



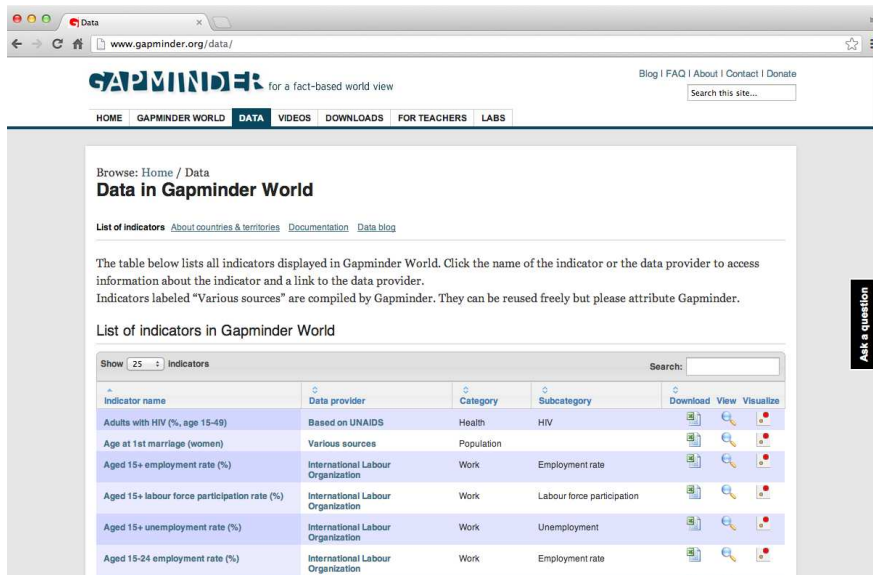
# Data resources

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



Browse: [Home](#) / [Data](#)

## Data in Gapminder World

[List of indicators](#) [About countries & territories](#) [Documentation](#) [Data blog](#)

The table below lists all indicators displayed in Gapminder World. Click the name of the indicator or the data provider to access information about the indicator and a link to the data provider.

Indicators labeled "Various sources" are compiled by Gapminder. They can be reused freely but please attribute Gapminder.

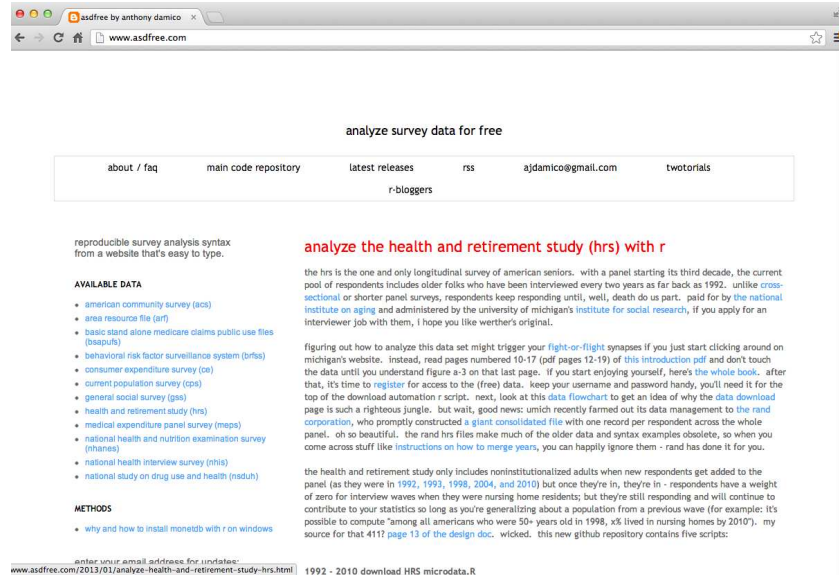
### List of indicators in Gapminder World

Show 25 Indicators Search:

