



# Unreasonable effectiveness of



**Kasia Kulma, Senior Data Scientist**



**@KKulma**



**kkulma@mango-solutions.com**

REPLACE WITH CONFIDENTIALITY STATEMENT

# What to expect?



Who am I / are we?



Why R?



What's possible?



How to learn R?

# About me and Mango



# About me



Evolutionary biologist



Created by Fauzan Adlima  
from Noun Project

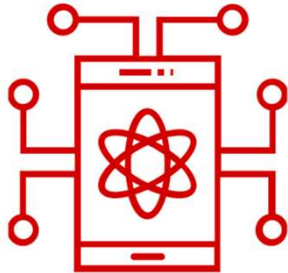
5+ years in various data science roles  
(R, Python, ML, web app dev, data analysis)



Created by Stephen Borengasser  
from Noun Project

R-lady, blogger, Open Source contributor, public speaker

# About Mango



Created by Creative Stall  
from Noun Project

## Data Science

Training  
Consulting  
App Dev



Created by Pravin Unagar  
from Noun Project

## Data Engineering

Analytic Infrastructure  
DevOps for Data  
Modernisation



Created by Boris  
from Noun Project

## Data Products

App Dev  
Validation  
Testing

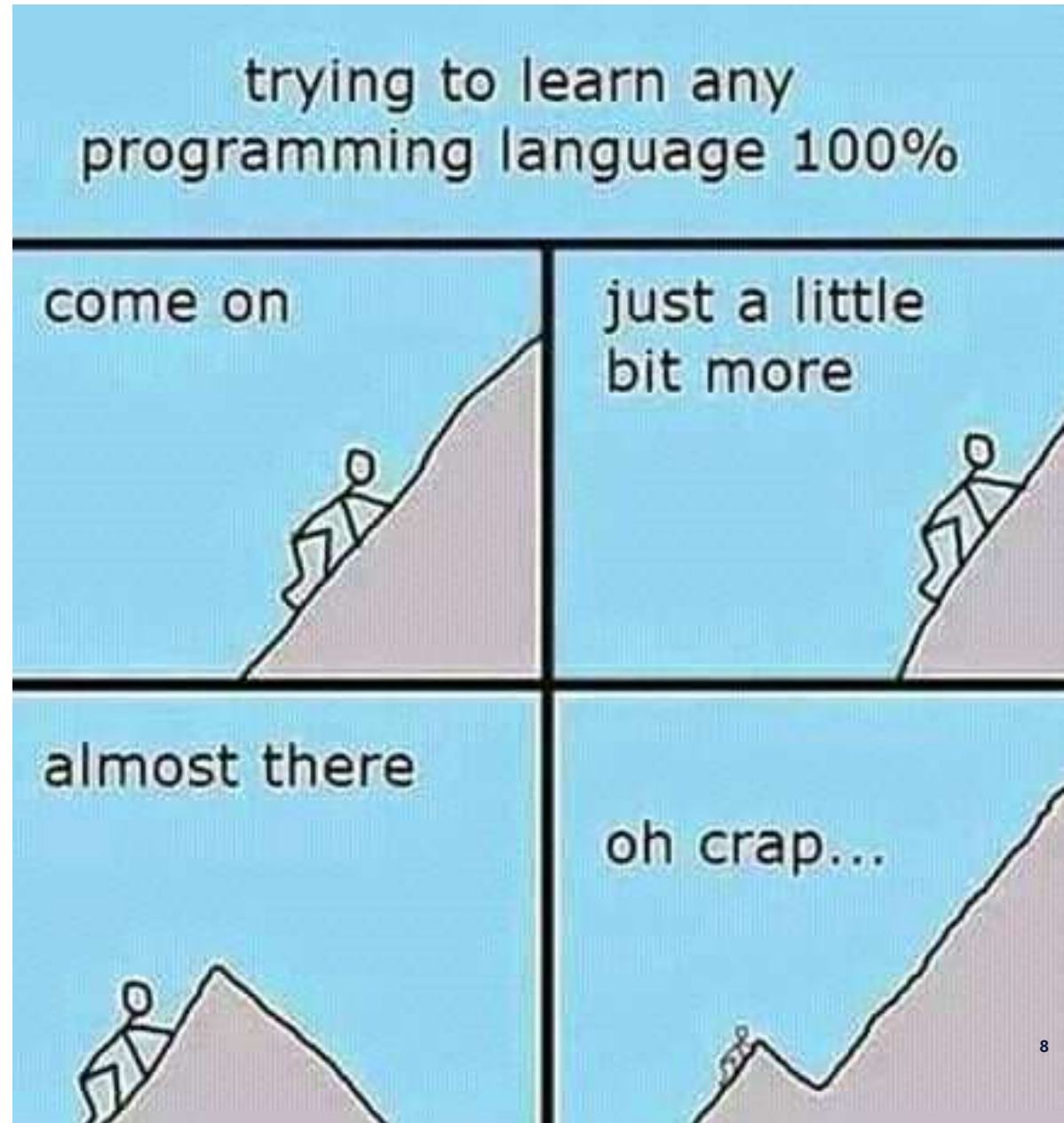
# Customers Across Sector



**Why even bother?**



# Why R?

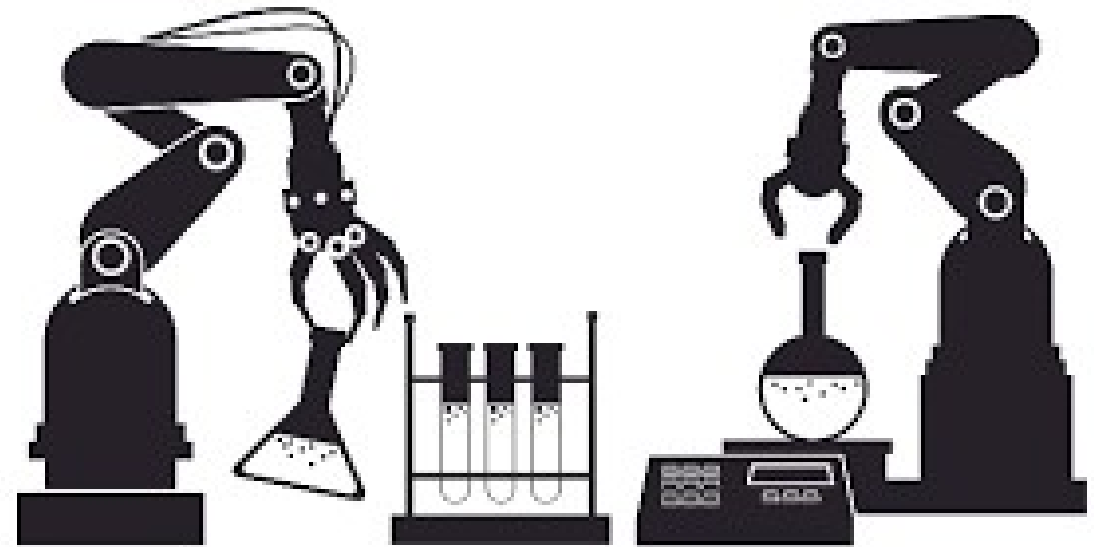






# Why R?

## Reproducibility





# Why R?

## Open source



# Why R?

## Platform-independent





## Why R?

**Connects with  
other languages**

# Why R?

## Documentation

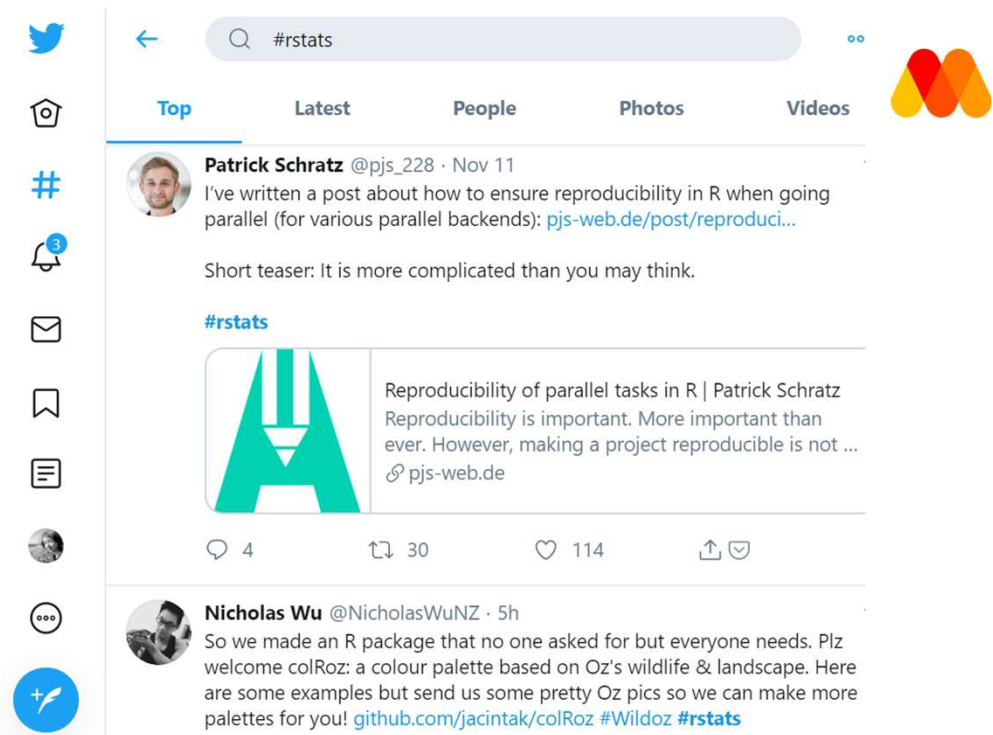
The screenshot shows the R Documentation website homepage. At the top left is the 'R Documentation' logo. On the right, there are links for 'R Enterprise Training', 'R package', 'Leaderboard', and 'Sign in'. The main heading reads 'Search all 18,535 CRAN, Bioconductor and GitHub packages.' Below this is a search bar with the placeholder text 'Search for packages, functions, etc.' and a 'Search' button. A link below the search bar says 'Or explore packages in one of the [Task Views](#).' At the bottom, there are three columns: 'Top 5 packages', 'Top 5 authors', and 'Newest packages'. The 'Newest packages' list includes: 1. AzureCognitive, 2. RAINBOWR, 3. Jaya, and 4. DealGPL570.

