

Lab 3b - UN Votes

Sebastian Montesinos

For class Thursday, Feb. 17

Lab Purpose

In the first lab activity, you examined a visualization that captured how the voting record of three randomly selected countries changed over time on a variety of issues. We will revisit the UN voting record data again in this lab, with a focus on learning how to change features such as shapes, colors, and line types with the **ggplot2** package.

You'll work on the lab in the company of classmates to help each other and share/see results with different options selected.

Data

The **unvotes** package provides three datasets that capture the voting history of countries in the United Nations General Assembly: **un_roll_calls**, **un_roll_call_issues**, and **un_votes**. Each of these datasets contains a variable called **rcid**, the roll call id, which can be used as a unique identifier to join the three datasets together.

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.

```
# head shows the first 6 observations by default  
# what does including the 4 do?  
head(un_votes, 4)
```

```
# A tibble: 4 x 4  
  rcid country      country_code vote  
  <dbl> <chr>         <chr>      <fct>  
1     3 United States US         yes  
2     3 Canada        CA         no  
3     3 Cuba          CU         yes  
4     3 Haiti         HT         yes
```

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

```
head(un_roll_calls, 4)
```

```
# A tibble: 4 x 9  
  rcid session importantvote date      unres  amend para short      descr
```

	<int>	<dbl>		<int>	<date>	<chr>	<int>	<int>	<chr>	<chr>
1	3	1	0	1946-01-01	R/1/66	1	0	AMENDMENTS, ~	"TO ~	
2	4	1	0	1946-01-02	R/1/79	0	0	SECURITY COU~	"TO ~	
3	5	1	0	1946-01-04	R/1/98	0	0	VOTING PROCE~	"TO ~	
4	6	1	0	1946-01-04	R/1/107	0	0	DECLARATION ~	"TO ~	

The `un_roll_call_issues` dataset contains (topic) classifications of roll call votes of the United Nations General Assembly. Many votes had no topic, and some have more than one.

```
head(un_roll_call_issues, 4)
```

```
# A tibble: 4 x 3
  rcid short_name issue
<int> <chr>      <fct>
1    77 me      Palestinian conflict
2   9001 me      Palestinian conflict
3   9002 me      Palestinian conflict
4   9003 me      Palestinian conflict
```

1 - Data prep

The code below prepares our data in several ways (you saw most of this code previously).

First, it combines our three datasets into one. Then, it limits the dataset to focus only on one of the six issues (“Human Rights”) and three of the countries. Finally, some wrangling is done so that we are only using records where there are more than 5 votes on an issue. You will be learning the details of the wrangling commands in the coming weeks. For now, just trust this does as specified.

Update the code below to select three countries of interest to you.

The country names should be spelled and capitalized exactly the same way as they appear in the dataset. A full list of the countries is provided in the UN country list at the end of this lab. The interactive data table is created by the *DT* package.

```
# ready data for plotting
unvotes_hr <- un_votes %>%
  inner_join(un_roll_calls, by = "rcid") %>%
  inner_join(un_roll_call_issues, by = "rcid") %>%
  filter(issue == "Human rights",
         country %in% c("San Marino",
                        "Iceland",
                        "Greece")) %>%
  group_by(country, year = year(date), issue) %>%
  summarize(votes = n(),
            percent_yes = mean(vote == "yes")) %>%
  filter(votes > 5)
```

This new data set is called `unvotes_hr` where the hr is for Human Rights, the issue we are focusing on.

Coding Details

What might be a reasonable name for a new data set if you wanted to focus on the issue: Economic development?

Solution: `unvotes_ed`

Saving the data set is different than what was done in the previous lab. There, since we only made one visual, we didn't really need to save it. Saving it is useful if we expect to re-use it several times.

What piece of the code does the assignment/saving of the new data set?

Solution: `unvotes_hr <- un_votes`

The `%>%` set of symbols occurs often here. Do you know what this symbol is? If so, what is it?

Solution: It carries the result from the line above as the argument for the next line.

(This will come up more with wrangling, but since it's here now, we can discuss it.)

The `issue == "Human Rights"` line pulls out only that issue. What do you think changing the `==` to `!=` would do?

Solution: This would change it to a 'not' operator that says to include all issues except human rights.

(If you change this to check, be sure to put it back to `==` before continuing and re-run the chunk.)

Code options - What would setting `eval = FALSE` in the code chunk do? Would this be wise if we plan to use the data set for visuals in the rest of the document that we want to show in the knitted pdf?

Solution: This would make it so the code chunk does not show in the knitted pdf. Since we will be using this code for the rest of the lab, having it display at the start of the PDF is a good idea.

2 - Shapes and sizes

From the reading you saw that ggplot2 objects start with choosing the data set and setting the aesthetics (which can range from x and y variables to color and beyond). After that, you need to add a *geom* and a *stat* to make a layer in the plot. Geoms describe the geometric objects being plotted and stats are the statistical transformation to be used.

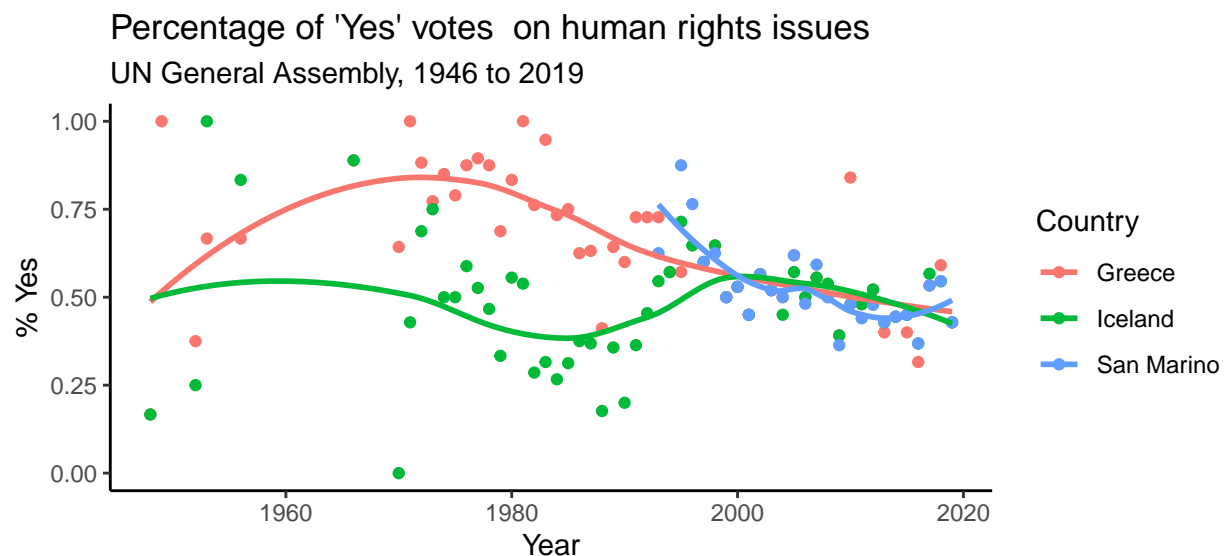
stat? Professor, we didn't see *stat* in the reading or in your examples! Every *geom* has a default *stat* (and vice versa), so you only need to specify one of them. For most of what we'll be doing, the *stat* is the identity function. That is, we plot the data as is. Thus, you really need to only worry about the *geoms* for now.

