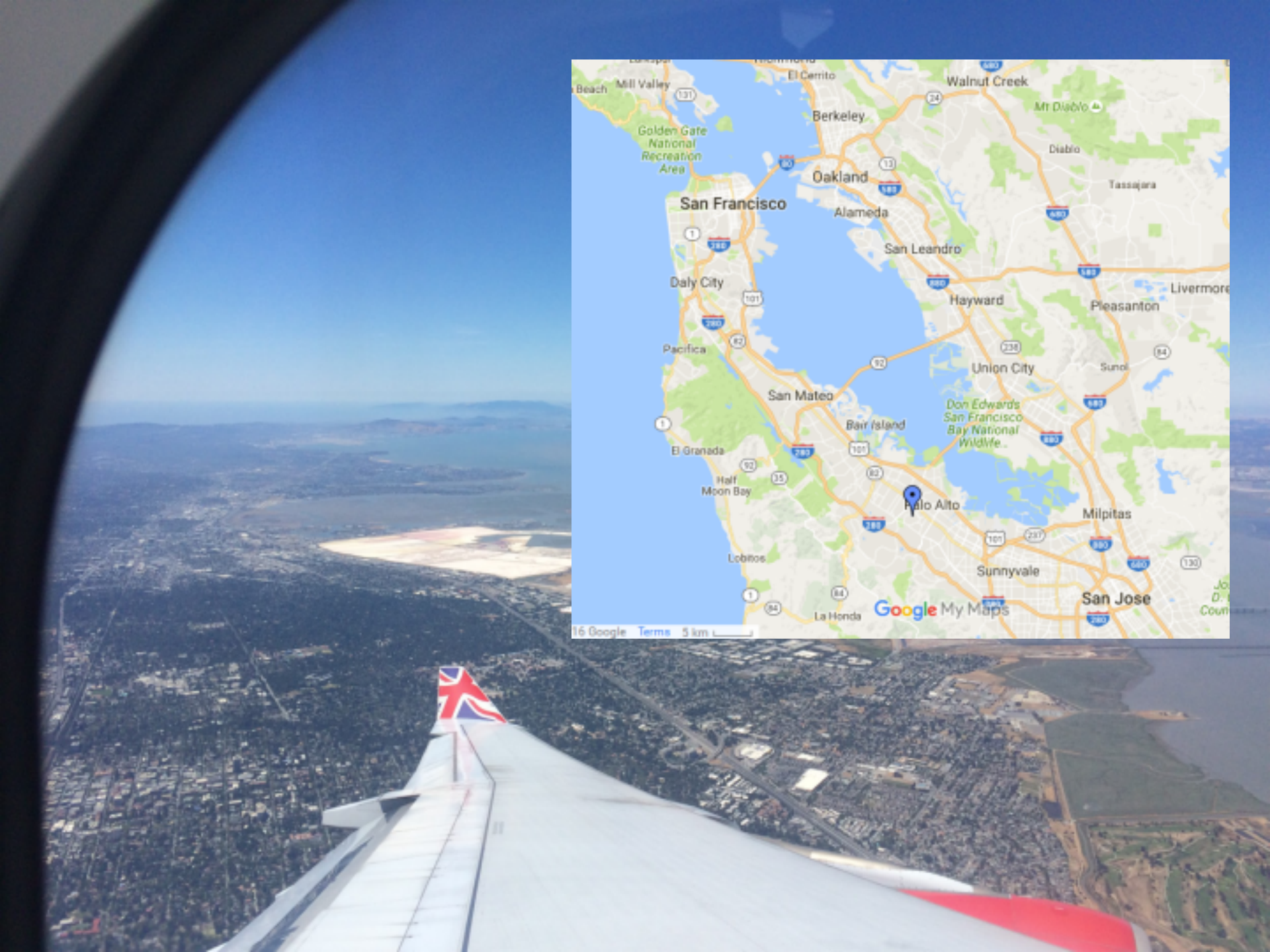
2nd Aug '16

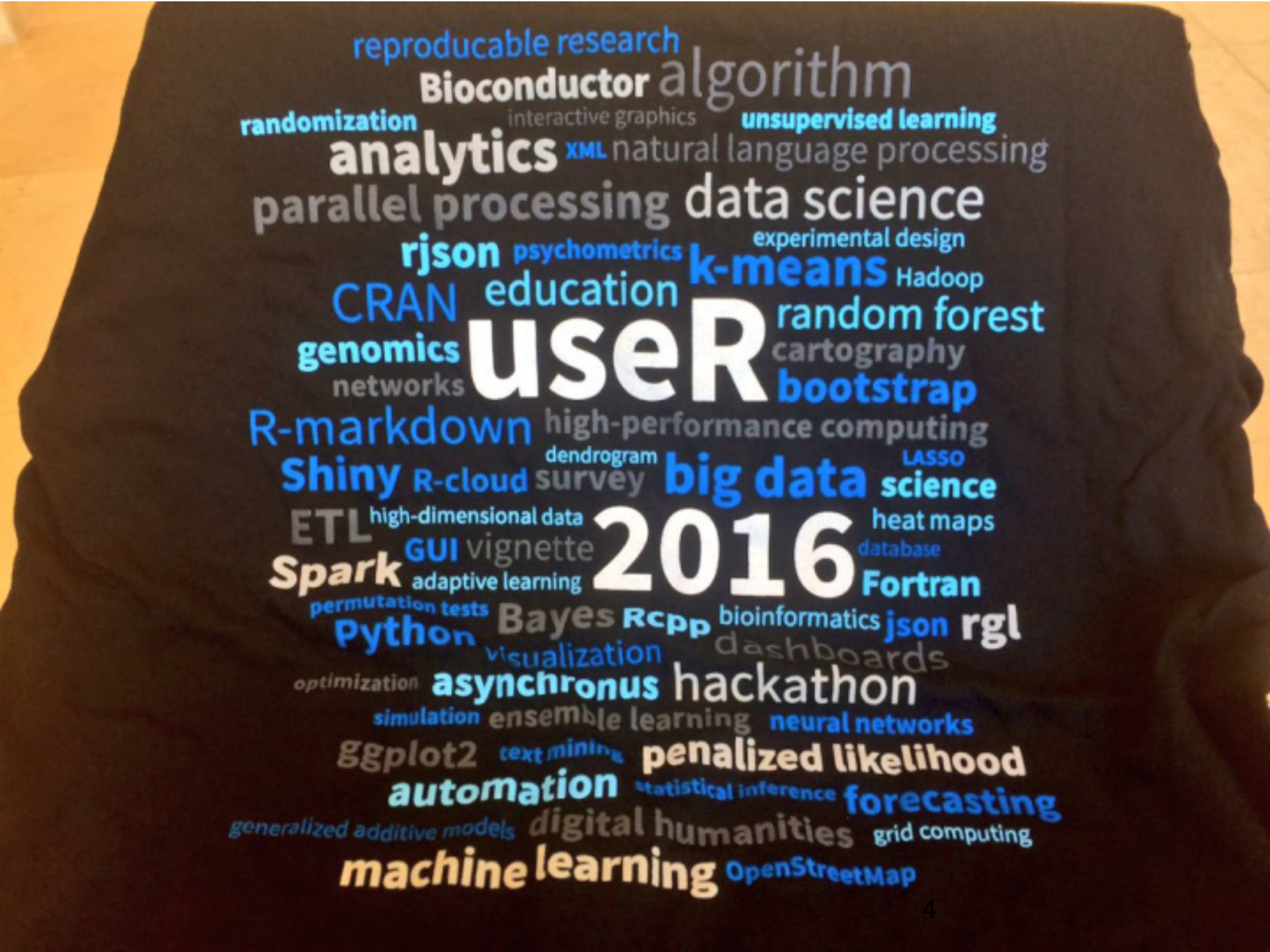
Review

Alice Daish & Hannah Frick

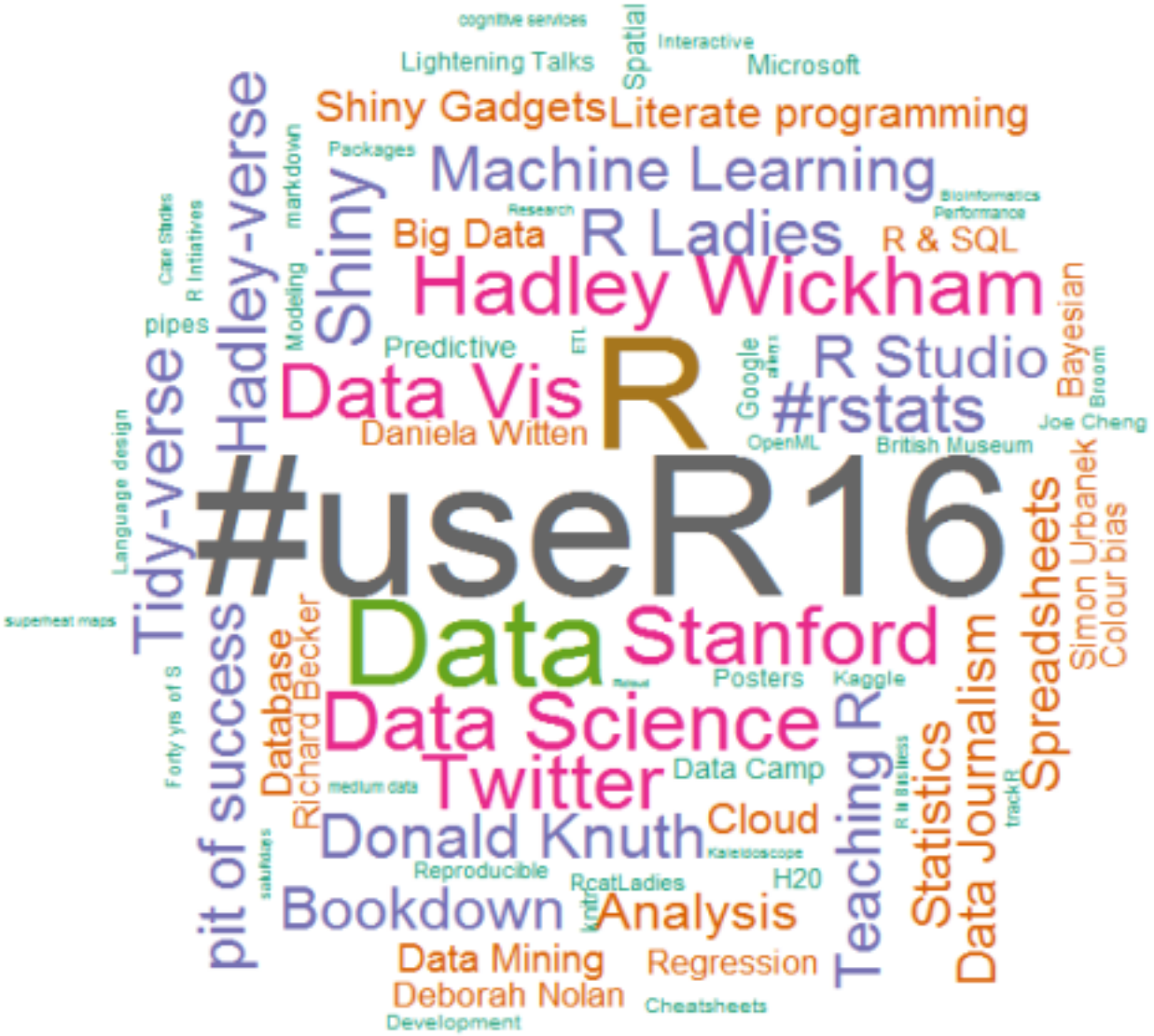
-Ladies London


El Cerrito
Walnut Creek
Mill Valley
Berkeley
Mt Diablo
Golden Gate National Recreation Area
Diablo
Oakland
Tassajara
San Francisco
Alameda
San Leandro
Daly City
Hayward
Pleasanton
Pacifica
Union City
San Mateo
Don Edwards San Francisco Bay National Wildlife
Bair Island
El Granada
Half Moon Bay
Palo Alto
Milpitas
Lobitos
Sunnyvale
San Jose
La Honda
Google My Maps
16 Google
Terms
5 km


USE R
2016
CONFERENCE


reproducable research
Bioconductor
algorithm
randomization
interactive graphics
unsupervised learning
analytics
XML
natural language processing
parallel processing
data science
experimental design
rjson
psychometrics
k-means
Hadoop
CRAN
education
random forest
genomics
useR
cartography
networks
bootstrap
R-markdown
high-performance computing
dendrogram
LASSO
Shiny
R-cloud
survey
big data
science
ETL
high-dimensional data
heat maps
GUI
vignette
2016
database
Spark
adaptive learning
Fortran
permutation tests
Bayes
Rcpp
bioinformatics
json
rgl
Python
visualization
dashboards
optimization
asynchronus
hackathon
simulation
ensemble learning
neural networks
ggplot2
text mining
penalized likelihood
automation
statistical inference
forecasting
generalized additive models
digital humanities
grid computing
machine learning
OpenStreetMap

Alice's useR2016 wordcloud :)



useR! 2016 Program at a glance

	Monday, June 27, 2016	Tuesday, June 28, 2016	Wednesday, June 29, 2016	Thursday, June 30, 2016
8:00-8:30	Registration (all day)	Registration (all day)	Registration (morning)	
8:30-9:00		Opening Session		
9:00-10:00	Morning Tutorials (including coffee break)	Richard Becker	Hadley Wickham	Deborah Nolan
