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

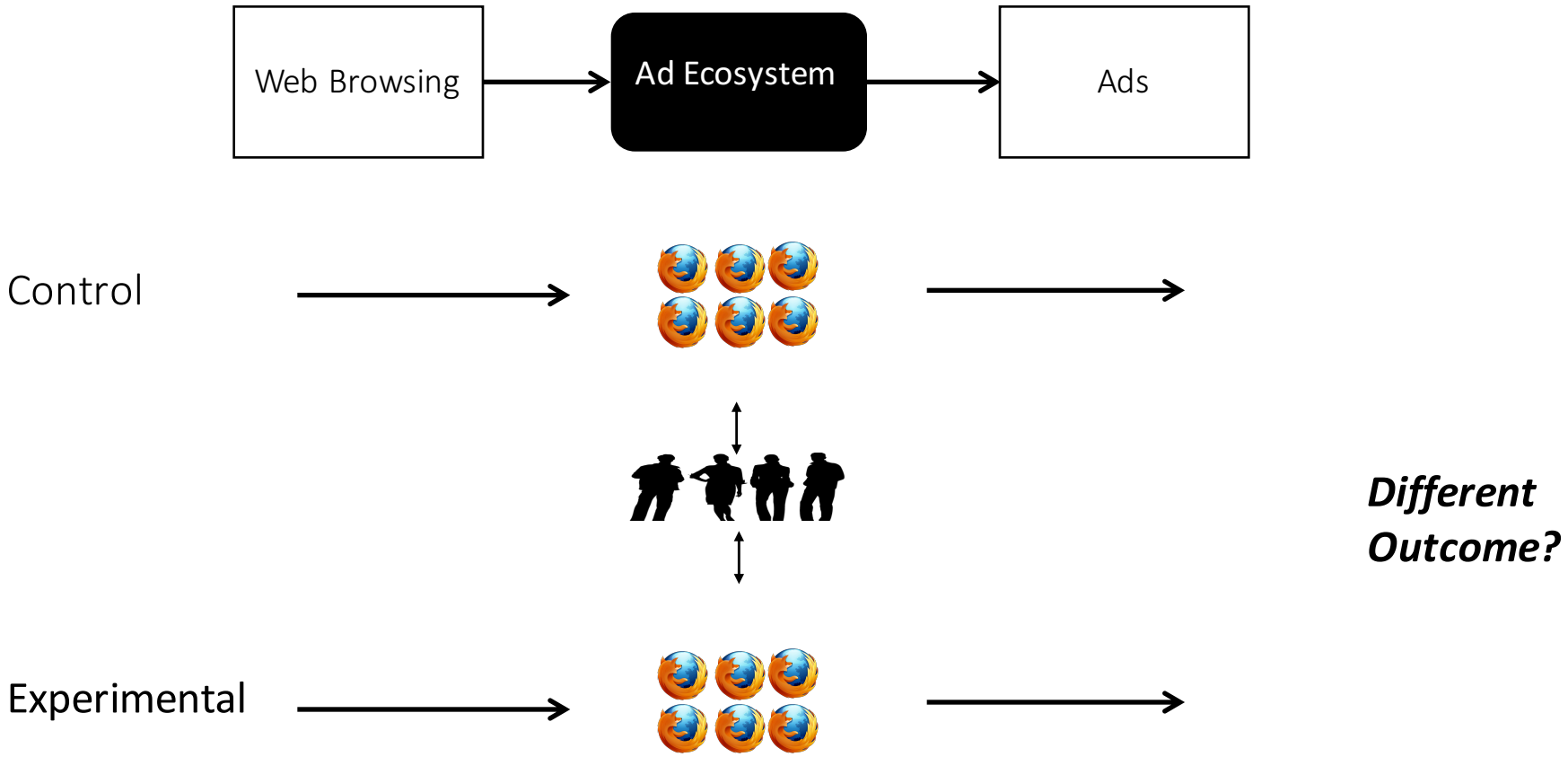
Homework 1 Statistics

Average	97
Median	100
Standard Deviation	5
Max	100
Min	82

Recitation on AdFisher

September 30, 2016

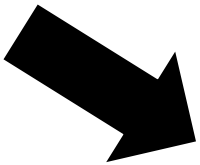
Motivation for building AdFisher



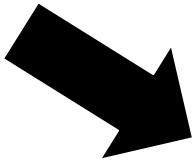
Slide Credit: Amit Datta

How AdFisher works

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



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



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

Github: https://github.com/tadatitam/info-flow-experiments

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 collects 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

