

Block 3: Homework Solutions

Henrique Sposito and Livio Silva-Muller

October 2022

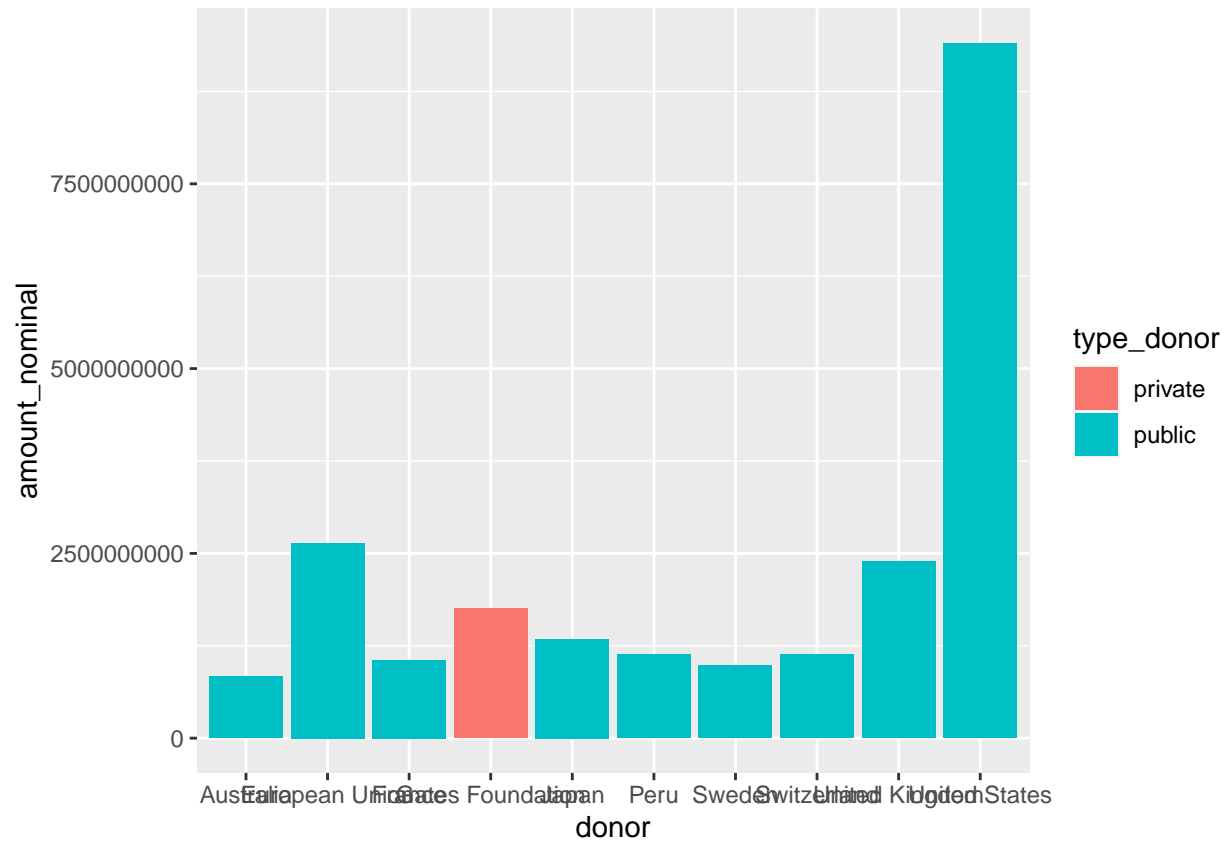
Question 1

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 (e.g. public or private).

```
q1 <- io_income_rs %>%
  drop_na(donor)%>%
  group_by(donor, type_donor)%>%
  summarize(amount_nominal = sum(amount_nominal, na.rm = TRUE)) %>%
  arrange(desc(amount_nominal))

q1 <- q1[1:10,]

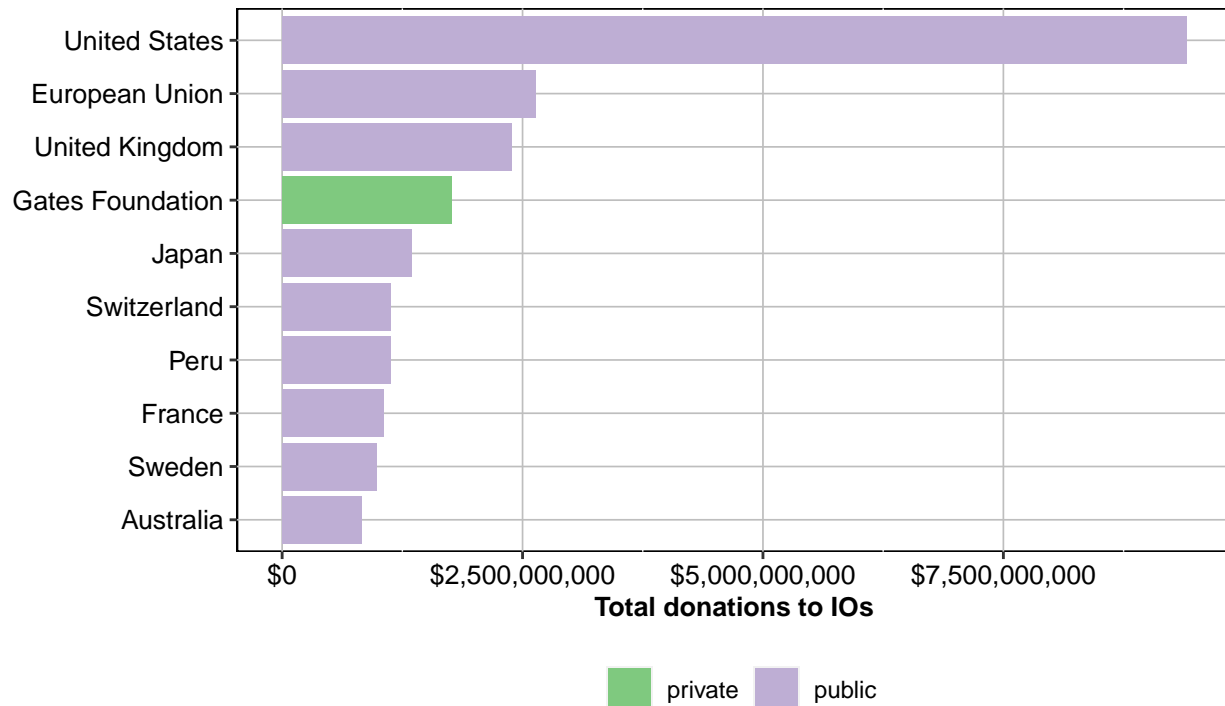
q1 %>%
  ggplot(mapping=aes(x=donor, y= amount_nominal, fill=type_donor))+
  geom_col() #minimal solution
```



```
q1 %>%
  ggplot(mapping=aes(x=reorder(donor, amount_nominal), y= amount_nominal, fill=type_donor))+
  geom_col()+
  scale_fill_brewer(type="qual", palette= "Accent", direction=1)+
  scale_y_continuous(labels=scales::dollar)+
  labs( x = NULL, y = "Total donations to IOs", title = "Top 10 donors from 2000 to 2020", #removing re
        subtitle=" composed by 4500 randomly sampled observations from dataset",
        caption = "Source: io_income_rs dataset")+
  theme(panel.background = element_rect("white", "black", .5, "solid"),
        panel.grid.major = element_line(color = "grey", size = 0.3, linetype = "solid"),
        axis.text = element_text(color = "black", size = 10),
        title = element_text(color = "black", size = 10, face = "bold"),
        legend.title = element_blank(),
        plot.subtitle = element_text(color = "black", size = 9, face = "plain"),
        legend.position = "bottom")+
  coord_flip() #flipping it to simplify reading
```

Top 10 donors from 2000 to 2020

composed by 4500 randomly sampled observations from dataset



Source: io_income_rs dataset

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.

