

# Read someone else's code

Silvie Cinková

2025-08-09

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.1      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.1
v purrr      1.0.4
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
billionaires_df <- read_tsv("../datasets_ATRIUM/billionaires_combined.tsv")
```

```
Rows: 28986 Columns: 21
```

```
-- Column specification -----
```

```
Delimiter: "\t"
```

```
chr (17): person, name.x, state, headquarters, source, industry, gender, las...
```

```
dbl (4): time, daily_income, age, birth_comb
```

```
i Use `spec()` to retrieve the full column specification for this data.
```

```
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
billionaires_2020 <- billionaires_df %>%
  filter(time == 2020) %>% select(daily_income, person, world_6region)
```

# 1 The task performed by the code

The data set compiles charts of world's billionaires between 2002 and 2020. This means that the same person can occur more than one time, but their details may differ. Plot billionaires in each world region (column `world_6region`). Use boxplots to plot the distribution of `daily_income` among all billionaires in the given world region and text labels with values from the column `person` for outliers. You may have to plot each region separately, that is, not break one plot into facets but run the/a plotting script individually for each world region.

Solution

```
library(ggrepel)
billionaires_outliers <- billionaires_2020 %>%
  group_by(world_6region) %>%
  mutate(Q3 = quantile(daily_income, 0.75),
         IQR_col = IQR(daily_income),
         outliers_above = Q3 + 1.5 * IQR_col) %>%
  mutate(is_outlier = if_else(condition = daily_income > outliers_above,
                              true = TRUE,
                              false = FALSE)) %>%
  arrange(desc(daily_income), world_6region)
world_6region_vec <- distinct(billionaires_outliers, world_6region) %>%
  arrange(world_6region) %>%
  pull()
outliers_above <- distinct(billionaires_outliers, world_6region, outliers_above) %>%
  arrange(world_6region) %>%
  pull()

