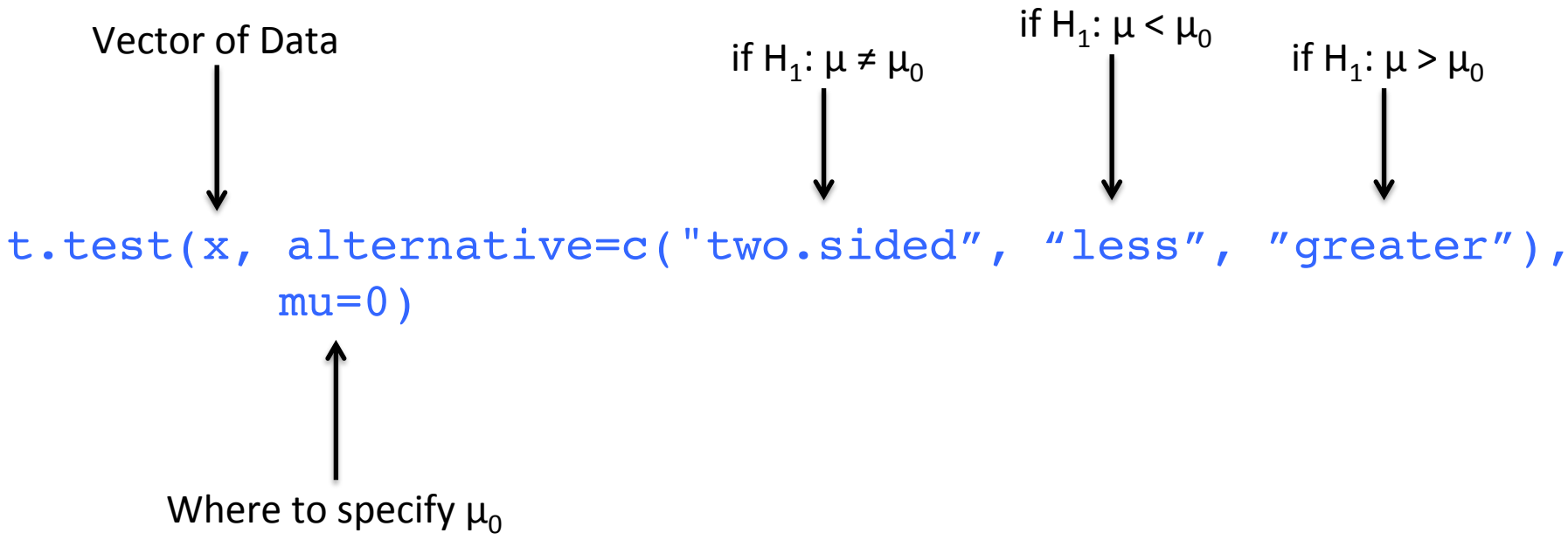
Using R for Hypothesis Tests

Keegan Korthauer
Department of Statistics
UW Madison

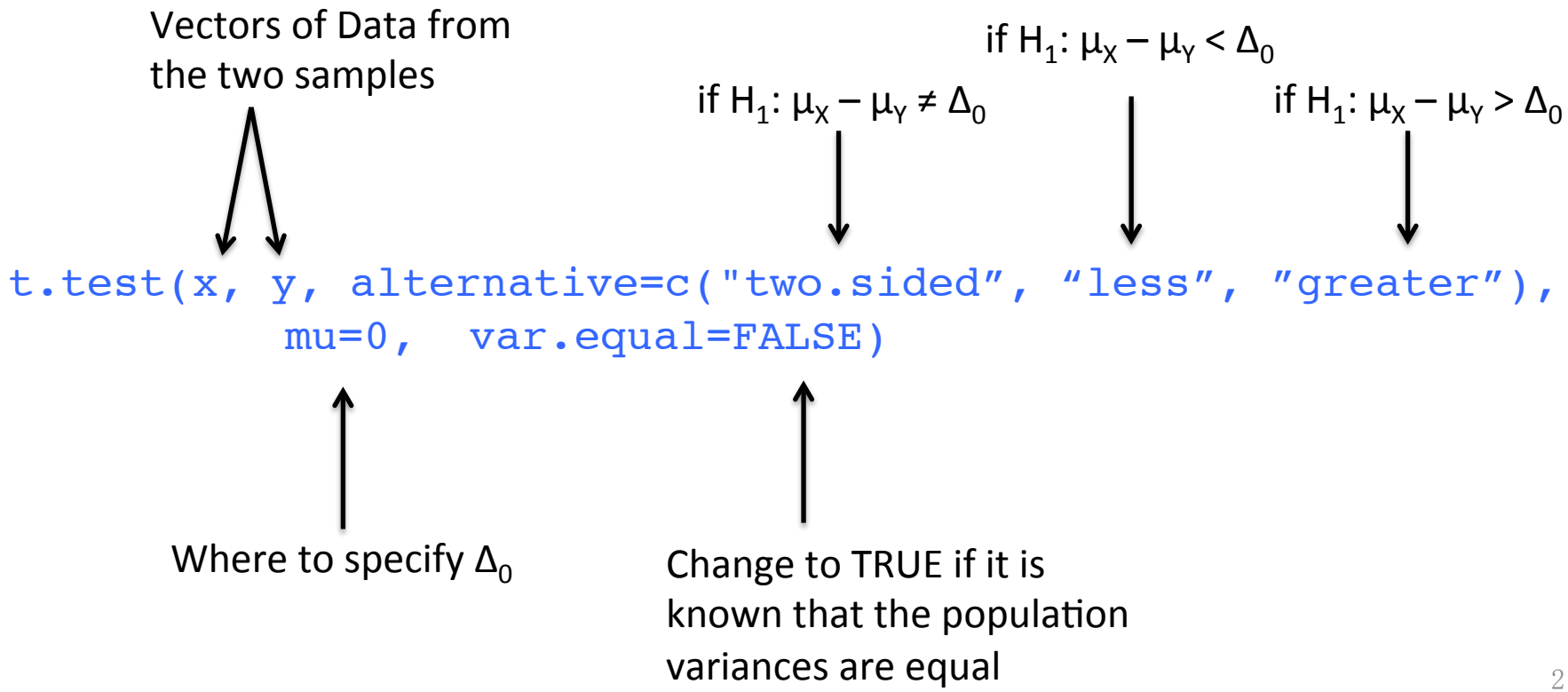
One Sample t-test

The `t.test()` function will perform a test for a population mean for a small (normally distributed) sample:



Two Sample t-test

The `t.test()` function will also perform a test for the difference in population means for two independent small (normally distributed) samples:



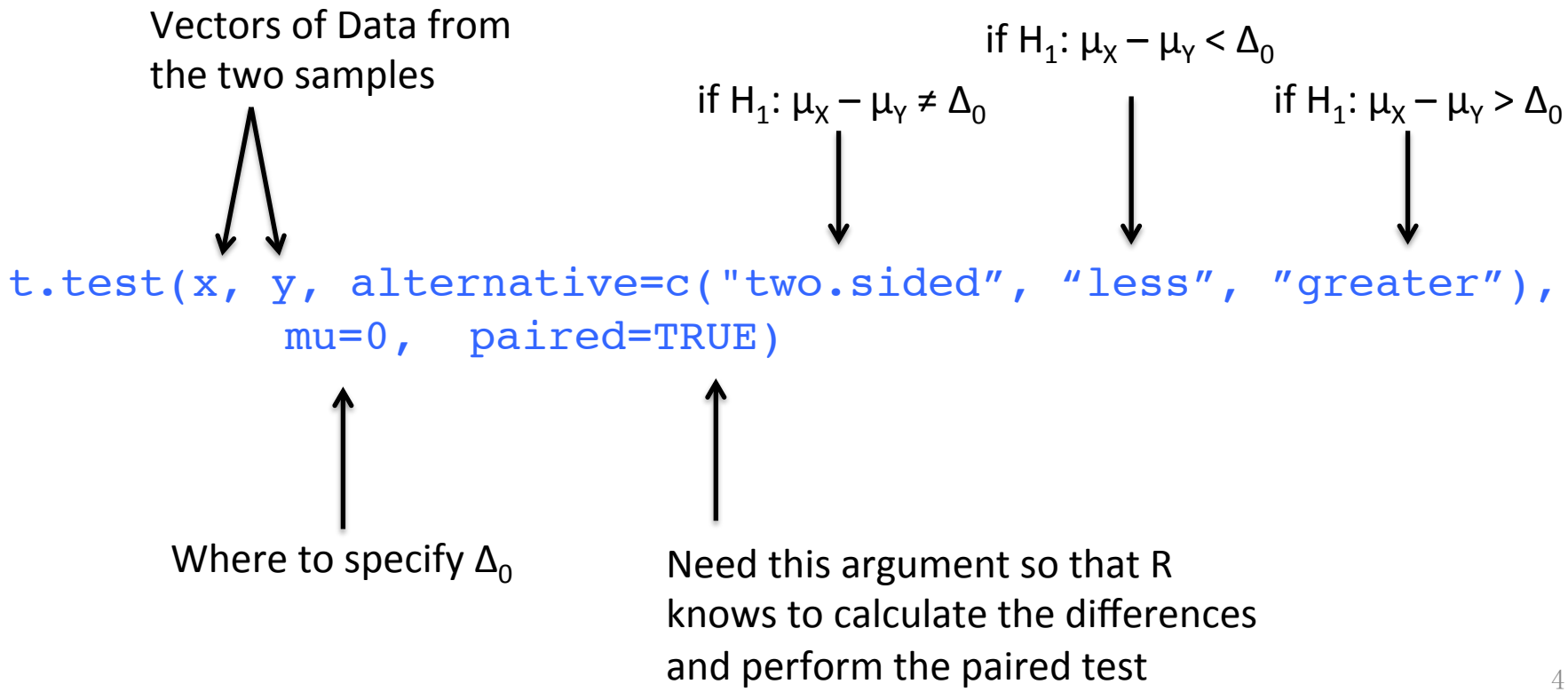
Example

```
> x <- c(12, 13, 15, 19, 20, 21, 27)
> y <- c(18, 23, 24, 30, 32, 35, 40)
> t.test(x, y, alternative="two.sided", mu=0,
var.equal=FALSE)

    Welch Two Sample t-test

data:  x and y
t = -3.0636, df = 10.636, p-value = 0.01118
alternative hypothesis: true difference in means is not
equal to 0
95 percent confidence interval:
 -18.444153  -2.984419
sample estimates:
mean of x mean of y
 18.14286  28.85714
```

