Recitation 4: Introduction to Research Methods for Politics

Dept. of Politics, NYU

POL-850

Spring 2020

Reminder

Homework 2 out! Due Wednesday 2/26, 5 pm **sharp**, in my mailbox:

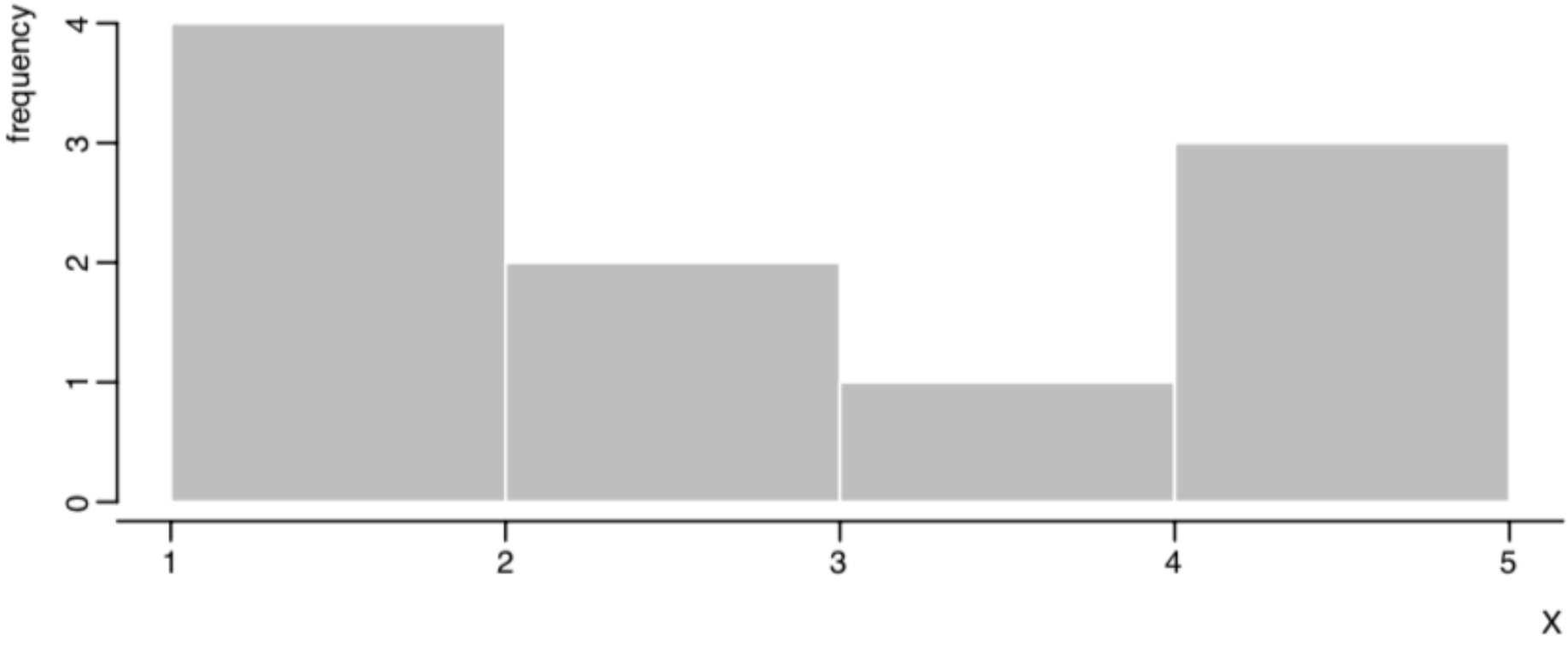
- Only typed work, no handwriting. R Markdown strongly suggested.
- When answering a question, please refer to which quantities/statistics it is based on.
- Tip: **do not leave blank answers**. Show us your work even if you deem your conclusions incorrect, please.

