

# Introduction to R

## 4.1 Constructing one Plot Step-by-Step

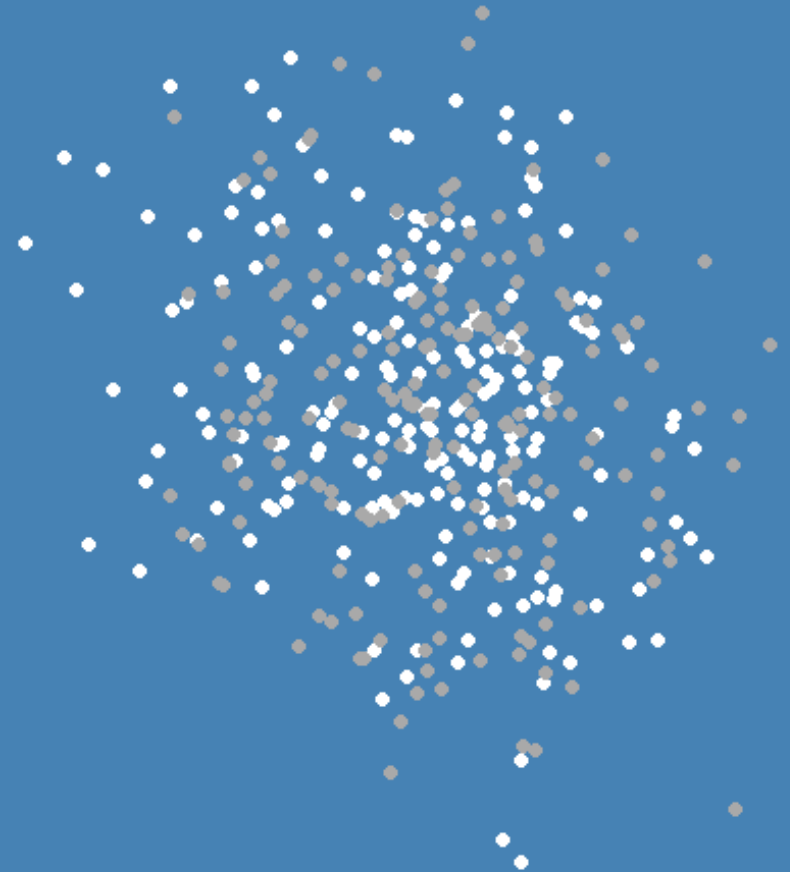
Grammar of graphics, Layered plots,  
Histograms

Lion Behrens, M.Sc.

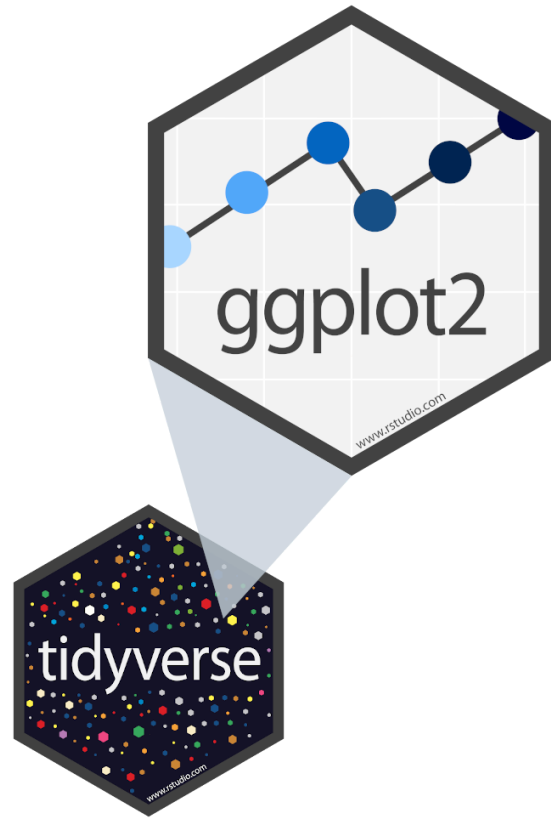


---

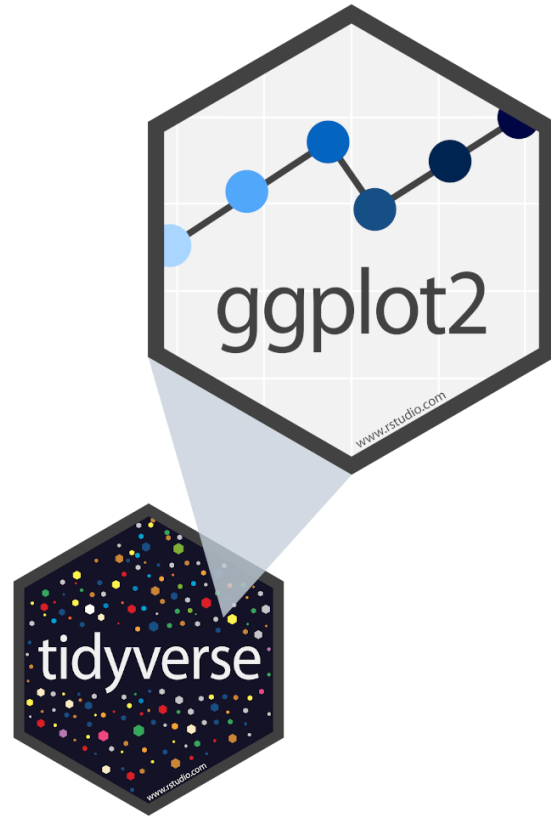
University of Mannheim  
Chair of Social Data Science and Methodology  
Chair of Quantitative Methods in the Social  
Sciences



# Package: ggplot2

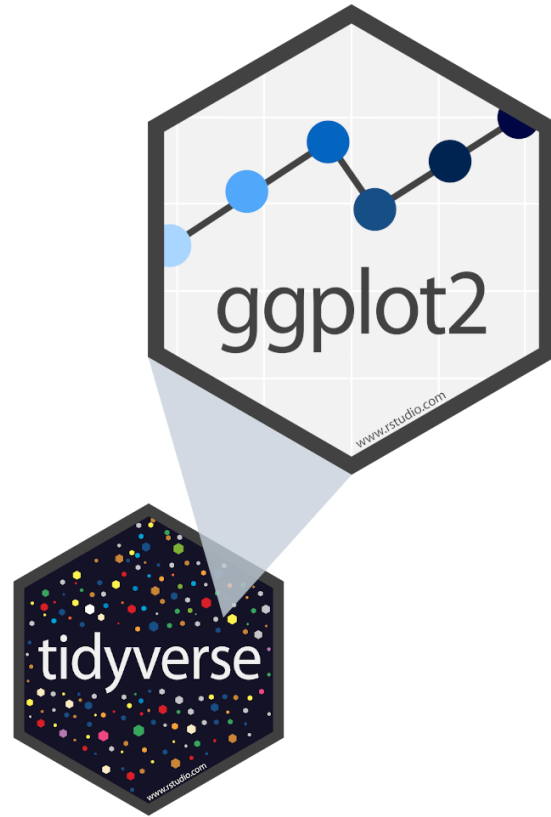


# Package: ggplot2



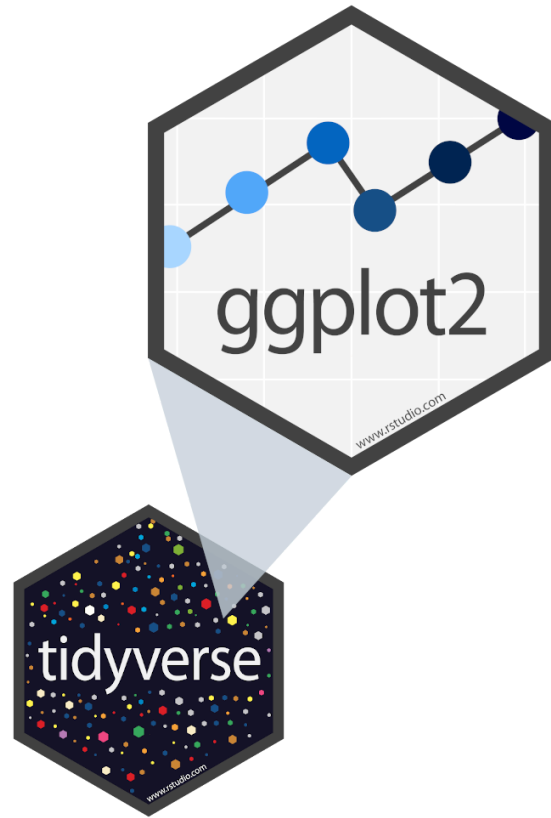
- `ggplot2` is a package that comes with the `tidyverse`

# Package: ggplot2



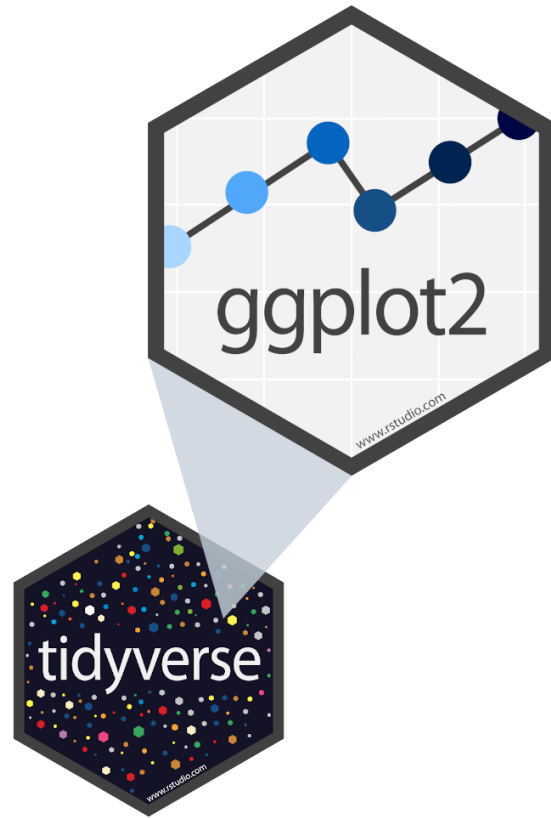
- `ggplot2` is a package that comes with the `tidyverse`
- Implementation of Wilkinson, Leland (1999) 'The Grammar of Graphics'

# Package: ggplot2



- `ggplot2` is a package that comes with the `tidyverse`
- Implementation of Wilkinson, Leland (1999) 'The Grammar of Graphics'
- Basic idea: Each plot can be **decomposed** into **several layers**
  - data
  - aesthetic mappings
  - geometric objects
  - scales
  - ...

# Package: ggplot2



- `ggplot2` is a package that comes with the `tidyverse`
- Implementation of Wilkinson, Leland (1999) 'The Grammar of Graphics'
- Basic idea: Each plot can be **decomposed** into **several layers**
  - data
  - aesthetic mappings
  - geometric objects
  - scales
  - ...
- Any plot can be **reproduced** by **specifying the correct layers** (components)

# Prerequisite: Data Wrangling Pipeline (I/III)

```
library(tidyverse)
ess10 <- haven::read_dta("./dat/ESS10.dta")
ess10 <- ess10 %>% # subset variables
  select(country = cntry, # sociodemographics
    gender = gndr,
    education_years = eduyrs,
    trust_social = ppltrst, # multidimensional trust
    trust_parliament = trstprl,
    trust_legalSys = trstlgl,
    trust_police = trstpplc,
    trust_politicians = trstplt,
    trust_parties = trstprt,
    trust_EP = trstep,
    trust_UN = trstun,
    left_right = lrscle, # attitudes
    life_satisfaction = stflife,
    pol_interest = polintr,
    voted = vote, # turnout
    party_choice = prtvtfr # party choice
  ) %>%
  mutate_at(c("country", "gender", "voted", "party_choice"), as.character) %>% # change types
  mutate_at("pol_interest", as.numeric) %>% # change types
  filter(country == "FR") # subset cases (only include France)
```

# Prerequisite: Data Wrangling Pipeline (II/III)

```
ess10 <- ess10 %>%  
  mutate(gender = recode_factor(gender,  
                                `1` = "Male",  
                                `2` = "Female"),  
         voted = recode_factor(voted,  
                               `1` = "Yes",  
                               `2` = "No",  
                               `3` = "Not eligible"),  
         party_choice = recode_factor(party_choice,  
                                       `1` = "Lutte Ouvrière",  
                                       `2` = "Nouv. Parti Anti-Capitaliste",  
                                       `3` = "Parti Communiste Français",  
                                       `4` = "La France Insoumise",  
                                       `5` = "Parti Socialiste",  
                                       `6` = "Europe Ecologie Les Verts",  
                                       `7` = "La République en Marche",  
                                       `8` = "Mouvement Démocrate",  
                                       `9` = "Les Républicains",  
                                       `10` = "Debout la France",  
                                       `11` = "Front National",  
                                       `12` = "Other",  
                                       `13` = "Blank",  
                                       `14` = "Null")  
  )
```



# Prerequisite: Data Wrangling Pipeline (III/III)

```
ess10 <- ess10 %>%  
  mutate(education_years = na_if(education_years, 114), # set 114 to missing  
         pol_interest = (pol_interest * -1) + 5, # invert scale  
         life_satisfaction = life_satisfaction + 1 # change scale to [1, 11]  
        ) %>%  
  drop_na(trust_politicians, gender, education_years,  
         life_satisfaction, pol_interest) # list-wise deletion of missings
```

# Our first plot: A Histogram

The [European Social Survey](#) Wave 10 asked people about their self-placement on a scale ranging from being [politically left](#) to being [politically right](#).

