A mosaic Sampler

Randall Pruim, Calvin College

Taking Advantage of Formulas – Numerical Summaries

```
> mean( age, data=HELP )
[1] 35.7

> mean( age ~ sex, data=HELP )
     sex    S   N Missing
1 female 36.3 107       0
2   male 35.5 346       0

> sd( age ~ sex + homeless, data=HELP )
     sex homeless    S   N Missing
1 female homeless 6.66  40       0
2   male homeless 8.61 169       0
3 female   housed 8.13  67       0
4   male   housed 6.71 177       0
```

Also works for `var()`, `median()`, `max()`, `min()`, `IQR()`, `sum()`, `prop()`, `count()`

Taking Advantage of Formulas – Testing

```
> faithful$long <- faithful$eruptions > 3
> binom.test( ~long , faithful )

Exact binomial test

data:  faithful$long
number of successes = 175, number of trials = 272,
p-value = 2.609e-06
alternative hypothesis: true probability of success is not equal t
95 percent confidence interval:
 0.583 0.700
sample estimates:
probability of success
                 0.643
```

These also work:

```
> binom.test( faithful$long )
> binom.test( faithful$eruptions > 3 )
> prop.test( faithful$eruptions > 3 )
```

Just the Facts Ma'am

R's output can sometimes be overly verbose for beginners.

```
> interval( t.test( age~sex , data=HELP ) )
mean in group female   mean in group male
             36.25                35.47
             lower                upper
             -0.88                 2.45

> pval( t.test( age~sex , data=HELP ) )
p.value
  0.354
```

(Remember this p-value for later.)

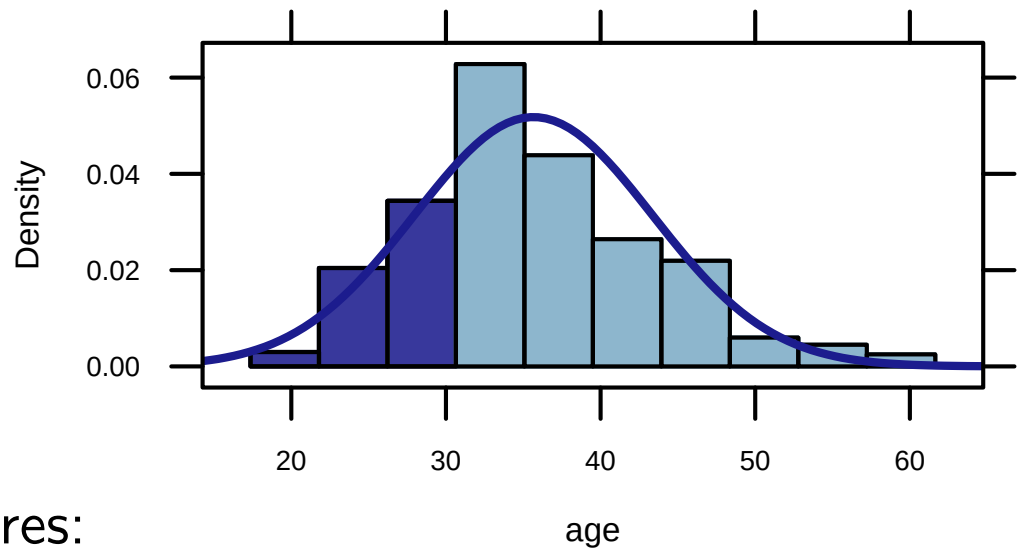
Xtra Output

Other times, it is too terse.

```
> xchisq.test( xtabs( ~ sex + substance, data=HELP ) )
Pearson's Chi-squared test

data:  xtabs(~sex + substance, data = HELP)
X-squared = 2.03, df = 2, p-value = 0.3631

  36.00    41.00    30.00
( 41.81) ( 35.90) ( 29.29)
[0.8068] [0.7236] [0.0173]
<-0.898> < 0.851> < 0.131>

 141.00   111.00    94.00
(135.19) (116.10) ( 94.71)
[0.2495] [0.2238] [0.0053]
< 0.500> <-0.473> <-0.073>

key:
observed
(expected)
[contribution to X-squared]
<residual>
```

Xtra Graphics

```
> xhistogram( ~age , data=HELP, fit='normal',
+             groups = age > 30)
```

