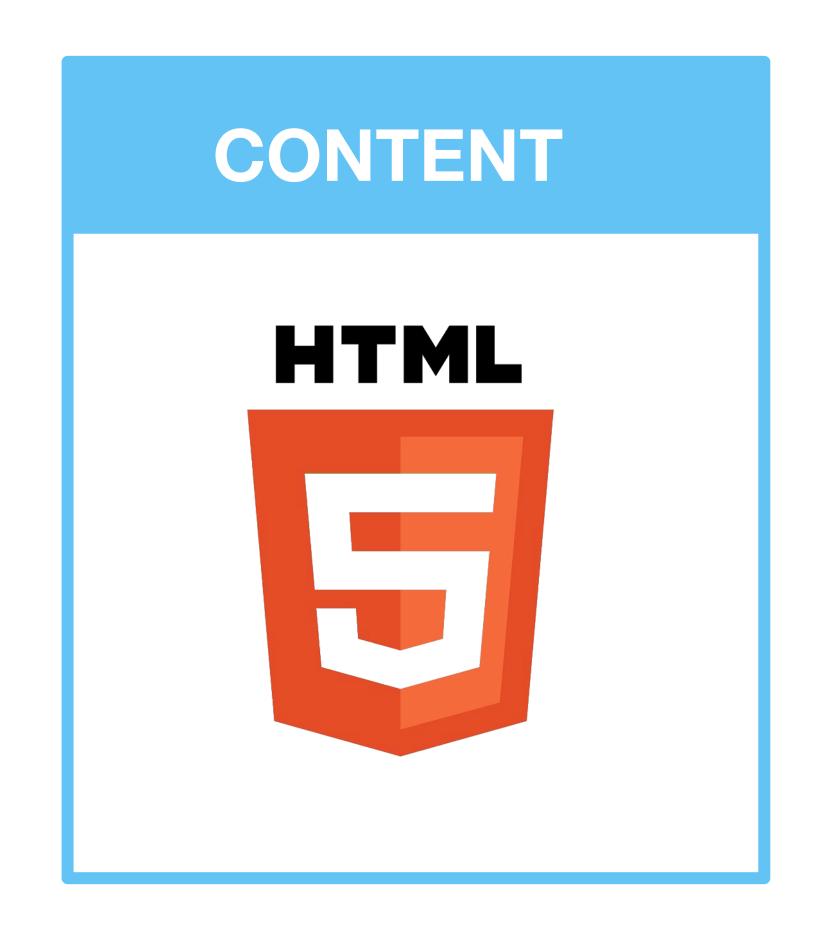
# HTML BASICS FOR DATA WARRIORS



# Germán Álvarez | Web Dev LT @ Ironhack

<u>linkedin.com/in/german-alvarez-dev</u>

# The web







#### **HTML** features



- HTML: Hypertext Markup Language
- 5th revision: semantic!
- Not a programming language
- Used by web pages and applications
- Understood by browsers
- <tagged> contents with attributes
- Hierarchical structures: parents, childs, siblings

# Metadata content

# Sectioning content

# Heading content

# Phrasing content

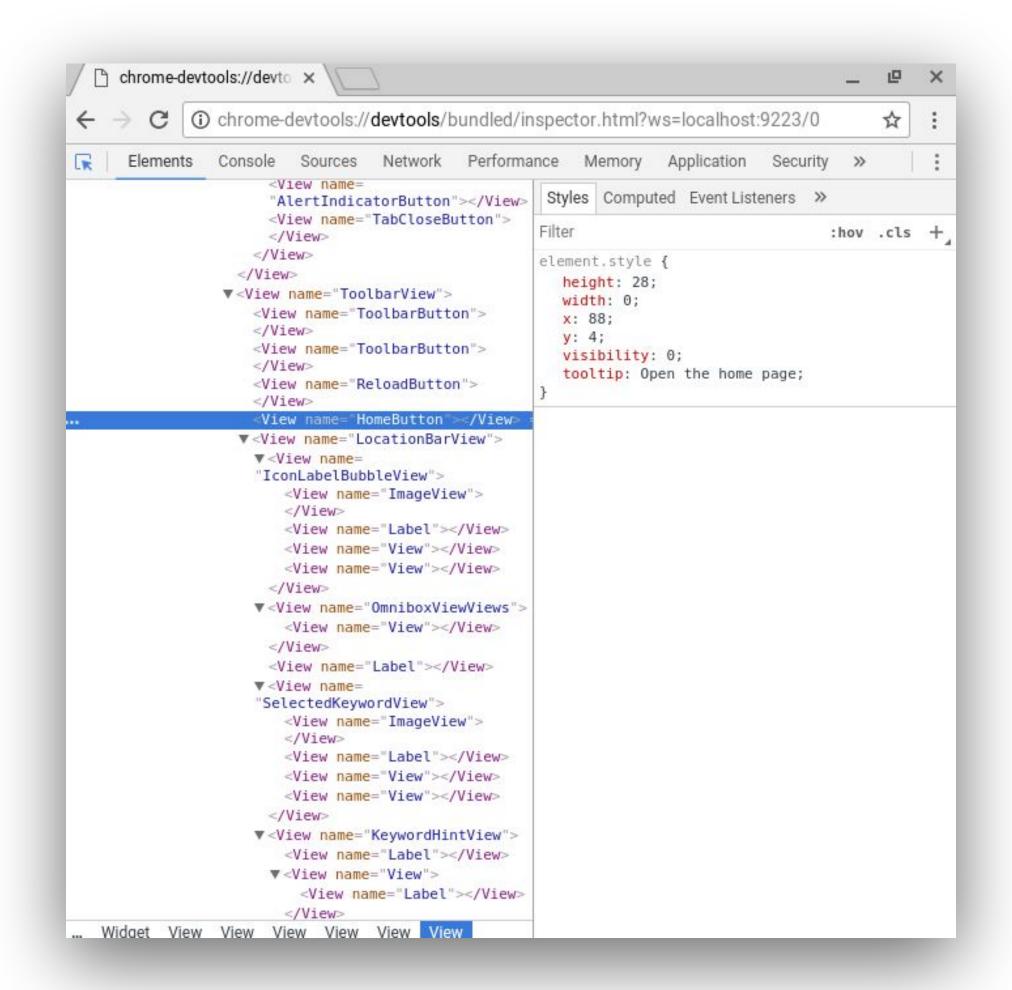
```
<abbr> <audio> <b> <bdo> <br> <but color="block"><abbr> <audio> <b <br/> <bdo> <br/> <br/> <bdo> <br/> <bdo> <br/> <br/> <bdo> <br/> <br/> <bdo> <br/> <br/> <br/> <br/> <br/> <bdo> <br/> <br/>
```

