Metadata mining of large collections of historical newspapers

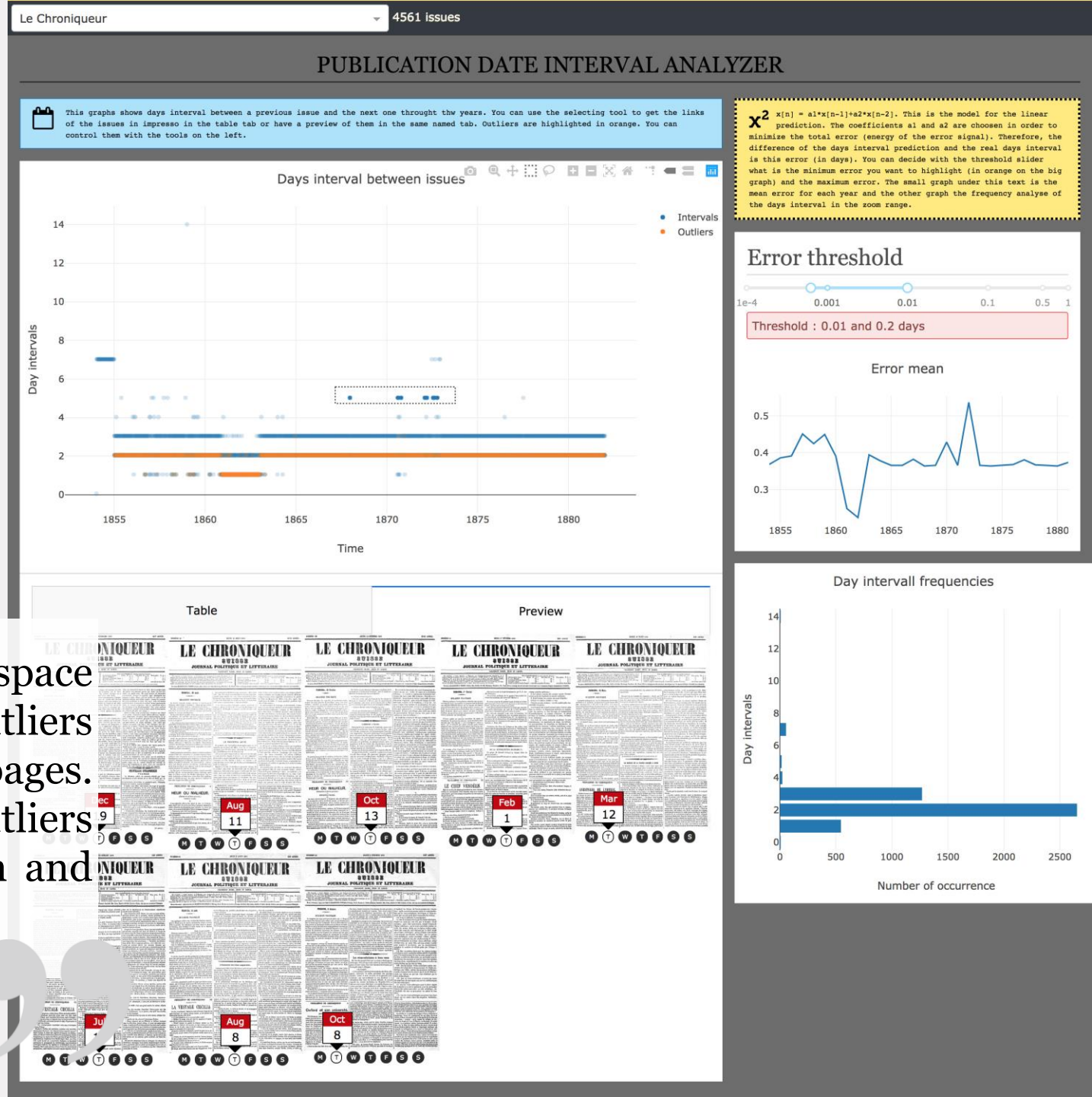
---

Supervisors : Maud Ehrmann and Matteo Romanello

Bachelor Project from Christian Gasser - student in Electrical and Electronics Engineering

« For my semester project, I created a tool that allows to analyse variation in publication date interval. It can detect outliers and shows the stability of the journal through the years. »

The tool is composed of a main space for view the intervals, see the outliers and show the selected front-pages. The lateral view contains the outliers controller, the mean error graph and the frequency analyzer.



