Lecture 11

408/508 *Computational Techniques for Linguists*

Today

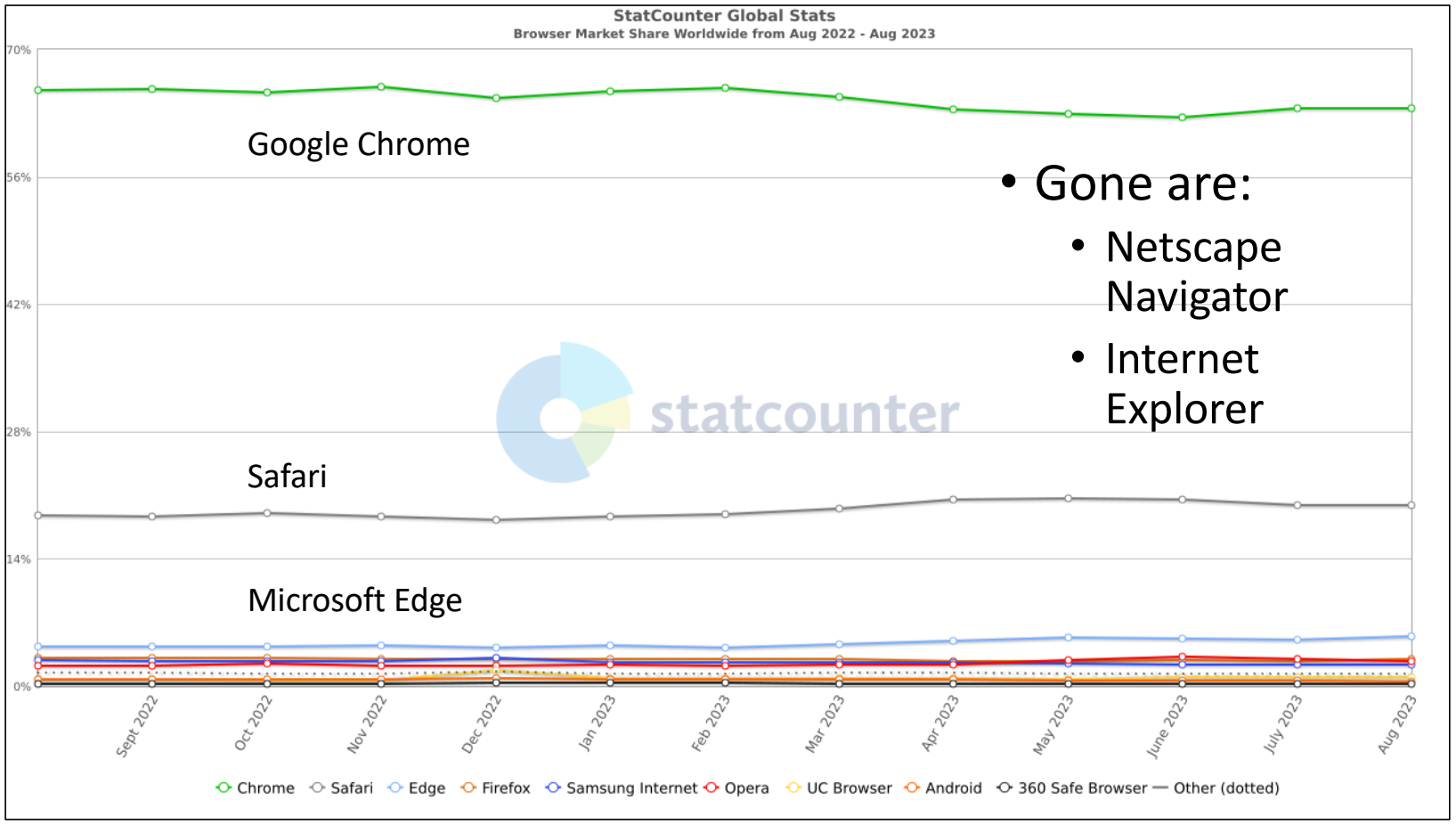
- New topic:
  - start with html5
    - `html`
    - `css`
    - `javascript`
  - *leading to building your own webserver*

Browser

- Nowadays browsers are very powerful in their own right (can compute locally, not just communicate with a webserver)

CSS (Cascading Style Sheets)
SVG (Scalable Vector Graphics)
- cf. HTML5 canvas
Javascript
- *programming language*
DOM (Domain Object Model)
- *programmatic access to documents*
Websockets
- An API to interact with regular programs

