

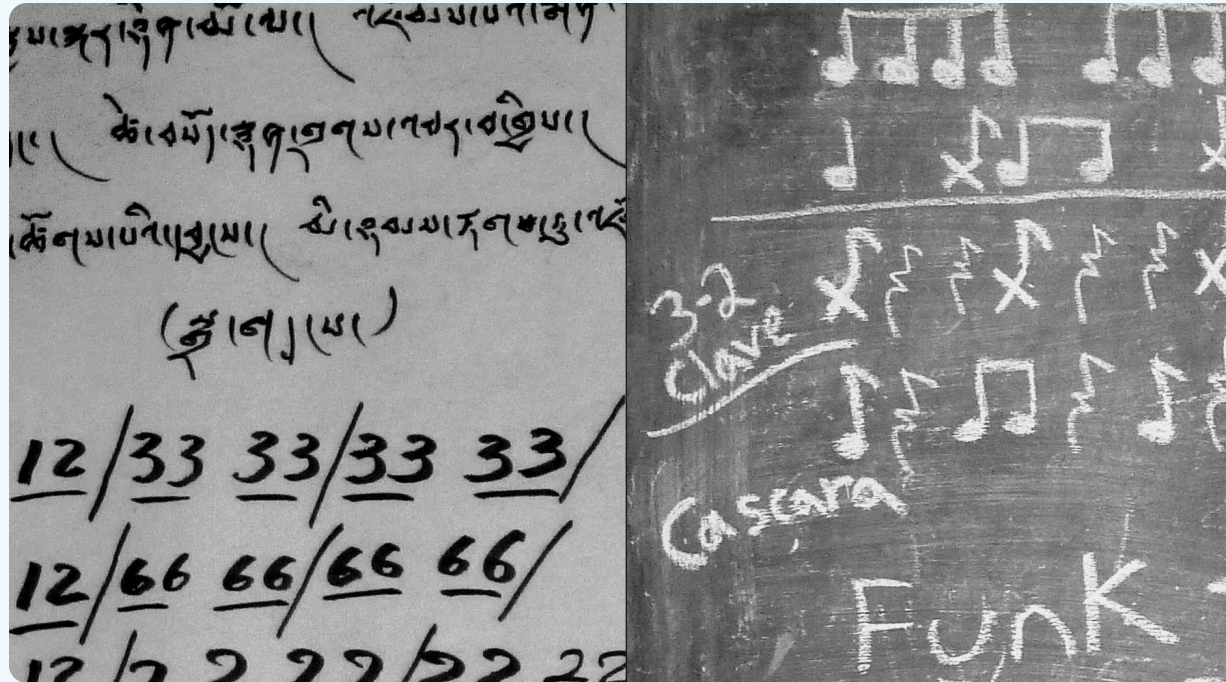
Scholarship Committee Meeting

Friday 8:09 AM

As I watch these presentations. I think I'm most interested in yours because I still don't entirely understand what your doing other than it sounds cool lol.

