

# STAT 503 Homework 4: STAT 101 Grades

*Andee Kaplan, Will Landau, Fangge Liu, Lindsay Rutter*

*March 27, 2015*

## Introduction

STAT 101 course instructors at Iowa State University usually claim their students are diverse. The undergrads who sign up have a wide variety of majors, backgrounds, perspectives, abilities, and levels of motivation. Visual and unsupervised analyses of homework grades may classify students in useful, insightful ways and even inform pedagogy.

We have three homework grade spreadsheets, each of which comes directly from Blackboard (Iowa State University 2015), Iowa State’s system for managing course materials and grades. Each dataset corresponds to a single semester of STAT 101: either fall 2013, fall 2014, or spring 2014. Each semester has six or seven sections of roughly one hundred students each. Every spreadsheet has roughly twenty variables, each of which corresponds to a homework grade (either percentage of points earned or NA for a missing assignment) or the average homework score with missing assignments removed.

## Missing values

A large fraction of homework scores appear as “NA”, or missing. Since the spreadsheets come directly from Blackboard, we assume that almost all NA’s correspond to homeworks that students failed to turn in, and only a small number, if any, are from bookkeeping errors. Before clustering, we need to impute these values, and for a imputation strategy, we look at patterns of missingness.

Figure 1 plots the number of missing values per student for all students. Most of the students missed few to no assignments, and some students missed several. This pattern is consistent with each semester and section.

Figures 2, 3, and 4 show each student’s pattern of missingness. General patterns vary significantly by semester and are mostly consistent by section within each semester.

From visual inspection, we can already partition the students in each semester into four groups. . .

## Acknowledgements

We would like to thank Dr. Cook for her advice on dealing with missing values and her imputation code. Also, we used the R packages ggplot2 (Wickham 2009), gridExtra (Auguie 2012), and reshape2 (Wickham 2007).

## References

- Auguie, Baptiste. 2012. *GridExtra: Functions in Grid Graphics*. <http://CRAN.R-project.org/package=gridExtra>.
- Iowa State University. 2015. “Blackboard Learn.” <https://bb.its.iastate.edu>.
- Wickham, Hadley. 2007. “Reshaping Data with the reshape Package.” *Journal of Statistical Software* 21 (12): 1–20. <http://www.jstatsoft.org/v21/i12/>.
- . 2009. *Ggplot2: Elegant Graphics for Data Analysis*. Springer New York. <http://had.co.nz/ggplot2/book>.