10:00-10:30		Coffee Break including Poster Exhibits (display at Sponsor Pavilion)		
10:30-11:00		• Bayesian (Barnes) • R in Business (Econ 140) • Bioinformatics (Lane) • Kaleidoscope (McCaw) • Performance (Siepr 130)	• Statistics Methods (Barnes) • Performance (Econ 140) • Bioinformatics (Lane) • Kaleidoscope (McCaw) • Lightning Talks (Siepr 120) • Database (Siepr 130)	• Teaching (Barnes) • Lightning Talks (Econ 140) • Statistics & Big Data (Lane) • Kaleidoscope (McCaw) • Graphics (Siepr 130)
11:00-12:00				
12:00:1:00	Lunch	Lunch including Poster Exhibits (displayed at Sponsor Pavilion)		
1:00-2:00	Afternoon Tutorials (including coffee break)	• Regression (Barnes) • R & Other Languages (Lane) • Kaleidoscope (McCaw) • Case Study (Siepr 120) • Teaching (Siepr 130)	• Reproducible Research (Barnes) • Generalized Mixed Models (Econ 140) • Sponsor Session (Lane) • Kaleidoscope (McCaw) • Spatial (Siepr 120) • Packages & Development (Siepr 130)	• Sponsor (Lane) • Lightning Talks (McCaw) • Lightning Talks (Siepr 130)
2:00-2:15				Simon Urbanek
2:15-2:30				
2:30-3:00		Coffee Break including Poster Session (displayed at Sponsor Pavilion)		
3:00-3:30				Closing Remarks
3:30-3:45		Donald Knuth	Daniela Witten	Conference ends
4:00-4:15	Short Break			
4:15-4:30	R Initiatives			
4:30-4:45		Short Break	Bus to Cruise (board by 5:00 latest)	
4:45-5:30		• Analytics (Barnes) • Miscellaneous (Econ 140) • Interactive (Lane) • Kaleidoscope (McCaw) • Miscellaneous (Siepr 120) • Machine Learning (Siepr 130)		
6:00-6:15			Cruise and Conference Dinner (Sponsored by Microsoft) Return to campus at 10:30	
6:30-8:30		Welcome Reception (Sponsored by RStudio)		

Machine Learning Tutorial

Erin LeDell @ledell (R-Ladies SF)

six popular supervised machine learning methods:

