

There's Always an API

Sometimes they make you work for it

Hi, I'm Matt Dennewitz

- VP Product, Pitchfork; Dir. Engineering, Wired
- I consult on data for baseball writers and MLB clubs
- @mattdennewitz on Twitter, Github

Agenda

- 101
- Your first scrape: Google Docs
- Interlude: HTML, JSON, XML, XPath
- Scaling up: Python
- What happens when the data isn't on the page?
- Advanced topics (time allowing)

What is scraping?

- Extracting information from a document
 - Rows from an HTML table
 - Text from a PDF
 - Images from Craigslist posts or museum websites
 - OCR'ing an image and reading its text
 - Spidering a website like Google

Tools

- Google Docs (surprise!)
- Chrome Developer Tools
- Python
- Scrapy

Strategy

1. "What do I want?"
2. Case the joint
3. Rob it just a little bit
4. Move in

“What do I want?”

- Envision the data you want, how you need it
 - “How will I scrape this data?” Script? Crawler?
 - “Do I need to scrape this more than once?”
 - “How do I need to shape the data?”
 - “What do I need to do to the data after I have it?”
Clean, verify, cross-link with another data set, ...?
 - “How/to where do I want to output the data?”

Case the joint

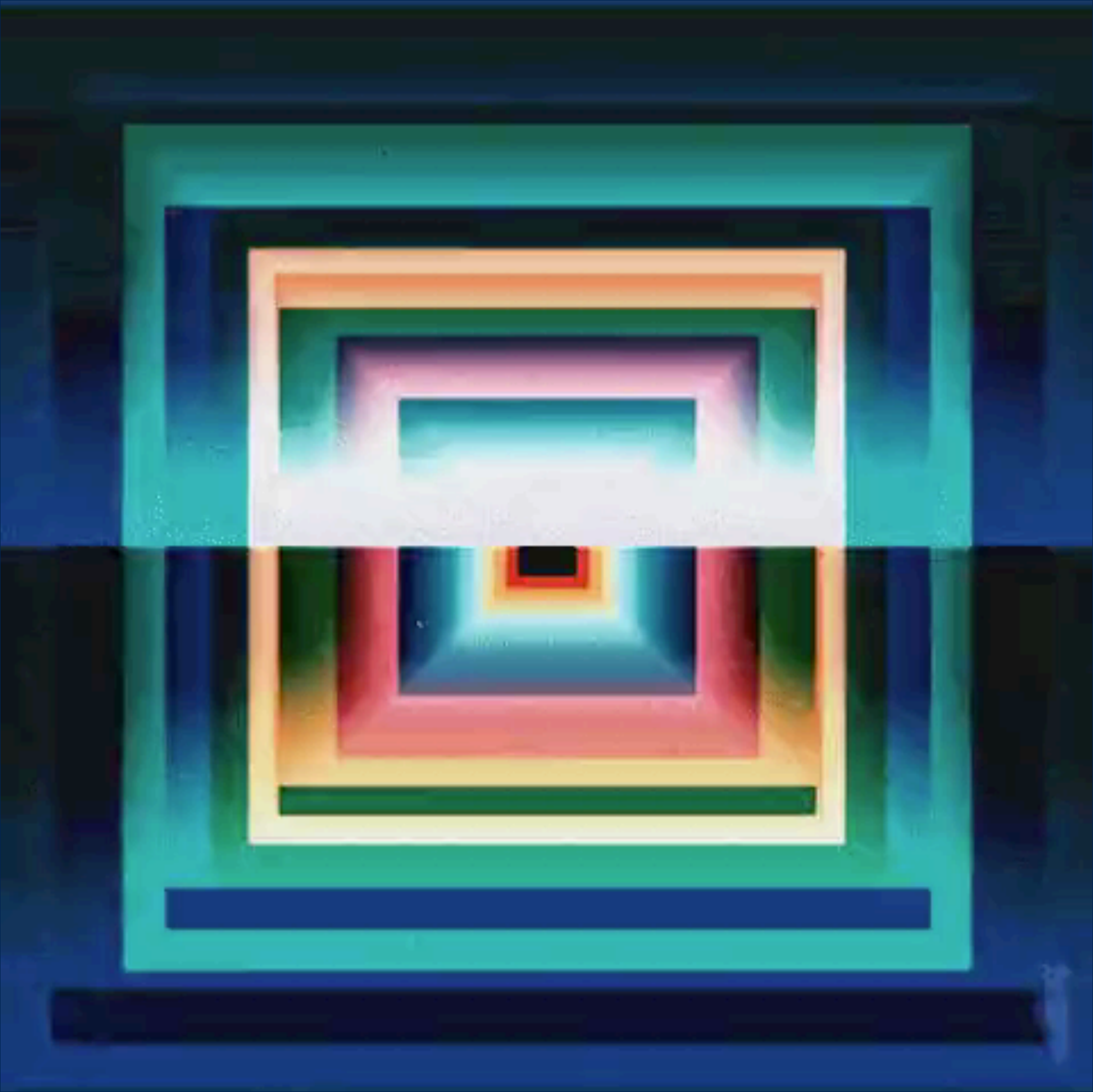
- Does the document seem scrape-ready? Does access come with preconditions?
 - Preconditions: password-protected? Online-only? Needs a special decoder?
- Look at how the data is presented in the document. Are there external dependencies, or is it self-contained?
 - External deps: more information on secondary pages, data in other spreadsheets or workbooks

Rob it just a little bit

- Prototype using a subset of the information
- Estimate how long scraping will take, determine imperative needs like throttling or a specific OS
- Validate your ideas about the data you wish to extract, correct bugs
 - Writing unit tests

Oceans 1101

- You've created a stable scraper which emits data in the format you want (CSV, JSON, XML, SQL, ...) to the location you want
- You understand its performance characteristics
- Go!