Many different *geoms* were explored in the reading. For scatterplots, we tend to use `geom_point` as a starting point.

The default symbol for `geom_point()` is a point. But you can change the symbol shape and size using the `shape =` and `size =` options, respectively, within the `geom_point()` function. Use the code below as a starting point to modify the plot in the questions that follow. Copy/paste the code so you can compare without having to edit.

```
# plot the data
ggplot(data = unvotes_hr, mapping = aes(x = year, y = percent_yes,
                                         color = country)) +

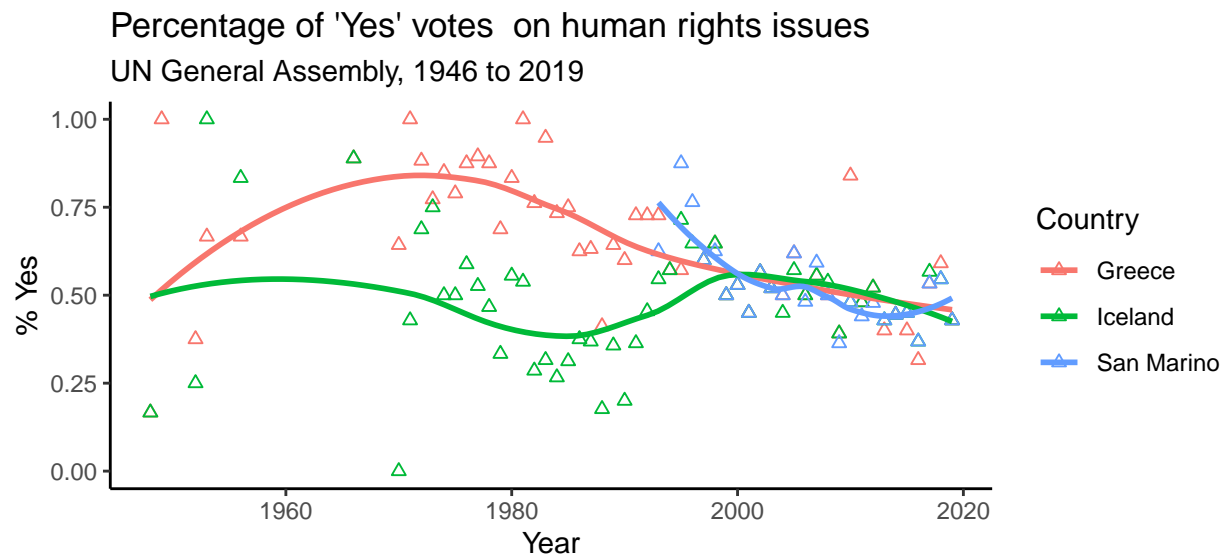
  geom_point() +
  geom_smooth(method = "loess", se = FALSE) +
  labs(title = "Percentage of 'Yes' votes on human rights issues",
       subtitle = "UN General Assembly, 1946 to 2019",
       y = "% Yes",
       x = "Year",
       color = "Country")
```



Try adding `shape = 2` to the `geom_point()` function. What happens?

Solution: The data points become triangles instead of circles

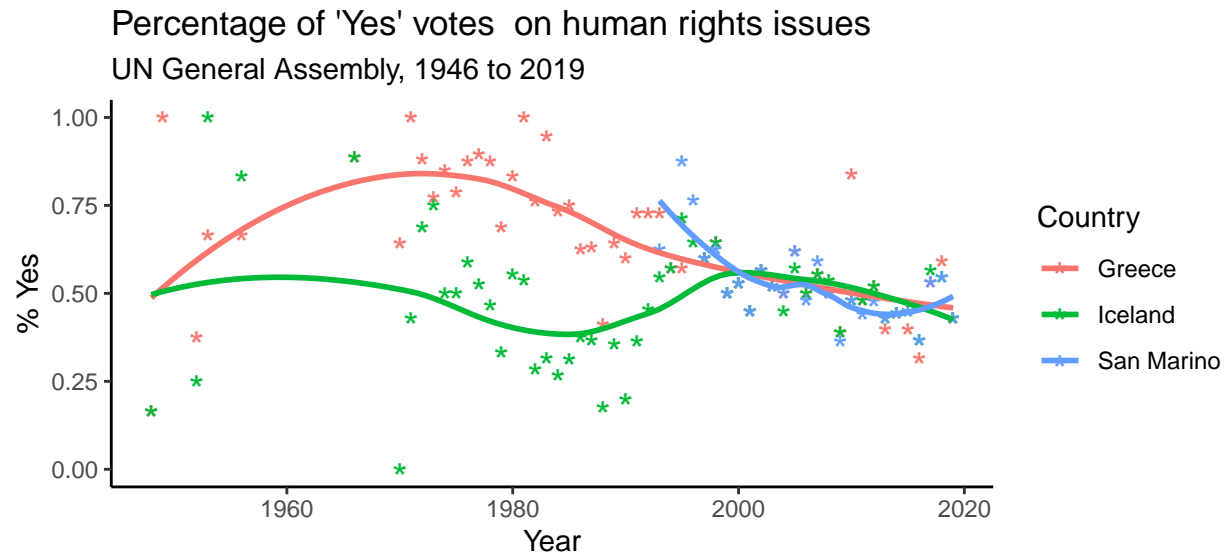
```
ggplot(data = unvotes_hr, mapping = aes(x = year, y = percent_yes,
                                         color = country)) +
  geom_point(shape = 2) +
  geom_smooth(method = "loess", se = FALSE) +
  labs(title = "Percentage of 'Yes' votes on human rights issues",
        subtitle = "UN General Assembly, 1946 to 2019",
        y = "% Yes",
        x = "Year",
        color = "Country")
```



What happens if you add `shape = "*", size = 5` to `geom_point()` instead?

