

# 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 = 0.5, size = 24),
plot.margin = unit(c(0.3, 0.3, 0.3, 0.3), "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
2	1	0	0	0	1	0
3	0	1	1	1	0	0
4	1	0	0	1	1	0
5	0	1	1	0	1	1
6	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))

```

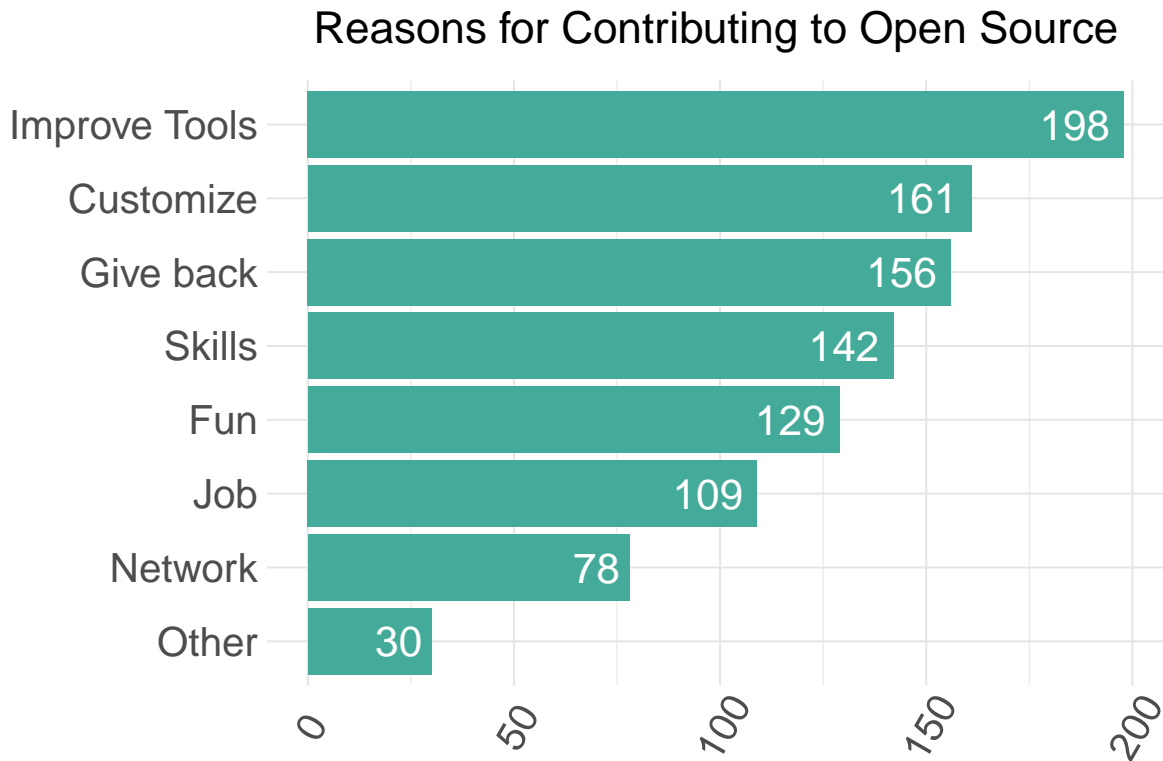
And make a plot, using a function in utils.R.

```

myplot <- basic_bar_chart(motivations_to_plot,
  x_var = "Motivation",
  y_var = "Count",
  title = "Reasons for Contributing to Open Source",
  horizontal = TRUE,
  show_bar_labels = TRUE,
  show_ticks_y = FALSE,
  color_index = 3,
  show_axis_title_y = FALSE,
  show_grid = TRUE
)

myplot

```



Save the plot if you wish.

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

## 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
```

