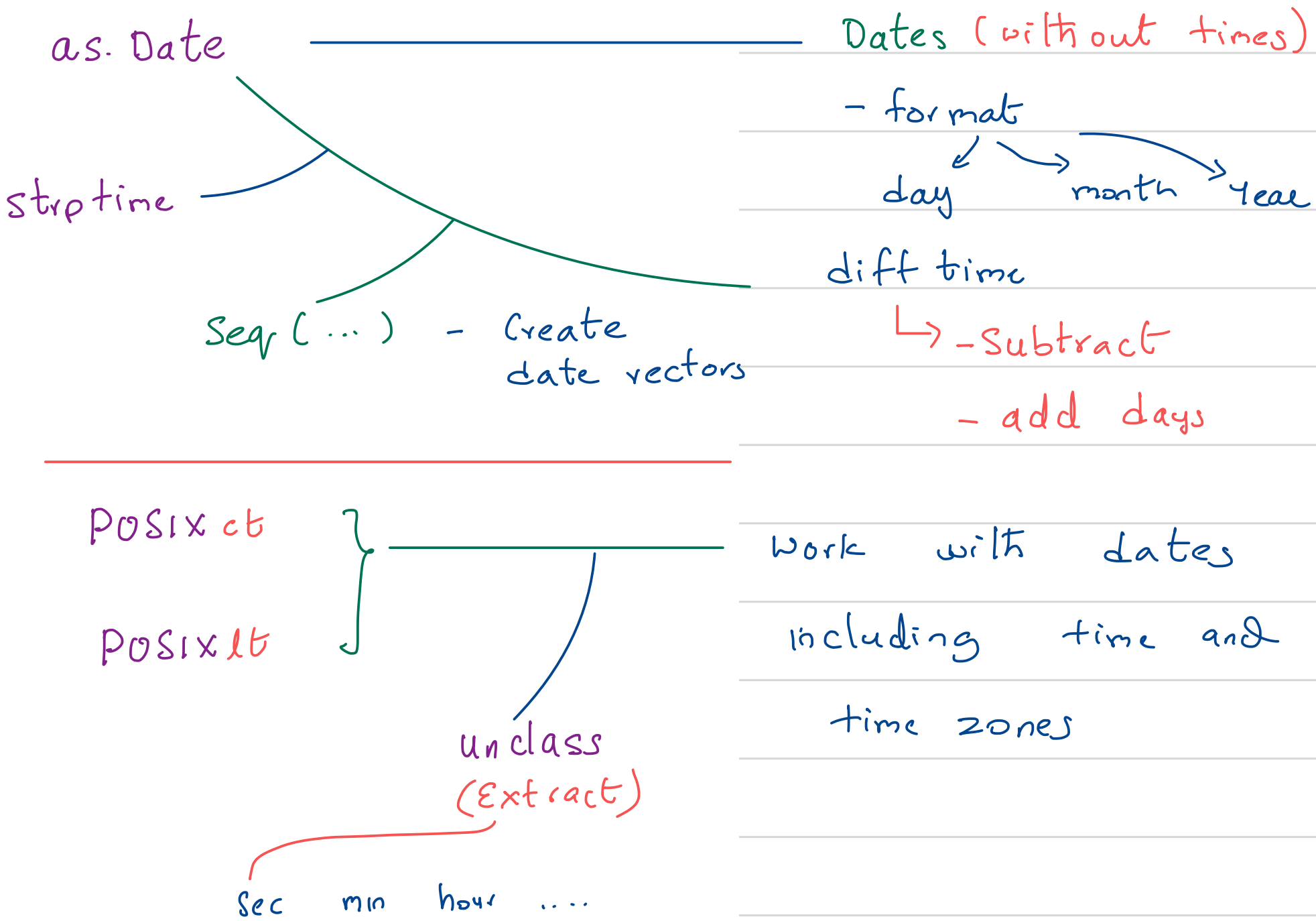
R- Date and Time



R- Date and Time

```
> require(chron)
> time1.c <- as.chron("2013-07-24 23:55:26")
> time1.c
[1] (07/24/13 23:55:26)
> time2.c <- as.chron("07/25/13 08:32:07", "%m/%d/%y %H:%M:%S")
> time2.c
[1] (07/25/13 08:32:07)
```

