

# Administrative

- Office Hours
  - Shifted to 4-5pm Eastern today due to a conflict
- Homework 1 graded
  - Generally well done
  - *Read the questions carefully*; most people lost points for not fully answering the question, not because they didn't understand how to do it
  - Grades, feedback on Blackboard
- Homework 2
  - Due next Monday, October 3, at 12 noon Eastern
- Project
  - Proposal revisions due 11:59pm Eastern ***on Blackboard*** today for those who were emailed feedback
  - Deliverables I due on October 31

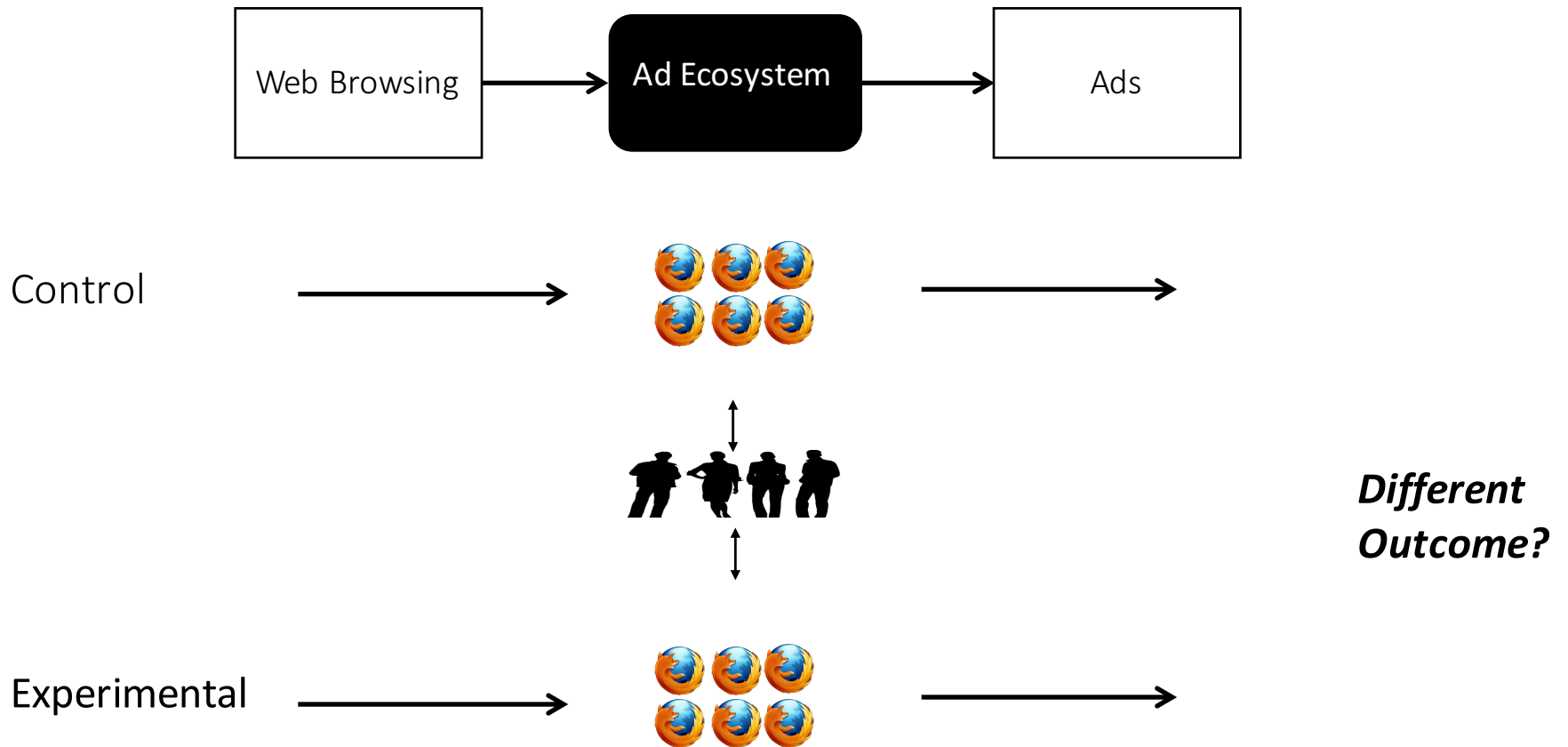
# Homework 1 Statistics

Average	97
Median	100
Standard Deviation	5
Max	100
Min	82

# Recitation on AdFisher

September 30, 2016

# Motivation for building AdFisher



# How AdFisher works

Simulate users having a particular interest by visiting webpages associated with that interest.

```
graph TD; A[Simulate users having a particular interest by visiting webpages associated with that interest.] --> B[Visit a webpage, collect ads, settings provided to the simulated users.]; B --> C[Automatically analyze the data to determine whether statistically significant differences between groups of users with different interests exist.];
```

Visit a webpage, collect ads, settings provided to the simulated users.

Automatically analyze the data to determine whether statistically significant differences between groups of users with different interests exist.

# Running experiments on AdFisher to determine if behavioral targeting occurs on the web

- Input:
  - Behaviors the control / experimental groups are to perform (e.g., visiting websites)
  - Measurements (e.g. ads) to collect afterwards
- AdFisher:
  - Takes input runs experiments collect measurements
  - Determines if there is a statistically significant difference between the data collected from the control and experimental groups

# Running AdFisher

Technical Details

# Installation & Running

- Installation
  - Instructions on Github
  - Tested on MacOS / Linux only
  - Windows users may want to use a VM to run AdFisher
- Running
  - Scripts to run experiments and collect data in `AdFisher/examples`



# Modifying AdFisher

Technical Details

# AdFisher/core

- AdFisher/core/adfisher.py
  - Runs the entire experiment
- AdFisher/core/driver
  - Stages the experiment and sets up the browser units that collect data
  - Stores measurements in a log file
- AdFisher/core/converter
  - Reads data written in the log file
  - Converts data to machine readable format for statistical analysis
- **AdFisher/core/web**
  - Codes that collect data (such as ads, news articles) from various websites
  - *Groups that want to collect data from various websites*
  - Familiarity with Selenium needed
- **AdFisher/core/analysis**
  - Determine if the difference between data collected from the control group and data collected from the experimental group is statistically significant
  - *Groups that want to implement different forms of statistical analysis*

