

Got data?

Tim Chartier

Depart. of Math & CS

DAVIDSON





Part 1: sports data



Data

- The sports data we used last time and in your homework came from:
<http://www.masseyratings.com>
- One file contains data on the games.
- Another file contains the names of the teams.



Game data

Each row of the file corresponds to a game and the columns contain the following data:

column 1 = date of game as measured as days since 1/1/0000

column 2 = date in YYYYMMDD format

column 3 = team 1 index

column 4 = team 1 home field (1 = home, -1 = away, 0 = neutral)

column 5 = team 1 score

column 6 = team 2 index

column 7 = team 2 home field (1 = home, -1 = away, 0 = neutral)

column 8 = team 2 score

Pile of numbers

Here is an example of the data.

735181,20121109, 345, 1, 117, 85, -1, 75

735181,20121109, 223, 1, 65, 271, -1, 58

735181,20121109, 190, 1, 97, 165, -1, 59

735181,20121109, 113, 1, 81, 90, -1, 76

Team data

Another file gives the enumeration of the teams.
For example,

- 1, Air_Force
- 2, Akron
- 3, Alabama
- 4, Alabama_A&M
- 5, Alabama_St
- 6, Albany_NY
- 7, Alcorn_St
- 8, American_Univ

Reading the numbers

Let's determine the game represented by the first line of the data file.

735181,20121109, 345, 1, 117, 85, -1, 75

735181,20121109, 223, 1, 65, 271, -1, 58

735181,20121109, 190, 1, 97, 165, -1, 59

735181,20121109, 113, 1, 81, 90, -1, 76

- The game was played on Nov. 9, 2012.
- It was played between teams 345 and 85.
- The score was 117-75.

Xavier vs. F. Dickinson

- Looking at the team file, we see that team 345 is Xavier and team 85 is Fairleigh Dickinson.

```
735181,20121109, 345, 1, 117, 85, -1, 75
```

```
735181,20121109, 223, 1, 65, 271, -1, 58
```

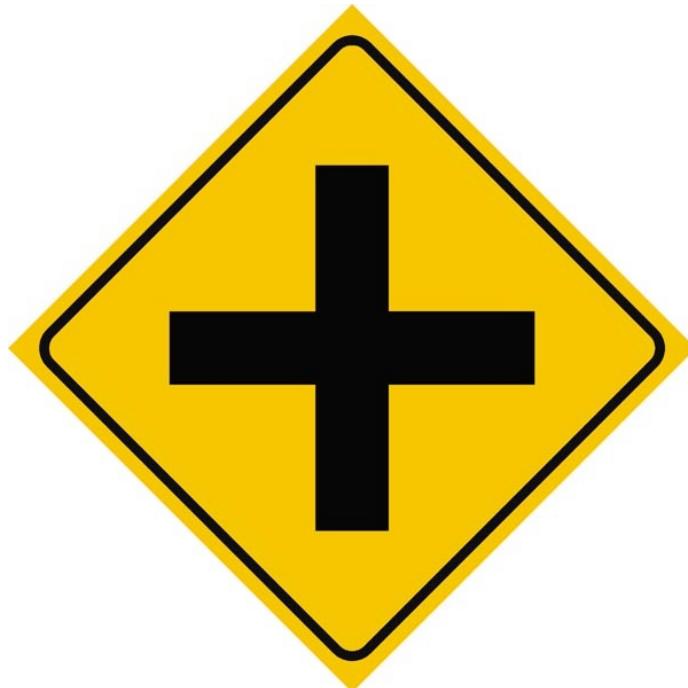
```
735181,20121109, 190, 1, 97, 165, -1, 59
```

```
735181,20121109, 113, 1, 81, 90, -1, 76
```

- Note the game was played at Xavier.

Data directions

Now, we see where to go to download such data for a variety of sports.



Getting Massey

- Go to <http://www.masseyratings.com>
- Click the “Data” button.

Popular **MASSEY RATING\$**

LANDING	RANK	SCHED	TOP 2
College I NCAA-D1 Baseball	Ratings	Games	Oregon St Florida St
Major League I MLB Baseball	Ratings	Games	Los Angeles Dodgers Cleveland Indians
Men I NCAA-D1 Basketball	Ratings	Games	Virginia Villanova
Women I NCAA-D1 Basketball	Ratings	Games	Connecticut Baylor
NBA Basketball	Ratings	Games	Houston Rockets Golden State Warriors
College I FBS Football	Ratings	Games	Alabama Georgia
NFL Football	Ratings	Games	Philadelphia Eagles Minnesota Vikings
NHL Hockey	Ratings	Games	Vegas Golden Knights Boston Bruins
NASCAR Cup Racing	Ratings	Games	Austin Dillon Darrell Wallace Jr
Domestic Pro I Top Soccer	Ratings	Games	Manchester City Barcelona

GOTickets has all the college football tickets and NCAA tournament tickets you need. Check them out!

