

Using EpiCurve

jean.pierre.decorps@gmail.com

2020-08-26

Package EpiCurve

Description

EpiCurve allows the user to create epidemic curves from case-based and aggregated data.

Details

The EpiCurve function creates a graph of number of cases by time of illness (for example date of onset). Each case is represented by a square. EpiCurve allows the time unit for the x-axis to have hourly, daily, weekly or monthly intervals. The hourly interval can be split into 1, 2, 3, 4, 6, 8 or 12 hour time units.

EpiCurve works on both case-based (one case per line) or aggregated data (where there is a count of cases for each date). With aggregated data, you need to specify the variable for the count of cases in the “freq” parameter.

With case-based (non-aggregated data), the date format for EpiCurve can be:

- hourly: YYYY-MM-DD HH:MM or YYYY-mm-DD HH:MM:SS
- daily: YYYY-MM-DD
- monthly: YYYY-MM

If the date format is daily or hourly, you can change and force the period for aggregation on the graph with the parameter “period” setted with “day”, “week” or “month”.

For aggregated data, the date formats can be as above, but they can also be weekly: YYYY-Wnn. Here, we need to specify how the data are aggregated in the parameter “period”. If we want to further aggregate the aggregated data for the epidemic curve (e.g. move from daily aggregated cases to weekly aggregated cases), we can specify the parameter “to.period”.

When the date format is hourly, the dataset is considered case-based, whether the “freq” parameter of the EpiCurve function is supplied or not.

The EpiCurve function

```
EpiCurve (  
  x,  
  date = NULL,  
  freq = NULL,  
  cutvar = NULL,  
  period = NULL,  
  to.period = NULL,  
  split = 1,  
  cutorder = NULL,  
  colors = NULL,  
  title = NULL,  
  xlabel = NULL,  
  ylabel = NULL,  
  note = NULL,  
  square = TRUE  
)
```

Arguments

| Parameter | Description |
|------------------|---|
| x | data.frame with at least one column with dates |
| date | character , name of date column |
| freq | character , name of a column with a value to display |
| cutvar | character , name of a column with factors |
| period | character , c(“hour”, “day”, “week”, “month”) |
| to.period | character , Convert date period to another period only for aggregated data. If period is “day”, to.period can be “week” or “month”. If period is “week”, to.period can be “month”. |
| split | integer , c(1,2,3,4,6,8,12) value for hourly split |
| cutorder | character vector of factors |
| colors | character , vector of colorss |
| title | character , title of the plot |
| xlabel | character , label for x axis |
| ylabel | character , label for y axis |
| note | character , add a note under the graph |
| square | boolean , If TRUE (default) squares are used to plot the curve, else if the number of cases is too hight please use square = FALSE. |

Depends

ggplot2, dplyr, ISOweek, scales, timeDate

Plot non-aggregated cases

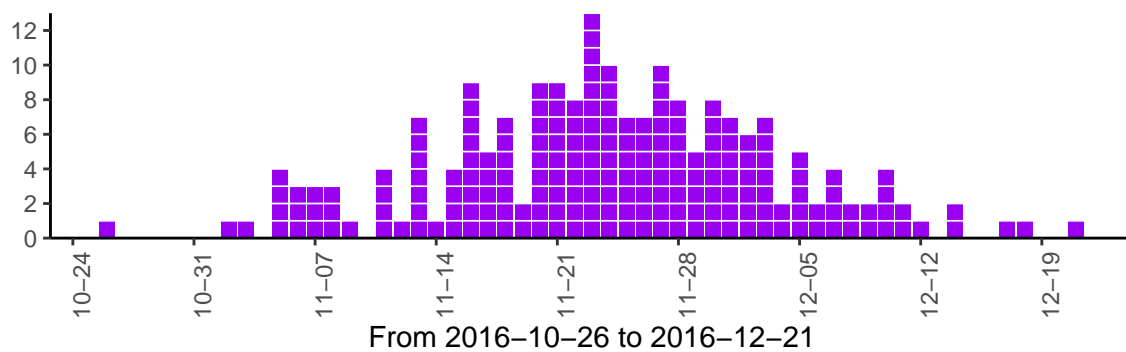
Daily - non-aggregated cases

```
DF <- read.csv("daily_unaggregated_cases.csv", stringsAsFactors=FALSE)
kable(head(DF, 12))
```

| UTS | V1 | V2 |
|------------|-------|-----|
| 2016-10-26 | 7.20 | 188 |
| 2016-11-02 | 7.03 | 95 |
| 2016-11-03 | 5.14 | 160 |
| 2016-11-05 | 9.89 | 165 |
| 2016-11-05 | 9.69 | 109 |
| 2016-11-05 | 4.15 | 154 |
| 2016-11-05 | 4.97 | 144 |
| 2016-11-06 | 8.97 | 187 |
| 2016-11-06 | 4.45 | 120 |
| 2016-11-06 | 6.60 | 116 |
| 2016-11-07 | 7.68 | 141 |
| 2016-11-07 | 10.08 | 126 |

```
EpiCurve(DF,
  date = "UTS", period = "day", colors = "#9900ef",
  xlabel=sprintf("From %s to %s", min(DF$UTS), max(DF$UTS)))
```

