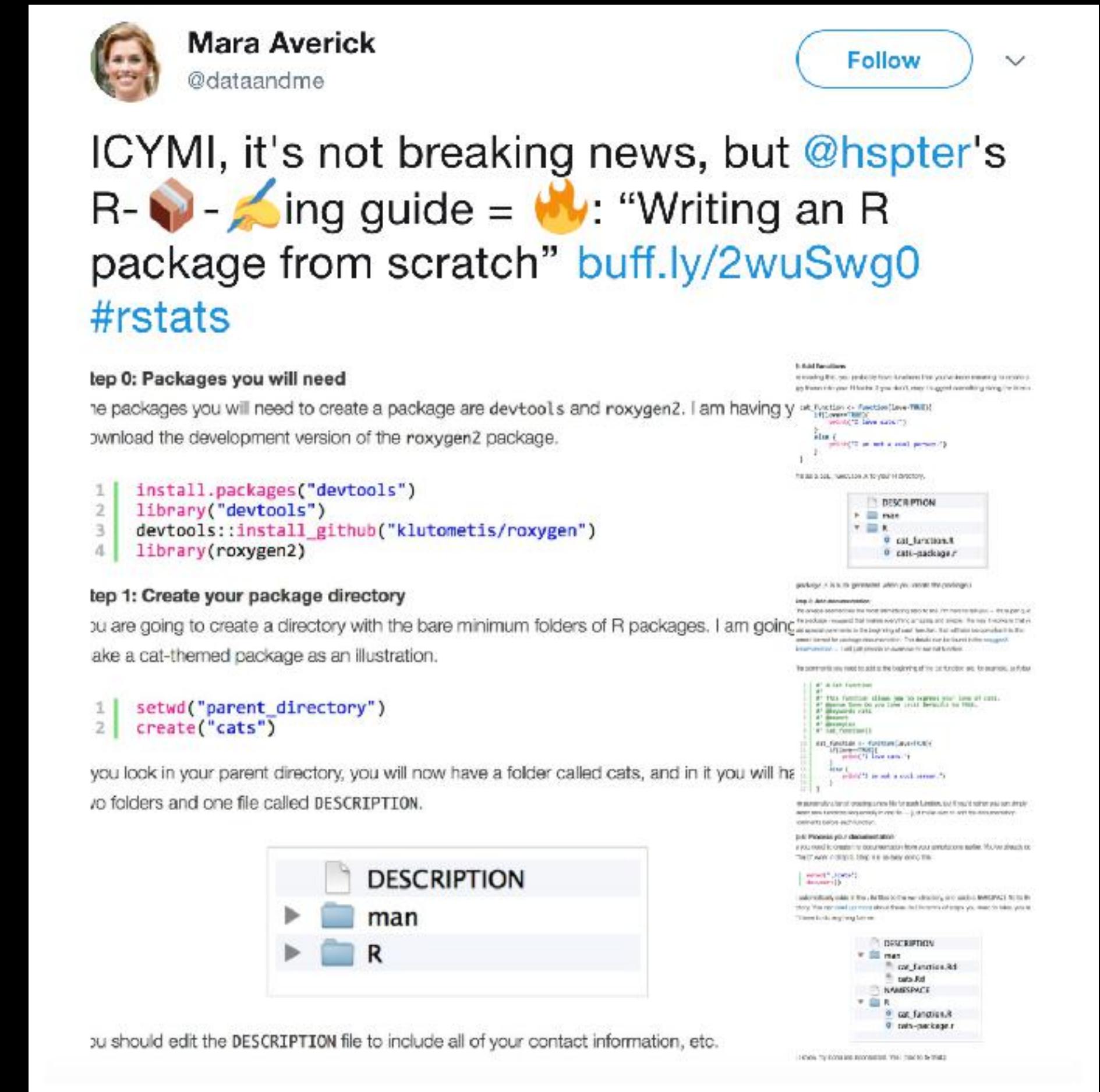




tradecraft

# other people's content...

# not necessarily new



# tweet anatomy



**Your Name**

@yourhandle

Cool tutorial using {**package-name**}:  
"Title of the tutorial" by **@author\_handle**  
[www.thelinktoit.com](http://www.thelinktoit.com) via **@sitename**  
**#relevanthash tag** **#maybeanother**

but with pretty pictures

# tweet anatomy

Mara Averick  
@dataandme

Follow

Another 🌟 chart resource: "ggplot2 extensions" (w/ more than shown) maintained by @Emaasit [buff.ly/2eJI4HA](https://buff.ly/2eJI4HA)  
#rstats #ggplot2 #dataviz

ggplot2 extensions - gallery

ggQC • [View](#)  
ggQC is used for single, faceted and multi-layered quality control charts.  
• author: Keith Gray  
• tags: ggplot2, kbase, Bioconductor, visualization, gglibaries

ggedit • [View](#)  
ggedit is aimed at interactively editing ggplot layers, scales and themes aesthetics.  
• author: josed  
• tags: visualization, interactive, shiny, general, gglibaries

ggforce • [View](#)  
ggforce is aimed at providing missing functionality to ggplot2 through the extension system introduced with ggplot2 v2.0.0.  
• author: thomaspk  
• tags: visualization, interactive, general, gglibaries

ggalt • [View](#)  
A complement of 'ggplot2', 'gridExtra' and 'grid' for 'ggalt'.  
• author: thomaspk  
• tags: visualization, general, gglibaries

ggraph • [View](#)  
A technique to make ggplot graphics interactive.  
• author: svanderwaag  
• tags: visualization, general, gglibaries

ggmluer • [View](#)  
Creates radial plots for visualizing evolutionary dynamics.  
• author: probostable  
• tags: visualization, evolution, dynamics, gglibaries

ggstance • [View](#)  
ggstance implements horizontal versions of common ggplot2 geoms.  
• author: Benoit  
• tags: visualization, general, gglibaries

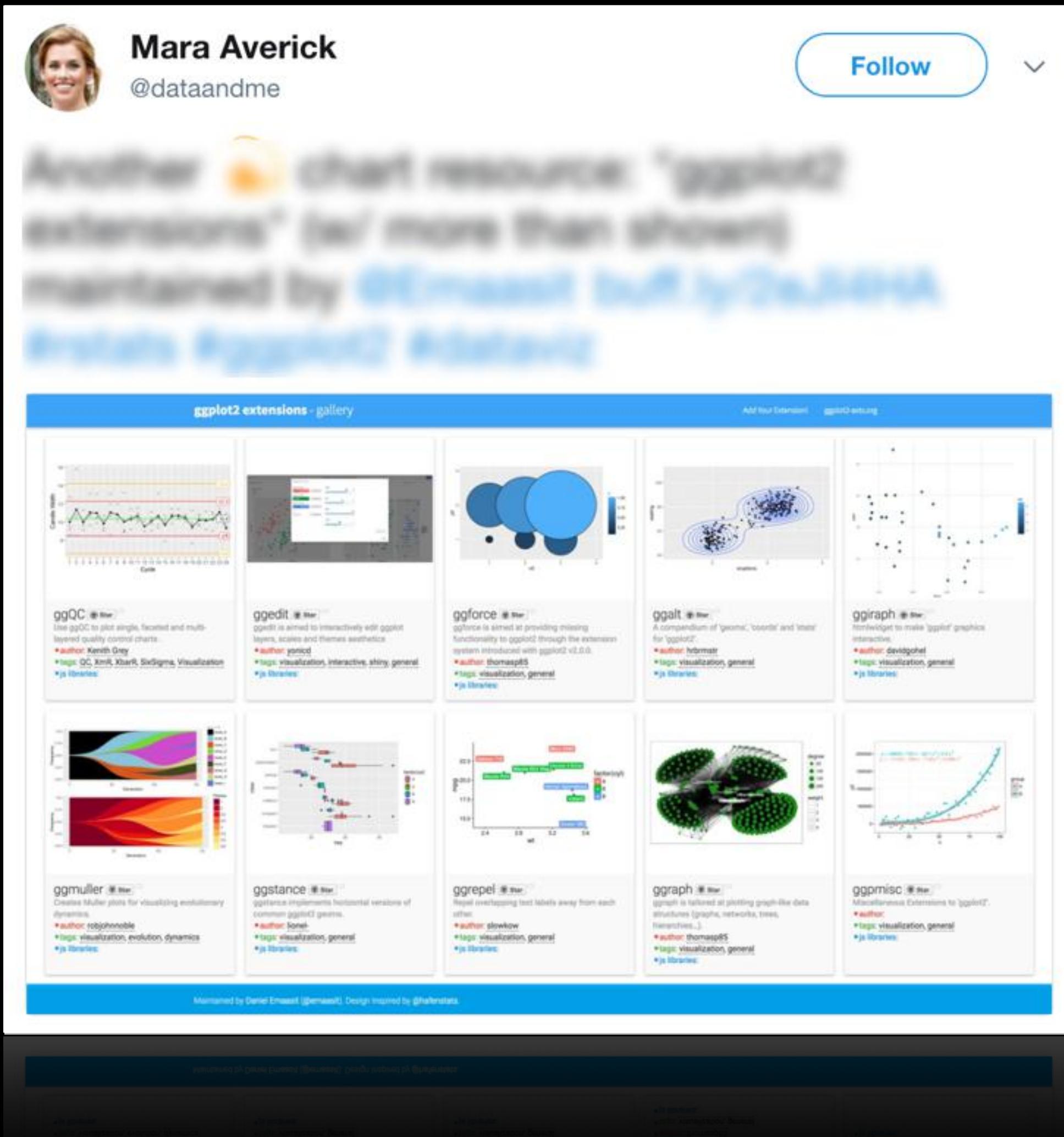
ggrepel • [View](#)  
Repel overlapping text labels away from each other.  
• author: jennybc  
• tags: visualization, general, gglibaries

ggraph • [View](#)  
A technique to render or plot graph-like data structures (graphs, networks, trees, hierarchies...).  
• author: thomaspk  
• tags: visualization, general, gglibaries

gpmrse • [View](#)  
Miscellaneous Extensions to ggplot2.  
• author:  
• tags: visualization, general, gglibaries

Maintained by Dale Ernest (Emaasit). Design inspired by @hadleywickham.

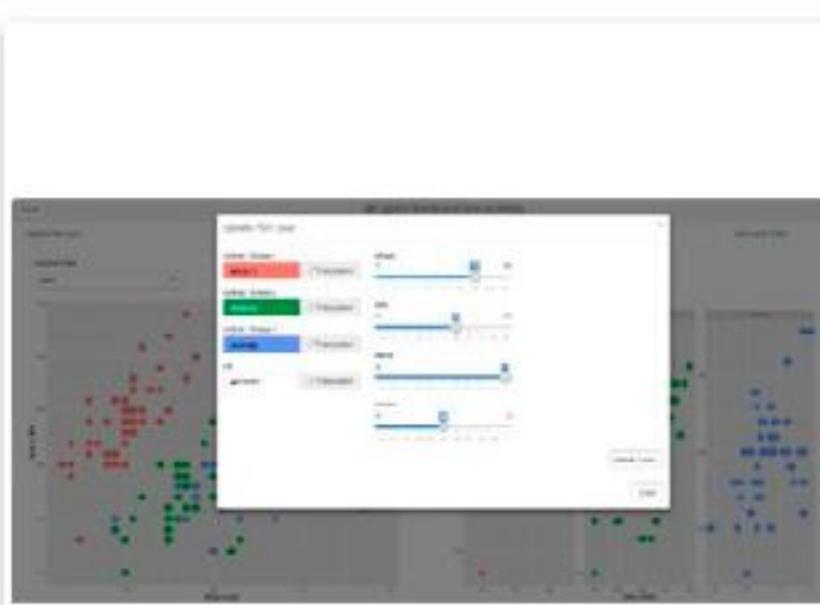
# tweet anatomy



**ggQC**

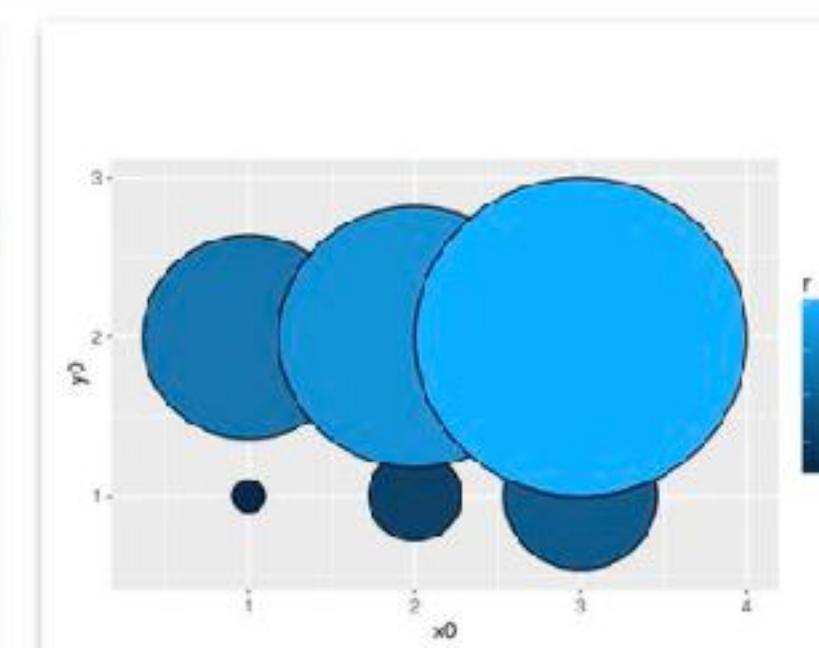
Use ggQC to plot single, faceted and multi-layered quality control charts .