Today's Games

Playoff Brackets

Ratings Archive

College Football Composite

College Basketball Composite

College Baseball Composite

College Sports Rankings

List of Colleges

Mascot Search

Data

S&P 500

Currencies

Pick your sport

Now, pick your sport. Let's pick men's basketball for 2017.

LEAGUE	YEAR
College Baseball	2017
College Basketball Women's	2017
College Basketball	2017
National Hockey League	2017
National Basketball Association	2017
National Football League	2016
National Basketball Association	2016
College Baseball	2016
College Baseball	2016
College Basketball Women's	2016

Pick your association

Select NCAA, among the various options.

[cb2017](#) > [ACCA](#) | [CCAA](#) | [CIS](#) | [NAIA](#) | [NCCAA](#) | **NCAA** | [NJCAA](#) | [Others](#) | [USCAA](#) | [CCCAA](#) | [NWAACC](#) |

[Prev](#) | [Next](#) | [All](#)

2017-02-21	@Ball St	79	E Michigan	72
2017-02-21	Kent	77	@Buffalo	69
2017-02-21	N Illinois	89	@C Michigan	66
2017-02-21	W Michigan	61	@Toledo	56 01
2017-02-21	Monmouth NJ	82	@Fairfield	62
2017-02-21	@Belmont	96	Trevecca Naz	45
2017-02-21	@Wichita St	109	Evansville	83
2017-02-21	Purdue	74	@Penn St	70 01
2017-02-21	Rhode Island	67	@La Salle	56
2017-02-21	Auburn	98	@LSU	75
2017-02-21	@Florida	81	South Carolina	66
2017-02-21	@Virginia Tech	71	Clemson	70
2017-02-21	@Baylor	60	Oklahoma	54

Division

Select NCAA I to get Division 1 scores.

[cb2017 < NCAA](#) [NCAA I](#) [NCAA II](#) | [NCAA III](#) |

[Prev](#) | [Next](#) | [All](#)

2017-02-21	@Ball St	79	E Michigan	72
2017-02-21	Kent	77	@Buffalo	69
2017-02-21	N Illinois	89	@C Michigan	66
2017-02-21	W Michigan	61	@Toledo	56 01
2017-02-21	Monmouth NJ	82	@Fairfield	62
2017-02-21	@Belmont	96	Trevecca Naz	45
2017-02-21	@Wichita St	109	Evansville	83
2017-02-21	Purdue	74	@Penn St	70 01
2017-02-21	Rhode Island	67	@La Salle	56
2017-02-21	Auburn	98	@LSU	75
2017-02-21	@Florida	81	South Carolina	66
2017-02-21	@Virginia Tech	71	Clemson	70
2017-02-21	@Baylor	60	Oklahoma	54
2017-02-21	@Bowling Green	66	Akron	65
2017-02-21	@Dayton	83	George Mason	70
2017-02-21	@Detroit	91	Cleveland St	83

Having it all

Click “All”.

[cb2017](#) | < NCAA | < NCAA | > Horizon | America East | Atlantic Coast | Atlantic Sun | Atlantic League | Metro Atlantic | Mid-American | Mid-Eastern AC | Missouri Val | Mountain West | Nor Belt | West Coast | Western Athletic | Summit Lg | Pac 12 | American Athletic |

[Prev](#) | [Next](#) | [All](#)

2017-02-21	@Ball St	79	E Michigan	72
2017-02-21	Kent	77	@Buffalo	69
2017-02-21	N Illinois	89	@C Michigan	66
2017-02-21	W Michigan	61	@Toledo	56 01
2017-02-21	Monmouth NJ	82	@Fairfield	62
2017-02-21	@Belmont	96	Trevecca Naz	45
2017-02-21	@Wichita St	109	Evansville	83
2017-02-21	Purdue	74	@Penn St	70 01
2017-02-21	Rhode Island	67	@La Salle	56
2017-02-21	Auburn	98	@LSU	75
2017-02-21	@Florida	81	South Carolina	66
2017-02-21	@Virginia Tech	71	Clemson	70