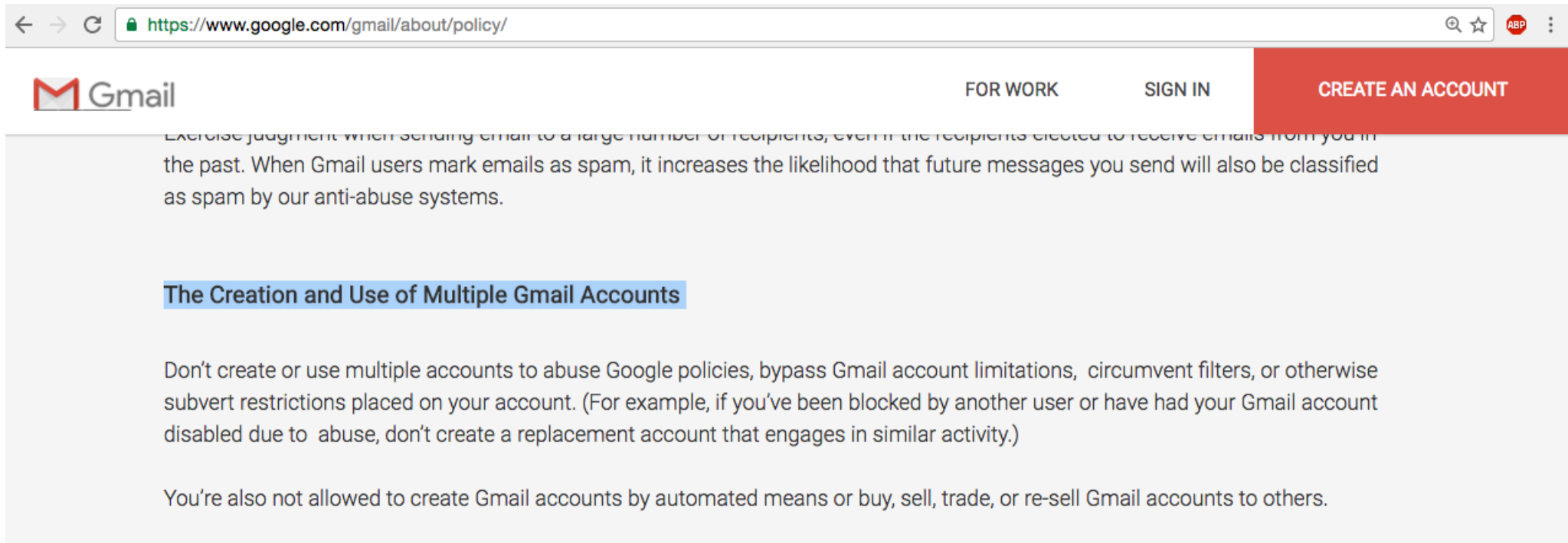
# Selenium Python

- Web Scraping tool that simulates a user surfing the Internet
- Limitations:
  - Works only on HTML elements
  - Does not work on XML, alert boxes, chrome://settings, firefox preferences page, etc

# Automated experiments and Site Policies

Technical Details

# Creating Multiple Accounts



The screenshot shows a web browser window with the address bar displaying <https://www.google.com/gmail/about/policy/>. The Gmail logo is visible on the left, and navigation links for 'FOR WORK', 'SIGN IN', and 'CREATE AN ACCOUNT' are on the right. The 'CREATE AN ACCOUNT' button is highlighted in red. Below the navigation bar, a paragraph of text discusses spam filtering. A section titled 'The Creation and Use of Multiple Gmail Accounts' is highlighted in blue. This section contains two paragraphs of text regarding the creation and use of multiple accounts.

Exercise judgment when sending email to a large number of recipients, even if the recipients elected to receive emails from you in the past. When Gmail users mark emails as spam, it increases the likelihood that future messages you send will also be classified as spam by our anti-abuse systems.

## The Creation and Use of Multiple Gmail Accounts

Don't create or use multiple accounts to abuse Google policies, bypass Gmail account limitations, circumvent filters, or otherwise subvert restrictions placed on your account. (For example, if you've been blocked by another user or have had your Gmail account disabled due to abuse, don't create a replacement account that engages in similar activity.)

You're also not allowed to create Gmail accounts by automated means or buy, sell, trade, or re-sell Gmail accounts to others.

# Creating Multiple Accounts

← → ↻ <https://www.linkedin.com/legal/user-agreement>

LinkedIn

[What is LinkedIn?](#) [Join Today](#) [Sign In](#)

## 2. Obligations

Here are some promises you make to us in this Agreement:

You're eligible to enter into this Agreement and you are at least our "Minimum Age."

### 2.1. Service Eligibility

To use the Services, you agree that: (1) you must be the "Minimum Age" (defined below) or older; (2) you will only have one LinkedIn account (and/or one SlideShare or Pulse account, if applicable), which must be in your real name; and (3) you are not already restricted by LinkedIn from using the Services.

"Minimum Age" means (a) 18 years old for the People's Republic of China, (b) 16 years old for the Netherlands, (c) 14 years old for the United States, Canada, Germany, Spain, Australia and South Korea, and (d) 13 years old for all other countries. However, if law requires that you must be older in order for LinkedIn to lawfully provide the Services to you (including the collection, storage and use of your information) then the Minimum Age is such older age. The Services are not for use by anyone under the age of 13.