1

1-1 TOOL

2

2-1

2-2

2-3

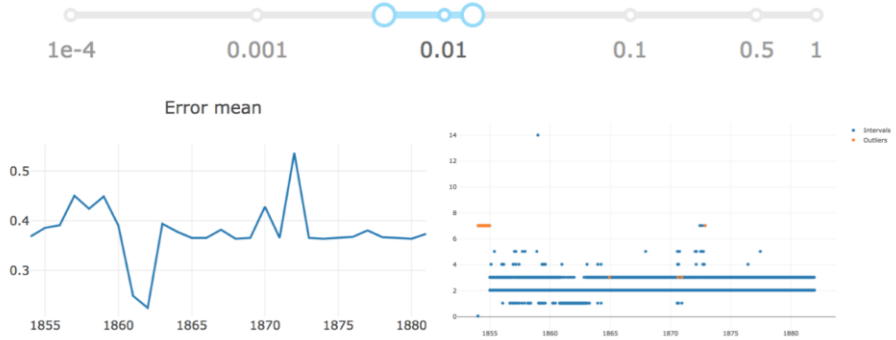
2-4

3

> For my error mean graph and the outlier tool I process the difference between the mean value of the two last intervals and the actual interval. I call that the error. Then I can plot the mean error and choose what point I want to hightlight with a minimum and maximum limit.

$$m_n = \frac{1}{2}x_{n-1} + \frac{1}{2}x_{n-2}$$

$$|e_n| = |m_n - x_n|$$



> In my programm, rather to choose the coefficient a half for each value I put two coefficients that are computet to minimize the error in order to not detect each Sunday for example...

