Google

Tools for open data

Filip Hráček

BigClean, Nov 2012

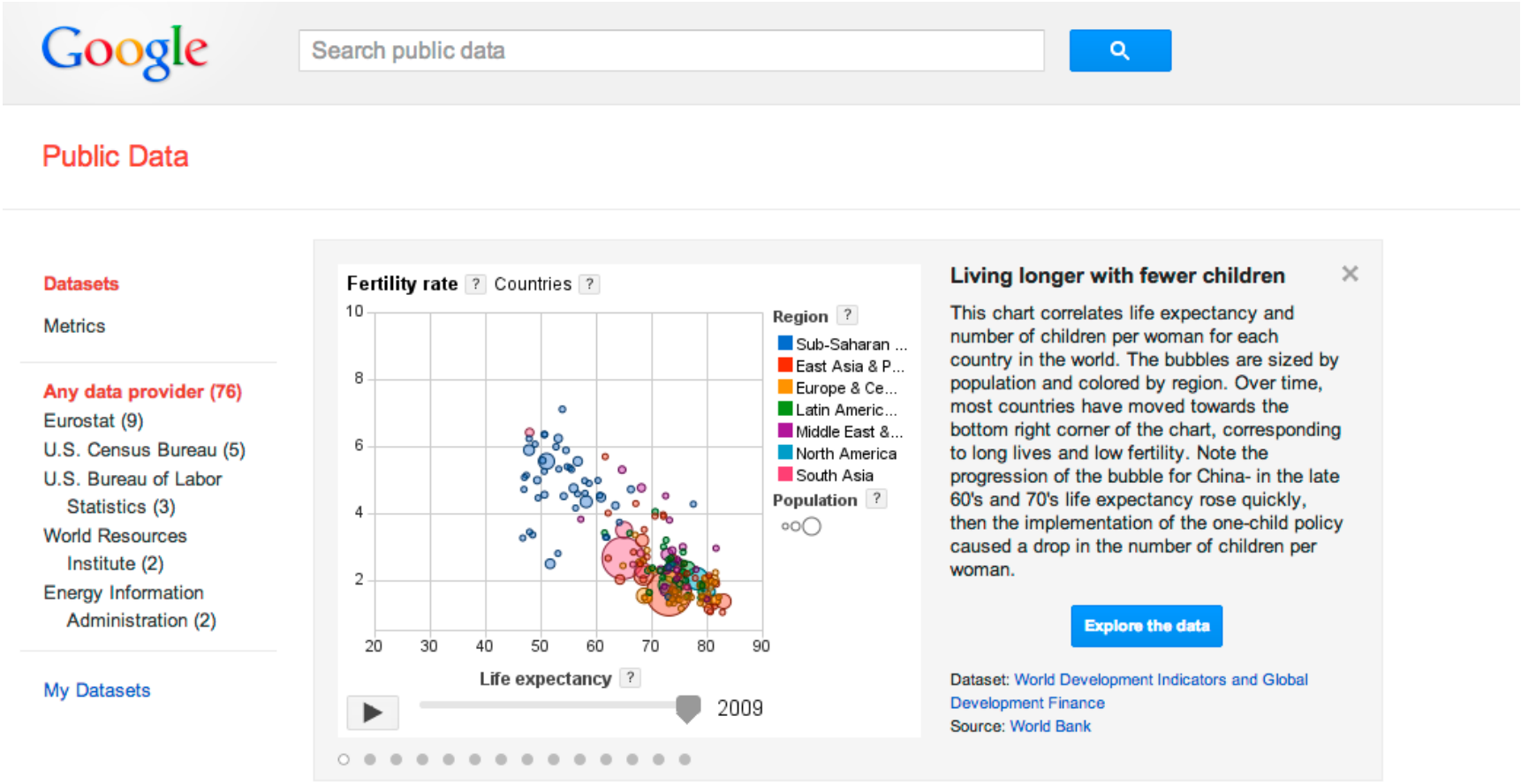
Big data is hard.

