



Lesson 3: Acquiring Data: Sources and Methods

The Promise of Open Data

"The data we generate, based on determinist assumptions and imperfect methodologies, could end up shaping populations and building worlds in their own image."

- Places Journal (on Hudson Yards)



Josh And Nicole

Sweet and Cozy Madrona Apartment

Seattle, WA, United States  (73)



Entire home/apt



2 Guests



1 Bedroom



1 Bed

\$95

Per Night

Check in

mm/dd/yyyy

Check out

mm/dd/yyyy

Guests

1 

Request to Book

What guests liked about this place

Private, Clean, Neat, Great location, Walkable, Hitch friendly, Pictures, Quiet

 Save to Wish List



Josh And Nicole

Sweet and Cozy Madrona Apartment

Seattle, WA, United States (73)



Entire home/apt



2 Guests



1 Bedroom



1 Bed

\$95

Per Night

Check in

Check out

Guests

mm/dd/yyyy

mm/dd/yyyy

1

Request to Book

What guests liked about this place

Residential

Cozy

Near restaurants and shops

Well-equipped kitchen

Private

Quiet

Save to Wish List

256 travelers saved this place

About this listing

We have a recently remodeled sweet, clean, cheery, MIL apartment on the first floor of our 1901 Craftsman home in the historic Madrona neighborhood. We are on the bus line, blocks from cafes, restaurants and lovely Lake Washington.

[Contact Host](#)

Business Travel

 This listing has essential business travel amenities. [Learn More](#)

The Space

Accommodates: 2

Check In: Anytime after 4PM

Bathrooms: 1

Check Out: 11AM

Bed type: **Real Bed**

Property type: Apartment

Bedrooms: 1

Room type: Entire home/apt

Beds: 1

House Rules

Amenities

[+ More](#)

Prices

Extra people: **No Charge**

Weekly discount: 17%

Cleaning Fee: \$40

Monthly discount: 0%

Security Deposit: \$150

Cancellation: **Moderate**

Description

The Space

The is a MIL apartment on the first floor of our 1901 Craftsman home. It has been recently remodeled with a full kitchen, heated floors in the bathroom and kitchen, a cozy gas stove, a queen sleep number bed, a propane grill, patio, TV, cable and wifi. The apartment is ideally suited to fit one or two people but a small family may be comfortable if the children sleep on the small sofas. Please let us know if you need extra bedding for a child.

Check in

Check out

Guests

mm/dd/yyyy

mm/dd/yyyy

1

[Request to Book](#) [Save to Wish List](#)

256 travelers saved this place

Email

Messenger

[... More](#)

Report this listing

73 Reviews ★★★★★

Search reviews

Summary

Accuracy

★★★★★

Location

★★★★★

Communication

★★★★★

Check In

★★★★★

Cleanliness

★★★★★

Value

★★★★★



Lucas

My wife, my little pug and I felt like home. The location is beautiful and really close to downtown. The house was perfect to spend a couple of days. I really recommend it ;-) Lucas

May 2016

Helpful



Tracy

Josh met us when we arrived. The apartment is in a lovely residential neighbourhood with great views and interesting architecture, both old and new. The place was comfy and private. We were very comfortable. There is a cute outdoor area. We didn't get a chance to use the BBQ!

April 2016

Helpful



Stephanie

We had a wonderful stay at this place! Clean, cozy, and in a great location. Would stay again!

March 2016

Helpful



Lorenz & Barbara

Josh and Nicole: We loved your place. Wonderful neighborhood and very comfy little flat- perfect base for our Seattle Symphony "side-by-side". Brigid's cello teacher, Olivia, lives near you apparently. We very much hope to return- Emily has a piano recital in June. We will be in touch! You were very gracious and helpful hosts! Thanks so much! (URL HIDDEN)Lefty & Barbara Schultz

Check in

Check out

Guests

mm/dd/yyyy

mm/dd/yyyy

1

Request to Book

Save to Wish List

256 travelers saved this place

Email

Messenger

... More

Report this listing

Your Host



Josh And Nicole

Seattle, Washington, United States · Member since May 2012

We are an adventurous couple with a love of the outdoors living in Seattle. We came here almost 20 years ago via Alaska and Ohio and are still discovering this awesome city. When not traveling or in Seattle we are in the Methow Valley at our cabin skiing, biking, hiking, fishing and reading with our toes in the river. We have our cabin to make special memories with our family and friends and love to share it with others who want to do the same.

Response rate: 100%

Response time: within a day

141

Reviews

3

References



Verified

[Contact Host](#)[Request to Book](#)[Save to Wish List](#)

256 travelers saved this place

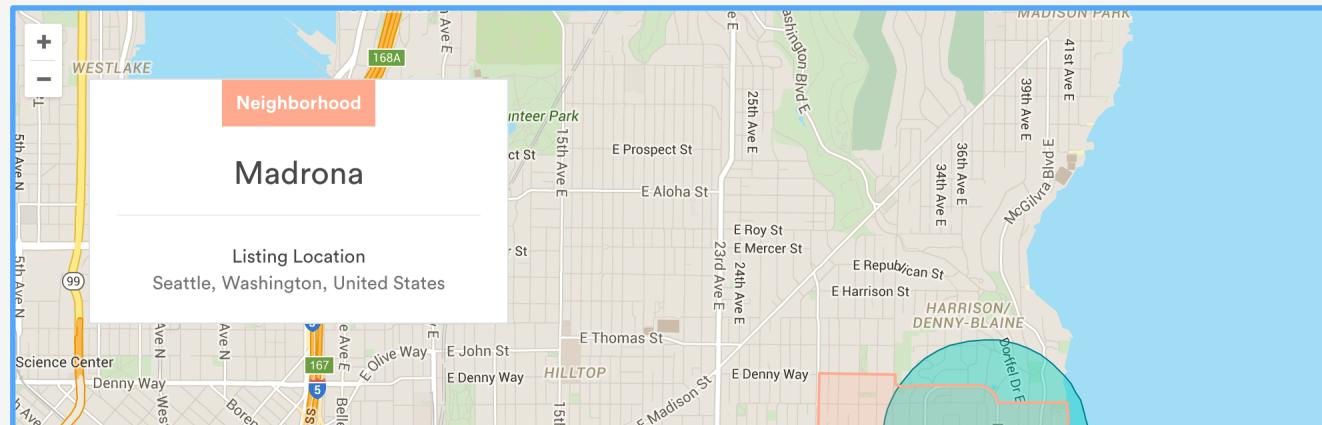
[Email](#)[Messenger](#)[... More](#)[Report this listing](#)

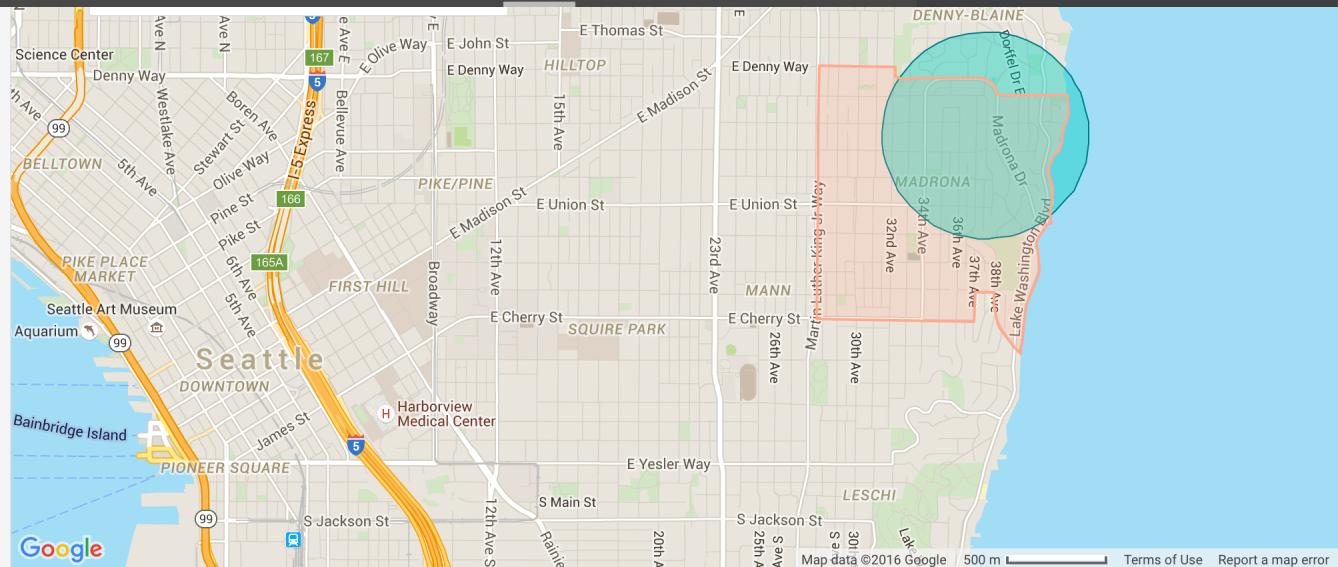
Neighborhood

Madrona

Listing Location

Seattle, Washington, United States





Recommendations!



