Google Tools for Data

Hal Varian
16 July 2012

Google

Outline

Tools

Docs: documents, spreadsheets, presentation

Refine: data cleansing

Fusion Tables: database and visualization

Google Scholar

Graphics and APIs

Motion charts: charting

Prediction API: machine learning for data in cloud

Chart Tools: interactive charts for web pages

Data

Insights for Search: query data time series

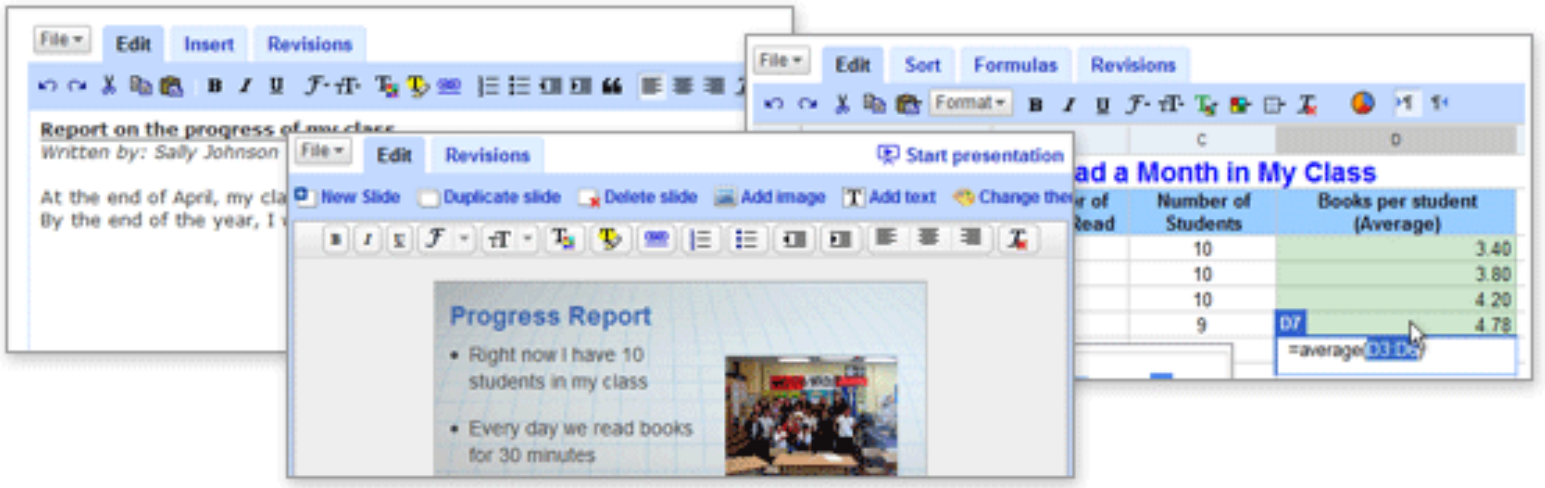
Correlate: query correlations

Ngrams: phrase occurrences in books

Public Data Explorer: find and examine public data

Google Consumer Surveys: run surveys for 10 cents/response

Google Docs



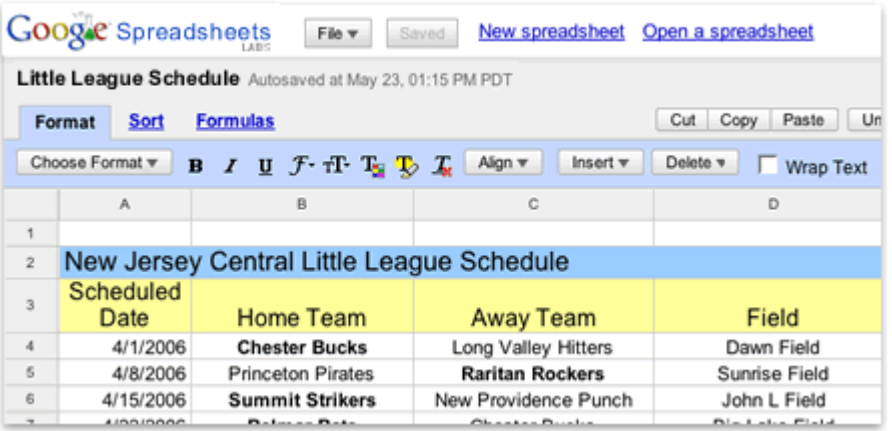
- Integrated suite of word processor, spreadsheet, slides, drawings
- Available in the cloud for multi-authored document creation

