

# R-Workshop-James

*James LeBreton with Rick Gilmore*

*2017-08-16 18:07:57*

```
path2data <- "../data/"
```

## PART 1: INSTALLATION, SETTINGS, AND DATA MANAGEMENT

### TOPIC 1: Projects & Directories in R Studio

```
getwd() #get the current working directory
```

```
## [1] "/Users/rick/github/psu-psychology/r-bootcamp/talks"
```

```
#setwd("~/Dropbox/James Work Files/R Workshop/2017") #change the working directory
```

Since ~/Dropbox/James Work Files/R Workshop/2017 is specific to James' computer, it won't work for others. When using an RStudio project, I don't change my working directory. Instead, I just make sure I give relevant functions information about the directories where other resources can be found.

### TOPIC 2: Installing Packages & Loading into Active Library of Resources

Install packages via syntax

```
# Can install by evaluating chunk, but not by "knitting"  
install.packages("multilevel") #Downloading a package to my computer  
#loading packages into working library  
library("multilevel")
```

---

### Understanding How R Searches for Information

```
search()  
detach(package:multilevel)  
search()
```

---

### Obtaining Help

```
#You may inquire about a function using any of the following:  
##If you know the exact name:  
?search  
help(search)
```

```
##If want to search by part of the name
apropos("searc")

## [1] ".rs.getCompletionsSearchPath" ".rs.objectsOnSearchPath"
## [3] ".rs.recursiveSearch"          ".rs.rpc.search"
## [5] "help.search"                  "hsearch_db"
## [7] "hsearch_db_concepts"          "hsearch_db_keywords"
## [9] "RSiteSearch"                  "search"
## [11] "searchpaths"

??sear
```

Another good source of help is StackOverflow.

## TOPIC 3: Data Types & Structures in R

### Numbers

```
x <- 2
x

## [1] 2

y = c(1:3); y

## [1] 1 2 3

z = c("Porsche 911", "Porsche 944", "Porsche 911", "BMW 335xi")
z

## [1] "Porsche 911" "Porsche 944" "Porsche 911" "BMW 335xi"

g=sqrt(x); g

## [1] 1.414214

is.numeric(x)

## [1] TRUE

is.numeric(z)

## [1] FALSE
```

---

### Strings

```
#String Data as character:
z

## [1] "Porsche 911" "Porsche 944" "Porsche 911" "BMW 335xi"

#String Data as factor:
z2=factor(z)
z2
```

```
## [1] Porsche 911 Porsche 944 Porsche 911 BMW 335xi
## Levels: BMW 335xi Porsche 911 Porsche 944
```

```
#Compute the Length of a String (or Numeric) Variable:
nchar(x)
```

```
## [1] 1
```

```
nchar(y)
```

```
## [1] 1 1 1
```

```
nchar(y)
```

```
## [1] 1 1 1
```

```
nchar(z)
```

```
## [1] 11 11 11 9
```

```
#nchar(z2) Throws error during rendering
```

---

## Logical Data

```
##Assumes values of TRUE or FALSE
###TRUE is considered equal to 1
###FALSE is considered equal to 0
TRUE*5
```

```
## [1] 5
```

```
sqrt(TRUE)
```

```
## [1] 1
```

```
t=TRUE
```

```
# you can test if a variable type is logical using:
is.logical(x)
```

```
## [1] FALSE
```

```
is.logical(t)
```

```
## [1] TRUE
```

```
# Logical data types also used as input to functions (see Day 2 examples)
2==2
```

```
## [1] TRUE
```

```
2==3
```

```
## [1] FALSE
```

---

## Vectors

```
#Vectors - 1 dimensional collections of same type data
v1=1:5; v1 #creating vector of numbers

## [1] 1 2 3 4 5
v2=c(1,2,3,4,5); v2

## [1] 1 2 3 4 5
v3=c("Porsche 911", "Ford Mustang GT", "Plymouth Baracuda", "Chevrolet Camaro", "Honda Pilot LX")
v1; v2; v3

## [1] 1 2 3 4 5
## [1] 1 2 3 4 5
## [1] "Porsche 911"      "Ford Mustang GT"    "Plymouth Baracuda"
## [4] "Chevrolet Camaro"  "Honda Pilot LX"

#Matrices - 2 dimensional collections of same type data
m=matrix(1:20, nrow=5); m

##      [,1] [,2] [,3] [,4]
## [1,]    1    6   11   16
## [2,]    2    7   12   17
## [3,]    3    8   13   18
## [4,]    4    9   14   19
## [5,]    5   10   15   20
```

---

## Arrays & Data Frames

```
#Arrays - multidimensional collection of same type data
#example of 3D array
a=array(1:20, dim=c(2,5,2)); a

## , , 1
##
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1    3    5    7    9
## [2,]    2    4    6    8   10
##
## , , 2
##
##      [,1] [,2] [,3] [,4] [,5]
## [1,]   11   13   15   17   19
## [2,]   12   14   16   18   20

#Creating a data frame from vectors
eng=c("Flat-6", "V-8", "V-8", "V-8", "V-6")
doors=c(2,2,2,2,4)
data1=data.frame(v2, v3, eng, doors)

# Viewing content of data framees
```

```
# Look at the "enviroment" tab in the upper left panel
# Click on one of the data frames listed under Data (e.g., "data1")
# Or, simply type:
```

```
data1
```

```
##   v2          v3    eng doors
## 1  1    Porsche 911 Flat-6    2
## 2  2   Ford Mustang GT      V-8    2
## 3  3 Plymouth Baracuda      V-8    2
## 4  4   Chevrolet Camaro      V-8    2
## 5  5   Honda Pilot LX       V-6    4
```

```
# Obtain a list of the variable names in a data frame
```

```
names(data1)
```

```
## [1] "v2"    "v3"    "eng"   "doors"
```

```
# Change the names of the variables in a data frame
```

```
data2=data.frame(id=v2, model=v3, eng=eng, doors=doors) #creates a new data frame
```

```
data1
```

```
##   v2          v3    eng doors
## 1  1    Porsche 911 Flat-6    2
## 2  2   Ford Mustang GT      V-8    2
## 3  3 Plymouth Baracuda      V-8    2
## 4  4   Chevrolet Camaro      V-8    2
## 5  5   Honda Pilot LX       V-6    4
```

```
data2
```

```
##   id          model    eng doors
## 1  1    Porsche 911 Flat-6    2
## 2  2   Ford Mustang GT      V-8    2
## 3  3 Plymouth Baracuda      V-8    2
## 4  4   Chevrolet Camaro      V-8    2
## 5  5   Honda Pilot LX       V-6    4
```

```
data3=data1 #make a copy of the original dataframe
```

```
install.packages("plyr")
```

```
library(plyr)
```

```
data3=rename(data3, replace=c("v2"="id","v3" = "model")) #renames specific variables
```

```
data3
```

```
names(data1)=c("id","model", "eng", "doors") #replaces names of all variables in existing data frame
```

```
data1
```

## TOPIC 4: Reading Data Files into R

### Reading Data - From R Data Sets

```
##List of avaiialble data sets
```

```
data()
```

```
library(multilevel)
```

```
#List data in the multilevel package
```

```
data(package="multilevel")
#load the univ data frame into R environment
data(univbct, package="multilevel")
d=univbct

#Confirm it is loaded as a data frame
class(d)

## [1] "data.frame"
```

---

## Saving data frames as comma-separated value (CSV)

```
#Saving a data frame as a .csv file (to be read into SPSS, Excel, Text Editor, etc.)
write.table(d, file = paste0(path2data, "d2.csv"), sep=",", row.names=F)
write.table(d, paste0(path2data, "d1.csv"), sep=",", row.names=FALSE)

#save the data as a text file to be read into SPSS
install.packages("foreign")
library("foreign")
write.foreign(univbct,
              datafile=paste0(path2data, "univbct.csv"),
              codefile=paste0(path2data, "univbct.sps"),
              package="SPSS")
file.show(paste0(path2data, "univbct.csv"))
file.show(paste0(path2data, "univbct.sps"))
```

---

## Reading data from SPSS

```
library("foreign")
demo1=read.spss(file=paste0(path2data, "demo1.sav"),
                use.value.labels=TRUE,
                to.data.frame=TRUE,
                use.missings=TRUE)
summary(demo1)
```

```
##      SUBNUM      TIME      BTN      COMPANY
##  Min.   : 1.00   Min.   :0   Min.   : 4.0   A       :246
##  1st Qu.: 75.75  1st Qu.:0   1st Qu.: 377.8  HHC      :210
##  Median :150.50  Median :1   Median :1022.0  B       :207
##  Mean   :150.50  Mean    :1   Mean   :1860.3  D       :114
##  3rd Qu.:225.25  3rd Qu.:2   3rd Qu.:3066.0  C       : 84
##  Max.   :300.00  Max.    :2   Max.   :4042.0  SVC      : 24
##                                     (Other): 15
##      MARITAL      GENDER      HOWLONG      RANK
##  Min.   :1.000   Min.   :1.000   Min.   :0.000   Min.   :11.00
##  1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:13.00
##  Median :2.000   Median :1.000   Median :2.000   Median :14.00
##  Mean   :1.711   Mean    :1.039   Mean    :2.371   Mean   :15.26
##  3rd Qu.:2.000   3rd Qu.:1.000   3rd Qu.:4.000   3rd Qu.:16.00
```

```
## Max. :5.000 Max. :2.000 Max. :5.000 Max. :32.00
## NA's :6 NA's :51 NA's :18 NA's :48
## EDUCATE AGE
## Min. :1.000 Min. :18.00
## 1st Qu.:2.000 1st Qu.:20.00
## Median :2.000 Median :24.00
## Mean :2.663 Mean :25.75
## 3rd Qu.:3.000 3rd Qu.:30.00
## Max. :6.000 Max. :44.00
## NA's :9 NA's :9
```

```
demo2=read.spss(file=paste0(path2data, "demo2.sav"),
  use.value.labels=T,
  to.data.frame=T,
  use.missings=FALSE)
summary(demo2) #oops, GENDER = 999 was a missing values code
```

```
## SUBNUM TIME BTN COMPANY MARITAL
## Min. :301 Min. :0 Min. : 4 A :156 Min. :1.000
## 1st Qu.:349 1st Qu.:0 1st Qu.: 404 HHC :144 1st Qu.:1.000
## Median :398 Median :1 Median :1022 B :141 Median :2.000
## Mean :398 Mean :1 Mean :1755 D : 69 Mean :1.756
## 3rd Qu.:447 3rd Qu.:2 3rd Qu.:3066 C : 42 3rd Qu.:2.000
## Max. :495 Max. :2 Max. :4042 SVC : 15 Max. :5.000
## (Other): 18 NA's :6
## GENDER HOWLONG RANK EDUCATE
## Min. : 1.00 Min. :0.000 Min. :11.0 Min. :1.00
## 1st Qu.: 1.00 1st Qu.:2.000 1st Qu.:13.0 1st Qu.:2.00
## Median : 1.00 Median :2.000 Median :14.0 Median :2.00
## Mean : 88.03 Mean :2.446 Mean :14.7 Mean :2.49
## 3rd Qu.: 1.00 3rd Qu.:3.000 3rd Qu.:15.0 3rd Qu.:2.00
## Max. :999.00 Max. :5.000 Max. :31.0 Max. :6.00
## NA's :6 NA's :27 NA's :3
## AGE
## Min. :18.00
## 1st Qu.:21.00
## Median :24.00
## Mean :25.68
## 3rd Qu.:29.00
## Max. :46.00
## NA's :3
```

```
demo2=read.spss(file=paste0(path2data, "demo2.sav"),
  use.value.labels=T,
  to.data.frame=T,
  use.missings=T)
names(demo1); names(demo2)
```

```
## [1] "SUBNUM" "TIME" "BTN" "COMPANY" "MARITAL" "GENDER" "HOWLONG"
## [8] "RANK" "EDUCATE" "AGE"
## [1] "SUBNUM" "TIME" "BTN" "COMPANY" "MARITAL" "GENDER" "HOWLONG"
## [8] "RANK" "EDUCATE" "AGE"
```

```
#Reading data (csv)
data1=read.csv(paste0(path2data, "data1.csv"), header=T)
```

```
data2=read.csv(paste0(path2data, "data2.csv"))
```

