Future Contributors

Overview

Plotting results of Q15: "What would make you more likely to participate in OSS projects?". This question was only visible to respondents who said they haven't yet contributed to OS, but would like to do so in the future.

Import packages and utilities

```
project_root <- here::here() # requires that you be somewhere in the
# project directory (not above it)
# packages
suppressMessages(source(file.path(project_root, "scripts/packages.R")))
# functions and objects used across scripts
suppressMessages(source(file.path(project_root, "scripts/utils.R")))
```

Load data

```
future <- load_qualtrics_data("clean_data/future_contributors_Q15.tsv")
other_quant <- load_qualtrics_data("clean_data/other_quant.tsv")
status <- load_qualtrics_data("clean_data/contributor_status_Q3.tsv")
head(future)
```

```
  Conferences and hackathons Computing environments Educational materials
1                          0                      0                     0
2                          0                      0                     0
3                          0                      0                     0
```

```
4                          0                     0                     0
5                          0                     0                     0
6                          0                     1                     1
  A learning community Sustainability grants Industry networking
1                    0                     0                    0
2                    0                     0                    0
3                    0                     0                    0
4                    0                     0                    0
5                    0                     0                    0
6                    1                     0                    0
  Academic job opportunities Other Help finding funding Legal support
1                          0     0                    0             0
2                          0     0                    0             0
3                          0     0                    0             0
4                          0     0                    0             0
5                          0     0                    0             0
6                          0     0                    0             0
  Mentoring programs
1                  0
2                  0
3                  0
4                  0
5                  0
6                  0
```

At this point, we COULD remove rows from participants who never saw this question, but since we're just tallying up the 1s, not the 0s, there's really no need.

Prepare data for plotting

Sum up counts for each solution.

```
to_plot <- data.frame(
  Solution = names(future),
  Count = unname(apply(future, 2, function(x) round(sum(x, na.rm = TRUE))))
)
to_plot
```

```
                    Solution Count
1  Conferences and hackathons    25
```

```
2      Computing environments    40
3       Educational materials    35
4        A learning community    37
5       Sustainability grants    22
6         Industry networking    29
7  Academic job opportunities    22
8                       Other     8
9        Help finding funding    21
10              Legal support    22
11         Mentoring programs    29
```

Plot

For visual clarity, let's remove the "Other" row.

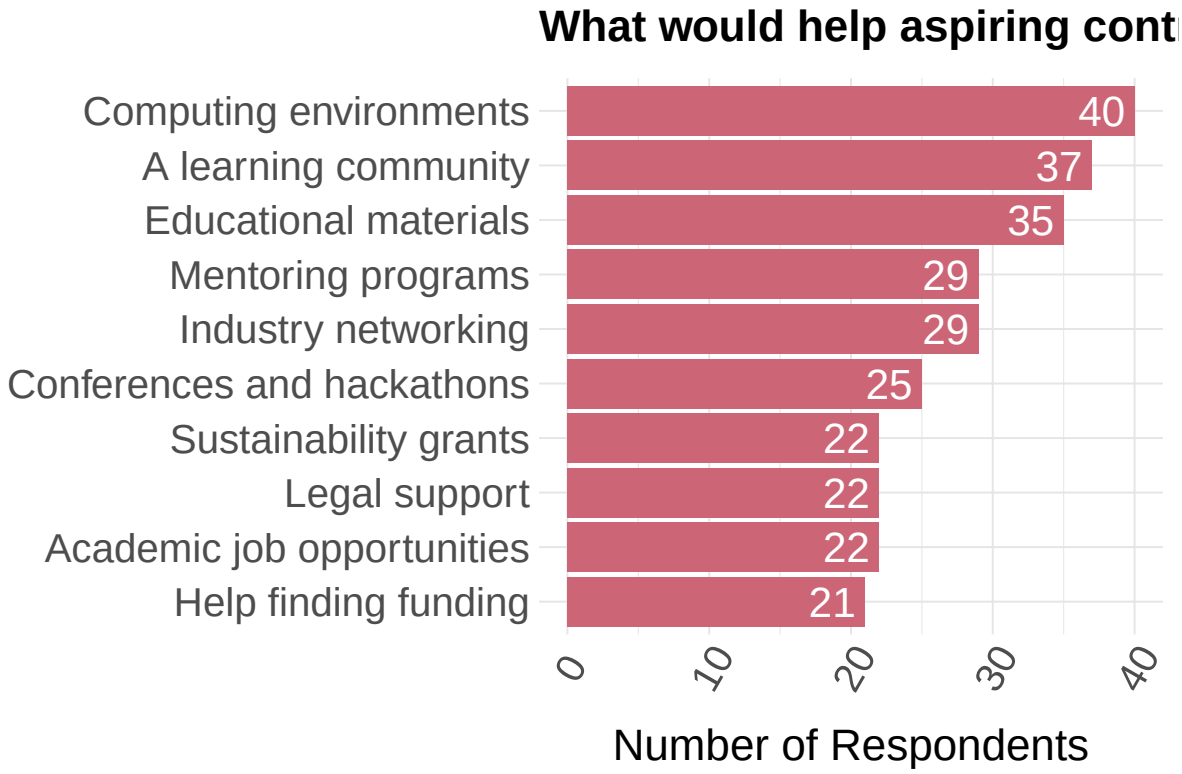
```
to_plot <- to_plot %>% filter(Solution != "Other")
```