1. Data gathering – hard
2. Data refinement – hard
3. Data analysis – super hard
4. Data sharing – hard

Data gathering

Public Data Explorer

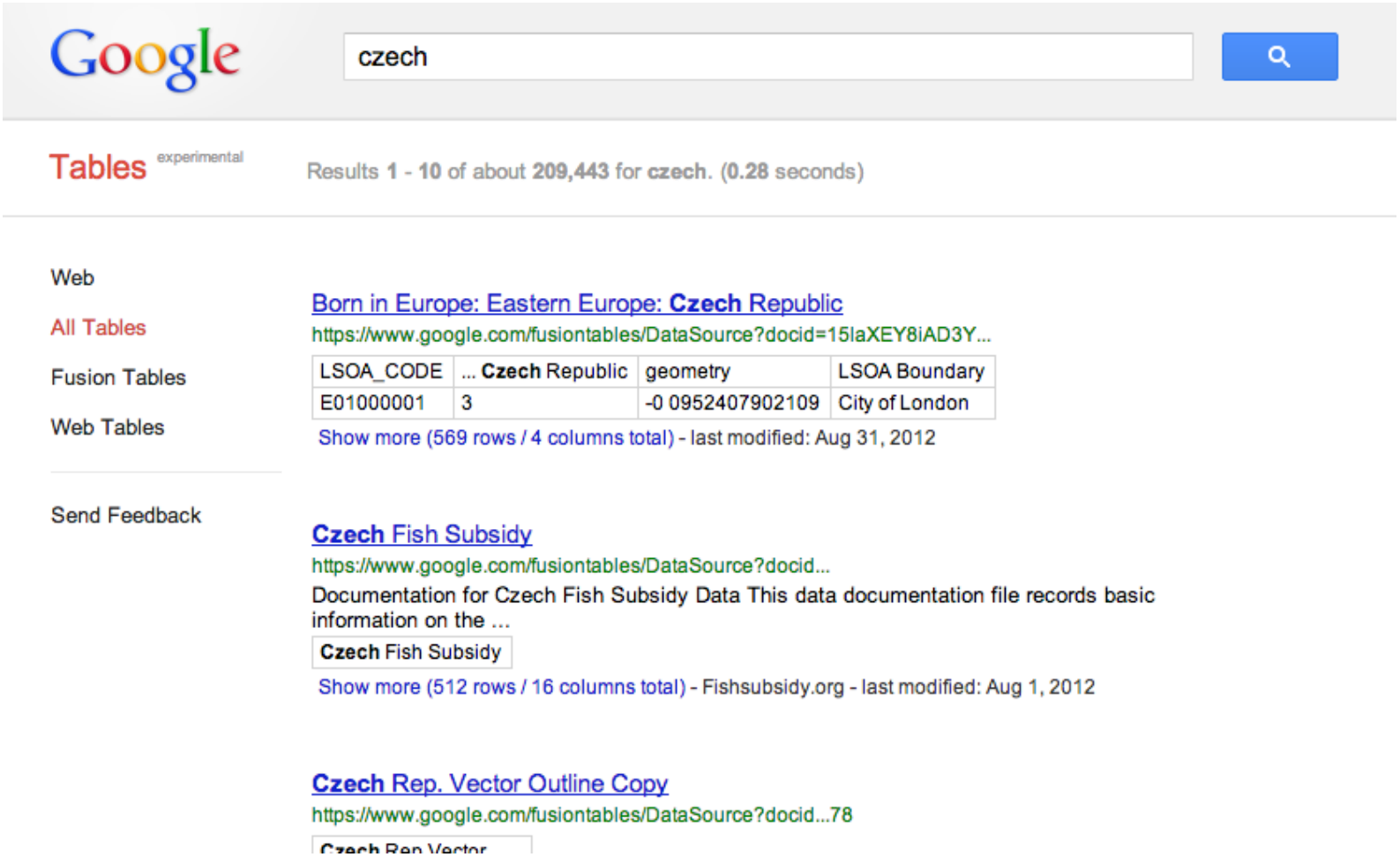
http://www.google.com/publicdata/directory



Google

Fusion Tables (Public)