Browsers

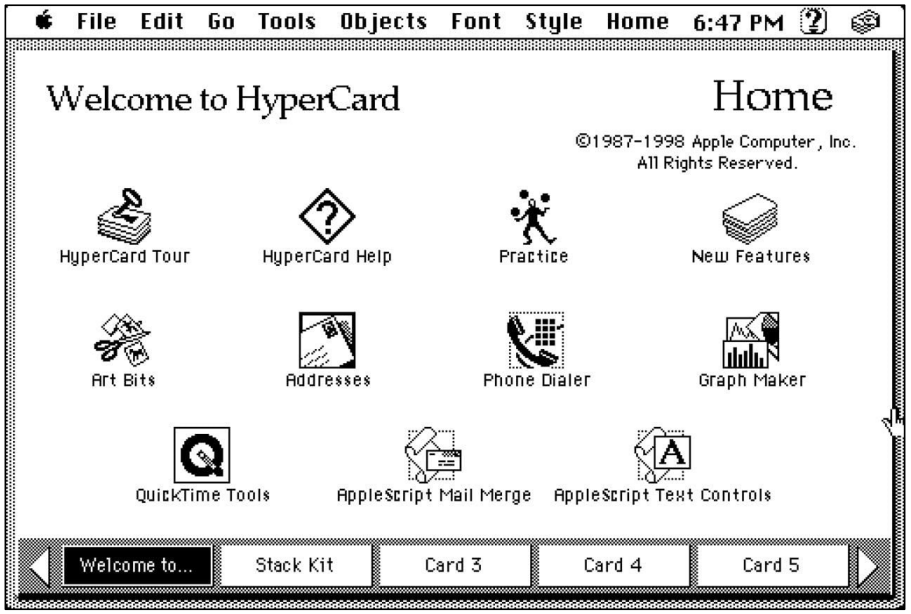


- Gone are:
  - Netscape Navigator
  - Internet Explorer

HTML

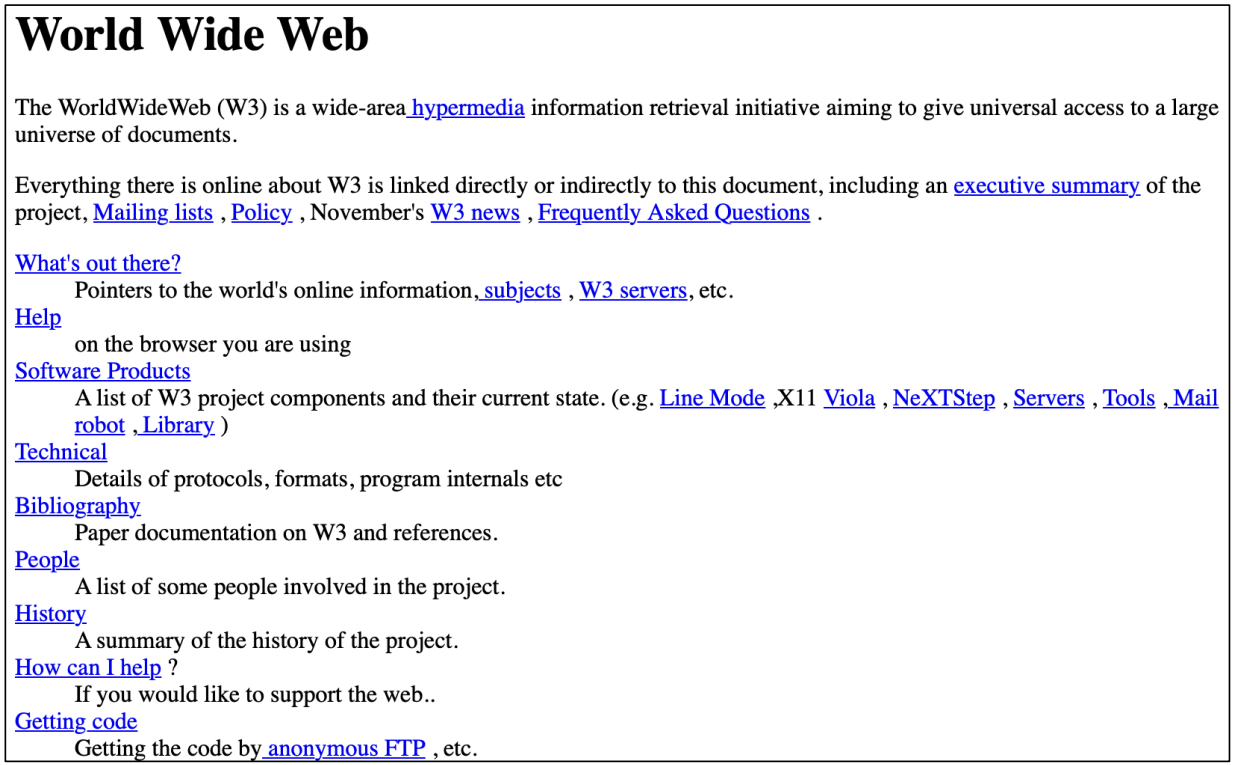
- HTML: Hypertext Markup Language
  - Web browser: can read and render pages written in HTML
  - **currently**: HTML5
- What is "hypertext"?
  - linked content (Nelson, 1963)
  - nowadays: selected text/images/video can have arbitrary associated actions
- Hypercard for the Macintosh (1987)
- World Wide Web (WWW) (1992)
  - World Wide Web Consortium (W3C)

Hypercard



- Before WWW
  - 1987
  - card stack with links
- Myst
  - 1993
  - a computer game
  - developed using Hypercard

World's first webpage



World Wide Web

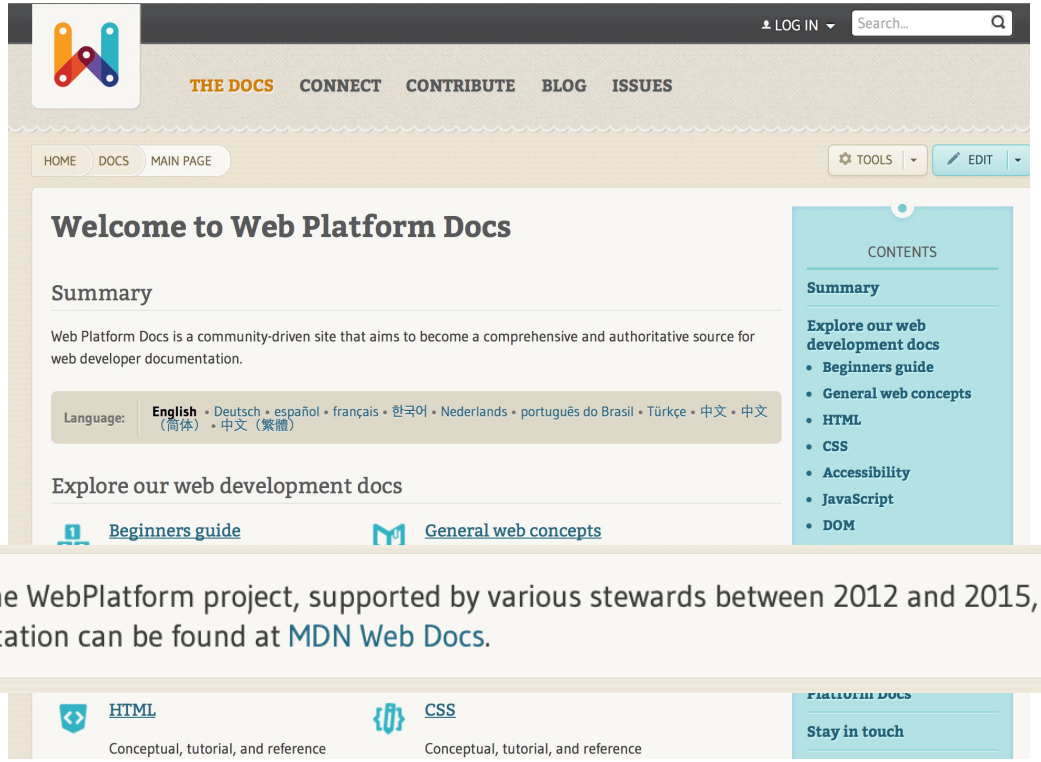
The WorldWideWeb (W3) is a wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including an executive summary of the project, Mailing lists , Policy , November's W3 news , Frequently Asked Questions .

What's out there?
Pointers to the world's online information, subjects , W3 servers, etc.
Help
on the browser you are using
Software Products
A list of W3 project components and their current state. (e.g. Line Mode ,X11 Viola , NeXTStep , Servers , Tools , Mail robot , Library)
Technical
Details of protocols, formats, program internals etc
Bibliography
Paper documentation on W3 and references.
People
A list of some people involved in the project.
History
A summary of the history of the project.
How can I help ?
If you would like to support the web..
Getting code
Getting the code by anonymous FTP , etc.

- At CERN:
  - August 6 1991
  - by Tim Berners-Lee
- Notice:
  - text, no graphics
  - no fancy fonts
  - no sidebars
  - no advertising
  - no Javascript

