Narrative

Yongzhen Lai

12/2/2020

2016-2020 Hong Kong Legislative Council Sessions

This project would analyze the everyday meeting sessions of the current term (2016-2020) of the legislative institution in the Hong Kong SAR, the Legislative Council (shortened as 'Legco'). It would analyze the Legco members' activities, mainly the frequency of raising motions and amendments, which denotes ordinary initiatives and procedural affairs, and especially decrees with legal effects, the 'bills'. Whether and how members' everyday meeting and billing activities have been related to their party affiliations is the interested issue in this project.

First, load in the library package.

```r
library(tidyverse)
library(rvest)
library(stringr)
library(purrr)
library(lubridate)
```

Preprocessing the Data: Motions

Insert the webpage of each meeting session. Since I find it runs nothing when using scrapping the link of the each meeting records with the function through using selector gadget, I have to paste the link one by one through the temporal order for running the result.

The original attempt: function(U){ page <- real_html(U) link <- html_nodes(page, ".vote a~ a")%>% html_attr("href") dframe <- data_frame("link"=link) return(dframe) }

Scrape the meeting records information from the official website of the 6th Hong Kong Legislative Council.

```r
#Formal start: Set the information scrapping function.
scrape_result<- function(URL){
  page <- read_xml(URL)
  date <- xml_nodes(page, "vote-date")%>%
    xml_text() %>%
    dmy()
  time <- xml_nodes(page, "vote-time")%>%
    xml_text()
  mover <- xml_nodes(page, "mover-en")%>%
    xml_text()
  type <- xml_nodes(page, "mover-type")%>%
    xml_text()
  motion <- xml_nodes(page, "motion-en")%>%
    xml_text()
  vote <- xml_nodes(page, "member vote")%>%
    xml_text()
  mechanism <- xml_nodes(page, "vote-separate-mechanism") %>%
    xml_text()
  result <- xml_nodes(page, "overall result")%>%
    xml_text()
  dframe <- data.frame("date"=date, "time"=time, "mover"=mover, "type"=type, "motion"=motion, "vote"=vo
  return(dframe)

}

#Select the meeting record of 2016 since the beginning of the 6th Legco, I originally scrape the links
#the 2016-2017 term.
result2016_17 <- c("https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20161109.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20161116.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20161123.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20161130.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20161207.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20161214.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170111.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170118.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170208.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170215.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170301.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170322.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170329.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170426.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170517.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170524.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170531.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170607.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170614.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170621.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170628.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170705.xml",
            "https://www.legco.gov.hk/yr16-17/chinese/counmtg/voting/cm_vote_20170712.xml")

#Map into a dataframe.
frame2016_17 <- map_dfr(result2016_17, scrape_result)
```

Paste the link with the meeting results of 2017-2018, 2018-2019, and 2019-2020, then map into a dataframe respectively.

```
result2017_18 <- c("https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20171018.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20171025.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20171101.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20171108.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20171115.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20171129.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20171206.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20171213.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180110.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180117.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180124.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180131.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180207.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180321.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180328.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180411.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180502.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180509.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180516.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180523.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180530.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180606.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180613.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180620.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180627.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180704.xml",
                "https://www.legco.gov.hk/yr17-18/chinese/counmtg/voting/cm_vote_20180711.xml")

frame2017_18 <- map_dfr(result2017_18, scrape_result)
```

2018-2019 data.

```
result2018_19 <- c("https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20181010.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20181024.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20181031.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20181107.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20181114.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20181121.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20181128.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20181205.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20181212.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20190116.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20190123.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20190130.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20190220.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20190320.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20190327.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20190403.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20190508.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20190515.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20190522.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20190529.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20190605.xml",
                "https://www.legco.gov.hk/yr18-19/chinese/counmtg/voting/cm_vote_20190626.xml")

frame2018_19 <- map_dfr(result2018_19, scrape_result)
```

2019-2020 data.

```
result2019_20 <- c("https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20191023.xml",
                "https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20191127.xml",
                "https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20191204.xml",
                "https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20191211.xml",
                "https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20191218.xml",
                "https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20200115.xml",
                "https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20200325a.xml",
                "https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20200429.xml",
                "https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20200506.xml",
                "https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20200513.xml",
                "https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20200520.xml",
                "https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20200527.xml",
                "https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20200603.xml",
                "https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20200610.xml",
                "https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20200617.xml",
                "https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20200708.xml",
                "https://www.legco.gov.hk/yr19-20/chinese/counmtg/voting/cm_vote_20200715.xml")

frame2019_20 <- map_dfr(result2019_20, scrape_result)
```