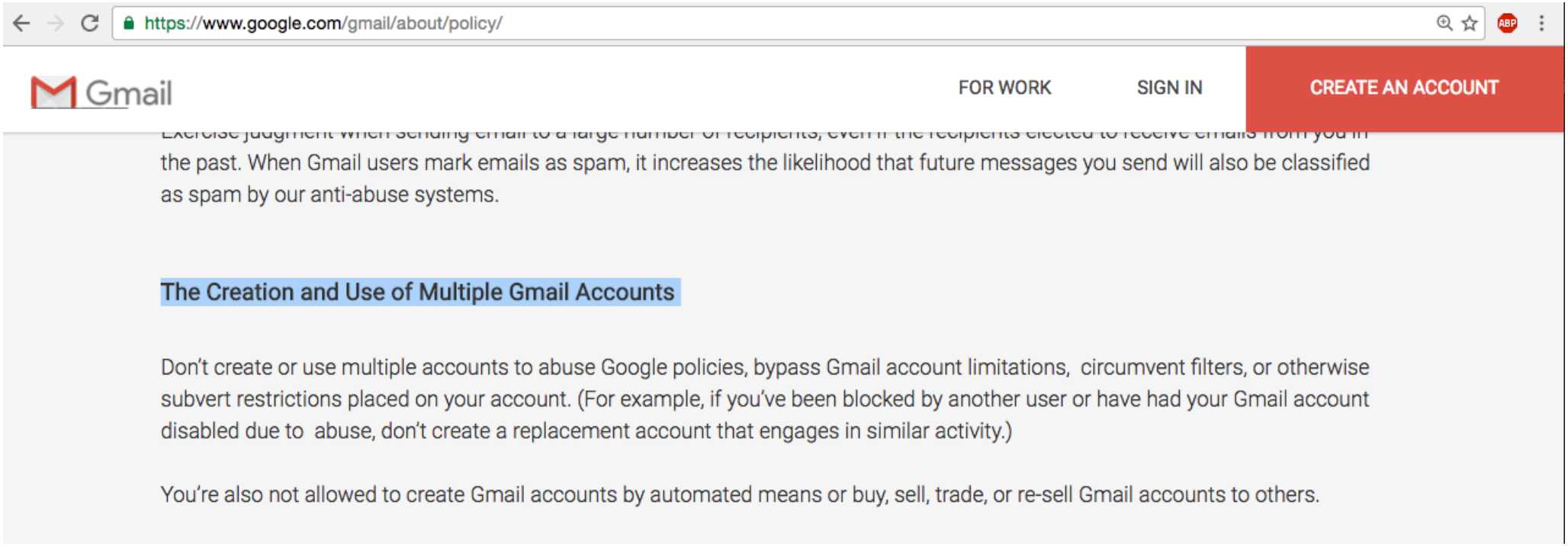
- Limitations:
  - Works only on HTML elements
  - Does not work on XML, alert boxes, chrome://settings, firefox preferences page, etc