Similar Listings



Guest House in Madison Vall...
Entire home/apt — 0.7 miles away

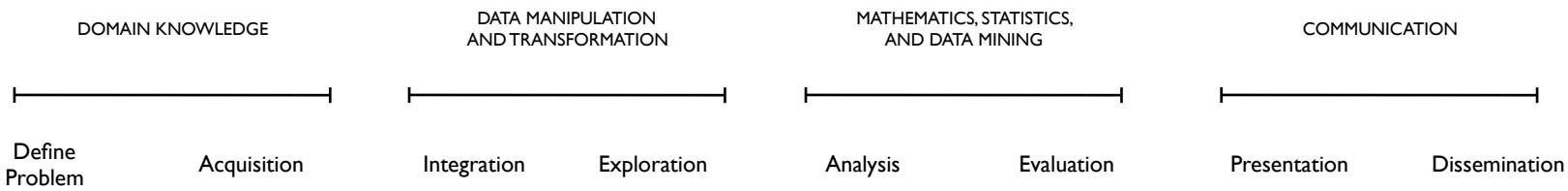


Private Apartment - Queen A...
Entire home/apt — 3.7 miles away

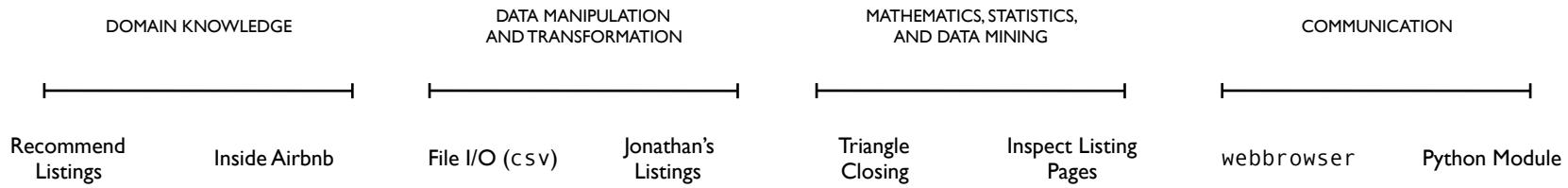


Perfect Location by Lake Wa...
Entire home/apt — 1.3 miles away

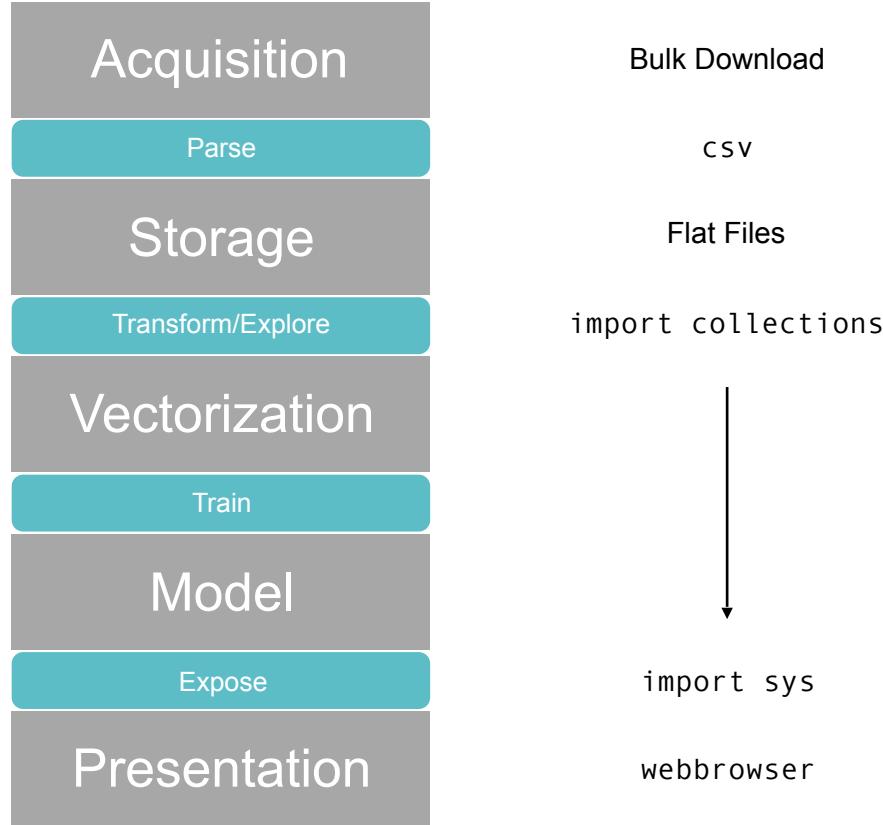
Process



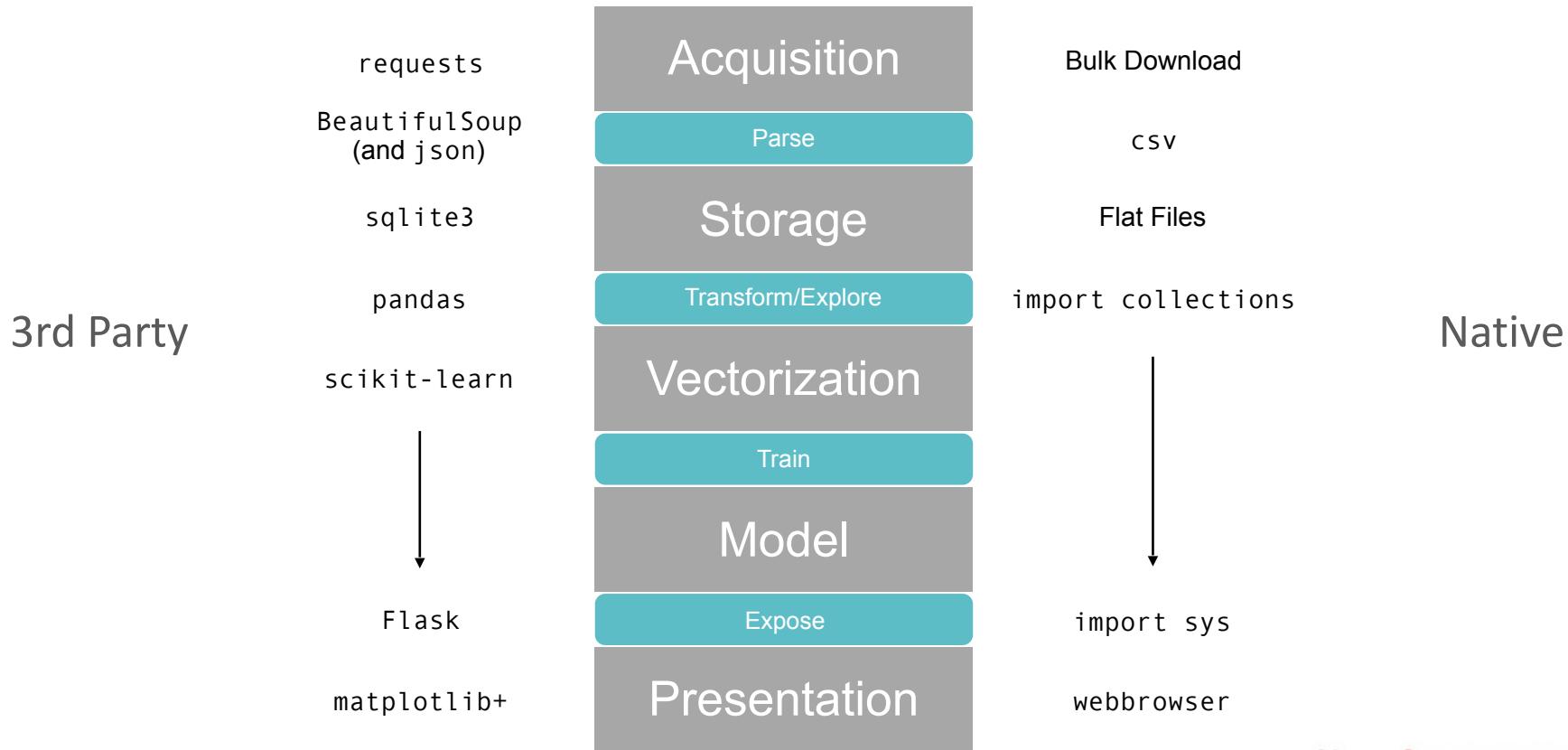
Process



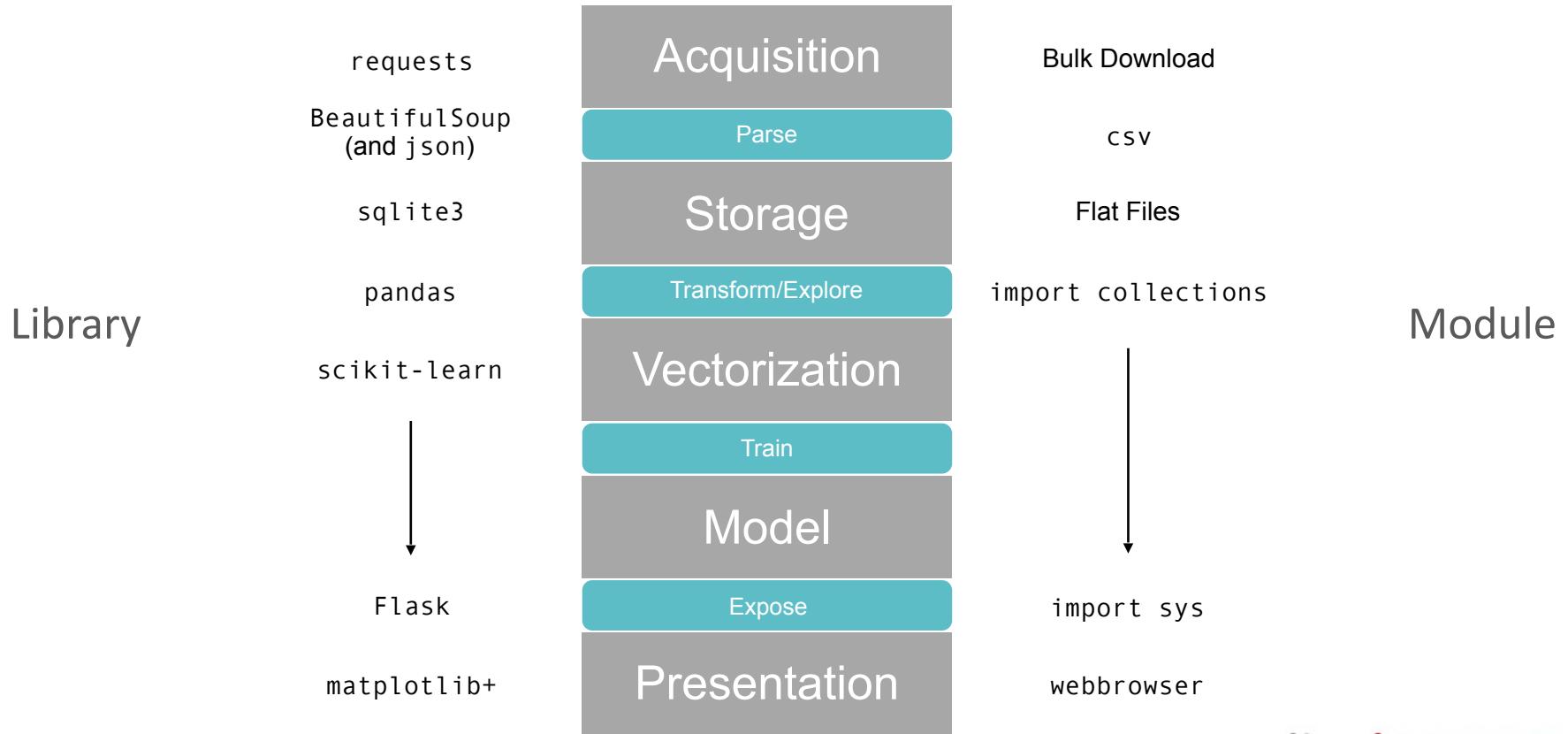
Process + Tools



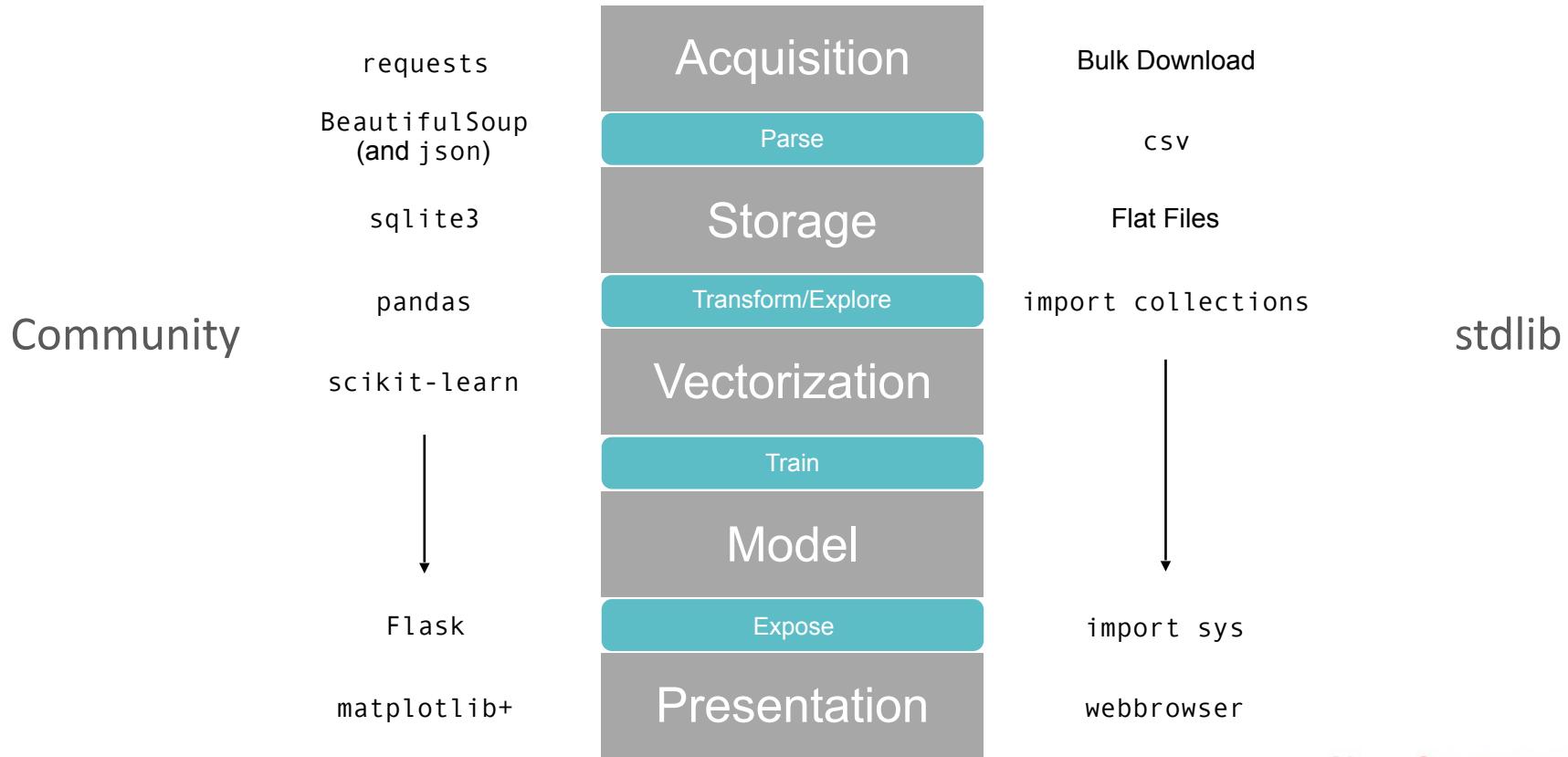
Process + Tools



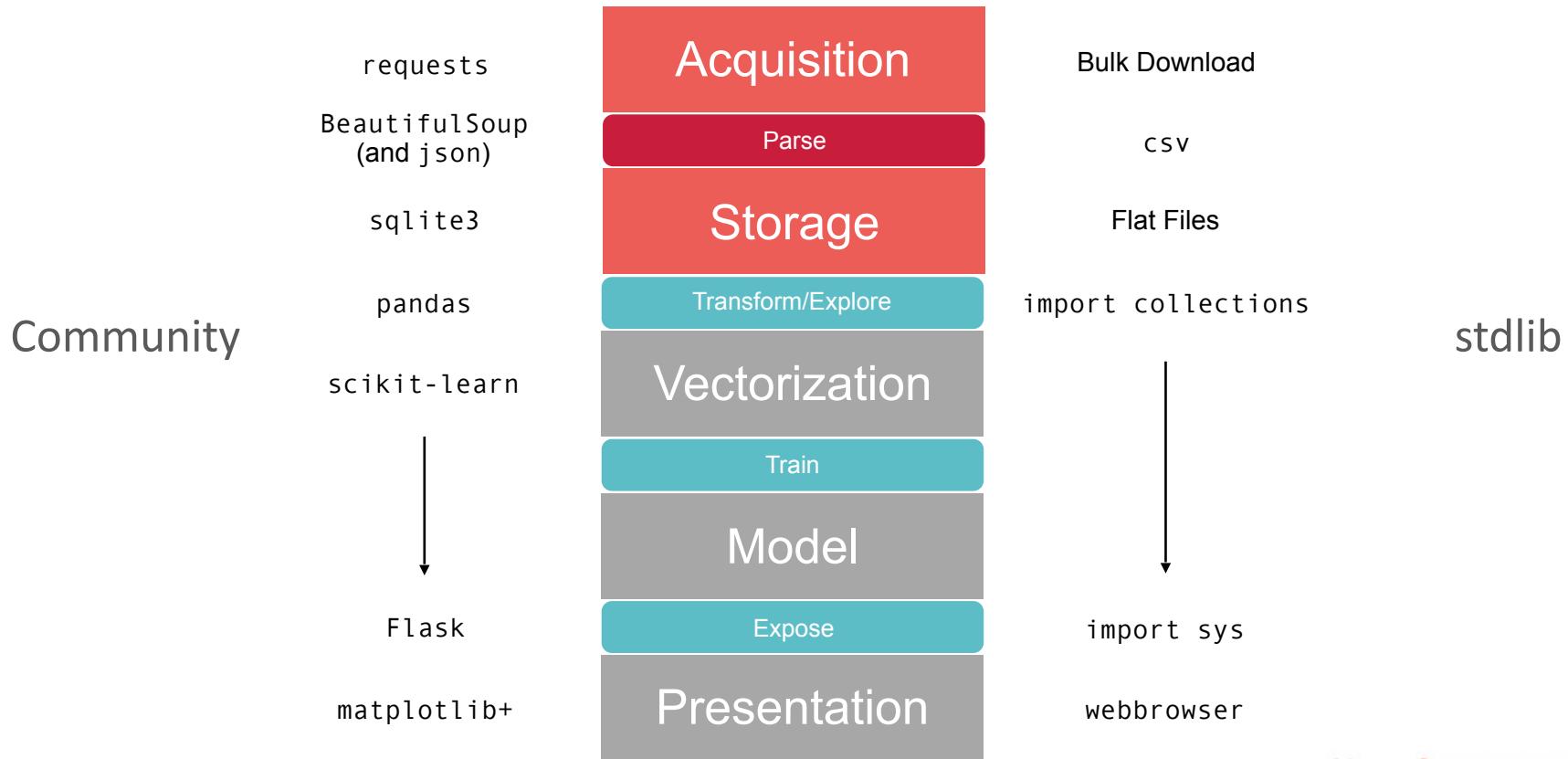
Process + Tools