Combine the dataframe together as the ‘raw’ dataframe of the 6th Legco.

```
#Combine the data frame.
frame_all <- rbind(frame2016_17, frame2017_18, frame2018_19, frame2019_20)

#Edit the names of Legco members.
#Extract the prefix of names.
frame_all$mover_edited <- sub("^(Dr|Mrs)", "", frame_all$mover)
#Set capitalized letter in first name and last name.
frame_all$mover_edited <- str_to_title(frame_all$mover_edited)
#trim the names orderly.
frame_all$mover_edited <- str_trim(frame_all$mover_edited, side = "left")

#arrange the data.
frame_all <- frame_all %>%
  group_by(motion)%>%
  arrange(date, time)
```

Build a dataframe of the voting record.

```
#Build a dataframe of voting result.
voting_record <- frame_all %>%
  group_by(date, time, motion) %>%
  count(vote) %>%
  rename("results"="n") %>%
  spread(vote, results)

#Set NA as the 0 of voting result.
voting_record[is.na(voting_record)] = 0

#Set the attendance rate and the approval rate of each bill.
voting_record <- voting_record %>%
  mutate(total = sum(Absent, Abstain, No, Present, Yes),
         attendance_rate = 1 - (Absent/total),
         attendance = sum(Abstain, No, Present, Yes),
         approval_rate = Yes/attendance)


#Round the percentage of rate figures.
voting_record$attendance_rate <- round(voting_record$attendance_rate, digits = 2)
voting_record$approval_rate <- round(voting_record$approval_rate, digits = 2)
voting_record%>%
  arrange(date, time)
```

Preprocessing the Data: Member Information

Here is the scraping link of the current members, but I don't think it is necessary to use it now.

```
scrape_infolink <- function(URL){
  page <- read_html(URL)
  namelink <- html_nodes(page, ".bio-member-detail-1 a") %>%
    html_attr("href")
  infopage <- paste0("https://www.legco.gov.hk/general/english/members/yr16-20/", namelink)
  vector_links <- as.vector(infopage)
  return(vector_links)
}

links <- scrape_infolink("https://www.legco.gov.hk/general/english/members/yr16-20/biographies.htm")
```

We can use scraping the link from wiki webpage to gather all information. I would choose the information of members' 'parties'(officially defined as 'political societies' in the local regulations) for analysis. As I assumed that HK legislative politics is mostly polarized through different party affiliations instead of other individualized background like different professionals when it comes to bill procedure and voting.

```
#Set the function for scrapping the names.
scrape_name <- function(U2){
  page <- read_html(U2)
  name <- html_nodes(page, ".fn a") %>%
    html_text()
  info <- data.frame("name"= name)
  return(info)
}
#Scrape the names of 6th Legco members.
member <- scrape_name("https://en.wikipedia.org/wiki/6th_Legislative_Council_of_Hong_Kong")

#Since there is difference of the list between names and party, as some Legco individual members have j
scrape_party <- function(U2){
  page <- read_html(U2)
  party <- html_nodes(page, "td:nth-child(6) a") %>%
    html_text()
  info <- data.frame("party"= party)
  return(info)
}
#Scrape the party.
party <- scrape_party("https://en.wikipedia.org/wiki/6th_Legislative_Council_of_Hong_Kong")
#Drop out the minor party information and list it again.
party <- unlist(party)
party <- party[-c(11, 14, 19, 30, 32, 38)]
party <- as.vector(party)
party <- data.frame(party)
#Combine the column of member and party information.
members <- cbind(member, party)
#Capitalized the first letter of names for the following process of joining the party information with
capitalized_names <- members %>%
  select(name) %>%
  unlist() %>%
  str_to_title() %>%
  data.frame(Capitalized)
#Add original name with the edited capitalized names.
members <- members %>%
  add_column(Capitalized)
```

Although Legco members had different party member status while some member were ‘independent’ status without formally joining any party, they are, unofficially but normally, largely categorized into two affiliation camps ideologically in HK local politics among media coverage or by information claimed by members themselves. I would use “Government” and “Opposition” and one claimed “Centrist” in denoting their camp affiliations relative to HKSAR governmental institutions through open source of party information, while names in the media coverage using to denote the camp identity are somewhat ideologically ‘controversial’.

