

Plotting Data

Fundamentals of R - Homework

Yuyu Chen

2022-10-20

R Markdown

October 2022

• All materials for the exercises below are available in the homework folder. • Please submit a PDF document generated from your R Markdown file containing your codes and plots. ## Income of International Organizations The 'io_income_rs_raw.csv' dataset contains funding data from some international organizations in Geneva.

Load libraries

```
library(tidyverse)
```

Please load the dataset, we will be working with it today. Tip: Adding titles, legends, and themes to your plots helps us help you! ### Load data

```
# load data
io_income_rs_raw <- read_csv("io_income_rs.csv")

## Rows: 4500 Columns: 5
## -- Column specification -----
## Delimiter: ","
## chr (3): donor, type_donor, issue_area
## dbl (2): year, amount_nominal
##
## 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.
```

Missing data

From the results above, we can see that there missing values in all the columns. This means we have to be care in removing these rows. To remove the rows with missing values, we take the following steps: 1. Check for rows with no donor values

2. Remove rows with no donor values

```
# remove rows where donor is NA i.e. donors cannot be identified
io_income_rs_cleaned <- io_income_rs_raw %>%
  filter(!is.na(io_income_rs_raw$donor))
```

3. Check to know how many rows are left with missing rows
4. From the result above, we want to retain these 2 rows by replacing the missing values for the `type_donor` variable with **public** and **private** for **South Korea** and **23Technology Limited, United Kingdom** respectively.
5. Update `io_income_rs_cleaned` dataset, with **only** those with no missing values.