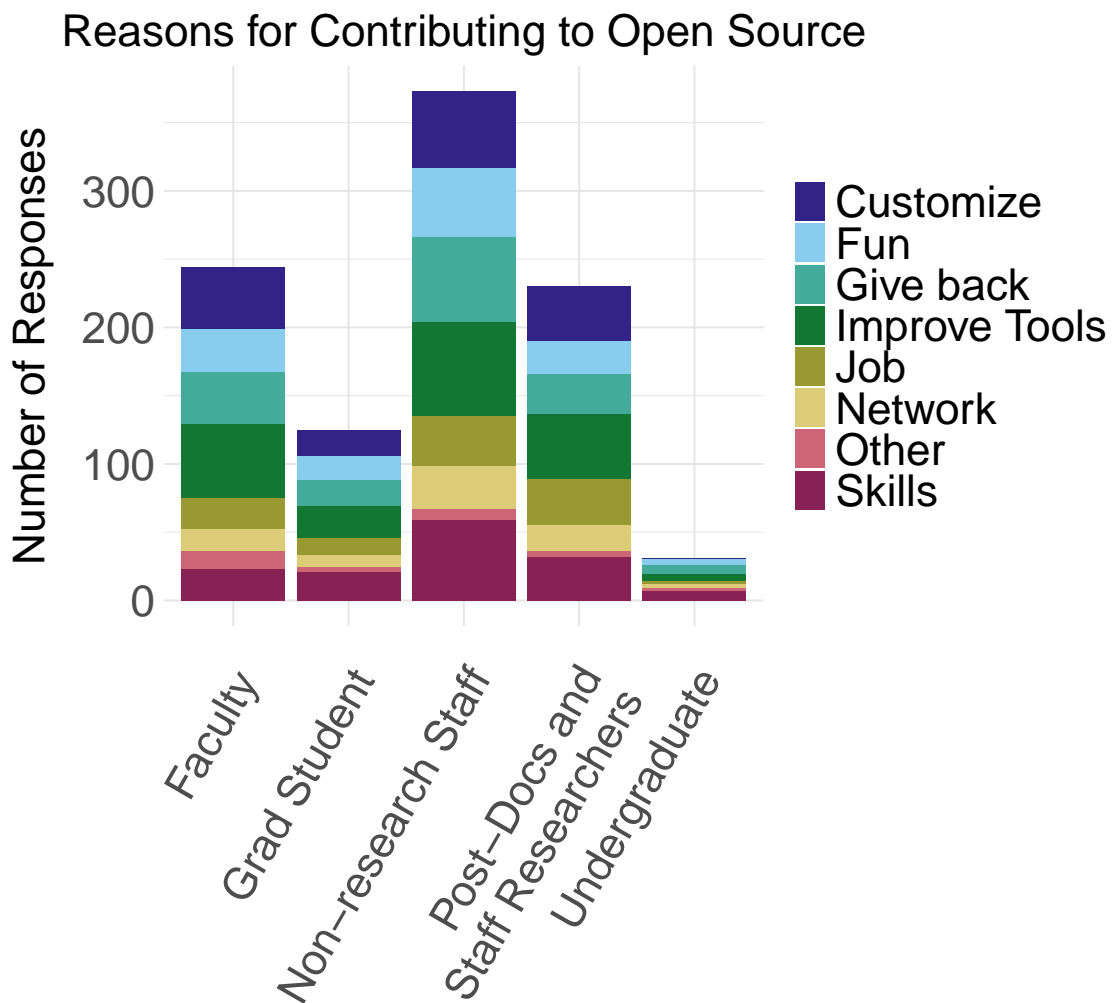
```
# A tibble: 40 x 3
  job_category Motivation    Count
  <chr>         <chr>      <int>
1 Faculty      Customize    45
2 Faculty      Fun          32
3 Faculty      Give back    38
4 Faculty      Improve Tools 54
5 Faculty      Job          23
6 Faculty      Network     16
7 Faculty      Other        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\nStaff Researchers"
[5] "Undergraduate"
```