- Classification and Regression Trees (CART)
- Random Forests (RF)
- Gradient Boosting Machines (GBM)
- Generalized Linear Models (GLM)
- Deep Neural Networks (DNN)
- Stacking / Super Learner (SL)

Here are some practical, related topics we will cover for each algorithm:

- Dimensionality Issues
- Sparsity
- Normalizaion
- Categorical Data
- Missing Data
- Class Imbalance
- Overfitting
- Software
- Scalability

https://github.com/ledell/useR-machine-learning-tutorial

Other Tutorials

- Using Git and GitHub with R, RStudio, and RMarkdown - Jennifer Bryan

- Handling and Analyzing Spatial, Spatiotemporal and Movement Data - Edzer Pebesma

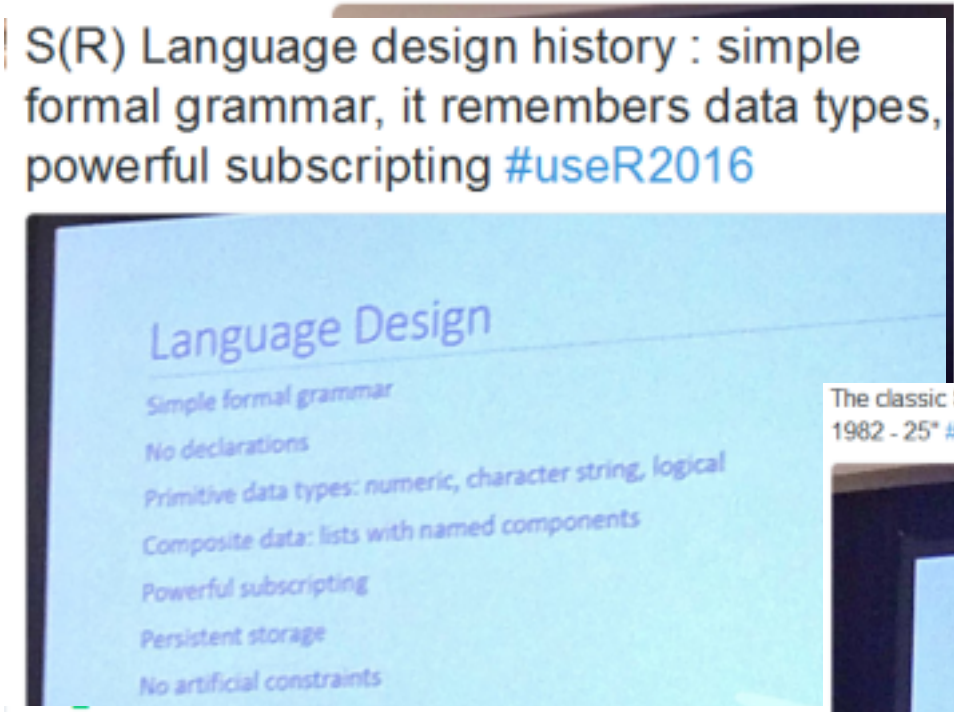
- Dynamic Documents with R Markdown - Yihui Xie

Keynotes

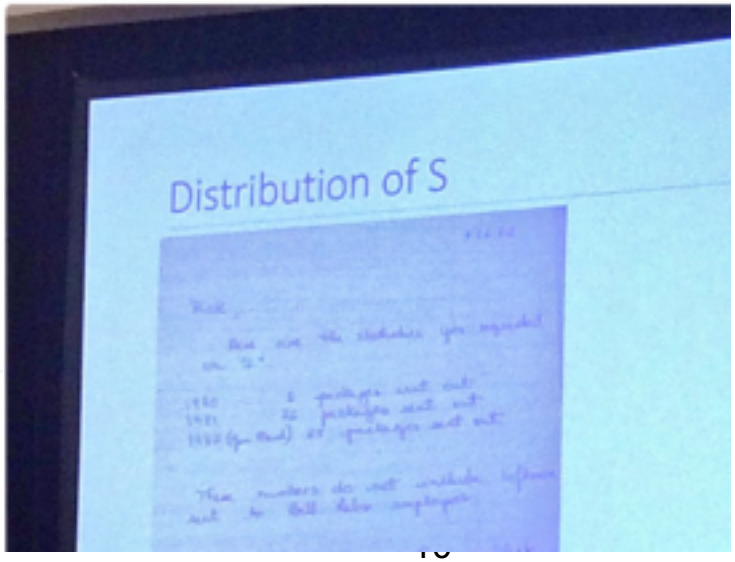
- Forty Years of S - Richard Becker
- Literate Programming - Donald Knuth
- Statistical Thinking in a Data Science Course - Deborah Nolan
- Tidy tools for data science - Hadley Wickham
- Flexible and Interpretable Regression Using Convex Penalties - Daniela Witten
- RCloud - Simon Urbanek

Forty years of S - Richard Becker

What was coding like back then? Well batch coding was punch cards and print outs #useR2016



The classic S download "letter of tape mail outs 1980 - 8, 1981 - 26, 1982 - 25" #useR2016



Hadley Wickham @hadleywickham · Jun 28
2 of the 5 original authors of S were women #rstats #user2016
109 151 •••

Alice Data @alice_data · Jun 28
Why has S survived 40yrs? "We built it for us, no junk code, re-written 3times, and then along came R" #useR2016

Literate Programming - Donald Knuth

Interactive graphics - Hadley Wickham

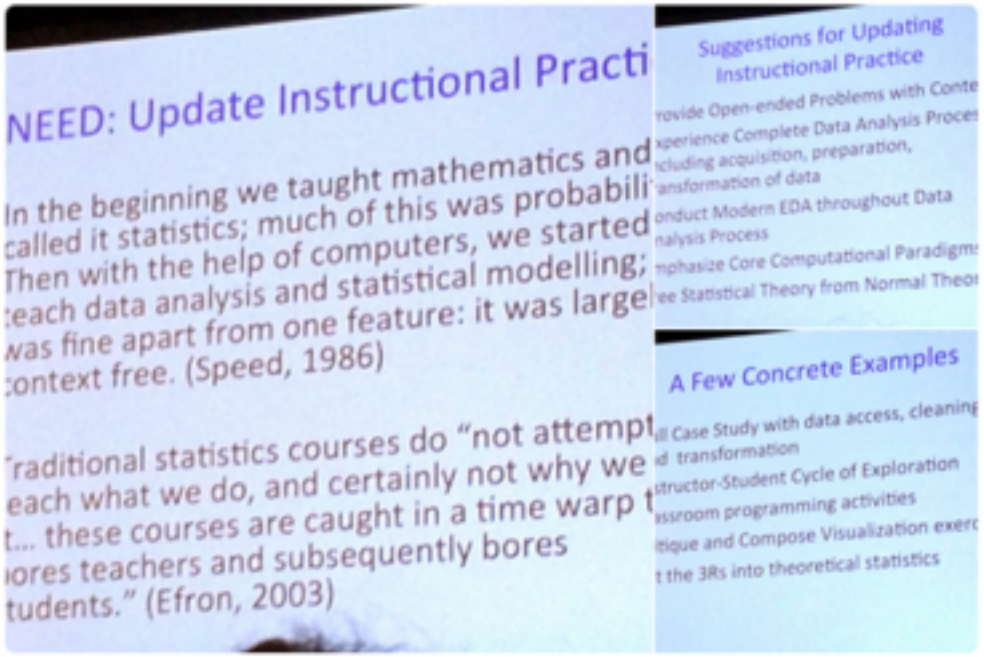
.@hadleywickham keynote at #useR2016
"my goal is to make a pit of success" by
tidy data and tidy tools #DataScience

Stats Thinking in Data Science Course - Deborah Nolan

Stats Thinking in Data Science Course - Deborah Nolan

Just imagine #stats course without simplified scenarios, canned data, non-coding, & always normal distrib #useR2016

Teach #stats in #DataScience update your practises & here R some suggestions & examples via Deborah Nolan #useR2016



