

Let them eat
cake!



Let them eat
cake!



Let them eat
cake first!



Imagine you're new to baking,
and you're in a baking class.
I'm going to present two
options for starting the class.
Which one gives you **better**
sense of the final product?

Pineapple and coconut sandwich cake



Pineapple and coconut sandwich cake

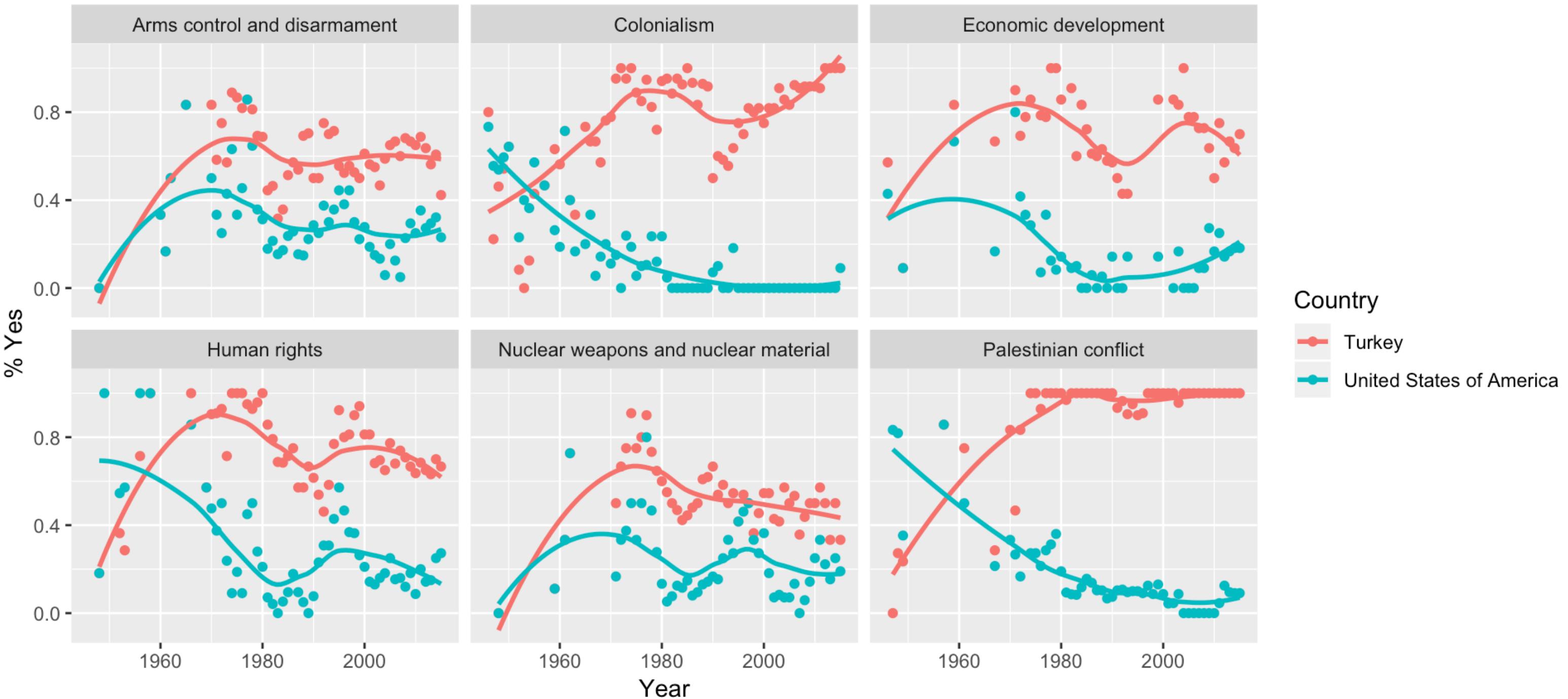




```
# Declare variables  
x ← 8  
y ← "monkey"  
z ← FALSE  
  
# Check class of x  
class(x)  
#> [1] "numeric"  
  
# Check class of y  
class(y)  
#> [1] "character"  
  
# Check class of z  
class(z)  
#> [1] "logical"
```



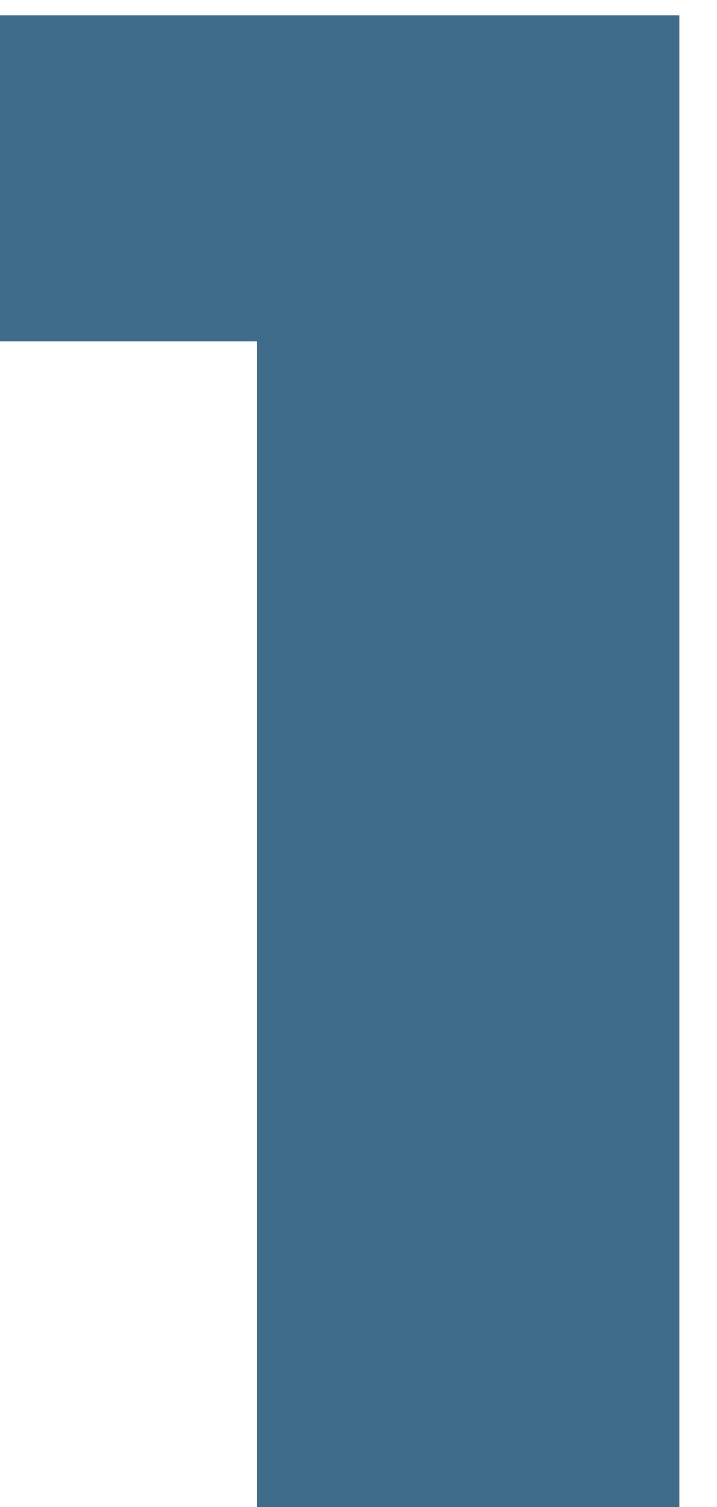
Percentage of 'Yes' votes in the UN General Assembly
1946 to 2015