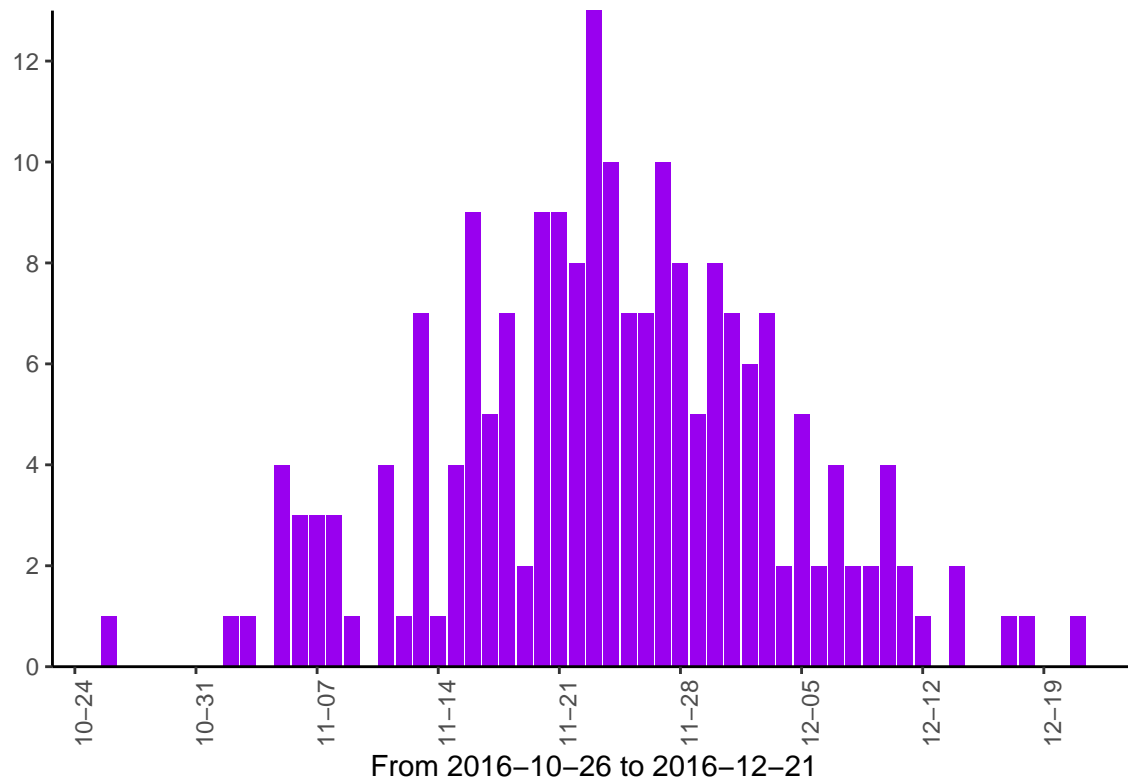
```
## Day
## 'data.frame':   57 obs. of  4 variables:
## $ Freq: num  1 0 0 0 0 0 0 1 1 0 ...
## $ Cut : chr  "1 cas" NA NA NA ...
## $ Day : chr  "2016-10-26" "2016-10-27" "2016-10-28" "2016-10-29" ...
## $ Date: Date, format: "2016-10-26" "2016-10-27" ...
```



With no squares

```
EpiCurve(DF,
  date = "UTS",
  period = "day",
  colors = "#9900ef",
  xlabel=sprintf("From %s to %s", min(DF$UTS), max(DF$UTS)),
  square = F)

## Day
## 'data.frame': 57 obs. of 4 variables:
## $ Freq: num 1 0 0 0 0 0 0 1 1 0 ...
## $ Cut : chr "1 cas" NA NA NA ...
## $ Day : chr "2016-10-26" "2016-10-27" "2016-10-28" "2016-10-29" ...
## $ Date: Date, format: "2016-10-26" "2016-10-27" ...
```

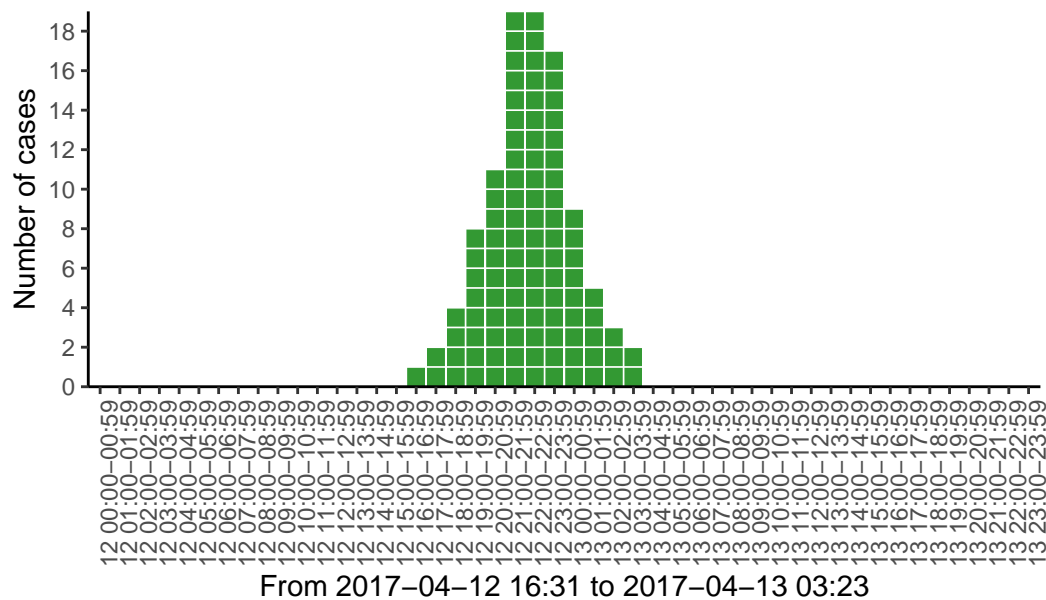


Hourly - non-aggregated cases

```
DF <- read.csv("hourly_unaggregated_cases.csv", stringsAsFactors=FALSE)
kable(head(DF, 12))
```