Spreadsheets - Jenny Bryan

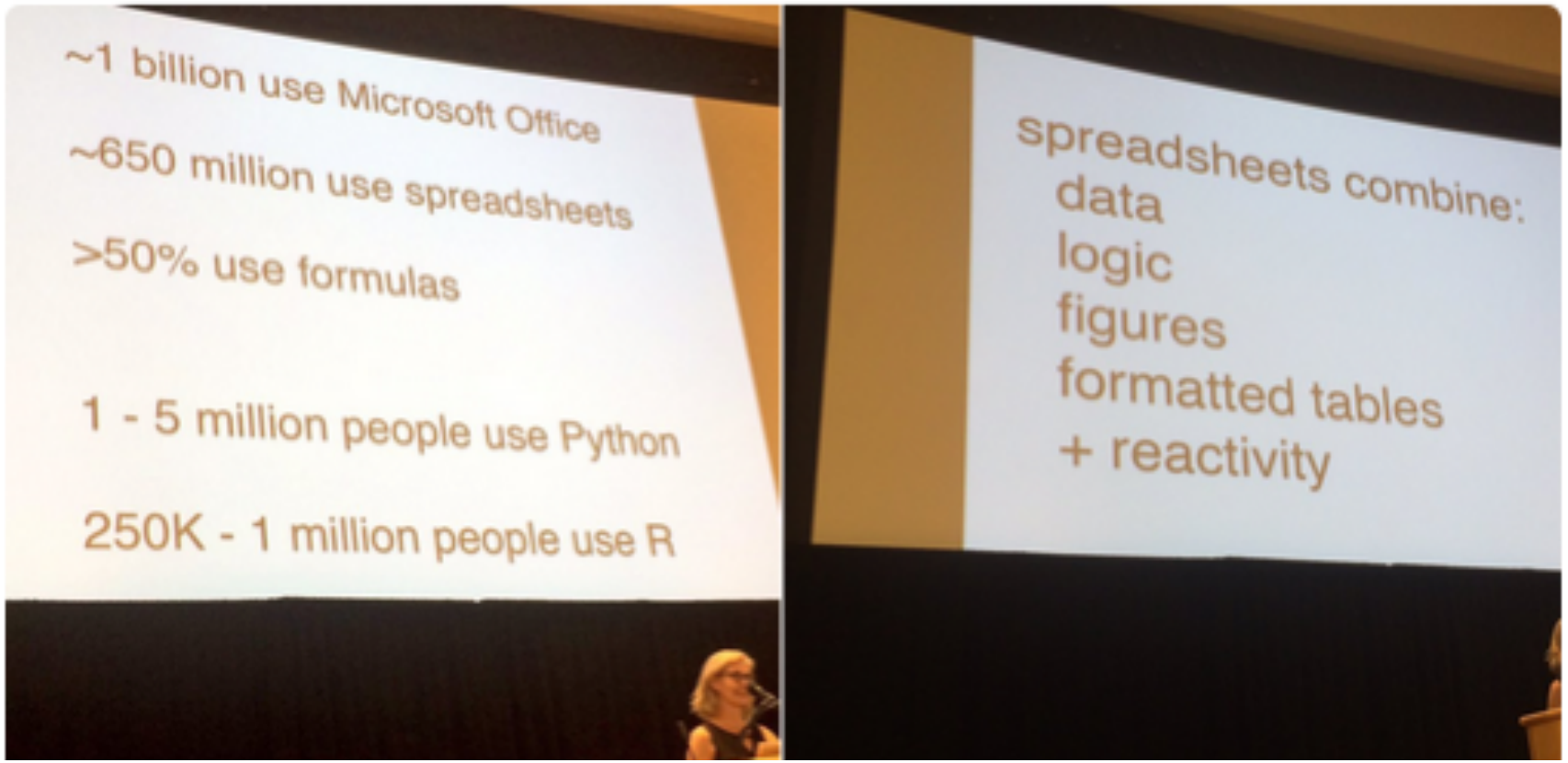
Alice Fav Presentation

Alice Data @alice_data · Jun 28

Ahh spreadsheets talk from @JennyBryan @STAT545 love it "some of my best friends use spreadsheets" #useR2016

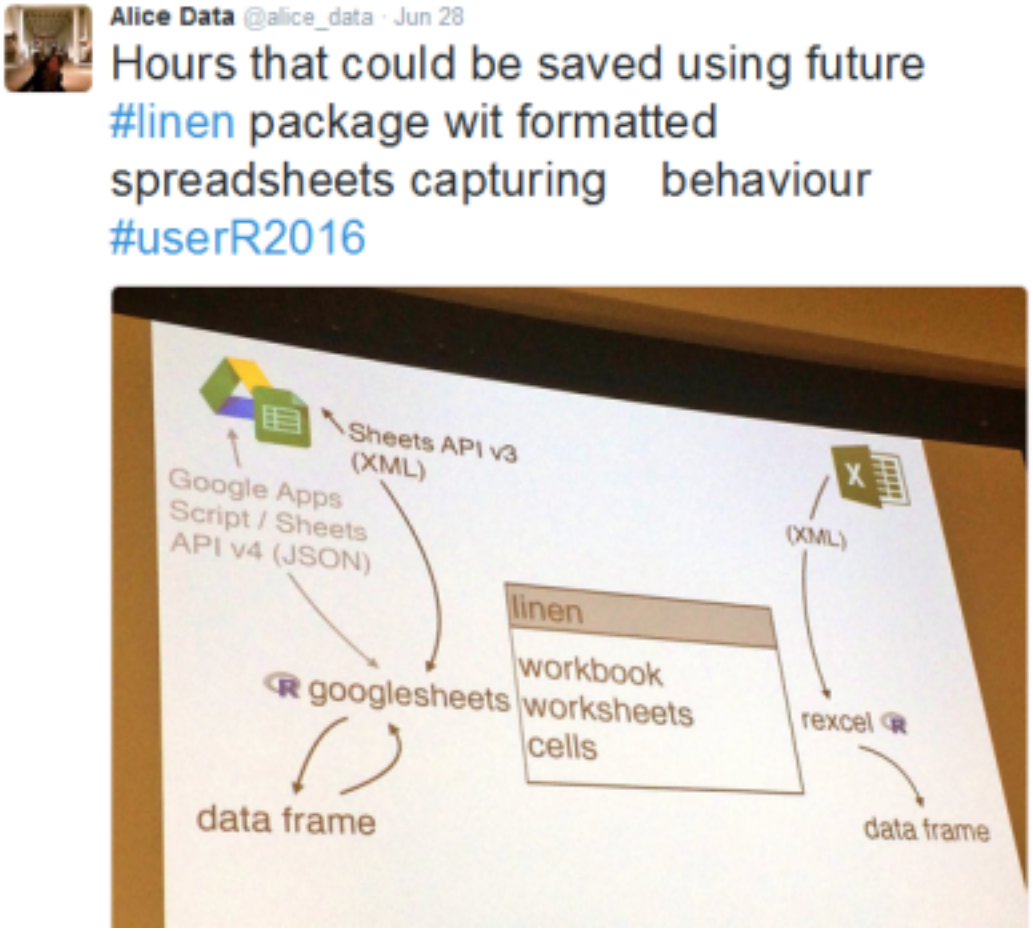
Alice Fav Presentation

#rstats are drastically outnumbered by ~650m using spreadsheets- why? bc it can combine data & is reactive #useR2016