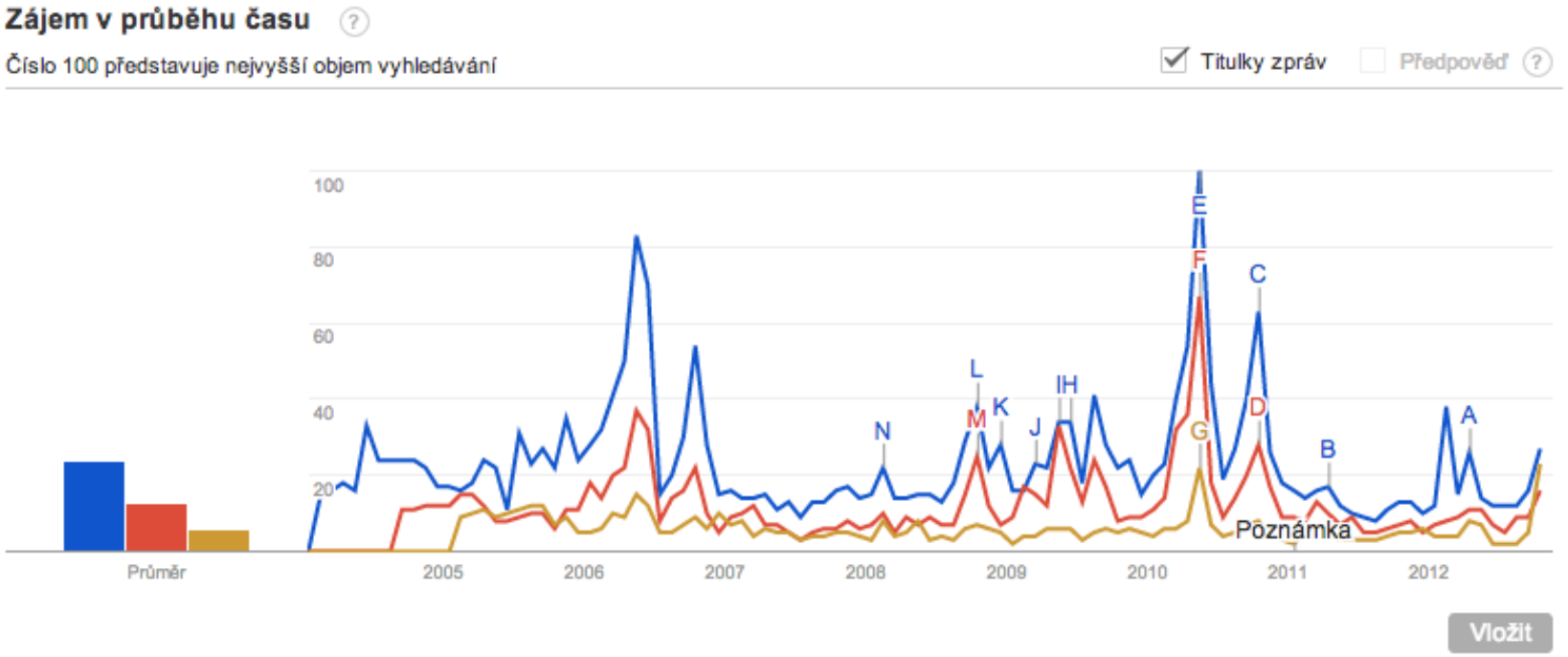
http://research.google.com/tables



Google

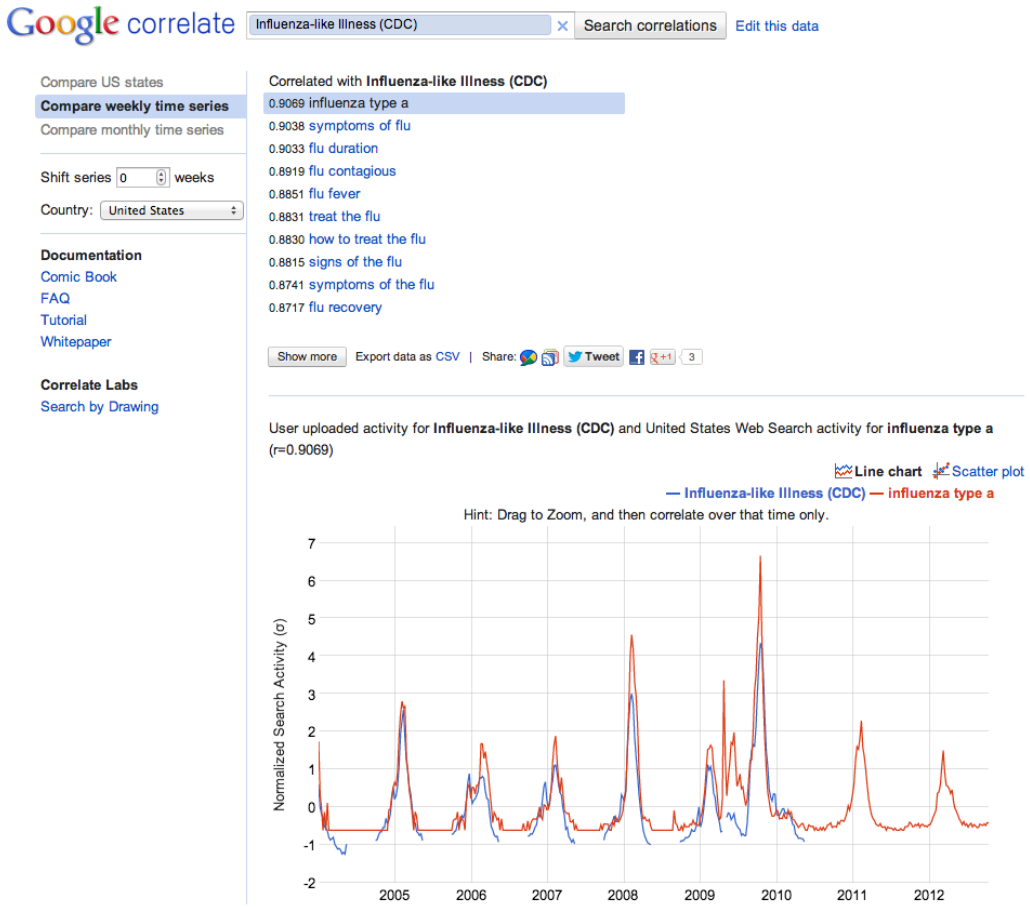
Google Trends

http://www.google.cz/trends



Google Trends Correlate

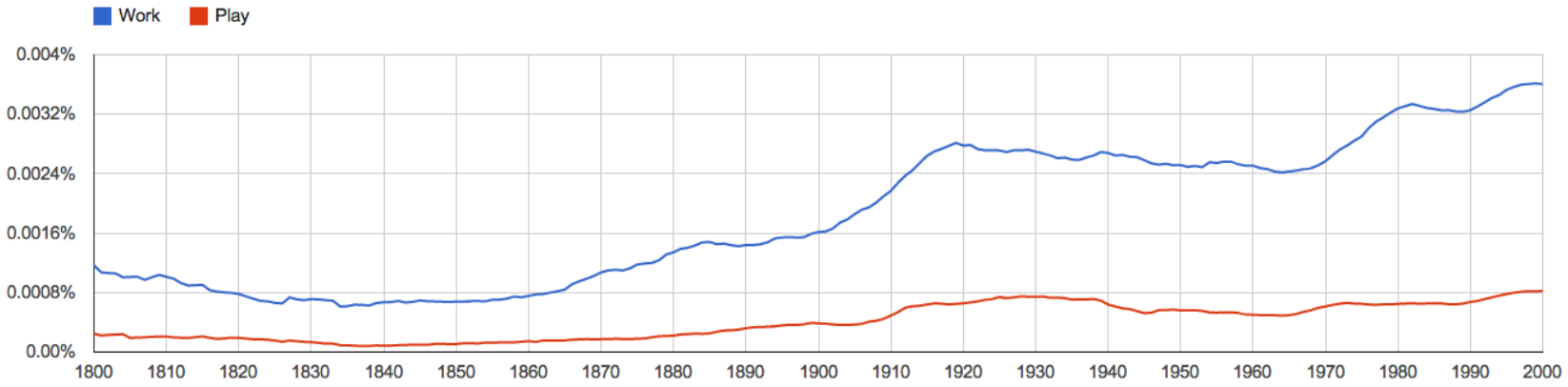