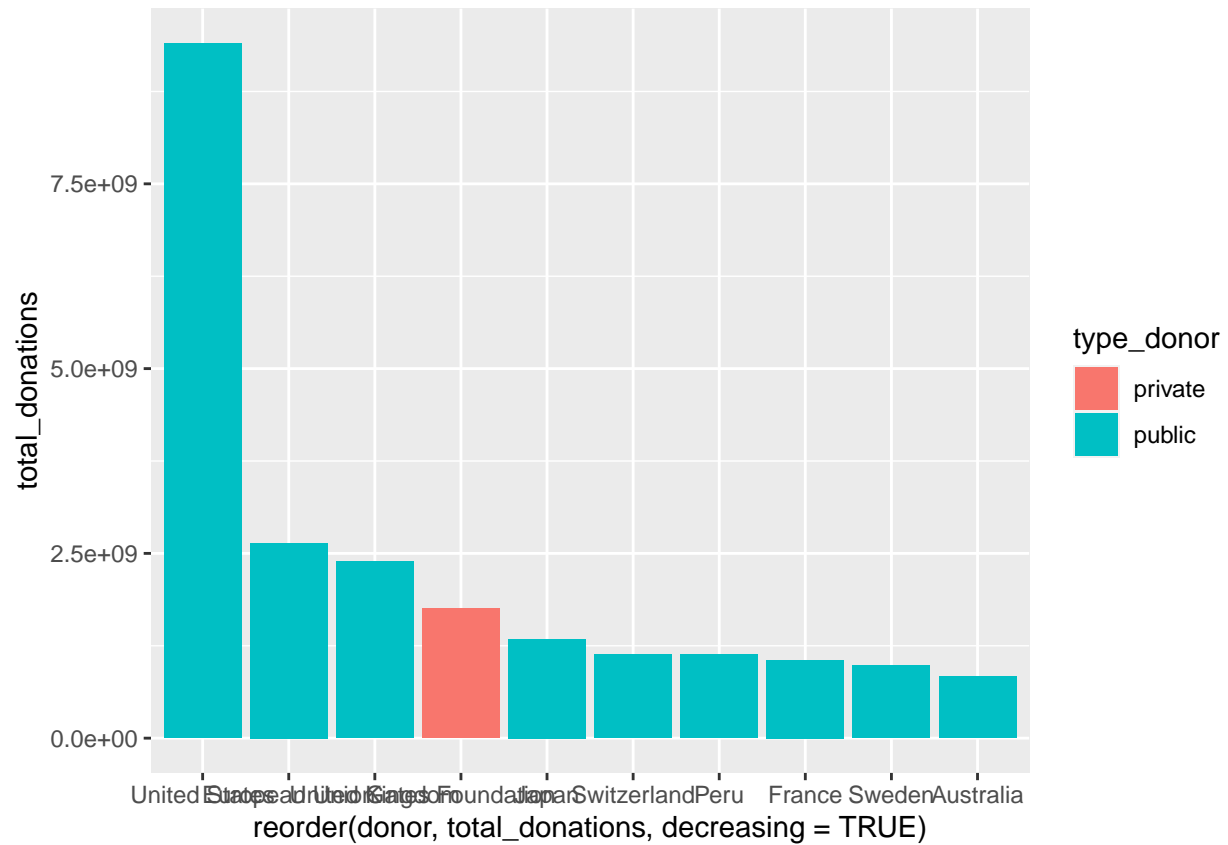
Question 1

Create a bar plot displaying the top 10 overall donors and their total donations to all international organizations. Please color the donors by their type (i.e. public or private).

```
# group data by donor and type_donor
top10_total_donations_by_donor <- io_income_rs_cleaned %>%
  select(donor, type_donor, amount_nominal) %>%
  group_by(donor, type_donor) %>%
  summarise(total_donations = sum(amount_nominal)) %>%
  ungroup() %>%
  top_n(10)
```

```
## 'summarise()' has grouped output by 'donor'. You can override using the
## '.groups' argument.
## Selecting by total_donations
```

```
ggplot(data = top10_total_donations_by_donor, aes(
  reorder(donor, total_donations, decreasing = TRUE),
  total_donations,
  fill = type_donor
)) +
  geom_bar(stat = "identity")
```

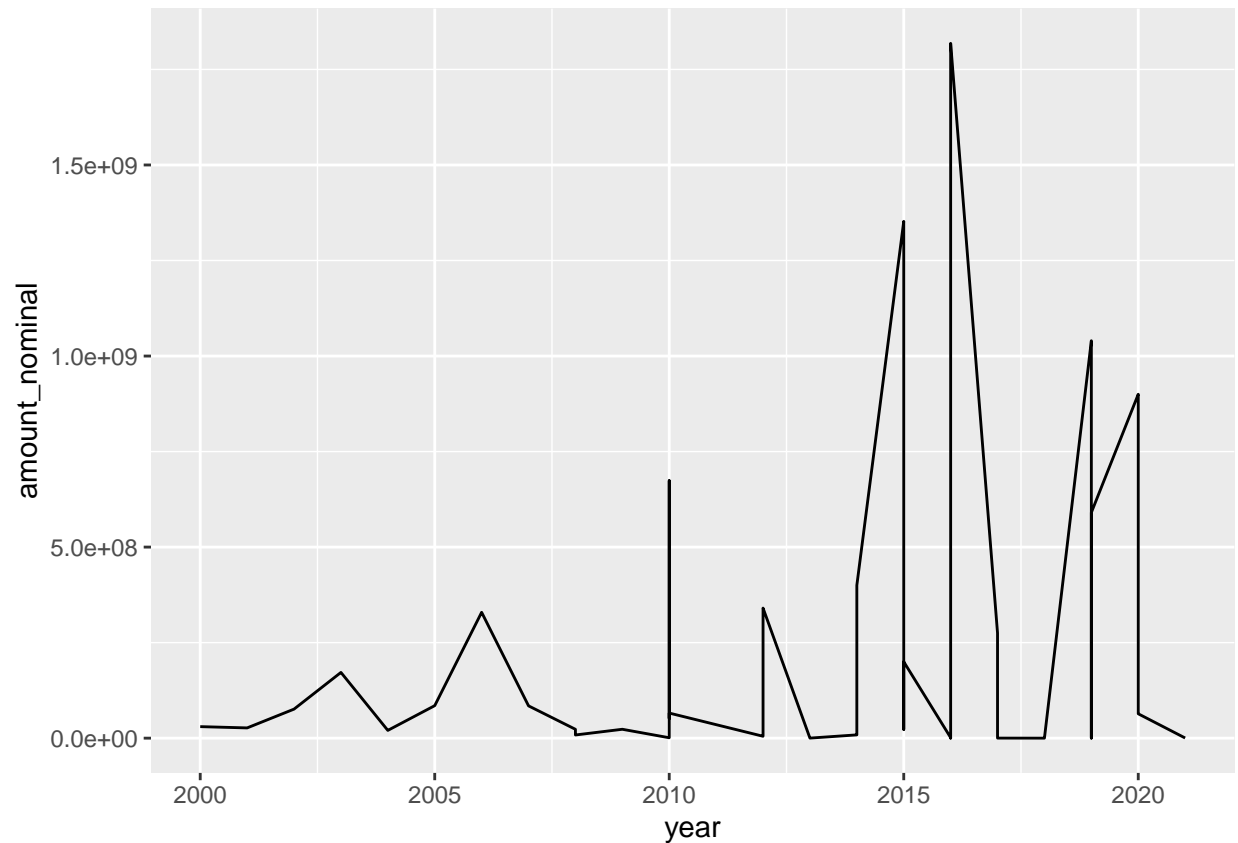


Question 2

Has the amount the US donates to all issue areas increased over time? Please illustrate this relationship in a line plot, colored by the respective issue areas.

```
data_barplot <- io_income_rs_cleaned %>%
  filter(donor == "United States")

data_barplot %>%
  ggplot(aes(x=year, y=amount_nominal))+
  geom_line()
```