Reorder factor levels based on count.

```
to_plot <- to_plot %>%
  mutate(Solution = fct_reorder(Solution, Count, .desc = FALSE))
```

And make a plot, using a function in utils.R.

```
myplot <- basic_bar_chart(to_plot,
  x_var = "Solution",
  y_var = "Count",
  title = "What would help aspiring contributors get started?",
  horizontal = TRUE,
  show_bar_labels = TRUE,
  show_ticks_y = FALSE,
  color_index = 7,
  show_axis_title_x = TRUE,
  show_axis_title_y = FALSE,
  show_grid = TRUE
)
myplot
```



Save the plot if you wish.

```
save_plot("future_contributors.tiff", 13, 6, p=myplot)
```

Look at jobs of respondents

NOW let's remove rows where the participant never saw the question. Also, remove rows where we have no data on the participant's job.

```
future_and_job <- future
future_and_job$job_category <- other_quant$job_category

n <- ncol(future)
future_cols <- names(future_and_job)[1:n]
last_col   <- names(future_and_job)[ncol(future_and_job)]

future_and_job <- future_and_job %>%
  filter(
```

```
    # drop rows where all of the future columns are 0
    !if_all(all_of(future_cols), ~ . == 0),
    # drop rows where the job column is ""
    .data[[last_col]] != ""
  )
```

Combine postdocs and other research staff for better visual clarity.

```
future_and_job <- future_and_job %>%
  mutate(
    job_category = recode(
      job_category,
      "Post-Doc" = "Postdocs and\nStaff Researchers",
      "Other research staff" = "Postdocs and\nStaff Researchers"
    )
  )
```

At this point I'm not going to bother with additional wrangling because I'm just interested in the jobs of the people who answered this question. But we are poised to see which jobs voted for which solutions, if needed.

```
jobs <- data.frame(table(future_and_job$job_category))
names(jobs) <- c("Job", "Count")
```

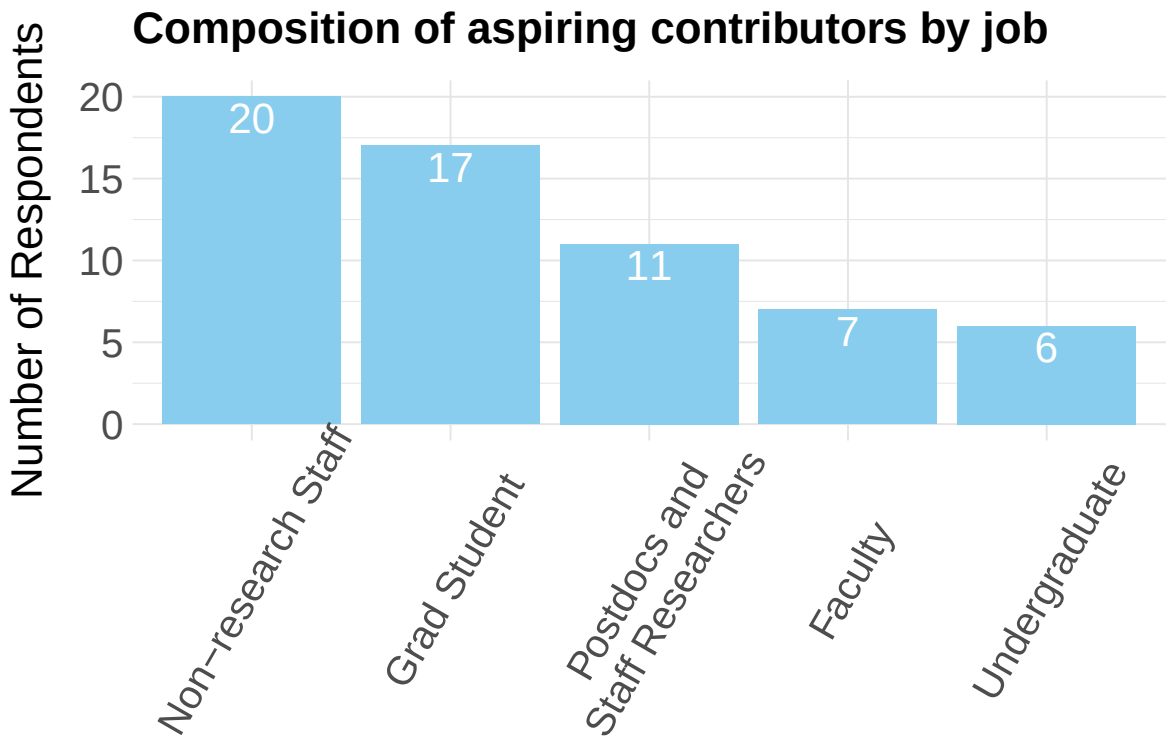
Reorder factor levels based on count.

```
jobs <- jobs %>%
  mutate(Job = fct_reorder(Job, Count, .desc = TRUE))
```

