Using EpiCurve

jp.decorps@epiconcept.fr

2017-08-24

Epiconcept is made up of a team of doctors, epidemiologists, data scientists and digital specialists. For more than 20 years, Epiconcept has contributed to the improvement of public health by writing software, carrying out epidemiological studies, research, evaluation and training to better detect, monitor and prevent disease and to improve treatment.

Epiconcept provides software, services and studies in the following areas:

- Software for managing public health programs,
- Secure cloud solutions for health data collection, reporting and processing,
- Research projects on vaccine preventable diseases, including measuring the effectiveness and impact of vaccines,
- Services in the field of epidemiology (protocols, analysis, training, etc.),
- Expertise in data analysis,
- Coaching and assistance to professionals in public health,
- Training (in software use and epidemiology: short and longer introductory modules, advanced courses, training through long-term practice).

To achieve such goals Epiconcept :

- Recognized research organization,
- Certified datacenter for hosting personal health data,
- Training organisation.

Epiconcept relies on :

- Its expertise in epidemiology
- Its IT expertise,
- Ethical values rooted in practice (responsibility and quality of services, data security and confidentiality, scientific independence, etc.),
- Capabilities to answer and anticipate tomorrow's challenges (Research - evaluation, e-health, Big Data, IoT, etc.),
- A desire to build long-term relationships with its clients and partners.

Its current customers and partners include some of the greatest names in the world such as: Santé Publique France (and many public health organizations around the world), WHO, ECDC, AFD, MSF, World Bank, etc.

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 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 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

)
```

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 colors
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

Depends

ggplot2, dplyr, ISOweek, scales, timeDate

Plot non-aggregated cases

Often epidemiologists work with non-aggregated data, so a dataset with one case on each line (case-based). A date is associated with each case. The date format for EpiCurve can be:

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

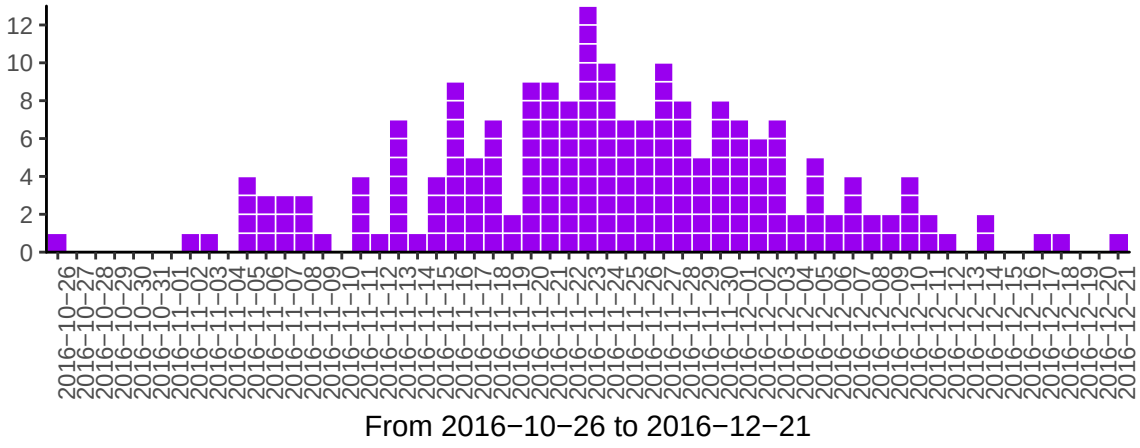
When the date format is hourly, the dataset is considered case-based, whether the "freq" parameter of the EpiCurve function is supplied or not. Whenever the parameter "freq" is not set, EpiCurve aggregates by date automatically.

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",
         color ="#9900ef",
         xlabel=sprintf("From %s to %s", min(DF$UTS), max(DF$UTS)))
```