http://www.google.com/trends/correlate



Google

Google Ngram Viewer

http://books.google.com/ngrams

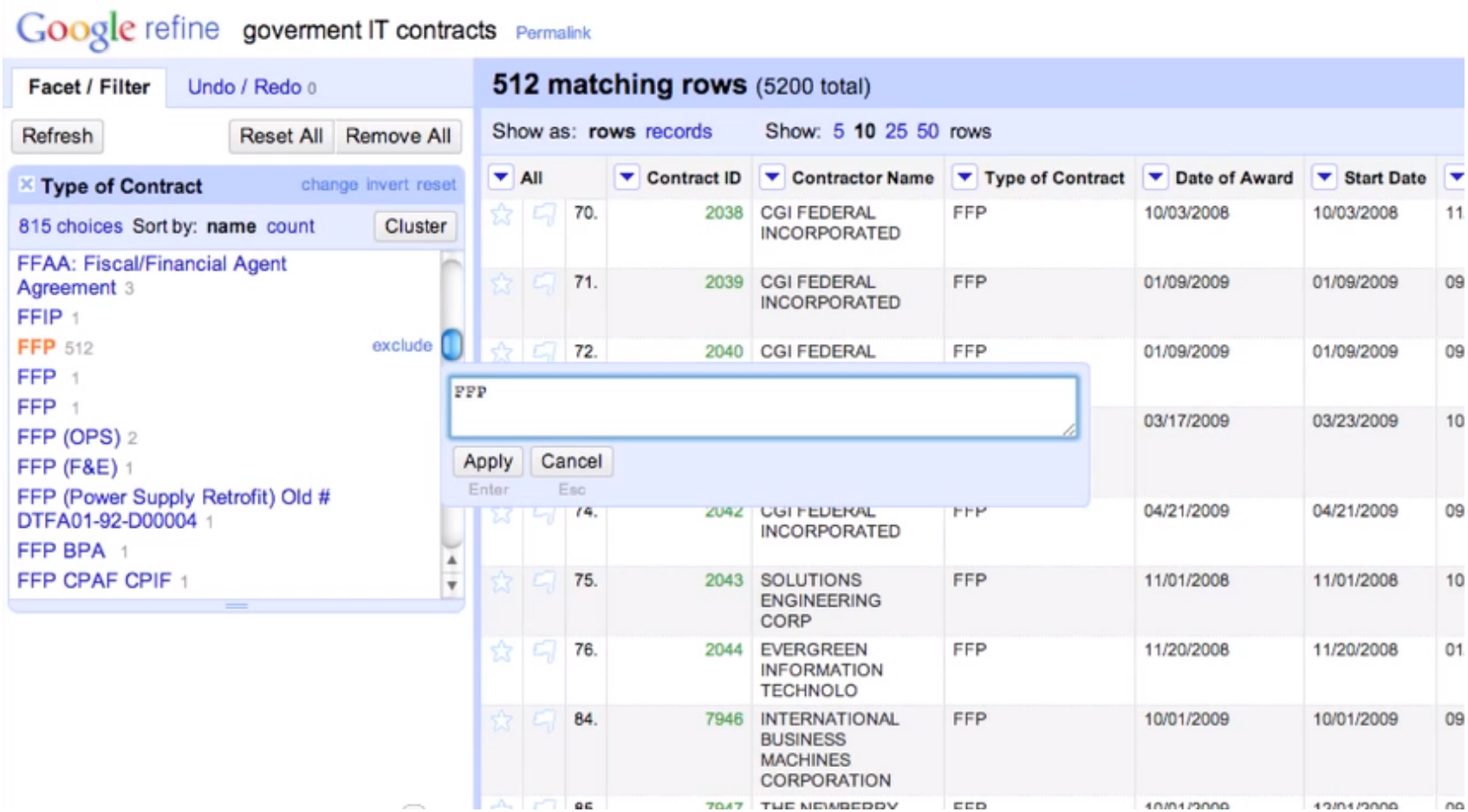


Data refinement

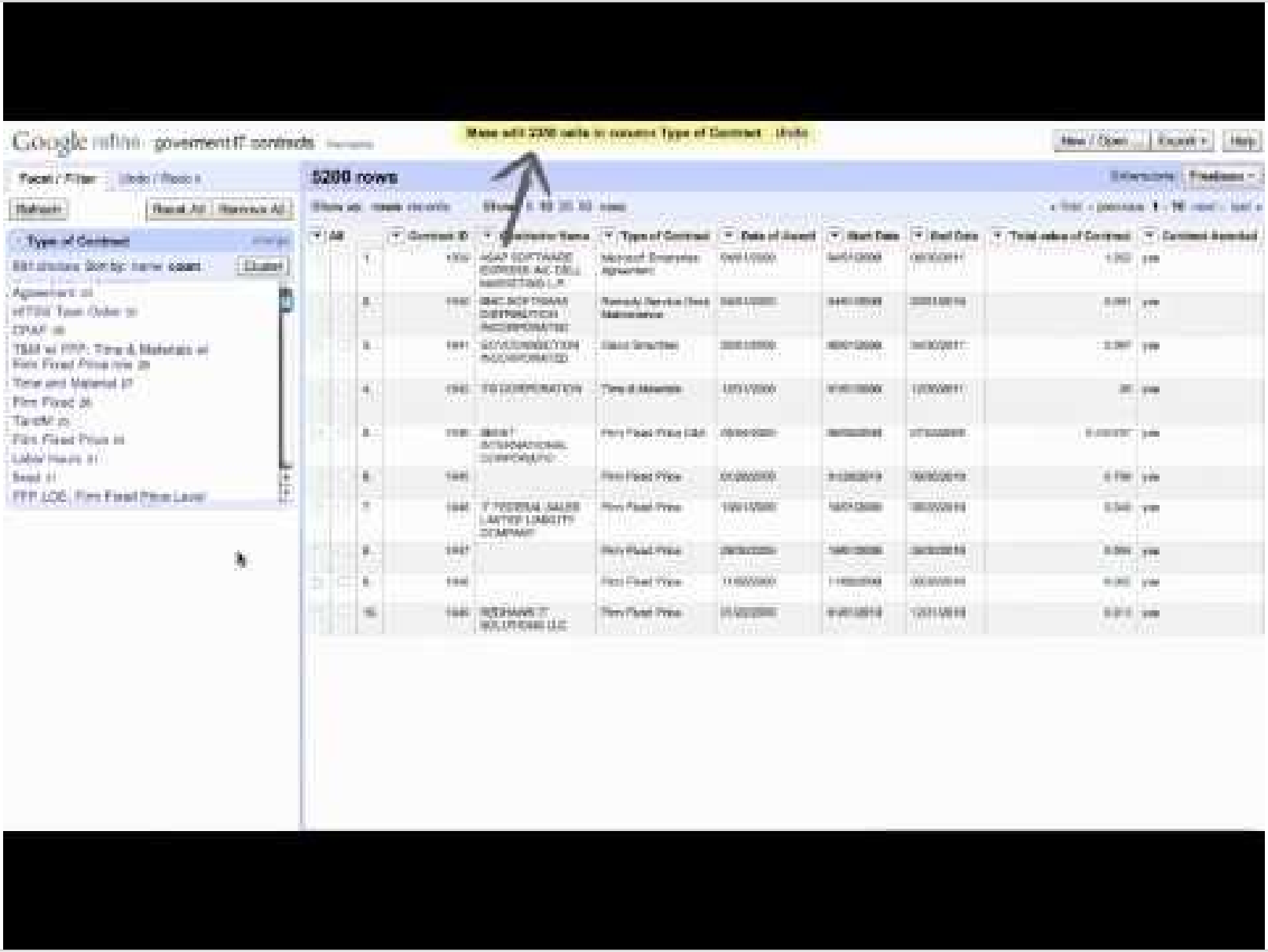
Google

Open Refine

http://code.google.com/p/google-refine/



Google


Google refine government IT contracts
Facet / Filter
Undo / Redo
Refresh
Reset All
Remove All
Type of Contract
Cluster
5200 rows
All
Contract ID
Contractor Name
Type of Contract
Date of Award
Start Date
End Date
Total value of Contract
Contract Awarded

