

EDLD652 Lab 2

Hyeonjin Cha

Rachael Latimer

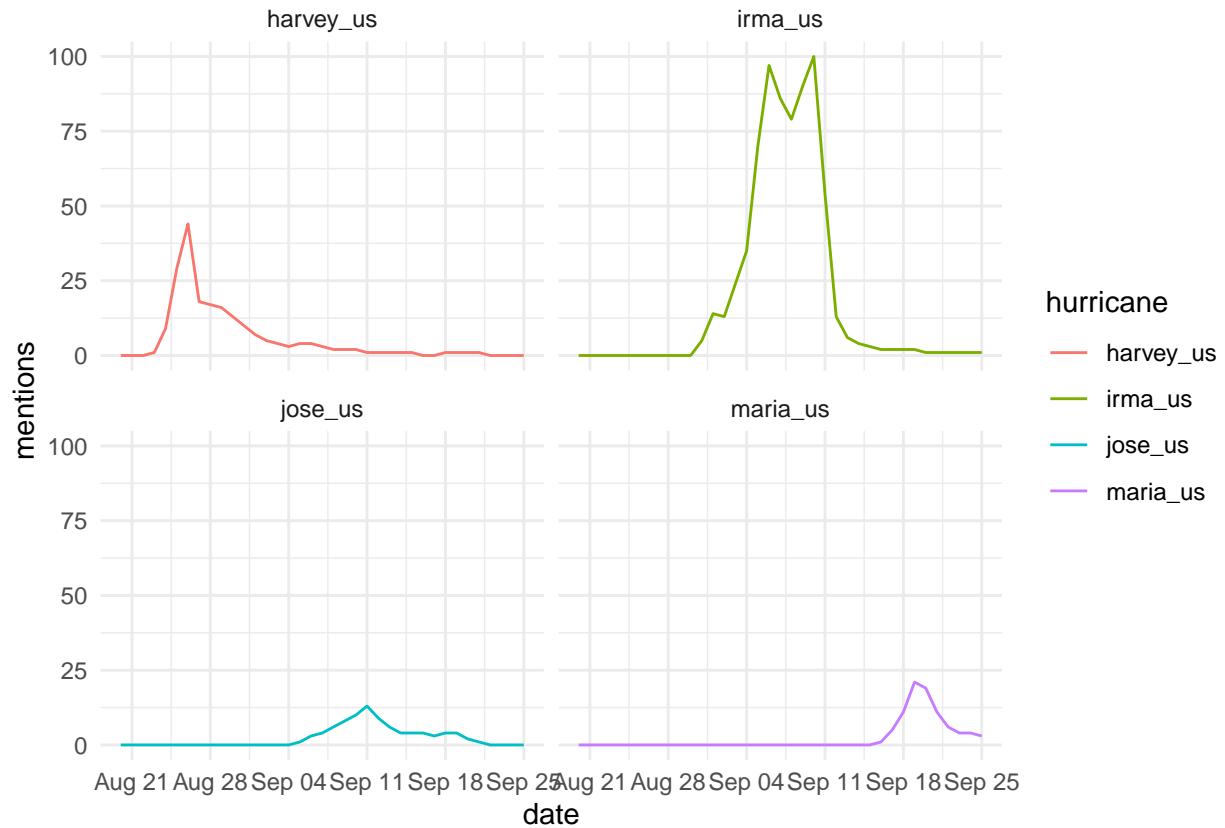
Tess Sameshima

2/3/2021

1. Google Trends Data

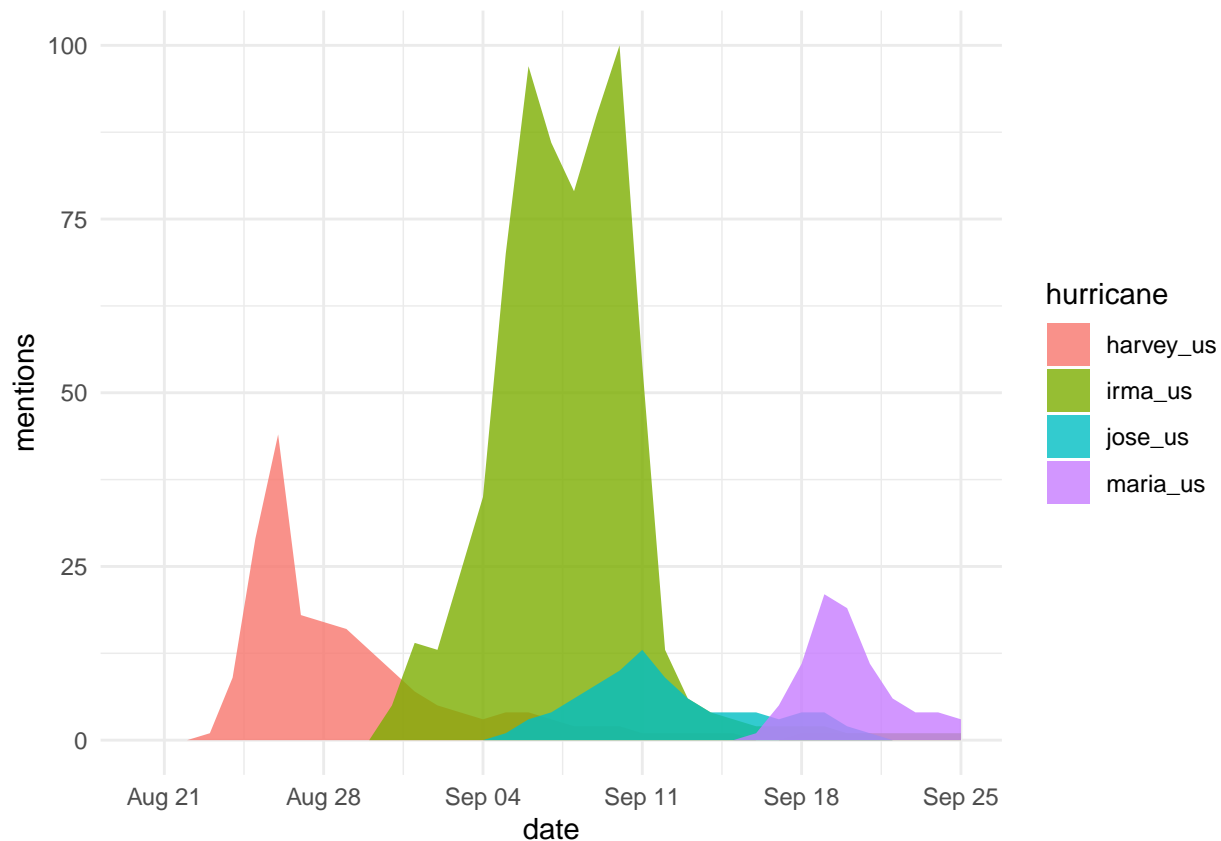
```
#dataset
google_trends_longer <- google_trends %>%
  pivot_longer(cols = starts_with("hurricane"),
               names_to = "hurricane",
               names_prefix = "hurricane_",
               values_to = "mentions")

#part 1
plot1 <- google_trends_longer %>%
  ggplot(aes(date, mentions, color = hurricane)) +
  geom_line() +
  facet_wrap(~hurricane) +
  theme_minimal()
plot1
```



```
#part2
plot2 <- google_trends_longer %>%