5
min

mins
principles

**start
with
cake**

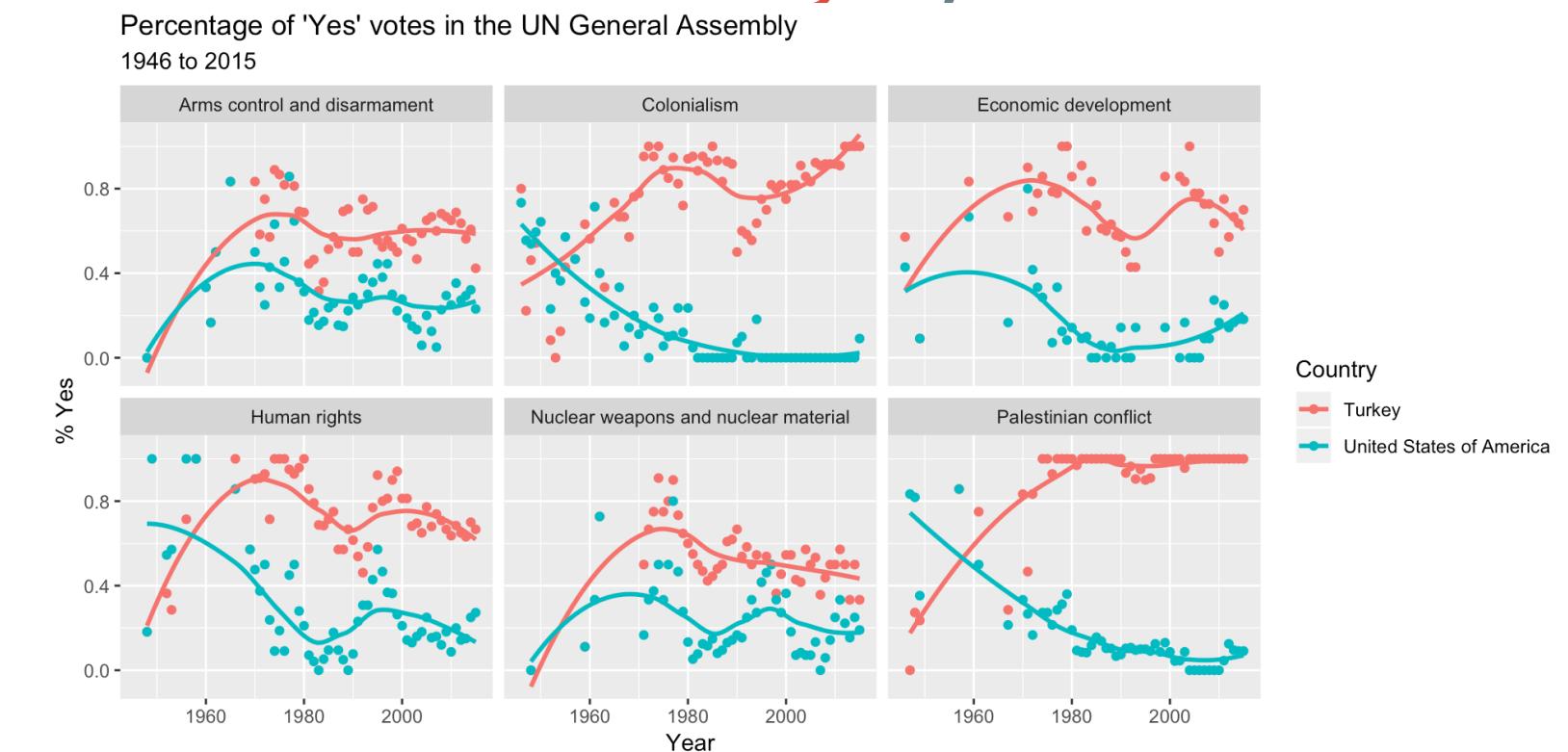


```

un_votes %>%
  filter(country %in% c("United States of America", "Turkey")) %>%
  inner_join(un_roll_calls, by = "rcid") %>%
  inner_join(un_roll_call_issues, by = "rcid") %>%
  group_by(country, year = year(date), issue) %>%
  summarize(
    votes = n(),
    percent_yes = mean(vote == "yes")
  ) %>%
  filter(votes > 5) %>% # only use records where there are more than 5 votes
  ggplot(mapping = aes(x = year, y = percent_yes, color = country)) +
    geom_point() +
    geom_smooth(method = "loess", se = FALSE) +
    facet_wrap(~ issue) +
    labs(
      title = "Percentage of 'Yes' votes in the UN General Assembly",
      subtitle = "1946 to 2015",
      y = "% Yes",
      x = "Year",
      color = "Country"
    )

```

rstd.io/eat-cake-5min

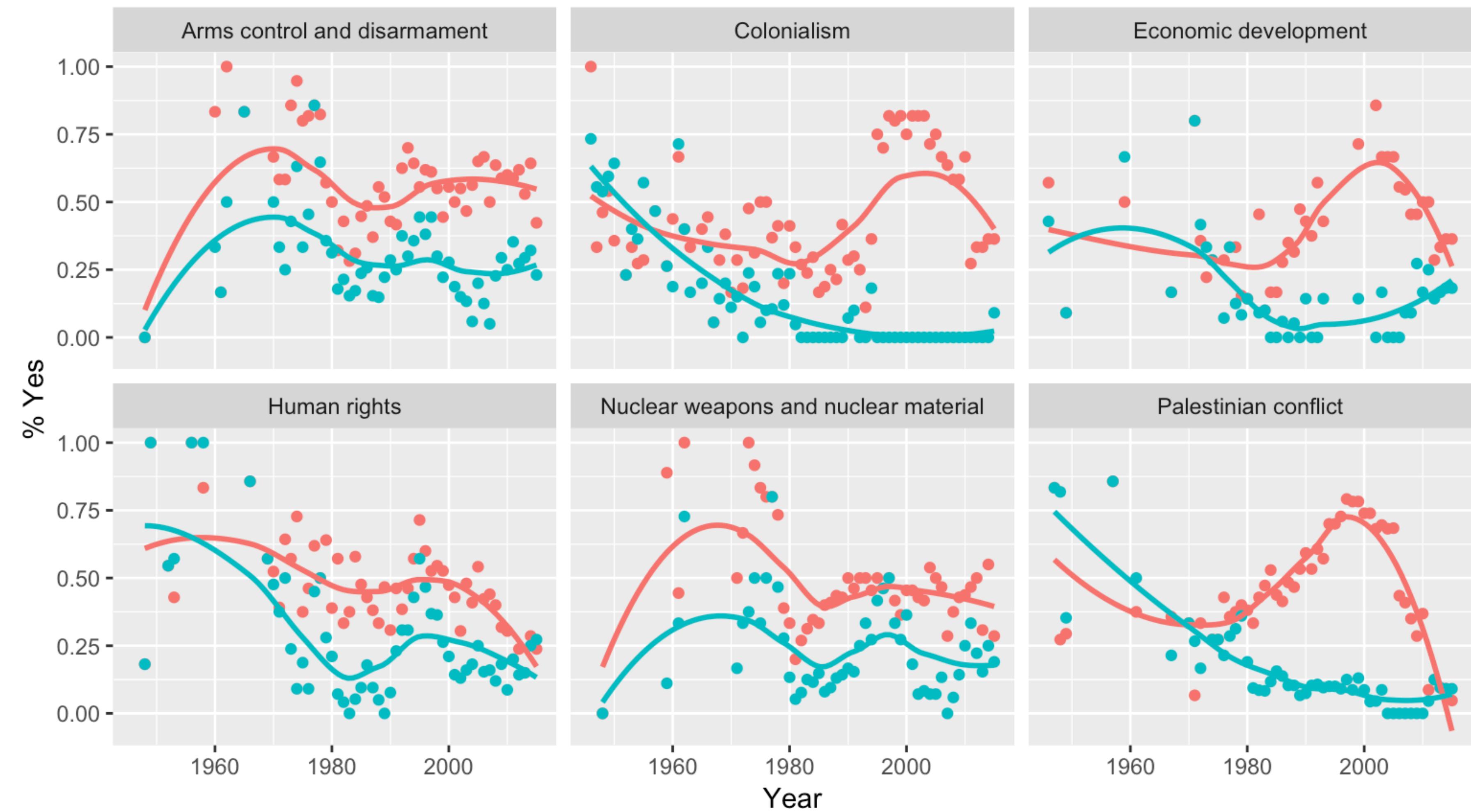


```
un_votes %>%  
filter(country %in% c("United States of America", "Turkey")) %>%  
inner_join(un_roll_calls, by = "rcid") %>%  
inner_join(un_roll_call_issues, by = "rcid") %>%  
group_by(country, year = year(date), issue) %>%  
summarize(  
  votes = n(),  
  percent_yes = mean(vote == "yes")  
) %>%  
filter(votes > 5) %>% # only use records where there are more than 5 votes  
ggplot(mapping = aes(x = year, y = percent_yes, color = country)) +  
  geom_point() +  
  geom_smooth(method = "loess", se = FALSE) +  
  facet_wrap(~ issue) +  
  labs(  
    title = "Percentage of 'Yes' votes in the UN General Assembly",  
    subtitle = "1946 to 2015",  
    y = "% Yes",  
    x = "Year",  
    color = "Country"  
)
```

```
un_votes %>%  
filter(country %in% c("United States of America", "Canada")) %>%  
inner_join(un_roll_calls, by = "rcid") %>%  
inner_join(un_roll_call_issues, by = "rcid") %>%  
group_by(country, year = year(date), issue) %>%  
summarize(  
  votes = n(),  
  percent_yes = mean(vote == "yes")  
) %>%  
filter(votes > 5) %>% # only use records where there are more than 5 votes  
ggplot(mapping = aes(x = year, y = percent_yes, color = country)) +  
  geom_point() +  
  geom_smooth(method = "loess", se = FALSE) +  
  facet_wrap(~ issue) +  
  labs(  
    title = "Percentage of 'Yes' votes in the UN General Assembly",  
    subtitle = "1946 to 2015",  
    y = "% Yes",  
    x = "Year",  
    color = "Country"  
)
```