Solution: Changes the shape of the data points and their size.

```
ggplot(data = unvotes_hr, mapping = aes(x = year, y = percent_yes,
                                         color = country)) +
  geom_point(shape = "*", size = 5) +
  geom_smooth(method = "loess", se = FALSE) +
  labs(title = "Percentage of 'Yes' votes on human rights issues",
        subtitle = "UN General Assembly, 1946 to 2019",
        y = "% Yes",
        x = "Year",
        color = "Country")
```



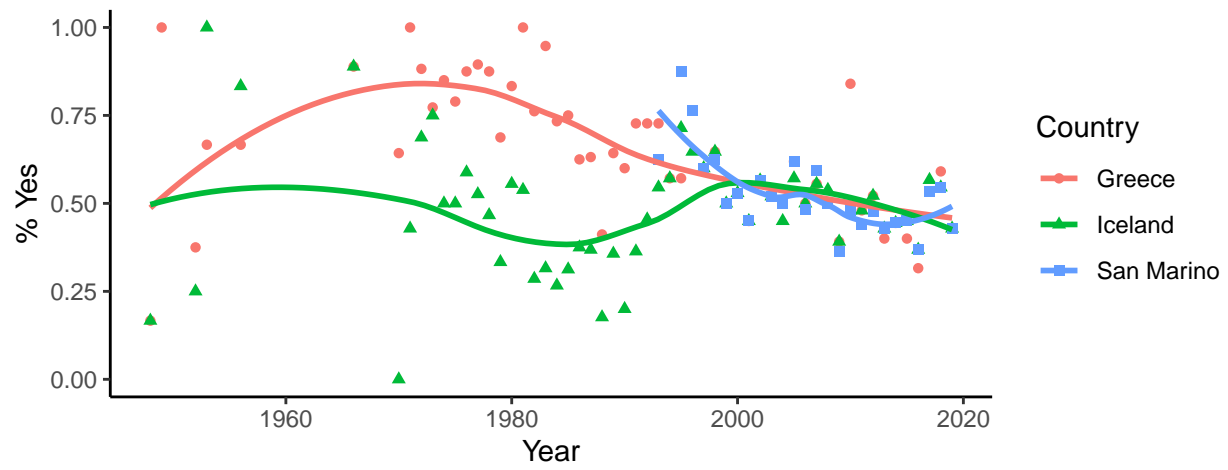
You can also specify different shapes for the different countries. Since country is a variable in our dataset, we can use the `aes()` function within `geom_point()` to specify different shapes for different countries.

See if you can figure out the correct syntax to use different shapes for your selected countries. Can you also figure out how to clean up the legend? (*Hint*: add something to the `labs()` function.)

Solution:

```
ggplot(data = unvotes_hr, mapping = aes(x = year, y = percent_yes,
                                         color = country)) +
  geom_point(aes(shape = country)) +
  geom_smooth(method = "loess", se = FALSE) +
  labs(title = "Percentage of 'Yes' votes on human rights issues",
       subtitle = "UN General Assembly, 1946 to 2019",
       y = "% Yes",
       x = "Year",
       color = "Country",
       shape = "Country")
```

Percentage of 'Yes' votes on human rights issues
UN General Assembly, 1946 to 2019



3 - Line types

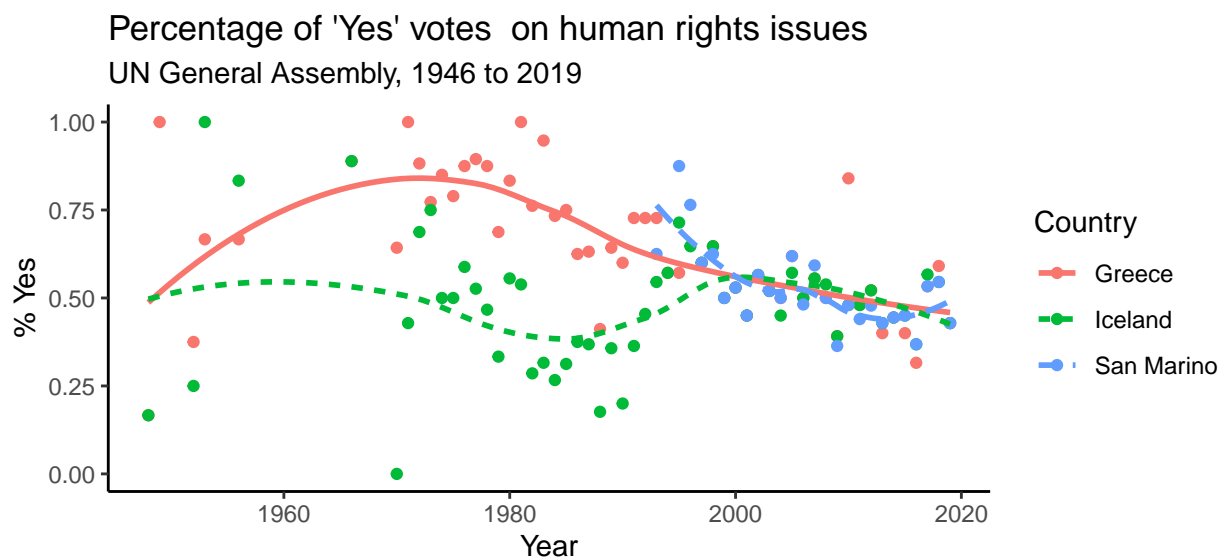
Often, we want to examine variable relationships and follow them with estimated best fitting curves or regression lines. To do that, we use `geom_smooth()`. It can do both smoothed lines (LOESS fits) or regression lines. It does smoothed lines by default (not regression).

The default line type for `geom_smooth()` is a solid line. You can change the line type and thickness using the `lty =` (or write out `linetype`) and `size =` options, respectively, within the `geom_smooth()` function. Copy and paste the code from your last figure as a starting point below.

Remove the different shapes for the countries. Then, update the figure using the `aes()` function within `geom_smooth()` so that each country has a different line type. Can you also figure out how to clean up the legend? (We removed the shapes so you can see the linetypes in the legend better.)