Data analysis

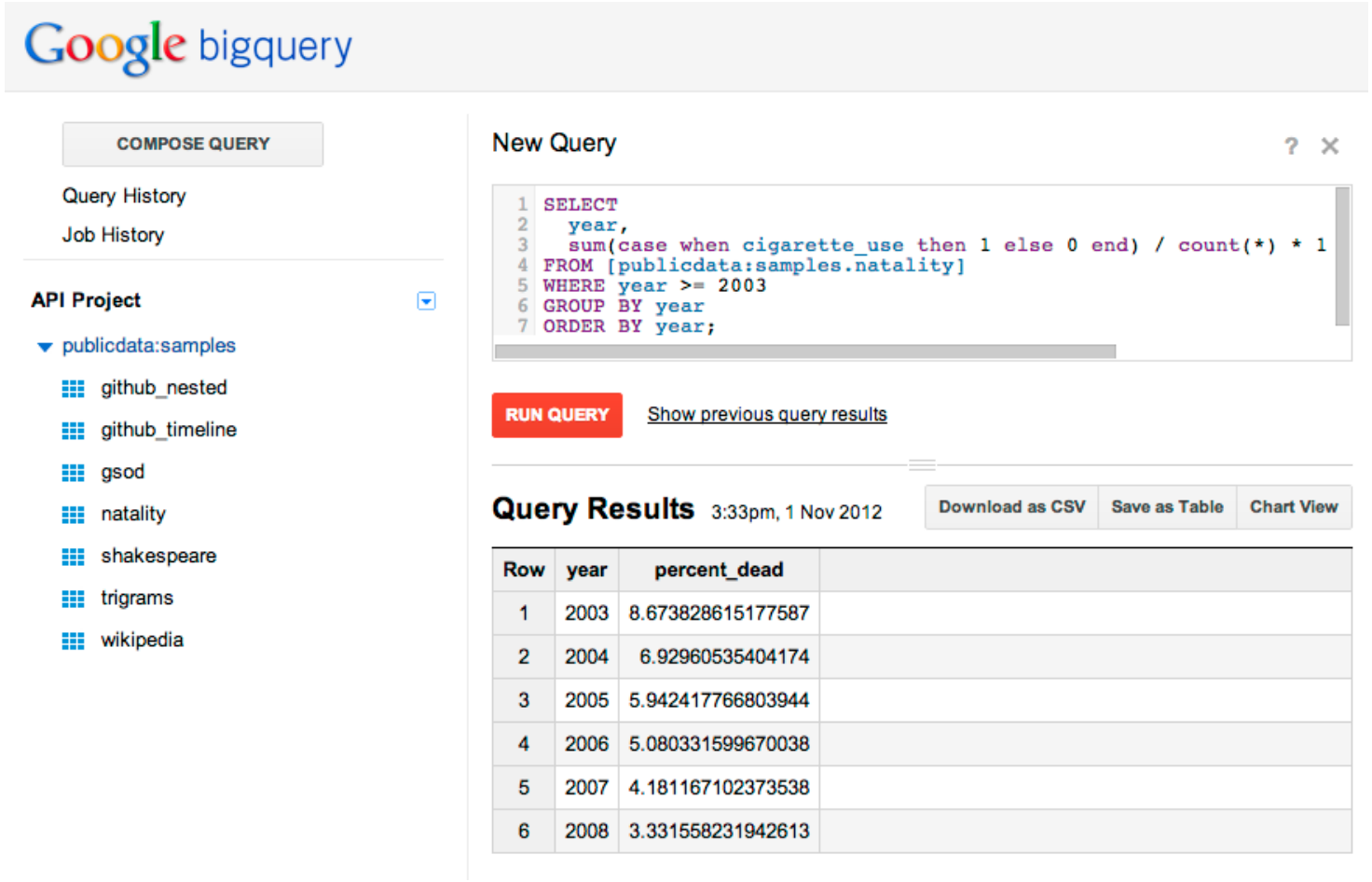
Big data.

