

Runtime Plots

Evan Bolyen

R Markdown

```
library(readr)
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(tidyr)
library(tibble)

div_df = read_tsv('results_diversity.tsv')

## Parsed with column specification:
## cols(
##   file = col_character(),
##   s = col_double(),
##   `h:m:s` = col_time(format = ""),
##   max_rss = col_double(),
##   max_vms = col_double(),
##   max_uss = col_double(),
##   max_pss = col_double(),
##   io_in = col_double(),
##   io_out = col_double(),
##   mean_load = col_double()
## )

div_df = div_df %>%
  separate(file, c(NA, 'context', 'iter'), '_', convert=TRUE) %>%
  add_column(focal=NA, type='diversity')

nb_df = read_tsv('results_neighbors.tsv')

## Parsed with column specification:
## cols(
##   file = col_character(),
##   s = col_double(),
##   `h:m:s` = col_time(format = ""),
##   max_rss = col_double(),
```

```

##   max_vms = col_double(),
##   max_uss = col_double(),
##   max_pss = col_double(),
##   io_in = col_double(),
##   io_out = col_double(),
##   mean_load = col_double()
## )

nb_df = nb_df %>%
  separate(file, c(NA, NA, 'focal', NA, NA, 'context', 'iter'), '_', convert=TRUE) %>%
  add_column(type='neighbors')

library(reshape2)

##
## Attaching package: 'reshape2'
## The following object is masked from 'package:tidyr':
##
##   smiths

bench_df = union(div_df, nb_df)
bench_df = melt(bench_df,
  id.vars=c('focal', 'context', 'iter', 'type', 'h:m:s', 'max_rss'),
  measure.vars = c('h:m:s', 'max_rss'),
  value.name='IGNORE')

## Warning: attributes are not identical across measure variables; they will be
## dropped

library(ggplot2)
library(patchwork)

base = ggplot(bench_df, aes(color=as.factor(context))) +
  theme_bw() + theme(panel.border = element_blank()) +
  scale_color_viridis_d(direction = -1)

p1 = base +
  geom_jitter(data=subset(bench_df, type=='diversity' & variable=='h:m:s'),
    aes(x=context, y=`h:m:s`), width=1000, height = 0, shape=21) +
  theme(legend.position = 'none',
    axis.text.x = element_blank(),
    axis.title.x = element_blank(),
    axis.ticks.x = element_blank(),
    axis.line.y = element_line()) +
  ggtitle("sample-diversity") +
  ylab("Runtime (h:m:s)") +
  scale_y_time(limits = c(0, 61285.00)) +
  scale_x_continuous(limits = c(0, 80000))
p2 = base +
  geom_jitter(data=subset(bench_df, type=='neighbors' & variable=='h:m:s'),
    aes(x=focal, y=`h:m:s`), width=10, height = 0, shape=21) +
  theme(axis.title.y = element_blank(),
    axis.text.x = element_blank(),
    axis.title.x = element_blank(),
    axis.ticks.x = element_blank(),

```

```

    legend.position = 'none',
    axis.line.y = element_line() +
ggtitle("sample-neighbors") +
scale_y_time(limits = c(0, 3681.00)) +
scale_x_continuous(limits = c(0, 750))
p3 = base +
geom_jitter(data=subset(bench_df, type=='diversity' & variable=='max_rss'),
            aes(x=context, y=max_rss), width=1000, height = 0, shape=21) +
theme(legend.position = 'none',
      axis.line.y = element_line(),
      axis.line.x = element_line()) +
ylab("Max RSS (MiB)") +
xlab("Context Seqs (N)") +
labs(color=expression(paste('Context\nSeqs (N)')))) +
scale_y_continuous(limits = c(0, 3072), breaks = seq(0, 3072, 1024)) +
scale_x_continuous(limits = c(0, 80000))
p4 = base +
geom_jitter(data=subset(bench_df, type=='neighbors' & variable=='max_rss'),
            aes(x=focal, y=max_rss), width=10, height = 0, shape=21) +
theme(axis.title.y = element_blank(),
      axis.text.y = element_blank(),
      axis.ticks.y = element_blank(),
      axis.line.x = element_line()) +
xlab("Focal Seqs (N)") +
labs(color=expression(paste('Context\nSeqs (N)')))) +
scale_y_continuous(limits = c(0, 3072), breaks = seq(0, 3072, 1024)) +
scale_x_continuous(limits = c(0, 750))

p1 + p2 + p3 + p4

```