  ggplot(aes(date, mentions, fill = hurricane)) +
  geom_area(position = "dodge", alpha = 0.8) +
  theme_minimal()
plot2
```

```
## Warning: Width not defined. Set with 'position_dodge(width = ?)'
```



```
#part3

# plot3 <- google_trends_longer %>%
#   ggplot(aes(date, mentions)) +
#   geom_area(aes(fill = hurricane)) +
#   scico:: scale_fill_scico(palette = "tokyo") +
#   theme_minimal

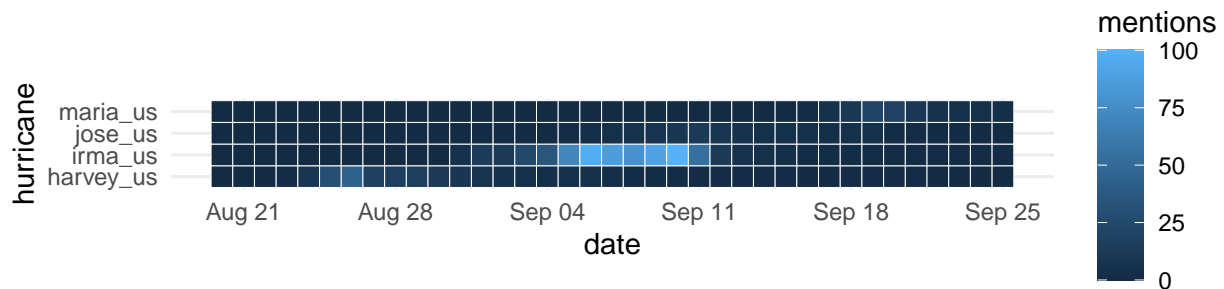
# #Error: Discrete value supplied to continuous scale
# How do I change the scales so color is mapped to a continuous scale? Hurricane is a categorical variable

# Maybe change the variable from categorical to numerical?
# google_trends_longer$hurricane = as.numeric(levels(google_trends_longer$hurricane))[google_trends_longer$hurricane]
# this didn't seem to work either... it all turned NA

#I asked Daniel for help; feel like an idiot not figuring out this simple but oh well
plot3 <- google_trends_longer %>%
  group_by(hurricane, date) %>%
  summarize(mentions = mean(mentions)) %>%
  ggplot(aes(date, hurricane)) +
  geom_tile(aes(fill = mentions,
                color = "white")) +
  coord_fixed() +
  theme_minimal()
```

```
## 'summarise()' regrouping output by 'hurricane' (override with '.groups' argument)
```

```
plot3
```



```
#part 4
```