The top part of the screenshot shows the 'Documentation' page on the R website. It features the R logo and a sidebar with links: '[Home]', 'Download CRAN', 'R Project' (About R, Logo, Contributors, What's New?, Reporting Bugs), and 'TEAMS' (What's this?, First 25 Users Free). The main content area is titled 'Documentation' and 'Document Collections, Journals and Proceedings'. It lists several resources: 'Browsable HTML versions of the manuals, help pages and NEWS for the developing versions of R "R-patched" and "R-devel", updated daily.', 'CRAN has a growing list of contributed documentation in a variety of languages.', 'A list of books and other publications related to R.', 'The Journal of Statistical Software features R-related articles like package guides on a regular basis.', and 'Proceedings from the International Workshops on Distributed Statistical Computing.'

The bottom part of the screenshot shows a Stack Overflow page for the tag '[r]'. The page title is 'Questions tagged [r]' with an 'Ask Question' button. A description of R is provided: 'R is a free, open-source programming language and software environment for statistical computing, bioinformatics, visualization, and general computing. Please provide minimal and reproducible example(s) along with the desired output. Use `dput()` for data and specify all non-base packages with `library()` calls. Do not embed pictures for data or code, use indented code blocks instead. For statistics related questions, use <https://stats.stackexchange.com>.' Below this, there are links for 'Learn more...', 'Top users', 'Synonyms (2)', and 'r jobs'. The page shows '314,596 questions' with filters for 'Newest', 'Active', 'Bountied', 'Unanswered', and 'More'. A specific question is visible: 'what is the best r code to make 10-days NDVI data sets to monthly data sets'. The question text is: 'I have a 10-days NDVI data sets for 15 years in raster form. I need it with monthly temporal resolution. There 36 images per year. How can I make it monthly from 10 days NDVI data sets using R?'. The question has 0 votes, 0 answers, and 5 views. It was asked 9 mins ago by user 'Istakhar Jahangir'.

# Why R?

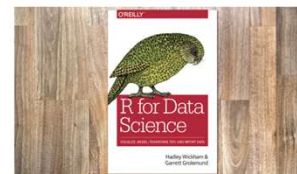
## Community



The screenshot shows a Twitter search for the hashtag #rstats. The search results are filtered to 'Top' posts. The first tweet is from Patrick Schratz (@pjs\_228) dated Nov 11. The tweet text reads: "I've written a post about how to ensure reproducibility in R when going parallel (for various parallel backends): [pjs-web.de/post/reproduci...](https://pjs-web.de/post/reproduci...)" followed by a short teaser: "Short teaser: It is more complicated than you may think." Below the text is a link to the #rstats hashtag and a preview of a blog post titled "Reproducibility of parallel tasks in R | Patrick Schratz". The preview text says: "Reproducibility is important. More important than ever. However, making a project reproducible is not ..." and includes the URL [pjs-web.de](https://pjs-web.de). The tweet has 4 replies, 30 retweets, and 114 likes. The second tweet is from Nicholas Wu (@NicholasWuNZ) dated 5h. The tweet text reads: "So we made an R package that no one asked for but everyone needs. Plz welcome colRoz: a colour palette based on Oz's wildlife & landscape. Here are some examples but send us some pretty Oz pics so we can make more palettes for you! [github.com/jacintak/colRoz](https://github.com/jacintak/colRoz) #Wildoz #rstats".

### Books

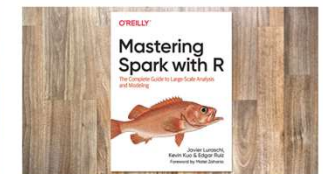
In addition to the software we've written at RStudio, we've also authored these guides to our tools.



