



NTNU – Trondheim
Norwegian University of
Science and Technology

Introduction to R

1. What is R
2. Command line
3. Calculations
4. Data types
5. Loops
6. Save and load data
7. Plotting
8. Analysis
9. NetCDF

- What is R
 - A “Language and environment for statistical computing and graphics”.
 - Free open source
 - Well documented
 - Command line based
 - URL: <http://www.r-project.org>

- **Command line.**

-On norstore-trd:

Load the module: \$ module load r

Start R \$ R

You will see a command prompt:

>

Quit: >q() or ctrl-d

Help:

Help for a function e.g. “mean”: ?mean

Searching:

help.search(“ploting”) or ??ploting



Open and save your workspace.

Save:

```
>save.image( file="myworkspace.RData",safe=TRUE)
```

Load:

```
>load("myworkspace.RData")
```

Calculations

Arithmetic:

+ , - , * , / , %% (Modulus) , %*% matr.mult

Statistical:

-rnorm(n,mean,sd) Normal distribution (random)

-mean(x), var(x) variance, sum(x), cumsum(x), etc

Math:

-sqrt(x),exp(),log(),sin(),cos() etc

String:

-substr(),paste(),strsplit(), etc



Data types

>12

[1]>12

>"String"

[1]>"String"

>TRUE

[1]>TRUE

>!TRUE

[1]>FALSE

>substr("My name is John" , 12 , 15)

[1]>"John"



Variables

>x<-5 or x=5

>x+10

>[1] 15

Show variables

>x or print(x)

List all variables

>ls()

Remove variables

>rm(x,y,z)



Vectors

```
>1:4
```

```
[1] 1 2 3 4
```

Concatenate

```
>c(1:4 , 5)
```

```
[1] 1 2 3 4 5
```

```
>c("One","two","three")
```

```
[1] "One" "two" "three"
```

Index

```
>num=c(5,4,3,2,1)
```

```
>num[2]
```

```
[1] 4
```



Matrices

```
>A=matrix( 1:9 , ncol=3 , nrow=3)
```

```
>A
```

```
      [,1] [,2] [,3]  
[1,]  1   4   7  
[2,]  2   5   8  
[3,]  3   6   9
```

Index: A[3,1] is 3

Matrix multiplication

```
B=matrix(9:1,ncol=3,nrow=3)
```

```
C=A % * % B
```

Inverse matrix: MASS::ginv

```
>Bi=MASS::ginv(B)
```

```
>C %*% Bi (Result is A matrix)
```

Loops.

For-loop:

```
x = array (1:10 , dim = c(10,1) )  
y = array (0 , dim = c(10,1) )  
for(i in 1:10){  
  y[i] =x[ i ] ^ 2  
}
```

Repeat loop (repeat until break)

>repeat expr

While loop

>while(condition) expr



- **Save and load data**

Load file: `>x=read.table("data.txt")`

Save file: `>write.table(x,"data.txt")`

PLOTTING

1. Array.

```
Y=array(1:100 , dim= c(100,2) )  
for (i in 1:100){  
    Y[i,2]=Y[i,1]^2  
}  
plot(Y)
```

2.Vector

```
x=1:100  
y=x^2  
plot(y~x)
```

Plot arguments

Line: `type="l"`

Colour: `col="red"`

Text:

Title: `main="Polynom"`

X-axis: `xlab="Time (sec)"`

Y-axis: `ylab="Temperature (C)"`

Ex

```
plot ( Y , type="l",col="red",main="Temperature graph",  
      +xlab="Time(sec)",ylab="Temperature(C)")
```

Other Plot:

`hist()`



Function

```
myfunction = function (x1, x2,...,xn){  
ret=x1+x2+x3+...+xn  
}
```

Ex

```
mysqr = function (x) {  
Ret = x * x  
}
```

```
>y=mysqr(2)
```

```
>y
```

```
[1] 4
```



Load your functions

1. Write your function(s) in a text editor as notepad or vim/emacs.
2. Save the text file under you project folder as eg. myfunctions.R
3. Start R and load your functions using:
>source("myfunctions.R") or
>source(file.choose()) and select your file
>y=myfunc1(x)

Analysis

Analysis models:

ACE, ACF, Adaboot, adaptive rejection sampling, AIC, ANCOVA,

- Anderson-Darling K -sample test, Ansari-Bradley test, APT, ARFIMA,
- ARIMA, ARMA, Aster models, Asymptotic regression, asymptotic
- statistics, AVAS, AWS, BACCO, bagging, Bartlett's test, Bayesian statistics,
- binomial test, boosting, Bootstrap, Box test, canonical correlations, chi-square
- contingency tables, factor analysis, Fisher's exact test, Fifty-fty MANOVA,
- Friedman's rank sum test, fuzzy clustering, Gibbs sampling, GLM, hierarchic
- clustering, imputation, Kalman filtering, K -means clustering, k-NN,
- Kolmogorov-Smirnov test, Kruskal-Wallis rank sum test, LDA, LOESS,
- logistic regression, logit models, LOWESS, MAD, Mahalanobis distance,
- MANCOVA, MANOVA, mixed models, multidimensional scaling, multiple
- comparison tests, non-linear optimization, non-linear regression, path
- modelling, PCA, PCR, Phillips-Perron test, PLSR, QDA, random effect
- models, running median smoothing, Shapiro-Wilk normality test, splines,
- time series, t test, varimax rotation, Wilcoxon signed rank test, . . .



Example. Test of a hypothesis.

For example, to test the hypothesis that a random sample of 100 people has been drawn from a population in which men and women are equal in frequency, the observed number of men and women would be compared to the theoretical frequencies of 50 men and 50 women. If there were 44 men and 56 women:

```
>x = c ( 44 , 56 )
```

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

Chi-squared test for given probabilities

data: x

X-squared = 1.44, df = 1, p-value = 0.2301

This probability is higher than conventional criteria for statistical significance (0.05), so normally we would not reject the null hypothesis that the number of men in the population is the same as the number of women



NetCDF and R

Open a netCDF file:

```
>library(ncdf)
```

```
>ex.nc = open.ncdf("test.nc")
```

```
>print(ex.nc)
```

```
>summary(ex.nc)
```

Get a variable from netCDF

```
x=get.var.ncdf(ex.nc,"Nmomentum1")
```