| UTS | X1 | X2 |
|------------------|-------|-----|
| 2017-04-12 16:31 | 5.17 | 166 |
| 2017-04-12 17:35 | 8.69 | 101 |
| 2017-04-12 17:38 | 6.81 | 140 |
| 2017-04-12 18:06 | 4.95 | 120 |
| 2017-04-12 18:36 | 10.92 | 189 |
| 2017-04-12 18:38 | 7.02 | 185 |
| 2017-04-12 18:43 | 8.03 | 175 |
| 2017-04-12 19:05 | 6.39 | 102 |
| 2017-04-12 19:11 | 4.61 | 126 |
| 2017-04-12 19:24 | 6.36 | 188 |
| 2017-04-12 19:37 | 7.80 | 112 |
| 2017-04-12 19:41 | 6.18 | 123 |

```
EpiCurve(DF,
  date = "UTS",
  period = "hour",
  split = 1,
  colors = "#339933",
  ylabel="Number of cases",
  xlabel=sprintf("From %s to %s", min(DF$UTS), max(DF$UTS)))
```

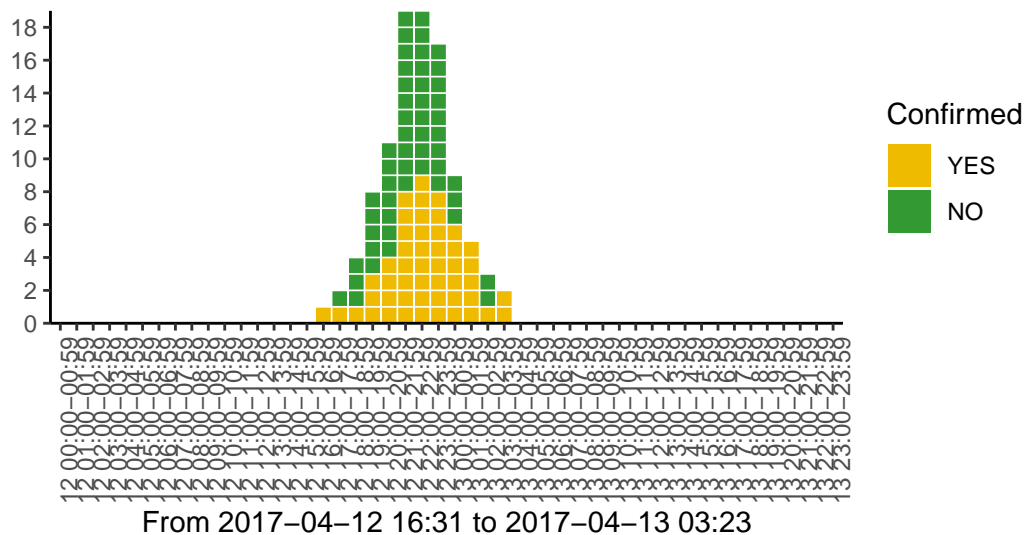


Hourly - non-aggregated cases with factors

```
DF <- read.csv("hourly_unaggregated_cases_factors.csv", stringsAsFactors=FALSE)
kable(head(DF, 12))
```

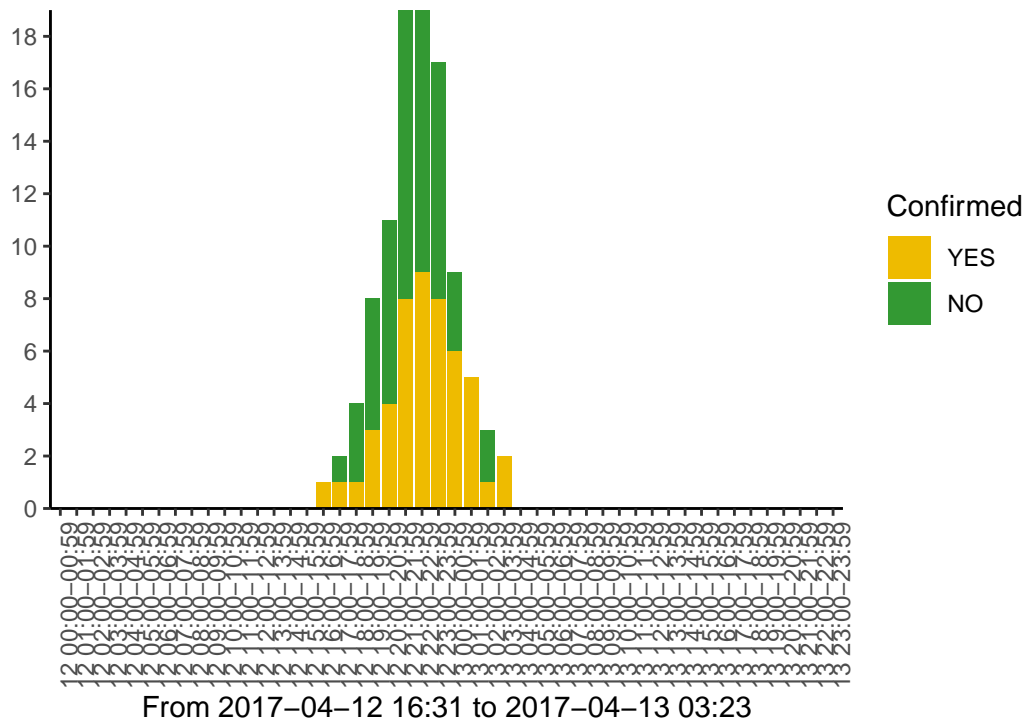
| UTS | X1 | X2 | Confirmed |
|------------------|-------|-----|-----------|
| 2017-04-12 16:31 | 5.17 | 166 | YES |
| 2017-04-12 17:35 | 8.69 | 101 | YES |
| 2017-04-12 17:38 | 6.81 | 140 | NO |
| 2017-04-12 18:06 | 4.95 | 120 | NO |
| 2017-04-12 18:36 | 10.92 | 189 | NO |
| 2017-04-12 18:38 | 7.02 | 185 | YES |
| 2017-04-12 18:43 | 8.03 | 175 | NO |
| 2017-04-12 19:05 | 6.39 | 102 | NO |
| 2017-04-12 19:11 | 4.61 | 126 | NO |
| 2017-04-12 19:24 | 6.36 | 188 | YES |
| 2017-04-12 19:37 | 7.80 | 112 | NO |
| 2017-04-12 19:41 | 6.18 | 123 | NO |

```
EpiCurve(DF,
  date = "UTS",
  period = "hour",
  split = 1,
  cutvar = "Confirmed",
  colors = c("#339933", "#eebb00"),
  xlabel=sprintf("From %s to %s", min(DF$UTS), max(DF$UTS)))
```