R for Data Science  
by Hadley Wickham and Garrett Grolemund



R Markdown: The Definitive  
Guide

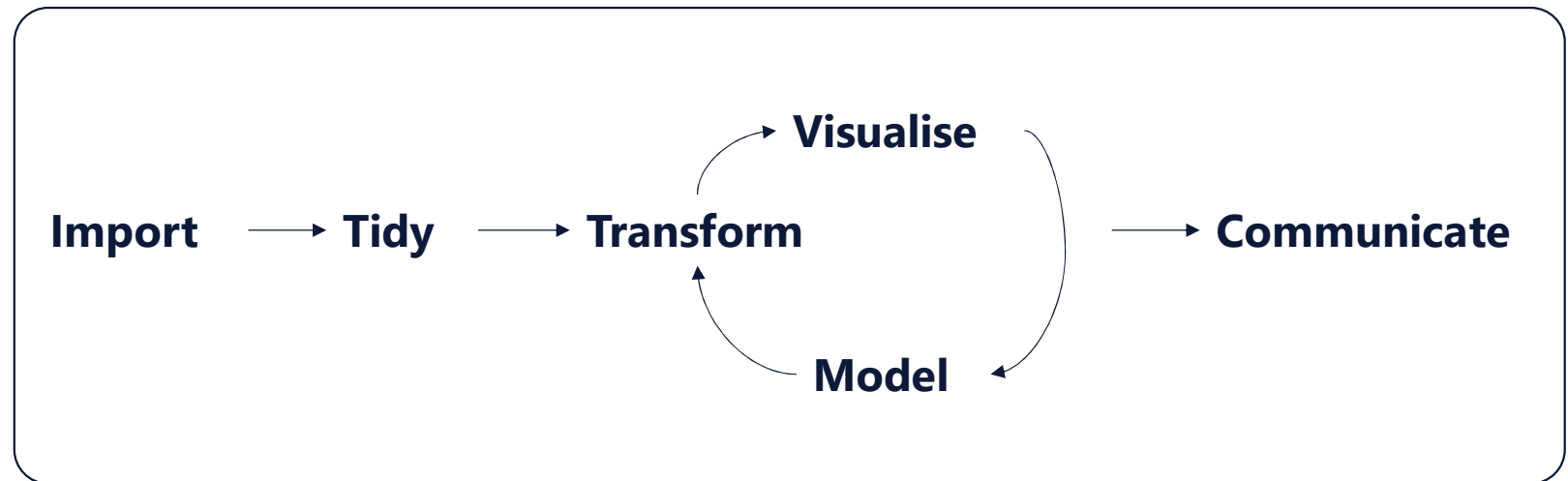


Mastering Apache Spark with R  
by Javier Luraschi, Kevin Kuo, Edgar Ruiz

**What's possible  
with R?**

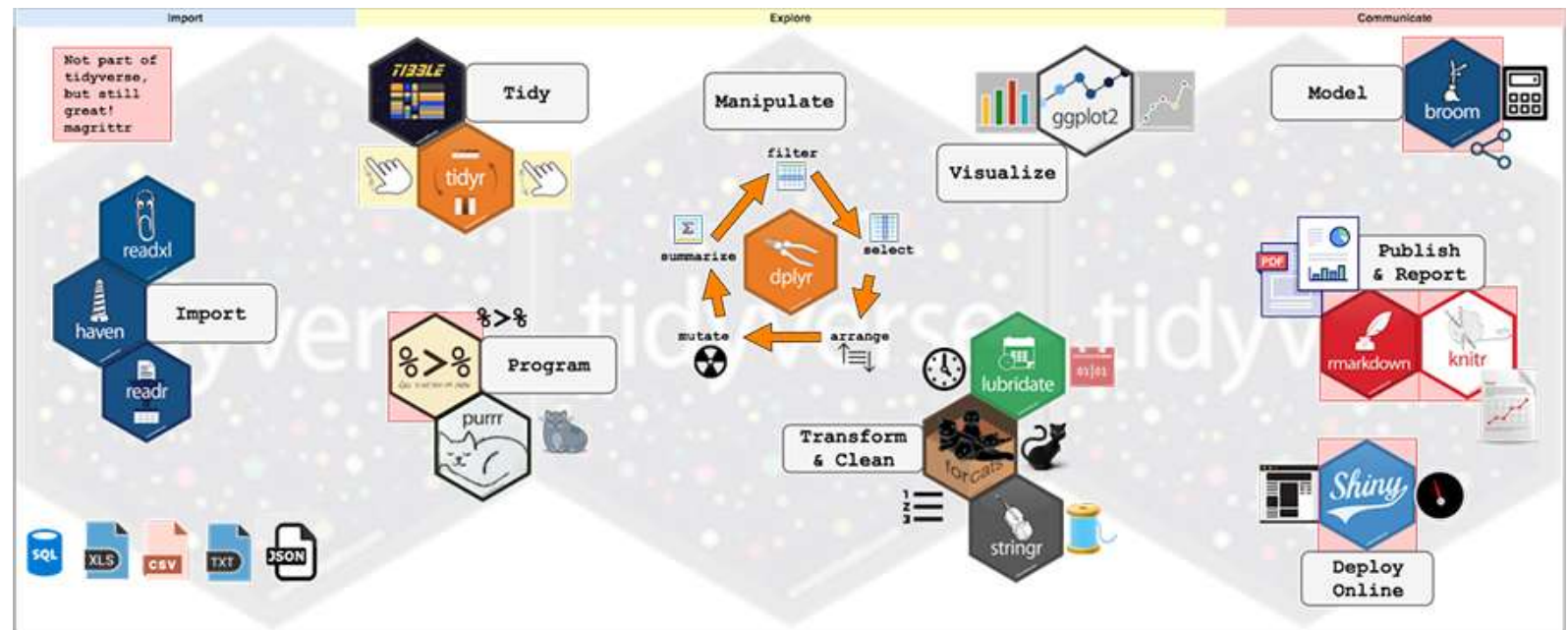


# R in the world of analytics





# R in the world of analytics



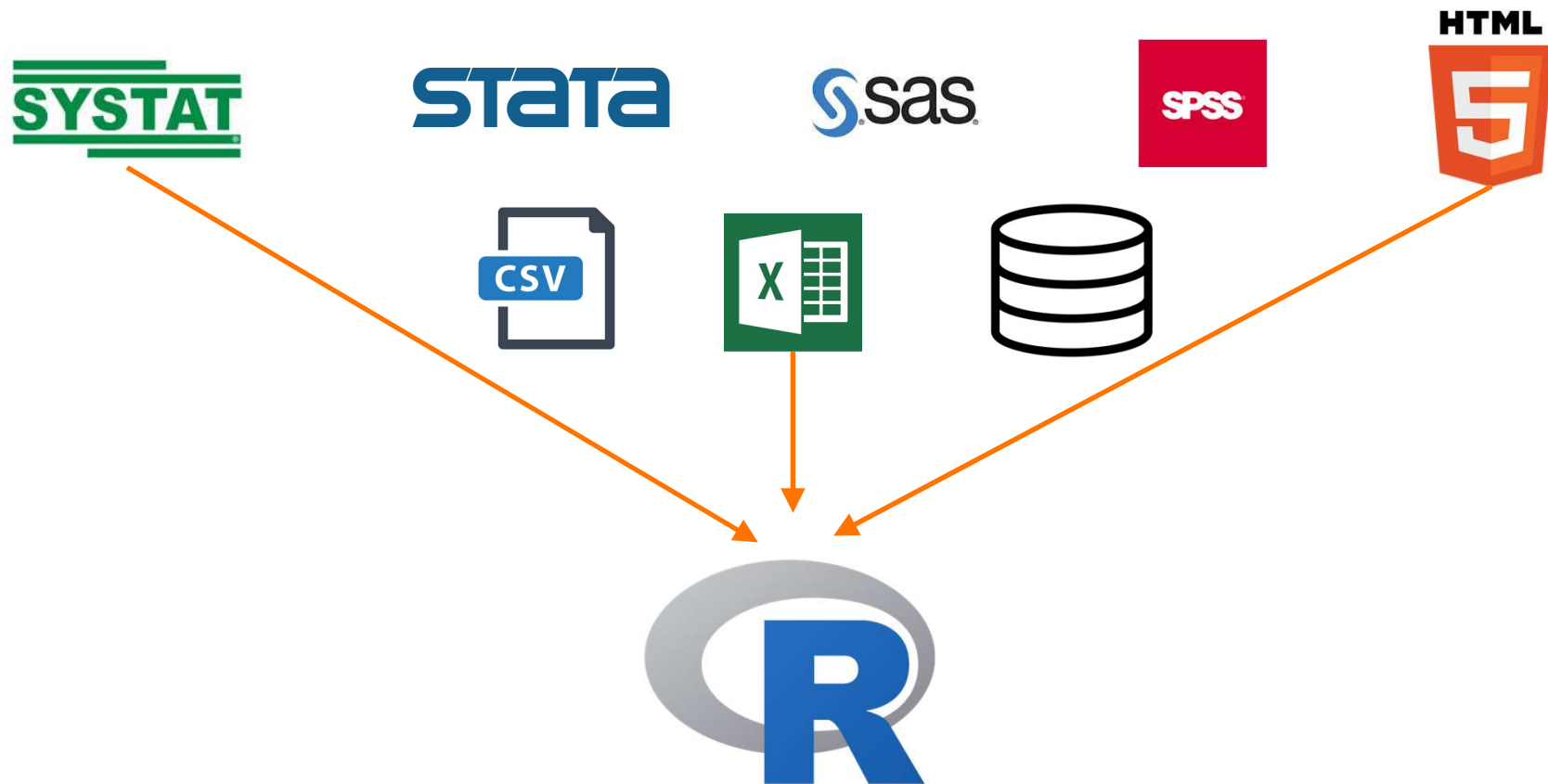
# R in the world of analytics



*Programs must be written for people to read and  
only incidentally for machines to execute*

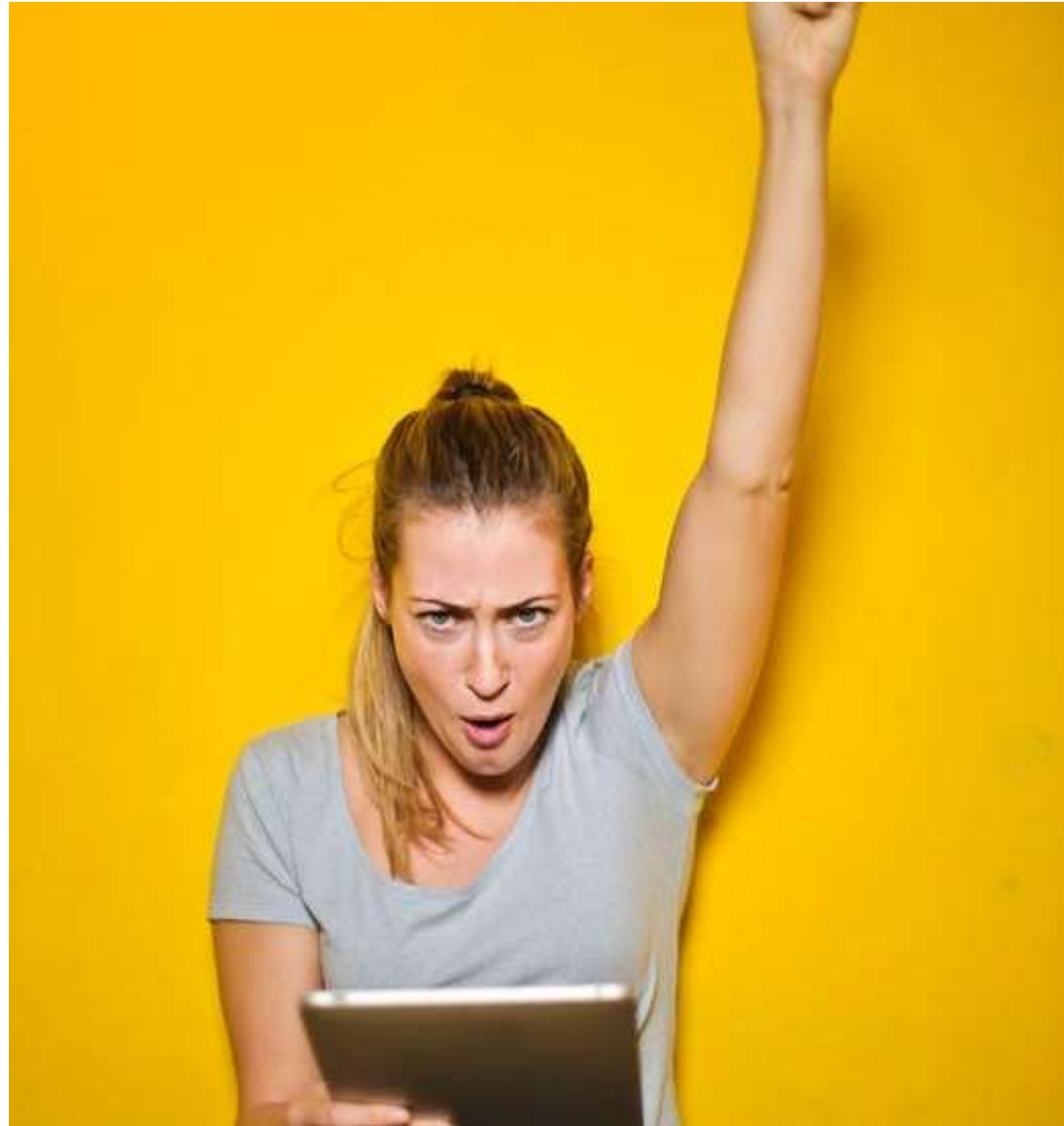
**Hal Abelson**

# Data import



# Data import

You can do it!



