

GRAPHICS FORMULA SYNTAX

```
goal ( y ~ x | z ,
      groups= w , data = mydata )
```

- y — y-axis variable (OPTIONAL)
- x — x-axis variable (REQUIRED)
- z — facet-by variable (OPTIONAL)
- w — color-by variable (OPTIONAL)

LEFT: `bwplot(wage~sex, data=CPS85)`
MIDDLE: `xypplot(wage~education | sex, data=CPS85)`
RIGHT: `xypplot(wage~education, groups=sex, data=CPS85)`

LEFT: `densityplot(~wage, data=CPS85)`
MIDDLE: `densityplot(~wage | sex, data=CPS85)`
RIGHT: `densityplot(~wage, groups=sex, data=CPS85)`

RANDOMIZATION AND ITERATION

RESAMPLE/BOOTSTRAP:

```
do(100)*mean(wage ~ sex, data=resample(CPS85))
```

RANDOM PERMUTATIONS:

```
do(100)*mean(wage ~ shuffle(sex), data=CPS85)
```

CONFIDENCE INTERVALS & STATISTICAL TESTS

```
t.test(wage ~ sex, data=CPS85)
prop.test(43, 100)

crosstab <- tally(~union+sex, data=CPS85)
chisq.test(crosstab) fisher.test(crosstab)

mod <- lm(wage ~ sector, data=CPS85)
Then ... anova(mod) TukeyHSD(mod) etc.

MODELING & COVARIATES
mod <- lm(wage ~ sex + education, data=CPS85)
summary(mod) or anova(mod) or confint(mod)

EXTRACT MODEL FUNCTION:
fun <- makeFun(mod)
fun(sex="F", education=10)

plotFun(fun(sex="F", education=x)) ~ x, x.lim=range(0,8))
```

A STUDENT'S GUIDE TO R is one of a series of books designed to help integrate modern computation into statistics courses. We refer to our approach as **computational statistics** because the availability of computation is shaping how statistics is done, taught, and understood. Computational statistics is a key component of **data science**, using data to answer questions and communicate results.

Other books in the series include:

- START MODELING WITH R
- START TEACHING WITH R
- START R IN CALCULUS

These materials have been shared with hundreds of statistics educators through workshops run under the auspices of Project Mosaic, CAUSE, the Mathematical Association of America, the American Statistical Association, the W.M. Keck Foundation, the Howard Hughes Medical Institute, and the US National Science Foundation.

