Package 'forwards'

October 13, 2022

Type Package

Title Data from Surveys Conducted by Forwards

Version 0.1.3

Description Anonymized data from surveys conducted by Forwards <https://forwards.github.io/>, the R Foundation task force on women and other under-represented groups. Currently, a single data set of responses to a survey of attendees at useR! 2016 <https://www.r-project.org/useR-2016/>, the R user conference held at Stanford University, Stanford, California, USA, June 27 - June 30 2016.

URL https://github.com/forwards/forwards

BugReports https://github.com/forwards/forwards/issues

License CC0

Encoding UTF-8

LazyData TRUE

Depends R (>= 2.10)

RoxygenNote 6.1.1

Suggests dplyr, FactoMineR, forcats, ggplot2, knitr, likert, rmarkdown, tidyr

VignetteBuilder knitr

NeedsCompilation no

Author Heather Turner [aut, cre],
Oliver Keyes [aut]

Maintainer Heather Turner <ht@heatherturner.net>

Repository CRAN

Date/Publication 2019-07-30 21:10:02 UTC

R topics documented:

---

forwards	*Data Released by Forwards*

Description

forwards provides data sets released by Forwards, the R Foundation task force on women and other under-represented groups.

---

useR2016	*Data From useR! 2016 Survey*

Description

This data set contains results from a survey conducted by Forwards of attendees at useR! 2016, the R user conference held at Stanford University, Stanford, California, June 27 - June 30 2016. Modifications made to anonymize the data are noted in Details.

Usage

```
useR2016
```

Format

A data frame with 449 records and 48 variables:

Q2 A factor with 3 levels: "Men", "Non-Binary/Unknown", "Women".

Q3 A factor with 2 levels: "> 35", "35 or under"

Q7 A factor with 2 levels: "Doctorate/Professional", "Masters or lower"

Q8 A factor with 2 levels: "Non-academic", "Academic"

Q11 A factor with 4 levels: "< 2 years", "2-5 years", "5-10 years", "> 10 years"

Q12 A factor with 2 levels: "Yes", "No"

Q13 A character vector with values "I use functions from existing R packages to analyze data" or NA

Q13_B A character vector with values "I write R code designed to make my work easier, such as loops or conditionals or functions" or NA

Q13_C A character vector with values "I write R functions for use by myself or my collaborators" or NA

Q13_D A character vector with values "I contribute to R packages (on CRAN or elsewhere)" or NA

Q13_E A character vector with values "I have written my own R package" or NA

Q13_F A character vector with values "I have written my own R package and released it on CRAN or Bioconductor (or shared it on GitHub, R-Forge or similar platforms)" or NA

Q14 A factor with 3 levels: "Primarily as part of a job or educational course;", "Primarily as a recreational activity, in your free time;", "For both recreational and job/educational purposes."

Q15 A factor with 5 levels: "Strongly disagree", "Disagree", "No opinion", "Agree", "Strongly agree"

Q15_B A factor with 5 levels: "Strongly disagree", "Disagree", "No opinion", "Agree", "Strongly agree"

Q15_C A factor with 5 levels: "Strongly disagree", "Disagree", "No opinion", "Agree", "Strongly agree"

Q15_D A factor with 5 levels: "Strongly disagree", "Disagree", "No opinion", "Agree", "Strongly agree"

Q16 A factor with 2 levels: "Yes", "No"

Q17 A factor with 21 levels: "Good for statistical analysis", "Good for working with biological data structures", ...

Q17_B A character vector of free text response for when Q17 == "Other (please specify)"

Q18 A factor with 2 levels: "Yes", "No"

Q19 A character vector with values "The R mailing lists" or NA

Q19_B A character vector with values "The #rstats hashtag on Twitter" or NA

Q19_C A character vector with values "The R StackOverflow queues" or NA

Q19_D A character vector with values "The R IRC channel" or NA

Q19_E A character vector with values "The rOpenSci mailing lists or chat forums" or NA

Q19_F A character vector with values "The Bioconductor support site" or NA

Q19_G A character vector with values "Other (please specify)" or NA

Q19_H A character vector of free text response for when Q19_G == "Other (please specify)"

Q20 A factor with 9 levels: "Twitter", "Facebook", "Google+", ...

Q20_B A character vector of free text response for when Q20 == "Other (please specify)"

Q21 A factor with 2 levels: "Yes", "No"

Q22 A factor with 5 levels: "A general user group", "A user group for women in R", "A user group within a university", "A user group within a company", "Other (please specify)"

Q22_B A character vector of free text response for when Q22 == "Other (please specify)"

Q23 A factor with 6 levels: "There is no group nearby/the group is inactive", "I am too busy", ...

Q24 A character vector with values "New R user group near me (specify location in comments box)" or NA

Q24_B A character vector with values "New R user group near me aimed at my demographic (specify relevant group in comments box)" or NA

Q24_C A character vector with values "Free local introductory R workshops" or NA

Q24_D A character vector with values "Paid local advanced R workshops" or NA

Q24_E A character vector with values "R workshop at conference in my domain (specify domain/conference in comments box)" or NA

Q24_F A character vector with values "R workshop aimed at my demographic (specify relevant group in comments box)" or NA

Q24_G A character vector with values "Mentoring (e.g. first CRAN submission/useR! abstract submission/GitHub contribution)" or NA

Q24_H A character vector with values "Training in non-English language (specify language in comments box)" or NA

Q24_I A character vector with values "Training that accommodates my disability (specify disability in comments box)" or NA

Q24_J A character vector with values "Online forum to discuss R-related issues" or NA

Q24_K A character vector with values "Online support group for my demographic (specify relevant group in comments box)" or NA

Q24_L A character vector with values "Special facilities at R conferences (give further detail in comments box)"

Details

This data set contains responses to the following questions from the survey of useR! 2016 attendees:

Q2 What is your gender?

