Multi-stage deterministic linkages and case definitions with diyar: : CHEAT SHEET
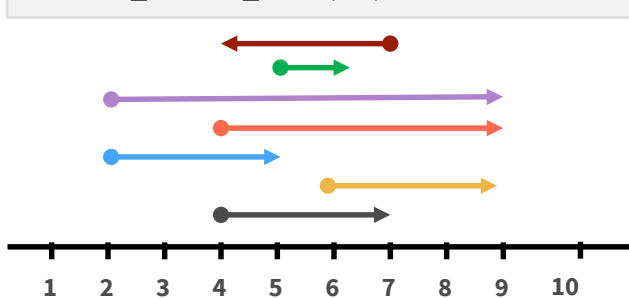
number_line objects

- A series of real numbers on a number line.
- S4 objects used in record and episode grouping

```
number_line(4, 7)
```
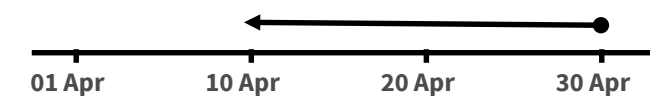

MANIPULATE NUMBER LINE OBJECTS

```
nl <- number_line(4, 7)
shift_number_line(nl,2)
shift_number_line(nl,-2)
expand_number_line(nl, 2, "end")
expand_number_line(nl, 2)
expand_number_line(nl, -1)
reverse_number_line(nl)
```



```
> number_line_width(nl)
[1] 3
> number_line_sequence(nl, 1)
[1] 4 5 6 7
> number_line_sequence(nl, .5)
[1] 4.0 4.5 5.0 5.5 6.0 6.5 7.0
> left_point(reverse_number_line(nl))
[1] 7
> start_point(reverse_number_line(nl))
[1] 4
> right_point(reverse_number_line(nl))
[1] 4
> end_point(reverse_number_line(nl))
[1] 7
```

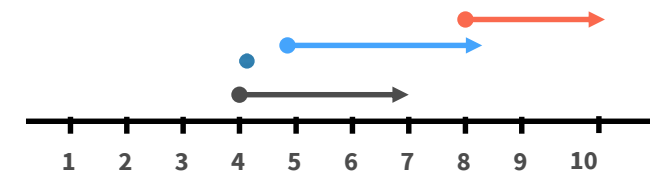Also supports objects that can be coerced to numeric values

```
dates <- c("30/04/2019", "10/04/2019")
dates <- as.Date(dates, "%d/%m/%Y")
number_line(dates[1], dates[2])
```


COMBINE NUMBER LINE OBJECTS
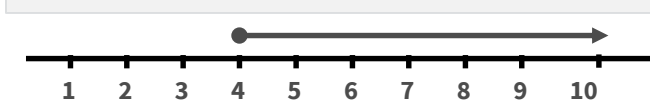
Overlapping number_line objects can be merged

```
n1 <- number_line(4,7)
n2 <- as.number_line(4)
n3 <- number_line(5,8)
n4 <- number_line(8,10)
nl <- c(n1, n2, n3, n4)
```



```
compress_number_line(nl)
```



```
compress_number_line(nl, collapse =T)
```
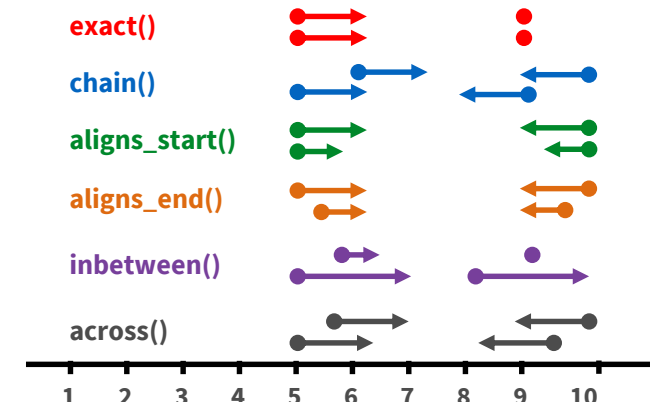