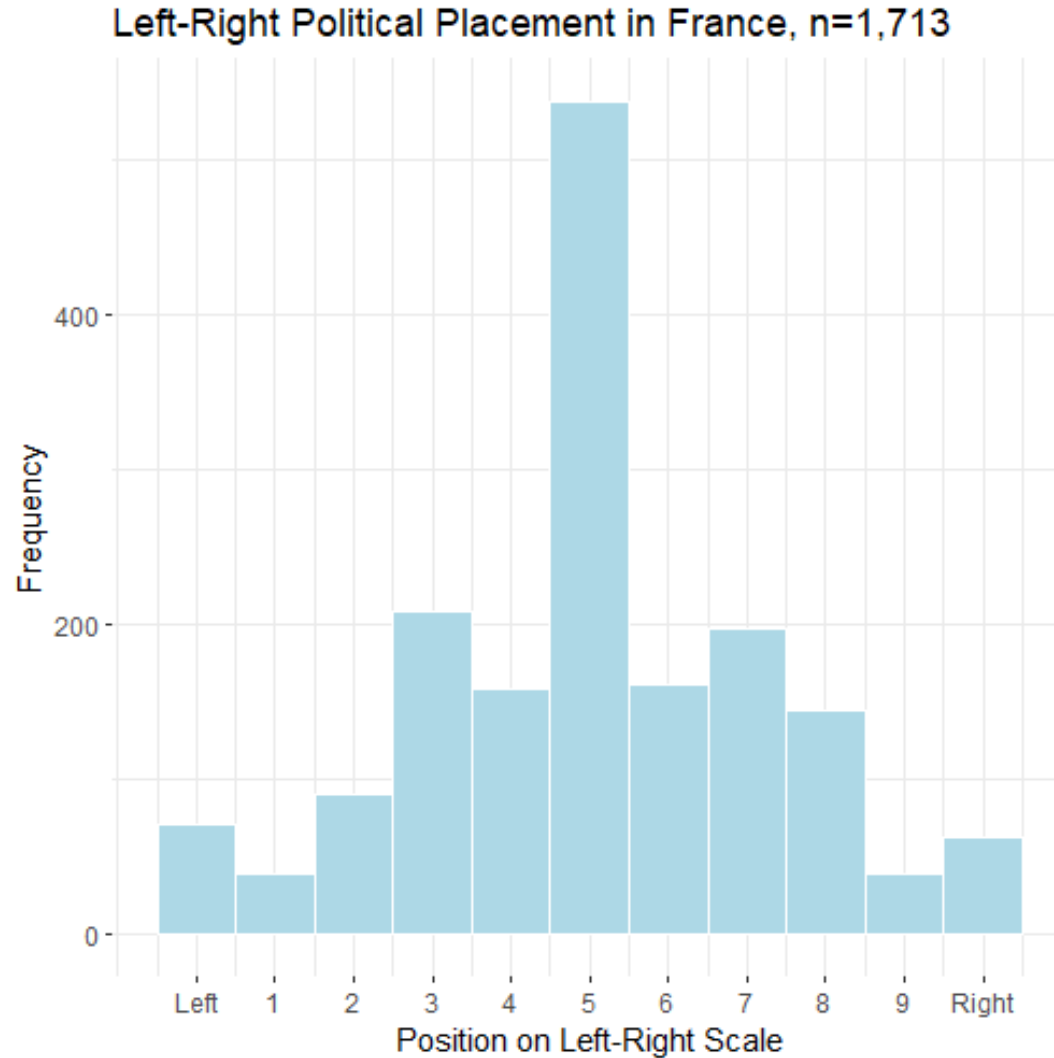
*In politics people sometimes talk of “left” and “right”. Using this card, where would you place yourself on this scale, where 0 means the left and 10 means the right?*

```
table(ess10$left_right)
```

```
##  
##    0    1    2    3    4    5    6    7    8    9   10  
##  71   39   90  209  159  538  162  198  145   39   63
```

- Let's visualize the distribution of this variable using a [histogram](#).
- We'll start with our familiar data from [France](#).
- Later, we'll expand and visualize this [cross-nationally](#).

# What We Are Trying To Get At



# Constructing a Plot Step-by-Step

First, let's initialize our plot.

```
ggplot()
```

# Constructing a Plot Step-by-Step

Second, let's add the first layer: our data.

```
ggplot()
```

# Constructing a Plot Step-by-Step

Second, let's add the first layer: our data.

```
ggplot(data = ess10)
```

# Constructing a Plot Step-by-Step

Now, let's specify the aesthetics across which we would like to plot.

```
ggplot(data = ess10)
```

# Constructing a Plot Step-by-Step

Now, let's specify the aesthetics across which we would like to plot.

For a histogram: Just one variable on the x-axis

```
ggplot(data = ess10)
```

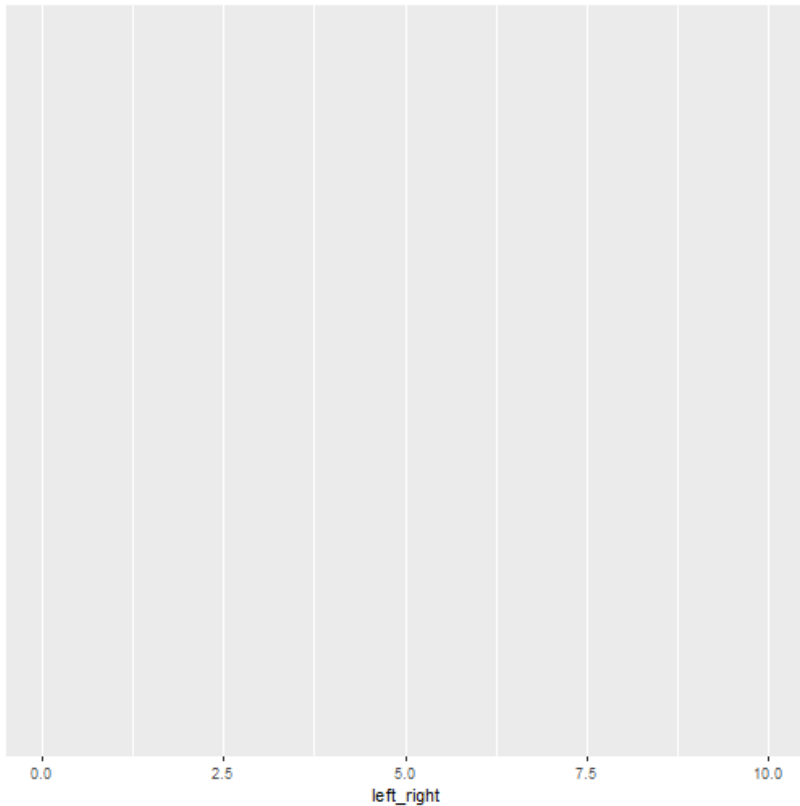


# Constructing a Plot Step-by-Step

Now, let's specify the aesthetics across which we would like to plot.

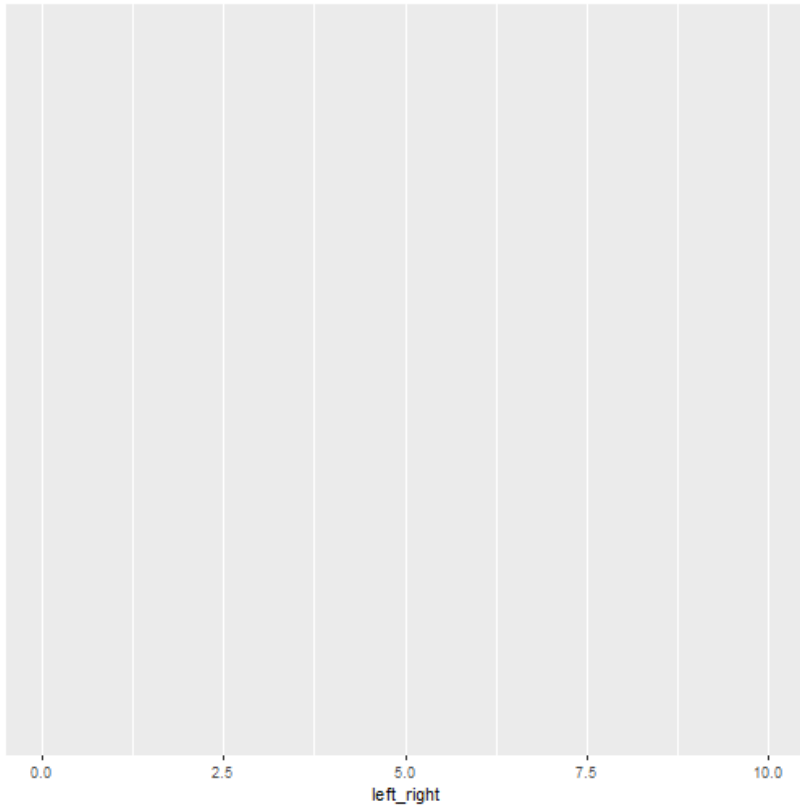
For a histogram: Just one variable on the x-axis

```
ggplot(data = ess10,  
       aes(x = left_right))
```



# Constructing a Plot Step-by-Step

Now, the basic dimensionality of our plot is set.

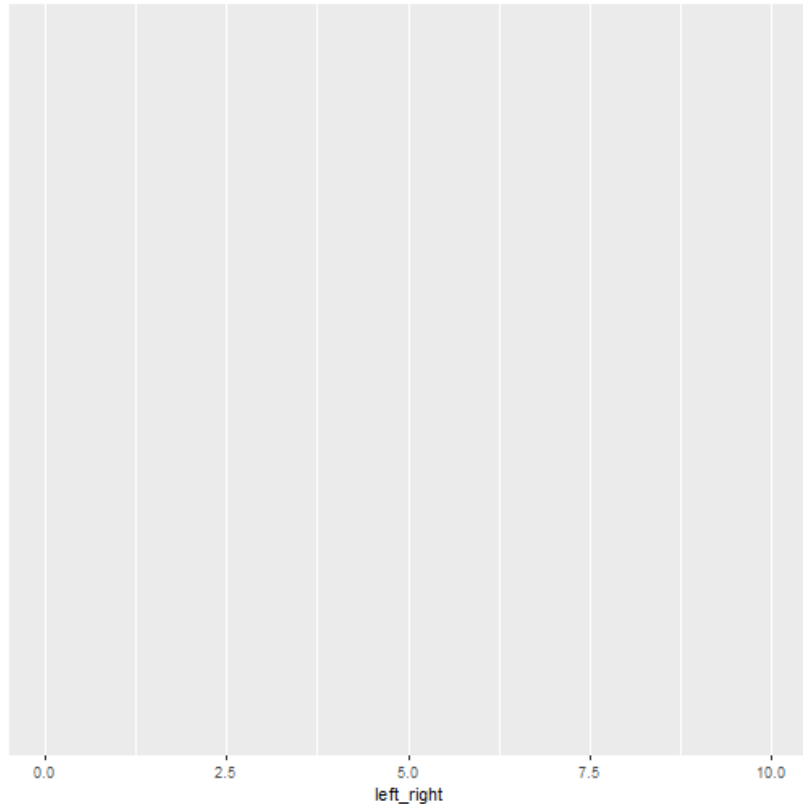


```
ggplot(data = ess10,  
       aes(x = left_right))
```

# Constructing a Plot Step-by-Step

Now, the basic dimensionality of our plot is set.