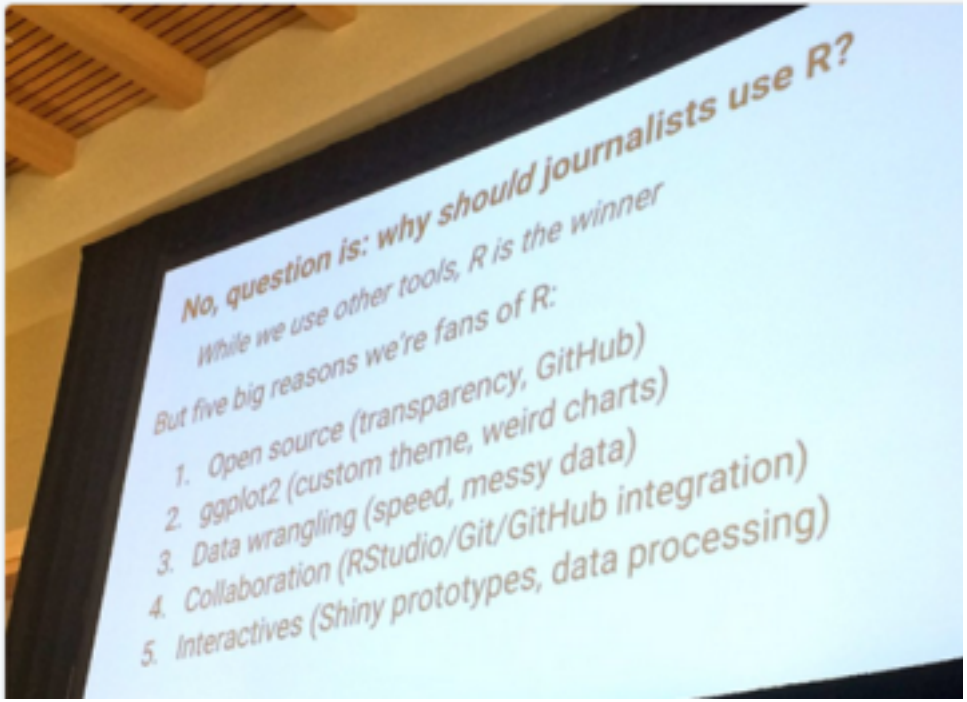
Alice Fav Presentation

Can't wait
for @rsheets
to be a reality



Data Journalism

Why do #datajournalist use #rstats? #useR2016

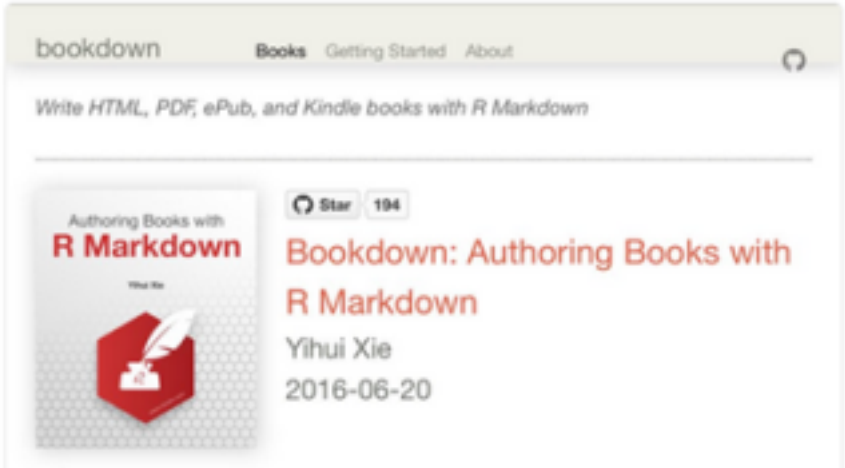


Data Journalism

Bookdown

Excellent @hadleywickham karaoking @xieyihui gif slides on authoring using #rmarkdown #useR2016

Check out #bookdown for writing books in #Rmarkdown bookdown.org #useR2016 #rstats



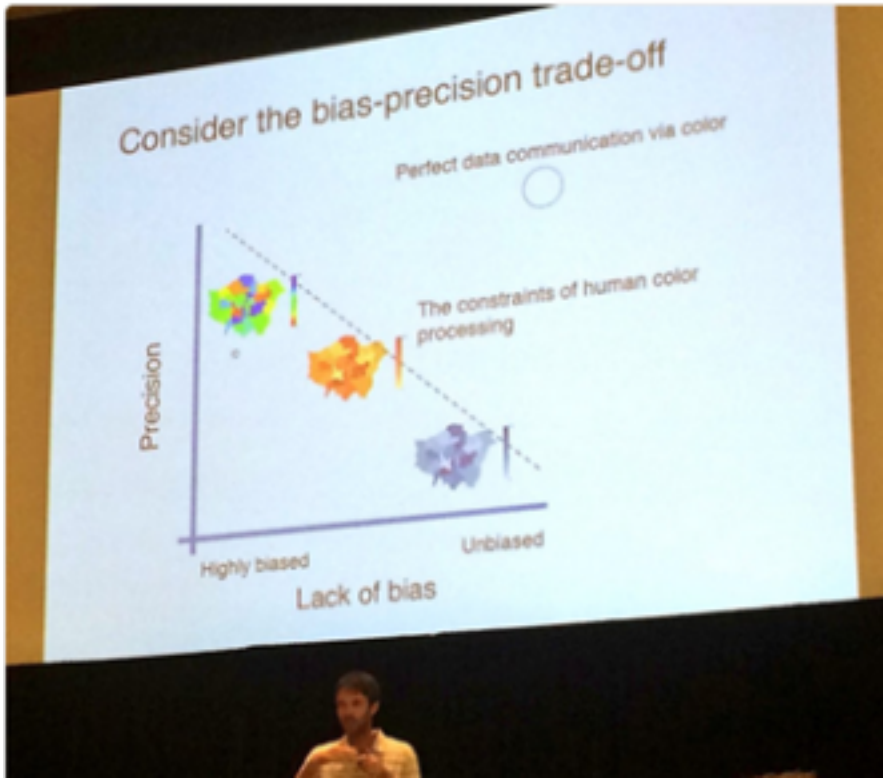
Careful with colours

Check your colours! If your worried about bias & precision interpretation of colours exists! #useR2016

@will_cornwell

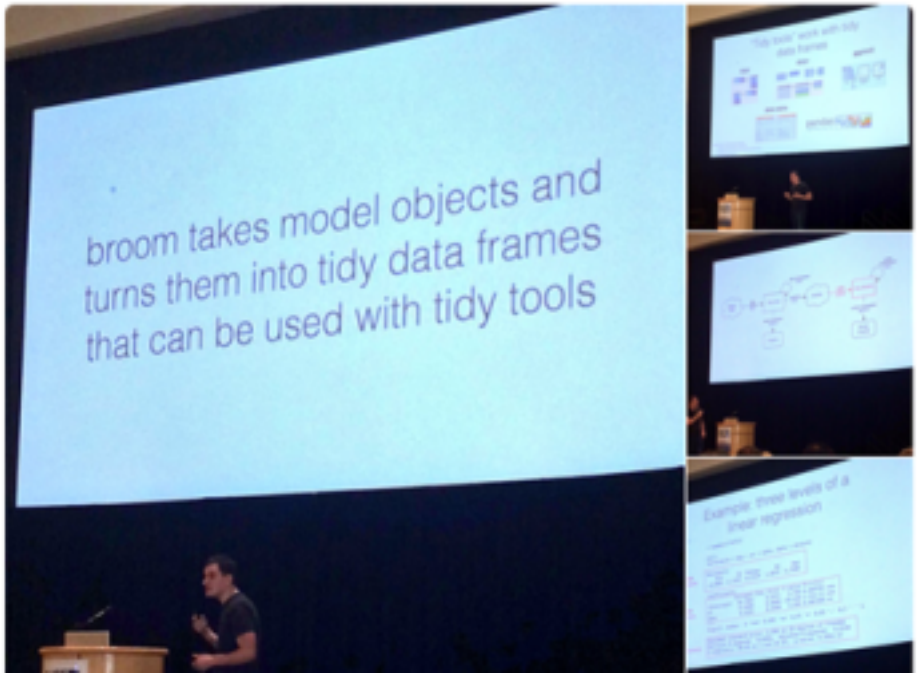
Wow learnt something new Russians have 2 types of Blue (light & dark) compare to others who only had one "blue" #dataviz #fact #useR2016

