

Projet final

Yun Shen Li (20210446)

2024-04-19

Trudeau et Harper: premiers ministres de deux Canada différents?

Contexte de la recherche

Justin Trudeau a hérité le Canada de Stephen Harper en 2015, un pays qui avait très étroitement échappé à une récession, d'après Greg Quinn de Bloomberg News. De plus, la Commission de vérité et réconciliation avait publié son rapport final qui avait présenté une conclusion dévastatrice sur la vérité du maltraitement des autochtones par le gouvernement canadien, un événement décrit comme l'un des deux moments politiques marquants de 2015 par Evan Solomon de Maclean's, l'autre étant évidemment que Justin Trudeau avait gagné les élections fédérales pour prendre le siège du 23ième Premier ministre du Canada. D'après Solomon, sa victoire avait ressuscité le Parti libéral parmi un atmosphère politique tendue, particulièrement à cause de la crise des réfugiés syriens et le racisme inévitable qui s'est aggravé avec. Cela dit, on avait l'impression que Trudeau était le jeune sauveur du Canada qui montrait sa vision pour un Canada libéral en accueillant lui-même les réfugiés à l'aéroport Pearson International, juste après sa victoire en novembre 2015. De ce fait, presque dix ans se sont écoulés après qu'il ait pris les rênes pour la première fois, et il est évident que son mandat a connu une histoire qui fluctue énormément comme des montagnes russes: une crise immobilière qui semble insoluble, une crise de réfugiés, la pandémie COVID-19 et ses conséquences, le conflit Israélo-palestinien, etc. . . bref, une multitude d'enjeux qui ne semblent pas pouvoir lui faciliter la vie. Étant étudiante d'économie et politique, je m'intéresse particulièrement aux cycles économiques et politiques d'un pays, et je me demande souvent pourquoi certains enjeux sociétaux, économiques, politiques, etc. . . ne quittent jamais un pays peu importe les politiques gouvernementaux. C'est pour cela que cette recherche se porte sur les enjeux les plus mentionnés par Trudeau et son prédécesseur, le 22ième Premier ministre canadien Stephen Harper, et par curiosité j'aimerais voir si je peux identifier les différences et leurs causes entre leurs agendas.

Question de recherche

Officiellement, ma question de recherche est la suivante: Est-ce qu'il y a une différence entre les enjeux les plus mentionnés par Trudeau et Harper lors de leur mandat respectif comme Premier ministre du Canada?

Présentation des données utilisés

J'ai employé les données téléchargées du LIPAD dans un fichier qui contient tous les transcriptions et données datant de 2006 jusqu'à 2019, l'année qu'ils ont arrêté de mettre à jours cette base de données. De plus, j'ai choisi 2006 à 2019 car c'est les années qui contiennent les mandats respectifs de Harper (2006 - 2015) et Trudeau (2015 - présent), donc j'ai pu rattrapper et préciser les données que j'utiliserai. Les données du LIPAD contiennent les dates, le nom de l'orateur ainsi que le discours transcrits de celui-ci et donc c'était la base parfaite pour ma recherche, en plus d'être très facilement manipulable et nettoyée pour pouvoir répondre à ma question de recherche. Ensuite, j'ai également téléchargé Lexicoder en anglais pour effectuer

mon analyse de dictionnaire: c'est une liste contenant des listes de mots catégorisés par un certain 'keyword' justement pour effectuer des analyses comme la mienne.

Après avoir importé la base de données et le dictionnaire, j'ai filtré pour avoir seulement les discours de Trudeau et Harper de 2006 à 2019, puis transformé tout le texte en minuscule pour faciliter l'analyse.

```
## -- dictionary and lipad data -- ##
lexicoder_en <- dictionary(file = "/Users/yoshili/fas_1001_Li/travail fin session/LTDjun2013/policy_age

data <- read_csv("/Users/yoshili/fas_1001_Li/travail fin session/lipad_merge_2006_2019.csv")

## -- cleaning -- ##
data_clean <- data |>
  select(speechtext, speakername, speechdate) |>
  mutate(speechtext = tolower(speechtext)) |>
  na.omit()
filtered_data <- data_clean |>
  filter(speakername %in% c("Justin Trudeau", "Stephen Harper"),
         year(speechdate) >= 2006 & year(speechdate) <= 2019)
```

