R, What is it good for?
Absolutely Everything

Jasmine Dumas
@jasdumas

Title homage: https://www.youtube.com/watch?v=dpWmlRNfLck the song War by Edwin Starr

“Seriously, it’s good for everything”

–Me

industry, academia & fun….everything

Good vs. Good

🔑

Key Presentation takeaways: Good does not mean great, but Good is better than bad. When we try to compare programming languages we tend to look at the surface components (popular developer influence or singular use cases or language development & design choices) and sometimes we forget the substantive (sometimes secondary) components of what can make a programming language appropriate for use, such as: **versatility, environment and inclusivity (and a strongly opinionated take).**

I'll highlight each of these themes in the presentation *to show and not tell* of why R is good for everything!

Hi Austin!

- Data Scientist @ Simple Finance
- Author of R packages: `ttbbeer`, `shinyLP`, `gramr`, `shinyGEO`
- Slides and Materials available on GitHub: jasdumas/talks/r-data-day-texas

What is this talk about?

This talk is about highlighting the practical, **real and precious** ways that R is useful and good for everything. The Real Gollum is tangible, mischievous, purposeful …

What is this talk not about?

… directly antithetical to James Franco's Fake Gollum: This talk is not about R vs. Python vs. Julia vs. Java vs. Cobol vs. Scala vs. Matlab vs. SAS….wait R is much better than SAS, but that's another debate for another time over beers. No hyped up topics just pure facts & opinions :)

I’m going to show examples of how R is versatile with some cool projects from different domains and objectives:

- bathroom tiles
- turkey chart
- gene expression
- map of GOT

Bathroom Tile Design

ggplot2, dplyr, data.table, magrittr

https://edwinth.github.io/blog/bathroom-with-r/

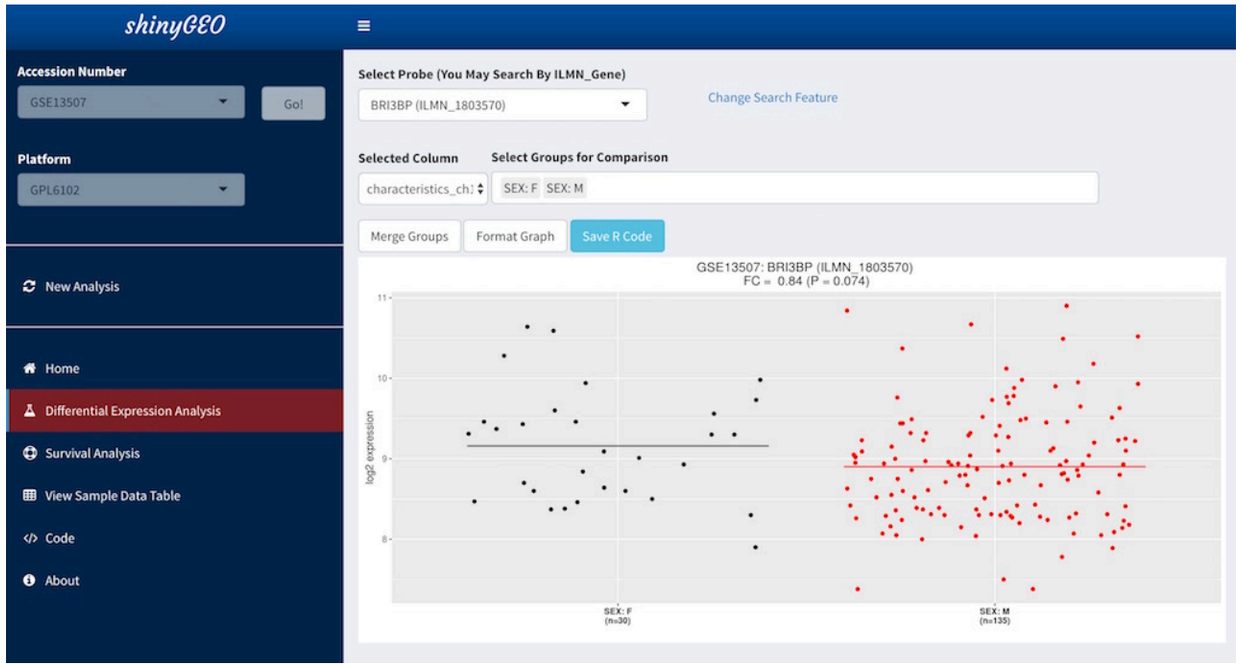
- this was created by Edwin Thoen on his blog: *That's so Random* which outlines how he choose to decorate his bathroom with a random tile design, rather than the mundane, ugly showroom examples he was presented with
- this really emphasizes how many components of our daily lives have a computational approach and could benefit from using R to solve those problems