The `geom_`-argument specifies the [type of geometric object](#) we want to [map to these aesthetics](#). From now on, each additional layer is added with a "+".

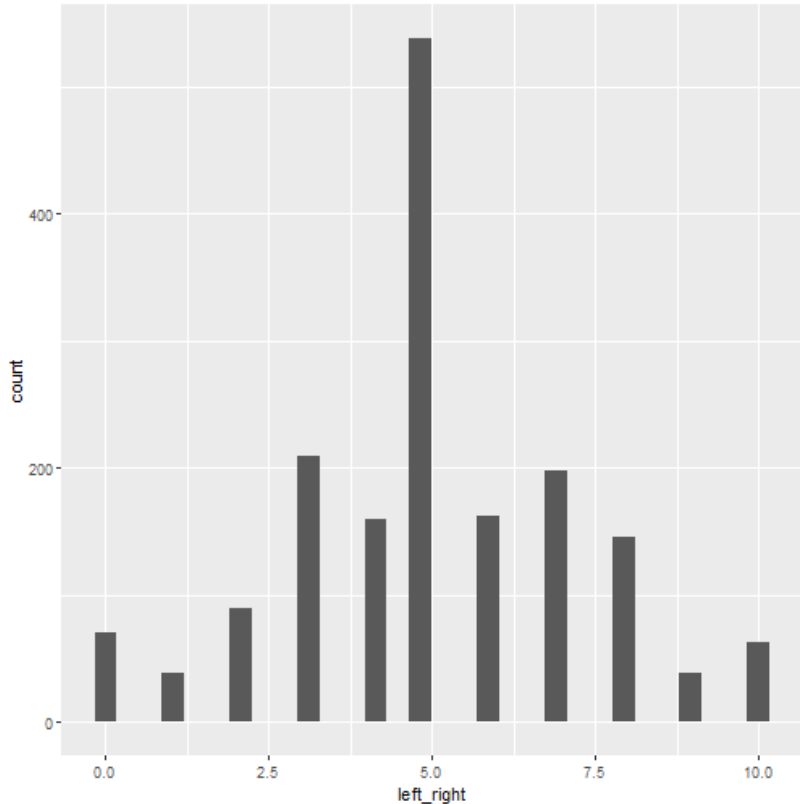


```
ggplot(data = ess10,  
       aes(x = left_right))
```

# Constructing a Plot Step-by-Step

Now, the basic dimensionality of our plot is set.

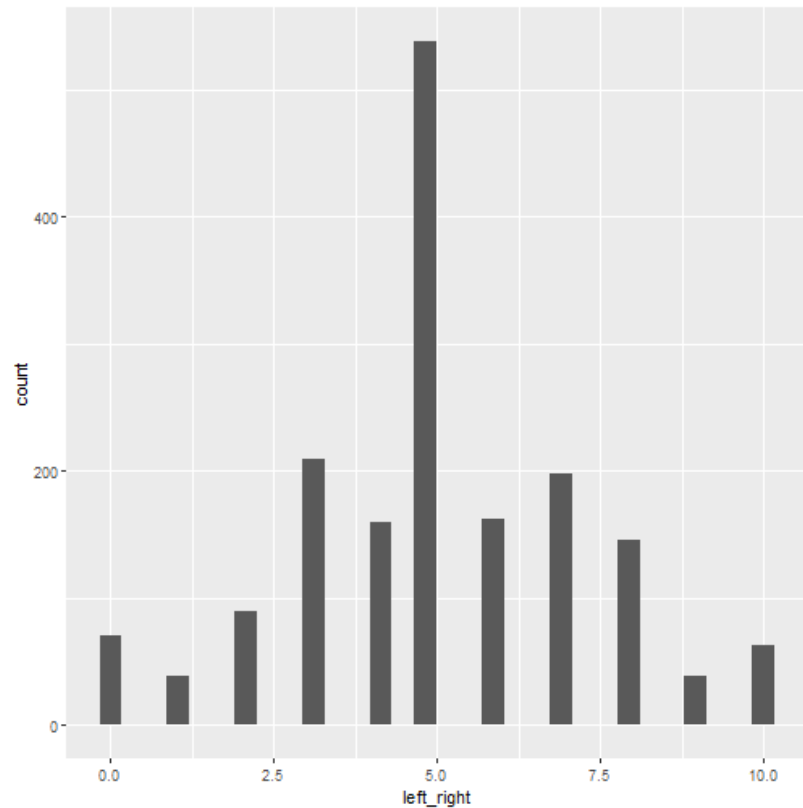
The `geom_`-argument specifies the [type of geometric object](#) we want to [map to these aesthetics](#). From now on, each additional layer is added with a "+".



```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram()
```

# Constructing a Plot Step-by-Step

In general, that's already the histogram that we wanted to plot. But let's customize it further to make it nicer.

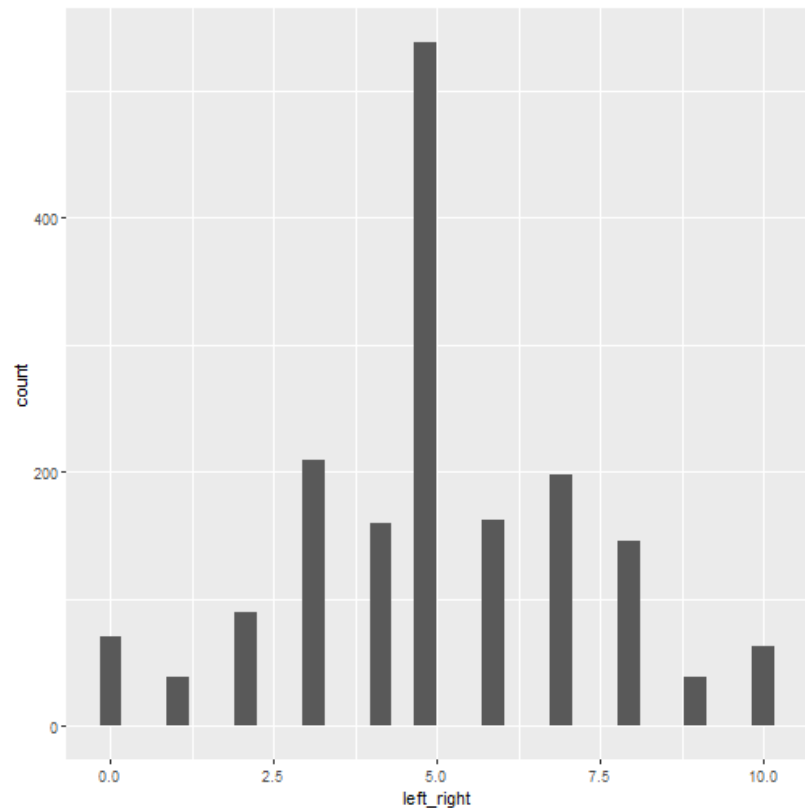


```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram()
```

# Constructing a Plot Step-by-Step

In general, that's already the histogram that we wanted to plot. But let's customize it further to make it nicer.

Let's start with removing whitespace between the individual bins.

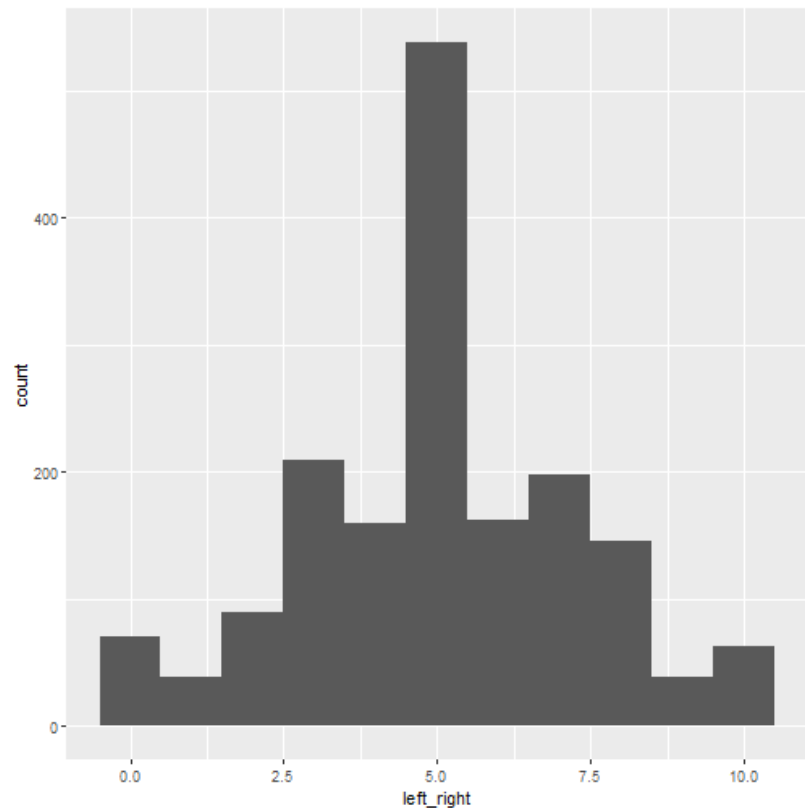


```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram()
```

# Constructing a Plot Step-by-Step

In general, that's already the histogram that we wanted to plot. But let's customize it further to make it nicer.

Let's start with removing whitespace between the individual bins.

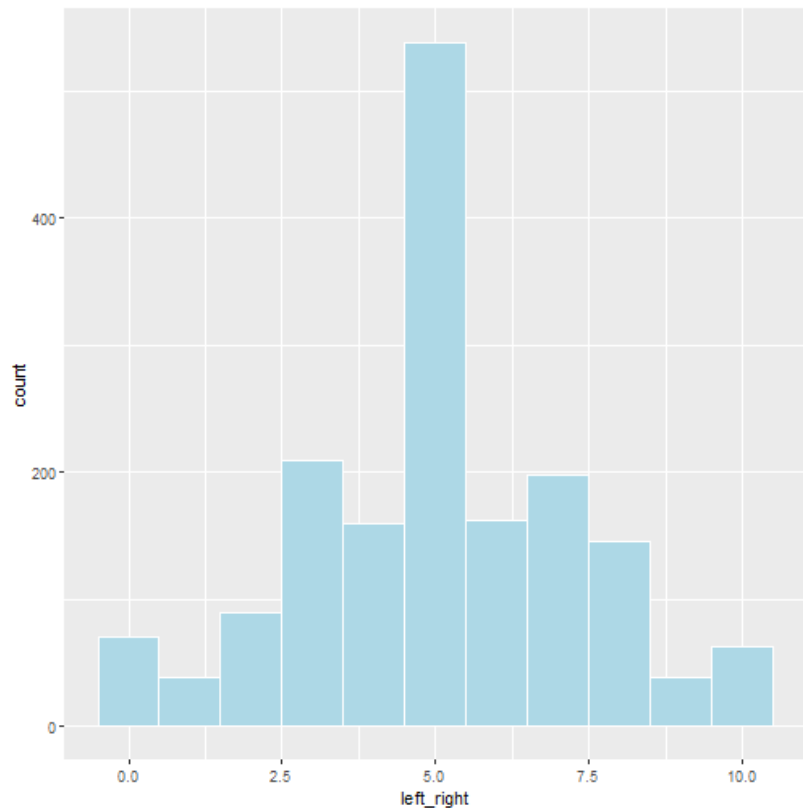


```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram(  
    bins = 11)
```

# Constructing a Plot Step-by-Step

In general, that's already the histogram that we wanted to plot. But let's customize it further to make it nicer.

And let's add a little color.