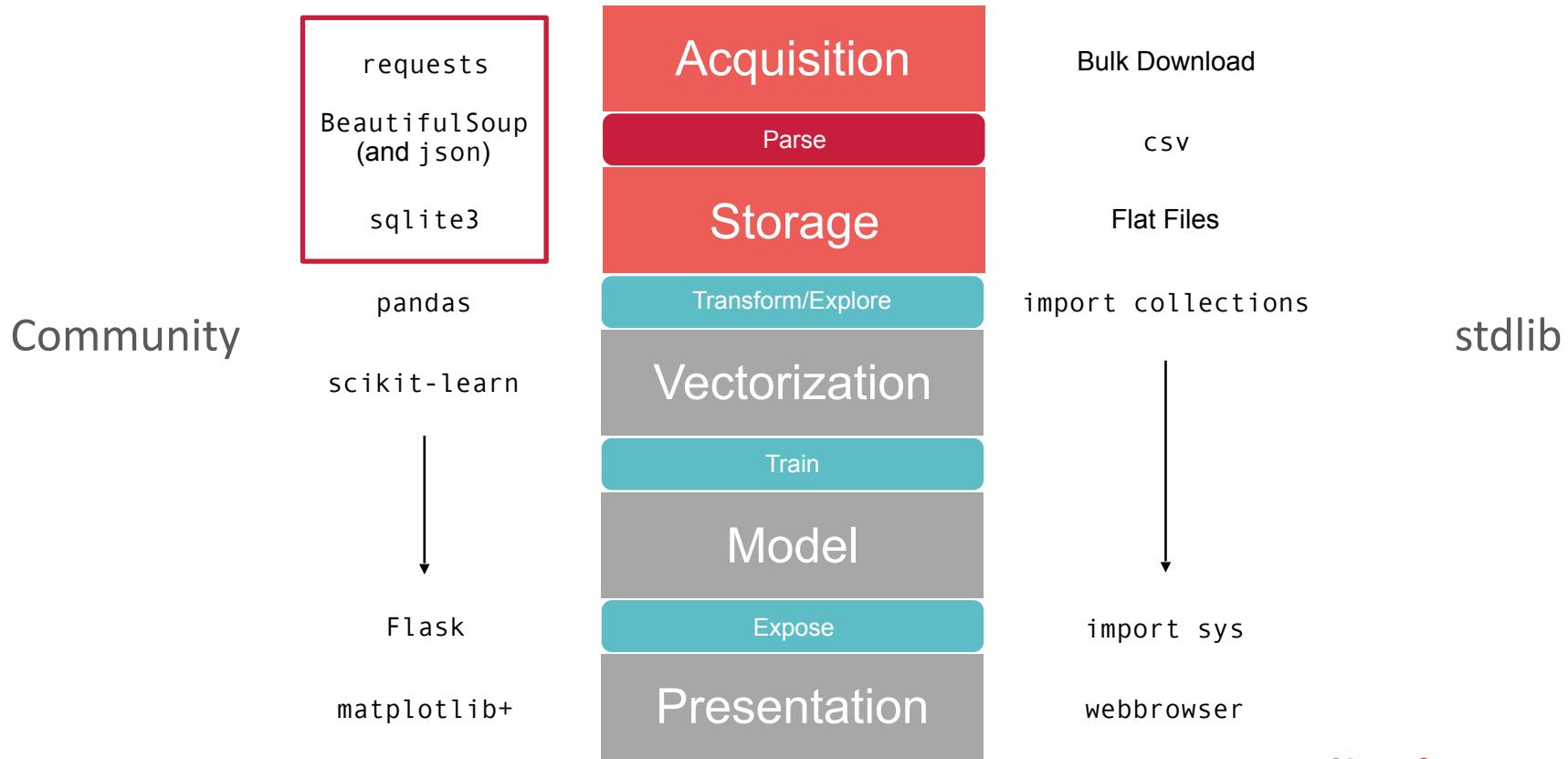
Process + Tools



Process + Tools



Process + Tools



Cheaper Beds, Better Breakfasts

Cheaper Beds, **Better** Breakfasts

**We will come back to the Cheaper beds
part...**

Aid users in their discovery of relevant listings

**For now, we are interested not in listings
themselves but the context around the
listing**

**Which neighborhoods are the most
explorable?**

**Which listings have the most lively and
active POI around them?**

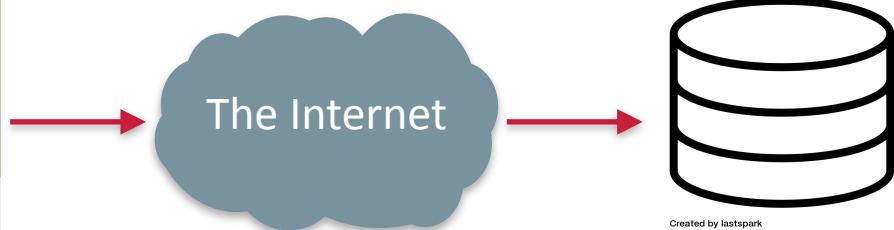
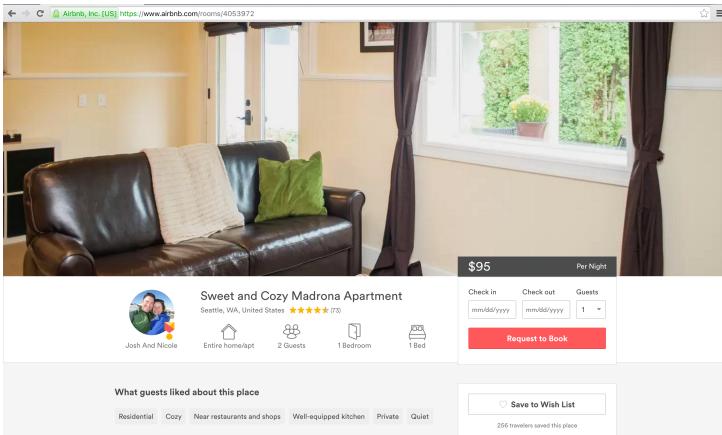
Common Data Sources

Type	Advantage	Disadvantage	Example
