

Building information visualizations with PubMed

Ed Sperr

Clinical Information Librarian

Augusta University/University of Georgia Medical Partnership

Why PubMed?

It's big enough to tell us something about the biomedical literature as a whole

It's free!

It's easy to develop against

PubMed and MEDLINE

MEDLINE is the carefully organized database from the National Library of Medicine that provides access to “journal articles in life sciences with a concentration on biomedicine”

Curated list of more than [5,600 journals](#)

Citations indexed with [MeSH controlled vocabulary](#)

Almost 24 million indexed citations

PubMed is that and in-process stuff, publisher-supplied records, books, etc. – a little more than 27 million records

Both are searched through the same interface

Why visualizations?

Hard (at least for some of us) to form mental pictures without them

They're pretty!

If they are *interactive*, they can give one a new way of interfacing with data

Basic Pattern

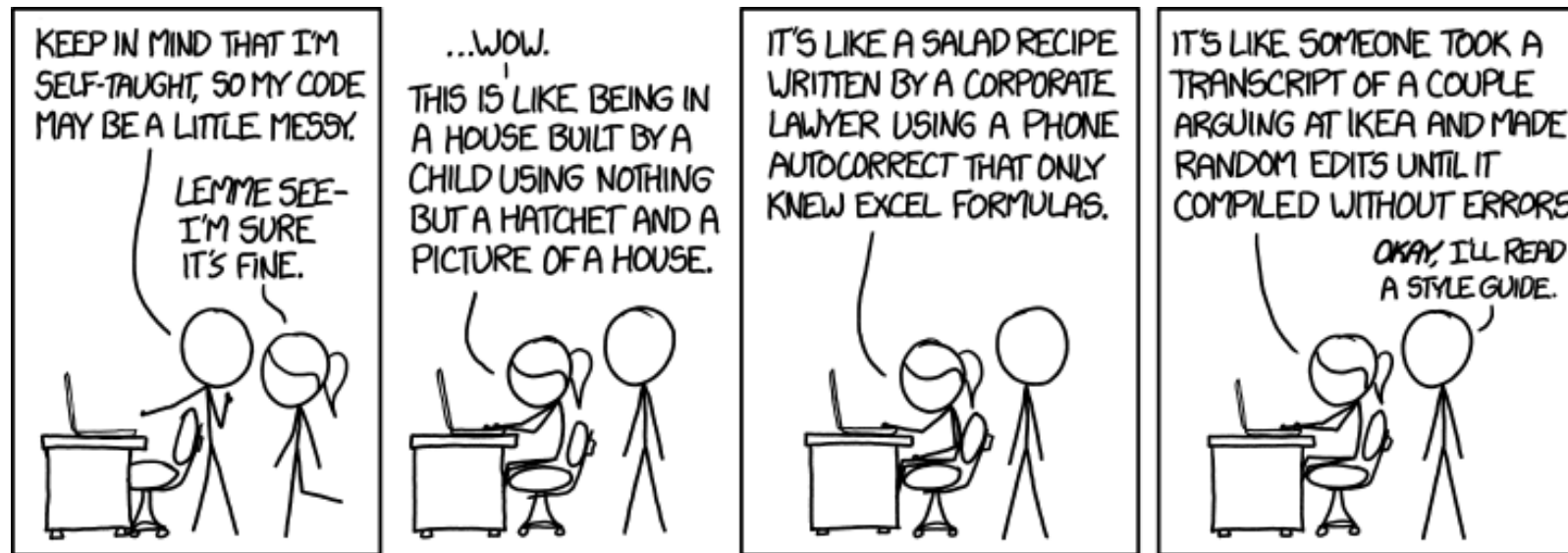
1. Send search to PubMed
2. Retrieve counts
3. Compare them to a baseline of some kind
4. Graph the difference
5. ???
6. Profit

Design Principles: Fast, Cheap and Out of Control

Use existing libraries whenever possible

Push as much out to the browser as possible

Working code will work just fine...