Solution:

```
ggplot(data = unvotes_hr, mapping = aes(x = year, y = percent_yes,
                                         color = country)) +
  geom_point() +
  geom_smooth(method = "loess", se = FALSE, (aes(lty = country))) +
  labs(title = "Percentage of 'Yes' votes on human rights issues",
       subtitle = "UN General Assembly, 1946 to 2019",
       y = "% Yes",
       x = "Year",
       color = "Country",
       lty = "Country")
```

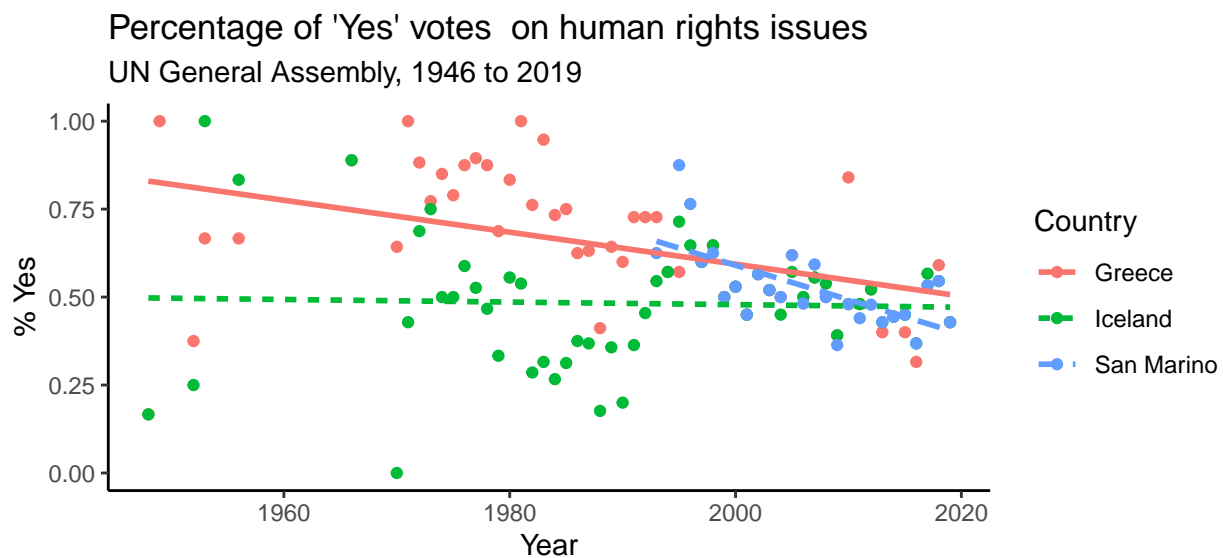


The `geom_smooth` function also allows you to fit regression lines. You have to set the method to “lm”.

Add regression lines instead of smoothed lines to the plot. Keep different line types. Does regression seem appropriate?

Solution: There is not a strong association between year and the % of yes vote so regression would not be appropriate.

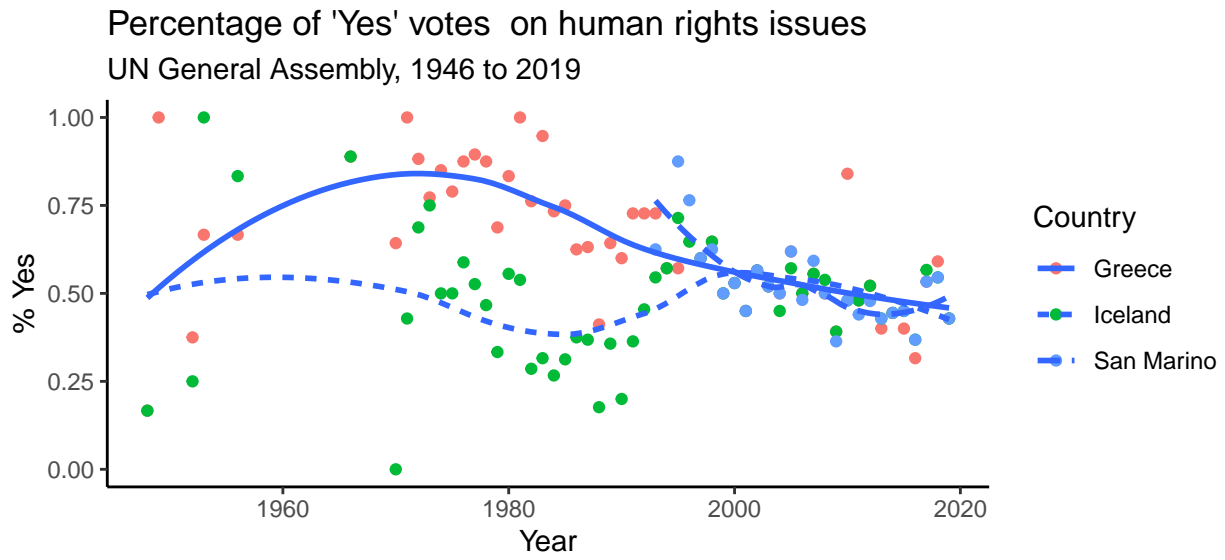

```
ggplot(data = unvotes_hr, mapping = aes(x = year, y = percent_yes,
                                         color = country)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, (aes(lty = country))) +
  labs(title = "Percentage of 'Yes' votes on human rights issues",
       subtitle = "UN General Assembly, 1946 to 2019",
       y = "% Yes",
       x = "Year",
       color = "Country",
       lty = "Country")
```



Return to smoothed lines without error bars, removing the regression lines. Here, we explore where the color aesthetic is set. Set the color aesthetic in the `geom_point` call, not in the overall aesthetic. Is the overall plot still as useful?

Solution: It is not as useful because in this case the color aesthetic is only applied to the points instead of both the points and the lines.

```
ggplot(data = unvotes_hr, mapping = aes(x = year, y = percent_yes)) +
  geom_point(aes(color = country)) +
  geom_smooth(method = "loess", se = FALSE, (aes(lty = country))) +
  labs(title = "Percentage of 'Yes' votes on human rights issues",
       subtitle = "UN General Assembly, 1946 to 2019",
       y = "% Yes",
       x = "Year",
       color = "Country",
       lty = "Country")
```