Documents

- Multiauthored documents in cloud
  - Version control
  - Access control
  - Math and drawings
- Move document back and forth between desktop and cloud
- Execute code directly from the cloud
  - LaTeX docs
  - R scripts

Spreadsheets

- Multiauthored calculations in the cloud
  - Collect data from multiple users
  - Collect data from Gmail forms
- Move from desktop to cloud and back
- Access control
- Interactive visualization
- Import data from spreadsheet in R
  - Example of loading R data

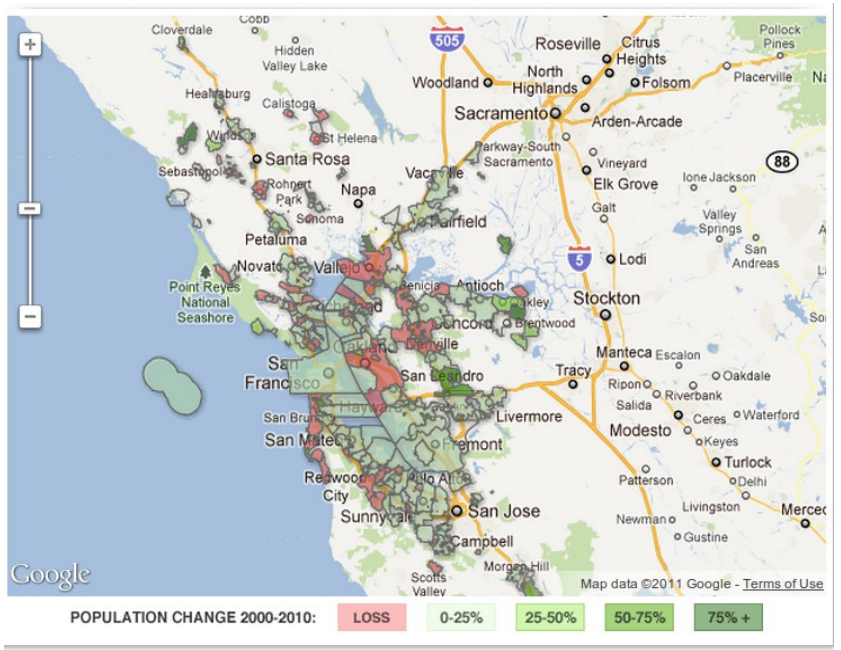


Presentations

- Multiauthored presentations in the cloud
- Move from desktop to cloud and back
- Access control
- Interactive visualization
- Version control

Fusion Tables

- Capabilities
  - Visualize and publish your data as maps, timelines and charts.
  - Host your data tables online.
  - Combine data from multiple people.
  - Export and import data
- Database in the cloud
  - Access control
  - Annotation



Refine

- Tool for working with messy data
  - Cleaning and normalizing data based on fuzzy matches
  - Transform from one format to another
  - Link to databases and spreadsheets
- Runs on desktop
- See video for more