1. ~~Text editor~~
2. ~~Excel~~
3. ~~Local database~~

Google

Google BigQuery

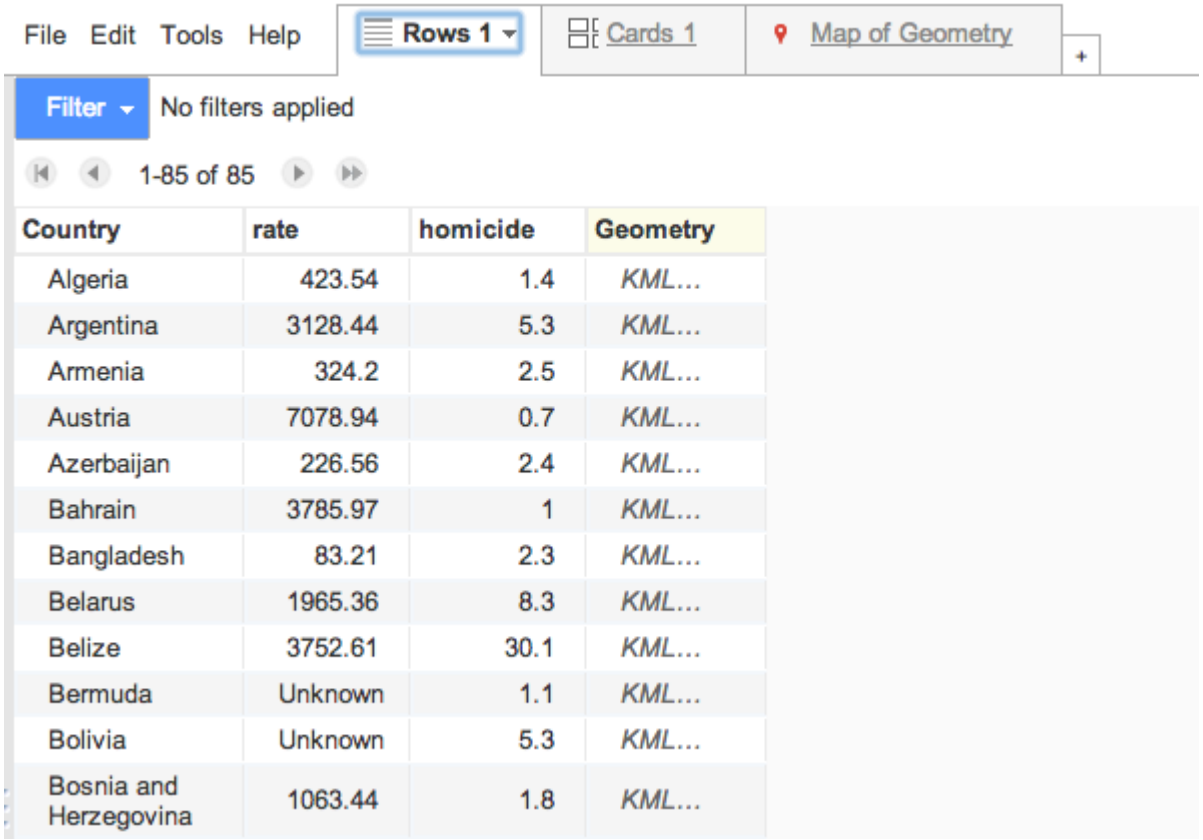
https://bigquery.cloud.google.com/



Google

Google Fusion Tables

http://www.google.com/fusiontables http://research.google.com/tables



Country	rate	homicide	Geometry
Algeria	423.54	1.4	KML...
Argentina	3128.44	5.3	KML...
Armenia	324.2	2.5	KML...
Austria	7078.94	0.7	KML...
Azerbaijan	226.56	2.4	KML...
Bahrain	3785.97	1	KML...
Bangladesh	83.21	2.3	KML...
Belarus	1965.36	8.3	KML...
Belize	3752.61	30.1	KML...
Bermuda	Unknown	1.1	KML...
Bolivia	Unknown	5.3	KML...
Bosnia and Herzegovina	1063.44	1.8	KML...