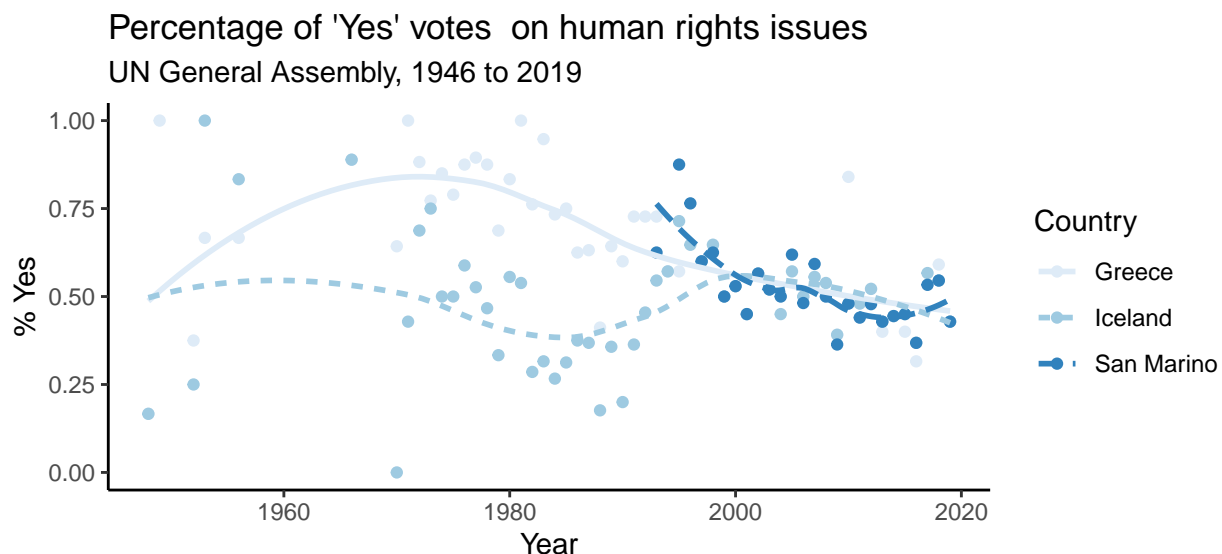
4 - Colors

There are many different ways to change the colors of points, lines, etc. in `ggplot()`. Today we'll explore just a few of them. We'll change colors both manually and using *color brewer*, which provides color schemes designed by a professional to help people choose good color schemes for their graphs.

Copy your smoothed lines by country code from the last exercise (first plot from that section). Add a layer `scale_color_brewer()` before the `labs()` layer (don't forget to add a plus sign). What happens?

Solution: The color of the lines/points change to different shades of blue.

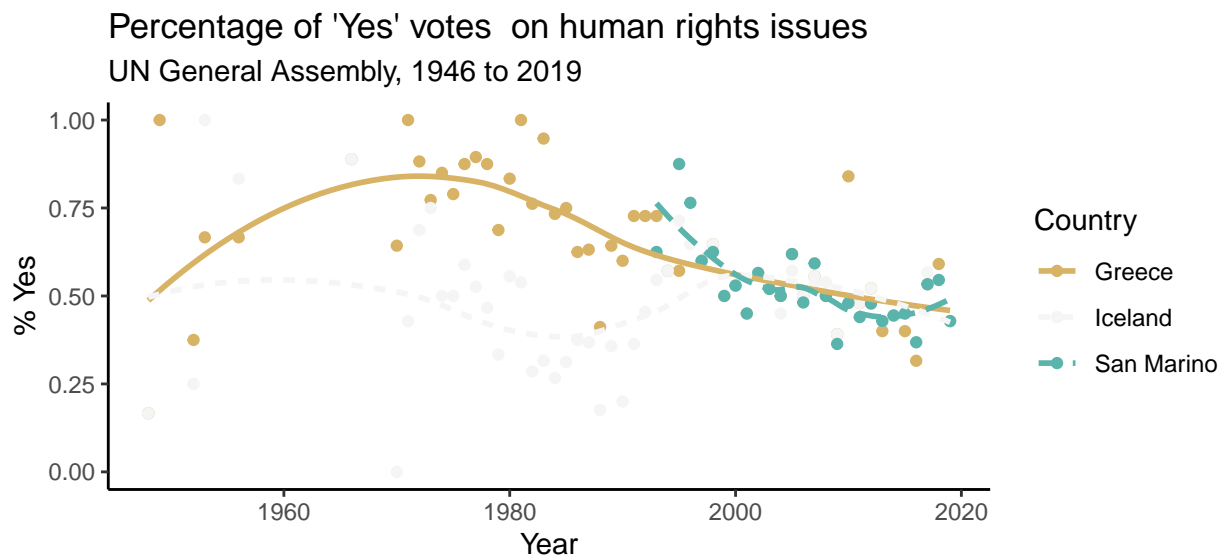
```
ggplot(data = unvotes_hr, mapping = aes(x = year, y = percent_yes,
                                         color = country)) +
  geom_point() +
  geom_smooth(method = "loess", se = FALSE, (aes(lty = country))) +
  scale_color_brewer() +
  labs(title = "Percentage of 'Yes' votes on human rights issues",
       subtitle = "UN General Assembly, 1946 to 2019",
       y = "% Yes",
       x = "Year",
       color = "Country",
       lty = "Country")
```



Within the `scale_color_brewer()` function, add the options `type = "div"` and `palette = 1`. Is this a good color scheme for this figure? Why or why not? (Can you think of a figure for which this would be a good color scheme?)

Solution: This is not a good color scheme for this figure because you cannot see the very light white line well. This could be a good color scheme when there is some ordering to the categories and where you are trying to emphasize that you are going from one category to the opposite category.

```
ggplot(data = unvotes_hr, mapping = aes(x = year, y = percent_yes,
                                         color = country)) +
  geom_point() +
  geom_smooth(method = "loess", se = FALSE, (aes(lty = country))) +
  scale_color_brewer(type = "div", palette = 1) +
  labs(title = "Percentage of 'Yes' votes on human rights issues",
       subtitle = "UN General Assembly, 1946 to 2019",
       y = "% Yes",
       x = "Year",
       color = "Country",
       lty = "Country")
```