# Data wrangling

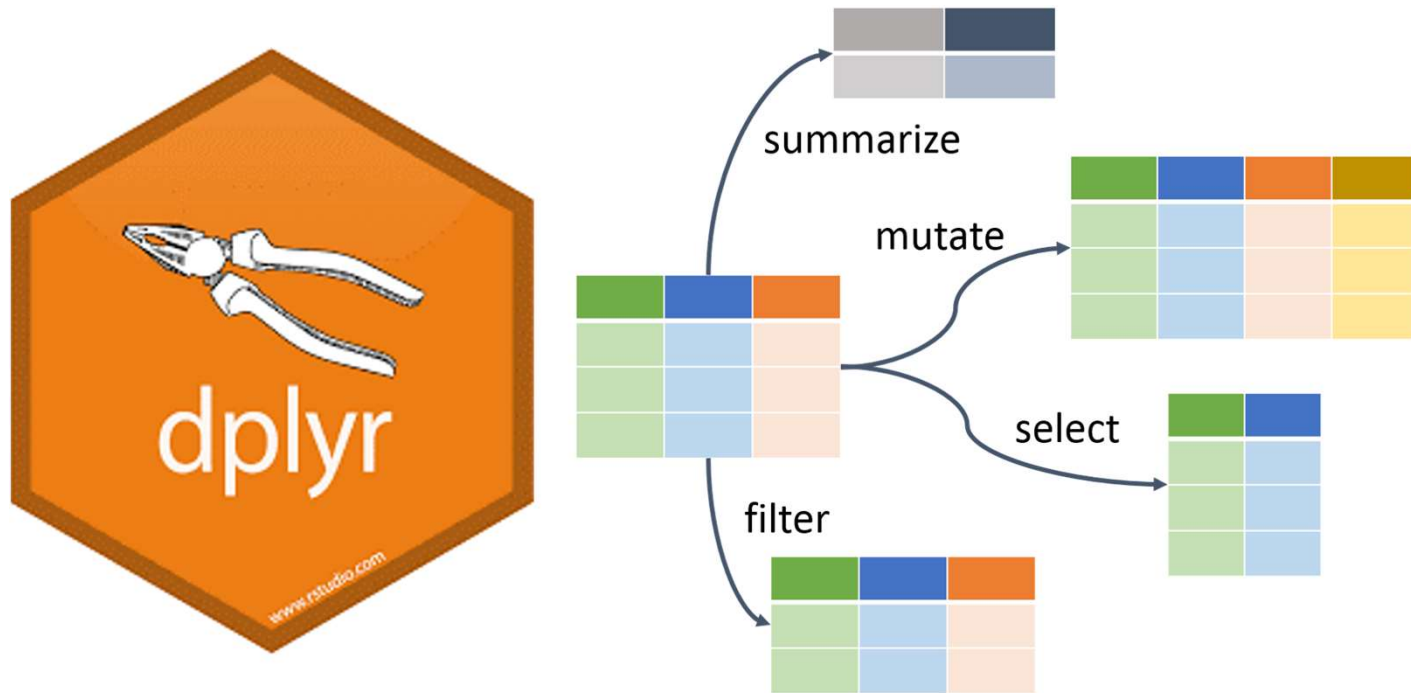


Image source: <https://towardsdatascience.com/data-manipulation-in-r-with-dplyr-3095e0867f75>

# Data wrangling

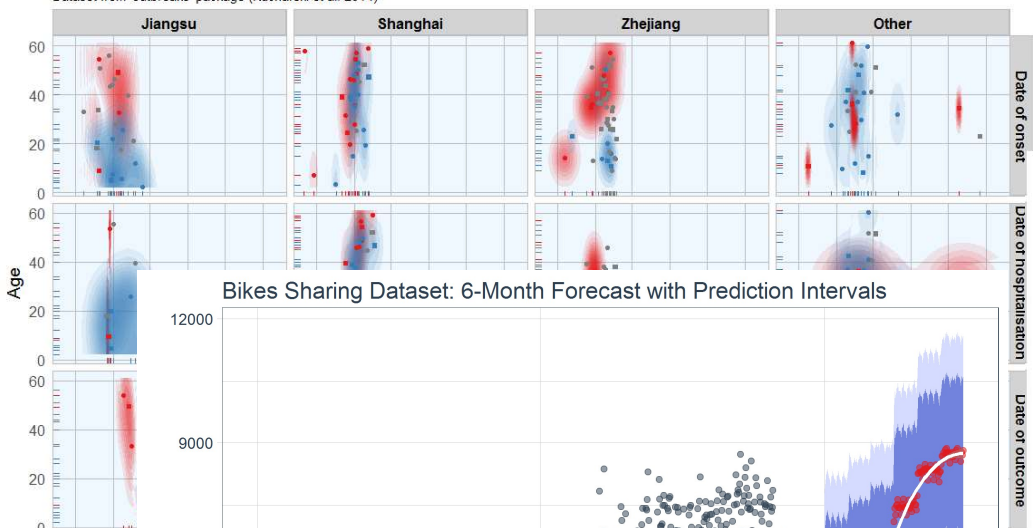
You can do it!



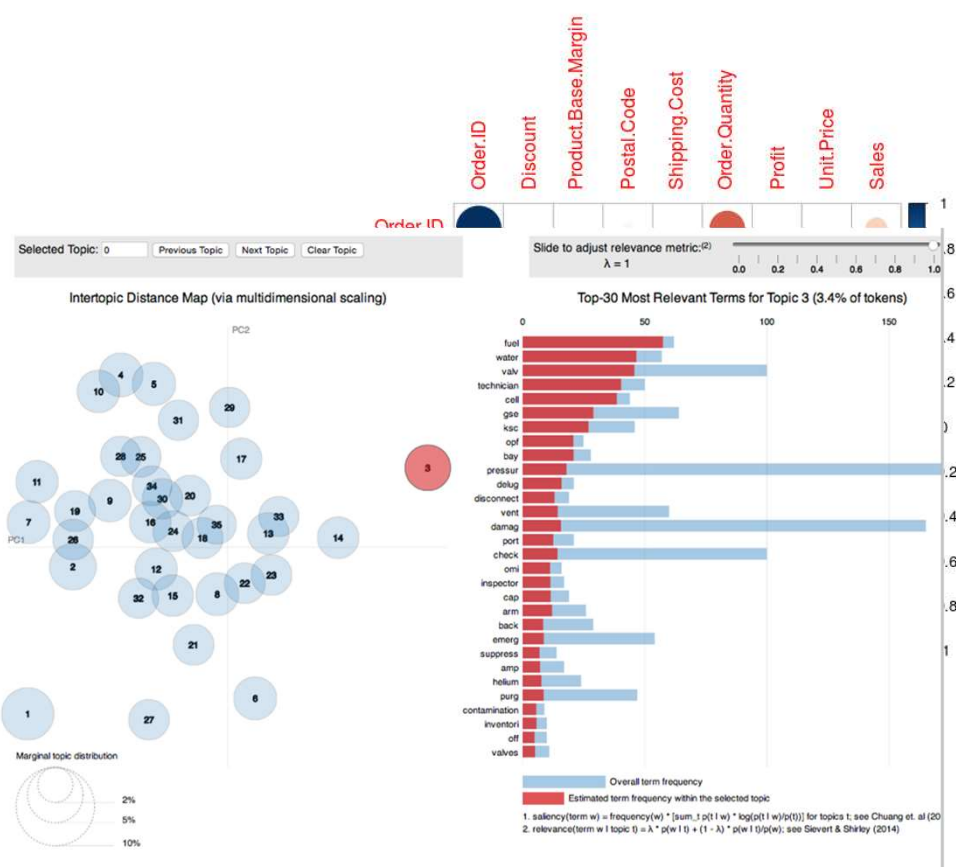
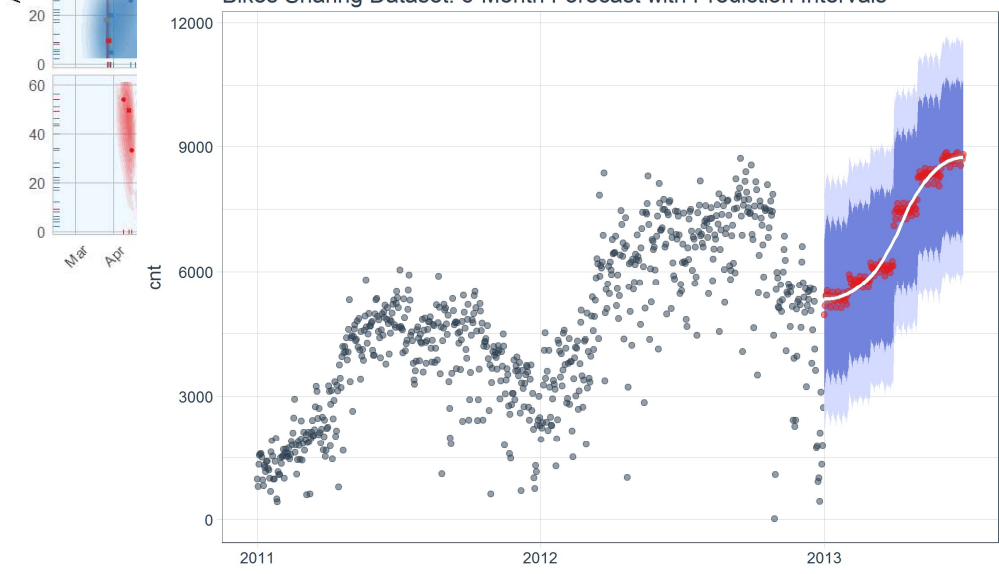
# Statistics / ML / AI