Reference



https://webplatform.github.io

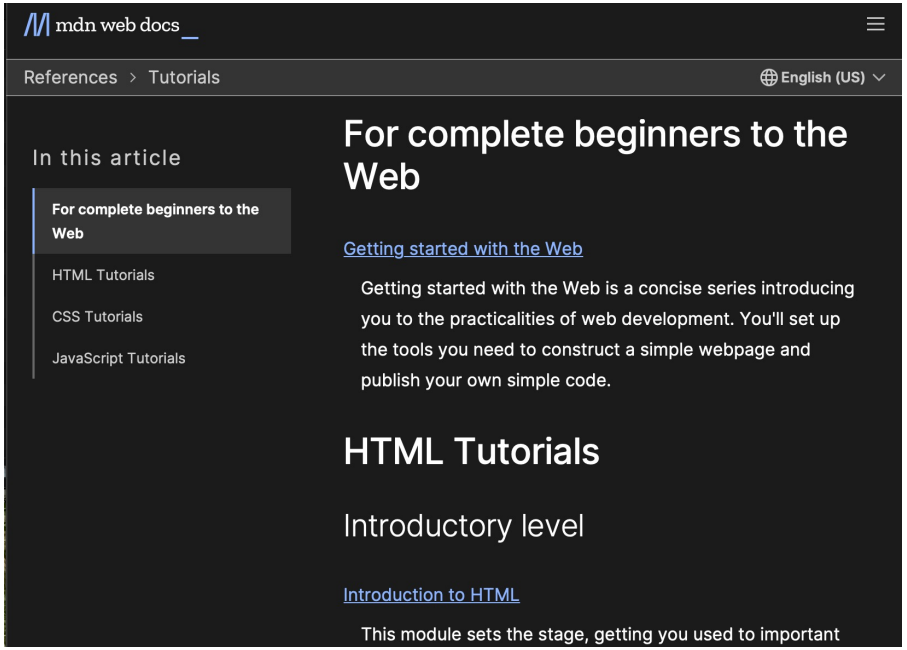
Now deprecated…

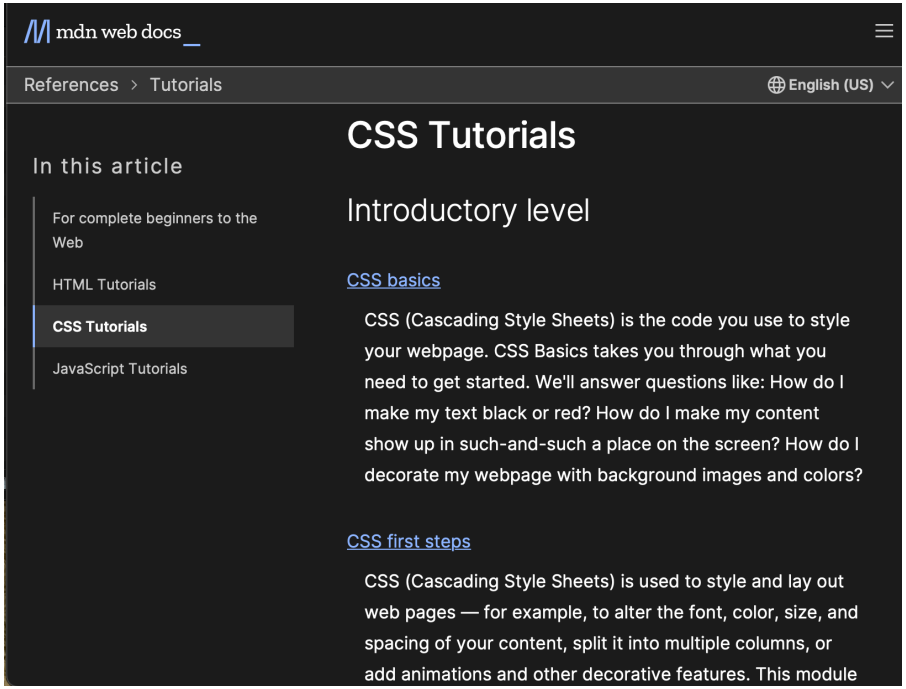
MDN


mdn_
https://developer.mozilla.org/en-US/
Resources for Developers,
by Developers_
Documenting web technologies, including CSS, HTML, and JavaScript, since 2005.

Reference

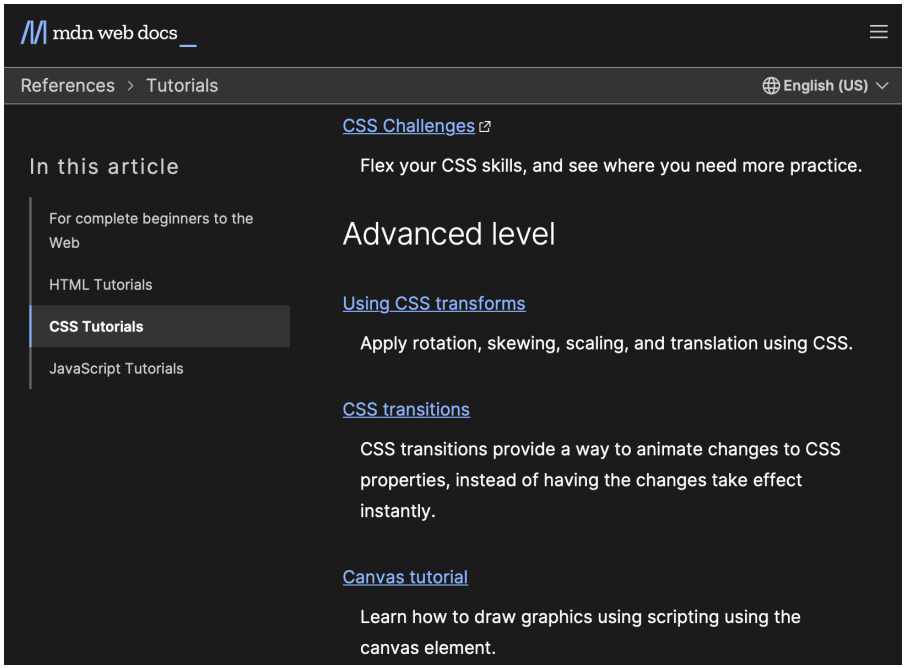
https://developer.mozilla.org/en-US/docs/Web/Tutorials

