```
---
#title: "NSS Data extraction"
#author: "SK JANE ALAM"
#Date: "13/06/2022"
#output: html_document
---

  install.packages("readr")
#Install and load required packages
library(dplyr)
library(readr)
library(tidyr)



#main extraction of NSSO data is done with the packages 'readr'


#Extraction of level 1 from NSSO 77th round




Dlevel1<-read_fwf(file="D:/NSS data/Nss77/Data/Visit1/r77s331v1L01.txt",
                  fwf_cols(centercd=c(1,3),    fsuslno=c(4,8),
                           round=c(9,10),      schedule=c(11,13),
                           sample=c(14,14),    sector=c(15,15),
                           region=c(16,18),    distcode=c(19,20),
                           stratum=c(21,22),   substratum=c(23,24),
                           subround=c(25,25),  fod=c(26,29),
```

```
                sssno=c(30,30),     hhno=c(31,32),
                visit=c(33,33),     level=c(34,35),
                filler=c(36,40),    infslno=c(41,42),
                responscd=c(43,43), surveycd=c(44,44),
                casualtycd=c(45,45),employcd1=c(46,49),
                employcd2=c(50,53), employcd3=c(54,57),
                sdate=c(58,63),     ddate=c(64,69),
                time=c(70,72),      noinvest=c(73,73),
                remarks1=c(74,74),  remarks2=c(75,75),
                remarks3=c(76,76),  remarks4=c(77,77),
                blank=c(78,126),    nsc=c(127,129),
                mlt=c(130,139)),
col_types=cols(centercd=col_character(),      fsuslno=col_character(),
          round=col_integer(),        schedule=col_character(),
          sample=col_character(),       sector=col_integer(),
          region=col_character(),   distcode=col_character(),
          stratum=col_character(),  substratum=col_character(),
          subround=col_character(), fod=col_character(),
          sssno=col_character(),    hhno=col_character(),
          visit=col_integer(),    level=col_character(),
          filler=col_character(),   infslno=col_character(),
          responscd=col_character(), surveycd=col_character(),
          casualtycd=col_character(), employcd1=col_character(),
          employcd2=col_character(), employcd3=col_character(),
          sdate=col_date(),     ddate=col_date(),
          time=col_character(),     noinvest=col_character(),
          remarks1=col_character(), remarks2=col_character(),
```

```
        remarks3=col_character(),  remarks4=col_character(),
        blank=col_character(),    nsc=col_number(),
        mlt=col_number()))
save(Dlevel1,file="Dlevel1.RData")

getwd()


#Extraction of level 6 from NSSO 77th round



Dlevel6<-read_fwf(file="D:/NSS data/Nss77/Data/Visit1/r77s331v1L06.txt",
          fwf_cols(centercd=c(1,3),    fsuslno=c(4,8),
                round=c(9,10),      schedule=c(11,13),
                sample=c(14,14),    sector=c(15,15),
                region=c(16,18),    distcode=c(19,20),
                stratum=c(21,22),   substratum=c(23,24),
                subround=c(25,25),  fod=c(26,29),
                sssno=c(30,30),     hhno=c(31,32),
                visit=c(33,33),     leve6=c(34,35),
                filler=c(36,39),    slnocrp=c(40,40),
                cropcd=c(41,44), unitcd=c(45,45),
                areairr=c(46,51),quantpirr=c(52,61),
                areaunirr=c(62,67), quantpunirr=c(68,77),
                totqua=c(78,87),
                nsc=c(127,129), mlt=c(130,139)),
```

```
        col_types=cols(centercd=col_character(),      fsuslno=col_character(),
                       round=col_integer(),        schedule=col_character(),
                       sample=col_character(),        sector=col_integer(),
                       region=col_character(),    distcode=col_character(),
                       stratum=col_character(),   substratum=col_character(),
                       subround=col_character(),  fod=col_character(),
                       sssno=col_character(),     hhno=col_character(),
                       visit=col_integer(),     level6=col_character(),
                       filler=col_character(),    slnocrp=col_character(),
                       cropcd=col_character(), unitcd=col_character(),
                       areairr=col_character(), quantpirr=col_character(),
                       areaunirr=col_character(), quantpunirr=col_character(),
                       totqua=col_character(),
                       nsc=col_number(),
                       mlt=col_number()))
save(Dlevel6,file="Dlevel6.RData")



##Creating weights from multiplier for each of these levels
#NSSO data though it is a sample survey, but it has wonderful value of weights
#through which you can exactly stipulate the entire results for country as a whole for whole results.
Dlevel1$weight<-Dlevel1$mlt/100
Dlevel2$weight<-Dlevel2$mlt/100
Dlevel3$weight<-Dlevel3$mlt/100
```

```
Dlevel4$weight<-Dlevel4$mlt/100

Dlevel5$weight<-Dlevel5$mlt/100

Dlevel6$weight<-Dlevel6$mlt/100

Dlevel7$weight<-Dlevel7$mlt/100



##generation of key for merging

Dlevel1$key<-paste0(Dlevel1$fsuslno,Dlevel1$sssno,Dlevel1$hhno)

Dlevel6$key<-paste0(Dlevel1$fsuslno,Dlevel1$sssno,Dlevel1$hhno)



Dlevel6$keyind<-paste0(Dlevel6$fsuslno,Dlevel6$sssno,Dlevel6$hhno,Dlevel6$slNo)

Dlevel7$keyind<-paste0(Dlevel7$fsuslno,Dlevel7$sssno,Dlevel7$hhno,Dlevel7$slNo)



##Merge two levels

Dlevel6_7 <-merge(Dlevel6,Dlevel7, by="keyind",all=TRUE)

##saving the merging files

save(Dlevel6_7,file="Dlevel6_7.RData")



##checking pattern of the data

table(Dlevel6$cropcd)

summarise(Dlevel6,n=mean(rate,na.rm=TRUE))
```

```
#average with the use of multiplier
Dlevel6_7%>%
  filter(cropcd==101)%>%
  group_by(whom_sold)%>%
  summarise(w=weighted.mean(rate,weight.x))

#count with multiplier
Dlevel6_7%>%
  filter(cropcode.x==101)%>%
  group_by(whom_sold)%>%
  count(sale_satis,wt=weight.x)


#{R use of crosstab with pollster package}
topline(Dlevel6_7,variable=whom_sold,weight=weight.x)
crosstab(Dlevel6_7,x=whom_sold, y=sale_satis,weight=weight.x)
crosstab_3way(Dlevel6_7,x=whom_sold, y=sale_satis, z=sssno.x, weight=weight.x)
```