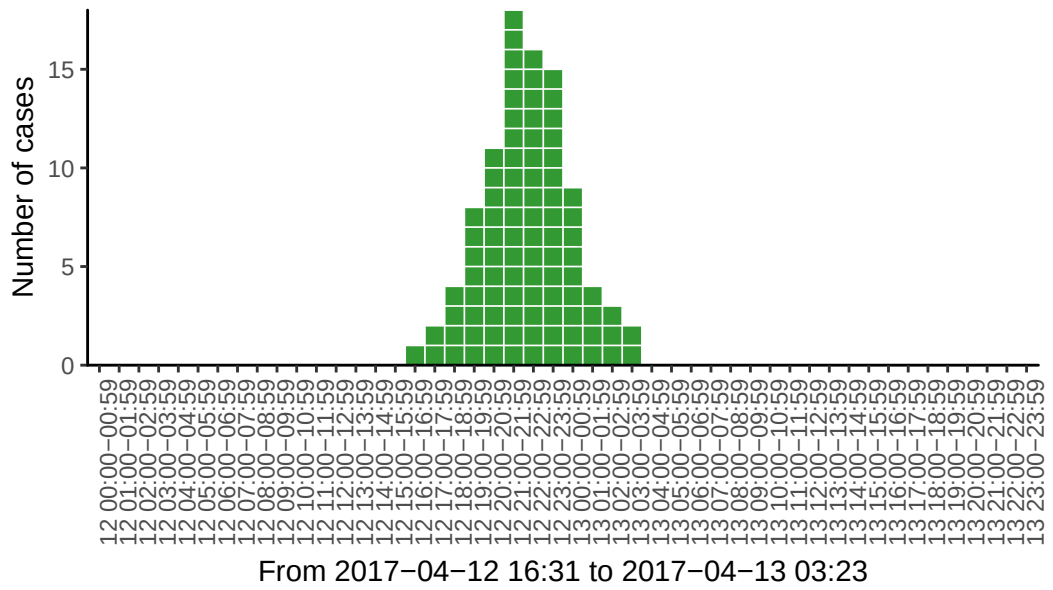


Hourly - non-aggregated cases

```
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,
         color ="#339933",
         ylabel="Number of cases",
         xlabel=sprintf("From %s to %s", min(DF$UTS), max(DF$UTS)))
```

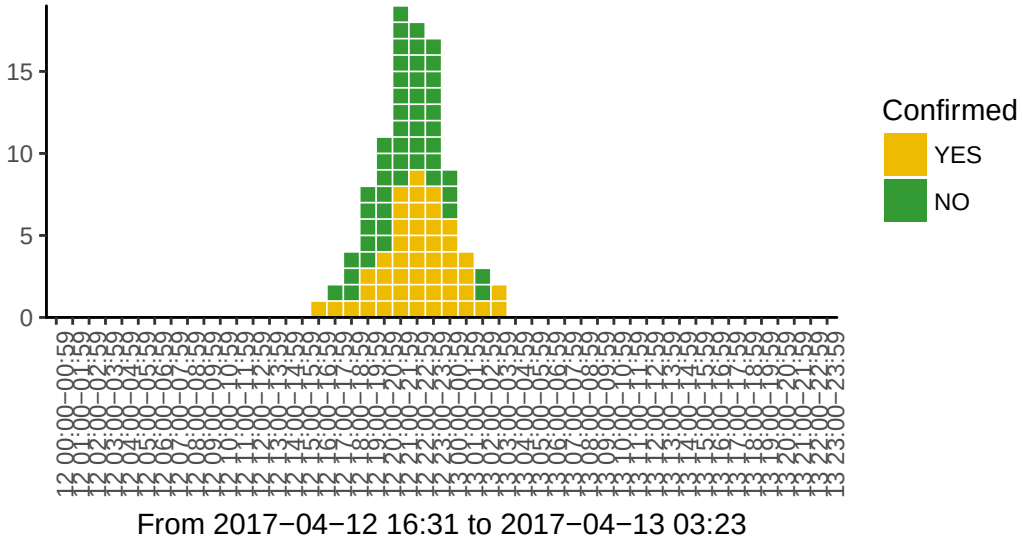


Hourly - non-aggregated cases with factors

```
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",
         color = c("#339933","#eebb00"),
         xlabel=sprintf("From %s to %s", min(DF$UTS), max(DF$UTS)))
```



