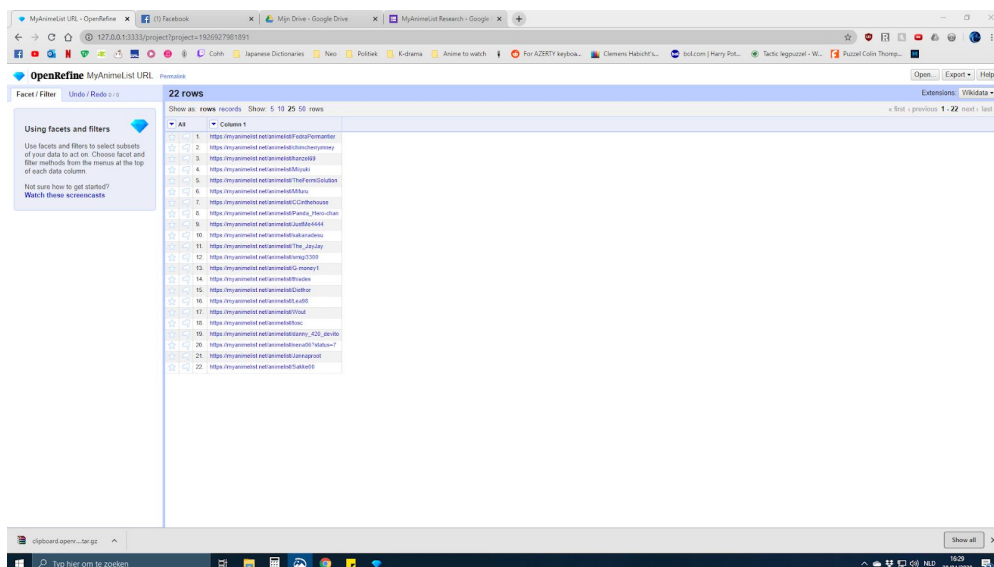
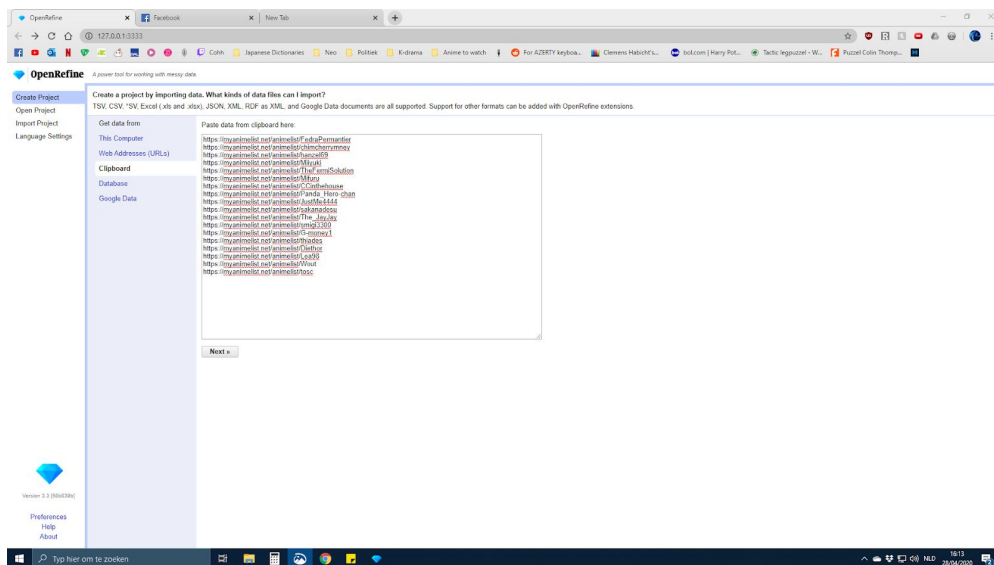