1. Mean error analysis
2. Publication interval : analysis example
3. History through thumbnail
4. Outliers detection

MENU

1

1-1

2

2-1 MEAN ERROR

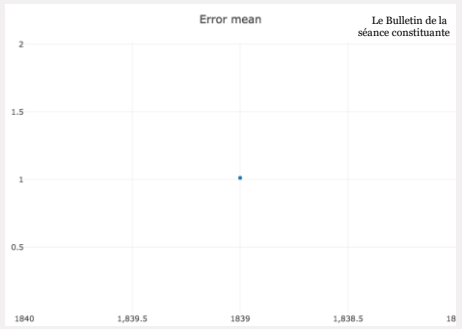
2-2

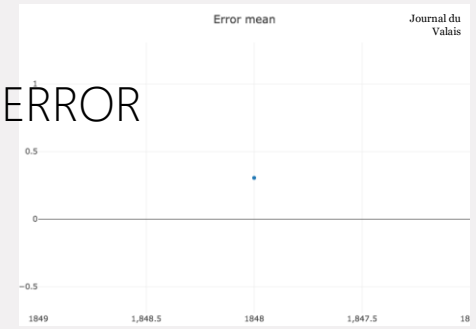
2-3

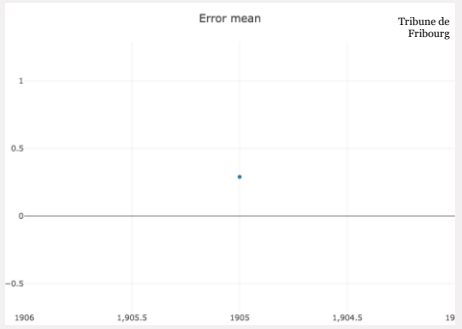
2-4

3

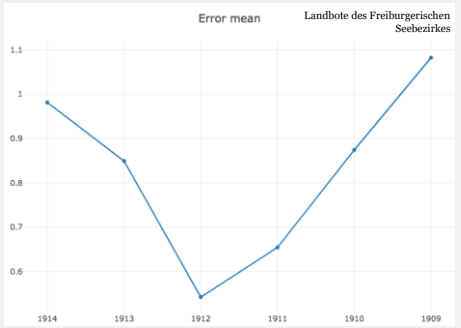
Unique point

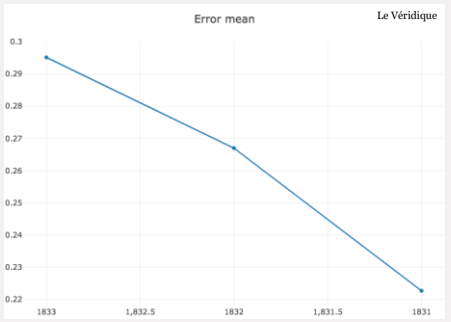


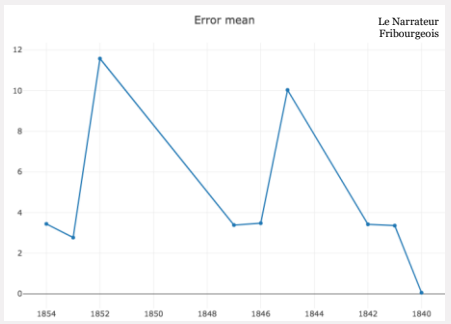




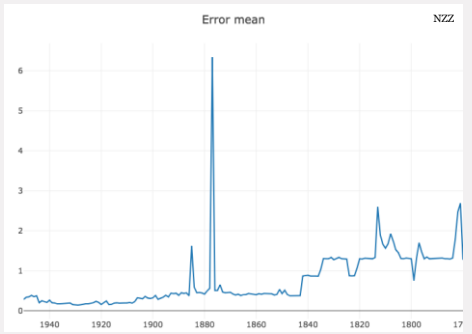
Less than 1000 publications

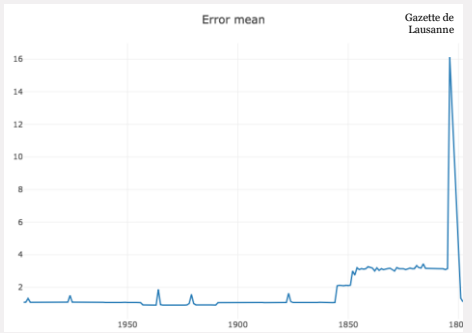




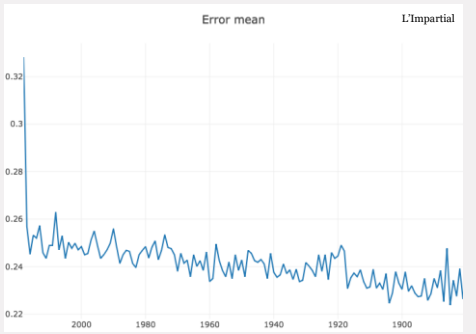


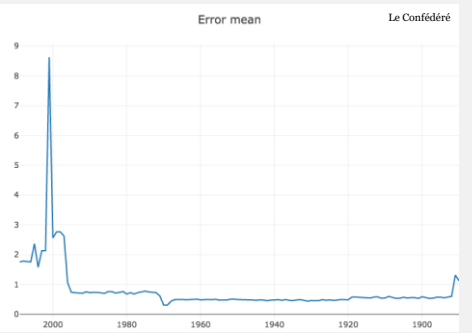
Convergent graphs





Divergent graphs

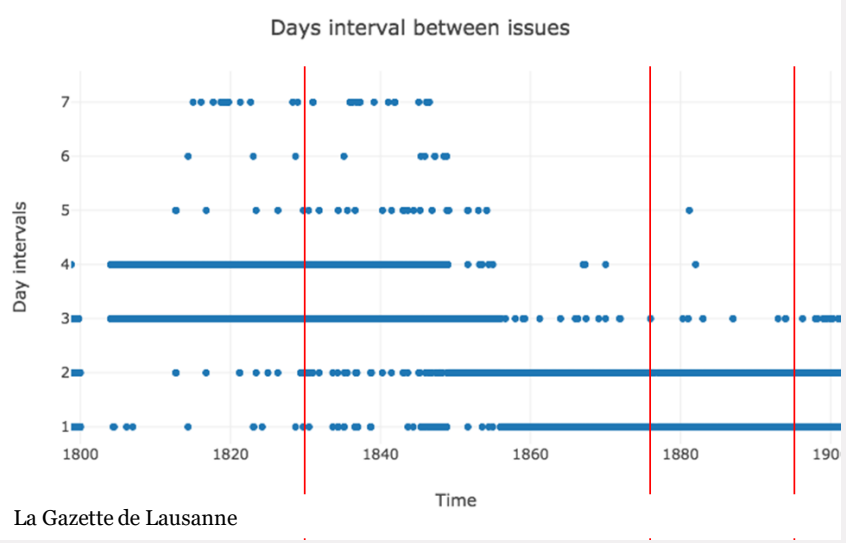




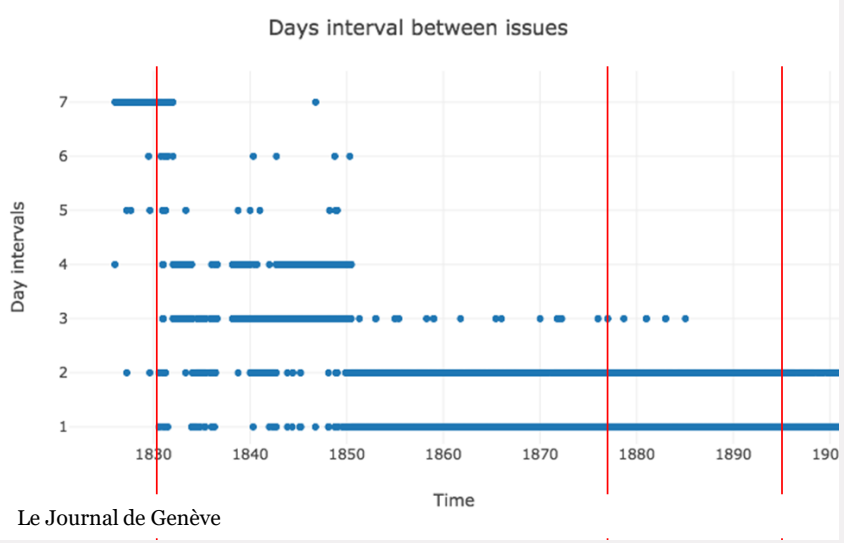
Time [years]

Année	Événement
1830	Invention du télégraphe
1876	Invention du téléphone
1895	Création de l'agence télégraphique suisse, ATS/SDA