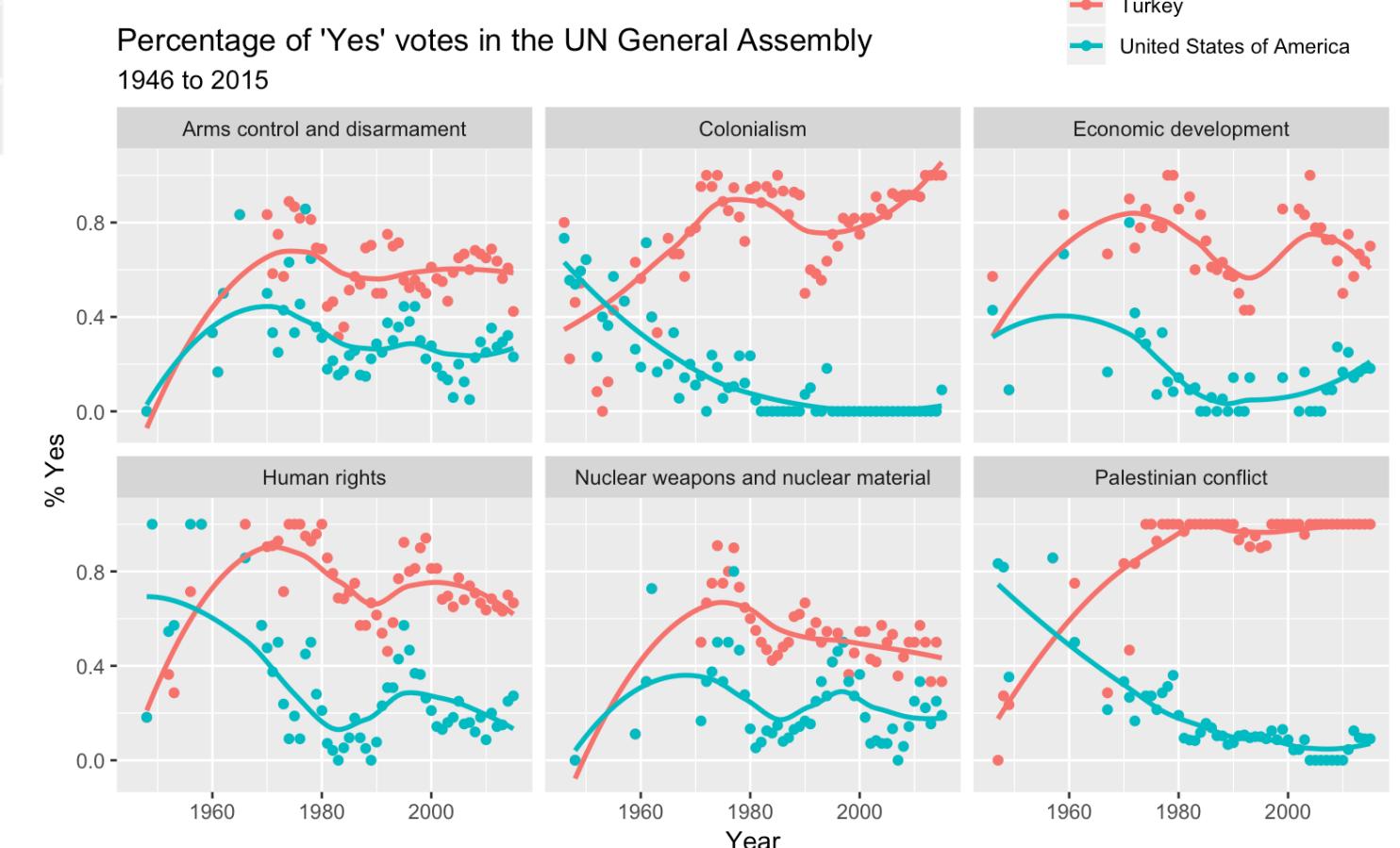
Percentage of 'Yes' votes in the UN General Assembly

1946 to 2015



Country

- Canada
- United States of America



cherish
day
one

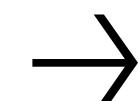




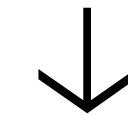
- ❑ Install R
- ❑ Install RStudio
- ❑ Install the following packages:
 - ❑ tidyverse
 - ❑ rmarkdown
 - ❑ ...
- ❑ Load these packages
- ❑ Install git



- ❑ Go to rstudio.cloud
 - ❑ Log in with your ID & pass
- > hello R!



R Studio®



Data
Analysis
References
Appendix

UN Votes

Mine Çetinkaya-Rundel

2018-09-26

Let's take a look at the voting history of countries in the United Nations General Assembly. We will be using data from the `unvotes` package. Additionally, we will make use of the `tidyverse` and `lubridate` packages for the analysis, and the `DT` package for interactive display of tabular output.

Data

The `unvotes` package provides three datasets we can work with: `un_roll_calls`, `un_roll_call_issues`, and `un_votes`. Each of these datasets contains a variable called `roid`, the roll call id, which can be used as a unique identifier to join them with each other.

- The `un_votes` dataset provides information on the voting history of the United Nations General Assembly. It contains one row for each country-vote pair.

`un_votes`

```
## # A tibble: 738,764 x 4
##   roid country      country_code vote
##   <int> <chr>        <chr>     <fct>
## 1 3 United States of America US      yes
## 2 3 Canada          CA      no
## 3 3 Cuba            CU      yes
## 4 3 Haiti           HT      yes
## 5 3 Dominican Republic DO      yes
## 6 3 Mexico          MX      yes
## 7 3 Guatemala       GT      yes
## 8 3 Honduras        HN      yes
## 9 3 El Salvador     SV      yes
## 10 3 Nicaragua       NI     yes
## # ... with 738,754 more rows
```

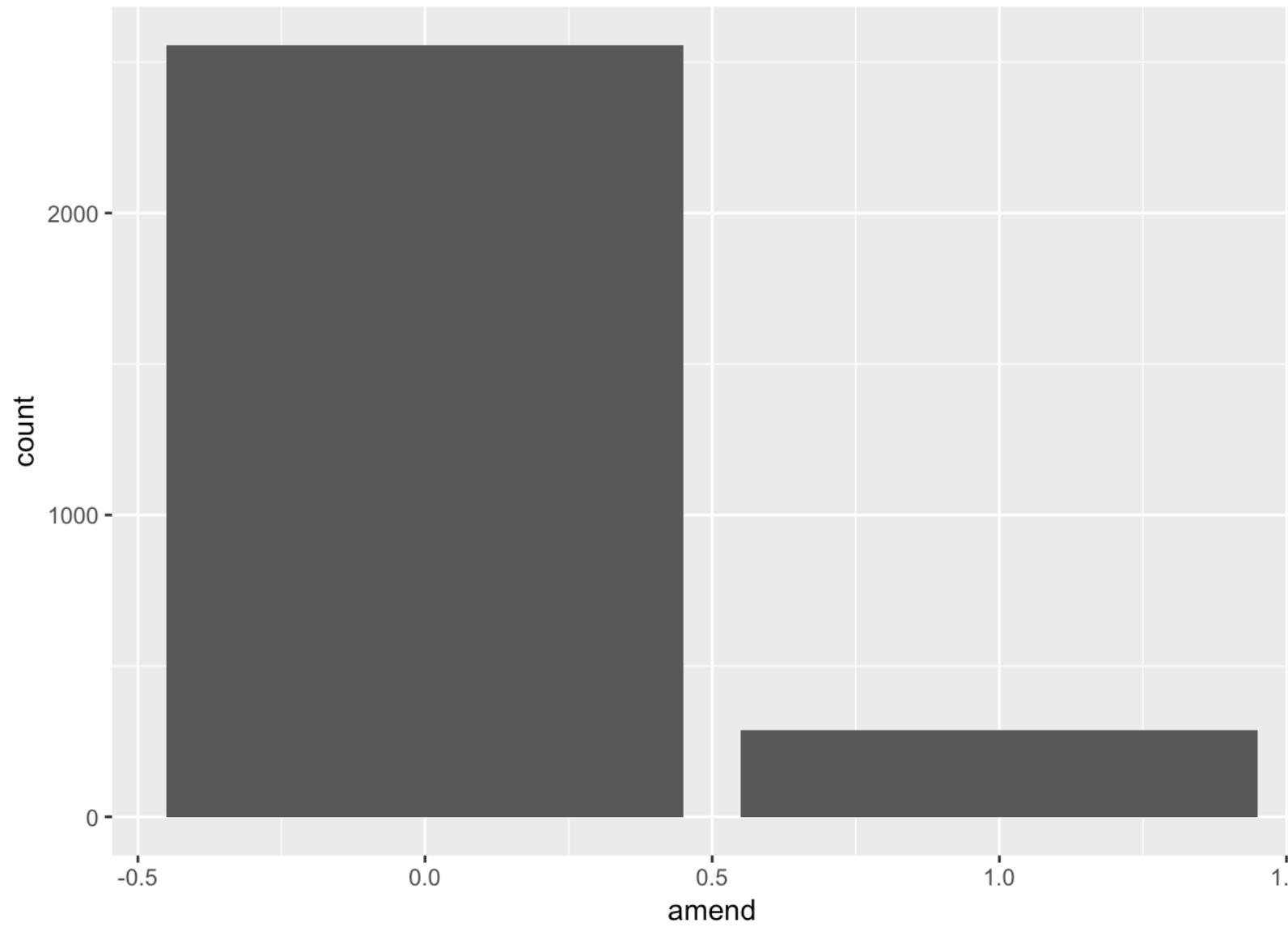
- The `un_roll_calls` dataset contains information on each roll call vote of the United Nations General Assembly.