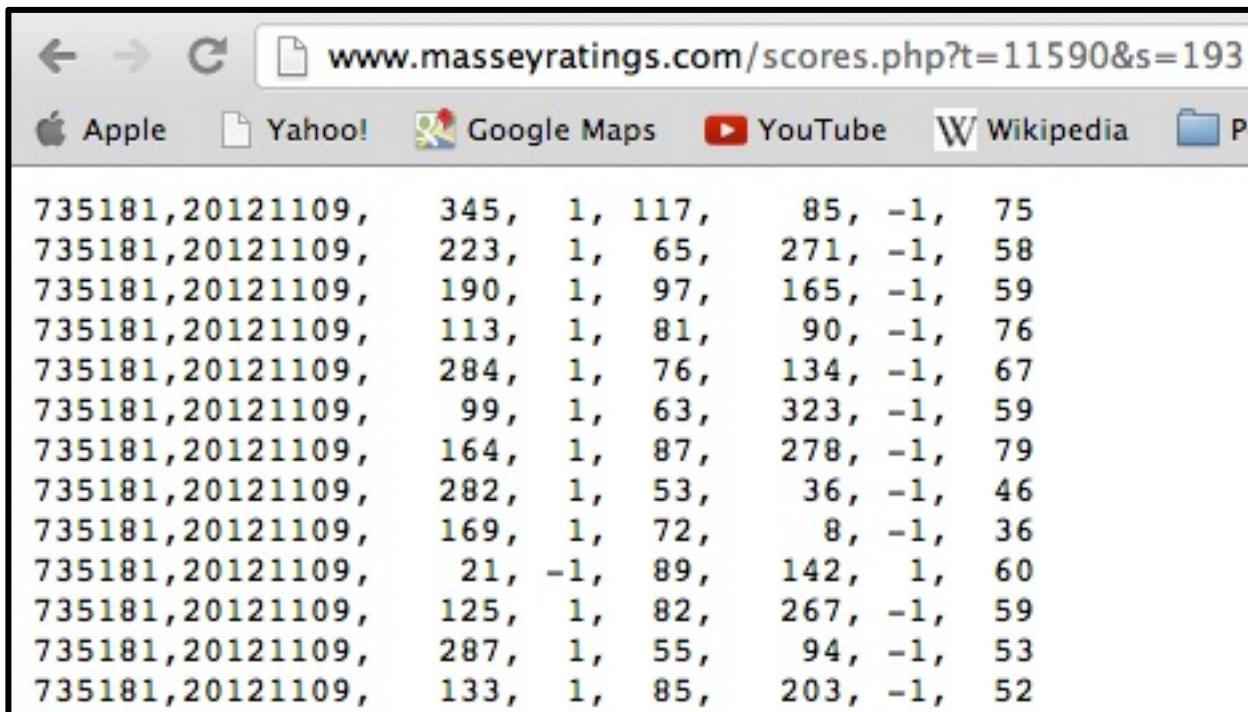
Playing games

Select “Intra” and “Matlab Games” and later “Matlab Teams”..



Having it all

- The data should be saved as a .txt text file.
- Note, you may need to “View Source” to see it in this form.



A screenshot of a web browser window displaying a list of data in a plain text format. The browser's address bar shows the URL www.masseyratings.com/scores.php?t=11590&s=1935. The page content consists of a single column of data, each row containing seven comma-separated values. The data appears to be a list of scores or ratings, likely for a tournament, with the first value being a unique identifier (735181) and the last value being a numerical score (e.g., 75, 58, 59, 76, 67, 59, 79, 46, 36, 60, 59, 53, 53).

735181,20121109,	345,	1,	117,	85,	-1,	75
735181,20121109,	223,	1,	65,	271,	-1,	58
735181,20121109,	190,	1,	97,	165,	-1,	59
735181,20121109,	113,	1,	81,	90,	-1,	76
735181,20121109,	284,	1,	76,	134,	-1,	67
735181,20121109,	99,	1,	63,	323,	-1,	59
735181,20121109,	164,	1,	87,	278,	-1,	79
735181,20121109,	282,	1,	53,	36,	-1,	46
735181,20121109,	169,	1,	72,	8,	-1,	36
735181,20121109,	21,	-1,	89,	142,	1,	60
735181,20121109,	125,	1,	82,	267,	-1,	59
735181,20121109,	287,	1,	55,	94,	-1,	53
735181,20121109,	133,	1,	85,	203,	-1,	52

U Turn

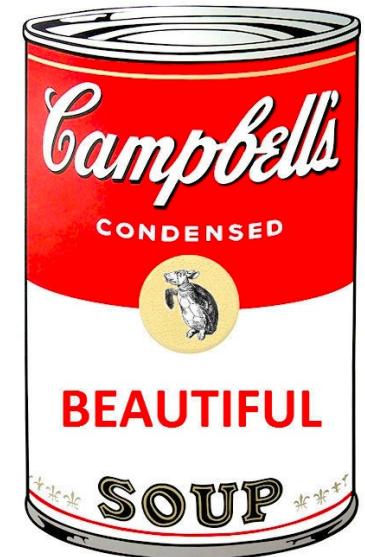
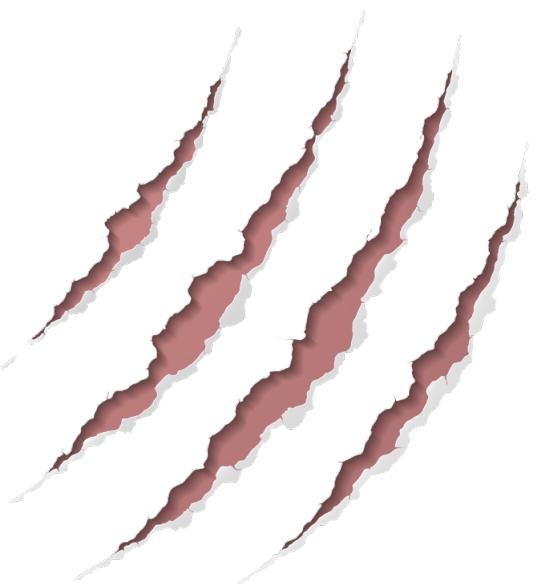
- It's your turn. Go to
- Click the “Data” button.
- And, try it.





