

0. Download R.

<http://www.r-project.org/>

1. Set the working directory: setwd

```
> setwd('E:/linguistics/r_tutorial')
> dir()
```

- Alternatively: under “File” menu -> “Change dir...” and browse the directory.

2. Create an object

a. Combine a small set of numbers or characters: assignment operators are <- or =

```
> data1<-c(1,2,3,4,1,2,3,4)
> data1
[1] 1 2 3 4 1 2 3 4
> summary(data1)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.00   1.75   2.50   2.50   3.25   4.00
```

- When you make up your object name, keep in mind that
 - R is case-sensitive; ‘data1’ and ‘Data1’ are different objects.
 - Do not begin your object name with a number.

b. Change the data mode into ‘character’ or ‘factor’: **as.character, as.factor**

```
> data1<-c("1","2","3","4","1","2","3","4")
> data1
[1] "1" "2" "3" "4" "1" "2" "3" "4"
```

- R will overwrite the old data without asking when they share the same object name.

```
> data1<-rep(1:4,2)
> data1
[1] 1 2 3 4 1 2 3 4
> data2<-as.character(data1)
> data2
[1] "1" "2" "3" "4" "1" "2" "3" "4"
> summary(data2)
   Length      Class      Mode
   8 character character
> data3<-as.factor(data1)
> data3
[1] 1 2 3 4 1 2 3 4
Levels: 1 2 3 4
> summary(data3)
 1 2 3 4
 2 2 2 2
```

c. Bind columns and rows: **cbind, rbind, data.frame**

```
> data1.1<- rep(1:3,4)
> data1.2<- rep(1:4,3)
> data1.1
[1] 1 2 3 1 2 3 1 2 3 1 2 3
```

```

> length(data1.2)
[1] 12
> data1.2
[1] 1 2 3 4 1 2 3 4 1 2 3 4
> data1.3<-rbind(data1.1,data1.2)
> data1.3
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
data1.1  1    2    3    1    2    3    1    2    3    1    2    3
data1.2  1    2    3    4    1    2    3    4    1    2    3    4
> cbind(data1.1,data1.2)
      data1.1 data1.2
[1,]      1      1
[2,]      2      2
[3,]      3      3
[4,]      1      4
[5,]      2      1
[6,]      3      2
[7,]      1      3
[8,]      2      4
[9,]      3      1
[10,]     1      2
[11,]     2      3
[12,]     3      4
> data1.4<- data.frame(rbind(data1.1,data1.2))
> dim(data1.4)
[1]  2 12
> data1.4
  X1 X2 X3 X4 X5 X6 X7 X8 X9 X10 X11 X12
data1.1 1 2 3 1 2 3 1 2 3 1 2 3
data1.2 1 2 3 4 1 2 3 4 1 2 3 4

```

d. Assign column names, and row names: **names, colnames, rownames**

```

> names(data1.4)
[1] "X1" "X2" "X3" "X4" "X5" "X6" "X7" "X8" "X9" "X10" "X11"
"X12"
> colnames(data1.4)
[1] "X1" "X2" "X3" "X4" "X5" "X6" "X7" "X8" "X9" "X10" "X11"
"X12"
> colnames(data1.4)<-letters[1:12]
> colnames(data1.4)
[1] "a" "b" "c" "d" "e" "f" "g" "h" "i" "j" "k" "l"
> rownames(data1.4)
[1] "data1.1" "data1.2"
> rownames(data1.4)<-c("A", "B")
> rownames(data1.4)
[1] "A" "B"

```

e. Write the object as a table: **write.table**

```

> data1.4

```

```

a b c d e f g h i j k l
A 1 2 3 1 2 3 1 2 3 1 2 3
B 1 2 3 4 1 2 3 4 1 2 3 4
> write.table(data1.4, "tutorial_data1.4.txt", sep=" ")
> write.table(data1.4, "tutorial_data1.4.txt", sep="\t")

```

3. Read in a data file: read.table, read.delim

```

> data4<-read.table("vowdata.txt", sep="", header=T)
> names(data4)
 [1] "sex" "sub" "vow" "dur" "f0" "F1" "F2" "F3" "F4" "F1.2"
[11] "F2.2" "F3.2" "F1.5" "F2.5" "F3.5" "F1.8" "F2.8" "F3.8"
> data4[1:3,1:7]
      sex sub vow dur  f0  F1  F2
m01ae  m m01  ae 323 174 663 2012
m02ae  m m02  ae 250 102 628 1871
m03ae  m m03  ae 344  99 605 1812
> summary(data4)

```

- <http://homepages.wmich.edu/~hillenbr/voweldata.html>
- Get further details on reading in data files.

```

> help("read.table")
> help.search("read.table")

```

- How to record the previous commands
 - a. Write them as a script
 - File-> New script
 - File-> Open script
 - b. Save them as a history
 - File -> Save History
 - File -> Load History

