



Data resources

Jeffrey Leek
Johns Hopkins Bloomberg School of Public Health

Open Government Sites

- United Nations <http://data.un.org/>
- U.S. <http://www.data.gov/>
 - [List of cities/states with open data](#)
- United Kingdom <http://data.gov.uk/>
- France <http://www.data.gouv.fr/>
- Ghana <http://data.gov.gh/>
- Australia <http://data.gov.au/>
- Germany <https://www.govdata.de/>
- Hong Kong <http://www.gov.hk/en/theme/psi/datasets/>
- Japan <http://www.data.go.jp/>
- Many more <http://www.data.gov/opendatasites>

Gapminder

The screenshot shows the Gapminder website at www.gapminder.org/data/. The page title is "Data In Gapminder World". It features a search bar and navigation links for Home, DATA, VIDEOS, DOWNLOADS, FOR TEACHERS, and LABS. Below the navigation is a table titled "List of indicators in Gapminder World" with columns for Indicator name, Data provider, Category, Subcategory, Download, View, and Visualize. The table lists several indicators such as "Adults with HIV (%; age 15-49)", "Age at first marriage (women)", "Aged 15+ employment rate (%)", etc.

Indicator name	Data provider	Category	Subcategory	Download	View	Visualize
Adults with HIV (%; age 15-49)	Based on UNAIDS	Health	HIV			
Age at first marriage (women)	Various sources	Population				
Aged 15+ employment rate (%)	International Labour Organization	Work	Employment rate			
Aged 15+ labour force participation rate (%)	International Labour Organization	Work	Labour force participation			
Aged 15+ unemployment rate (%)	International Labour Organization	Work	Unemployment			
Aged 16-24 employment rate (%)	International Labour Organization	Work	Employment rate			

<http://www.gapminder.org/>

Survey data from the United States

The screenshot shows a web browser window with the URL www.asdfree.com. The page title is "analyze survey data for free". The main content is titled "analyze the health and retirement study (hrs) with r". It includes a brief introduction about the HRS survey, a "AVAILABLE DATA" section with links to various datasets, and a "METHODS" section. At the bottom, there's a note about using R and a copyright notice.

analyze survey data for free

about / faq main code repository latest releases rss sydnemico@gmail.com tutorials
+blog+rs

reproducible survey analysis syntax from a website that's easy to type.

AVAILABLE DATA

- [available community survey \(acs\)](#)
- [area resource file \(arf\)](#)
- [long-term care and long-term care claims public use file \(lccpuf\)](#)
- [berkman-litnick longitudinal survey \(bls\)](#)
- [university student life survey \(usls\)](#)
- [veteran population survey \(vps\)](#)
- [general social survey \(gss\)](#)
- [health and retirement study \(hrs\)](#)
- [medicare expenditure panel survey \(meps\)](#)
- [national health and nutrition examination survey \(nhes\)](#)
- [national health interview survey \(nhis\)](#)
- [national study on drug use and health \(nsud\)](#)

METHODS

- [why and how to read r-metabase with r-metabase](#)

enter your email address for updates
www.asdfree.com/2012/04/analyze-health-and-retirement-study-with-r.html | 1992 - 2010 download RIRI microdata.R

<http://www.asdfree.com/>

Infochimps Marketplace

The screenshot shows the Infochimps Marketplace homepage. At the top, there's a navigation bar with links for Home, Solutions, Platform, Resources, Community, Company, Blog, and Request a Demo. Below the navigation is a search bar labeled "Search for data" and a "Log in" button. The main content area features a section titled "Data Marketplace" with a sub-section titled "Geo APIs". It includes a brief description and a "Learn more" link. Next to it is a "Social APIs" section with its own description and a "Learn more" link. Below these sections is a "Top Tags" section containing a grid of tags: locations, list, businesses, geo, location, database, places, phone, stores, hours, business-information, phone-numbers, place-names, store-hours, aggregate, population, census, government, demographics, america, health, maps, state, care, toxic, facilities, geonames, facility, release, school, community, and More... .

<http://www.infochimps.com/marketplace>

Kaggle

kaggle™

[Sign Up](#) [About](#) [Hosting Center](#) [All Competitions](#) [Users](#) [Forums](#) [Wiki](#) [Blog](#) [Data Science Jobs](#)

What's in your data?

Participate in competitions

Kaggle is an arena where you can match your data science skills against a global cadre of experts in statistics, mathematics, and machine learning. Whether you're a world-class algorithm wizard competing for prize money or a novice looking to learn from the best, here's your chance to jump in and geek out, for fame, fortune, or fun.

[Join as a participant](#)

(Need convincing?)

Create a competition

Kaggle is a platform for data prediction competitions that allows organizations to post their data and have it scrutinized by the world's best data scientists. In exchange for a prize, winning competitors provide the algorithms that beat all other methods of solving a data crunching problem. Most data problems can be framed as a competition.

[Learn more about hosting](#)

<http://www.kaggle.com/>

Collections by data scientists

- Hilary Mason <http://bitly.com/bundles/hmason/1>
- Peter Skomoroch <https://delicious.com/pskomoroch/dataset>
- Jeff Hammerbacher <http://www.quora.com/Jeff-Hammerbacher/Introduction-to-Data-Science-Data-Sets>
- Gregory Piatetsky-Shapiro <http://www.kdnuggets.com/gps.html>
- <http://blog.mortardata.com/post/67652898761/6-dataset-lists-curated-by-data-scientists>

More specialized collections

- [Stanford Large Network Data](#)
- [UCI Machine Learning](#)
- [KDD Nuggets Datasets](#)
- [CMU Statlib](#)
- [Gene expression omnibus](#)
- [ArXiv Data](#)
- [Public Data Sets on Amazon Web Services](#)

Some API's with R interfaces

- [twitter](#) and [twitteR](#) package
- [figshare](#) and [rfigshare](#)
- [PLoS](#) and [rplos](#)
- [rOpenSci](#)
- [Facebook](#) and [RFacebook](#)
- [Google maps](#) and [RGoogleMaps](#)