Besides, although there would be some claimed affiliations in recent years, which can still largely categorized into the binary differentiation which has shown in their meeting and voting behavior.

```
#Denote the affiliation camps.
members$affiliation = ifelse(members$party %in% c("BPA", "Liberal", "DAB", "FTU", "NPP", "New Forum","F
                             ifelse(members$party %in% c("Prof Commons", "Civic", "LSD", "Democratic", 
                                    ifelse(members$Capitalized %in% c("Joseph Lee", "Eddie Chu", "Shiu 
                                           ifelse(members$Capitalized %in% c("Chan Kin-Por", "Paul Tse"
                                                  "Centrist"))))

#Revise certain name information to fit in the meeting record dataframe, as some members have both Engl
members$Capitalized[45]="Chu Hoi-Dick"
members$Capitalized[41]="Ir Dr Lo Wai-Kwok"
members$Capitalized[24]="Charles Peter Mok"
```

```
#Group by party for a quick check.
members <- members %>%
  group_by(affiliation, party) %>%
  arrange (party)
```

Data Cleaning and Processing

Here is to refine the dataframe based on the previous two dataframes, the meeting record dataframe and the members’ dataframe.

Select the information from the raw frame to construct a new dataframe called “frame_motion”.

```
frame_motion <- frame_all %>%
#Select all variables without "vote".
  select(date, time, mover_edited, type, motion, separate_mechanism, voting_result) %>%
#Distinct it.
  distinct() %>%
  rename("mover"="mover_edited")%>%
  #factorize the character.
  mutate_if(is.character, as.factor) %>%
  arrange(date, time) %>%
  #Mutate the year.
  mutate(year=year(date))
```

Join the member information into the motion frame, called it “frame_edited”.

```
frame_edited <- frame_motion %>%
  left_join(members, by =c("mover"="Capitalized")) %>%
  select(-name) %>%
  arrange(date) %>%
  mutate(year=year(date))

#Rename the NA information, as NA without party affiliations denotes that motions raised by the executi
```

```
frame_edited[is.na(frame_edited)] <- "Official"
#Factorize again.
frame_edited <- frame_edited %>%
  mutate_if(is.character, as.factor)
```

Categorize the motion type.

```
#Ordinary motions, including proposals without legal bindings.
motion <- data.frame(motion = str_which(frame_edited$motion, "^MOTION|^\"MOTION|^FIRST|^SECOND|^THIRD|A

#Ordinary amendments.
amendment <- data.frame(motion = str_which(frame_edited$motion, "^AMENDMENT"), motion_type = rep("amend

#Bills and related proposed amendments, with legal bindings after authorization through the Legco.
bill <- data.frame(motion = str_which(frame_edited$motion, "BILL"), motion_type = rep("bill", times=leng

#resolutions.
resolution <- data.frame(motion = str_which(frame_edited$motion, "^PROPOSED"), motion_type = rep("resolu

#Combine together.
motions <- rbind(motion, amendment, bill, resolution)

#Set the id number for each motion in the edited frame, in order for joining the categorization into th
frame_edited$id <- seq.int(nrow(frame_edited))

#Join the frame.
frame_edited <- frame_edited %>%
  left_join(motions, by =c("id"= "motion"))
#Arrange the frame by motion type and find out the NA.
frame_edited <- frame_edited %>%
  arrange(desc(motion_type))
#Rename the NAs.
frame_edited$motion_type[c(958,959,960,961)]= "resolution"
frame_edited$motion_type[c(964,965,966,967)]= "motion"
frame_edited$motion_type[c(962,963)]= "others"
#Factorize again.
frame_edited$motion_type <- as.factor(frame_edited$motion_type)
#Count for checking.
frame_edited %>%
  arrange(motion_type)
```

Check whether there are duplicated categorizations, and it is reasonable that some motions could related to bills, which could be still for using analysis.

```
frame_edited %>%
  select (motion, motion_type, id) %>%
  count (id, motion) %>%
  mutate(duplicate = ifelse(n>1, TRUE, FALSE)) %>%
  filter(duplicate == TRUE)
```

Join the voting record into the edited frame.

```
frame_motion <- frame_motion %>%
  left_join(voting_record, by =c("date", "time", "motion"))
#Arrange it.
frame_motion %>%
  arrange(date, time)
```

Filter out the bills, as the vital message and largest share of categories in the Legco.

```
bills <- frame_edited %>%
  filter(motion_type == "bill") %>%
  mutate(subtitle = str_extract(motion, "^[^-]+")) %>%
  arrange(date)
```

Data Analysis

Vital Step! For the analysis of the data information, and we could have a look on what Legco members have done in the last five years!