E-Utilities

API to NCBI databases

Maintained by NLM

Simple syntax for calls – easy to implement in many environments

[https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=pubmed
&term=*diabetes*&retmode=*json*](https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=pubmed&term=diabetes&retmode=json)

Brand-new documentation at <https://dataguide.nlm.nih.gov/>

Google Charts

Free

Simple

Works across browsers and probably(?) will be supported for a while

Lots of different iterations on the same theme

Once you figure out one, easier to do another

https://developers.google.com/chart/interactive/docs/quick_start

Case one: Distribution by MeSH category or Subheading

MeSH headings are classes under 16 main branches

“Anatomy”, “Humanities”, etc

Many times MeSH headings will be further qualified with a subheading

“Analysis”, “Diagnosis”, “Psychology”, etc.

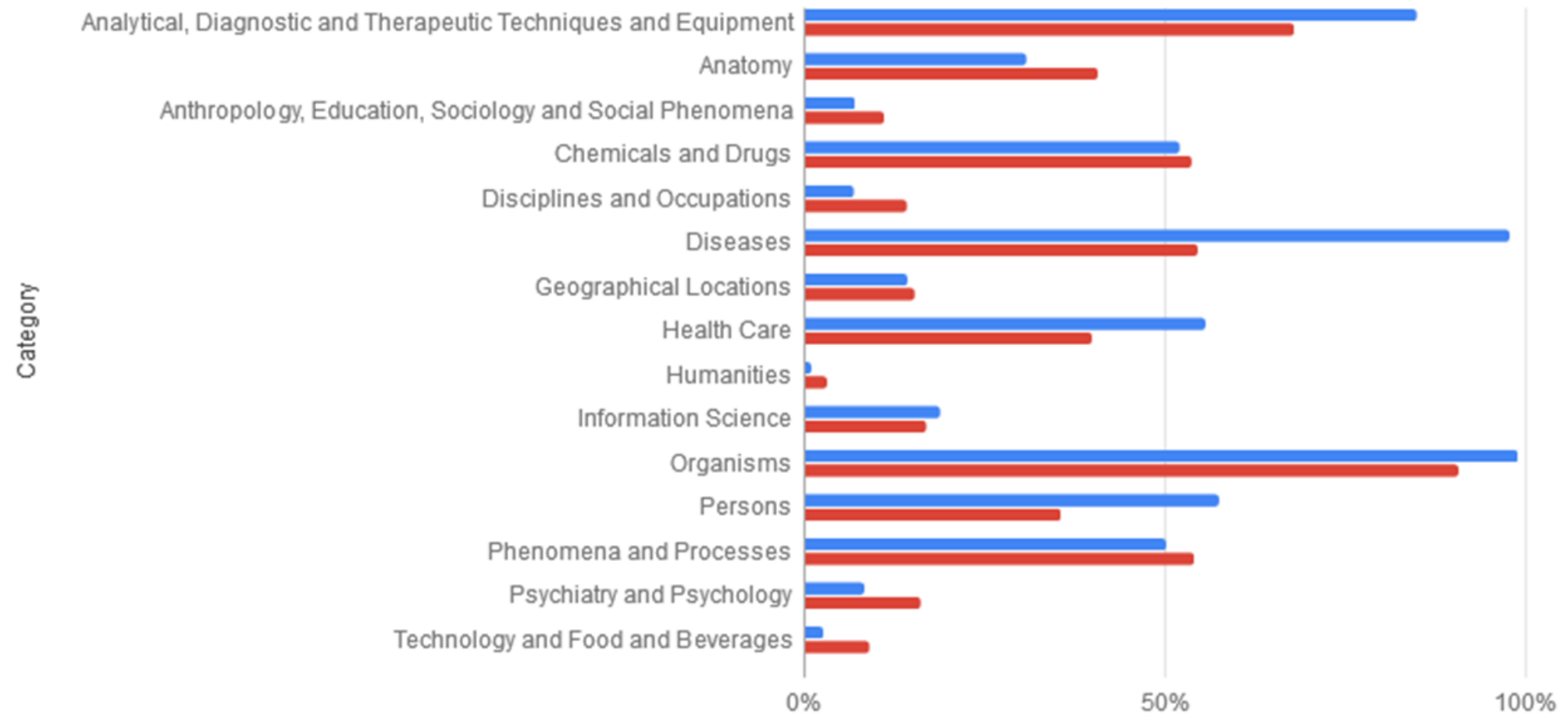
Do the *proportions* of headings under each tell us something about our search?

