Connectome 2023 reproducibility survey analysis

Mateus Harrington

Load packages

```r
# list packages
pkg <- c("tidyverse", "here", "gt", "viridis")
# install any missing packages and load them
lapply(pkg, function(pkg) {
  if (!require(pkg, character.only = TRUE)) {
    install.packages(pkg)
    library(pkg, character.only = TRUE)
  } else {
    library(pkg, character.only = TRUE)
  }
})
```

Read in data

```r
# read in data
df <-
  readr::read_csv(
    here::here(
      "04_data/001_connectome_2023/2023-10-13_reproducibility-sentiment-analysis.csv"
    ),
    show_col_types = FALSE
  )
gt::gt(df)
```

question	unfamilar	aware	experienced	actively_using
Version control	2	11	2	14
Unit Tests	16	7	2	3
Containerisation	13	3	2	8
Virtual Environments	2	11	1	14
Workflow Managers	12	7	2	7
CI/CD	20	3	1	3

```r
# pivot to long format
df_long <- df %>%
  pivot_longer(cols = -question, names_to = "experience", values_to = "count")

gt::gt(head(df_long))
```

question	experience	count
Version control	unfamilar	2
Version control	aware	11
Version control	experienced	2
Version control	actively_using	14
Unit Tests	unfamilar	16
Unit Tests	aware	7

```r
numeric_cols <- sapply(df, is.numeric)

# Get the sum of each numeric column
sums <- colSums(df[, numeric_cols])
sums
```

```
      unfamilar           aware     experienced  actively_using
             65              42              10              49
```

```r
df_long %>%
  group_by(question) %>%
  summarise(total_responses = sum(count)) %>%
  gt::gt()
```

question	total_responses
CI/CD	27
Containerisation	26
Unit Tests	28
Version control	29
Virtual Environments	28
Workflow Managers	28

```r
# Create the ggplot with dodged bars and colourblind-friendly palette
plot <- ggplot(df_long, aes(x = question, y = count, fill = experience)) +
  geom_col(position = "dodge") +
  scale_fill_viridis_d() +  # Use viridis colour palette
  labs(x = "Question", y = "Count") +
  theme_bw() +
  # Rotate x-axis labels by 45 degrees
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        text = element_text(size = 26))
plot

# Save plot
ggsave(here::here("06_figures/001_connectome_2023_analysis/count_plot.svg"),
       plot, device = "svg")
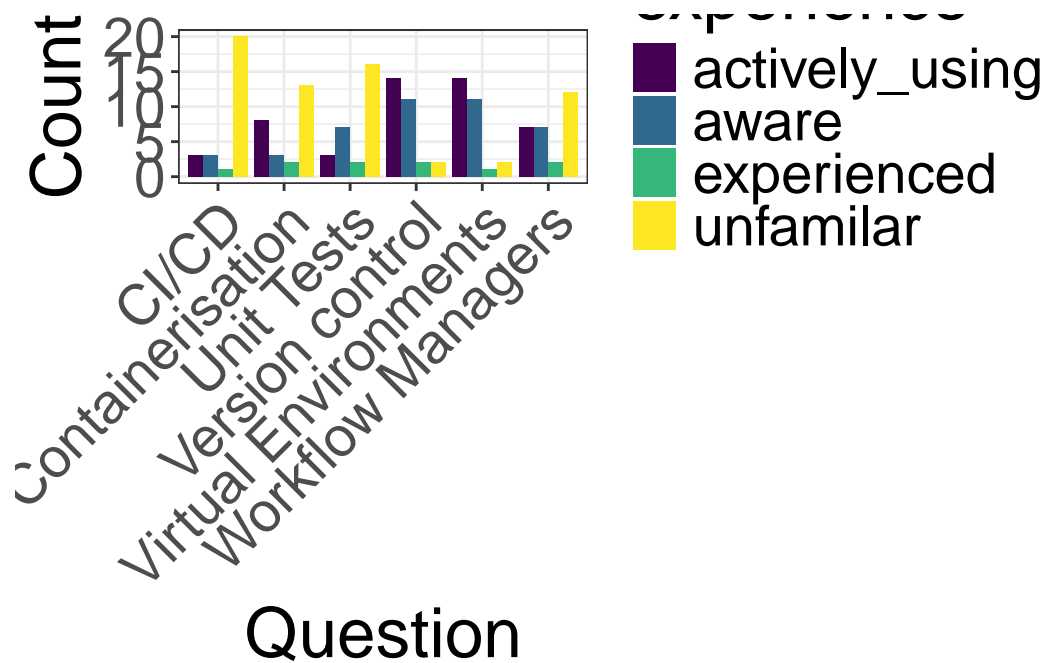```

```
Saving 5.5 x 3.5 in image
```

Figure 1: Counts of responses to reproducibility questions gathered at connectome 2023.

Seems like most folk aren't familiar with CI/CD, containers, unit tests and workflow managers (Figure 1). Most people are aware off/actively using version control and virtual environments though. One wonders to what extend they're fully utilising even those though!