First, we could plot the information of motion frequencies from different parties and affiliations.

```
#As the motion frame doesn't include duplicate information of motions
frame_motion <- frame_motion %>%
  left_join(members, by=c("mover"="Capitalized"))

frame_motion[is.na(frame_motion)] <- "Official"

frame_motion %>%
  group_by(year) %>%
  count(affiliation)

#Group the motion frame by counting.
motion_record <- frame_motion %>%
  select(year, affiliation, party, mover) %>%
  group_by(year) %>%
  add_count(affiliation, name="freq_affi_motions")%>%
  group_by(year, affiliation) %>%
  add_count(mover, name = "motions") %>%
  add_count(party, name = "freq_party_motions") %>%
  arrange(year)
#In order to avoid messiness of data, I make a sum data of party unit and affiliation unit respectively
party_record <- motion_record %>%
  select(year, affiliation, party, freq_party_motions)%>%
  arrange(year, desc(freq_party_motions))

#Affiliation record.
affi_record <- motion_record %>%
  select(year, affiliation, freq_affi_motions)%>%
  arrange(year, desc(freq_affi_motions))

#Plot the motion record in terms of parties(including the executive sector).
party_freq <- ggplot(party_record, aes(x=party, y=freq_party_motions, fill=affiliation))+
  geom_col(position="dodge")+
  labs(title = "Motion Frequency of the 6th Hong Kong Legislative Council(By Parties)", x="year", y="fre
  scale_fill_manual("Affiliations", values=c("Centrist"="green", "Government"="blue", "Opposition"="orar
  coord_flip()+
  facet_grid(rows = vars(year), scales = "free_y",
             space = "free_y")

ggsave("party_freq.png", plot = party_freq)
```

```
#Plot the motion record in terms of affiliations(including the executive sector).
affil_freq <- ggplot(affi_record, aes(x=year, y=freq_affi_motions, fill=affiliation))+
  geom_col(position="dodge")+
  coord_cartesian()+
  labs(title = "Motion Frequency of the 6th Hong Kong Legislative Council(By affiliation)", x="year", y
  scale_fill_manual("Affiliations", values=c("Centrist"="green", "Government"="blue", "Opposition"="ora

ggsave("affiliation_freq.png", plot = affil_freq)
```

Plot the motion type record, and we find that motions related to bills have been the vital part in the 6th Legco meeting.

```
mo_type_record <- frame_edited %>%
  select(year, affiliation, party, motion_type) %>%
  group_by(year, affiliation) %>%
  add_count(motion_type, name="affi_motion_type")%>%
  arrange(year)

mo_type <- ggplot(mo_type_record, aes(x=motion_type, y=affi_motion_type, fill=affiliation))+
  geom_col(position="dodge")+
  labs(title = "Motion type of the 6th Hong Kong Legislative Council(By Parties)", x="year", y="frequen
  scale_fill_manual("Affiliations", values=c("Centrist"="green", "Government"="blue", "Opposition"="ora
  coord_flip()+
  facet_grid(rows = vars(year), scales = "free_y",
           space = "free_y")

ggsave("motion_type.png", mo_type)
```

Plot the frequency of raising motions, amendments related to bills. Categorize the bills and we could find some patterned things inside it.

```
#Count the frequency of bill in different issues.
bills <- bills %>%
   group_by(subtitle) %>%
   add_count(name = "bill_frequency") %>%
   group_by(subtitle) %>%
   add_count(affiliation, name = "affi_frequency")


#Pick out bills that have been experienced with more than 5 motions for plotting.
bill_freq <- bills %>%
  filter(bill_frequency > 5) %>%
  ggplot (aes(x= subtitle, y= affi_frequency, fill=affiliation))+
  geom_col(position = "dodge")+
  coord_fixed(ratio=10)+
  theme(axis.text = element_text(size = 3),
        aspect.ratio = 3/10,
        legend.text = element_text(size=3),
        legend.title = element_text(size=3),
        axis.title = element_text(size=4),
        plot.title = element_text(size=6),
        legend.key.size = unit(0.3, "cm"))+
  labs(title="Amendments Regarding Bill Debating", x="Bills on the 6th Legislative Council", y="Frequen
  scale_fill_manual("Affiliations", values=c("Centrist"="green", "Government"="blue", "Opposition"="ora
  coord_flip()
bill_freq

ggsave("frequency of amendments of bills.jpeg", bill_freq)
```