Data as a Service (DaaS)	Well structured, cleaned, and enriched	Often cost money and limited to specific services	Factual, ProPublica
Bulk Download	Easy to acquire	Can be hard to incorporate new data	Inside Airbnb
API	Parameterized queries	Rate limits and single data point access	Twitter, Foursquare
Web Scraping	Anything can be scraped from the internet	Hard to parse, unstructured, and unpredictable	Airbnb

<https://github.com/hopelessoptimism/awesome-resources/blob/master/datasets.md>

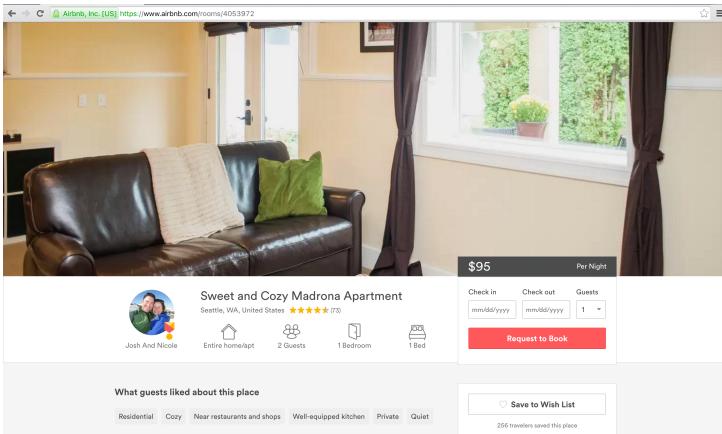
<https://pinboard.in/u:hopelessoptimism/t:dataset/>

