####CONDITIONAL STATEMENT & LOOPING

##IF

x=1

if (x==2)

{

print ("x=2")

}

##IF-ELSE

x=1

if (x==2)

{

print ("x = 2")

}else

{

print ("x != 2")

}

##FOR

for (i in 1:4)

{

print (i)

}

#WHILE

i=0

while (i<4)

{

i=i+1

print (i)

}

##REPEAT

i=0

repeat

{

i=i+1

print (i)

if (i==4) break

}

##APPLY (matrix/array)

matrix (1:10 , nrow =2) -> a;a

apply (a,1,mean)

apply (a,2, mean)

#SAPPLY&LAPPY (vector/list)

matrix (2:11 , nrow =2) -> a;a

matrix (1:10 , nrow =2) -> b;b

c <- list (a,b);c

sapply(c,mean) #vec/mat

lapply(c,mean) #list

##TAPPLY (factor)

data(iris)

head(iris)

attach(iris)

tapply(Sepal.Width,Species,mean)

####DATA MANIPULATION

a <- c (10 ,20 ,15 ,43 ,76 ,41 ,25 ,46) # numeric

# Factor ’sex ’

b <- factor (c("m","f","m","f","m","f","m","f"))

# siblings , numeric

c <- c(2 ,5 ,8 ,3 ,6 ,1 ,5 ,6)

myframe <- data.frame (a,b,c)

colnames (myframe) <- c("Age","Sex","Siblings")

print (myframe)

str(myframe)

myframe[ ,1] #call column 1

myframe["Age"] #call column Age

myframe$Age #call column Age

myframe [3,3]=2 #change value

myframe [,-2] #access all column except column 2

attach(myframe) #add obj to search path

Sex

detach("myframe") #del obj from search path

subset(myframe,myframe$Age>30) #4 entries

mean(subset(myframe$Age,myframe$Sex=="m"))

mean(subset(myframe$Age,myframe$Sex=="f"))

myframe[(myframe$Sex=="m")&(myframe$Age>30),] #males over 30

myframe[(myframe$Sex=="m")|(myframe$Age>30),] #male or over 30

myframe=cbind(myframe,"Income"=c(1700,2100,2300,2050,2800,1450,3400,2000)) #add column

head(myframe)

myframe$Income=NULL #del col

head(myframe)

x=c(2,3,5,2,5,6,7,3)

sort(x)

order(x)

rank(x)

myframe[order(myframe$Sex,myframe$Age),]

##na.rm=TRUE to not use NA (missing value) in calculating

##is.na(data) to detect missing value

####VISUALIZATION

data=read.csv("F:/Data W2.csv",header=TRUE, sep=",")

head(data)

str(data)

View(data)

library(ggplot2)

# Bar Chart cat.age

ggplot(data,aes(x=category.age,fill=category.age))+geom\_bar()+theme\_bw()+

scale\_fill\_brewer(palette="Set2")

#Histo G1

ggplot(data,aes(x=G1))+

geom\_histogram(aes(y=..density..),binwidth=1,colour="black",fill="white")+

geom\_density(alpha=.1,fill="#00FFFF")

#Pie Chart Sex

library(tidyverse)

sex.c=count(data,Sex)

sex.c=sex.c %>% mutate(perc=round(n\*100/sum(n),1))

sex.c=sex.c %>%

arrange(desc(Sex)) %>%

mutate(y\_pos = cumsum(perc)-0.5\*perc)

sex.c %>%

ggplot(aes(x="",perc, fill=Sex)) +

geom\_bar(width=1,stat="identity",color="white",alpha=.5) +

coord\_polar("y", start=0)+

scale\_fill\_manual(values = rainbow(11)) +

theme\_void()

#Boxplot

ggplot(data,aes(freetime,G1,fill=freetime))+geom\_boxplot(width=0.6)+

stat\_summary(fun=mean,geom="point",size=2,color="darkred")+theme\_bw()+

theme(legend.position="none")+labs(title="Boxplot", subtitle="G1 by Free Time",

x="Free Time",y="G1")+

scale\_fill\_manual(values=c("#8ECAE6","#219EBC","#FD9E02","#FB8500","#FFB703"))+

theme(legend.position="none")

#Scatter G1~G2

ggplot(data, aes(x=G1, y=G2)) +

geom\_point()+

geom\_smooth(method=lm)+theme\_classic()

#Corr

library("ggpubr")

ggscatter(data, x = "G1", y = "G2",

add = "reg.line", conf.int = TRUE,

cor.coef = TRUE, cor.method = "pearson",

xlab = "G1", ylab = "G2")

shapiro.test(data$G1)

cor.test(data$G1, data$G2, method="pearson")

#####DESCRIPTIVE STATISTICS

x=c(1:10);x

y=c(11:20);y

#####MEAN

jml=0

rata=function(x)

{

n=length(x)

for (i in 1:n)

{

jml=jml+x[i]

}

rata=jml/n

list(mean=rata)

}

rata(x)

mean(x)

rata(data$G1)

mean(data$G1)

#####VARIANCE

jml=0

jk=0

varians=function(x)

{

n=length(x)

for (i in 1:n)

{

jml=jml+x[i]

}

rata=jml/n

ss=matrix(n,1)

for (i in 1:n)

{

ss[i]=(x[i]-rata)^2

jk=jk+ss[i]

}

varians=jk/(n-1)

list (variance=varians)

}

varians(x)

var(x)

varians(data$G1)

var(data$G1)

#####COVARIANCE

sxy=0

kov=function(x,y)

{

n=length(x)

xbar=mean(x)

ybar=mean(y)

xy=matrix(n,1)

for(i in 1:n)

{

xy[i]=(x[i]-xbar)\*(y[i]-ybar)

sxy=sxy+xy[i]

}

kov=sxy/(n-1)

list(covariance=kov)

}

kov(x,y)

cov(x,y)

kov(data$G1, data$G2)

cov(data$G1, data$G2)

#####SORT

z=c(2,4,7,3,9,10,1)

urut=function(x)

{

n=length(x)

for (i in 2:n)

{

a=x[i]

j=i-1

while(x[j]>a && j>0)

{

x[j+1]=x[j]

j=j-1

x[j+1]=a

}

}

list(urut=x, min=x[1], max=x[n])

}

urut(z)

sort(z)

#####CORRELATION

kor=function(x,y)

{

n=length(x)

xx=matrix(n,1)

yy=matrix(n,1)

xy=matrix(n,1)

sxx=0

syy=0

sxy=0

sx=0

sy=0

for(i in 1:n)

{

xx[i]=x[i]^2

sxx=sxx+xx[i]

yy[i]=y[i]^2

syy=syy+yy[i]

xy[i]=x[i]\*y[i]

sxy=sxy+xy[i]

sx=sx+x[i]

sy=sy+y[i]

}

ssxy=(n\*sxy)-(sx\*sy)

ssxx=(n\*sxx)-(sx)^2

ssyy=(n\*syy)-(sy)^2

r=ssxy/sqrt(ssxx\*ssyy)

list(correlation=r)

}

kor(x,y)

cor(x,y)