Thanksgiving Gantt Chart

googlesheets, lubridate, plotly

http://livefreeordichotomize.com/2017/11/12/thanksgiving-gantt-chart/

- this was created by Lucy D'Agostino McGowan who is also speaking today (actually right now) on the 2nd Floor - Amphitheatre 204
- Really interesting approach to visualizing a common problem on Thanksgiving with multiple dishes using a few R packages

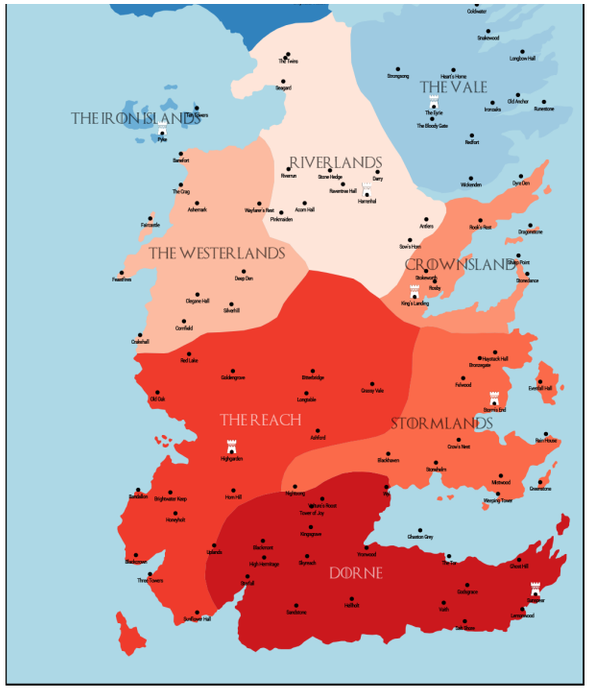


Gene Expression

`shiny, ggplot2, survival, shinyjs`

https://gdancik.github.io/bioinformatics/

- this was co-developed by myself, during Google Summer of Code as a web application to enable Bioinformaticians to perform differential expression on survival analysis with ease
- shiny as a web framework has empowered those with statistical know-how and limited Frontend skills to create interactive analyses



Map of Westeros

`Hmisc, rgdal, tmap, RColorBrewer`

https://www.gokhanciflikli.com/post/game-of-thrones/

- this was made by Gokhan Ciflikli, originally as a Quantitative Text Analysis to mine the scripts of GOT however things get more fun when a shapefile for Westeros exists form the Cartographers' Guild website
- Wide range of GIS and mapping packages for fictional and real maps

Examples of how R is more than an language but also an environment which makes it good for everything.

Fully planned and coherent system

https://www.r-project.org/about.html
The term “environment” is intended to describe R as a fully planned and coherent system, rather than an incremental concoction of overly specific and inflexible tools, as is frequently the case with other data analysis software, ala pandas for python.

And by *planned and coherent*, I’m really speaking to:
- R being focused on statistical techniques, data manipulation, calculations and graphical displays.

Storage

Small, medium, big data

R being memory bound is a frequent (and often contentious) discussion point in our ever expanding world for Big data, however most of your data needs (Small, medium, big data) can be addressed with R using a variety of packages developed to take advantage of parallel computing [multi-threading]. Some of those packages are: ff, bigmemory, data.table, xgboost, Rcpp or rJava.