<https://esperr.github.io/mesh-cat-graph/> and

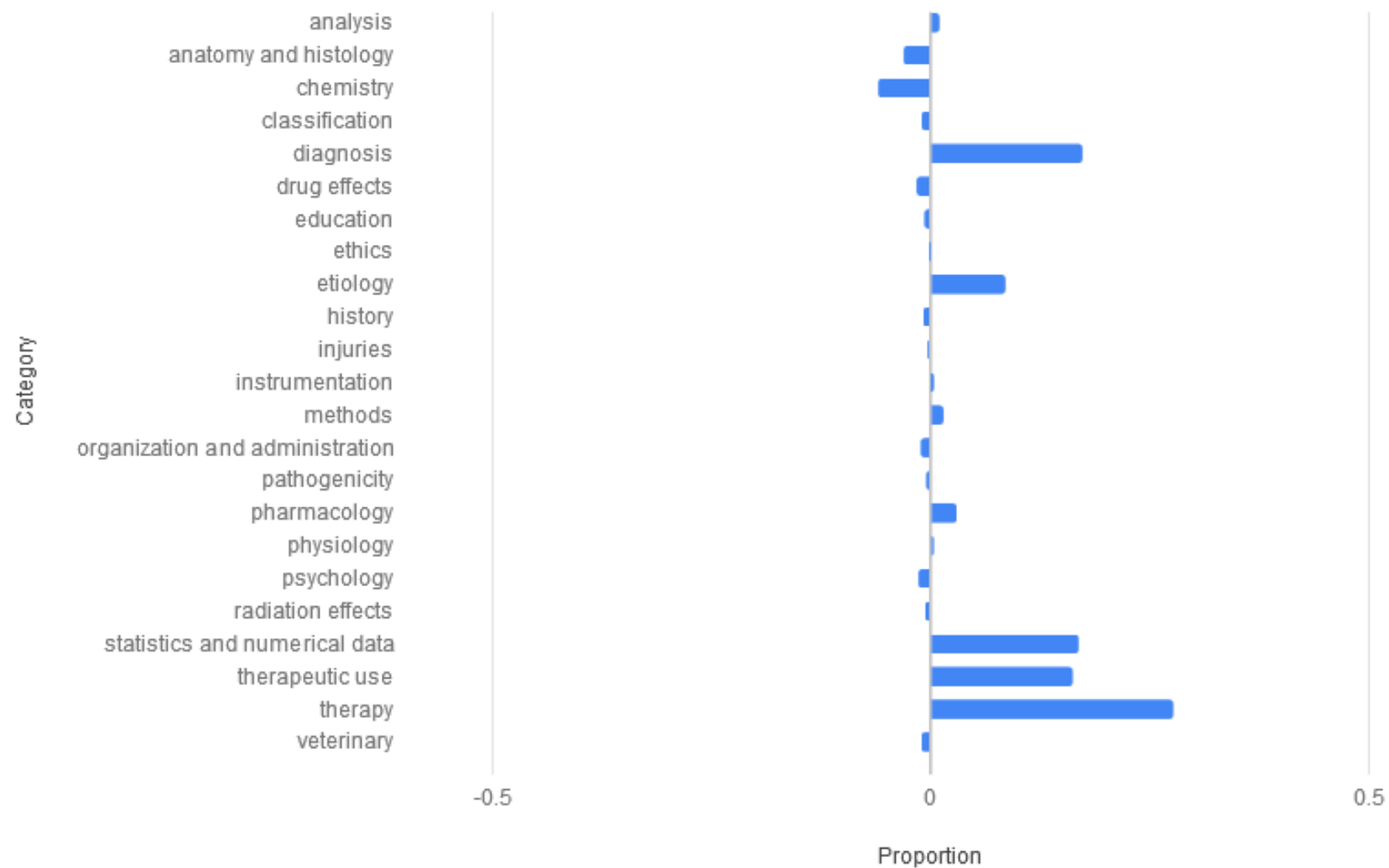
<https://esperr.github.io/mesh-subhead-graph/>