Google
Sheets

Part 2: Scrape data



Got data?

- Data - fundamental part of most ranking.
- Let's learn to grab our own so we can create our own datasets.



Easy access

- Some datasets are already available and ready for download.
- Here are 2 sources on government data:
 - <https://www.data.gov/> - home of the U.S. Government's open data
 - The Opportunity Project -
<http://opportunity.census.gov/> - launched to catalyze the creation of new digital tools that use federal and local data in the U.S. to empower communities

Dead end?

- What if you want something else?
- What do you do then?



Method 1



Google
Sheets

Step 1

- Find a web site with a table of data.
- Let's use NFL data at: <http://bit.ly/2dtQrqt>

The screenshot shows the NFL.com homepage with a navigation bar featuring team logos and links for Fantasy, News, Video, Scores, Schedules, Standings, Stats, Watch Games, Tickets, and Shop. Below the navigation is a search bar and a link to NFL Network. The main content area is titled 'STATISTICS' and displays a table of 'Passing' statistics for the 2016 Regular Season. The table includes columns for Rk, Player, Team, Pos, Comp, Att, Pct, Att/G, Yds, Avg, Yds/G, TD, Int, 1st, 1st%, Lng, 20+, 40+, Sck, and Rate. The top seven players listed are Drew Brees, Matt Ryan, Kirk Cousins, Aaron Rodgers, Philip Rivers, Matthew Stafford, and Joe Flacco. To the right of the table is a sidebar titled 'OTHER LINKS' with links to Hall of Fame, Record & Fact Book, and Rule Book. At the bottom right is an advertisement for FedEx and the NFL.

Rk	Player	Team	Pos	Comp	Att	Pct	Att/G	Yds	Avg	Yds/G	TD	Int	1st	1st%	Lng	20+	40+	Sck	Rate
1	Drew Brees	NO	QB	471	673	70.0	42.1	5,208	7.7	325.5	37	15	264	39.2	98T	70	10	27	101.7
2	Matt Ryan	ATL	QB	373	534	69.9	33.4	4,944	9.3	309.0	38	7	238	44.6	76T	69	17	37	117.1
3	Kirk Cousins	WAS	QB	406	606	67.0	37.9	4,917	8.1	307.3	25	12	225	37.1	80T	69	13	23	97.2
4	Aaron Rodgers	GB	QB	401	610	65.7	38.1	4,428	7.3	276.8	40	7	222	36.4	66T	57	10	35	104.2
5	Philip Rivers	SD	QB	349	578	60.4	36.1	4,386	7.6	274.1	33	21	214	37.0	59	57	16	36	87.9
6	Matthew Stafford	DET	QB	388	594	65.3	37.1	4,327	7.3	270.4	24	10	217	36.5	73T	52	10	37	93.3
7	Joe Flacco	BAL	QB	436	672	64.9	42.0	4,317	6.4	269.8	20	15	204	30.4	95T	40	11	33	83.5

Import table

- Open Google Sheets
- Go to the cell where you want the table to be entered.



Step 2

Type

```
=IMPORTHTML(url,"table",1)
```

where the index identifies which table or list as defined in the HTML source should be returned.