`un_roll_calls`

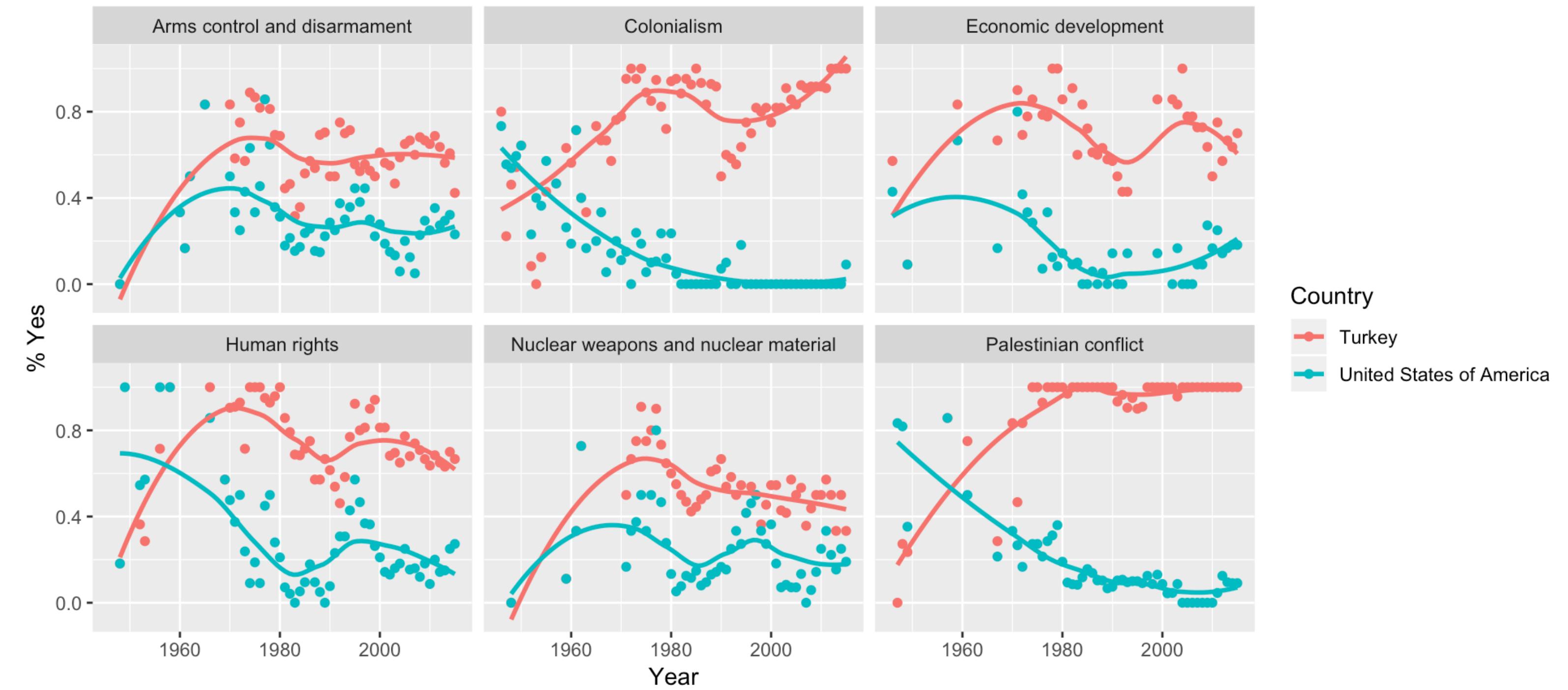
```
## # A tibble: 5,429 x 9
##   roid session importantvote date      unres amend para short descr
##   <int> <dbl> <dbl> <date>    <chr> <dbl> <dbl> <dbl> <chr>
## 1 3     1     1     0 1946-01-01 8/1/66    1     0 AMEN_ TO ADD_
## 2 4     1     1     0 1946-01-02 8/1/79    0     0 SECU_ TO ADD_
## 3 5     1     1     0 1946-01-04 8/1/98    0     0 VOTI_ TO AD_
```

**skip
baby
steps**

73



Percentage of 'Yes' votes in the UN General Assembly
1946 to 2015



How to draw an owl

1.



2.



1. Draw some circles

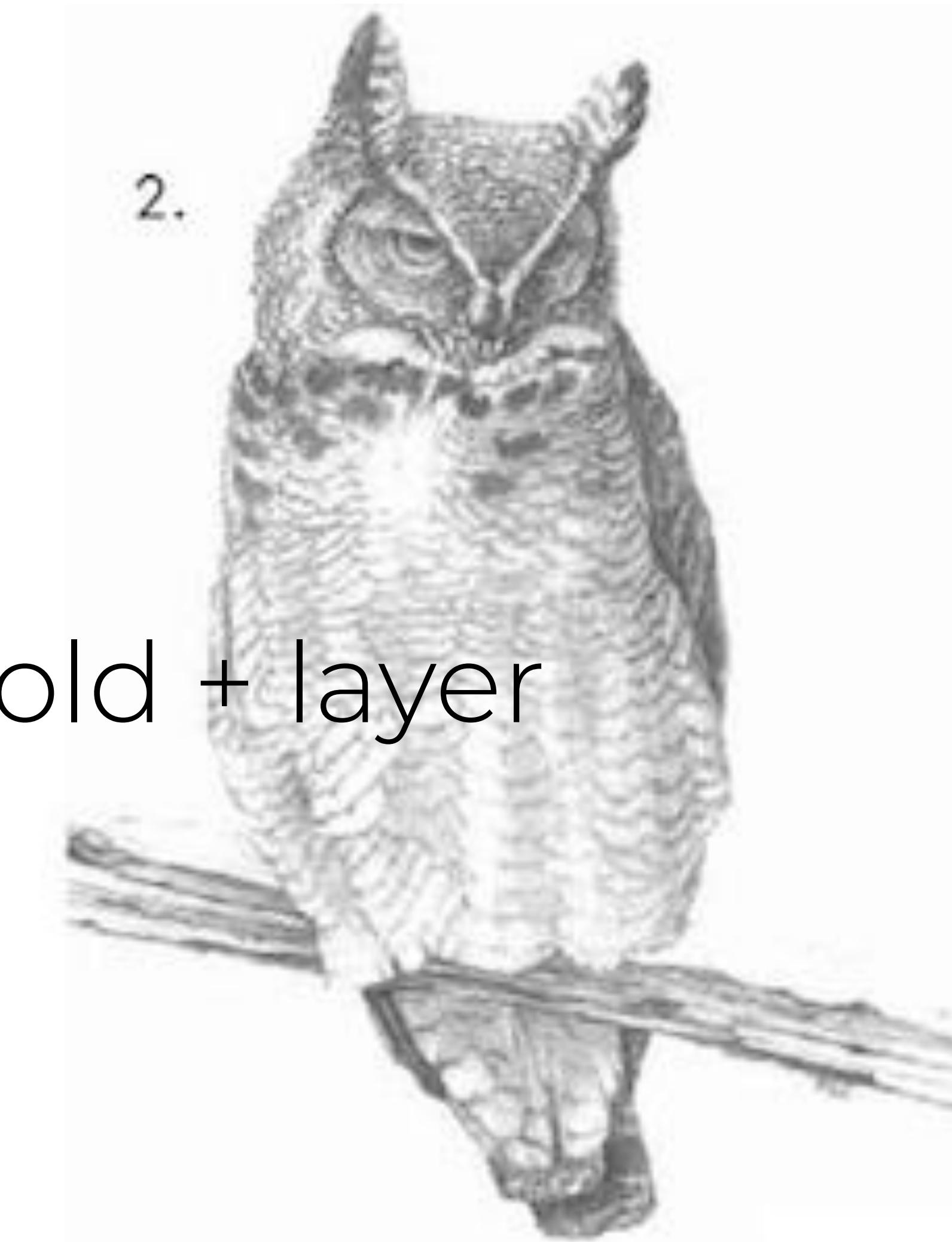
2. Draw the rest of the fucking owl

How to draw an owl

1.



2.