With no squares

```
EpiCurve(DF,
  date = "UTS",
  period = "hour",
  split = 1,
  cutvar = "Confirmed",
  colors = c("#339933", "#eebb00"),
  xlabel=sprintf("From %s to %s", min(DF$UTS), max(DF$UTS)),
  square = FALSE)
```



Plot aggregated data

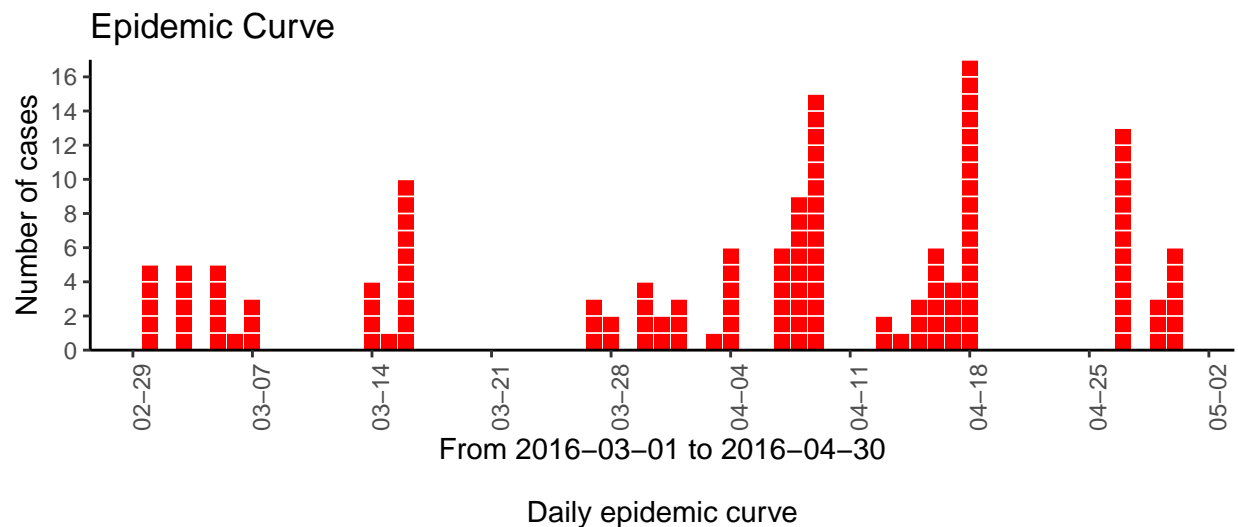
Daily

Without factors

| date | value |
|------------|-------|
| 2016-03-01 | 5 |
| 2016-03-03 | 5 |
| 2016-03-05 | 5 |
| 2016-03-06 | 1 |
| 2016-03-07 | 3 |
| 2016-03-14 | 4 |
| 2016-03-15 | 1 |
| 2016-03-16 | 10 |
| 2016-03-27 | 3 |
| 2016-03-28 | 2 |
| 2016-03-30 | 4 |
| 2016-03-31 | 2 |
| 2016-04-01 | 3 |
| 2016-04-03 | 1 |
| 2016-04-04 | 6 |
| 2016-04-07 | 6 |
| 2016-04-08 | 9 |
| 2016-04-09 | 15 |
| 2016-04-13 | 2 |
| 2016-04-14 | 1 |
| 2016-04-15 | 3 |
| 2016-04-16 | 6 |
| 2016-04-17 | 4 |
| 2016-04-18 | 17 |
| 2016-04-27 | 13 |
| 2016-04-29 | 3 |
| 2016-04-30 | 6 |

```
EpiCurve(DF,
  date = "date",
  freq = "value",
  period = "day",
  ylabel="Number of cases",
  xlabel=sprintf("From %s to %s", min(DF$date), max(DF$date)),
  title = "Epidemic Curve",
  note = "Daily epidemic curve")
```

```
## Day
## 'data.frame': 61 obs. of 4 variables:
## $ Freq: num 5 0 5 0 5 1 3 0 0 0 ...
## $ Cut : chr "1 cas" NA "1 cas" NA ...
## $ Day : chr "2016-03-01" "2016-03-02" "2016-03-03" "2016-03-04" ...
## $ Date: Date, format: "2016-03-01" "2016-03-02" ...
```