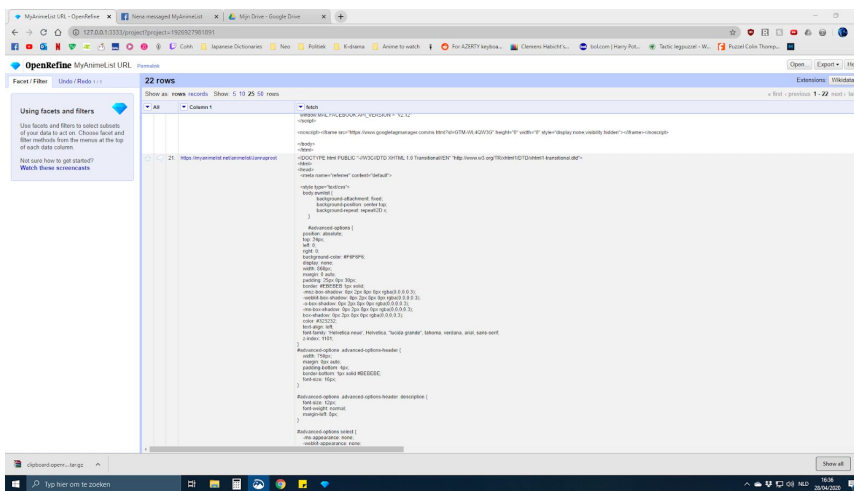
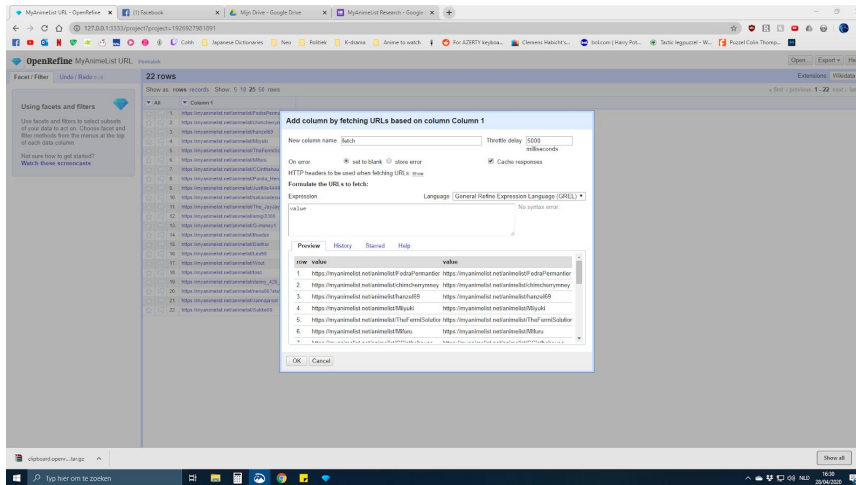
Undo / Redo 1649 / 1649

You might be wondering why our group ended up performing 1649 commands on OpenRefine. Well, the first reason would be our process of 'trial & error', as we slowly tried to figure out the correct formulas and commands to transform a list of URLs into a block of incomprehensible text and finally into neatly organized columns of information.

The second reason is because of the mistakes we had to clean up in the organized data. For example, we had to change a large amount of strange symbols which had found their way into the titles of several anime. An example would be the anime "Lucky☆Star", where OpenRefine was not able to reproduce the star symbol correctly, and it ended up with "Luckyã†Star". There were over 50 anime titles we ended up changing manually. After hours (9, to be specific) we finally had a clean, organized collection of anime titles, viewer status and scores, etc., which allowed us to create graphs on Tableau Public for easy access to the viewer.



We started off by copying each URL to the Clipboard to create our project, ending up with one column of 22 rows, each containing the link to a MAL user's Anime List.



The next step was to fetch the URLs, which we simply did by creating a new column called "fetch" using the 'fetch' function OpenRefine provides. This gave us a very extensive list of information per URL, including (for us) unnecessary information like CSS.

