Revision:2 (Date:5/9/24)

Unit-II

Unit-II Data Wrangling

#tidyr commands

#gather(),spread(),unite(),seperate()

```
Temperature=data.frame(city=c("Pune","Nasik","Jalgaon","Pune","Nasik","Jalgaon","Pune","Nasik","Jalg
aon"),
                       Year=c(2010,2010,2010,2011,2011,2011,2012,2012,2012),
                       temperature=c(22,23,24,31,32,33,41,42,43))
View(Temperature)
```

#data frame

```
A=data.frame(country=c("India","US","FR"),
                         P2011=c(23,24,25),
                         P2012=c(32,33,34),
                         P2013=c(21,22,23))
View(A)
```

#gather command(#wide to long)

```
B=gather(A,"Year","n",2:4)
View(B)
```

#spread command(#long to wide)

```
C=spread(B,Year,n)
View(C)
```

```
library(tidyr)# to install package

new=spread(Temperature,Year,temperature)

View(new)


#gather command

new10=gather(new,"year","n",2:4)

View(new10)

new11=spread(new10,year,n)

View(new11)



# New dataframe

date=as.Date(2001-2-23,2004-1-22)

View(date)


# How to write date in R

# Package is lubridate.


install.packages("lubridate")



dates = c("05/27/84", "07/07/05")
```

```
ActualDates = as.Date(dates,
                      format = "%m/%d/%y")
View(ActualDates)


Temperature1=data.frame(city=c("Pune","Nasik","Jalgaon","Pune","Nasik","Jalgaon","Pune","Nasik","Jal
gaon"),
                        Year=c(2010,2010,2010,2011,2011,2011,2012,2012,2012),
                        temperature=c(22,23,24,31,32,33,41,42,43),
                        date=c('1/1/2021', '4/5/2021',
                             '2/3/2021', '11/21/2011', '10/11/2011', '11/21/2030',
'11/11/2022','11/12/2022','11/21/2010'))


View(Temperature1)


Temperature1[order(as.Date(Temperature1$date, format="%m/%d/%Y")), ]
View(Temperature1)


#seperate command
library(tidyr)


Temperature2=separate(Temperature1,date,c("date","month","year"),sep="/")
View(Temperature2)


#Unite command
Temperature3=unite(Temperature2,"date",date,month,year,sep="/")
View(Temperature3)
```

```
#Deplyr commands

#select(),filter(),mutate,summarise


# New Data frame

Fruits=c("orange","apple","banana","grapes")

Colour=c("orange","red","yellow","green")

price=c(60,120,40,50)

taste=c("tangy","sweet","sweet","sour")

favorite=c("ram","shyam","tiku","piku")


fruit1=data.frame(Fruits=c("orange","apple","banana","grapes"),

                  Colour=c("orange","red","yellow","green"),

                  price=c(60,120,40,50),

                  taste=c("tangy","sweet","sweet","sour"),

                  favorite=c("ram","shyam","tiku","piku"))

View(fruit1)


#select command


fruit2=select(fruit1,Fruits,Colour)

View(fruit2)


fruit3=select(fruit1,price:taste)

View(fruit3)
```

```
# filter command

fruit4=filter(fruit1,price>=50)

View(fruit4)


fruit5=filter(fruit1,price>=50,Fruits %in% c("orange","apple"))

View(fruit5)


#data frame

fruit=data.frame(Fruits=c("orange","apple","banana","grapes"),

                 Colour=c("orange","red","yellow","green"),

                 Mondayprice=c(60,120,40,50),

                 Tuesdayprice=c(80,152,90,70),

                 taste=c("tangy","sweet","sweet","sour"),

                 favorite=c("ram","shyam","tiku","piku"))

View(fruit)


#mutate command


fruit6=mutate(fruit,differenceprice=Tuesdayprice-Mondayprice)

View(fruit6)


#summarise command

fruit7=summarise(fruit,median=median(Mondayprice),variance=var(Mondayprice))

View(fruit7)
```

```
#arrange command

fruits8=arrange(fruit,Mondayprice)

View(fruits8)


fruits9=arrange(fruit,desc(Mondayprice))

View(fruits9)
```