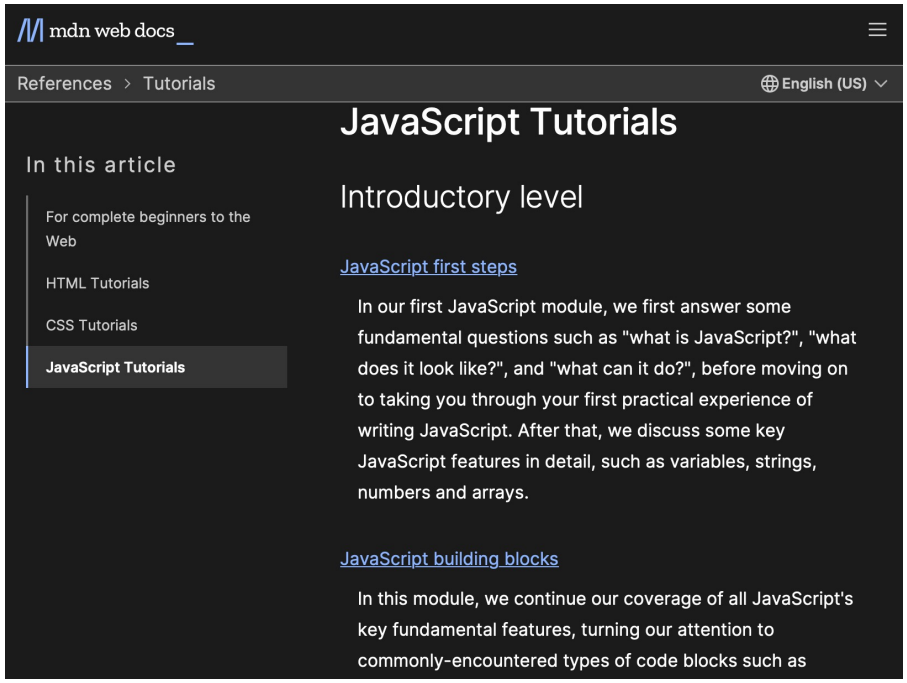




Reference

https://developer.mozilla.org/en-US/docs/Web/Tutorials

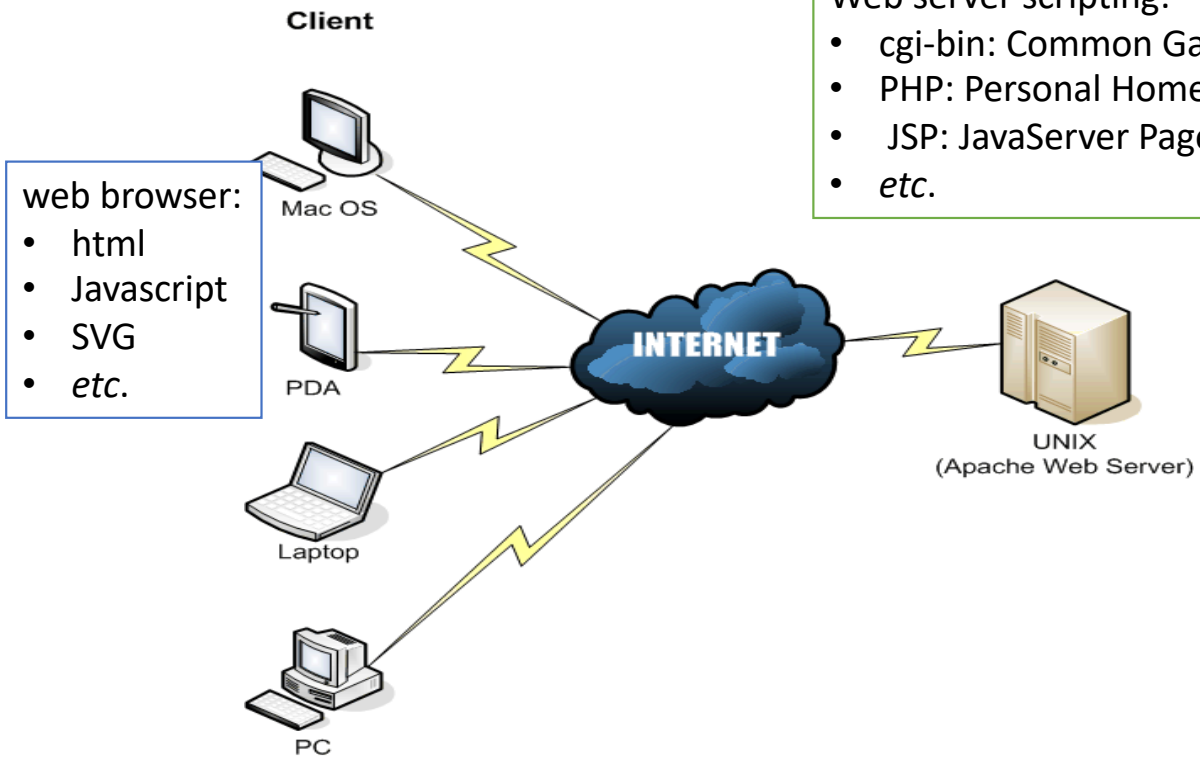




Client-side web development

- **HTML**:
  - structure of content
- CSS (cascading style sheets):
  - presentation
- Javascript
  - scripting language
- DOM (document object model):
  - hierarchical representation of webpage
- SVG (scalable vector graphics):
  - 2D graphical objects and methods

Client/server model



web browser:
- html
- Javascript
- SVG
- *etc*.

Web server scripting:
- cgi-bin: Common Gateway Interface
- PHP: Personal Home Page
- JSP: JavaServer Pages
- *etc*.

http://www.visualbuilder.com/jsp/tutorial/introduction-to-jsp/

HTML

boilerplate inserted by my Aquamacs editor:

```
 1 <!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML//EN">
 2 <html> <head>
 3 <title></title>
 4 </head>
 5
 6 <body>
 7 <h1></h1>
 8
 9
10
11 <hr>
12 <address></address>
13 <!-- hhmts start --><!-- hhmts end -->
14 </body> </html>
```

- First line:
  `<!DOCTYPE HTML>`
  - signifies HTML5
- Tags:
  `<tag> … </tag>`
  - `html`
  - `head`: `title`, `style`, javascript definitions etc.
  - `body`: body of the document
  - `h1`: heading level 1 (1-6)
  - `address`: contact information
- "*self-closing*" tags:
  - `hr`: horizontal rule
  - `br`: line break optional: `<br/>`
- optionally paired:
  - `<p> .. </p>`: paragraph
- comment:
  - `<!-- … -->`

HTML

- hypertext (link):
  - text (text presented in blue)
- URL: uniform resource locator
  - Examples:
    - http://sandiway.arizona.edu/
    - http://nlp.stanford.edu:8080/parser/
    - https://netid.arizona.edu/newid.php (PHP)
    - http://localhost/perl/test.pl (mod_perl program)
  - Format:
    - protocol://host(:port)/path
    - protocol://host(:port)/path?query
      - protocol = http (hypertext transfer protocol)
      - port = TCP/IP port number

HTML

- Images:
  - <img src=URL>
  - attribute: src (required)
  - value: URL (or filename etc.) (jpg, gif, png supported, see note below)
  - attribute: alt (supposed to be required)
  - value: text
  - attribute: height
  - value: pixels
  - attribute: width
  - value: pixels
  - attribute: align (not in HTML5)
  - value: top|bottom|middle|left|right
- Can embed:
  - `<a href=URL><img src=URL></a>`
- **Note:** http://en.wikipedia.org/wiki/Comparison_of_web_browsers#Image_format_support

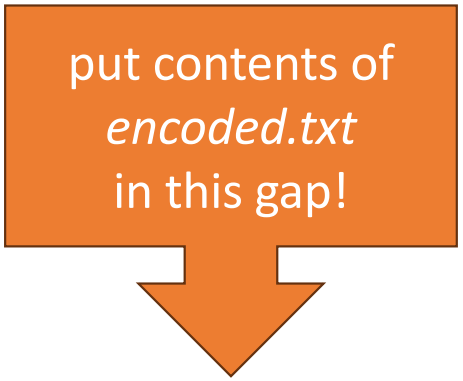
HTML

- Images can be embedded inside the file via base64 encoding:

```
<img width="70px"
src="
goAAAANSUhEUgAAAIIAAAB6CAYAAABzwouJA
AAMFmlDQ1BJQ0MgUHJvZmlsZQAASImVVwdUk
8kWnr+kEJJQAqFICb0JUqQLhF4FpIONkAQIJ
UJCULGjiwquXUSxoqsiCq4FkLViwYII9v5AR
EVZFws2VN6kgK6vnXfPmX++3Ln3zncnd+bMA
KBiz8nLy0FVAcgVFohigv1YSckpLFInUAJag
A5owIbDFef5RkdHACjD/d/l/S2ASPvrttJY/
zr+X0WNxxdzAUCiIU7jibm5EB8GANfm5okKA
CC0Qr3J …">
```

HTML

- On Ubuntu:
  - `base64 imagefile > encoded.txt`
- macOS:
  - `base64 –i imagefile –o encoded.txt`
- base64 in Windows:
  - `certutil –encode imagefile encoded.txt`
- HTML file:
  - `<img width="200px" src=" ">`