- overlap methods

exact()
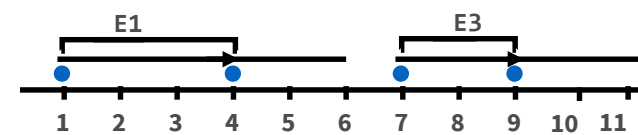chain()
aligns_start()
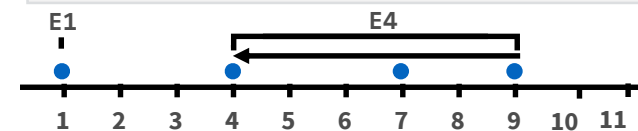aligns_end()
inbetween()
across()

Episode grouping

EPISODES FROM POINTS IN TIME

```
dates <- c("01","04","07","09")
dates <- paste(dates,"04/2019",sep= "/")
dates <- as.Date(dates, "%d/%m/%Y")

fixed_episodes(dates, case_length = 5)
```



```
fixed_episodes(dates, case_length = 5,
from_last = T)
```
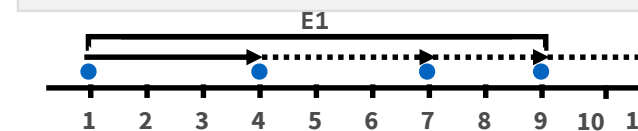


```
fixed_episodes(dates, case_length = 5,
episode_unit = "hours")
```
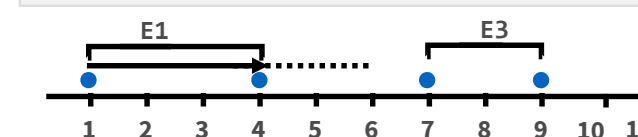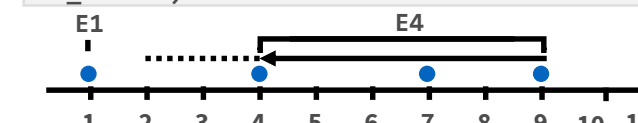

ROLLING EPISODES

```
rolling_episodes(dates, case_length = 5)
```



```
rolling_episodes(dates, case_length = 5,
recurrence_length =2)
```
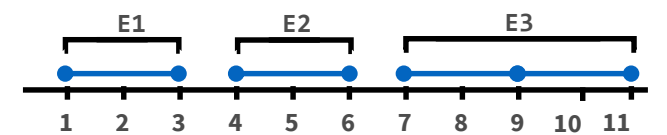


```
rolling_episodes(dates, case_length = 5,
recurrence_length =2, from_last = T,
to_s4 = T)
```


EPISODES FROM PERIODS IN TIME

```
p <- as.number_line(dates)
p <- expand_number_line(p,2, "end")
periods <- p

fixed_episodes(periods,case_length = 0)
```



```
fixed_episodes(periods, case_length = 1)
```
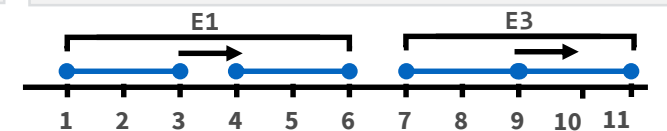


```
rolling_episodes(periods, case_length = 1)
```
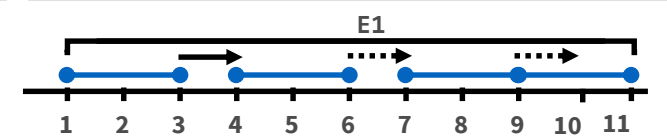

CONTROL CASE ASSIGNMENT

```
fixed_episodes(dates, case_length =
c(1,3,2,2))
```
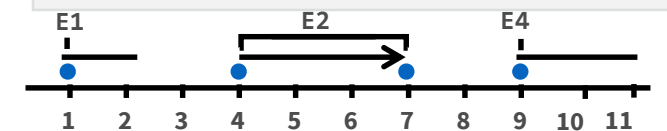


```
fixed_episodes(dates, case_length =5,
custom_sort=c(1,0,1,1))
```
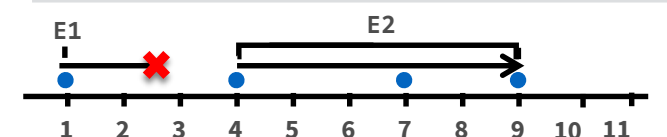

solid lines – case window
dashed lines – recurrence windows
solid end – start of an episode
arrow head – end of an episode