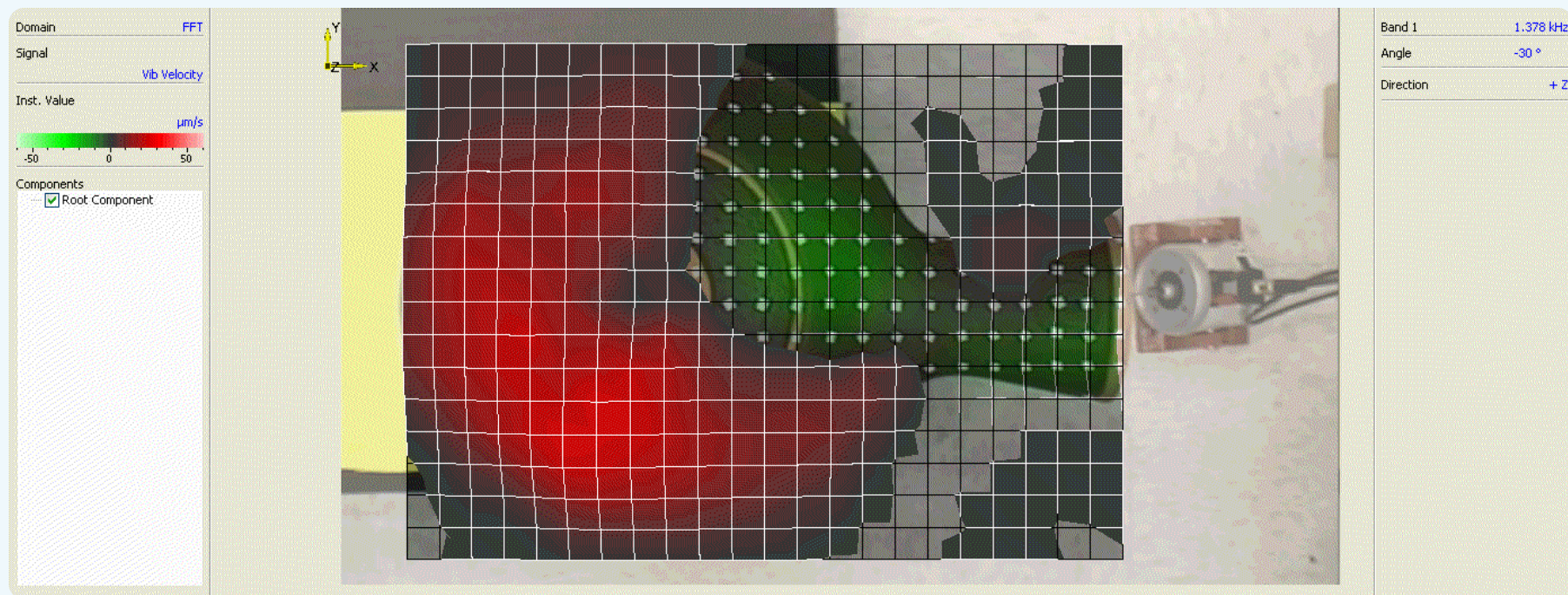
lol that makes two of us

Background



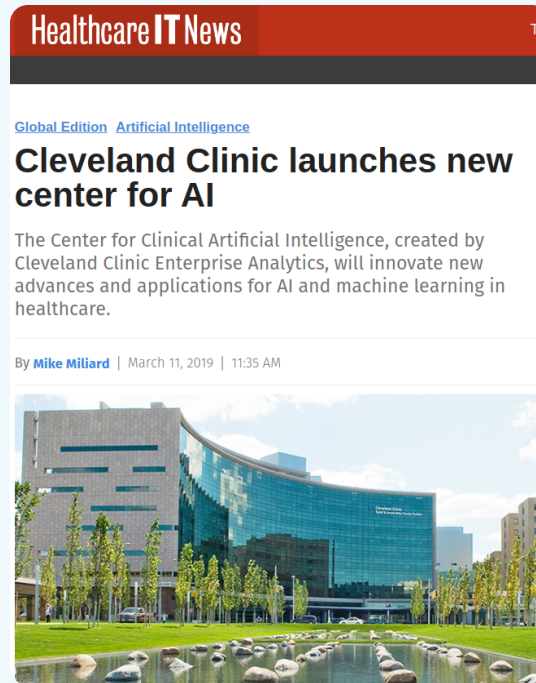
undergrad = anthropology + musicology + activism == ethnomusicology

Background



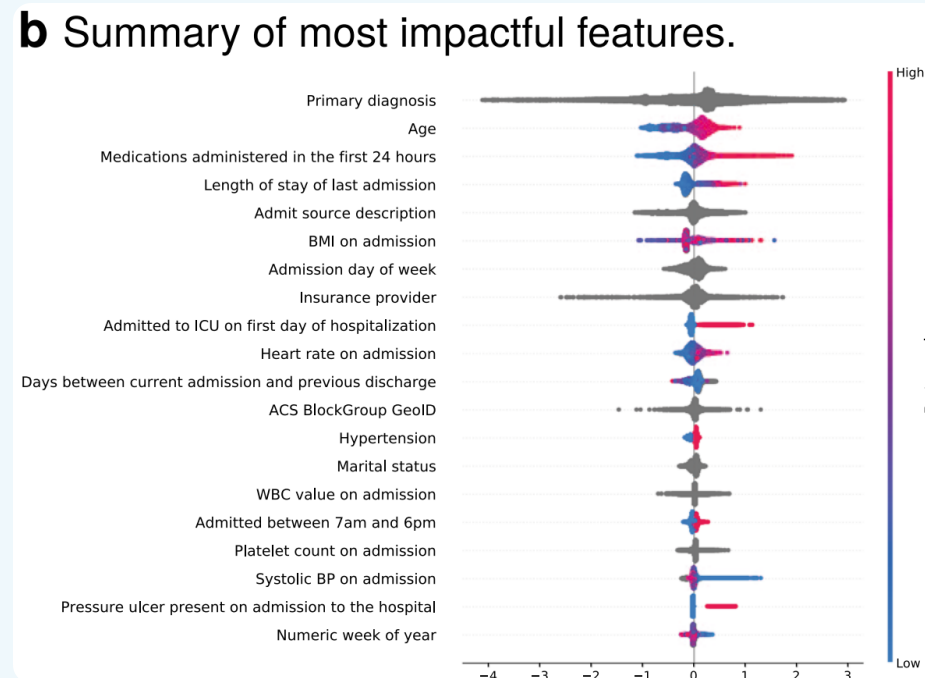
ethnomusicology + lasers + MATLAB == ?

