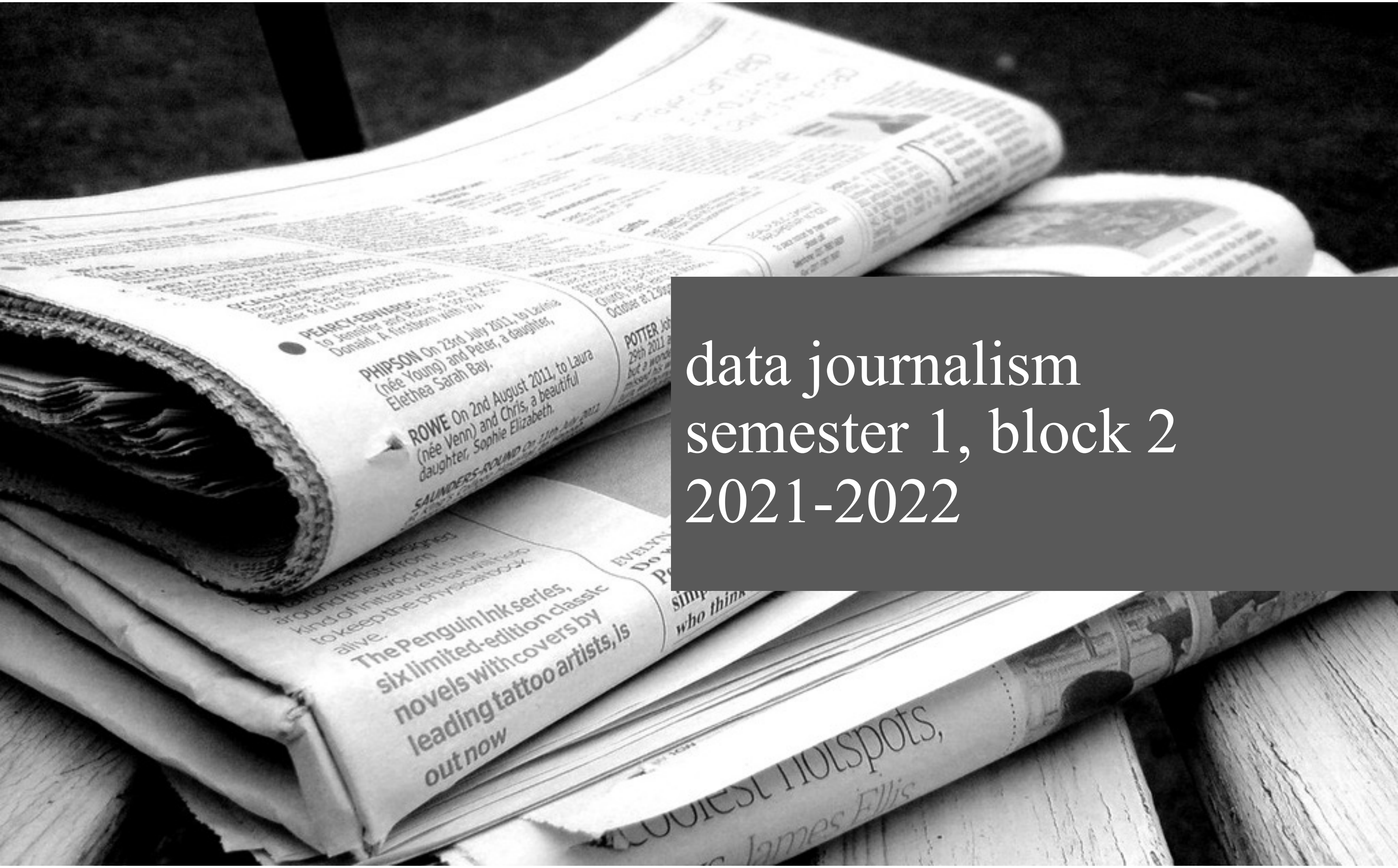

PEARCY-EDWARDS
PHIPSON On 23rd July 2011, to Lavinia (née Young) and Peter, a daughter, Elethea Sarah Bay.
ROWE On 2nd August 2011, to Laura (née Venn) and Chris, a beautiful daughter, Sophie Elizabeth.
POTTER
The Penguin Ink series, six limited-edition classic novels with covers by leading tattoo artists, is out now
Coolest Hotspots,
James Ellis

data journalism
semester 1, block 2
2021-2022

first things first

- Introductions
    - Of us
    - Of you: name & “grab & gab”
        - find an object that allows you to tell us something about yourself
- An overview of the course
- What is data journalism?
- Ensuring we’re operational

official goals for you

You should be able to demonstrate that you…

- are able to find an interesting and compelling story in a dataset;
- are able to apply basic python programming techniques learned in this course to process, analyze, and visualize the data;
- are able to translate these skills and techniques into an original piece of data journalism of 750-1000 words;
- are able, in groups, to identify an online tool relevant to data journalism and to teach their classmates how to use that tool in a short presentation and handout.

Note: scalability!

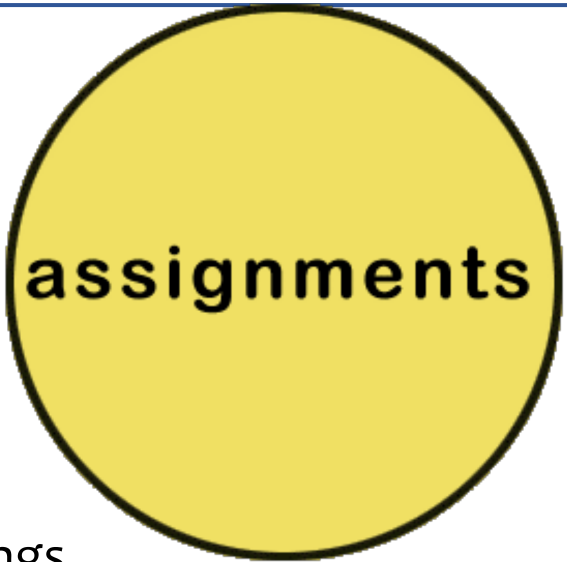
goals, more simply

- Learn a few techniques, get inspired
- Actually use some tools
- Think about this vis-à-vis real journalistic stories
- Know where to go to learn more
- Balance practice with the other academic courses
    - Think about journalistic product for Mundus thesis, too