http://selenium-python.readthedocs.io/

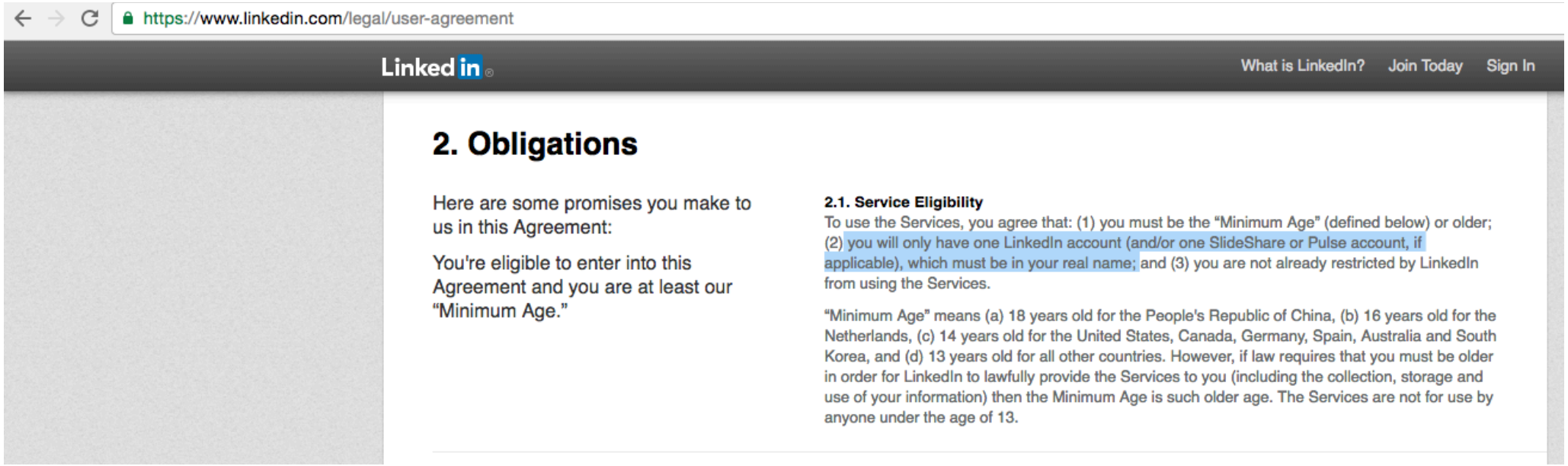
Automated experiments and Site Policies

Technical Details

Creating Multiple Accounts

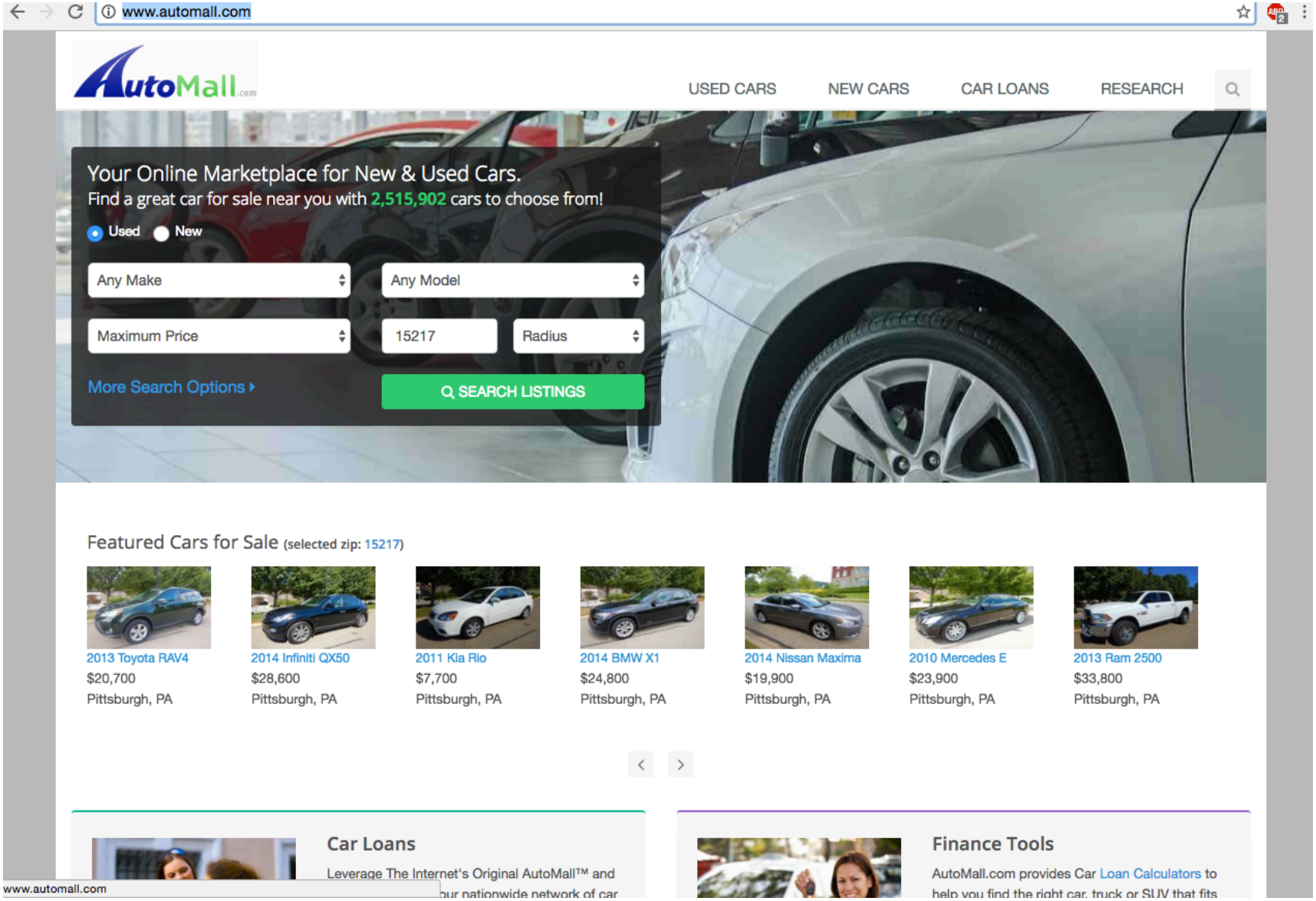


Creating Multiple Accounts