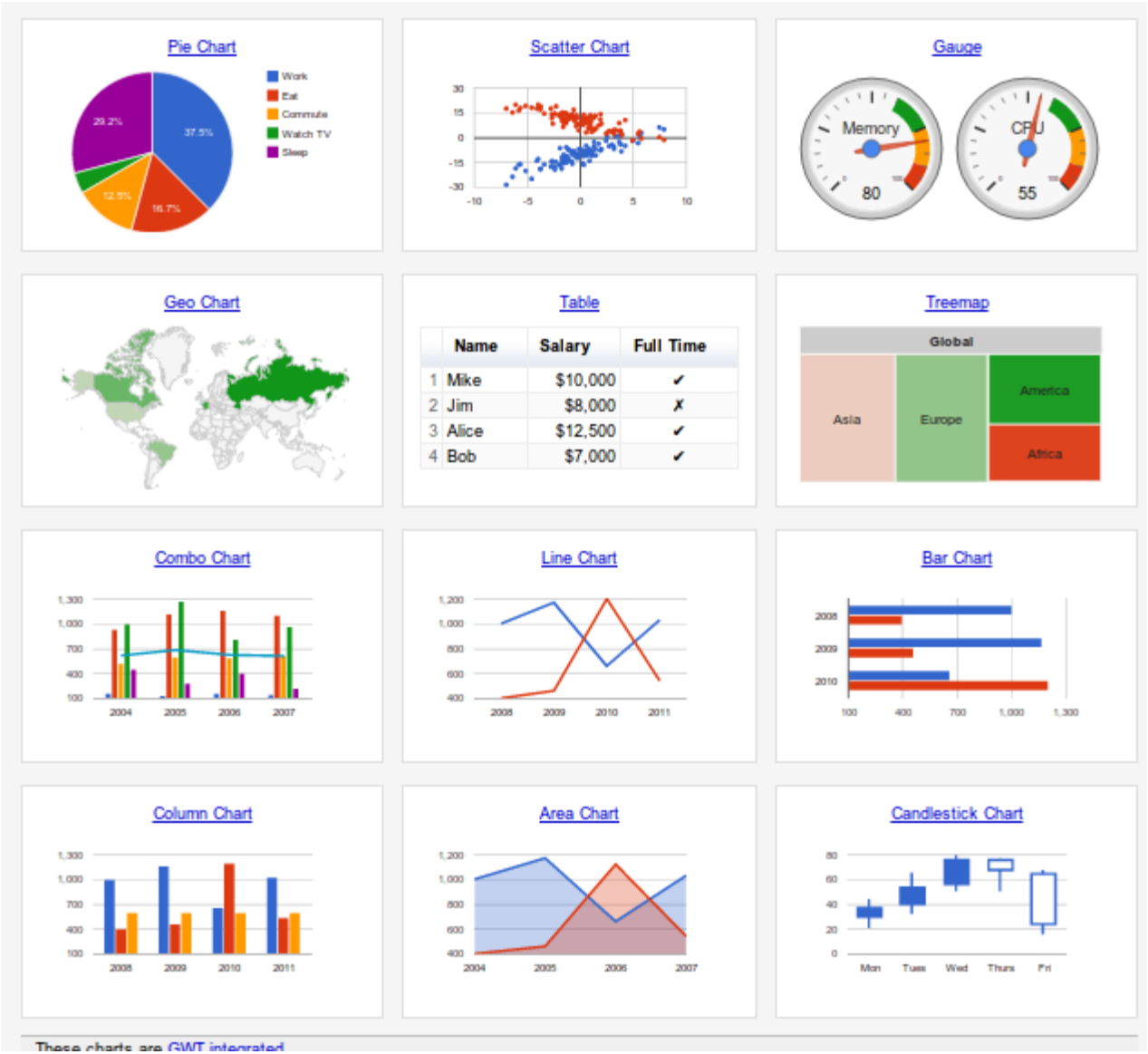
Motion charts

- Accessible from Spreadsheets, Chart Tools, R, and Public Data Explorer

Prediction API

- Google's cloud-based machine learning tools can help analyze your data to add the following features to your applications:
  - Customer sentiment analysis
  - Spam detection
  - Message routing decisions
  - Upsell opportunity analysis
  - Document and email classification
  - Diagnostics
  - Churn analysis
  - Suspicious activity identification
  - Recommendation systems

Chart Tools for Web sites



Google Search Insights

Google Insights for Search beta

hal@google.com | My Account | Help | Sign out | Download as CSV | English (US)

Compare by	Search terms	Filter
◉ Search terms ○ Locations ○ Time Ranges	Tip: Use quotation marks to match an exact phrase. ("table tennis") ● hangover + Add search term	Web Search United State / All subregic / All metros Dec / 2008 - Feb / 2009 Reset All Categories Search

Web Search Interest: hangover

United States, Dec 2008 - Feb 2009

Categories: Food & Drink (25-50%), Health (10-25%), Entertainment (0-10%), Local (0-10%), more...