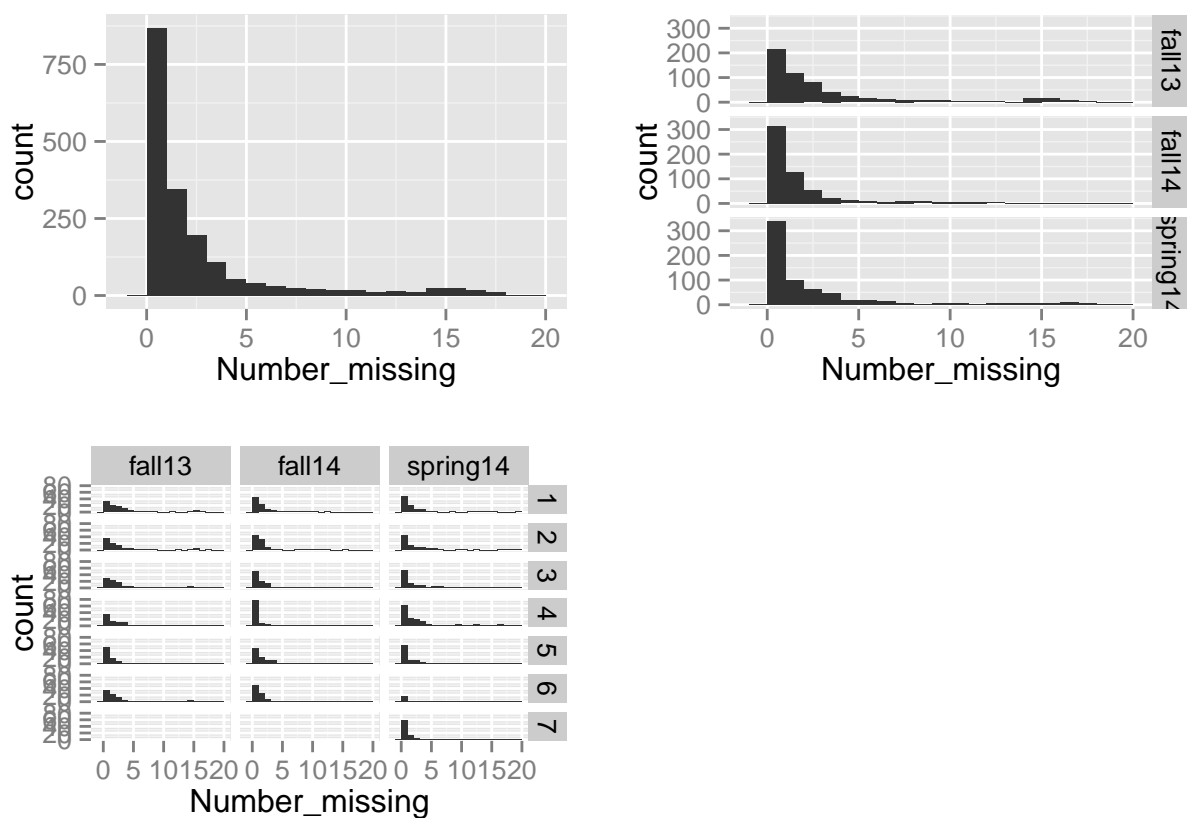


Figure 1: number of missing values per student for all students. The top right panel facets by semester, and the bottom left panel facets by semester and section number. Most of the students missed few to no assignments, and some students missed several. This pattern is consistent with each semester and section.