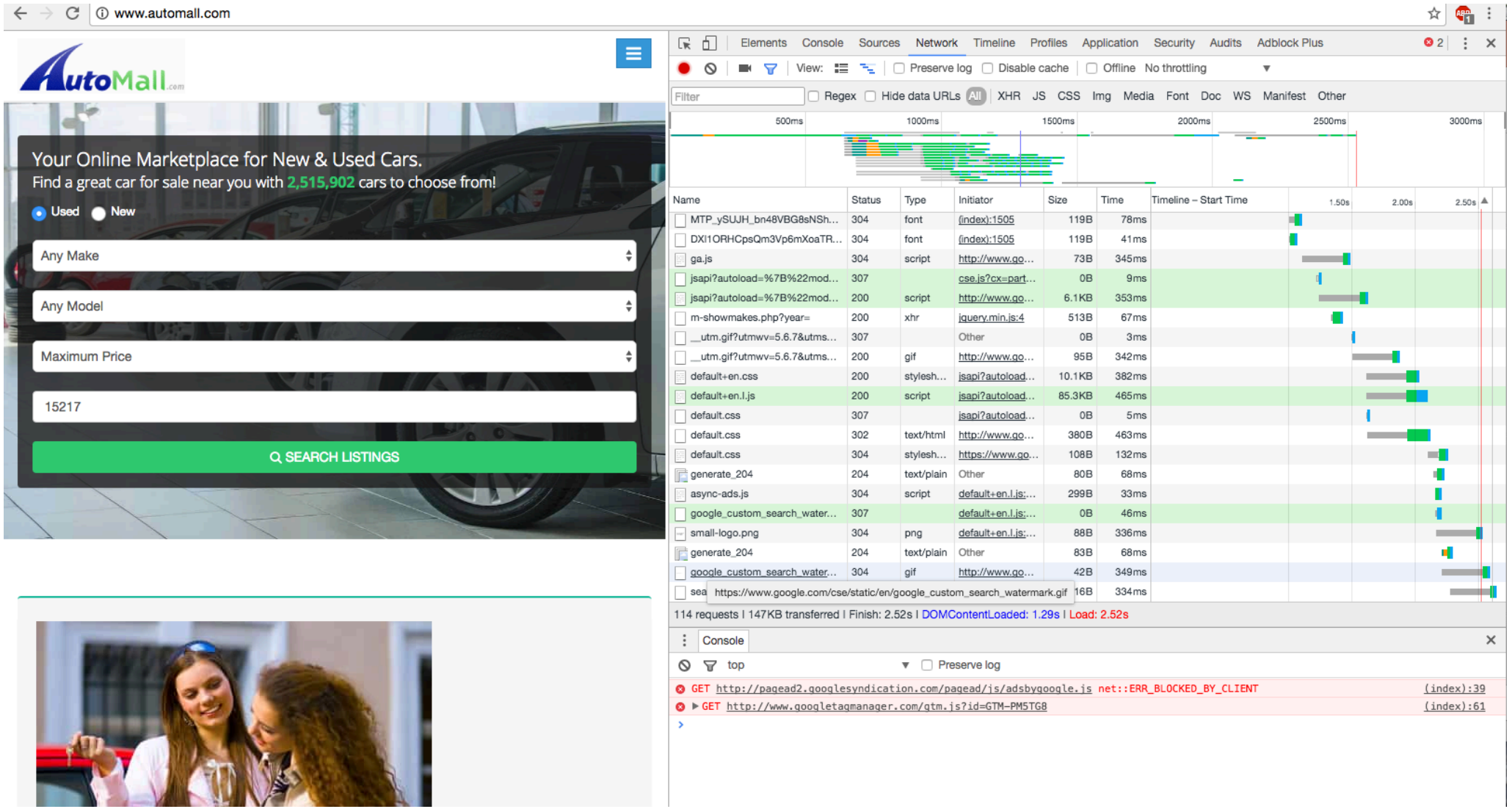


Number of requests a website expects

#Requests / unit time



#requests / unit time



http://blog.hubspot.com/marketing/reduce-http-requests#sm.0001t43gs9739cxeq3916bv5w0cmy