Présentation des méthodes utilisés

Comme mentionné ci-dessus, j'ai effectué une analyse de dictionnaire sur les transcriptions des discours de Trudeau et Harper à la Chambre de Communes pour identifier les enjeux qu'ils mentionnaient le plus et pour voir s'il y a une différence entre eux. J'ai eu beaucoup de misère en manipulant les données et en essayant avec un des exemples vu dans le tp3 du cours, mais finalement j'ai réussi en transformant le dictionnaire Lexicoder en liste, puis ensuite j'ai pioché les 'keywords' du dictionnaire pour pouvoir utiliser pivot_longer().

```
## -- troubleshooting -- ##
str(lexicoder_en) #list of lists

keyword_list <- unlist(lexicoder_en)
lexicoder_df <- data.frame(
  topic = rep(names(keyword_list), lengths(keyword_list)),
  keywords = unlist(keyword_list)
)

lexicoder_long <- lexicoder_df |>
  pivot_longer(cols = starts_with("keywords"),
               names_to = "keyword_index",
               values_to = "keyword")
lexicoder_long$topic <- gsub("\\d+", "", lexicoder_long$topic)

head(lexicoder_long)

lexicoder_long <- lexicoder_long |>
  select(-keyword_index)
```

Pour effectuer mon analyse, j'ai d'abord utilisé le 'tokenization' pour pouvoir matcher les mots clés dans la transcription des discours:

```

#tokenization
filtered_data <- filtered_data |>
  mutate(tokens = strsplit(speechtext, "\\s+"))

#matching keywords
keyword_counts <- lapply(lexicoder_long$keyword, function(keyword) {
  sapply(filtered_data$tokens, function(tokens) {
    sum(tokens %in% keyword)
  })
})
})

