Chapter 14 Sharing Information

Making analysis outputs usable **1**

Outputs **1**
- Reports 1
- MISP events 1

Output components: visualisations **2**
- Understanding time series 2
- Understanding relative sizes 4
- Understanding connections 6
- Others 7

Making analysis outputs usable

Data science, data analysis, starts and ends with human beings. We can do beautiful analysis, but if we don’t make it accessible to the people who need to take action from it, then we haven’t done our job.

There's no point building things without thinking about the end users, so let’s talk about outputs. The ways we present the data we produce, and how we do that, including the forms/ formats some of the people we interact with are used to, what good visualisations in this space look like \(and how to create them\), and how to get those outputs to the right people.

Outputs

Reports

What we get out of Orient is an incident report, containing a summary, narratives, techniques, artifacts and objects.

MISP events

We get a misp event that we can share with other groups either directly or by email, via their threat intelligence tools etc.

We added a few other things to MISP for this.

- Object types for common social media platforms, and code to load these into MISP using single-line commands in Slack, because speed is everything in a tactical response.
- New relationship types, to make the graphs that users can traverse in MISP richer.
- Taxonomies to cover things like types of threat actor.

Output components: visualisations

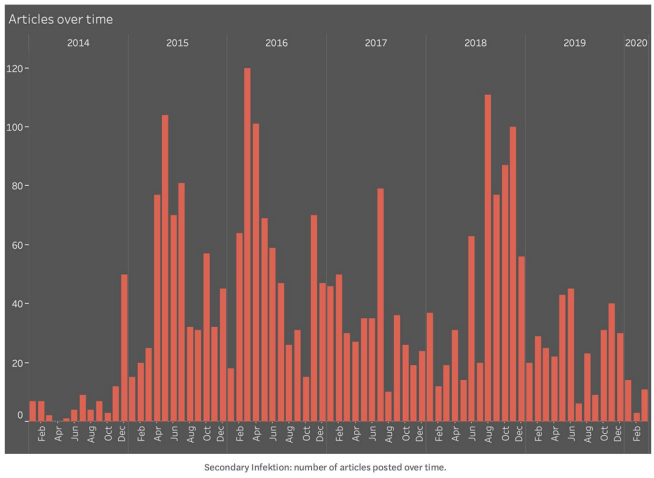
Eyeballing the data, looking at statistics, and examining machine learning outputs are good, but part of getting to know data, and explaining it to other people is being able to look at it visually. There's a lot of work on data visualisation \(read "Storytelling with Data" to see it done well\), so this section is looking at what disinformation people do with visuals.

Good places to look for what "that chart is" include

- All Charts - python visuals (most data scientists use Python)
- A Periodic Table of Visualization Methods - periodic table of visualisations

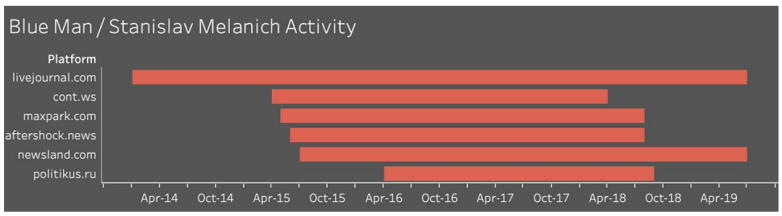
Understanding time series

Disinformation operations happen over time, so time-based plots can be useful tools. The humble bargraph \(or column plot\) is really useful for this. Almost every visualisation tool has this as an option \(e.g. [https://matplotlib.org/3.2.1/api/_as_gen/matplotlib.pyplot.bar.html](https://matplotlib.org/3.2.1/api/_as_gen/matplotlib.pyplot.bar.html#)\)



(Sekondary Infektion report, 2020\)

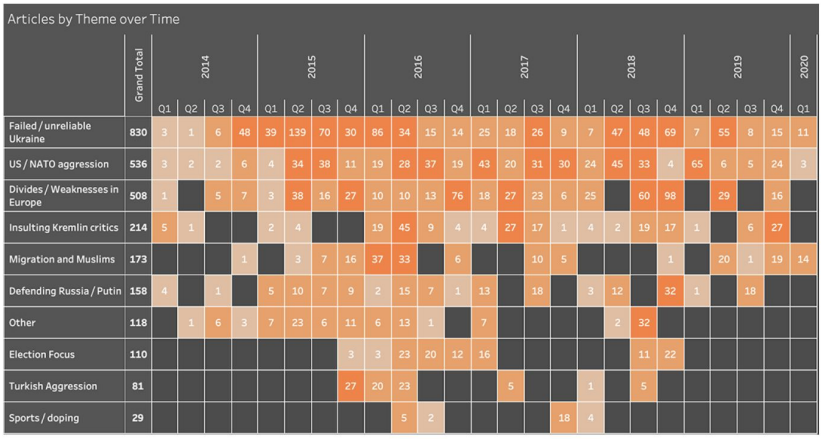
Bar graphs and line plots can be used for showing a range of entities over time.