- **author:** Kenneth Grey
- **tags:** QC, XmR, XbarR, SixSigma, Visualization
- **js libraries:**

**ggedit**

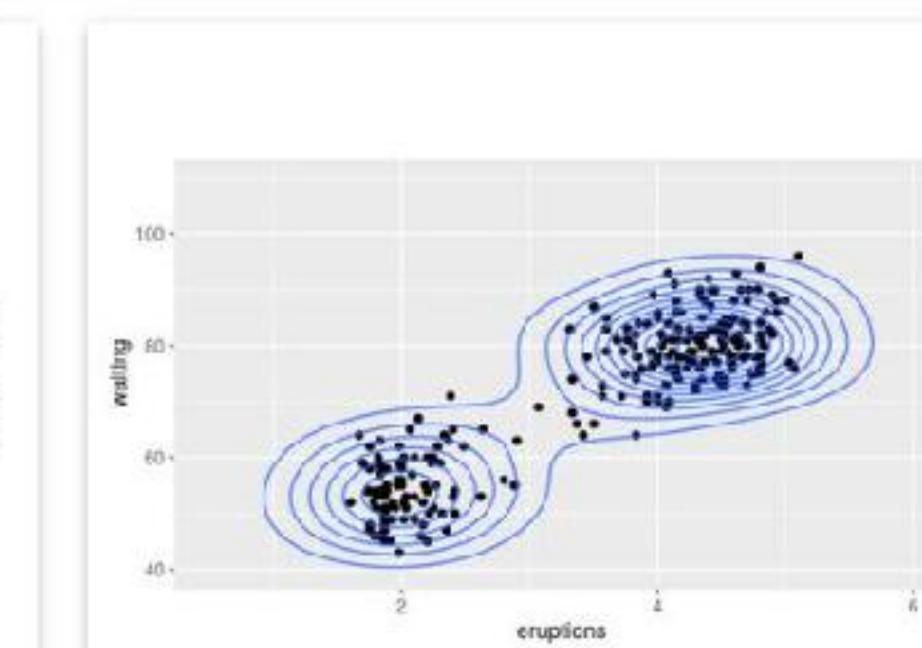
ggedit is aimed to interactively edit ggplot layers, scales and themes aesthetics

- **author:** yonied
- **tags:** visualization, interactive, shiny, general
- **js libraries:**

**ggforce**

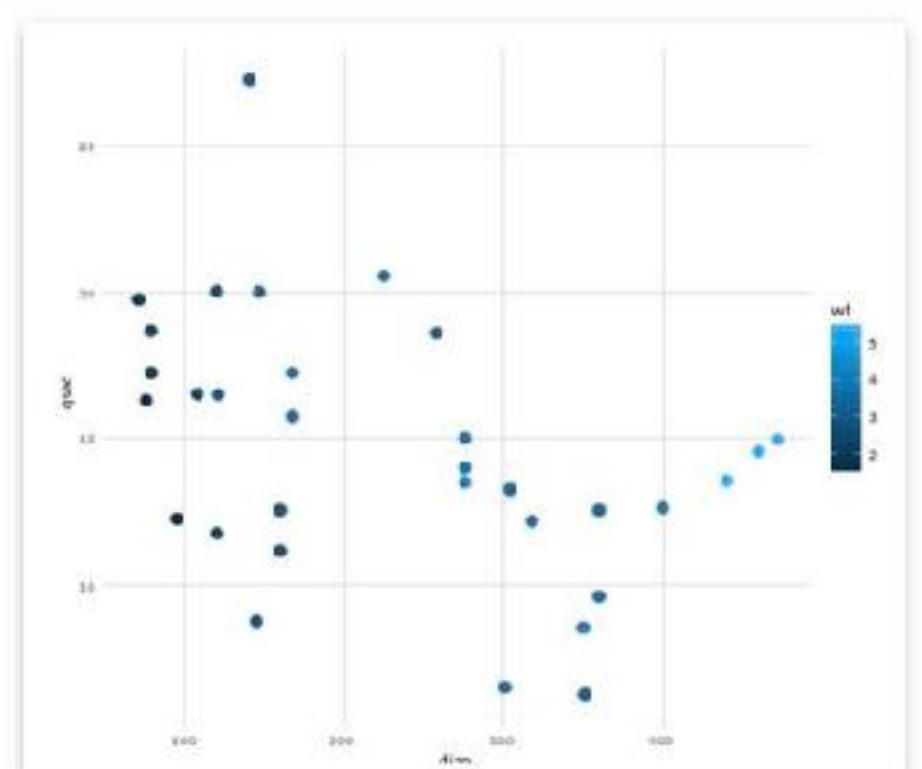
ggforce is aimed at providing missing functionality to ggplot2 through the extension system introduced with ggplot2 v2.0.0.

- **author:** thomasp85
- **tags:** visualization, general
- **js libraries:**

**ggalt**

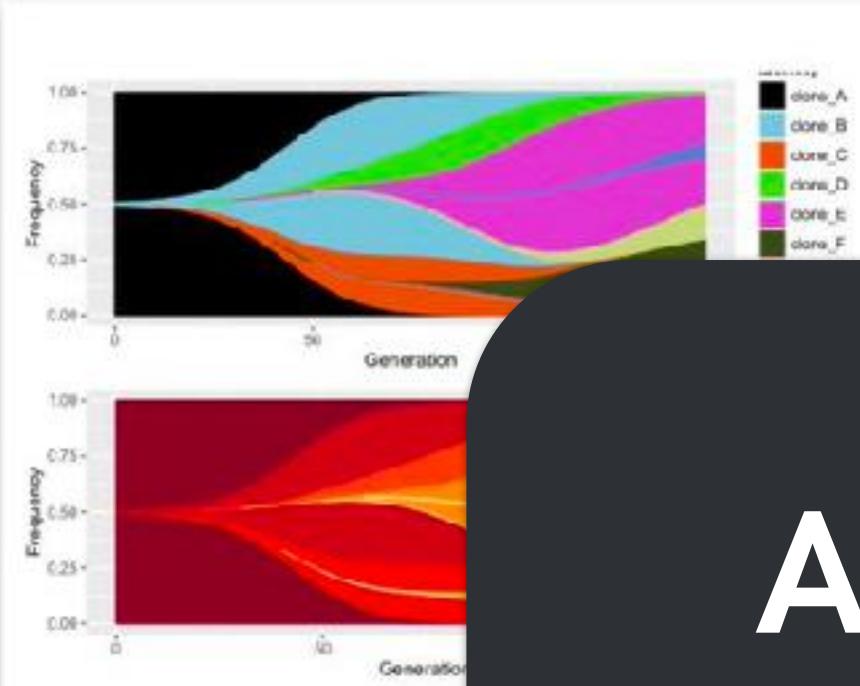
A compendium of 'geoms', 'coords' and 'stats' for 'ggplot2'.

- **author:** hrbrmstr
- **tags:** visualization, general
- **js libraries:**

**ggiraph**

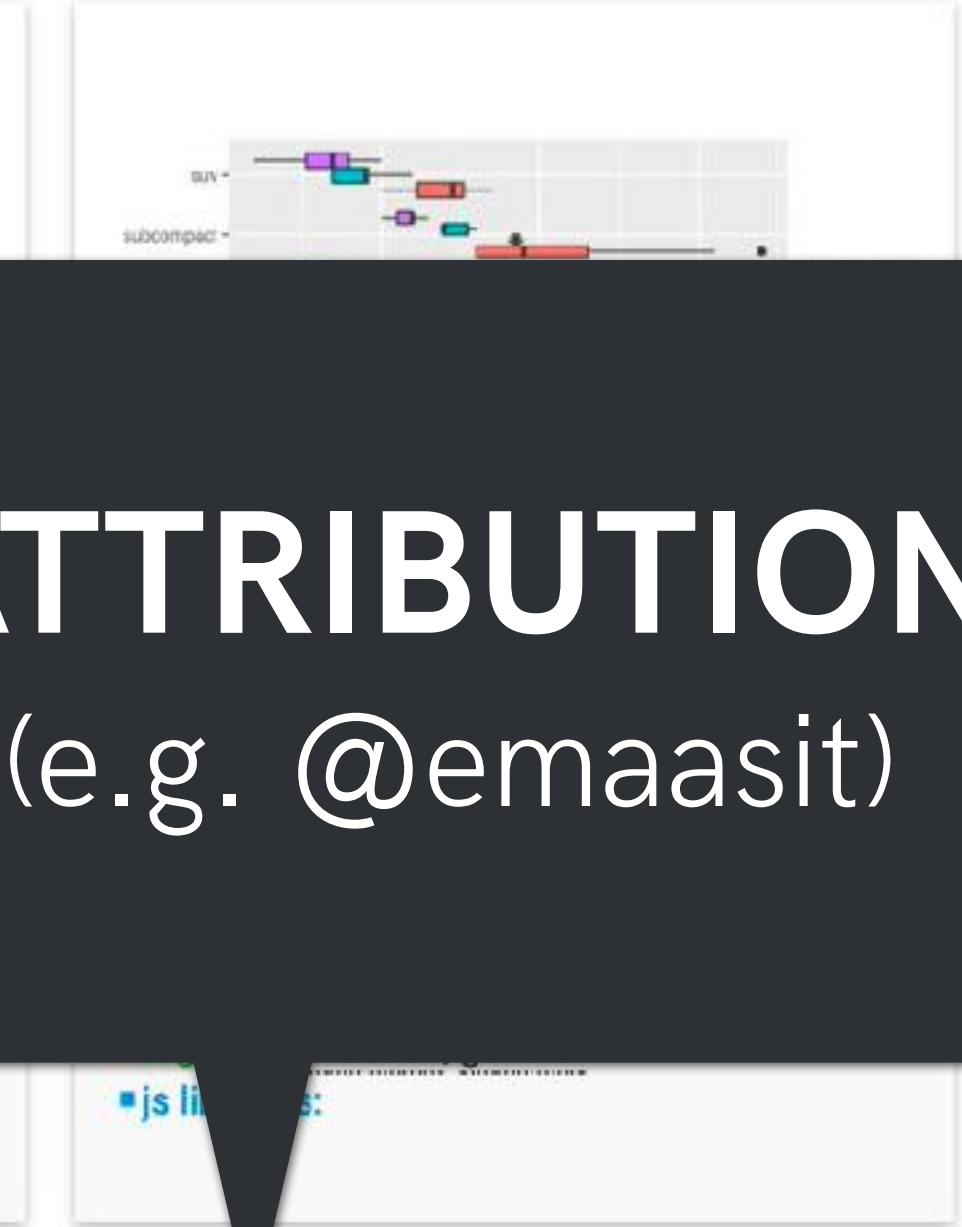
htmlwidget to make 'ggplot' graphics interactive.

- **author:** davidgohel
- **tags:** visualization, general
- **js libraries:**

**ggmuller**

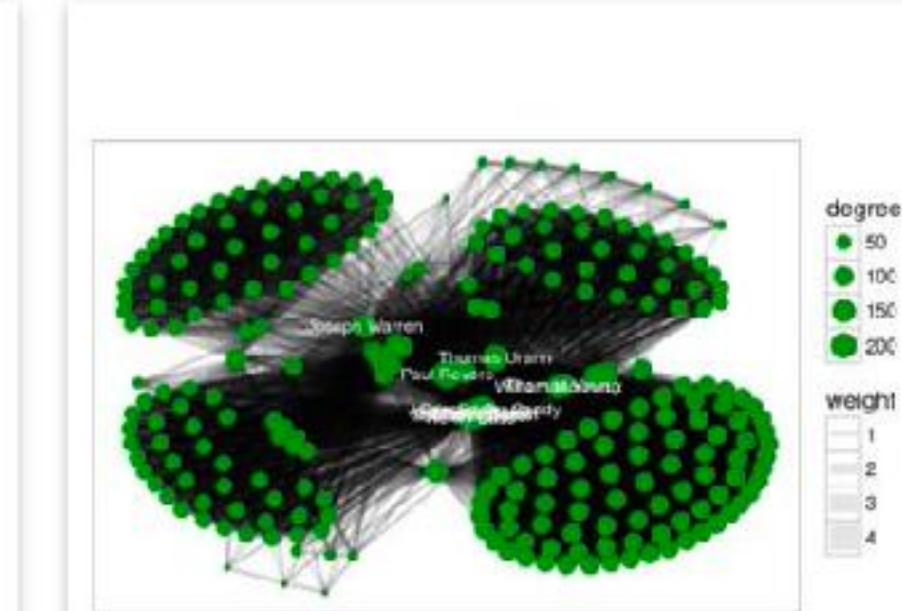
Creates Muller plots for visualizing dynamics.

- **author:** robjohnnoble
- **tags:** visualization, evolution, dynamics
- **js libraries:**

**ggtext**

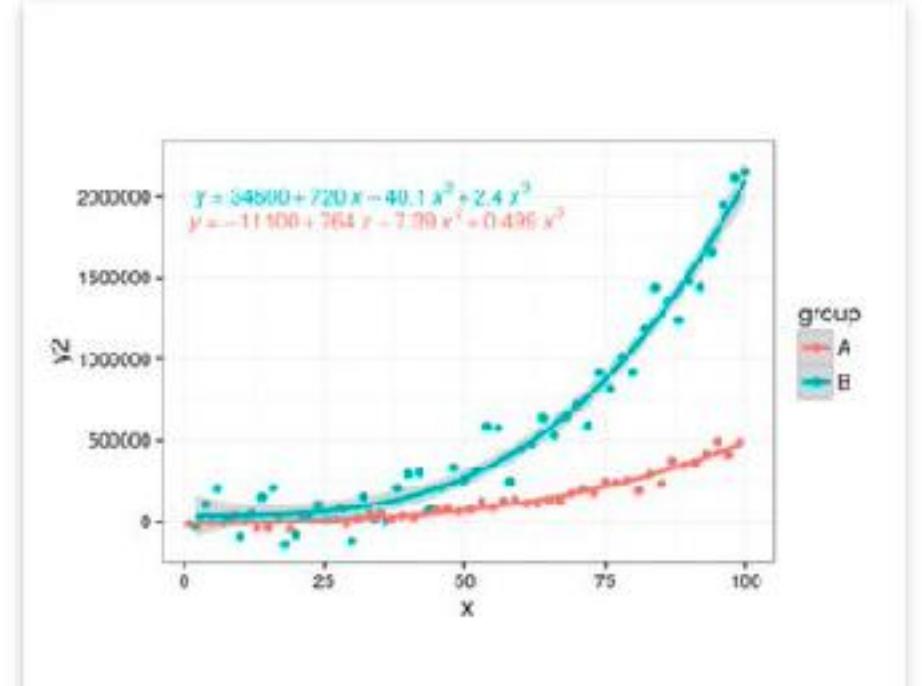
Plotting text labels away from each other without overlap.

- **author:** hadley
- **tags:** visualization, general
- **js libraries:**

**ggraph**

ggraph is tailored at plotting graph-like data structures (graphs, networks, trees, hierarchies...).

- **author:** thomasp85
- **tags:** visualization, general
- **js libraries:**

**ggpmisc**

Miscellaneous Extensions to 'ggplot2'.

- **author:**
- **tags:** visualization, general
- **js libraries:**

# ATTRIBUTION

(e.g. @emaasit)

# the tweet != the content

Mara Averick  
@dataandme

💡 thought I knew the ↔ of readxl greatness, & then I 💬 @JennyBryan's "readxl webinar"... 🚨🏆 (slides, too) [buff.ly/2uwo16M](http://buff.ly/2uwo16M) #rstats

The image shows a Twitter post from Mara Averick (@dataandme) featuring a tweet from Jennifer Bryan (@JennyBryan) and several supporting images. The tweet reads: "💡 thought I knew the ↔ of readxl greatness, & then I 💬 @JennyBryan's \"readxl webinar\"... 🚨🏆 (slides, too) [buff.ly/2uwo16M](http://buff.ly/2uwo16M) #rstats". The supporting images include a corkboard background, the 'readxl' hex logo, a flowchart of the 'readxl' workflow, and a slide from Jennifer Bryan's webinar titled '1. readxl is all about the data rectangle'.

# the tweet != the content



← embrace "directed forgetting"

# tweet Rs

- Read (or run)
- Relatable
- Retrievable
- Relevant
- Real

you do you 

\* completely arbitrary rules I made up

# you never know...

 **Hadley Wickham** • @hadleywickham • 18 Nov 2017  
If you're using databases with shiny, make sure to check out the pool 🍃:  
[blog.rstudio.com/2017/11/17/poo...](http://blog.rstudio.com/2017/11/17/poo...) Will make your life much easier! #rstats

7 120 395

 **Mara Averick** • @dataandme • 18 Nov 2017  
Never forget to check the pool...



2 19

# you never know...



**Matthew Dickinson**

@midickinson

Follow



Replies to [@dataandme](#) [@hadleywickham](#)

If it weren't for the gif I would have totally just skipped over knowing about this package that looks to be a huge benefit to me.  
Thanks!

