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

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...

Interest over time

Learn what these numbers mean
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

You create online surveys to gain consumer insight.

People complete questions to access premium content.

Publishers get paid as their visitors answer.

You get nicely aggregated and analyzed data.

You create online surveys to gain consumer insight.

Which version of my new logo will people like better? How much are dog owners willing to pay for an organic cotton leash? Is my brand awareness growing over time? We all have nagging questions about our own products, companies, and industries. Now it’s easy to get answers and make major decisions with your consumers’ behavior and preferences in mind.

Write your own survey questions or customize existing templates. Target the entire US internet population or specify a custom audience you’re after, whether it’s 25-34 year olds, coffee drinkers, or pet owners. Pick how many responses you want and your survey is ready to go.
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

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

Results for respondents with demographics (754 responses)

Weighted by Age, Gender, Region

<table>
<thead>
<tr>
<th>Inferred Gender</th>
<th>Insights</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun Male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sun Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sun Male</td>
<td>Compare</td>
<td></td>
</tr>
<tr>
<td>Sun Female</td>
<td>Compare</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inferred Age</th>
<th>Insights</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65+</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Geography</th>
<th>Insights</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sun</td>
<td>Compare</td>
<td></td>
</tr>
<tr>
<td>Sun</td>
<td>Compare</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Urban Density</th>
<th>Insights</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun Urban</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sun Suburban</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The same: 36.1% (+3.9/-4.7)
Worse off: 30.1% (+3.7/-3.4)
Better off: 19.7% (+3.3/-3.0)
I don't know: 14.2% (+3.1/-2.6)

All (754):
The same: 36.1% (+3.9/-3.7)
Worse off: 30.1% (+3.7/-3.4)
Better off: 19.7% (+3.3/-3.0)
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 retail 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.)