for (i in seq_along(world_6region_vec)) {
  all_oneregion <- billionaires_outliers %>%
    filter(world_6region == world_6region_vec[i])
  outliers_oneregion <- all_oneregion %>%
    filter(is_outlier == TRUE)
  y_axis_offset_for_outlier_label <- range(all_oneregion$daily_income) %>%
    diff()
  y_axis_offset_for_outlier_label <- y_axis_offset_for_outlier_label * 0.05
  if (nrow(outliers_oneregion) == 0){
    cat(world_6region_vec[i], "has no outliers. I will plot all names.\n")
  }
  p <- ggplot() +
    geom_boxplot(data = all_oneregion,
                mapping = aes(y = daily_income, x = 1),
                color = "purple") +
    geom_text_repel(mapping = aes(y = daily_income,
                                x = 1,
                                label = person,
                                size = daily_income),
                  data = all_oneregion,
                  max.overlaps = 100,
```

[illegible]

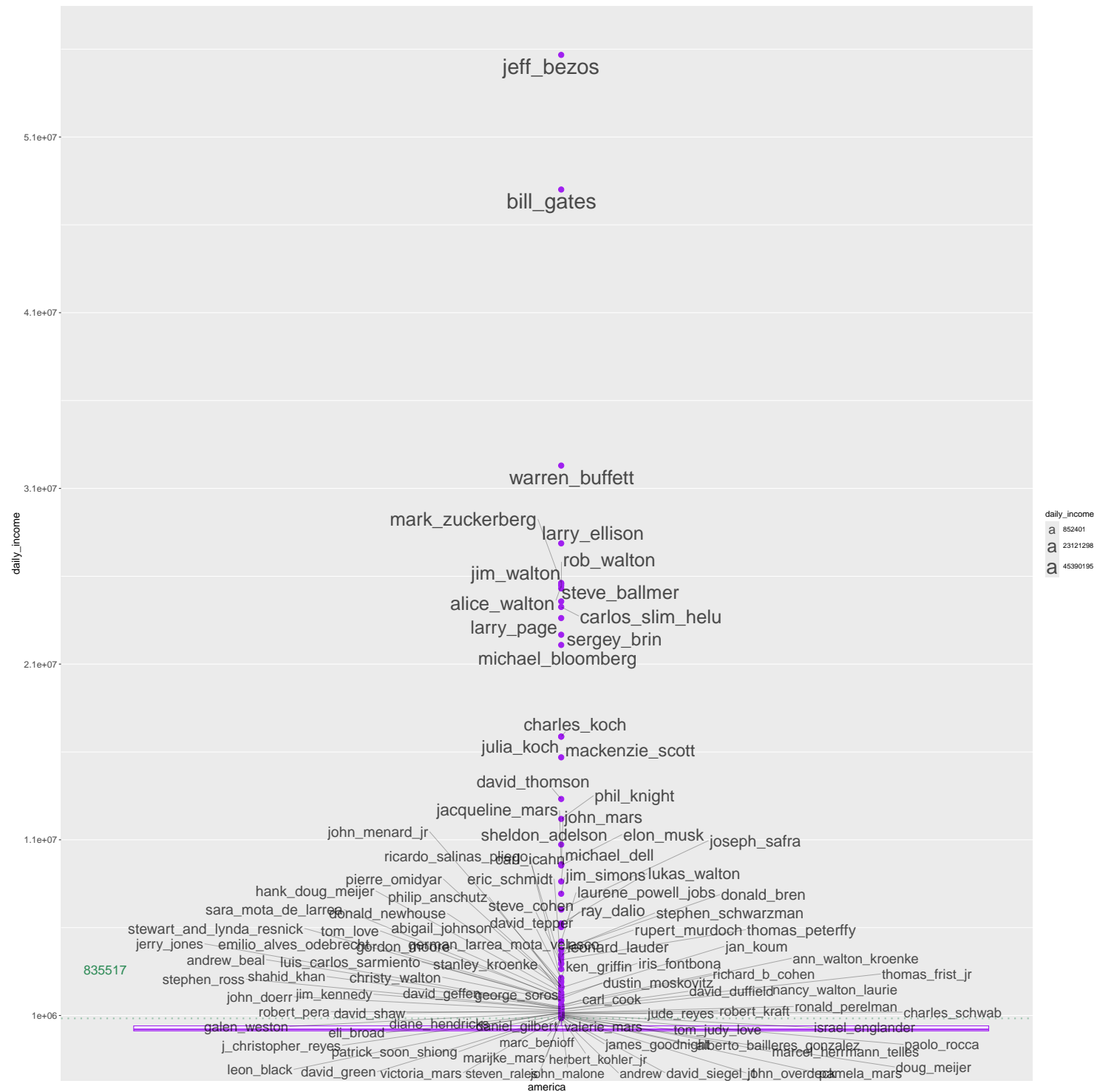
```