Now we can find out the relation between attendance rate and approval rate of passing a bill or related amendments, without concerning separate_mechanism. I would not save some pictures due to image problems that some names of bills needed to change in order to fit the space of the facet. We could read it in the Rnotebooks

```
#Combine the voting record frame with the bill frame.
bills %>%
  left_join(voting_record, by=c("date", "time", "motion"))


#Pick out bills experienced with more than 15 motions. What type of bills that have been discussed and
bill_anal_1 <- bills %>%
  filter(bill_frequency > 15) %>%
  ggplot (aes(x= attendance_rate, y= approval_rate, color=affiliation, size = separate_mechanism, shape
  geom_point()+
  coord_fixed(ratio=2/3)+
  scale_color_manual("Affiliation", values = c("Centrist"="green", "Government"="blue", "Opposition"="o
  scale_size_manual("Separate Mechanism", values = c("No" = 1, "Yes"= 2))+
  scale_shape_manual("Voting Result", values = c("Negatived"=17, "Passed"=19))+
  theme(strip.text = element_text(size = 5))+
  labs(title="Attendance and Approval of Bills and Amendments", x="Attendance Rate", y="Approval Rate")
  facet_wrap(~subtitle)

ggsave("Bill Passing-1.png", bill_anal_1)

#Pick out bills experienced with less than 15 but more than 5 motions.
bill_anal_2 <- bills %>%
  filter(bill_frequency < 15, bill_frequency > 5) %>%
  ggplot (aes(x= attendance_rate, y= approval_rate, color=affiliation, shape = voting_result))+
  geom_point(size = 2)+
  scale_color_manual("Affiliation", values = c("Centrist"="green", "Government"="blue", "Opposition"="o
  scale_shape_manual("Voting Result", values = c("Negatived"=17, "Passed"=19))+
  theme(strip.background = element_blank(),
        strip.text = element_text(size=3),
        aspect.ratio = 2/3)+
  labs(title="Attendance and Approval of Bills and Amendments", x="Attendance Rate", y="Approval Rate")
  facet_wrap(~subtitle)

ggsave("Bill Passing-2.png", bill_anal_2)

#Analyzing the relation of attendance rate and the approval rate, and we can see some patterns.
bill_anal_3 <- bills %>%
  filter(bill_frequency>0) %>%
  ggplot (aes(x= attendance_rate, y= approval_rate, color=affiliation, size = separate_mechanism, shape
  geom_point()+
  coord_fixed(ratio=1/2)+
  geom_smooth(method = lm)+
  scale_color_manual("Affiliation", values = c("Centrist"="green", "Government"="blue", "Opposition"="o
  scale_size_manual("Separate Mechanism", values = c("No" = 1, "Yes"=2))+
  scale_shape_manual("Voting Result", values = c("Negatived"=17, "Passed"=19))+
  labs(title="Attendance and Approval of Bills and Related Amendments", x="Attendance Rate", y="Approval

ggsave("relation between approval rate and attendance rate of passing bills.png", bill_anal_3)
```

```
#Analyze the relation in terms of different years.
bill_anal_4 <- bills %>%
  filter(bill_frequency>0) %>%
  ggplot(aes(x= attendance_rate, y= approval_rate, color=affiliation, shape = voting_result))+
  geom_point()+
  labs(title="Attendance and Approval of Bills and Related Amendments (By year)", x="Attendance Rate", y
  scale_color_manual("Affiliation", values = c("Centrist"="green", "Government"="blue", "Opposition"="or
  scale_shape_manual("Voting Result", values = c("Negatived"=17, "Passed"=19))+
  geom_smooth(method= lm)+
  facet_wrap(~year)

ggsave("billanalysis4.png", bill_anal_4)
```