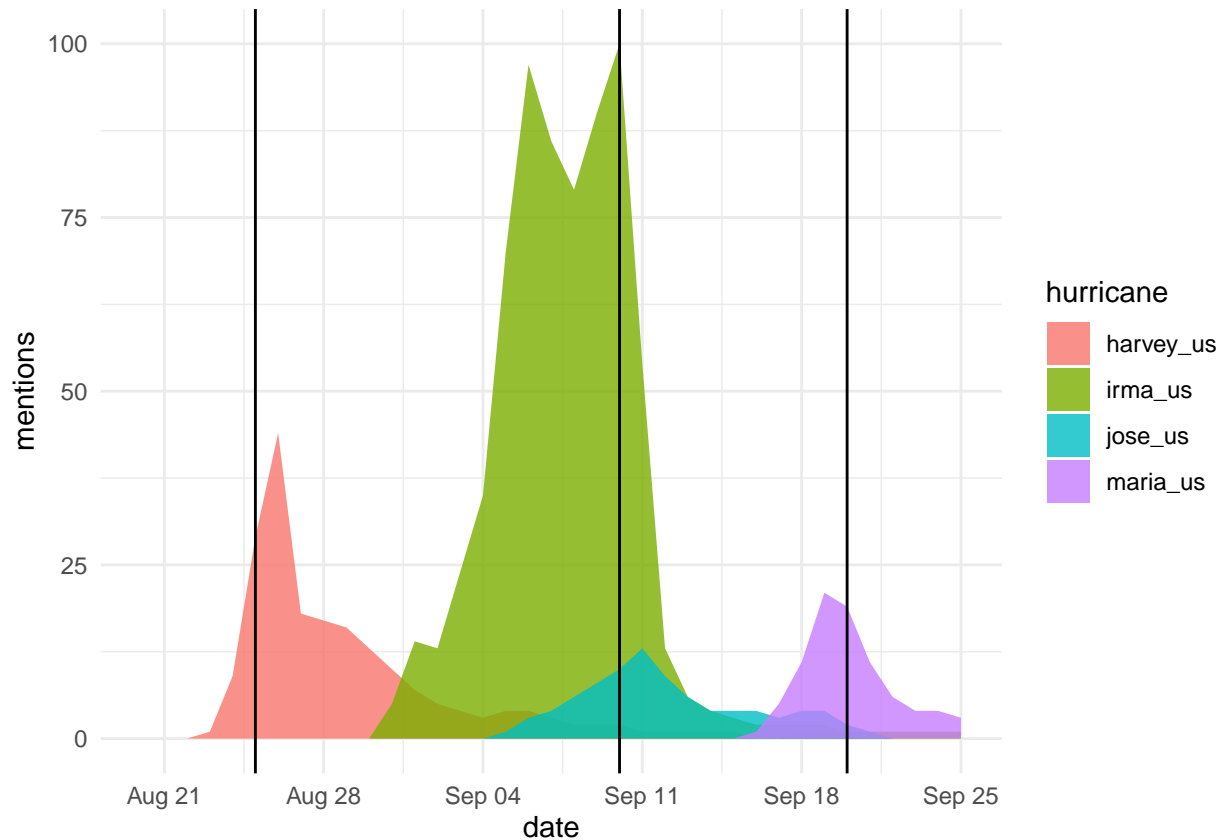
```
# google_trends_longer_landfall <- google_trends_longer %>%
#   mutate(landfall = 0) %>%
#   mutate(landfall = ifelse(
#     hurricane == "harvey_us" & date == "2017-08-25" |
#     hurricane == "irma_us" & date == "2017-09-10" |
#     hurricane == "maria_us" & date == "2017-09-20",
#     1, 0))
#probably not the way Daniel intended....
```

```
landfall <- data.frame(hurricane = c("harvey_us", "irma_us", "maria_us"),
                      date = as.Date(c("2017-08-25", "2017-09-10", "2017-09-20")),
                      Ref = c("Harvey landfall", "Irma landfall", "Maria landfall"),
                      stringsAsFactors = FALSE)
```

```
plot4 <- google_trends_longer %>%
  ggplot(aes(date, mentions, fill = hurricane)) +
  geom_area(position = "dodge", alpha = 0.8) +
  geom_vline(data = landfall, aes(xintercept=as.numeric(date[c(1,2,3)]))) +
  theme_minimal()
```