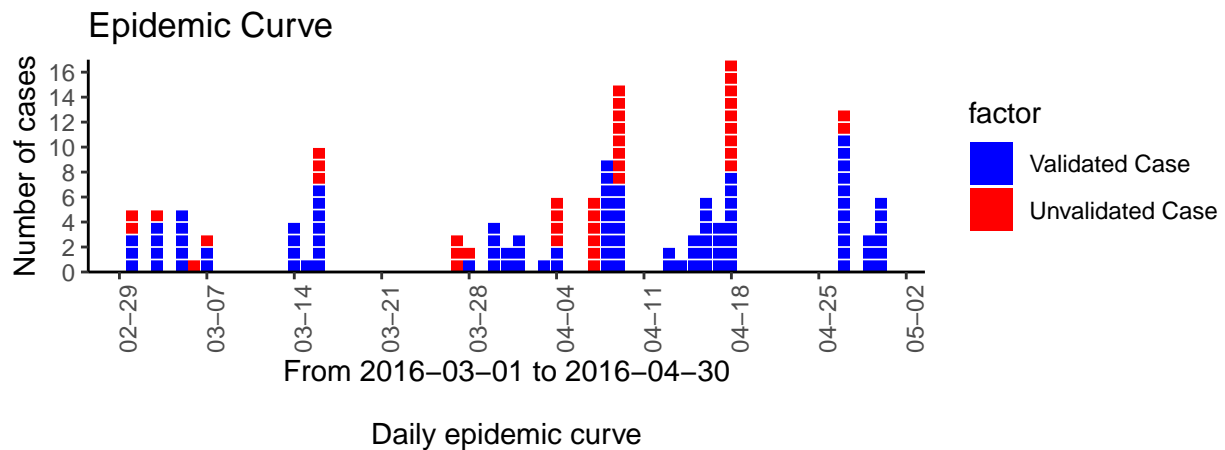


With factors

| date | value | factor |
|------------|-------|------------------|
| 2016-03-01 | 3 | Validated Case |
| 2016-03-01 | 2 | Unvalidated Case |
| 2016-03-03 | 4 | Validated Case |
| 2016-03-03 | 1 | Unvalidated Case |
| 2016-03-05 | 5 | Validated Case |
| 2016-03-06 | 1 | Unvalidated Case |
| 2016-03-07 | 2 | Validated Case |
| 2016-03-07 | 1 | Unvalidated Case |
| 2016-03-14 | 4 | Validated Case |
| 2016-03-15 | 1 | Validated Case |
| 2016-03-16 | 7 | Validated Case |
| 2016-03-16 | 3 | Unvalidated Case |
| 2016-03-27 | 3 | Unvalidated Case |
| 2016-03-28 | 1 | Validated Case |
| 2016-03-28 | 1 | Unvalidated Case |
| 2016-03-30 | 4 | Validated Case |
| 2016-03-31 | 2 | Validated Case |
| 2016-04-01 | 3 | Validated Case |
| 2016-04-03 | 1 | Validated Case |
| 2016-04-04 | 2 | Validated Case |
| 2016-04-04 | 4 | Unvalidated Case |
| 2016-04-07 | 6 | Unvalidated Case |
| 2016-04-08 | 9 | Validated Case |
| 2016-04-09 | 7 | Validated Case |
| 2016-04-09 | 8 | Unvalidated Case |
| 2016-04-13 | 2 | Validated Case |
| 2016-04-14 | 1 | Validated Case |
| 2016-04-15 | 3 | Validated Case |
| 2016-04-16 | 6 | Validated Case |
| 2016-04-17 | 4 | Validated Case |
| 2016-04-18 | 8 | Validated Case |
| 2016-04-18 | 9 | Unvalidated Case |
| 2016-04-27 | 11 | Validated Case |
| 2016-04-27 | 2 | Unvalidated Case |
| 2016-04-29 | 3 | Validated Case |
| 2016-04-30 | 6 | Validated Case |

```
EpiCurve(DF,
  date = "date",
  freq = "value",
  cutvar = "factor",
  period = "day",
  ylabel="Number of cases",
  xlabel=sprintf("From %s to %s", min(DF$date), max(DF$date)),
  title = "Epidemic Curve",
  note = "Daily epidemic curve")
```

```
## Day
## 'data.frame': 70 obs. of 4 variables:
## $ Freq: num 3 2 0 4 1 0 5 1 2 1 ...
## $ Cut : Factor w/ 2 levels "Unvalidated Case",...: 2 1 NA 2 1 NA 2 1 2 1 ...
## $ Day : chr "2016-03-01" "2016-03-01" "2016-03-02" "2016-03-03" ...
## $ Date: Date, format: "2016-03-01" "2016-03-01" ...
```



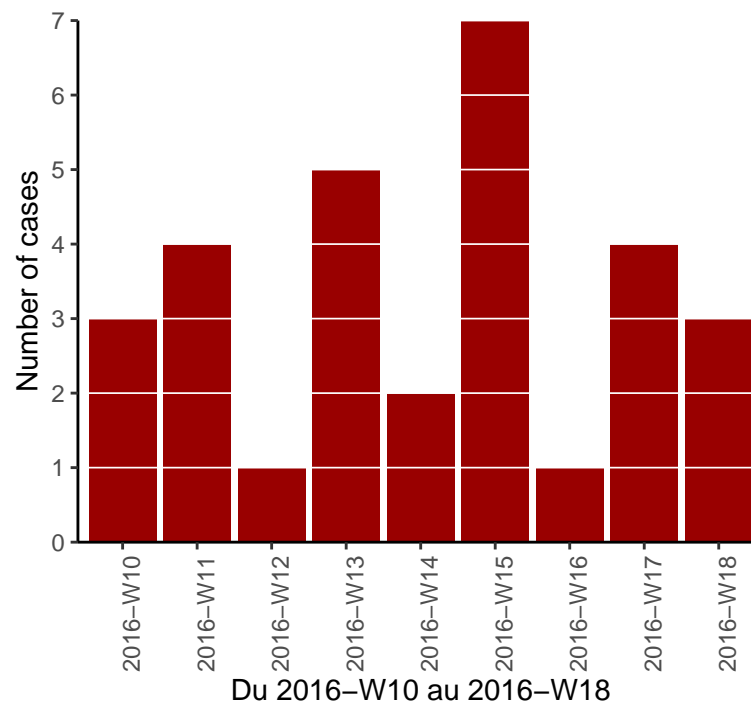
Weekly

Without factors

| date | value |
|----------|-------|
| 2016-W10 | 3 |
| 2016-W11 | 4 |
| 2016-W12 | 1 |
| 2016-W13 | 5 |
| 2016-W14 | 2 |
| 2016-W15 | 7 |
| 2016-W16 | 1 |
| 2016-W17 | 4 |
| 2016-W18 | 3 |

```
EpiCurve(DF,  
  date = "date",  
  freq = "value",  
  period = "week",  
  colors=c("#990000"),  
  ylabel="Number of cases",  
  xlabel=sprintf("Du %s au %s", min(DF$date), max(DF$date)),  
  title = "Epidemic Curve\n")
```