max(all_oneregion$daily_income) * 10^(-6)) * 10^6),
by = 10^7))) +
geom_hline(yintercept = outliers_above[i],
  color = "seagreen",
  linewidth = 1,
  linetype=3,
  alpha = 0.4) +
annotate(geom = "text",
  x = 0.6,
  y = outliers_above[i] + y_axis_offset_for_outlier_label,
  label = outliers_above[i],
  color = "seagreen",
  size = 6) +
theme(axis.text = element_text(size = 12),
  axis.title = element_text(size = 14))
cat(world_6region_vec[i], "\n")
print(p)
ggsave(plot = p, filename = paste0(
  "../my_output_files/outliers_billionaires-",
  world_6region_vec[i], ".pdf"),
  width = 7 * 2.2)
}
}

```

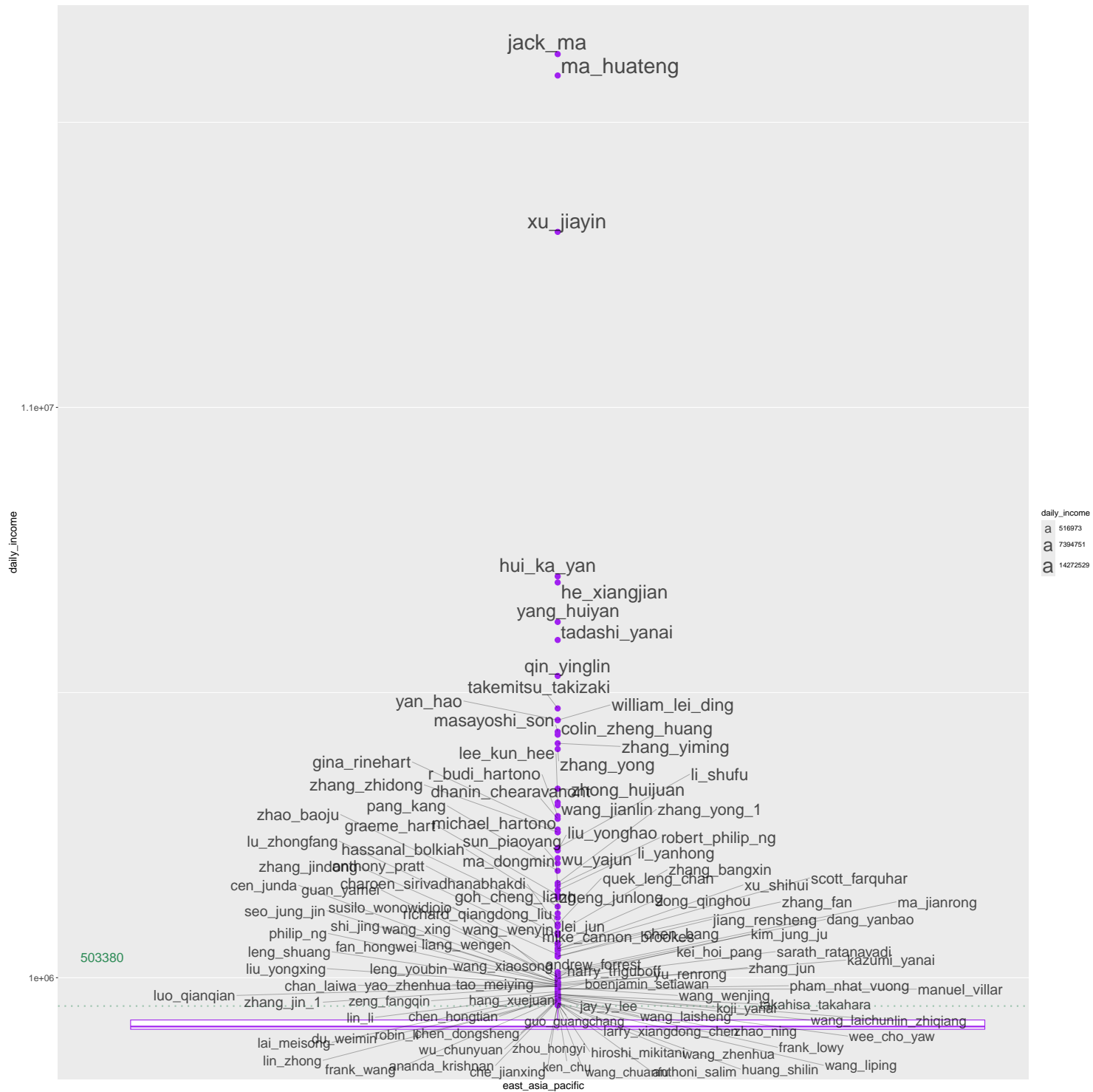
america

Saving 15.4 x 20 in image



east\_asia\_pacific