Background



automated quantitative differential diagnosis, explainable ML

Background



hello anthro, my old friend

Career goals

clinical oncologist

head/neck/esophagus

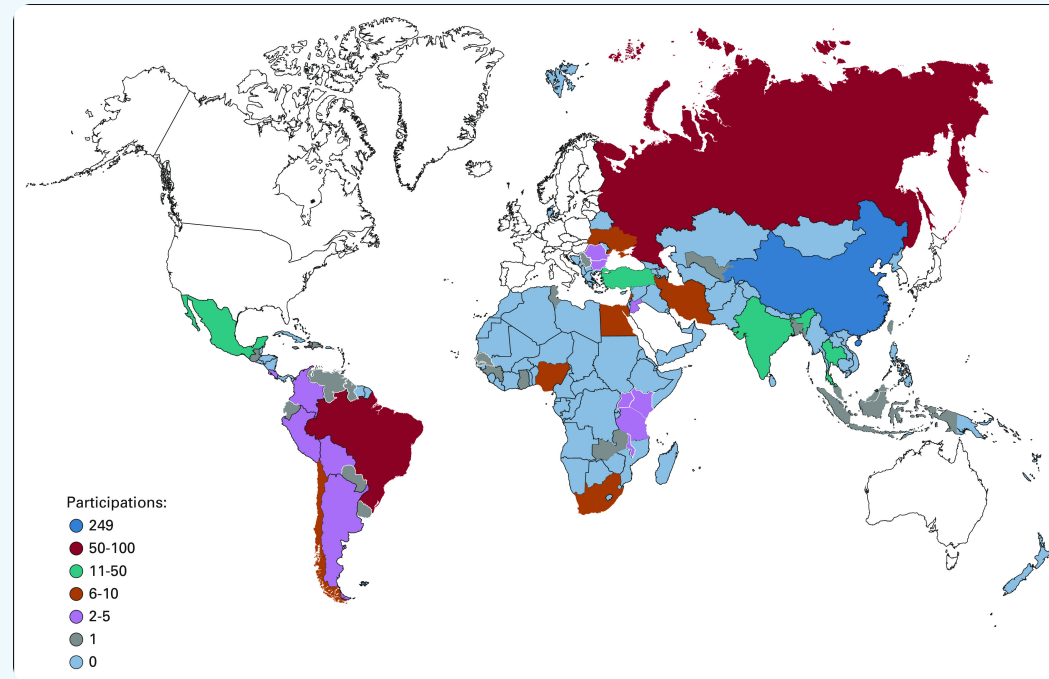
data scientist

computational anthropology, metascience

(☐ + ☐)

Current project
([click here](#))

Where in the world is hematology?



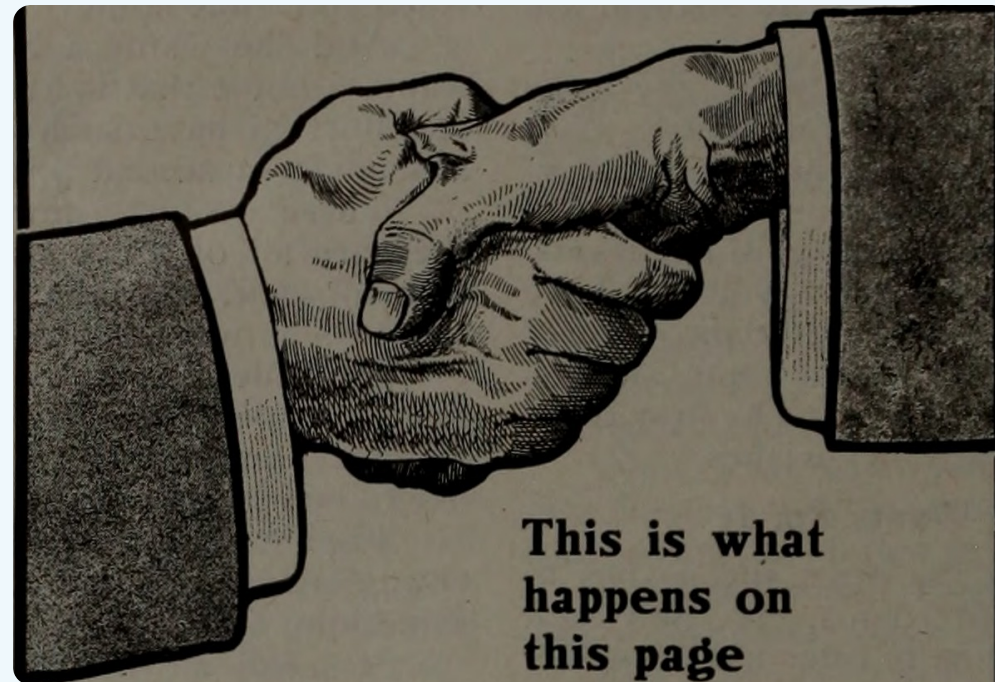
~5000 abstracts reviewed, ~500 had author from LMICs

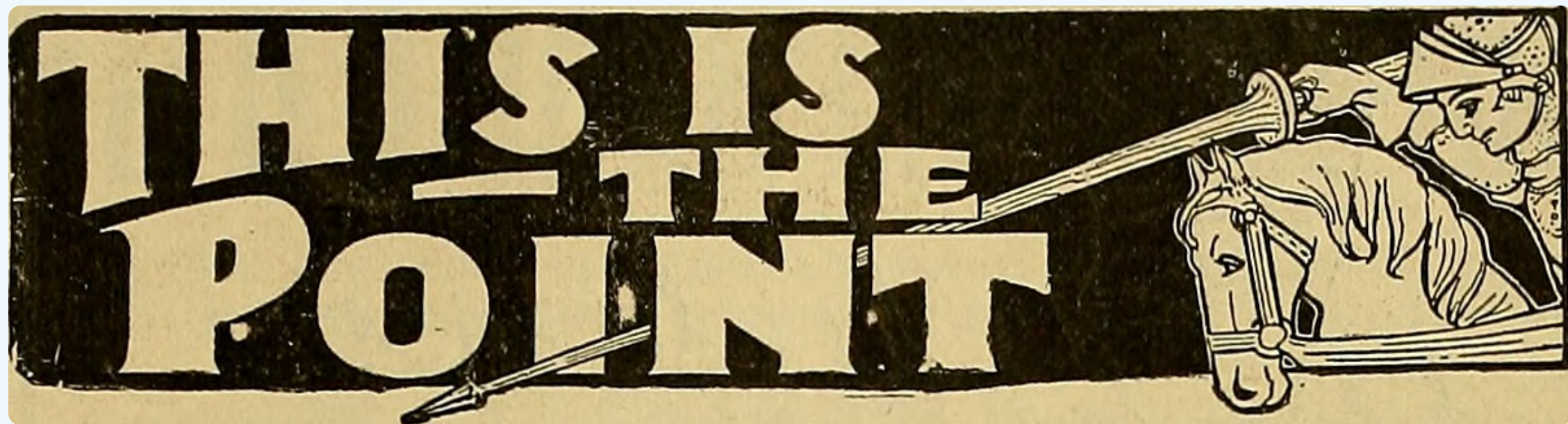
Right [patient, drug, time, dose, route, *place*]