Indicator name	Data provider	Category	Subcategory	Download	View	Visualize
Adults with HIV (% age 15-49)	Based on UNAIDS	Health	HIV			
Age at 1st 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 15-24 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 address bar displaying "www.asdfree.com". The page title is "analyze survey data for free". Below the title is a navigation bar with links: "about / faq", "main code repository", "latest releases", "rss", "ajdamico@gmail.com", and "twotutorials". The main content area is divided into two columns. The left column has a heading "reproducible survey analysis syntax from a website that's easy to type." followed by a section "AVAILABLE DATA" with a list of survey names: "american community survey (acs)", "area resource file (arf)", "basic stand alone medicare claims public use files (bsapufs)", "behavioral risk factor surveillance system (brfss)", "consumer expenditure survey (ces)", "current population survey (cps)", "general social survey (gas)", "health and retirement study (hrs)", "medical expenditure panel survey (meps)", "national health and nutrition examination survey (nhanes)", "national health interview survey (nhis)", and "national study on drug use and health (nsduh)". Below this is a section "METHODS" with a link "why and how to install monedit with r on windows". The right column has a heading "analyze the health and retirement study (hrs) with r" followed by a paragraph about the HRS survey and a link to a "data flowchart". At the bottom of the page, there is a footer with the text "1992 - 2010 download HRS microdata.R".

analyze survey data for free

about / faq   main code repository   latest releases   rss   ajdamico@gmail.com   twotutorials

r-bloggers

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

**AVAILABLE DATA**

- american community survey (acs)
- area resource file (arf)
- basic stand alone medicare claims public use files (bsapufs)
- behavioral risk factor surveillance system (brfss)
- consumer expenditure survey (ces)
- current population survey (cps)
- general social survey (gas)
- health and retirement study (hrs)
- medical expenditure panel survey (meps)
- national health and nutrition examination survey (nhanes)
- national health interview survey (nhis)
- national study on drug use and health (nsduh)

**METHODS**

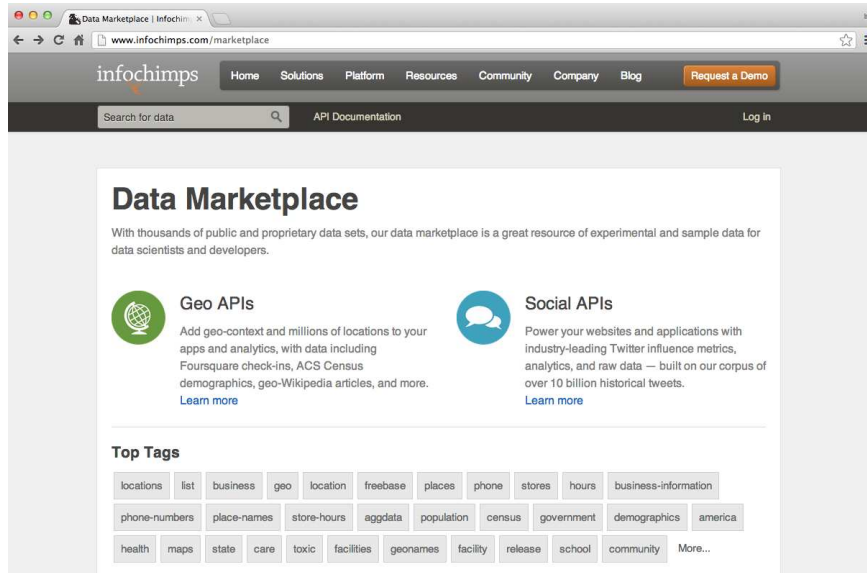
- why and how to install monedit with r on windows

enter your email address for updates:  
[www.asdfree.com/2013/01/analyze-health-and-retirement-study-hrs.html](http://www.asdfree.com/2013/01/analyze-health-and-retirement-study-hrs.html)

1992 - 2010 download HRS microdata.R

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

# Infochimps Marketplace



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

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



# Some API's with R interfaces

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