Data sharing

Google

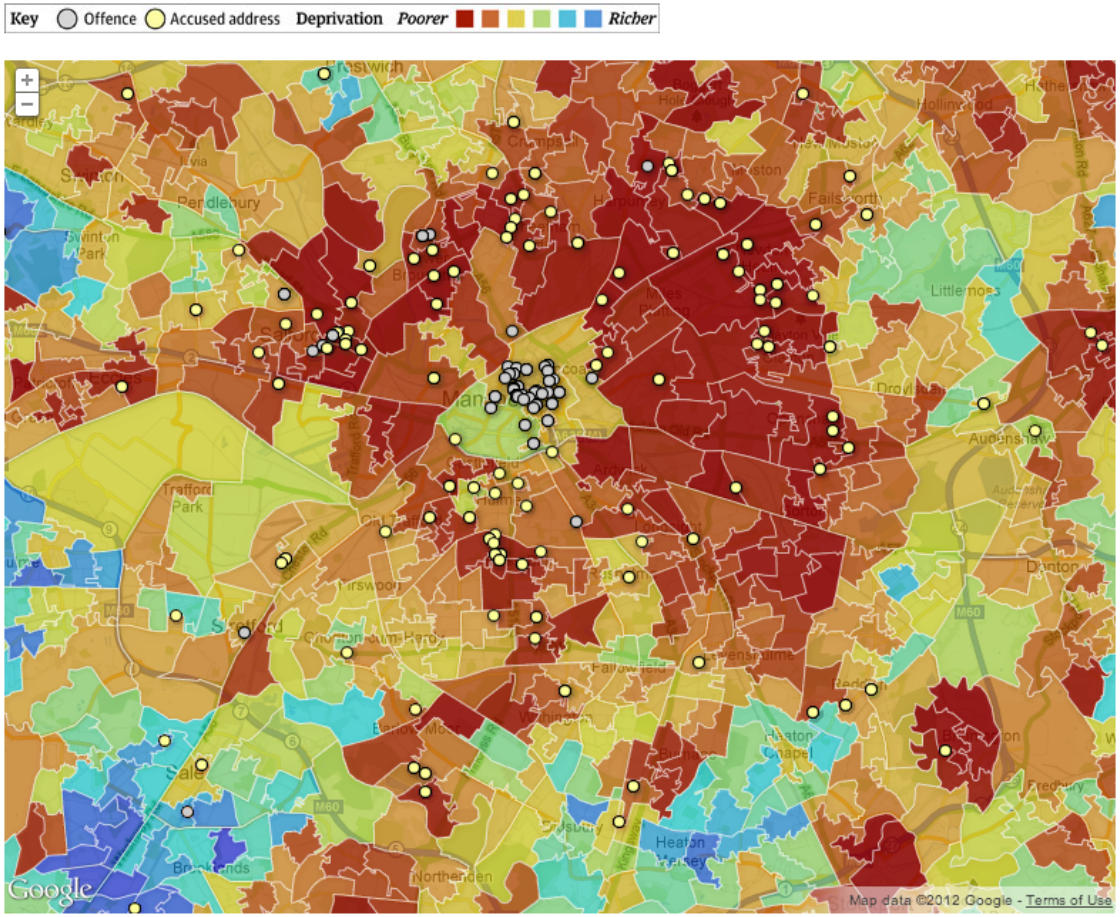
Google Fusion Tables

http://www.google.com/fusiontables

Google

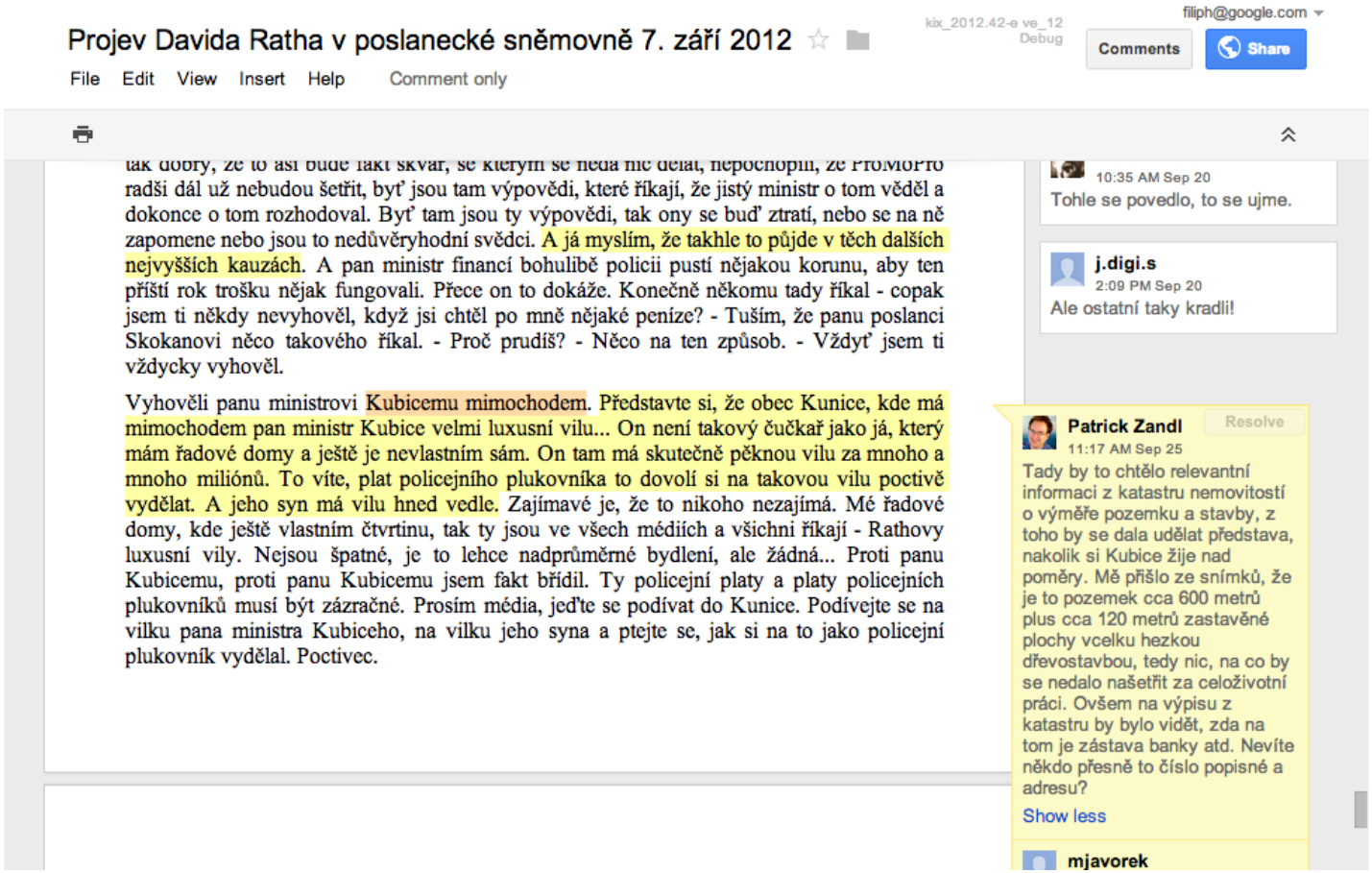
Google Fusion Tables – in the wild

http://www.guardian.co.uk/news/datablog/interactive/2011/aug/16/riots-poverty-map



Google Docs – in the wild

http://data.blog.ihned.cz/c1-57386250-hledejte-s-nami-fakta-v-projevu-davida-ratha



Google

Big thank you!

(for your attention)

Filip Hráček,
Google Czech Republic

Google

Links

- Public Data Explorer
  - Import public data form
- Google Correlate
  - Explanation (Comics)
- Ngram Viewer
  - Advanced use