Require package chron

Creating times in chron

Specified format

R- Date and Time

time1.c = [1] (07/24/13 23:55:26)

time2.c = [1] (07/25/13 08:32:07)

Chron

```
> dates(time1.c)
    day
07/24/13
```

Extracting date

```
> time2.c > time1.c
[1] TRUE
```

Compare times

```
> time1.c + 10
[1] (08/03/13 23:55:26)
```

add days

```
> time2.c - time1.c
[1] 08:36:41
```

subtraction

```
> difftime(time2.c, time1.c, units = "hours")
Time difference of 8.611389 hours
```

```
> as.chron("2013-03-10 08:32:07") - as.chron("2013-03-09 23:55:26")
[1] 08:36:41
> as.chron("2013-03-10 08:32:07") - as.chron("2013-03-09 23:55:26")
[1] 08:36:41
```

Does not adjust for Daylight Savings time

R- read.csv and names

```
> decdf= read.csv(file="    Master.csv", header=TRUE)
> head(decdf,2)
  Sno        District State.P.No Age.In.Years    Sex
1   1      Kalaburagi          6           76   Male
2   2 Chikkaballapura         53           70 Female
                   Description Symptoms Co.Morbidities
1 Travel history to Saudi Arabia     <NA>  HTN & Asthama
2        Travel history to Mecca     <NA>           <NA>
   DOA  DOD    MB.Date Notes
1 <NA> <NA> 2020-03-13  <NA>
2 <NA> <NA> 2020-03-26  <NA>
```

Daily data on deceased from the COVID-19 bulletins of GoK.

```
> names(decdf) = c("Sno","District","Pid","Age","Sex",
+                  "Description","Symptoms","CMB",
+                  "DOA","DOD","MB.Date","Notes")
> head(decdf,2)
  Sno        District Pid Age    Sex
1   1      Kalaburagi   6  76   Male
2   2 Chikkaballapura  53  70 Female
                   Description Symptoms           CMB
1 Travel history to Saudi Arabia     <NA> HTN & Asthama
2        Travel history to Mecca     <NA>          <NA>
   DOA  DOD    MB.Date Notes
1 <NA> <NA> 2020-03-13  <NA>
2 <NA> <NA> 2020-03-26  <NA>
```

Rename the variables

Exercise :- [TRY]

- decdf $ Age
- mode - ?
  class - ?
- summary (decdf)

OBJECTIVES :-

(1) - Introduction to dplyr
- we will review basic commands

- Use online resources after that to understand the package.

(2) - Understand factors and levels.

R- Basic Introduction to Dplyr– filter

```
> library(dplyr)
> f100=filter(decdf, Age>100)
> f100F=filter(decdf, Age>100 & Sex=="Female")
> head(f100,2)
    Sno        District     Pid Age  Sex Description
1  3277 Bengaluru Urban  180841 102 Male         ILI
2 17972 Bengaluru Rural 1361618 102 Male        SARI
        Symptoms      CMB        DOA        DOD    MB.Date
1   Fever, Cough CKD, IHD 2020-08-08 2020-08-08 2020-08-10
2 Breathlessness  DM, HTN 2021-04-24 2021-04-25 2021-05-08
  Notes
1  <NA>
2  <NA>
> head(f100F,2)
    Sno        District     Pid Age    Sex Description
1 27273 Bengaluru Urban 2360283 102 Female        SARI
2 37190          Kodagu 2947715 101 Female        SARI
                     Symptoms CMB        DOA        DOD
1               Breathlessness   - 2021-05-11 2021-05-25
2 Fever, Cough, Breathlessness   - 2021-08-01 2021-08-26
     MB.Date Notes
1 2021-05-27  <NA>
2 2021-08-27  <NA>
```

filter() - use it to subset a dataframe

- retains rows that satisfy the condition.