Percentage of results in each category for "myocardial infarction"

Your search
All MEDLINE



Proportion of subheadings for "myocardial infarction" compared to baseline



Made with Mesh Subheading Graph: <http://esperr.github.io/mesh-subhead-graph/>

Difference between proportions of results for "cancer" and "heart disease" in each category



Made with Mesh Category Graph: <http://esperr.github.io/mesh-cat-graph/>

cancer | heart disease

Case two: citations on a choropleth

MeSH headings include not only topic areas but *geographic regions*

Around 15% of all indexed records are tagged with the name of a continent, country or city

For real coverage, probably want to throw in some text search as well

Too slow to do everything at once, so we split up the work

<https://esperr.github.io/mapping-medline/>

Mapping MEDLINE

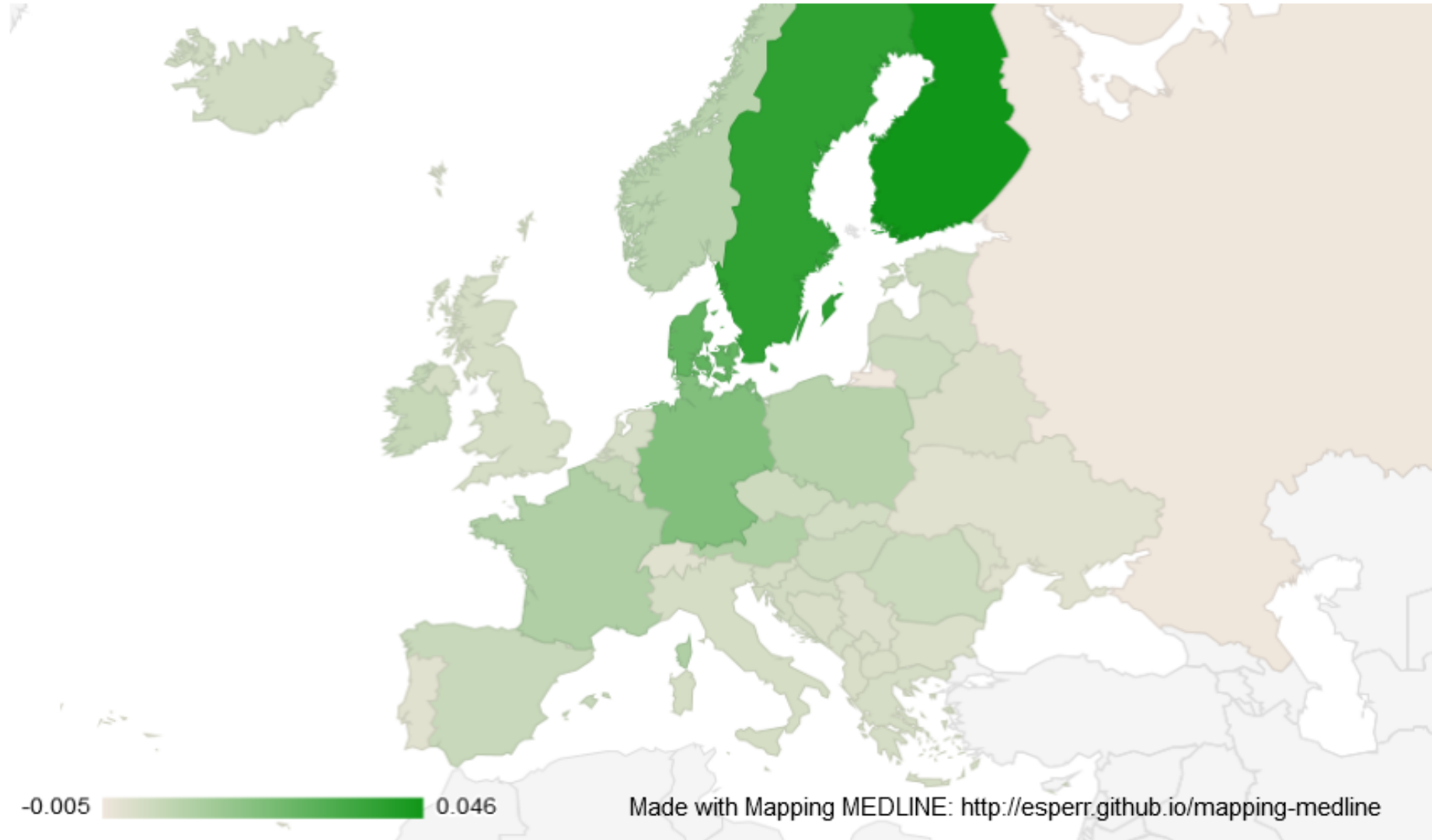
10,335 geographically tagged results for 'Diabetes Mellitus, Type 1'