Think about the bias-precision trade-off #rstats #dataviz #useR2016



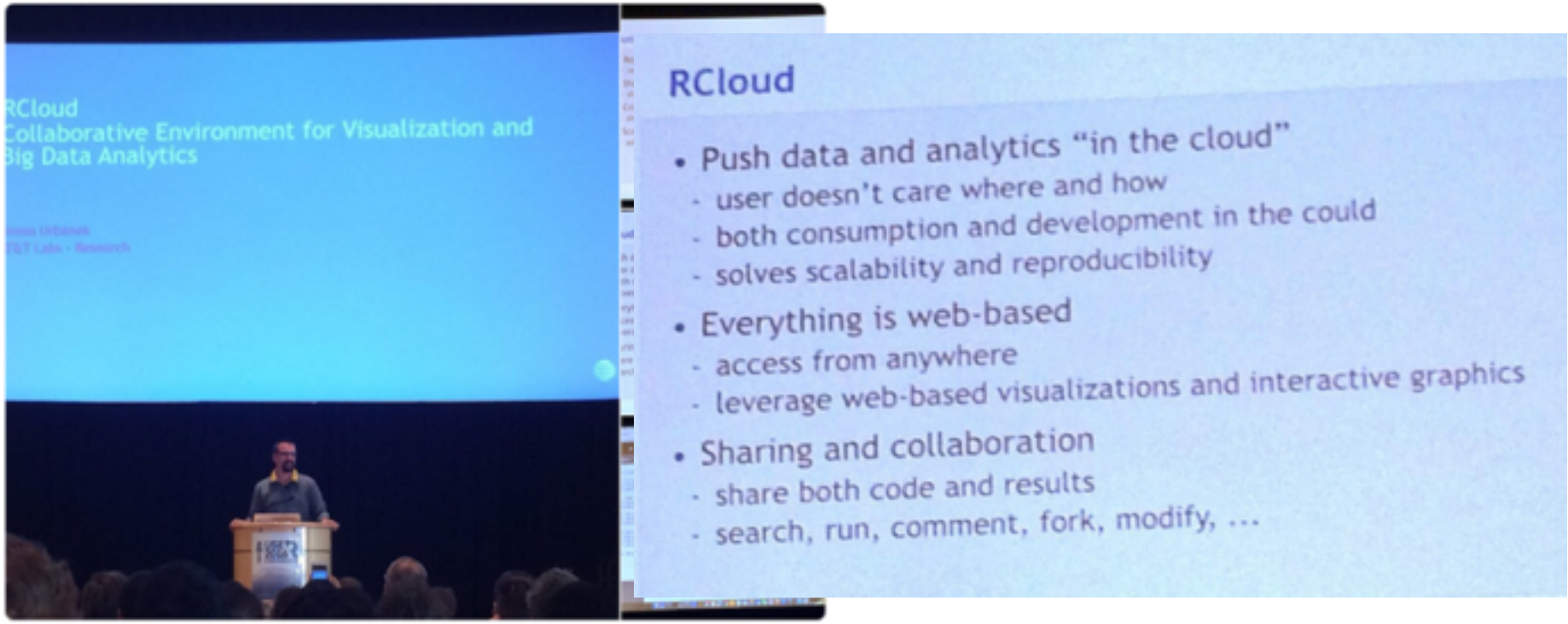
Broom

Love #broom ! I need to do some tidying I make a lot of mess! messy hacker! #useR2016



RCloud

Final #useR2016 keynote from Simon Urbanek "RCloud" - #rstats
anywhere anytime via the cloud !



Try Kaggle

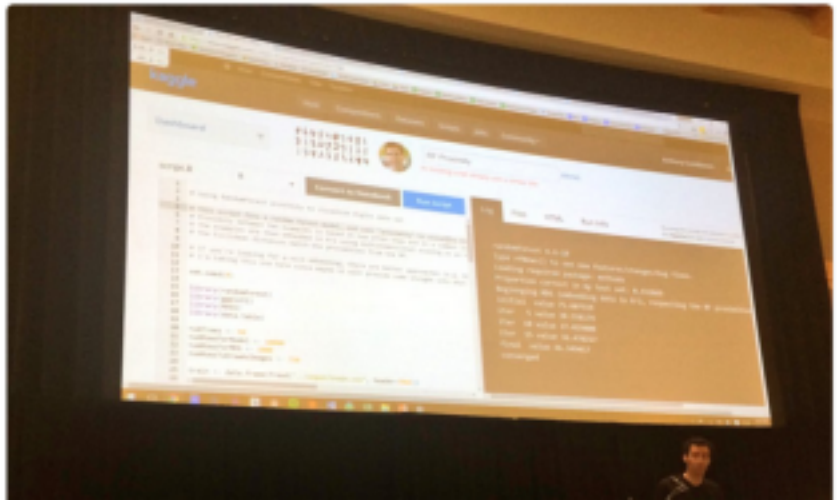
Learning about @kaggle 1000 #machine learning comp per month and #rstat is the second most used scripts #useR2016

.@kaggle supportive of dataset sharing, reproducibility & being a resource to see what other people have done wit #rstats packages #useR2016

1 1

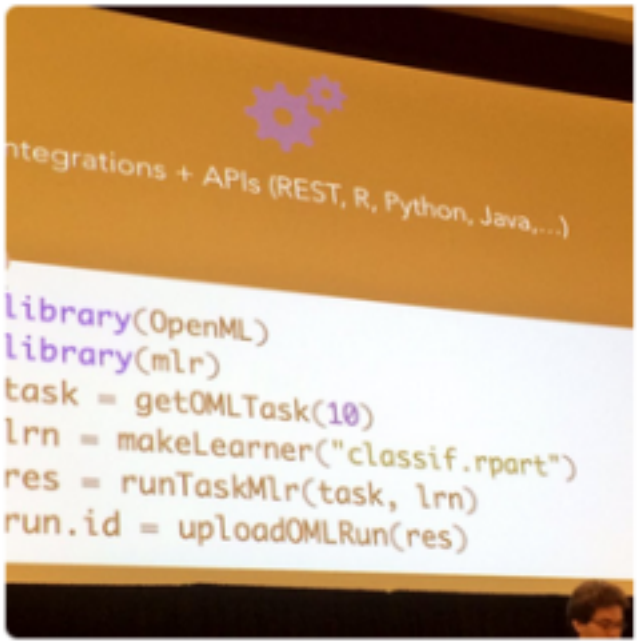
Alice Data @alice_data · Jun 28

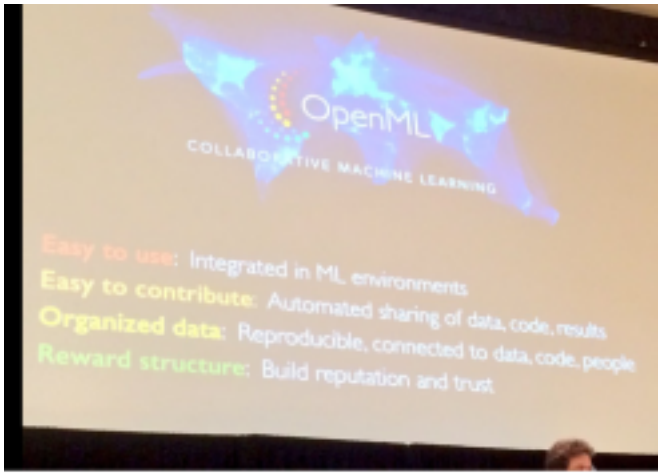
.@antgoldbloom CEO @kaggle with a community of 450k #datascientist with a one click fork of scripts #useR2016



OpenML

How to get started with open machine learning in #rstats #OpenML openml.org #useR2016