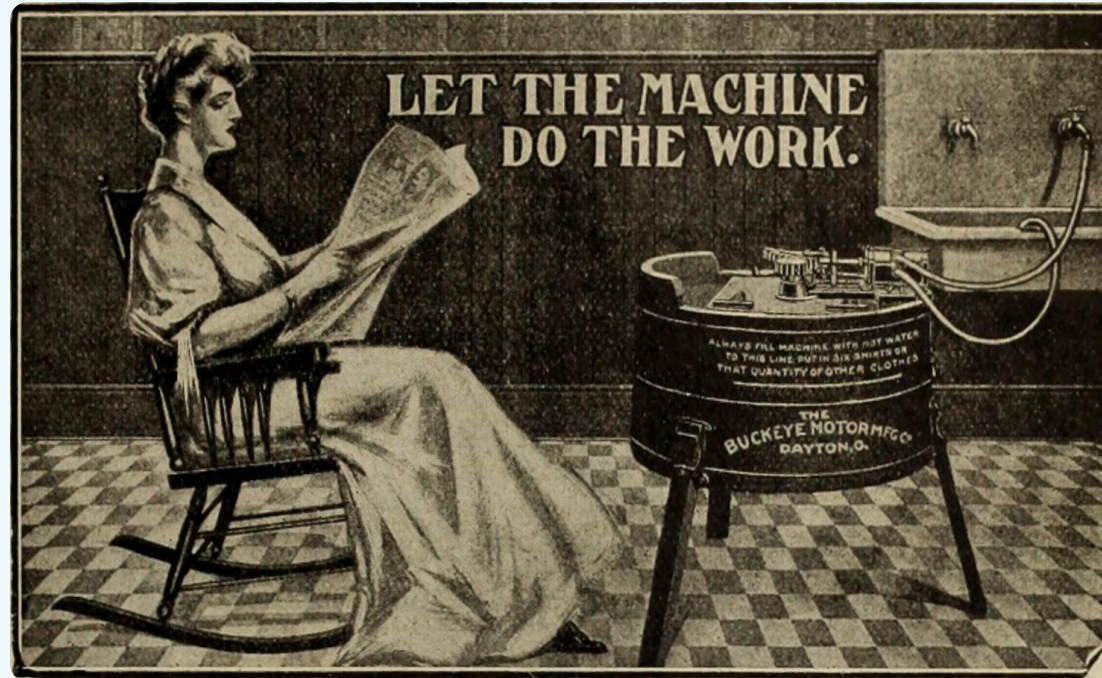
Léon Tshilolo, MD, PhD. ASH 2018, speaking on hydroxyurea in Sub-Saharan Africa.