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

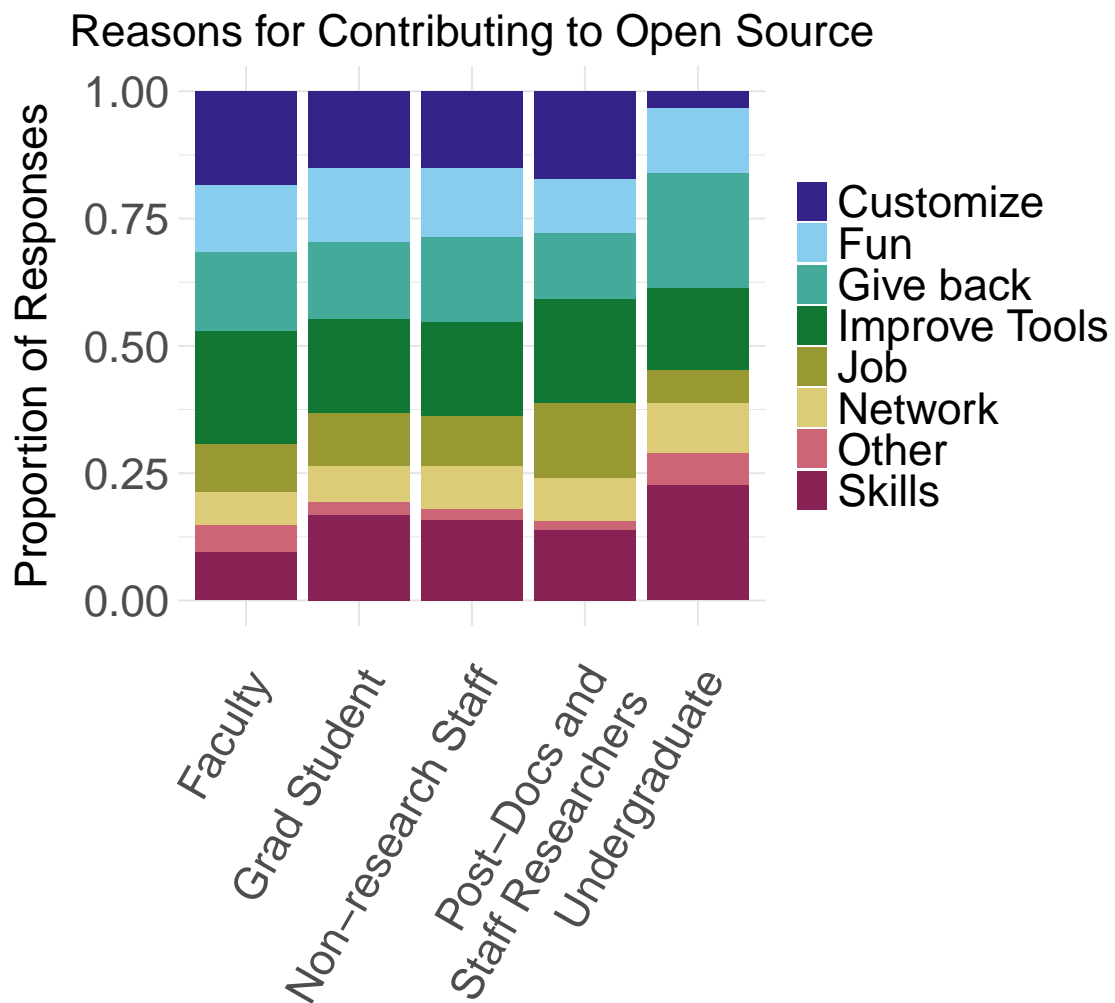
```
stacked_plot_raw <- stacked_bar_chart(big_counts_df,
  x_var = "job_category",
  y_var = "Count",
  fill = "Motivation",
  title = "Reasons for Contributing to Open Source",
)
stacked_plot_raw
```