The Web

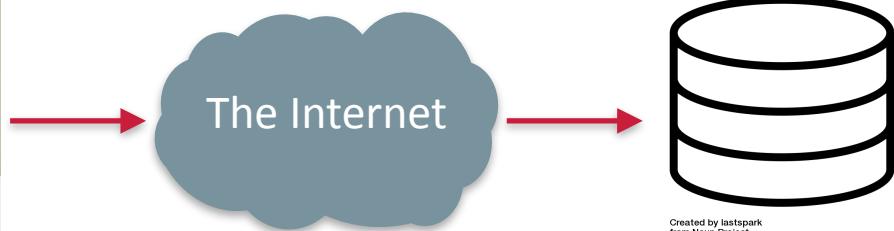


Created by lastspark
from Noun Project

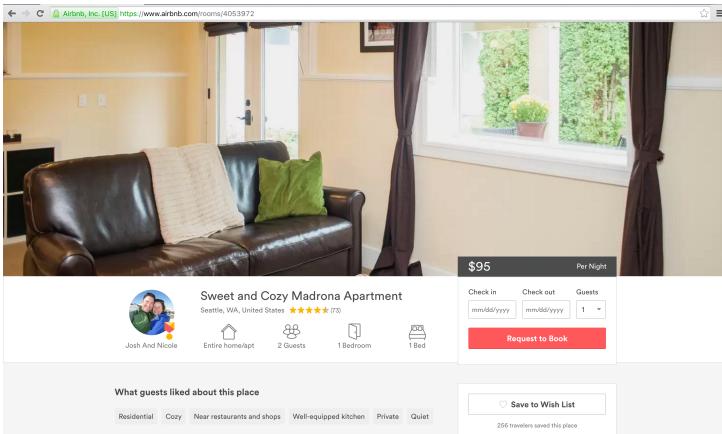
The Web



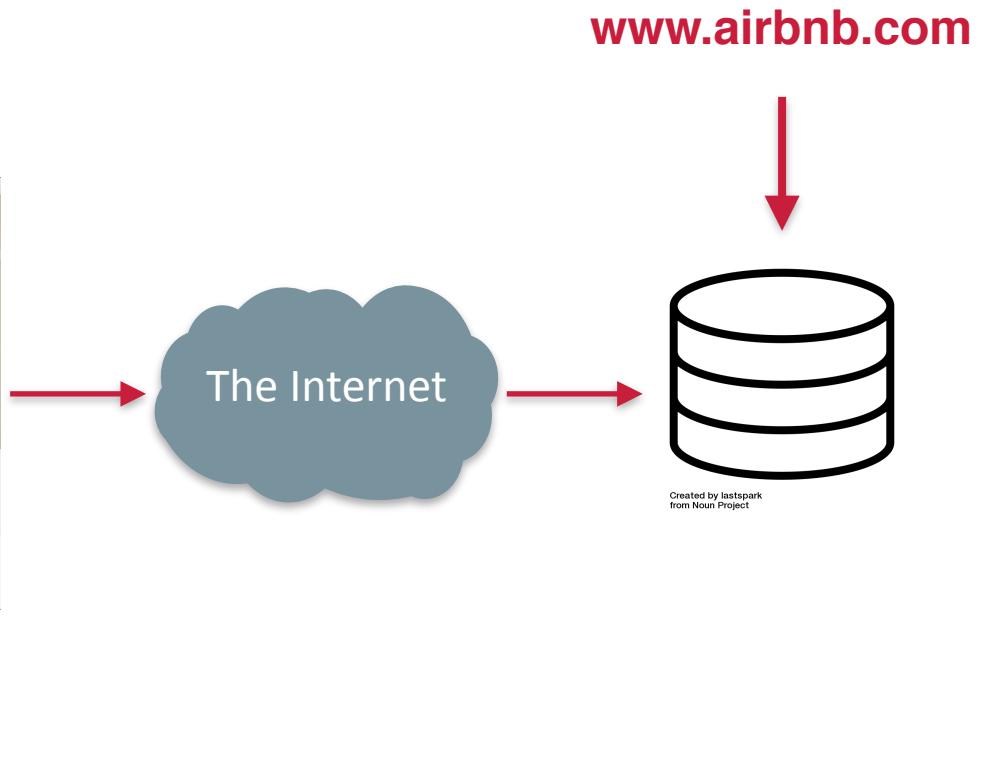
Jonathan's Web Browser



The Web



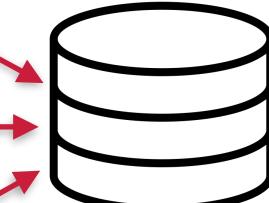
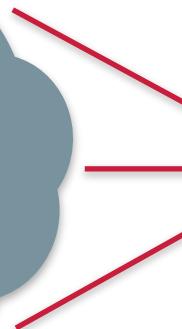
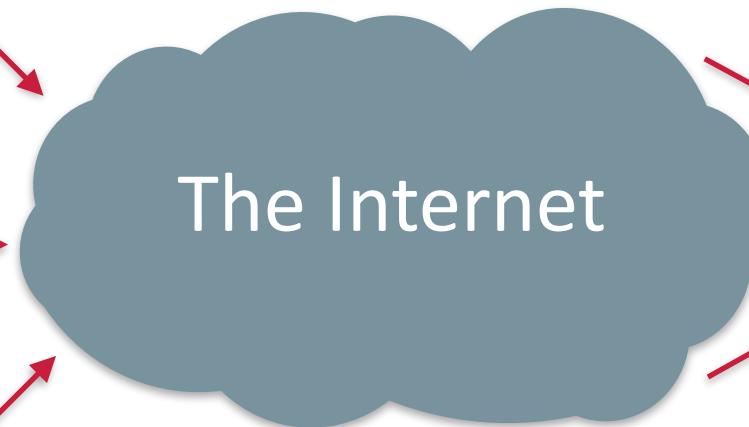
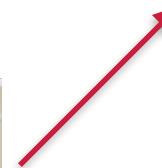
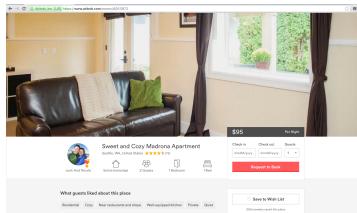
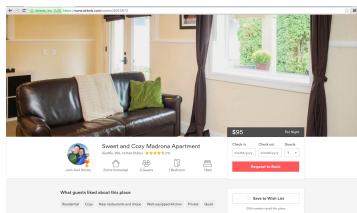
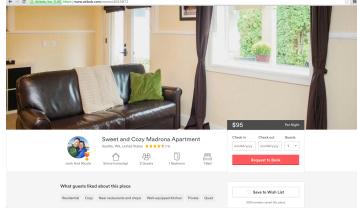
Jonathan's Web Browser



www.airbnb.com

Created by lastspark
from Noun Project

The Web



Created by lastspark
from Noun Project

What is a Protocol Anyway?

An agreement on how to communicate

What is a Protocol Anyway?

Syntax

- Format
- Order of messages sent/received

Semantics

- What messages mean
- Actions taken during communication

HTTP: The Protocol of the Web

Syntax

```
GET /index.html HTTP/1.1
```

```
Host: www.example.com
```

Semantics

- Go to the server
www.example.com
- Retrieve /index.html

Deconstructing a URL

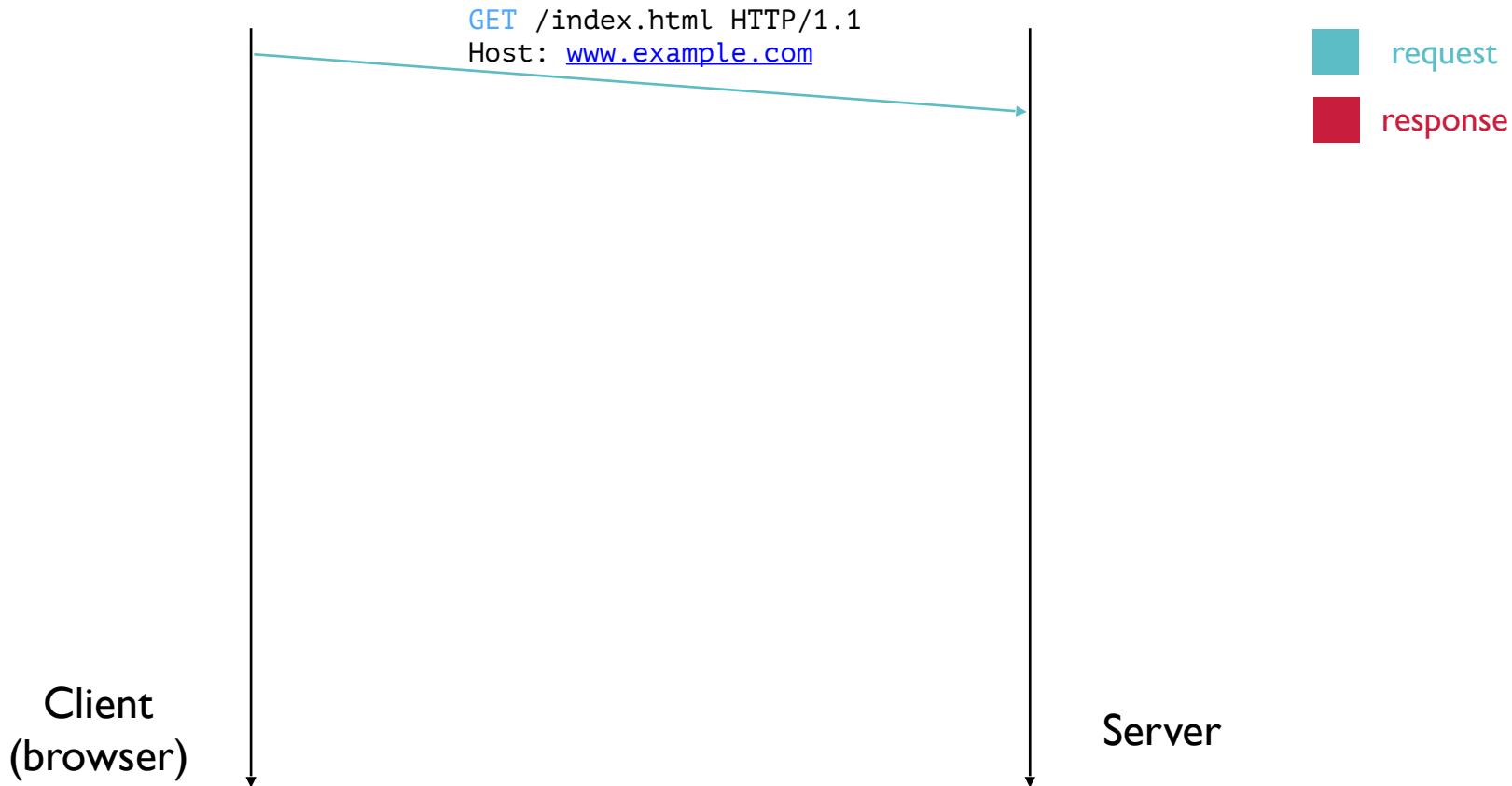
A Uniform Resource Locator (URL), commonly informally termed a web address (which term is not defined identically)[1] is a reference to a web resource that specifies its location on a computer network and a mechanism for retrieving it.