Number of requests  
a website expects



# #Requests / unit time

The screenshot shows the AutoMall.com website. At the top, there is a navigation bar with links for USED CARS, NEW CARS, CAR LOANS, and RESEARCH. A search bar is located on the right side of the navigation bar. Below the navigation bar, there is a large banner image of a white car. Overlaid on this banner is a search form with the following elements:

- Text: "Your Online Marketplace for New & Used Cars. Find a great car for sale near you with 2,515,902 cars to choose from!"
- Radio buttons for "Used" (selected) and "New".
- Dropdown menus for "Any Make" and "Any Model".
- Input fields for "Maximum Price" and "15217".
- A dropdown menu for "Radius".
- A link for "More Search Options" and a green "SEARCH LISTINGS" button.

Below the search form, there is a section titled "Featured Cars for Sale (selected zip: 15217)". This section displays a grid of seven car listings, each with a small image, the car's make and model, its price, and its location (Pittsburgh, PA):

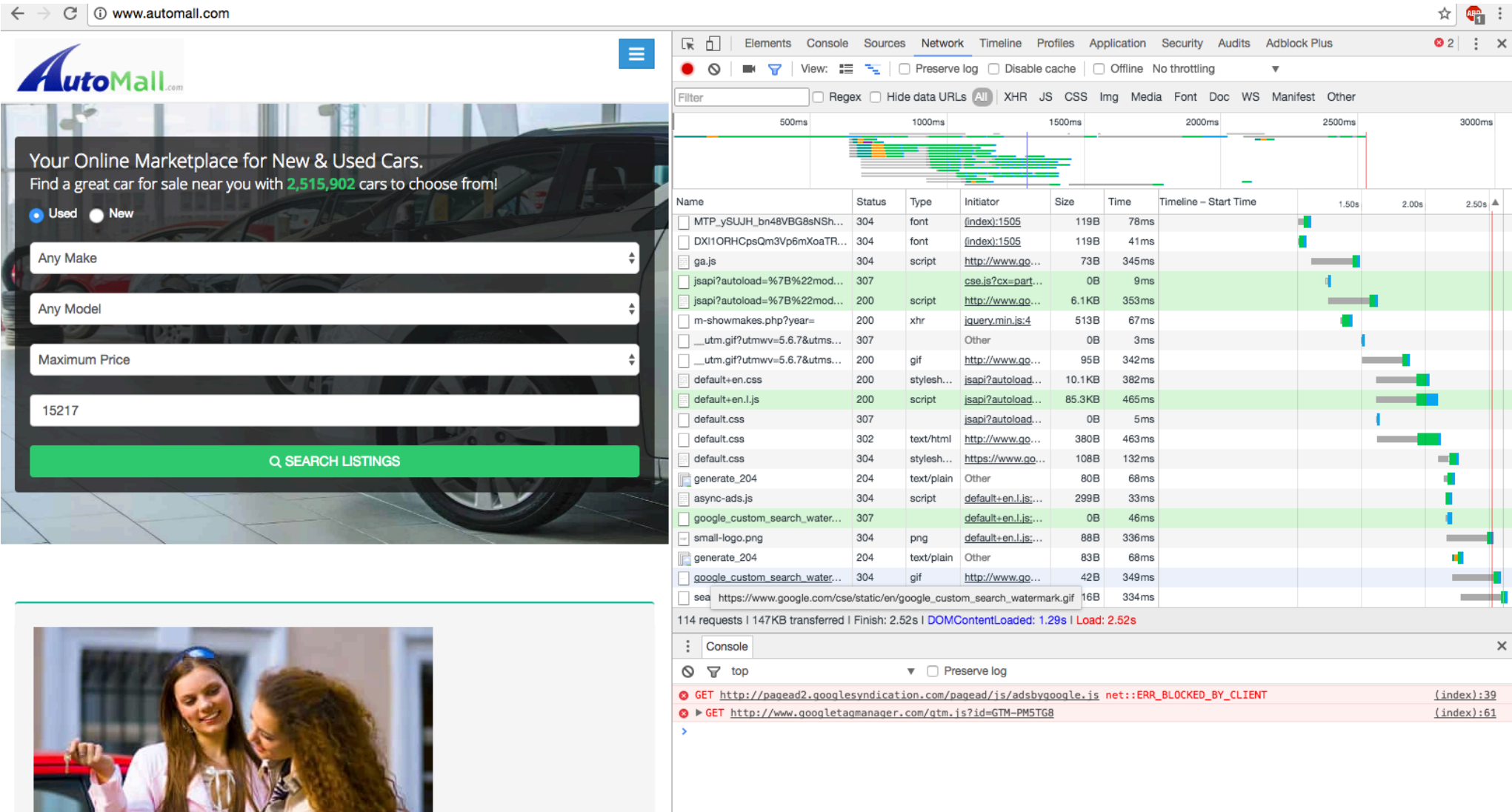
Year	Make	Model	Price	Location
2013	Toyota	RAV4	\$20,700	Pittsburgh, PA
2014	Infiniti	QX50	\$28,600	Pittsburgh, PA
2011	Kia	Rio	\$7,700	Pittsburgh, PA
2014	BMW	X1	\$24,800	Pittsburgh, PA
2014	Nissan	Maxima	\$19,900	Pittsburgh, PA
2010	Mercedes	E	\$23,900	Pittsburgh, PA
2013	Ram	2500	\$33,800	Pittsburgh, PA