*#Now click on "Environment" tab and the "data1" dataframe*  
*#NA (not available) is automatically inserted by R for any missing data*  

```
head(data1) # display first 6 cases
```

```
## SUBNUM TIME JOBSAT1 COMMIT1 READY1 JOBSAT2 COMMIT2 READY2 JOBSAT3
## 1 1 0 1.666667 1.666667 2.75 1 1.666667 1 3
## 2 1 1 1.666667 1.666667 2.75 1 1.666667 1 3
## 3 1 2 1.666667 1.666667 2.75 1 1.666667 1 3
## 4 2 0 3.666667 1.666667 3.00 4 1.333333 2 4
## 5 2 1 3.666667 1.666667 3.00 4 1.333333 2 4
## 6 2 2 3.666667 1.666667 3.00 4 1.333333 2 4
## COMMIT3 READY3 JSAT COMMIT READY
## 1 3.000000 3.00 1.666667 1.666667 2.75
## 2 3.000000 3.00 1.000000 1.666667 1.00
## 3 3.000000 3.00 3.000000 3.000000 3.00
## 4 1.333333 1.75 3.666667 1.666667 3.00
## 5 1.333333 1.75 4.000000 1.333333 2.00
## 6 1.333333 1.75 4.000000 1.333333 1.75
```

```
tail(data1) # display last 6 cases
```

```
## SUBNUM TIME JOBSAT1 COMMIT1 READY1 JOBSAT2 COMMIT2 READY2 JOBSAT3
## 895 299 0 2.333333 4.333333 2.75 3.666667 3.666667 2.75 4
## 896 299 1 2.333333 4.333333 2.75 3.666667 3.666667 2.75 4
## 897 299 2 2.333333 4.333333 2.75 3.666667 3.666667 2.75 4
## 898 300 0 2.666667 3.666667 2.25 2.666667 4.333333 3.75 3
## 899 300 1 2.666667 3.666667 2.25 2.666667 4.333333 3.75 3
## 900 300 2 2.666667 3.666667 2.25 2.666667 4.333333 3.75 3
## COMMIT3 READY3 JSAT COMMIT READY
## 895 3.333333 3.00 2.333333 4.333333 2.75
## 896 3.333333 3.00 3.666667 3.666667 2.75
## 897 3.333333 3.00 4.000000 3.333333 3.00
## 898 4.333333 3.25 2.666667 3.666667 2.25
## 899 4.333333 3.25 2.666667 4.333333 3.75
## 900 4.333333 3.25 3.000000 4.333333 3.25
```

```
summary(data1) # display summary
```

```
## SUBNUM TIME JOBSAT1 COMMIT1
## Min. : 1.00 Min. :0 Min. : 1.000 Min. : 1.000
## 1st Qu.: 75.75 1st Qu.:0 1st Qu.: 2.667 1st Qu.: 3.333
## Median :150.50 Median :1 Median : 3.667 Median : 3.667
## Mean :150.50 Mean :1 Mean : 49.763 Mean : 46.794
## 3rd Qu.:225.25 3rd Qu.:2 3rd Qu.: 4.000 3rd Qu.: 4.333
## Max. :300.00 Max. :2 Max. :999.000 Max. :999.000
##
## READY1 JOBSAT2 COMMIT2 READY2
## Min. : 1.00 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.: 2.75 1st Qu.:2.667 1st Qu.:3.000 1st Qu.:2.750
## Median : 3.25 Median :3.333 Median :3.667 Median :3.250
## Mean : 56.18 Mean :3.272 Mean :3.498 Mean :3.176
## 3rd Qu.: 3.75 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:3.750
```



```
## Max. :999.00 Max. :5.000 Max. :5.000 Max. :5.000
## NA's :66 NA's :48 NA's :54
## JOBSAT3 COMMIT3 READY3 JSAT
## Min. :1.000 Min. :1.333 Min. :1.000 Min. :1.000
## 1st Qu.:3.000 1st Qu.:3.000 1st Qu.:2.750 1st Qu.:2.667
## Median :3.333 Median :3.667 Median :3.250 Median :3.333
## Mean :3.355 Mean :3.556 Mean :3.241 Mean :3.308
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000
## Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000
## NA's :51 NA's :48 NA's :48 NA's :53
## COMMIT READY
## Min. :1.000 Min. :1.000
## 1st Qu.:3.000 1st Qu.:2.750
## Median :3.667 Median :3.250
## Mean :3.573 Mean :3.161
## 3rd Qu.:4.000 3rd Qu.:3.750
## Max. :5.000 Max. :5.000
## NA's :45 NA's :50
```

```
summary(data2)
```

```
## SUBNUM TIME JOBSAT1 COMMIT1 READY1
## Min. :301 Min. :0 Min. :1.000 Min. :1.000 Min. :1.00
## 1st Qu.:349 1st Qu.:0 1st Qu.:2.667 1st Qu.:3.000 1st Qu.:2.25
## Median :398 Median :1 Median :3.333 Median :3.667 Median :3.00
## Mean :398 Mean :1 Mean :3.137 Mean :3.543 Mean :2.92
## 3rd Qu.:447 3rd Qu.:2 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:3.50
## Max. :495 Max. :2 Max. :5.000 Max. :5.000 Max. :4.75
## NA's :39 NA's :45 NA's :48
## JOBSAT2 COMMIT2 READY2 JOBSAT3
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:2.667 1st Qu.:3.000 1st Qu.:2.500 1st Qu.:3.000
## Median :3.333 Median :3.667 Median :3.000 Median :3.333
## Mean :3.207 Mean :3.422 Mean :3.007 Mean :3.313
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:3.750 3rd Qu.:4.000
## Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000
## NA's :24 NA's :21 NA's :33 NA's :45
## COMMIT3 READY3 JSAT COMMIT
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:3.000 1st Qu.:2.750 1st Qu.:2.667 1st Qu.:3.000
## Median :3.667 Median :3.250 Median :3.333 Median :3.667
## Mean :3.508 Mean :3.165 Mean :3.219 Mean :3.490
## 3rd Qu.:4.000 3rd Qu.:3.750 3rd Qu.:4.000 3rd Qu.:4.000
## Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000
## NA's :36 NA's :57 NA's :36 NA's :34
## READY
## Min. :1.00
## 1st Qu.:2.50
## Median :3.25
## Mean :3.03
## 3rd Qu.:3.75
## Max. :5.00
## NA's :46
```