2013 Influenza A H7N9 cases in China  
Dataset from 'outbreaks' package (Kucharski et al. 2014)

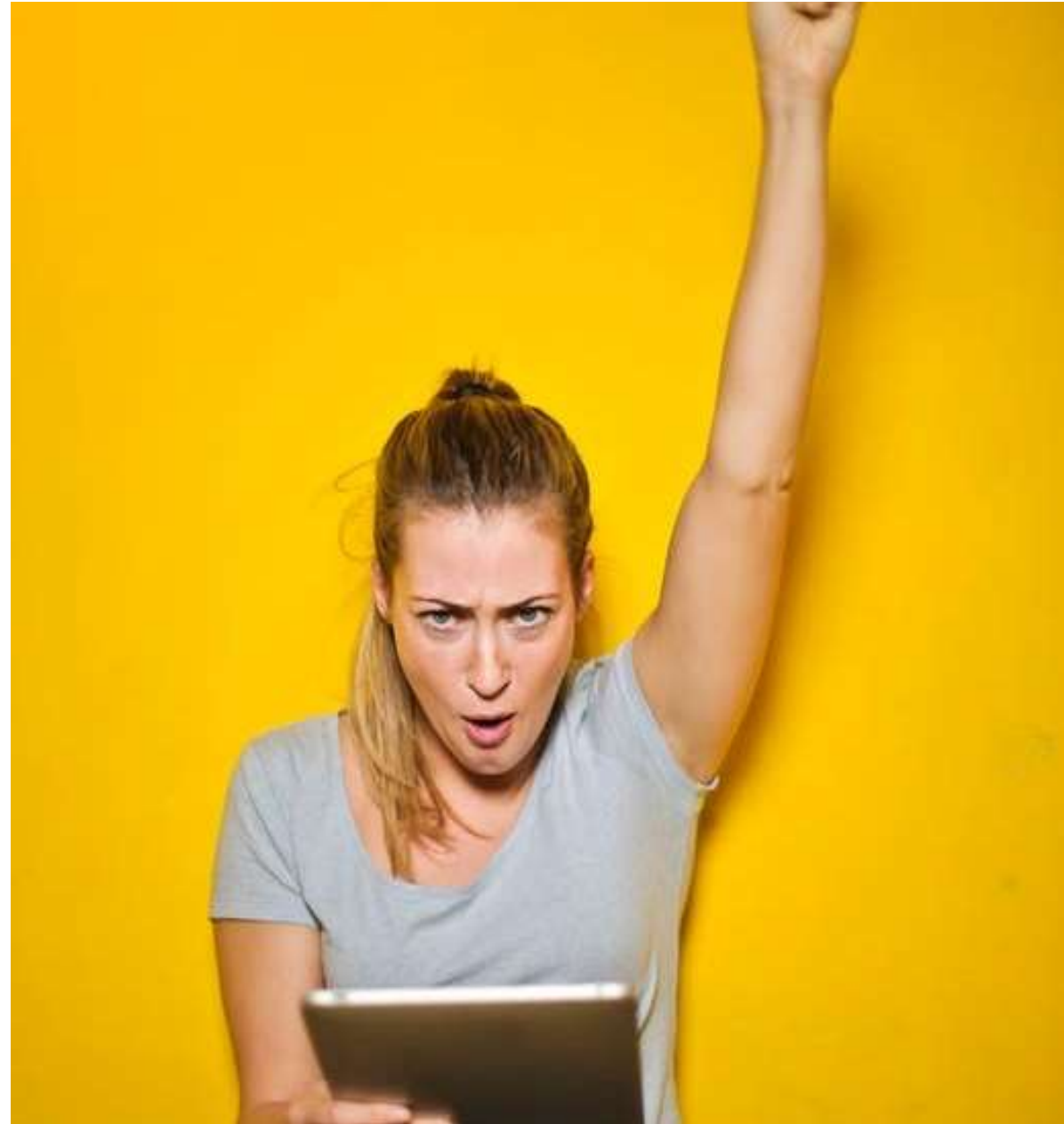


Bikes Sharing Dataset: 6-Month Forecast with Prediction Intervals



# Statistics / ML / AI

You can do it!





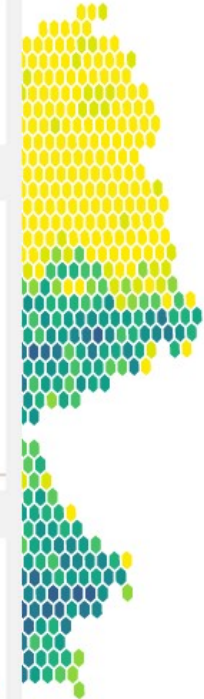
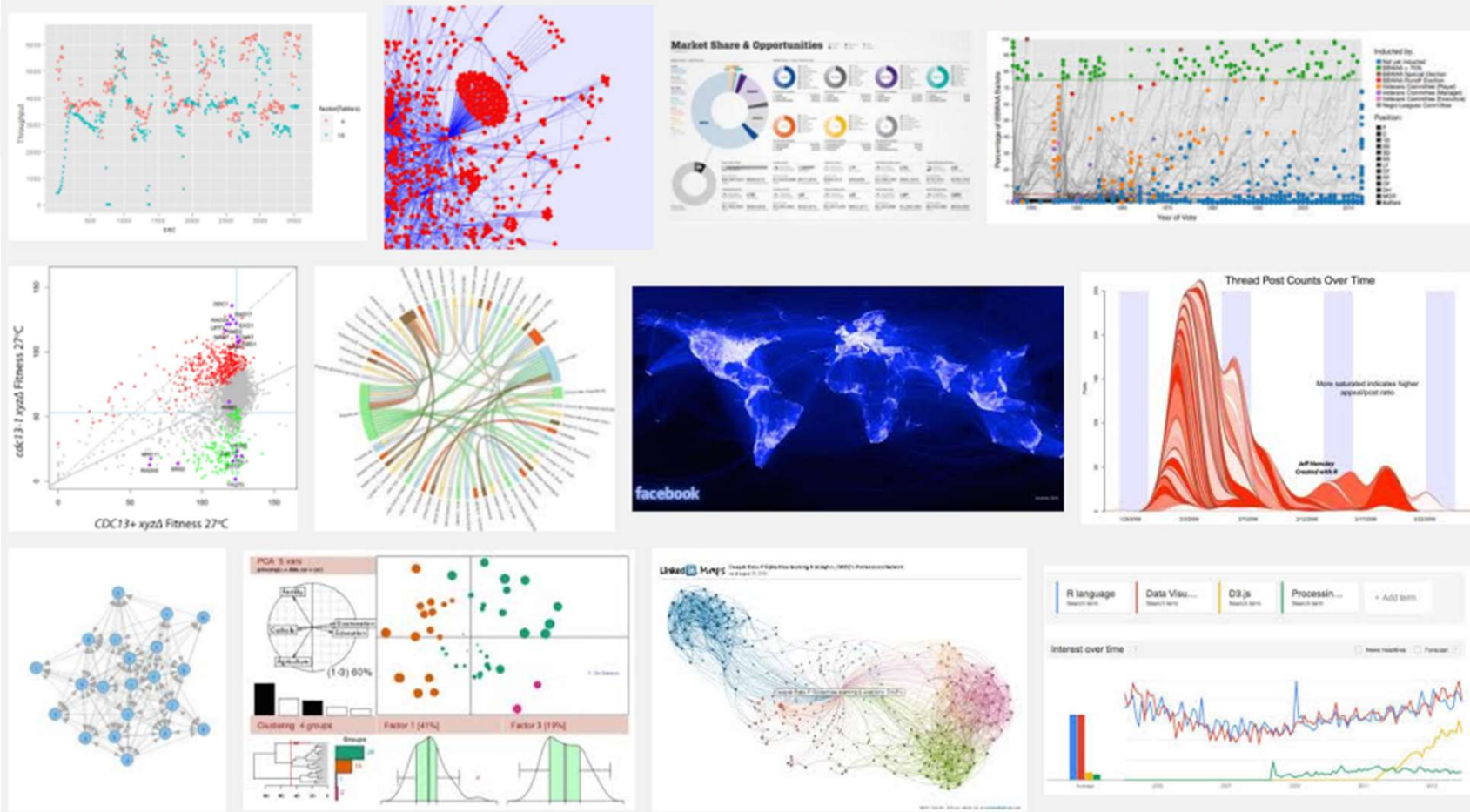
# Data visualization

each link is black with  
transparency 100/127 (79%)  
# color definition  
# R,G,B,alpha(127)  
blackweat = 0,0,0,100