NFL QB passing data  

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

   100%  \$  .0  .00  Arial  10  B  I  S  A    

fx | =importhtml("http://bit.ly/2dtQrqt","table",1)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Rk	Player	Team	Pos	Comp	Att	Pct	Att/G	Yds	Avg	Yds/G	TD	Int	1st	1st%	Lng	20+	40+	Sck	Rate
2	1	Drew Brees	NO	QB	471	673	70	42.1	5,208	7.7	325.5	37	15	264	39.2	98T	70	10	27	101.7
3	2	Matt Ryan	ATL	QB	373	534	69.9	33.4	4,944	9.3	309	38	7	238	44.6	76T	69	17	37	117.1
4	3	Kirk Cousins	WAS	QB	406	606	67	37.9	4,917	8.1	307.3	25	12	225	37.1	80T	69	13	23	97.2
5	4	Aaron Rodgers	GB	QB	401	610	65.7	38.1	4,428	7.3	276.8	40	7	222	36.4	66T	57	10	35	104.2
6	5	Philip Rivers	SD	QB	349	578	60.4	36.1	4,386	7.6	274.1	33	21	214	37	59	57	16	36	87.9
7	6	Matthew Stafford	DET	QB	388	594	65.3	37.1	4,327	7.3	270.4	24	10	217	36.5	73T	52	10	37	93.3
8	7	Joe Flacco	BAL	QB	436	672	64.9	42	4,317	6.4	269.8	20	15	204	30.4	95T	40	11	33	83.5
9	8	Andrew Luck	IND	QB	346	545	63.5	36.3	4,240	7.8	282.7	31	13	206	37.8	64T	61	8	41	96.4
10	9	Carson Palmer	ARI	QB	364	597	61	39.8	4,233	7.1	282.2	26	14	218	36.5	80T	48	6	40	87.2
11	10	Russell Wilson	SEA	QB	353	546	64.7	34.1	4,219	7.7	263.7	21	11	193	35.3	59	51	11	41	92.6
12	11	Andy Dalton	CIN	QB	364	563	64.7	35.2	4,206	7.5	262.9	18	8	204	36.2	86T	53	13	41	91.8
13	12	Jameis Winston	TB	QB	345	567	60.8	35.4	4,090	7.2	255.6	28	18	214	37.7	45T	43	4	35	86.1
14	13	Eli Manning	NYG	QB	377	598	63	37.4	4,027	6.7	251.7	26	16	190	31.8	75T	46	11	21	86
15	14	Derek Carr	OAK	QB	357	560	63.8	37.3	3,937	7	262.5	28	6	189	33.8	75T	50	8	16	96.7
16	15	Blake Bortles	JAX	QB	368	625	58.9	39.1	3,905	6.2	244.1	23	16	206	33	51T	43	4	34	78.8
17	16	Sam Bradford	MIN	QB	395	552	71.6	36.8	3,877	7	258.5	20	5	197	35.7	71T	49	8	37	99.3
18	17	Ben Roethlisberger	PIT	QB	328	509	64.4	36.4	3,819	7.5	272.8	29	13	184	36.1	72T	59	8	17	95.4
19	18	Carson Wentz	PHI	QB	379	607	62.4	37.9	3,782	6.2	236.4	16	14	191	31.5	73T	39	6	33	79.3
20	19	Dak Prescott	DAL	QB	311	459	67.8	28.7	3,667	8	229.2	23	4	182	39.7	83T	38	8	25	104.9
21	20	Tom Brady	NE	QB	291	432	67.4	36	3,554	8.2	296.2	28	2	163	37.7	79T	45	8	15	112.2
22	21	Cam Newton	CAR	QB	270	510	52.9	34	3,509	6.9	233.9	19	14	172	33.7	88T	42	8	36	75.8
23	22	Alex Smith	KC	QB	328	489	67.1	32.6	3,502	7.2	233.5	15	8	181	37	80T	39	8	28	91.2

+  NFL QB page 1 ▾

WARNING

- Not everything can be grabbed in this way!
- This can be quite empowering and for some rather addictive!



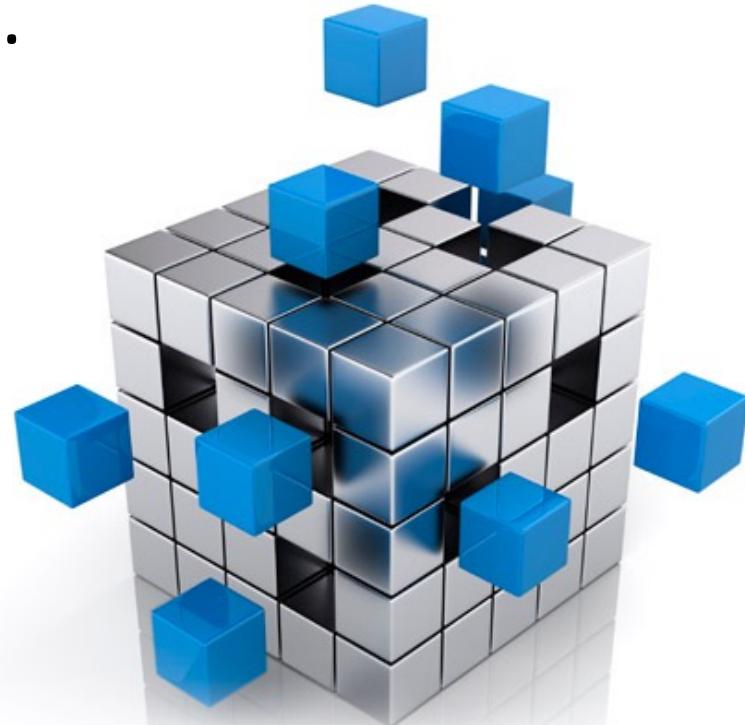
U Turn

- Search some sites and try downloading data.
- Google Sheets works well with tables in Wikipedia, for instance.
- Does this technique grab the data you want?



Enrich

Another way to mine data is take an existing data set and then enrich (or add) data to make a dataset of greater value.



Picture credit: <http://www.decisionpace.com/wp-content/uploads/2014/09/content-img-solutions.jpg>

DATA OF THRONES

W List of Game of Thrones episodes

Secure https://en.wikipedia.org/wiki/List_of_Game_of_Thrones_episodes

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

WIKIPEDIA The Free Encyclopedia

Main page Contents Featured content Current events Random article Donate to Wikipedia Wikipedia store

Interaction Help About Wikipedia Community portal Recent changes Contact page

Tools What links here Related changes Upload file Special pages Permanent link Page information Wikidata item Cite this page

