

# Using R Markdown in Introductory Statistics

Ben Baumer<sup>1</sup>

<sup>1</sup> Smith College  
Northampton, MA

USCOTS 2013  
Cary, NY  
May 17th, 2013



SMITH COLLEGE

# Student Workflow in Intro Stats

- Computation is essential
  - ▶ Ideal Tool: stat package of your choice
  - ▶ R with `mosaic`
- Written analysis is imperative
  - ▶ Ideal Tool: word processor of your choice
  - ▶ Word? GoogleDocs? LibreOffice?  $\text{\LaTeX}$ ?
- How to combine the two?

COPY - AND - PASTE!



# Student Workflow in Intro Stats

- Computation is essential
  - ▶ Ideal Tool: stat package of your choice
  - ▶ R with `mosaic`
- Written analysis is imperative
  - ▶ Ideal Tool: word processor of your choice
  - ▶ Word? GoogleDocs? LibreOffice?  $\text{\LaTeX}$ ?
- How to combine the two?

COPY - AND - PASTE!



# Student Workflow in Intro Stats

- Computation is essential
  - ▶ Ideal Tool: stat package of your choice
  - ▶ R with `mosaic`
- Written analysis is imperative
  - ▶ Ideal Tool: word processor of your choice
  - ▶ Word? GoogleDocs? LibreOffice? L<sup>A</sup>T<sub>E</sub>X?
- How to combine the two?

COPY - AND - PASTE!



# Student Workflow in Intro Stats

- Computation is essential
  - ▶ Ideal Tool: stat package of your choice
  - ▶ R with `mosaic`
- Written analysis is imperative
  - ▶ Ideal Tool: word processor of your choice
  - ▶ Word? GoogleDocs? LibreOffice?  $\text{\LaTeX}$ ?
- How to combine the two?

COPY - AND - PASTE!



# Why is that bad?

- Not reproducible
  - ▶ Difficult or impossible to follow
  - ▶ Easy to forget how to retrace steps
- Not logical
  - ▶ Separates analysis from computation
  - ▶ Little or no connection between data and analysis
- Not necessarily honest
  - ▶ Allows fudging
  - ▶ Permits selective reporting



# Why is that bad?

- Not reproducible
  - ▶ Difficult or impossible to follow
  - ▶ Easy to forget how to retrace steps
- Not logical
  - ▶ Separates analysis from computation
  - ▶ Little or no connection between data and analysis
- Not necessarily honest
  - ▶ Allows fudging
  - ▶ Permits selective reporting



# Why is that bad?

- Not reproducible
  - ▶ Difficult or impossible to follow
  - ▶ Easy to forget how to retrace steps
- Not logical
  - ▶ Separates analysis from computation
  - ▶ Little or no connection between data and analysis
- Not necessarily honest
  - ▶ Allows fudging
  - ▶ Permits selective reporting





# R Markdown

- Simple, free, open source, easy-to-learn markup syntax
- Text & R code  $\Rightarrow$  HTML
  - ▶ R commands alongside the output from that command
  - ▶ Plots embedded into a single file
- Supports some  $\text{\LaTeX}$
- One file, one workflow
- Implementation: RStudio with `knitr`

# At Smith

- Fall 2012
  - ▶ MTH 245: Intro Prob. & Stats (5 credits, 42 students)
  - ▶ MTH 247: Regression (33 students)
- Spring 2013
  - ▶ MTH 241: Intro Prob. & Stats (4 credits, 3 × 25 students)
- Fall 2013
  - ▶ MTH 292: Data Science (4 credits, 22 students?)
  - ▶ Python hooks?
- (almost) All homeworks and projects completed in Markdown
- Building institutional knowledge
  - ▶ 100+ students on campus with Markdown experience
  - ▶ 6 Stat TAs trained and experienced with R Markdown
- Collaborations poster exploring attitudes towards Markdown
  - ▶ More on this at JSM Roundtable



