SNL Network/Collaboration Visualization
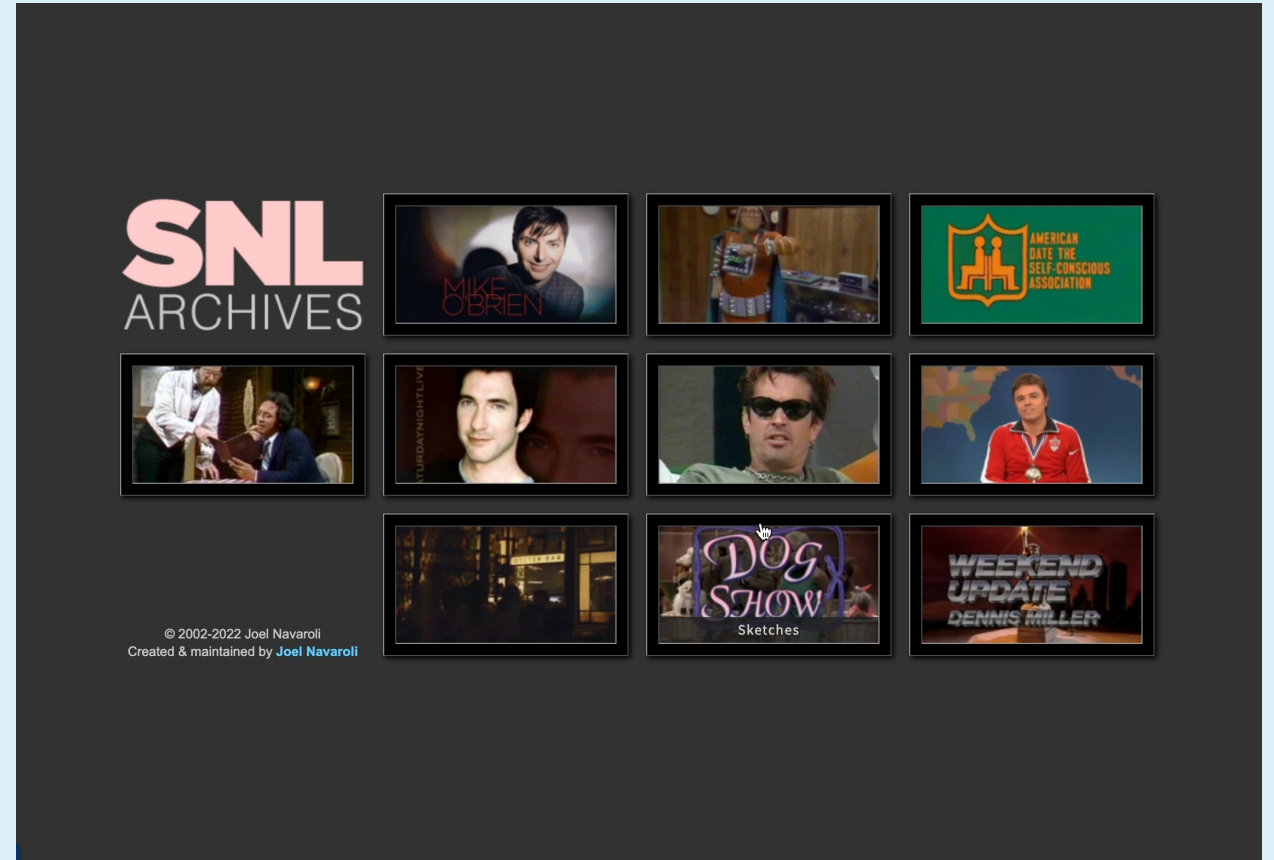
Tali Zacks

Idea Sparked

- I recently rewatched Brooklyn Nine-Nine, a show starring Andy Samberg (2005–2012), and noticed random appearances from a slew of SNL alumni, including:
  - Fred Armisen (2002–2013) : Mlep(clay)nos (the clay is silent)
  - Maya Rudolph (2000–2007) : Karen Haas
  - Vanessa Bayer (2010–2017) : Debbie Fogle
- It made me think about all the collaborations that occur between SNL cast members and how frequently include one another in their projects
  - Adam Sandler (1990–1995) movies featuring Rob Schneider (1990–1994) and David Spade (1990–1996)
  - Amy Poehler (2001–2008) and Tina Fey (1997–2006)
- I began digging into how I might be able to explore these connections

SNL Archives

- © 2002-2022 Joel Navaroli
  Created & maintained by **Joel Navaroli**

- GitHub and Kaggle Datasets created by scraping snlarchives.net
  - Hendrik Hilleckes (@hhllcks, hllcks@gmail.com, blog.hhllcks.de)
  - Colin Morris (http://colinmorris.github.io/)

