

STATISTICAL GRAPHICS

Distribution of 1 Variable:

```
histogram(~wage, data=CPS85)
densityplot(~wage, data=CPS85)
freqpolygon(~wage, data=CPS85)
```

Scatter plot: xyplot(wage ~ educ, data=CPS85)

Compare distribution by group:

```
bwplot(wage ~ sex, data=CPS85)
```

Can use groups=sex as an argument to xyplot()

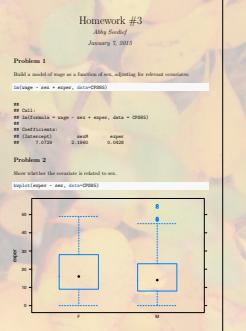
```
densityplot(), or freqpolygon()
```

R MARKDOWN DOCUMENTS

```
## title: "Homework #3"
## author: "Abby Seidoff"
## 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.
lm(wage ~ sex + exper, data=CPS85)
```
## Call:
## lm(wage ~ sex + exper, data = CPS85)
## Coefficients:
## (Intercept) sex        exper  
##             2.308       0.0408 
```

Problem 2
Show whether the covariate is related to sex.
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

## BASIC STATISTICAL TESTS

Difference between two means

```
res <- t.test(wage ~ sex, data=CPS85, mu=1.50)
```

Difference between two proportions

```
res <- prop.test(sex ~ union, data=CPS85)
```

For terse output use pval(res) or confint(res).

## LINEAR MODELS

```
res <- lm(wage ~ sex + educ, data=CPS85)
```

For lm() use summary(res), anova(res), pval(res)

or confint(res).

## RANDOMIZATION AND ITERATION

RESAMPLE/BOOTSTRAP:

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

RANDOM PERMUTATIONS:

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

1000 trials of flipping 6 coins, count heads

```
flips <- do(1000) * rflip(6)
tally(~heads, data=flips)
```

10000 trials of adding three dice