overview of the course

- Part 1: gathering data
  - Week 1: basic techniques, finding data, ethics
- Part 2: processing and analyzing data
  - Weeks 2-3, working with numbers & texts, data wrangling
- Part 3: visualizing & presenting data
  - Weeks 4-7, visualizations & creating your own data journalism piece

working towards your final project



- Step 1: gathering data
  - Finding datasets
  - Find stories in the data
- Step 2: analyzing data
  - Conducting an analysis on data of your choice
  - Output: Jupyter notebook with description of findings
- Final project (70% of your grade)
  - Conduct analyses, visualize data, and create a data journalistic piece
  - Step 1 and step 2 are feedback moments along the way

group tool sharing assignment

- Group assignment: teaching tools

  (30% of your grade)

  - Selecting an online tool & teaching it to your classmates
  - Output: 10 mins. presentation, example & Quick Guide handout

weekly schedule

- Consult the ‘syllabus’ document on Canvas
- Thursdays more hands-on, Tuesdays more discussion/lecture (but not exclusively)
- And Wednesdays: office hours!
    - https://canvas.uva.nl/courses/24061/pages/office-hours?module_item_id=1095120

so, what *IS* data journalism?

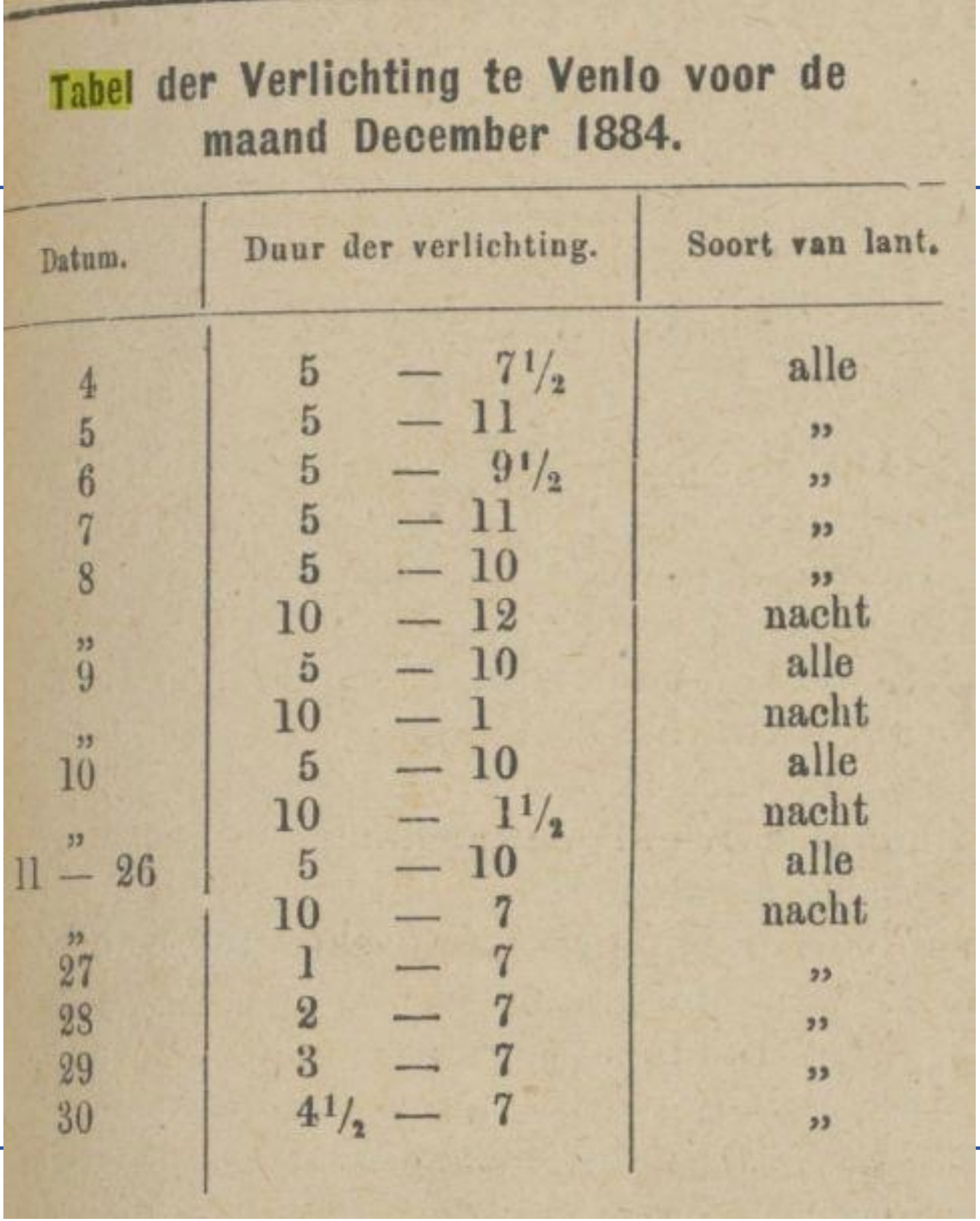
- How would you define it?

so, what *IS* data journalism?

- "The ways in which journalists can explore and make use of data-sets, which ranges from the use of infographics… to the analysis and investigation of raw data sources (Knight 2015)" (Cushion et al., 2017, p. 1199)
- "Obtaining, reporting on, curating and publishing data in the public interest" (Stray, 2011, quoted in Coddington, 2015, p. 334)
  - Not necessarily (always) linked to investigative journalism
  - Visualization as a core practice
- "Social science done on a deadline" (Steve Doig, ASU, 2012)

is it new?

Venloosch Weekblad, 06-12-1884



Tabel der Verlichting te Venlo voor de maand December 1884.