Myopia



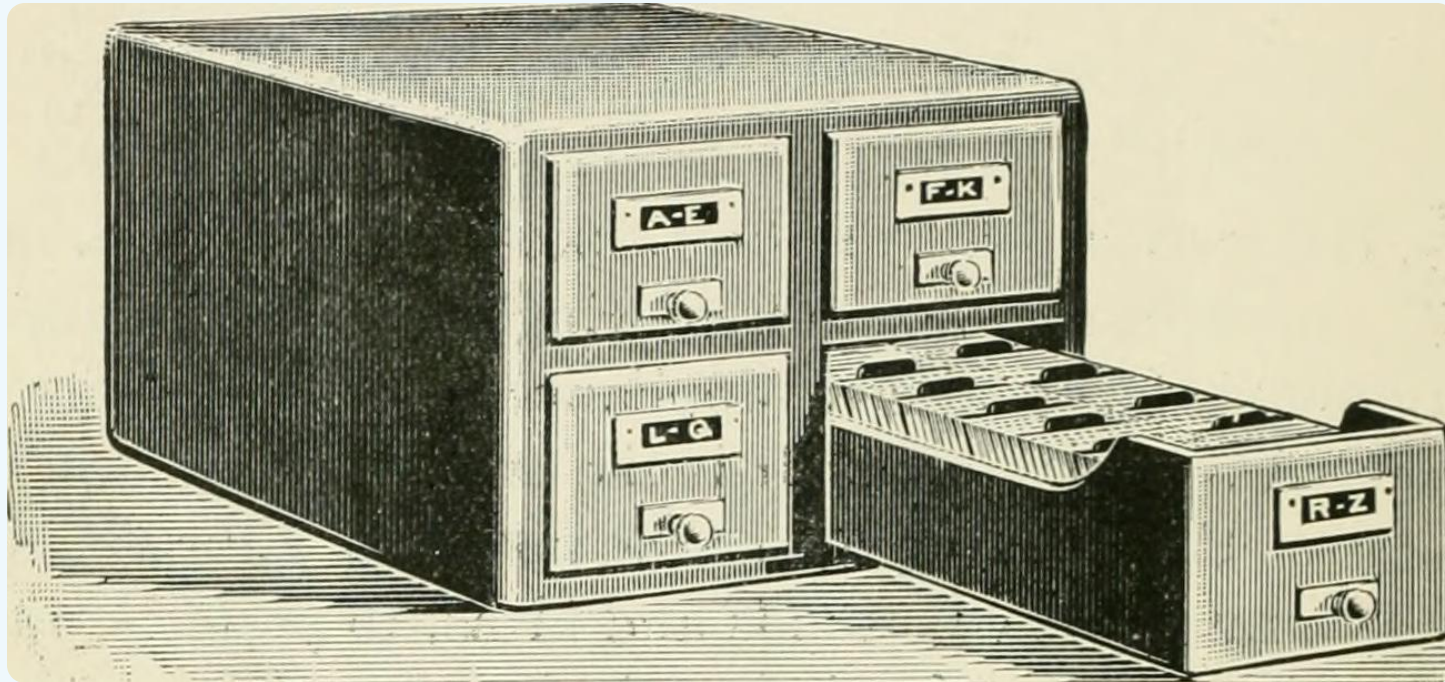


Current project



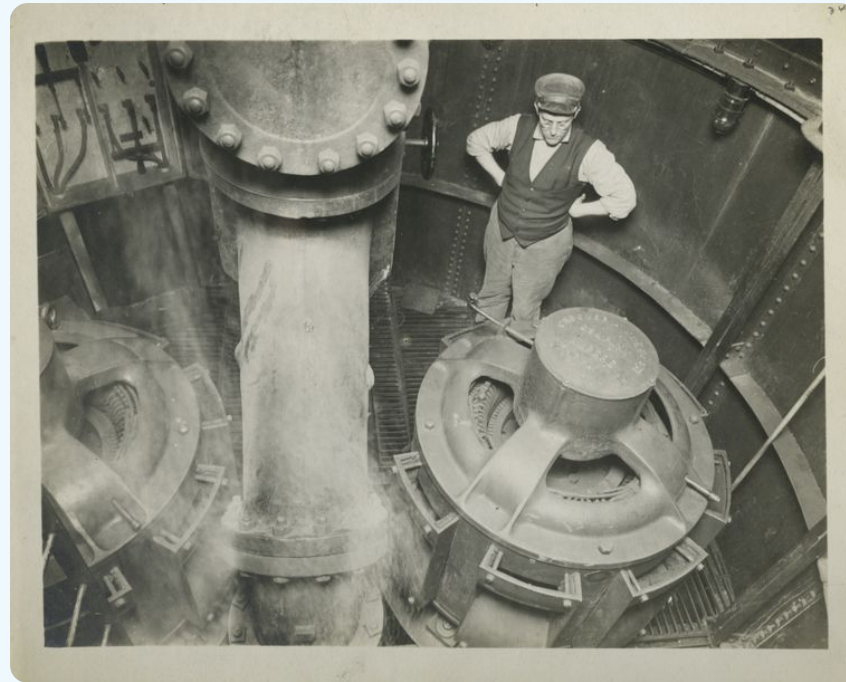
automate the extraction and structuring of all ~100k abstracts

Current project



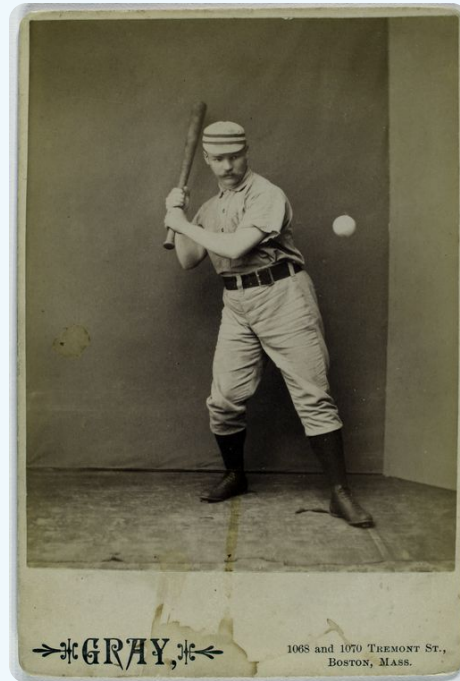
keep it public, keep it safe

Current project



analysis pipeline

Current project



BYOA (bring your own analysis)