We rounded this down to the nearest integer; p-values will differ slightly than "by-hand" method

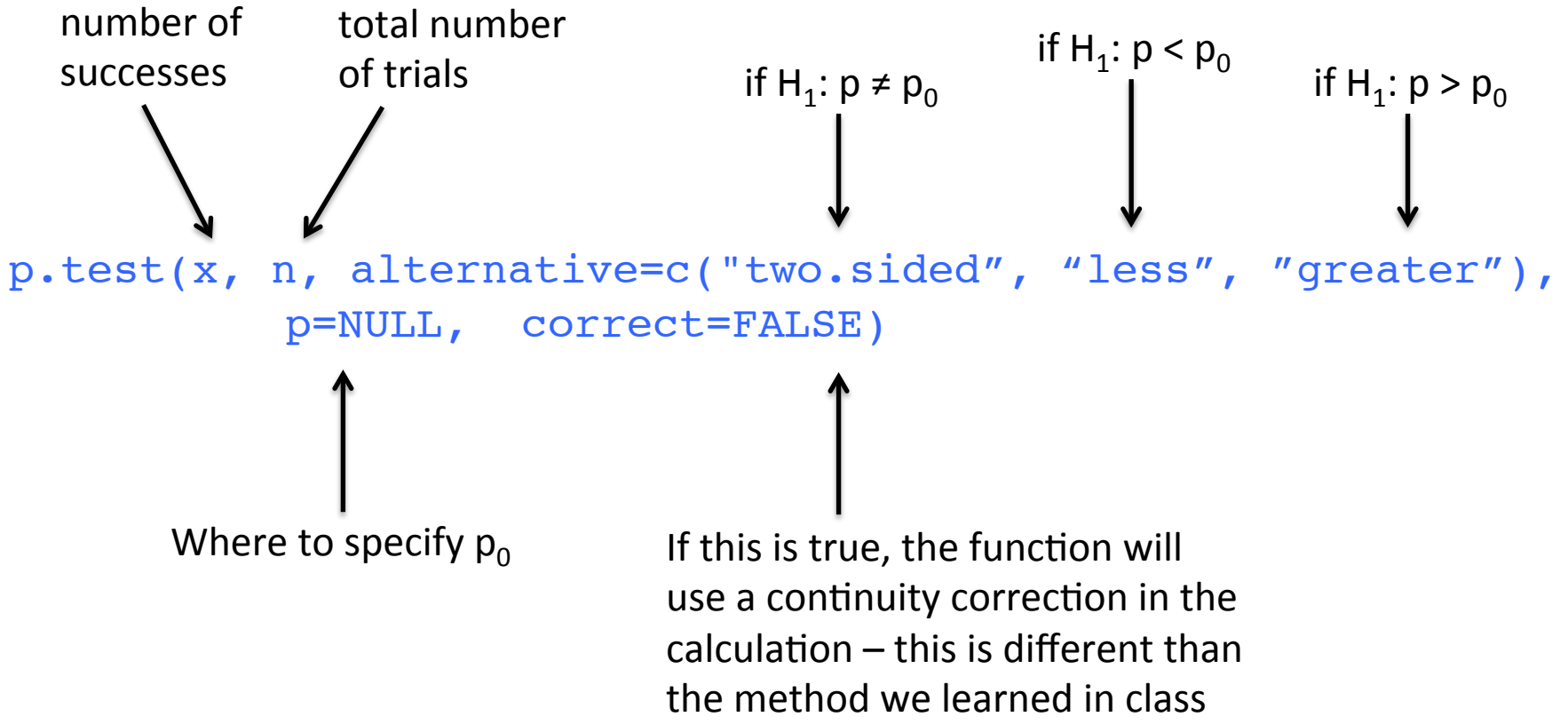
Paired Sample t-test

The `t.test()` function will also perform a test for the difference in population means for two