

# EPI Info CDC

Nick Lauerman

3/16/2020

## Contents

<b>Libraries Used</b>	<b>1</b>
<b>Load CDC data</b>	<b>2</b>
Read CSV File . . . . .	2
Clean data and calculate cumulative number of cases . . . . .	2
<b>Data</b>	<b>2</b>
<b>Visualize all data</b>	<b>5</b>
<b>Filter to remove incomplete reporting</b>	<b>8</b>
Visualize . . . . .	8

## Libraries Used

```
library(tidyverse)
```

```
## -- Attaching packages -----
## v ggplot2 3.3.0    v purrr  0.3.3
## v tibble  2.1.3    v dplyr  0.8.5
## v tidyr   1.0.2    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.5.0

## -- Conflicts -----
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##     date
```

## Load CDC data

### Read CSV File

```
cdc <- read_csv(file = "./data/CDC_data.csv")

## Warning: Missing column names filled in: 'X3' [3]

## Parsed with column specification:
## cols(
##   Date = col_character(),
##   cases = col_double(),
##   X3 = col_logical(),
##   `16 mar data` = col_double(),
##   `17 mar data` = col_double(),
##   `18 mar data` = col_double(),
##   `19 mar data` = col_double(),
##   `20 mar data` = col_double(),
##   `23 mar data` = col_double(),
##   `24 mar data` = col_double(),
##   `25 mar data` = col_double()
## )
```

### Clean data and calculate cumulative number of cases

```
names(cdc)[1:2] <- c("Date",
                    "Number of new cases")
cdc$cum <- cumsum(cdc$`Number of new cases`)
cdc$Date <- as.Date(cdc$Date,
                  format = "%d-%b-%y")
```

## Data

```
cdc %>%
```

```
data.frame
```

```
##      Date Number.of.new.cases X3 X16.mar.data X17.mar.data X18.mar.data
## 1 2020-01-12          0 NA          0          0          0
## 2 2020-01-13          0 NA          0          0          0
## 3 2020-01-14          2 NA          2          2          2
## 4 2020-01-15          0 NA          0          0          0
## 5 2020-01-16          1 NA          1          1          1
## 6 2020-01-17          0 NA          0          0          0
## 7 2020-01-18          0 NA          0          0          0
## 8 2020-01-19          0 NA          0          0          0
## 9 2020-01-20          1 NA          1          1          1
## 10 2020-01-21         1 NA          1          1          1
## 11 2020-01-22         2 NA          1          1          1
## 12 2020-01-23          0 NA          0          0          0
## 13 2020-01-24          1 NA          1          1          1
## 14 2020-01-25          3 NA          3          3          3
## 15 2020-01-26          0 NA          0          0          0
## 16 2020-01-27          0 NA          0          0          0
```

## 17	2020-01-28	3 NA	2	2	2
## 18	2020-01-29	1 NA	1	1	1
## 19	2020-01-30	1 NA	1	1	1
## 20	2020-01-31	0 NA	0	1	1
## 21	2020-02-01	1 NA	1	1	1
## 22	2020-02-02	4 NA	1	1	1
## 23	2020-02-03	0 NA	0	0	0
## 24	2020-02-04	0 NA	0	0	0
## 25	2020-02-05	0 NA	0	0	0
## 26	2020-02-06	2 NA	1	1	1
## 27	2020-02-07	1 NA	0	0	0
## 28	2020-02-08	1 NA	1	1	1
## 29	2020-02-09	0 NA	0	0	0
## 30	2020-02-10	3 NA	3	4	4
## 31	2020-02-11	6 NA	1	3	3
## 32	2020-02-12	0 NA	1	1	1
## 33	2020-02-13	4 NA	3	3	3
## 34	2020-02-14	3 NA	2	2	2
## 35	2020-02-15	7 NA	0	1	1
## 36	2020-02-16	3 NA	0	1	1
## 37	2020-02-17	13 NA	5	5	5
## 38	2020-02-18	14 NA	10	9	9
## 39	2020-02-19	8 NA	6	6	6
## 40	2020-02-20	12 NA	7	8	8
## 41	2020-02-21	23 NA	11	14	14
## 42	2020-02-22	23 NA	14	17	17
## 43	2020-02-23	23 NA	13	14	14
## 44	2020-02-24	50 NA	34	38	38
## 45	2020-02-25	39 NA	16	23	23
## 46	2020-02-26	73 NA	30	40	40
## 47	2020-02-27	69 NA	19	33	33
## 48	2020-02-28	99 NA	40	50	50
## 49	2020-02-29	91 NA	40	47	47
## 50	2020-03-01	166 NA	53	66	66
## 51	2020-03-02	180 NA	51	68	68
## 52	2020-03-03	214 NA	48	71	71
## 53	2020-03-04	237 NA	58	71	71
## 54	2020-03-05	257 NA	74	57	57
## 55	2020-03-06	333 NA	43	57	57
## 56	2020-03-07	352 NA	79	54	54
## 57	2020-03-08	498 NA	46	51	51
## 58	2020-03-09	696 NA	50	70	70
## 59	2020-03-10	662 NA	15	36	36
## 60	2020-03-11	792 NA	2	25	25
## 61	2020-03-12	1055 NA	0	4	4
## 62	2020-03-13	1113 NA	NA	2	2
## 63	2020-03-14	1058 NA	NA	0	0
## 64	2020-03-15	1001 NA	NA	0	0
## 65	2020-03-16	926 NA	NA	NA	NA
## 66	2020-03-17	467 NA	NA	NA	NA
## 67	2020-03-18	302 NA	NA	NA	NA
## 68	2020-03-19	187 NA	NA	NA	NA
## 69	2020-03-20	55 NA	NA	NA	NA
## 70	2020-03-21	13 NA	NA	NA	NA