At the bottom of the page, there are two promotional banners:

- Car Loans:** "Leverage The Internet's Original AutoMall™ and our nationwide network of car..."
- Finance Tools:** "AutoMall.com provides Car Loan Calculators to help you find the right car, truck or SUV that fits..."



# #requests / unit time



The image displays a screenshot of the AutoMall.com website on the left and the Chrome DevTools Network tab on the right. The website shows a search interface for cars with filters for 'Used' or 'New', 'Any Make', 'Any Model', and 'Maximum Price' (set to 15217). A green button labeled 'SEARCH LISTINGS' is visible. The DevTools Network tab shows a timeline of requests, with a table listing various resources and their performance metrics.

Name	Status	Type	Initiator	Size	Time	Timeline - Start Time	1.50s	2.00s	2.50s
MTP_ySUJH_bn48VBG8sNSh...	304	font	(index):1505	119B	78ms				
DXI1ORHCpsQm3Vp6mXoaTR...	304	font	(index):1505	119B	41ms				
ga.js	304	script	http://www.go...	73B	345ms				
jsapi?autoload=%7B%22mod...	307	script	cse.js?cx=part..	0B	9ms				
jsapi?autoload=%7B%22mod...	200	script	http://www.go...	6.1KB	353ms				
m-showmakes.php?year=	200	xhr	jquery.min.js:4	513B	67ms				
__utm.gif?utmwv=5.6.7&utms...	307	Other		0B	3ms				
__utm.gif?utmwv=5.6.7&utms...	200	gif	http://www.go...	95B	342ms				
default+en.css	200	stylesh...	jsapi?autoload...	10.1KB	382ms				
default+en.l.js	200	script	jsapi?autoload...	85.3KB	465ms				
default.css	307	script	jsapi?autoload...	0B	5ms				
default.css	302	text/html	http://www.go...	380B	463ms				
default.css	304	stylesh...	https://www.go...	108B	132ms				
generate_204	204	text/plain	Other	80B	68ms				
async-ads.js	304	script	default+en.l.js:...	299B	33ms				
google_custom_search_water...	307	script	default+en.l.js:...	0B	46ms				
small-logo.png	304	png	default+en.l.js:...	88B	336ms				
generate_204	204	text/plain	Other	83B	68ms				
google_custom_search_water...	304	gif	http://www.go...	42B	349ms				
sea	https://www.google.com/cse/static/en/google_custom_search_watermark.gif	16B			334ms				

114 requests | 147KB transferred | Finish: 2.52s | DOMContentLoaded: 1.29s | Load: 2.52s

Console

```
top
```

GET http://pagead2.googlesyndication.com/pagead/js/adsbygoogle.js net::ERR\_BLOCKED\_BY\_CLIENT (index):39

GET http://www.googletagmanager.com/gtm.js?id=GTM-PM5TG8 (index):61

<http://blog.hubspot.com/marketing/reduce-http-requests#sm.0001t43gs9739cxeq3916bv5w0cm>