paired samples (differences normally distributed):



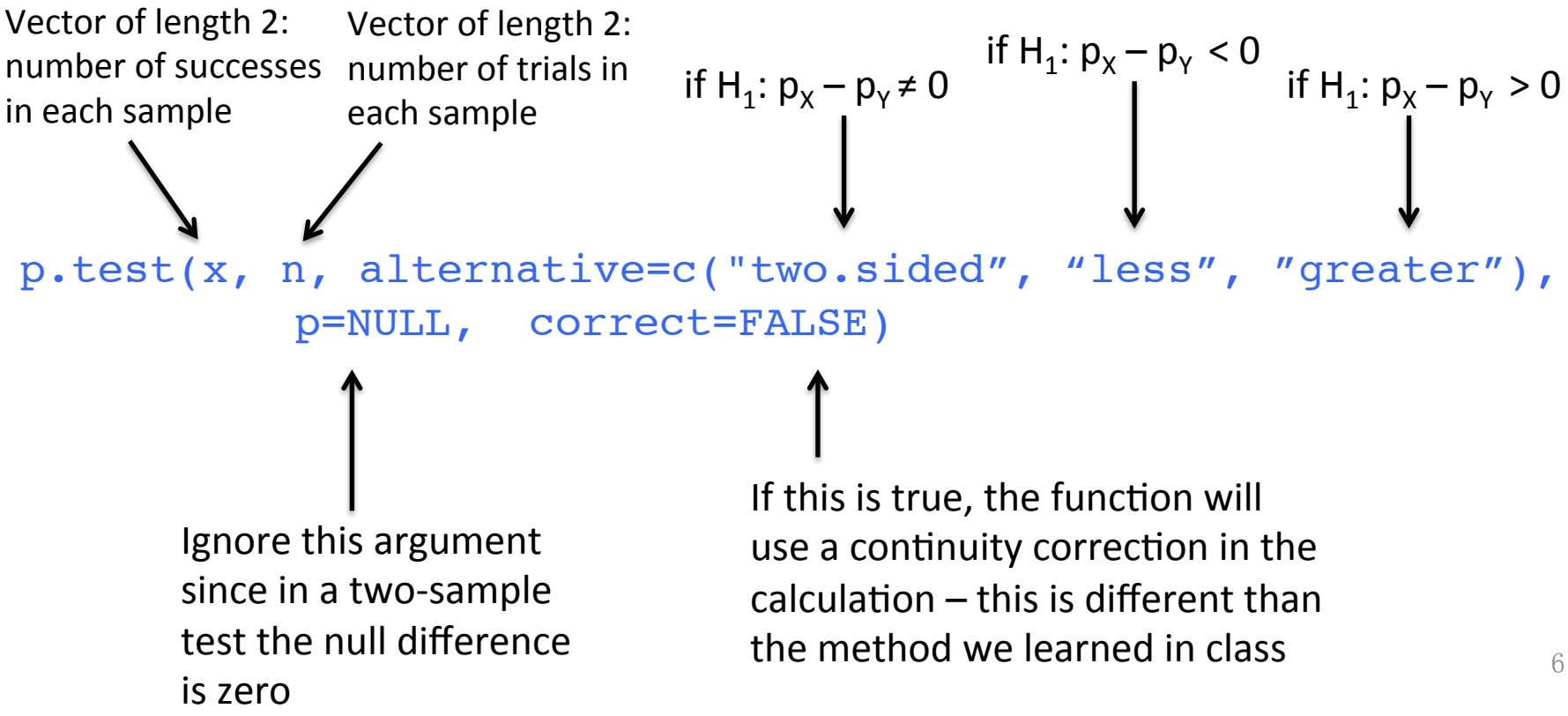
One Sample Test of Proportions

The `p.test()` function will perform a test for a population proportion for a large sample (at least 10 successes and failures):