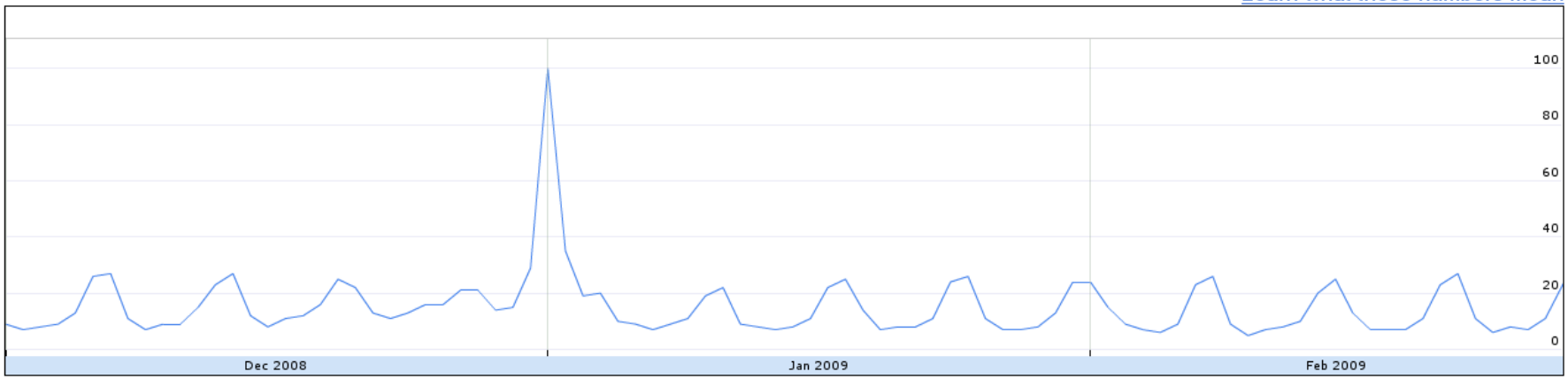
Totals

hangover 15

Interest over time

☐ Forecast ☐ News headlines

Learn what these numbers mean



Embed this chart

Google Search Insights

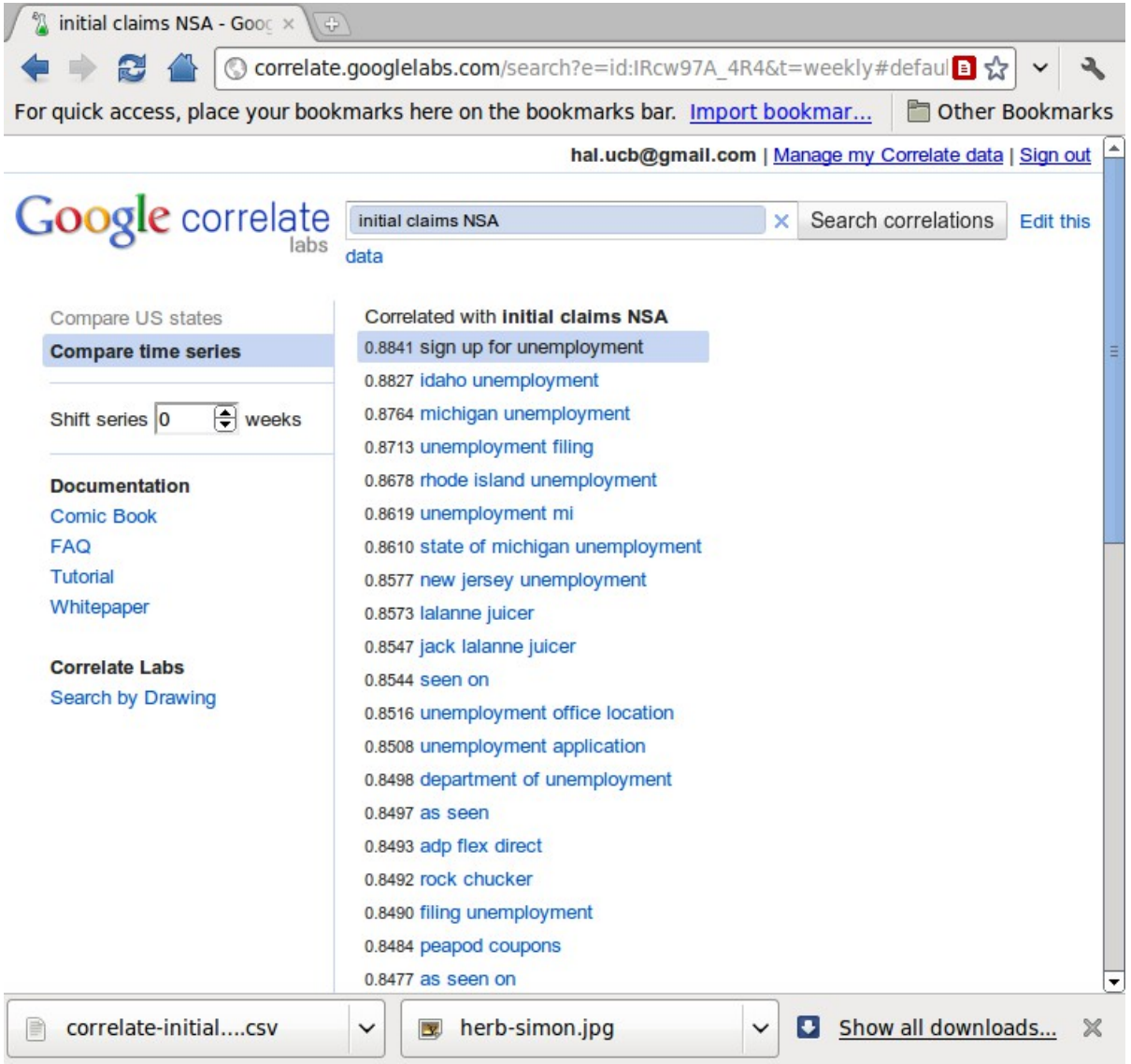
Download search volume index by country/category/term on weekly or daily basis.