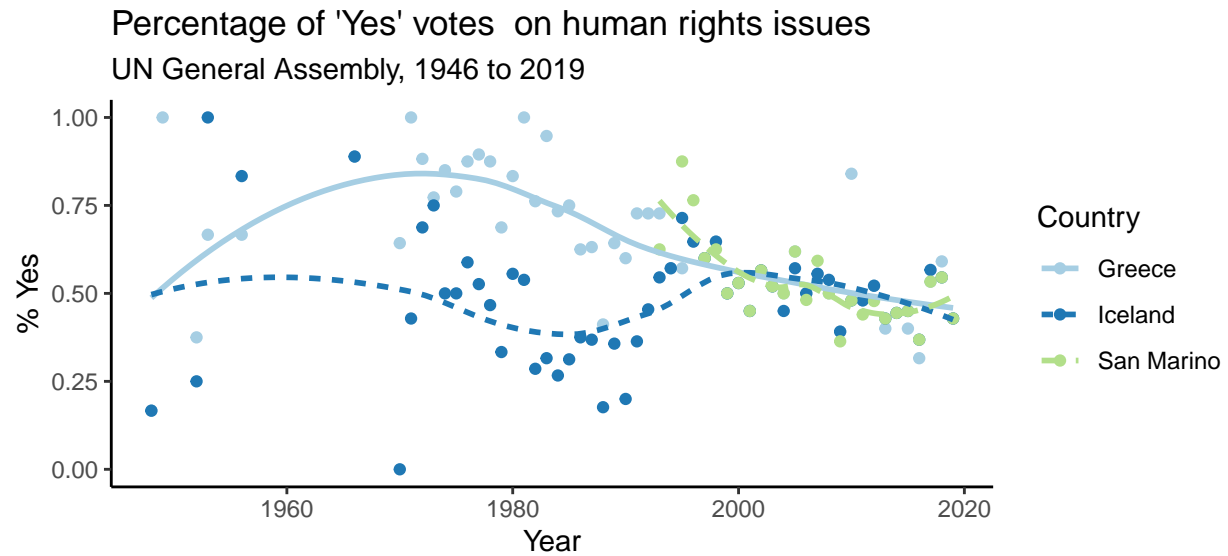


Check out the scale brewer reference manual for more information about sequential, diverging, and qualitative color schemes available from Color Brewer. The color brewer site is also helpful to visualize the different palettes (also shown in Figure 2.11 of MDSR).

Does a sequential, diverging, or qualitative color scheme make sense for this figure? Update the figure with one of the palettes from the appropriate scheme.

Solution: A qualitative color scheme makes the most sense because we are just looking at three particular categories that are not ordered in any way.

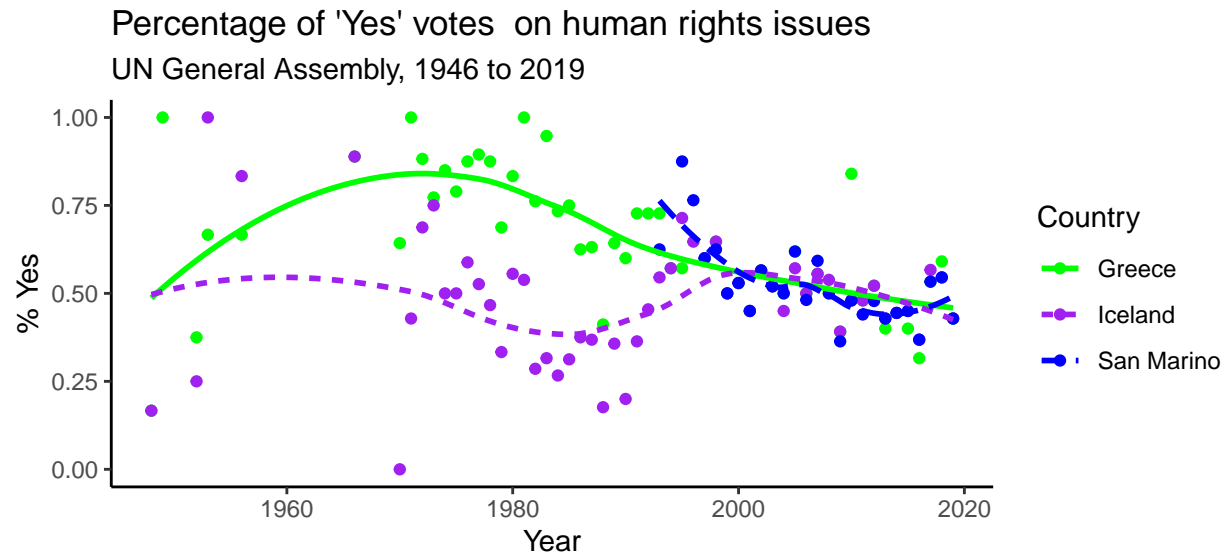
```
ggplot(data = unvotes_hr, mapping = aes(x = year, y = percent_yes,
                                         color = country)) +
  geom_point() +
  geom_smooth(method = "loess", se = FALSE, (aes(lty = country))) +
  scale_color_brewer(type = "qual", palette = 3) +
  labs(title = "Percentage of 'Yes' votes on human rights issues",
       subtitle = "UN General Assembly, 1946 to 2019",
       y = "% Yes",
       x = "Year",
       color = "Country",
       lty = "Country")
```



Replace the `scale_color_brewer()` line with `scale_color_manual(values = c("green", "purple", "blue"))`. Notice there are three colors I'm specifying, one for each of the three countries.

Solution:

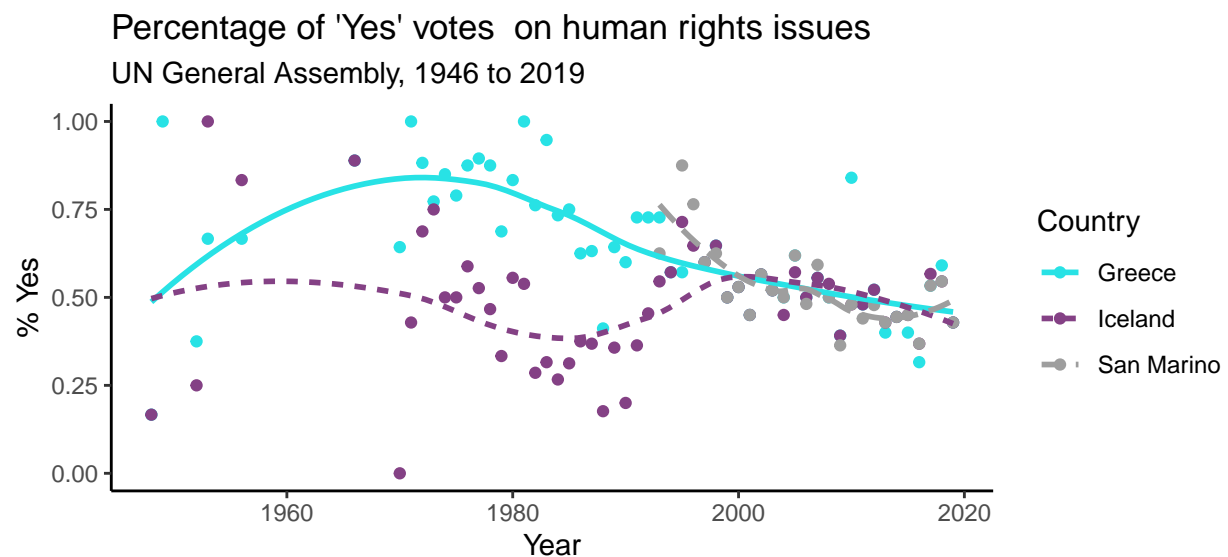
```
ggplot(data = unvotes_hr, mapping = aes(x = year, y = percent_yes,
                                         color = country)) +
  geom_point() +
  geom_smooth(method = "loess", se = FALSE, (aes(lty = country))) +
  scale_color_manual(values = c("green", "purple", "blue")) +
  labs(title = "Percentage of 'Yes' votes on human rights issues",
       subtitle = "UN General Assembly, 1946 to 2019",
       y = "% Yes",
       x = "Year",
       color = "Country",
       lty = "Country")
```



Don't like those colors? You can be more exact by specifying hex color codes. Try replacing "green", "purple" and "blue" with "#05a05a", "#844185", "#024a81", respectively.