## Handling missing values

```
#Note: I used 999 to represent missing data for JOBSAT1 COMMIT1 and READY1
#R needs to be told that 999 is not a legitimate value, but is user-defined missing value
data1$JOBSAT1[data1$JOBSAT1==999]=NA #Explain what the heck this means!
data1$COMMIT1[data1$COMMIT1==999]=NA
data1$READY1[data1$READY1==999]=NA
summary(data1)
```

```
##      SUBNUM      TIME      JOBSAT1      COMMIT1
##  Min.   : 1.00   Min.   :0   Min.   :1.000   Min.   :1.000
## 1st Qu.: 75.75   1st Qu.:0   1st Qu.:2.667   1st Qu.:3.000
## Median :150.50   Median :1   Median :3.333   Median :3.667
## Mean   :150.50   Mean    :1   Mean   :3.297   Mean   :3.663
## 3rd Qu.:225.25   3rd Qu.:2   3rd Qu.:4.000   3rd Qu.:4.000
## Max.   :300.00   Max.    :2   Max.   :5.000   Max.   :5.000
##                                     NA's   :42   NA's   :39
##      READY1      JOBSAT2      COMMIT2      READY2
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
## 1st Qu.:2.500   1st Qu.:2.667   1st Qu.:3.000   1st Qu.:2.750
## Median :3.000   Median :3.333   Median :3.667   Median :3.250
## Mean   :3.066   Mean   :3.272   Mean   :3.498   Mean   :3.176
## 3rd Qu.:3.750   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:3.750
## Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
## NA's   :48     NA's   :66     NA's   :48     NA's   :54
##      JOBSAT3      COMMIT3      READY3      JSAT
##  Min.   :1.000   Min.   :1.333   Min.   :1.000   Min.   :1.000
## 1st Qu.:3.000   1st Qu.:3.000   1st Qu.:2.750   1st Qu.:2.667
## Median :3.333   Median :3.667   Median :3.250   Median :3.333
## Mean   :3.355   Mean   :3.556   Mean   :3.241   Mean   :3.308
## 3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000
## Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
## NA's   :51     NA's   :48     NA's   :48     NA's   :53
##      COMMIT      READY
##  Min.   :1.000   Min.   :1.000
## 1st Qu.:3.000   1st Qu.:2.750
## Median :3.667   Median :3.250
## Mean   :3.573   Mean   :3.161
## 3rd Qu.:4.000   3rd Qu.:3.750
## Max.   :5.000   Max.   :5.000
## NA's   :45     NA's   :50
```

```
summary(data2)
```

```
##      SUBNUM      TIME      JOBSAT1      COMMIT1      READY1
##  Min.   :301   Min.   :0   Min.   :1.000   Min.   :1.000   Min.   :1.00
## 1st Qu.:349   1st Qu.:0   1st Qu.:2.667   1st Qu.:3.000   1st Qu.:2.25
## Median :398   Median :1   Median :3.333   Median :3.667   Median :3.00
## Mean   :398   Mean   :1   Mean   :3.137   Mean   :3.543   Mean   :2.92
## 3rd Qu.:447   3rd Qu.:2   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:3.50
## Max.   :495   Max.   :2   Max.   :5.000   Max.   :5.000   Max.   :4.75
##                                     NA's   :39   NA's   :45   NA's   :48
##      JOBSAT2      COMMIT2      READY2      JOBSAT3
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
```

```
## 1st Qu.:2.667 1st Qu.:3.000 1st Qu.:2.500 1st Qu.:3.000
## Median :3.333 Median :3.667 Median :3.000 Median :3.333
## Mean :3.207 Mean :3.422 Mean :3.007 Mean :3.313
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:3.750 3rd Qu.:4.000
## Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000
## NA's :24 NA's :21 NA's :33 NA's :45
## COMMIT3 READY3 JSAT COMMIT
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:3.000 1st Qu.:2.750 1st Qu.:2.667 1st Qu.:3.000
## Median :3.667 Median :3.250 Median :3.333 Median :3.667
## Mean :3.508 Mean :3.165 Mean :3.219 Mean :3.490
## 3rd Qu.:4.000 3rd Qu.:3.750 3rd Qu.:4.000 3rd Qu.:4.000
## Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000
## NA's :36 NA's :57 NA's :36 NA's :34
## READY
## Min. :1.00
## 1st Qu.:2.50
## Median :3.25
## Mean :3.03
## 3rd Qu.:3.75
## Max. :5.00
## NA's :46
```

*#The above can be tedious if you have a large number of variables*

### it is easier if you copy & paste code

*#Or, if 999 doesn't hold any meaning for ANY of the variables*

```
data1=read.csv(paste0(path2data, "data1.csv"), na.strings=c(".", "999","9","-9"))
summary(data1)
```

```
## SUBNUM TIME JOBSAT1 COMMIT1 READY1
## Min. : 1 Min. :0 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.: 76 1st Qu.:0 1st Qu.:2.667 1st Qu.:3.000 1st Qu.:2.500
## Median :151 Median :1 Median :3.333 Median :3.667 Median :3.000
## Mean :151 Mean :1 Mean :3.297 Mean :3.663 Mean :3.066
## 3rd Qu.:226 3rd Qu.:2 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:3.750
## Max. :300 Max. :2 Max. :5.000 Max. :5.000 Max. :5.000
## NA's :3 NA's :42 NA's :39 NA's :48
## JOBSAT2 COMMIT2 READY2 JOBSAT3
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:2.667 1st Qu.:3.000 1st Qu.:2.750 1st Qu.:3.000
## Median :3.333 Median :3.667 Median :3.250 Median :3.333
## Mean :3.272 Mean :3.498 Mean :3.176 Mean :3.355
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:3.750 3rd Qu.:4.000
## Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000
## NA's :66 NA's :48 NA's :54 NA's :51
## COMMIT3 READY3 JSAT COMMIT
## Min. :1.333 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:3.000 1st Qu.:2.750 1st Qu.:2.667 1st Qu.:3.000
## Median :3.667 Median :3.250 Median :3.333 Median :3.667
## Mean :3.556 Mean :3.241 Mean :3.308 Mean :3.573
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000
## Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000
## NA's :48 NA's :48 NA's :53 NA's :45
```

```
##      READY
## Min.   :1.000
## 1st Qu.:2.750
## Median :3.250
## Mean   :3.161
## 3rd Qu.:3.750
## Max.   :5.000
## NA's   :50
```

*#OR, you could write a function*

```
my999isNA=function(x) {x[x==999]=NA; x}
```

*#Now we will apply this missing data function to the proper variables in data2*

*#To do this, we use the "lapply" function which allows us to apply the same function over a list or array*

```
data1=read.csv(paste0(path2data, "data1.csv")) #reread data1 as a data.frame with missing data
names(data1)
```

```
## [1] "SUBNUM" "TIME" "JOBSAT1" "COMMIT1" "READY1" "JOBSAT2" "COMMIT2"
## [8] "READY2" "JOBSAT3" "COMMIT3" "READY3" "JSAT" "COMMIT" "READY"
```

```
summary(data1)
```

```
##      SUBNUM      TIME      JOBSAT1      COMMIT1
## Min.   : 1.00   Min.   :0      Min.   : 1.000   Min.   : 1.000
## 1st Qu.: 75.75   1st Qu.:0      1st Qu.: 2.667   1st Qu.: 3.333
## Median :150.50   Median :1      Median : 3.667   Median : 3.667
## Mean   :150.50   Mean   :1      Mean   : 49.763   Mean   : 46.794
## 3rd Qu.:225.25   3rd Qu.:2      3rd Qu.: 4.000   3rd Qu.: 4.333
## Max.   :300.00   Max.   :2      Max.   :999.000   Max.   :999.000
##
##      READY1      JOBSAT2      COMMIT2      READY2
## Min.   : 1.00   Min.   :1.000   Min.   :1.000   Min.   :1.000
## 1st Qu.: 2.75   1st Qu.:2.667   1st Qu.:3.000   1st Qu.:2.750
## Median : 3.25   Median :3.333   Median :3.667   Median :3.250
## Mean   : 56.18   Mean   :3.272   Mean   :3.498   Mean   :3.176
## 3rd Qu.: 3.75   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:3.750
## Max.   :999.00   Max.   :5.000   Max.   :5.000   Max.   :5.000
##      NA's      NA's      NA's      NA's
##      :66      :48      :48      :54
##      JOBSAT3      COMMIT3      READY3      JSAT
## Min.   :1.000   Min.   :1.333   Min.   :1.000   Min.   :1.000
## 1st Qu.:3.000   1st Qu.:3.000   1st Qu.:2.750   1st Qu.:2.667
## Median :3.333   Median :3.667   Median :3.250   Median :3.333
## Mean   :3.355   Mean   :3.556   Mean   :3.241   Mean   :3.308
## 3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000
## Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
## NA's   :51     NA's   :48     NA's   :48     NA's   :53
##      COMMIT      READY
## Min.   :1.000   Min.   :1.000
## 1st Qu.:3.000   1st Qu.:2.750
## Median :3.667   Median :3.250
## Mean   :3.573   Mean   :3.161
## 3rd Qu.:4.000   3rd Qu.:3.750
## Max.   :5.000   Max.   :5.000
```

```
## NA's :45      NA's :50
```

```
data1[3:5]=lapply(data1[3:5],my999isNA)
summary(data1)
```

```
##      SUBNUM      TIME      JOBSAT1      COMMIT1
## Min.   : 1.00   Min.   :0   Min.   :1.000   Min.   :1.000
## 1st Qu.: 75.75   1st Qu.:0   1st Qu.:2.667   1st Qu.:3.000
## Median :150.50   Median :1   Median :3.333   Median :3.667
## Mean   :150.50   Mean    :1   Mean   :3.297   Mean   :3.663
## 3rd Qu.:225.25   3rd Qu.:2   3rd Qu.:4.000   3rd Qu.:4.000
## Max.   :300.00   Max.    :2   Max.   :5.000   Max.   :5.000
##                                     NA's   :42   NA's   :39
##      READY1      JOBSAT2      COMMIT2      READY2
## Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
## 1st Qu.:2.500   1st Qu.:2.667   1st Qu.:3.000   1st Qu.:2.750
## Median :3.000   Median :3.333   Median :3.667   Median :3.250
## Mean   :3.066   Mean    :3.272   Mean   :3.498   Mean   :3.176
## 3rd Qu.:3.750   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:3.750
## Max.   :5.000   Max.    :5.000   Max.   :5.000   Max.   :5.000
## NA's   :48     NA's   :66     NA's   :48     NA's   :54
##      JOBSAT3      COMMIT3      READY3      JSAT
## Min.   :1.000   Min.   :1.333   Min.   :1.000   Min.   :1.000
## 1st Qu.:3.000   1st Qu.:3.000   1st Qu.:2.750   1st Qu.:2.667
## Median :3.333   Median :3.667   Median :3.250   Median :3.333
## Mean   :3.355   Mean    :3.556   Mean   :3.241   Mean   :3.308
## 3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000
## Max.   :5.000   Max.    :5.000   Max.   :5.000   Max.   :5.000
## NA's   :51     NA's   :48     NA's   :48     NA's   :53
##      COMMIT      READY
## Min.   :1.000   Min.   :1.000
## 1st Qu.:3.000   1st Qu.:2.750
## Median :3.667   Median :3.250
## Mean   :3.573   Mean    :3.161
## 3rd Qu.:4.000   3rd Qu.:3.750
## Max.   :5.000   Max.    :5.000
## NA's   :45     NA's   :50
```