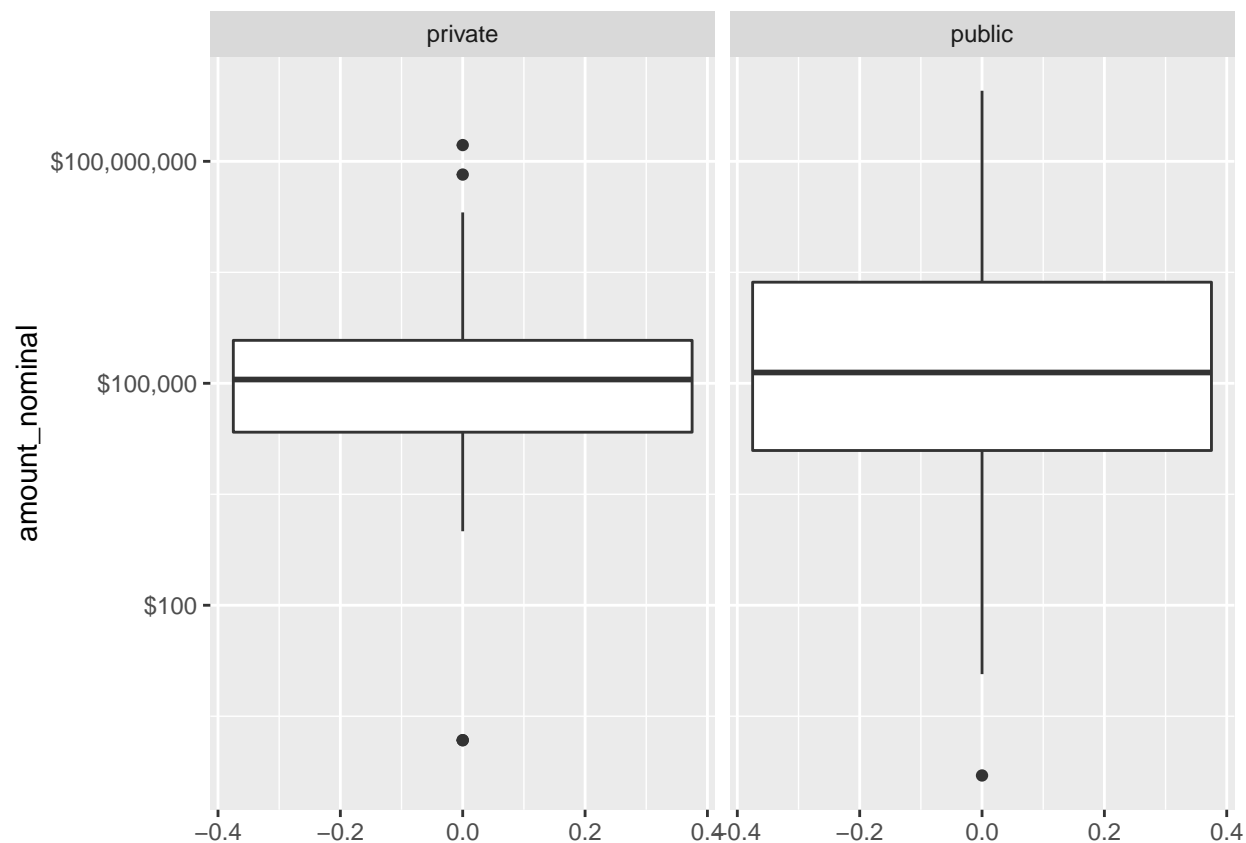
Question 3

Plot the distribution of all donations in the year 2000 and in the year 2020 comparing public and private donors.

(Tip: box plots are great for distributions! Please treat year as factor and scale the nominal amount using “scale_y_log10(labels = scales::dollar)”).

```
#filter dataset for year in 2000 and 2020
data_boxplot <- io_income_rs_cleaned %>%
  filter(year %in% c(2000, 2020))

data_boxplot %>%
  ggplot(aes(y=amount_nominal)) +
  geom_boxplot() +
  scale_y_log10(labels = scales::dollar)+
  facet_wrap( ~ type_donor)
```



Are there outliers for either of these types of donors in 2000? Yes, there outliers

What about in 2020? Yes. Only one outlier

Question 4

Create four scatter plots, one for each issue area, containing all donation by year per donor type (Tip: use “`facet_wrap(~issue_area)`”). The shape of the points should reflect the donor type (Tip: use “`geom_point(aes(shape = donor_type))`”). In each facet, please add a smoothed line to show the direction of the relationship (TIP: use “`geom_smooth()`”). The smoothed lines should be colored “red”.

1.

```
data_sp <- io_income_rs_cleaned %>%
  group_by(year, type_donor, issue_area) %>%
  summarise(donations = sum(amount_nominal)) %>%
  ungroup()
```

‘`summarise()`’ has grouped output by ‘year’, ‘type_donor’. You can override
using the ‘.groups’ argument.

```
data_sp %>%
  ggplot(aes(y = donations, x = year)) +
  geom_point(aes(shape = type_donor)) +
  geom_smooth(color="red")+
  
```

```
scale_y_log10(labels = scales::dollar)+
facet_wrap( ~ issue_area)
```

```
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```