Epidemic Curve

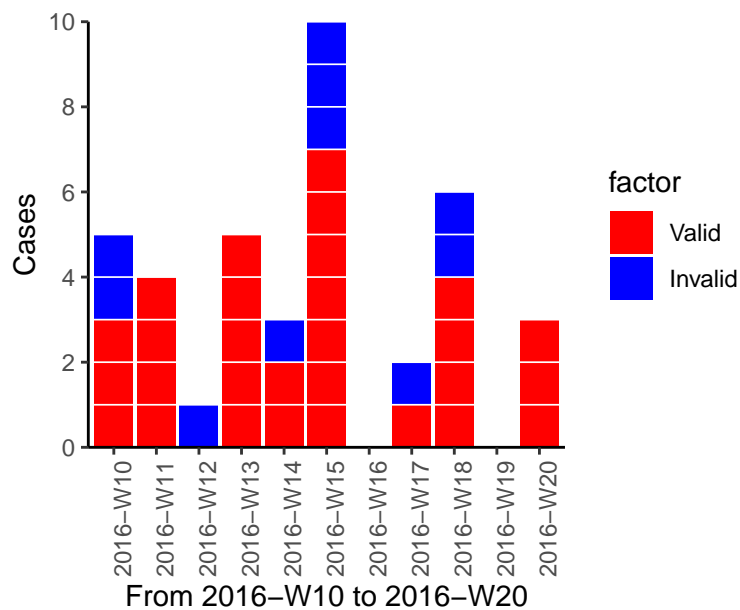


With factors

| date | value | factor |
|----------|-------|---------|
| 2016-W10 | 3 | Valid |
| 2016-W10 | 2 | Invalid |
| 2016-W11 | 4 | Valid |
| 2016-W12 | 1 | Invalid |
| 2016-W13 | 5 | Valid |
| 2016-W14 | 2 | Valid |
| 2016-W14 | 1 | Invalid |
| 2016-W15 | 7 | Valid |
| 2016-W15 | 3 | Invalid |
| 2016-W17 | 1 | Valid |
| 2016-W17 | 1 | Invalid |
| 2016-W18 | 4 | Valid |
| 2016-W18 | 2 | Invalid |
| 2016-W20 | 3 | Valid |

```
EpiCurve(DF,  
  date = "date",  
  freq = "value",  
  period = "week",  
  cutvar = "factor",  
  colors=c("Blue", "Red"),  
  ylabel="Cases",  
  xlabel=sprintf("From %s to %s", min(DF$date), max(DF$date)),  
  title = "Epidemic Curve\n")
```

Epidemic Curve



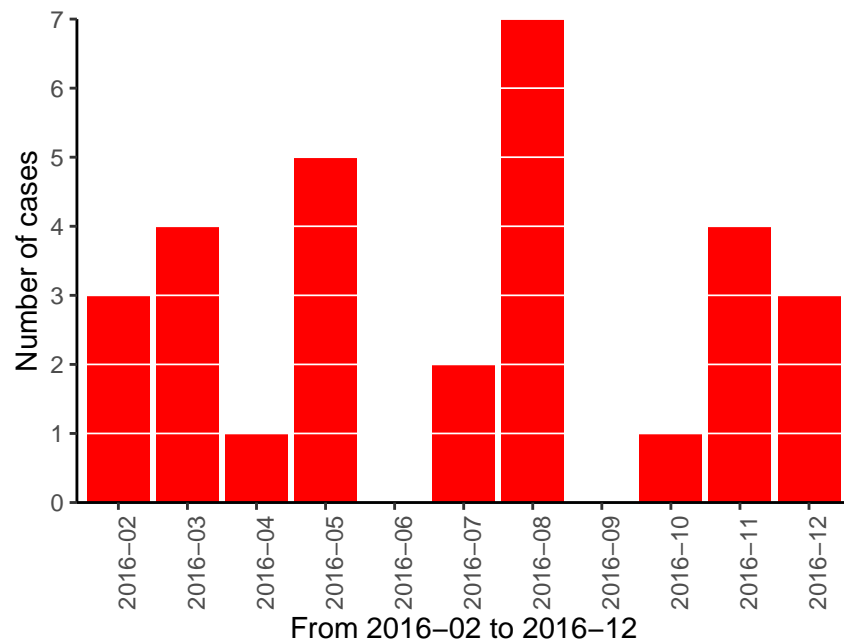
Monthly

Without factors

| date | value |
|---------|-------|
| 2016-02 | 3 |
| 2016-03 | 4 |
| 2016-04 | 1 |
| 2016-05 | 5 |
| 2016-07 | 2 |
| 2016-08 | 7 |
| 2016-10 | 1 |
| 2016-11 | 4 |
| 2016-12 | 3 |

```
EpiCurve(DF,  
  date = "date",  
  freq = "value",  
  period = "month",  
  ylabel="Number of cases",  
  xlabel=sprintf("From %s to %s", min(DF$date), max(DF$date)),  
  title = "Epidemic Curve\n")
```