Print/export Create a book Download as PDF

Printable version

List of *Game of Thrones* episodes

From Wikipedia, the free encyclopedia

GAME OF THRONES

Game of Thrones is an American fantasy drama television series created by David Benioff and D. B. Weiss. The series is based on the *A Song of Ice and Fire* novels by author George R. R. Martin. The series takes place on the fictional continents of Westeros and Essos and chronicles the power struggles among noble families as they fight for control of the Iron Throne of the Seven Kingdoms. The series starts when House Stark, led by Lord Eddard "Ned" Stark (Sean Bean) is drawn into schemes against King Robert Baratheon (Mark Addy) when the Hand of the King Jon Arryn (Robert's chief advisor) dies mysteriously.^[1]

The series premiered on April 17, 2011, on HBO. David Benioff and D. B. Weiss both serve as executive producers along with Carolyn Strauss, Frank Doelger, Bernadette Caulfield and George R. R. Martin.^{[2][3]} Filming for the series has taken place in a number of locations, including Croatia, Northern Ireland, Iceland and Spain.^{[4][5][6]} Episodes are broadcast on Sunday at 9:00 pm Eastern Time,^{[7][8]} and the episodes are between 50 and 81 minutes in length.^[9] The first seven seasons are available on DVD and Blu-ray.

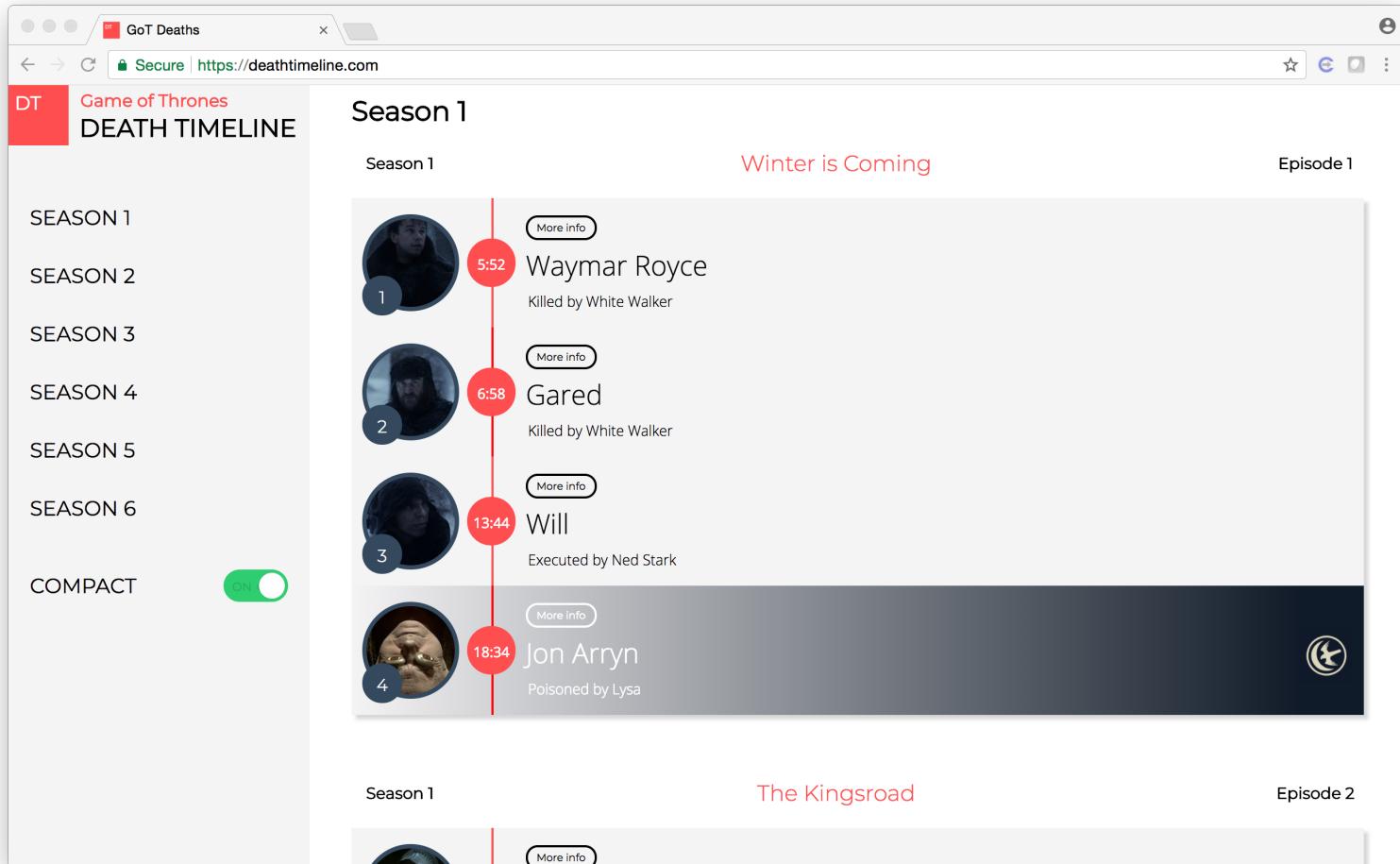
Episodes [edit]

Season 1 (2011) [edit]

