# R reminders for exercise

Read .txt files with `read.delim`.

If R doesn't recognize numeric data, use `data.matrix()` to transform numeric parts of a data frame to something that R thinks `is.numeric`.

`cor(x,y)` computes the <Pearson> correlation between two vectors.

# sample() in R

**`sample(x, n, replace, prob)`**

x: vector from which we sample n random elements.

If x is an integer, it is viewed as shorthand for the vector 1:x

Default for **`replace`** is FALSE:  for bootstrapping, must change this.

**`prob`** *weights* the sampling:

**`sample(c(0,1),10,replace=TRUE)`** `--` same as binomial

# bootstrapping in R

Sample can also be used to resample rows of a data matrix or data frame.

```
df1<-data.frame(matrix(rnorm(20), nrow=10))
```

```
    X1         X2
1 -0.5996083  0.80017687
2 -0.1294107 -0.16393097
3  0.8867361  1.24291877
4 -0.1513960 -0.93438506
5  0.3297912  0.39370865
6 -3.2273228  0.40363146
7 -0.7717918 -0.88643672
8  0.2865486 -1.31893760
9 -1.2205120  0.02884391
10 0.4345504 -0.43212979
```

```
df1[sample(nrow(df1), 3), ]
```

```
    X1              X2
10 0.4345504 -0.4321298
5  0.3297912  0.3937087
8  0.2865486 -1.3189376
```

# bootstrap exercise

Save and load into R the data set “dbtable.txt” from today’s date in the Schedule on the course web page.

Recall:

```
medians=replicate(1000, {
  indices=sample(...);
  median(#resampled vector)
})
```

What is the correlation between a1c and glucose in this data set?

What is the 95% confidence interval for the correlation?

Send answers in private chat in Piazza.