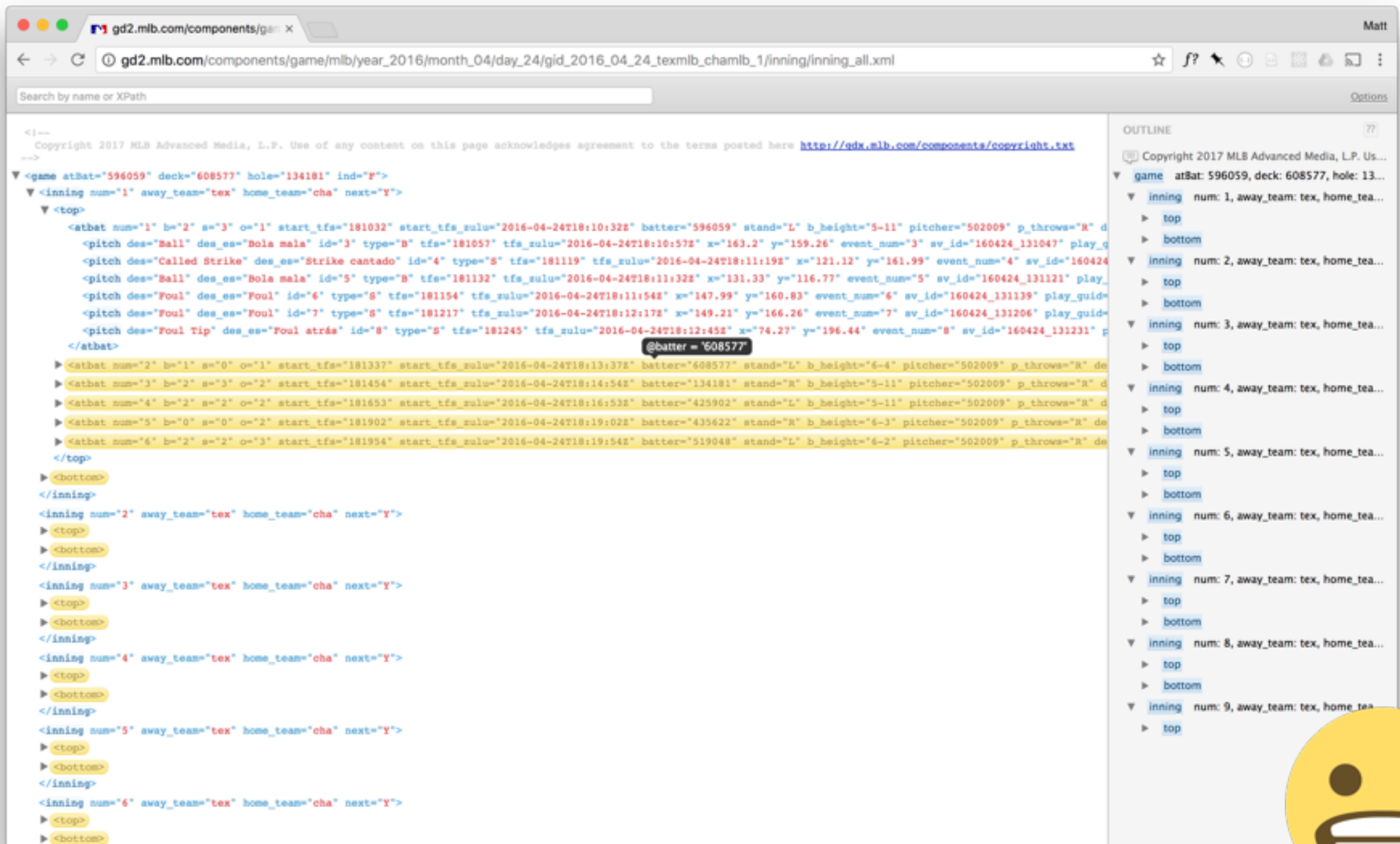
Interlude: formats

- Data is distributed in mercilessly innumerable formats
- The Big Three of web scraping
 - HTML
 - JSON
 - XML

Formats: XML

- eXtensible Markup Language
- Well-structured, self-validating, predictable
- Pedantic, though not with its charms

Formats: XML



gd2.mlb.com/components/game/mlb/year_2016/month_04/day_24/gid_2016_04_24_texmlb_chamlb_1/inning/inning_all.xml