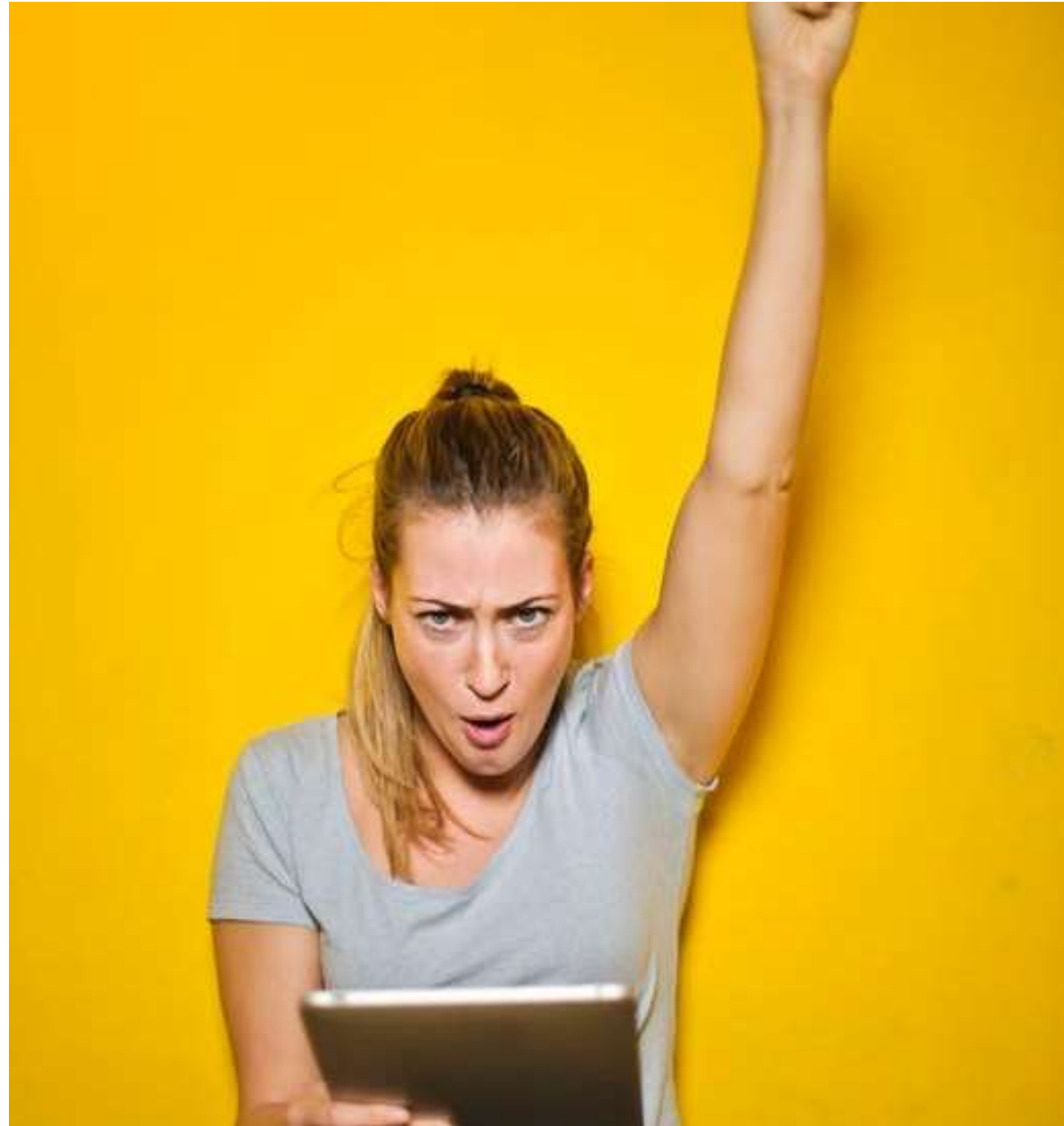
-ach

-a -aa -ah



# Data visualization

You can do it!

