

```

---
title: "Data Transformation and Wrangling"
author: "Gulfam"
date: "12/20/2021"
output: html_document
---
```{r}

library(tidyverse)

library(tidyr)

#data_types
numeric
character
logical
factor

###step-1-set your home directory(ctrl+shift+H)
###step-2-import your data in R(readxl , readcsv)
###step-3-data structure and types
###step-4-data transformation and tidy tips and tricks

#how to confirm class of variables

a <- 76

is.numeric(a) #true

is.character(a) #false

is.numeric(iris$Sepal.Length) #true

is.factor(iris$Species) #true

#how to transform data

num <- '3.2'
class(num)
num <- as.numeric(num)
class(num)

num1 <- c(1,2,3,4,5,6)
class(num1)
num1 <- as.logical(num1)
class(num1)
num1

num1 <- c(1,2,3,0,4,5,6,-1) #only zero becomes false

```

```
class(num1)
num1 <- as.logical(num1)
class(num1)
num1
is.logical(num1)
```

```
#data_types are more imp than data itself.....
#data structure is more important than data significance....
```

```
x <- diamonds
x
head(x)
str(x)
view(x)
```

```
class(x$carat)
class(x$price)
```

```
as.data.frame(x)
as_tibble(x)
glimpse(x)
```

```
library(dyplr)
```

```
x$carat <- as.numeric(x$carat)
class(x$carat)
```

```
vec1 <- c(1,2,3,4,4,5,6,6,7,7,76,5,5)
class(vec1)
vec2 <- as.factor(vec1)
class(vec2)
```

```
#continuous variables
numeric
integer
string character
vector
factor
```

```
###Data transformation
tidy means to collect data and tranform it into usable form###
x <- CO2
glimpse(x)
as_tibble(x)

view(CO2)
```

```

#if some treatment or result values are repeating themselves take them as
factor
if only two parameters are there in a column take them as logical

x1 <- chickwts
x1 <- as.factor(chickwts$feed)
x1

###manipulating a data

install.packages('dslabs')
library(dslabs)

x <- diamonds

#add a column
as_tibble(x)
x <- mutate(x , a=x+y+z)
x <- mutate(x , a2 = a^2)
x
view(x)
x <- mutate(x , mean=mean(x+y+z))
as_tribble(x)
x
view(x)

###sub-setting
#using filter

as_tibble(x)

x_a1 <- filter(x , x==4)
as_tibble(x_a1)
view(x_a1)

x_a2 <- filter(x , cut=='Ideal') #subsetting a character
view(x_a2)

#subsetting using select function

as_tibble(x)
x1 <- select(x , cut , color , x) #selecting different columns
as_tibble(x1)

x <- murders
as_tibble(x)

x1 <- filter(x , population == 9535483)
view(x1)
x2 <- filter(x ,state == 'California')
view(x2)

```

```

###less code for more output
###dplyr %>%

diamonds

x <- diamonds %>%
 filter(cut == 'Good') %>%
 select(color , clarity , depth) %>%
 mutate(b= depth * 2)

x %>% mutate(c = b + depth)

library(tidyverse)

#summarise function

x <- mtcars
as_tibble(x)

summarise(mtcars , avg=mean(mpg))
x %>% summarise(avg = mean(mpg))

#group_by

mtcars %>%
 group_by(gear,cyl) %>% #multiple grouping
 summarise(mean1 = mean(mpg) , mean2 = mean(hp))

#starwars

view(starwars)
x <- starwars %>%
 rowwise() %>% #do process rowwise

 mutate(film_count = length(films))
view(x)

#manipulating
#filter

filter(mtcars , mpg > 20)

#distinct

```

```

x <- mtcars
distinct(x , gear)
view(x)

?distinct

#slice

x2 <- slice(x , 10:15) #rowwise

X2 <- slice_sample(x , n=10, replace = T)

x2 <- slice_min(x , mpg , prop = .25)

x2 <- slice_head(x , n = 7)

#order data

arrange(x , mpg)

arrange(x , desc(mpg))

#add_rows

view(cars)
cars <- add_row(cars , speed = 4 , dist = 7)

cars <- add_row(cars , speed = 4)

cars[is.na(cars)] = 0 #update NA values

#manipulation of variables

x <- pull(mtcars , wt) #column into vectors

x1 <- select(mtcars , mpg , wt) #extract columns

x2 <- relocate(mtcars , mpg , cyl , .after = last_col())

x3 <- select(mtcars , mpg:drat)

#manipulate multiple variavlebs at once

```

```
x4 <- summarise(mtcars , across(everything() , mean)) #mean of every
variable

y <- rowMeans(mtcars) %>%
 mutate(mtcars , mean = y)

a <- mtcars %>% # means by rows and adding it into the data
 mutate(mean = rowMeans(mtcars))

x1 <- mutate(mtcars , gpm = 1/mpg)

x2 <- transmute(mtcars , gpm= 1/mpg)

x3 <- rename(mtcars , gpm = mpg)
```