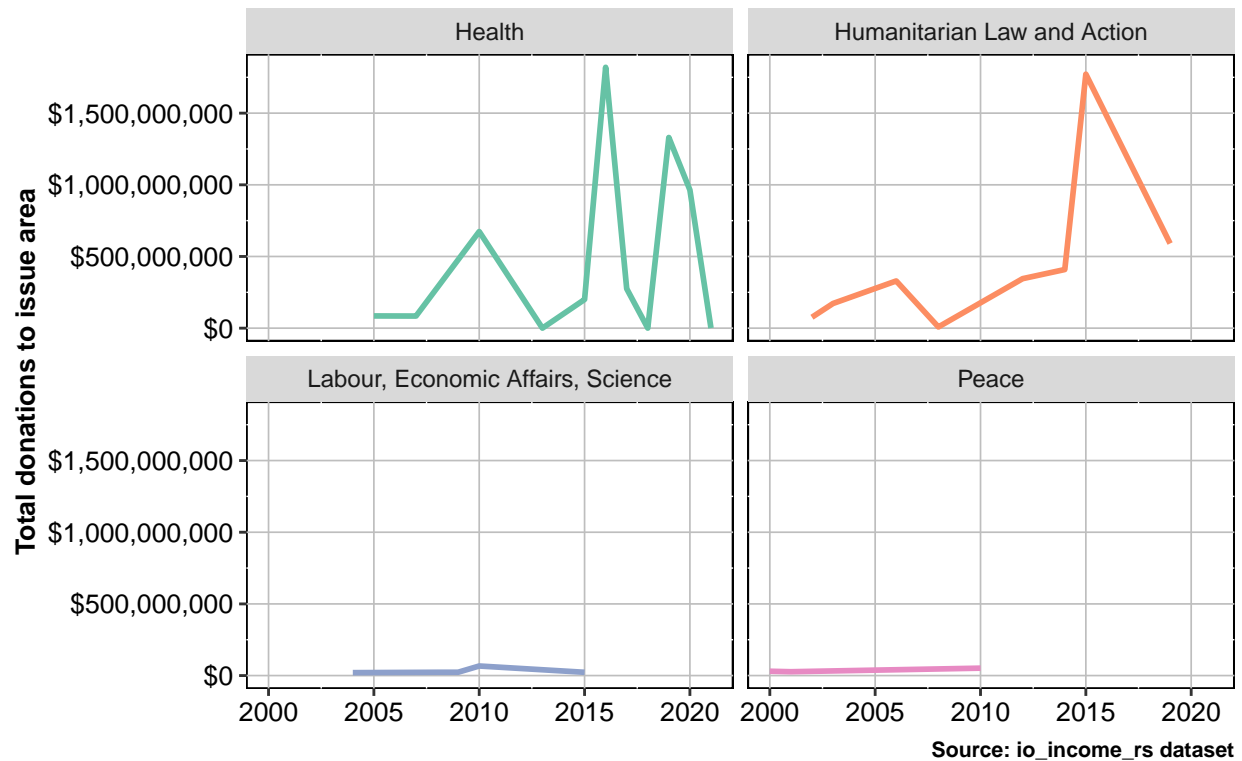
```
io_income_rs %>%
  filter(donor== "United States") %>%
  group_by(year, issue_area)%>%
  summarize(amount_nominal = sum(amount_nominal, na.rm = TRUE)) %>%
  ggplot(mapping=aes(x=year, y=amount_nominal, color=issue_area))+
  geom_line() #minimal solution
```



```
io_income_rs %>%
  filter(donor== "United States") %>%
  group_by(year, issue_area)%>%
  summarize(amount_nominal = sum(amount_nominal, na.rm = TRUE)) %>%
  ggplot(mapping=aes(x=year, y=amount_nominal, color=issue_area))+
  geom_line(size=1)+
  scale_color_brewer(type="qual", palette= "Set2", direction=1)+
  scale_y_continuous(labels=scales::dollar)+
  labs( x = NULL, y = "Total donations to issue area", title = "Donations by the United States since 2000",
        subtitle=" composed by 4500 randomly sampled observations from dataset",
        caption = "Source: io_income_rs dataset")+
  theme(panel.background = element_rect("white", "black", .5, "solid"),
        panel.grid.major = element_line(color = "grey", size = 0.3, linetype = "solid"),
        axis.text = element_text(color = "black", size = 10),
        title = element_text(color = "black", size = 10, face = "bold"),
        legend.title = element_blank(),
        plot.subtitle = element_text(color = "black", size = 9, face = "plain"),
        legend.position = "none")+
  facet_wrap(~issue_area)
```