- drop the "NA" rows. [-unlike subset()]

R- Basic Introduction to Dplyr-mutate

```
> decdf= mutate(decdf,
+       reporting.time= as.Date(decdf$MB.Date)-as.Date(decdf$DOD))
> head(decdf,2)
  Sno        District Pid Age    Sex
1   1      Kalaburagi   6  76   Male
2   2 Chikkaballapura  53  70 Female
                   Description Symptoms           CMB
1 Travel history to Saudi Arabia     <NA> HTN & Asthama
2        Travel history to Mecca     <NA>          <NA>
   DOA  DOD    MB.Date Notes reporting.time
1 <NA> <NA> 2020-03-13  <NA>        NA days
2 <NA> <NA> 2020-03-26  <NA>        NA days
> decdf= mutate(decdf,
+                 Month= months(as.Date(decdf$MB.Date)))
> head(decdf,2)
  Sno        District Pid Age    Sex
1   1      Kalaburagi   6  76   Male
2   2 Chikkaballapura  53  70 Female
                   Description Symptoms           CMB
1 Travel history to Saudi Arabia     <NA> HTN & Asthama
2        Travel history to Mecca     <NA>          <NA>
   DOA  DOD    MB.Date Notes reporting.time Month
1 <NA> <NA> 2020-03-13  <NA>        NA days March
2 <NA> <NA> 2020-03-26  <NA>        NA days March
```

mutate() -

add new variables

and preserve

existing ones.

Exercise :-

testdf = transmute(

decdf,

decdf$test=0)

add variables

and drop existing

ones.

R- Basic Introduction to Dplyr - Slicing and Distinct

```
> DT=distinct(decdf, Age)
> head(DT,2)
  Age
1  76
2  70
```

```
> SL=slice(decdf,10:12)
> head(SL,2)
  Sno        District Pid Age  Sex Description
1  10      Vijayapura 257  69 Male        <NA>
2  11 Chikkaballapura 250  65 Male        <NA>
                                        Symptoms
1                                           <NA>
2 H1N1 positive & COPD with obstructive sleep apnea
       CMB        DOA        DOD    MB.Date Notes
1     <NA>       <NA>       <NA> 2020-04-14  <NA>
2 DM & HTN 2020-04-13 2020-04-15 2020-04-15  <NA>
  reporting.time Month
1        NA days April
2         0 days April
```

Select the unique or distinct rows from a dataframe

Exercise (option)

distinct (decdf, Age, .keep_all = TRUE)

- allow us to keep other variables.

- index rows by location

- select ∫ remove

Exercise : - duplicate

R- Basic Introduction to Dplyr- Group and Summarise

```
> GS=group_by(decdf, Sex)
> head(GS,2)
# A tibble: 2 x 14
# Groups:   Sex [2]
    Sno District          Pid   Age Sex    Description Symptoms
  <int> <chr>           <int> <dbl> <chr>  <chr>       <chr>
1     1 Kalaburagi          6    76 Male   Travel his~ <NA>
2     2 Chikkaballapura    53    70 Female Travel his~ <NA>
# ... with 7 more variables: CMB <chr>, DOA <chr>,
#   DOD <chr>, MB.Date <chr>, Notes <chr>,
#   reporting.time <drtn>, Month <chr>
```

← Specifies the grouping

group_by( )

- groups data by one or more variables
- head( ) - will not display grouping

```
> summarise(GS, mean(Age, na.rm=TRUE))
# A tibble: 10 x 2
   Sex             `mean(Age, na.rm = TRUE)`
   <chr>                               <dbl>
 1 F                                    65.2
 2 Female                               60.7
 3 M                                    66.1
 4 M E23                                71
 5 Male                                 60.7
 6 N                                    39
 7 O                                    67
 8 Other categ ogy                      53
 9 TG                                   65
10 <NA>                                 51.8
```

- Summarise multiple values into a single value.
- Used in conjuction with another function.

calculates mean across various groups in Sex variable.

