R Forwards

Julie Josse
R Forwards taskforce on women and other underrepresented groups
Presentation

Aim: improve the participation and experience of underrepresented groups in the R community.

- R foundation task force set up in December 2015 to address the underrepresentation of women (Rwomen)
- Rebranded in January 2017 to accommodate more underrepresented groups: LGBT, minority ethnic groups, people with disabilities

Website: https://forwards.github.io/
- Tweets: @R_Forwards, Facebook
Subteams


Forwards teleconf/ slackathons (virtual asynchronous meetings on slack) on alternate months - Github Repo

- **Community**: General outreach to help people from under-represented groups get into R. Gina Griffin, Kevin O’Brien

- **On-ramps**: Creating paths for users to develop their skills and make contributions to the R/BioConductor ecosystem. Jenny Bryan, Michael Lawrence

- **Social Media**: Posting to Twitter and/or Facebook, soliciting blog posts and publishing them, maintaining website. Tracy Shen, David Smith

- **Teaching**: Materials and workshops for under-represented groups. Isabella Gollini, Di Cook.

- **Conferences**: Liaising with OC/PC on policies and inclusion initiatives. **Code of conducts**, Childcare, Meet the diversity scholars, Conference buddies, R newbies session Heather Turner, Julie Josse
Survey subteam

Run and analyse community surveys, collect data (data page). Get input from the community on obstacles and ideas for improvements. Madlene Hamilton, Carolin Strobl, Jasmine Dumas

- Data monitoring: % women google summer of code mentors, students; R editors journals; ISC proposals, packages 2016: 11.4% (genderize package/manual)

Goal: numbers at least be comparable to computer science figures (>20%) mathematical/natural sciences (30-40%)
useR!2016 survey

First survey during useR!2016 participants (June 27-30). N = 455/899

**Demography**: gender, age, ethnic groups, country, education, employment, full-time/ part-time job, caregiver for children or adult dependents?

**R programming**:

Q11 How long have you been using R for?

Q12 Did you have previous programming experience before beginning to use R?

Tick any that apply among: Q13A I use functions from existing R packages to analyze data/
Q13B I write R code designed to make my work easier, such as loops or conditionals or functions/ Q13C I write R functions for use by myself or my collaborators/ Q13D I contribute to R packages (on CRAN or elsewhere)/ Q13E I have written my own R package/ Q13F I have written my own R package and released it on CRAN or Bioconductor (or shared it on GitHub, R-Forge or similar platforms)/

Q14 Do you use R as a recreational activity, primarily as part of a job or both?
useR!2016 survey

Q15 How much do you agree or disagree with the following statements? Writing R is fun/considered cool by my peers?/ a monotonous task/ is difficult

Q16 Would you recommend R as a programming language to learn?

Q17 What would be your number one argument for/against learning R?

R community

Q18 Do you consider yourself part of the R community?

Q19 Which of the following resources do you use for support? (The R mailing lists Twitter, R StackOverflow, IRC channel, etc..

Q21 Do you attend R user group meetings in your local area? ...

- **Blog posts**: mapping useRs, users-relationship-with-r, useRs participation in the community
- **Reports**: Non-response in useR! 2016 Survey, useR!2016 participants and R programming: a multivariate analysis
- **R journal paper** in preparation Stella Bollmann, Dianne Cook, Rudolf Debelak, Jasmine Dumas, John Fox, Julie Josse, Oliver Keyes, Carolin Strobl, Heather Turner.
Results
General findings

- Demographics: experienced R users (83% > 2 years)

Under representation of non whites and LGTB

- UseRs relationship with R:
R programming

Multiple Correspondence Analysis: dimensionality reduction. Similarities between individuals, relationship between variables

- Two categories are close when individuals who have selected the first category also tend to take the other category
R programming

3 clusters: The experienced users (38%): more than 10 years’ experience, in depth knowledge of R programming, write their own package. Mainly men, academic, with doctorate. Use R both for their job and for pleasure.

The intermediate users (57%): for less than 2 years, use functions from existing packages, do not write their own functions. Mainly women, people from industry, undergraduate or master degree, more from US. Use R for their jobs

The curious users (3%): they have discovered R very recently. They don’t have an opinion that is already formed. (9 people with less than 6 months experience).
R programming

Men are more highly represented than women among the most advanced users.

Age effect?
R community

Q18 Do you consider yourself part of the R community?

Q19 Which of the following resources do you use for support? (The R mailing lists, The #rstats hashtag on Twitter, StackOverflow queues, The R IRC channel, The rOpenSci mailing lists or chat forums, The Bioconductor support site)

Q20 What would be your preferred medium for R community news?

Q21 Do you attend R user group meetings in your local area?

Q24 Which of the following would make you more likely to participate in the R community, or improve your experience? Tick any that apply: A New R user group near me, New R user group near me aimed at my demographic, Free local introductory R workshops, Paid local advanced R workshops, R workshop at conference in my domain, R workshop aimed at my demographic, Online forum to discuss R-related issues, etc..

Q25 What other ideas do you have for improving the R community?
R community

2 main clusters: **Do not feel part of the R community.** (35%) Do not attend RUG because feel inexperimented. **Would like to attend workshops/RUG/online support if close to them (geographic/ demographic).** Use blogs to get help. Prefered medium for R community news: Facebook or mailing lists. **Do not use Twitter,** do not want to use it. More women, master/undergraduate degree, young (men and women).

**Feel part of the R community.** (61%) Do not have time to participate to meetings. Attend general RUG meetings. Not interested in free workshops, online demos. Use Twitter. Twitter preferred medium for R community news. More men, in academia with a doctorate.
Conclusion
Survey

- Missing values/ Coding issues “yes” “no” answers for checkboxes
- Cautious about over-interpretation: women are younger
- Logistic regression: gender differences in R packages writing caused by lengths of R usage, employment in academia and a feeling of belonging to the R community. Ask women the fear in developing packages
- Other results on our website (a small group who do not like R)
- useR! attendees are not representative from the R community as a whole (depends on the location - Stanford)
- People willing to participate more
- Looking forwards to the new data
Activities of Rforwards

- Development workshops for women engaged in research. Melbourne, Auckland, US, Europe. Workshops for teenage girls to encourage an interest in programming.

- Send emails to invite friends submitting abstracts, attending conferences

- Ideas for improving R community, promote welcoming culture: a webpage introducing the community and how it is organized, publish guide for running local group, team participation in challenges, mentorship, joint event with other data sciences groups, online conferences, ...

Lot of work! Thank you for the efforts to improve diversity and inclusion