

# Motivations for contributing to OS: plots

## Overview

This script makes some plots from Q6, which is about participants' reasons for contributing to open source.

## 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")))
```

## Define functions

```
line_plot <- function(df, x_var, y_var, title) {
  p <- ggplot(df, aes(x = .data[[x_var]], y = .data[[y_var]])) +
    geom_point(size = 4) + # Adjust dot size
    scale_y_continuous(labels=scales::percent) +
    labs(
      title = title
    ) +
    theme(
      axis.title.x = element_blank(),
      axis.title.y = element_blank(),
```

```

axis.text.x = element_text(angle = 60, vjust = 0.6, size = 24),
axis.text.y = element_text(size = 24),
axis.ticks.x = element_blank(),
axis.ticks.y = element_blank(),
legend.title = element_blank(),
plot.title = element_text(
  hjust = 0,
  face = "bold",
  size = 24,
  margin = margin(b = 15)
),
plot.margin = unit(c(1, 1, 1, 3), "cm"),
panel.grid = element_line(linetype = "solid", color = "gray90"),
panel.background = element_blank()
)
return(p)
}

```

## Load data

```

motivations <- load_qualtrics_data("clean_data/motivations_Q6.tsv")
other_quant <- load_qualtrics_data("clean_data/other_quant.tsv")

```

Wrangle into some handy data frames we'll use later. I'm pretty sure that all these cleaning steps are unnecessary, but they make me feel better.

```

motivations_job_staff <- cbind(motivations, other_quant$job_category)
# Rename columns
names(motivations_job_staff)[length(names(motivations_job_staff))] <- "job_category"
motivations_job_staff <- cbind(motivations_job_staff, other_quant$staff_categories)
names(motivations_job_staff)[length(names(motivations_job_staff))] <- "staff_category"
head(motivations_job_staff)

```

	Job	Improve	Tools	Customize	Network	Give back	Skills	Fun	Other
1	1		1	1	1	1	1	1	0
2	0		1	1	1	0	1	0	0
3	0		1	1	0	0	1	1	0
4	1		1	1	0	1	0	0	0
5	0		1	1	0	1	1	1	0

```

6   0          0          0          0          0          0          0          0
                job_category staff_category
1             Faculty
2            Post-Doc
3 Other research staff
4             Faculty
5             Faculty
6 Other research staff

```

```

# Remove any rows where the job_category or staff_category are missing
motivations_job_staff_clean <- exclude_empty_rows(motivations_job_staff, strict=TRUE)
# Remove rows of all 0s
motivations_job_staff_clean <- motivations_job_staff_clean %>%
  filter(!if_all(Job:Other, ~ .x == 0))
head(motivations_job_staff_clean)

```

	Job	Improve	Tools	Customize	Network	Give back	Skills	Fun	Other
1	0		1	1	0	0	1	1	0
2	1		0	0	0	1	0	0	0
3	0		1	1	1	0	0	0	0
4	1		0	0	1	1	0	0	0
5	0		1	1	0	1	1	1	1
6	1		1	1	1	1	1	1	1
		job_category			staff_category				
1	Non-research Staff				Other				
2	Non-research Staff	DevOps or System Administration							
3	Non-research Staff	DevOps or System Administration							
4	Non-research Staff		Information Technology (IT)						
5	Non-research Staff	DevOps or System Administration							
6	Non-research Staff				Other				

```

# Do the same, but dropping staff categories (e.g. IT)
motivations_job <- subset(motivations_job_staff, select=-staff_category)
# Remove any rows where the job_category is missing
motivations_job_clean <- exclude_empty_rows(motivations_job, strict=TRUE)
# Remove rows of all 0s
motivations_job_clean <- motivations_job_clean %>%
  filter(!if_all(Job:Other, ~ .x == 0))

```

## Basic bar plot of contributor motivations

Sum up counts for each motivation.

```
motivations_to_plot <- data.frame(
  Motivation = names(motivations),
  Count = unname(apply(motivations, 2, function(x) round(sum(x, na.rm = TRUE)))))
)
head(motivations_to_plot)
```

	Motivation	Count
1	Job	109
2	Improve Tools	198
3	Customize	161
4	Network	78
5	Give back	156
6	Skills	142

Reorder factor levels based on count.