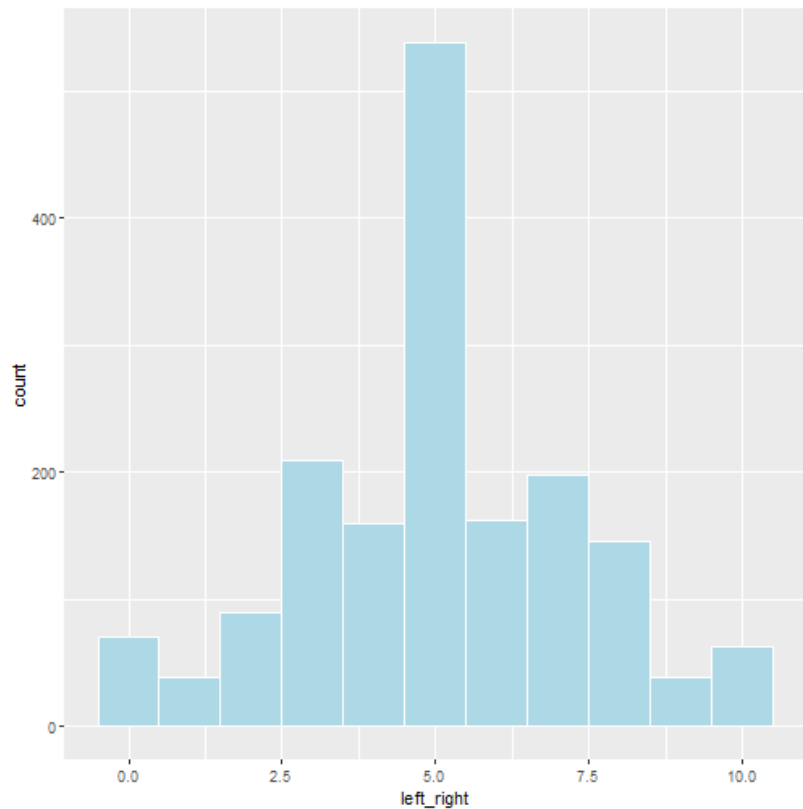
```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram(  
    bins = 11, color = "white",  
    fill = "lightblue")
```



# Constructing a Plot Step-by-Step

In general, that's already the histogram that we wanted to plot. But let's customize it further to make it nicer.

The x-axis scale is currently taken as continuous, which in reality it is not. Let's fix this.

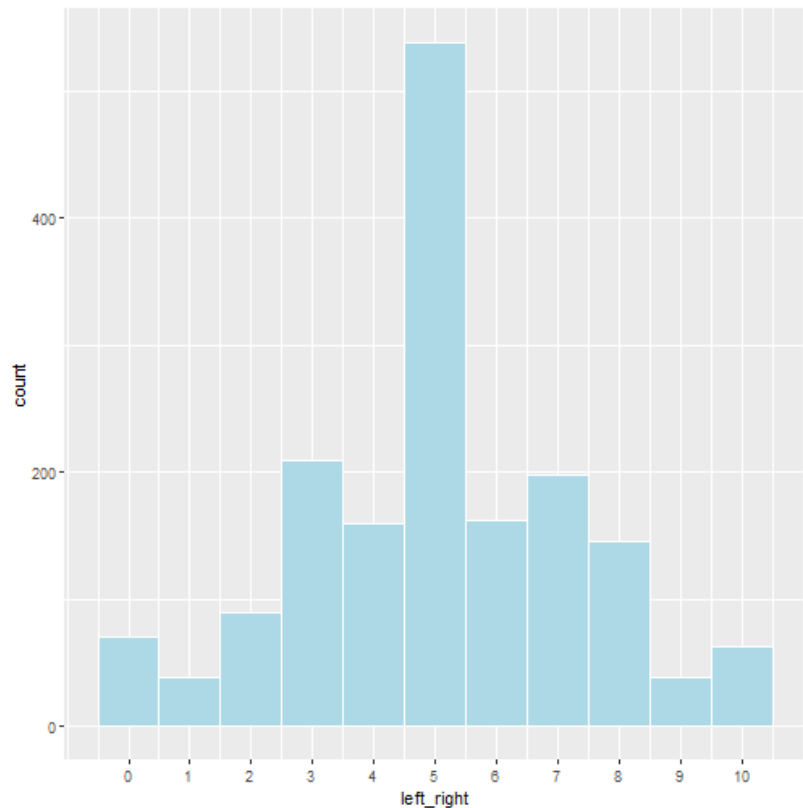


```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram(  
    bins = 11, color = "white",  
    fill = "lightblue")
```

# Constructing a Plot Step-by-Step

In general, that's already the histogram that we wanted to plot. But let's customize it further to make it nicer.

The x-axis scale is currently taken as continuous, which in reality it is not. Let's fix this.

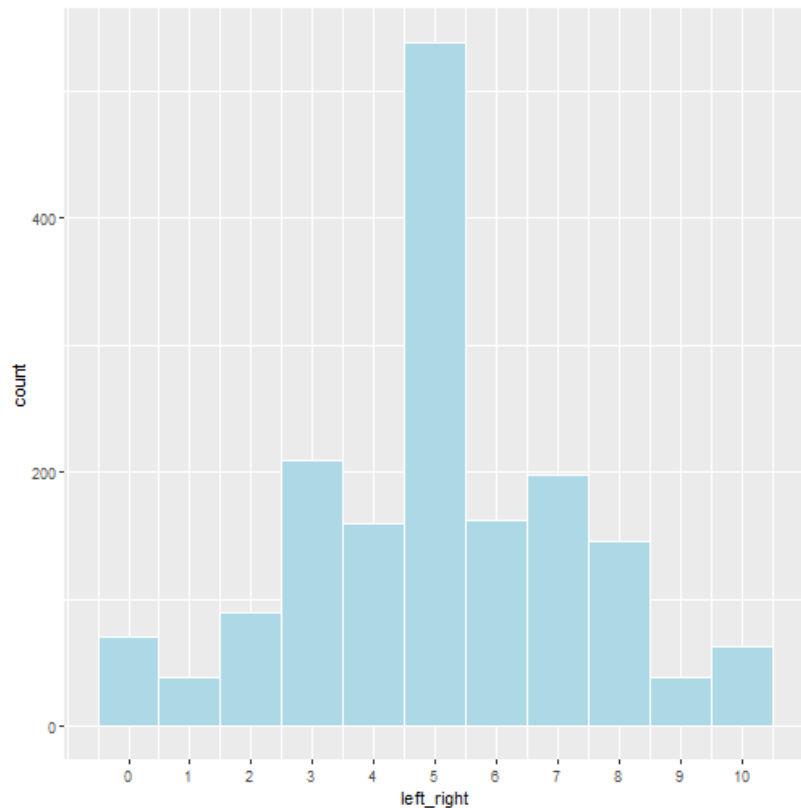


```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram(  
    bins = 11, color = "white",  
    fill = "lightblue") +  
  scale_x_continuous(  
    breaks = seq(0,10,1),  
    labels = seq(0,10,1))
```

# Constructing a Plot Step-by-Step

In general, that's already the histogram that we wanted to plot. But let's customize it further to make it nicer.

Where on the scale is "left", where is "right" located? Let's add this information.

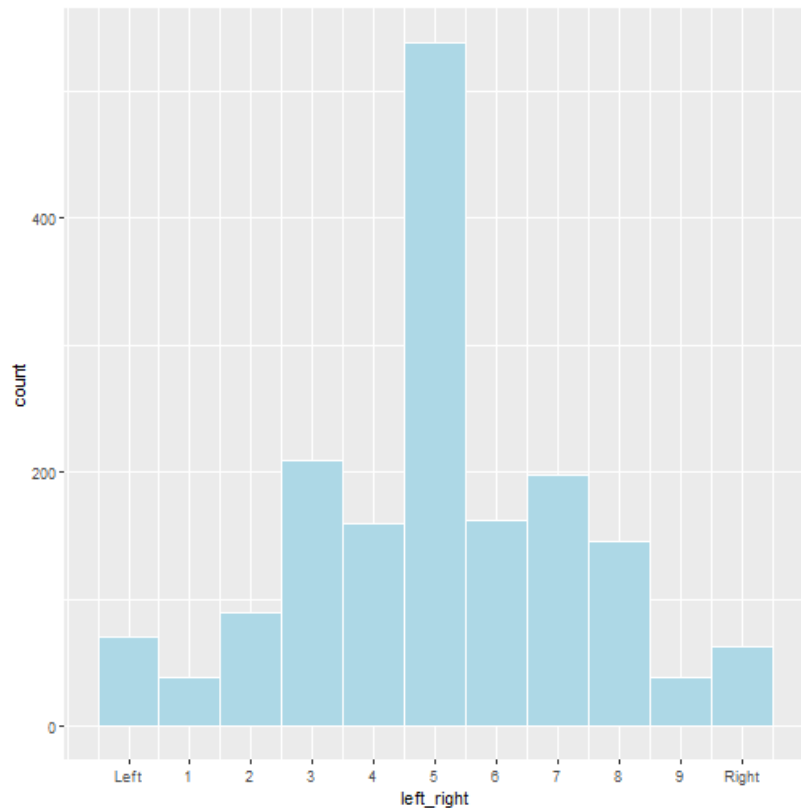


```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram(  
    bins = 11, color = "white",  
    fill = "lightblue") +  
  scale_x_continuous(  
    breaks = seq(0,10,1),  
    labels = seq(0,10,1))
```

# Constructing a Plot Step-by-Step

In general, that's already the histogram that we wanted to plot. But let's customize it further to make it nicer.

Where on the scale is "left", where is "right" located? Let's add this information.

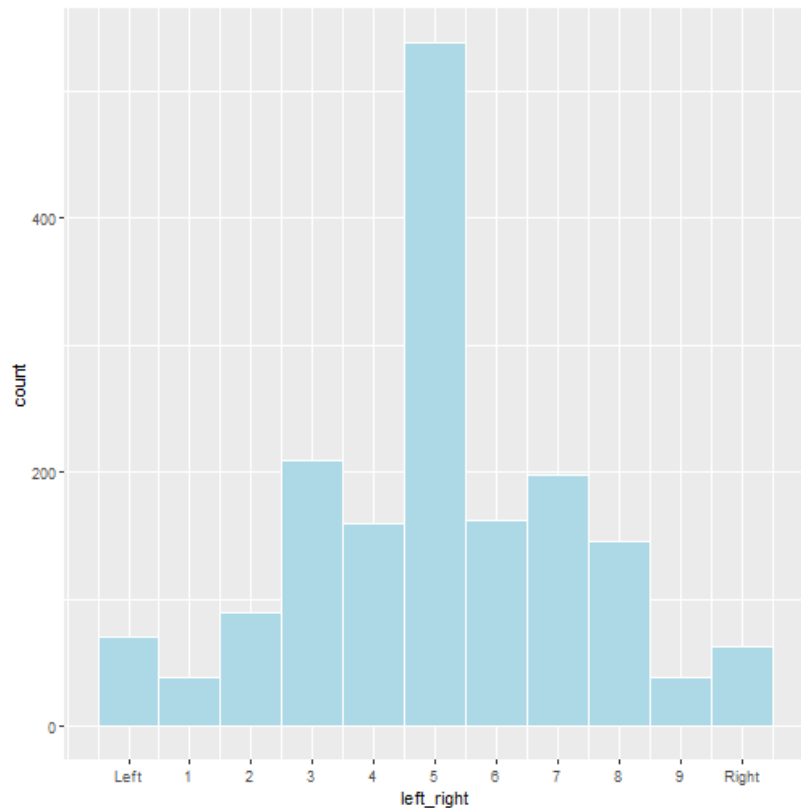


```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram(  
    bins = 11, color = "white",  
    fill = "lightblue") +  
  scale_x_continuous(  
    breaks = seq(0,10,1),  
    labels = c("Left", seq(1,9,1), "Right"))
```

# Constructing a Plot Step-by-Step

In general, that's already the histogram that we wanted to plot. But let's customize it further to make it nicer.

Let's name the x- and y-axis intuitively and add a title.