Search by name or XPath

Copyright 2017 MLB Advanced Media, L.P. Use of any content on this page acknowledges agreement to the terms posted here <http://gd2.mlb.com/components/copyright.txt>

```
<!--
Copyright 2017 MLB Advanced Media, L.P. Use of any content on this page acknowledges agreement to the terms posted here http://gd2.mlb.com/components/copyright.txt
-->
<game atBat="596059" deck="608577" hole="134181" ind="F">
  <inning num="1" away_team="tex" home_team="cha" next="Y">
    <top>
      <atbat num="1" b="2" s="3" o="1" start_tfs="181032" start_tfs_zulu="2016-04-24T18:10:32Z" batter="596059" stand="L" b_height="5-11" pitcher="502009" p_throws="R" d
      <pitch des="Ball" des_es="Bola mala" id="3" type="B" tfs="181057" tfs_zulu="2016-04-24T18:10:57Z" x="163.2" y="159.26" event_num="3" sv_id="160424_131047" play_g
      <pitch des="Called Strike" des_es="Strike cantado" id="4" type="S" tfs="181119" tfs_zulu="2016-04-24T18:11:19Z" x="121.12" y="161.99" event_num="4" sv_id="160424
      <pitch des="Ball" des_es="Bola mala" id="5" type="B" tfs="181132" tfs_zulu="2016-04-24T18:11:32Z" x="131.33" y="116.77" event_num="5" sv_id="160424_131121" play_
      <pitch des="Foul" des_es="Foul" id="6" type="S" tfs="181154" tfs_zulu="2016-04-24T18:11:54Z" x="147.99" y="160.83" event_num="6" sv_id="160424_131139" play_guid
      <pitch des="Foul" des_es="Foul" id="7" type="S" tfs="181217" tfs_zulu="2016-04-24T18:12:17Z" x="149.21" y="166.26" event_num="7" sv_id="160424_131206" play_guid
      <pitch des="Foul Tip" des_es="Foul atrás" id="8" type="S" tfs="181245" tfs_zulu="2016-04-24T18:12:45Z" x="74.27" y="196.44" event_num="8" sv_id="160424_131231" p
    </atbat>
    <@batter = "608577">
      <atbat num="2" b="1" s="0" o="1" start_tfs="181337" start_tfs_zulu="2016-04-24T18:13:37Z" batter="608577" stand="L" b_height="6-4" pitcher="502009" p_throws="R" de
      <atbat num="3" b="2" s="3" o="2" start_tfs="181454" start_tfs_zulu="2016-04-24T18:14:54Z" batter="134181" stand="R" b_height="5-11" pitcher="502009" p_throws="R" d
      <atbat num="4" b="2" s="2" o="2" start_tfs="181653" start_tfs_zulu="2016-04-24T18:16:53Z" batter="425902" stand="L" b_height="5-11" pitcher="502009" p_throws="R" d
      <atbat num="5" b="0" s="0" o="2" start_tfs="181902" start_tfs_zulu="2016-04-24T18:19:02Z" batter="435622" stand="R" b_height="6-3" pitcher="502009" p_throws="R" de
      <atbat num="6" b="2" s="2" o="3" start_tfs="181954" start_tfs_zulu="2016-04-24T18:19:54Z" batter="519048" stand="L" b_height="6-2" pitcher="502009" p_throws="R" de
    </top>
    <bottom>
    </bottom>
  </inning>
  <inning num="2" away_team="tex" home_team="cha" next="Y">
    <top>
    </top>
    <bottom>
    </bottom>
  </inning>
  <inning num="3" away_team="tex" home_team="cha" next="Y">
    <top>
    </top>
    <bottom>
    </bottom>
  </inning>
  <inning num="4" away_team="tex" home_team="cha" next="Y">
    <top>
    </top>
    <bottom>
    </bottom>
  </inning>
  <inning num="5" away_team="tex" home_team="cha" next="Y">
    <top>
    </top>
    <bottom>
    </bottom>
  </inning>
  <inning num="6" away_team="tex" home_team="cha" next="Y">
    <top>
    </top>
    <bottom>
    </bottom>
  </inning>
  <inning num="7" away_team="tex" home_team="cha" next="Y">
    <top>
    </top>
    <bottom>
    </bottom>
  </inning>
  <inning num="8" away_team="tex" home_team="cha" next="Y">
    <top>
    </top>
    <bottom>
    </bottom>
  </inning>
  <inning num="9" away_team="tex" home_team="cha" next="Y">
    <top>
    </top>
    <bottom>
    </bottom>
  </inning>
</game>
```

OUTLINE

- Copyright 2017 MLB Advanced Media, L.P. Us...
- game atBat: 596059, deck: 608577, hole: 13...
- inning num: 1, away_team: tex, home_tea...
 - top
 - bottom
- inning num: 2, away_team: tex, home_tea...
 - top
 - bottom
- inning num: 3, away_team: tex, home_tea...
 - top
 - bottom
- inning num: 4, away_team: tex, home_tea...
 - top
 - bottom
- inning num: 5, away_team: tex, home_tea...
 - top
 - bottom
- inning num: 6, away_team: tex, home_tea...
 - top
 - bottom
- inning num: 7, away_team: tex, home_tea...
 - top
 - bottom
- inning num: 8, away_team: tex, home_tea...
 - top
 - bottom
- inning num: 9, away_team: tex, home_tea...
 - top
 - bottom

Formats: HTML

- Hypertext something something something
- XML-like, without the upside
- Needs stronger class of parser to heal broken code
- Less predictable, far more susceptible to changes in the wind

Formats: HTML

<p>

1.

```
<strong><span class="playerdef"><a href="http://  
www.baseballprospectus.com/card/card.php?id=102123">Alex Reyes</  
a></span>, RHP, <span class="teamdef"><a href="http://  
www.baseballprospectus.com/team_audit.php?team=SLN"  
target="blank">St. Louis Cardinals</a></span></strong><br>
```

```
Scouting Report: <a href="http://www.baseballprospectus.com/  
article.php?articleid=30958">LINK</a>
```

</p>



Formats: JSON

- JavaScript Object Notation
- Data objects with simple primitives: int, double, string, boolean, object (key/value pairs), array (untyped), null.
- Requires waaaaaaay less parsing, much easier to serialize
- No schemas, but validation tools exist
- Has taken over for XML in web data transmission

Formats: JSON

```
{  
  "prospect_year": "2017",  
  "player_id": 643217,  
  "player_first_name": "Andrew",  
  "player_last_name": "Benintendi",  
  "rank": 1,  
  "position": "OF",  
  "preseason100": 1,  
  "preseason20": 1,  
  "team_file_code": "BOS",  
}
```



Bonus: XPath

- XPath is a way to query XML (and HTML)
- It's got a super goofy syntax
- Very powerful, essential for scraping the web

Bonus: XPath

- **XPath:** `//table/tbody/tr`

- **HTML** (fragment):

```
<table>
  <thead>
    <tr>
      <th>Name</th><th>HR</th><th>SB</th>
    </tr>
  </thead>
  <tbody>
    <tr><td>Mike Trout</td><td>40</td><td>40</td></tr>
  </tbody>
</table>
```

