

# Reproducible research and data science

Julian

June 16, 2015

# Importance of reproducible research

- Basic reports/summaries
- Consistent mistakes
- Presenting results
- Replicating results

# Application of reproducible research

- Funding
  - Figuring out dead ends
  - Agencies/foundations require audits
- Accessibility:
  - Journals: raw data, analyses, figures
  - Supplemental data
  - Public resource
- Fraud detection
  - Benford's law/distribution
  - Irregularities in LaCour 2014

# 'Practical' applications of reproducible research

- Platform independent
- Job applications
- Tutorials
- Demonstrating skills
- Showcasing projects
- Publishing/portfolio

# Resources

- LaTeX: wonderful, sharp learning curve
- R + RStudio
  - Sweave: can use LaTeX + R, Python, etc.
  - knitr: (R) Markdown + R code, Python, etc.
- Python
  - Jupyter/IPython: Markdown + Python/IPython + 'magics'
  - Eclipse
- Git/GitHub
  - Sharing data publicly
  - Tracking changes/version control
  - Portfolio/public persona