Donations by the United States since 2000 by issue area

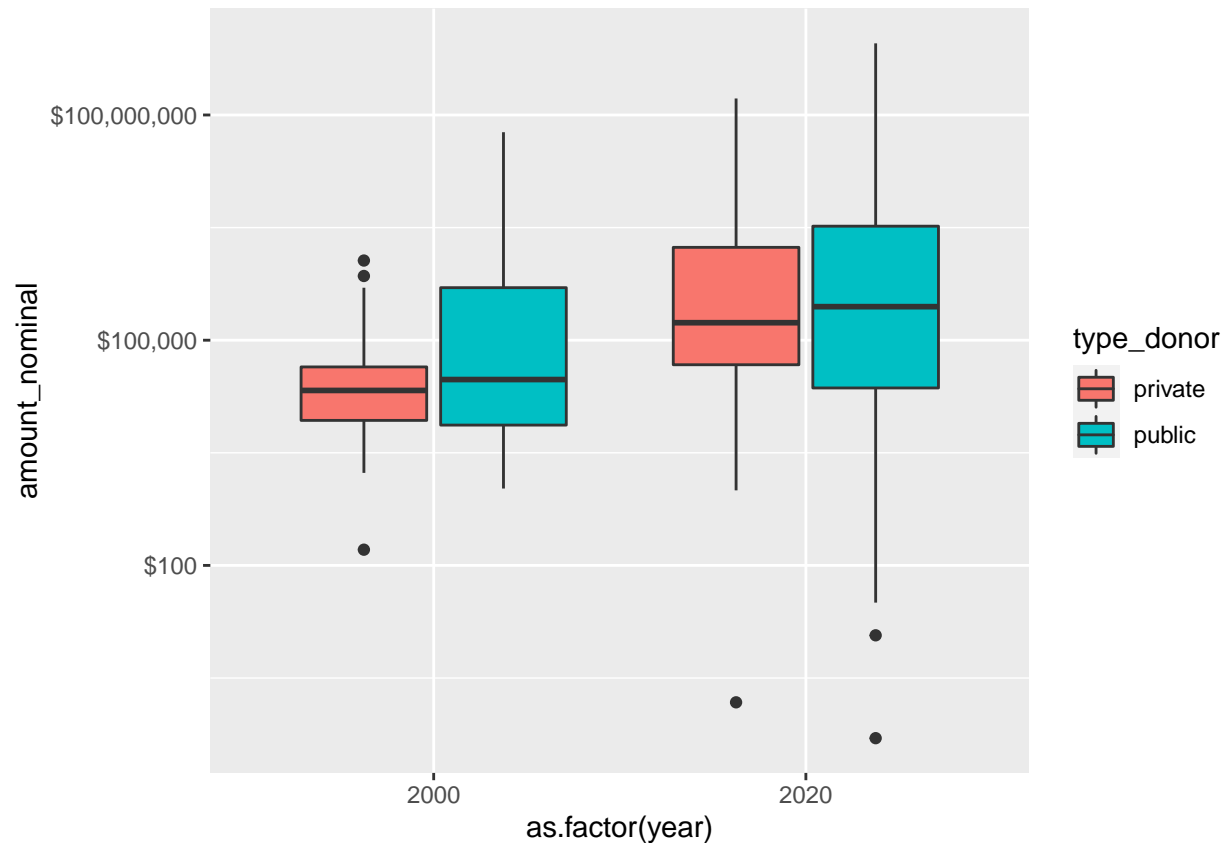
composed by 4500 randomly sampled observations from dataset



Question 3

Plot the distribution of all donations in the year 2000 and in the year 2020 comparing public and private donors. Are there outliers for either of these types of donors in 2000? What about in 2020? (Tip: box plots are great for distributions, you might have to `as.factor(year)` and scale the `amount_nominal` in `log10`)