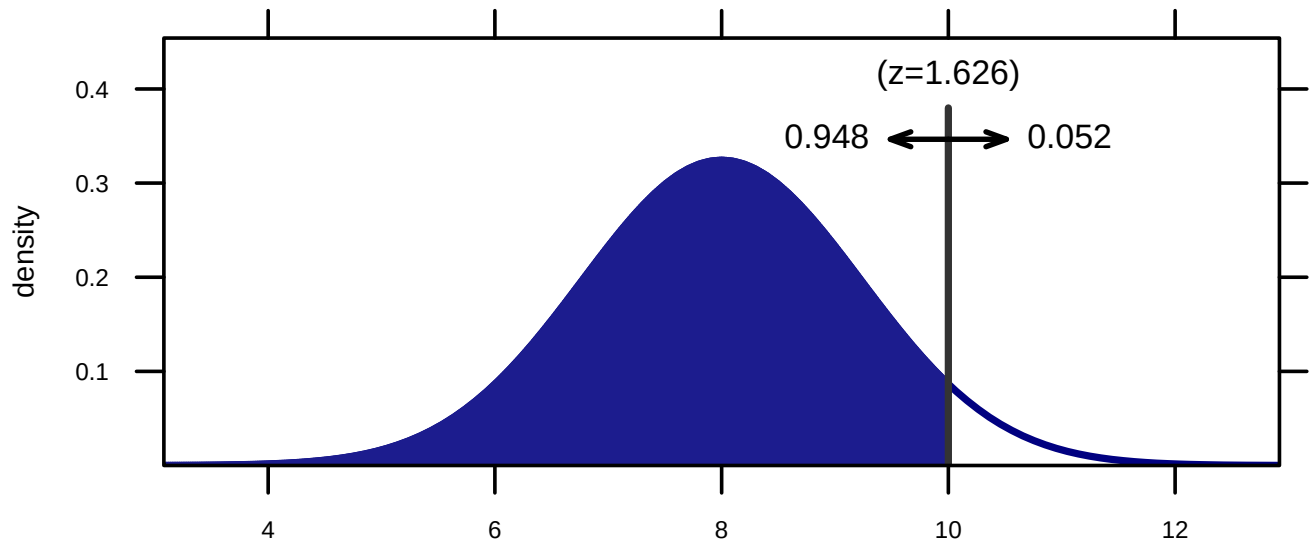


Other features:

- Easy horizontal and vertical reference lines.
- Uses `fitdistr` from `MASS` for fitting.

More Xtras

```
> xpnorm( 10, mean = 8, sd = 1.23 )
If X ~ N(8,1.23), then

P(X <= 10) = P(Z <= 1.626) = 0.948
P(X >  10) = P(Z >  1.626) = 0.052

[1] 0.948
```



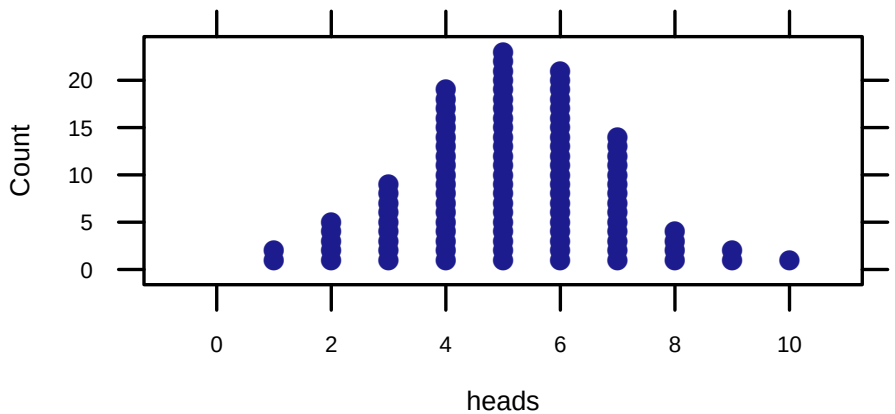
Tossing Coins

```
> rflip(10)
Flipping 10 coins [ Prob(Heads) = 0.5 ] ...

H T T H H H H H T H

Result: 7 heads.
```

This can be used for simulations early in the course, before students know what a binomial distribution is. We just need a way to replicate the coin tossing easily.

Tossing Coins

```
> rflip(10)
Flipping 10 coins [ Prob(Heads) = 0.5 ] ...

T T T T H T T H H H
> x <- do(100) * rflip(10); head(x, 2)
   n heads tails
1 10     2     8
2 10     5     5

> dotPlot( ~heads, data = x, breaks=seq(-0.5, 10.5, by=1 ) )
```



do()ing the mosiac shuffle()

```
> do(1) * lm( age ~ sex, HELP )
  Intercept sexmale sigma r-squared
1      36.3  -0.784  7.71   0.00187

> null.dist <- do(1000) * lm( age ~ shuffle(sex), HELP )
> head(null.dist, 1)
  Intercept sexmale sigma r-squared
1      35.3   0.452  7.72  0.000621

> with(null.dist, perctable( abs(sexmale) > 0.784 ) )
FALSE  TRUE
 62.3  37.7
```

Data is a Distribution, Too

We have implemented functions paralleling `rnorm()`, `pnorm()`, and `qnorm()` for data distributions.

```
> rdata( 10, HELP$age )
 [1] 21 36 41 33 36 21 38 47 28 21
> pdata( 30, HELP$age )
[1] 0.256
> qdata( .50, HELP$age )
50%
 35
```

Calculus

Differentiation:

```
> f <- D(A * sin(x + B) ~ x,
+        A = 1, B = 0)


> f( pi )
[1] -1
> f( pi, A=3, B=pi )
[1] 3


> randx <- runif(4, -pi, pi)


> f(randx) - cos(randx)
[1] 0 0 0 0
```