```
motivations_to_plot <- motivations_to_plot %>%
  mutate(Motivation = fct_reorder(Motivation, Count, .desc = FALSE))
```

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

```
basic_plot <- basic_bar_chart(motivations_to_plot,
  x_var = "Motivation",
  y_var = "Count",
  title = "Reasons for Contributing to\nOpen Source",
  horizontal = TRUE,
  show_bar_labels = TRUE,
  show_ticks_y = FALSE,
  color_index = 3,
  show_axis_title_x = TRUE,
  show_axis_title_y = FALSE,
  show_grid = TRUE,
  margin_vals = c(1, 1, 1, 1)
)
```

Save the plot if you wish.

```
save_plot("motivations_overall.tiff", 10, 6, p=basic_plot)
```

Save the data to a file for fine-tuning of plot.

```
write_df_to_file(  
  motivations_to_plot,  
  file.path("data_for_plots/motivations_basic_bar.tsv")  
)
```

## Stacked bar plots of motivations by role

Now let's make some stacked bar plots of motivations by role (job category). We'll make two: one with the absolute number of responses, and one where all roles are normalized to 1, so we can see the relative proportions of each motivation.

For visual clarity, let's combine post-docs and other research staff into one category.

```
counts_long <- motivations_job_clean %>%  
  pivot_longer(  
    cols      = -c(job_category),  
    names_to  = "Motivation",  
    values_to = "Value"  
  ) %>%  
  group_by(job_category, Motivation) %>%  
  summarise(  
    Count = sum(Value, na.rm = TRUE),  
    .groups = "drop"  
  )  
counts_long
```

```
# A tibble: 48 x 3  
  job_category Motivation   Count  
  <chr>        <chr>       <int>  
1 Faculty      Customize    45  
2 Faculty      Fun          32  
3 Faculty      Give back   38  
4 Faculty      Improve Tools 54  
5 Faculty      Job          23  
6 Faculty      Network      16  
7 Faculty      Other         13
```

```

8 Faculty      Skills          23
9 Grad Student Customize     19
10 Grad Student Fun         18
# i 38 more rows

unique(counts_long$job_category)

[1] "Faculty"           "Grad Student"        "Non-research Staff"
[4] "Other research staff" "Post-Doc"           "Undergraduate"

big_counts_df <- counts_long %>%
  # lump Post-Doc & Other research staff together
  mutate(
    job_category = case_when(
      job_category %in% c("Post-Doc", "Other research staff") ~
        "Post-Docs and\nStaff Researchers",
      TRUE ~ job_category
    )
  ) %>%
  # re-group and sum up counts
  group_by(job_category, Motivation) %>%
  summarise(
    Count = sum(Count),
    .groups = "drop"
  )
big_counts_df