How about the relations regarding ordinary motions (without concerning separate mechanism: it is a special voting calculation procedure: voting results are calculated and aggregated within members of direct election and indirect election separately, and the approval status only happened if the "yes" votes are more than "no" in both aggregations instead of directly total aggregation. In this way, some motions appeared high approval rate might failed to be passed are subjected to the separate mechanism)?

```
#Join the edited frame with the voting result.
frame_edited <- frame_edited %>%
  left_join(voting_record, by=c("date", "time", "motion")) %>%
  arrange(date, time)

#Plot the attendance and approval rate of all motions in each year.
motion_anal_1 <- frame_edited %>%
  ggplot(aes(x= attendance_rate, y= approval_rate, color=affiliation, shape = voting_result))+
  geom_point()+
  coord_fixed(ratio=1/2.4)+
  labs(title="Attendance and Approval of All Motions (By year)", x="Attendance Rate", y="Approval Rate")
  scale_color_manual("Affiliation", values = c("Centrist"="green", "Government"="blue", "Opposition"="or
  scale_shape_manual("Voting Result", values = c("Negatived"=17, "Passed"=19))+
  facet_wrap(~year)

ggsave("motions by year.png", motion_anal_1)

#Plot motions except the bill.
motion_anal_2 <- frame_edited %>%
  filter(motion_type == "amendment" | motion_type == "motion" | motion_type =="resolution" | motion_type
  ggplot(aes(x= attendance_rate, y= approval_rate, color=affiliation, size = separate_mechanism, shape =
  geom_point()+
  coord_fixed(ratio=1/2.4)+
  scale_color_manual("Affiliation", values = c("Centrist"="green", "Government"="blue", "Opposition"="or
  scale_size_manual("Separate Mechanism", values = c("No" = 1, "Yes"=2.5))+
  scale_shape_manual("Voting Result", values = c("Negatived"=17, "Passed"=19))+
  labs(title="Attendance and Approval of Motions Except Bills", x="Attendance Rate", y="Approval Rate")

ggsave("motions(except bills).png", motion_anal_2)
```

```
#Plot all motions in the 6th term.
motion_anal_3 <- frame_edited %>%
  ggplot(aes(x= attendance_rate, y= approval_rate, color=affiliation, size = separate_mechanism, shape
  geom_point()+
  coord_fixed(ratio=1/2.4)+
  scale_color_manual("Affiliation", values = c("Centrist"="green", "Government"="blue", "Opposition"="o
  scale_size_manual("Separate Mechanism", values = c("No" = 1, "Yes"=2.5))+
  scale_shape_manual("Voting Result", values = c("Negatived"=17, "Passed"=19))+
  labs(title="Attendance and Approval of All Motions", x="Attendance Rate", y="Approval Rate")

ggsave("motions in the 6th term.png", motion_anal_3)
```

How about taking separate mechanism into consideration? One for example. (Only motions raised by the Officials or related to government would not be subject to the separate mechanism, which increased the probability of passing the initiative raised by the executive branch). (Unfortunately, several bugs appeared while grabbing the voting data as it is not reasonable to denote a voting result as 'pass' if the approval rate is lower than 0.5. Problem might appear where data was revealed "Yes" in the dataframe when grabbing "No" in the xml, but generally, the graph largely shows the 'dominance of executive branch' in Hong Kong politics).

```
motion_anal_4 <- frame_edited %>%
  filter(separate_mechanism == "No") %>%
  ggplot(aes(x= attendance_rate, y= approval_rate, color=affiliation, shape = voting_result))+
  geom_point()+
  coord_fixed(ratio=1/2.4)+
  scale_color_manual("Affiliation", values = c("Centrist"="green", "Government"="blue", "Opposition"="o
  scale_shape_manual("Voting Result", values = c("Negatived"=17, "Passed"=19))+
  labs(title="Attendance and Approval of All Motions", x="Attendance Rate", y="Approval Rate")

ggsave("motions without separate mechanism.png", motion_anal_4)
```

Write the csv for some important dataframe.

```
write.csv(voting_record, "voting.csv")
write.csv(frame_edited, "type of motions.csv")
write.csv(bills, "bills.csv")
write.csv(frame_motion, "pure record of motions.csv")
```

Reflections

There are still some 'bugs' in this project. The first one is about the repeated work of scraping the raw data from the website, as I use the function for scrapping attempt but runs nothing. I hope it would be the next stage's work in refining my job. The second is the 'bugs' after scrapping the voting results for building a dataframe as I found finally after making plots.