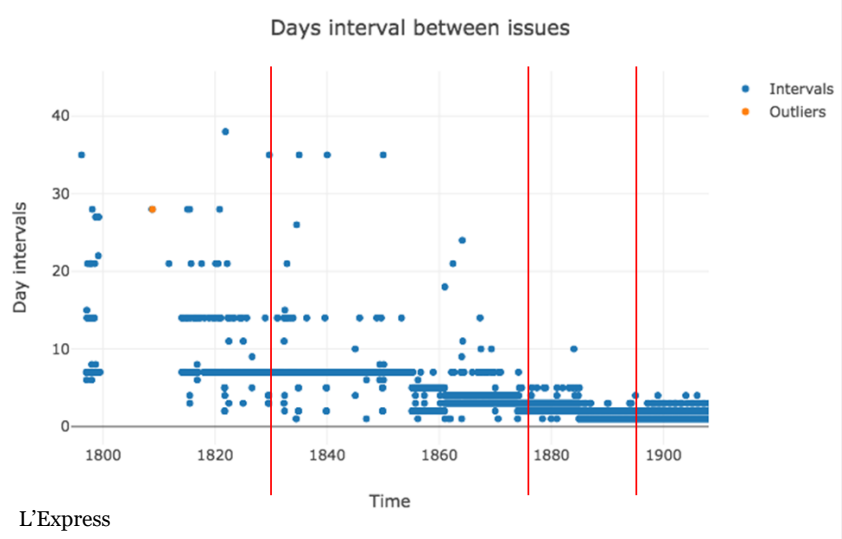
Journal	$\Delta t_{telegraph}$	Δt_{phone}	Δt_{ats}	$t_{telegraph}$	t_{phone}	t_{ats}
GDL	19	34	15	1849	1910	1910
JDG	2	22	3	1832	1898	1898
EXP	25	8	-	1855	1884	-
NZZ	5	5	-	1835	1881	-