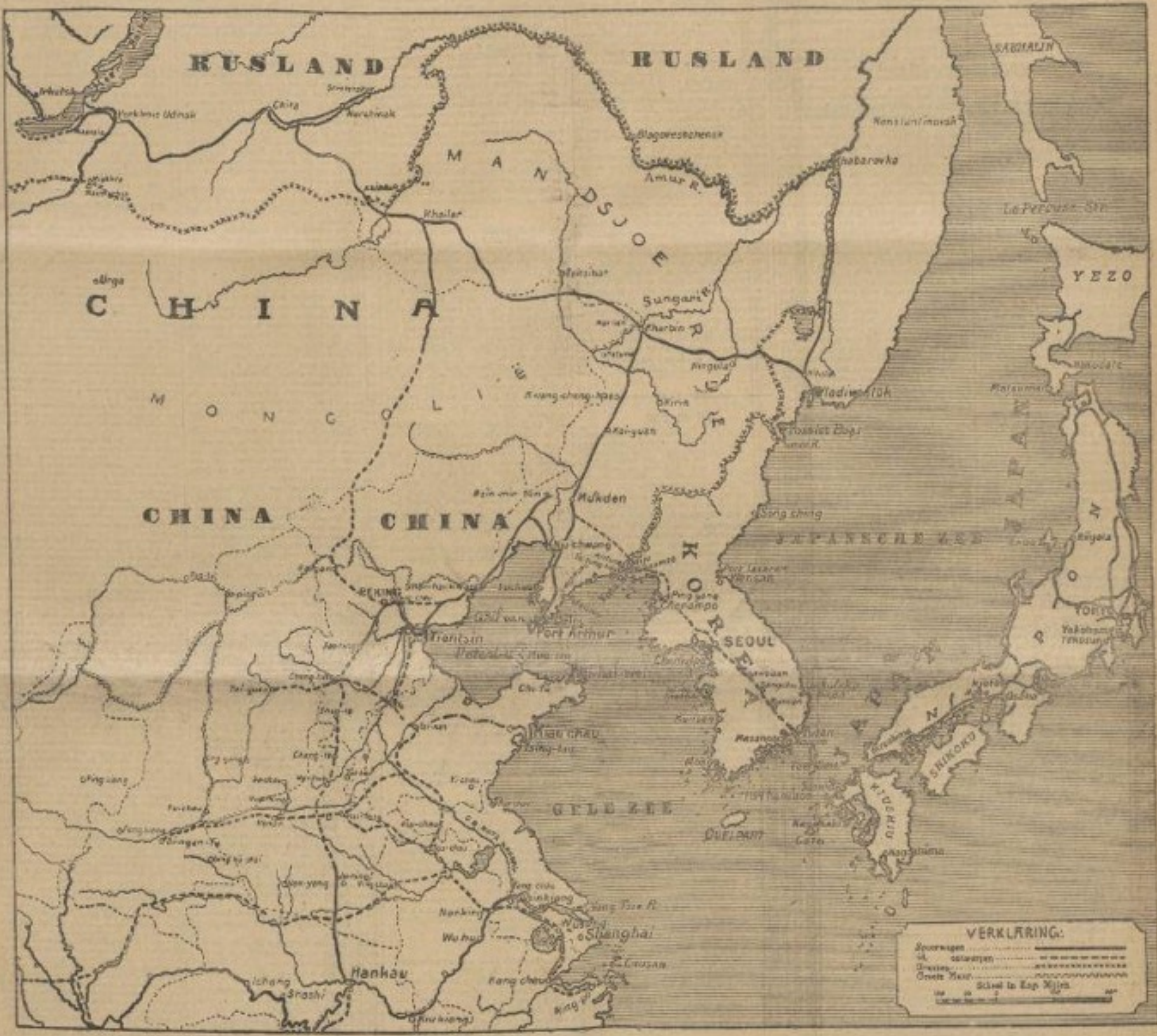
Datum.	Duur der verlichting.	Soort van lant.
4	5 — 7½	alle
5	5 — 11	”
6	5 — 9½	”
7	5 — 11	”
8	5 — 10	”
”	10 — 12	nacht
9	5 — 10	alle
”	10 — 1	nacht
10	5 — 10	alle
”	10 — 1½	nacht
11 — 26	5 — 10	alle
”	10 — 7	nacht
27	1 — 7	”
28	2 — 7	”
29	3 — 7	”
30	4½ — 7	”

ALGEMEEN HANDELSBLAD van ZATERDAG 13 FEBRUARI 1904 — Avondblad. — Derde Blad 9

AVONDBLAD.

Derde Blad.

KAART VAN HET OORLOGSTERREIN.