1. Histograms (DSS 3.4.4)

Histograms

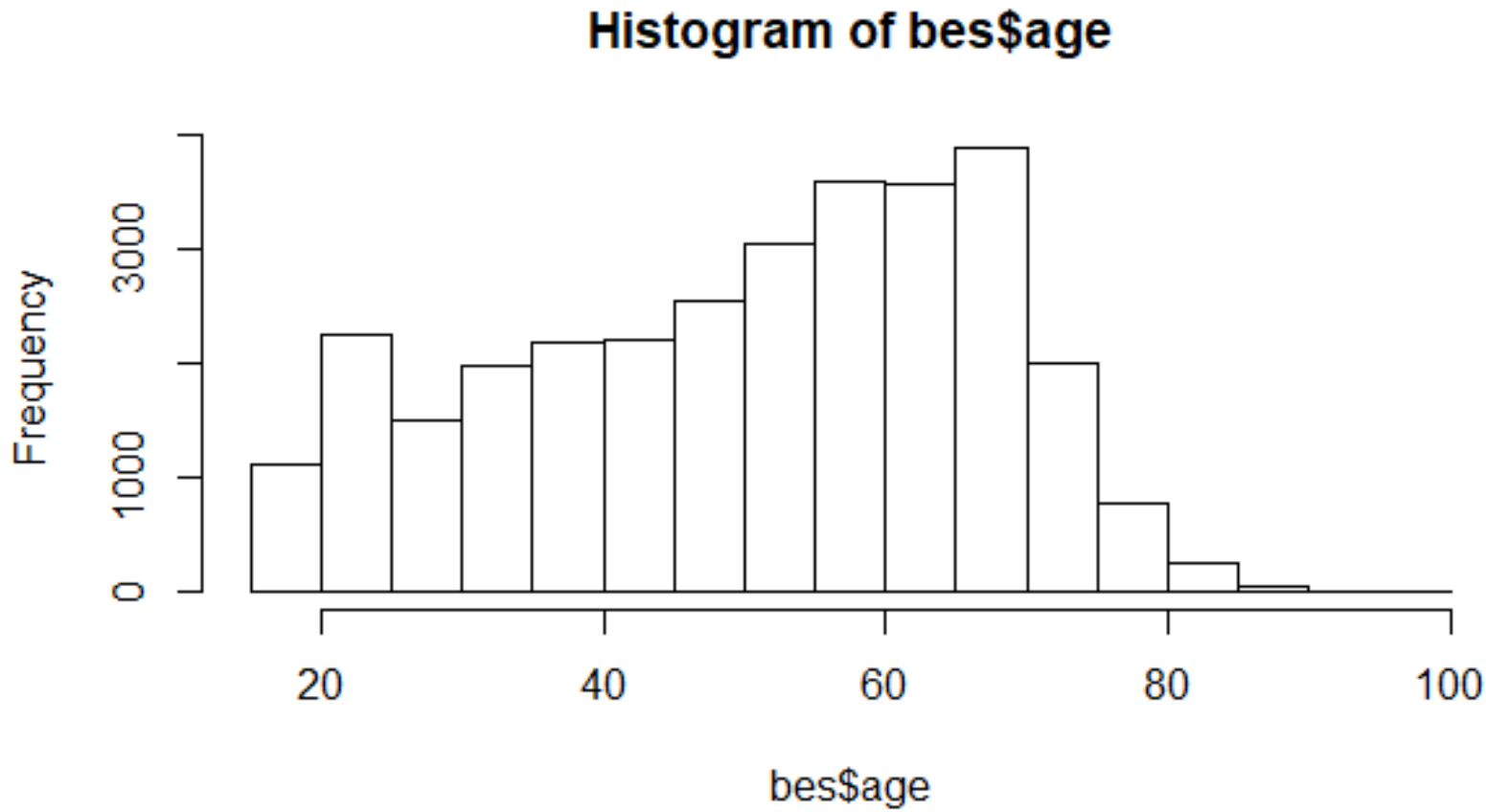
- A **histogram** is a graphical representation of a variable's distribution, made up of bins (rectangles) of different heights
- The position of the bins along the x-axis (the horizontal axis) indicates the interval of values
- The height of the bins represents how often the variable takes the values in the corresponding interval
- Use hist(data$variable)

For example, if $X = \{1.1, 1.1, 1.2, 1.3, 2.2, 2.6, 3.3, 4.3, 4.3, 4.8\}$, the histogram of X is:



A Basic Example w/ Brexit Data

- Produce a density histogram of the variable age:
- hist(bes$age)