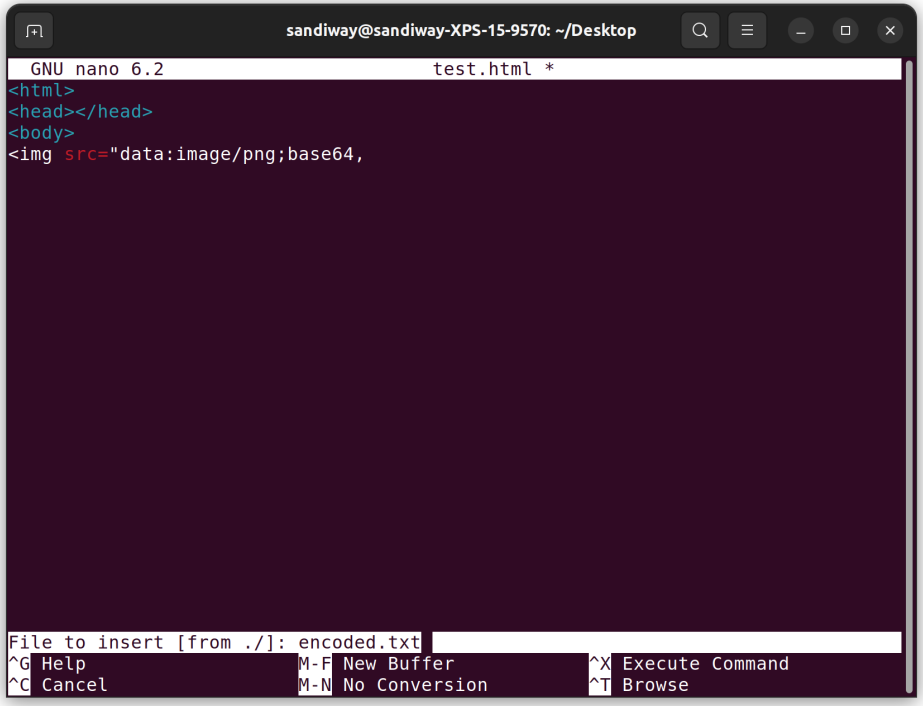


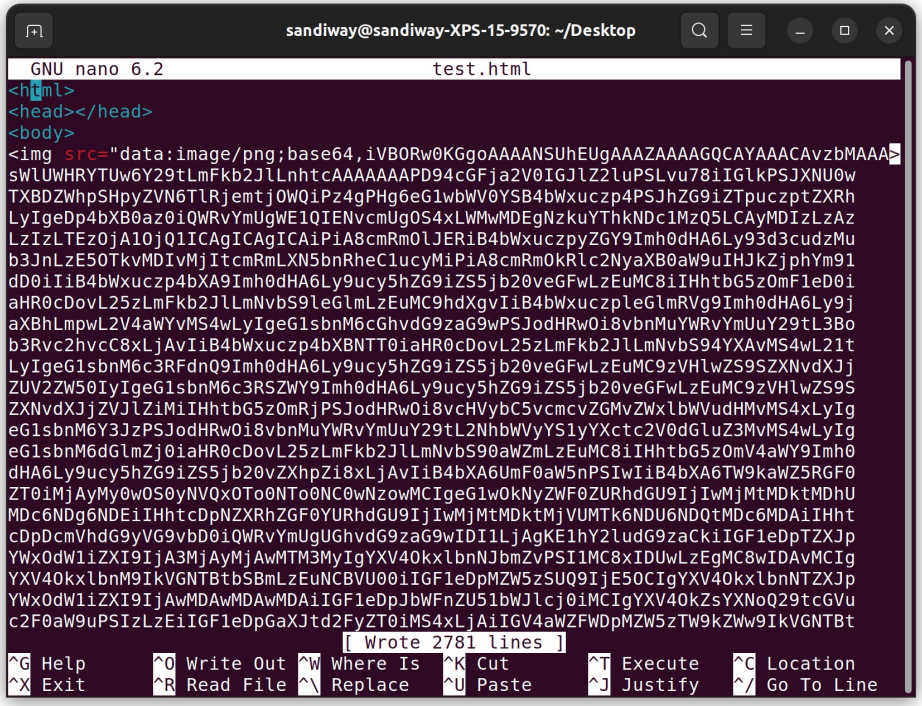
HTML

- Example (macOS):

```
$ ls -l sandiway.png
-rw-r--r--@ 1 sandiway  staff  158266 Sep 25 19:45 sandiway.png
$ base64 -i sandiway.png -o encoded.txt
$ ls -l encoded.txt
-rw-r--r--  1 sandiway  staff  211025 Sep 25 19:47 encoded.txt
```

HTML


sandiway@sandiway-XPS-15-9570: ~/Desktop
GNU nano 6.2 test.html *
<html>
<head></head>
<body>
<img src="
File to insert [from ./]: encoded.txt
^G Help M-F New Buffer ^X Execute Command
^C Cancel M-N No Conversion ^T Browse


sandiway@sandiway-XPS-15-9570: ~/Desktop
GNU nano 6.2 test.html
<html>
<head></head>
<body>
<img src="
sWlUWHRYTUw6Y29tLmFkb2JlLnhtcAAAAAAAPD94cGFja2V0IGJlZ2luPSLvu78iIGlkPSJXNU0w
TXBDZWhpSHpyZVN6TlRjemtjOWQiPz4gPHg6eG1wbWV0YSB4bWxuczp4PSJhZG9iZTpucptZXRh
LyIgeDp4bXB0az0iQWRvYmUgWE1QIENvcmUgOS4xLWMwMDEgNzkuYThkNDc1MzQ5LCAyMDIzLzAz
LzIzLTEzOjA1OjQ1ICAgICAgICAiPiA8cmRmOlJERiB4bWxuczpyZGY9Imh0dHA6Ly93d3cudzMu
b3JnLzE5OTkvMDIvMjItcmRmLXN5bnRheC1ucyMiPiA8cmRmOkRlc2NyaXB0aW9uIHJkZjphYm91
dD0iIiB4bWxuczp4bXA9Imh0dHA6Ly9ucy5hZG9iZS5jb20veGFwLzEuMC8iIHhtbG5zOmF1eD0i
aHR0cDovL25zLmFkb2JlLmNvbS9leGlmLzEuMC9hdXgvIiB4bWxuczpleGlmRVg9Imh0dHA6Ly9j
aXBhLmpwL2V4aWYvMS4wLyIgeG1sbnM6cGhvdG9zaG9wPSJodHRwOi8vbnMuYWRvYmUuY29tL3Bo
b3Rvc2hvcC8xLjAvIiB4bWxuczp4bXBNTT0iaHR0cDovL25zLmFkb2JlLmNvbS94YXAvMS4wL21t
LyIgeG1sbnM6c3RFdnQ9Imh0dHA6Ly9ucy5hZG9iZS5jb20veGFwLzEuMC9zVHlwZS9SZXNvdXJj
ZUV2ZW50IyIgeG1sbnM6c3RSZWY9Imh0dHA6Ly9ucy5hZG9iZS5jb20veGFwLzEuMC9zVHlwZS9S
ZXNvdXJjZVJlZiMiIHhtbG5zOmRjPSJodHRwOi8vcHVybC5vcmcvZGMvZWxlbWVudHMvMS4xLyIg
eG1sbnM6Y3JzPSJodHRwOi8vbnMuYWRvYmUuY29tL2NhbWVyYS1yYXctc2V0dGluZ3MvMS4wLyIg
eG1sbnM6dGlmZj0iaHR0cDovL25zLmFkb2JlLmNvbS90aWZmLzEuMC8iIHhtbG5zOmV4aWY9Imh0
dHA6Ly9ucy5hZG9iZS5jb20vZXhpZi8xLjAvIiB4bXA6UmF0aW5nPSIwIiB4bXA6TW9kaWZ5RGF0
ZT0iMjAyMy0wOS0yNVQxOTo0NTo0NC0wNzowMCIgeG1wOkNyZWF0ZURhdGU9IjIwMjMtMDktMDhU
MDc6NDg6NDEiIHhtcDpNZXRhZGF0YURhdGU9IjIwMjMtMDktMjVUMTk6NDU6NDQtMDc6MDAiIHht
cDpDcmVhdG9yVG9vbD0iQWRvYmUgUGhvdG9zaG9wIDI1LjAgKE1hY2ludG9zaCkiIGF1eDpTZXJp
YWxOdW1iZXI9IjA3MjAyMjAwMTM3MyIgYXV4OkxlbnNJbmZvPSI1MC8xIDUwLzEgMC8wIDAvMCIg
YXV4OkxlbnM9IkVGNTBtbSBmLzEuNCBVU00iIGF1eDpMZW5zSUQ9IjE5OCIgYXV4OkxlbnNTZXJp
YWxOdW1iZXI9IjAwMDAwMDAwMDAiIGF1eDpJbWFnZU51bWJlcj0iMCIgYXV4OkZsYXNoQ29tcGVu
c2F0aW9uPSIzLzEiIGF1eDpGaXJtd2FyZT0iMS4xLjAiIGV4aWZFWDpMZW5zTW9kZWw9IkVGNTBt
[Wrote 2781 lines]
^G Help ^O Write Out ^W Where Is ^K Cut ^T Execute ^C Location
^X Exit ^R Read File ^\ Replace ^U Paste ^J Justify ^/ Go To Line