```

Mes étapes d'analyse: J'ai filtré les données pour avoir seulement la date du discours et la transcription pour identifier les enjeux mentionnés qui correspondaient avec les mots clés de Lexicoder. Les résultats sont placés dans un dataframe qui contient les variables 'topic' et 'keyword'.

```

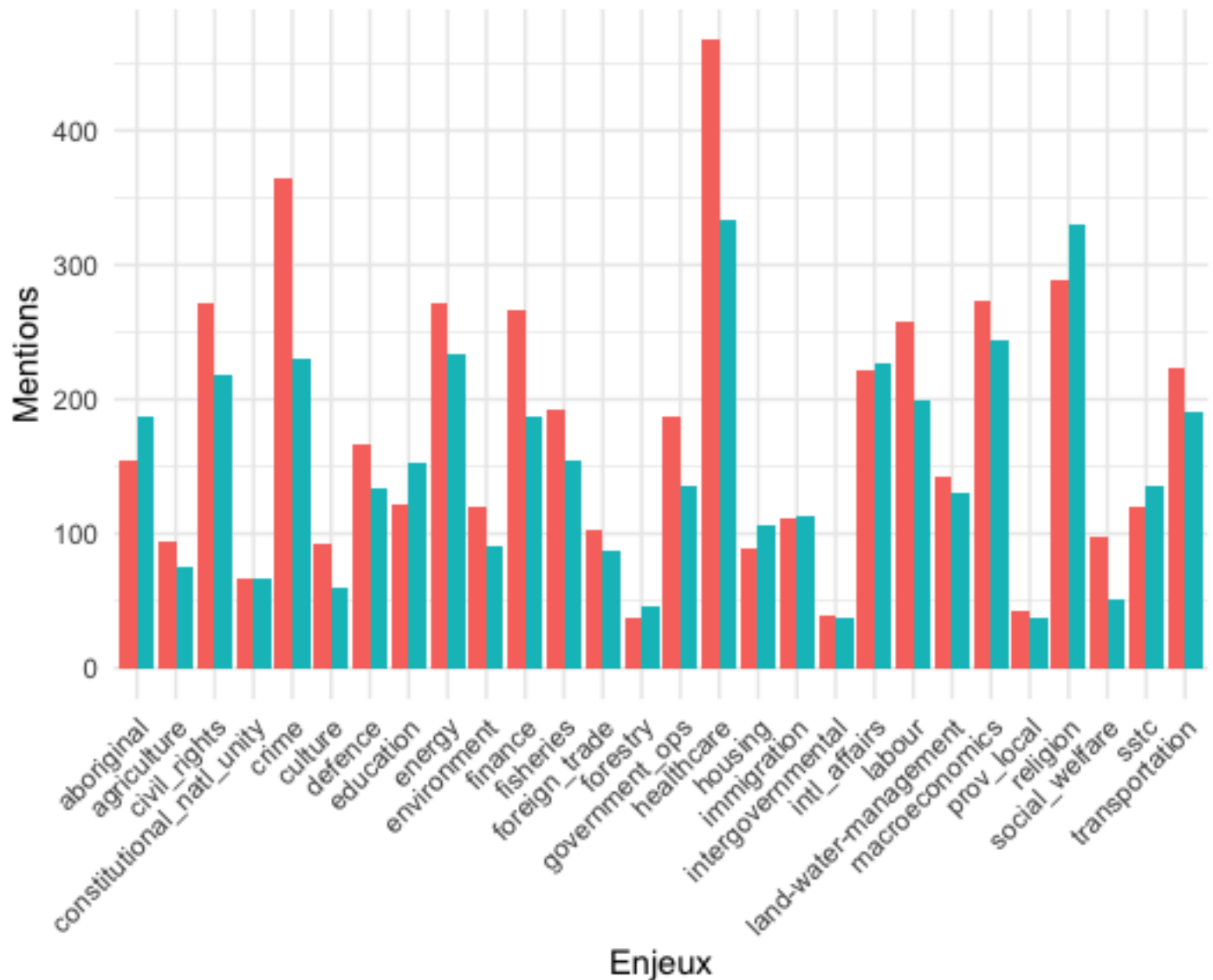
#analysis
results <- data.frame(
  speechdate = filtered_data$speechdate,
  speakername = filtered_data$speakername,
  topic = lexicoder_long$topic,
  keyword_count = unlist(keyword_counts)
) |>
  group_by(speakername, topic) |>
  summarise(total_keywords = sum(keyword_count))

```

Pour visualiser mes résultats, j'ai choisi de faire des graphiques à barres: le premier est une comparaison directe entre les enjeux les plus mentionnés par Trudeau et Harper au cours de 2006 - 2019.

```
knitr::include_graphics("/Users/yoshili/fas_1001_Li/graphbar1.png")
```

Enjeux mentionnés par Harper et Trudeau au cours de leurs m



```
#visualisation
ggplot(results, aes(x = topic, y = total_keywords, fill = speakername)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Enjeux mentionnés par Harper et Trudeau au cours de leurs mandats",
        x = "Enjeux",
        y = "Mentions") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

On peut donc remarquer que l'assurance médicale est l'enjeu le plus mentionné par Trudeau au cours des années, avec d'autres mentions principales comme le taux de criminalité, les droits civils, l'environnement, l'éducation et la finance. La distribution pour les enjeux mentionnés par Harper sont beaucoup plus éparpillés, mais il se concentraient beaucoup sur l'assurance médicale, la religion, l'énergie et le taux de criminalité.

Pour approfondir ma recherche, j'ai effectué la même analyse textuelle mais cette fois-ci, j'ai séparé les

données par année de mandat: pour Harper j'ai utilisé les données de 2006 à 2015, et puis de 2015 à la dernière mise à jours de LIPAD en 2019. J'ai séparé les données de Trudeau et Harper en utilisant la même méthode de tokenization et matching de mots clés.