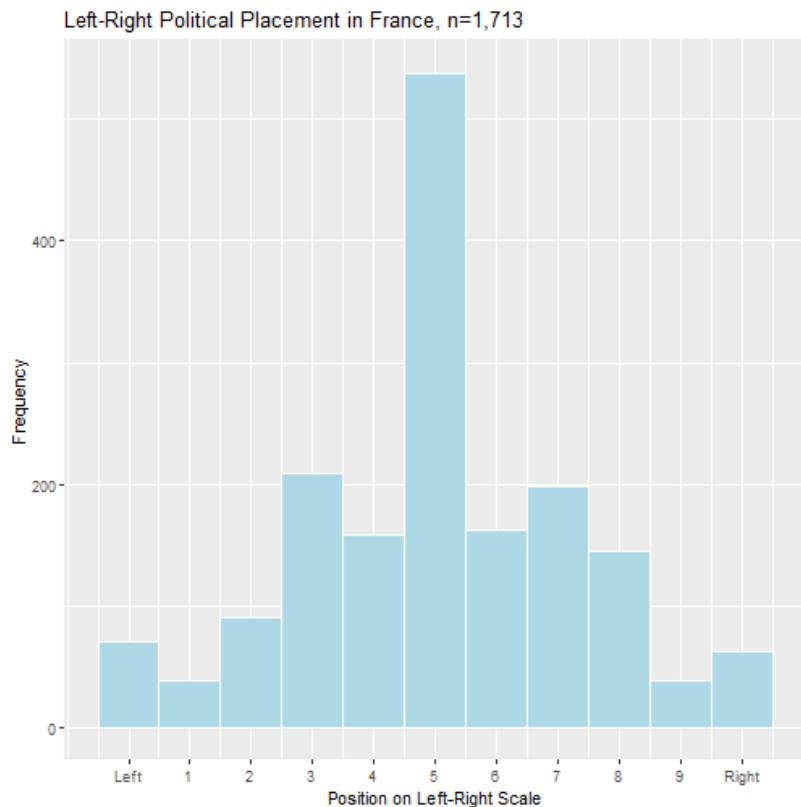


```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram(  
    bins = 11, color = "white",  
    fill = "lightblue") +  
  scale_x_continuous(  
    breaks = seq(0,10,1),  
    labels = c("Left", seq(1,9,1), "Right"))
```

# Constructing a Plot Step-by-Step

In general, that's already the histogram that we wanted to plot. But let's customize it further to make it nicer.

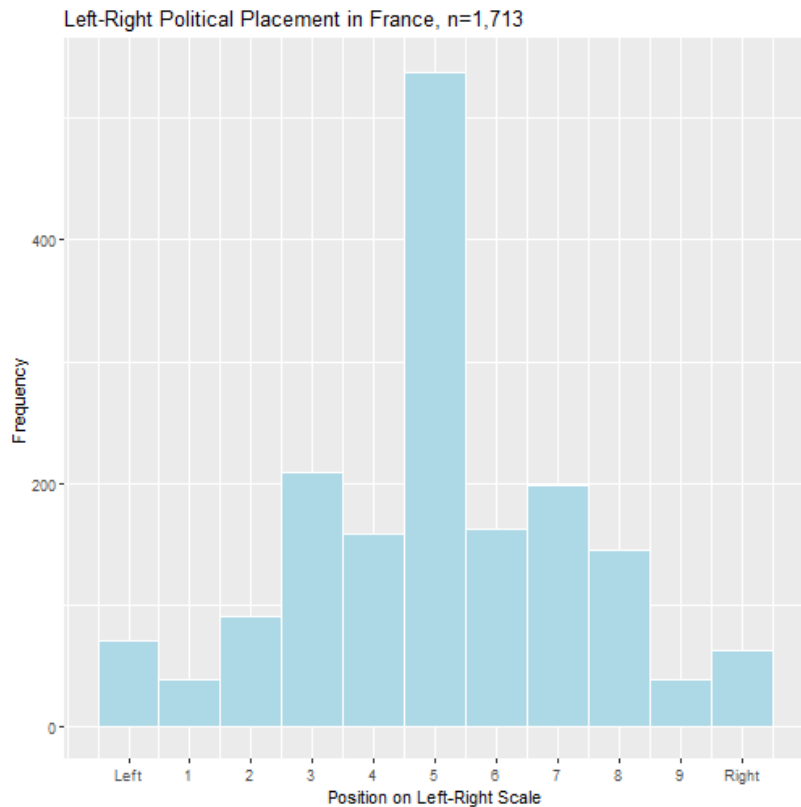
Let's name the x- and y-axis intuitively and add a title.



```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram(  
    bins = 11, color = "white",  
    fill = "lightblue") +  
  scale_x_continuous(  
    breaks = seq(0,10,1),  
    labels = c("Left", seq(1,9,1), "Right")) +  
  labs(x = "Position on Left-Right Scale",  
       y = "Frequency",  
       title = "Left-Right Political Placement")
```

# Constructing a Plot Step-by-Step

That looks great already!

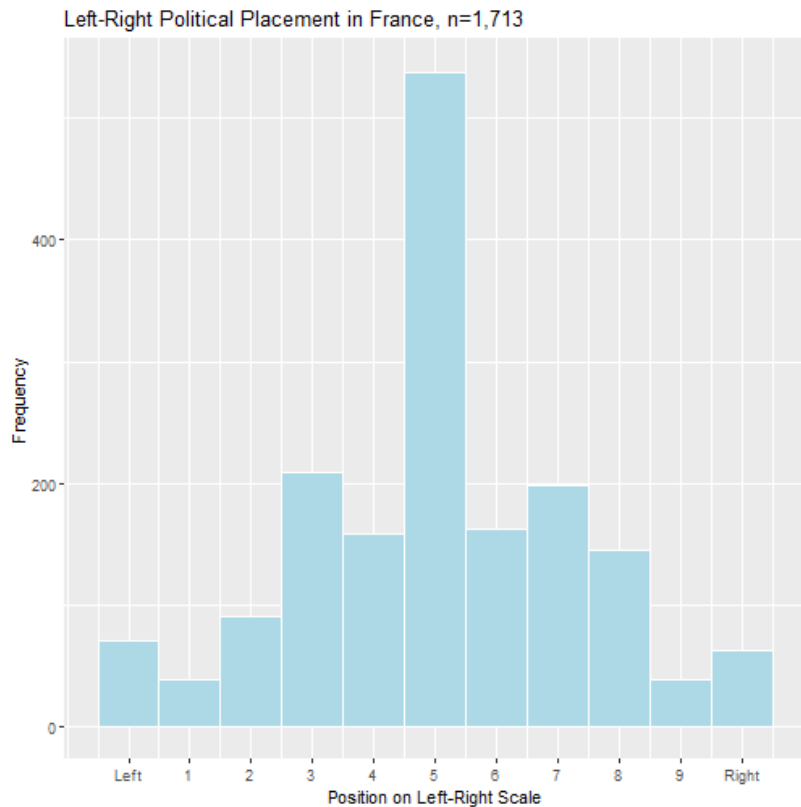


```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram(  
    bins = 11, color = "white",  
    fill = "lightblue") +  
  scale_x_continuous(  
    breaks = seq(0,10,1),  
    labels = c("Left", seq(1,9,1), "Right")) +  
  labs(x = "Position on Left-Right Scale",  
       y = "Frequency",  
       title = "Left-Right Political Placement")
```

# Constructing a Plot Step-by-Step

That looks great already!


Let's learn about themes.



```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram(  
    bins = 11, color = "white",  
    fill = "lightblue") +  
  scale_x_continuous(  
    breaks = seq(0,10,1),  
    labels = c("Left", seq(1,9,1), "Right")) +  
  labs(x = "Position on Left-Right Scale",  
       y = "Frequency",  
       title = "Left-Right Political Placement")
```




# Constructing a Plot Step-by-Step

ggplot2 3.4.0   Reference   News ▾   Articles ▾   Extensions      

## Complete themes

Source: [R/theme-defaults.r](#)



These are complete themes which control all non-data display. Use `theme()` if you just need to tweak the display of an existing theme.

### Usage

```
theme_grey(  
  base_size = 11,  
  base_family = "",  
  base_line_size = base_size/22,  
  base_rect_size = base_size/22  
)  
  
theme_gray(  
  base_size = 11,  
  base_family = "",  
  base_line_size = base_size/22,  
  base_rect_size = base_size/22  
)  
  
theme_bw(  
  base_size = 11,  
  base_family = "",  
  base_line_size = base_size/22,  
  base_rect_size = base_size/22  
)
```

#### ON THIS PAGE

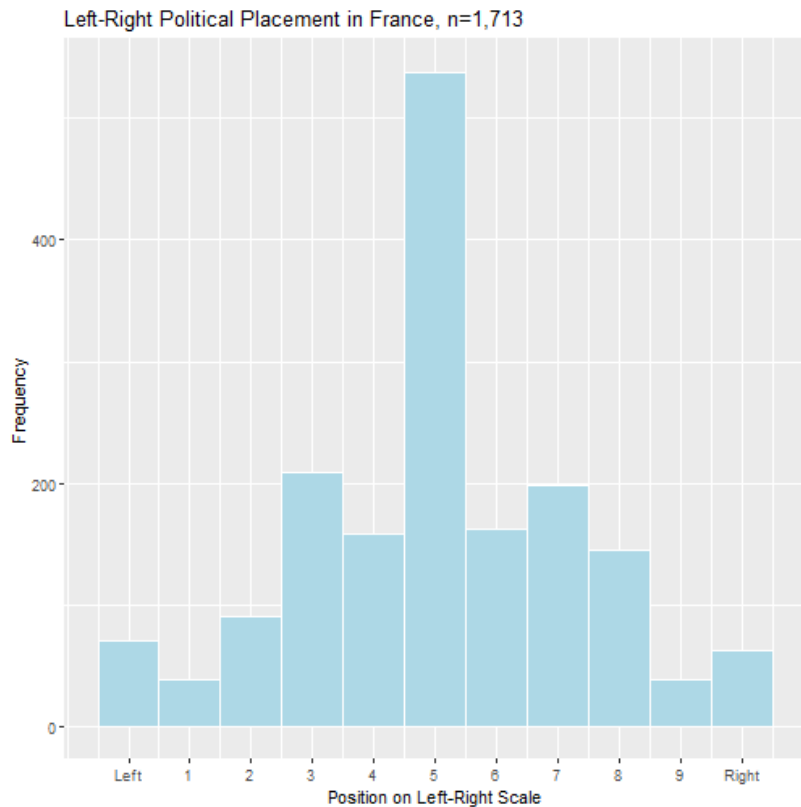
- Usage
- Arguments
- Details
- Examples

Link: <https://ggplot2.tidyverse.org/reference/ggtheme.html>.

# Constructing a Plot Step-by-Step

That looks great already!

Let's choose `theme_bw()`.

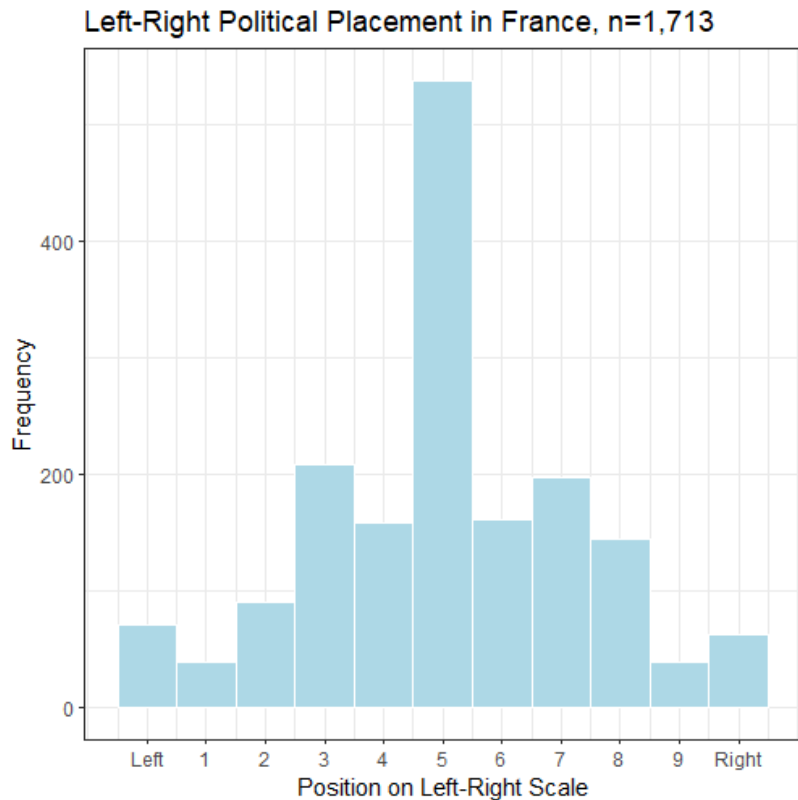


```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram(  
    bins = 11, color = "white",  
    fill = "lightblue") +  
  scale_x_continuous(  
    breaks = seq(0,10,1),  
    labels = c("Left", seq(1,9,1), "Right")) +  
  labs(x = "Position on Left-Right Scale",  
       y = "Frequency",  
       title = "Left-Right Political Placement")
```

# Constructing a Plot Step-by-Step

That looks great already!

