

# Data Management in R

Ista Zahn

Harvard MIT Data Center

November 2nd, 2012



**The Institute**  
*for* Quantitative Social Science  
at Harvard University

- 1 Introduction
- 2 Data management overview
- 3 Wrap-up

# Topic

- 1 Introduction
- 2 Data management overview
- 3 Wrap-up

# Materials and setup

- USERNAME: dataclass
- PASSWORD: dataclass
- Find class materials at

<http://projects.iq.harvard.edu/rtc/RDatMan>

- Download the zip file at the bottom of the page and unzip on your desktop!

# Topic

- 1 Introduction
- 2 Data management overview
- 3 Wrap-up

# Some thoughts on data management

*It's often said that 80% of the effort of analysis is in data cleaning. Despite the amount of time it takes, there has been little research on how to clean data well.*

*–Hadley Wickham,*

*<http://vita.had.co.nz/papers/tidy-data.pdf>*

*Another thing that parenting and the "dirty jobs" of research have in common is that nobody gets taught how to do it. Parents just learn on the job and researchers typically have to do likewise when it comes to learning how to manage their data.*

*–Paul Murrell,*

*<http://www.stat.auckland.ac.nz/~paul/ItDT/HTML/>*

# Topic

- 1 Introduction
- 2 Data management overview
- 3 **Wrap-up**

# Help us make this workshop better!

- Please take a moment to fill out a very short feedback form
- These workshops exist for you – tell us what you need!
- <http://tinyurl.com/R-intro-feedback>



# Additional resources

- IQSS workshops:  
[http://projects.iq.harvard.edu/rtc/filter\\_by/workshops](http://projects.iq.harvard.edu/rtc/filter_by/workshops)
- MIT workshops: <http://libraries.mit.edu/sites/news/check-complete-listing-5/9319/>
- Software (all free!):
  - R and R package download: <http://cran.r-project.org>
  - Rstudio download: <http://rstudio.org>
  - ESS (emacs R package): <http://ess.r-project.org/>
- Getting help:
  - Documentation and tutorials:  
<http://cran.r-project.org/other-docs.html>
  - Recommended R packages by topic:  
<http://cran.r-project.org/web/views/>
  - Mailing list: <https://stat.ethz.ch/mailman/listinfo/r-help>
  - StackOverflow: <http://stackoverflow.com/questions/tagged/r>