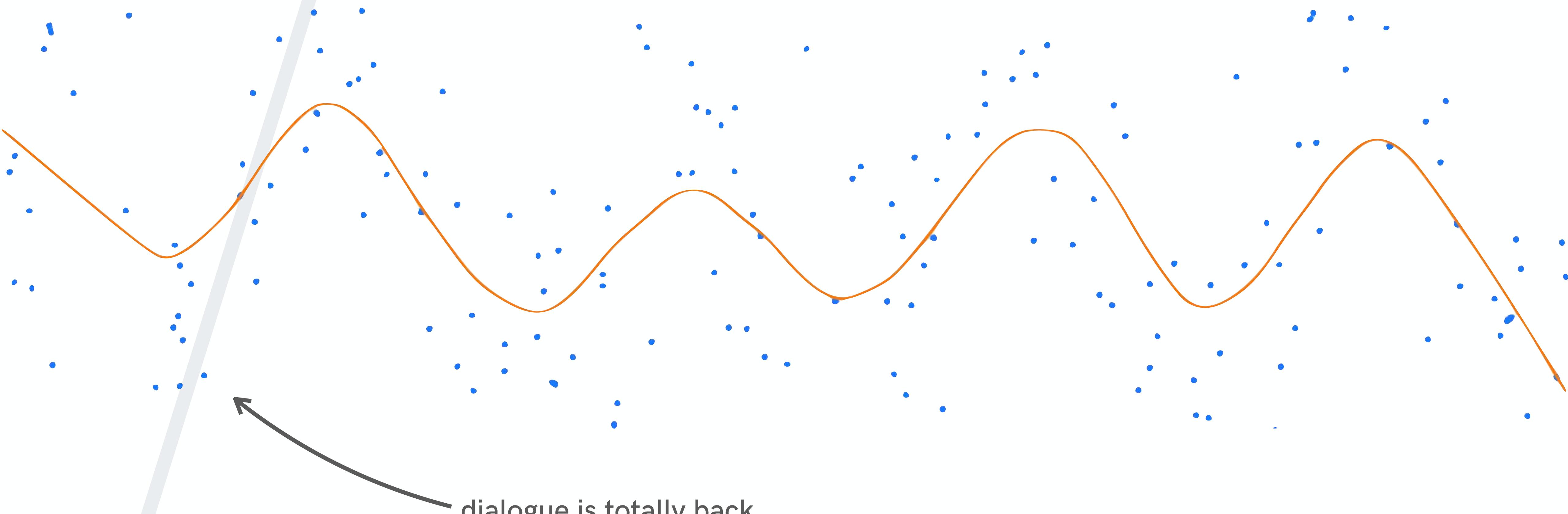
6:13 PM - 18 Nov 2017

# leaRning out loud



OMG, I just learned a thing!

# leaRning out loud



dialogue is totally back...

Roger Peng and Hilary Parker are the new Boethius and Lady Philosophy  
(check out *Not So Standard Deviations*)

# The newcomer's paradox...



*When you ask for help, some friendly soul will no doubt tell you that "it's easy, just do foo, bar and baz." Except for you, it is not easy, there may be no documentation for foo, bar is not doing what it is supposed to be doing and what is this baz thing anyway with its eight disambiguation entries on Wikipedia?*

— Leslie Hawthorne

"You'll Eventually Know Everything They've Forgotten." In *Open Advice: FOSS: What We Wish We Had Known When We Started*, edited by Lydia Pintscher, 29–32.

# Contributing to FOSS

## WHAT HOLDS PEOPLE BACK?

- “*I can't write code.*”



# Contributing to FOSS

## WHAT HOLDS PEOPLE BACK?

- “*I can't write code.*”
- “*I'm not really good at this.*”
- “*I'd just be a burden.*”



# Contributing to FOSS

## WHAT HOLDS PEOPLE BACK?

- “*I can't write code.*”
- “*I'm not really good at this.*”
- “*I'd just be a burden.*”
- “*They already have enough people smarter than me.*”



# Key FOSS actors

- maintainers
- contributors
- consumers/users
- sustainers

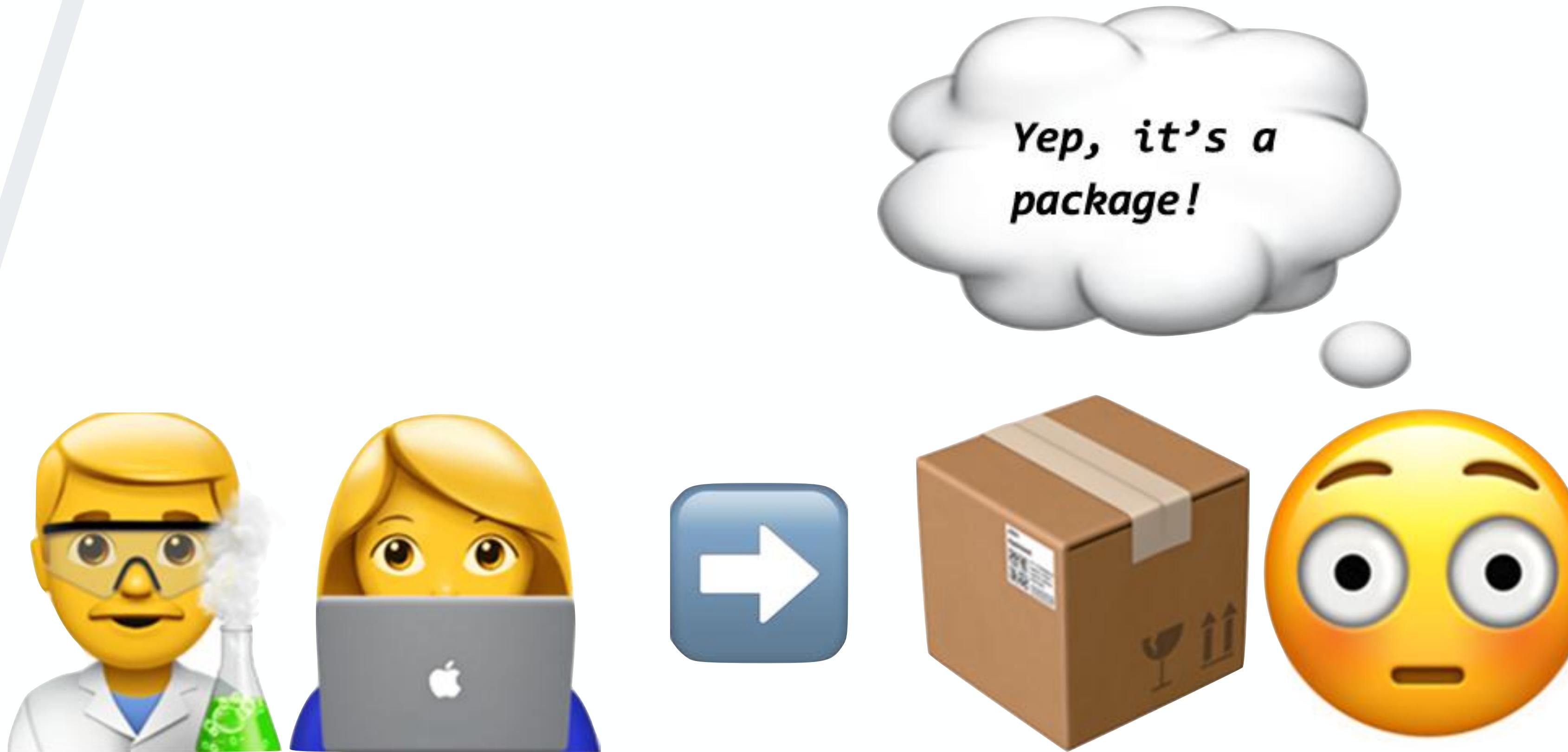
GitHub HQ (SF) | June 19, 2017

# Sustain

A ONE DAY CONVERSATION  
for Open Source Software sustainers

## THE REPORT

# require(n00bs)



# require(n00bs)

When assigning reviewers to a submission, we aim to pair experienced reviewers with new ones, or reviewers with expertise on a package's programming methods with those experienced in its field of application.



## **How rOpenSci uses Code Review to Promote Reproducible Science**

[ropensci.org](http://ropensci.org) highlighted with Highly

# Contribute documentation

“Innocence lost is not easily regained. The designer simply cannot predict the problems people will have, the misinterpretations that will arise, and the errors that will get made.”

— Donald Norman, *The Design of Everyday Things*

# the thoughtful user



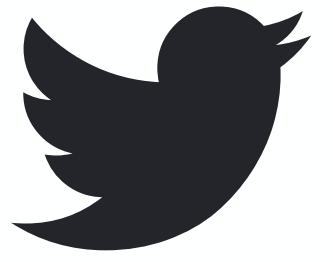
# asking questions

*The most useless problem statement that one can face is "it doesn't work", yet we seem to get it far too often.*

- Thiago Maciera



# where to ask



Twitter

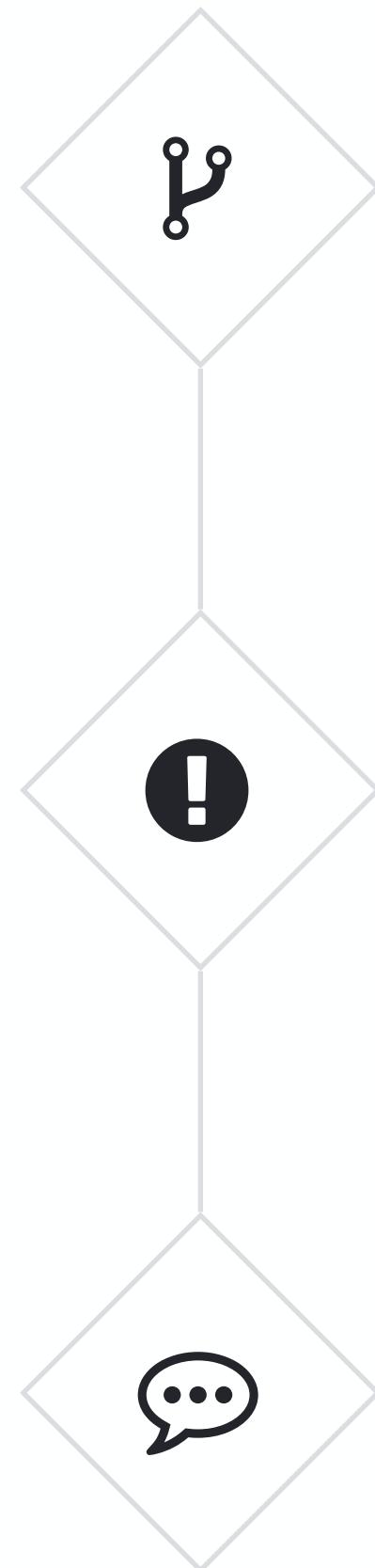


RStudio Community



StackOverflow

# Ways to contribute



## PULL REQUESTS

Contributing code/making fixes.

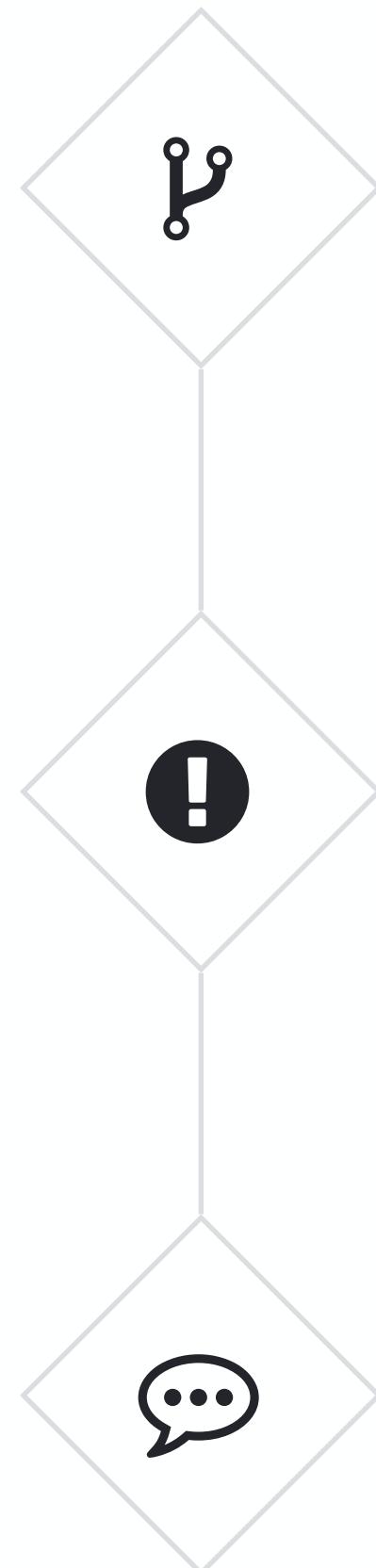
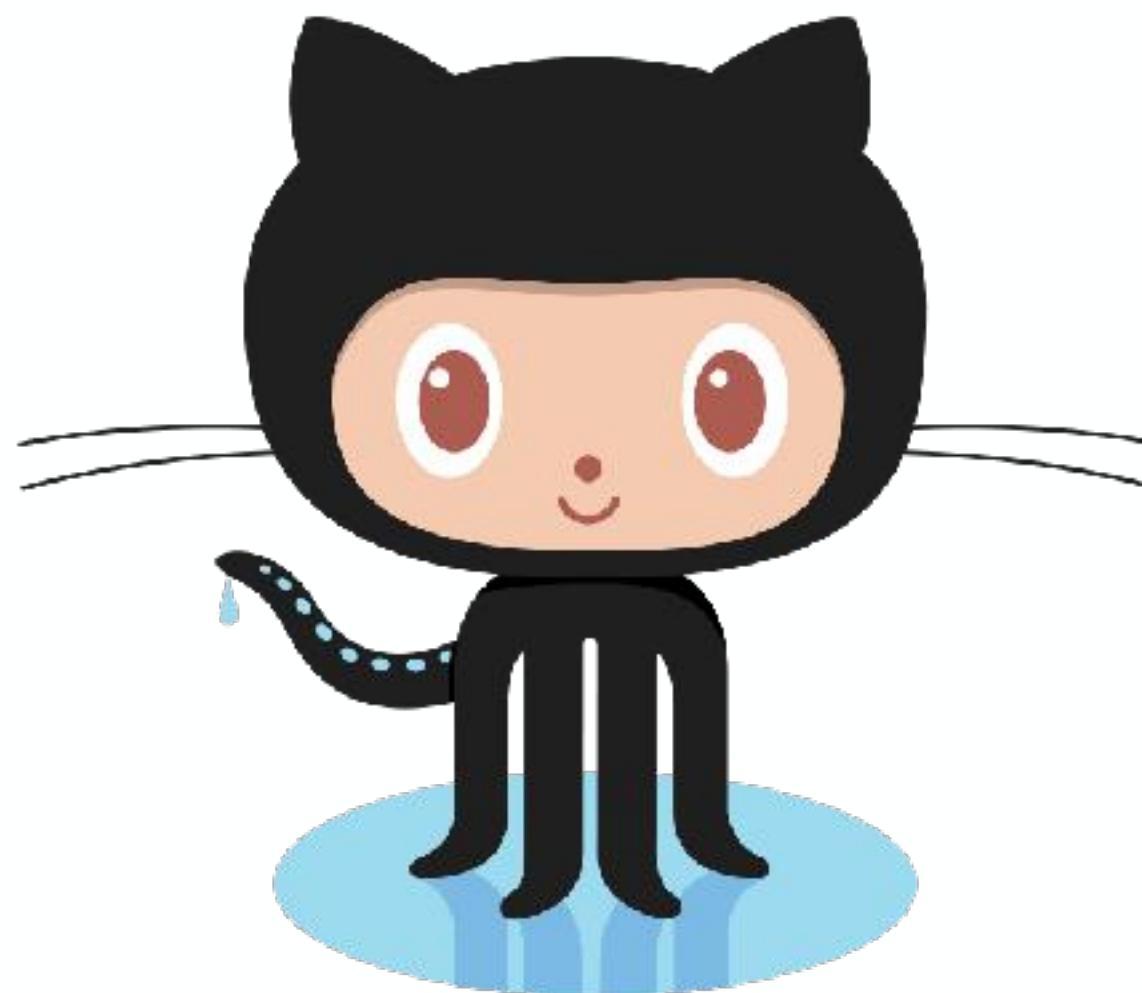
## ISSUES

Identifying a problem, trying your best to isolate its source.