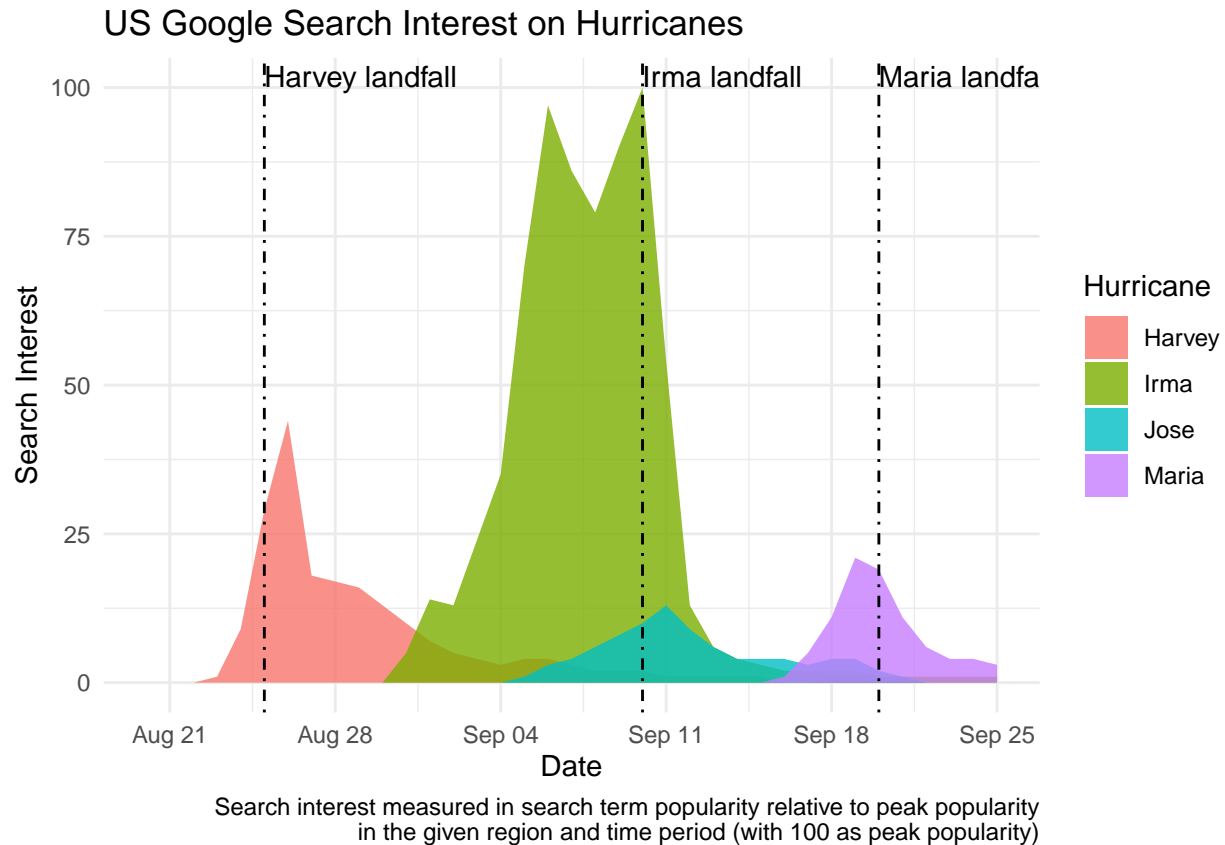
```
plot4
```

```
## Warning: Width not defined. Set with 'position_dodge(width = ?)'
```



```
#part 5
plot5 <- google_trends_longer %>%
  ggplot(aes(date, mentions, fill = hurricane)) +
  geom_area(position = "dodge", alpha = 0.8) +
  geom_vline(data = landfall, aes(xintercept=as.numeric(date[c(1,2,3)])), linetype=4) +
  geom_text(mapping = aes(x = date,
                          y = 100,
                          label = Ref,
                          hjust = 0,
                          vjust = 0),
            data = landfall) +
  labs(title = "US Google Search Interest on Hurricanes",
       x = "Date",
       y = "Search Interest",
       label = "Hurricane",
       caption = "Search interest measured in search term popularity relative to peak popularity
in the given region and time period (with 100 as peak popularity)") +
  scale_fill_discrete(name = "Hurricane",
                     breaks=c("harvey_us", "irma_us", "jose_us", "maria_us"),
                     labels=c("Harvey", "Irma", "Jose", "Maria")) +
  theme_minimal()
plot5
```

```
## Warning: Width not defined. Set with 'position_dodge(width = ?)'
```