4. Picturize the data: hist, plot (barplot, piechart...)

a. One-dimensional figure: histogram (hist)

```

> data4$f0
> hist(data4$f0, xlim=c(50, 350), xlab="Fundamental Frequency (Hz)", main="F0 data: Hillenbrand et al")
> savePlot("tutorial_hist", type="emf")

```

- Make a subset: **subset**

```

> m<-subset(data4, sex=="m")
> w<- subset(data4, sex=="w")
> hist(m$f0, xlim=c(50, 350), ylim=c(0, 300), breaks=seq(50,350, 20), xlab="Fundamental Frequency (Hz)", main="F0 data: Hillenbrand et al", col="sky blue")
> hist(w$f0, xlim=c(50, 350), breaks=seq(50,350, 20), add=T, density=5)
> legend("topleft", c("male", "female"), fill=c("sky blue", "black"), density=c(0, 5))

```

b. Two-dimensional figure: scatterplot (plot)

```

> plot(data4$F2, data4$F1)
> plot(data4$F2, data4$F1, xlim=c(3500,500), ylim=c(1500,200))

```

```
> plot(data4$F2, data4$F1, xlim=c(3500,500), ylim=c(1500,200), xlab="F2 (Hz)", ylab="F1 (Hz)")
```

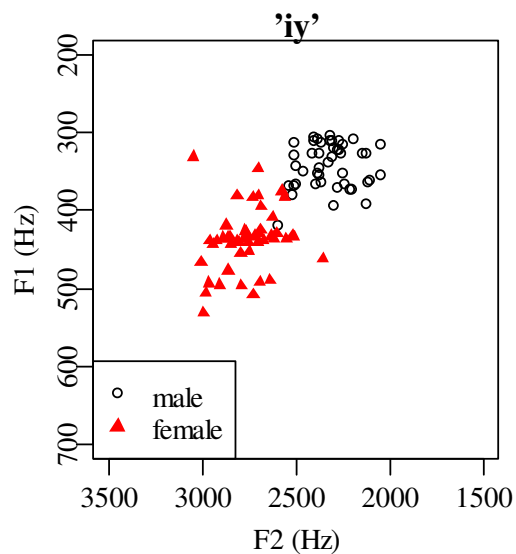
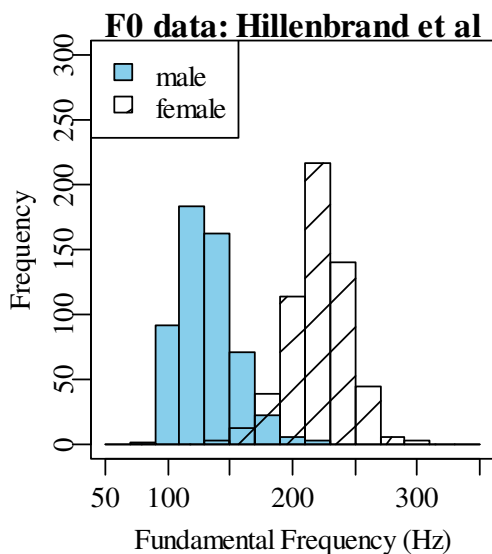
c. Sort out gender

```
> plot(m$F2[m$vow=="iy"], m$F1[m$vow=="iy"], xlim=c(3500,1500), ylim=c(700,200),
xlab="F2 (Hz)", ylab="F1 (Hz)", main="'iy'")
> points(w$F2[w$vow=="iy"], w$F1[w$vow=="iy"], col="red", pch=17, cex=0.8)
> legend("bottomleft", c("male", "female"), pch=c(21,17), col=c("black","red"))
```

- Different plot characters in R: plot(1:25, 1:25, pch=c(1:25))
- **pch** also can be any character or symbol such as "a" or "?".

d. Define a graphic window: **windows**

```
> windows(width=6,height=3,pointsize=12)
> par(family="serif", mfrow=c(1,2), oma=rep(0,4), mar=c(3,3,1,1), mgp=c(1.8,0.5,0))
# histogram
> hist(m$f0, xlim=c(50, 350), ylim=c(0, 300), breaks=seq(50,350, 20), xlab="Fundamental
Frequency (Hz)", main="F0 data: Hillenbrand et al", col="sky blue")
> hist(w$f0, xlim=c(50, 350), breaks=seq(50,350, 20), add=T, density=5)
> legend("topleft", c("male", "female"), fill=c("sky blue", "black"), density=c(0, 5));box()
# scatterplot
> plot(m$F2[m$vow=="iy"], m$F1[m$vow=="iy"], xlim=c(3500,1500), ylim=c(700,200),
xlab="F2 (Hz)", ylab="F1 (Hz)", main="'iy'")
> points(w$F2[w$vow=="iy"], w$F1[w$vow=="iy"], col="red", pch=17, cex=0.8)
> legend("bottomleft", c("male", "female"), pch=c(21,17), col=c("black","red"))
```



5. Statistics in R

a. t.test, linear regression (lm), chisq.test, cor.test, aov.