Main article: *Game of Thrones (season 1)*

No. overall	No. in season	Title	Directed by	Written by	Original air date	U.S. viewers (millions)
1	1	"Winter Is Coming"	Tim Van Patten	David Benioff & D. B. Weiss	April 17, 2011	2.22 ^[19]
2	2	"The Kingsroad"	Tim Van Patten	David Benioff & D. B. Weiss	April 24, 2011	2.20 ^[20]
3	3	"Lord Snow"	Brian Kirk	David Benioff & D. B. Weiss	May 1, 2011	2.44 ^[21]
4	4	"Cripples, Bastards, and Broken Things"	Brian Kirk	Bryan Cogman	May 8, 2011	2.45 ^[22]
5	5	"The Wolf and the Lion"	Brian Kirk	David Benioff & D. B. Weiss	May 15, 2011	2.56 ^[23]
6	6	"A Golden Crown"	Daniel Minahan	Story by: David Benioff & D. B. Weiss Teleplay by: Ian Brennan and David	May 22, 2011	2.44 ^[24]

DATA OF THRONES



DATA OF THRONES

author ≠ David Benioff ⇒ no major death
& D. B. Weiss



data ⇒ data asset

Combining the right datasets can create a valuable asset in itself

The screenshot shows the Wikipedia article for the "List of Game of Thrones episodes". It includes a brief summary of the show, a table of contents for Season 1, and a detailed table of episodes from Season 1. The table has columns for episode number, overall rank, title, director, writer, original air date, and U.S. viewership.

No.	No. in overall season	Title	Directed by	Written by	Original air date	U.S. viewers (millions)
1	1	"Winter Is Coming"	Tim Van Patten	David Benioff & D. B. Weiss	April 17, 2011	2.22 ^[1]
2	2	"The Kingroad"	Tim Van Patten	David Benioff & D. B. Weiss	April 24, 2011	2.20 ^[2]
3	3	"Lord Snow"	Brian Kirk	David Benioff & D. B. Weiss	May 1, 2011	2.44 ^[21]
4	4	"Cripples, Bastards, and Broken Things"	Brian Kirk	Bryan Cogman	May 8, 2011	2.45 ^[22]
5	5	"The Wolf and the Lion"	Brian Kirk	David Benioff & D. B. Weiss	May 15, 2011	2.58 ^[23]



The screenshot shows a timeline visualization titled "Game of Thrones DEATH TIMELINE". It tracks character deaths across the first seven seasons. A vertical red line represents time, with specific death moments marked by red circles. Character portraits are shown next to their names and the cause of death. The interface includes a sidebar for navigating seasons and a compact mode toggle.

Season 1

- 1 5:52 Waymar Royce Killed by White Walker
- 2 6:58 Gared Killed by White Walker
- 3 13:44 Will Executed by Ned Stark
- 4 18:34 Jon Arryn Poisoned by Lysa

Season 2

Season 3

Season 4

Season 5

Season 6

Season 7

Winter is Coming

The Kingsroad



t r e s a t a

Excel 2 combine

VLOOKUP is a helpful command. (The “False” at the end indicate an exact match.)

Excel 2 combine

Pivot tables are also helpful. Suppose we have data like that on the left. We can tally the data in a pivot table.

	A	B	C	D	E	F
1	Order ID	Product	Category	Amount	Date	Country
2	1	Carrots	Vegetables	\$4,270	1/6/2016	United States
3	2	Broccoli	Vegetables	\$8,239	1/7/2016	United Kingdom
4	3	Banana	Fruit	\$617	1/8/2016	United States
5	4	Banana	Fruit	\$8,384	1/10/2016	Canada
6	5	Beans	Vegetables	\$2,626	1/10/2016	Germany
7	6	Orange	Fruit	\$3,610	1/11/2016	United States
8	7	Broccoli	Vegetables	\$9,062	1/11/2016	Australia
9	8	Banana	Fruit	\$6,906	1/16/2016	New Zealand
10	9	Apple	Fruit	\$2,417	1/16/2016	France
11	10	Apple	Fruit	\$7,421	1/16/2016	Canada

	A	B	C
1	Country	(All)	
2			
3	Row Labels	Sum of Amount	
4	Apple	191257	
5	Banana	340295	
6	Beans	57281	
7	Broccoli	142439	
8	Carrots	136945	
9	Mango	57079	
10	Orange	104438	
11	Grand Total	1029734	
12			

Why not python?

Why use Excel (or Google Sheets)? You can use python but, in some cases, spreadsheets can be quick and useful.



