



WORKING WITH WEB DATA IN R

**CSS**

Oliver Keyes  
Instructor



# CSS example

## CSS

```
.class_a {  
  color: black;  
}  
  
.class_b {  
  color: red;  
}
```

## HTML

```
<a class = "class_a" href = "http://en.wikipedia.org/"> This is black </a>  
  
<a class = "class_b" href = "http://en.wikipedia.org/"> This is red </a>
```



# CSS versus XPATH

- CSS scraping looks for class or id names (e.g. "class\_a")
- Much like XPATH
- ...but grabs multiple things
- Still use `html_node()` (or `html_nodes()`), but with `css` argument.



WORKING WITH WEB DATA IN R

**Let's practice!**



WORKING WITH WEB DATA IN R

# **Final case study: Introduction**

Charlotte Wickham  
Instructor



# What we'll cover

Extracting an infobox from a Wikipedia page



# What we'll cover

Extracting an infobox from a Wikipedia page

1. Get the XML content of a Wikipedia page through API



# What we'll cover

Extracting an infobox from a Wikipedia page

1. Get the XML content of a Wikipedia page through API
2. Extract the infobox from the page





# What we'll cover

Extracting an infobox from a Wikipedia page

1. Get the XML content of a Wikipedia page through API
2. Extract the infobox from the page
3. Clean it up and turn it into a data frame



# What we'll cover

Extracting an infobox from a Wikipedia page

1. Get the XML content of a Wikipedia page through API
2. Extract the infobox from the page
3. Clean it up and turn it into a data frame
4. Turn it into a function



WORKING WITH WEB DATA IN R

**Let's practice!**



WORKING WITH WEB DATA IN R

# Wrap Up

Oliver Keyes & Charlotte Wickham  
Instructors



# Wrap up

- Downloading and reading flat files
- Designing and using API clients
- Web scraping: XPATHs & CSS