# Embeded content

# Tabulated content

#### **Chrome Dev Tools**

## PRACTICE: website inspection



## Selecting techniques

#### Basic

Tag selecting Class selecting ID selecting

#### Advanced

Descendants
Direct descendants
Combined
Siblings
Has attribute
Attribute value
Child number

## JS / BS equivalence

#### JS

```
document.querySelectorAll("p")

document.querySelectorAll("p a")

document.querySelectorAll("p > a")

document.querySelectorAll("p > a:nth-of-type(2)")
```

#### BS

```
soup.select("p")
soup.select("p a")
soup.select("p > a")
soup.select("p > a:nth-of-type(2)")
```

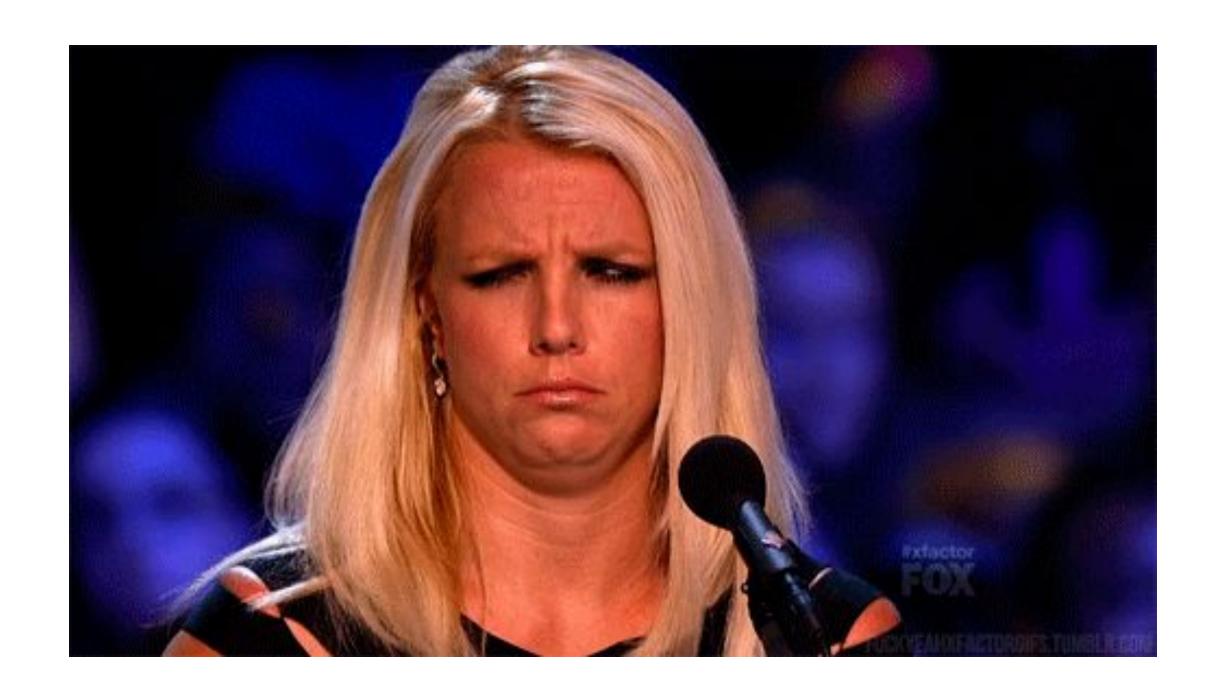
#### Selection techniques

PRACTICE: website cloning & scraping



rcdb.com

# Ask me!



Questions? GO!

# THANK YOU!