- **Result:** `<tr><td>Mike Trout</td><td>40</td><td>40</td></tr>`

Bonus: XPath

- **XPath:** `//span[@class="playerdef"]/text()`
- **HTML:**
`<p>1. Eloy Jiminez, 0F, ...</p>`
- **Result:** "Eloy Jiminez"

Ok, time to scrape

Google Docs

- Fire up Google Docs, start a new spreadsheet
- `IMPORTXML` and `IMPORTHTML` are your friends
- Let's look at `IMPORTHTML`

IMPORTHTML

- Allows you put pull in specific list or tabular data from a web page
- Syntax:
`=IMPORTHTML(url, <"list" or "table">, [index])`

IMPORTHTML

- ESPN Home Run Tracker
- Syntax:
`=importhtml("http://
www.hittrackeronline.com/?perpage=1000",
"table", 17)`
- “Give me the 16th table on the page” (0-based indexing)

IMPORTHTML

IMPORTHTML Example - Google Docs

Secure <https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDlDGA1x5ovwUI8amXv9V95OnY/edit#gid=1937097544>

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

mattdennewitz@gmail.com

Comments Share

*Pages: * 1*, 2, 3, 4, 5, 6 Next

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
3	Date	Video	Path	Hitter	Hitter Team	Pitcher	Pitcher Team	INN	Ballpark	Type/Luck	True Dist.	Speed Off Bat	Elev. Angle	Horiz. Angle
4	10/02/16	Video	View	Contreras, Willsc	CHC	Wood, Blake	CIN	8	Great Americ...	PL		425	102.6	29.6
5	10/02/16	Video	View	Susac, Andrew	MIL	Rusin, Chris	COL	10	Coors Field	ND		417	109	25.3
6	10/02/16	Video	View	Cuthbert, Cheslo	KC	Tomlin, Josh	CLE	5	Kauffman Sta...	JE/L		413	101.4	29.4
7	10/02/16	Video	View	Gomes, Yan	CLE	Kennedy, Ian	KC	3	Kauffman Sta...	PL		412	105.6	27.8
8	10/02/16	Video	View	Santana, Doming	MIL	Estevez, Carlos	COL	8	Coors Field	ND		411	109.4	25.5
9	10/02/16	Video	View	Marte, Jefry	LAA	Gustave, Jandel	HOU	7	Angel Stadiu...	JE		406	104.1	23.2
10	10/02/16	Video	View	Drury, Brandon	ARI	Hand, Brad	SD	8	Chase Field	JE		405	112.1	20
11	10/02/16	Video	View	Carpenter, Matt	STL	Bastardo, Antoni	PIT	6	Busch Stadiu...	JE		402	100.8	31.7
12	10/02/16	Video	View	Sano, Miguel	MILN	Sale, Chris	CHW	3	U.S. Cellula...	PL		400	101.3	29.1
13	10/02/16	Video	View	Espinosa, Danny	WSH	Brice, Austin	MIA	5	Nationals Pa...	PL		399	101	27.8
14	10/02/16	Video	View	Telis, Tomas	MIA	Scherzer, Max	WSH	5	Nationals Pa...	JE		398	98.9	31
15	10/02/16	Video	View	Buxton, Byron	MILN	Sale, Chris	CHW	1	U.S. Cellula...	ITP/L		394	100.7	23.7
16	10/02/16	Video	View	Jaso, John	PIT	Broxton, Jonatha	STL	7	Busch Stadiu...	PL		394	101.5	29.3
17	10/02/16	Video	View	Montero, Miguel	CHC	Iglesias, Raisel	CIN	9	Great Americ...	PL		394	102.7	35.8
18	10/02/16	Video	View	Blackmon, Charl	COL	Cravy, Tyler	MIL	1	Coors Field	PL		389	102	39.5
19	10/02/16	Video	View	Ramirez, Hanley	BOS	Sanchez, Aaron	TOR	7	Fenway Park	ND		388	104.4	32.6
20	10/02/16	Video	View	Hood, Destin	MIA	Scherzer, Max	WSH	4	Nationals Pa...	JE/L		387	97.6	28.8
21	10/02/16	Video	View	Lucroy, Jonathan	TEX	Whitley, Chase	TB	4	Globe Life P...	JE		386	99.8	26
22	10/02/16	Video	View	Wielers, Matt	BAL	Layne, Tommy	NYN	6	Yankee Stadi...	ND		376	102	34.3
23	10/02/16	Video	View	Wielers, Matt	BAL	Cessa, Luis	NYN	4	Yankee Stadi...	PL		371	100	35.6
24	10/02/16	Video	View	Vogt, Stephen	OAK	Hernandez, Felix	SEA	1	Safeco Field	JE		353	96	30.1
25	10/02/16	Video	View	Travis, Devon	TOR	Price, David	BOS	5	Fenway Park	PL		350	96.3	34.5
26	10/02/16	Video	View	McCann, Brian	NYN	Gausman, Kevin	BAL	4	Yankee Stadi...	PL		333	94.3	38
27	10/01/16	Video	View	Carter, Chris	MIL	Ottavino, Adam	COL	10	Coors Field	ND		429	106.7	35.2

+ Home run tracker Brooks pitch values Explore

IMPORTHTML

- Brooks Baseball Player Pitch Logs

- Syntax:

```
=IMPORTHTML("http://  
www.brooksbaseball.net/pfxVB/  
tabdel_expanded.php?  
pitchSel=458584&game=gid_2016_06_27_bosm  
lb_tbamlb_1/  
&s_type=&h_size=700&v_size=500",  
"table")
```

IMPORTHTML

IMPORTHTML Example - Google Sheets

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv9V95OnY/edit#gid=0

IMPORTHTML Example

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Secure https://docs.google.com/spreadsheets/d/1Dma4hX1uh7ko8JdhYgDIDGA1x5ovwUI8amXv