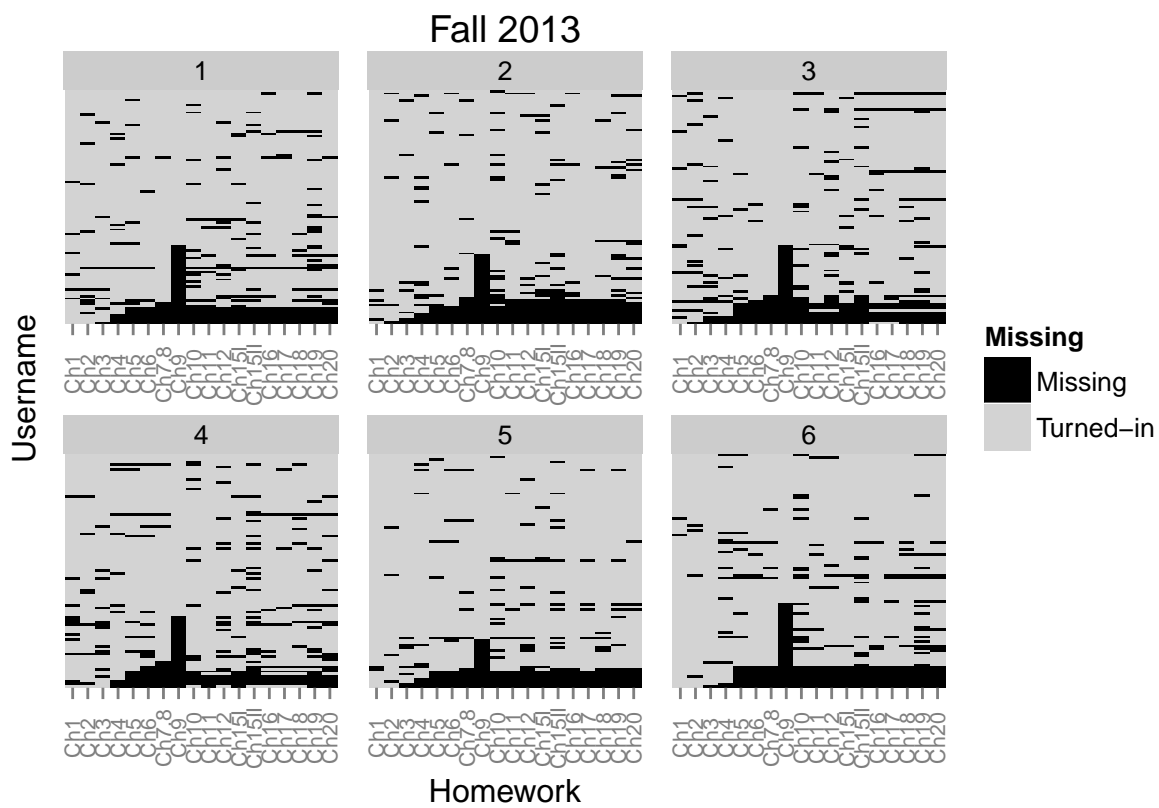


Figure 2: fall 2013

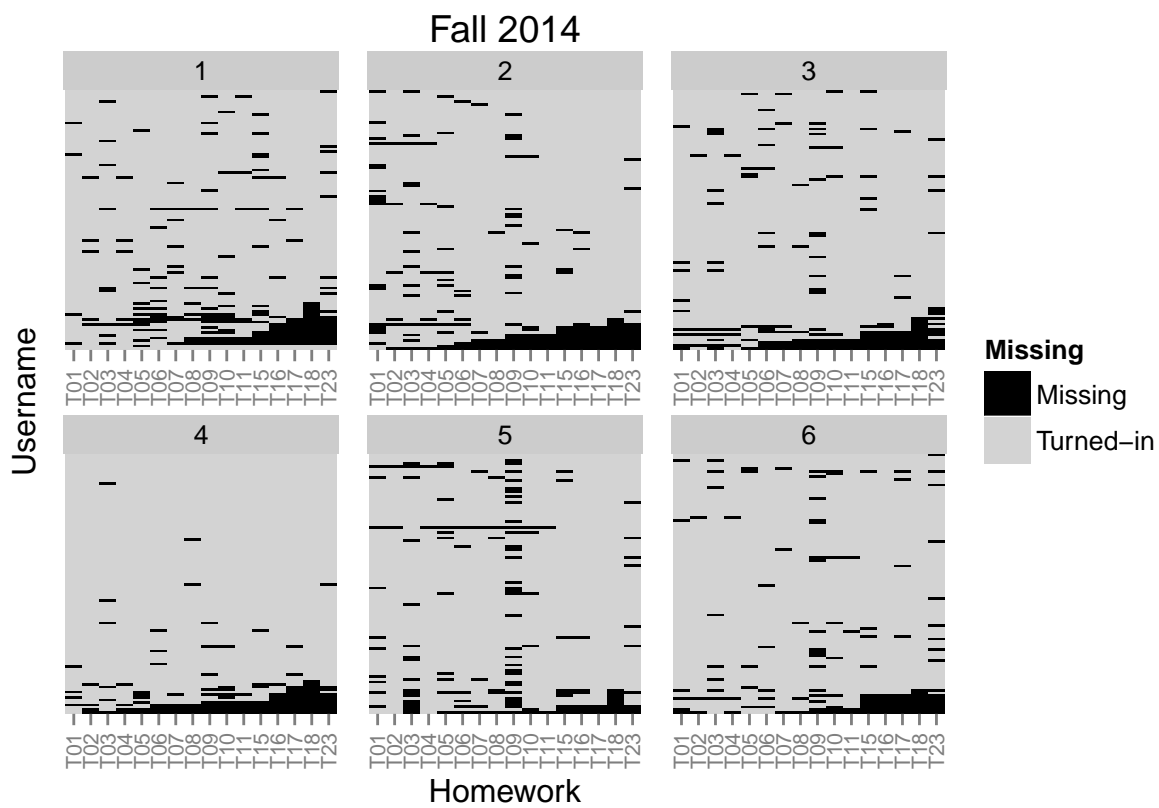


Figure 3: fall 2014

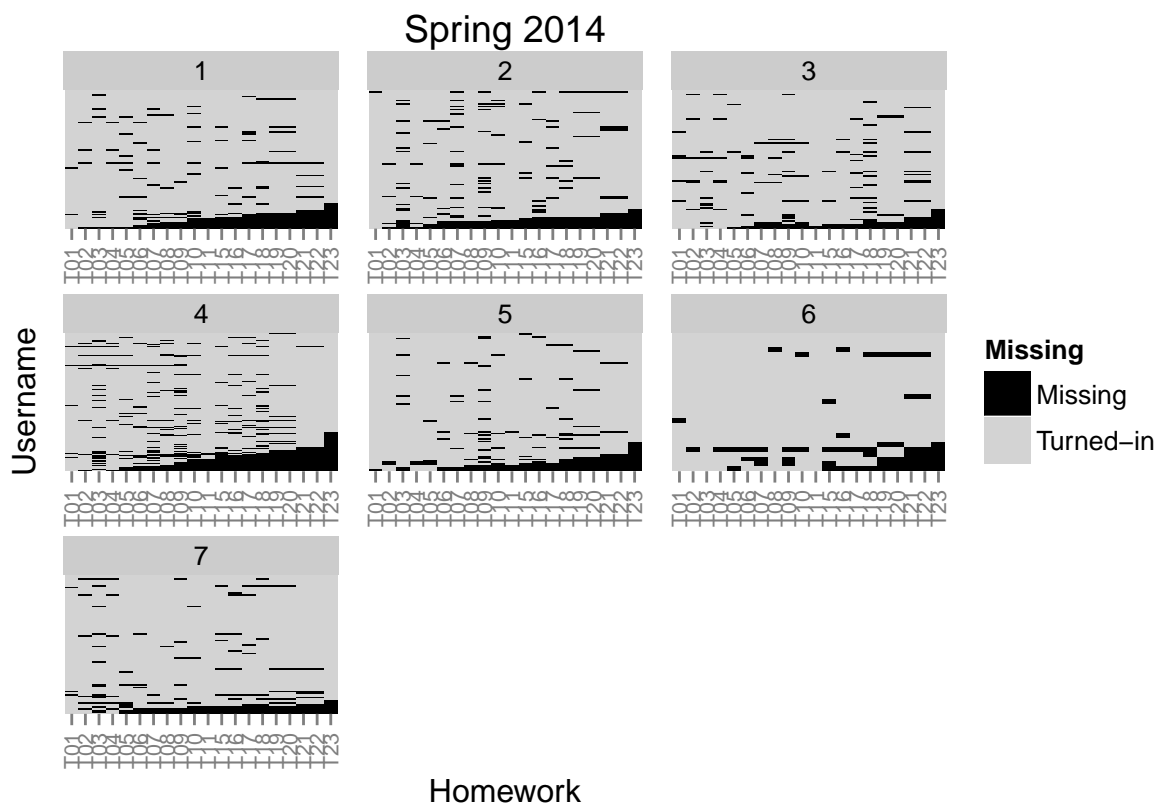


Figure 4: spring 2014