See more useful features here

Multi-stage deterministic linkages and case definitions with diyar: : **CHEAT SHEET**

Record grouping

- **Multi-stage deterministic linkage**

- **Relevance of each stage controlled by** `criteria`

- **Use** `sub_criteria` **for additional matching conditions**

- **Missing data handled with alternative matching** `criteria`

- **Group records separately within subsets of a dataset with** `strata`
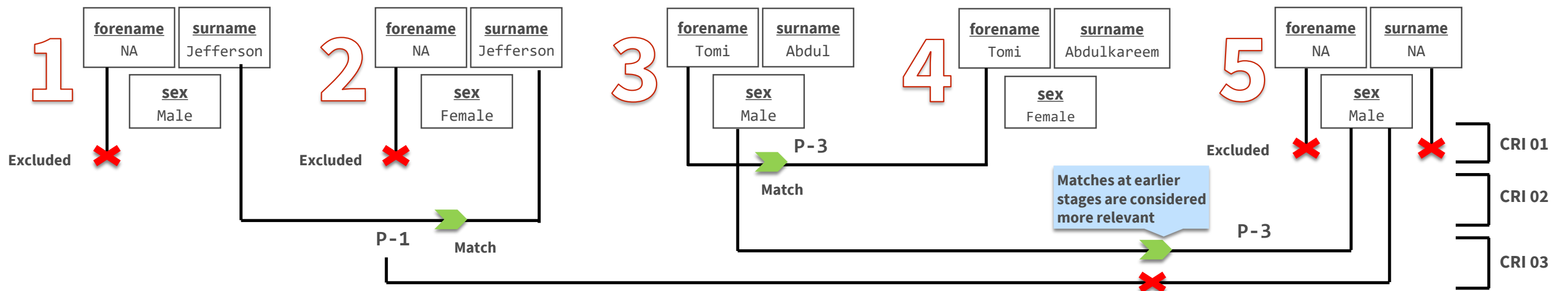
```
record_group(diyar::patient_list_2, sn = rd_id, criteria = c(forename, surname, sex))
```

1 forename: NA | surname: Jefferson | sex: Male — Excluded ✖

2 forename: NA | surname: Jefferson | sex: Female — Excluded ✖

3 forename: Tomi | surname: Abdul | sex: Male — P-3 / Match

4 forename: Tomi | surname: Abdulkareem | sex: Female

5 forename: NA | surname: NA | sex: Male — Excluded ✖

P-1 Match

P-3

CRI 01
CRI 02
CRI 03

> **Matches at earlier stages are considered more relevant**

```
df <- Opes[c("department","date_of_birth","db_pt1","db_pt2","db_pt3")]
df$age <- dmy("02/11/2019") -  dmy(df$date_of_birth)
df$age <- round(as.numeric(df$age)/365)
```

```
df$age_range <- number_line(df$age, df$age +5, gid=df$age)
record_group(df, criteria = c(department, department),
sub_criteria = list("s1"=c("age_range"), "s2"=c("db_pt1","db_pt2","db_pt3")),
display = FALSE)
```

P-2 (CRI 02)

1 Dept.: Procure. | DoB: 23/03/1986 | Age: 34

2 Dept.: Security | DoB: 23/03/1986 | Age: 34

3 Dept.: Security | DoB: 23/03/1968 | Age: 52

4 Dept.: Publish. | DoB: 01/02/1985 | Age: 40 — P-4 (CRI 01)

5 Dept.: Publish. | DoB: 02/01/1985 | Age: 35

6 Dept.: Publish. | DoB: 11/03/1964 | Age: 56 — P-6 (CRI 01)

7 Dept.: Publish. | DoB: 11/03/1964 | Age: 56

8 Dept.: Procure. | DoB: 11/10/1985 | Age: 34

P-1 (CRI 01)

> **Match a range of values using** `number_line` **objects**