Add column based on column fetch

New column name:

On error: set to blank store error copy value from original column

Expression: `value.parseHtml().select("table")` Language: General Refine Expression Language (GREL) ▾

No syntax error.

Preview History Starred Help

row	value	value.parseHtml().select("tabl ...
1.	<DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd"> <html> <head> <meta name="referrer" content="default"> <style type="text/css"> body,owlstat { background:linear	[<table class="list-table" data-items=" [{""status":1,"score":0,"File","anime_num_episodes":50,i s=6b4bdf71060d554a5004ec639750678",& 08-18","anime_end_date_string":& q 31-19","days_string":null,"stor "status":1,"score":9,"Shinnurden","anime_num_episodes&

OK Cancel

Facet / Filter Undo / Redo 2/2 **22 rows** Extensions: Wikidata ▾

Show as: rows records Show: 5 10 25 50 rows

Filter:

0 Create project

1 Create column fetch at index 1 by fetching URLs based on column Column 1 using expression grel.value

2 Create new column parse based on column fetch by filling 22 rows with `grel.value.parseHtml().select("table").join("")`

<table class="list-table" data-items="
 [{""status":1,"score":0,"tags","is_rewatching":0,"num_watched_episodes&
 File","anime_num_episodes":50,"anime_airing_status":2,"anime_id":37023,"anime_studios","anime_licensors&
 s=6b4bdf71060d554a5004ec639750678","is_added_to_list":false,"anime_media_type_string":"TV","anime_ppaa_rating_string
 Shippuden","anime_num_episodes":500,"anime_airing_status":2,"anime_id":1735,"anime_studios","anime_licensors&
 s=13","anime_start_date_string":null,"anime_finish_date_string":null,"anime_start_date_string":"02-15-07","anime_end_date_string
 Piece","anime_num_episodes":0,"anime_airing_status":1,"anime_id":21,"anime_studios","anime_licensors&
 s=13","anime_start_date_string":null,"anime_finish_date_string":null,"anime_start_date_string":"10-20-98","anime_end_date_string
 (,"status":2,"score":0,"tags","is_rewatching":0,"num_watched_episodes":12,"anime_title":&
 s=c974ba83e2a0b6f6c2516eb2e5f7","is_added_to_list":false,"anime_media_type_string":"TV","anime_ppaa_rating_string
 Kyoushitsu","anime_num_episodes":22,"anime_airing_status":2,"anime_id":24833,"anime_studios","anime_licensors&
 s=13","anime_start_date_string":null,"anime_finish_date_string":null,"anime_start_date_string":"01-10-15","anime_end_date_string
 Exorcist","anime_num_episodes":25,"anime_airing_status":2,"anime_id":1919,"anime_studios","anime_licensors&
 s=13","anime_start_date_string":null,"anime_finish_date_string":null,"anime_start_date_string":"04-17-11","anime_end_date_string
 hen","anime_num_episodes":12,"anime_airing_status":2,"anime_id":33506,"anime_studios","anime_licensors&
 s=13","anime_start_date_string":null,"anime_finish_date_string":null,"anime_start_date_string":"01-07-17","anime_end_date_string
 (,"status":2,"score":9,"tags","is_rewatching":0,"num_watched_episodes":12,"anime_title":&
 s=680bd219893455198372cabee51b7","is_added_to_list":false,"anime_media_type_string":"TV","anime_ppaa_rating_string
 Noboy","anime_num_episodes":1,"anime_airing_status":2,"anime_id":1686,"anime_studios","anime_licensors&
 s=atfb29297ee680bc7bcd9d9d247","is_added_to_list":false,"anime_media_type_string":"Movie","anime_ppaa_rating_string
 Machi","anime_num_episodes":12,"anime_airing_status":2,"anime_id":31043,"anime_studios","anime_licensors&
 s=ab4f46017309045762e44e32d8","is_added_to_list":false,"anime_media_type_string":"TV","anime_ppaa_rating_string
 Jukyou","anime_num_episodes":4,"anime_airing_status":2,"anime_id":15037,"anime_studios","anime_licensors&
 s=47e0ae65e12a5f18415c71cab405599","is_added_to_list":false,"anime_media_type_string":"OVA","anime_ppaa_rating_string
 Imasu","anime_num_episodes":13,"anime_airing_status":2,"anime_id":37587,"anime_studios","anime_licensors&
 s=23742580480ab005a6827524845","is_added_to_list":false,"anime_media_type_string":"TV","anime_ppaa_rating_string
 hen","anime_num_episodes":1,"anime_airing_status":2,"anime_id":34103,"anime_studios","anime_licensors&
 s=12c6e9a805bae271bf0c61b7082cb1","is_added_to_list":false,"anime_media_type_string":"Special","anime_ppaa_rating_string
 hen","anime_num_episodes":12,"anime_airing_status":2,"anime_id":32189,"anime_studios","anime_licensors&
 s=cdff050a4178537471f65a26d17","is_added_to_list":false,"anime_media_type_string":"TV","anime_ppaa_rating_string
 hen","anime_num_episodes":11,"anime_airing_status":2,"anime_id":33028,"anime_studios","anime_licensors&
 s=71dab47bc8a2d4760da9d8e11a2954","is_added_to_list":false,"anime_media_type_string":"TV","anime_ppaa_rating_string
 Animation","anime_num_episodes":13,"anime_airing_status":2,"anime_id":16592,"anime_studios","anime_licensors&
 s=8a04073466778ac0e95d774175970","is_added_to_list":false,"anime_media_type_string":"TV","anime_ppaa_rating_string
 Wonderland","anime_num_episodes":12,"anime_airing_status":2,"anime_id":6880,"anime_studios","anime_licensors&
 s=