## TOPIC 5: Merging Data Files

```
#Merging data by adding variables (e.g, two data.frames, demo1 + data1)
dd1=merge(demo1,data1, by="SUBNUM")
dd1=merge(demo1,data1, by=c("SUBNUM","TIME"), all=TRUE)

dd2=merge(demo2,data2, by=c("SUBNUM","TIME"), all=TRUE)
summary(dd1)
```

```
##      SUBNUM      TIME      BTN      COMPANY
## Min.   : 1.00   Min.   :0   Min.   : 4.0   A       :246
## 1st Qu.: 75.75   1st Qu.:0   1st Qu.: 377.8   HHC     :210
## Median :150.50   Median :1   Median :1022.0   B       :207
## Mean   :150.50   Mean    :1   Mean   :1860.3   D       :114
## 3rd Qu.:225.25   3rd Qu.:2   3rd Qu.:3066.0   C       : 84
## Max.   :300.00   Max.    :2   Max.   :4042.0   SVC     : 24
```

```

##                                     (Other): 15
##      MARITAL          GENDER      HOWLONG      RANK
## Min.   :1.000   Min.   :1.000   Min.   :0.000   Min.   :11.00
## 1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:13.00
## Median :2.000   Median :1.000   Median :2.000   Median :14.00
## Mean   :1.711   Mean   :1.039   Mean   :2.371   Mean   :15.26
## 3rd Qu.:2.000   3rd Qu.:1.000   3rd Qu.:4.000   3rd Qu.:16.00
## Max.   :5.000   Max.   :2.000   Max.   :5.000   Max.   :32.00
## NA's   :6       NA's   :51     NA's   :18     NA's   :48
##      EDUCATE          AGE      JOBSAT1      COMMIT1
## Min.   :1.000   Min.   :18.00   Min.   :1.000   Min.   :1.000
## 1st Qu.:2.000   1st Qu.:20.00   1st Qu.:2.667   1st Qu.:3.000
## Median :2.000   Median :24.00   Median :3.333   Median :3.667
## Mean   :2.663   Mean   :25.75   Mean   :3.297   Mean   :3.663
## 3rd Qu.:3.000   3rd Qu.:30.00   3rd Qu.:4.000   3rd Qu.:4.000
## Max.   :6.000   Max.   :44.00   Max.   :5.000   Max.   :5.000
## NA's   :9       NA's   :9       NA's   :42     NA's   :39
##      READY1          JOBSAT2      COMMIT2      READY2
## Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
## 1st Qu.:2.500   1st Qu.:2.667   1st Qu.:3.000   1st Qu.:2.750
## Median :3.000   Median :3.333   Median :3.667   Median :3.250
## Mean   :3.066   Mean   :3.272   Mean   :3.498   Mean   :3.176
## 3rd Qu.:3.750   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:3.750
## Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
## NA's   :48     NA's   :66     NA's   :48     NA's   :54
##      JOBSAT3          COMMIT3      READY3      JSAT
## Min.   :1.000   Min.   :1.333   Min.   :1.000   Min.   :1.000
## 1st Qu.:3.000   1st Qu.:3.000   1st Qu.:2.750   1st Qu.:2.667
## Median :3.333   Median :3.667   Median :3.250   Median :3.333
## Mean   :3.355   Mean   :3.556   Mean   :3.241   Mean   :3.308
## 3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000
## Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
## NA's   :51     NA's   :48     NA's   :48     NA's   :53
##      COMMIT          READY
## Min.   :1.000   Min.   :1.000
## 1st Qu.:3.000   1st Qu.:2.750
## Median :3.667   Median :3.250
## Mean   :3.573   Mean   :3.161
## 3rd Qu.:4.000   3rd Qu.:3.750
## Max.   :5.000   Max.   :5.000
## NA's   :45     NA's   :50

```

summary(dd2)

```

##      SUBNUM      TIME      BTN      COMPANY      MARITAL
## Min.   :301   Min.   :0   Min.   : 4   A       :156   Min.   :1.000
## 1st Qu.:349   1st Qu.:0   1st Qu.: 404   HHC     :144   1st Qu.:1.000
## Median :398   Median :1   Median :1022   B       :141   Median :2.000
## Mean   :398   Mean   :1   Mean   :1755   D       : 69   Mean   :1.756
## 3rd Qu.:447   3rd Qu.:2   3rd Qu.:3066   C       : 42   3rd Qu.:2.000
## Max.   :495   Max.   :2   Max.   :4042   SVC     : 15   Max.   :5.000
##                                     (Other): 18   NA's   :6
##      GENDER      HOWLONG      RANK      EDUCATE
## Min.   :1.000   Min.   :0.000   Min.   :11.0   Min.   :1.00
## 1st Qu.:1.000   1st Qu.:2.000   1st Qu.:13.0   1st Qu.:2.00

```

```
## Median :1.000 Median :2.000 Median :14.0 Median :2.00
## Mean :1.022 Mean :2.446 Mean :14.7 Mean :2.49
## 3rd Qu.:1.000 3rd Qu.:3.000 3rd Qu.:15.0 3rd Qu.:2.00
## Max. :2.000 Max. :5.000 Max. :31.0 Max. :6.00
## NA's :51 NA's :6 NA's :27 NA's :3
## AGE JOBSAT1 COMMIT1 READY1
## Min. :18.00 Min. :1.000 Min. :1.000 Min. :1.00
## 1st Qu.:21.00 1st Qu.:2.667 1st Qu.:3.000 1st Qu.:2.25
## Median :24.00 Median :3.333 Median :3.667 Median :3.00
## Mean :25.68 Mean :3.137 Mean :3.543 Mean :2.92
## 3rd Qu.:29.00 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:3.50
## Max. :46.00 Max. :5.000 Max. :5.000 Max. :4.75
## NA's :3 NA's :39 NA's :45 NA's :48
## JOBSAT2 COMMIT2 READY2 JOBSAT3
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:2.667 1st Qu.:3.000 1st Qu.:2.500 1st Qu.:3.000
## Median :3.333 Median :3.667 Median :3.000 Median :3.333
## Mean :3.207 Mean :3.422 Mean :3.007 Mean :3.313
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:3.750 3rd Qu.:4.000
## Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000
## NA's :24 NA's :21 NA's :33 NA's :45
## COMMIT3 READY3 JSAT COMMIT
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:3.000 1st Qu.:2.750 1st Qu.:2.667 1st Qu.:3.000
## Median :3.667 Median :3.250 Median :3.333 Median :3.667
## Mean :3.508 Mean :3.165 Mean :3.219 Mean :3.490
## 3rd Qu.:4.000 3rd Qu.:3.750 3rd Qu.:4.000 3rd Qu.:4.000
## Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000
## NA's :36 NA's :57 NA's :36 NA's :34
## READY
## Min. :1.00
## 1st Qu.:2.50
## Median :3.25
## Mean :3.03
## 3rd Qu.:3.75
## Max. :5.00
## NA's :46
```

Merging data by adding rows (subjects)

```
#let's combine dd1 with dd2
#when you have IDENTICAL columns in both data sets you may use rbind
names(dd1); names(dd2)
```

```
## [1] "SUBNUM" "TIME" "BTN" "COMPANY" "MARITAL" "GENDER" "HOWLONG"
## [8] "RANK" "EDUCATE" "AGE" "JOBSAT1" "COMMIT1" "READY1" "JOBSAT2"
## [15] "COMMIT2" "READY2" "JOBSAT3" "COMMIT3" "READY3" "JSAT" "COMMIT"
## [22] "READY"

## [1] "SUBNUM" "TIME" "BTN" "COMPANY" "MARITAL" "GENDER" "HOWLONG"
## [8] "RANK" "EDUCATE" "AGE" "JOBSAT1" "COMMIT1" "READY1" "JOBSAT2"
## [15] "COMMIT2" "READY2" "JOBSAT3" "COMMIT3" "READY3" "JSAT" "COMMIT"
## [22] "READY"
```

```
dd3=rbind(dd1,dd2)
summary(dd3)
```

```
##      SUBNUM      TIME      BTN      COMPANY      MARITAL
## Min.   : 1   Min.   :0   Min.   : 4   A       :402   Min.   :1.000
## 1st Qu.:124   1st Qu.:0   1st Qu.: 404   HHC      :354   1st Qu.:1.000
## Median :248   Median :1   Median :1022   B       :348   Median :2.000
## Mean   :248   Mean   :1   Mean   :1819   D       :183   Mean   :1.729
## 3rd Qu.:372   3rd Qu.:2   3rd Qu.:3066   C       :126   3rd Qu.:2.000
## Max.   :495   Max.   :2   Max.   :4042   SVC      : 39   Max.   :5.000
##                                     (Other): 33   NA's    :12
##      GENDER      HOWLONG      RANK      EDUCATE
## Min.   :1.000   Min.   :0.0   Min.   :11.00   Min.   :1.000
## 1st Qu.:1.000   1st Qu.:1.0   1st Qu.:13.00   1st Qu.:2.000
## Median :1.000   Median :2.0   Median :14.00   Median :2.000
## Mean   :1.033   Mean   :2.4   Mean   :15.04   Mean   :2.595
## 3rd Qu.:1.000   3rd Qu.:4.0   3rd Qu.:16.00   3rd Qu.:3.000
## Max.   :2.000   Max.   :5.0   Max.   :32.00   Max.   :6.000
## NA's    :102   NA's    :24   NA's    :75   NA's    :12
##      AGE      JOBSAT1      COMMIT1      READY1
## Min.   :18.00   Min.   :1.000   Min.   :1.000   Min.   :1.00
## 1st Qu.:21.00   1st Qu.:2.667   1st Qu.:3.000   1st Qu.:2.50
## Median :24.00   Median :3.333   Median :3.667   Median :3.00
## Mean   :25.72   Mean   :3.235   Mean   :3.617   Mean   :3.01
## 3rd Qu.:30.00   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:3.75
## Max.   :46.00   Max.   :5.000   Max.   :5.000   Max.   :5.00
## NA's    :12   NA's    :81   NA's    :84   NA's    :96
##      JOBSAT2      COMMIT2      READY2      JOBSAT3
## Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
## 1st Qu.:2.667   1st Qu.:3.000   1st Qu.:2.500   1st Qu.:3.000
## Median :3.333   Median :3.667   Median :3.250   Median :3.333
## Mean   :3.246   Mean   :3.468   Mean   :3.109   Mean   :3.338
## 3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:3.750   3rd Qu.:4.000
## Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
## NA's    :90   NA's    :69   NA's    :87   NA's    :96
##      COMMIT3      READY3      JSAT      COMMIT
## Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
## 1st Qu.:3.000   1st Qu.:2.750   1st Qu.:2.667   1st Qu.:3.000
## Median :3.667   Median :3.250   Median :3.333   Median :3.667
## Mean   :3.537   Mean   :3.212   Mean   :3.273   Mean   :3.540
## 3rd Qu.:4.000   3rd Qu.:3.750   3rd Qu.:4.000   3rd Qu.:4.000
## Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
## NA's    :84   NA's    :105   NA's    :89   NA's    :79
##      READY
## Min.   :1.00
## 1st Qu.:2.50
## Median :3.25
## Mean   :3.11
## 3rd Qu.:3.75
## Max.   :5.00
## NA's    :96
```