2. Replicating “National cable news networks” Plot

```
tv_states_longer <- tv_states %>%
  pivot_longer(cols = florida:puerto_rico,
    names_to = "state",
    values_to = "percent")

TVlines<- data.frame(state = c("florida", "texas", "puerto_rico","florida"),
  date = as.Date(c("2017-08-25", "2017-09-10", "2017-09-20", "2017-10-01")),
  RefTV = c("Harvey landfall", "Irma landfall", "Maria landfall", "Las Vegas shooting"),
  stringsAsFactors = FALSE)

d <- data.frame(state = c("texas", "florida", "puerto_rico"),
  date=as.Date(c("2017-08-28", "2017-09-10", "2017-10-01")),
  percent=c(1,1.5,0.40),
  name = c("Texas", "Florida", "Puerto Rico"),
  stringsAsFactors = FALSE)

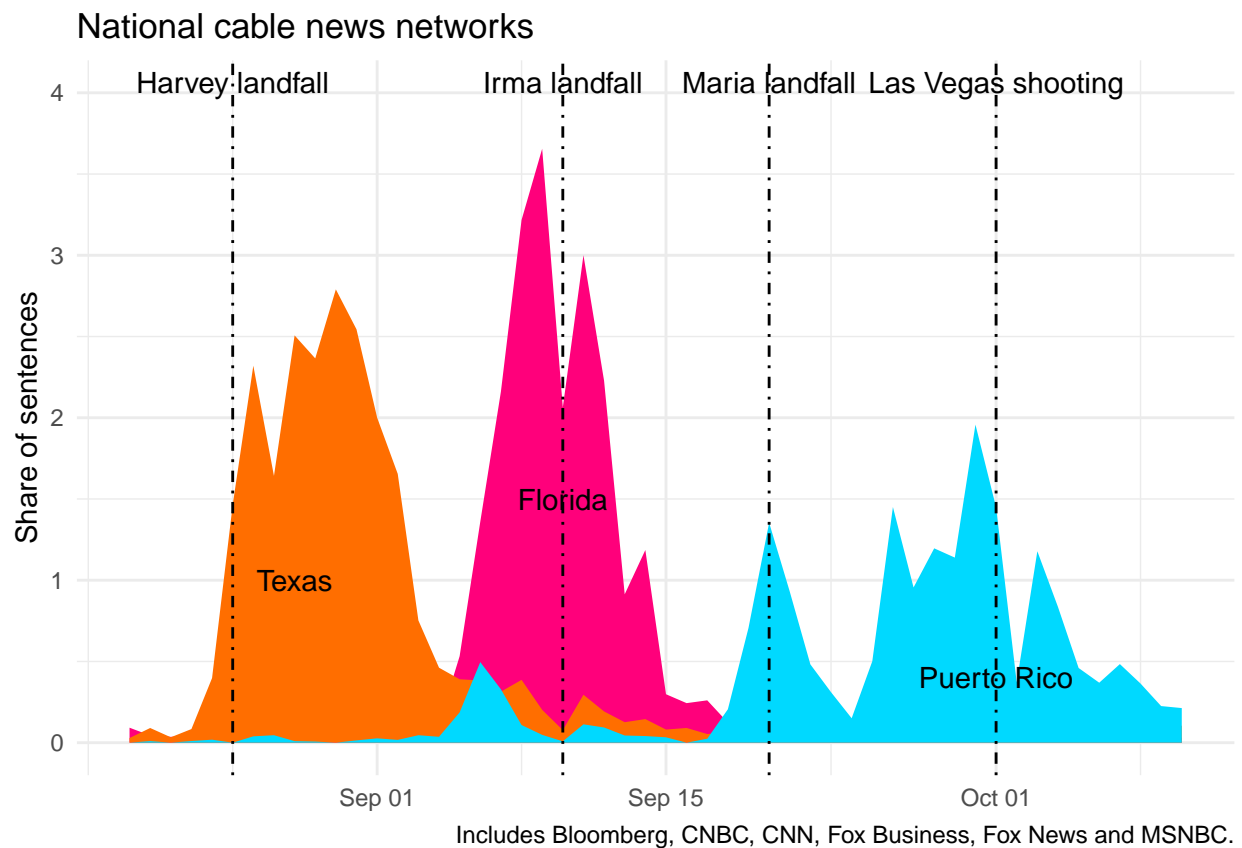
TVplot <- tv_states_longer %>%
  mutate(state = fct_relevel(state, "florida", "texas", "puerto_rico")) %>%
  ggplot(aes(date, percent, fill = state)) +
  guides(fill = FALSE) +
  geom_area(position = "dodge") +
  geom_vline(data = TVlines, aes(xintercept=as.numeric(date[c(1,2,3,4)])), linetype=4) +
```

```
geom_text(mapping = aes(x = date,
                        y = 4,
                        label = RefTV,
                        hjust = "center",
                        vjust = 0),
          data = TVlines) +
  scale_fill_manual(values = c("#ff007b", "#ff6e00", "#00d9ff")) +
  labs(title = "National cable news networks",
       x = NULL,
       y = "Share of sentences",
       caption = "Includes Bloomberg, CNBC, CNN, Fox Business, Fox News and MSNBC.") +

  theme_minimal() +
  geom_text(data=d, mapping=aes(x=date, y=percent, label=name))
```