Google Docs

- Useful for pulling in single tables, or keeping everything in a spreadsheet
- Data doesn't always exist in a single place
 - Spread across several pages
 - Spread across several files or APIs

Google Docs

- Useful for pulling in single tables, or keeping everything in a spreadsheet
- Data doesn't always exist in a single place
 - Spread across several pages
 - Spread across several files or APIs
- Automate as much as you can

Python time

- Beautiful language. Transcendental even.
- Robust ecosystem for handling data parsing, cleaning, making net requests, etc
- A+ community
- Runs anywhere



Python time

- I'm going to use two non-standard packages today:
 - `lxml`, for HTML parsing and cleaning
 - `requests`, for HTTP fetching



Strategy (again)

1. "What do I want?"
2. Case the joint
3. Rob it just a little bit
4. Move in

Strategy (again)

1. **“What do I want?”**: prospect rankings from BP, MLB, Baseball America
2. Case the joint
3. Rob it just a little bit
4. Move in

Strategy (again)

1. **“What do I want?”**: prospect rankings from BP, MLB
2. **Case the joint**: BP has dirty HTML. MLB loads a JSON file.
3. Rob it just a little bit
4. Move in

Strategy (again)

1. **“What do I want?”**: prospect rankings from BP, MLB
2. **Case the joint**: BP has dirty HTML. MLB loads a JSON file.
3. **Rob it just a little bit**: Get a feel for BP and BA’s HTML structure, examine MLB’s JSON file.
4. Move in

Strategy (again)

1. **“What do I want?”**: prospect rankings from BP, MLB.
2. **Case the joint**: BP has dirty HTML. MLB loads a JSON file.
3. **Rob it just a little bit**: Get a feel for BP and BA’s HTML structure, examine MLB’s JSON file.
4. **Move in**: Write three scripts, one for each.

Strategy (again)

- Fields to export:
 - Name
 - Rank
 - List type ("BP", "MLB", ...)
 - System ID (MLBAM ID, BP player ID, ...)

BP

- <http://www.baseballprospectus.com/article.php?articleid=31160>

BP

- First thing to do is inspect the source
 - Is there a pattern in the HTML you can engineer for, or an attribute you can target?
- Let's head to the console! Right click on the one of the capsules, and click "Inspect"

BP



BP

- Yes! Player data is in a paragraph tag, `<p>`, which contains a `` with class "playerdef"
- Get used to talking like this
- Using XPath, we can target that `` and walk up to its parent element, `<p>`, which gives us access to the whole player capsule

