

Hello

## LOAD PACKAGES

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

---

## ESSENTIAL R SYNTAX

```
Function & arguments:           rflip(10)  
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 for how y “depends on” by x, or “is modeled by” x using variables in mydata

Examples:

```
favstats(homeless~sex, data=HELPrct)  
| .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=HELPrct,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=HELPrct)  
  
| [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, "scatter")
```

Other types: "boxplot" "violin"

"frequency" "density" "frequency polygon"

---

## 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()
```

---

## RMARKDOWN DOCUMENTS

```
---
title: "Homework #3"
author: "Abby Seedief"
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)
```
```

### Homework #3

Abby Seedief  
January 7, 2015

#### Problem 1

Build a model of wage as a function of sex, adjusting for relevant covariates.

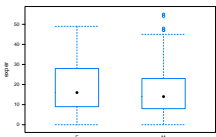
```
lm(wage ~ sex + exper, data=CPS85)
```

```
## Call:
## lm(formula = wage ~ sex + exper, data = CPS85)
##
## Coefficients:
## (Intercept)      sex      exper
##    7.0729      2.1960      0.0428
```

#### 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)
```



---

## 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 + educ, data=CPS85)
```

```
summary(mod) or anova(mod) or confint(mod)
```

## EXTRACT MODEL FUNCTION:

```
fun <- makeFun(mod)
```

```
fun(sex="F",educ=10)
```

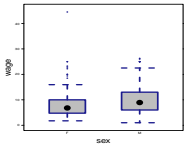
```
plotFun(fun(sex="F",educ=x)} ~ x,x.lim=range(0,8))
```

---

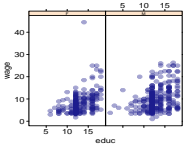
## GRAPHICS FORMULA SYNTAX

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

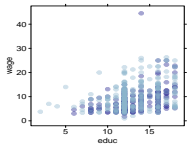
- `y` — y-axis variable (OPTIONAL)
- `x` — x-axis variable (**REQUIRED**)
- `z` — facet-by variable (OPTIONAL)
- `w` — color-by variable (OPTIONAL)



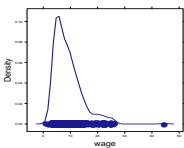
LEFT: `bwplot(wage~sex, data= CPS85)`



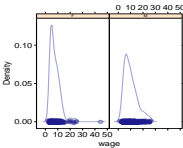
MIDDLE: `xyplot(wage~educ | sex, data= CPS85)`



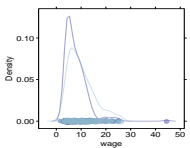
RIGHT: `xyplot(wage~educ, groups=sex, data=CPS85)`



LEFT: `densityplot(~wage, data= CPS85)`



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



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