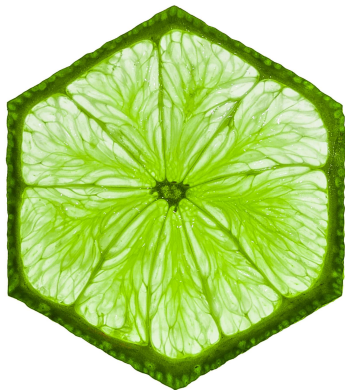


# Explainable Machine Learning / XAI

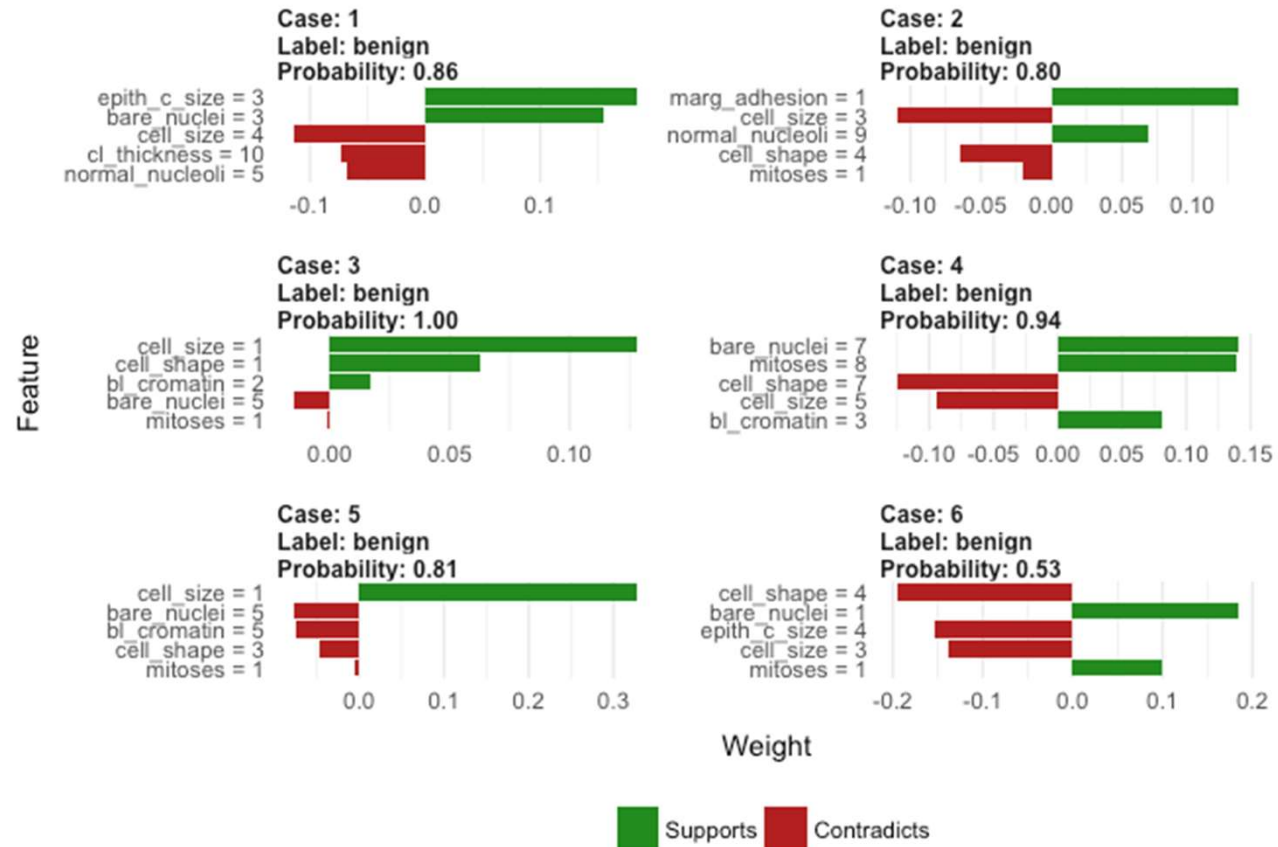


Understand not **what** machines predict  
but **why**

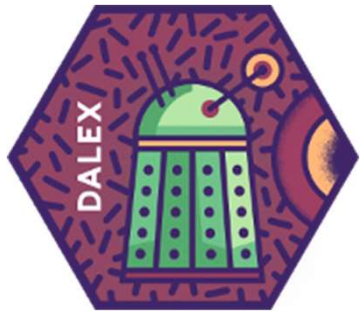
# Explainable Machine Learning / XAI



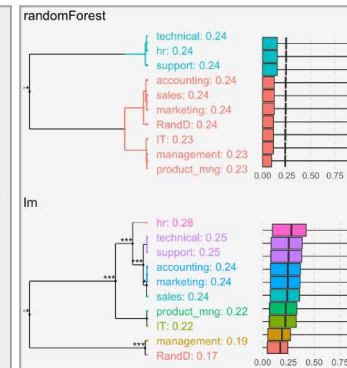
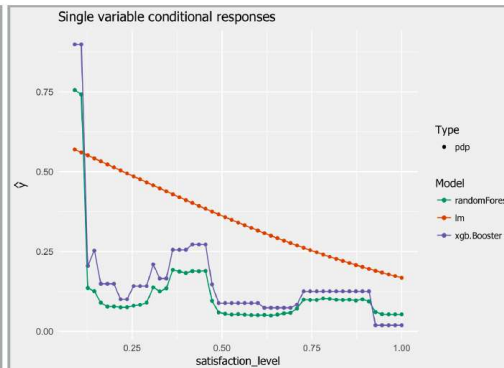
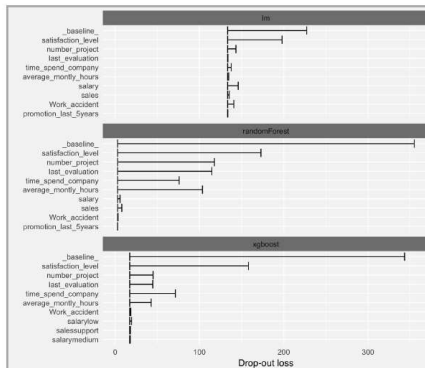
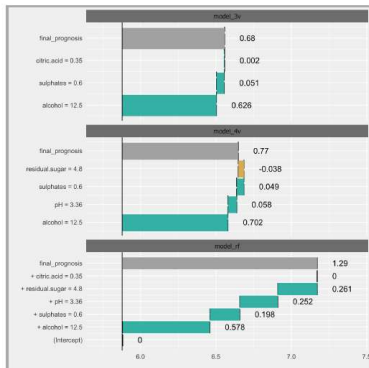
LIME



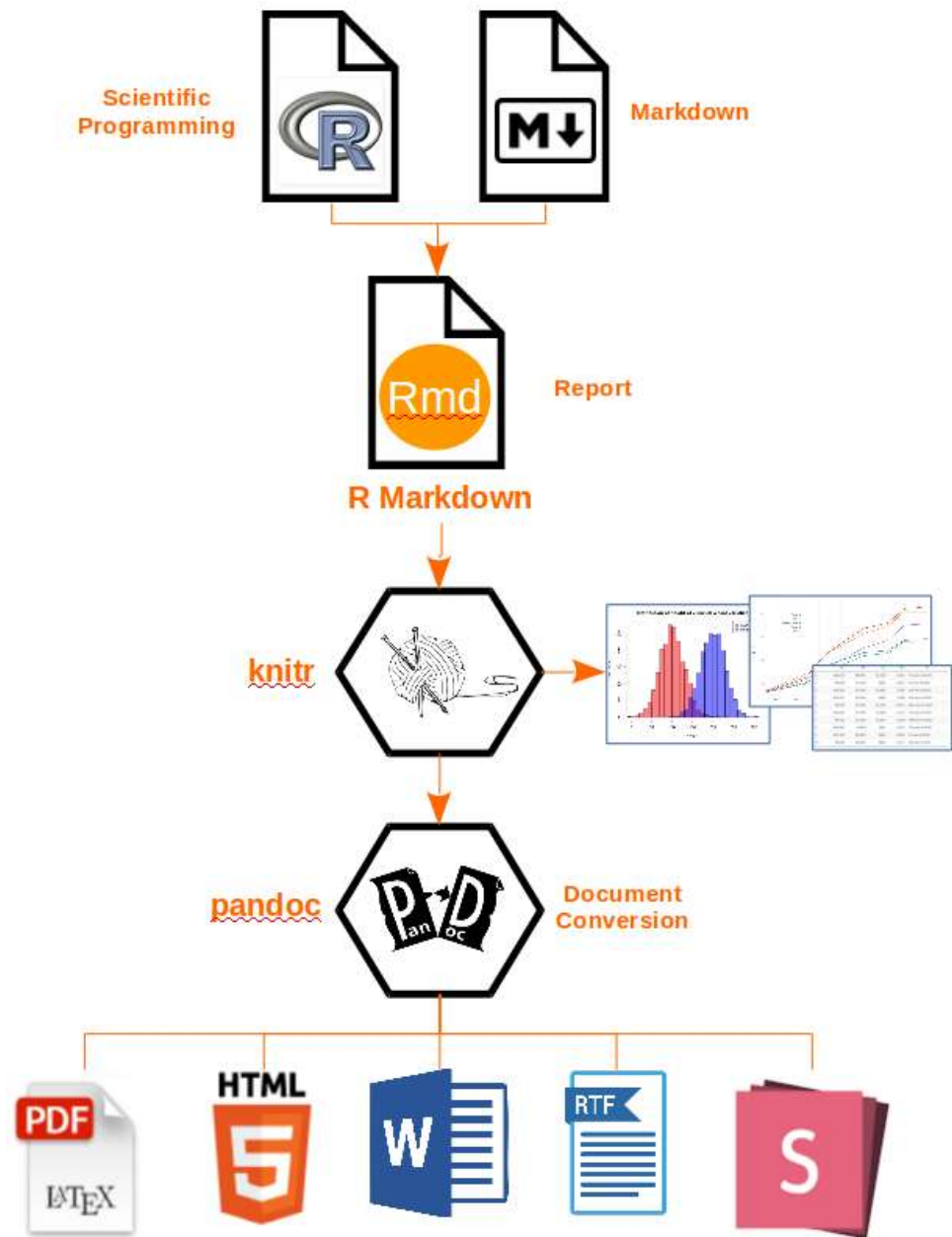
# Explainable Machine Learning / XAI



DALEX

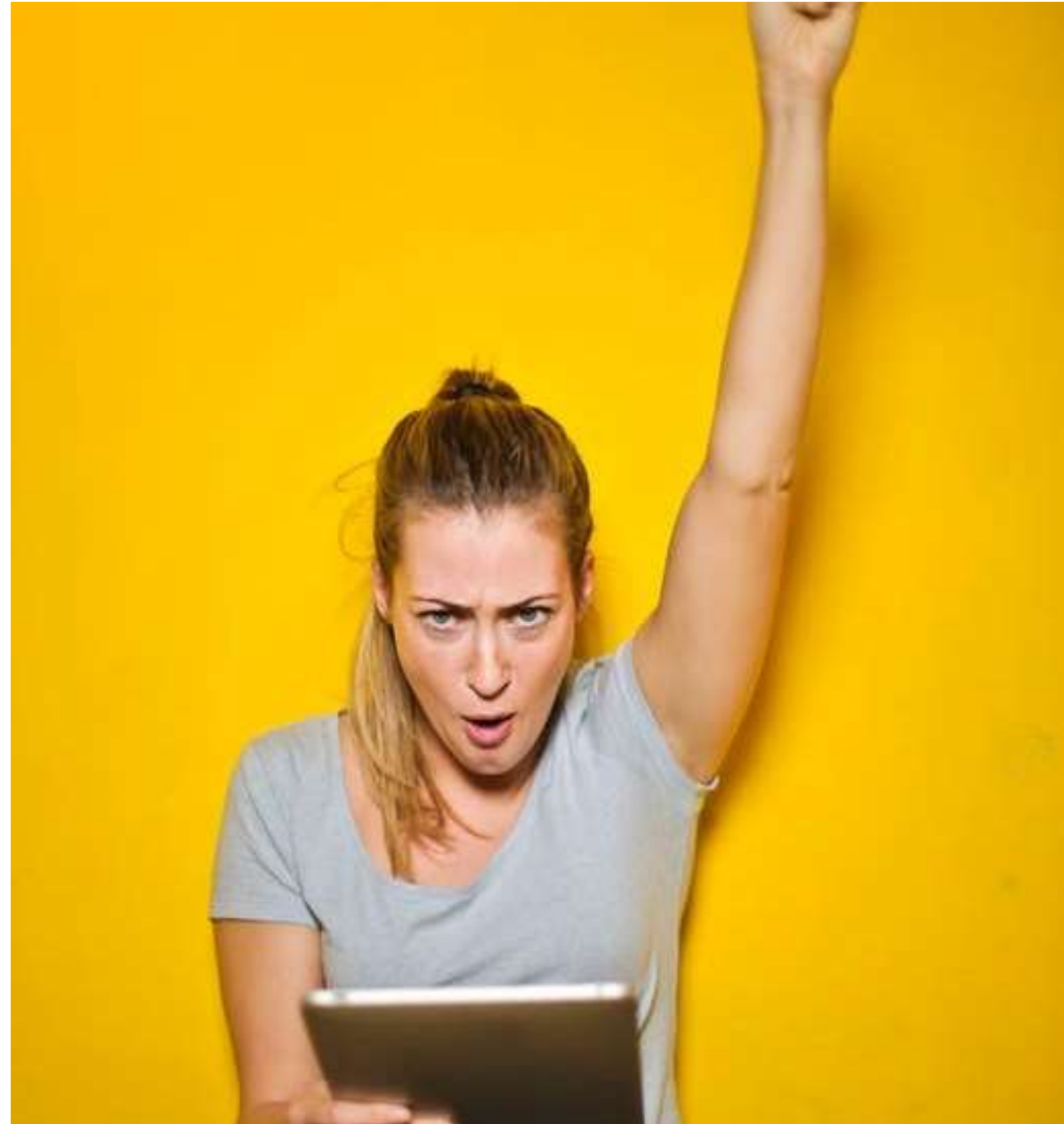


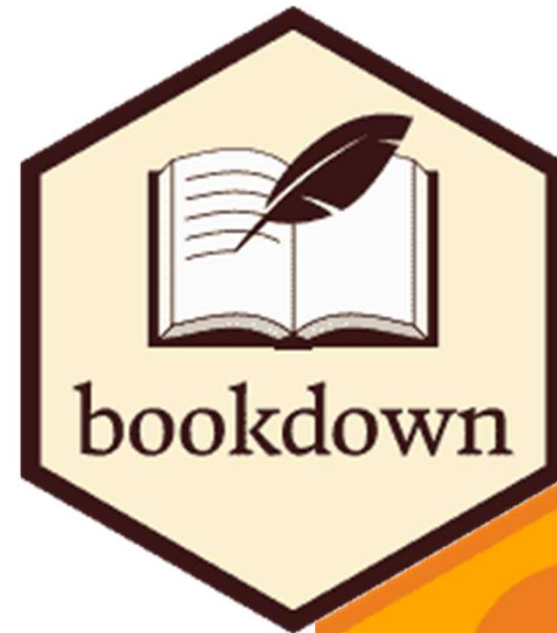
# Automated reporting



# Automated reporting

You can do it!