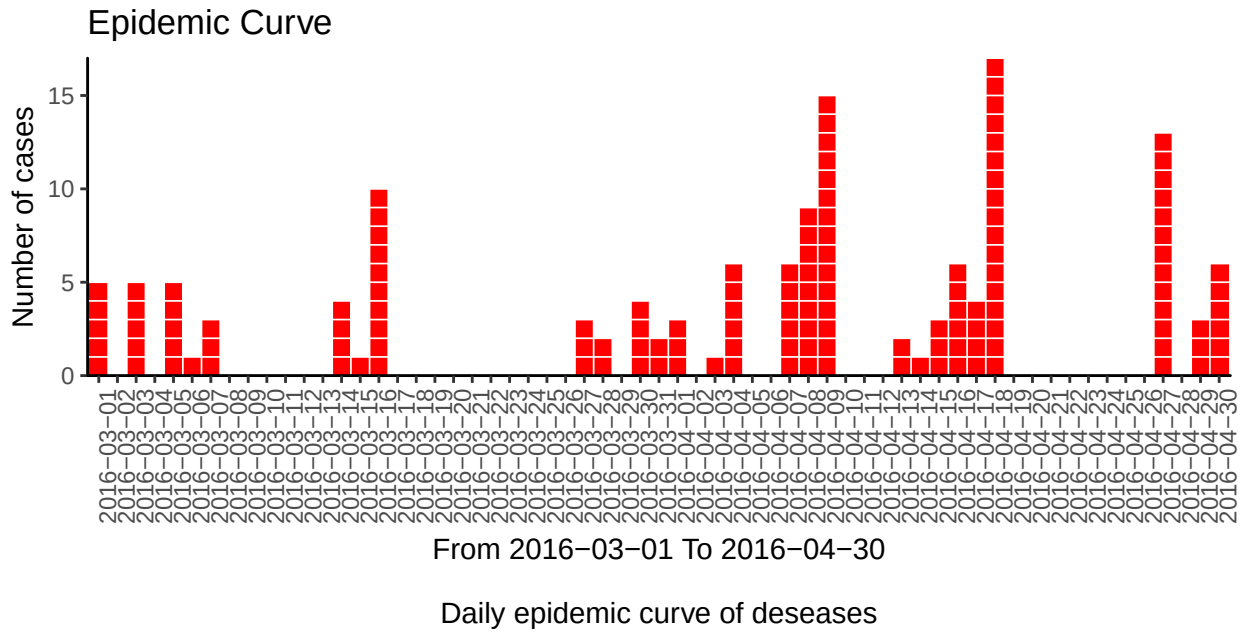
Plot aggregated data

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

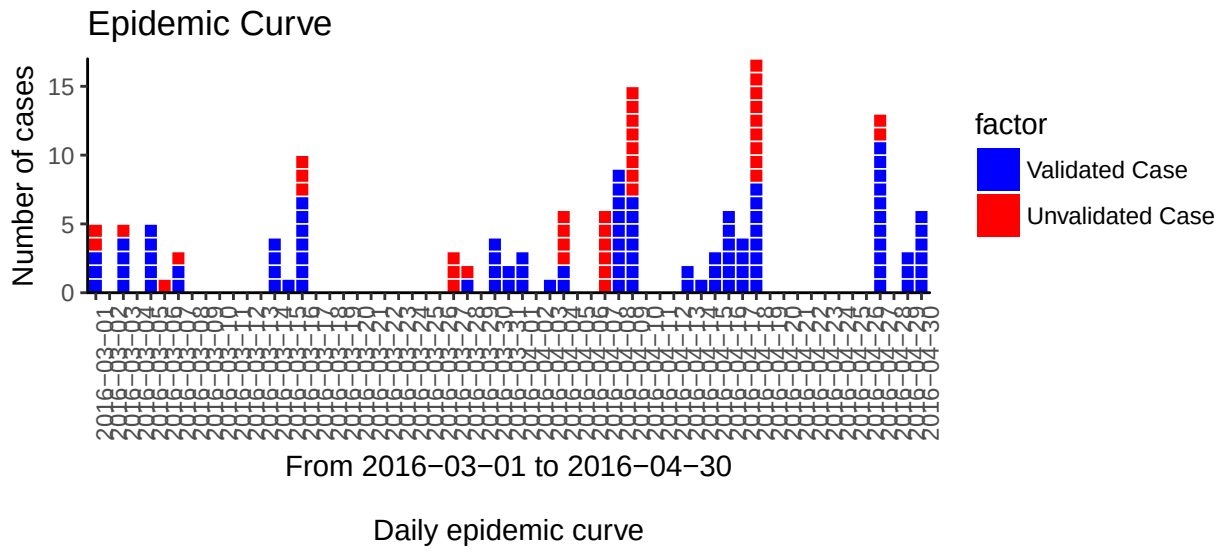
```
EpiCurve(DF,
         date = "date",
         freq = "value",
         period = "day",
         ylabel="Number of cases",
         xlabel=sprintf("From %s To %s", min(DF$date), max(DF$date)),
         title = "Epidemic Curve",
         note = "Daily epidemic curve of deseases")
```



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")
```

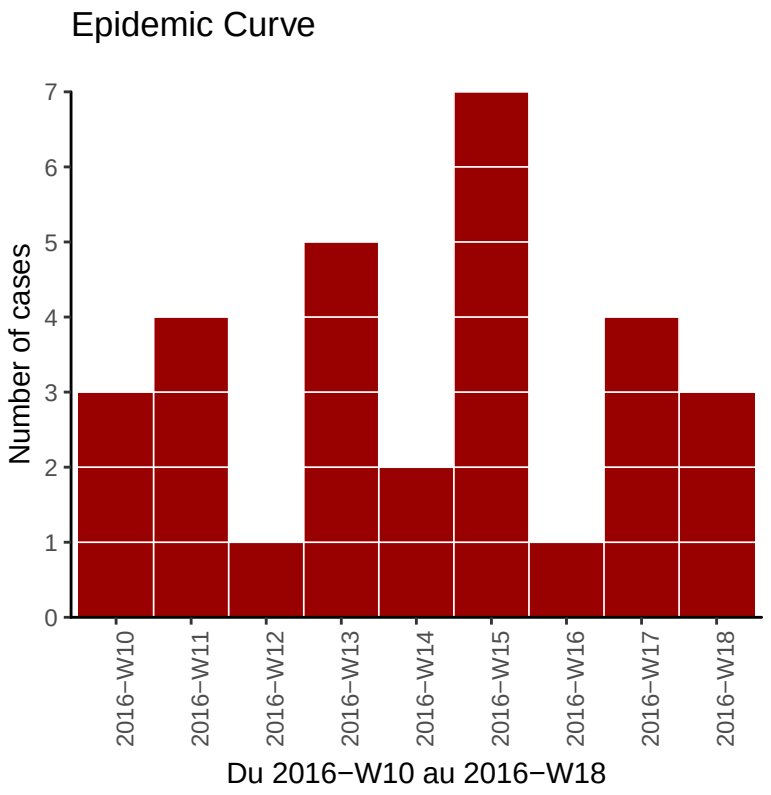


Weekly