127.0.0.1:3333/project?project=2376613817894

OpenRefine MAL Permalk Open... Export Help

Facet / Filter Undo / Redo 2/3 **25 rows** Extensions: Wikidata ▾

Show as: rows records Show: 5 10 25 50 rows

Using facets and filter your data to act filter methods from of each data column

Not sure how to get Watch these ops

Custom text transform on column Parse

Expression: `value.unescape('html')` Language: General Refine Expression Language (GREL) ▾

No syntax error.

Preview History Starred Help

0","days_string":null,"storage_string":","priority_string":"Low",
 05-10","anime_end_date_string":"06-21-
 10","anime_airing_status":6,"tags","is_rewatching":0,"num_watched_episodes":1,"anime_title":
 (,"status":2,"score":6,"tags","is_rewatching":0,"num_watched_episodes":1,"anime_title":
 s=1406d63ee993ae2791324ef5224889","is_added_to_list":false,"anime_media_type_string":"TV","anime_ppaa_rating_string
 13","anime_start_date_string":null,"anime_finish_date_string":null,"anime_start_date_string":"05-31-
 09-16","anime_end_date_string":"06-25-
 16","anime_airing_status":2,"anime_id":13,"days_string":null,"storage_string":null,"priority_string":"Low",
 (,"status":2,"score":7,"tags":null,"is_rewatching":0,"num_watched_episodes":1,"anime_title":
 Note","anime_num_episodes":37,"anime_airing_status":2,"anime_id":1535,"anime_studios","anime_licensors&
 s=af1b3e737c52d27e20808072ac","is_added_to_list":false,"anime_media_type_string":"TV","anime_ppaa_rating_string":"R",&
 Paradise","anime_num_episodes":12,"anime_airing_status":2,"anime_id":29233,"anime_studios","anime_licensors&
 s=6e87c74d77d80142219d2ae58a33","is_added_to_list":false,"anime_media_type_string":"TV","anime_ppaa_rating_string":"R",&
 Lovest","anime_num_episodes":12,"anime_airing_status":2,"anime_id":17513,"anime_studios","anime_licensors&
 s=af3034c747c41c0c10b6c72","is_added_to_list":false,"anime_media_type_string":"TV","anime_ppaa_rating_string":"R",&
 Tail","anime_num_episodes":175,"anime_airing_status":2,"anime_id":6702,"anime_studios","anime_licensors&
 s=13","anime_start_date_string":null,"anime_finish_date_string":null,"anime_start_date_string":"10-09-98","anime_end_date_string":"03-30-19&
 (2014)","anime_num_episodes":102,"anime_airing_status":2,"anime_id":22043,"anime_studios","anime_licensors&
 s=