Planned analyses

trends

geographic

chronologic

topologic

ontologic

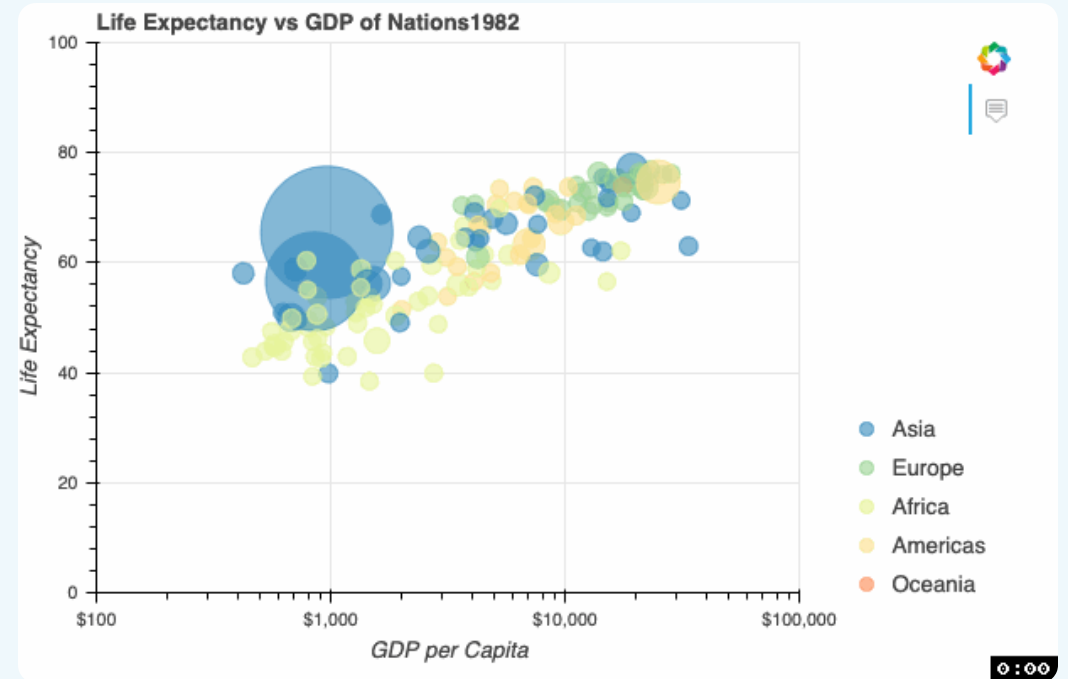
biographic

disparities

e.g. overlay maps of the disease
itself (incidence, morbidity, etc.)
and research about that disease

funding/resource allocation

link with National Clinical Trial



h/t Hans Rosling

Next steps

finalize data structure

rewrite web scraper

redundant data storage

analyses (interactive)

automatic updates

abstract submission to ASCO (Feb 2024), ASH
(Aug 2024)

manuscript (mid 2024)

Mentorship committee

Clinical Mentors

Dr Jennifer Choe (HN)

Dr Mike Gibson (HN/UGI)

alternate clinics

Research Mentors

Dr Travis Osterman (primary, local, EMR,
structured pt journeys)

1:1 meetings at least monthly

Dr Jeremy Warner (second primary, abroad,
HemOnc.org, ontology)

1:1 meetings at least q2mo

Other current projects

multiomics for sarcoma (PI: Osterman)

extreme biometry database (validated extremes in lab and radiographic parameters, to be used for data cleaning and exploration)

Future directions in fellowship

Goal 1: Develop clinical expertise in HN/UGI

Goal 2: Develop funding sources

early/training awards - e.g. [T15](#)

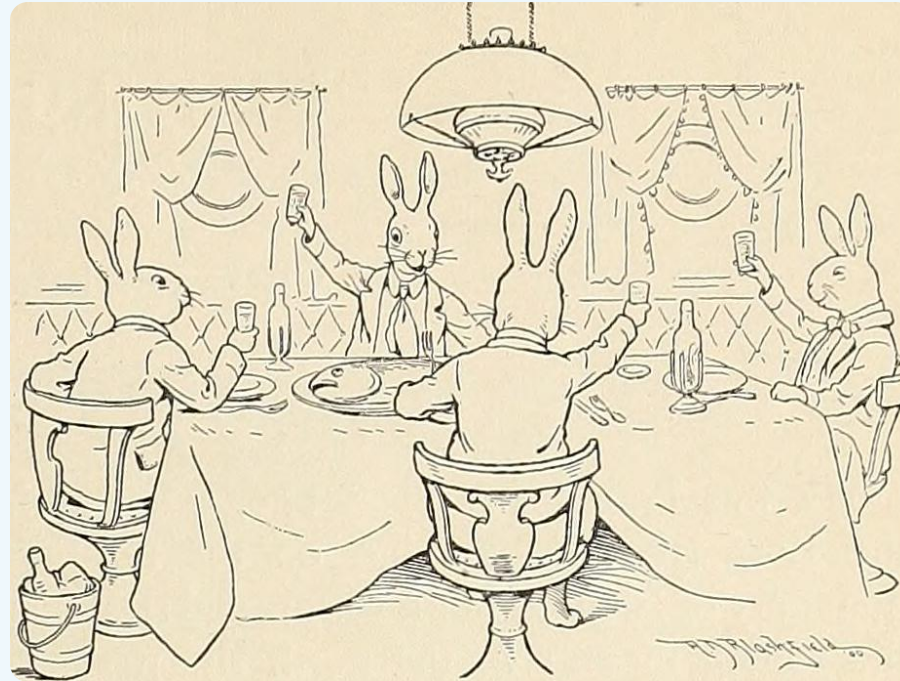
later awards - e.g. [ITCR](#) U01, U24, etc.

Links to relevant, current available FOA:

[Enhancement & Dissemination](#), [Sustainment](#),
[Biomedical Data Repository](#), [Biomedical Knowledgebase](#)

Goal 3: ?Pursue clinical informatics fellowship

Discuss!