```
## -- analyse par année -- ##
trudeau_data <- filtered_data |>
  filter(speakername == "Justin Trudeau") |>
  rename(keyword = tokens) |>
  unnest(keyword)
harper_data <- filtered_data |>
  filter(speakername == "Stephen Harper") |>
  rename(keyword = tokens) |>
  unnest(keyword)

joined_data_trudeau <- inner_join(trudeau_data, lexicoder_long, by = "keyword", relationship = "many-to-many")
joined_data_harper <- inner_join(harper_data, lexicoder_long, by = "keyword", relationship = "many-to-many")

#fun to aggregate by year and topic
aggregate_year <- function(data) {
  data <- data |>
    mutate(year = year(as.Date(speechdate))) |>
    group_by(year, topic) |>
    summarise(total_mentions = n()) |>
    ungroup()

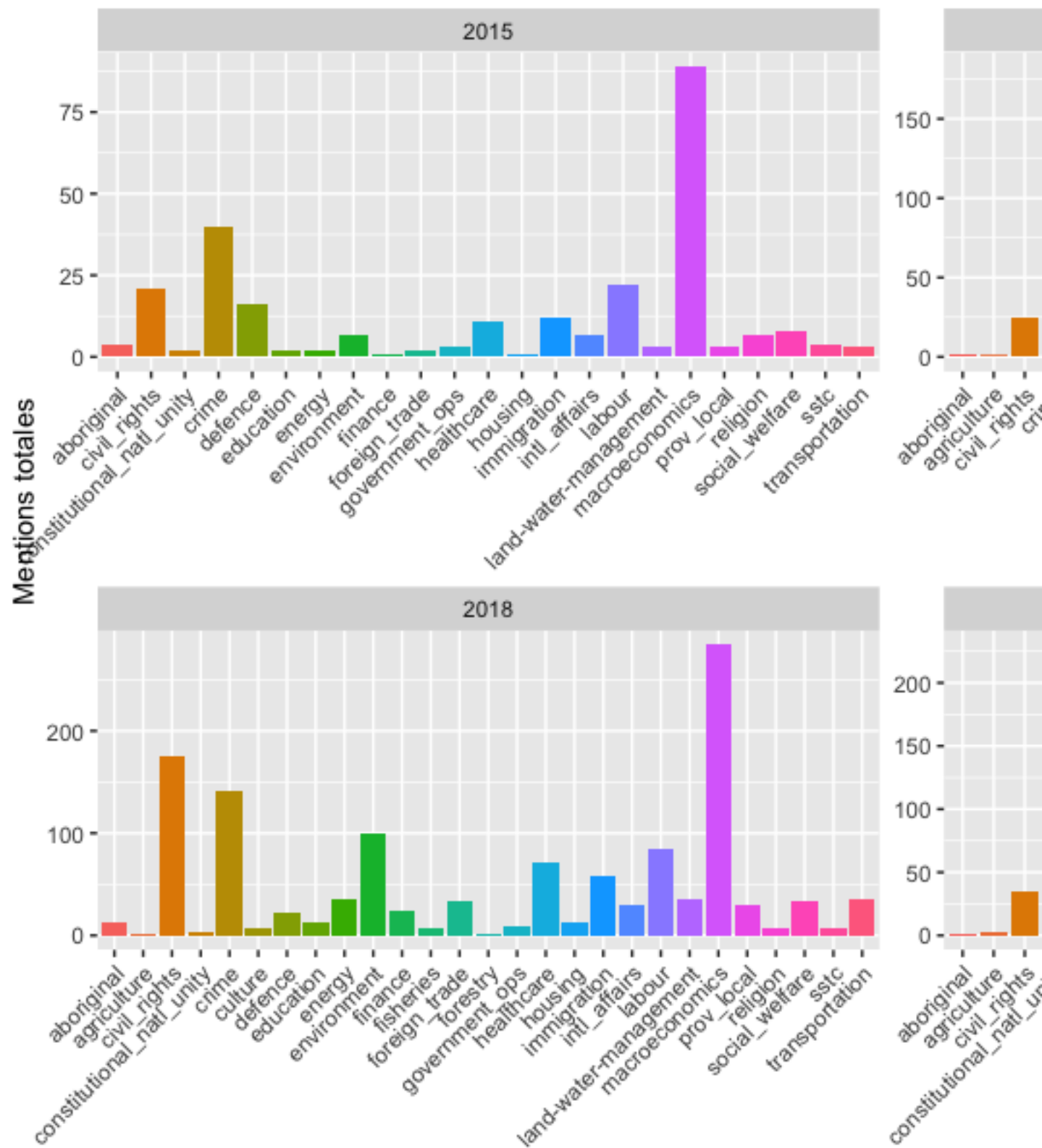
  return(data)
}

#putting allat together
trudeau_all <- aggregate_year(joined_data_trudeau)
harper_all <- aggregate_year(joined_data_harper)
```

Visualisation Justin Trudeau