## COMMENTS

Help maintainers answer questions, triage issues. Help newcomers learn how to ask better questions (e.g. the art of the reprex).

# Ways to contribute



## PULL REQUESTS

Contributing code/making fixes.

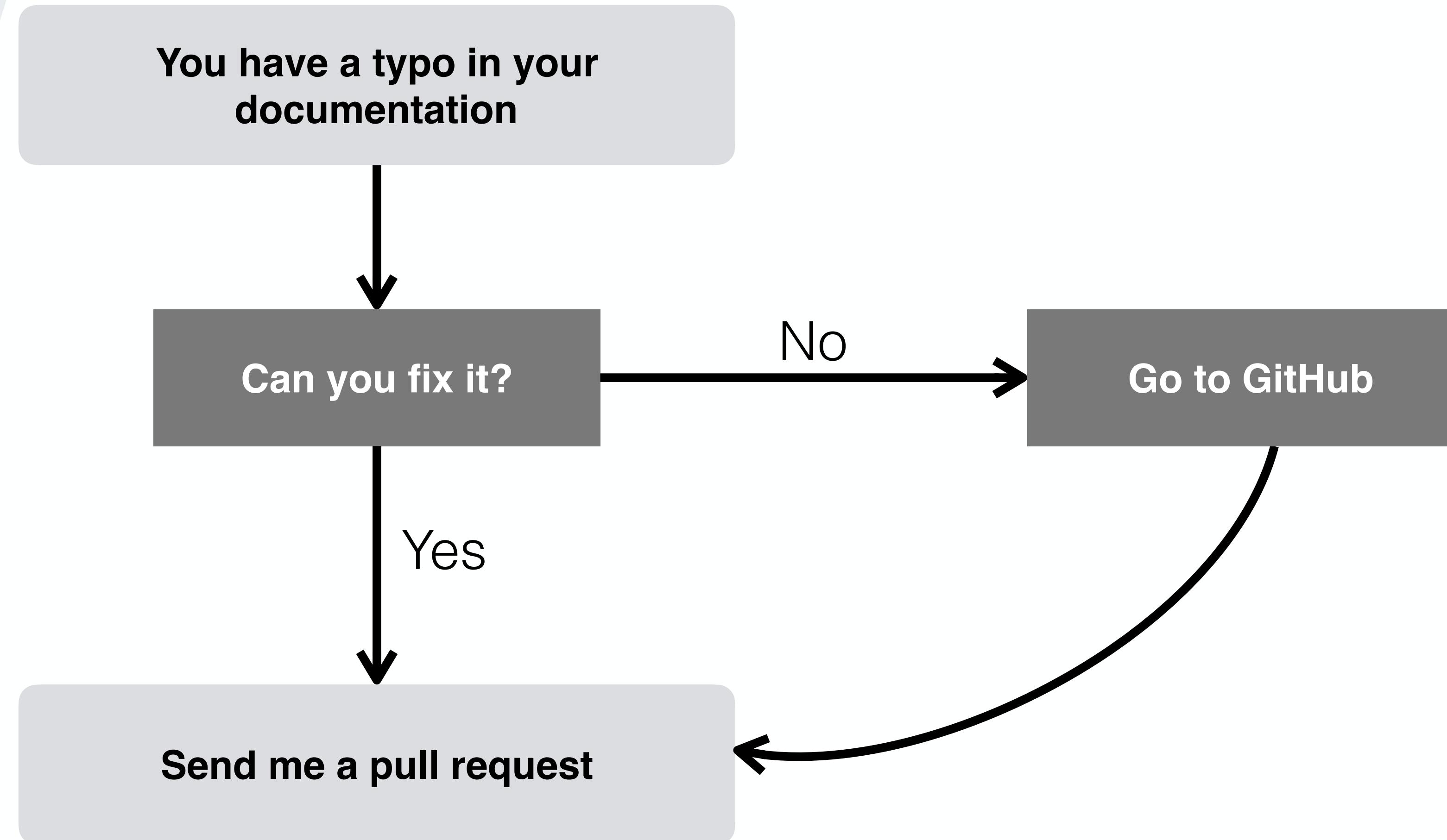
## ISSUES

Identifying a problem, trying your best to isolate its source.

## COMMENTS

Help maintainers answer questions, triage issues. Help newcomers learn how to ask better questions (e.g. the art of the reprex).

# typos



Adapted from: You Do Not Need to Tell Me I Have A Typo in My Documentation by Yihui Xie

# My first “contribution”

hadley / r4ds

Watch ▾ 150 Star 1,072 Fork 1,323

Code Issues 29 Pull requests 35 Projects 0 Wiki Insights

Fixed spelling of purrr on line 121 (#208) Browse files

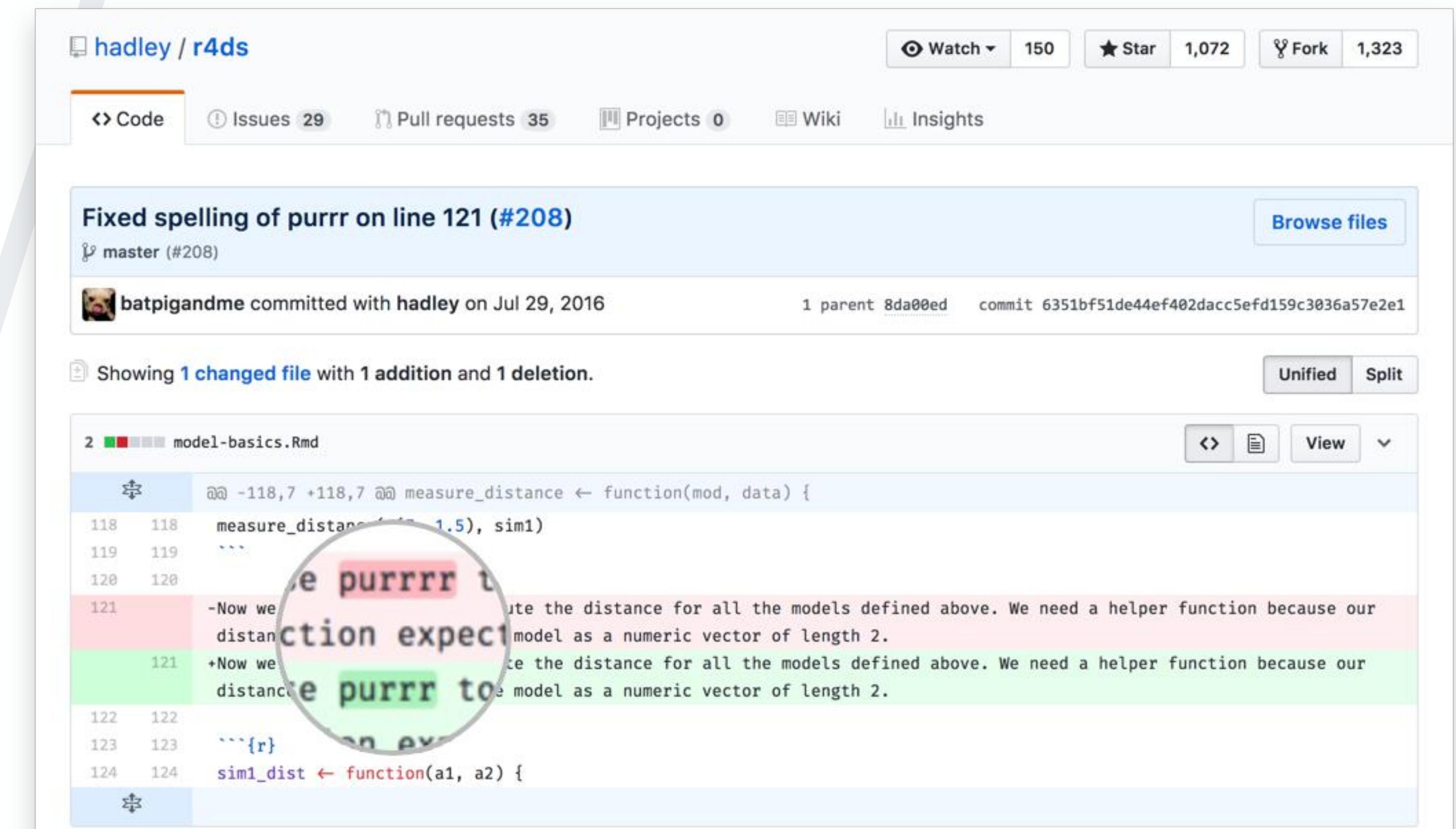
master (#208)

batpigandme committed with hadley on Jul 29, 2016 1 parent 8da00ed commit 6351bf51de44ef402dacc5efd159c3036a57e2e1

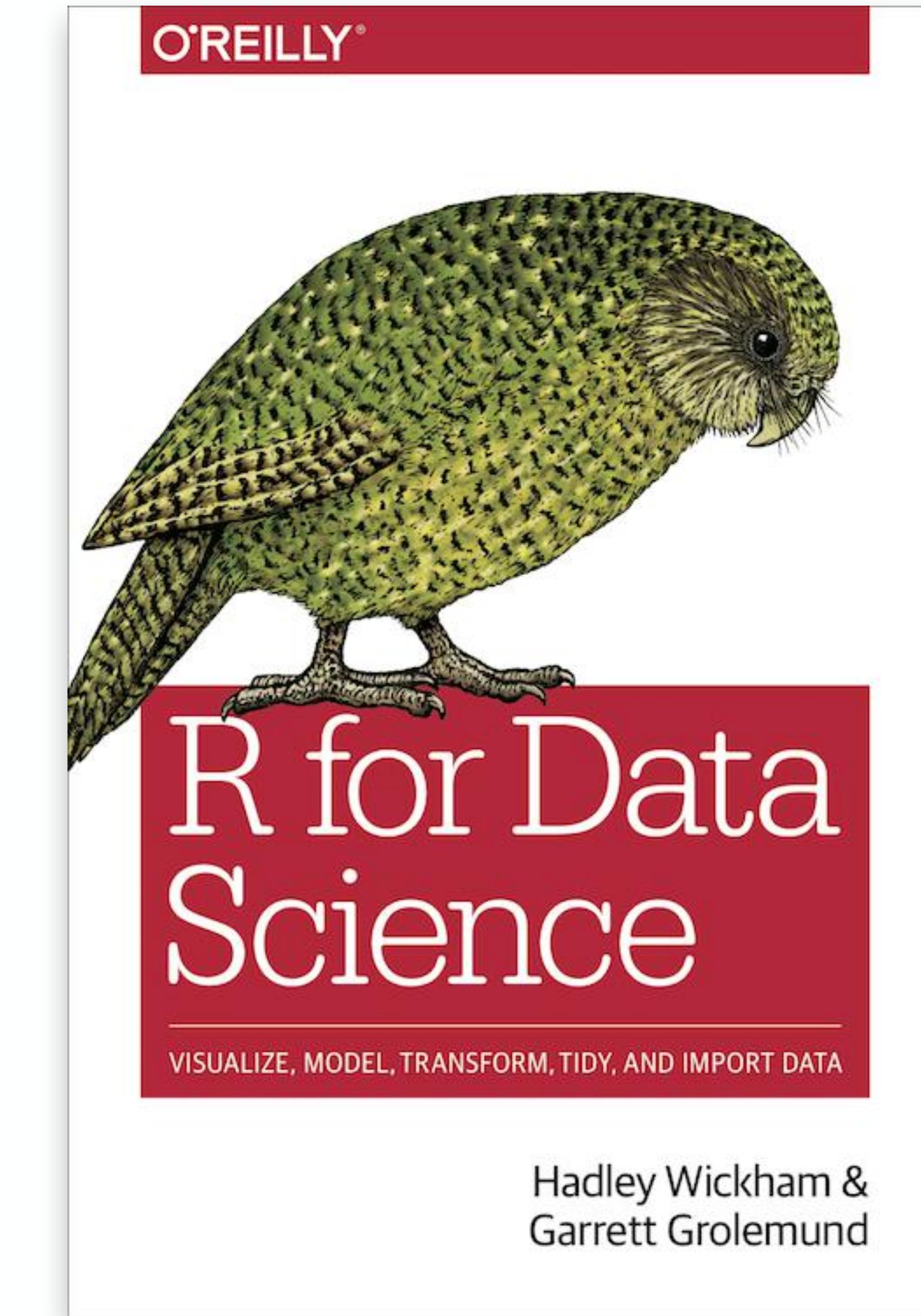
Showing 1 changed file with 1 addition and 1 deletion. Unified Split