(Select a region below to see the map)

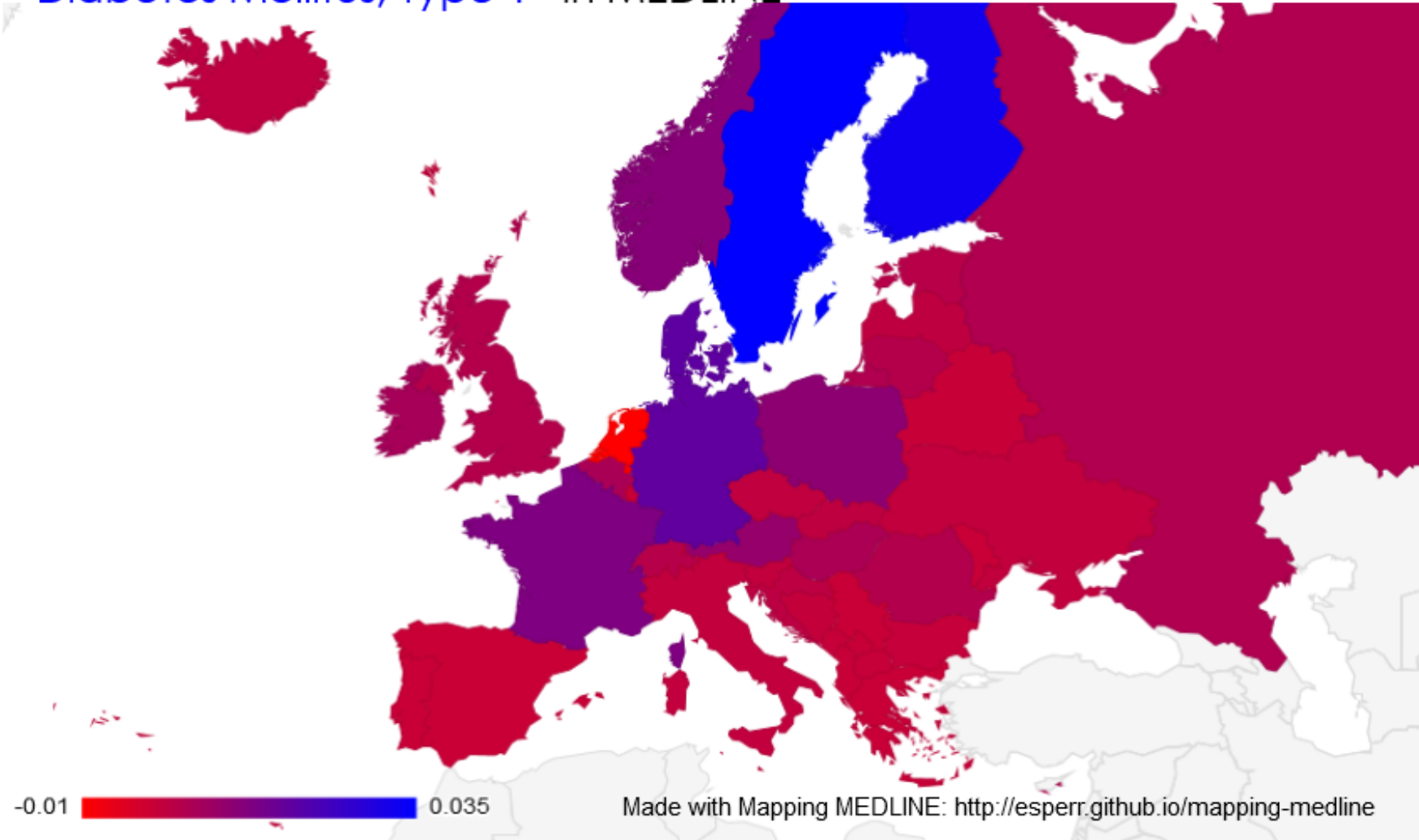
Africa 540	Asia 1,790
Caribbean Region 72	Central America 4
Europe 5,348	North America 2,462
Oceania 445	South America 301