Data classified by category, can disambiguate

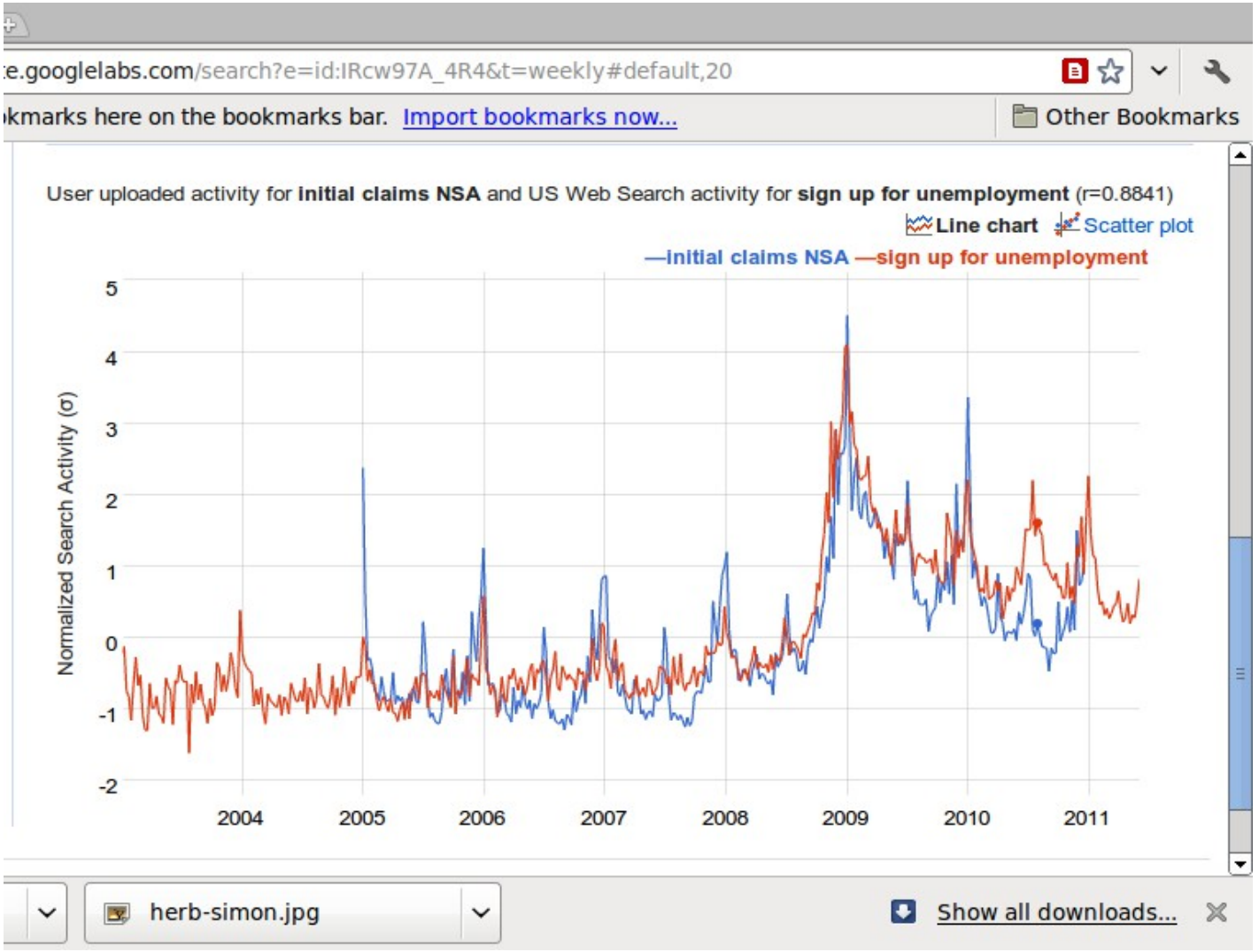
Can be useful in “nowcasting” economic indicators