moar slides follow, use prn

Challenges

Structure, standards

<https://schema.datacite.org>

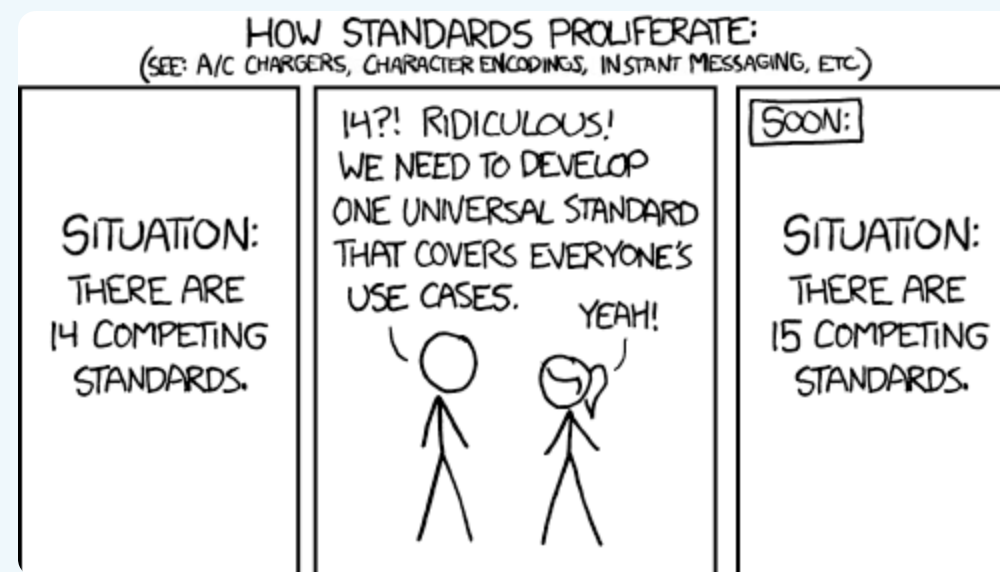
FAIR

Findable

Accessible

Interoperable

Reusable



Challenges

robots.txt

<https://ashpublications.org/robots.txt>

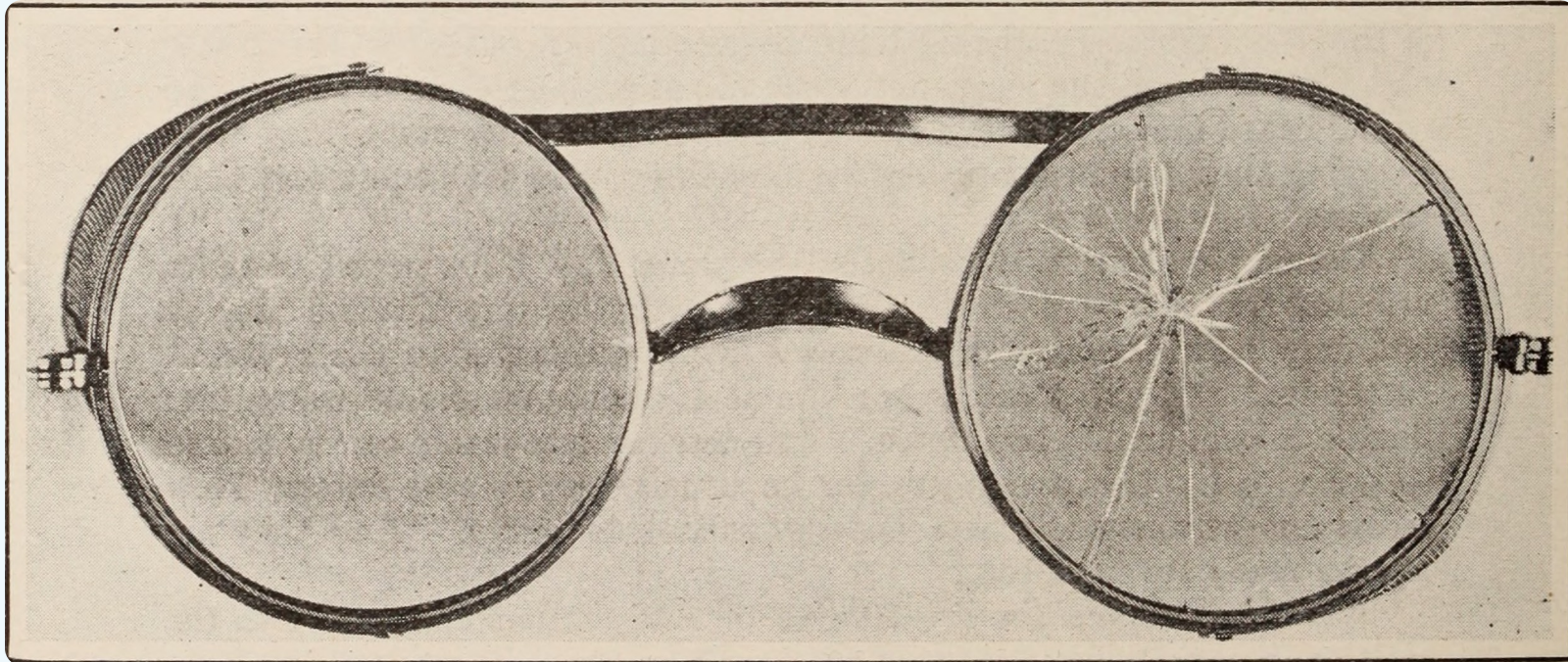
cf hiQ vs LinkedIn

```
< > M M C A https://ashpublications.org/robots.txt

User-agent: 008
Disallow: /

User-agent: *
Disallow: /bin/
Disallow: /config/
Disallow: /app_data/
Disallow: /d11s/
Disallow: /App_Themes/
Disallow: /signin.aspx
Disallow: /searchresults.aspx
Disallow: /advancedsearch.aspx
Disallow: /DownloadFile/
Disallow: /DownloadImage.aspx
Disallow: /Citation/Download
Disallow: /Shibboleth.sso/
Disallow: /SignInShibboleth.aspx
Disallow: /lookup/external-ref
Disallow: /user/logout
Disallow: /cgi/pmidlookup
Disallow: /Themes/Client/app/img/favicons/manifest.json
Disallow: /lookup/google-scholar
Disallow: /store/*
Allow: /cassette.axd/stylesheets/
Allow: /cassette.axd/script/
Allow: /cassette.axd/manifest.xml
Disallow: /*.axd
Disallow: /DownloadFile/DownloadImage.aspx
Disallow: /Citation/Download
Disallow: /Shibboleth.sso/SLO/Redirect
Disallow: /sign-in
Disallow: /my-account
Disallow: /specificToClient/
Sitemap: https://ashpublications.org/data/sitemap/Site_1/sitemap_books_sitel.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1/sitemap_gbos_sitel.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000001/sitemap_gbos_sitel1000001.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000002/sitemap_gbos_sitel1000002.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000005/sitemap_gbos_sitel1000005.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000007/sitemap_gbos_sitel1000007.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000009/sitemap_gbos_sitel1000009.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000011/sitemap_gbos_sitel1000011.xml
Sitemap: https://ashpublications.org/data/sitemap/sitemap.xml
```