model-basics.Rmd

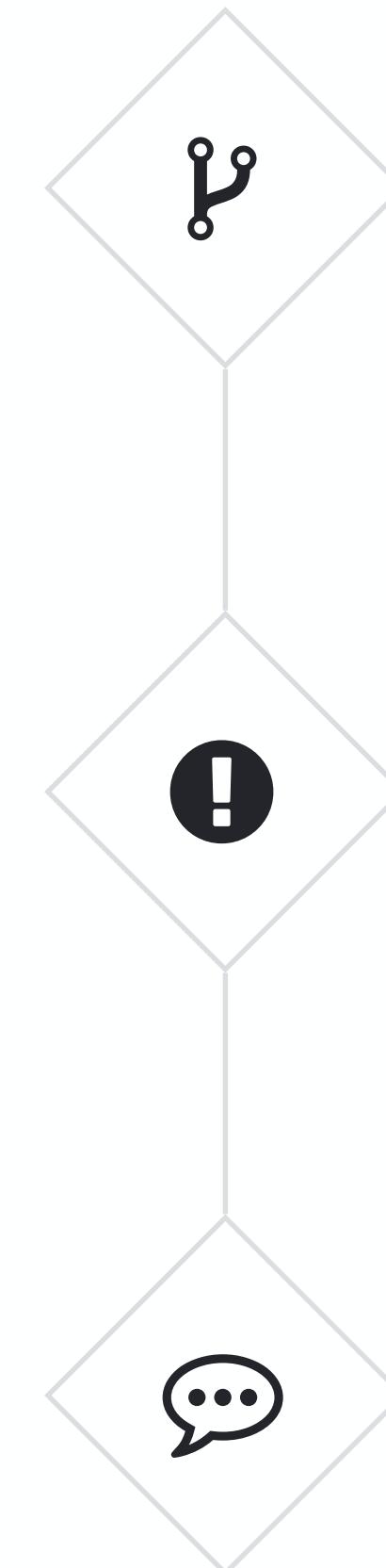
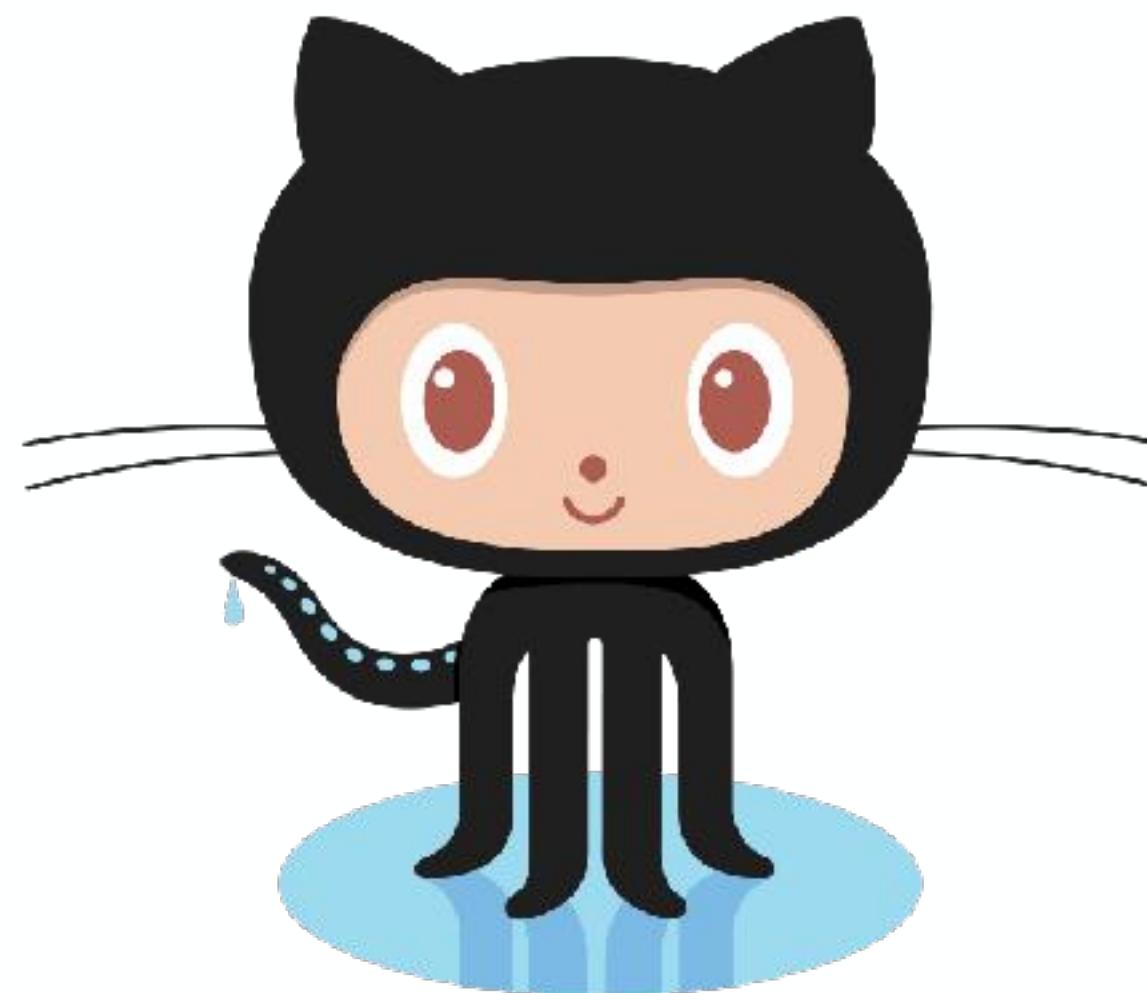
@@ -118,7 +118,7 @@ measure\_distance <- function(mod, data) {  
 118 118 measure\_distance('1.5), sim1)  
 119 119 ...  
 120 120  
 121 -Now we  
 121 +Now we  
 121 -distan  
 121 +distance  
 121 -e purrrr t  
 121 +ction expec  
 121 -e purrr to  
 121 +e purrr to  
 122 122 ...{r}  
 123 123 sim1\_dist <- function(a1, a2) {  
 124 124  
 124 124



# My first “contribution”



# Ways to contribute



## PULL REQUESTS

Contributing code/making fixes.

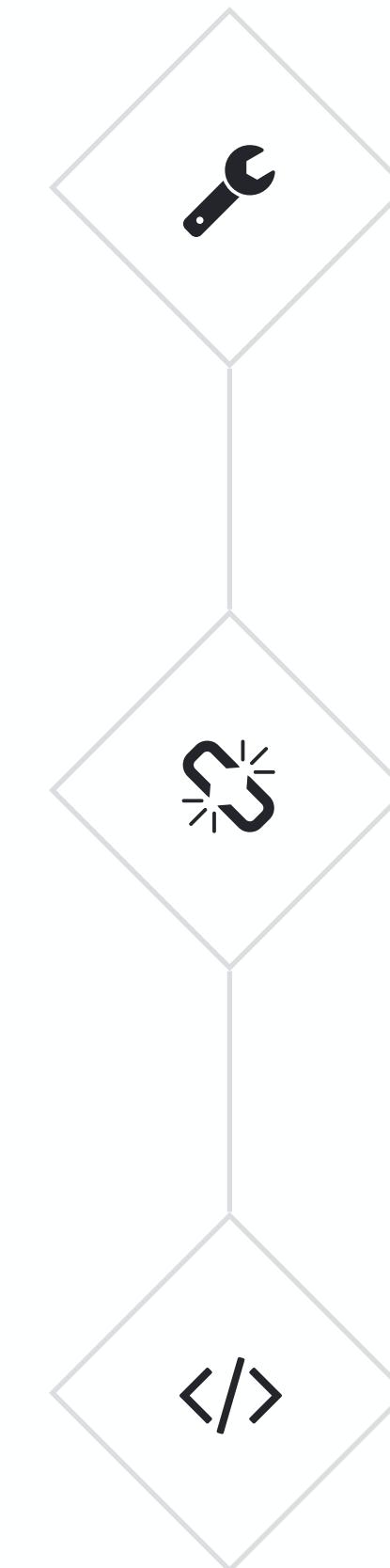
## ISSUES

Identifying a problem, trying your best to isolate its source.

## COMMENTS

Help maintainers answer questions, triage issues. Help newcomers learn how to ask better questions (e.g. the art of the reprex).

# the anatomy of an issue



**PROBLEM DESCRIPTION**

**EXPECTED BEHAVIOUR**

**MINIMAL REPRODUCIBLE EXAMPLE**

# the anatomy of an issue

Please review the [guidelines for contributing](#) to this repository.

 Title

Write Preview AA B i “ “ “ “ @

### Expected behavior

### Actual behavior

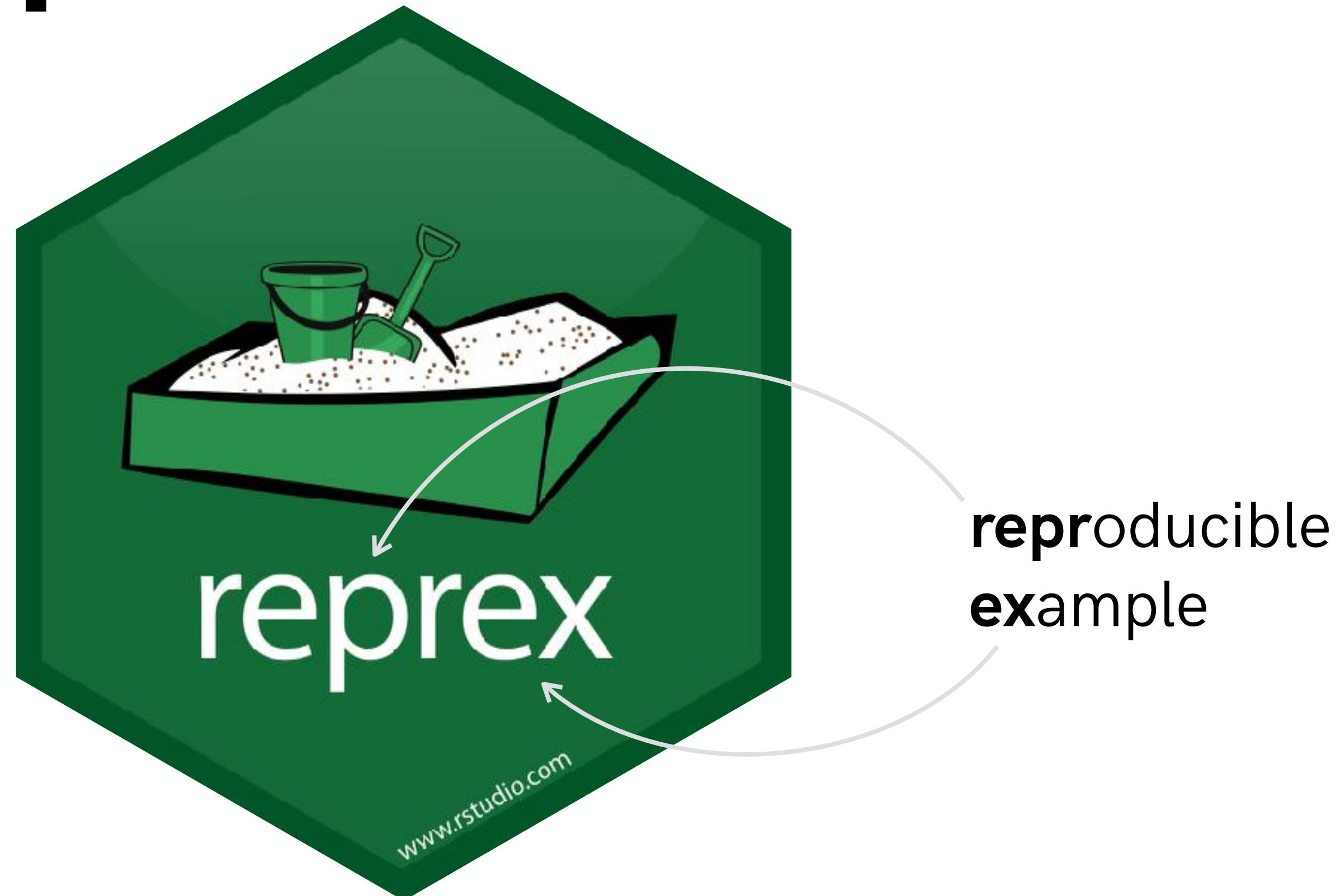
### Steps to reproduce the behavior

Attach files by dragging & dropping, [selecting them](#), or pasting from the clipboard.

MD Styling with Markdown is supported

**Submit new issue**

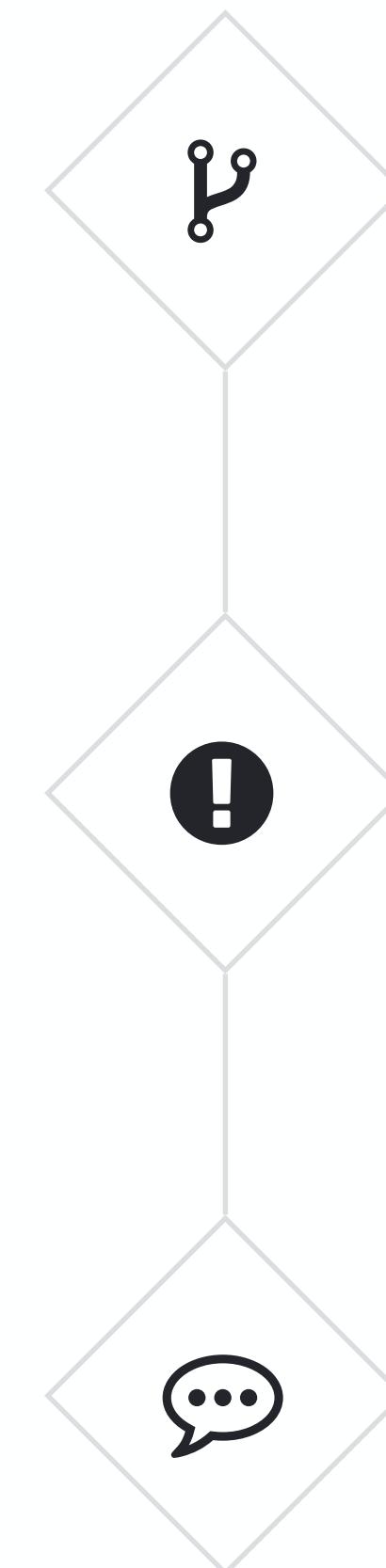
# the magic of reprex



# Keys to reprex-cellence

- ✓ Code that **actually runs**
- ✓ Code that **doesn't have to be run**
- ✓ Code that **can be easily run**

# Ways to contribute



## PULL REQUESTS

Contributing code/making fixes.

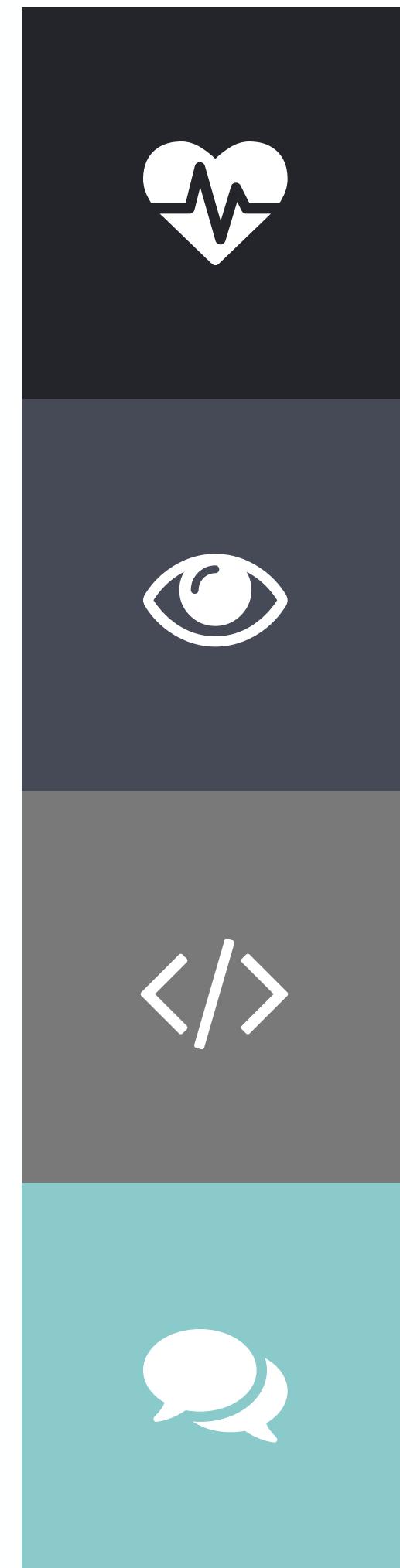
## ISSUES

Identifying a problem, trying your best to isolate its source.