Note: Use the 'Back to region list' link to return to this list

Proportion of MEDLINE citations for 'Diabetes Mellitus, Type 1 ' in Europe



Proportion of 'Diabetes Mellitus, Type 2' vs
'Diabetes Mellitus, Type 1' in MEDLINE



Case three: Citations over time

We *could* search each year individually, but that would take forever

Is there something else we could do?

Urlfetch and regex to the rescue!

<https://esperr.github.io/pubmed-by-year/>

[Psychiatry: Benchmarks for the H-Index.](#)
Rittenbach K.
doi: 10.1007/s40596-016-0656-2. [Epub ahead of print]

[Biosensors in Italy: A Review of the 2015 Literature.](#)
Natale CD, Neri G, Seeber R, Tajani A.
doi: 10.3390/s17040868. Review.

[Oropouche research: impact on the surveillance of emerging arboviruses](#)

Ospina JA, Patiño-Barbosa AM, Rodríguez-Morales AJ.
Feb 28 [revised 2017 Mar 17];6:194. doi: 10.12688/1000research.10936.2. eCollection

IC Article

[Lung, and Blood Institute Data Repository](#)
Wagner EL, Goldfarb ME, Hitchcock DM, Giffen CA.
doi: 10.1056/NEJMsa1603542. [Epub ahead of print]

(1006)
[Diagnosis/Broad \(475\)](#)
[Etiology/Broad \(1196\)](#)
[Georgia Regents University Reese Library, GA \(6268\)](#)
[Prognosis/Broad \(457\)](#)
[Therapy/Broad \(992\)](#)

[Manage Filters](#)

Results by year



Related searches

[bibliometric analysis diabetes](#)
[research bibliometric analysis](#)