- Unemployment
- Auto sales
- Real estate
- Travel planning

Google Correlate

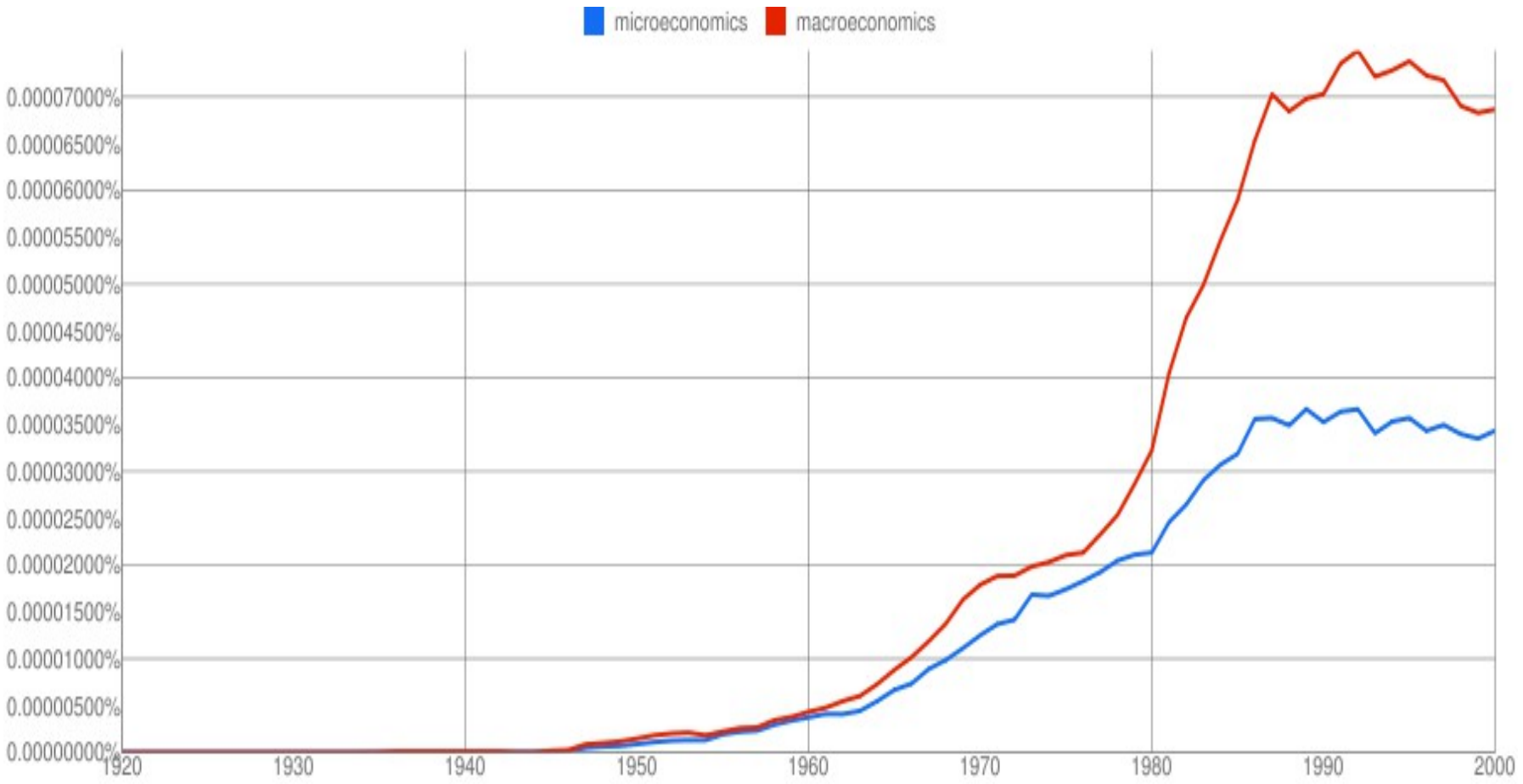


Initial claims for unemployment



nGrams

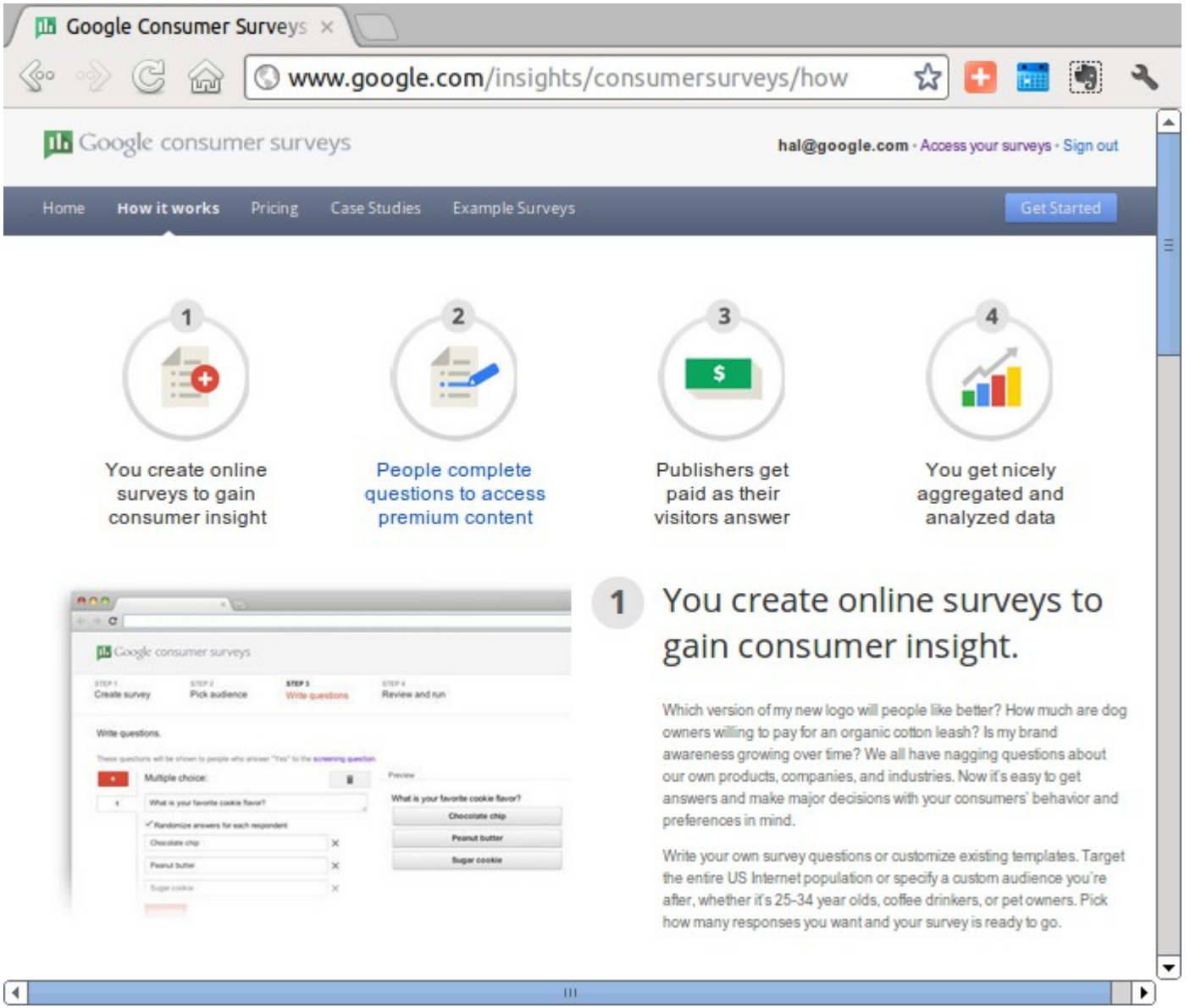
Data from 20 million books



Public Data Explorer

- XML schema for public data
  - World Bank, World Development Indicators
  - International Monetary Fund, September 2011 World Economic Outlook
  - OECD Factbook 2010
  - Unemployment in Europe (monthly)
  - Broadband penetration in Europe
  - Government Debt in Europe
  - Infectious Disease Outbreaks
  - Unemployment in the U.S.
- Visualization tools