## COMMENTS

Help maintainers answer questions, triage issues. Help newcomers learn how to ask better questions (e.g. the art of the reprex).

# Hints for happy code contribution in the tidyverse



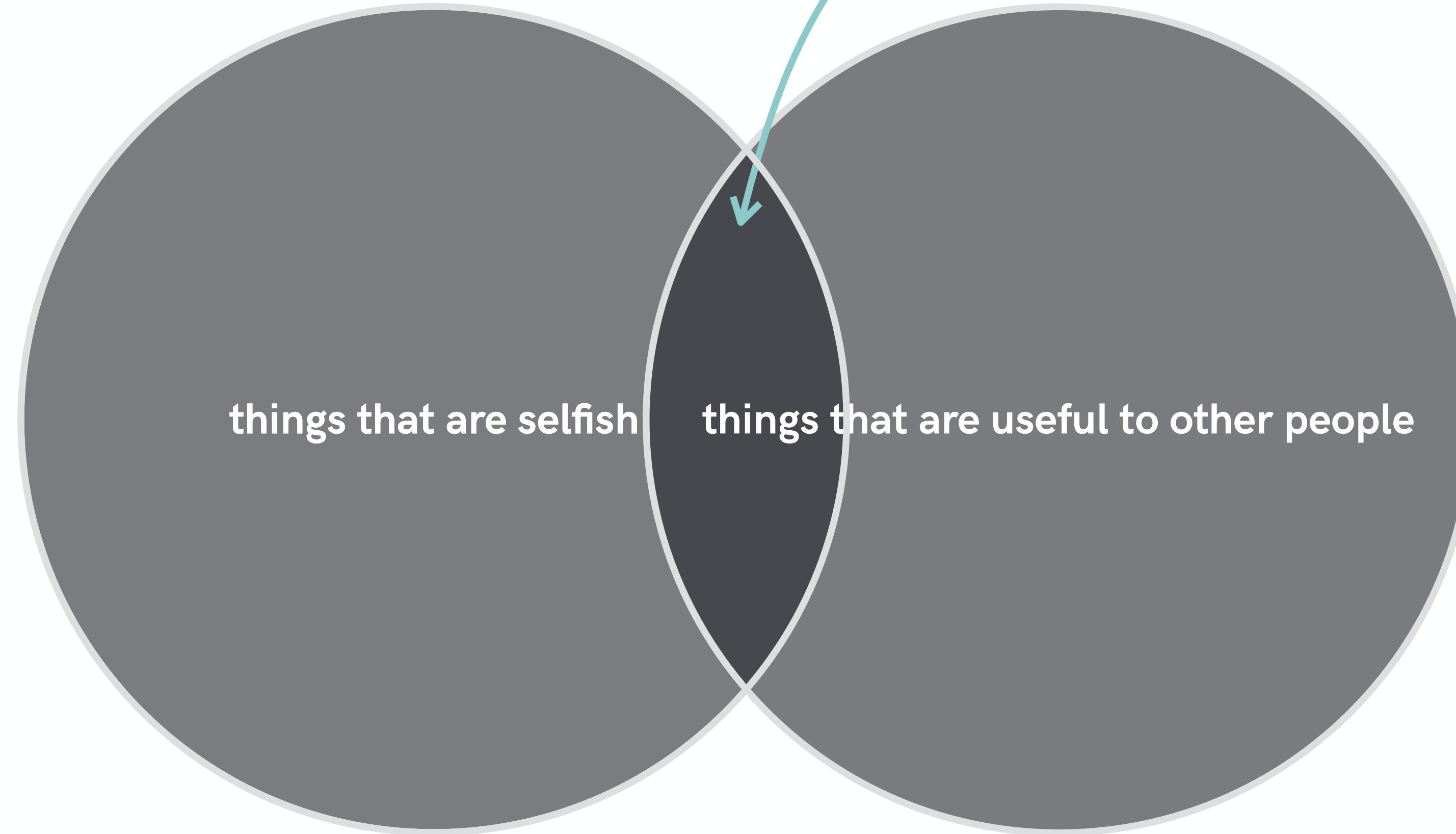
GET THE PULSE OF A PROJECT

WATCH THE REPO

READ THE CODE

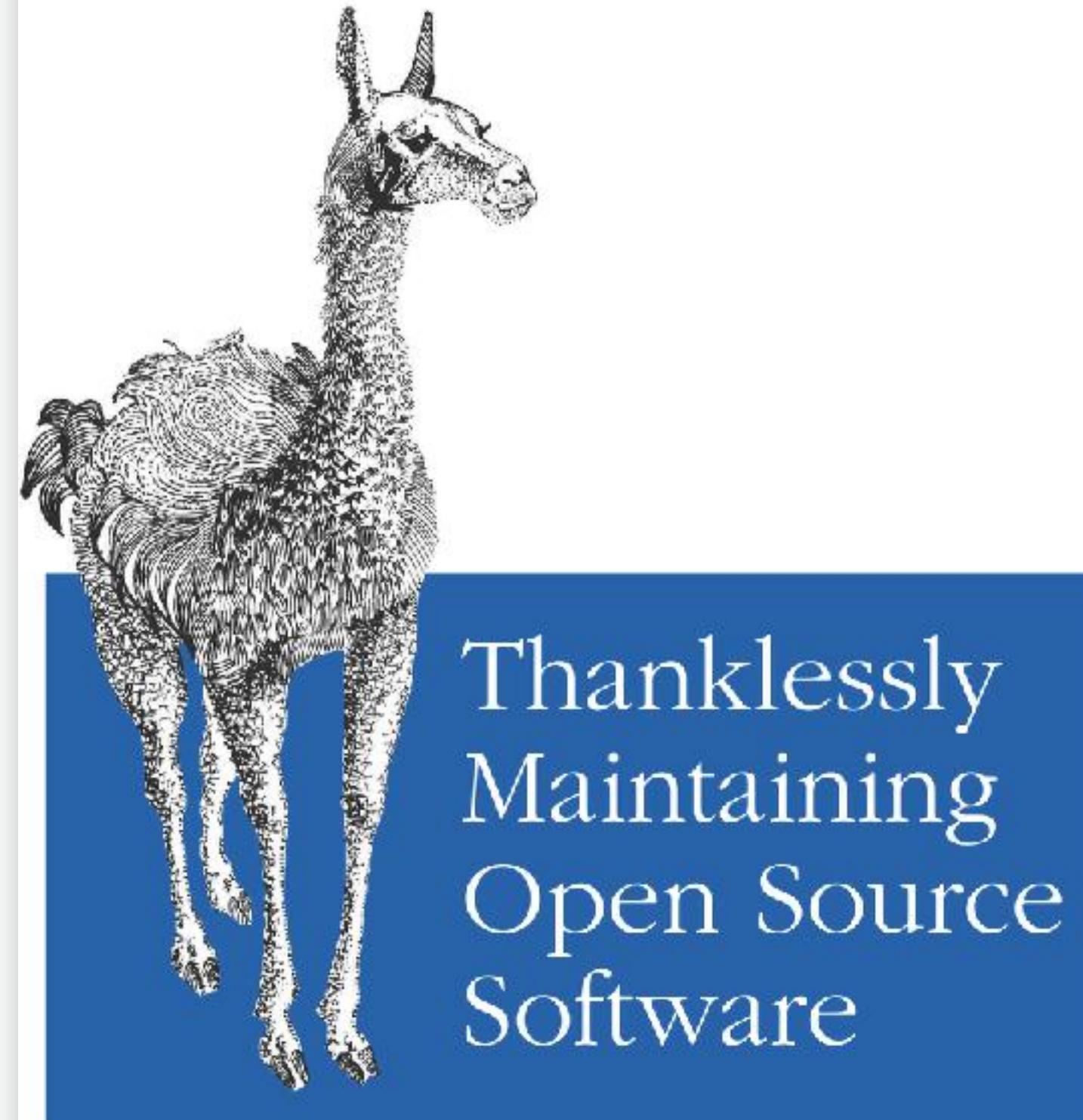
DISCUSS YOUR IDEAS

# wait, but why?



# giving back

*Acting out of the goodness of your heart, or something*



O RLY?

@ThePracticalDev

# need help getting started?

## Save the date: tidyverse developer day

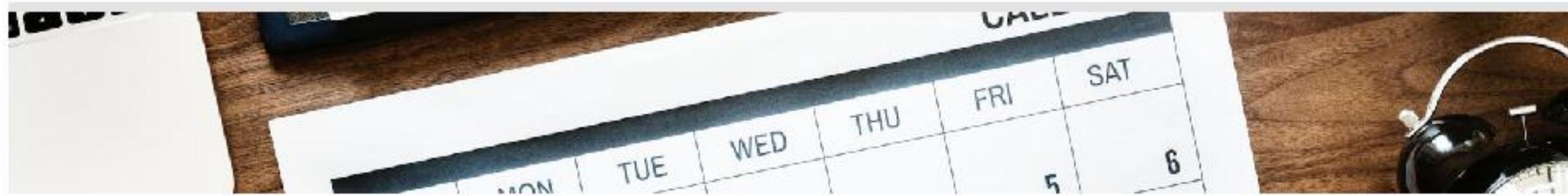


Photo by rawpixel

### Save the date 📅

On the Saturday following `rstudio::conf`, we'll be holding our first ever **tidyverse developer day**, and you're invited! So, if you're interested, plan on being in Austin on the 19th of January. The venue is not settled yet, but the intent is to make it convenient for people who've chosen lodging based on the `rstudio::conf` location.

### Squad goals 🎯

There will be more details to come, but the idea is to have a day where we can work together on anything ranging from submitting your first pull request, to working on your own package. The tidyverse team will be there, so we can help you hit the ground running and/or get over any stumbling blocks that you encounter. Don't have any ideas for something to work on? No problem! We'll be tagging issues in advance to make sure there's lots to do for any- and everyone, regardless of level of expertise.