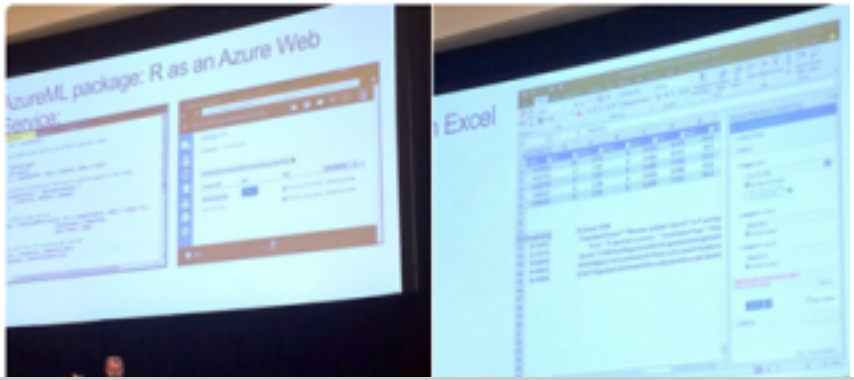


Alice Data @alice_data · Jun 28
Imagine a world where we collaboratively on web scale and real time do #MachineLearning #OpenML #useR2016

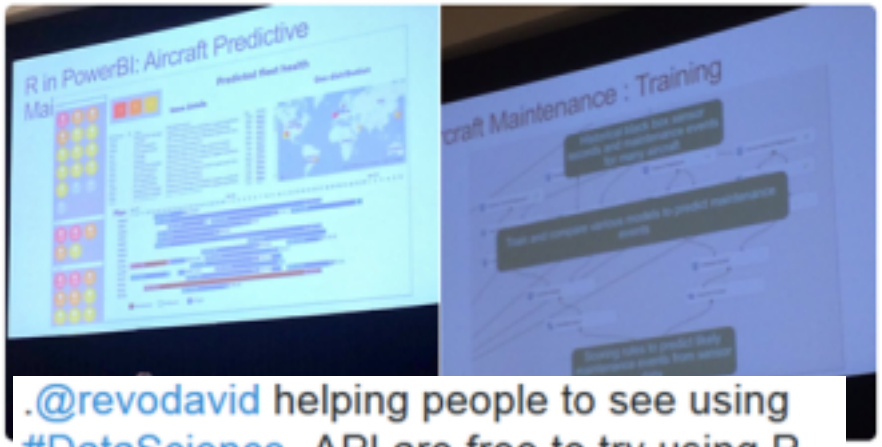
R and Microsoft

#rstats and @Microsoft building MicrosoftR making R fast @revodavid #useR2016

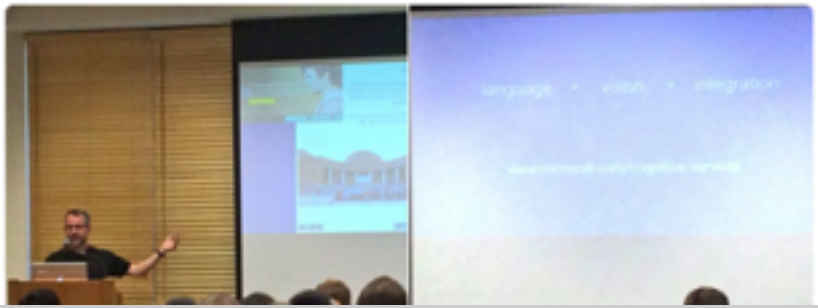
This is amazing from #rstats as web service and then doom into excel #useR2016



#powerBI driven by #rstats and then workflow creating R in Azure ML and turn it into API #useR2016

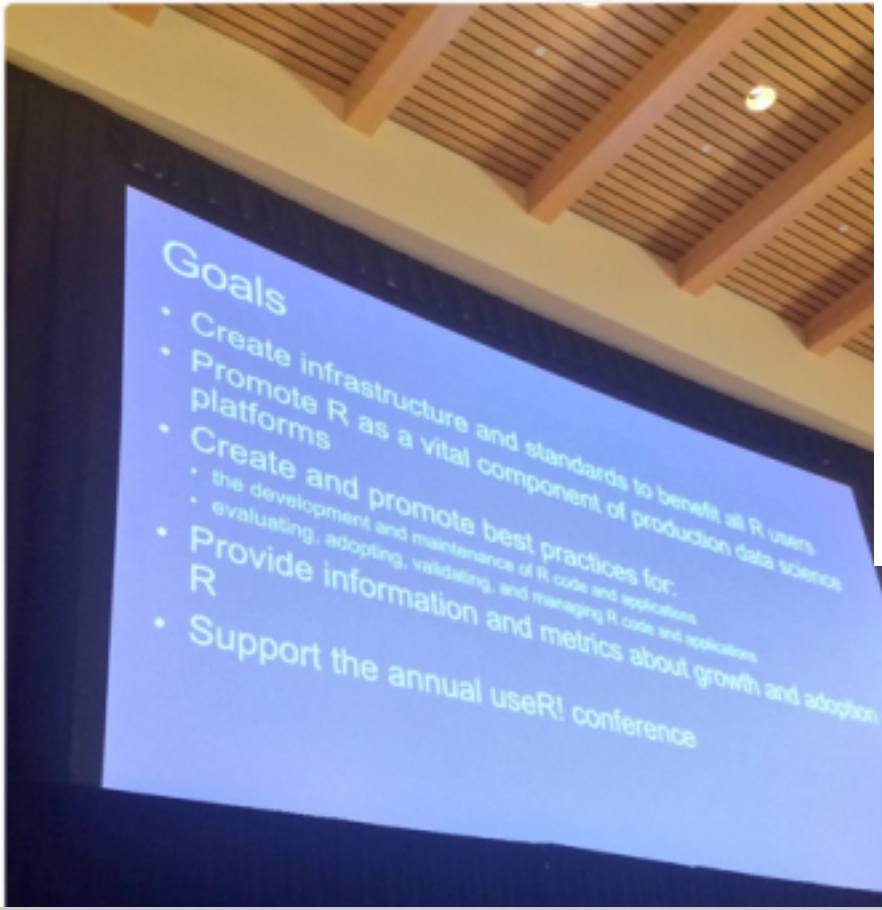


.@revodavid helping people to see using #DataScience -API are free to try using R microsoft.com/cognitive-serv… #useR2016



R-Consortium

Goals of @RConsortium #useR2016



What is @RConsortium ? joseph rickert @microsoft talking about the drive from business of #rstats growth #user2016

Top areas of discussion

- Teaching R
- Tidy data / tidyverse
- Cloud
- R Consortium & Sponsors
- Machine Learning
- R-Ladies

Data Camp

Improve your skills - and your career

No matter what industry you're in, learning how to analyze and understand your data is critical. That's why DataCamp provides you with the tools to learn the data science skills you need to start your own data projects.

Check out @DataCamp excellent searchable interface for #rstats packages & teaching rdocumentation.org #useR2016

Resources

https://channel9.msdn.com/Events/useR-international-R-User-conference/useR2016
http://user2016.org/
http://user2016.org/files/abs-book.pdf

#useR2016
@user_stanford

Resources

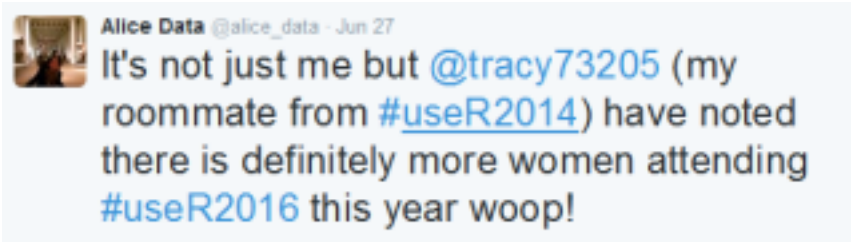
https://channel9.msdn.com/Events/useR-international-R-User-conference/useR2016
http://user2016.org/
http://user2016.org/files/abs-book.pdf