HTML

- Other text element style tags (inline; semantic):
  - `<em> … </em>` — *italics*
  - `<strong> .. </strong>` — **bold**
  - `<tt> … </tt>` — monospaced
  - `<code> … </code>`
- Older style tags (specifies "look" or presentation):
  - `<b> … </b>` — **bold**
  - `<i> … </i>` — *italics*
- block-level:
  - <pre> … </pre> — preformatted

The **HTML `<pre>` element** represents preformatted text which is to be presented exactly as written in the HTML file. The text is typically rendered using a non-proportional ("monospace") font. Whitespace inside this element is displayed as written.

https://developer.mozilla.org/en-US/docs/Web/HTML/Element/pre

HTML

HTML CSS

```
<pre>
  L           TE
    A       A
      C    V
       R A
       DOU
       LOU
      REUSE
      QUE TU
      PORTES
    ET QUI T'
    ORNE O CI
     VILISÉ
    OTE-  TU VEUX
```

Output

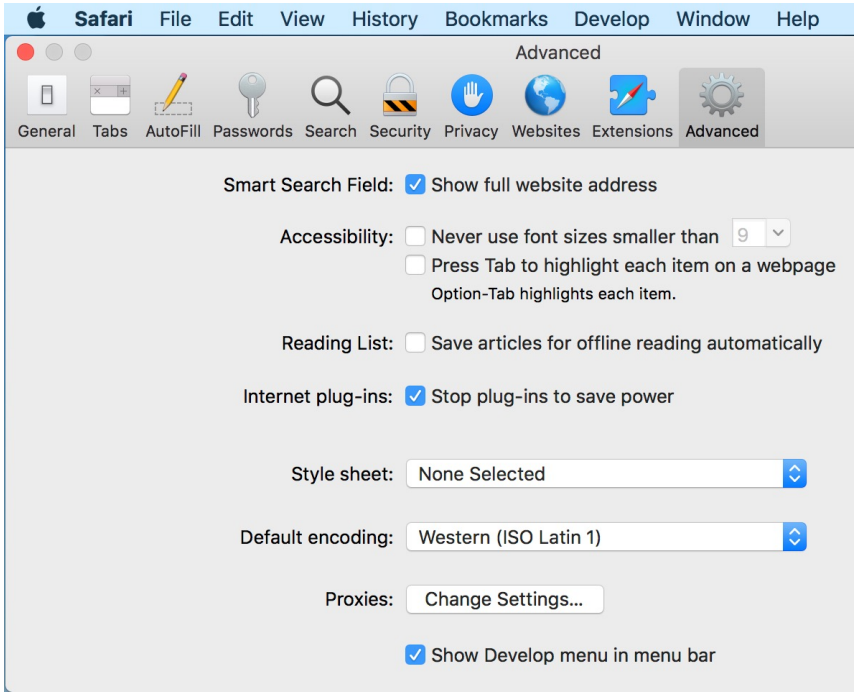
```
L           TE
  A       A
    C    V
     R A
     DOU
     LOU
    REUSE
    QUE TU
    PORTES
  ET QUI T'
  ORNE O CI
   VILISÉ
  OTE-  TU VEUX
   LA    BIEN
  SI      RESPI
          RER       - Apollinaire
```

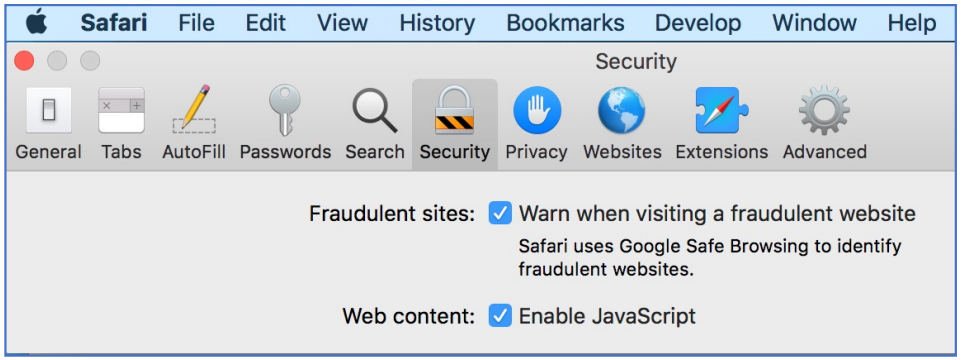
HTML

- Lists:
  - list item: `<li> … </li>`
  - ordered lists: `<ol> … </ol>`
  - unordered lists: `<ul> … </ul>`
  - **note**: can be nested arbitrarily deep

Debugging

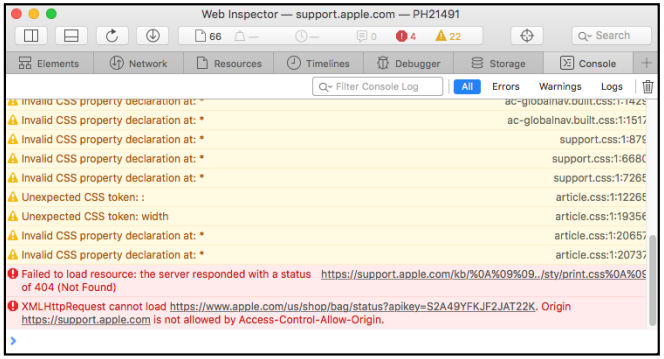
- In Safari:
  - Preferences > Advanced > **Show Develop menu in menu bar**