R- Basic Introduction to Dplyr-Sampling Random Rows

```
> sample_n(decdf, size = 2)
    Sno        District     Pid Age    Sex Description
1 25432 Bengaluru Urban 1775128  69 Female         ILI
2 39129 Bengaluru Urban 3379259  45   Male         ILI
      Symptoms      CMB        DOA        DOD    MB.Date
1        Fever      HTN 2021-05-04 2021-05-11 2021-05-24
2 Fever, Cough DM, IHD 2022-01-19 2022-01-21 2022-02-04
  Notes reporting.time    Month
1  <NA>         13 days      May
2  <NA>         14 days February
```

sample() – select random rows from a dataframe.

Selects 2 rows at random.

```
> sample_frac(decdf, size = 0.0001)
   Sno       District     Pid Age    Sex Description
1 20748 Bengaluru Urban 1421748  69   Male        SARI
2 31577          Haveri 2724392  35 Female        SARI
3   942 Bengaluru Urban   20896  68   Male         ILI
4  8581      Kalaburagi  574615  68   Male        SARI
                    Symptoms     CMB        DOA
1       Cough, Breathlessness       - 2021-04-27
2 Fever, Cough, Breathlessness DM, HTN 2021-06-04
3                      Fever      DM 2020-06-28
4 Fever, Cough, Breathlessness      CA 2020-09-23
        DOD    MB.Date Notes reporting.time     Month
1 2021-05-06 2021-05-14  <NA>         8 days       May
2 2021-06-05 2021-06-06  <NA>         1 days      June
3 2020-06-29 2020-07-16  <NA>        17 days      July
4 2020-09-26 2020-09-27  <NA>         1 days September
```

Selects 0.0001 fraction of rows at random.

Exercise :- slice_sample

R- Basic Introduction to Dplyr- Count and Order

```
> count(decdf, Month)
       Month     n
1      April  2974
2     August  4108
3   December   430
4   February   385
5    January   843
6       July  3561
7       June  6049
8      March   239
9        May 13599
10  November   737
11   October  2593
12 September  3643
13      <NA>   111
```

tallies observation according to the group Specified.

Exercise: Use group_by and count across years separately

```
> orderdf=arrange(decdf, Age)
> head(orderdf,2)
    Sno        District     Pid    Age    Sex Description
1 14253 Bengaluru Urban 1260623 0.0000   Male        SARI
2 20970      Ramanagara 2032210 0.0082 Female        SARI
        Symptoms CMB        DOA        DOD    MB.Date Notes
1 Breathlessness HTN 2021-04-21 2021-04-23 2021-04-24  <NA>
2 Breathlessness   - 2021-05-07 2021-05-10 2021-05-14  <NA>
  reporting.time Month
1         1 days April
2         4 days   May
```

arranges the rows by the variable.

Exercise :- - character variables
- descending
- multiple variables

R- Basic Introduction to Dplyr-Pipe

```
> filteredData <- filter(decdf, Month != "September" )
> groupedData <- group_by(filteredData, Month)
> summarise(groupedData, mean(Age, na.rm = TRUE))
```

```
> decdf %>%
+   filter(Month != "September") %>%
+   group_by(Month) %>%
+   summarise(mean(Age, na.rm = TRUE))
# A tibble: 11 x 2
   Month     `mean(Age, na.rm = TRUE)`
   <chr>                         <dbl>
 1 April                          61.3
 2 August                         61.3
 3 December                       64.8
 4 February                       65.1
 5 January                        63.6
 6 July                           60.0
 7 June                           59.6
 8 March                          65.9
 9 May                            59.4
10 November                       64.2
11 October                        63.6
```

Exercise :- Do the computation.

3 calculations

- %>% — pipe operator can be used to chain code
- useful for performing several operations