# Web Data Collection with R
## How HTML Works

Peter Meißner / 2016-02-29 – 2016-03-04 / ECPR WSMT

How HTML works . . .

**How HTML works . . .**

# How HTML works ... basics

```
<html>
  <head>
    <title>PageTitle</title>
  </head>
  <body>
    <p>Hallo World.</p>
  <body>
</html>
```

- ► HTML is one of the possible formats to get back by server
- ► HTML is plain text
- ► HTML is markup (Hyper Text Markup Language)
    - ► tags and nodes
    - ► attributes
    - ► content
- ► HTML is tree structured

# How HTML works . . . some special features

- ▶ tags have predefined meaning
    - ▶ e.g. `<p>...</p>` for paragraph or `<a href="...">...</a>` for links
- ▶ includs further external ressources via, e.g.:
    - ▶ `<link ... href="...">`
    - ▶ `<script src="..."></script>`
    - ▶ `<img src="...">`
- ▶ HTML forms (e.g. search bar)
    - ▶ gather information to be send to server later on `<form><input>...</...`
- ▶ CSS (Cascading Style Sheets) `<p class="redintro">...`
- ▶ Javascript
    - ▶ a computer language (like e.g. R)
    - ▶ understood and executed by browser
    - ▶ usually manipulating the HTML (tree) received by server