

# 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

```
data <- load_qualtrics_data("deidentified_no_qual.tsv")
```

## Wrangle data

```
future <- data %>%
  select(
    starts_with("future_contributors")
  )
```

This is one of those questions where the column value is either an empty string, indicating that the person did not select that option, or it is the option itself, e.g. “Accessible conferences or hackathons”, indicating that they did select it. The numbering system is weird—each column starts with the prefix “future\_contributors” (since that’s what I named the question in Qualtrics), but the column numbers don’t make much sense. They are: future\_contributors\_1 future\_contributors\_4 future\_contributors\_7 future\_contributors\_8 future\_contributors\_9 future\_contributors\_10 future\_contributors\_11 future\_contributors\_12 future\_contributors\_22 future\_contributors\_23 future\_contributors\_24. There were only ten options (+ “Other”), so I don’t know why the numbers go up to 24. Whatever. It doesn’t really matter, since we will rename the columns based on the entries.

## Define codes

```
codenames <- c(
  "Accessible conferences" = "Conferences/hackathons",
  "Access to free" = "Computing environments",
  "Educational materials" = "Educational materials",
  "An open source discussion group" = "Learning community",
  "Dedicated grants" = "Dedicated grants",
  "Networking opportunities" = "Industry networking",
  "Job" = "Academic job opportunities",
  "Other " = "Other",
  "Assistance identifying" = "Help finding funding",
  "Legal" = "Legal support",
  "A mentor" = "Mentoring programs"
)
```

Wrangle the data using several functions from my utilities script `scripts/utils.R`.

```
future <- shorten_long_responses(future, codenames)

# Remove any columns that are all NA or empty strings
# (Which means nobody selected that response)
future <- exclude_empty_columns(future)
# Remove any rows that are all NA or empty strings
```

```
# (The participant did not answer the question
# (because they're not eligible))
future <- exclude_empty_rows(future)

future <- rename_cols_based_on_entries(future)
future <- make_df_binary(future)
head(future)
```

	Conferences/hackathons	Computing environments	Educational materials
6	0	1	1
8	0	0	0
15	0	0	0
25	0	1	0
27	0	1	1
29	0	1	0

	Learning community	Dedicated grants	Industry networking
6	1	0	0
8	1	0	1
15	0	0	0
25	1	1	0
27	1	0	0
29	1	0	0

	Academic job opportunities	Other	Help finding funding	Legal support
6	0	0	0	0
8	0	0	0	0
15	0	1	0	0
25	0	0	0	0
27	0	1	0	0
29	1	0	0	0

	Mentoring programs
6	0
8	1
15	0
25	0
27	0
29	1

Now we 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))))
)
head(to_plot)
```

	Solution	Count
1	Conferences/hackathons	25
2	Computing environments	40
3	Educational materials	35
4	Learning community	37
5	Dedicated grants	22
6	Industry networking	29

## Plot

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

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

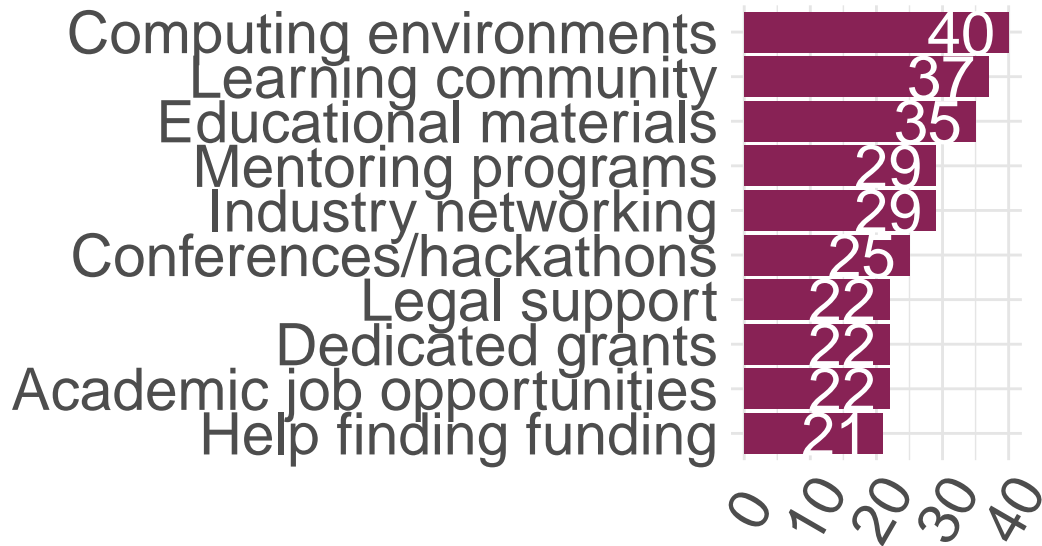
Reorder factor levels based on count.

```
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 = 8,
  show_axis_title_y = FALSE,
  show_grid = TRUE
)
myplot
```

## What would help aspiring contri



Save the plot if you wish.

```
save_plot("future_contributors.tiff", 12, 8, p=myplot)
```

## Look at jobs of respondents

```
future_and_job <- data %>%
  select(
    starts_with("future_contributors") |
    starts_with("job_category")
  )

future_and_job <- shorten_long_responses(
  future_and_job,
  c("Other research staff" = "Other research staff")
)

# Remove rows where they didn't answer the question
future_and_job <- future_and_job %>%
  filter(if_any(future_contributors_1:future_contributors_24, ~ .x != ""))
```

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")
```

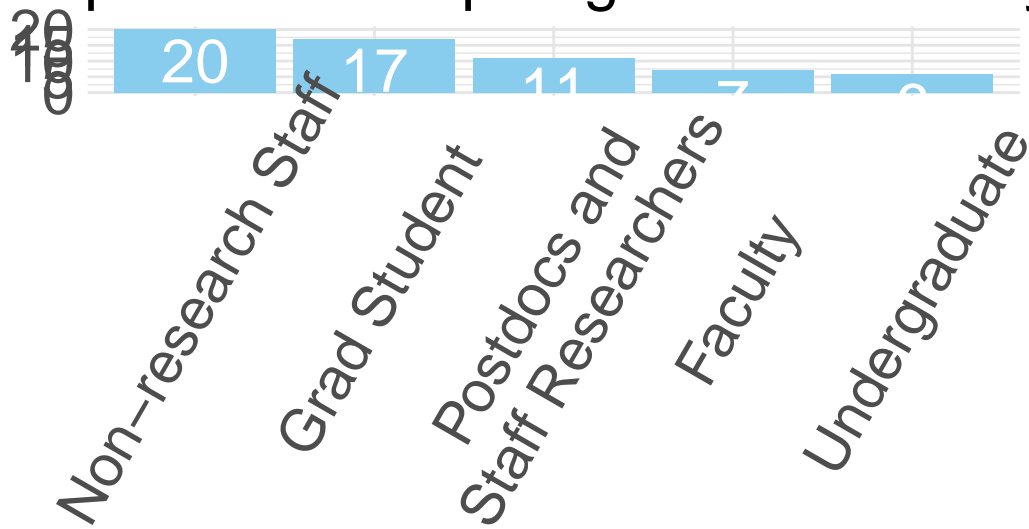
Reorder factor levels based on count.

```
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 = FALSE,
  show_grid = TRUE
)
jobs_plot
```

Composition of aspiring contributors by



Save the plot if you wish.

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