On error: keep original set to blank store error Re-transform up to 10 times until no change

OK Cancel

We managed to remove all useless information by parsing the necessary information with the RegEx `value.parseHtml().select("table")`, resulting in this massive block of text containing the information we would then split up in several columns. To clean up the ‘"’, we used the RegEx `value.unescape('html')`.

Split column Parse 1 into several columns

How to Split Column

by separator
 Separator regular expression
 Split into columns at most (leave blank for no limit)

by field lengths

List of integers separated by commas, e.g., 5, 7, 15

After Splitting

Guess cell type
 Remove this column

The screenshot shows the OpenRefine MAL interface with 21 records displayed in a table. The table has the following columns: Title, Num_Episodes, Airing Status, Anime-ID, Media Type, Mpaas Rating, Status, and Score. The records include titles like IDOLISH?, Kaguya-sama wa Kokurasetai: Tensai-tachi no Renai Zunousen, Kimi ni Todoke, and many others.

Title	Num_Episodes	Airing Status	Anime-ID	Media Type	Mpaas Rating	Status	Score
IDOLISH?	17	False	33899	TV	PG-13	Plan to Watch	0
Kaguya-sama wa Kokurasetai: Tensai-tachi no Renai Zunousen	12	False	37999	TV	PG-13	Plan to Watch	0
Kimetsu no Yaiba Movie: Mugen Ressha-hen	1	Not yet aired	40456	Movie	R	Plan to Watch	0
Kimi ni Todoke	25	False	6045	TV	PG-13	Plan to Watch	0
Kobayashi-san Chi no Maid Dragon 2nd Season	0	Not yet aired	39247	TV	PG-13	Plan to Watch	0
Kono Subarashi Sekai ni Shukufuku wo!	10	False	30831	TV	PG-13	Plan to Watch	0
Lord El-Melloi II Sei no Jikembo: Rail Zeppelin Grace Note - Hakamori to Neko to Majutsushi	1	False	38936	Special	PG-13	Plan to Watch	0
Mahou Shoujo Ore	12	False	36002	TV	PG-13	Plan to Watch	0
NHK ni Youkoso!	24	False	1210	TV	R+	Plan to Watch	0
Persona 5 the Animation	26	False	36023	TV	R	Plan to Watch	0
Plastic Memories	13	False	27775	TV	PG-13	Plan to Watch	0
Re:Zero kara Hajimeru Isekai Seikatsu 2nd Season	0	Not yet aired	30587	TV	R	Plan to Watch	0
ReLIFE	13	False	30015	TV	PG-13	Plan to Watch	0
Sakamichi no Apollon	12	False	12531	TV	PG-13	Plan to Watch	0
Shingeki! Kyojin Chuugakkou	12	False	31374	TV	PG-13	Plan to Watch	0
Tengen Toppa Gurren Lagann	27	False	2001	TV	PG-13	Plan to Watch	0
Tsurune: Kazemai Koukou Kyuusoubu	13	False	36053	TV	PG-13	Plan to Watch	0
Yakusoku no Neverland 2nd Season	0	Not yet aired	39617	TV	R	Plan to Watch	0

Finally, we created new columns by splitting the column containing the block of text into several columns with the separator ‘,’. At last, we had a clear and structured table which was very easy to work with. After a lot of cleaning up, we were certain that all the information present was correct and legible and ready to create graphs with.