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
```

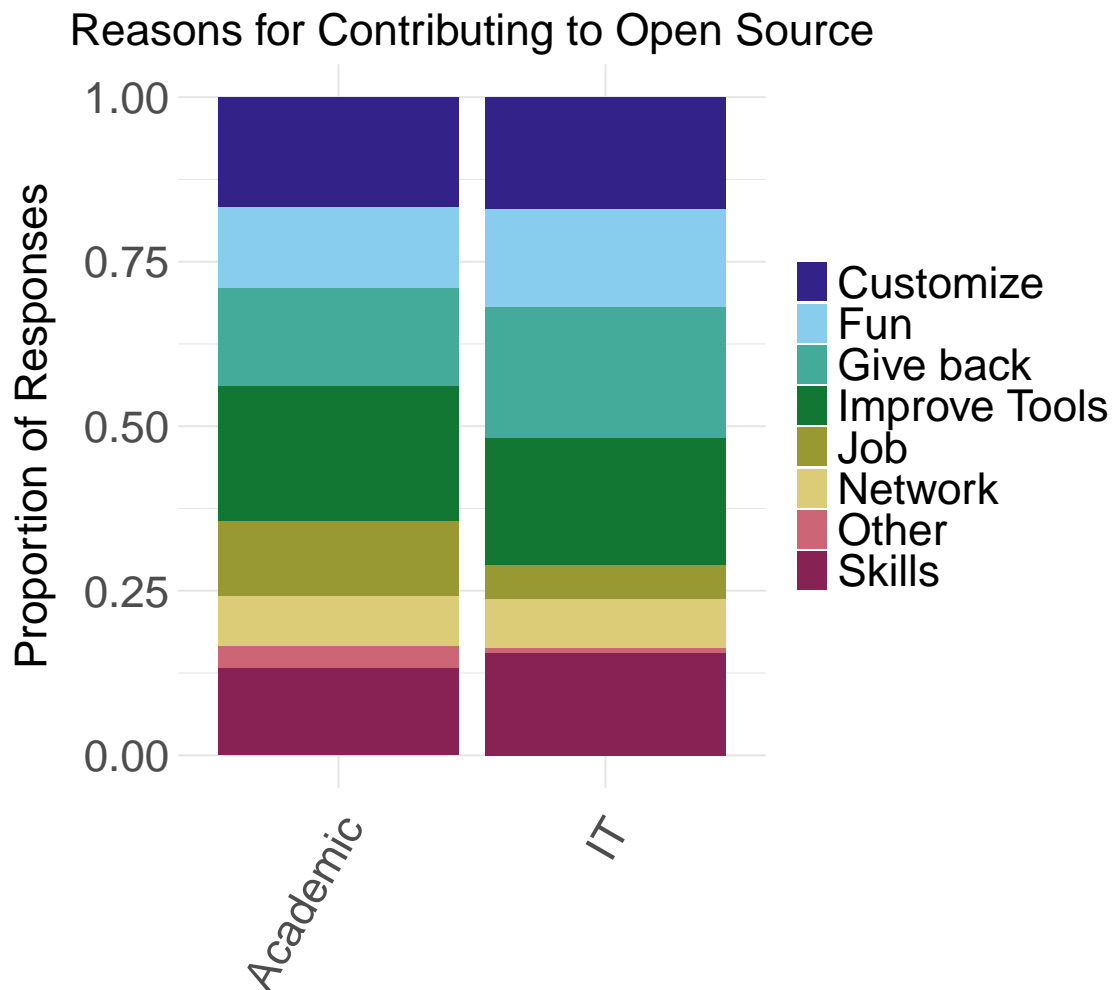
Plot

```
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)
```

## Line 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.