## 71	2020-03-22		8	NA	NA	NA	NA
## 72	2020-03-23		5	NA	NA	NA	NA
## 73	2020-03-24		0	NA	NA	NA	NA
## 74	2020-03-25		0	NA	NA	NA	NA
##	X19.mar.data	X20.mar.data	X23.mar.data	X24.mar.data	X25.mar.data	cum	
## 1	0	0	0	0	0	0	0
## 2	0	0	0	0	0	0	0
## 3	2	2	2	2	2	2	2
## 4	0	0	0	0	0	0	2
## 5	1	1	1	1	1	1	3
## 6	0	0	0	0	0	0	3
## 7	0	0	0	0	0	0	3
## 8	0	0	0	0	0	0	3
## 9	1	1	1	1	1	1	4
## 10	1	1	1	1	1	1	5
## 11	1	1	1	1	1	2	7
## 12	0	0	0	0	0	0	7
## 13	1	1	1	1	1	1	8
## 14	3	3	3	3	3	3	11
## 15	0	0	0	0	0	0	11
## 16	0	0	0	0	0	0	11
## 17	3	3	3	3	3	3	14
## 18	1	1	1	1	1	1	15
## 19	1	1	1	1	1	1	16
## 20	0	0	0	0	0	0	16
## 21	1	1	1	1	1	1	17
## 22	1	1	2	3	4	4	21
## 23	0	0	0	0	0	0	21
## 24	0	0	0	0	0	0	21
## 25	0	0	0	0	0	0	21
## 26	1	2	2	2	2	2	23
## 27	0	0	0	0	1	1	24
## 28	1	1	1	1	1	1	25
## 29	0	0	0	0	0	0	25
## 30	4	4	3	3	3	3	28
## 31	5	5	6	6	6	6	34
## 32	0	0	0	0	0	0	34
## 33	4	4	4	4	4	4	38
## 34	1	3	3	3	3	3	41
## 35	5	5	6	6	7	7	48
## 36	3	3	3	3	3	3	51
## 37	9	9	12	12	13	13	64
## 38	10	10	14	14	14	14	78
## 39	6	6	7	7	8	8	86
## 40	10	10	12	12	12	12	98
## 41	18	20	23	23	23	23	121
## 42	22	22	22	22	23	23	144
## 43	18	18	20	21	23	23	167
## 44	42	43	50	50	50	50	217
## 45	33	34	35	37	39	39	256
## 46	52	52	61	64	73	73	329
## 47	43	47	55	58	68	68	398
## 48	66	66	82	87	98	98	497
## 49	55	57	64	70	91	91	588

## 50	97	103	113	125	165	754
## 51	91	95	120	147	180	934
## 52	108	121	146	179	213	1148
## 53	117	125	159	183	235	1385
## 54	111	119	159	204	257	1642
## 55	124	139	193	249	331	1975
## 56	104	120	176	280	349	2327
## 57	121	140	214	379	496	2825
## 58	163	194	307	559	692	3521
## 59	140	172	302	523	659	4183
## 60	123	174	331	628	791	4975
## 61	75	122	275	741	1048	6030
## 62	54	95	269	802	1105	7143
## 63	23	55	238	643	1051	8201
## 64	13	35	203	455	991	9202
## 65	2	12	173	369	916	10128
## 66	0	8	90	181	455	10595
## 67	0	0	44	82	281	10897
## 68	NA	0	21	41	180	11084
## 69	NA	NA	2	12	52	11139
## 70	NA	NA	NA	2	13	11152
## 71	NA	NA	NA	1	5	11160
## 72	NA	NA	NA	0	4	11165
## 73	NA	NA	NA	NA	0	11165
## 74	NA	NA	NA	NA	NA	11165