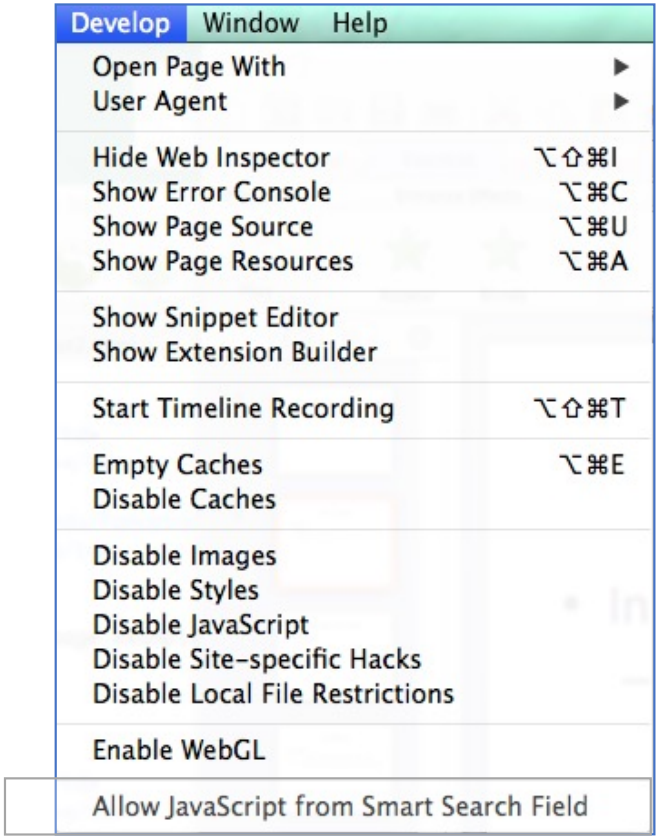




Debugging

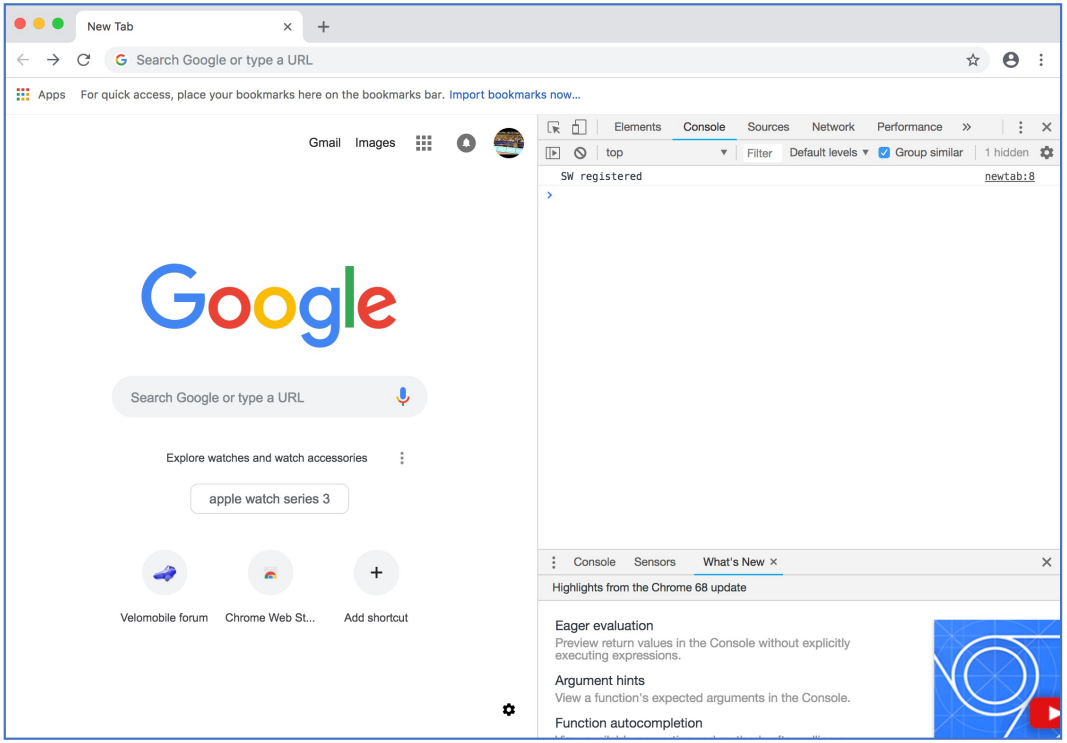
- In the Develop menu:
  - *Allow JavaScript from Smart Search Field*
    - **`javascript:`** ok in address bar
  - Show Page Source
  - Show Web Inspector