- Wikipedia

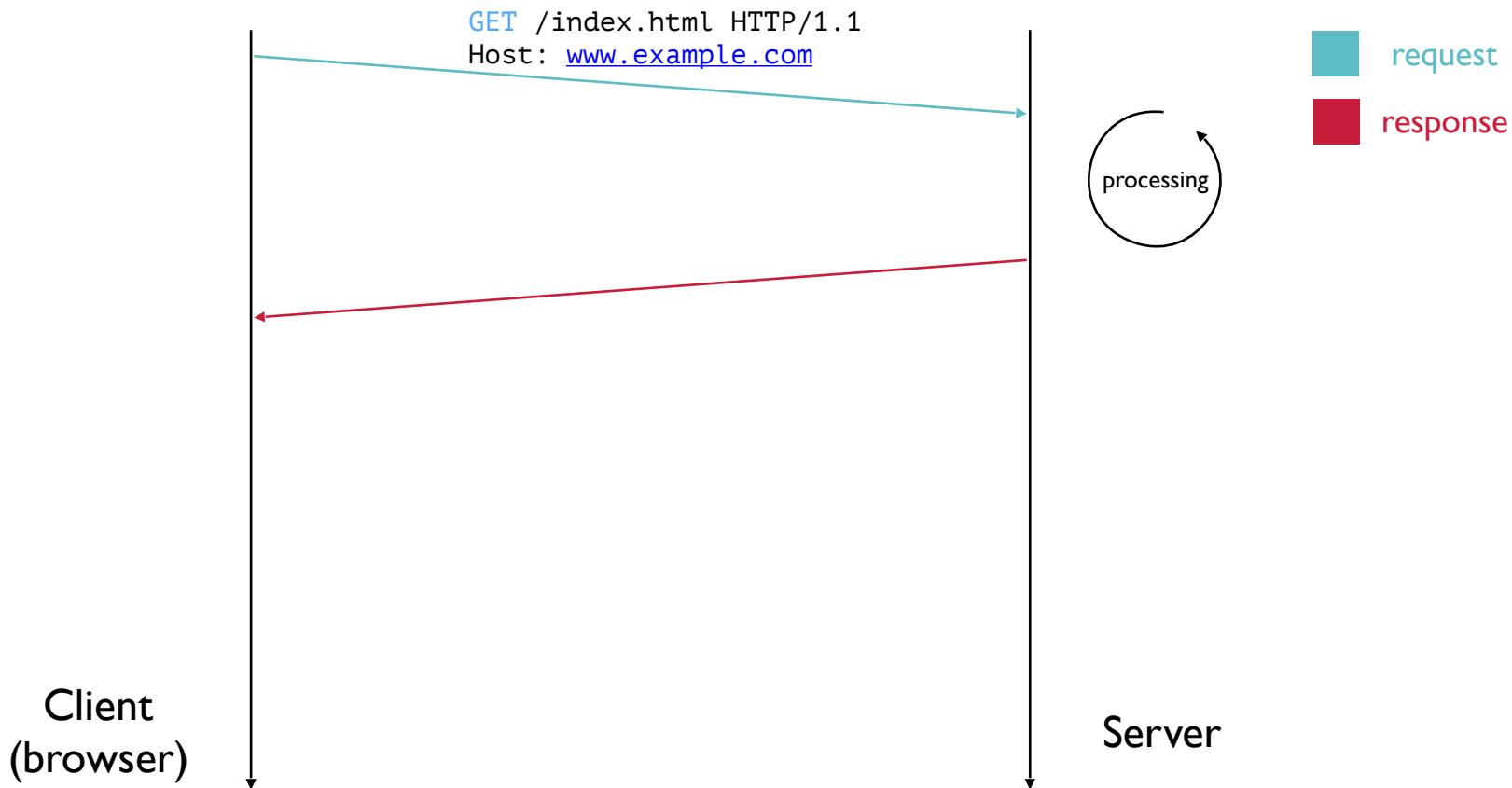
Deconstructing a URL



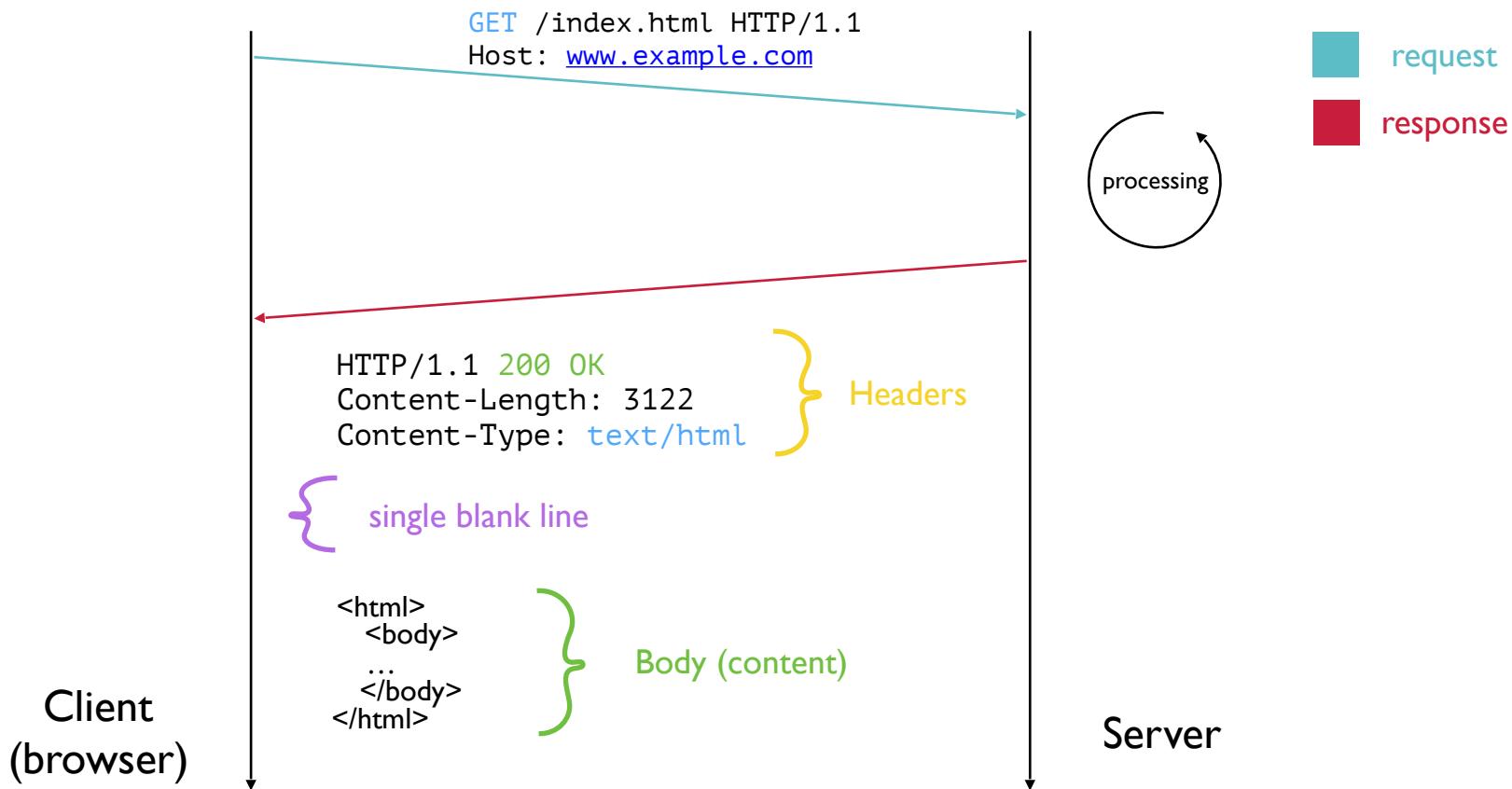
HTTP Life Cycle

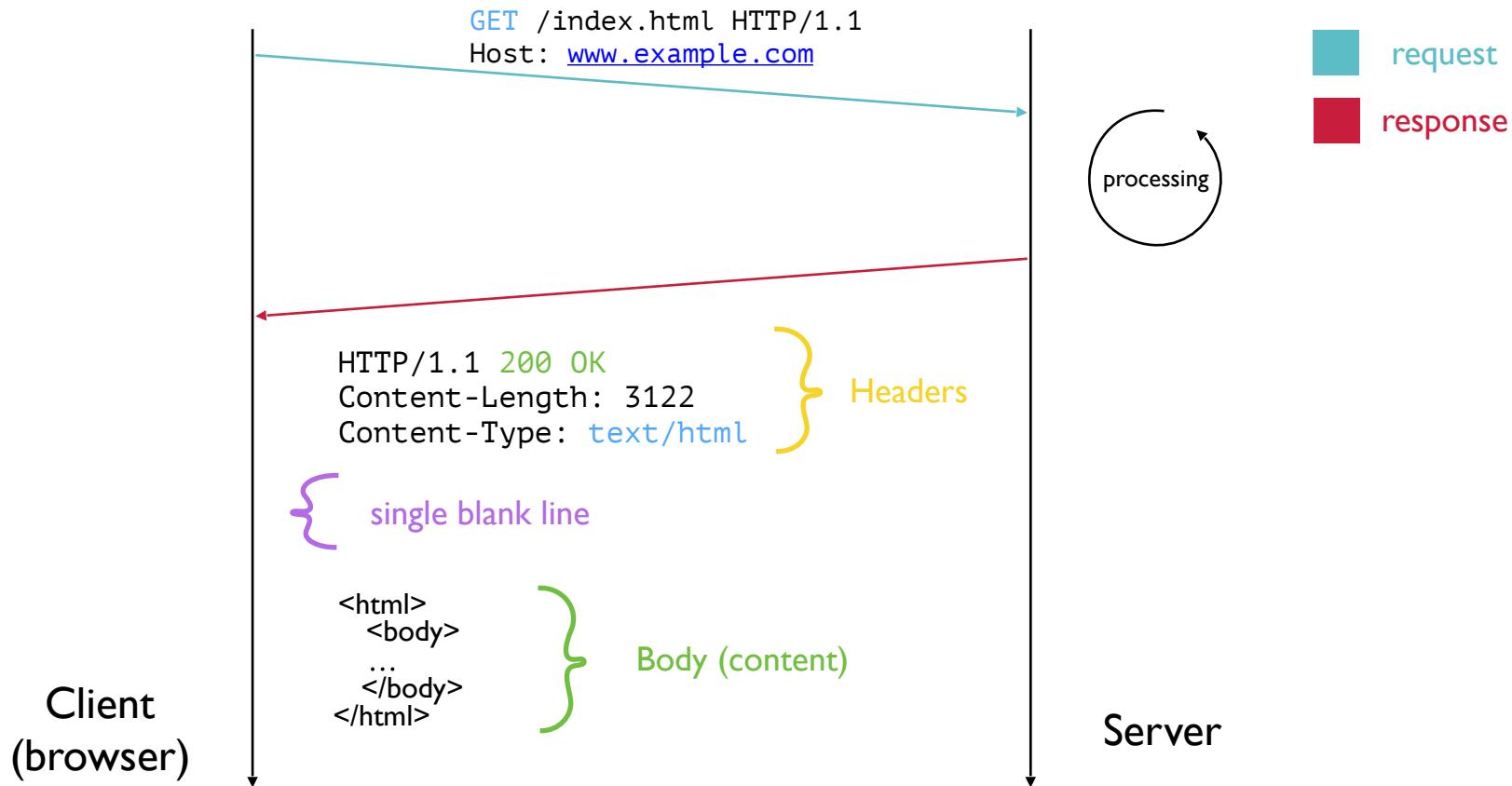


HTTP Life Cycle

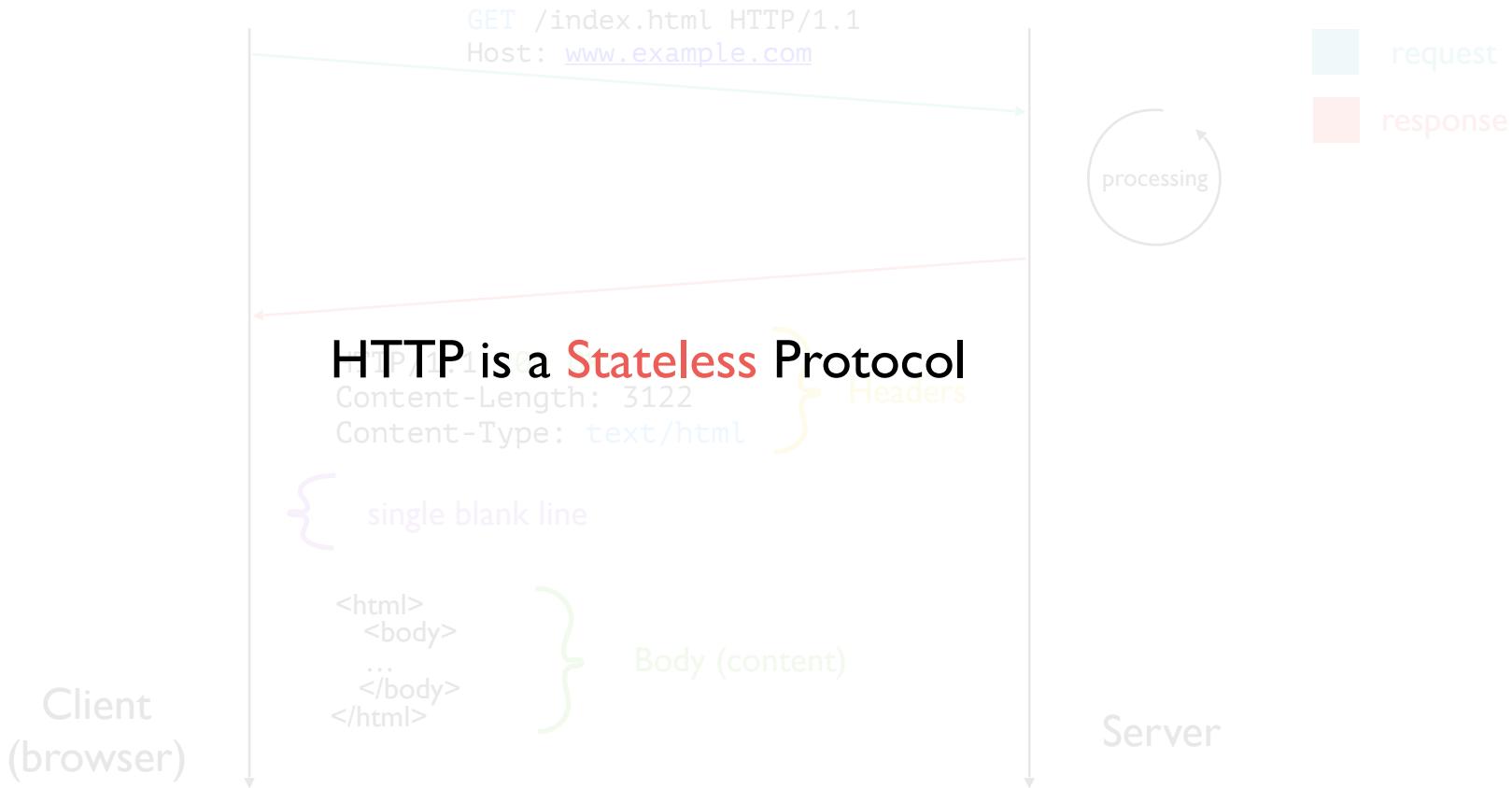


HTTP Life Cycle

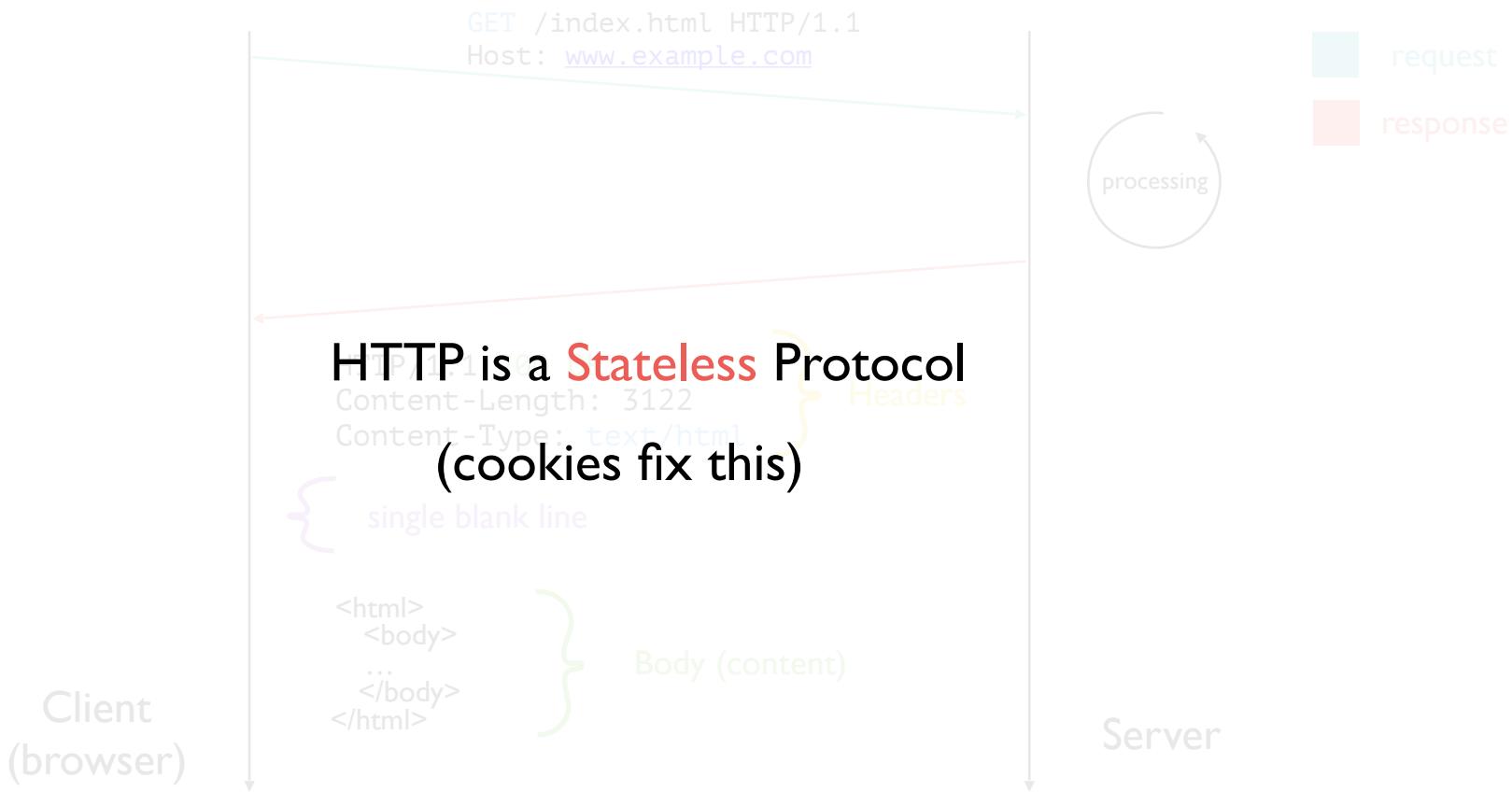




HTTP Life Cycle

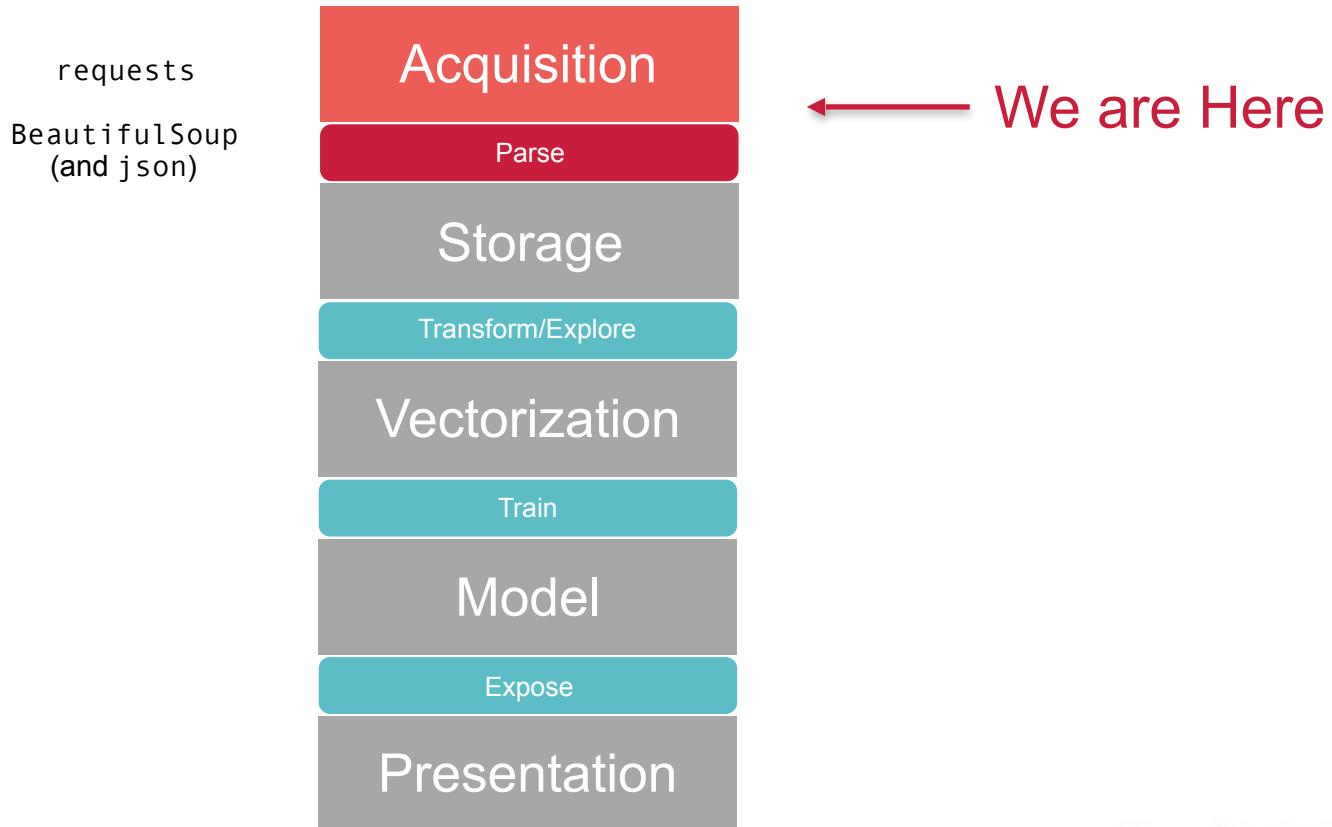


HTTP Life Cycle



Let's see that in action (in Python)

Process + Tools



Data Formats

Structured

- Easy to analyze
- Can be combined with other datasets
- Example: API responses and Bulk downloads

Unstructured

- Ubiquitous
- Can be hard to parse
- Often rich/complex data
- Example: Text and Images

Hierarchical Formats

XML

- Superset of HTML
- Can represent complex entities: tags, nesting, with attributes