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

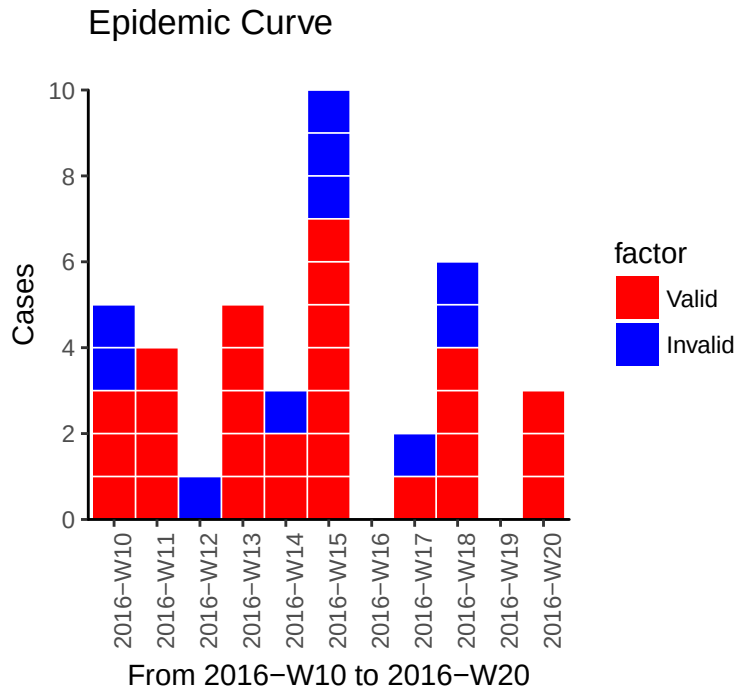
```
EpiCurve(DF,
         date = "date",
         freq = "value",
         period = "week",
         color=c("#990000"),
         ylabel="Number of cases",
         xlabel=sprintf("Du %s au %s", min(DF$date), max(DF$date)),
         title = "Epidemic Curve\n")
```



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",
         color=c("Blue", "Red"),
         ylabel="Cases",
         xlabel=sprintf("From %s to %s", min(DF$date), max(DF$date)),
         title = "Epidemic Curve\n")
```