TVplot

Warning: Width not defined. Set with 'position_dodge(width = ?)'



2. Using Comic_characters Dataset

```
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##      chisq.test, fisher.test

comic_characters <- comic_characters
N <- 23272

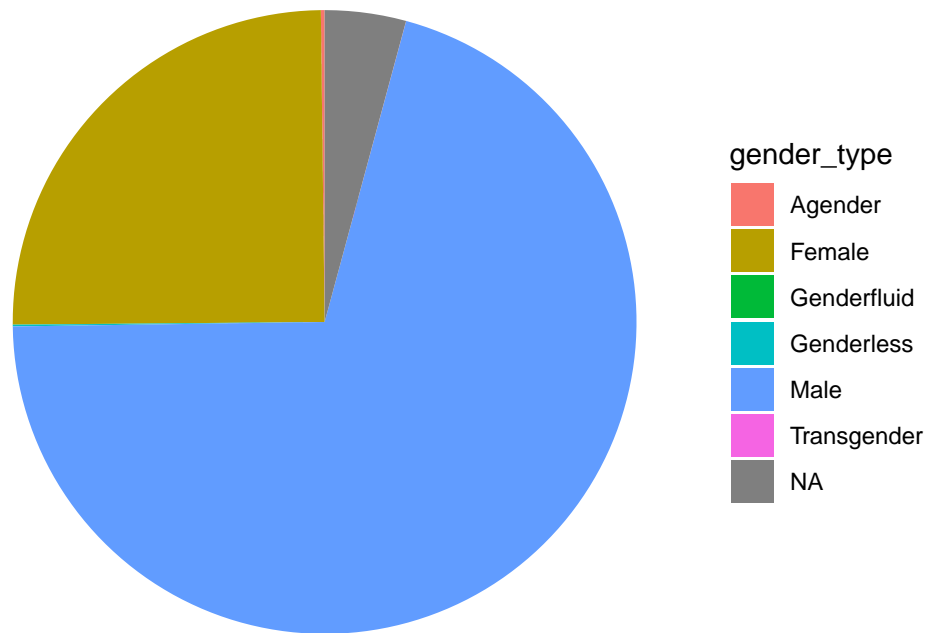
#pie chart of appearance count by first appearnce date
comic_characters_gender <- comic_characters %>%
  count(gender_type = sex) %>%
  mutate(gender_ratio = n/N) #devide each sex count value by total

comic_characters_gender$gender_type <- gsub(" Characters", "", comic_characters_gender$gender_type)

comic_plot_1 <- comic_characters_gender %>%
  ggplot(aes("", gender_ratio, fill = gender_type)) +
    geom_bar(stat = "identity", width = 5) +
    coord_polar("y", start = 0) +
    theme_void() +
    labs(title = "Ratio of Gender Types of Comic Book Characters",
         x = "Gender Ratio",
         y = NULL,
         caption = "Includes DC & Marvel Characters from 1938 to 2013")
    #scale_color_discrete(name = "Gender Types")

comic_plot_1
```