```

# 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)

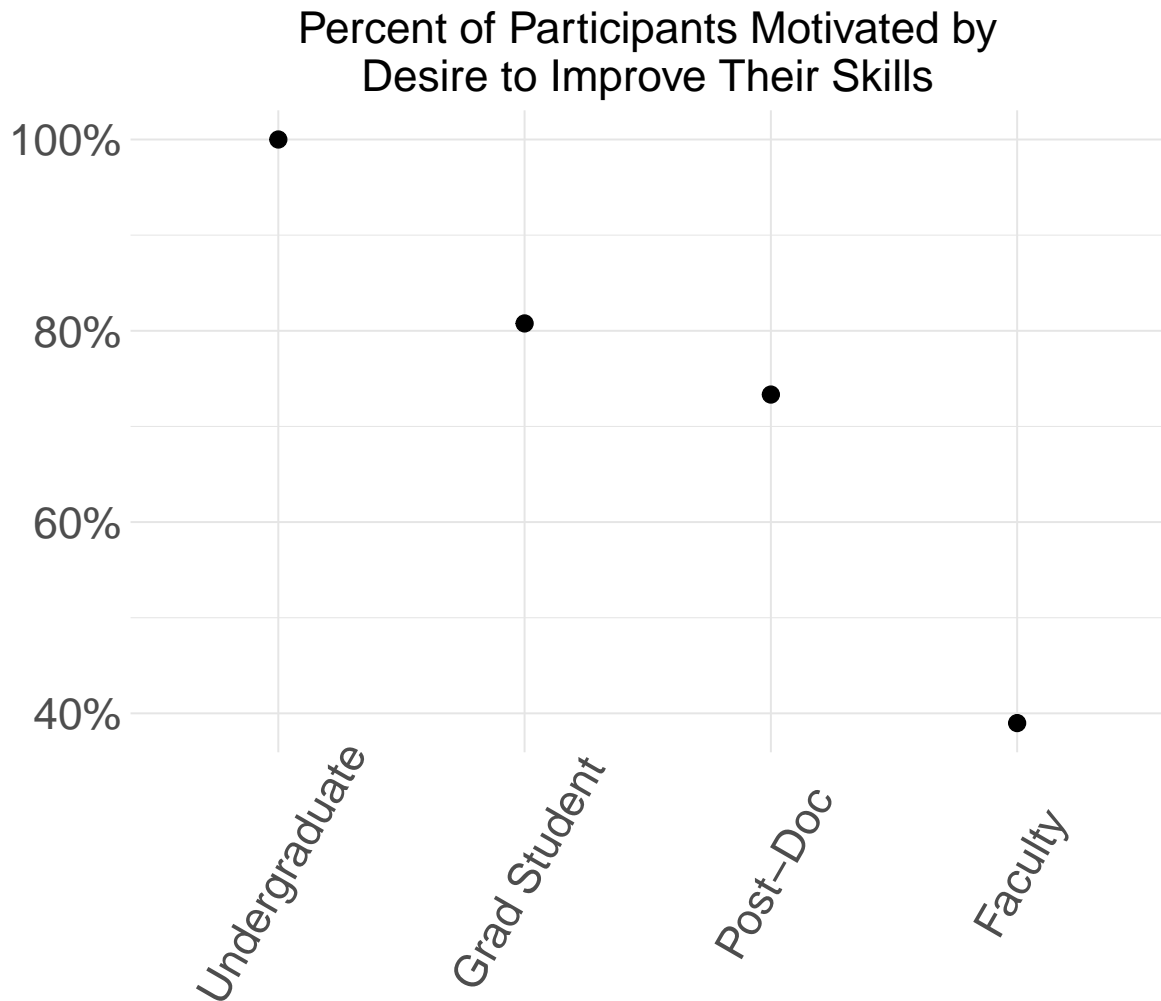
```

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

```



Save

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

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)