Proportion of citations in PubMed

proportion for each search by year, 1945 to 2017



Made with PubMed by Year: <http://esperr.github.io/pubmed-by-year>

Case four: Venn diagrams

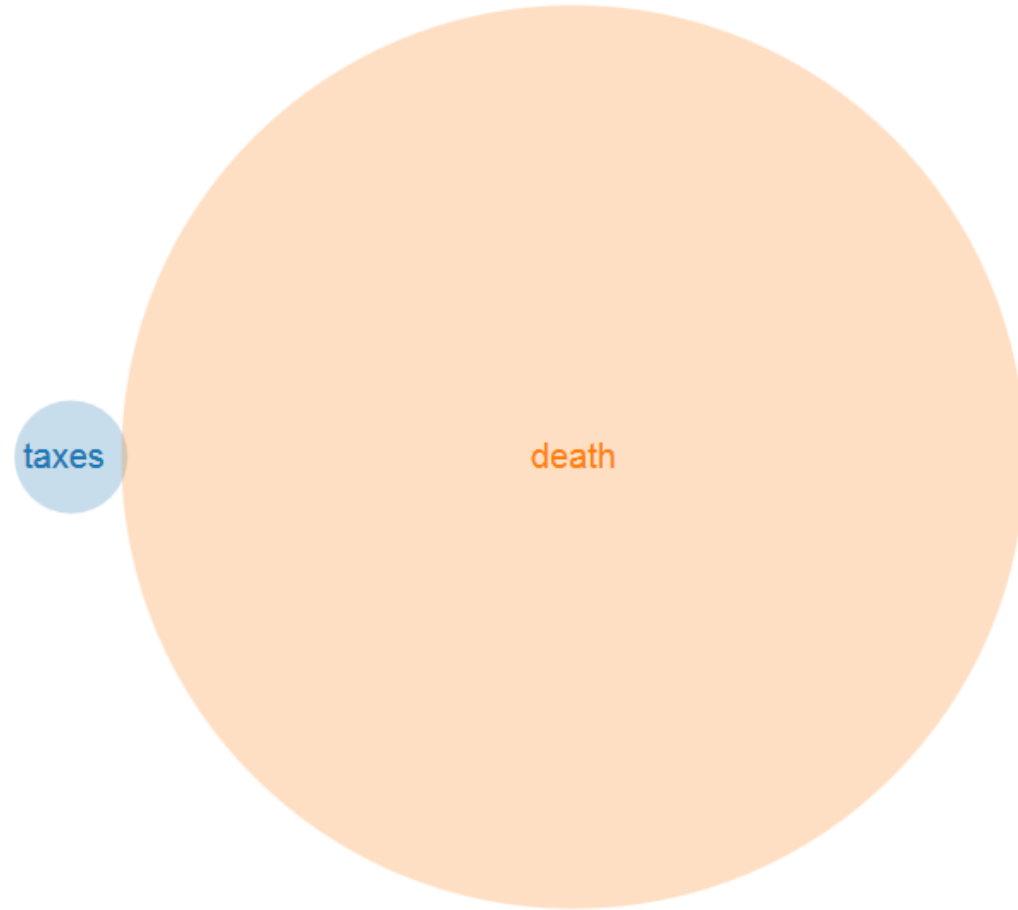
Librarians love Venn diagrams!

Another way of seeing beyond numbers

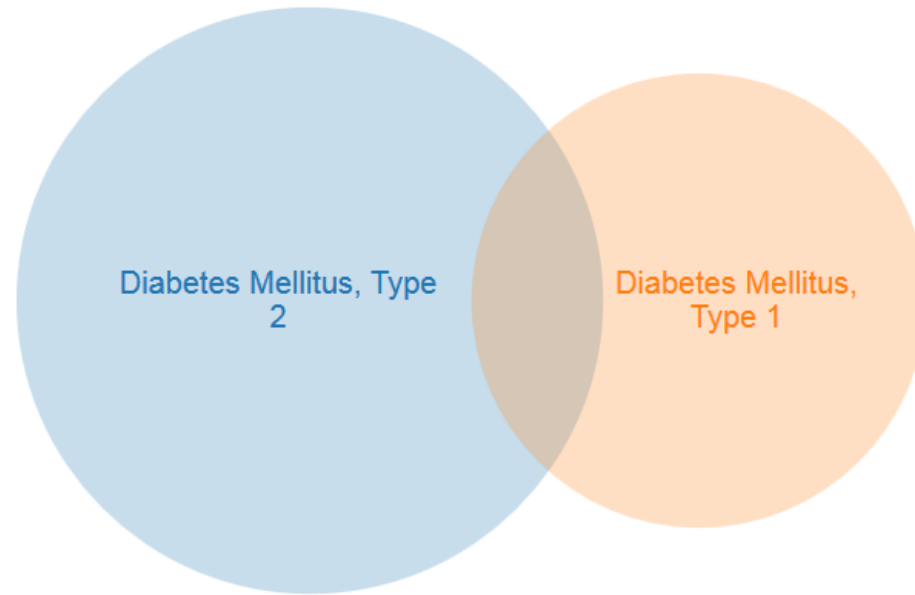
Does require a little [d3](#) (with the help of [venn.js](#))

<https://pubvenn.appspot.com/>

death and taxes



Diabetes Mellitus, Type 2 AND Diabetes Mellitus, Type 1



Questions?

<https://esperr.github.io/visualizingpubmed/>
esperr@uga.edu