```
scores <- do(10000)*sum(resample(1:6,size=3))
```

```
freqpolygon(~result, data=scores)
```



**START TEACHING WITH R** provides tactics, activities, and techniques to help teach effectively with modern computation. Too often, the emphasis is on statistical computation: using computers to carry out calculations from the pre-computer era. **START TEACHING WITH R** promotes another approach: building a pedagogy that's based in the capabilities of modern computing to enhance understanding of statistical concepts. It's part of a series of books designed to help educators use computing to teach more effectively.

Other books in the series include:

- **START MODELING WITH R**
- **A COMPENDIUM OF COMMANDS FOR TEACHING STATISTICS USING R**

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.

**Randall Pruim**, **Nicholas Horton**, and **Daniel Kaplan** are professors at Calvin College, Amherst College, and Macalester College respectively. They are co-PIs on the National Science Foundation supported Project MOSAIC (NSF DUE-0920350) which aims to help teachers unify modeling, statistics, computation and mathematics.



PRUIM



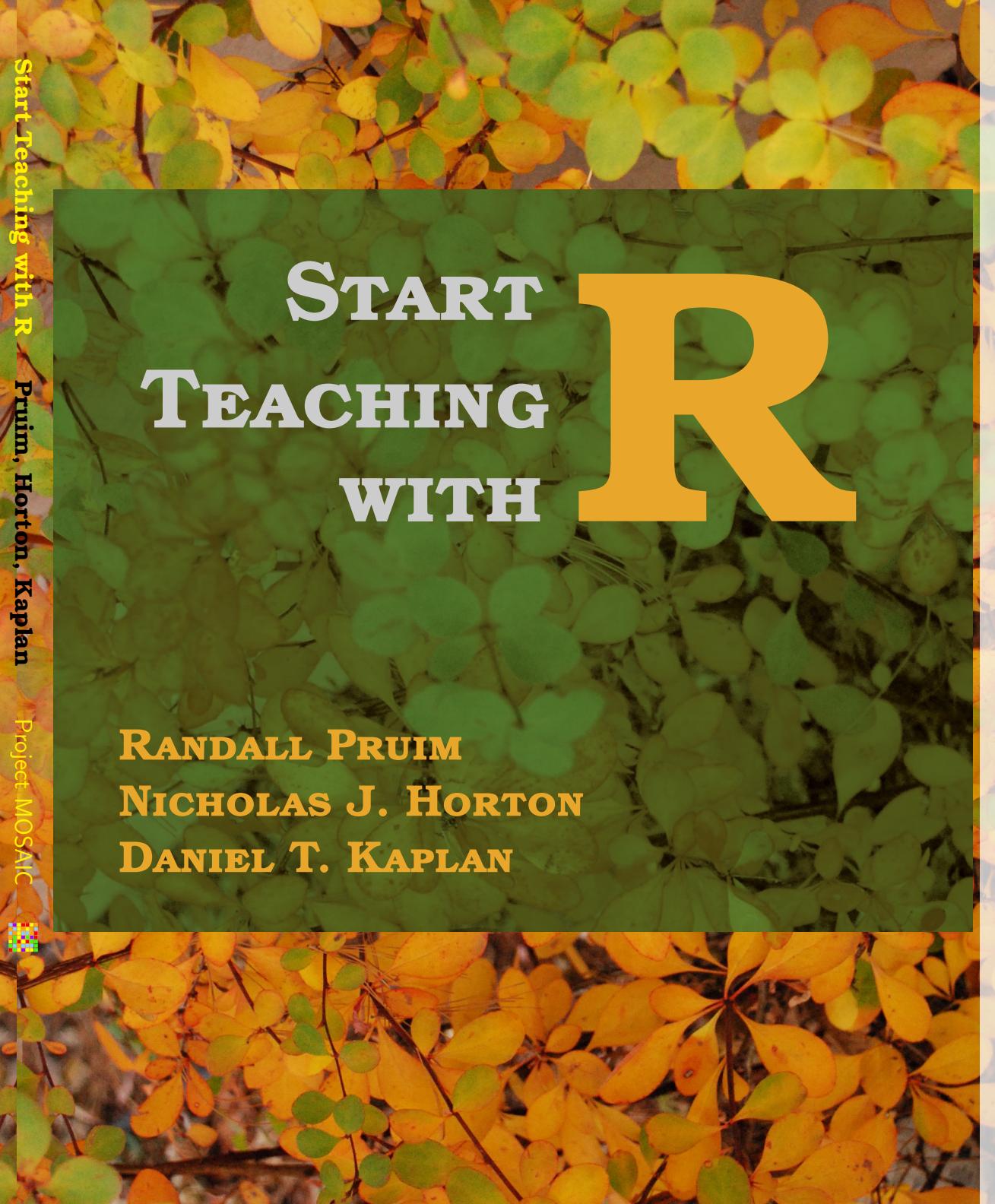
HORTON



KAPLAN

## Some other books by the authors:

*Foundations and Applications of Statistics: An Introduction Using R* (RJP),  
*Statistical Modeling: A Fresh Approach* (DTK), *Using R for Data Management, Statistical Analysis & Graphics* (NJH & KK)



## LOAD PACKAGES

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

## ESSENTIAL R SYNTAX

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

```
rflip(10)
rflip(10, prob=0.3)
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 for how y "depends on" by x, or "is modeled by" x using variables in mydata  
Examples:

```
favstats(homeless~sex, data=HELPrc)
| .group min Q1 median Q3 max mean ...
| 1 female 21 31 35 40.5 58 36.25234 ...
| 2 male 19 30 35 40.0 60 35.46821 ...
```

```
quantile(age~sex,data=HELPrc,p=c(.2,.8))
| .group 20% 80%
| 1 female 30 42.8
| 2 male 29 41.0
```

Only one variable? It goes to right of ~  
mean(~age, data=HELPrc)

```
| [1] 35.65342
```

## DATA FRAMES

Number of rows:  
nrow(CPS85)

Names of variables:  
names(CPS85)

Add a new variable to a data frame

```
res <- mutate(CPS85, yearly=wage*2000)
```

Drop a variable from a data frame

```
res <- select(CPS85, -married)
```

Extract cases meeting a criterion

```
res <- filter(CPS85, sector=="manag")
```

Random sample of 50 cases

```
mysamp <- sample(CPS85, size=50)
```

File reading and writing

```
myData <- read.file("URL or filename")
```

```
write.csv(myData, "filename.csv")
```

## GRAPHICS INTERACTIVELY (IN RSTUDIO)

```
mplot(CPS85, format="scatter")
```

Other types: "boxplot" "violin"

"frequency" "density" "frequency polygon"