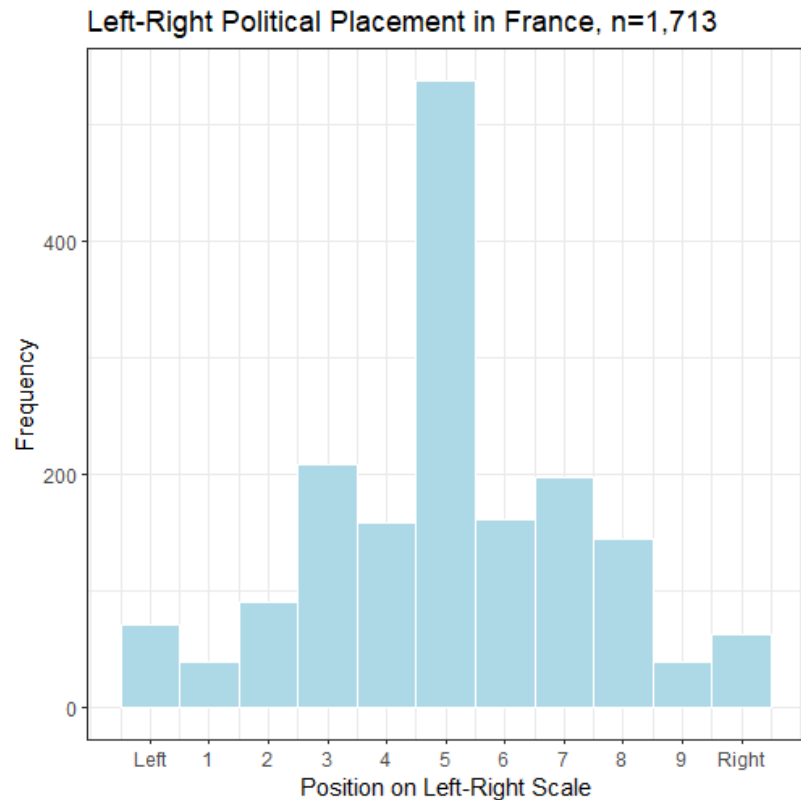
Let's choose `theme_bw()`.



```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram(  
    bins = 11, color = "white",  
    fill = "lightblue") +  
  scale_x_continuous(  
    breaks = seq(0,10,1),  
    labels = c("Left", seq(1,9,1), "Right")) +  
  labs(x = "Position on Left-Right Scale",  
       y = "Frequency",  
       title = "Left-Right Political Placement") +  
  theme_bw(base_size=15)
```

# Constructing a Plot Step-by-Step

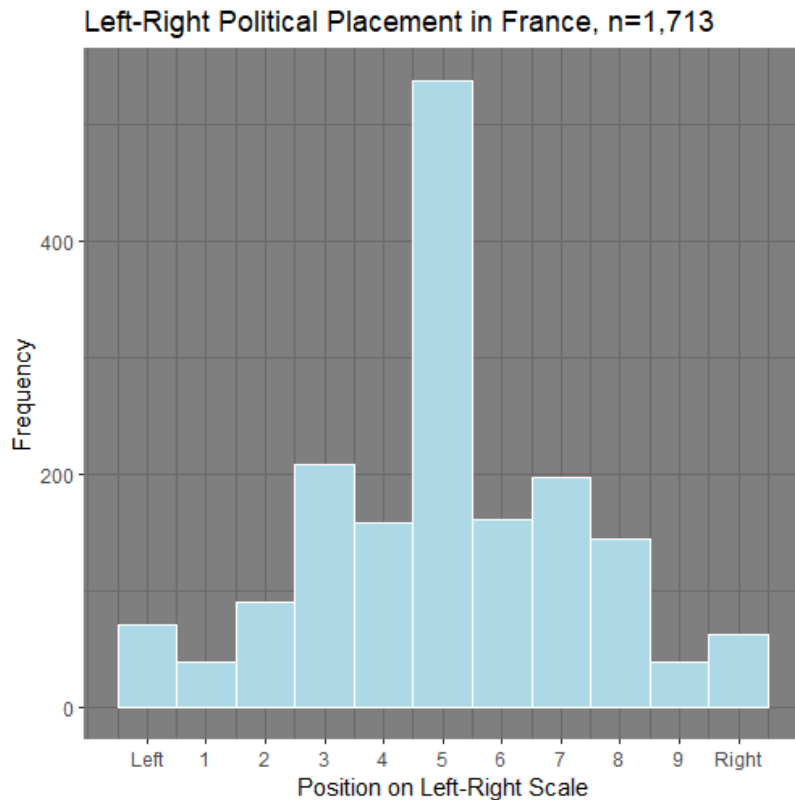
What about `theme_dark()`?



```
ggplot(data = ess10,
       aes(x = left_right)) +
  geom_histogram(
    bins = 11, color = "white",
    fill = "lightblue") +
  scale_x_continuous(
    breaks = seq(0,10,1),
    labels = c("Left", seq(1,9,1), "Right")) +
  labs(x = "Position on Left-Right Scale",
       y = "Frequency",
       title = "Left-Right Political Placement") +
  theme_bw(base_size=15)
```

# Constructing a Plot Step-by-Step

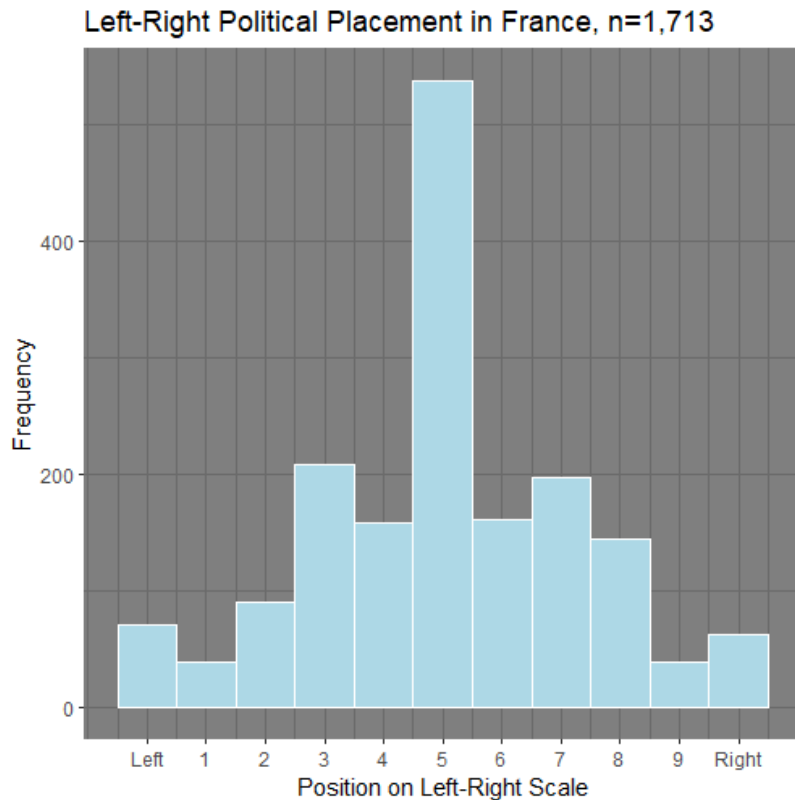
What about `theme_dark()`?



```
ggplot(data = ess10,
       aes(x = left_right)) +
  geom_histogram(
    bins = 11, color = "white",
    fill = "lightblue") +
  scale_x_continuous(
    breaks = seq(0,10,1),
    labels = c("Left", seq(1,9,1), "Right")) +
  labs(x = "Position on Left-Right Scale",
       y = "Frequency",
       title = "Left-Right Political Placement") +
  theme_dark(base_size=15)
```

# Constructing a Plot Step-by-Step

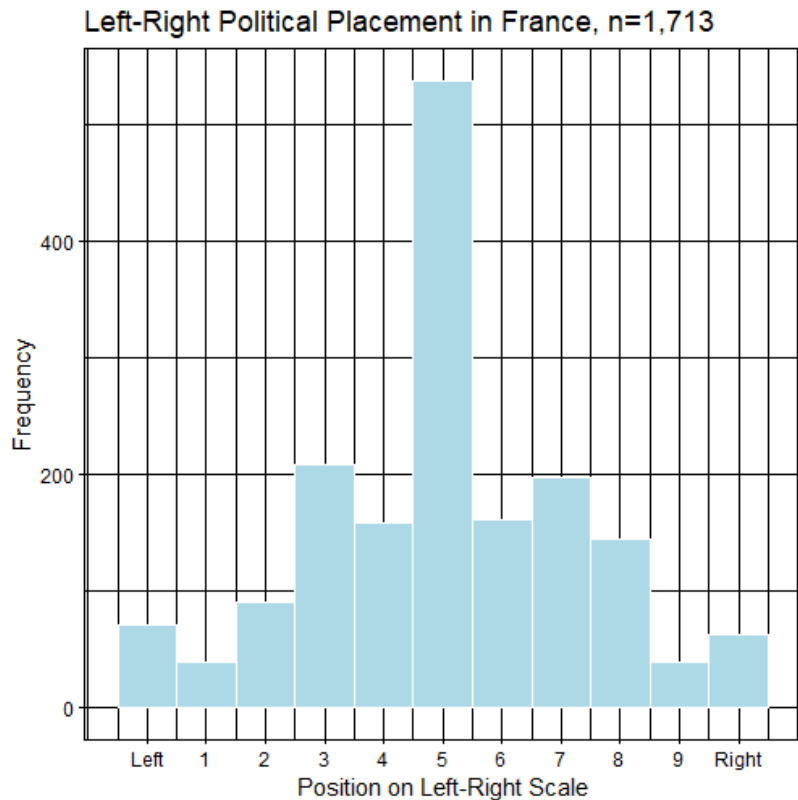
Or `theme_linedraw()`?



```
ggplot(data = ess10,
       aes(x = left_right)) +
  geom_histogram(
    bins = 11, color = "white",
    fill = "lightblue") +
  scale_x_continuous(
    breaks = seq(0,10,1),
    labels = c("Left", seq(1,9,1), "Right")) +
  labs(x = "Position on Left-Right Scale",
       y = "Frequency",
       title = "Left-Right Political Placement") +
  theme_dark(base_size=15)
```

# Constructing a Plot Step-by-Step

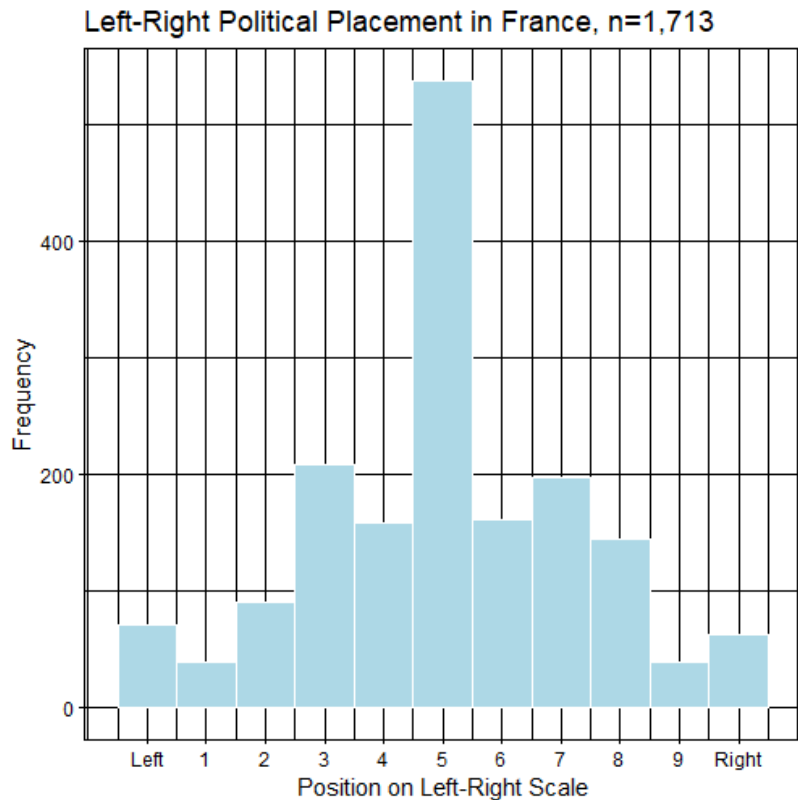
Or `theme_linedraw()`?



```
ggplot(data = ess10,
       aes(x = left_right)) +
  geom_histogram(
    bins = 11, color = "white",
    fill = "lightblue") +
  scale_x_continuous(
    breaks = seq(0,10,1),
    labels = c("Left", seq(1,9,1), "Right")) +
  labs(x = "Position on Left-Right Scale",
       y = "Frequency",
       title = "Left-Right Political Placement") +
  theme_linedraw(base_size=15)
```

# Constructing a Plot Step-by-Step

Let's stick to `theme_bw()`.