# At Smith

- Fall 2012
  - ▶ MTH 245: Intro Prob. & Stats (5 credits, 42 students)
  - ▶ MTH 247: Regression (33 students)
- Spring 2013
  - ▶ MTH 241: Intro Prob. & Stats (4 credits, 3 × 25 students)
- Fall 2013
  - ▶ MTH 292: Data Science (4 credits, 22 students?)
  - ▶ Python hooks?
- (almost) All homeworks and projects completed in Markdown
- Building institutional knowledge
  - ▶ 100+ students on campus with Markdown experience
  - ▶ 6 Stat TAs trained and experienced with R Markdown
- Collaborations poster exploring attitudes towards Markdown
  - ▶ More on this at JSM Roundtable



# At Smith

- Fall 2012
  - ▶ MTH 245: Intro Prob. & Stats (5 credits, 42 students)
  - ▶ MTH 247: Regression (33 students)
- Spring 2013
  - ▶ MTH 241: Intro Prob. & Stats (4 credits, 3 × 25 students)
- Fall 2013
  - ▶ MTH 292: Data Science (4 credits, 22 students?)
  - ▶ Python hooks?
- (almost) All homeworks and projects completed in Markdown
- Building institutional knowledge
  - ▶ 100+ students on campus with Markdown experience
  - ▶ 6 Stat TAs trained and experienced with R Markdown
- Collaborations poster exploring attitudes towards Markdown
  - ▶ More on this at JSM Roundtable



# At Smith

- Fall 2012
  - ▶ MTH 245: Intro Prob. & Stats (5 credits, 42 students)
  - ▶ MTH 247: Regression (33 students)
- Spring 2013
  - ▶ MTH 241: Intro Prob. & Stats (4 credits, 3 × 25 students)
- Fall 2013
  - ▶ MTH 292: Data Science (4 credits, 22 students?)
  - ▶ Python hooks?
- (almost) All homeworks and projects completed in Markdown
- Building institutional knowledge
  - ▶ 100+ students on campus with Markdown experience
  - ▶ 6 Stat TAs trained and experienced with R Markdown
- Collaborations poster exploring attitudes towards Markdown
  - ▶ More on this at JSM Roundtable



# At Smith

- Fall 2012
  - ▶ MTH 245: Intro Prob. & Stats (5 credits, 42 students)
  - ▶ MTH 247: Regression (33 students)
- Spring 2013
  - ▶ MTH 241: Intro Prob. & Stats (4 credits, 3 × 25 students)
- Fall 2013
  - ▶ MTH 292: Data Science (4 credits, 22 students?)
  - ▶ Python hooks?
- (almost) All homeworks and projects completed in Markdown
- Building institutional knowledge
  - ▶ 100+ students on campus with Markdown experience
  - ▶ 6 Stat TAs trained and experienced with R Markdown
- Collaborations poster exploring attitudes towards Markdown
  - ▶ More on this at JSM Roundtable



# At Smith

- Fall 2012
  - ▶ MTH 245: Intro Prob. & Stats (5 credits, 42 students)
  - ▶ MTH 247: Regression (33 students)
- Spring 2013
  - ▶ MTH 241: Intro Prob. & Stats (4 credits,  $3 \times 25$  students)
- Fall 2013
  - ▶ MTH 292: Data Science (4 credits, 22 students?)
  - ▶ Python hooks?
- (almost) All homeworks and projects completed in Markdown
- Building institutional knowledge
  - ▶ 100+ students on campus with Markdown experience
  - ▶ 6 Stat TAs trained and experienced with R Markdown
- Collaborations poster exploring attitudes towards Markdown
  - ▶ More on this at JSM Roundtable



# Examples

- Illustration ([http://www.rstudio.com/ide/docs/authoring/using\\_markdown](http://www.rstudio.com/ide/docs/authoring/using_markdown))
- Lectures Notes (<http://www.math.smith.edu/~bbaumer/mth247/labs/logistic.html>)
- Homework Solutions ([http://www.math.smith.edu/~bbaumer/uscots/hw4\\_solutions.html](http://www.math.smith.edu/~bbaumer/uscots/hw4_solutions.html))
- Student Project (<http://www.math.smith.edu/~bbaumer/uscots/group-d-submit.html>)
- RPubs (<http://rpubs.com>)