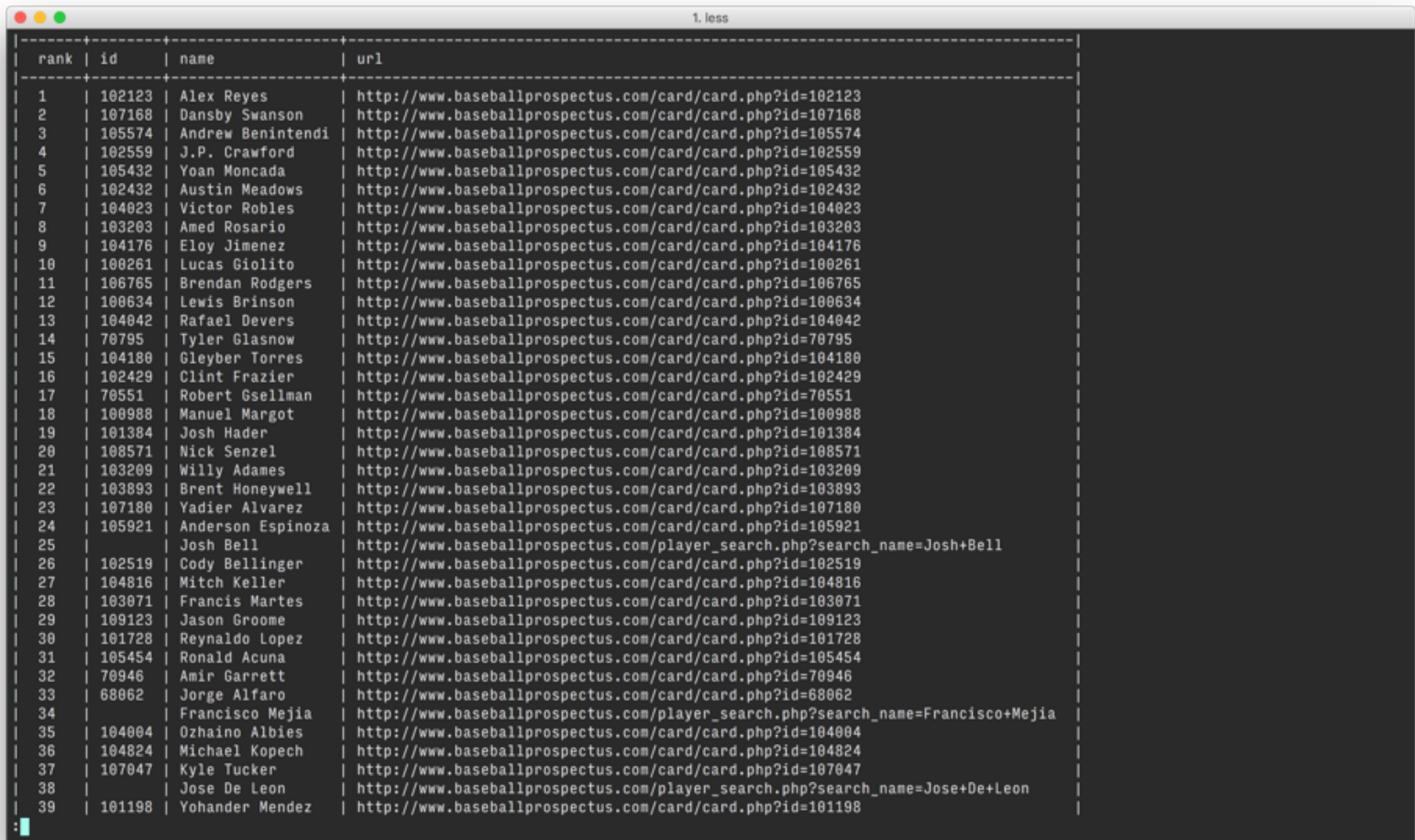
BP

- Beware: the “playerdef” class could be used anywhere. We need to find a reasonable scope for our XPath.
- Luckily for us, player capsules are in a <div> with class “article”, and that structure appears only once per article page across BP.
- XPath: `//div[@class="article"]//span[@class="playerdef"]/..`
- What else?

BP

- Code: <https://github.com/mattddennewitz/sloan-scraping/blob/master/bp-top-101-2017.py>
- Output: <https://github.com/mattddennewitz/sloan-scraping/blob/master/bp-2017.csv>

BP



A screenshot of a terminal window with a dark background and light gray text. The window title bar at the top shows three colored circles (red, yellow, green) on the left and the text "1. less" in the center. The terminal displays a table with four columns: "rank", "id", "name", and "url". The table is enclosed in a dashed border. The data is as follows:

rank	id	name	url
1	102123	Alex Reyes	http://www.baseballprospectus.com/card/card.php?id=102123
2	107168	Dansby Swanson	http://www.baseballprospectus.com/card/card.php?id=107168
3	105574	Andrew Benintendi	http://www.baseballprospectus.com/card/card.php?id=105574
4	102559	J.P. Crawford	http://www.baseballprospectus.com/card/card.php?id=102559
5	105432	Yoan Moncada	http://www.baseballprospectus.com/card/card.php?id=105432
6	102432	Austin Meadows	http://www.baseballprospectus.com/card/card.php?id=102432
7	104023	Victor Robles	http://www.baseballprospectus.com/card/card.php?id=104023
8	103203	Amed Rosario	http://www.baseballprospectus.com/card/card.php?id=103203
9	104176	Eloy Jimenez	http://www.baseballprospectus.com/card/card.php?id=104176
10	100261	Lucas Giolito	http://www.baseballprospectus.com/card/card.php?id=100261
11	106765	Brendan Rodgers	http://www.baseballprospectus.com/card/card.php?id=106765
12	100634	Lewis Brinson	http://www.baseballprospectus.com/card/card.php?id=100634
13	104042	Rafael Devers	http://www.baseballprospectus.com/card/card.php?id=104042
14	70795	Tyler Glasnow	http://www.baseballprospectus.com/card/card.php?id=70795
15	104180	Gleyber Torres	http://www.baseballprospectus.com/card/card.php?id=104180
16	102429	Clint Frazier	http://www.baseballprospectus.com/card/card.php?id=102429
17	70551	Robert Gsellman	http://www.baseballprospectus.com/card/card.php?id=70551
18	100988	Manuel Margot	http://www.baseballprospectus.com/card/card.php?id=100988
19	101384	Josh Hader	http://www.baseballprospectus.com/card/card.php?id=101384
20	108571	Nick Senzel	http://www.baseballprospectus.com/card/card.php?id=108571
21	103209	Willy Adames	http://www.baseballprospectus.com/card/card.php?id=103209
22	103893	Brent Honeywell	http://www.baseballprospectus.com/card/card.php?id=103893
23	107180	Yadier Alvarez	http://www.baseballprospectus.com/card/card.php?id=107180
24	105921	Anderson Espinoza	http://www.baseballprospectus.com/card/card.php?id=105921
25		Josh Bell	http://www.baseballprospectus.com/player_search.php?search_name=Josh+Bell
26	102519	Cody Bellinger	http://www.baseballprospectus.com/card/card.php?id=102519
27	104816	Mitch Keller	http://www.baseballprospectus.com/card/card.php?id=104816
28	103071	Francis Martes	http://www.baseballprospectus.com/card/card.php?id=103071
29	109123	Jason Groome	http://www.baseballprospectus.com/card/card.php?id=109123
30	101728	Reynaldo Lopez	http://www.baseballprospectus.com/card/card.php?id=101728
31	105454	Ronald Acuna	http://www.baseballprospectus.com/card/card.php?id=105454
32	70946	Amir Garrett	http://www.baseballprospectus.com/card/card.php?id=70946
33	68062	Jorge Alfaro	http://www.baseballprospectus.com/card/card.php?id=68062
34		Francisco Mejia	http://www.baseballprospectus.com/player_search.php?search_name=Francisco+Mejia
35	104004	Ozhaino Albies	http://www.baseballprospectus.com/card/card.php?id=104004
36	104824	Michael Kopech	http://www.baseballprospectus.com/card/card.php?id=104824
37	107047	Kyle Tucker	http://www.baseballprospectus.com/card/card.php?id=107047
38		Jose De Leon	http://www.baseballprospectus.com/player_search.php?search_name=Jose+De+Leon
39	101198	Yohander Mendez	http://www.baseballprospectus.com/card/card.php?id=101198


BP

- What did we do?
- Inspected the page
- Found critical path to data, wrote supporting XPaths
- Scripted collecting and outputting the data


MLB

- <http://m.mlb.com/prospects/2017>

MLB

- Again, start by inspecting the source
- Try to find “Benintendi” or “Moncada”
- 

MLB

- Again, start by inspecting the source
- Try to find “Benintendi” or “Moncada” in the HTML
-  “uhh”

