

```

---
title: "FoR Homework"
author: "Krithi Ramaswamy"
date: "20.10.2022"
output:
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <<http://rmarkdown.rstudio.com>>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

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).

```

```{r}
library(dplyr)
library(tidyr)
library(ggplot2)
library(ggpubr)

importing the data
io_income_rs <- read_csv("Desktop/MINT338/Material for week 4-20221014/
io_income_rs.csv")
View(io_income_rs)
data <- io_income_rs

NA has been considered as not applicable by R, changing to character
variable
data$donor= ifelse(is.na(data$donor),"NA",data$donor)

omitting the other NAs
data=na.omit(data)

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).

{r}
grouping by donor and adding the donation amount to get the total, then
arranging it into descending order and slicing out the first 10 data
top10= data %>% group_by(donor, type_donor) %>%

```

```

 summarise(sum_donation=sum(amount_nominal))
#arranging in descending order
top10=top10%>%
 arrange(desc(sum_donation))
taking the top 10
top10= top10[1:10,]

generating the bar plot for top 10 overall donors
ggplot(data=top10, aes(x=donor, y=sum_donation, fill=type_donor),las=2) +
geom_bar(stat="identity", position=position_dodge())+ theme(axis.text.x =
element_text(angle = 90))+ labs(title = "Overall Donations of Top 10
Donors", # add a title
 x = "Donor", # add a label for x axis
 y = "Donation Amount", # add a label for y axis
 fill = "Type of Donor") + # sub legend for fill
 scale_fill_manual(values = c("blue", "grey")) + # manually set colors
 theme_classic()
```

```

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.

```

```{r}
#filtering the US from donors
US= data %>% filter(donor=="United States")
sorting by year
US= US[order(US$year),]
#calculating the mean
US1 = US %>% group_by(year,issue_area) %>%
summarise(mean=mean(amount_nominal))
line plot, colored by the respective issue areas
US1 %>%
 ggplot(aes(x=year, y=mean, group=issue_area, color=issue_area)) +
 geom_line()+ ggtitle("US Donations arranged by Issue Areas") +
 labs(x = "Year", # add
 y = "Amount Donated") + # add caption
 theme_classic()
```

```

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! Please treat year as factor and scale the nominal amount using “scale_y_log10(labels = scales::dollar)”)

```

```{r}
filtering for public and private
box = data %>% filter(type_donor %in% c("public","private"))

filtering for the year 2000
box2000=box %>% filter(year == 2000)

```

```
box plot 1
box_plot_2000=ggplot(box2000, aes(x = type_donor, y = amount_nominal, color
=type_donor)) + geom_boxplot()+scale_y_log10(labels = scales::dollar)+
ggtitle("Box Plot for distribution of all donations in the year 2000")
```

```
filtering for the year 2020
box2020=box %>% filter(year == 2020)
```

```
box plot 2
box_plot_2020=ggplot(box2020, aes(x = type_donor, y = amount_nominal, color
=type_donor)) + geom_boxplot()+scale_y_log10(labels = scales::dollar)+
ggtitle("Box Plot for distribution of all donations in the year 2020")
```

```
joining the 2 box plots
geom_boxplot(box_plot_2000, box_plot_2020,
 labels = c("2000", "2020") +
 ncol = 1, nrow = 2)
```

```
#I am getting an error message here but I can't understand how to fix the
mistake
```

```
```
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”.
```{r}
```

```
ggplot(data, aes(x = year, y = amount_nominal))+
 geom_point(aes(shape = type_donor)) +
 facet_wrap(~issue_area) +
 geom_smooth(colour="red")+ggtitle("Scatter Plot for all donation by year
per donor type") +
 labs(x = "Year", y = "Amount Donated")
```
```