The Blue Man's activity across different platforms, 2014-2019.

![\(Sekondary Infektion report, 2020\)]

If the value range is too large to show easily \(e.g. there's a mix of very small and very large values that you can't easily plot on one axis\), heatmaps might be more appropriate. https://python-graph-gallery.com/91-customize-seaborn-heatmap/

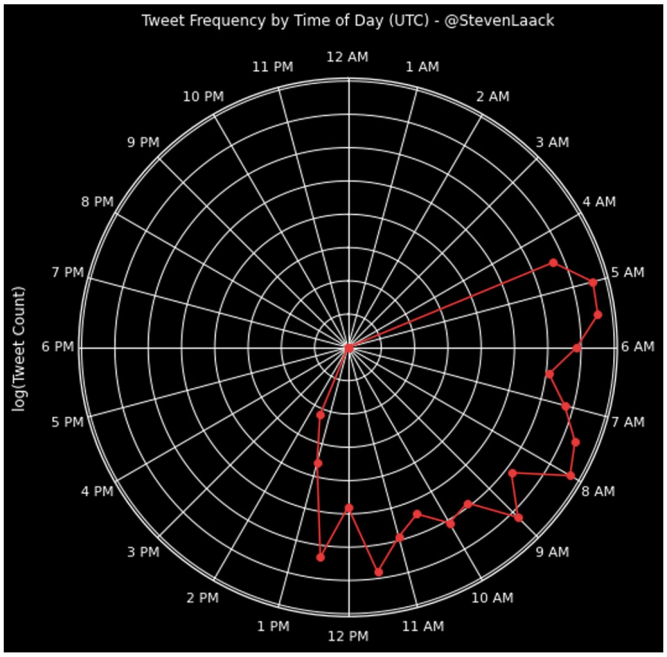


	Grand Total	2014				2015				2016				2017				2018				2019				2020
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
Failed / unreliable Ukraine	830	3	1	6	48	39	139	70	30	86	34	15	14	25	18	26	9	7	47	48	69	7	55	8	15	11
US / NATO aggression	536	3	2	2	6	4	34	38	11	19	28	37	19	43	20	31	30	24	45	33	4	65	6	5	24	3
Divides / Weaknesses in Europe	508	1		5	7	3	38	16	27	10	10	13	76	18	27	23	6	25		60	98		29		16	
Insulting Kremlin critics	214	5	1			2	4			19	45	9	4	4	27	17	1	4	2	19	17	1		6	27	
Migration and Muslims	173				1		3	7	16	37	33		6			10	5				1		20	1	19	14
Defending Russia / Putin	158	4		1		5	10	7	9	2	15	7	1	13		18		3	12		32	1		18		
Other	118		1	6	3	7	23	6	11	6	13	1		7					2	32						
Election Focus	110								3	3	23	20	12	16						11	22					
Turkish Aggression	81								27	20	23				5			1		5						
Sports / doping	29									5	2						18	4								

Secondary Infektion: main themes over time

![\(Sekondary Infektion report, 2020\)

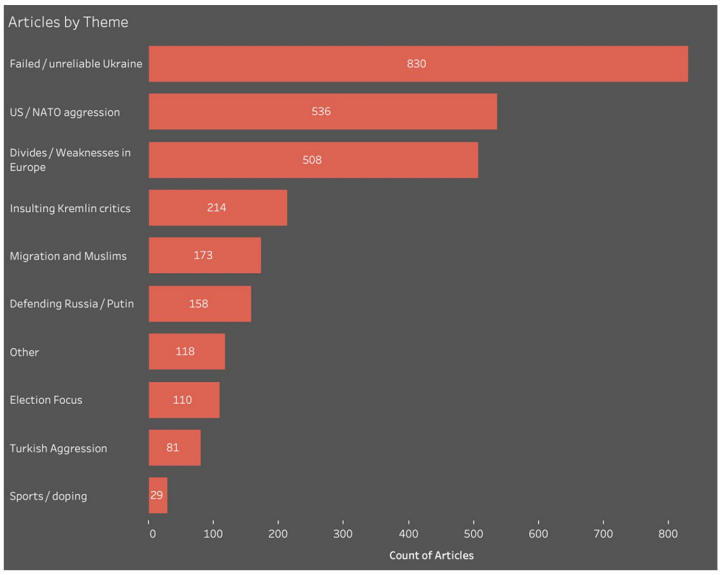
The use of spider plots for 24-hour data is good too, because they don't have a "start" or "end" time, making it easier to compare different diurnal patterns.