- Not as wide support from programming languages
- More prone to parsing errors
- Human readable

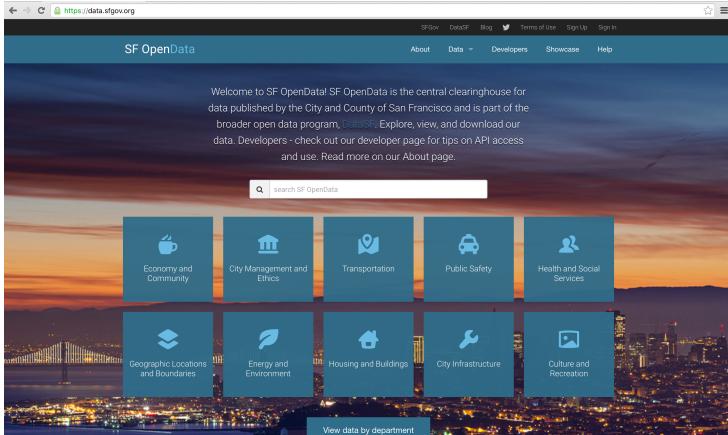
JSON

- Ubiquitous programming language support
- Smaller size of objects
- Easy to serialize/deserialize
- Not as optimized as binary formats
- Human Readable

File Formats

Type	Advantage	Structure	Example
CSV	Well suited to tabular processing	Flat	Bulk downloads
JSON	Ubiquitous programming language support	Hierarchical	API responses
XML/HTML	Can represent complex data models	Hierarchical	Web Pages
Binary	No limits on complexity of structure or state	Monolithic	PDF, Images, Proprietary Formats

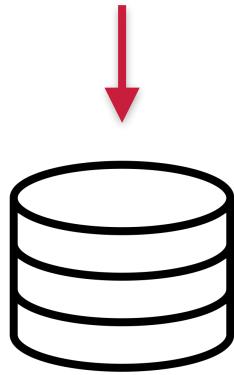
The Web



Jonathan's Web Browser



SF OpenData
Web Server



APIs

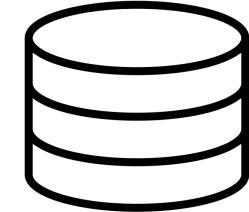
```
jonathan@ndinu$ python download_data.py
```



Jonathan's Script



SF OpenData
“database”



Created by lastspark
from Noun Project

Steps to Get Started with Foursquare

1. Create a Foursquare Account
2. Create an App to get credentials (keep these secret/hidden!)
3. Explore Endpoints from Documentation
4. Make (authenticated) requests!
5. Attribution

APIs

- Documentation/Endpoints
- Parameters/Query
- Data Limiting (results per page)
- Versioning
- Attribution and Policies

Nuances of APIs

- Authentication (key based, OAuth)
- Rate Limiting (requests per day/hour/minute)
- Data Limiting (results per page)
- Be Ethical and Respectful (always read the terms!)