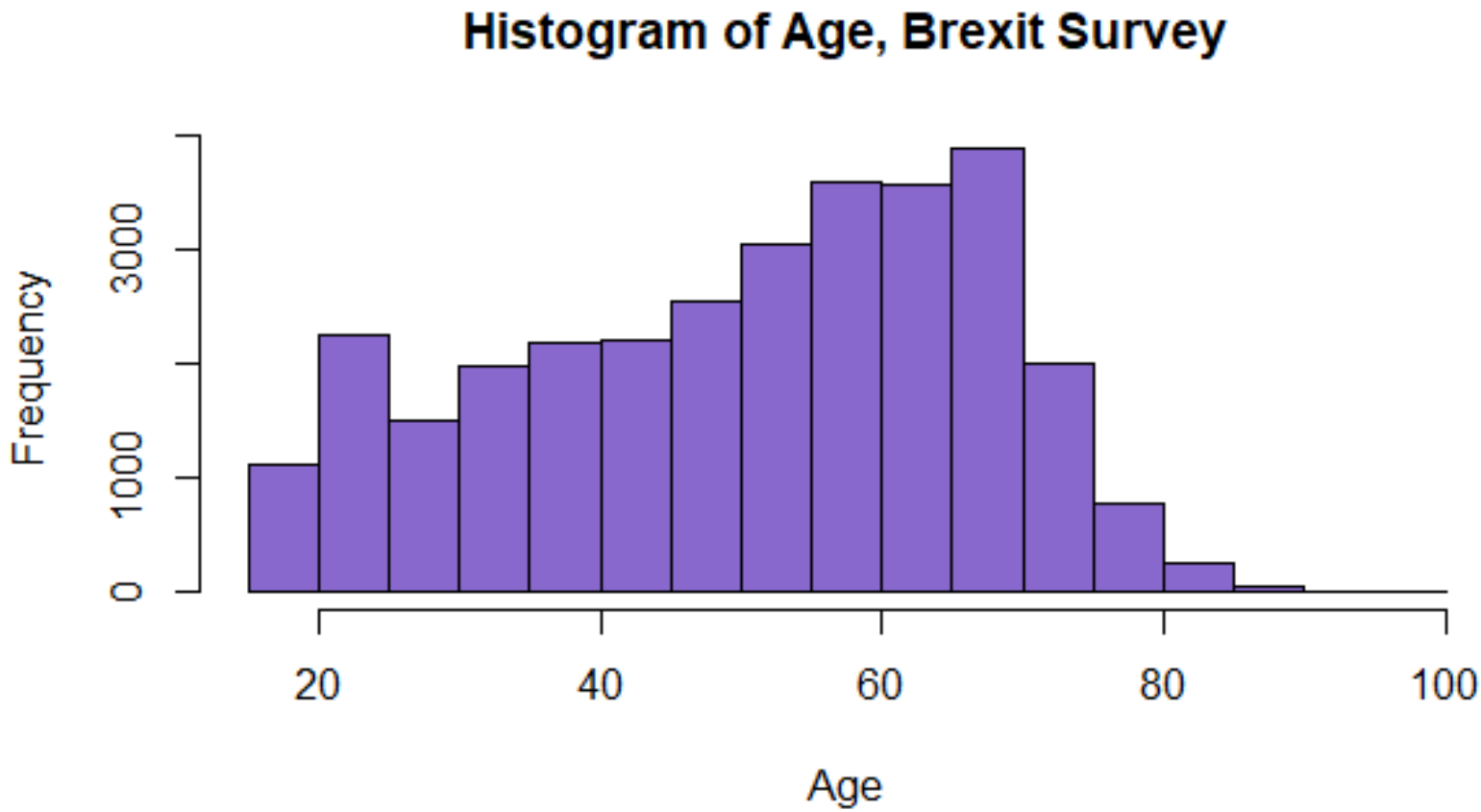


Making It Nicer

- Can add title to graph: hist(data$variable, main = "Insert Title Here")
- Can add x and y axis labels to graph: hist(data$variable, main = "Insert Title Here", xlab = "X Axis Label", ylab = "Y Axis Label")
- Can add vertical line to graph at a specified value or a result of a function (mean, median, etc.): abline(v = mean(data$variable))
- Can add horizontal line to graph at a specified value or a result of a function (mean, median, etc.): abline(h = mean(data$variable))