SNL Archive Dataset

- **actors = pd.read_csv('SNL_data/actors.csv')**
- appearances = pd.read_csv('SNL_data/appearances.csv')
- casts = pd.read_csv('SNL_data/casts.csv')
- characters = pd.read_csv('SNL_data/characters.csv')
- episodes = pd.read_csv('SNL_data/episodes.csv')
- hosts = pd.read_csv('SNL_data/hosts.csv')
- impressions = pd.read_csv('SNL_data/impressions.csv')
- seasons = pd.read_csv('SNL_data/seasons.csv')
- sketches = pd.read_csv('SNL_data/sketches.csv')
- tenure = pd.read_csv('SNL_data/tenure.csv')
- titles = pd.read_csv('SNL_data/titles.csv')

	aid	url	type	gender
0	Kate McKinnon	/Cast/?KaMc	cast	female
1	Alex Moffat	/Cast/?AlMo	cast	male
2	Ego Nwodim	/Cast/?EgNw	cast	unknown
...	...	...	...	...
2304	Janis Ian	/Guests/?2	guest	female
2305	Valri Bromfield	/Guests/?5	guest	unknown

IMDb Datasets

- IMDb (Internet Movie Database) allows any user to download their HUGE, messy datasets and use them under the

- I downloaded:
  - *names.basics.tsv (691.9 MB)*
  - *title.basics.tsv (757.3 MB)*
  - *title.pricipals.tsv (2.2 GB)*

  from https://datasets.imdbws.com/

- Licensing Information: Subsets of IMDb data are available for access to customers for personal and non-commercial use

names.basics.tsv

- nconst (string) - alphanumeric unique identifier of the name/person

- primaryName (string)– name by which the person is most often credited

- birthYear – in YYYY format

- deathYear – in YYYY format if applicable, else '\N'

- primaryProfession (array of strings)– the top-3 professions of the person

- knownForTitles (array of tconsts) – titles the person is known for

names.basics.tsv

	nconst	primaryName	birthYear	deathYear	primaryProfession	knownForTitles
0	nm0000001	Fred Astaire	1899	1987	soundtrack,actor,miscellaneous	tt0050419,tt0053137,tt0031983,tt0072308
1	nm0000002	Lauren Bacall	1924	2014	actress,soundtrack	tt0071877,tt0037382,tt0117057,tt0038355
2	nm0000003	Brigitte Bardot	1934	\N	actress,soundtrack,music_department	tt0054452,tt0056404,tt0057345,tt0049189
...	...	...	...	...	...	...
11554308	nm9993718	Aayush Nair	\N	\N	cinematographer	\N
11554309	nm9993719	Andre Hill	\N	\N	NaN	\N

title.basics.tsv

- tconst (string) - alphanumeric unique identifier of the title
- titleType (string) – the type/format of the title (e.g. movie, short, tvseries, tvepisode, video, etc)
- primaryTitle (string) – the more popular title / the title used by the filmmakers on promotional materials at the point of release
- originalTitle (string) - original title, in the original language
- isAdult (boolean) - 0: non-adult title; 1: adult title
- startYear (YYYY) – represents the release year of a title. In the case of TV Series, it is the series start year
- endYear (YYYY) – TV Series end year. '\N' for all other title types
- runtimeMinutes – primary runtime of the title, in minutes
- genres (string array) – includes up to three genres associated with the title

title.basics.tsv

	tconst	titleType	primaryTitle	originalTitle	isAdult	startYear	endYear	runtimeMinutes	genres
0	tt0000001	short	Carmencita	Carmencita	0	1894	\N	1	Documentary,Short
1	tt0000002	short	Le clown et ses chiens	Le clown et ses chiens	0	1892	\N	5	Animation,Short
...	...	...	...	...	...	...	...	...	...
885 418 3	tt9916856	short	The Wind	The Wind	0	2015	\N	27	Short
885 418 4	tt9916880	tvEpisode	Horrid Henry Knows It All	Horrid Henry Knows It All	0	2014	\N	10	Adventure,Animation,Comedy

title.pricipals.tsv

- tconst (string) - alphanumeric unique identifier of the title
- ordering (integer) – a number to uniquely identify rows for a given titleId
- nconst (string) - alphanumeric unique identifier of the name/person
- category (string) - the category of job that person was in
- job (string) - the specific job title if applicable, else '\N'
- characters (string) - the name of the character played if applicable, else '\N'

title.pricipals.tsv

	tconst	ordering	nconst	category	job	characters
0	tt0000001	1	nm1588970	self	\N	["Self"]
1	tt0000001	2	nm0005690	director	\N	\N
2	tt0000001	3	nm0374658	cinematographer	director of photography	\N
...	...	...	...	...	...	...
49901290	tt9916880	4	nm10535738	actress	\N	["Horrid Henry"]
49901291	tt9916880	5	nm0996406	director	principal director	\N
49901292	tt9916880	6	nm1482639	writer	\N	\N
49901293	tt9916880	7	nm2586970	writer	books	\N
49901294	tt9916880	8	nm1594058	producer	producer	\N