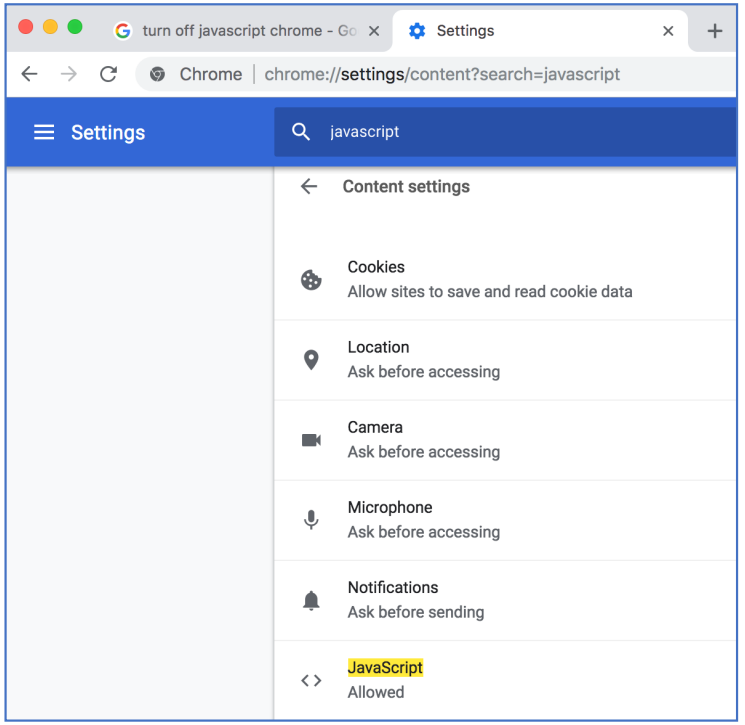




Debugging

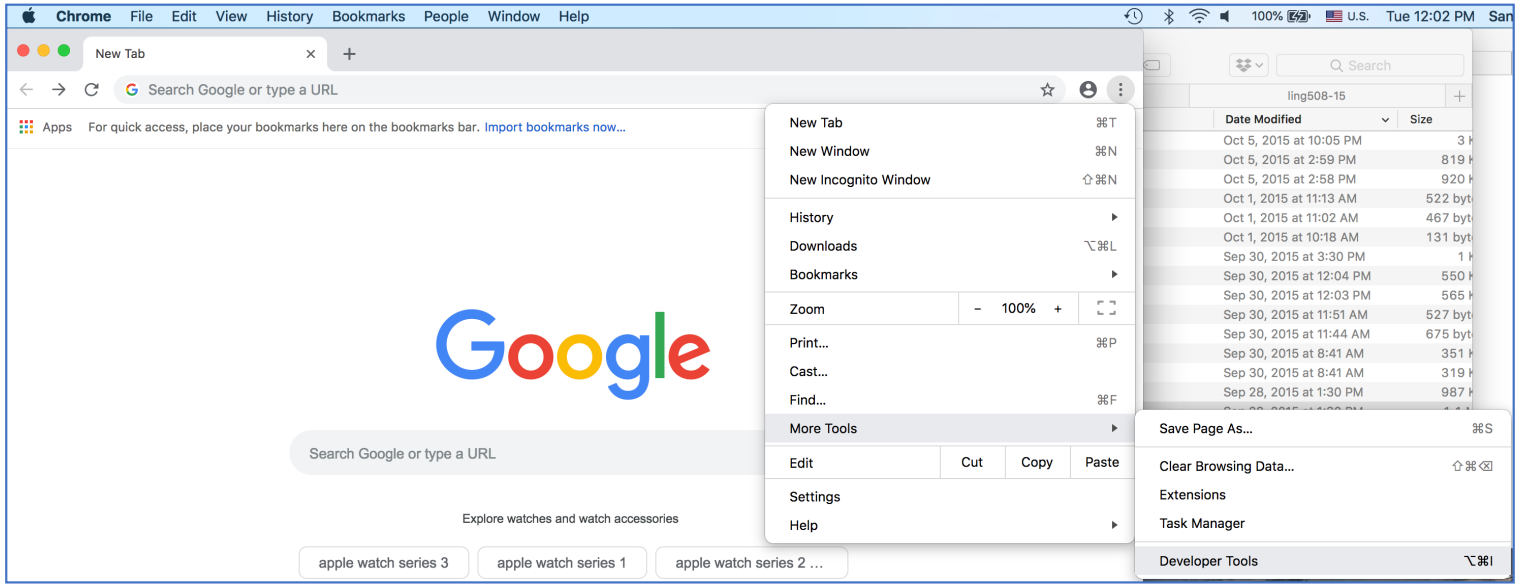
- Developer Tools and Javascript in Google Chrome:





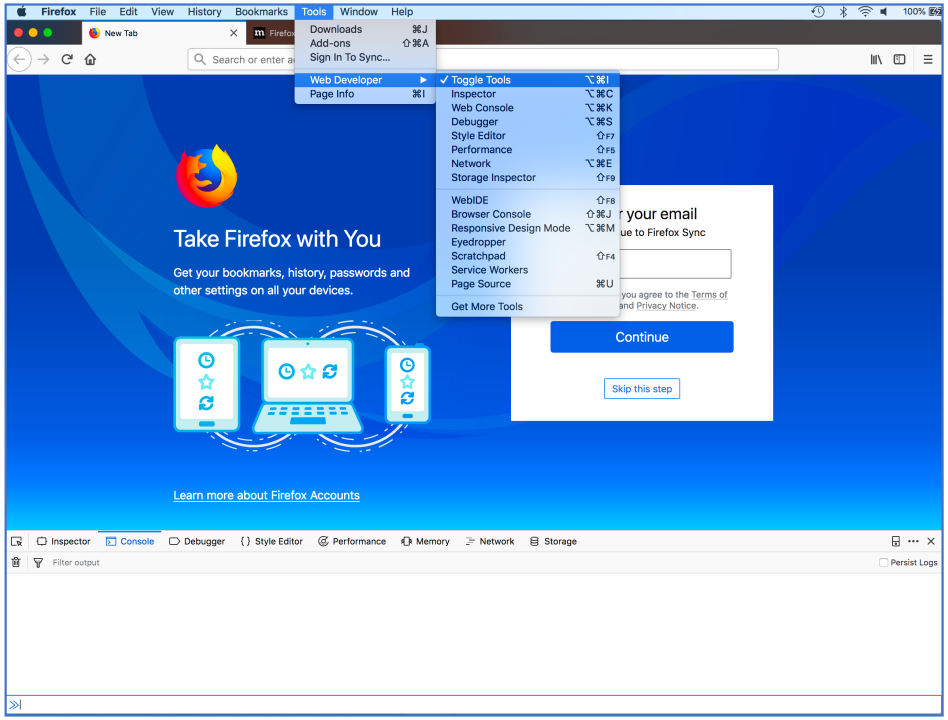
Debugging

- In Google Chrome:



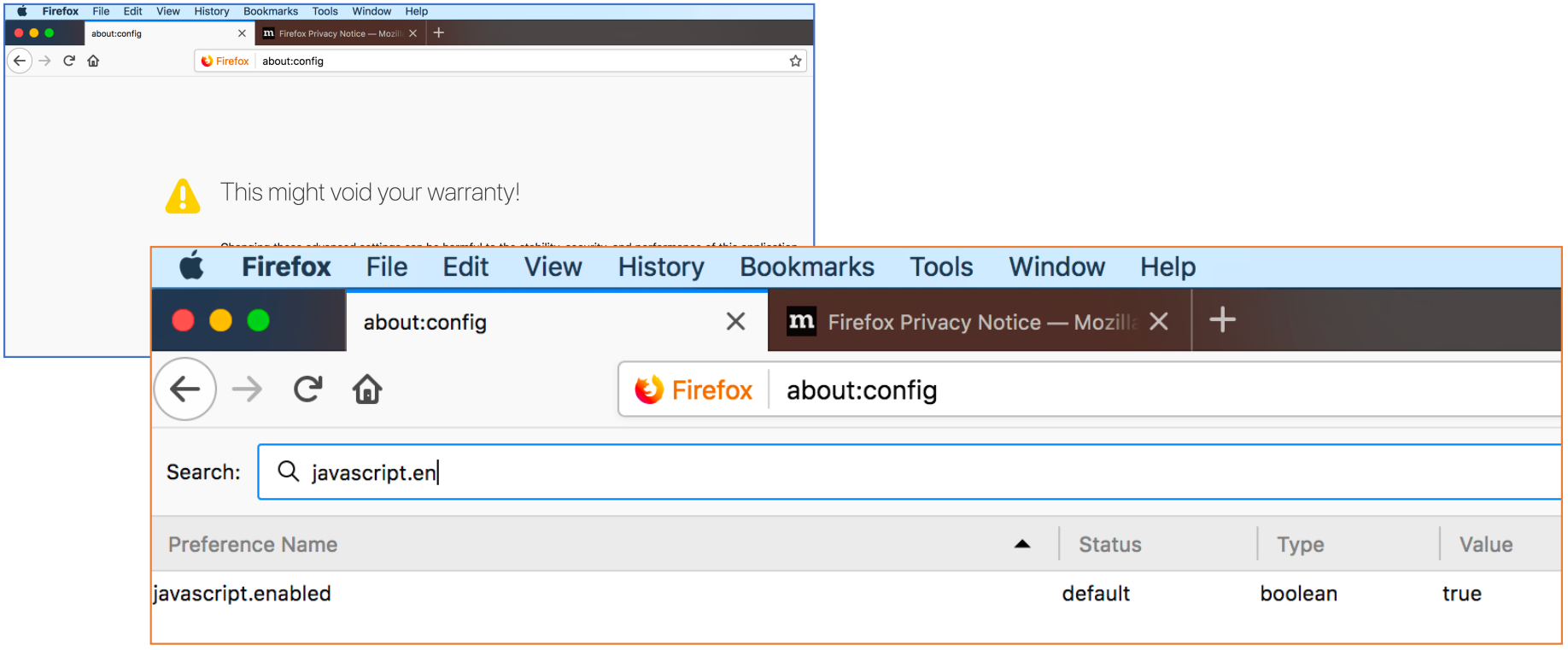
Debugging

- Firefox:



Debugging

- Firefox:



Ungraded Homework Exercise

- Build your own html file.
- Play with the tags
- Embed a picture of yourself encoded in base64