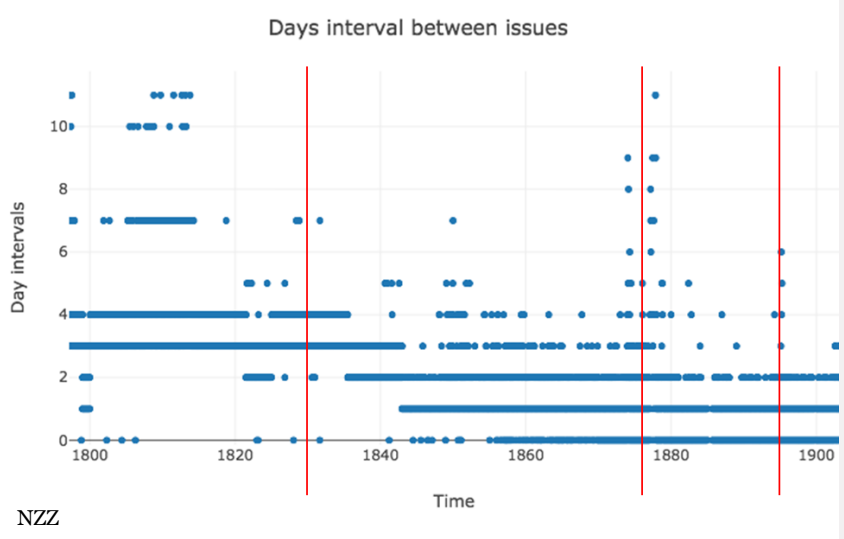
La Gazette de Lausanne



Le Journal de Genève

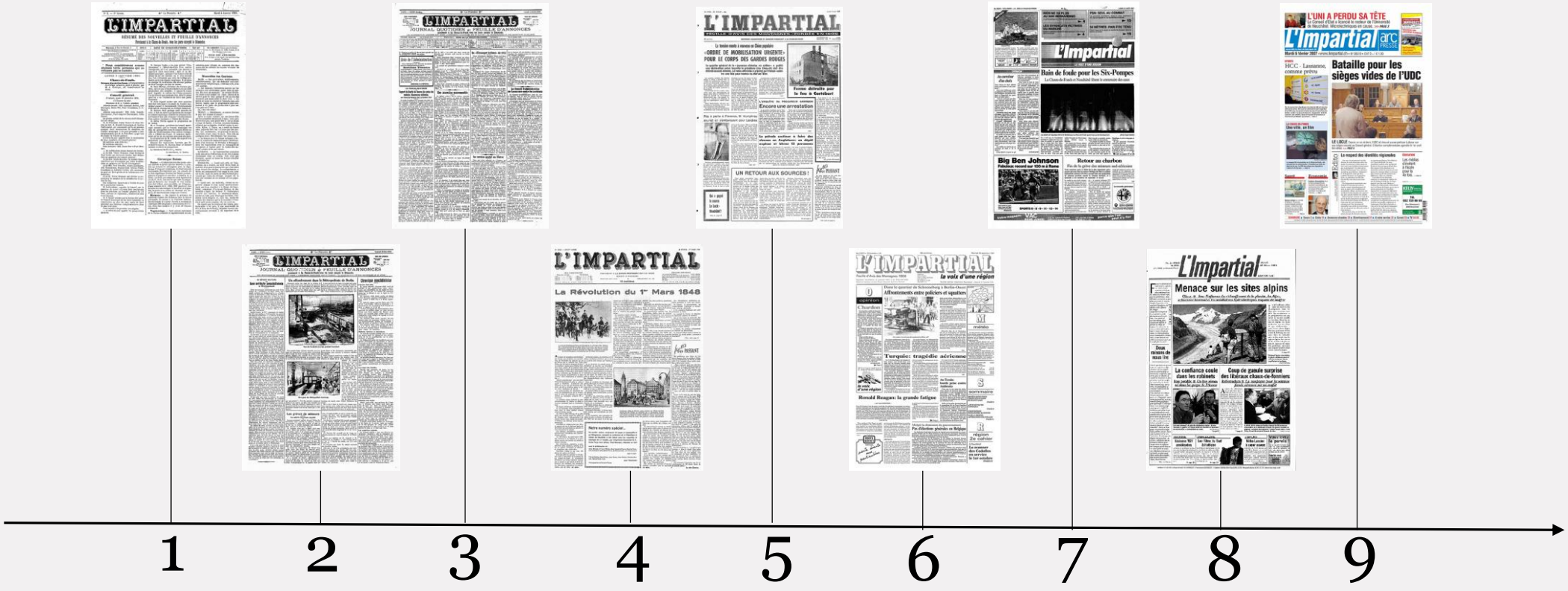


L'Express

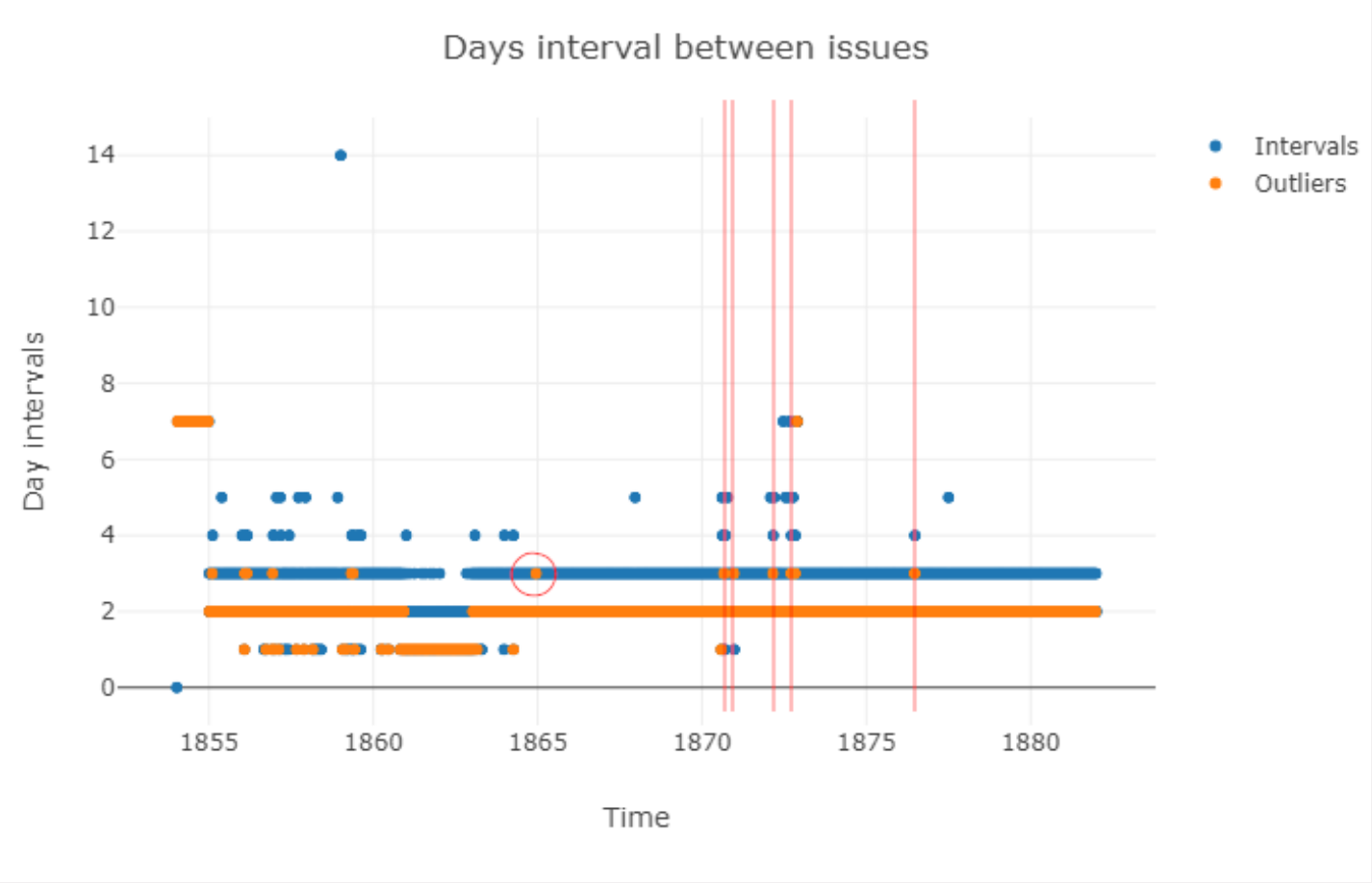


NZZ

1
1-1
2
2-1
2-2
2-3 THUMBNAIL
2-4
3



1	4/1/1881	First issue
2	29/3/1912	Publication of pictures are frequent
3	4/10/1912	Last time we see the train timetable under the title
4	28/2/1948	New design for the century of Neuchâtel
5	3/4/1967	New design after fusion with «Feuille d'Avis des Montagnes»
6	23/9/1981	New formula
7	31/8/1987	First time color
8	20/3/1987	New design to distinguish «L'Impartial» and «L'Express»
9	6/2/2007	Creation of the branding «Arc Presse»



> Projection of the points on horizontal lines are not interesting. But single points yes!