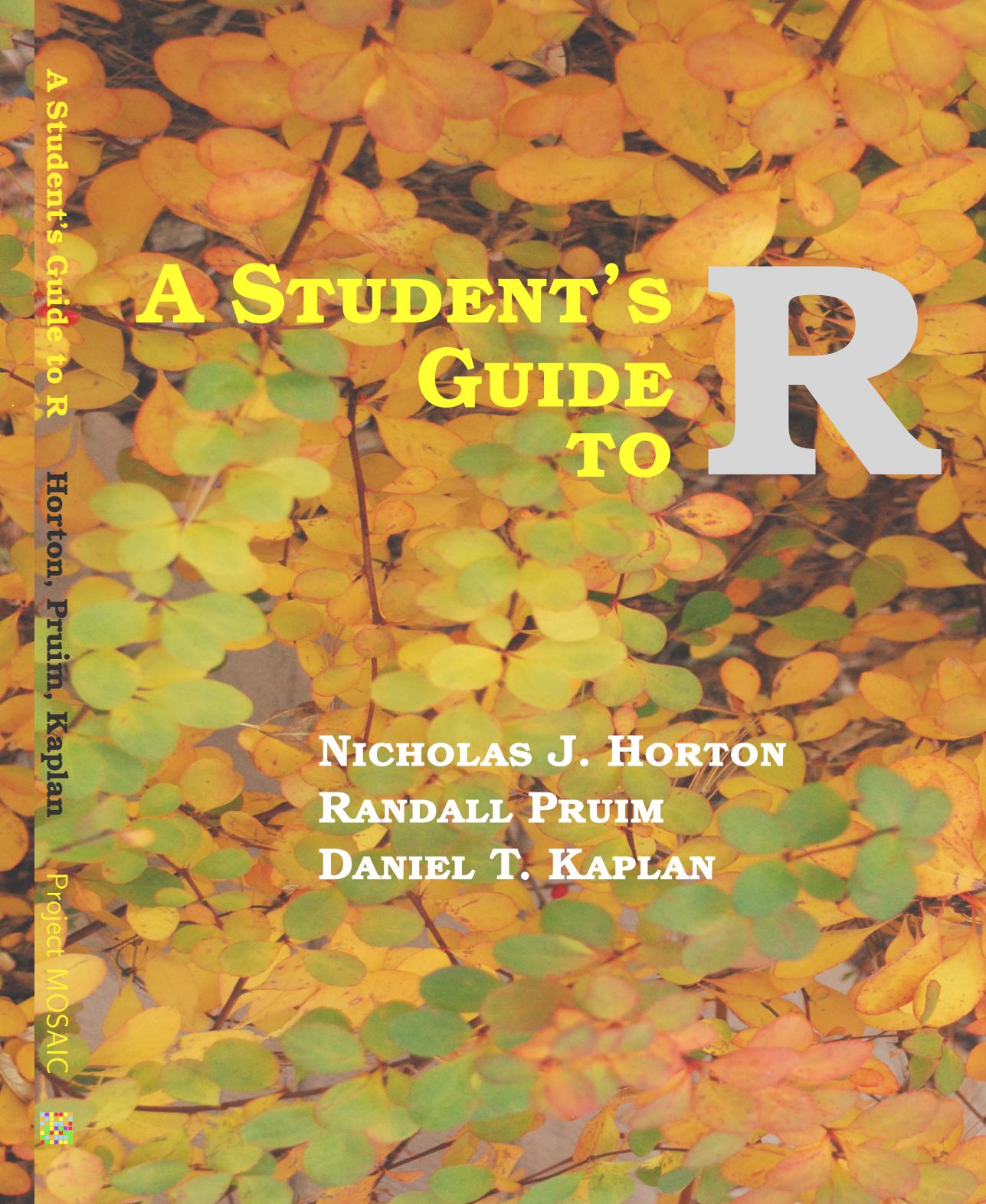
Nicholas Horton, Randall Pruim, and Daniel Kaplan teach statistics, mathematics, and computation at Amherst College, Calvin College, and Macalester College respectively. They are co-PIs on the National Science Foundation support Project MOSAIC (NSF DUE-0920350).

HORTON PRUIM KAPLAN

Other books by the authors:

Using R for Data Management, Statistical Analysis and Graphics (2nd edition) (NJH & KK)
Foundations and Applications of Statistics: An Introduction Using R (RJP),
Gems of Theoretical Computer Science (US & RJP), *Understanding Nonlinear Dynamics* (DTK), *Statistical Modeling: A Fresh Approach* (DTK), *Start R in Calculus* (DTK), *Data Computing* (DTK)

ISBN 9780983965886
90000 >
9 780983 965886



LOAD PACKAGES

```
require(mosaic)
require(mosaicData)
```

ESSENTIAL R SYNTAX

Function & arguments:
Optional arguments: `rflip(10, prob=0.3)`
Assignment: `x <- rflip(10, prob=0.3)`

FORMULA INTERFACE

Used for graphics, statistics, inference, and modeling operations.

```
goal ( y ~ x , data = mydata )
```

Read as: Calculate goal using mydata for y “broken down” by x, or “modeled by” x. Examples:

```
mean(age~homeless, data=HELPrc)
```

homeless	housed
36.4	35.0

```
quantile(age~sex,data=HELPrc,p=c(.2,.8))
```

.group	20%	80%
1 female	30	42.8
2 male	29	41.0

```
tally(homeless~sex, data=HELPrc)
```

sex		
homeless	female	male
homeless	0.374	0.488
housed	0.626	0.512

R MARKDOWN DOCUMENTS

Homework #3
Abby Seidie
January 7, 2015

```
## title: "Homework #3"
## author: "Abby Seidie"
## date: "January 7, 2015"
## output: pdf_document

## {r include=FALSE}
## require(mosaic)
## require(mosaicData)
## 

## Problem 1
Build a model of wage as a function of sex, adjusting for relevant covariates.
## {r}
## lm(wage ~ sex + exper, data=CPS85)

## Problem 2
Show whether the covariate is related to sex.
## {r}
bwplot(exper ~ sex, data=CPS85)
```

Compile to any of HTML, PDF, or Word.
See mosaic plain template through RStudio menu:
FILE/NEW FILE/RMARKDOWN/FROM TEMPLATE