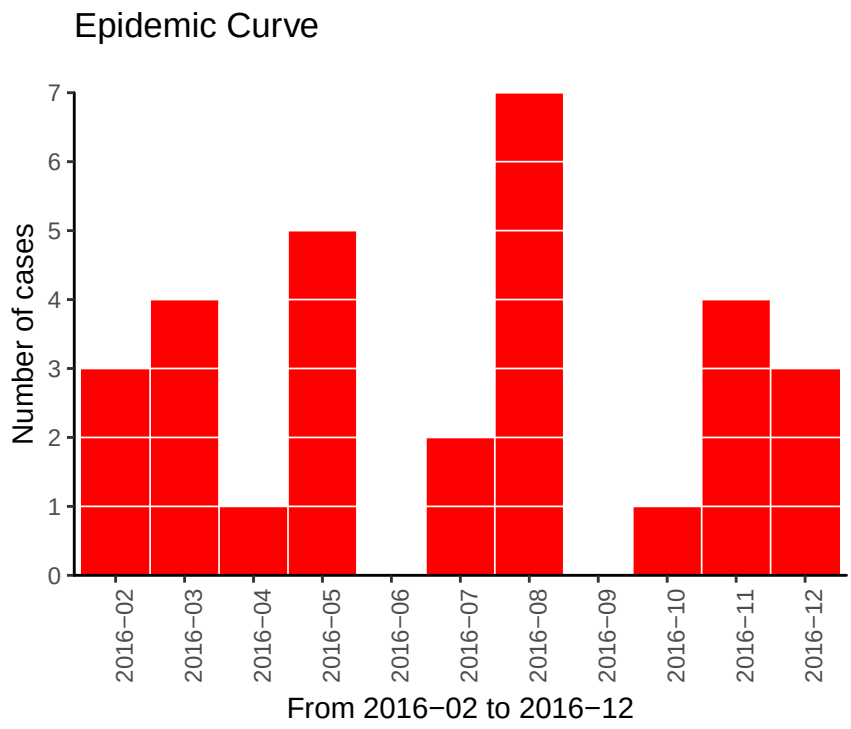


Monthly

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

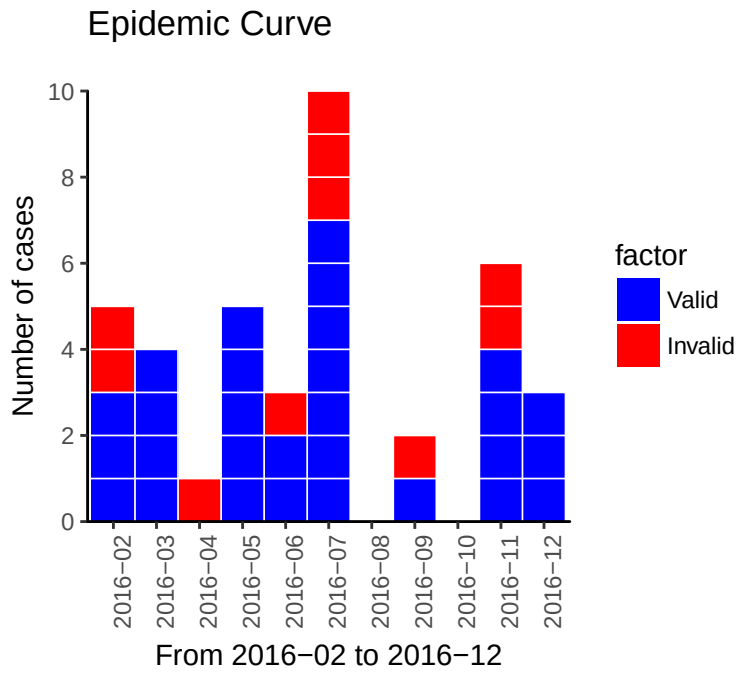
```
EpiCurve(DF,
         date = "date",
         freq = "value",
         period = "month",
         ylabel="Number of cases",
         xlabel=sprintf("From %s to %s", min(DF$date), max(DF$date)),
         title = "Epidemic Curve\n")
```



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")
```

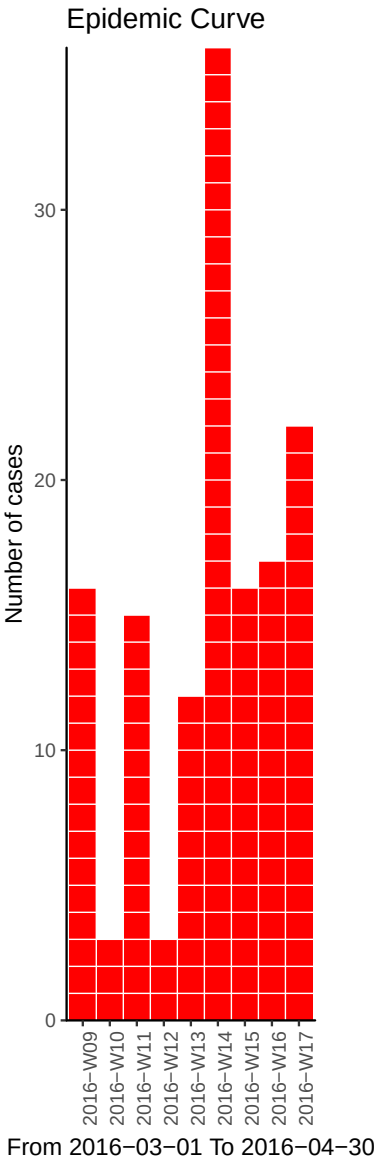