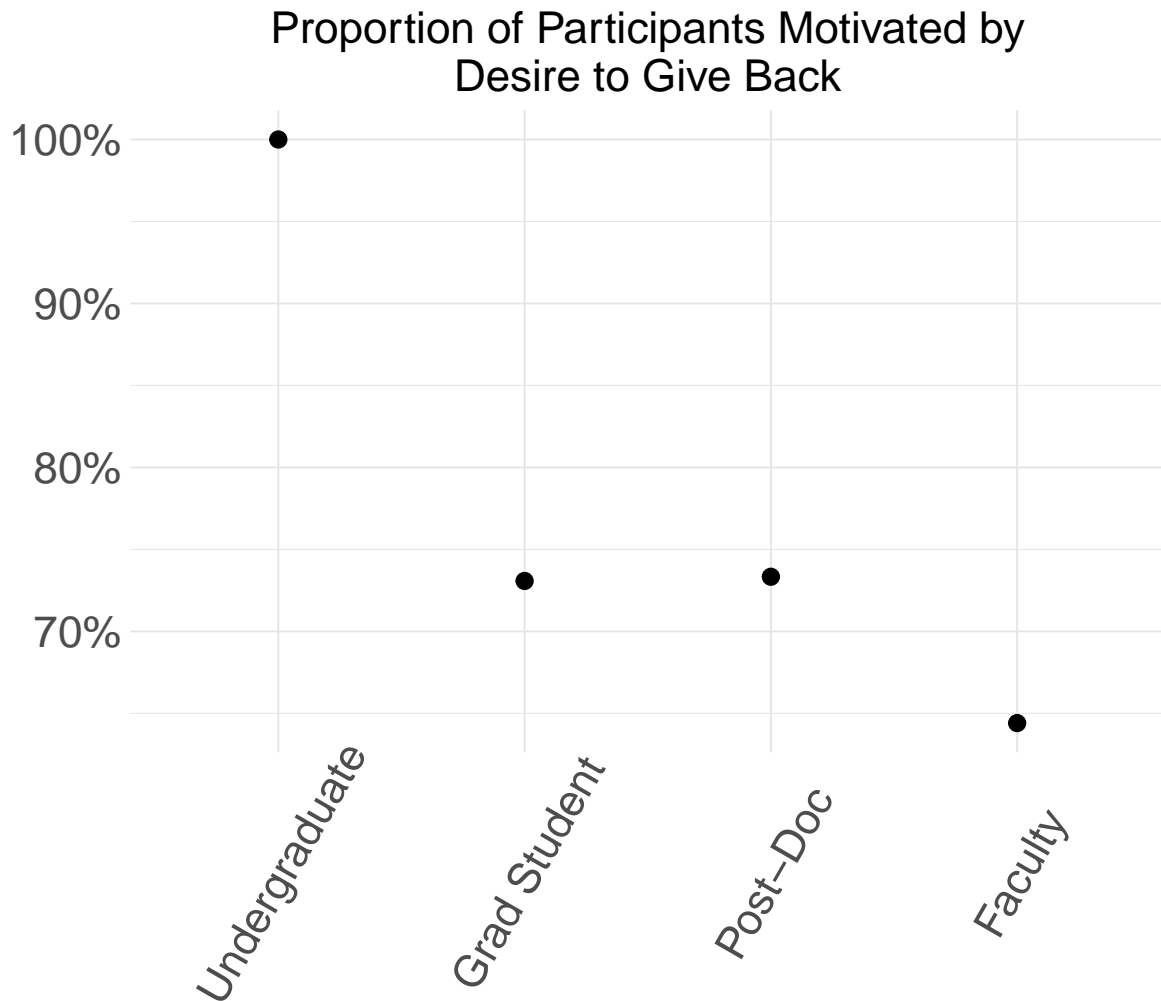
```

Plot and visualize

```

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

```



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 Sequoia 15.4.1

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 stats graphics grDevices datasets utils methods  
[8] base

other attached packages:

[1] treemap_2.4-4	tidyr_1.3.1	stringr_1.5.1
[4] scales_1.4.0	readr_2.1.5	pwr_1.3-0
[7] patchwork_1.3.0	mvabund_4.2.1	languageserver_0.3.16
[10] here_1.0.1	gttools_3.9.5	fpc_2.2-13
[13] forcats_1.0.0	factoextra_1.0.7	ggplot2_3.5.2
[16] dplyr_1.1.4	corrplot_0.95	cluster_2.1.8.1

loaded via a namespace (and not attached):

[1] gtable_0.3.6	xfun_0.52	ggrepel_0.9.6
[4] processx_3.8.6	lattice_0.22-6	callr_3.7.6
[7] tzdb_0.5.0	vctrs_0.6.5	ps_1.9.1
[10] generics_0.1.4	stats4_4.4.2	parallel_4.4.2
[13] flexmix_2.3-20	tibble_3.2.1	DEoptimR_1.1-3-1
[16] pkgconfig_2.0.3	data.table_1.17.6	RColorBrewer_1.1-3
[19] lifecycle_1.0.4	compiler_4.4.2	farver_2.1.2
[22] statmod_1.5.0	httpuv_1.6.16	htmltools_0.5.8.1
[25] class_7.3-22	yaml_2.3.10	later_1.4.2
[28] pillar_1.10.2	prabclus_2.3-4	MASS_7.3-61
[31] diptest_0.77-1	mclust_6.1.1	mime_0.13
[34] robustbase_0.99-4-1	tidyselect_1.2.1	digest_0.6.37
[37] stringi_1.8.7	purrr_1.0.4	kernlab_0.9-33
[40] labeling_0.4.3	rprojroot_2.0.4	fastmap_1.2.0
[43] grid_4.4.2	colorspace_2.1-1	cli_3.6.5
[46] magrittr_2.0.3	utf8_1.2.5	withr_3.0.2
[49] promises_1.3.3	tweedie_2.3.5	rmarkdown_2.29
[52] igraph_2.1.4	nnet_7.3-19	modeltools_0.2-24

[55]	<code>hms_1.1.3</code>	<code>shiny_1.11.0</code>	<code>evaluate_1.0.3</code>
[58]	<code>knitr_1.50</code>	<code>rlang_1.1.6</code>	<code>Rcpp_1.0.14</code>
[61]	<code>xtable_1.8-4</code>	<code>gridBase_0.4-7</code>	<code>glue_1.8.0</code>
[64]	<code>xml2_1.3.8</code>	<code>renv_1.1.4</code>	<code>jsonlite_2.0.0</code>
[67]	<code>R6_2.6.1</code>		