*#when you have different columns in your data, you can use rbind.fill*  
*#first let's compute some extra variables and add them to dd1*



```

#Computing new variables in an existing data.frame
dd1$STAY=dd1$JSAT+dd1$COMMIT
#dd3=rbind(dd1,dd2) doesn't work because of differing colums
?rbind.fill
install.packages("plyr")
library(plyr)

dd3=plyr::rbind.fill(dd1,dd2)
head(dd3); tail(dd3)

```

```

##      SUBNUM TIME  BTN COMPANY MARITAL GENDER HOWLONG RANK EDUCATE AGE
## 1      1      0 1022     HHC      1      1      2  12      2  20
## 2      1      1 1022     HHC      1      1      2  12      2  20
## 3      1      2 1022     HHC      1      1      2  12      2  20
## 4     10      0 3066      C      1      1      3  13      2  22
## 5     10      1 3066      C      1      1      3  13      2  22
## 6     10      2 3066      C      1      1      3  13      2  22
##      JOBSAT1  COMMIT1 READY1  JOBSAT2  COMMIT2 READY2  JOBSAT3  COMMIT3 READY3
## 1 1.666667 1.666667  2.75 1.000000 1.666667      1      3      3      3
## 2 1.666667 1.666667  2.75 1.000000 1.666667      1      3      3      3
## 3 1.666667 1.666667  2.75 1.000000 1.666667      1      3      3      3
## 4 1.000000 3.666667  2.00 1.333333 3.333333      3      3      3      3
## 5 1.000000 3.666667  2.00 1.333333 3.333333      3      3      3      3
## 6 1.000000 3.666667  2.00 1.333333 3.333333      3      3      3      3
##      JSAT  COMMIT  READY
## 1 1.666667 1.666667  2.75
## 2 1.000000 1.666667  1.00
## 3 3.000000 3.000000  3.00
## 4 1.000000 3.666667  2.00
## 5 1.333333 3.333333  3.00
## 6 3.000000 3.000000  3.00

##      SUBNUM TIME  BTN COMPANY MARITAL GENDER HOWLONG RANK EDUCATE AGE
## 1480    494      0 4042      B      1      1      0  11      2  22
## 1481    494      1 4042      B      1      1      0  11      2  22
## 1482    494      2 4042      B      1      1      0  11      2  22
## 1483    495      0 1022      B      1      1      3  13      2  19
## 1484    495      1 1022      B      1      1      3  13      2  19
## 1485    495      2 1022      B      1      1      3  13      2  19
##      JOBSAT1  COMMIT1 READY1  JOBSAT2  COMMIT2 READY2  JOBSAT3  COMMIT3
## 1480      3 4.333333  3.75 2.333333 3.333333  3.25 3.666667 4.000000
## 1481      3 4.333333  3.75 2.333333 3.333333  3.25 3.666667 4.000000
## 1482      3 4.333333  3.75 2.333333 3.333333  3.25 3.666667 4.000000
## 1483      4 4.000000  3.75 4.000000 4.000000  4.00 3.000000 3.333333
## 1484      4 4.000000  3.75 4.000000 4.000000  4.00 3.000000 3.333333
## 1485      4 4.000000  3.75 4.000000 4.000000  4.00 3.000000 3.333333
##      READY3      JSAT  COMMIT  READY
## 1480  4.00 3.000000 4.333333  3.75
## 1481  4.00 2.333333 3.333333  3.25
## 1482  4.00 3.666667 4.000000  4.00
## 1483  3.25 4.000000 4.000000  3.75
## 1484  3.25 4.000000 4.000000  4.00
## 1485  3.25 3.000000 3.333333  3.25

```

## Deleting a variable from a data frame

```
#let's delete STAY from the previous dd3 data.frame
names(dd3)

## [1] "SUBNUM" "TIME" "BTN" "COMPANY" "MARITAL" "GENDER" "HOWLONG"
## [8] "RANK" "EDUCATE" "AGE" "JOBSAT1" "COMMIT1" "READY1" "JOBSAT2"
## [15] "COMMIT2" "READY2" "JOBSAT3" "COMMIT3" "READY3" "JSAT" "COMMIT"
## [22] "READY"

dd4=dd3[c(1,2,3:22)]
names(dd4)

## [1] "SUBNUM" "TIME" "BTN" "COMPANY" "MARITAL" "GENDER" "HOWLONG"
## [8] "RANK" "EDUCATE" "AGE" "JOBSAT1" "COMMIT1" "READY1" "JOBSAT2"
## [15] "COMMIT2" "READY2" "JOBSAT3" "COMMIT3" "READY3" "JSAT" "COMMIT"
## [22] "READY"

#Renaming a variable in a data.frame
#let's rename HOWLONG to TENURE and MARITAL to STATUS
dd4=plyr::rename(dd4, c(HOWLONG="TENURE", MARITAL="STATUS"))
names(dd4)

## [1] "SUBNUM" "TIME" "BTN" "COMPANY" "STATUS" "GENDER" "TENURE"
## [8] "RANK" "EDUCATE" "AGE" "JOBSAT1" "COMMIT1" "READY1" "JOBSAT2"
## [15] "COMMIT2" "READY2" "JOBSAT3" "COMMIT3" "READY3" "JSAT" "COMMIT"
## [22] "READY"
```

---

## Recoding variables

```
#Categorical Variables: recode sex into a different, dummy variable
#Only "factor" type variables are assigned value labels
dd4$GENDER2=plyr::revalue(as.factor(dd4$GENDER), c("1"="male","2"="female"))
dd4$GENDER3=(dd4$GENDER-1)
class(dd4$GENDER)

## [1] "numeric"

class(dd4$GENDER2)

## [1] "factor"

class(dd4$GENDER3)

## [1] "numeric"

#recode Likert-type items/scales
###let's reverse the overall score on COMMIT so that high scores = more likely to leave
dd4$LEAVE=6-dd4$COMMIT
```

## TOPIC 6: Summarizing & Visualizing Data Frames

### Central Tendency

```
mean(dd3$JSAT); median(dd3$JSAT)

## [1] NA
## [1] NA
mean(dd3$JSAT,na.rm=TRUE); median(dd3$JSAT,na.rm=TRUE)

## [1] 3.272923
## [1] 3.333333
#Dispersion
var(dd3$JSAT,na.rm=T)

## [1] 0.8622181
sd(dd3$JSAT,na.rm=T)

## [1] 0.928557
min(dd3$JSAT, na.rm=T)

## [1] 1
max(dd3$JSAT,na.rm=T)

## [1] 5
summary(dd3$JSAT,na.rm=T)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##      1.000   2.667   3.333   3.273   4.000   5.000     89
quantile(dd3$JSAT,probs=c(.1,.2,.3,.4,.5,.6,.7,.8,.9),na.rm=T)

##      10%      20%      30%      40%      50%      60%      70%      80%
## 2.000000 2.333333 3.000000 3.000000 3.333333 3.666667 4.000000 4.000000
##      90%
## 4.333333
```

### Alternative: Hmisc

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

Hmisc::describe(dd4)

## dd4
##
## 25 Variables      1485 Observations
## -----
## SUBNUM
##      n missing distinct    Info    Mean    Gmd    .05    .10
## 1485      0      495      1    248   165.1   25.2   50.0
##    .25    .50    .75    .90    .95
```

```

##      124.0      248.0      372.0      446.0      470.8
##
## lowest :      1      2      3      4      5, highest: 491 492 493 494 495
## -----
## TIME
##      n missing distinct      Info      Mean      Gmd
##    1485      0      3      0.889      1      0.8895
##
## Value      0      1      2
## Frequency    495    495    495
## Proportion 0.333 0.333 0.333
## -----
## BTN
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1485      0      16      0.965    1819    1566      4     104
##      .25      .50      .75      .90      .95
##      404    1022    3066    4042    4042
##
## Value      0    100    120    140    300    400    700    1000    1010    1020
## Frequency    141     15     42     30    123     48      6     66     21    288
## Proportion 0.095 0.010 0.028 0.020 0.083 0.032 0.004 0.044 0.014 0.194
##
## Value      2000    2010    3070    4000    4040
## Frequency     36     51    435     18    165
## Proportion 0.024 0.034 0.293 0.012 0.111
## -----
## COMPANY
##      n missing distinct
##    1485      0      8
##
## Value      A      B      C      D      F      HHC      REC      SVC
## Frequency    402    348    126    183     15    354     18     39
## Proportion 0.271 0.234 0.085 0.123 0.010 0.238 0.012 0.026
## -----
## STATUS
##      n missing distinct      Info      Mean      Gmd
##    1473     12      5      0.79    1.729    0.745
##
## Value      1      2      3      4      5
## Frequency    603    768     21     60     21
## Proportion 0.409 0.521 0.014 0.041 0.014
## -----
## GENDER
##      n missing distinct      Info      Mean      Gmd
##    1383    102      2      0.094    1.033    0.063
##
## Value      1      2
## Frequency   1338     45
## Proportion 0.967 0.033
## -----
## TENURE
##      n missing distinct      Info      Mean      Gmd
##    1461     24      6      0.949     2.4    1.747
##

```