```
knitr::include_graphics("/Users/yoshili/fas_1001_Li/Trudeau.png")
```

Enjeux mentionnés par Justin Trudeau (2016 - 2019)



```

#visualisation trudeau
trudeau_pm <- trudeau_all |>
  filter(year >= 2015 & year <= 2019)

ggplot(data = trudeau_pm, aes(x = topic, y = total_mentions, fill = topic)) +
  geom_bar(stat = "identity") +
  facet_wrap( ~ year, scales = "free") +
  labs(title = "Enjeux mentionnés par Justin Trudeau (2016 - 2019)",
        x = "Enjeux",
        y = "Mentions totales") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "none")

```

On observe qu'il mentionne principalement des enjeux économiques, et ces mentions s'empilent de plus en plus au cours des années. Les mentions pour les droits civiles et le taux de criminalité augmentent aussi, avec 2019 l'année où il a eu un agenda plus diversifié, même comparé aux enjeux économiques qui restent dominant.

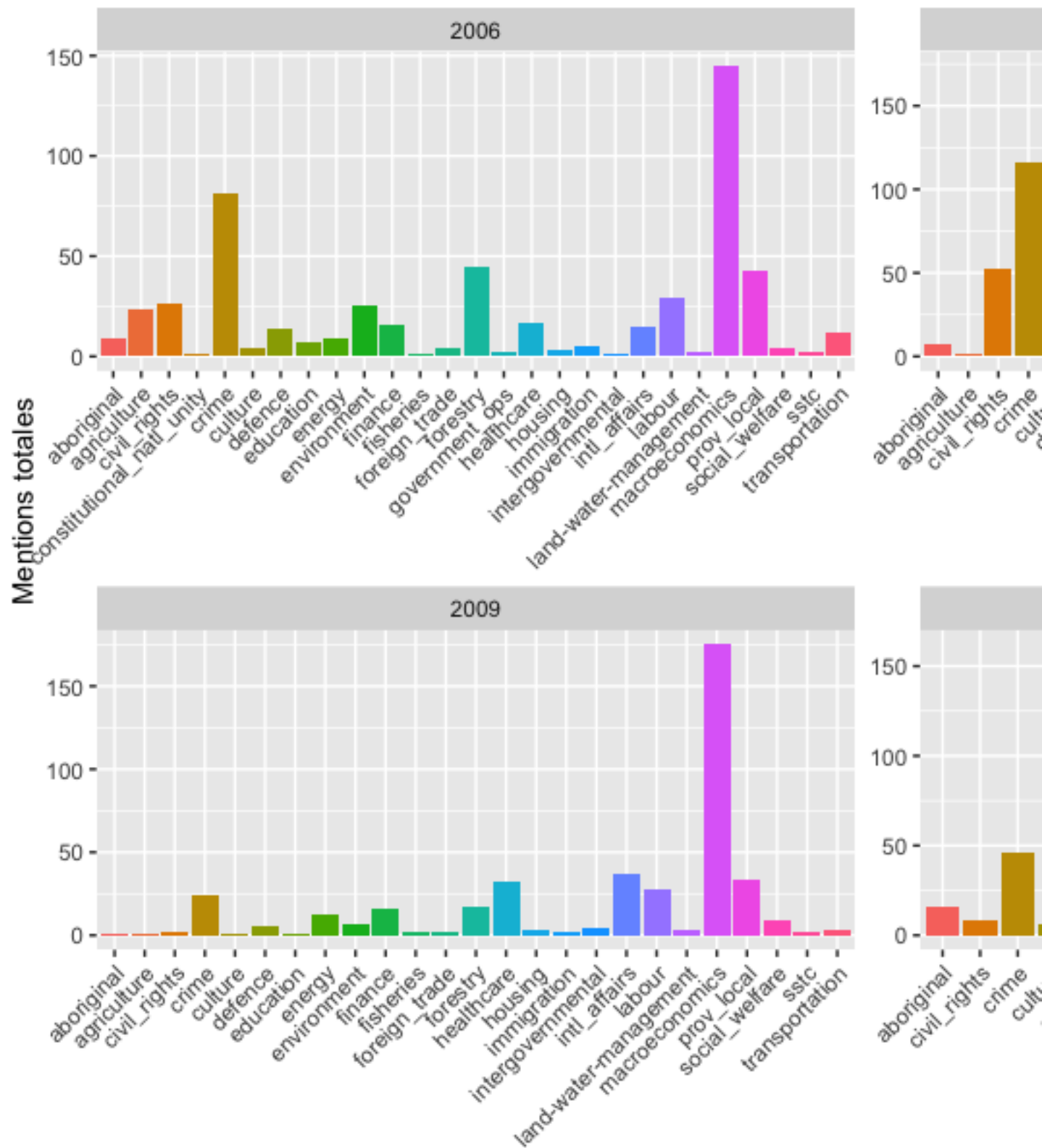
Visualisation Stephen Harper