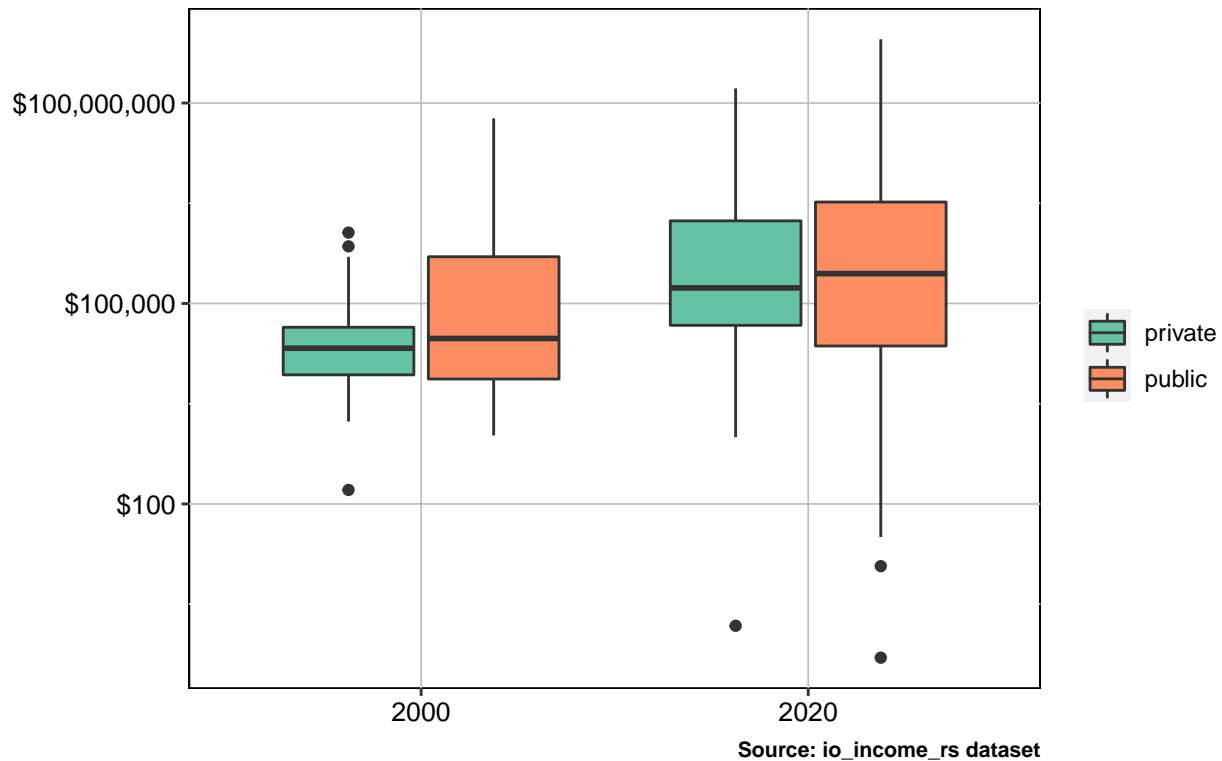
```
io_income_rs %>%
  drop_na(type_donor) %>%
  filter(year==2000 | year == 2020) %>%
  ggplot(mapping=aes(x=as.factor(year), y=amount_nominal, fill=type_donor))+
  geom_boxplot()+
  scale_y_log10(labels=scales::dollar) #minimal solution
```



```
io_income_rs %>%
  drop_na(type_donor) %>%
  filter(year==2000 | year == 2020) %>%
  ggplot(mapping=aes(x=as.factor(year), y=amount_nominal, fill=type_donor))+
  geom_boxplot()+
  scale_y_log10(labels=scales::dollar)+
  scale_fill_brewer(type="qual", palette= "Set2", direction=1)+
  labs( x = NULL, y = NULL, title = "Distribution of donations by donor type in 2000 and 2020", #removing subtitle=" composed by 4500 randomly sampled observations from dataset",
  caption = "Source: io_income_rs dataset")+
  theme(panel.background = element_rect("white", "black", .5, "solid"),
  panel.grid.major = element_line(color = "grey", size = 0.3, linetype = "solid"),
  axis.text = element_text(color = "black", size = 10),
  title = element_text(color = "black", size = 10, face = "bold"),
  legend.title = element_blank(),
  plot.subtitle = element_text(color = "black", size = 9, face = "plain"),
  legend.position = "right")
```