```

## Value      0      1      2      3      4      5
## Frequency  216   159   495   225   147   219
## Proportion 0.148 0.109 0.339 0.154 0.101 0.150
## -----
## RANK
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1410      75      15    0.972    15.04    2.979      12      12
##      .25      .50      .75      .90      .95
##      13      14      16      21      22
##
## Value      11      12      13      14      15      16      17      18      19      21
## Frequency   21    147    324    264    279    114     84     18      3     54
## Proportion 0.015 0.104 0.230 0.187 0.198 0.081 0.060 0.013 0.002 0.038
##
## Value      22      23      24      31      32
## Frequency   51     42      3      3      3
## Proportion 0.036 0.030 0.002 0.002 0.002
## -----
## EDUCATE
##      n missing distinct      Info      Mean      Gmd
##    1473      12      6    0.617    2.595    0.9586
##
## Value      1      2      3      4      5      6
## Frequency    9   1068    99   117   168    12
## Proportion 0.006 0.725 0.067 0.079 0.114 0.008
## -----
## AGE
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1473      12      29    0.994    25.72    6.715      19      19
##      .25      .50      .75      .90      .95
##      21      24      30      35      37
##
## lowest : 18 19 20 21 22, highest: 42 43 44 45 46
## -----
## JOBSAT1
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1404      81      13    0.983    3.235    1.104    1.333    1.667
##      .25      .50      .75      .90      .95
##    2.667    3.333    4.000    4.333    4.667
##
## Value      1.000000 1.333333 1.666667 2.000000 2.333333 2.666667 3.000000
## Frequency      48      39      63      96      78     102     180
## Proportion    0.034    0.028    0.045    0.068    0.056    0.073    0.128
##
## Value      3.333333 3.666667 4.000000 4.333333 4.666667 5.000000
## Frequency     156     141     315      87      54      45
## Proportion    0.111    0.100    0.224    0.062    0.038    0.032
## -----
## COMMIT1
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1401      84      13    0.982    3.617    0.9408    2.000    2.333
##      .25      .50      .75      .90      .95
##    3.000    3.667    4.000    4.667    5.000
##

```

```

## Value      1.000000 1.333333 1.666667 2.000000 2.333333 2.666667 3.000000
## Frequency      12      9      21      45      57      75      165
## Proportion    0.009    0.006    0.015    0.032    0.041    0.054    0.118
##
## Value      3.333333 3.666667 4.000000 4.333333 4.666667 5.000000
## Frequency     189     222     282     102     102     120
## Proportion    0.135    0.158    0.201    0.073    0.073    0.086
## -----
## READY1
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1389      96      17      0.99      3.01    0.9286      1.50      1.75
##      .25      .50      .75      .90      .95
##      2.50      3.00      3.75      4.00      4.00
##
## Value      1.00  1.25  1.50  1.75  2.00  2.25  2.50  2.75  3.00  3.25
## Frequency      36   33   33   45   66   78   108   141   177   204
## Proportion 0.026 0.024 0.024 0.032 0.048 0.056 0.078 0.102 0.127 0.147
##
## Value      3.50  3.75  4.00  4.25  4.50  4.75  5.00
## Frequency     105   117   183    36    18     6     3
## Proportion 0.076 0.084 0.132 0.026 0.013 0.004 0.002
## -----
## JOBSAT2
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1395      90      13      0.978      3.246      1.041      1.333      2.000
##      .25      .50      .75      .90      .95
##      2.667      3.333      4.000      4.000      4.667
##
## Value      1.000000 1.333333 1.666667 2.000000 2.333333 2.666667 3.000000
## Frequency      51      30      54      75      99      84      174
## Proportion    0.037    0.022    0.039    0.054    0.071    0.060    0.125
##
## Value      3.333333 3.666667 4.000000 4.333333 4.666667 5.000000
## Frequency     177     168     348      63      33      39
## Proportion    0.127    0.120    0.249    0.045    0.024    0.028
## -----
## COMMIT2
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1416      69      13      0.981      3.468      0.9529      1.667      2.333
##      .25      .50      .75      .90      .95
##      3.000      3.667      4.000      4.667      5.000
##
## Value      1.000000 1.333333 1.666667 2.000000 2.333333 2.666667 3.000000
## Frequency      39      18      30      18      57      93      207
## Proportion    0.028    0.013    0.021    0.013    0.040    0.066    0.146
##
## Value      3.333333 3.666667 4.000000 4.333333 4.666667 5.000000
## Frequency     213     207     291      96      75      72
## Proportion    0.150    0.146    0.206    0.068    0.053    0.051
## -----
## READY2
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1398      87      17      0.989      3.109      0.9311      1.50      2.00
##      .25      .50      .75      .90      .95

```

```

##      2.50      3.25      3.75      4.00      4.25
##
## Value      1.00  1.25  1.50  1.75  2.00  2.25  2.50  2.75  3.00  3.25
## Frequency    24   30   30   39   69   75   105   75   216   162
## Proportion 0.017 0.021 0.021 0.028 0.049 0.054 0.075 0.054 0.155 0.116
##
## Value      3.50  3.75  4.00  4.25  4.50  4.75  5.00
## Frequency   162  162  156   39   18   15   21
## Proportion 0.116 0.116 0.112 0.028 0.013 0.011 0.015
## -----
## JOBSAT3
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1389      96      13    0.972    3.338    0.941    1.667    2.000
##      .25      .50      .75      .90      .95
##    3.000    3.333    4.000    4.333    4.667
##
## Value      1.000000 1.333333 1.666667 2.000000 2.333333 2.666667 3.000000
## Frequency      24      33      21      72      72      69      279
## Proportion    0.017    0.024    0.015    0.052    0.052    0.050    0.201
##
## Value      3.333333 3.666667 4.000000 4.333333 4.666667 5.000000
## Frequency     183     138     351      60      42      45
## Proportion    0.132    0.099    0.253    0.043    0.030    0.032
## -----
## COMMIT3
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1401      84      13    0.974    3.537    0.8182    2.000    2.667
##      .25      .50      .75      .90      .95
##    3.000    3.667    4.000    4.333    4.667
##
## Value      1.000000 1.333333 1.666667 2.000000 2.333333 2.666667 3.000000
## Frequency      9      9      21      33      42      54      261
## Proportion    0.006    0.006    0.015    0.024    0.030    0.039    0.186
##
## Value      3.333333 3.666667 4.000000 4.333333 4.666667 5.000000
## Frequency     204     234     315     102      48      69
## Proportion    0.146    0.167    0.225    0.073    0.034    0.049
## -----
## READY3
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1380     105      17    0.986    3.212    0.8964    1.50    2.00
##      .25      .50      .75      .90      .95
##    2.75      3.25      3.75      4.00      4.25
##
## Value      1.00  1.25  1.50  1.75  2.00  2.25  2.50  2.75  3.00  3.25
## Frequency    12   24   42   36   36   39   87   102   237   144
## Proportion 0.009 0.017 0.030 0.026 0.026 0.028 0.063 0.074 0.172 0.104
##
## Value      3.50  3.75  4.00  4.25  4.50  4.75  5.00
## Frequency   168  114  231   48   21   21   18
## Proportion 0.122 0.083 0.167 0.035 0.015 0.015 0.013
## -----
## JSAT
##      n missing distinct      Info      Mean      Gmd      .05      .10

```

```

##      1396      89      13      0.978      3.273      1.032      1.333      2.000
##      .25      .50      .75      .90      .95
##      2.667      3.333      4.000      4.333      4.667
##
## Value      1.000000 1.333333 1.666667 2.000000 2.333333 2.666667 3.000000
## Frequency      41      34      46      81      83      85      211
## Proportion      0.029      0.024      0.033      0.058      0.059      0.061      0.151
##
## Value      3.333333 3.666667 4.000000 4.333333 4.666667 5.000000
## Frequency      172      149      338      70      43      43
## Proportion      0.123      0.107      0.242      0.050      0.031      0.031
## -----
## COMMIT
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      1406      79      13      0.979      3.54      0.9079      2.000      2.667
##      .25      .50      .75      .90      .95
##      3.000      3.667      4.000      4.667      5.000
##
## Value      1.000000 1.333333 1.666667 2.000000 2.333333 2.666667 3.000000
## Frequency      20      12      24      32      52      74      211
## Proportion      0.014      0.009      0.017      0.023      0.037      0.053      0.150
##
## Value      3.333333 3.666667 4.000000 4.333333 4.666667 5.000000
## Frequency      202      221      296      100      75      87
## Proportion      0.144      0.157      0.211      0.071      0.053      0.062
## -----
## READY
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      1389      96      17      0.989      3.11      0.924      1.50      2.00
##      .25      .50      .75      .90      .95
##      2.50      3.25      3.75      4.00      4.25
##
## Value      1.00 1.25 1.50 1.75 2.00 2.25 2.50 2.75 3.00 3.25
## Frequency      24      29      35      40      57      64      100      106      210      170
## Proportion 0.017 0.021 0.025 0.029 0.041 0.046 0.072 0.076 0.151 0.122
##
## Value      3.50 3.75 4.00 4.25 4.50 4.75 5.00
## Frequency      145      131      190      41      19      14      14
## Proportion 0.104 0.094 0.137 0.030 0.014 0.010 0.010
## -----
## GENDER2
##      n missing distinct
##      1383      102      2
##
## Value      male female
## Frequency      1338      45
## Proportion 0.967 0.033
## -----
## GENDER3
##      n missing distinct      Info      Sum      Mean      Gmd
##      1383      102      2      0.094      45      0.03254      0.063
## -----
## LEAVE

```



```
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1406      79      13    0.979      2.46    0.9079      1.000      1.333
##      .25      .50      .75      .90      .95
##    2.000      2.333      3.000      3.333      4.000
##
## Value      1.000000 1.333333 1.666667 2.000000 2.333333 2.666667 3.000000
## Frequency      87      75      100      296      221      202      211
## Proportion      0.062      0.053      0.071      0.211      0.157      0.144      0.150
##
## Value      3.333333 3.666667 4.000000 4.333333 4.666667 5.000000
## Frequency      74      52      32      24      12      20
## Proportion      0.053      0.037      0.023      0.017      0.009      0.014
## -----
```

Alternative: psych