Challenges



missing, incomplete, wrong

Challenges



Web scraping is the bug. Not sure who the other guy is.

Challenges (also a feature)

home / abstracts_small

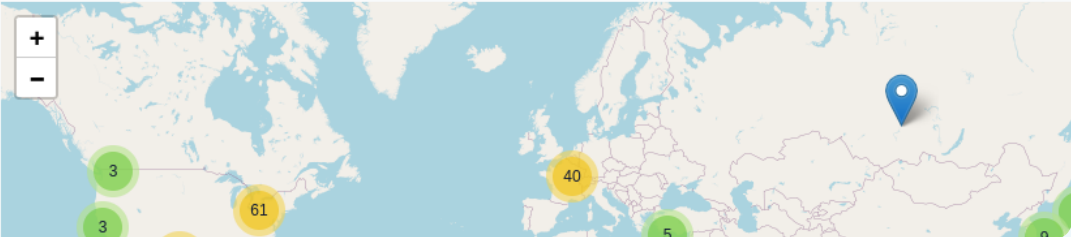
abstracts_small

Custom SQL query returning more than 1,000 rows ([hide](#))

```
1 select doi, article_title, article_date, session_type, topics,  
   author_names, author_affiliations, first_author_latitude,  
   first_author_longitude from abstracts order by doi limit 5000
```

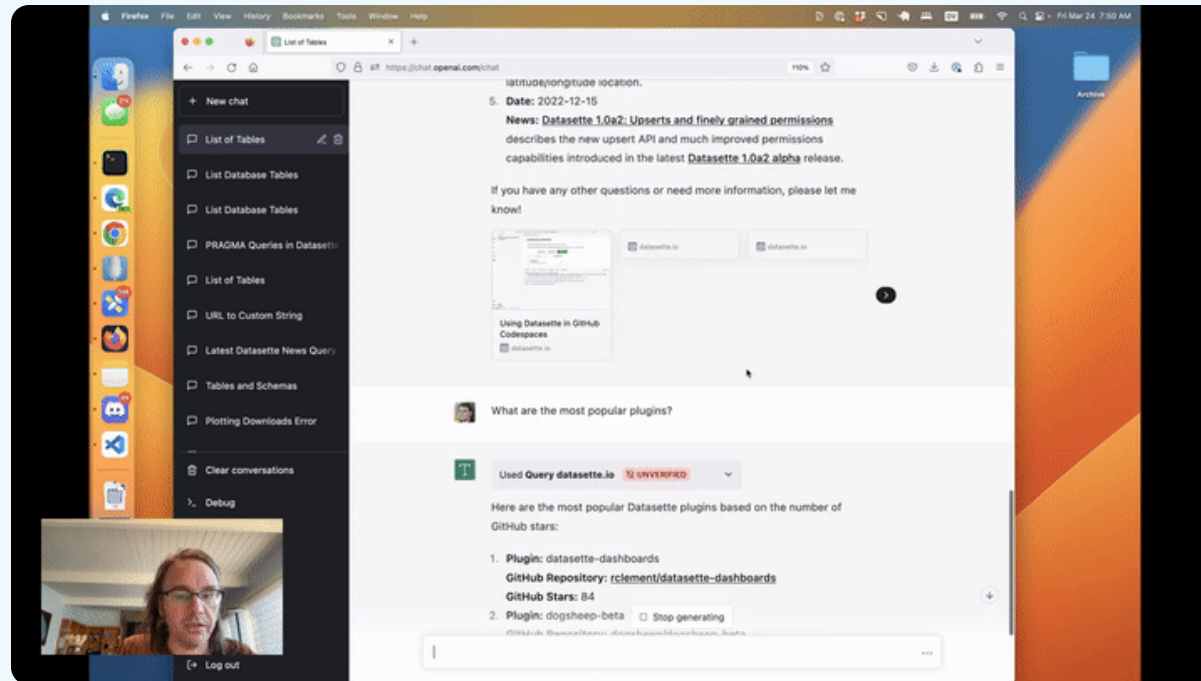
[Format SQL](#) [Run SQL](#)

This data as [json](#), [CSV](#)



If you can write SQL, you can analyze the data.

Challenges (also a feature)



ChatGPT to the rescue?