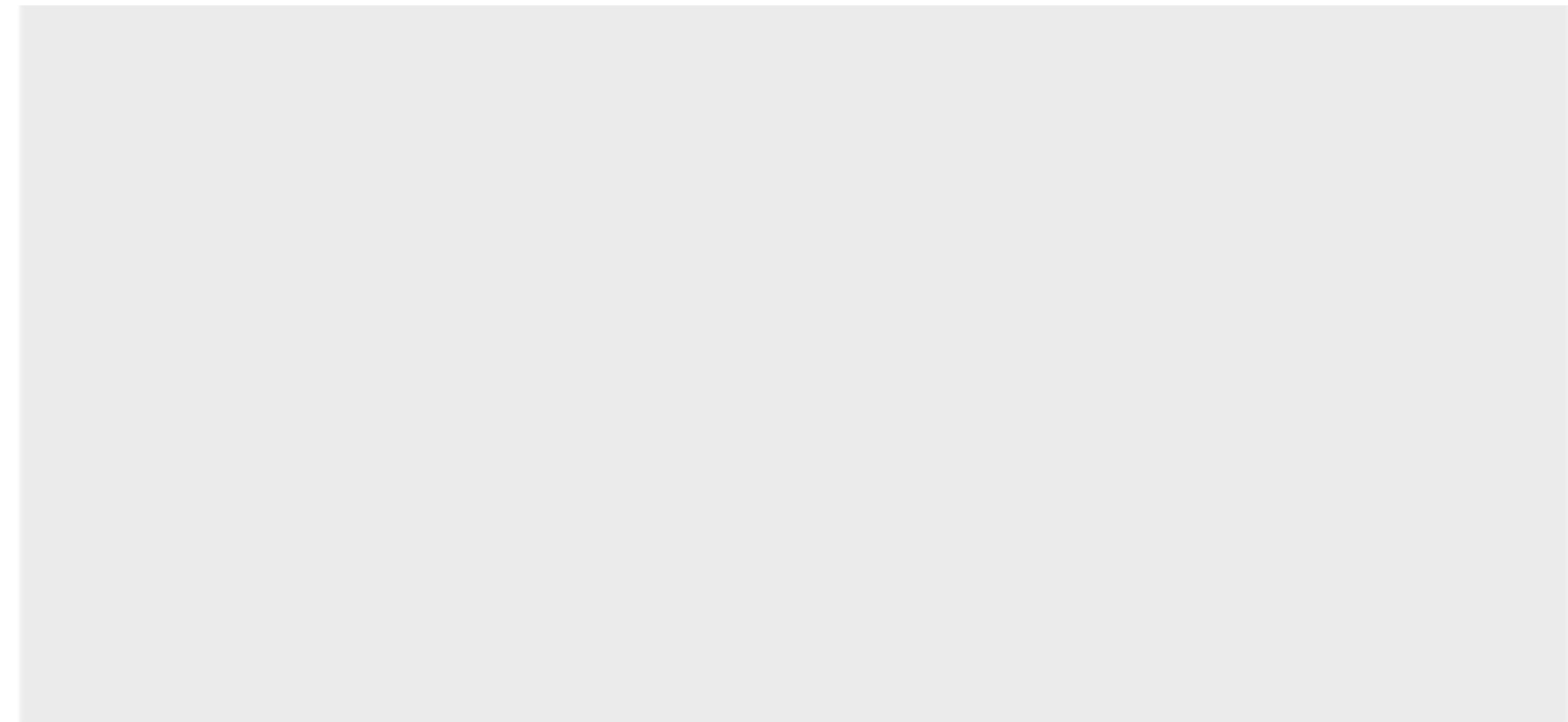


scaffold + layer

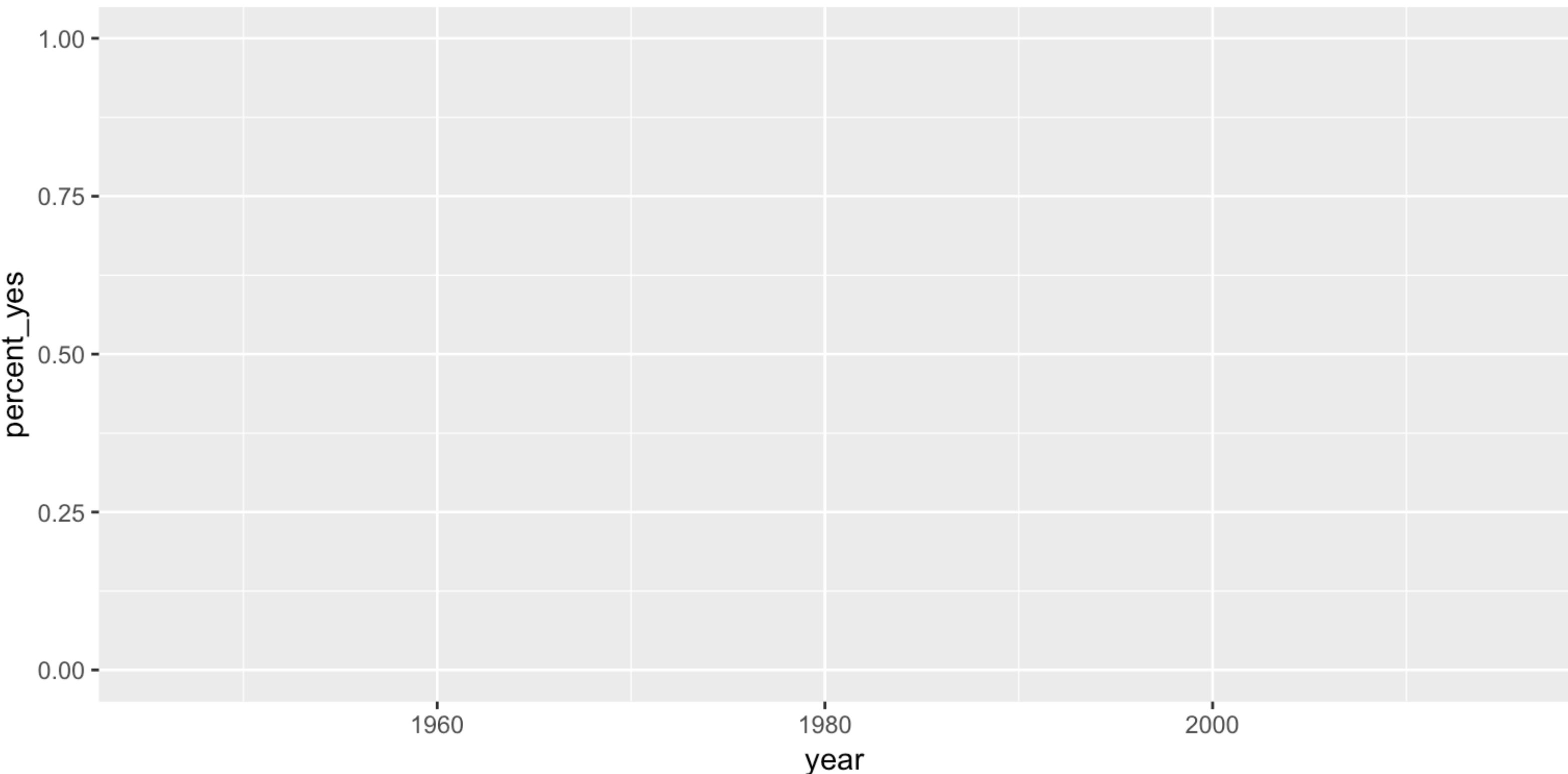
1. Draw some circles

2. Draw the rest of the fucking owl

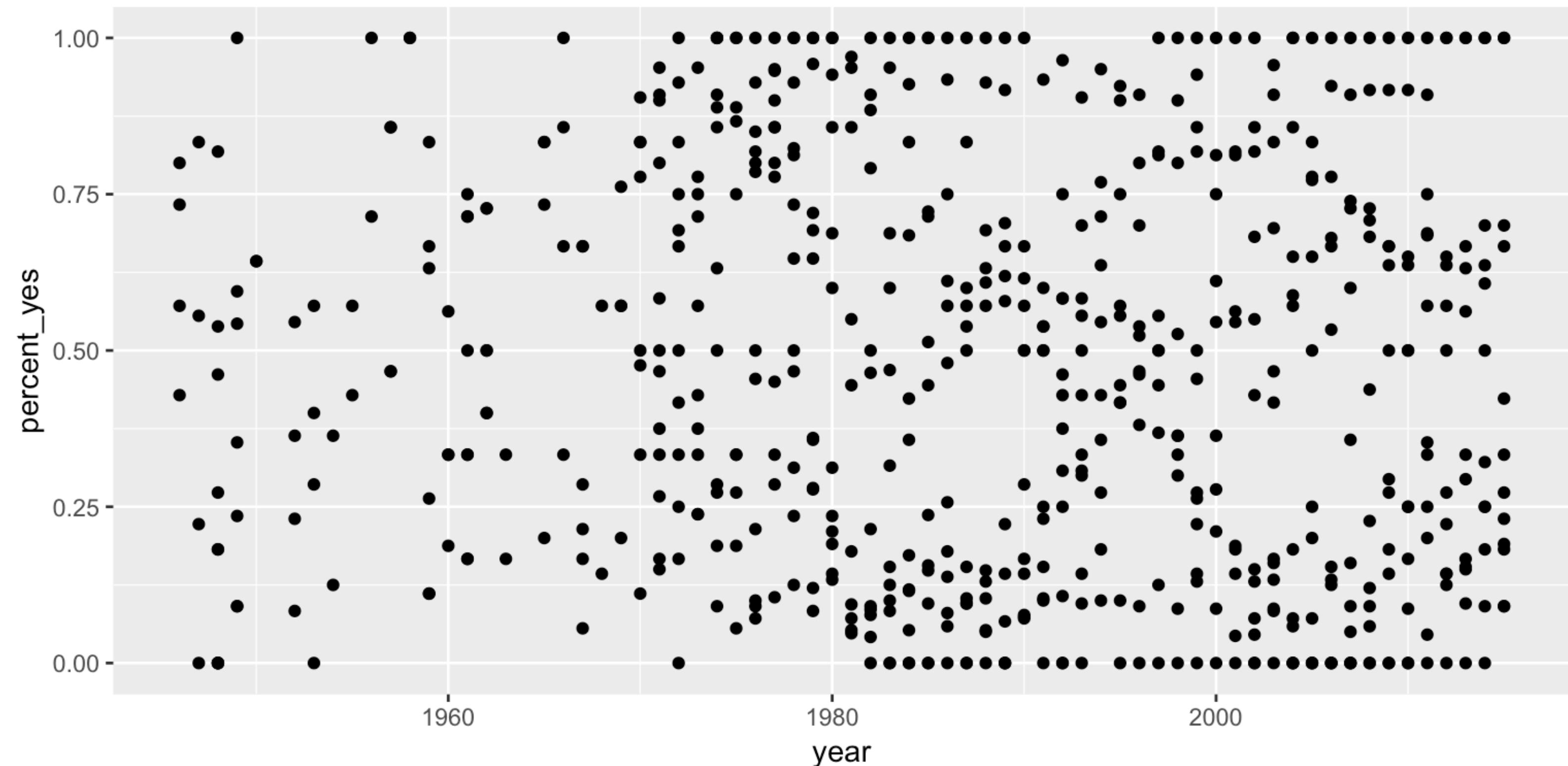
```
ggplot(data = un_votes_joined)
```



```
ggplot(data = un_votes_joined,  
       mapping = aes(x = year, y = percent_yes))
```