Algemeen Handelsblad, 13-02-1904

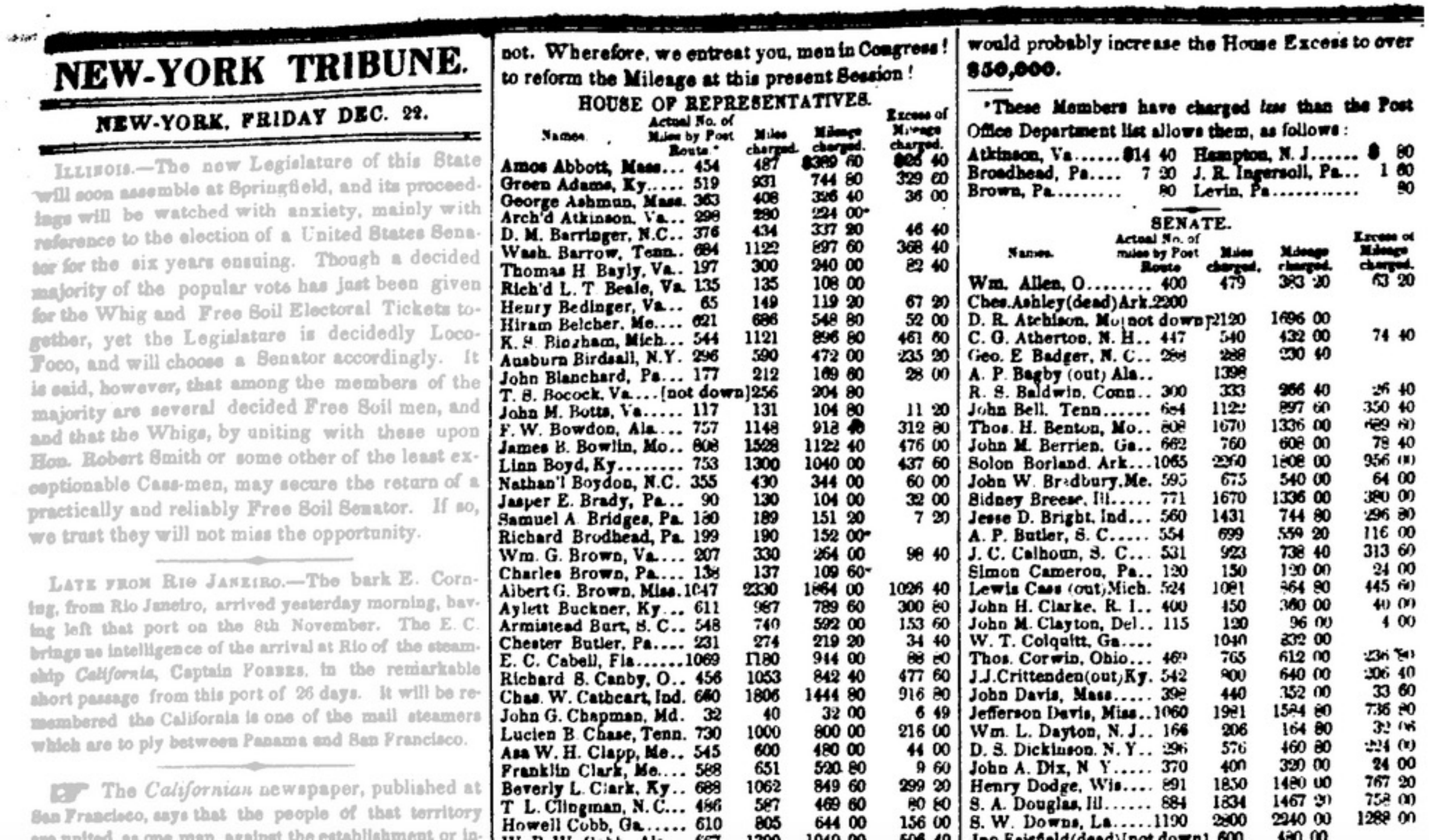
The Mileage of Congress

Published in The New York Tribune, Dec. 22, 1848

In the mid-1800s, Horace Greeley was the popular and controversial editor of The New York Tribune. For a few months starting in late 1848 he was also a congressman from New York. During that time, he produced this investigative data story, which accuses Abraham Lincoln, among many others, of taking too much money for mileage to and from the Capitol. Related Story »

https://projects.propublica.org/graphics/greeley

PAGE | TEXT



NEW-YORK TRIBUNE.

NEW-YORK, FRIDAY DEC. 22.

ILLINOIS.—The new Legislature of this State will soon assemble at Springfield, and its proceedings will be watched with anxiety, mainly with reference to the election of a United States Senator for the six years ensuing. Though a decided majority of the popular vote has just been given for the Whig and Free Soil Electoral Tickets together, yet the Legislature is decidedly Loco-Foco, and will choose a Senator accordingly. It is said, however, that among the members of the majority are several decided Free Soil men, and that the Whigs, by uniting with these upon Hon. Robert Smith or some other of the least exceptionable Cass-men, may secure the return of a practically and reliably Free Soil Senator. If so, we trust they will not miss the opportunity.

LATE FROM RIO JANEIRO.—The bark E. Corning, from Rio Janeiro, arrived yesterday morning, having left that port on the 8th November. The E. C. brings us intelligence of the arrival at Rio of the steamship *California*, Captain FORBES, in the remarkable short passage from this port of 26 days. It will be remembered the California is one of the mail steamers which are to ply between Panama and San Francisco.

☞ The *Californian* newspaper, published at San Francisco, says that the people of that territory are united, as one man, against the establishment or in-

not. Wherefore, we entreat you, men in Congress! to reform the Mileage at this present Session!

HOUSE OF REPRESENTATIVES.

Names.	Actual No. of Miles by Post Route.*	Miles charged.	Mileage charged.	Excess of Mileage charged.
Amos Abbott, Mass...	454	487	$389 60	$26 40
Green Adams, Ky.....	519	931	744 80	329 60
George Ashmun, Mass.	363	408	326 40	36 00
Arch'd Atkinson, Va...	298	280	224 00*	
D. M. Barringer, N.C..	376	434	337 20	46 40
Wash. Barrow, Tenn..	684	1122	897 60	368 40
Thomas H Bayly, Va..	197	300	240 00	82 40
Rich'd L. T Beale, Va.	135	135	108 00	
Henry Bedinger, Va...	65	149	119 20	67 20
Hiram Belcher, Me....	621	686	548 80	52 00
K. S. Bingham, Mich...	544	1121	896 80	461 60
Ausburn Birdsall, N.Y.	296	590	472 00	235 20
John Blanchard, Pa....	177	212	169 60	28 00
T. S. Bocock, Va....[not down]		256	204 80	
John M. Botts, Va......	117	131	104 80	11 20
F. W. Bowdon, Ala....	757	1148	918 40	312 80
James B. Bowlin, Mo..	808	1528	1122 40	476 00
Linn Boyd, Ky.........	753	1300	1040 00	437 60
Nathan'l Boydon, N.C.	355	430	344 00	60 00
Jasper E. Brady, Pa...	90	130	104 00	32 00
Samuel A. Bridges, Pa.	180	189	151 20	7 20
Richard Brodhead, Pa.	199	190	152 00*	
Wm. G. Brown, Va....	207	330	264 00	98 40
Charles Brown, Pa....	138	137	109 60*	
Albert G. Brown, Miss.	1047	2330	1864 00	1026 40
Aylett Buckner, Ky...	611	987	789 60	300 80
Armistead Burt, S. C..	548	740	592 00	153 60
Chester Butler, Pa.....	231	274	219 20	34 40
E. C. Cabell, Fla.......	1069	1180	944 00	88 80
Richard S. Canby, O..	456	1053	842 40	477 60
Chas. W. Cathcart, Ind.	640	1806	1444 80	916 80
John G. Chapman, Md.	32	40	32 00	6 49
Lucien B Chase, Tenn.	730	1000	800 00	216 00
Asa W. H. Clapp, Me..	545	600	480 00	44 00
Franklin Clark, Me....	588	651	520 80	9 60
Beverly L. Clark, Ky..	688	1062	849 60	299 20
T L. Clingman, N. C...	486	587	469 60	80 80
Howell Cobb, Ga......	610	805	644 00	156 00
W. R. W. Cobb, Ala...	667	1300	1040 00	506 40