Full Cast WikiTable

Table of *Saturday Night Live* cast members

Performer	Time on *SNL*	No. of seasons	Repertory Player	Featured Player	Middle Group	"Weekend Update" Anchor	Hosted	Best of...	Writer
Fred Armisen	2002–2013	11	✓	✓			✓		
Aristotle Athari	2021–present	1		✓					
Dan Aykroyd	1975–1979	4	✓			✓	✓	✓	✓
Peter Aykroyd	1980	1		✓					✓
Morwenna Banks	1995	1	✓						
Vanessa Bayer	2010–2017	7	✓	✓					
Jim Belushi	1983–1985	2	✓						✓
John Belushi	1975–1979	4	✓					✓	✓
Beck Bennett	2013–2021	8	✓	✓					
Jim Breuer	1995–1998	3	✓						
Paul Brittain	2010–2012	2		✓					
A. Whitney Brown	1986–1991	6		✓					✓
Aidy Bryant	2012–present	10	✓	✓					
Beth Cahill	1991–1992	1		✓					
Dana Carvey	1986–1993	7	✓					✓	✓
Chevy Chase	1975–1976	2	✓			✓	✓	✓	✓
Michael Che	2014–present	8	✓	✓		✓			✓
Ellen Cleghorne	1991–1995	4	✓		✓				
George Coe	1975	1	✓						
Billy Crystal	1984–1985	1	✓			✓		✓	✓
Jane Curtin	1975–1980	5	✓			✓			
Joan Cusack	1985–1986	1	✓						
Pete Davidson	2014–present	8	✓	✓					
Tom Davis	1977–1980	3		✓					✓

- I then came across a Wikipedia page with a table featuring
  - *Every SNL cast member (all 159)*
  - *Their time on SNL*
  - *Number of seasons*
  - *And some columns with checkmarks denoting various roles on the show*
- I downloaded the table using an online WikiTable to csv converter
  - *Didn't keep the checkmark columns intact*
  - *I copied the table contents into Excel and used Find & Replace to change every checkmark into a "Yes"*

```
In [8]: wiki_cast_excel = pd.read_csv('wiki_cast_excel.csv')
```

```
In [9]: wiki_cast_excel.iloc[29]
```

```
Out[9]: Performer                      Brian Doyle-Murray
        Time on SNL                     1980;1981—1982
        No. of seasons                               2
        Repertory Player                           Yes
        Featured Player                            Yes
        Middle Group                               NaN
        "Weekend Update" Anchor                    Yes
        Hosted                                     NaN
        Writer                                     Yes
        Name: 29, dtype: object
```

```
In [10]: wiki_cast = wiki_cast_excel.fillna('No')
         wiki_cast
```

Out[10]:

	Performer	Time on SNL	No. of seasons	Repertory Player	Featured Player	Middle Group	"Weekend Update" Anchor	Hosted	Writer
0	Fred Armisen	2002–2013	11	Yes	Yes	No	No	Yes	No
1	Aristotle Athari	2021–present	1	No	Yes	No	No	No	No
2	Dan Aykroyd	1975–1979	4	Yes	No	No	Yes	Yes	Yes
3	Peter Aykroyd	1980	1	No	Yes	No	No	No	Yes
4	Morwenna Banks	1995	1	Yes	No	No	No	No	No
...	...	...	...	...	...	...	...	...	...
154	Casey Wilson	2008–2009	2	No	Yes	No	No	No	No
155	Fred Wolf	1996–1996	2	No	Yes	No	No	No	Yes
156	Bowen Yang	2019–present	3	Yes	Yes	No	No	No	Yes
157	Sasheer Zamata	2014–2017	4	Yes	Yes	No	No	No	No
158	Alan Zweibel	1980	1	No	Yes	No	No	No	Yes

159 rows × 9 columns

Combining DataFrames

- In order to combine the IMDb and Wikipedia DataFrames, I needed to set the indices to the actors' names in both