1

1-1

2

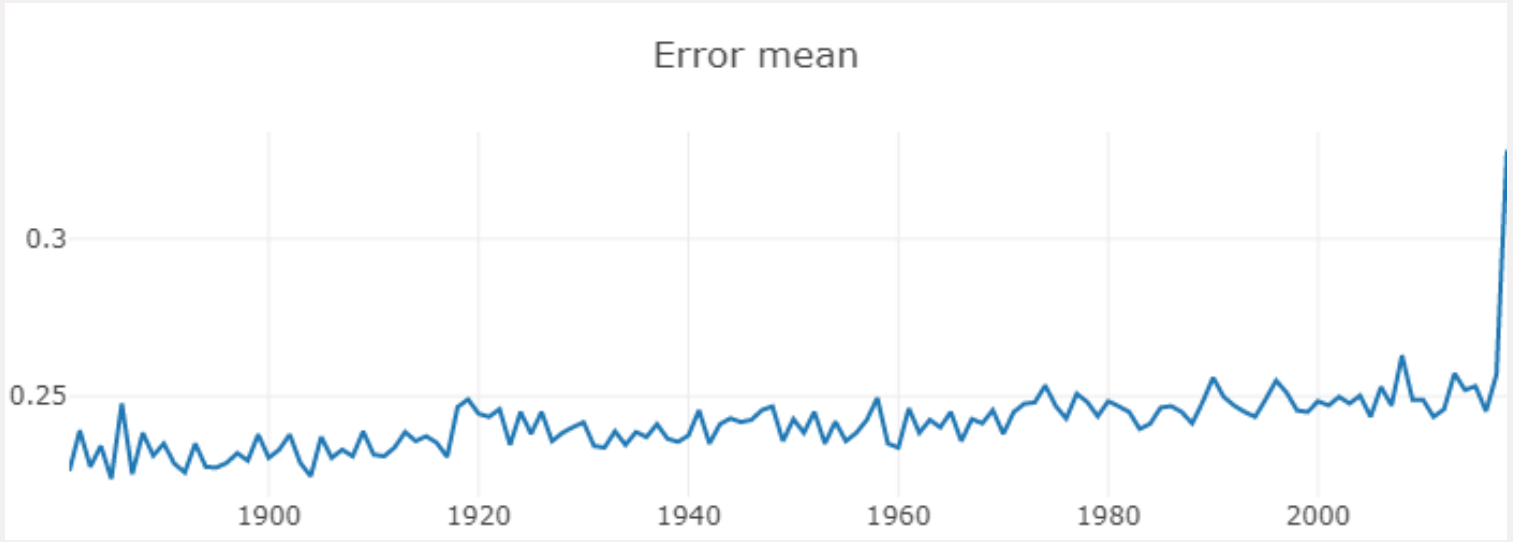
2-1

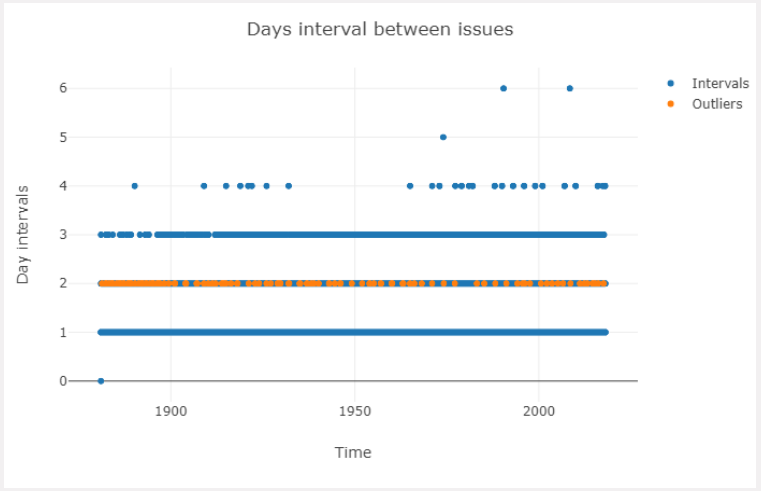
2-2

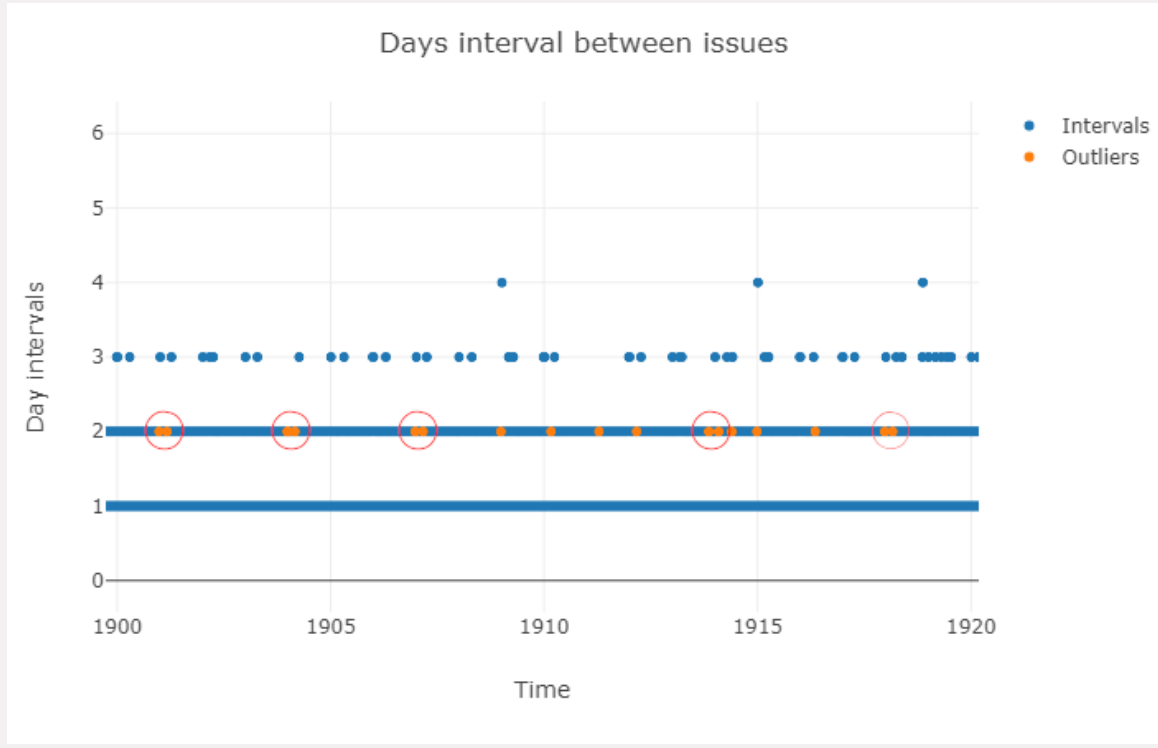
2-3

2-4 OUTLIERS

3







1

1-1

2

2-1

2-2

2-3

2-4 OUTLIERS

3

LES JOURNAUX REUNIS
des Montagnes Neuchâteloises
L'IMPARTIAL - L'EFFORT - LA FEUILLE D'AVIS DES MONTAGNES