# Thank You

<http://bit.ly/not-science>



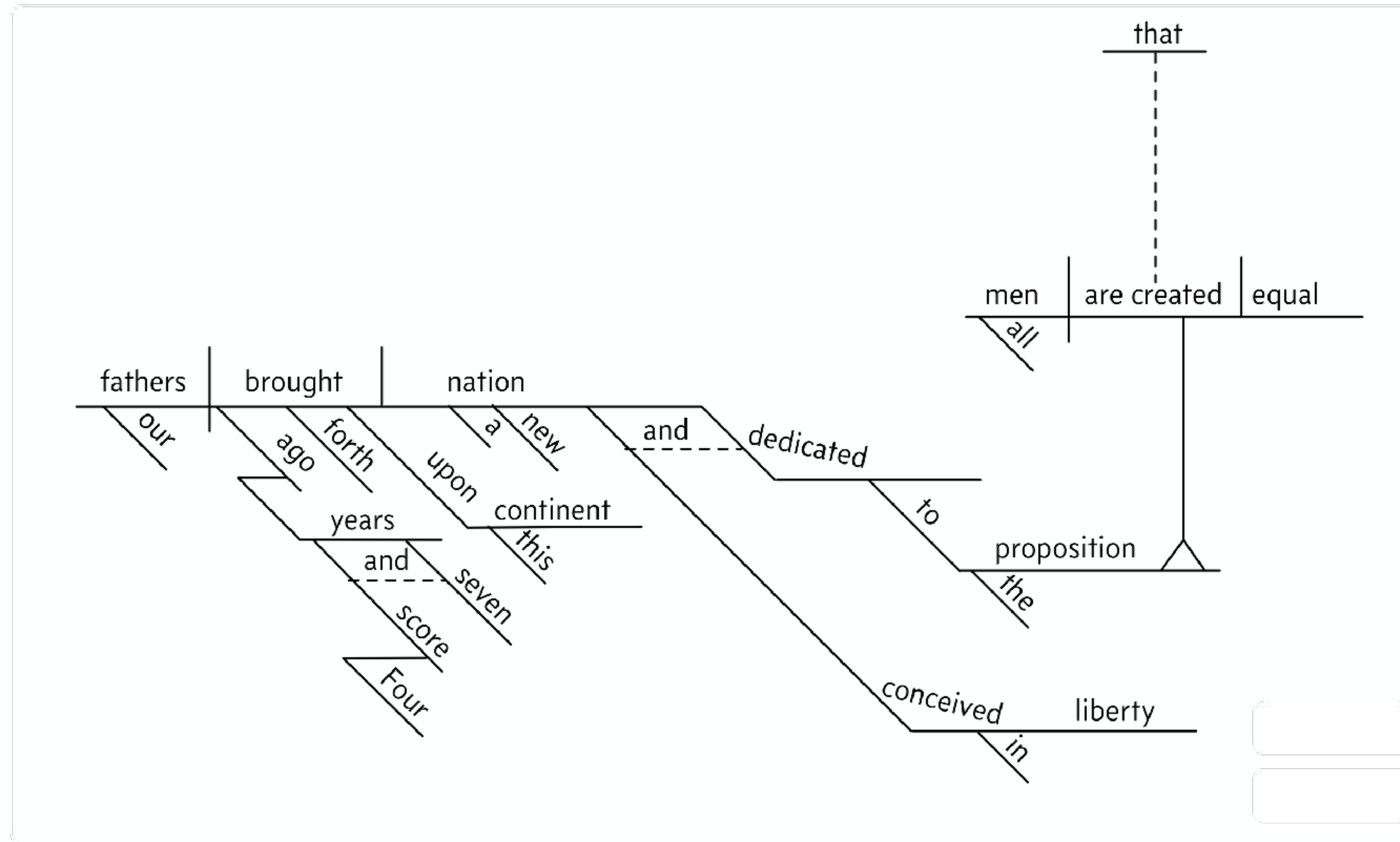
# Works cited

- Ross, Sydney. 1962. "Scientist: The Story of a Word." *Annals of Science* 18 (2): 65–85. doi:10.1080/00033796200202722
- Donoho, David. 2017. "50 Years of Data Science." *Journal of Computational and Graphical Statistics* 26 (4). Taylor & Francis: 745–66. doi:10.1080/10618600.2017.1384734.
- Park, Dong Huk, Lisa Anne Hendricks, Zeynep Akata, Bernt Schiele, Trevor Darrell, and Marcus Rohrbach. 2016. "Attentive Explanations: Justifying Decisions and Pointing to the Evidence." arXiv, <<https://arxiv.org/abs/1612.04757>>
- Traweek, Sharon. 1988. *Beamtimes and Lifetimes: The World of High Energy Physics*. Cambridge, MA: Harvard University Press.
- Hofstadter, Douglas, and Emmanuel Sanders. 2013. *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*. 1st Ed. New York: Basic Books.
- Foucault, Michel. 1973. *The Order of Things: An Archaeology of the Human Sciences*. Edited by Trans Smith, A.M.S. New York Vintage. New York: Vintage Books.
- Paxtyn Merten (14 Feb. 2018.). What can data visualization learn from feminism? Storybench. Retrieved from <http://www.storybench.org/can-data-visualization-learn-feminism/>
- Diagram translated by Benjamin Heller of the University of Michigan and designed by Marketa Kubacakova.(6 Nov. 2013.). The Encyclopedia of Diderot & d'Alembert Collaborative Translation Project. Quod.lib.umich.edu. Retrieved from <https://quod.lib.umich.edu/d/did/tree.html>
- The Encyclopedia Diderot and d'Alembert - Departmental Archives of the Ardennes. *Archives.cd08.fr*. Retrieved from <http://archives.cd08.fr/article.php?laref=1387&titre=l-encyclopedie-diderot-et-d-alembert>
- Thomas Sprat (1635-1713) - The History of the Royal Society of London for the Improving of Natural Knowledge 1667. Royalcollection.org.uk. <https://www.royalcollection.org.uk/collection/1057783/the-history-of-the-royal-society-of-london-for-the-improving-of-natural-knowledge>
- Whewell, William. (anonymously) 1834. The Quarterly Review, 51, 58-61.
- Thos. H. Huxley 1894. Letter to J.T. Carrington, editor of Science-Gossip, in Ross 1962.
- Hardwicke's science-gossip : an illustrated medium of interchange and gossip for students and lovers of nature. London : Robert Hardwicke, 1866- <https://www.biodiversitylibrary.org/bibliography/1953>
- Merton, Robert K. (1973) [1942]. "The Normative Structure of Science." In *The Sociology of Science: Theoretical and Empirical Investigations*, 267–78. University of Chicago Press (1979). 
- Latour, Bruno, and Steve Woolgar. 1979. *Laboratory Life: The Construction of Scientific Facts*. Beverly Hills: Sage Publications.
- Norman, Donald A., and Pieter Jan Stappers. 2015. "DesignX: Complex Sociotechnical Systems." *She Ji: The Journal of Design, Economics, and Innovation* 1 (2): 83–106. doi:10.1016/j.sheji.2016.01.002
- NIST image: This work is in the public domain in the United States because it is a work prepared by an officer or employee of the United States Government as part of that person's official duties under the terms of Title 17, Chapter 1, Section 105 of the US Code.
- Isaac Newton, *Philosophiae Naturalis Principia Mathematica*, Cambridge 1713. <[http://www.people.vcu.edu/~mjmurphy/history\\_of\\_science/newton/newton\\_principia.html](http://www.people.vcu.edu/~mjmurphy/history_of_science/newton/newton_principia.html)>
- Sparrow, B., J. Liu, and D. M. Wegner. 2011. "Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips." *Science* 333 (6043): 776–78. doi:10.1126/science.1207745.
- Dudo, Anthony. 2015. "Scientists, the Media, and the Public Communication of Science." *Sociology Compass* 9 (9): 761–75. doi:10.1111/soc4.12298.
- Relman, Arnold S. 1981. "The Ingelfinger Rule." *New England Journal of Medicine* 305 (14): 824–26. doi:10.1056/nejm198110013051408.
- Pintscher, Lydia, ed. 2012. *Open Advice: FOSS: What We Wish We Had Known When We Started*. <<http://open-advice.org/>>
- "You'll Eventually Know Everything They've Forgotten." In *Open Advice: FOSS: What We Wish We Had Known When We Started*, 29–32.

# **Analogies and a-ha! moments**

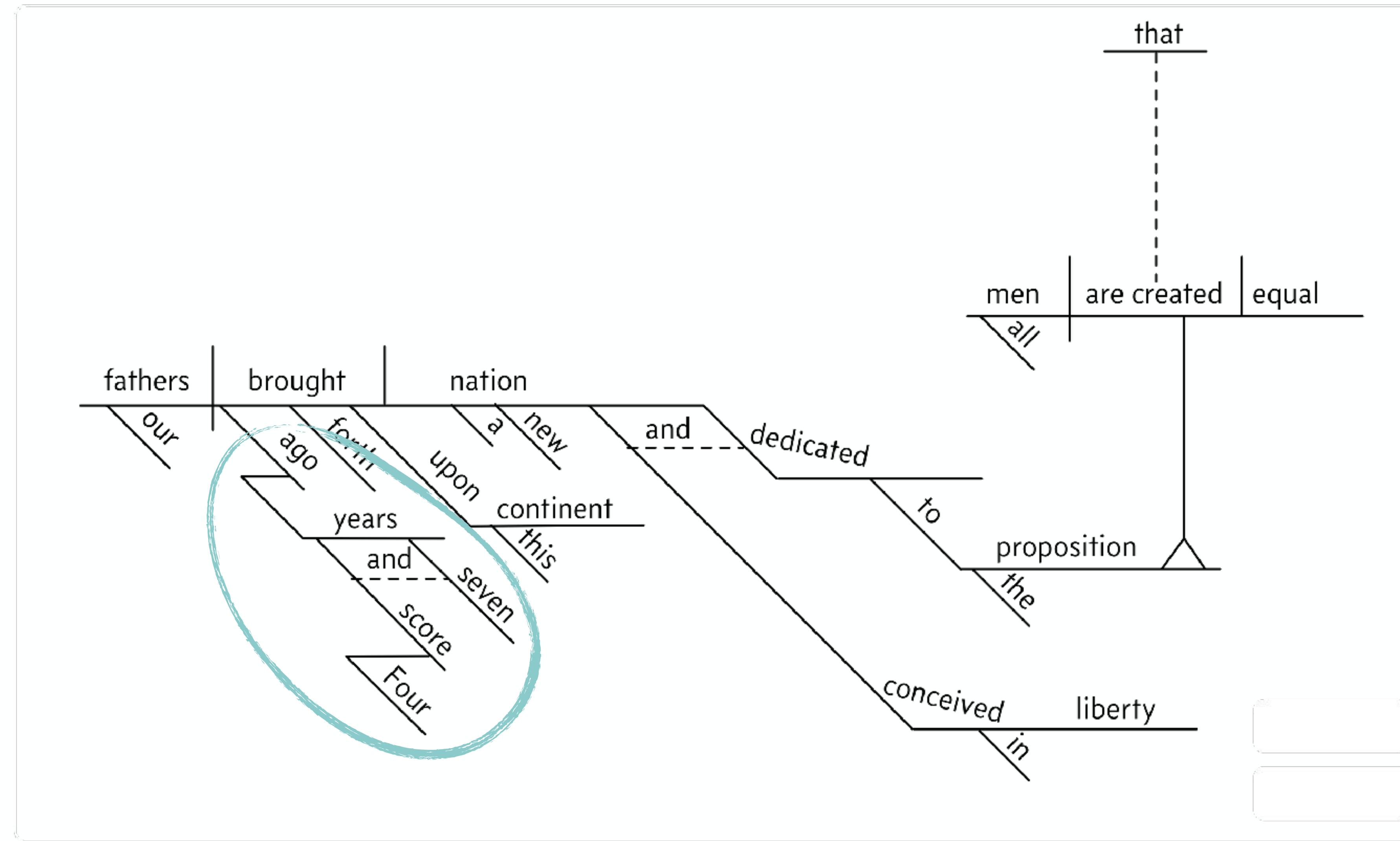
*Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal.*

— Abraham Lincoln, *The Gettysburg Address* (1863)



*Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal.*

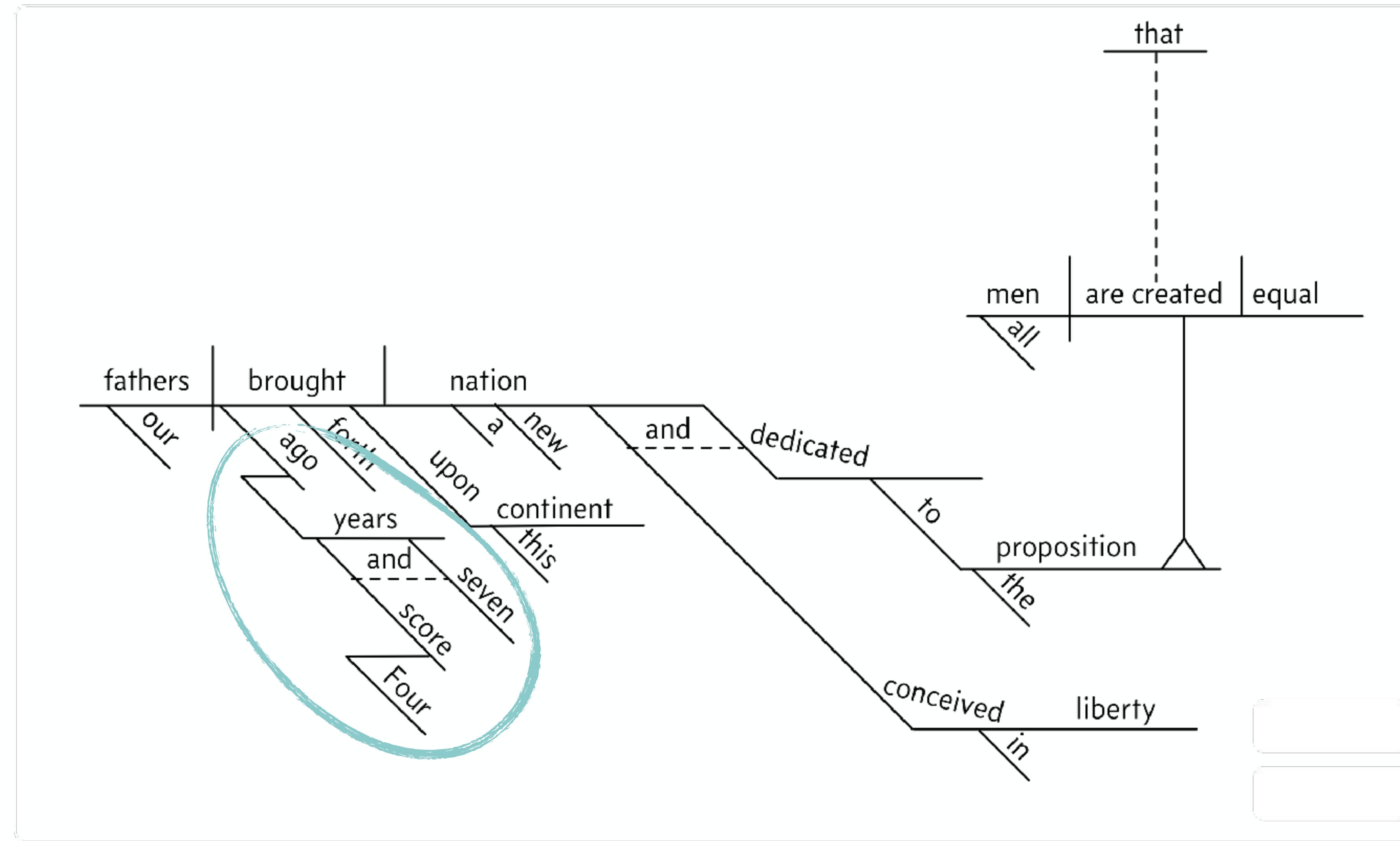
— Abraham Lincoln, *The Gettysburg Address* (1863)



*Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal.*

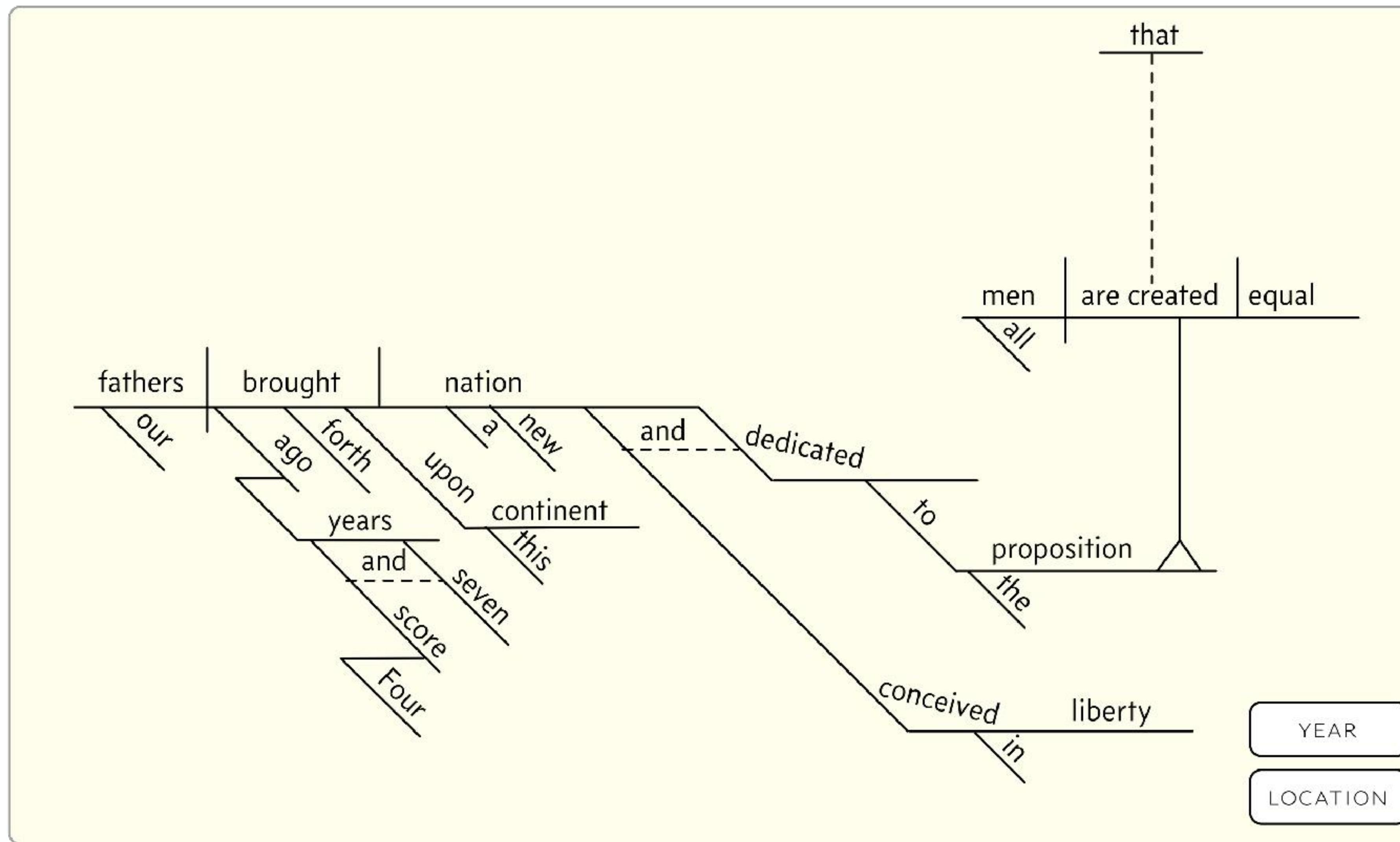
— Abraham Lincoln, *The Gettysburg Address* (1863) ←

env't year = 1863



*Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal.*

— Abraham Lincoln, *The Gettysburg Address* (1863)

