More on Factors and Data Frames

Semester Evaluation Components

- Assignments - 10%
- Quiz - 15%
- Project - 20%
- Mid term - 25%
- End term - 30%

Using Factors in Data Frames

- tab1 <- read.table(file='house_copy.txt',header = T)
- str(tab1)
- table(tab1$region)
- table(tab1$availability)
- table(tab1$sale.price)
- tab1 <- read.table(file='house_copy.txt',header = T, sep = "", colClasses = c('double','double','factor','factor'))
- str(tab1)
- table(tab1$region)
- tab1 <- read.table(file='house_copy.txt',header = T, stringsAsFactors = TRUE)
- str(tab1)
- tab1$region
- tab1$region <- factor(tab1$region, levels=c("E",'W','N','S'))
- tab1$region
- str(tab1)
-

The conversion into factor has to be reassigned to be permanent

- tab1 <- read.table(file='house_copy.txt',header = T)
- str(tab1)
- as.factor(tab1$region)
- str(tab1)
- tab1$region <- as.factor(tab1$region)
- str(tab1)

Re-ordering Factor levels

- x <- factor(c("yes","yes", "no", "yes", "no"))
- x
- x <- factor(x, levels=c('yes','no')); x

- row.names(df1) <- paste0('row',1:nrow(df1))
- colnames(df1) <- paste0('column',1:ncol(df1))

How to operate on specific columns based on some condition of columns of a data frame

```
tab1 <- read.table(file='house_copy.txt',header = T)

startsWith(colnames(tab1),’a')

tab1[,startsWith(colnames(tab1),’a')]

colnames(tab1)

colnames(tab1)[startsWith(colnames(tab1),’a’)]


#selecting based on column names

for (cname in colnames(tab1)[startsWith(colnames(tab1),'a')]){

   print(tab1[,cname])

}


#selecting based on column datatype

for (cname in colnames(tab1)[sapply(tab1, is.numeric)]){

  print(tab1[,cname])

}
```