A nos lecteurs

Notes d'un Passant

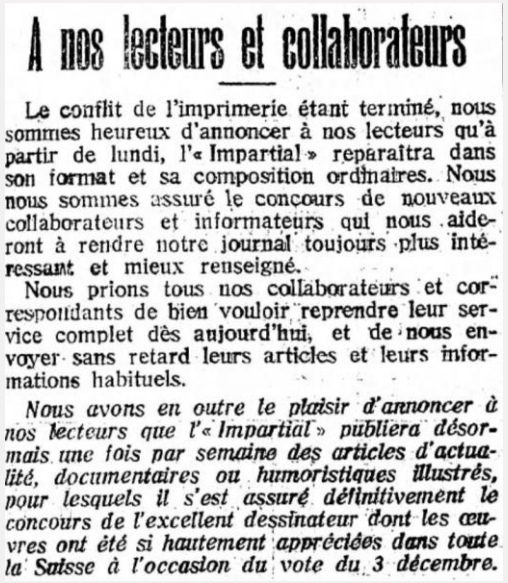
La leçon d'arithmétique

A propos des tâches sociales

Nous aussi, nous voulons l'assurance-vieillesse

A travers l'actualité

La tragédie grecque

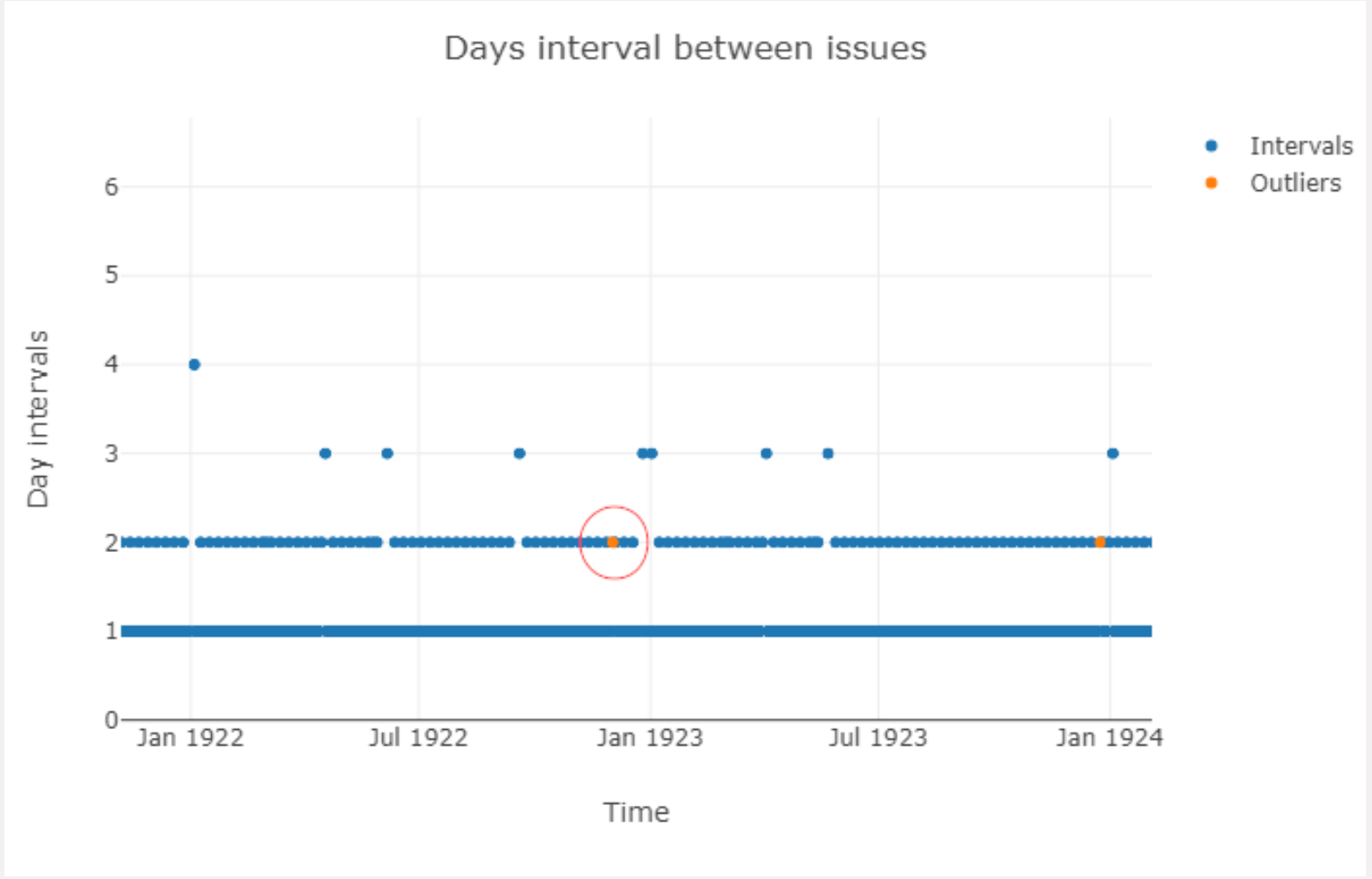


A nos lecteurs et collaborateurs

Le conflit de l'imprimerie étant terminé, nous sommes heureux d'annoncer à nos lecteurs qu'à partir de lundi, l'« Impartial » reparaîtra dans son format et sa composition ordinaires. Nous nous sommes assuré le concours de nouveaux collaborateurs et informateurs qui nous aideront à rendre notre journal toujours plus intéressant et mieux renseigné.

Nous prions tous nos collaborateurs et correspondants de bien vouloir reprendre leur service complet dès aujourd'hui, et de nous envoyer sans retard leurs articles et leurs informations habituels.

Nous avons en outre le plaisir d'annoncer à nos lecteurs que l'« Impartial » publiera désormais une fois par semaine des articles d'actualité, documentaires ou humoristiques illustrés, pour lesquels il s'est assuré définitivement le concours de l'excellent dessinateur dont les œuvres ont été si hautement appréciées dans toute la Suisse à l'occasion du vote du 3 décembre.



Questions ?