Making it Nicer

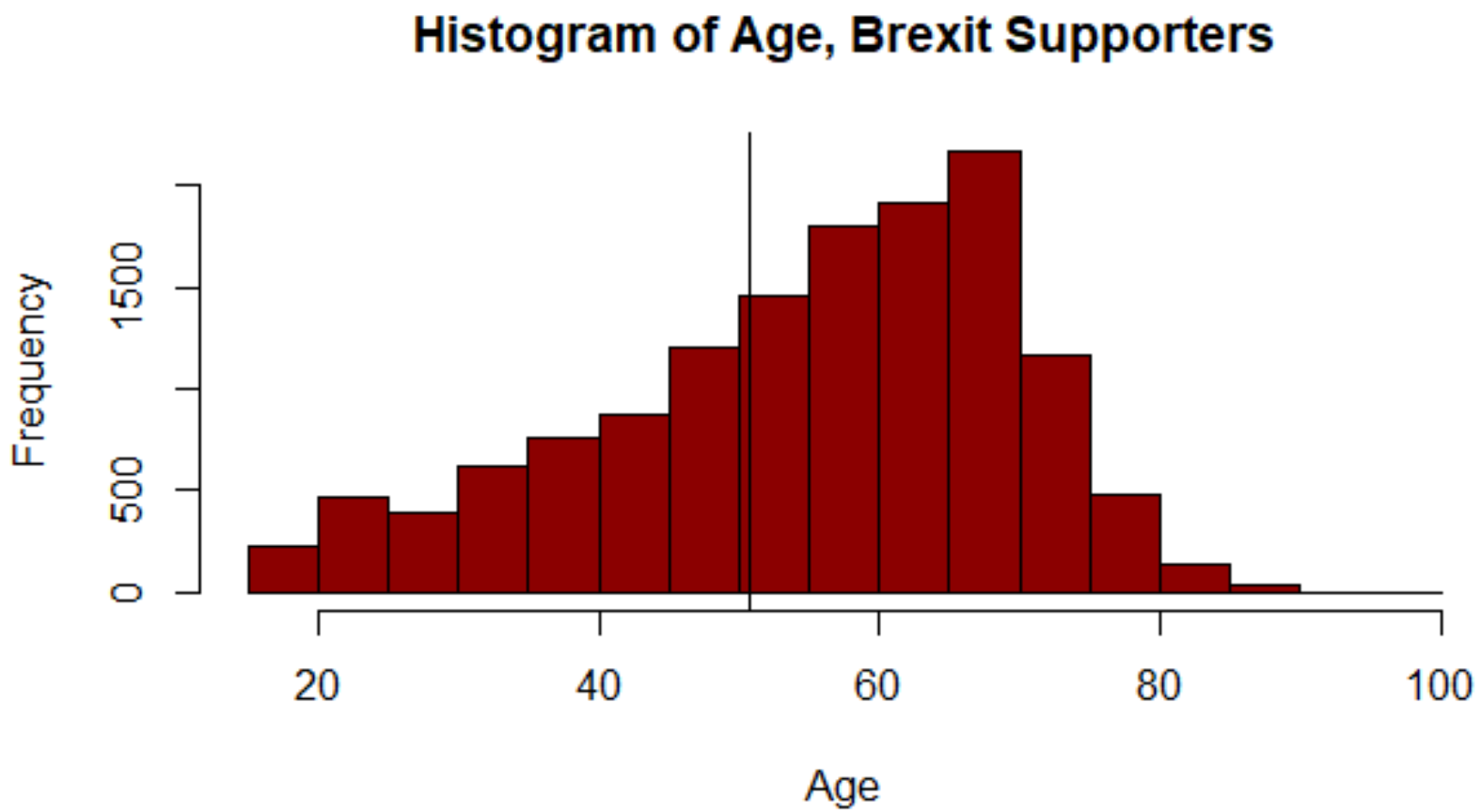
- hist(bes$age, main = "Histogram of Age, Brexit Survey", xlab = "Age", col = "mediumpurple3")



Selecting Groups and Adding Reference Lines

- Most of the times, we are interested in looking at the distribution of a variable by values of another
- Example: we may want to see how age varies by Brexit support, to understand whether supporters are younger/older, on average:
- hist(bes*age*[*bes*leave == 1], main = "Histogram of Age, Brexit Survey", xlab = "Age", col = "red4")
- Then, **outside of the hist function**, we can add a reference line to the mean age in the survey:
- abline(v=mean(bes$age, na.rm = TRUE))