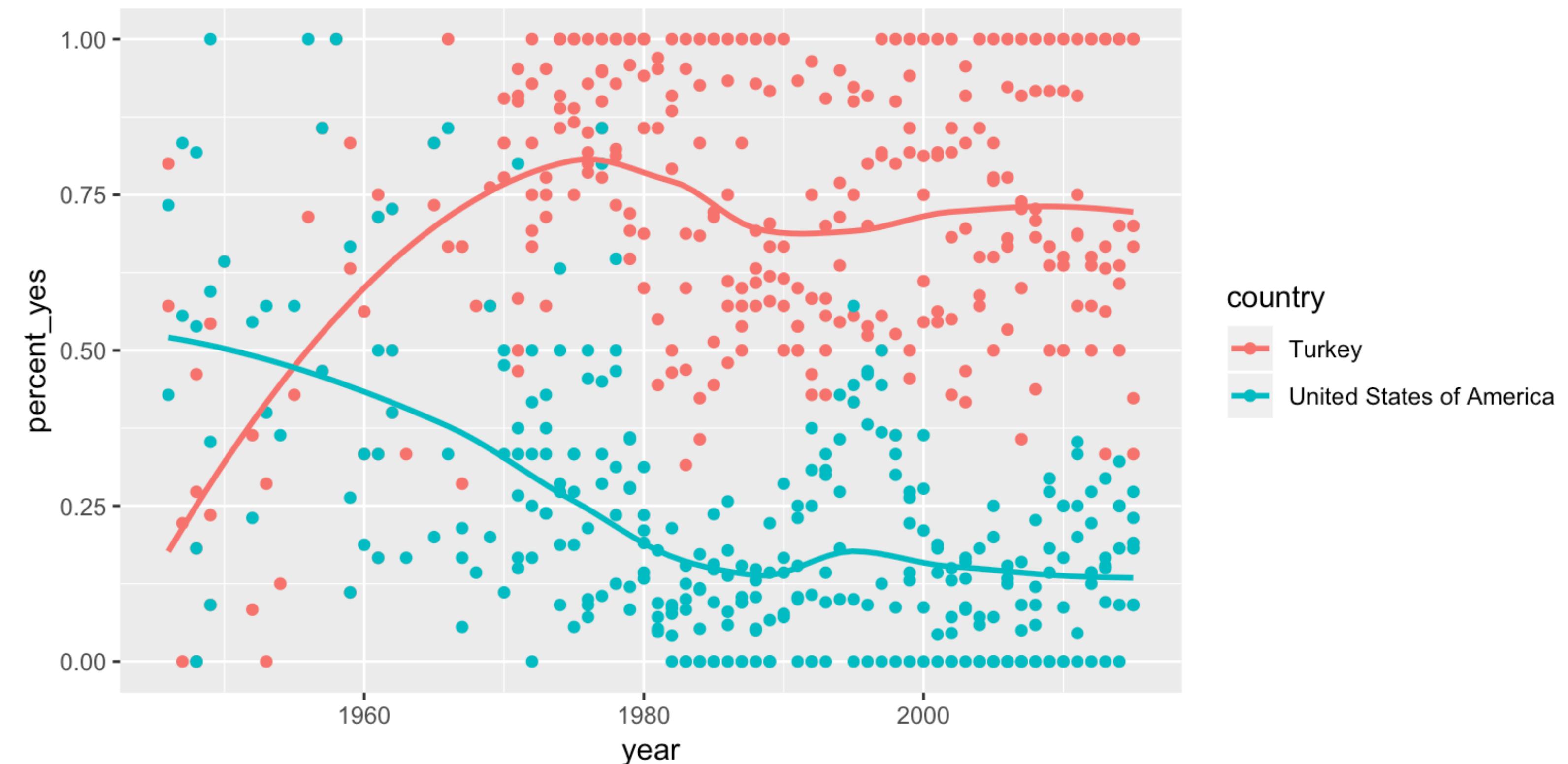
```
ggplot(data = un_votes_joined,  
       mapping = aes(x = year, y = percent_yes)) +  
  geom_point()
```



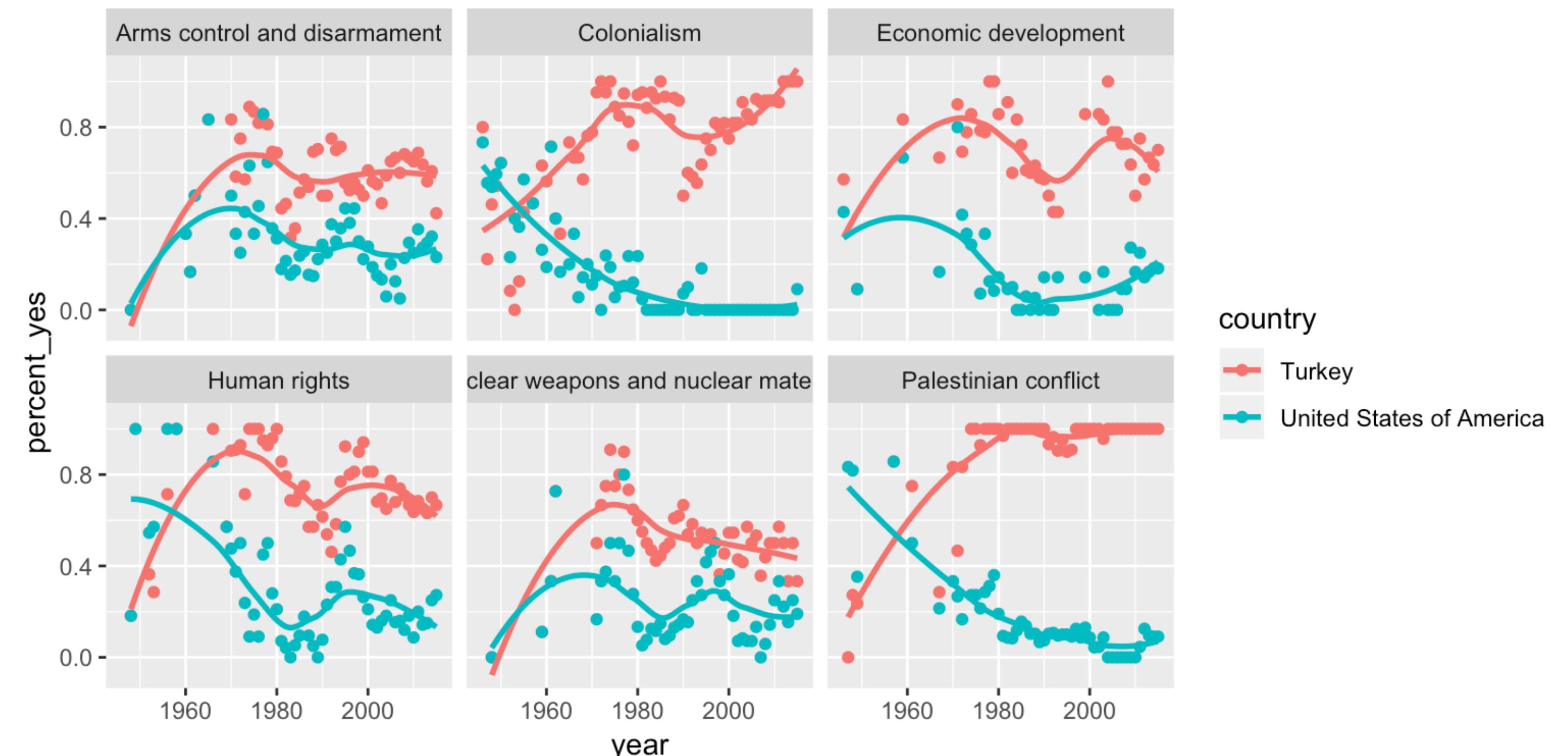
```
ggplot(data = un_votes_joined,  
       mapping = aes(x = year, y = percent_yes, color = country)) +  
geom_point()
```



```
ggplot(data = un_votes_joined,  
       mapping = aes(x = year, y = percent_yes, color = country)) +  
  geom_point() +  
  geom_smooth(method = "loess", se = FALSE)
```