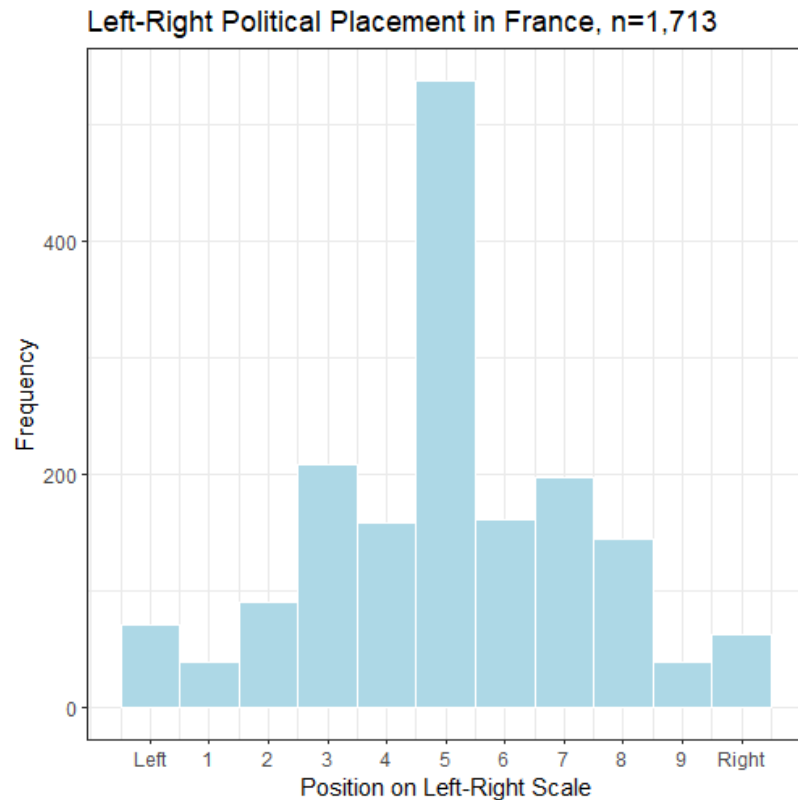


```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram(  
    bins = 11, color = "white",  
    fill = "lightblue") +  
  scale_x_continuous(  
    breaks = seq(0,10,1),  
    labels = c("Left", seq(1,9,1), "Right")) +  
  labs(x = "Position on Left-Right Scale",  
       y = "Frequency",  
       title = "Left-Right Political Placement") +  
  theme_linedraw(base_size=15)
```



# Constructing a Plot Step-by-Step

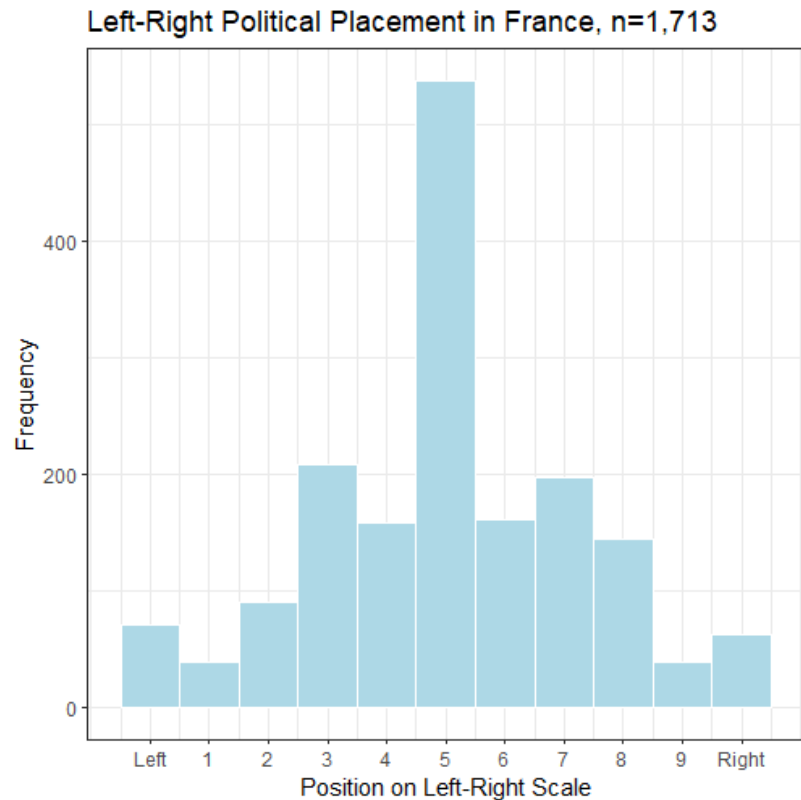
Let's stick to `theme_bw()`.



```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram(  
    bins = 11, color = "white",  
    fill = "lightblue") +  
  scale_x_continuous(  
    breaks = seq(0,10,1),  
    labels = c("Left", seq(1,9,1), "Right")) +  
  labs(x = "Position on Left-Right Scale",  
       y = "Frequency",  
       title = "Left-Right Political Placement") +  
  theme_bw(base_size=15)
```

# Constructing a Plot Step-by-Step

Some people don't like the box around the plot.

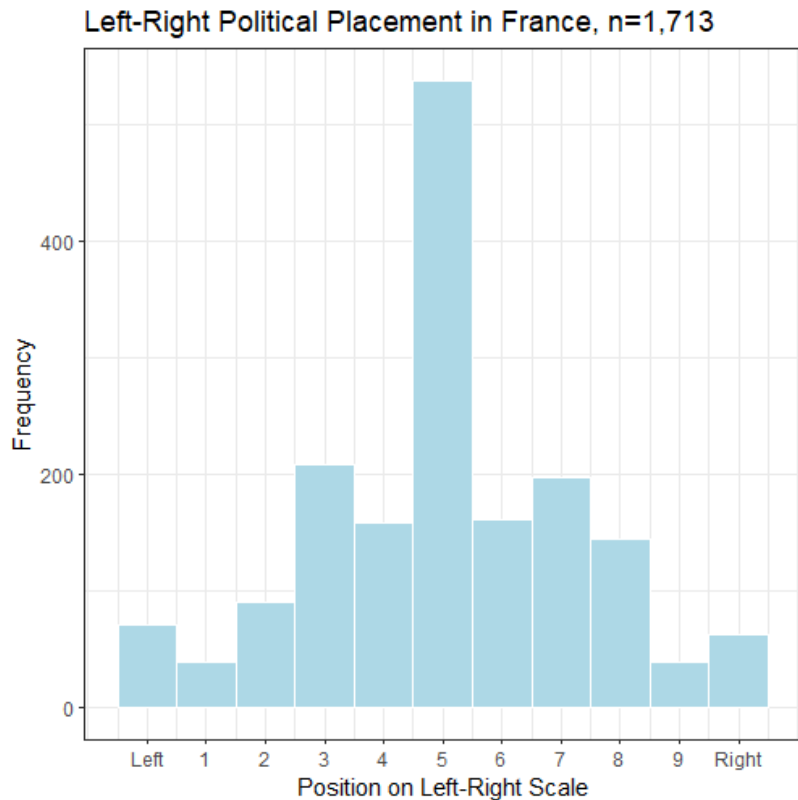


```
ggplot(data = ess10,  
       aes(x = left_right)) +  
  geom_histogram(  
    bins = 11, color = "white",  
    fill = "lightblue") +  
  scale_x_continuous(  
    breaks = seq(0,10,1),  
    labels = c("Left", seq(1,9,1), "Right")) +  
  labs(x = "Position on Left-Right Scale",  
       y = "Frequency",  
       title = "Left-Right Political Placement") +  
  theme_bw(base_size=15)
```

# Constructing a Plot Step-by-Step

Some people don't like the box around the plot.

An additional `theme()` layer lets us further customize our chosen theme.

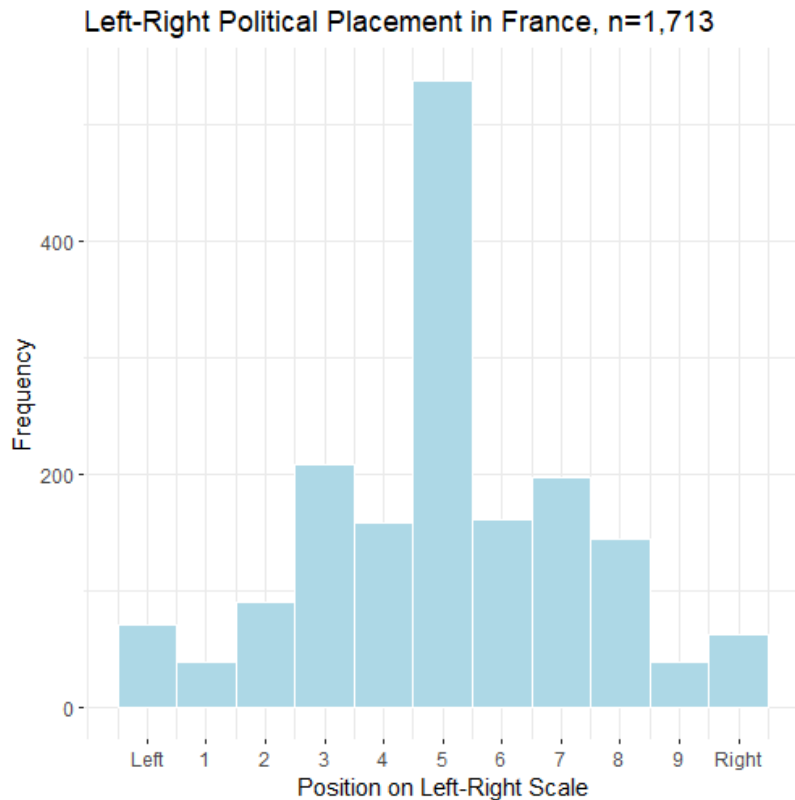


```
ggplot(data = ess10,
       aes(x = left_right)) +
  geom_histogram(
    bins = 11, color = "white",
    fill = "lightblue") +
  scale_x_continuous(
    breaks = seq(0,10,1),
    labels = c("Left", seq(1,9,1), "Right")) +
  labs(x = "Position on Left-Right Scale",
       y = "Frequency",
       title = "Left-Right Political Placement") +
  theme_bw(base_size=15)
```

# Constructing a Plot Step-by-Step

Some people don't like the box around the plot.