Solution:

```
ggplot(data = unvotes_hr, mapping = aes(x = year, y = percent_yes,
                                         color = country)) +
  geom_point() +
  geom_smooth(method = "loess", se = FALSE, (aes(lty = country))) +
  scale_color_manual(values = c("#05a05a", "#844185", "#024a81")) +
  labs(title = "Percentage of 'Yes' votes on human rights issues",
       subtitle = "UN General Assembly, 1946 to 2019",
       y = "% Yes",
       x = "Year",
       color = "Country",
       lty = "Country")
```

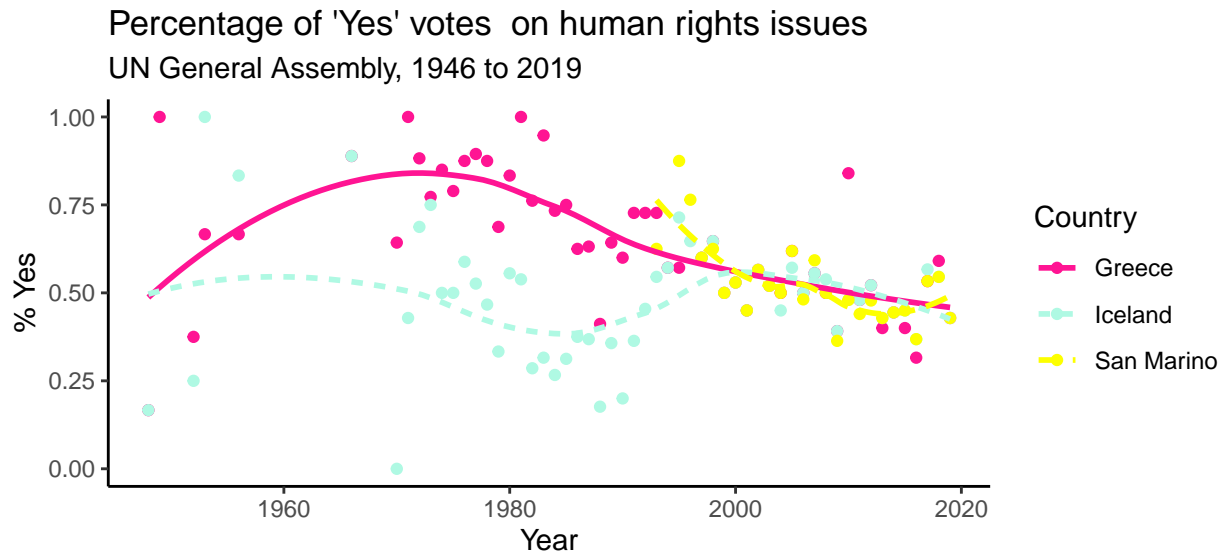


To identify hex codes for more colors, check out: color-hex.com.

Create one last figure with three colors of your choice assigned manually.

Solution:

```
ggplot(data = unvotes_hr, mapping = aes(x = year, y = percent_yes,
                                         color = country)) +
  geom_point() +
  geom_smooth(method = "loess", se = FALSE, (aes(lty = country))) +
  scale_color_manual(values = c("#ff1493", "#aff9e3", "#fdff00")) +
  labs(title = "Percentage of 'Yes' votes on human rights issues",
       subtitle = "UN General Assembly, 1946 to 2019",
       y = "% Yes",
       x = "Year",
       color = "Country",
       lty = "Country")
```



5 - Other Geoms

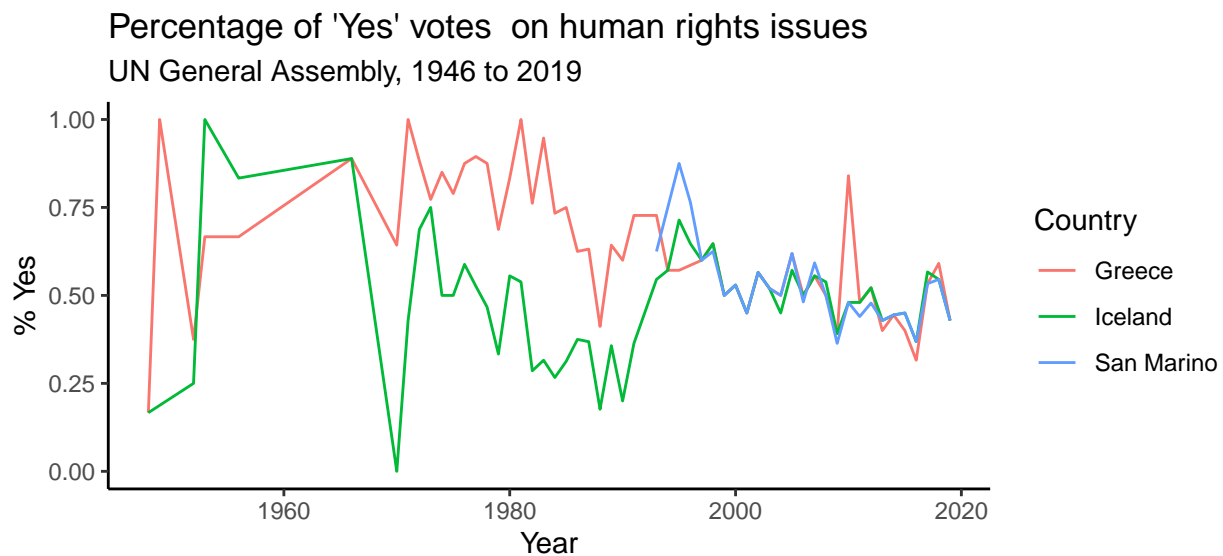
While `geom_point` and `geom_smooth` will do a lot for us, you'll see other geoms that can be useful as well, depending on what you aim to do.

Let's go back to our (almost) original plot to explore a few. This plot has the smoothed lines removed.

Change the `geom_point()` to `geom_line()`. What happens?