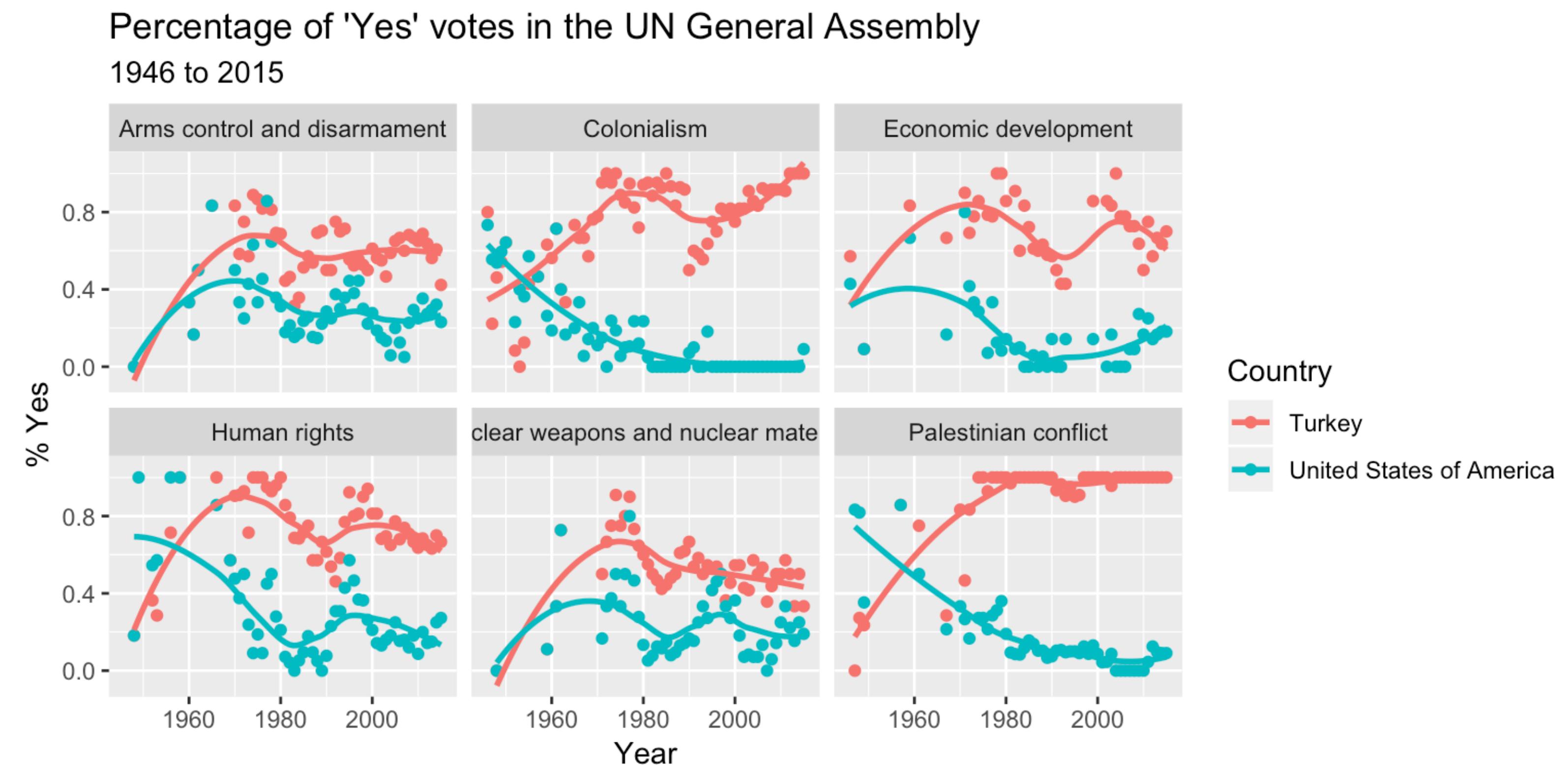
```
ggplot(data = un_votes_joined,  
       mapping = aes(x = year, y = percent_yes, color = country)) +  
  geom_point() +  
  geom_smooth(method = "loess", se = FALSE) +  
  facet_wrap(~ issue)
```



```

ggplot(data = un_votes_joined,
       mapping = aes(x = year, y = percent_yes, color = country)) +
  geom_point() +
  geom_smooth(method = "loess", se = FALSE) +
  facet_wrap(~ issue) +
  labs(
    title = "Percentage of 'Yes' votes in the UN General Assembly",
    subtitle = "1946 to 2015",
    y = "% Yes",
    x = "Year",
    color = "Country"
)

```



hide
the
veggies





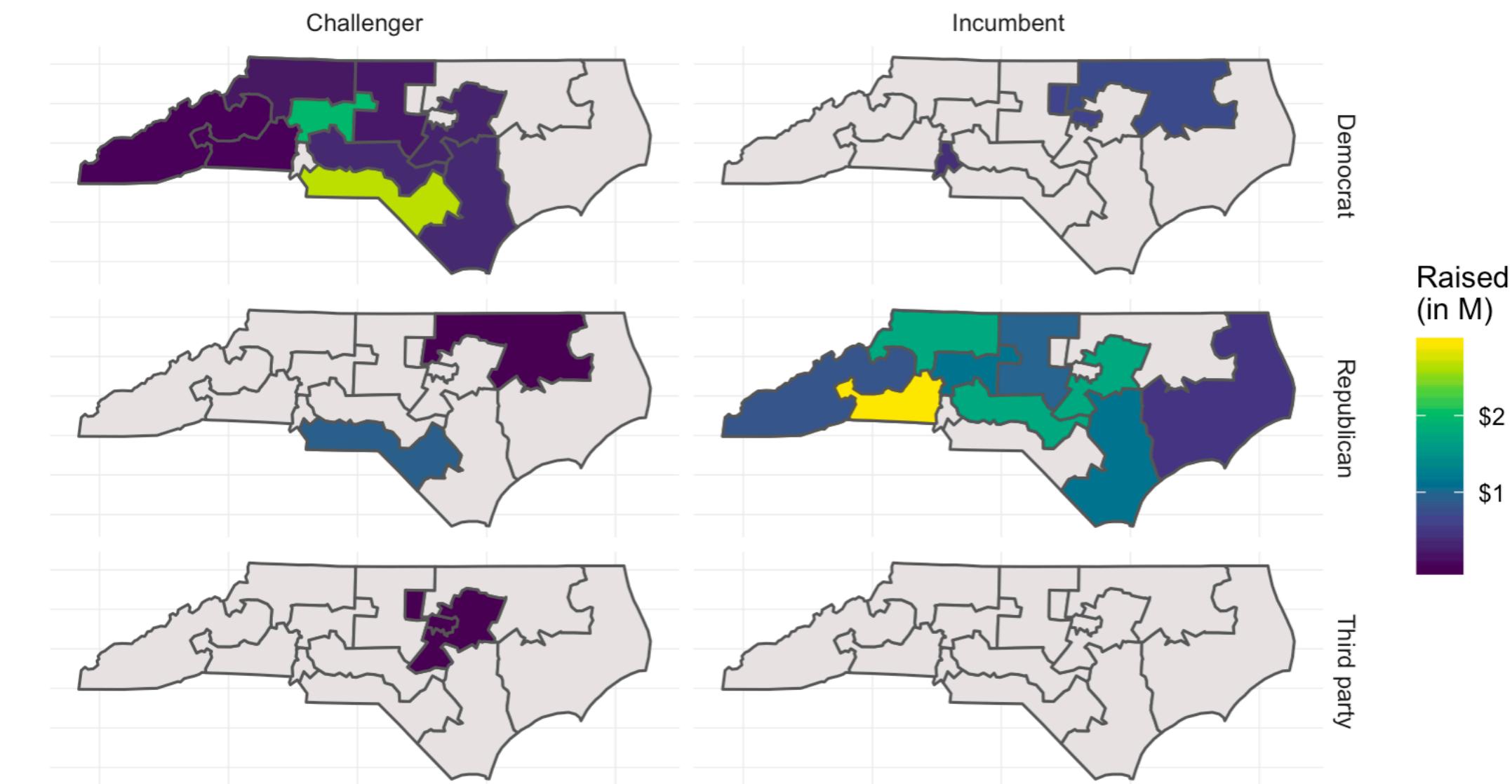
- Topic: Web scraping
- Tools:
 - **rvest**
 - regular expressions



□ Today we start with this:

□ and end with this:

Political contributions for 2018 NC Congressional Races
as of 9/30/2018

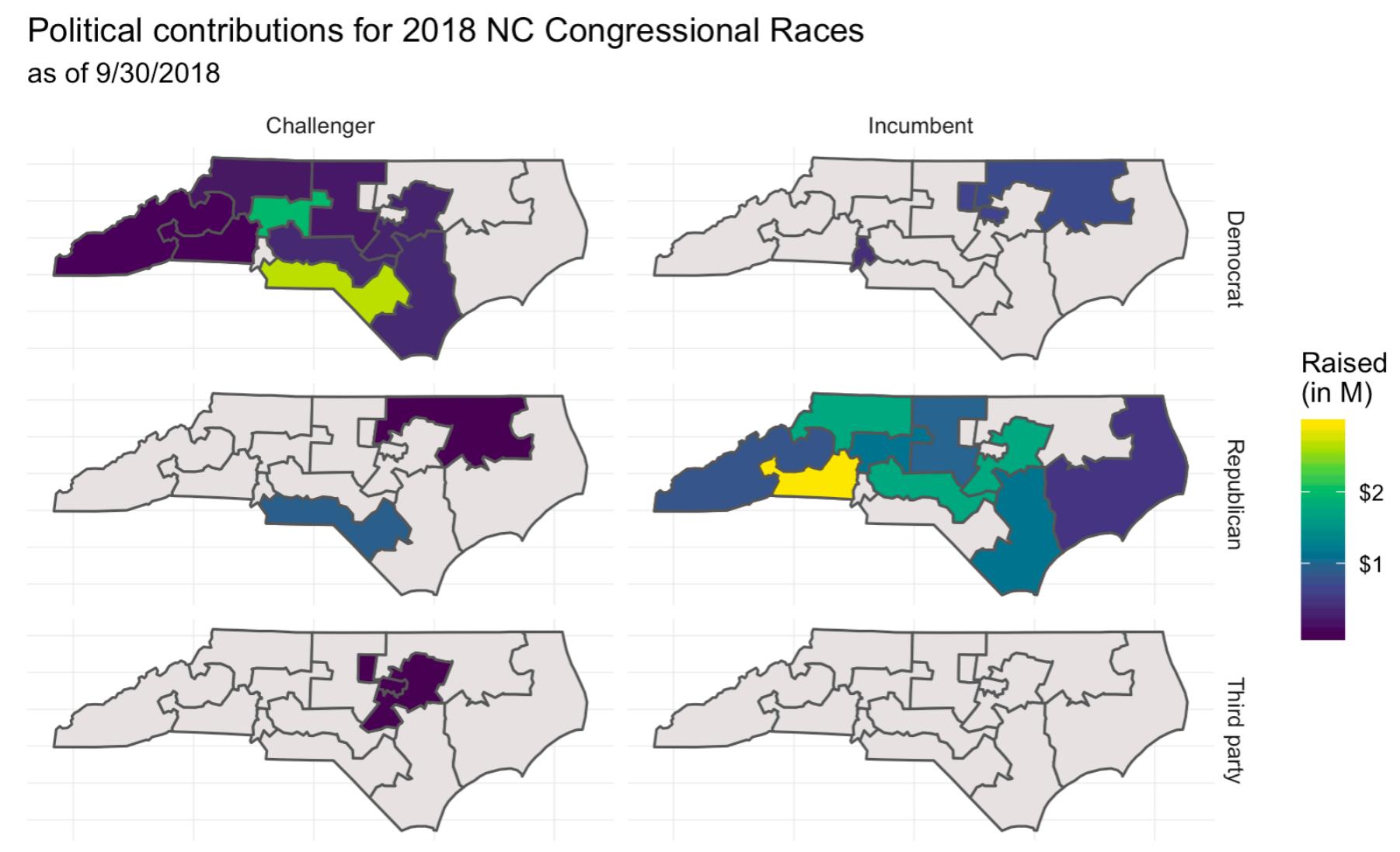


- **Ex 1:** Scrape the table off the web and save as a data frame.
-
- **Ex 2:** What other information do we need represented as variables in the data to obtain the desired facets?

Candidate	Raised	Spent	Cash on Hand	Last Report
G K Butterfield (D) • Incumbent	\$714,219	\$797,700	\$560,416	10/17/2018
Roger Allison (R)	\$28,314	\$27,817	\$497	10/17/2018

to

▲	candidate_info	raised	spent	cash_on_hand	last_report	race
1	G K Butterfield (D) • Incumbent	714219	797700	560416	2018-10-17	North Carolina District 01
2	Roger Allison (R)	28314	27817	497	2018-10-17	North Carolina District 01



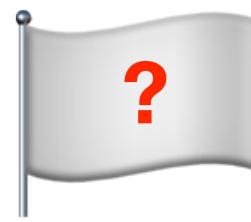
- **Lesson:** “Just enough” string parsing and regular expressions to go from

▲	candidate_info
1	G K Butterfield (D) • Incumbent
2	Roger Allison (R)

▲	candidate_name	party	status
1	G K Butterfield	Democrat	Incumbent
2	Roger Allison	Republican	Challenger

leverage
the
ecosystem





```
t.test(mtcars$mpg ~ mtcars$vs)

# Welch Two Sample t-test
# data: mtcars$mpg by mtcars$vs
# t = -4.6671, df = 22.716,
# p-value = 0.0001098
# alternative hypothesis:
# true difference in means is not equal to 0
# 95 percent confidence interval:
# -11.462508 -4.418445
# sample estimates:
# mean in group 0 mean in group 1
# 16.61667 24.55714
```



```
library(tidyverse)
```

```
library(infer)
```

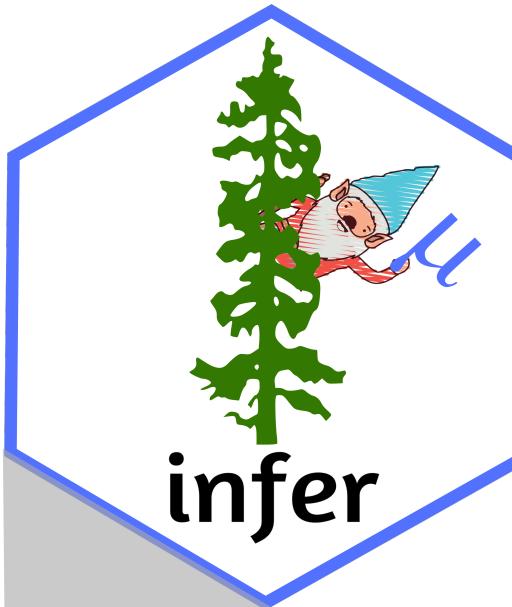
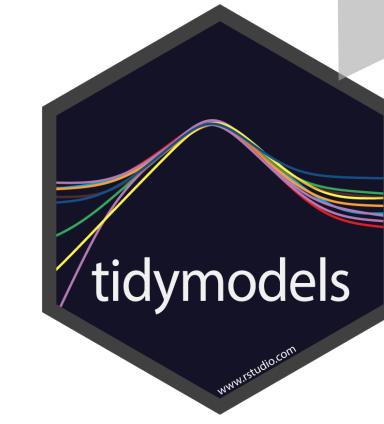
```
mtcars %>%
```

```
  mutate(vs = as.factor(vs)) %>%
  specify(mpg ~ vs) %>%
  generate(reps = 1000,
            type = "bootstrap") %>%
  calculate(stat = "diff in means",
            order = c("0", "1")) %>%
```

```
summarise(
```

```
  l = quantile(stat, 0.025),
  u = quantile(stat, 0.975)
)
```

```
#      l      u
# -11.1 -4.85
```





start with cake



cherish day one



skip baby steps



hide the veggies

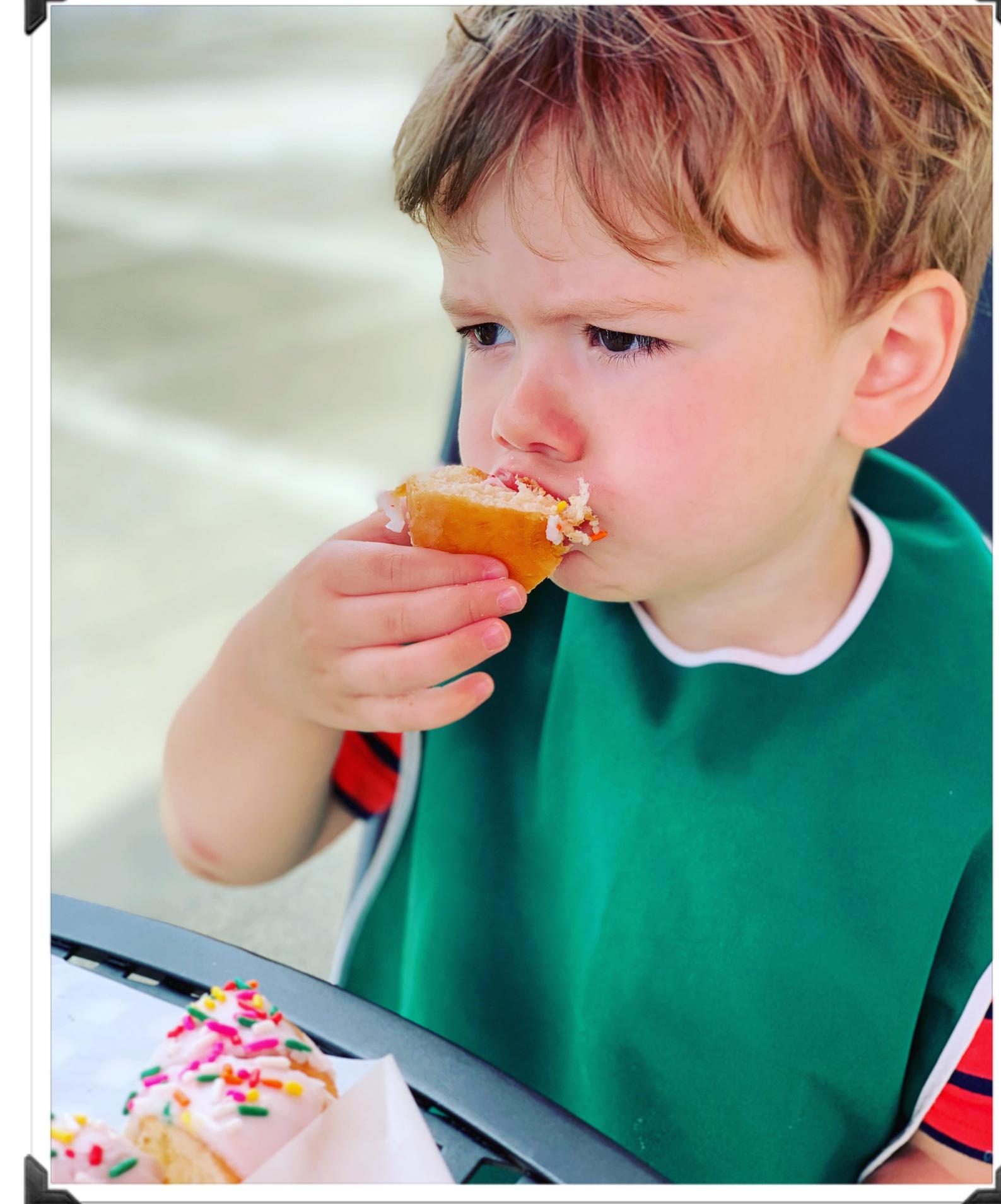


leverage the ecosystem

Let them
eat cake
(first)!*

↳ rstd.io/eat-cake-5min

* You can tell
them all about the
ingredients later!



@minebocek
mine-cetinkaya-rundel
mine@rstudio.com