![\(Sekondary Infektion report, 2020\)]

Understanding relative sizes

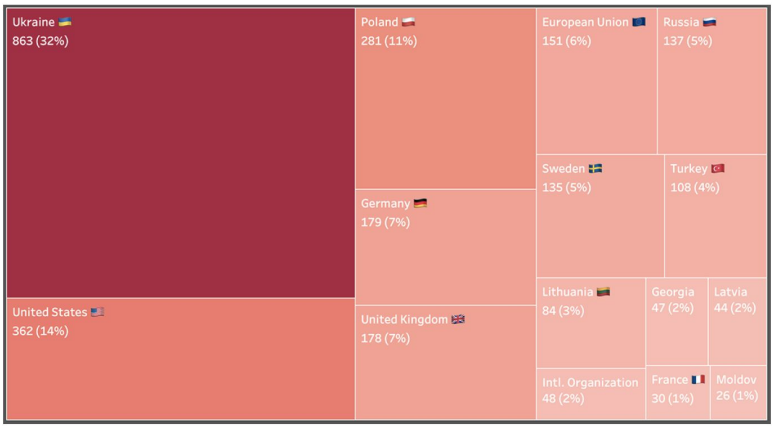
Bargraphs can do this too.
http://python-graph-gallery.com/barplot/



Breakdown of Secondary Infektion articles by theme and number.

![\(Sekondary Infektion report, 2020\)]

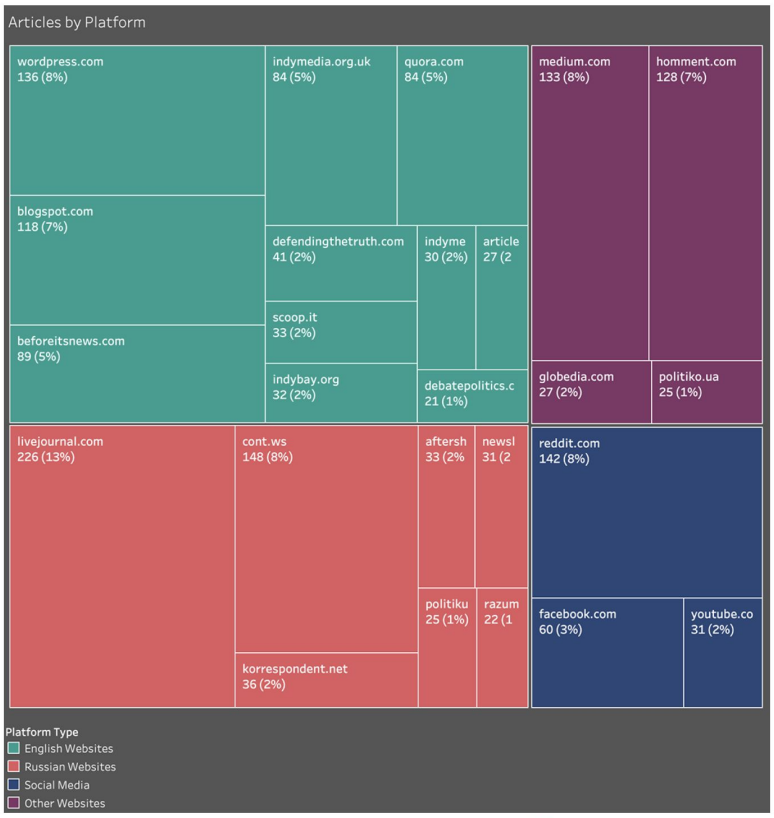
Treemaps show relative sizes as areas. https://python-graph-gallery.com/200-basic-treemap-with-python/



Countries mentioned or targeted by Secondary Infektion, total number of stories.