# A tibble: 40 x 3
  job_category Motivation   Count
  <chr>        <chr>       <int>
  1 Faculty     Customize     45
  2 Faculty     Fun           32
  3 Faculty     Give back     38
  4 Faculty     Improve Tools 54
  5 Faculty     Job           23
  6 Faculty     Network        16
  7 Faculty     Other          13
  8 Faculty     Skills          23
  9 Grad Student Customize     19
 10 Grad Student Fun         18
# i 30 more rows

```

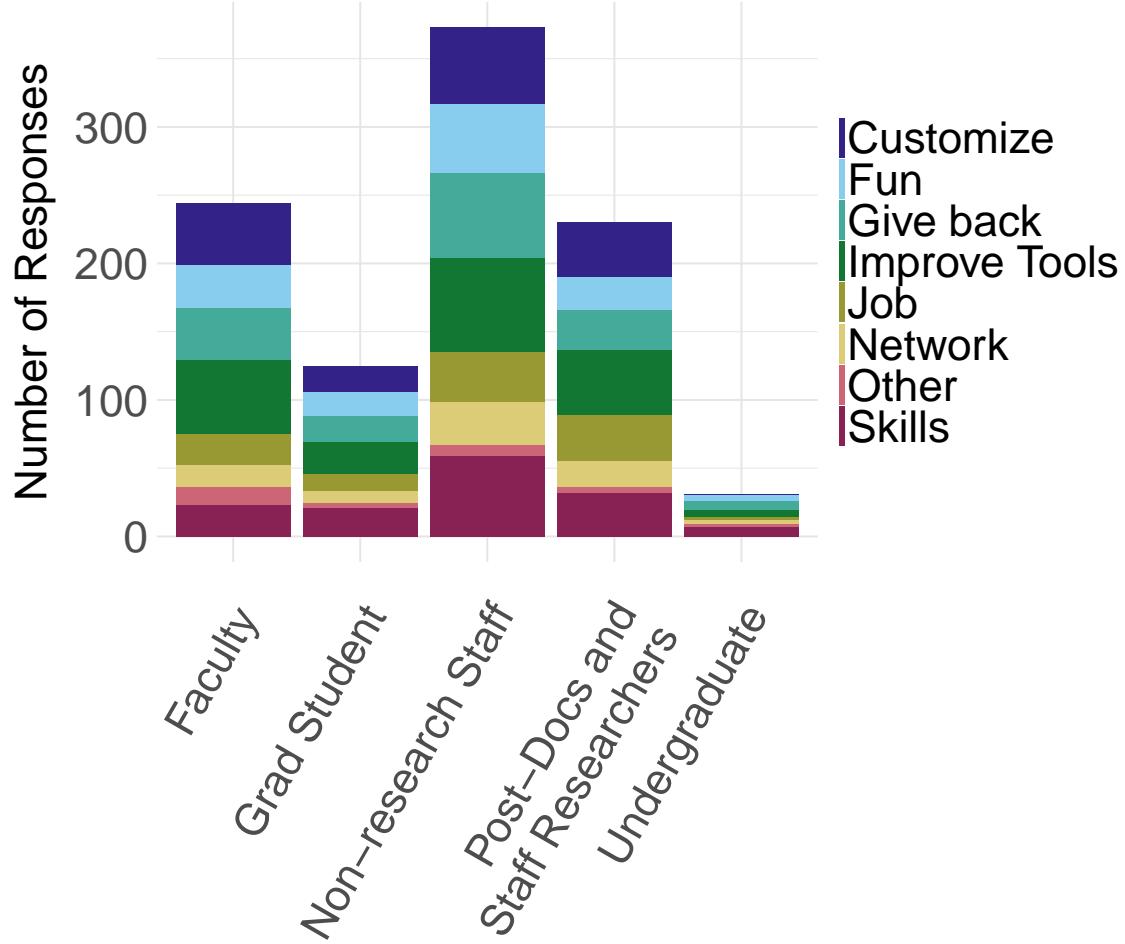
```
unique(big_counts_df$job_category)
```