Two Sample Test of Proportions

The `p.test()` function will also perform a test for the difference in population proportions for two independent large samples (at least 10 successes and failures):



Example

```
> # voted for candidate A in two districts
> x   <- c(120, 334)
> n <- c(450, 1067)
> prop.test(x, n, correct=FALSE)

    2-sample test for equality of proportions without continuity
    correction

data:  x out of n
X-squared = 3.2439, df = 1, p-value = 0.07169
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.095793043  0.003072018
sample estimates:
   prop 1    prop 2
0.2666667 0.3130272
```

Function reports the squared value of the Z test statistic we learned

Chi-square Test

The `chisq.test()` function will perform a Chi square test of independence/homogeneity:

Table of Data

```
chisq.test(x)
```

Get a table of data read into an object `x` by using the function

```
matrix(data, nrow, ncol, byrow=FALSE)
```

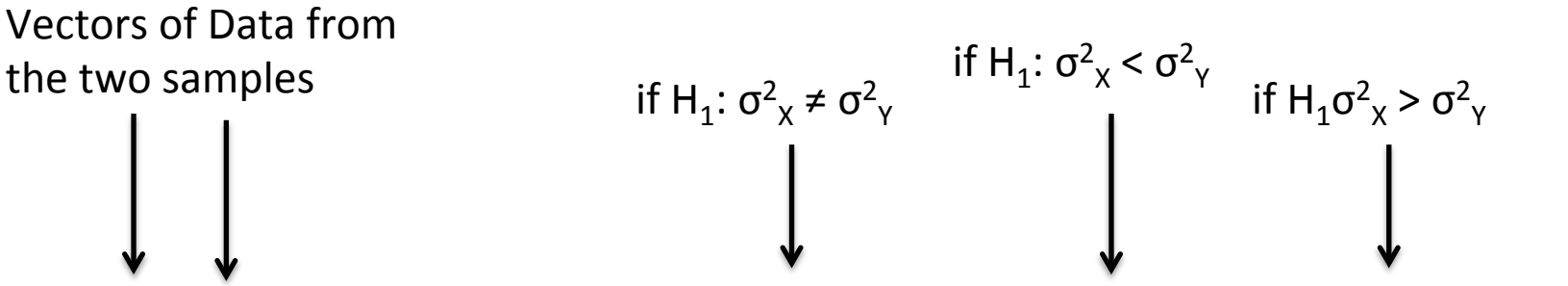
which fills an `nrow` by `ncol` matrix, by column, from the vector data. (Change the default `byrow=FALSE` to `byrow=TRUE` to fill by row instead.)

Example

```
> titanic <- matrix(c(212, 202, 118, 178, 673, 123,
+ 167, 528), nrow=2, ncol=4, byrow=TRUE)
> #check to make sure we read the table in correctly
> titanic
     [,1] [,2] [,3] [,4]
[1,]  212  202  118  178
[2,]  673  123  167  528
> chisq.test(titanic)

   Pearson's Chi-squared test

data:  titanic
X-squared = 187.7932, df = 3, p-value < 2.2e-16
```

Test for Equality of Variance

The `var.test()` function will test for the equality of variance for samples from two normal populations




```
var.test(x, y, alternative=c("two.sided" "less", "greater"))
```

Resources

- Work through the examples in John Gillett's R notes (note that section numbers refer to a different text)
  - http://pages.cs.wisc.edu/~jgillett/224/R/6/
  - http://pages.cs.wisc.edu/~jgillett/224/R/7/

- Note that we didn't talk about a function to perform large-sample z-tests
  - There isn't one, because (1) the t-test is used much more often and (2) as the sample gets large, the t-test will be almost the same as the z-test

Next

- Intro to Correlation
- Friday: HW 9 due and Sample Exam 2 posted
- Monday: Review for Exam 2
- Wednesday: Exam 2