Converted period (aggragated cases)

"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

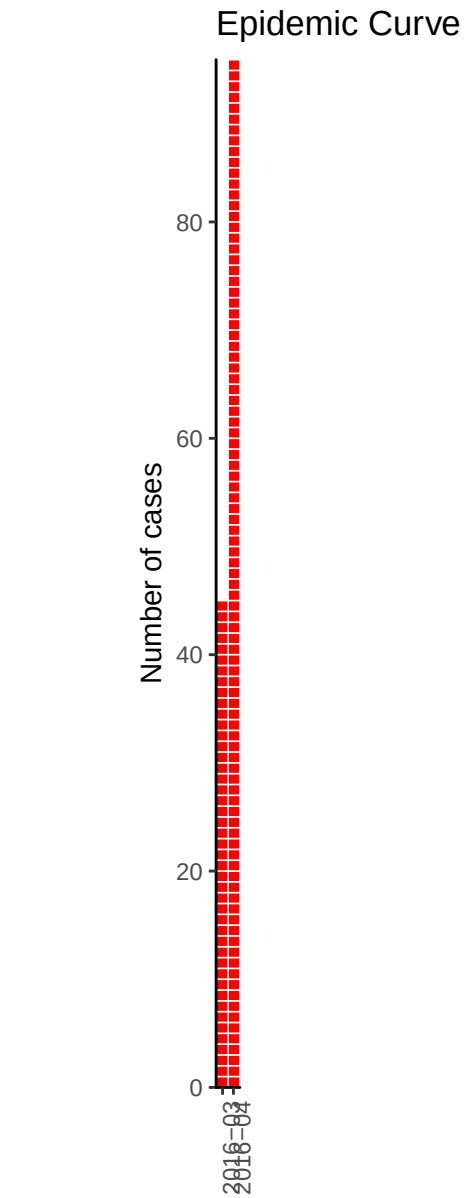
```
EpiCurve(DF,
         date = "date",
         freq = "value",
         period = "day",
         to.period = "week",
         ylabel="Number of cases",
         xlabel=sprintf("From %s To %s", min(DF$date), max(DF$date)),
         title = "Epidemic Curve",
         note = "Daily epidemic curve of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