And make a plot, using a function in utils.R.

```
jobs_plot <- basic_bar_chart(jobs,
  x_var = "Job",
  y_var = "Count",
  title = "Composition of aspiring contributors by job",
  horizontal = FALSE,
  show_bar_labels = TRUE,
  show_ticks_y = FALSE,
  color_index = 2,
  show_axis_title_y = TRUE,
```

```
  show_grid = TRUE
)
jobs_plot
```



Save the plot if you wish.

```
save_plot("future_contributors_jobs.tiff", 12, 8, p=jobs_plot)
```

```
sessionInfo()
```

```
R version 4.4.2 (2024-10-31)
Platform: aarch64-apple-darwin20
Running under: macOS Sequoia 15.6.1

Matrix products: default
BLAS:   /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRlapack.dylib;

locale:
```

```
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

time zone: America/Los_Angeles
tzcode source: internal

attached base packages:
[1] tools     grid      stats     graphics  grDevices datasets  utils
[8] methods   base

other attached packages:
 [1] treemapify_2.5.6      tidyr_1.3.1           svglite_2.2.1
 [4] stringr_1.5.1         scales_1.4.0          readr_2.1.5
 [7] pwr_1.3-0             patchwork_1.3.2       ordinal_2023.12-4.1
[10] lme4_1.1-37           Matrix_1.7-1          languageserver_0.3.16
[13] here_1.0.1            gtools_3.9.5          ggforce_0.5.0
[16] FSA_0.10.0            fpc_2.2-13            forcats_1.0.0
[19] factoextra_1.0.7      ggplot2_3.5.2         emmeans_1.11.2
[22] dplyr_1.1.4           corrplot_0.95         ComplexHeatmap_2.22.0
[25] cluster_2.1.8.1       BiocManager_1.30.26

loaded via a namespace (and not attached):
 [1] Rdpack_2.6.4        rlang_1.1.6         magrittr_2.0.3
 [4] clue_0.3-66         GetoptLong_1.0.5    matrixStats_1.5.0
 [7] compiler_4.4.2      flexmix_2.3-20      systemfonts_1.2.3
[10] png_0.1-8           callr_3.7.6         vctrs_0.6.5
[13] pkgconfig_2.0.3     shape_1.4.6.1       crayon_1.5.3
[16] fastmap_1.2.0       labeling_0.4.3      rmarkdown_2.29
[19] ggfittext_0.10.2    tzdb_0.5.0          ps_1.9.1
[22] nloptr_2.2.1        purrr_1.1.0         xfun_0.53
[25] modeltools_0.2-24   jsonlite_2.0.0      tweenr_2.0.3
[28] parallel_4.4.2      prabclus_2.3-4      R6_2.6.1
[31] stringi_1.8.7       RColorBrewer_1.1-3  boot_1.3-31
[34] diptest_0.77-2      numDeriv_2016.8-1.1 estimability_1.5.1
[37] Rcpp_1.1.0          iterators_1.0.14    knitr_1.50
[40] IRanges_2.40.1      splines_4.4.2       nnet_7.3-19
[43] tidyselect_1.2.1    yaml_2.3.10         doParallel_1.0.17
[46] codetools_0.2-20    processx_3.8.6      lattice_0.22-6
[49] tibble_3.3.0        withr_3.0.2         evaluate_1.0.4
[52] polyclip_1.10-7     xml2_1.4.0          circlize_0.4.16
[55] mclust_6.1.1        kernlab_0.9-33      pillar_1.11.0
[58] renv_1.1.5          foreach_1.5.2       stats4_4.4.2
[61] reformulas_0.4.1    generics_0.1.4      rprojroot_2.1.1
[64] S4Vectors_0.44.0    hms_1.1.3           minqa_1.2.8
```

```
[67] xtable_1.8-4         class_7.3-22         glue_1.8.0
[70] robustbase_0.99-4-1 mvtnorm_1.3-3        rbibutils_2.3
[73] colorspace_2.1-1     nlme_3.1-166         cli_3.6.5
[76] textshaping_1.0.1    gtable_0.3.6         DEoptimR_1.1-4
[79] digest_0.6.37        BiocGenerics_0.52.0 ucminf_1.2.2
[82] ggrepel_0.9.6        rjson_0.2.23         farver_2.1.2
[85] htmltools_0.5.8.1    lifecycle_1.0.4      GlobalOptions_0.1.2
[88] MASS_7.3-61
```