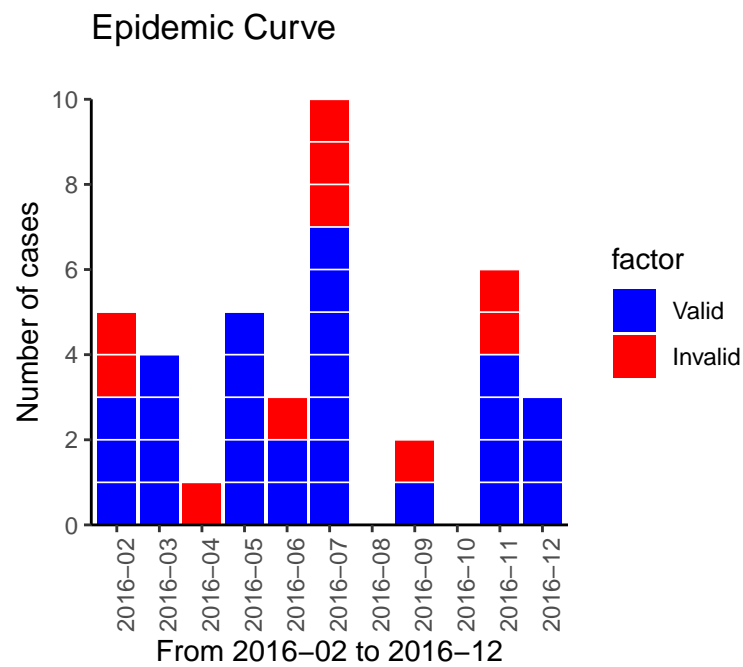
Epidemic Curve



With factors

| date | value | factor |
|---------|-------|---------|
| 2016-02 | 3 | Valid |
| 2016-02 | 2 | Invalid |
| 2016-03 | 4 | Valid |
| 2016-04 | 1 | Invalid |
| 2016-05 | 5 | Valid |
| 2016-06 | 2 | Valid |
| 2016-06 | 1 | Invalid |
| 2016-07 | 7 | Valid |
| 2016-07 | 3 | Invalid |
| 2016-09 | 1 | Valid |
| 2016-09 | 1 | Invalid |
| 2016-11 | 4 | Valid |
| 2016-11 | 2 | Invalid |
| 2016-12 | 3 | Valid |

```
EpiCurve(DF,  
  date = "date",  
  freq = "value",  
  cutvar = "factor",  
  period = "month",  
  ylabel="Number of cases",  
  xlabel=sprintf("From %s to %s", min(DF$date), max(DF$date)),  
  title = "Epidemic Curve\n")
```

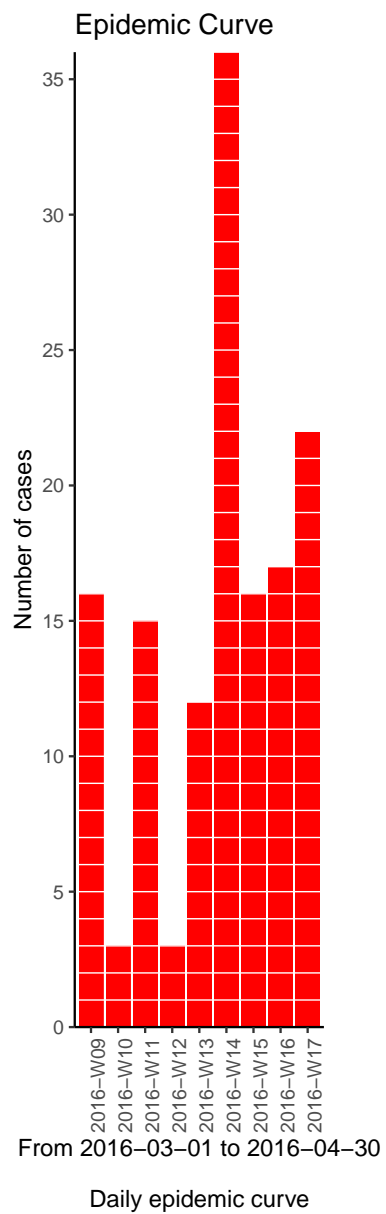


Converted period (aggragated cases)

“day” to “week”