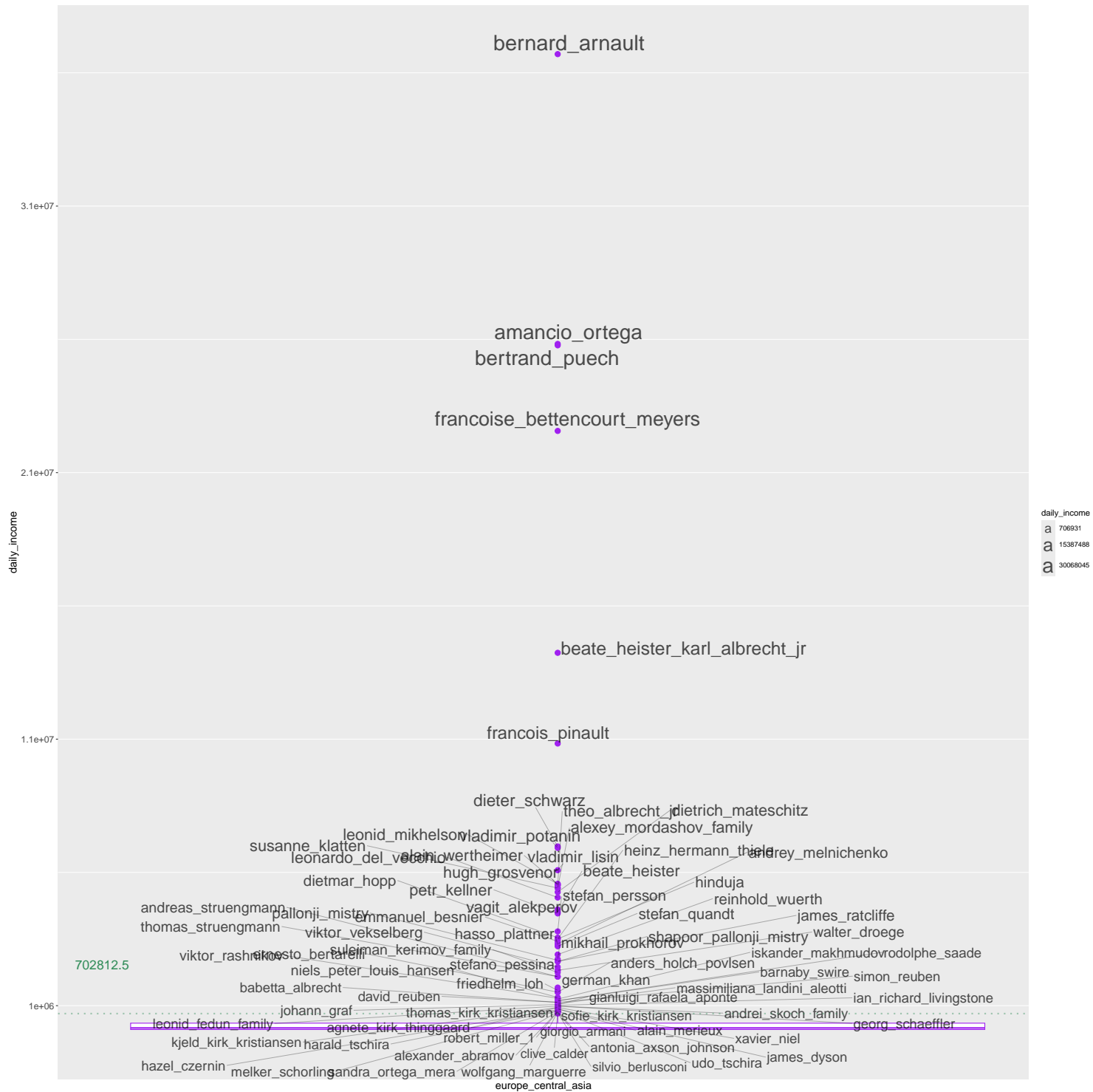
Saving 15.4 x 20 in image



europa\_central\_asia

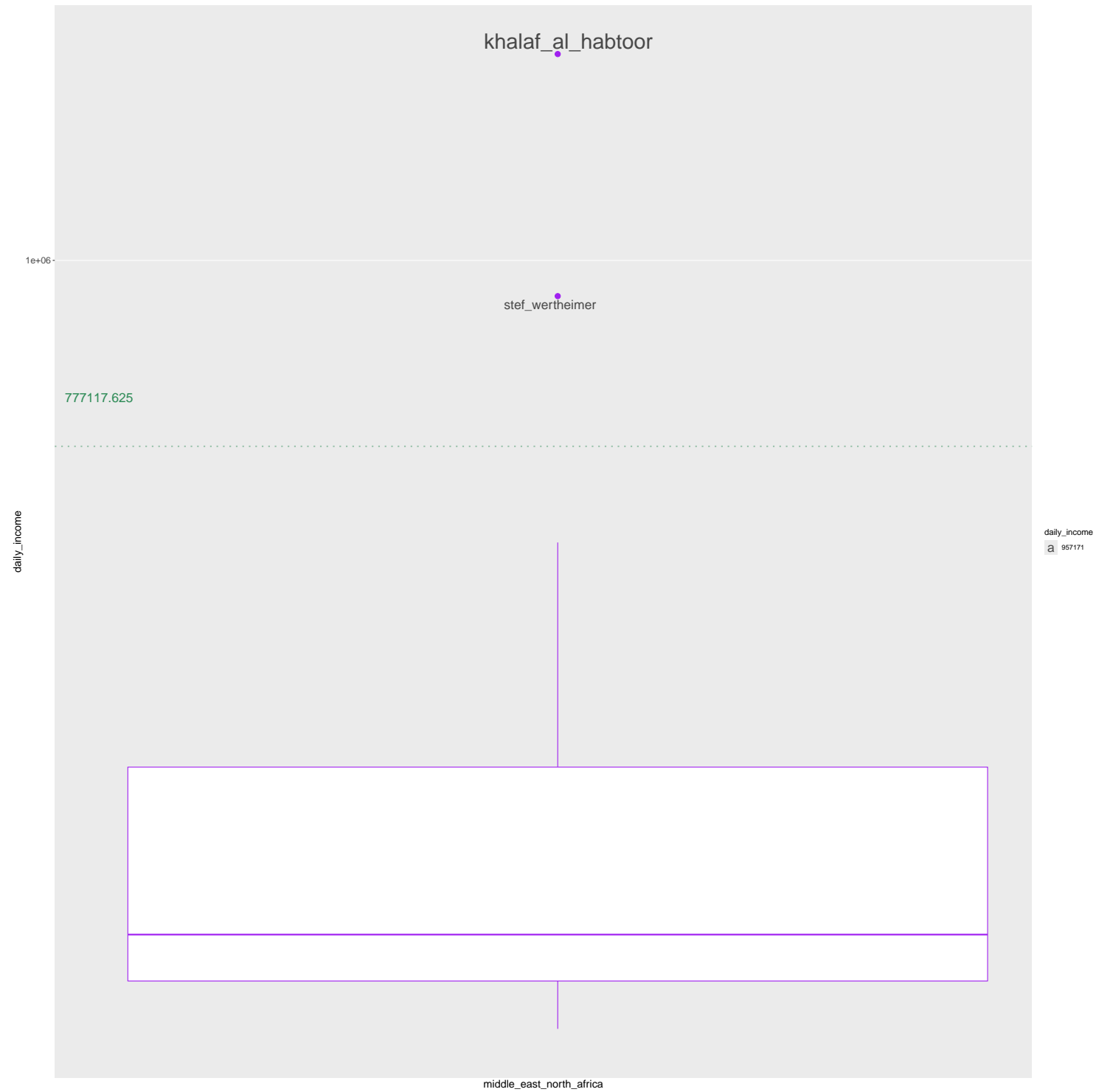
Saving 15.4 x 20 in image





middle\_east\_north\_africa

Saving 15.4 x 20 in image



south\_asia

Saving 15.4 x 20 in image

sub\_saharan\_africa has no outliers. I will plot all names.

Saving 14 x 20 in image

