Software quality analysis of fourteen hydrological model

Arnald Puy

Contents

1 Preliminary functions

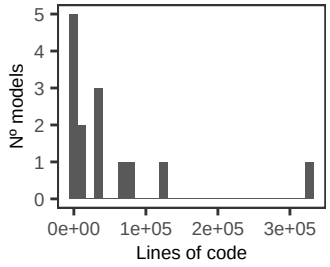
```
#   PRELIMINARY FUNCTIONS ###########################################################

sensobol::load_packages(c("data.table", "tidyverse", "openxlsx", "scales",
                          "cowplot"))

# Create custom theme ----------------------------------------------------------

theme_AP <- function() {
  theme_bw() +
    theme(panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.background = element_rect(fill = "transparent",
                                           color = NA),
          legend.key = element_rect(fill = "transparent",
                                    color = NA),
          strip.background = element_rect(fill = "white"),
          legend.text = element_text(size = 7.3),
          axis.title = element_text(size = 10),
          legend.key.width = unit(0.4, "cm"),
          legend.key.height = unit(0.4, "cm"),
          legend.key.spacing.y = unit(0, "lines"),
          legend.box.spacing = unit(0, "pt"),
          legend.title = element_text(size = 7.3),
          axis.text.x = element_text(size = 7),
          axis.text.y = element_text(size = 7),
          axis.title.x = element_text(size = 7.3),
          axis.title.y = element_text(size = 7.3),
          plot.title = element_text(size = 8),
          strip.text.x = element_text(size = 7.4),
          strip.text.y = element_text(size = 7.4))
}

# Select color palette ---------------------------------------------------------

selected.palette <- "Darjeeling1"
```

2 Dataset

```
# READ IN DATASET ##################################################################

dt <- data.table(read.xlsx("./datasets/results_sqa.xlsx"))
cols <- colnames(dt)
color_languages <- c("fortran" = "steelblue", "python" = "lightgreen")
```

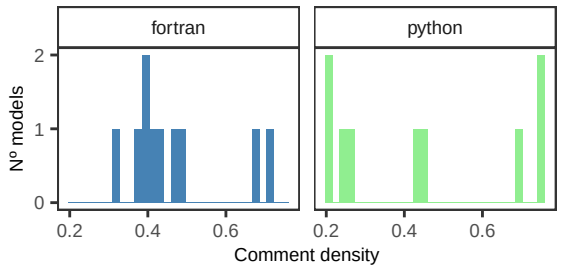
```
# PLOT LINES OF CODE ############################################################

plot_lines_code <- dt[, .(total_lines_code = sum(lines_code)), model] %>%
  ggplot(., aes(total_lines_code)) +
  geom_histogram() +
  labs(x = "Lines of code", y = "Nº models") +
  theme_AP()

plot_lines_code
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# PLOT COMMENT DENSITY ##########################################################

plot_comment_density <- dt[, .(total_lines_code = sum(lines_code),
       total_lines_comments = sum(lines_comments)), .(model, language)] %>%
  .[, comment_density:= total_lines_comments / total_lines_code] %>%
  ggplot(., aes(comment_density, fill = language)) +
  geom_histogram() +
  facet_wrap(~language) +
  scale_y_continuous(breaks = breaks_pretty(n = 3)) +
  scale_fill_manual(values = color_languages) +
  labs(x = "Comment density", y = "Nº models") +
  theme_AP() +
  theme(legend.position = "none")

plot_comment_density
```

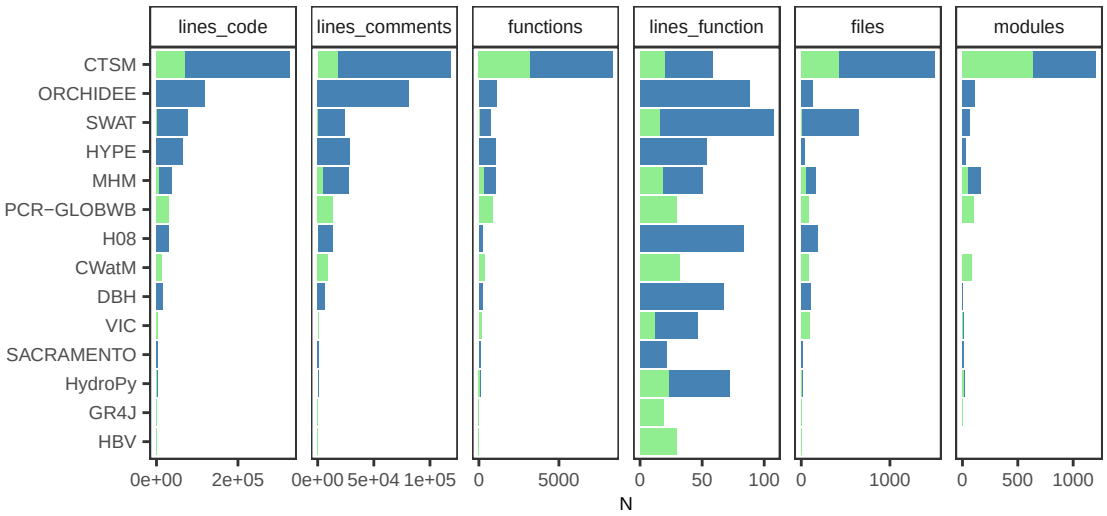
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
model_ordered <- dt[, sum(lines), model] %>%
  .[order(V1)] %>%
  .[, model]
```

```
col_names <- colnames(dt)

facet_order <- c("lines", "lines_code", "lines_comments", "functions",
                 "lines_function", "files", "modules")

plot_per_model <- melt(dt, measure.vars = col_names[-c(1, length(col_names))]) %>%
  .[, variable:= factor(variable, levels = facet_order)] %>%
  .[, model:= factor(model, levels = model_ordered)] %>%
  .[!variable == "lines"] %>%
  ggplot(., aes(model, value, fill = language)) +
  geom_col() +
  coord_flip() +
  scale_y_continuous(breaks = breaks_pretty(n = 2)) +
  scale_fill_manual(values = color_languages) +
  facet_wrap(~ variable, ncol = 7, scales = "free_x") +
  labs(x = "", y = "N") +
  theme_AP() +
  theme(legend.position = "none")

plot_per_model
```

```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (`geom_col()`).
```




```
# MERGE PLOTS ######################################################################

top <- plot_grid(plot_lines_code, plot_comment_density + labs(x = "Comment density", y  = ""),
          labels = "auto", rel_widths = c(0.4, 0.6))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
plot_grid(top, plot_per_model, ncol = 1, labels = c("", "c"), rel_heights = c(0.4, 0.6))
```

```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (`geom_col()`).
```