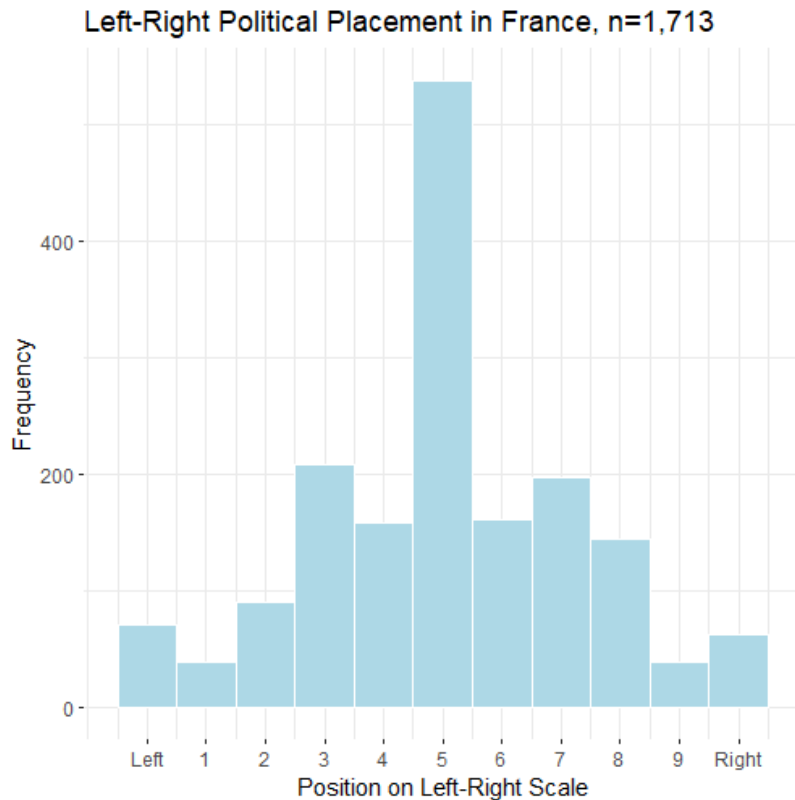
An additional `theme()` layer lets us further customize our chosen theme.



```
ggplot(data = ess10,
       aes(x = left_right)) +
  geom_histogram(
    bins = 11, color = "white",
    fill = "lightblue") +
  scale_x_continuous(
    breaks = seq(0,10,1),
    labels = c("Left", seq(1,9,1), "Right")) +
  labs(x = "Position on Left-Right Scale",
       y = "Frequency",
       title = "Left-Right Political Placement") +
  theme_bw(base_size=15) +
  theme(panel.border = element_blank())
```

# Constructing a Plot Step-by-Step

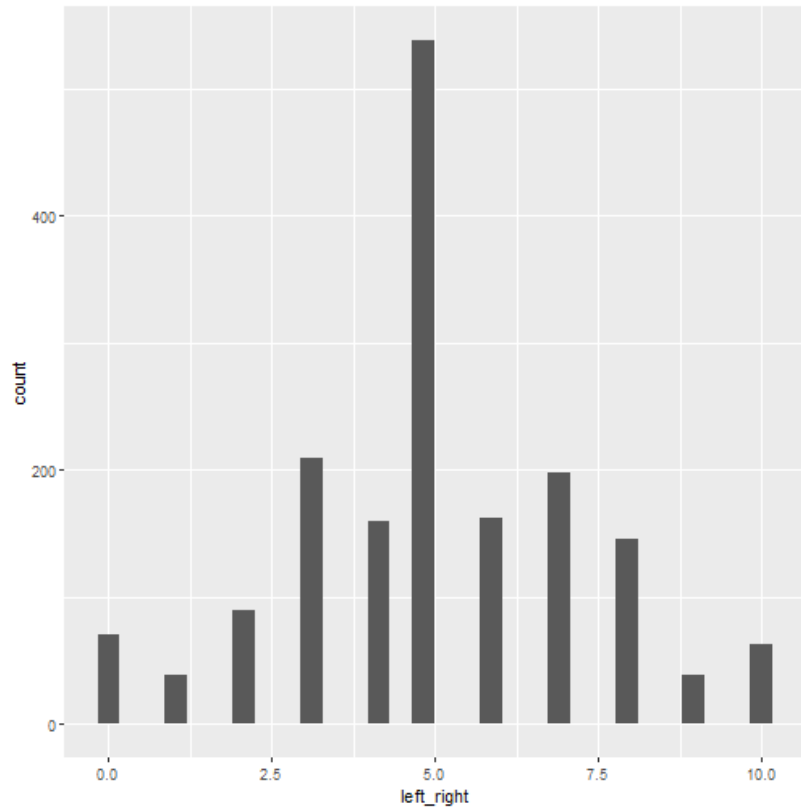
Voilà!



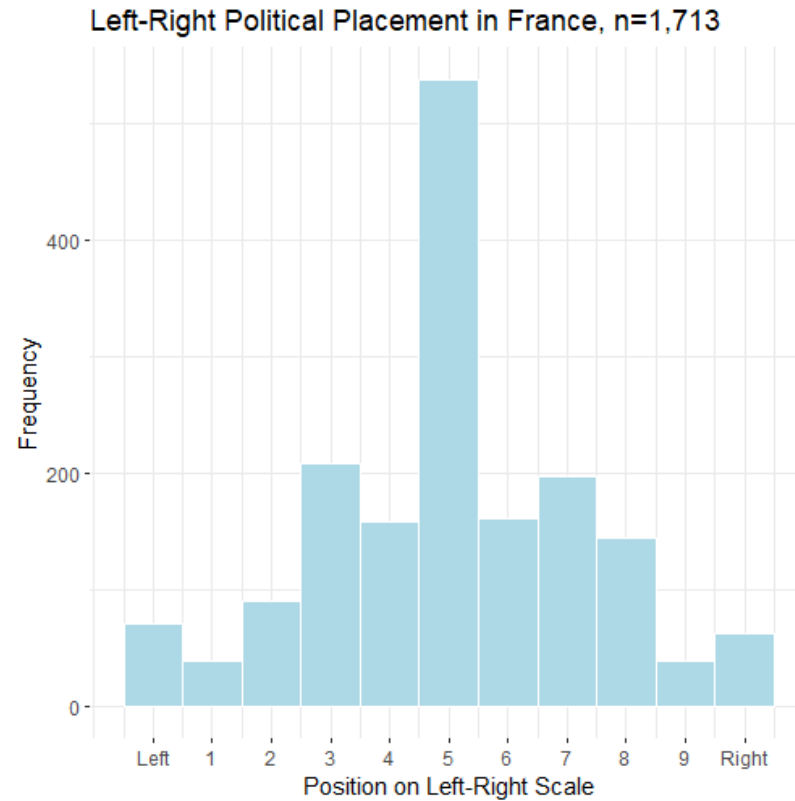
```
ggplot(data = ess10,
       aes(x = left_right)) +
  geom_histogram(
    bins = 11, color = "white",
    fill = "lightblue") +
  scale_x_continuous(
    breaks = seq(0,10,1),
    labels = c("Left", seq(1,9,1), "Right")) +
  labs(x = "Position on Left-Right Scale",
       y = "Frequency",
       title = "Left-Right Political Placement") +
  theme_bw(base_size=15) +
  theme(panel.border = element_blank())
```

# Constructing a Plot Step-by-Step

## How it started



## How it's going



# Steps for Plotting: Summary

## A. Specify necessary layers

1. Data.
2. Aesthetics (*axis definitions*).
3. Geometric objects to use (*plot type*).

## B. Customize your plot

1. Shape, color and size of geometric objects.
2. Fine-tune your x-scale and y-scale (axis ticks, ticks labels).
3. Label your axes, assign a title to your plot.
4. Choose a ggtheme.
5. Customize your theme.

# A Note on Facets

- What if you would like to produce the **same plot** for **different groups** in your data?
- For example, what if we want to **compare** the distribution of **left-right positions across countries**?
- `facet_wrap()` is the way to go.

```
ess10 <- haven::read_dta("./dat/ESS10.dta")
ess10 <- ess10 %>% # subset variables
  select(country = cntry, # sociodemographics
         left_right = lrscale, # attitudes
         ) %>%
  mutate_at(c("country"), as.character)
```



# A Note on Facets

Let's first store our plot in an object called `p`.

```
p <- ggplot(data = ess10,
  aes(x = left_right)) +
  geom_histogram(bins = 11, color = "white",
    fill = "lightblue") +
  scale_x_continuous(
    breaks = seq(0,10,1),
    labels = c("Left", seq(1,9,1), "Right")) +
  labs(x = "Position on Left-Right Scale",
    y = "Frequency",
    title = "Left-Right Political Placement Across Europe") +
  theme_bw(base_size=15) +
  theme(panel.border = element_blank())
```

# A Note on Facets

Now, let's span this plot `p` across the individual groups in the variable `country`.

```
p <- ggplot(data = ess10,
  aes(x = left_right)) +
  geom_histogram(bins = 11, color = "white",
    fill = "lightblue") +
  scale_x_continuous(
    breaks = seq(0,10,1),
    labels = c("Left", seq(1,9,1), "Right")) +
  labs(x = "Position on Left-Right Scale",
    y = "Frequency",
    title = "Left-Right Political Placement Across Europe") +
  theme_bw(base_size=15) +
  theme(panel.border = element_blank())

p + facet_wrap(~ country, ncol = 5)
```

# A Note on Facets

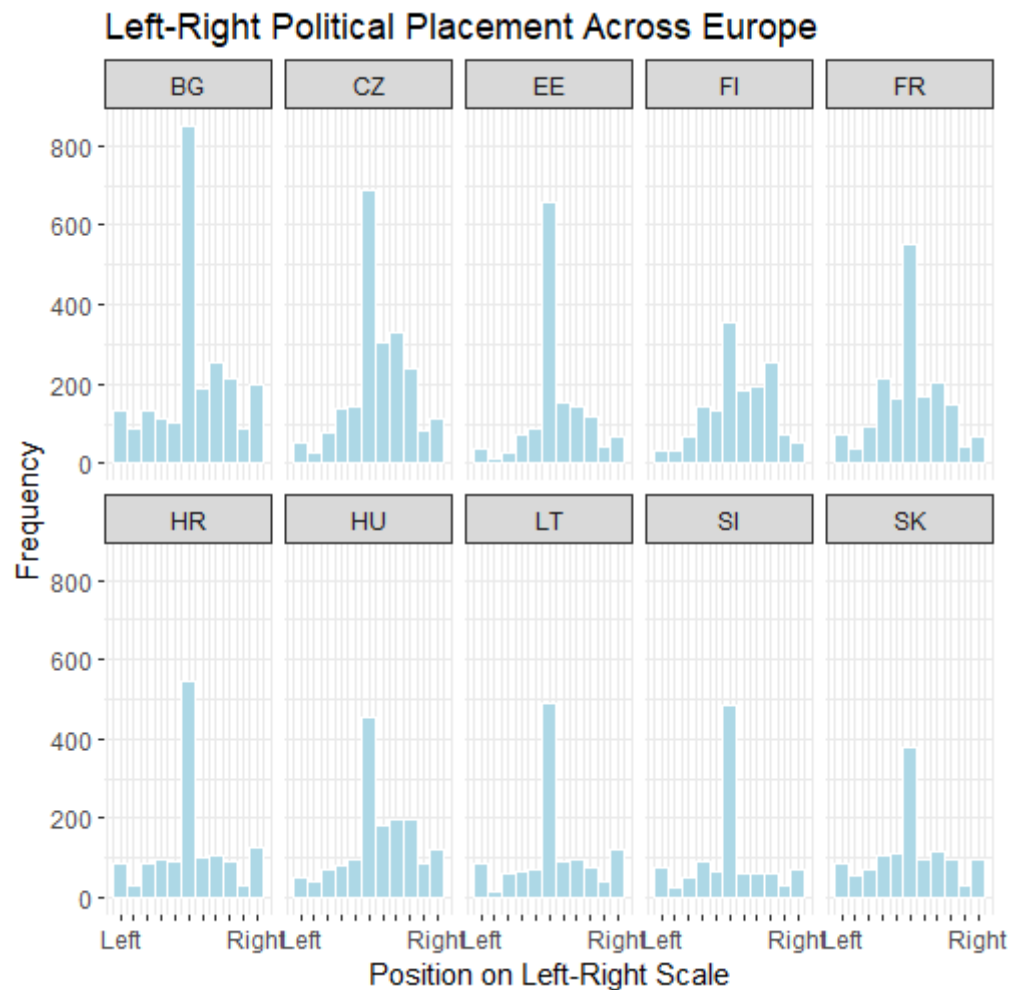
Now, let's span this plot `p` across the individual groups in the variable `country`.

```
p <- ggplot(data = ess10,
  aes(x = left_right)) +
  geom_histogram(bins = 11, color = "white",
    fill = "lightblue") +
  scale_x_continuous(
    breaks = seq(0,10,1),
    labels = c("Left", seq(1,9,1), "Right")) +
  labs(x = "Position on Left-Right Scale",
    y = "Frequency",
    title = "Left-Right Political Placement Across Europe") +
  theme_bw(base_size=15) +
  theme(panel.border = element_blank())
```

```
p + facet_wrap(~ country, ncol = 5)
```



# A Note on Facets



# References

Parts of this course are inspired by the following resources:

- Wickham, Hadley and Garrett Grolemund, 2017. *R for Data Science - Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly.
- Bahnsen, Oke and Guido Ropers, 2022. *Introduction to R for Quantitative Social Science*. Course held as part of the GESIS Workshop Series.
- Breuer, Johannes and Stefan Jünger, 2021. *Introduction to R for Data Analysis*. Course held as part of the GESIS Summer School in Survey Methodology.
- Teaching material developed by Verena Kunz, David Weyrauch, Oliver Rittmann and Viktoriia Semenova.