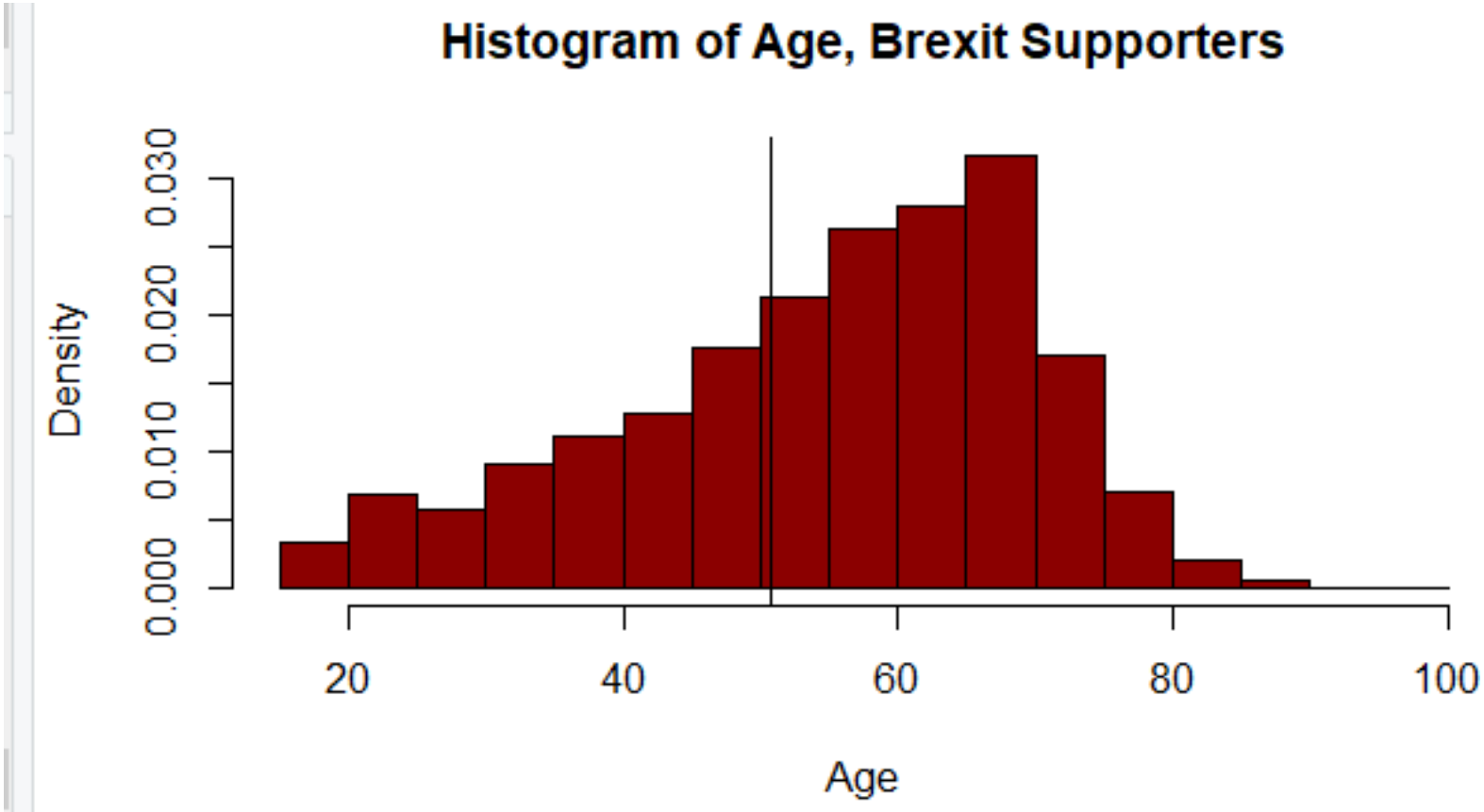
Selecting Groups and Adding Reference Lines



Density histograms

- hist() shows the frequency of a variable by default, but density histograms are especially useful for comparing groups with substantially different numbers of observations
- In a **density histogram**, the area of each bin (rectangles) is equivalent to the proportion of observations that fall in that bin
- The areas of the blocks sum to 1 (100%)
- hist(data$variable, freq = FALSE)

Density Histograms



2. Scatterplots (DSS 3.5.1)

Visualizing Relationships

Our research questions usually involve relationships between variables:

- Does small class size increase students' achievement?
- Does criminal record decrease callback rate?
- You name it...