Google Consumer Surveys



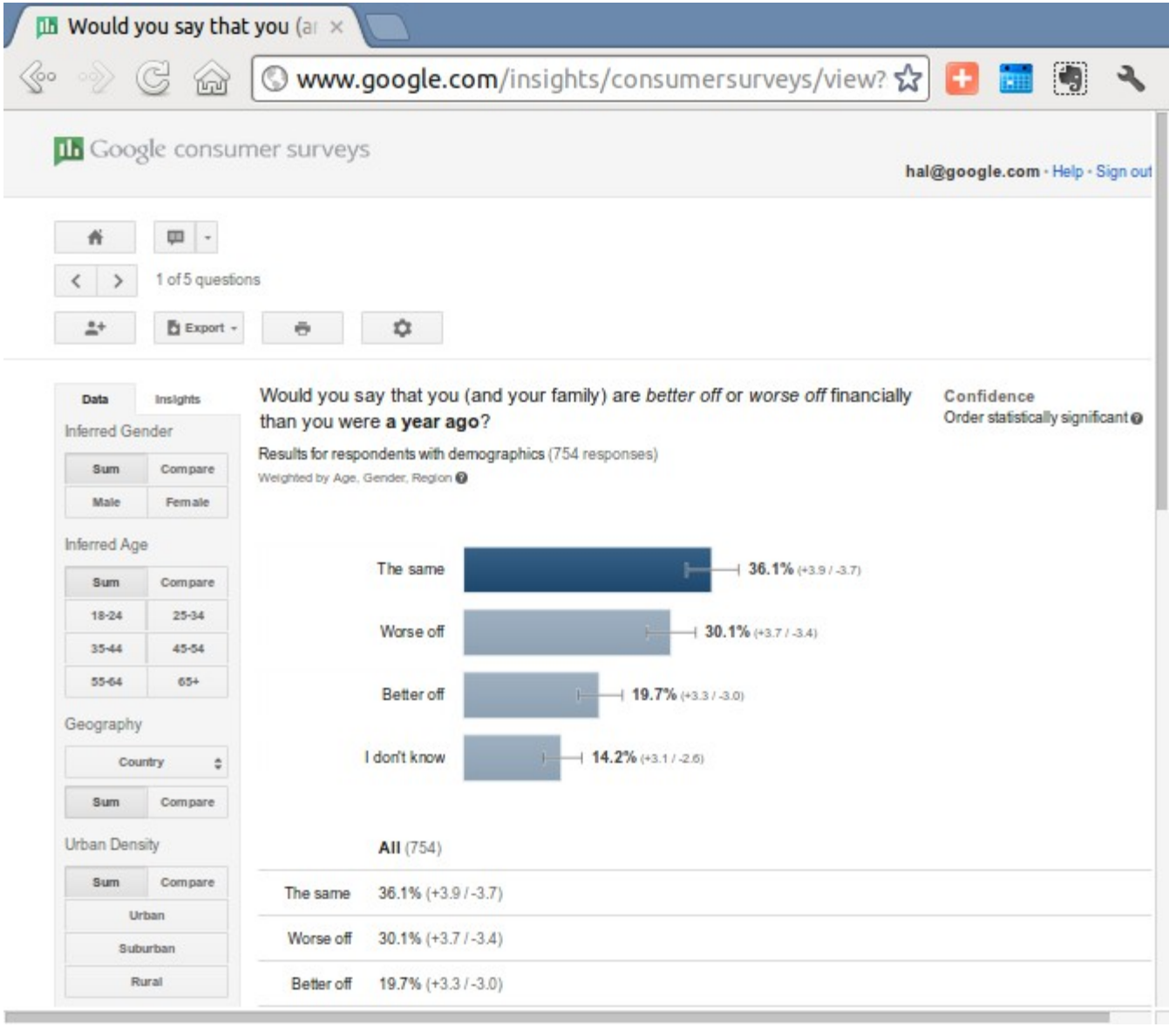
Example survey questions

Consumer sentiment. Would you say that you (and your family) are better off or worse off financially than you were a year ago?

Game theory. A and B are on a TV game show with $100 to divide. A offers B $x. If B accepts the division is carried out. If B refuses each gets zero.

Valuation. Imagine you lived in a world without free search engines like Google and Bing. How much would you pay for a subscription to a search engine?

Consumer Surveys, results



Other web data

Existing

- Billion prices project (MIT)
- Mastercard Spending Pulse
- Monster Employment Index
- Intuit Small Business Employment Index
- Zillow Real Estate Market Reports

Potential

- Wal-Mart, Target, K-Mart retails sales
- Price indices from retail data
- Package delivery data from UPS, FedEx

Policy issues for discussion

For business “real time” is more important than “historical consistency”

- Changes in definitions makes life difficult for researchers

What are incentives for private sector to provide data?

- Profit motive (Mastercard, Visa)
- Brand identity, thought leadership (Intuit, Monster, Zillow, Google)
- Financial reporting to investors (FedEx, UPS, retail)

Some forms of data are subject to manipulation

- Search data, non-transactional price data

Econometric issues

- Short time series, but many predictors (fat regression)
- Model averaging v economic science

Data Requests to Google

Google likes to release data to everyone or no one...

- “Managing the world's information...”
- Very hard to do custom requests
- Clearance, external contracts, engineering resources, privacy, attention, etc.

Very interested in supporting research community

- Feature requests for products
- Google Faculty Grants – access on-site data (Mt View, Cambridge, New York, Pittsburgh, Ann Arbor, Chicago, Boulder, Irvine, Venice CA, London, Paris, etc.)