Distribution of donations by donor type in 2000 and 2020

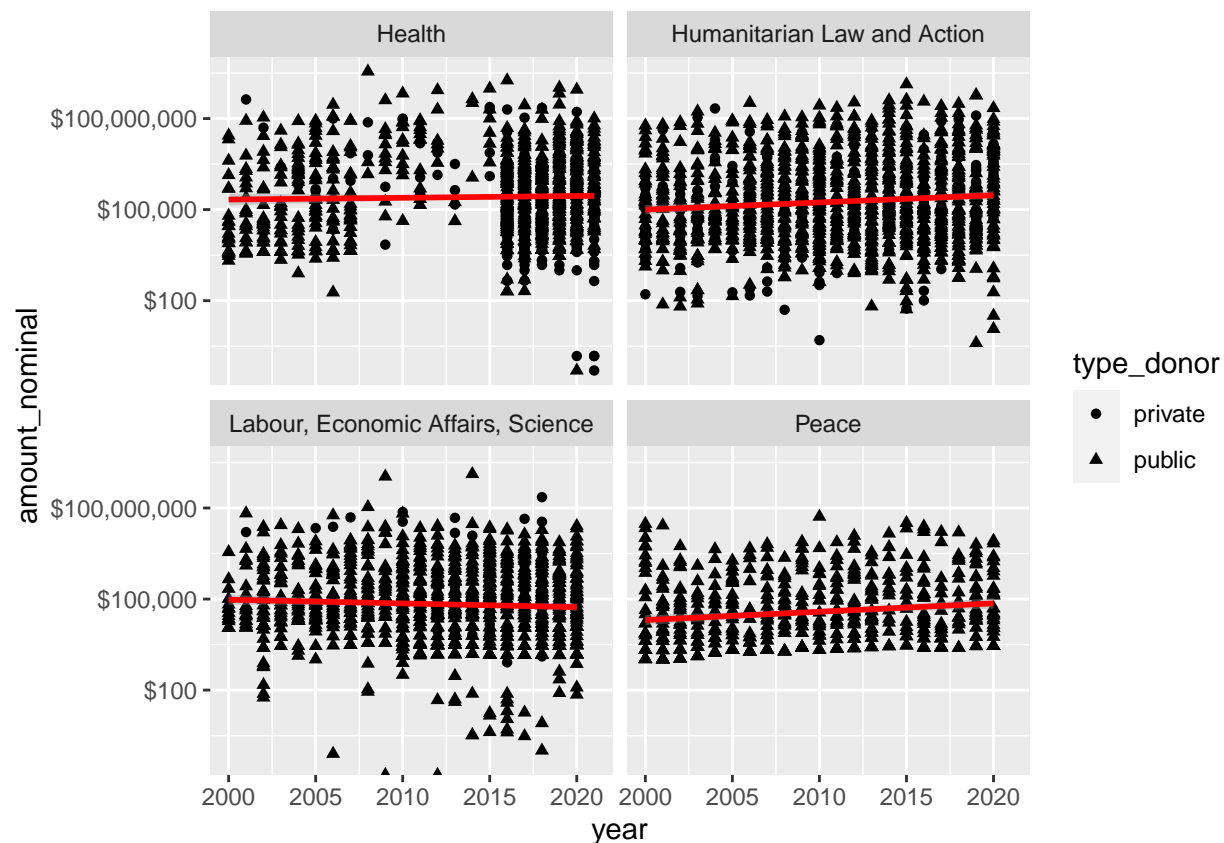
composed by 4500 randomly sampled observations from dataset



Question 4

We want one scatter plot for each issue area containing all donation by year per donor type. The shape of the points should reflect the donor type. In each facet, we want a smoothed line to show the direction of the relationship (TIP: `geom_smooth(method="lm")`). This line should be colored "red".

```
io_income_rs %>%  
  drop_na(type_donor)%>%  
  ggplot(mapping=aes(x=year, y=amount_nominal))+  
  geom_point(aes(shape=type_donor))+  
  geom_smooth(method = "lm", color="red")+  
  scale_y_log10(labels=scales::dollar)+  
  facet_wrap(~issue_area) #minimal
```



```
io_income_rs %>%
  drop_na(type_donor)%>%
  ggplot(mapping=aes(x=year, y=amount_nominal))+
  geom_point(aes(shape=type_donor), position="jitter", alpha=.5)+ #position jitter adds random noise to
  geom_smooth(method = "lm", color="red")+
  scale_y_log10(labels=scales::dollar)+
  facet_wrap(~issue_area)+
  labs( x = NULL, y = NULL, title = "Donations to issues areay by year per type since 2000", #removing
        subtitle=" composed by 4500 randomly sampled observations from dataset",
        caption = "Source: io_income_rs dataset")+
  theme(panel.background = element_rect("white", "black", .5, "solid"),
        panel.grid.major = element_line(color = "grey", size = 0.3, linetype = "solid"),
        axis.text = element_text(color = "black", size = 10),
        title = element_text(color = "black", size = 10, face = "bold"),
        legend.title = element_blank(),
        plot.subtitle = element_text(color = "black", size = 9, face = "plain"),
        legend.position = "right")
```


Donations to issues areay by year per type since 2000

composed by 4500 randomly sampled observations from dataset