R can also handle different data formats (csv, json, spss, sas, text, proprietary excel …)

https://jasdumas.github.io/tech-short-papers/Reading_large_datasets.html

Calculations

Familiar binary and logical operators

https://www.statmethods.net/management/operators.html
R has a suite of operators for calculations on arrays, and matrices which lends itself really well to R's vectorized nature.

There are also some handy pipe themed operators: %>%, %<>%, :%in% to reduce code clutter and allow for nice integration with the tidyverse suite of packages.

Data Analysis Tools

Tidyverse and more…

R has a large, coherent, integrated collection of intermediate tools for data analysis, made possible by extension of R's base capabilities via packages on CRAN or GitHub.
Some of the packages include: tidyverse, dplyr, data.table, string (string manipulation), lubridate (date/time manipulation), sf (GIS/Spatial data), bioconductor (Bioinformatics), tidyquant (Financial), Environmental, Surveys…and many more.

Graphs

Publication-ready and interactive

R has graphical facilities for data analysis and display either on-screen or on hardcopy. This is one of my favorite aspects of the R environment because even out-the-box with base R, you are on your way to creating data visualizations that illuminate large data sets.

R is also great for interactive graphics through packages such as: plotly and htmlwidgets

pretty graphs FTW! ggplot2! plotly! htmlwidgets!

Virtual Environments

R + Docker, Packrat

Like you know and are use to with python. R has capabilities to work nicely with supporting cloud based computation (data storage on AWS) as to enable repeatable and portable analyses across machines.

R in the cloud: https://hub.docker.com/u/rocker/
https://www.andrewheiss.com/blog/2017/04/27/super-basic-practical-guide-to-docker-and-rstudio/

I’m going to speak to examples of how R is inclusive, welcoming, and diverse which goes to increase R’s position as being good for everything. This theme is not always at the forefront of new developers thoughts when first selecting programming languages to learn but is proving to be a high indicator of community maturity.

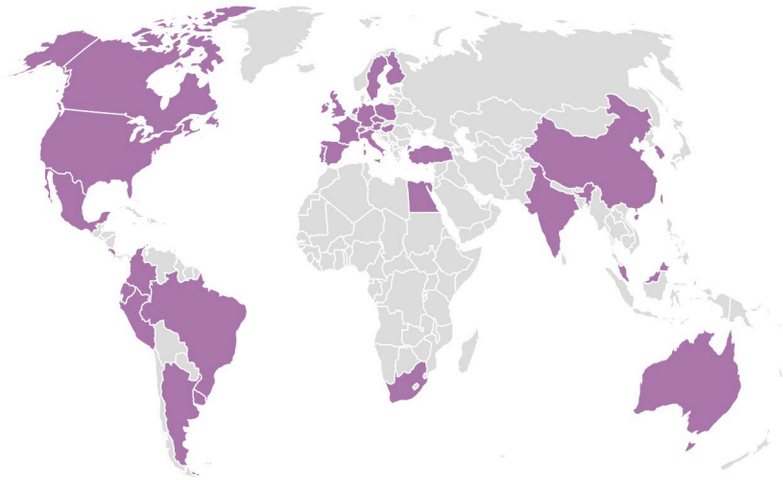
Forwards

http://forwards.github.io/

The Forwards is a Taskforce established by the R Foundation in 2015 to address the underrepresentation of groups such as LGBT, minority ethnic groups, and people with disabilities in the R community.

There a sub teams within the Forwards task group which focus on: community, conference, new useRs, Social media, surveys, and teaching

http://forwards.github.io/about/



R-Ladies

https://rladies.org/

R-ladies is a world-wide organization to promote gender diversity in the r community.

Its an amazing organization of inclusive and genuine people. And there a number of R-Ladies (and founder Gabriela!!!) on the speaker panel today!

#rstats

https://twitter.com/hashtag/rstats

The impact of the twitter R community! Sometimes social media can be a drain however, when anyone new to the community ask about how they can learn about events, solicit help I always point them to using the #rstats hashtag for a method of connection to other R Programmers.