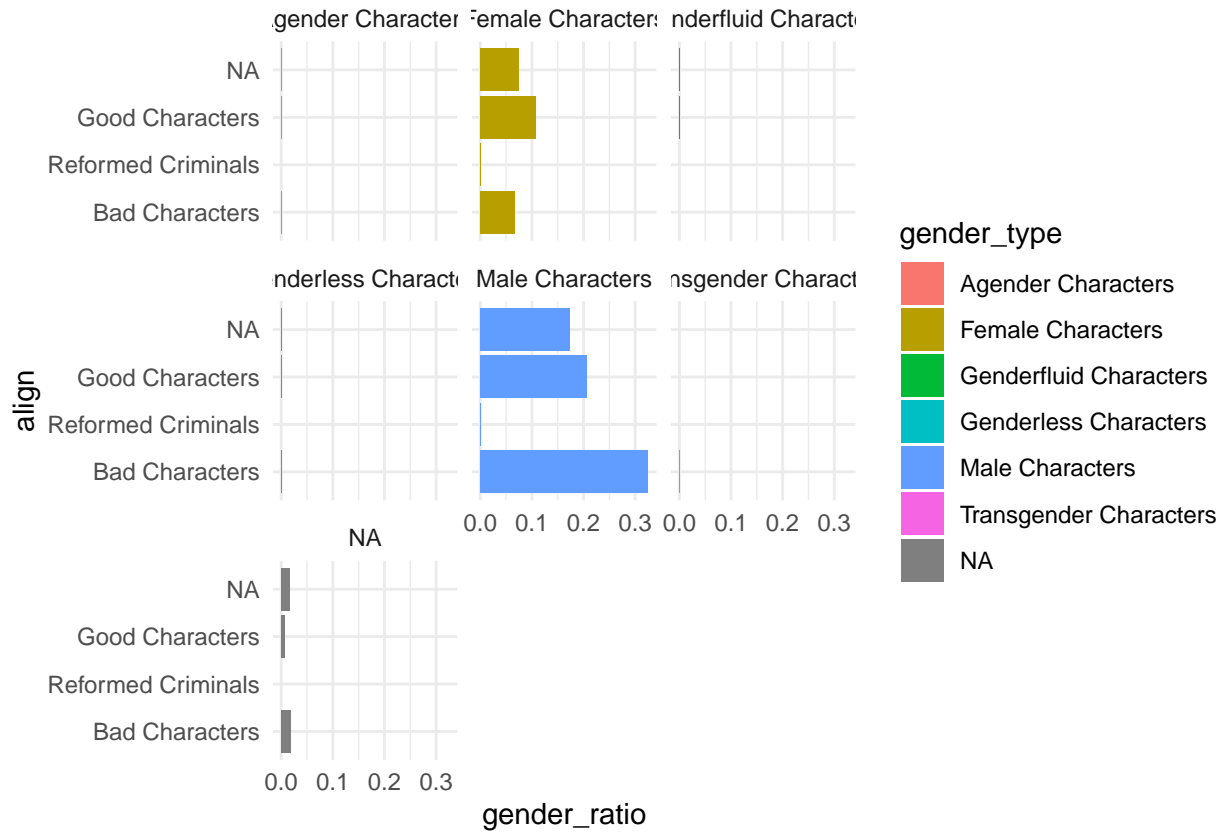

Ratio of Gender Types of Comic Book Characters



Includes DC & Marvel Characters from 1938 to 2013

```
#Bar plots for type of gender ratio by alignment
comic_plot_2 <- comic_characters %>%
  count(gender_type = sex, align) %>%
  mutate(gender_ratio = n/N) %>%
#tabyl(gender_type, align) %>% #returns dataframe with counts with sex as row and align as column,
  ggplot(aes(gender_ratio, align, fill = gender_type)) +
    geom_col() +
    facet_wrap(~gender_type) +
    theme_minimal()

comic_plot_2
```



```
comic_plot_3 <- comic_characters_gender %>%
  ggplot(aes(gender_type, gender_ratio, fill = gender_type)) +
  geom_bar(stat="identity")

comic_plot_3 +
  labs(title = "Ratio of Gender Types of Comic Book Characters",
       subtitle = "Includes DC & Marvel Characters from 1938 to 2013",
       x = "Type",
       y = "Ratio",
       label = "Type")
```

Ratio of Gender Types of Comic Book Characters
Includes DC & Marvel Characters from 1938 to 2013