would probably increase the House Excess to over **$50,000.**

*These Members have charged *less* than the Post Office Department list allows them, as follows:

Atkinson, Va......	$14 40	Hampton, N. J......	$ 80
Broadhead, Pa....	7 20	J. R. Ingersoll, Pa...	1 60
Brown, Pa.........	80	Levin, Pa.............	80

SENATE.

Names.	Actual No. of miles by Post Route	Miles charged.	Mileage charged.	Excess of Mileage charged.
Wm. Allen, O.........	400	479	383 20	63 20
Ches. Ashley (dead) Ark.	2200			
D. R. Atchison, Mo [not down]		2120	1696 00	
C. G. Atherton, N. H..	447	540	432 00	74 40
Geo. E Badger, N. C..	288	288	230 40	
A. P. Bagby (out) Ala..		1398		
R. S. Baldwin, Conn..	300	333	266 40	26 40
John Bell, Tenn.......	684	1122	897 60	350 40
Thos. H. Benton, Mo..	808	1670	1336 00	689 60
John M. Berrien, Ga..	662	760	608 00	78 40
Solon Borland, Ark....	1065	2260	1808 00	956 00
John W. Bradbury, Me.	595	675	540 00	64 00
Sidney Breese, Ill......	771	1670	1336 00	380 00
Jesse D. Bright, Ind....	560	1431	744 80	296 80
A. P. Butler, S. C.....	554	699	559 20	116 00
J. C. Calhoun, S. C...	531	923	738 40	313 60
Simon Cameron, Pa..	120	150	120 00	24 00
Lewis Cass (out) Mich.	524	1081	864 80	445 60
John H. Clarke, R. I..	400	450	360 00	40 00
John M. Clayton, Del..	115	120	96 00	4 00
W. T. Colquitt, Ga....		1040	832 00	
Thos. Corwin, Ohio...	469	765	612 00	236 80
J.J. Crittenden (out) Ky.	542	800	640 00	206 40
John Davis, Mass.....	392	440	352 00	33 60
Jefferson Davis, Miss..	1060	1981	1584 80	736 80
Wm. L. Dayton, N. J..	166	206	164 80	32 00
D. S. Dickinson, N. Y..	296	576	460 80	224 00
John A. Dix, N Y.....	370	400	320 00	24 00
Henry Dodge, Wis....	891	1850	1480 00	767 20
S. A. Douglas, Ill......	884	1834	1467 20	758 00
S. W. Downs, La.....	1190	2800	2240 00	1288 00
Jno Fairfield (dead) [not down]		600	480 00	

so, why now?

what has changed, societally/politically?

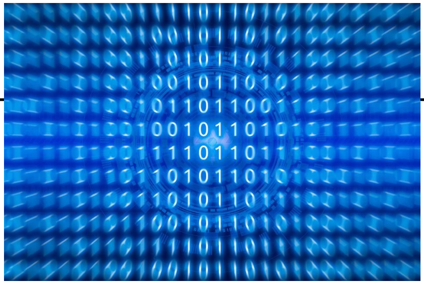
- Think Blumler & Kavanagh's "3rd age"

- Modernization
- Individualization
- Secularization
  - Fragmentation of society
  - Loss of traditional structures
  - Increased emphasis on personal aspirations, making own way
- Economization
- Aestheticization
  - Rise of monetary, image-based, free market values
- Rationalization
  - Increased demand for evidence, research
- Mediatization
  - Increasing centrality of media in social processes

Source: Blumler, J. G., & Kavanagh, D. (1999). The third age of political communication: Influences and features. *Political Communication, 16*(3), 209-230.

what has changed, technologically?

- More Data
  - Digital vs. analog
  - Digital traces
  - Internet of Things
  - Open government
  - Leaks
  - …



- Better Facilities
  - More storage
  - Online services
  - Better algorithms
  - User-friendlier
  - Hand-held devices
  - …

data journalism as new ways of…

- Finding stories
  - e.g. in datasets that didn’t exist before
  - e.g. in datasets that weren’t manageable/accessible before
  - e.g. in networks of journalists & outlets, more collaborative
- Telling stories
  - e.g. in interactive ways allowing personalization
  - e.g. using visualization tools that are much more engaging and appealing to news consumers
  - e.g. in transparent ways; data made available to public

problem is…

- A lot of actual data journalism is (still) poorly done
- Cushion et al. study examined “statistical references”
    - Use of figures, or statements which related to figures, in news items that “could realistically be used to make *statistical comparisons* relevant to the story (across time, borders, etc.) or *inferences about a wider situation*, even if the comparison or inference was not always made explicit” (p. 1201)
        - Yes: “cost of solar power is tumbling” (based on & refers to comparative data-sets)
        - No: “new contract will create 1000 jobs and cost £300 million” (figures in isolation, no comparisons referred to)
    - Thoughts about this operationalization?