Daily epidemic curve of deseases

“day” to “month”

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



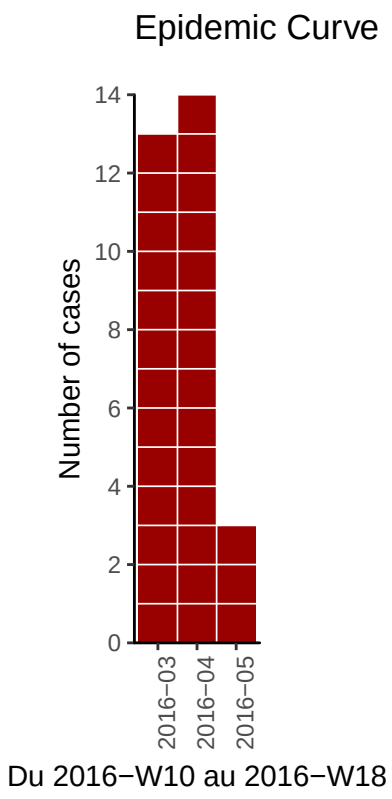
From 2016−03−01 To 2016−04−30

Daily epidemic curve of deseases

“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

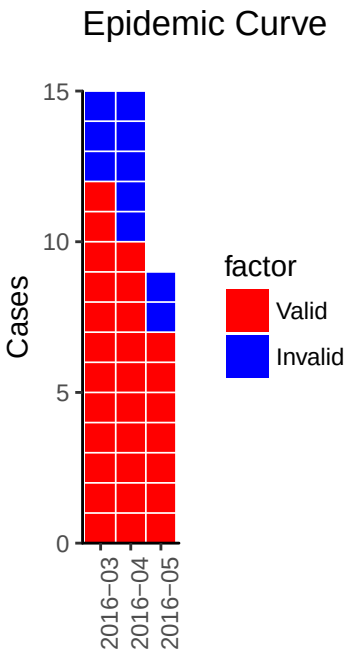
```
EpiCurve(DF,
         date = "date",
         freq = "value",
         period = "week",
         to.period = "month",
         color=c("#990000"),
         ylabel="Number of cases",
         xlabel=sprintf("Du %s au %s", min(DF$date), max(DF$date)),
         title = "Epidemic Curve\n")
```



"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",
         color=c("Blue", "Red"),
         ylabel="Cases",
         xlabel=sprintf("From %s to %s", min(DF$date), max(DF$date)),
         title = "Epidemic Curve\n")
```



From 2016−W10 to 2016−W20