Q3 In what year were you born?

Q7 What is the highest level of education you have completed?

Q8 What is your current (primary) employment status?

Q11 How long have you been using R for?

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

Q13 Which of the following do you do? Tick any that apply. (Responses stored in Q13 to Q13_F.)

- I use functions from existing R packages to analyze data
- I write R code designed to make my work easier, such as loops or conditionals or functions
- I write R functions for use by myself or my collaborators
- I contribute to R packages (on CRAN or elsewhere)
- I have written my own R package
- 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:

- Primarily as part of a job or educational course;
- Primarily as a recreational activity, in your free time;
- For both recreational and job/educational purposes.

Q15 How much do you agree or disagree with the following statements? (Responses stored in Q15 to Q15_D.)

- Writing R is fun
- Writing R is considered cool or interesting by my peers
- Writing R is a monotonous task
- Writing R is difficult

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

Q17 What would be your number one argument for/against learning R? (fixed responses in `Q17`, other specified responses in `Q17_B`)

Q18 Do you consider yourself part of the R community?

Q19 Which of the following resources do you use for support? Select all that apply. (Fixed responses stored in `Q19` to `Q19_G`, other specified responses in `Q19_H`.)

- The R mailing lists
- The #rstats hashtag on Twitter
- The R StackOverflow queues
- The R IRC channel
- The rOpenSci mailing lists or chat forums
- The Bioconductor support site
- Other (please specify)

Q20 What would be your preferred medium for R community news (e.g. events, webinars, opportunities)? (Fixed responses in `Q20`, other specified responses in `Q20_B`.)

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

Q22 If you do: what type of user group is it? (Fixed responses in `Q22`, other specified responses in `Q22_B`.)

Q23 If you do not: why not?

Q24 Which of the following would make you more likely to participate in the R community, or improve your experience? Tick any that apply. (Fixed responses stored in `Q24` to `Q24_L`.)

Various measures were taken to protect anonymity of the respondents and avoid disclosure of sensitive information. In particular the following questions/variables are completely excluded:

Q1 What did you register as at useR! 2016?

Q4 To what racial or ethnic group(s) do you identify?

Q5 In what country do you currently reside?

Q6 Do you identify as LGBT (Lesbian, Gay, Bisexual, Asexual and/or Transgender)?

Q9 Is your current job:

- Full-time
- Part-time
- I am not currently employed

Q10 Are you a caregiver for children or adult dependents on a regular basis?

Q23_B Specific reason for not attending a user group

Q24_M Specific location/demographic/domain/language etc for which the respondent would like a user group/workshop/other support

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

Q26 Do you have any feedback for the survey authors?

Summaries of all these variables have been presented in blog posts (see references). Q1, Q9 and Q10 were used in multivariate analyses (see references) but Q9 and Q10 did not feature in the interpretation and Q1 has inconsistencies with Q8. For the latter we give priority to Q8, the employment status of respondents at the time they completed the survey.

Of the remaining variables, we consider Q2, Q3, Q7, Q8, Q11, and Q13_F to be implicit identifiers (key variables). These variables were modified to achieve 3-anonymity, i.e. the smallest subgroup identifiable from combinations of these variables is at least of size 3. In particular, the following modifications were made

Q2 Non-binary grouped with missing; all other key variables for this group suppressed (set to NA).

Q3 Year of birth converted to approximate age groups: "> 35" and "35 and under"; age group suppressed for 14 individuals.

Q7 Highest education level aggregated to two groups: "Doctorate/Professional" and "Masters and under"; highest education level suppressed for 3 individuals.

Q8 Employment status aggregated to three groups: "Non-academic" (includes employment in industry, government, non-profit, self-employed) and "Academic" (includes retired, unemployed, student).

Q11 Length of R usage aggregated to four groups: combined groups corresponding to shortest times into "< 2 years" group.

Q13_F Suppressed for two individuals.

In addition specific values containing personal/personally identifiable information were suppressed in Q19_H, Q22_B and Q23_B.

Author(s)

Heather Turner and Oliver Keyes

References

Bollmann, S., Cook, D., Debelak, R., Dumas, J., Fox, J., Josse, J., Keyes, O., Strobl, C. and Turner, H. (2017) Mapping useRs https://forwards.github.io/blog/2017/01/13/mapping-users/.

Bollmann, S., Cook, D., Debelak, R., Dumas, J., Fox, J., Josse, J., Keyes, O., Strobl, C. and Turner, H. (2017) useRs Relationship with R https://forwards.github.io/blog/2017/03/11/users-relationship-with-r/.

Josse, J. and Turner, H. (2017) useR! 2016 participants and R programming: a multivariate analysis https://forwards.github.io/docs/mca_programming_user2016_survey/.

Josse, J. and Turner, H. (2017) useR! 2016 participants and the R community: a multivariate analysis https://forwards.github.io/docs/mca_community_user2016_survey/.

Examples

```
# cross-tabulate age and length of time using R
xtabs(~ Q3 + Q11, data = useR2016)

# fit a logistic regression with "contribute to or write packages" predicted by
# gender, length of R usage, employment status, and community belonging
response <- with(useR2016,
    ifelse(!is.na(Q13_D) | !is.na(Q13_E) | !is.na(Q13_F), 1, 0))
glm(response ~ Q2 + Q11 + Q8 + Q18, data = useR2016)
```

Index