MLB

- Websites love to load data asynchronously.
 - LOVE to
- Let's head to the Inspector's Network panel to poke around and find the source
 - In Chrome: Ctrl+Shift+I (Windows) or Cmd+Opt+I (Mac), then select "Network"

MLB

- Websites love to load data asynchronously.
 - LOVE to
- Let's head to the Inspector's Network panel to poke around and find the source
 - In Chrome: Ctrl+Shift+I (Windows) or Cmd+Opt+I (Mac), then select "Network"
- Let's start by looking under "XHR", the typical place to look for dynamically loaded data

MLB

MLB.com 2017 Prospect Watch

m.mlb.com/prospects/2017

Top 100 Prospects 10 by Position 50 Draft 30 by Team 30 International PDP Archive

Jump to Player Search 30689 COMMENTS

1 ANDREW BENINTENDI OF

2 YOAN MONCADA 2B

3 GLEYBER TORRES SS

4 DANSBY SWANSON SS

5

6

7

8

Elements Security Sources Network Timeline Profiles Application Audits Console

View: [Icons] Preserve log Disable cache Offline No throttling

Filter [] Reg [] Hide data URLs All XHR JS CSS Img Media Font Doc WS Manifest Other

Name	Status	Type	Initiator	Size	Time	Waterfall	10.00 s
NTU6ODgtNWNjZmE0NmUjJkOGM0ZTQ2ZjM4NmNmM2Y2YmQ0NzQ4OTE2YmVhNmE4ZTQwMDM5M2M3D.json	200	xhr	intercept.js:180	1.2 KB	243 ms		
livefyre_mlb.jsp?tokenType=collection&url=http%3A%2F%2Fm.mlb.com%2Fprospects%2F2017&title=MLB.com%25202020...	200	xhr	jquery.js:4	751 B	20 ms		
playerProspects.json	304	xhr	bundler?is=http://i.bamstatic.com/...	215 B	149 ms		
playerProspects.json	200	xhr	bundler?is=http://i.bamstatic.com/...	811 KB	530 ms		
playerProspects.json	200	xhr	bundler?is=http://i.bamstatic.com/...	835 KB	570 ms		
playerProspects.json	200	xhr	bundler?is=http://i.bamstatic.com/...	286 KB	137 ms		
playerProspects.json	200	xhr	bundler?is=http://i.bamstatic.com/...	254 KB	149 ms		

16 / 182 requests | 2.2 MB / 3.4 MB transferred | Finish: 13.16 s | DOMContentLoaded: 801 ms | Load: 5.15 s

MLB

- “playerProspects.json” looks promising
- We know it’s a JSON file
- The filename is a pretty dead giveaway
- When we open it up, it has a ton of prospect data

MLB

- Here, we have a JSON file
- Let's inspect the structure to find exactly what attributes we would like to scrape
- Fast-forward: the "prospect_players" key has prospects for all teams! And it has the Top 100 under the "prospects" key.

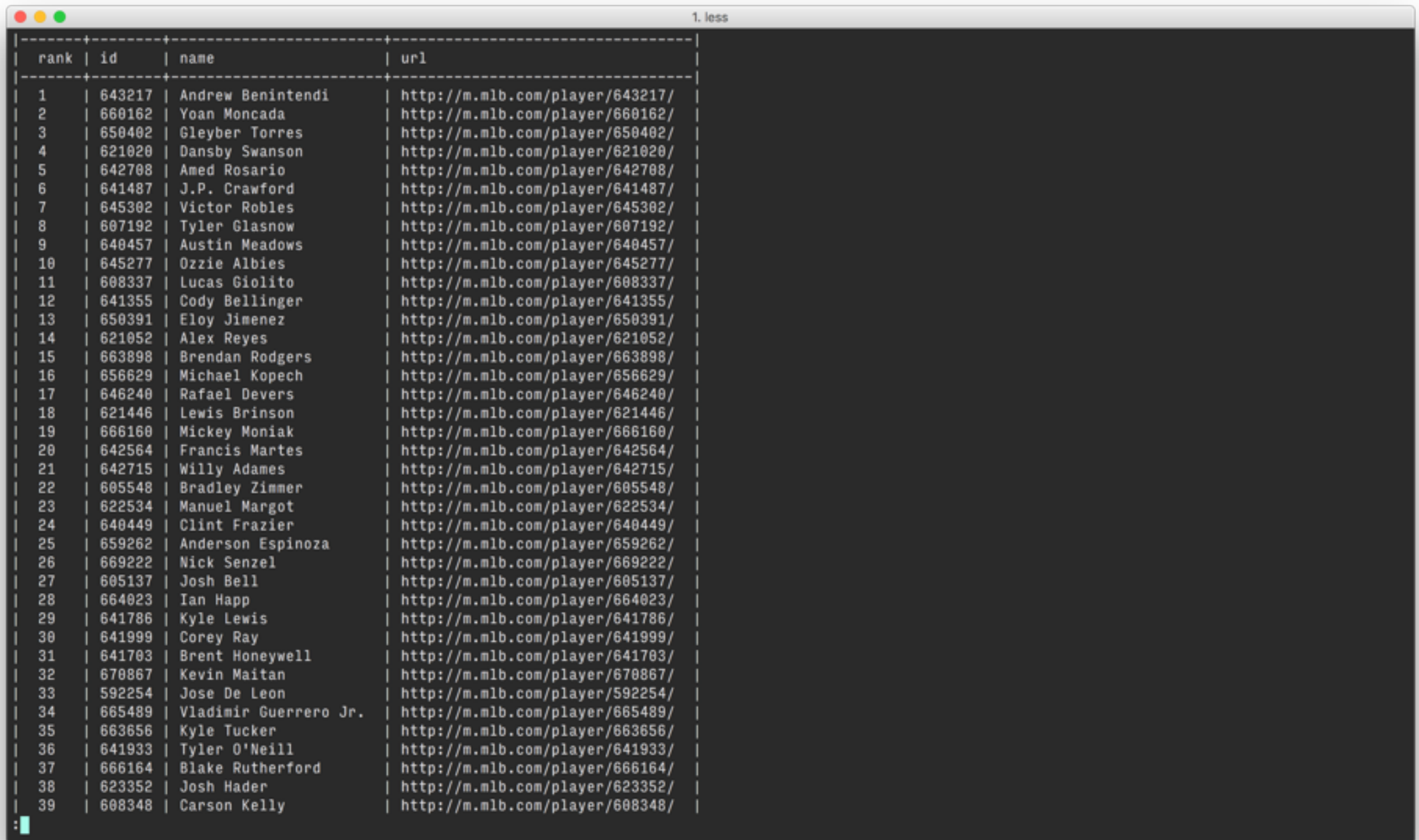
MLB

```
{  
  "prospect_year": "2017",  
  "player_id": 643217,  
  "player_first_name": "Andrew",  
  "player_last_name": "Benintendi",  
  "rank": 1,  
  "position": "OF",  
  "preseason100": 1,  
  "preseason20": 1,  
  "team_file_code": "BOS",  
}
```