Source: Cushion, S., Lewis, J., & Callaghan, R. (2017). Data journalism, impartiality and statistical claims. *Journalism Practice*, *11*(10), 1198-1215.

cushion et al.: clarity & context

- References can be:
  - Made in passing, or without clarifying data or providing source/context
  - Made more explicit, but without explaining broader comparative picture or data quality
  - Made clearly, with explanation of data or some reference to method (still a low bar)

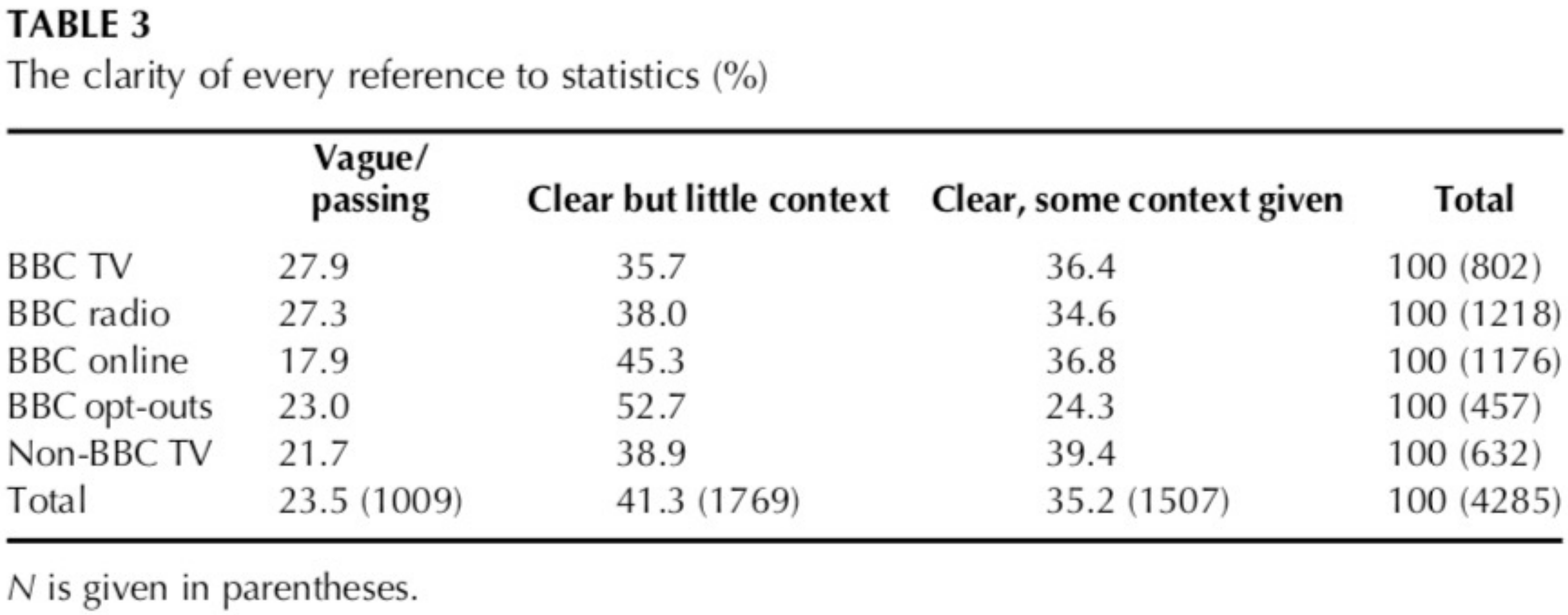


TABLE 3

The clarity of every reference to statistics (%)

	Vague/ passing	Clear but little context	Clear, some context given	Total
BBC TV	27.9	35.7	36.4	100 (802)
BBC radio	27.3	38.0	34.6	100 (1218)
BBC online	17.9	45.3	36.8	100 (1176)
BBC opt-outs	23.0	52.7	24.3	100 (457)
Non-BBC TV	21.7	38.9	39.4	100 (632)
Total	23.5 (1009)	41.3 (1769)	35.2 (1507)	100 (4285)

N is given in parentheses.

Source: Cushion, S., Lewis, J., & Callaghan, R. (2017). Data journalism, impartiality and statistical claims. *Journalism Practice*, *11*(10), 1198-1215.

cushion et al.: topics*

episodic framing?

cultivation theory?