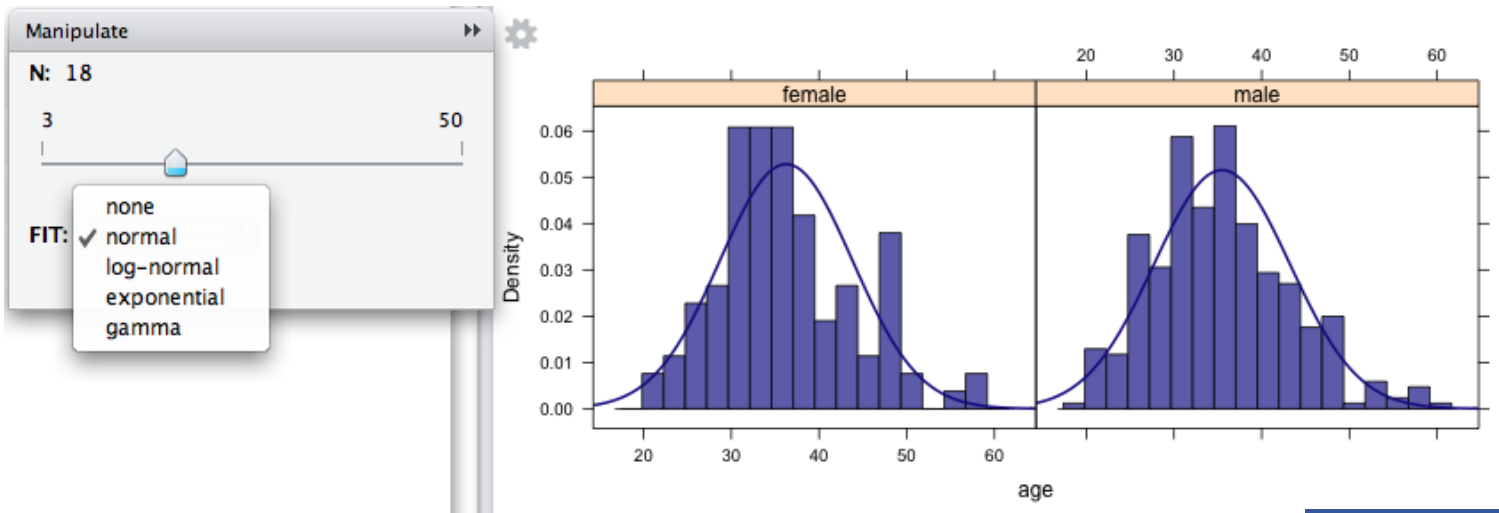
Anti-differentiation:

```
> F <- antiD( dnorm(x) ~ x )
> # F(0) == 0 by default
> F (randx) - pnorm(randx)
[1] -0.5 -0.5 -0.5 -0.5

> # Using G(-Inf) == 0 gives pdf
> G <- antiD( dnorm(x) ~ x ,
+             from=-Inf)


> G (randx) - pnorm(randx)
[1] -2.13e-10 -2.13e-10 -2.13e-10
[4] -2.13e-10


> G(2)
[1] 0.977
```

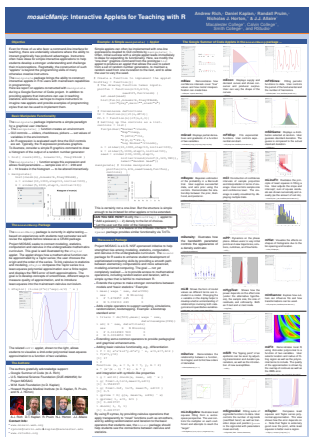
But Wait, There's More! – mosaicManip

The mosaicManip package takes advantage of the manipulate package from RStudio to provide interactive applets for Statistics and Calculus.

```
> mhistogram( ~ age | sex, data=HELP )
```



- In alpha testing now.
- Come see our poster for many more examples.



The mosaic team

R Pruim

Calvin C

D Kaplan

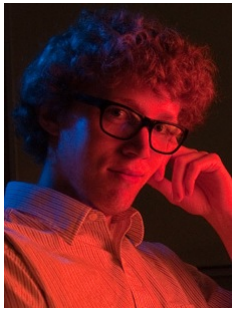
Macalaster C

N Horton

Smith C

JJ Alaire

RStudio



A Rich

Macalester C

You

http://www.mosaic-web.org

https://r-forge.r-project.org/projects/mosaic/

The mosaic team

R Pruim

Calvin C

D Kaplan

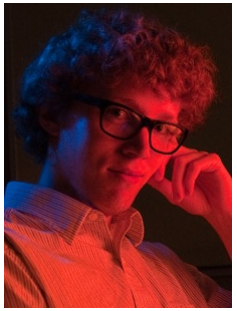
Macalaster C

N Horton

Smith C

JJ Alaire

RStudio



A Rich

Macalester C

You

http://www.mosaic-web.org

https://r-forge.r-project.org/projects/mosaic/