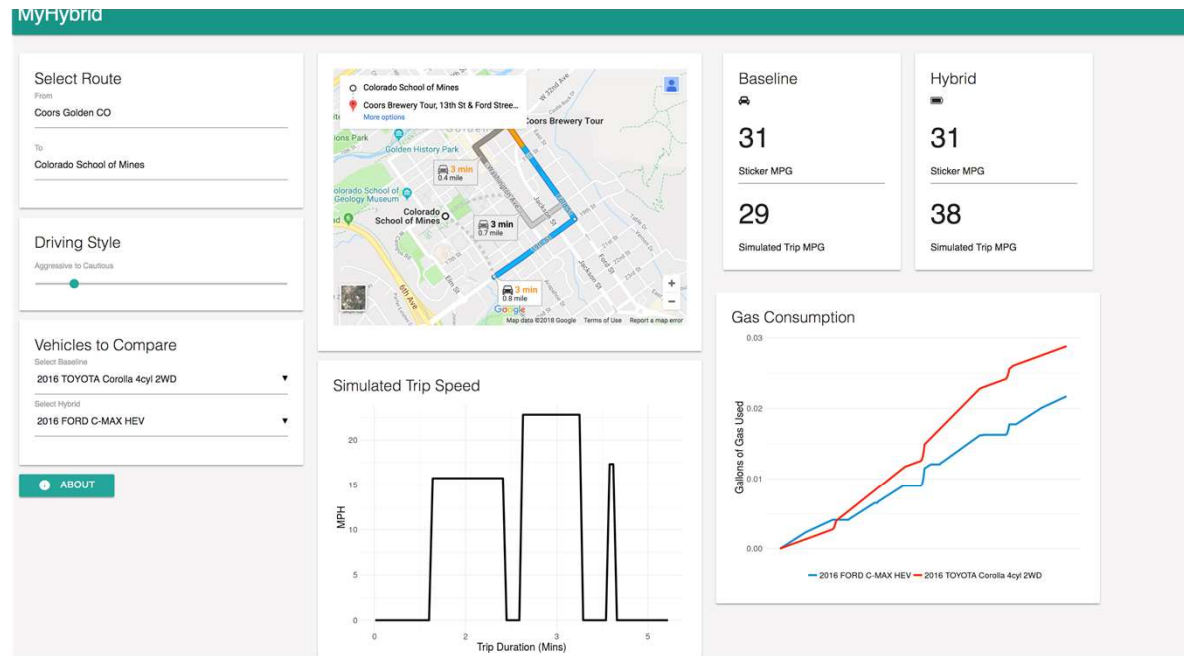




## Writing books and publishing blogs

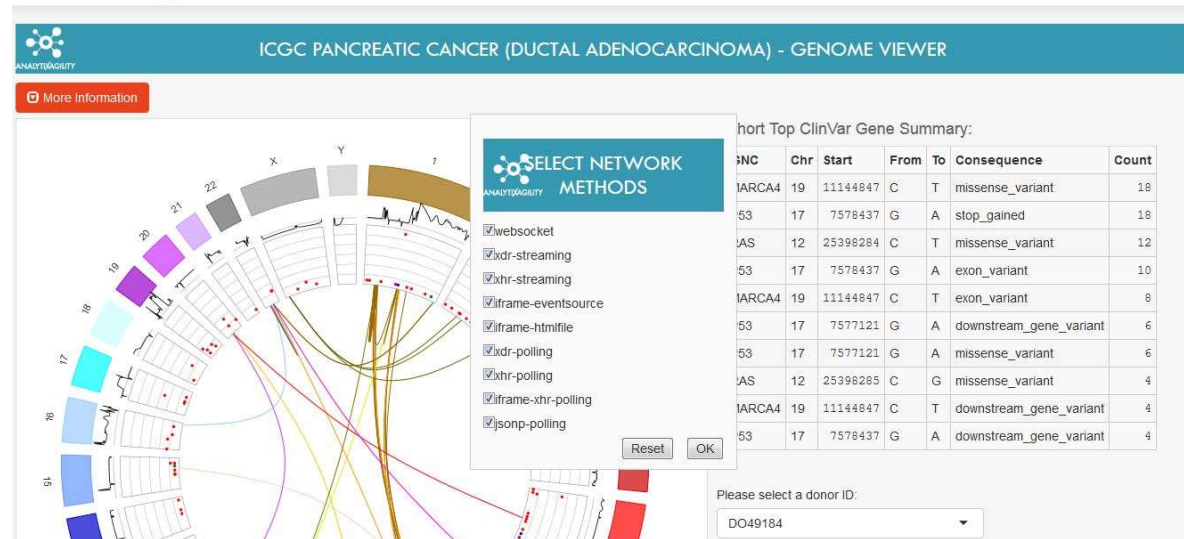


# Shiny apps and dashboards



Shiny from R Studio

[Back to Gallery](#)



# How to start with R?





# Free resources

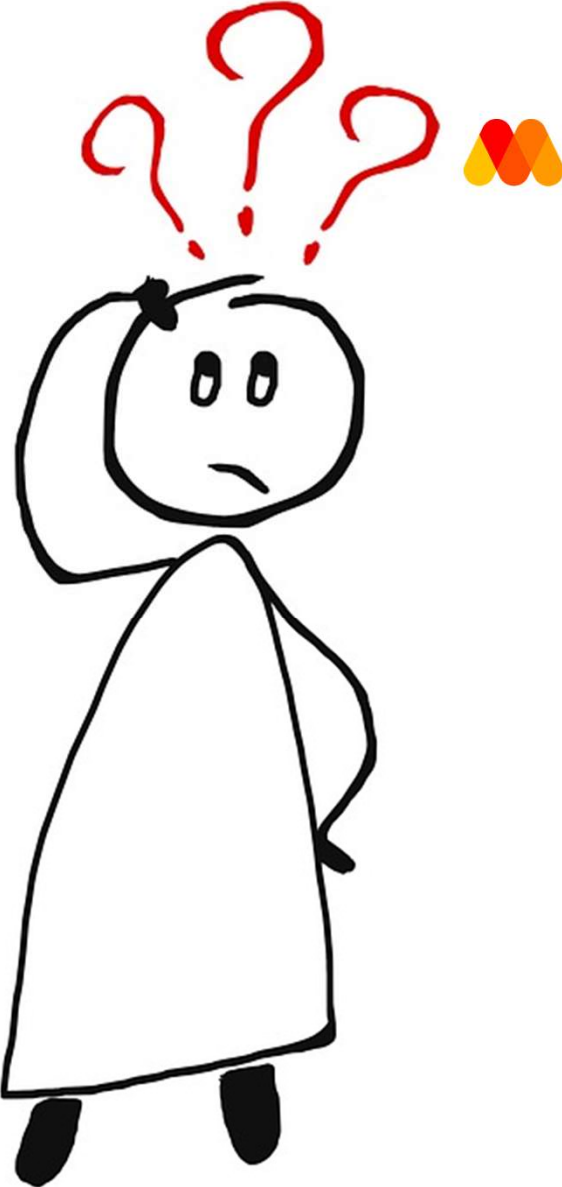
- Free online books, e.g.  
<https://rstudio.com/resources/books/>
- swirl package
- Google 'Learn R for free'
- Blog posts
- YouTube
- <https://awesome-r.com/>
- Twitter (#rstats)
- Meetups (LondonR, R-ladies London)

## Take-home message

- There are lots of good reasons to learn R
- R is intuitive and powerful
- There are lots of free resources out there
- Nothing should stop you from starting to use it NOW



**Questions?**






We empower organisations to make informed decisions, using advanced analytics and AI/ML techniques to meet their objectives and deliver data-driven value. We do this through our unique combination of people, values and strategic priorities.


---

---

 @MangoTheCat

 [www.mango-solutions.com](http://www.mango-solutions.com)

 [info@mango-solutions.com](mailto:info@mango-solutions.com)

 +44 1249 705450

### **Chippenham Office**

Mango Solutions  
2 Methuen Park  
Chippenham  
SN14 0GB

### **London Office**

Dawson House  
5 Jewry Street  
London  
EC3N 2EX