```
[1] "Faculty"                      "Grad Student"  
[3] "Non-research Staff"          "Post-Docs and  
[5] "Undergraduate"                Staff Researchers"
```

Create a plot with the absolute number of responses on the y-axis.

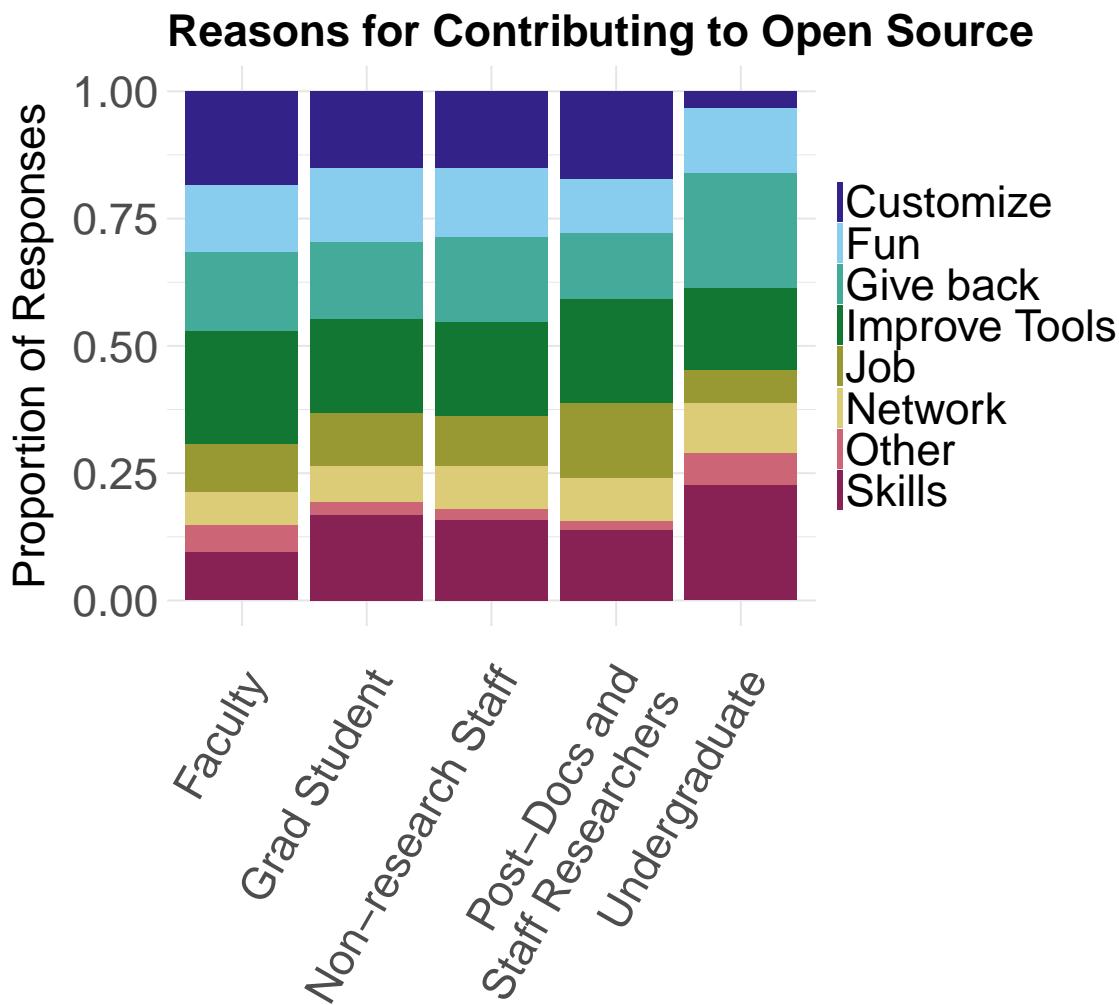
```
stacked_plot_raw <- stacked_bar_chart(big_counts_df,  
  x_var = "job_category",  
  y_var = "Count",  
  fill = "Motivation",  
  title = "Reasons for Contributing to Open Source",  
)  
stacked_plot_raw
```

## Reasons for Contributing to Open Source



Create another plot where all jobs are scaled to 1.

```
stacked_plot_proportional <- stacked_bar_chart(big_counts_df,
  x_var = "job_category",
  y_var = "Count",
  fill = "Motivation",
  title = "Reasons for Contributing to Open Source",
  proportional = TRUE
)
stacked_plot_proportional
```



I like the proportional one better. I'll save it.

```
save_plot("motivations_stacks.tiff", 12, 9, p=stacked_plot_proportional)
```

## IT vs. Academics

Request from Greg: What about IT vs. academics? (Students, Teachers, and Researchers)

```
it <- motivations_job_staff_clean %>%
  filter(staff_category == "Information Technology (IT)")
```

```

it_counts <- it %>%
  pivot_longer(
    cols = -c(job_category, staff_category),
    names_to = "Motivation",
    values_to = "Value"
  ) %>%
  select(Motivation, Value) %>%
  group_by(Motivation) %>%
  summarise(
    Count = sum(Value, na.rm = TRUE),
    .groups = "drop"
  )
it_counts$Role <- "IT"

academics <- motivations_job_clean %>%
  filter(
    job_category %in%
      c(
        "Faculty",
        "Post-Doc",
        "Other research staff",
        "Grad Student",
        "Undergraduate"
      )
  )
acad_counts <- academics %>%
  pivot_longer(
    cols = -job_category,
    names_to = "Motivation",
    values_to = "Value"
  ) %>%
  select(Motivation, Value) %>%
  group_by(Motivation) %>%
  summarise(
    Count = sum(Value, na.rm = TRUE),
    .groups = "drop"
  )
acad_counts$Role <- "Academic"