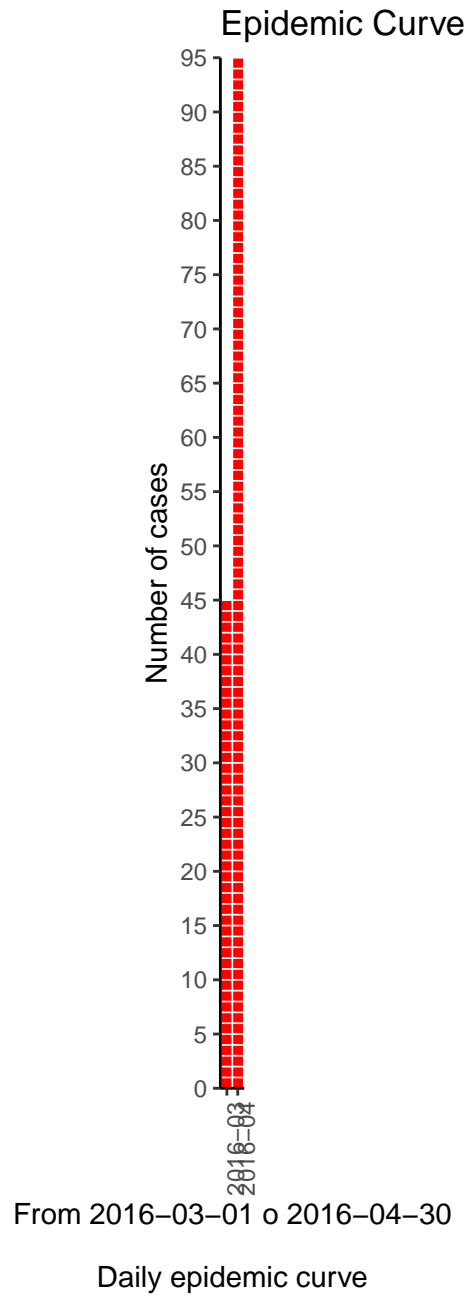
| date | value |
|------------|-------|
| 2016-03-01 | 5 |
| 2016-03-03 | 5 |
| 2016-03-05 | 5 |
| 2016-03-06 | 1 |
| 2016-03-07 | 3 |
| 2016-03-14 | 4 |
| 2016-03-15 | 1 |
| 2016-03-16 | 10 |
| 2016-03-27 | 3 |
| 2016-03-28 | 2 |
| 2016-03-30 | 4 |
| 2016-03-31 | 2 |
| 2016-04-01 | 3 |
| 2016-04-03 | 1 |
| 2016-04-04 | 6 |
| 2016-04-07 | 6 |
| 2016-04-08 | 9 |
| 2016-04-09 | 15 |
| 2016-04-13 | 2 |
| 2016-04-14 | 1 |
| 2016-04-15 | 3 |
| 2016-04-16 | 6 |
| 2016-04-17 | 4 |
| 2016-04-18 | 17 |
| 2016-04-27 | 13 |
| 2016-04-29 | 3 |
| 2016-04-30 | 6 |

```
EpiCurve(DF,  
  date = "date",  
  freq = "value",  
  period = "day",  
  to.period = "week",  
  ylabel="Number of cases",  
  xlabel=sprintf("From %s to %s", min(DF$date), max(DF$date)),  
  title = "Epidemic Curve",  
  note = "Daily epidemic curve")
```



“day” to “month”

```
EpiCurve(DF,  
  date = "date",  
  freq = "value",  
  period = "day",  
  to.period = "month",  
  ylabel="Number of cases",  
  xlabel=sprintf("From %s o %s", min(DF$date), max(DF$date)),  
  title = "Epidemic Curve",  
  note = "Daily epidemic curve")
```

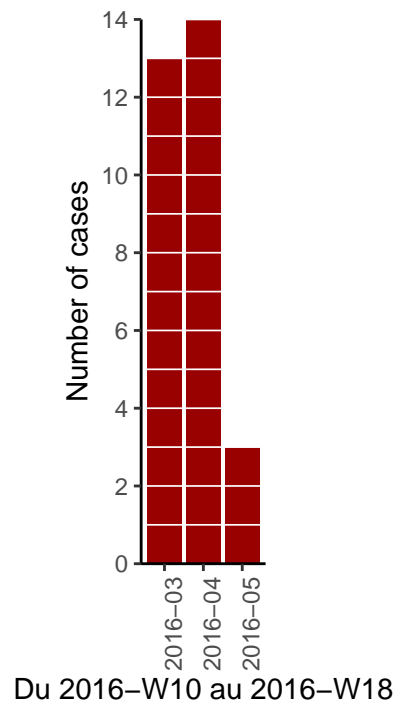


“week” to “month”

| date | value |
|----------|-------|
| 2016-W10 | 3 |
| 2016-W11 | 4 |
| 2016-W12 | 1 |
| 2016-W13 | 5 |
| 2016-W14 | 2 |
| 2016-W15 | 7 |
| 2016-W16 | 1 |
| 2016-W17 | 4 |
| 2016-W18 | 3 |

```
EpiCurve(DF,  
  date = "date",  
  freq = "value",  
  period = "week",  
  to.period = "month",  
  colors=c("#990000"),  
  ylabel="Number of cases",  
  xlabel=sprintf("Du %s au %s", min(DF$date), max(DF$date)),  
  title = "Epidemic Curve\n")
```

Epidemic Curve

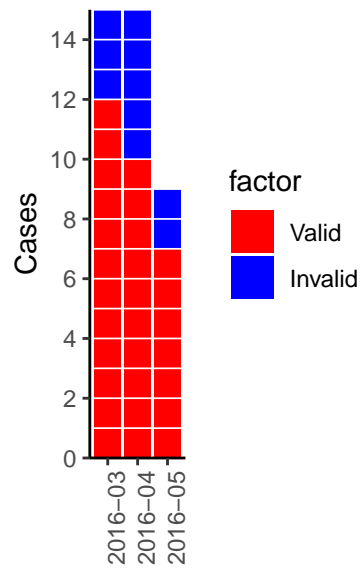


“week” to “month” with factors

| date | value | factor |
|----------|-------|---------|
| 2016-W10 | 3 | Valid |
| 2016-W10 | 2 | Invalid |
| 2016-W11 | 4 | Valid |
| 2016-W12 | 1 | Invalid |
| 2016-W13 | 5 | Valid |
| 2016-W14 | 2 | Valid |
| 2016-W14 | 1 | Invalid |
| 2016-W15 | 7 | Valid |
| 2016-W15 | 3 | Invalid |
| 2016-W17 | 1 | Valid |
| 2016-W17 | 1 | Invalid |
| 2016-W18 | 4 | Valid |
| 2016-W18 | 2 | Invalid |
| 2016-W20 | 3 | Valid |

```
EpiCurve(DF,  
  date = "date",  
  freq = "value",  
  period = "week",  
  to.period = "month",  
  cutvar = "factor",  
  colors=c("Blue", "Red"),  
  ylabel="Cases",  
  xlabel=sprintf("From %s to %s", min(DF$date), max(DF$date)),  
  title = "Epidemic Curve\n")
```

Epidemic Curve



From 2016-W10 to 2016-W20