Solution: The data is plotted using changes in lines rather than discrete points.

```
# plot the data
ggplot(data = unvotes_hr, mapping = aes(x = year, y = percent_yes,
                                         color = country)) +
  geom_line() +
  labs(title = "Percentage of 'Yes' votes on human rights issues",
       subtitle = "UN General Assembly, 1946 to 2019",
       y = "% Yes",
       x = "Year",
       color = "Country")
```



(`geom_path` gives the same results as `geom_line` here due to observations already being in chronological order in the data set.)

For these next geoms, our data isn't set up in such a way that they really make sense, but I want you to see if you can figure out what each does, and how you might use them for other data sets in the future.

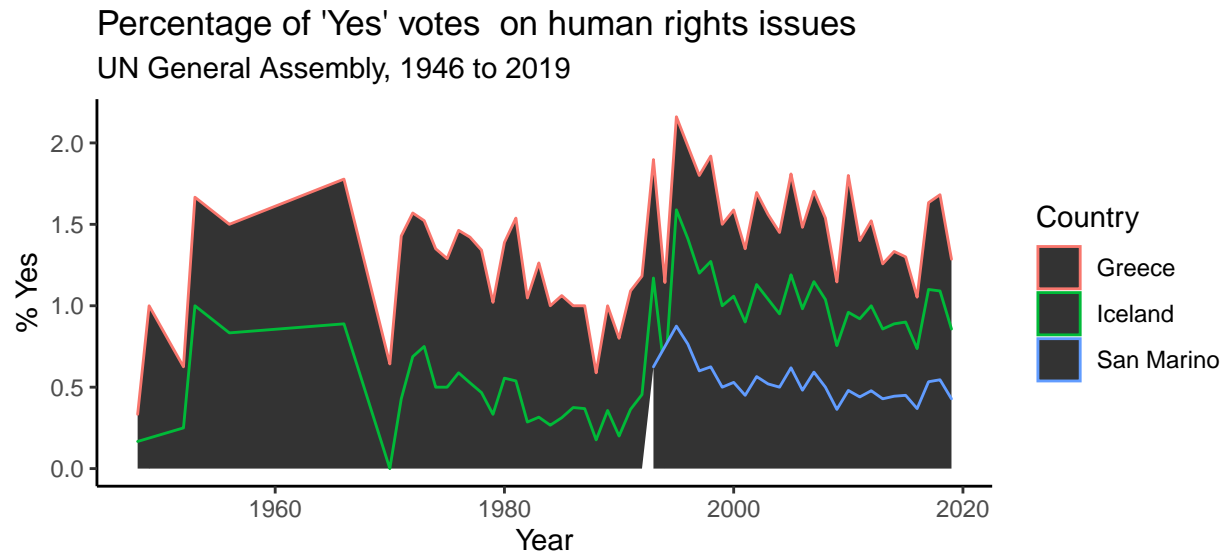
Copy over the code from above and use `geom_area` instead of `geom_line`. What happens?

Solution: Designed for when values are fractions of a whole, adds up data and shades under area.

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



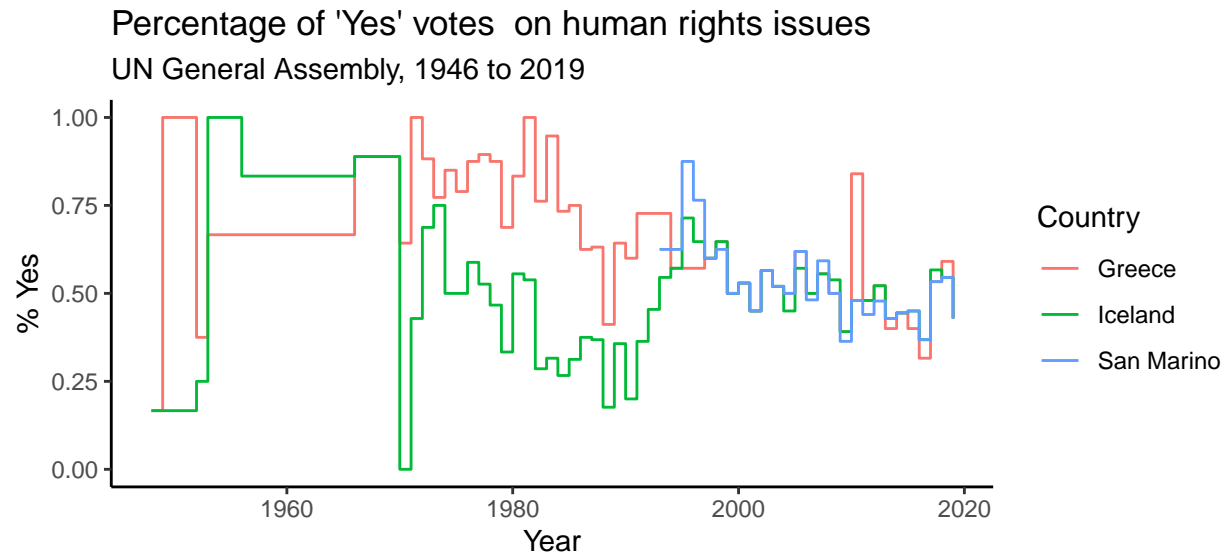
```
labs(title = "Percentage of 'Yes' votes on human rights issues",
      subtitle = "UN General Assembly, 1946 to 2019",
      y = "% Yes",
      x = "Year",
      color = "Country")
```



Now try `geom_step`. What happens?

Solution: Creates a step graph with straight lines being drawn between the points.

```
ggplot(data = unvotes_hr, mapping = aes(x = year, y = percent_yes,
                                          color = country)) +
  geom_step() +
  labs(title = "Percentage of 'Yes' votes on human rights issues",
        subtitle = "UN General Assembly, 1946 to 2019",
        y = "% Yes",
        x = "Year",
        color = "Country")
```



Many geoms will show up if you search. Some are designed for univariate plots, etc. Feel free to experiment around to see what exists. For example, you might explore a violin plot over a boxplot.

References

1. David Robinson (2017). `unvotes`: United Nations General Assembly Voting Data. R package version 0.2.0. <https://CRAN.R-project.org/package=unvotes>.
2. Erik Voeten “Data and Analyses of Voting in the UN General Assembly” Routledge Handbook of International Organization, edited by Bob Reinalda (published May 27, 2013).
3. Much of the analysis has been modeled on the examples presented in the `unvotes` package vignette.

UN country list

Below is a list of countries in the dataset:

```
un_votes %>%  
  select(country) %>%  
  arrange(country) %>%  
  distinct() %>%  
  datatable()
```