it_academics <- rbind(it_counts, acad_counts)
it_academics

```

```
# A tibble: 16 x 3
  Motivation     Count Role
  <chr>        <int> <chr>
1 Customize      23 IT
2 Fun            20 IT
3 Give back      27 IT
4 Improve Tools  26 IT
5 Job             7 IT
6 Network         10 IT
7 Other            1 IT
8 Skills           21 IT
9 Customize      105 Academic
10 Fun            78 Academic
11 Give back      94 Academic
12 Improve Tools 129 Academic
13 Job            72 Academic
14 Network         47 Academic
15 Other           22 Academic
16 Skills          83 Academic
```

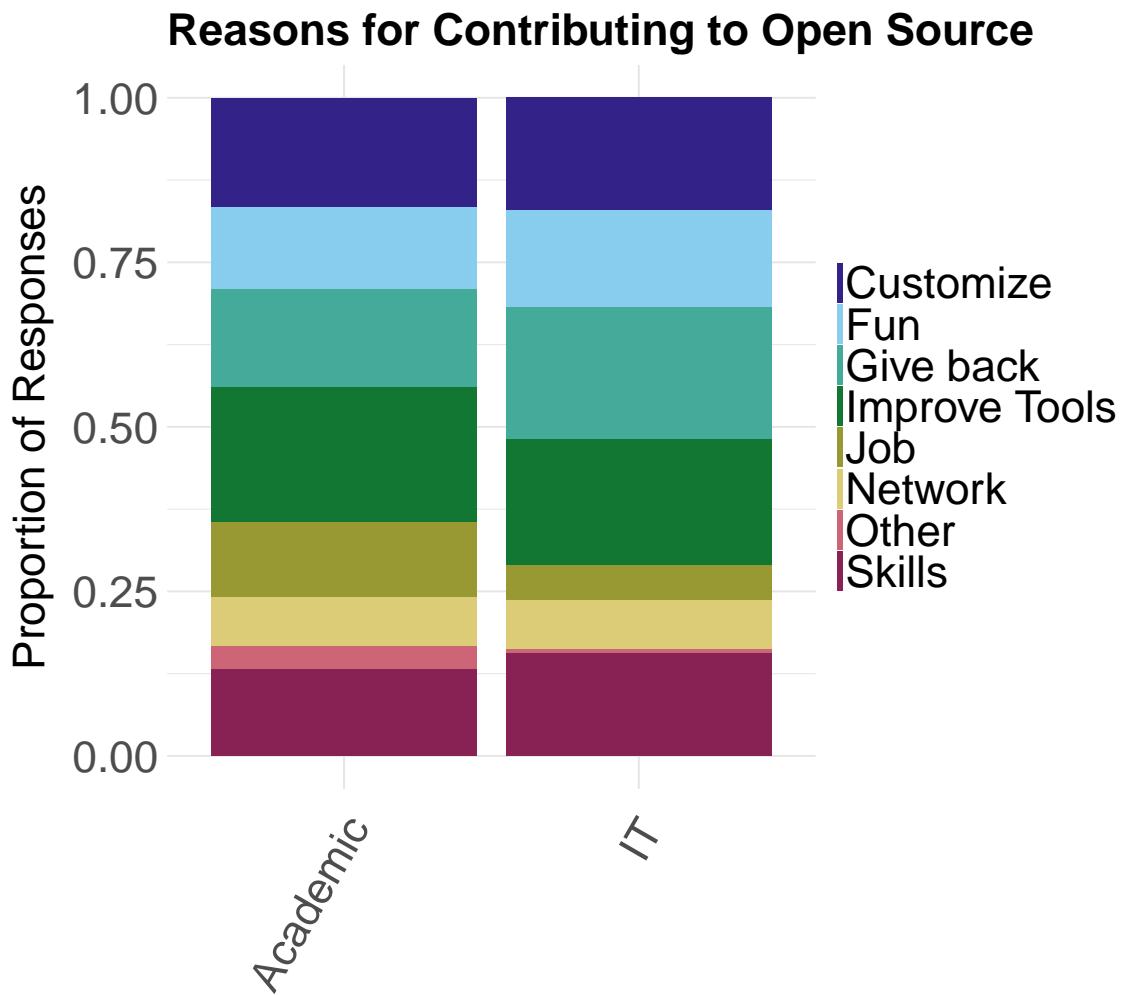
Plot

```
stacked_plot_raw_it <- stacked_bar_chart(
  it_academics,
  x_var = "Role",
  y_var = "Count",
  fill = "Motivation",
  title = "Reasons for Contributing to Open Source",
)
```

```
stacked_plot_proportional_it <- stacked_bar_chart(
  it_academics,
  x_var = "Role",
  y_var = "Count",
  fill = "Motivation",
  title = "Reasons for Contributing to Open Source",
  proportional = TRUE
)
```

Visualize