```

knitr::include_graphics("/Users/yoshili/fas_1001_Li/harper1.png")

```

Enjeux mentionnés par Stephen Harper (2006 - 2010)




```

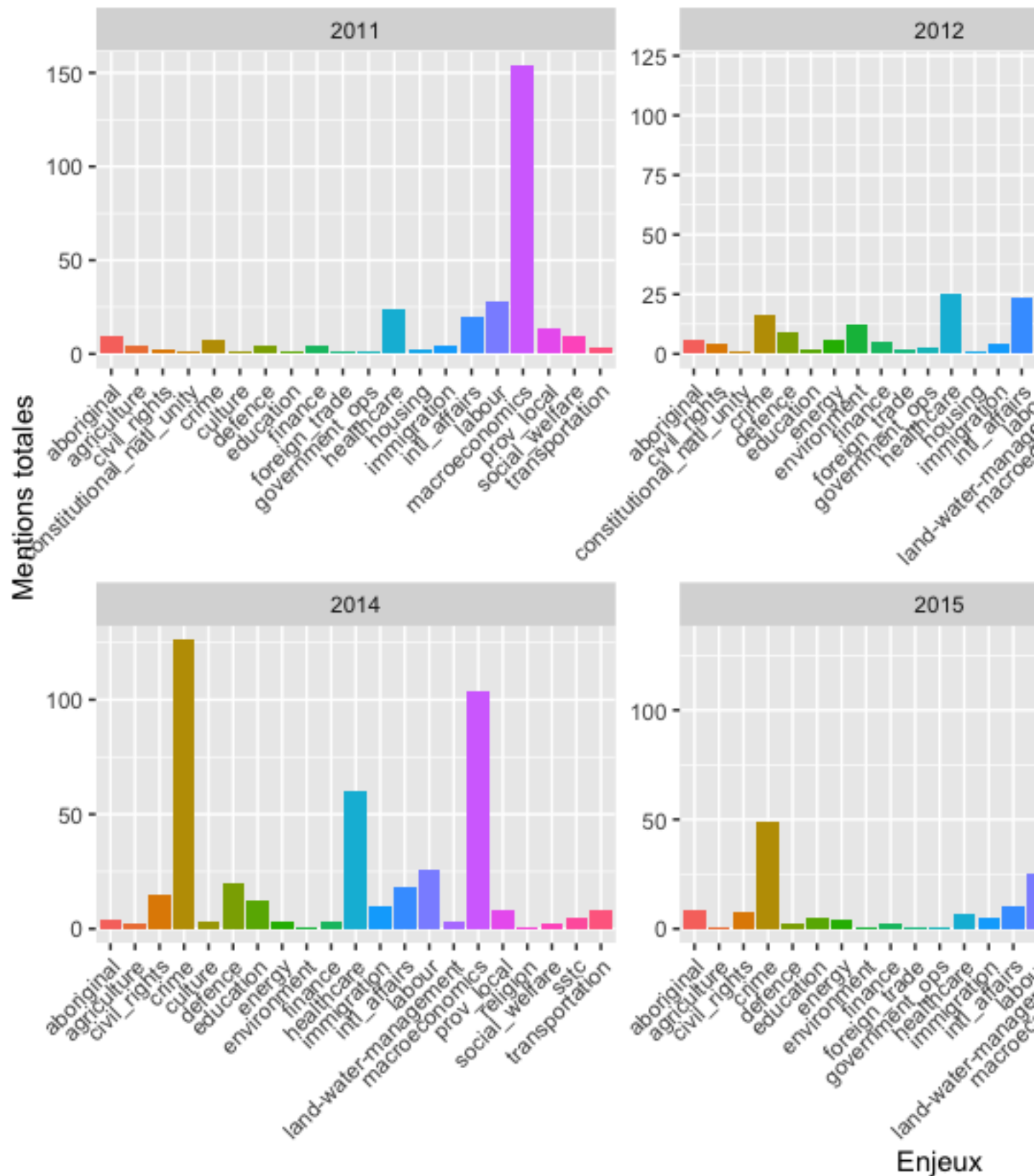
#visualisation harper
harper_pm_1 <- harper_all |>
  filter(year >= 2006 & year <= 2010)
ggplot(data = harper_pm_1, aes(x = topic, y = total_mentions, fill = topic)) +
  geom_bar(stat = "identity") +
  facet_wrap( ~ year, scales = "free") +
  labs(title = "Enjeux mentionnés par Stephen Harper (2006 - 2010)",
        x = "Enjeux",
        y = "Mentions totales") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "none")

```

Cette première visualisation contient les enjeux mentionnés par Harper durant ses deux mandats comme chef de gouvernement minoritaire. Tout comme Trudeau, il s'est focalisé principalement sur les enjeux économiques, avec le taux de criminalité en deuxième place mais à une fréquence beaucoup moins élevée.