Sekondary Infektion report, 2020

Colours are an extra, useful, dimension on most plots.

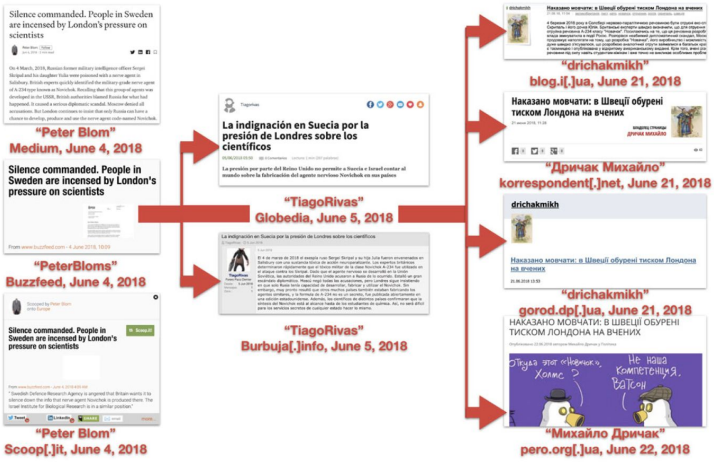


Use of platforms: where Secondary Infektion posted the most.[179]

Sekondary Infektion report, 2020

Understanding connections

Really simple graphics \(think powerpoint\) can help explain the connections between objects.

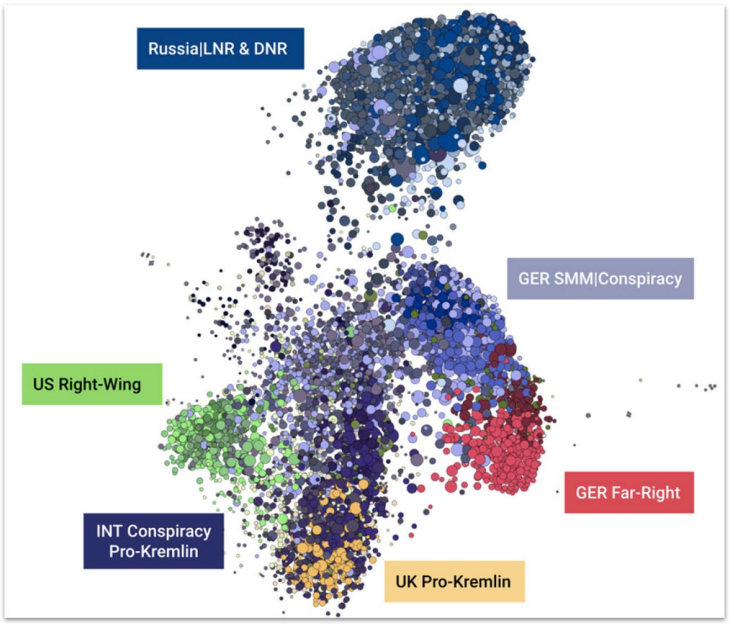


A simplified diagram of the spread of the article by assets commanded by SecondaryInfektion, from the English-language persona "Peter Blom" or "Peter Bloms" (left), through the Spanish-language persona "Tiago Rivas" (center) and the Ukrainian-language persona Михайло Дричак ("Mikhailo

Sekondary Infektion report, 2020: nice use of arrows

Simply gridding out pages or accounts with the same visuals or information can be really powerful if you're describing a network.

Graph diagrams show a large number of nodes and the connections between them - the "snurfball" images that we sometimes show to explain where the influencers are in an incident. Tools like Gephi produce these, with a little work \(and liberal use of things like the Force Atlas 2 algorithm to make the network structure easier to see\). Graphika produces "network maps" - one explanation is "The circles represent individual Twitter accounts. The volume of the circles represent influence by following, while the colours represent political ideologies.". This also looks like graph diagrams.



Graphika network map of the Secondary Infektion Twitter assets' followers and the followers of significant amplifiers, mapped April 2020.

Others

Some visualisations are hard to classify - is this a network diagram or the output from a dimension reduction algorithm? \(dimension reduction = a type of machine learning algorithm that takes a set of objects that exist in many dimensions, and flattens it so it's easy to see - usually as a two-dimensional plot\).

Specialist text analysis: look at things like Scattertext
https://towardsdatascience.com/hkprotest-visualizing-state-troll-tweets-from-chinas-disinformation-campaign-1dc4bcab437d