Method 2



Application program interface (**API**) is a set of routines, that among other things, allows you to grab data.

numbers

- Here is a site with a list of public access APIs.
<http://numbersapi.com/>
- We will pull data with the Jupyter notebook:
apiNumbers.ipynb.
- We will access the api with an HTML call

<http://numbersapi.com/42>



apiNumbers.py

```
# Reading the http://numbersapi.com/
import requests
num = 210
url = "http://numbersapi.com/" + repr(num)
response = requests.get(url)
print(response.text)
```

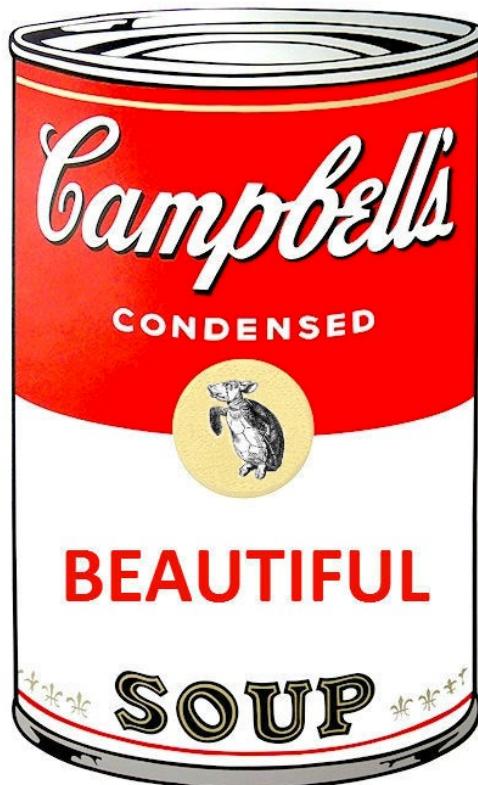
Try it yourself

sunrise, sunset

- Here is a site with a list of public access APIs.
<http://sunrise-sunset.org/api>
- Let's take a look at this site.
- A key is seeing the form of the output, which, in this case, is JSON.
- We will pull data with the Jupyter notebook: apiSunset.ipynb.



Method 3



python™

pip install beautifulsoup4

Grab a table

- You can easily grab a table of data, like we did in Google Sheets.
- For an example, see the notebook:

grabTable.ipynb



python™

Targeted scrape

- You can easily grab a table of data, like we did in Google Sheets.
- For an example, see the notebook:

scrapeNFL.ipynb



python™

Scraping Wikipedia

- This is how wikipediaNetwork.py that you used earlier worked.
- BeautifulSoup enabled us to grab the webpages on the site.
- You may wish to look at that code, although it is more advanced than what we are doing here.



U Turn

Take some time to download the notebooks
and play with what we've learned.



Want more?

The following is optional. You may want to try Step 2 and see what you think.



Step 1: Library access

The first step is to be sure you have the correct libraries installed by typing the following where you code in python:

pip install beautifulsoup4

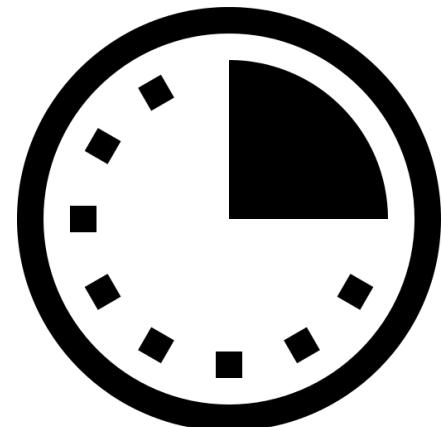
pip install lxml

pip install requests



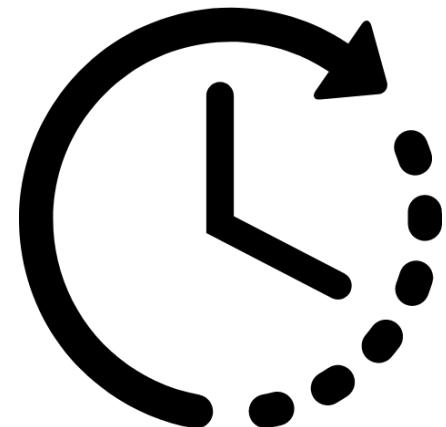
Step 2: 4-minute tutorial

- Next, walk through this tutorial on learning to scrape with python in 4 minutes. That may be generous but at least we can see what multiple of 4 it takes!
- Tutorial: <https://bit.ly/2tVB3ym>



Step 3: U-scrape

- If you have time left or have interest, take a deeper dive by watching this tutorial via YouTube (about 45 minutes in length).
- Tutorial: <https://bit.ly/2thGwPY>



Homework – 1

- Read *Bracketology: How can math help?*
- In 2-3 paragraphs describe what you found most interesting about the paper. Further, do you enjoy the application part of the paper or the theoretical development or both?

Homework – 2

Scape a dataset using python and/or Google Sheets. You may also wish to combine datasets, creating enriched data. In your Word document, explain your dataset to someone who might want to use it. What's there? Further, indicate if you used Google Sheets or python. Finally, include the dataset with your uploaded files to Schoology. (If you download a significant amount of data, just upload a small sample.)

Homework – 3

The draft of your research proposal is due. In it, you will propose a topic for your research, including what application they intend to rank, which ranking methods you will use or adapt, and what data exists or will be created. Be prepared to give a 2-3 minute description of your topic to the group during our next group session. You may do this entirely verbally or accompanied with slides. You will receive feedback on your draft so it can be further developed for your first individual session. This paper should be 2-3 pages long.