#useR2016
@user_stanford

Next year @UseR_Brussels
July 4th - 7th, 2017

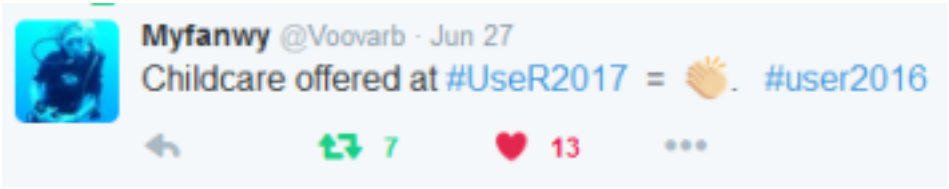
R-Initiatives

RWomenTaskforce



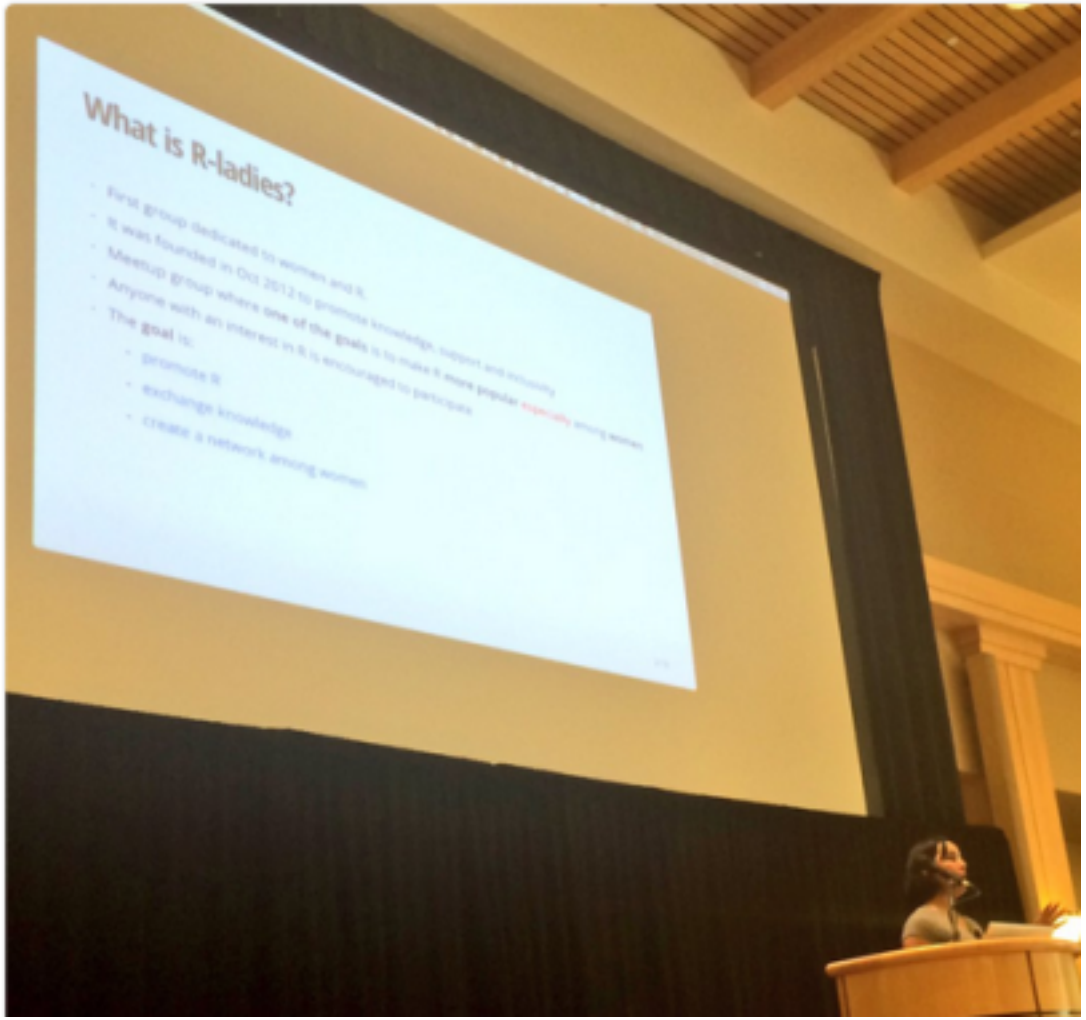
http://rpubs.com/hturner/useR2016

Progress graph by @HeathrTurnr @RWomenTaskforce great progress! #useR2016 #equality



San Francisco Sisters

Why was @RladiesSF was set up? They now have 700 #rladies



R-Ladies London presentation

R-Ladies Coding Club @RLadiesLondon · Jun 27

At @useR_Stanford? Check out @RLadiesLondon introducing ourselves to the #rstats community!! schedule.user2016.org/event/7BVF/r-l...

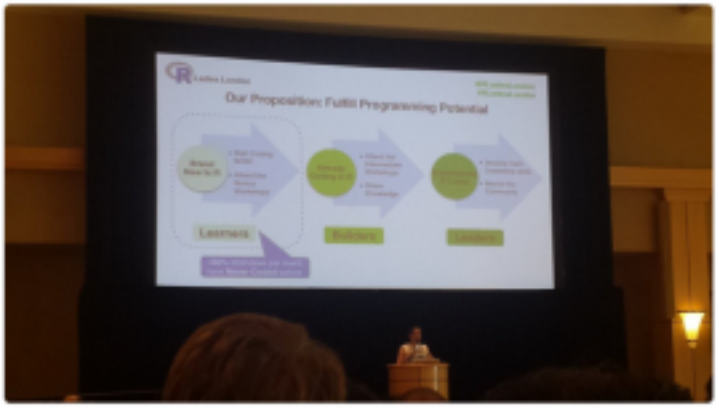
You and Hannah Frick

Matt Ritchie @mritchieau · Jun 28

@RLadiesLondon, 4 months old has 400+ members, 40+ participants per event. Tour de #Rstats a popular beginner workshop #useR2016

Scott Chamberlain @sckottie · Jun 28

#useR2016 fulfill programming potential by @RLadiesLondon



SanghaChick @SanghaChick · Jun 28

@RLadiesLondon got me all inspired.#rstats #useR2016

Great Response, and What can we do?

Lou Bajuk-Yorgan @LouBajuk · Jun 28

Kudos to all the great work being done by @RLadiesLondon #useR2016

Karl Broman @kwbroman · Jun 28

"what can we do?"

- "actively encourage women to program"
- "better job descriptions – no 'rock star'"
- "*Do* talk technical" #UseR2016

13 34

Mine CetinkayaRundel @minebocek · Jun 27

Inspired by the RLadies talks at #useR2016, anyone who wants to join forces to get RLadies RTP off the ground? #rstats

15 23

Next Steps

- R-Ladies London
- R-Ladies SF
- Women in R Taskforce
- Hadley Wickham

- Discuss : R-Ladies the success from SF and LDN

Next Steps

- R-Ladies London
- R-Ladies SF
- Women in R Taskforce

Project title : “R-Ladies Alignment and Global Expansion”

Funding request : $10,000

R-Ladies Global

Alignment

- Mission Statement, Organisational Structure, Code of Conduct, Logo, Github, Slack
- R-Ladies global website

Expansion

- “R-Ladies in a box”
- US: New York, Boston, Houston, Research Triangle Park
- UK: Manchester, Bristol, Scotland
- Australia: Melbourne, Sydney

R-Ladies Global

Alignment

- Mission Statement, Organisational Structure, Code of Conduct, Logo, Github, Slack
- R-Ladies global website

FINGERS CROSSED!

- "R-Ladies in a box"
- US: New York, Boston, Houston, Research Triangle Park
- UK: Manchester, Bristol, Scotland
- Australia: Melbourne, Sydney

Women in Machine Learning and Data Science

WiMLDS is a community of women interested in machine learning and data science. We host events which include talks by prominent data scientists, lightning talks, technical workshops and networking events. Our members include engineers, technologists, statisticians, students and many other professionals who work in data science or would like to transition into this field. Our mission is to build a community of women in data science by means of connecting and educating our members with support from the Bay Area, NYC, Chicago, North Carolina and Boulder tech community. If you're interested in starting a new chapter please reach out to us.

http://wimlds.org or erin@wimlds.org

rladiescodingclub@gmail.com

@rladieslondon