MLB

- Using Python's out-of-box JSON parser, we can easily parse this file and extract players
- Code: `https://github.com/mattddennewitz/sloan-scraping/blob/master/mlb-top-100-2017.py`
- Output: `https://github.com/mattddennewitz/sloan-scraping/blob/master/mlb-2017.csv`

MLB



A terminal window titled "1. less" displays a list of MLB players. The data is presented in a table with four columns: rank, id, name, and url. The list contains 39 entries, starting with Andrew Benintendi at rank 1 and ending with Carson Kelly at rank 39. Each entry includes a unique ID and a corresponding URL from MLB's website.

rank	id	name	url
1	643217	Andrew Benintendi	http://m.mlb.com/player/643217/
2	660162	Yoan Moncada	http://m.mlb.com/player/660162/
3	650402	Gleyber Torres	http://m.mlb.com/player/650402/
4	621020	Dansby Swanson	http://m.mlb.com/player/621020/
5	642708	Amed Rosario	http://m.mlb.com/player/642708/
6	641487	J.P. Crawford	http://m.mlb.com/player/641487/
7	645302	Victor Robles	http://m.mlb.com/player/645302/
8	607192	Tyler Glasnow	http://m.mlb.com/player/607192/
9	640457	Austin Meadows	http://m.mlb.com/player/640457/
10	645277	Ozzie Albies	http://m.mlb.com/player/645277/
11	608337	Lucas Giolito	http://m.mlb.com/player/608337/
12	641355	Cody Bellinger	http://m.mlb.com/player/641355/
13	650391	Eloy Jimenez	http://m.mlb.com/player/650391/
14	621052	Alex Reyes	http://m.mlb.com/player/621052/
15	663898	Brendan Rodgers	http://m.mlb.com/player/663898/
16	656629	Michael Kopech	http://m.mlb.com/player/656629/
17	646240	Rafael Devers	http://m.mlb.com/player/646240/
18	621446	Lewis Brinson	http://m.mlb.com/player/621446/
19	666160	Mickey Moniak	http://m.mlb.com/player/666160/
20	642564	Francis Martes	http://m.mlb.com/player/642564/
21	642715	Willy Adames	http://m.mlb.com/player/642715/
22	605548	Bradley Zimmer	http://m.mlb.com/player/605548/
23	622534	Manuel Margot	http://m.mlb.com/player/622534/
24	640449	Clint Frazier	http://m.mlb.com/player/640449/
25	659262	Anderson Espinoza	http://m.mlb.com/player/659262/
26	669222	Nick Senzel	http://m.mlb.com/player/669222/
27	605137	Josh Bell	http://m.mlb.com/player/605137/
28	664023	Ian Happ	http://m.mlb.com/player/664023/
29	641786	Kyle Lewis	http://m.mlb.com/player/641786/
30	641999	Corey Ray	http://m.mlb.com/player/641999/
31	641703	Brent Honeywell	http://m.mlb.com/player/641703/
32	670867	Kevin Maitan	http://m.mlb.com/player/670867/
33	592254	Jose De Leon	http://m.mlb.com/player/592254/
34	665489	Vladimir Guerrero Jr.	http://m.mlb.com/player/665489/
35	663656	Kyle Tucker	http://m.mlb.com/player/663656/
36	641933	Tyler O'Neill	http://m.mlb.com/player/641933/
37	666164	Blake Rutherford	http://m.mlb.com/player/666164/
38	623352	Josh Hader	http://m.mlb.com/player/623352/
39	608348	Carson Kelly	http://m.mlb.com/player/608348/

Recap

- We've used the four step approach to plan for consistent output across disparate systems
- We've used tools like the Inspector to probe for data
- We've written very simple yet powerful scripts in Python to download prospect lists
- We've streamlined the data into a consistent shape
- Our scripts are easily reusable

Next steps

- Since we were clever and included system IDs, we can tie it all together using a baseball player ID registry
 - Chadwick Register
 - Smart Fantasy Baseball
 - Crunchtime

Tools

- Hopefully there's time to talk about this

Tools

- requests: A beautiful HTTP library
- lxml: A beautiful XML and HTML parsing library. Tricky to install on Windows, binaries are available.
- BeautifulSoup: another A+ HTML parser
- Scrapy: a very robust Python framework for crawling websites

Code

- The code and output from this session is online at: <https://github.com/mattdennewitz/2017-sloan-data-scraping>

Thanks!

- Questions?
- If we have some time left, we could try a bit of live coding
- If you have very specific scraping questions, find me after and let's talk