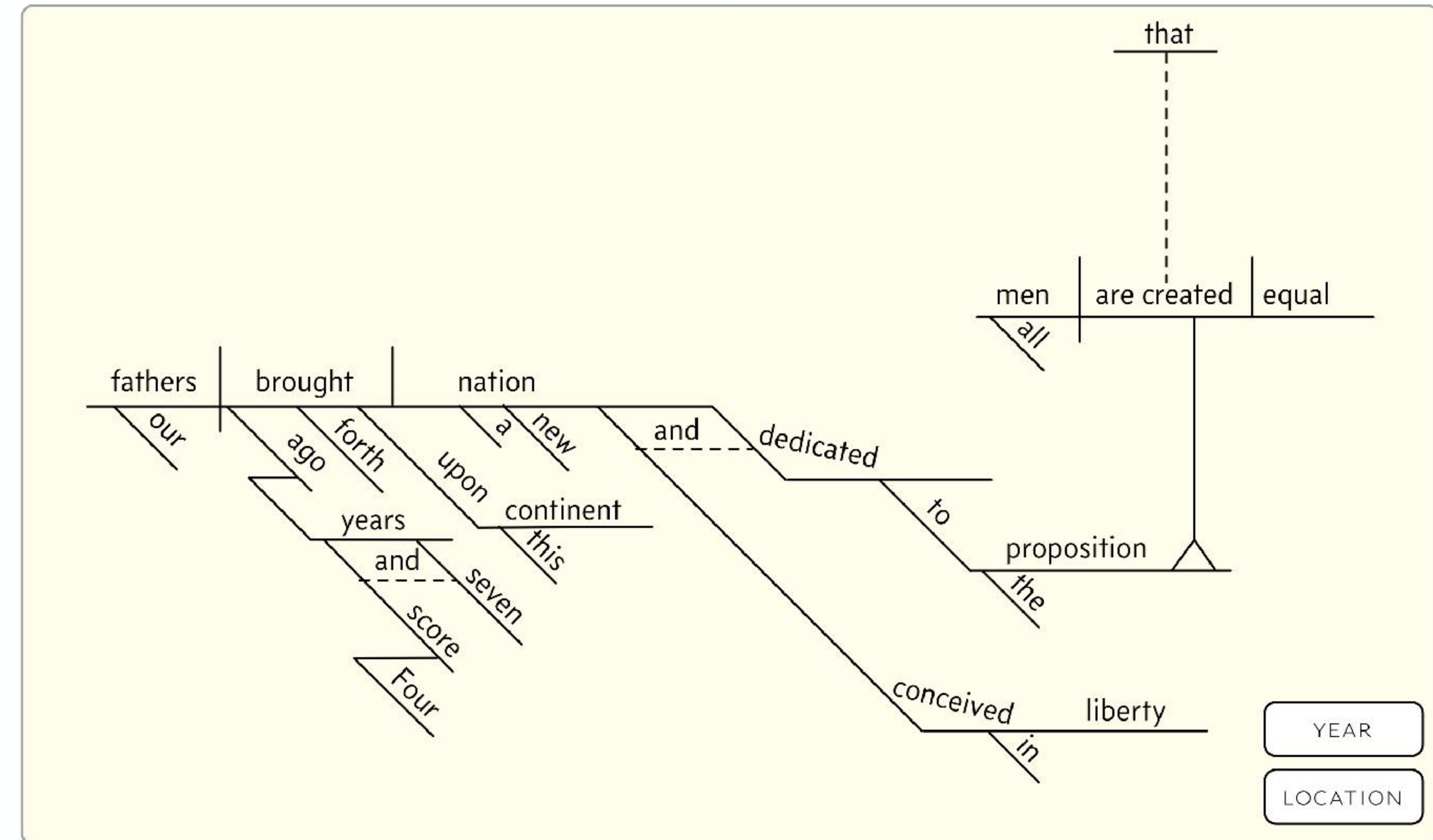


*Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal.*

— Abraham Lincoln, *The Gettysburg Address* (1863)

$$(4 \cdot \text{score}) + 7 = 87$$

$$1863 - 87 = 1776$$

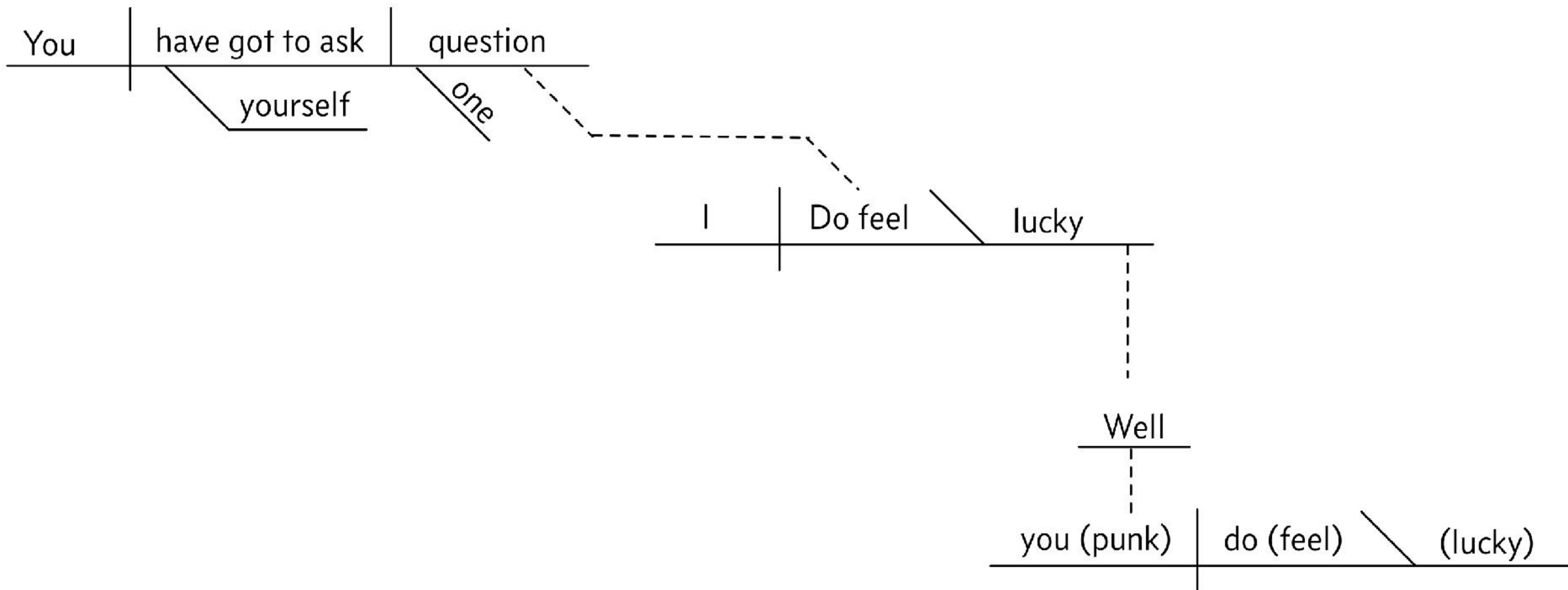


# Execution environments with Dirty Harry



You've got to ask yourself one question:  
‘*Do I feel lucky?*’ Well, do ya, punk?

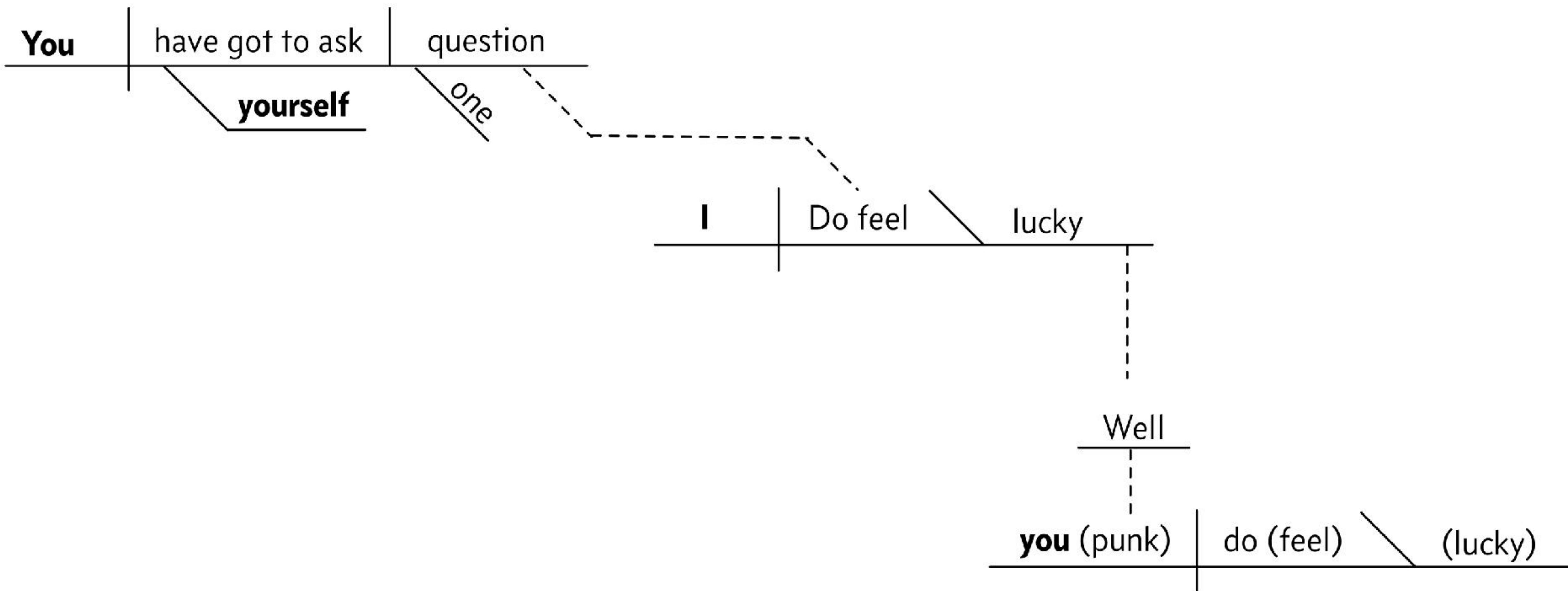
Harry Callahan, *Dirty Harry*



Adapted from Pop Chart Lab

You've got to ask yourself one question:  
‘*Do I feel lucky?*’ Well, do ya, punk?

Harry Callahan, *Dirty Harry*



Adapted from Pop Chart Lab

subjective pronoun  
(2nd person singular)

You

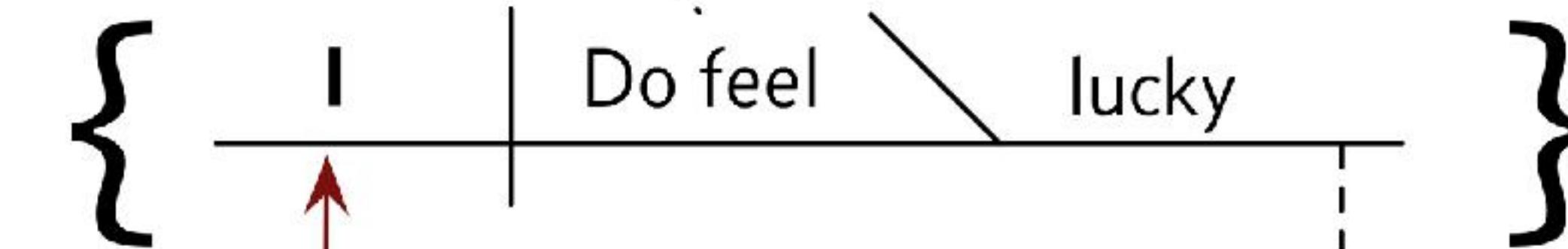
have got to ask

**yourself**

reflexive pronoun  
(2nd person singular)

question

one



nominative pronoun  
(1st person singular)

subjective pronoun  
(2nd person singular)

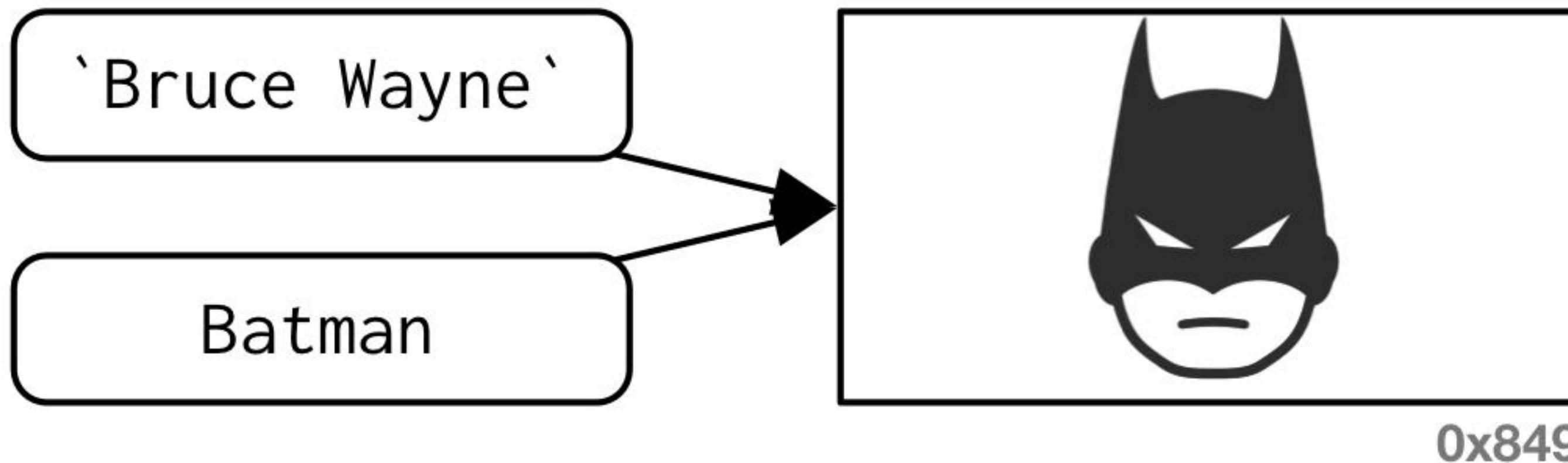
Harry Callahan, *Dirty Harry*

You've got to ask yourself one question:  
'*Do I feel lucky?*' Well, do ya, punk?

Well

→ **you** (punk) | do (feel) | (lucky)

# names and values



# names and values

