

```
#PO 841
#10/8/2014
#Lab 5
```

```
#=====
#DATA MANIPULATION: IF/ELSE and RECODE
#=====
```

```
#The "if/else" and "recode" functions will come in very useful from here on out, as we'll be working with large data sets that we may want to manipulate in order to perform certain analyses.
```

```
#Let's assume we have a matrix of 20 students' scores on each of three exams, with each column representing one student:
```

```
scores<-sapply(1:20, function(x) rnorm(3,50,12))
```

```
scores
```

```
means<-colMeans(scores) #Figure out the mean score for all the exams for each student
```

```
means #We now have a vector of 20 different students' final grades
```

```
grades<-ifelse(means>=50,"P","F") #Let's assign Pass/Fail grades to each student, assuming that passing means an average score >- 50 and failing means anything under 50. The "if/else" function tells R to apply a function selectively, based on a specified condition. We're telling R: IF a student has a mean score of greater-than-or-equal-to 50, then assign a grade of P; if anything ELSE, assign a grade of F.
```

```
grades
```

```
#In essence, we can use the "if/else" function to manipulate and reclassify data:
```

```
#Let's say a researcher has administered a survey of 100 people and has asked them to rate Obama's performance on a scale of 1-7 (1=very much disapprove, 2=disapprove, 3=slightly disapprove, 4=neutral, 5=slightly approve, 6=approve, 7=very much approve)
```

```
results<-sample(7,100, replace=T)
```

```
results
```

```
#Now say you want to use this researcher's data for your own study, and you want to classify everyone who rated Obama a 5 or above as "approving" of him and everyone who rated Obama a 4 or below as "not approving"
```

```
obama<-ifelse(results>=5, "Yes", "No")
```

```
obama
```

```
#You may also want to create a binary variable, where 0=not approving and
```

```
1=approving, so that you can use this approval variable in a regression later on:
```

```
obama<-ifelse(results>=5,1,0)
```

```
obama
```

```
#Let's say you only want to analyze those respondents who felt very strongly about Obama, i.e., those who gave him either a 1 or a 7. Everyone else you want to discard from your data set (by turning their responses into NA):
```

```
obama<-ifelse(results %in% c(2:6), NA, results)
```

```
obama
```

```
sum(obama==1, na.rm=T) #How many people rated Obama a 1?  
sum(obama==7, na.rm=T) #How many rated him a 7?
```

#We can also use the "recode" function to manipulate data in much the same way. But this function requires installing a package (a collection of functions) that an R user wrote in order to perform certain types of operations (this is the benefit of R's being an open-source program!):

```
install.packages("car")  
library(car)
```

#Let's recode all people who rated Obama a 5 or above as "Approve," those who rated him a 3 or below as "Disapprove," and those who rated him a 4 as "Neutral":

```
obama<-recode(results, '1:3="Disapprove"; 4="Neutral"; 5:7="Approve"')  
obama
```

#What if we wanted to use "recode" to classify only those who rated Obama 6 or 7 as "Approve," and wanted to discard everything else?

```
obama<-recode(results, '6:7="Approve"; else=NA')  
obama
```

#What proportion of our respondents approve of Obama if we define "approval" as a rating of a 6 or a 7?

```
p<-(sum(obama=="Approve", na.rm=T))/length(obama)  
p
```

#How would we do this with the "if/else" function?:

```
obama<-ifelse(results>=6, "approve", NA)  
obama
```