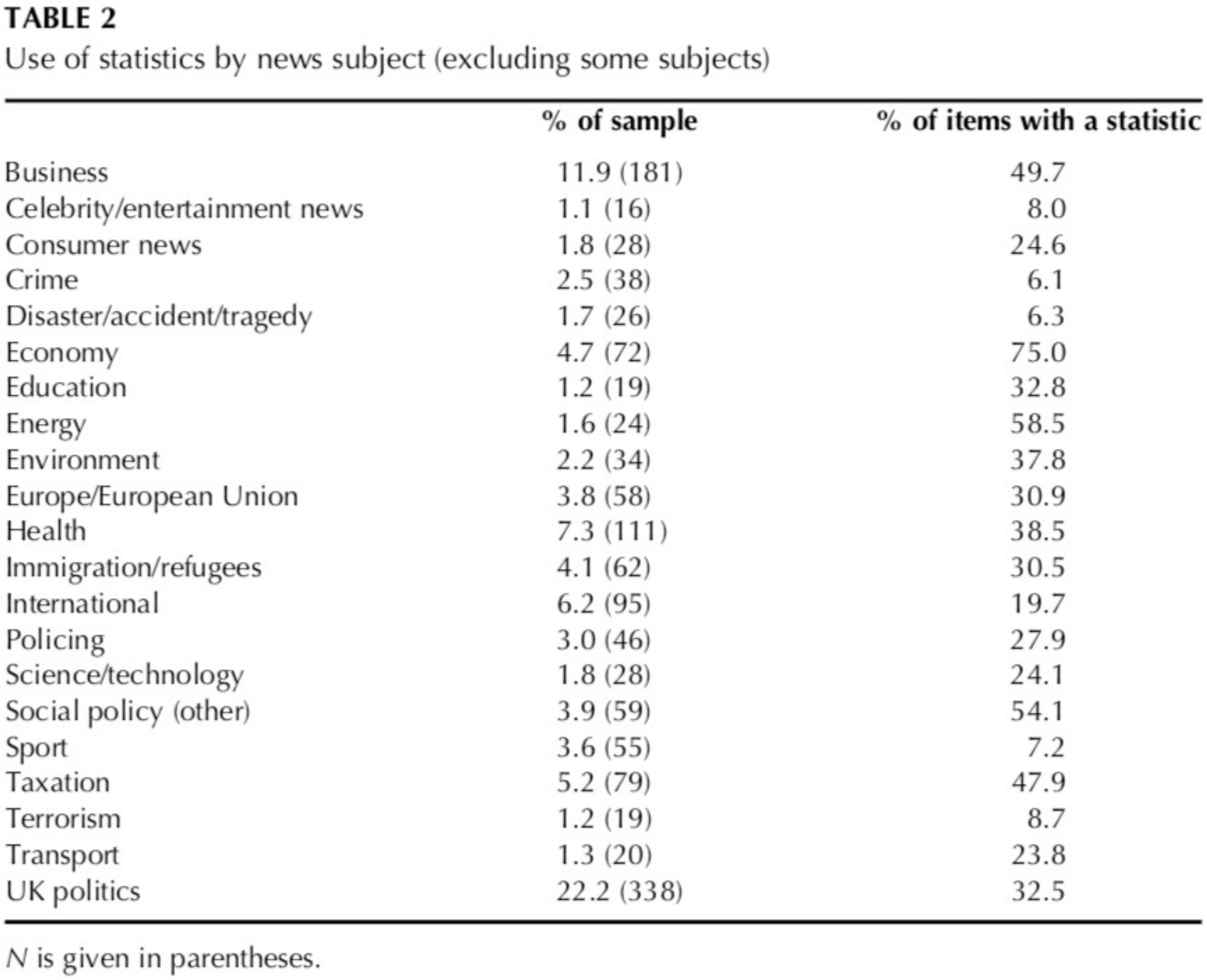


TABLE 2

Use of statistics by news subject (excluding some subjects)

	% of sample	% of items with a statistic
Business	11.9 (181)	49.7
Celebrity/entertainment news	1.1 (16)	8.0
Consumer news	1.8 (28)	24.6
Crime	2.5 (38)	6.1
Disaster/accident/tragedy	1.7 (26)	6.3
Economy	4.7 (72)	75.0
Education	1.2 (19)	32.8
Energy	1.6 (24)	58.5
Environment	2.2 (34)	37.8
Europe/European Union	3.8 (58)	30.9
Health	7.3 (111)	38.5
Immigration/refugees	4.1 (62)	30.5
International	6.2 (95)	19.7
Policing	3.0 (46)	27.9
Science/technology	1.8 (28)	24.1
Social policy (other)	3.9 (59)	54.1
Sport	3.6 (55)	7.2
Taxation	5.2 (79)	47.9
Terrorism	1.2 (19)	8.7
Transport	1.3 (20)	23.8
UK politics	22.2 (338)	32.5

N is given in parentheses.

Source: Cushion, S., Lewis, J., & Callaghan, R. (2017). Data journalism, impartiality and statistical claims. *Journalism Practice*, *11*(10), 1198-1215.

cushion et al.: sources of the references

- Journalists by far dominant; external sources often not even identified

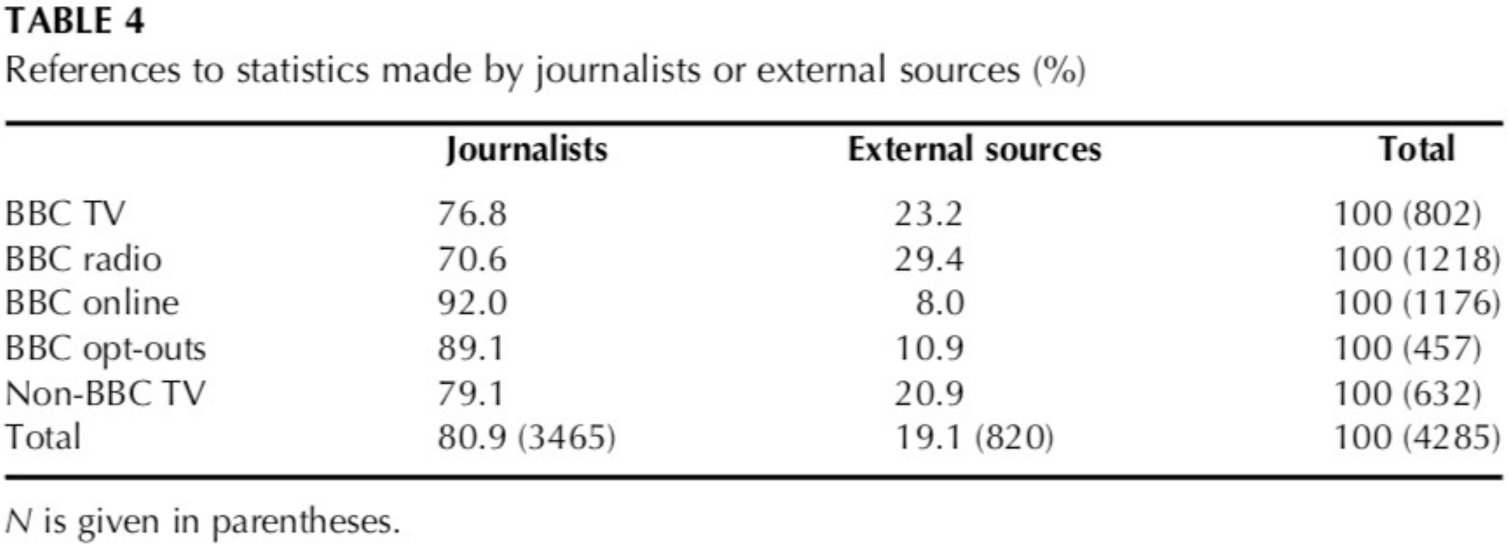


TABLE 4

References to statistics made by journalists or external sources (%)

	Journalists	External sources	Total
BBC TV	76.8	23.2	100 (802)
BBC radio	70.6	29.4	100 (1218)
BBC online	92.0	8.0	100 (1176)
BBC opt-outs	89.1	10.9	100 (457)
Non-BBC TV	79.1	20.9	100 (632)
Total	80.9 (3465)	19.1 (820)	100 (4285)

N is given in parentheses.