```
detach("package:Hmisc")
install.packages("psych")
library(psych)
```

```
psych::describe(dd4,na.rm=T)
```

```
##      vars      n      mean      sd median trimmed      mad min  max range
## SUBNUM      1 1485    248.00   142.94    248.00    248.00   183.84    1  495   494
## TIME        2 1485      1.00     0.82      1.00      1.00     1.48    0    2     2
## BTN         3 1485  1818.73  1403.35  1022.00  1767.29  1509.29    4 4042  4038
## COMPANY*    4 1485      3.26     2.10      2.00      3.12     1.48    1    8     7
## STATUS      5 1473      1.73     0.80      2.00      1.61     0.00    1    5     4
## GENDER      6 1383      1.03     0.18      1.00      1.00     0.00    1    2     1
## TENURE      7 1461      2.40     1.56      2.00      2.38     1.48    0    5     5
## RANK        8 1410     15.04     3.02     14.00     14.51     1.48   11   32    21
## EDUCATE     9 1473      2.59     1.09      2.00      2.37     0.00    1    6     5
## AGE        10 1473     25.72     6.12     24.00     25.01     5.93   18   46    28
## JOBSAT1     11 1404      3.24     0.98      3.33      3.29     0.99    1    5     4
## COMMIT1     12 1401      3.62     0.85      3.67      3.66     0.99    1    5     4
## READY1      13 1389      3.01     0.83      3.00      3.07     0.74    1    5     4
## JOBSAT2     14 1395      3.25     0.94      3.33      3.31     0.99    1    5     4
## COMMIT2     15 1416      3.47     0.87      3.67      3.52     0.49    1    5     4
## READY2      16 1398      3.11     0.83      3.25      3.15     0.74    1    5     4
## JOBSAT3     17 1389      3.34     0.85      3.33      3.39     0.99    1    5     4
## COMMIT3     18 1401      3.54     0.75      3.67      3.57     0.49    1    5     4
## READY3      19 1380      3.21     0.81      3.25      3.27     0.74    1    5     4
## JSAT        20 1396      3.27     0.93      3.33      3.33     0.99    1    5     4
## COMMIT      21 1406      3.54     0.83      3.67      3.58     0.49    1    5     4
## READY       22 1389      3.11     0.83      3.25      3.16     0.74    1    5     4
## GENDER2*    23 1383      1.03     0.18      1.00      1.00     0.00    1    2     1
## GENDER3     24 1383      0.03     0.18      0.00      0.00     0.00    0    1     1
## LEAVE       25 1406      2.46     0.83      2.33      2.42     0.49    1    5     4
##
##      skew kurtosis      se
## SUBNUM    0.00     -1.20   3.71
## TIME       0.00     -1.50   0.02
## BTN        0.20     -1.48  36.42
## COMPANY*   0.52     -1.11   0.05
```

```
## STATUS      1.67      3.97  0.02
## GENDER      5.26     25.73  0.00
## TENURE      0.19     -0.85  0.04
## RANK        1.84      4.18  0.08
## EDUCATE     1.50      0.75  0.03
## AGE         0.92      0.16  0.16
## JOBSAT1    -0.47     -0.50  0.03
## COMMIT1    -0.48      0.14  0.02
## READY1     -0.47     -0.26  0.02
## JOBSAT2    -0.60     -0.23  0.03
## COMMIT2    -0.69      0.64  0.02
## READY2     -0.41     -0.09  0.02
## JOBSAT3    -0.55      0.16  0.02
## COMMIT3    -0.48      0.69  0.02
## READY3     -0.50      0.09  0.02
## JSAT       -0.55     -0.20  0.02
## COMMIT     -0.57      0.56  0.02
## READY      -0.46     -0.09  0.02
## GENDER2*    5.26     25.73  0.00
## GENDER3     5.26     25.73  0.00
## LEAVE       0.57      0.56  0.02
```

```
psych::describe(dd4,na.rm=F)
```

```
##      vars      n      mean      sd  median trimmed      mad min  max range
## SUBNUM      1 1032  249.52  143.06  248.50  250.14  182.36  1  495  494
## TIME        2 1032    1.00    0.82    1.00    1.00    1.48  0    2    2
## BTN         3 1032 1805.08 1400.20 1022.00 1750.15 1509.29  4 4042 4038
## COMPANY*    4 1032    3.29    2.08    2.50    3.16    2.22  1    8    7
## STATUS      5 1032    1.75    0.81    2.00    1.63    0.00  1    5    4
## GENDER      6 1032    1.03    0.16    1.00    1.00    0.00  1    2    1
## TENURE      7 1032    2.42    1.58    2.00    2.41    1.48  0    5    5
## RANK        8 1032   15.14    3.18   14.00   14.61    1.48 11   32   21
## EDUCATE     9 1032    2.65    1.13    2.00    2.44    0.00  1    6    5
## AGE        10 1032   25.68    6.03   24.00   24.99    5.93 18   45   27
## JOBSAT1     11 1032    3.20    0.99    3.33    3.26    0.99  1    5    4
## COMMIT1     12 1032    3.63    0.84    3.67    3.66    0.99  1    5    4
## READY1      13 1032    3.02    0.79    3.00    3.08    0.74  1    5    4
## JOBSAT2     14 1032    3.23    0.93    3.33    3.29    0.99  1    5    4
## COMMIT2     15 1032    3.47    0.86    3.67    3.53    0.49  1    5    4
## READY2      16 1032    3.13    0.83    3.25    3.18    0.74  1    5    4
## JOBSAT3     17 1032    3.28    0.87    3.33    3.34    0.99  1    5    4
## COMMIT3     18 1032    3.52    0.76    3.67    3.54    0.49  1    5    4
## READY3      19 1032    3.21    0.81    3.25    3.26    0.74  1    5    4
## JSAT        20 1032    3.24    0.93    3.33    3.30    0.99  1    5    4
## COMMIT      21 1032    3.54    0.82    3.67    3.58    0.49  1    5    4
## READY       22 1032    3.12    0.81    3.25    3.17    0.74  1    5    4
## GENDER2*    23 1032    1.03    0.16    1.00    1.00    0.00  1    2    1
## GENDER3     24 1032    0.03    0.16    0.00    0.00    0.00  0    1    1
## LEAVE       25 1032    2.46    0.82    2.33    2.42    0.49  1    5    4
##
##      skew kurtosis      se
## SUBNUM -0.04    -1.18  4.45
## TIME    0.00    -1.50  0.03
## BTN     0.24    -1.47 43.59
## COMPANY* 0.49    -1.17  0.06
```

```
## STATUS      1.71      4.12  0.03
## GENDER      5.93     33.18  0.00
## TENURE      0.17     -0.90  0.05
## RANK        1.83      4.08  0.10
## EDUCATE     1.33      0.21  0.04
## AGE         0.91      0.17  0.19
## JOBSAT1     -0.46     -0.61  0.03
## COMMIT1     -0.42      0.05  0.03
## READY1      -0.47     -0.08  0.02
## JOBSAT2     -0.55     -0.39  0.03
## COMMIT2     -0.68      0.68  0.03
## READY2      -0.41     -0.18  0.03
## JOBSAT3     -0.54      0.06  0.03
## COMMIT3     -0.46      0.70  0.02
## READY3      -0.47      0.19  0.03
## JSAT        -0.52     -0.33  0.03
## COMMIT      -0.52      0.52  0.03
## READY       -0.44     -0.03  0.03
## GENDER2*    5.93     33.18  0.00
## GENDER3     5.93     33.18  0.00
## LEAVE       0.52      0.52  0.03
```

```
psych::describe(na.omit(dd4))
```

```
##          vars      n    mean      sd median trimmed      mad min  max range
## SUBNUM      1 1032  249.52  143.06  248.50  250.14  182.36  1  495   494
## TIME        2 1032    1.00    0.82    1.00    1.00    1.48  0    2     2
## BTN         3 1032 1805.08 1400.20 1022.00 1750.15 1509.29  4 4042  4038
## COMPANY*    4 1032    3.29    2.08    2.50    3.16    2.22  1    8     7
## STATUS      5 1032    1.75    0.81    2.00    1.63    0.00  1    5     4
## GENDER      6 1032    1.03    0.16    1.00    1.00    0.00  1    2     1
## TENURE      7 1032    2.42    1.58    2.00    2.41    1.48  0    5     5
## RANK        8 1032   15.14    3.18   14.00   14.61    1.48 11   32    21
## EDUCATE     9 1032    2.65    1.13    2.00    2.44    0.00  1    6     5
## AGE        10 1032   25.68    6.03   24.00   24.99    5.93 18   45    27
## JOBSAT1     11 1032    3.20    0.99    3.33    3.26    0.99  1    5     4
## COMMIT1     12 1032    3.63    0.84    3.67    3.66    0.99  1    5     4
## READY1      13 1032    3.02    0.79    3.00    3.08    0.74  1    5     4
## JOBSAT2     14 1032    3.23    0.93    3.33    3.29    0.99  1    5     4
## COMMIT2     15 1032    3.47    0.86    3.67    3.53    0.49  1    5     4
## READY2      16 1032    3.13    0.83    3.25    3.18    0.74  1    5     4
## JOBSAT3     17 1032    3.28    0.87    3.33    3.34    0.99  1    5     4
## COMMIT3     18 1032    3.52    0.76    3.67    3.54    0.49  1    5     4
## READY3      19 1032    3.21    0.81    3.25    3.26    0.74  1    5     4
## JSAT        20 1032    3.24    0.93    3.33    3.30    0.99  1    5     4
## COMMIT      21 1032    3.54    0.82    3.67    3.58    0.49  1    5     4
## READY       22 1032    3.12    0.81    3.25    3.17    0.74  1    5     4
## GENDER2*    23 1032    1.03    0.16    1.00    1.00    0.00  1    2     1
## GENDER3     24 1032    0.03    0.16    0.00    0.00    0.00  0    1     1
## LEAVE       25 1032    2.46    0.82    2.33    2.42    0.49  1    5     4
##
##          skew kurtosis      se
## SUBNUM   -0.04    -1.18  4.45
## TIME      0.00    -1.50  0.03
## BTN       0.24    -1.47 43.59
## COMPANY*  0.49    -1.17  0.06
```

```
## STATUS      1.71      4.12  0.03
## GENDER      5.93     33.18  0.00
## TENURE       0.17     -0.90  0.05
## RANK         1.83      4.08  0.10
## EDUCATE      1.33      0.21  0.04
## AGE          0.91      0.17  0.19
## JOBSAT1     -0.46     -0.61  0.03
## COMMIT1     -0.42      0.05  0.03
## READY1      -0.47     -0.08  0.02
## JOBSAT2     -0.55     -0.39  0.03
## COMMIT2     -0.68      0.68  0.03
## READY2      -0.41     -0.18  0.03
## JOBSAT3     -0.54      0.06  0.03
## COMMIT3     -0.46      0.70  0.02
## READY3      -0.47      0.19  0.03
## JSAT        -0.52     -0.33  0.03
## COMMIT      -0.52      0.52  0.03
## READY       -0.44     -0.03  0.03
## GENDER2*    5.93     33.18  0.00
## GENDER3     5.93     33.18  0.00
## LEAVE        0.52      0.52  0.03
```

## Simple Distributions