Therefore, we want an effective way to visualize a variable as **a function of another**. Scatterplots serve this purpose.

Understanding Scatterplots

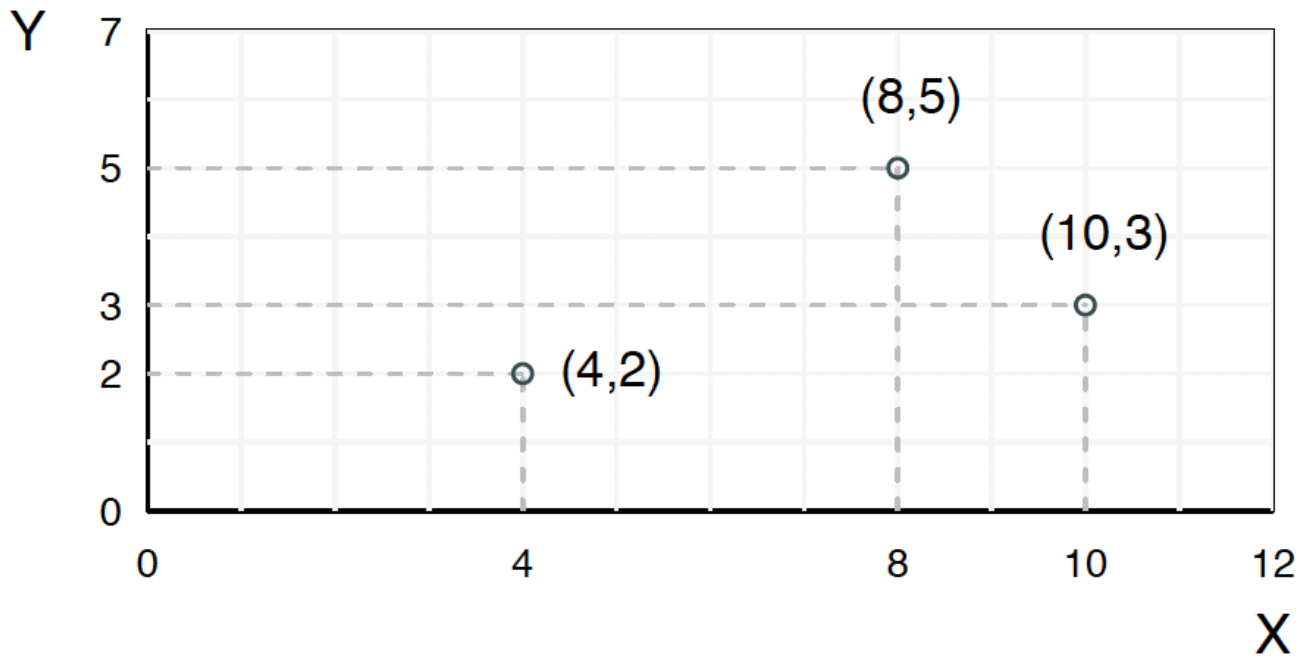
Scatterplots feature three main elements:

- A y-axis, measuring our **outcome variable** (Y)
- An x-axis, measuring our **explanatory variable** (X)
- A **cloud** of points, each of which represents an observation in our data. The overall shape and orientation of the cloud informs us about the correlation between Y and X.

(a) The two variables of interest:

i	X	Y
1	4	2
2	8	5
3	10	3

(b) The scatter plot:



A Practical Example

variable	description
name	name of the district
leave	vote share received by the leave camp (in percents)
high_education	proportion of residents with an undergraduate degree, professional qualification, or equivalent (in percents)

Plot Code Syntax

- Order matters!
- plot(data$x_variable, data$y_variable)
- Title, axis labels and other things can be added in the same way as for histograms
- Can add line type to `abline`
- E.g., lty="dashed"

How Does it Look?