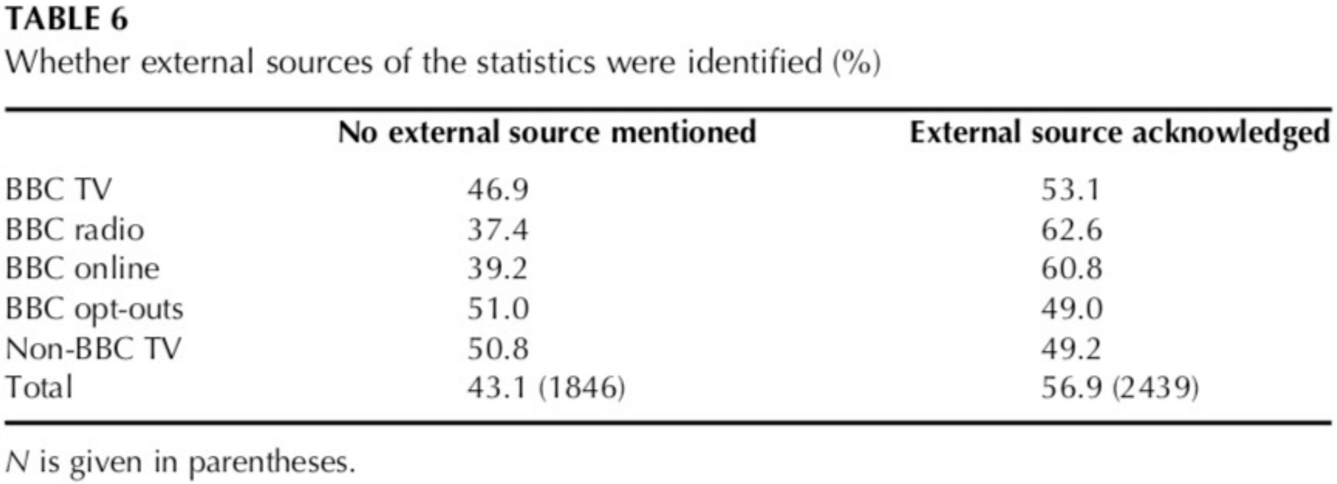


TABLE 6

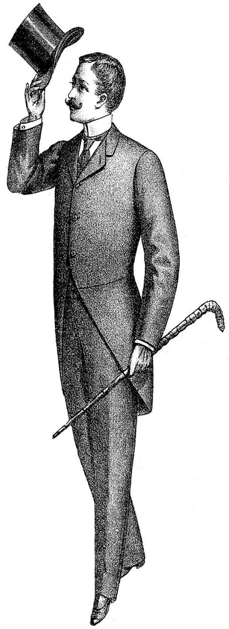
Whether external sources of the statistics were identified (%)

	No external source mentioned	External source acknowledged
BBC TV	46.9	53.1
BBC radio	37.4	62.6
BBC online	39.2	60.8
BBC opt-outs	51.0	49.0
Non-BBC TV	50.8	49.2
Total	43.1 (1846)	56.9 (2439)

N is given in parentheses.

ıs. *Journalism Practice*, *11*(10), 1198-1215.

reinforcing institutional voices?



- Elite sources referenced, and found some partisan imbalance
- But, can this be interpreted another way?
  - As demanding statistics of officials, to support their arguments?
  - Yes, but… only 4.2% of references attributed to government sources were challenged or contextualized by journalists
  - Often, a competing stat is given without any judgment or analysis

Source: Cushion, S., Lewis, J., & Callaghan, R. (2017). Data journalism, impartiality and statistical claims. *Journalism Practice*, *11*(10), 1198-1215.

their solution?

- 1. Diversifying the range of sources
- 2. Rethinking the practice of "impartiality"
- 3. Questioning news values that privilege drama over context
- "Part of the function of good journalism is to communicate *what the weight of the evidence tells us*" (p. 1213)

Source: Cushion, S., Lewis, J., & Callaghan, R. (2017). Data journalism, impartiality and statistical claims. *Journalism Practice*, *11*(10), 1198-1215.

points of discussion

- What do **you** see as the *purpose* of journalism in this data-rich era?
  - How has that changed from 50, even 20 years ago?
  - And how does it depend on (national/local) context?
- “Using data, the job of journalists shifts its main focus from being the first ones to report to being the ones telling us what a certain development might actually mean” (DJH, introduction)
  - Data journalism: goal to help public enhance its own understanding of public issues (Coddington’s distinction)
- What kind of *expertise* is needed to fulfill this purpose? Are journalists experts? Were they ever?

what of codes of ethics? must they be unique?

a. **Does the information serve a journalistic and public purpose**? To what extent? The data must at least serve both of these purposes to be posted online.

b. **Who could be harmed by the information? To what extent**? Are there risks to a person's private life from elements of the data? What is the potential impact of data that may be erroneous or out of date?

c. **Are there alternatives that would maximize the public purpose such as combining with information from other databases**? **Are there alternatives that would minimize harm, such as aggregating personal data instead of using individual names and addresses**?

d. **Can the data be verified? Have reasonable steps been taken to verify the accuracy of the data**? Can people in the database be notified before publication? What can be done to enable correction of data errors identified after publication?

Source: Craig, D., Ketterer, S., & Yousuf, M. (2017). To post or not to post: Online discussion of gun permit mapping and the development of ethical standards in data journalism. *Journalism & Mass Communication Quarterly, 94*. 168-188

summary

- Data journalism a fuzzy term
- But there is something new, unique to it
- Requires expertise, skills
- Potentially changing the role & purpose of journalists?
- We hope you see the relevance of this & these skills, particularly

let’s try Jupyter Notebook

- Go to our class “book” and download the first notebook:

https://fhopp.github.io/data_journalism//content/intro.html

for thursday

- Make sure you know how to open Anaconda and how to download/access notebooks
- Consult the cheat sheet if unsure.
- Ideally: finish up the python tutorials by the end of the week