```
stacked_plot_proportional_it
```



Save

```
save_plot("motivations_stacks_it_academics.tiff", 8, 8, p=stacked_plot_proportional_it)
```

**Simple bar plot of just “doing my job”**

```
it_job <- it %>%
  select("Job")
```

```
it_job$Role <- "IT"  
head(it_job)
```

```
Job Role  
1 1 IT  
2 1 IT  
3 0 IT  
4 0 IT  
5 0 IT  
6 0 IT
```

```
dim(it_job)
```

```
[1] 33 2
```

```
# Everyone except non-research staff  
academics_job <- academics %>%  
  select("Job")  
academics_job$Role <- "Academic"  
head(academics_job)
```

```
Job      Role  
1 1 Academic  
2 0 Academic  
3 0 Academic  
4 1 Academic  
5 0 Academic  
6 0 Academic
```

```
dim(academics_job)
```

```
[1] 147 2
```

```
it_job_prop_yes <- sum(it_job$Job) / nrow(it_job)  
acad_job_prop_yes <- sum(academics_job$Job) / nrow(academics_job)  
  
it_acad_job_data <- data.frame(  
  role = c("IT", "Academic"),
```

```

proportion_yes = c(
  it_job_prop_yes,
  acad_job_prop_yes
)
)

it_acad_job_plot <- basic_bar_chart(it_acad_job_data,
  x_var = "role",
  y_var = "proportion_yes",
  title = "Percent of Respondents who said\n'Developing open source products\nis part of my job' by role",
  show_bar_labels = FALSE,
  show_ticks_y = FALSE,
  show_axis_title_y = FALSE,
  show_axis_title_x = FALSE,
  show_grid = TRUE,
  percent = TRUE,
  margin_vals = c(3, 1, 1, 1)
)

save_plot("acad_it_simple.tiff", 10, 8, p=it_acad_job_plot)

```

Save the data to a file for fine-tuning of plot.

```

write_df_to_file(
  it_acad_job_data,
  file.path("data_for_plots/motivations_it_acad.tsv")
)

```

## Scatter plots for particular motivations

All 7 undergraduates selected “Skills” and “Give back” as motivations. This made me curious about whether these motivations decrease as we get older and advance in our careers. Let’s make some line plots to investigate.

### Skills

```

# Drop staff categories (e.g. IT)
motivations_job <- subset(motivations_job_staff, select=-staff_category)
# Remove any rows where the job_category is missing
motivations_job_clean <- exclude_empty_rows(motivations_job, strict=TRUE)
# Remove rows of all 0s
motivations_job_clean <- motivations_job_clean %>%
  filter(!if_all(Job:Other, ~ .x == 0))

skills_by_role_grouped <- motivations_job_clean %>%
  group_by(job_category) %>%
  summarise(
    n_yes = sum(Skills == 1), # number of 1s
    n_tot = n(), # total rows
    Proportion = n_yes / n_tot
  )