```
#Frequency Counts
table(dd4$COMPANY)
```

```
##
##   A   B   C   D   F HHC REC SVC
## 402 348 126 183  15 354  18  39
```

```
#Proportions
prop.table(table(dd4$COMPANY))
```

```
##
##           A           B           C           D           F           HHC
## 0.27070707 0.23434343 0.08484848 0.12323232 0.01010101 0.23838384
##           REC           SVC
## 0.01212121 0.02626263
```

```
#Rounding proportions to 3 decimals
round(prop.table(table(dd4$COMPANY)),3)
```

```
##
##   A   B   C   D   F   HHC   REC   SVC
## 0.271 0.234 0.085 0.123 0.010 0.238 0.012 0.026
```

```
#Percentages
100*(prop.table(table(dd4$COMPANY)))
```

```
##
##           A           B           C           D           F           HHC           REC
## 27.070707 23.434343  8.484848 12.323232  1.010101 23.838384  1.212121
##           SVC
##  2.626263
```

---

```
#Cross Tabs & Simple Tables
```

```
#install.packages("gmodels")
```

```
library(gmodels)
```

```
CrossTable(dd4$GENDER,dd4$COMPANY,chisq=TRUE,format="SPSS")
```

```
## Warning in chisq.test(t, correct = FALSE, ...): Chi-squared approximation
## may be incorrect
```

```
##
```

```
## Cell Contents
```

```
## |-----|
## |                Count |
## | Chi-square contribution |
## |                Row Percent |
## |                Column Percent |
## |                Total Percent |
## |-----|
```

```
##
```

```
## Total Observations in Table: 1383
```

```
##
```

```
##          | dd4$COMPANY
```

dd4\$GENDER	A	B	C	D	F	HHC	REC
1	357	321	111	165	9	321	18
	0.023	0.181	0.042	0.037	0.010	0.148	0.020
	26.682%	23.991%	8.296%	12.332%	0.673%	23.991%	1.345%
	95.968%	99.074%	94.872%	98.214%	100.000%	94.690%	100.000%
	25.813%	23.210%	8.026%	11.931%	0.651%	23.210%	1.302%
2	15	3	6	3	0	18	0
	0.693	5.396	1.263	1.113	0.293	4.404	0.586
	33.333%	6.667%	13.333%	6.667%	0.000%	40.000%	0.000%
	4.032%	0.926%	5.128%	1.786%	0.000%	5.310%	0.000%
	1.085%	0.217%	0.434%	0.217%	0.000%	1.302%	0.000%
Column Total	372	324	117	168	9	339	18
	26.898%	23.427%	8.460%	12.148%	0.651%	24.512%	1.302%

```
##
```

```
##
```

```
## Statistics for All Table Factors
```

```
##
```

```
##
```

```
## Pearson's Chi-squared test
```

```
## -----
## Chi^2 = 15.42045    d.f. = 7    p = 0.03097201
```

```
##
```

```
##
```

```
##
```

```
## Minimum expected frequency: 0.2928416
```

```
## Cells with Expected Frequency < 5: 4 of 16 (25%)
```

```
table(dd4$GENDER,dd4$COMPANY)
```

```
##
##      A    B    C    D    F HHC REC SVC
##  1 357 321 111 165   9 321 18 36
##  2  15   3   6   3   0  18  0  0
```

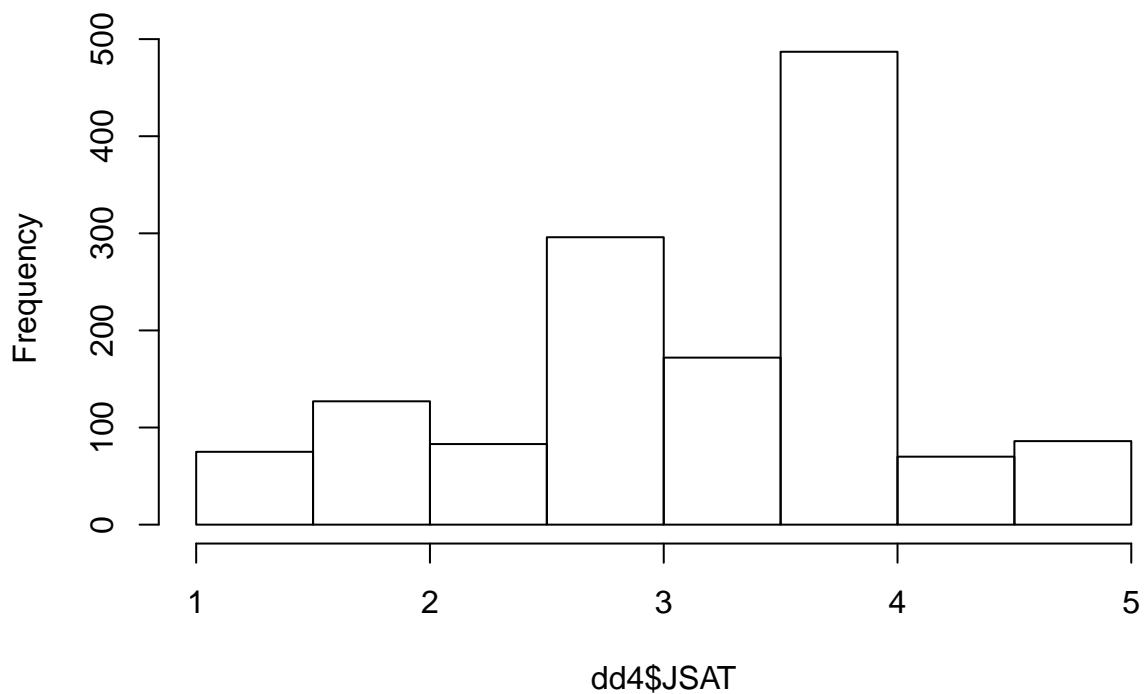
```
prop.table(table(dd4$GENDER,dd4$COMPANY))
```

```
##
##              A              B              C              D              F
##  1 0.258134490 0.232104121 0.080260304 0.119305857 0.006507592
##  2 0.010845987 0.002169197 0.004338395 0.002169197 0.000000000
##
##              HHC              REC              SVC
##  1 0.232104121 0.013015184 0.026030369
##  2 0.013015184 0.000000000 0.000000000
```

```
#Histograms
```

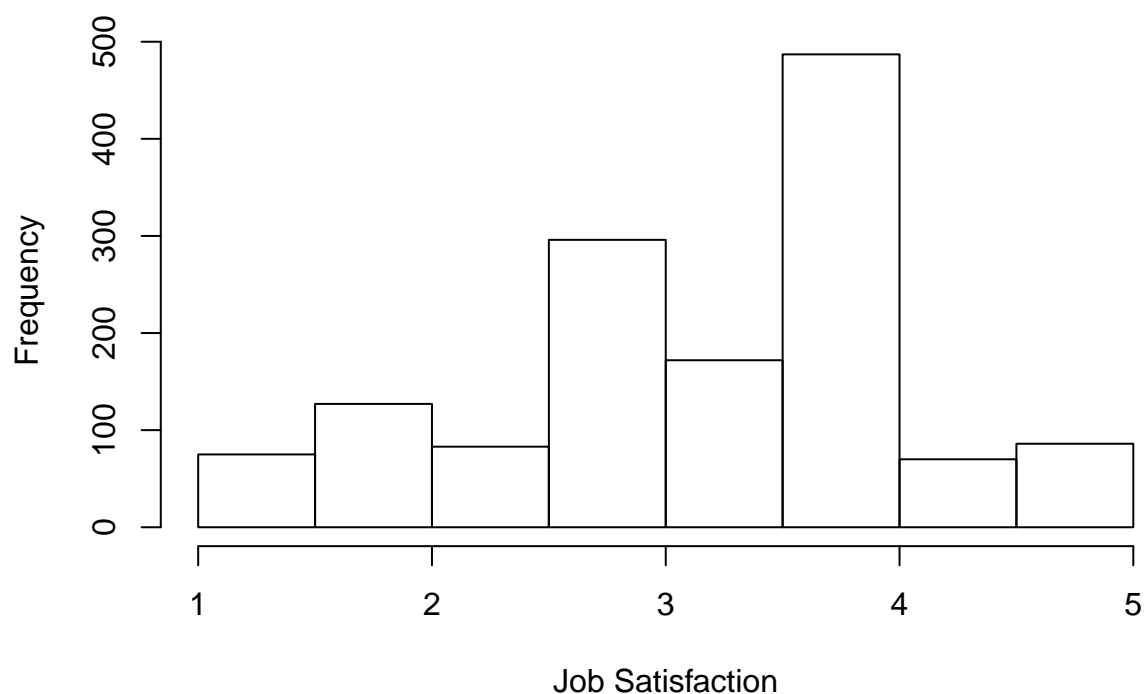
```
hist(dd4$JSAT)
```

**Histogram of dd4\$JSAT**



```
hist(dd4$JSAT, main="Job Satisfaction Histogram",xlab="Job Satisfaction" )
```

## Job Satisfaction Histogram



Correlations using `cor` (part of stats) or `rcorr` (part of Hmisc)

```
cor(dd4[,20:22],use="complete.obs")
```

```
##           JSAT   COMMIT   READY
## JSAT      1.0000000 0.5373179 0.5093204
## COMMIT    0.5373179 1.0000000 0.4610560
## READY     0.5093204 0.4610560 1.0000000
```

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

```
Hmisc::rcorr(as.matrix(dd4[,c(20:22)]))
```

```
##           JSAT COMMIT READY
## JSAT      1.00   0.54  0.51
## COMMIT    0.54   1.00  0.46
## READY     0.51   0.46  1.00
```

```
##
```

```
## n
```

```
##           JSAT COMMIT READY
## JSAT      1396   1385  1369
## COMMIT    1385   1406  1375
## READY     1369   1375  1389
```

```
##
```

```
## P
```

```
##           JSAT COMMIT READY
## JSAT           0      0
## COMMIT    0      0
```

```
## READY    0    0
```

## Popular Packages

multilevel

lme4 & nlme

plyr

ggplot2

reshape2

Rcmdr

Hmisc