```
knitr::include_graphics("/Users/yoshili/fas_1001_Li/harper 2.png")
```

Enjeux mentionnés par Stephen Harper (2011 - 2015)



```

harper_pm_2 <- harper_all |>
  filter(year >= 2011 & year <= 2015)
ggplot(data = harper_pm_2, aes(x = topic, y = total_mentions, fill = topic)) +
  geom_bar(stat = "identity") +
  facet_wrap(~ year, scales = "free") +
  labs(title = "Enjeux mentionnés par Stephen Harper (2006 - 2010)",
       x = "Enjeux",
       y = "Mentions totales") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "none")

```

Cette deuxième visualisation contient les enjeux mentionnés durant son seul mandat majoritaire, et on voit que les enjeux économiques sont encore très dominants dans les discours de Harper, avec un pic sur les enjeux de taux de criminalité en 2013 et 2014.

Conclusion de l'étude

D'après les données et les graphiques présentés, les enjeux économiques préoccupaient énormément les deux premiers ministres, ainsi que le taux de criminalité en deuxième mandat, mais, encore une fois, ceci est à un degré moins élevé. Pour Harper, je présume que le pic des mentions économiques en 2009 est à cause de la crise financière. Pour Trudeau, ses mentions dominantes du système de santé ont dû s'empiler spécialement avec la pandémie de la COVID-19. On remarque également que les deux premiers ministres mentionnaient beaucoup les enjeux qui affectent les Canadiens directement comme le taux de criminalité, l'éducation, la religion et les transports. La différence marquante est que Trudeau mentionne beaucoup plus les enjeux liés aux droits civils et le taux de criminalité comparé à Harper, une différence qui pourrait être causée par leur association à des partis plus ou moins opposés sur l'axe politique, même si l'atmosphère politique canadienne est beaucoup moins polarisée qu'aux États-Unis par exemple.