skills_by_role_grouped <- skills_by_role_grouped %>%
  # Remove staff categories
  filter(!job_category %in% c("Non-research Staff", "Other research staff")) %>%
  # Keep only the relevant columns
  select(job_category, Proportion) %>%
  # Order the factor levels
  mutate(job_category = factor(job_category,
    levels = c(
      "Undergraduate",
      "Grad Student",
      "Post-Doc",
      "Faculty"
    ),
    ordered = TRUE
  )) %>%
  arrange(job_category)

```

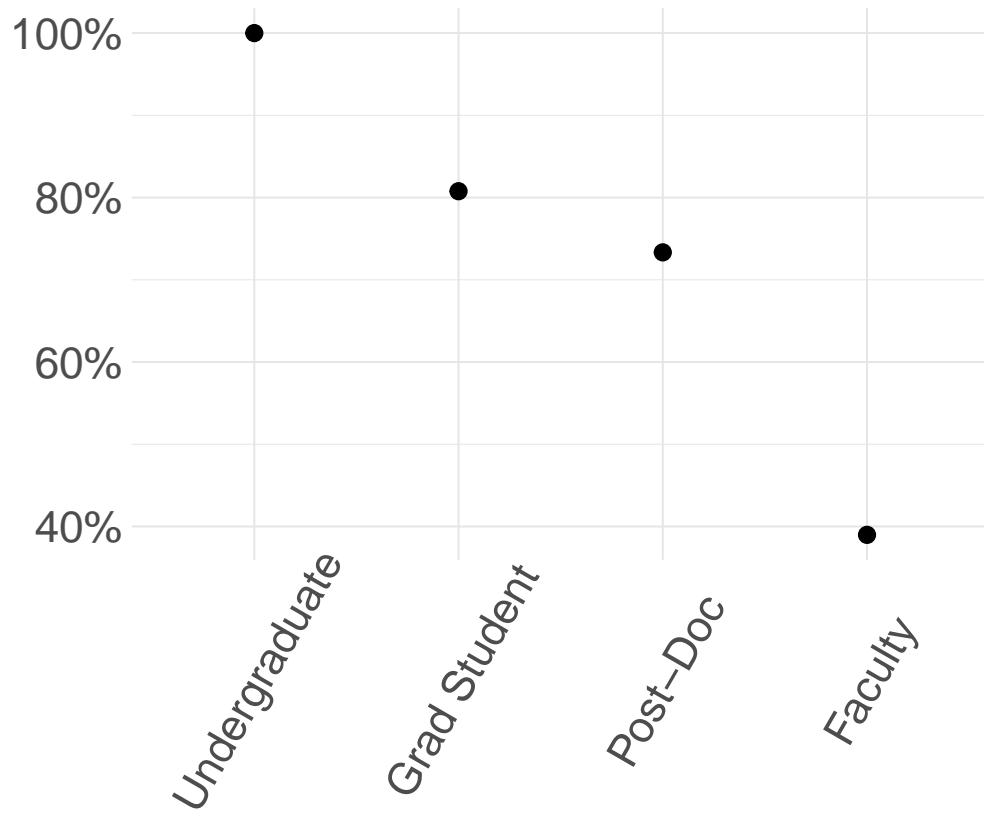
Plot and visualize

```

skills_plot <- line_plot.skills_by_role_grouped,
x_var = "job_category",
y_var = "Proportion",
title = "Percent of Participants Motivated by\nDesire to Improve Their Skills"
)
skills_plot

```

## Percent of Participants Motivated by Desire to Improve Their Skills



Save

```
save_plot("motivations_skill_by_role.tiff", 10, 8, p=skills_plot)
```

Save the data to a file for fine-tuning of plot.

```
write_df_to_file(  
    skills_by_role_grouped,  
    file.path("data_for_plots/motivations_skills.tsv")  
)
```

## Give back

What about giving back?

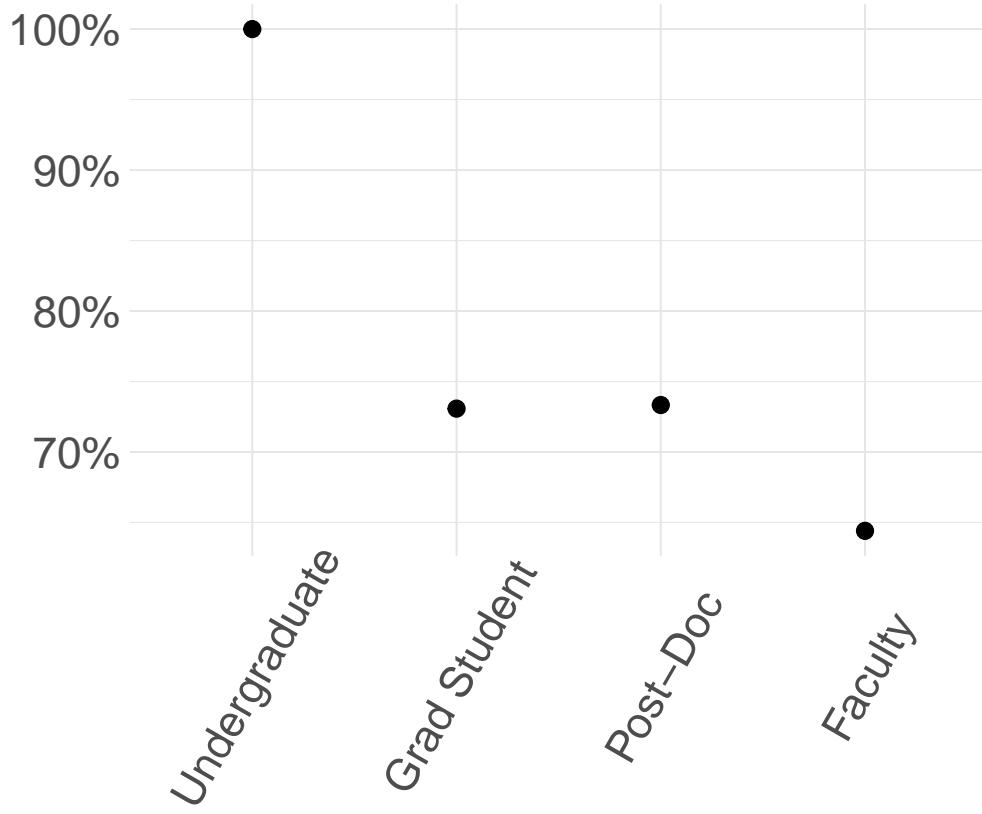
```
give_by_role_grouped <- motivations_job_clean %>%
  group_by(job_category) %>%
  summarise(
    n_yes = sum(`Give back` == 1), # number of 1s
    n_tot = n(), # total rows
    Proportion = n_yes / n_tot
  )