R-Bloggers or RWeekly

https://www.r-bloggers.com/
https://rweekly.org/

R-Bloggers (https://www.r-bloggers.com/) and RWeekly (https://rweekly.org/) are nice aggregators so you don't sweat keeping up with tons of blogs

Blogging is live artifacts for sharing - paraphrasing from David Robinson (@drob) who is also speaking this afternoon, if you have some cool R analyses on your local computer, is essentially useless compared to anything else in the world (blog posts, papers, open source code, tweets even)

rOpenSci

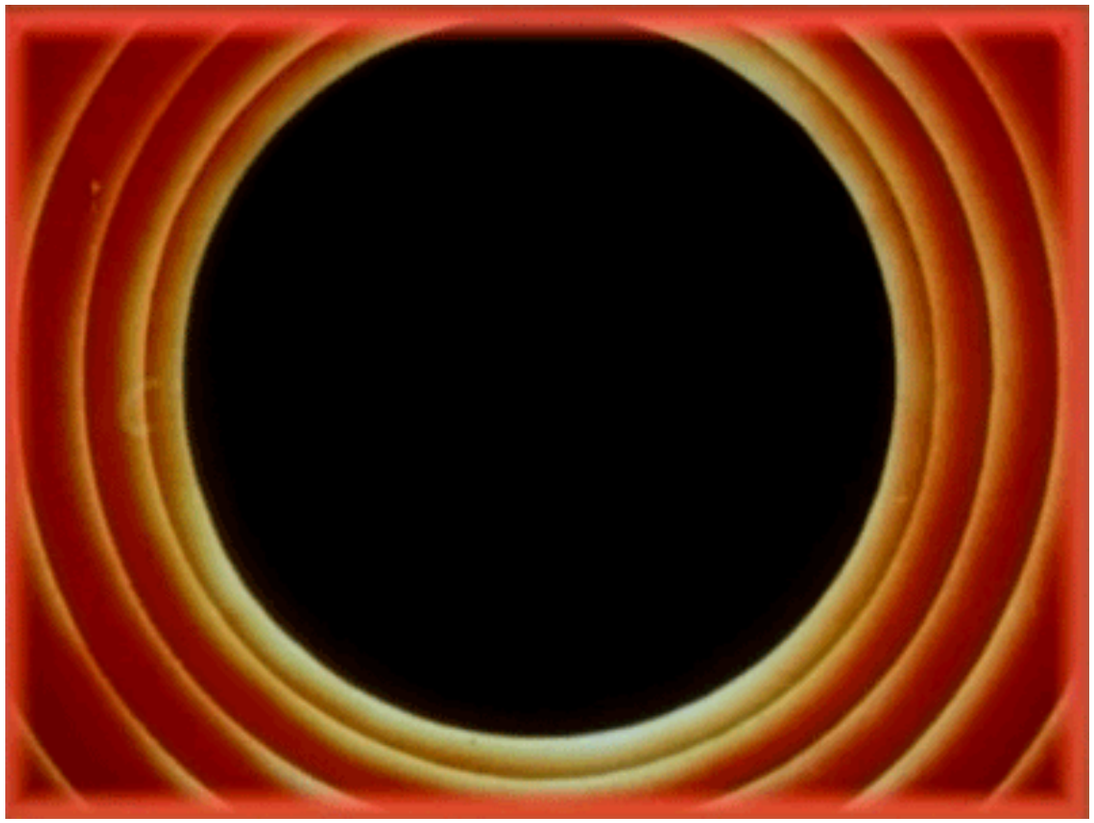
https://ropensci.org/

rOpenSci builds software with a community of users and developers, and help to educate scientists about transparent research practices including: hosting a annual unconference, packages review and community support with open calls - many ways in being advocates for the R community.
https://ropensci.org/

R is good for...

Wrapping things up with a recap…I tried to hammer across several points about R being versatile in domain application, a fully planned environment alongside the language and R being welcoming and inclusive. I would say that makes R good for..

EVERYTHING



Questions?