give_by_role_grouped <- give_by_role_grouped %>%
  # Remove staff categories
  filter(!job_category %in% c("Non-research Staff", "Other research staff")) %>%
  # Keep only the relevant columns
  select(job_category, Proportion) %>%
  # Order the factor levels
  mutate(job_category = factor(job_category,
    levels = c(
      "Undergraduate",
      "Grad Student",
      "Post-Doc",
      "Faculty"
    ),
    ordered = TRUE
  )) %>%
  arrange(job_category)
```

Plot and visualize

```
#Function defined in this notebook
give_plot <- line_plot(give_by_role_grouped,
  x_var = "job_category",
  y_var = "Proportion",
  title = "Proportion of Participants Motivated by\nDesire to Give Back"
)
give_plot
```

## Proportion of Participants Motivated by Desire to Give Back



Save

```
save_plot("motivations_giveback_by_role.tiff", 8, 6, p=give_plot)
```

## Session Info

```
sessionInfo()
```

```
R version 4.4.2 (2024-10-31)
Platform: aarch64-apple-darwin20
```

```
Running under: macOS 26.1
```

```
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; 1

locale:
[1] C.UTF-8/C.UTF-8/C.UTF-8/C/C.UTF-8/C.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    tidyverse_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    utf8_1.2.6
[19] rmarkdown_2.29      ggrepittext_0.10.2 tzdb_0.5.0
[22] ps_1.9.1           nloptr_2.2.1      purrr_1.1.0
[25] xfun_0.53          modeltools_0.2-24 jsonlite_2.0.0
[28] tweenr_2.0.3       parallel_4.4.2    prabclus_2.3-4
[31] R6_2.6.1            stringi_1.8.7     RColorBrewer_1.1-3
[34] boot_1.3-31         diptest_0.77-2    numDeriv_2016.8-1.1
[37] estimability_1.5.1 Rcpp_1.1.0        iterators_1.0.14
[40] knitr_1.50          IRanges_2.40.1     splines_4.4.2
[43] nnet_7.3-19         tidyselect_1.2.1   yaml_2.3.10
```

```
[46] doParallel_1.0.17    codetools_0.2-20   processx_3.8.6
[49] lattice_0.22-6      tibble_3.3.0     withr_3.0.2
[52] evaluate_1.0.4      polyclip_1.10-7  xml2_1.4.0
[55] circlize_0.4.16    mclust_6.1.1    kernlab_0.9-33
[58] pillar_1.11.0       renv_1.1.5      foreach_1.5.2
[61] stats4_4.4.2       reformulas_0.4.1 generics_0.1.4
[64] rprojroot_2.1.1    S4Vectors_0.44.0 hms_1.1.3
[67] minqa_1.2.8        xtable_1.8-4    class_7.3-22
[70] glue_1.8.0          robustbase_0.99-4-1 mvtnorm_1.3-3
[73] rbibutils_2.3      colorspace_2.1-1 nlme_3.1-166
[76] cli_3.6.5          textshaping_1.0.1 gtable_0.3.6
[79] DEoptimR_1.1-4     digest_0.6.37   BiocGenerics_0.52.0
[82] ucminf_1.2.2       ggrepel_0.9.6   rjson_0.2.23
[85] farver_2.1.2        htmltools_0.5.8.1 lifecycle_1.0.4
[88] GlobalOptions_0.1.2 MASS_7.3-61
```