## Visualize all data

```
epi <- ggplot(data = cdc)
```

```
epi + geom_point(aes(x = Date,
  y = `Number of new cases`))+
#   geom_line(aes(x = Date,
#     y = `Number of new cases`),
#     linetype = 2) +
  geom_smooth(aes(x = Date,
    y = `Number of new cases`),
    color = "red",
    fill = "blue") +
  labs(y = "Cases",
    title = "Number of New Cases of COVID-19 Reported to the CDC")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
epi + geom_col(aes(x=Date,
  y=`Number of new cases`),
  fill= "blue") +
  geom_smooth(aes(x=Date,
    y=`Number of new cases`),
    color = "red") +
  labs(y = "Cases",
    title = "Number of New Cases of COVID-19 Reported to the CDC")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

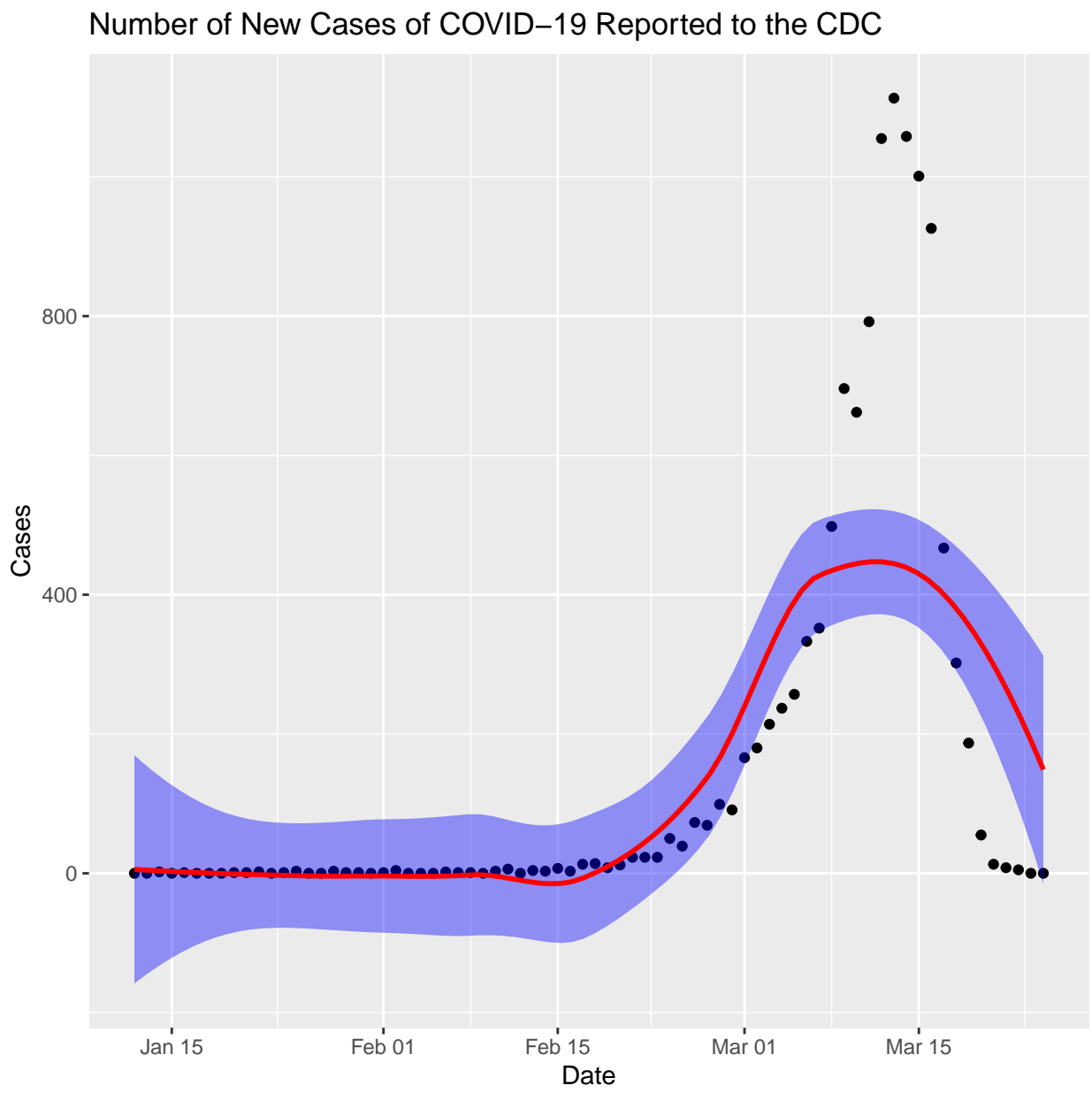


Figure 1: Epi curve 1

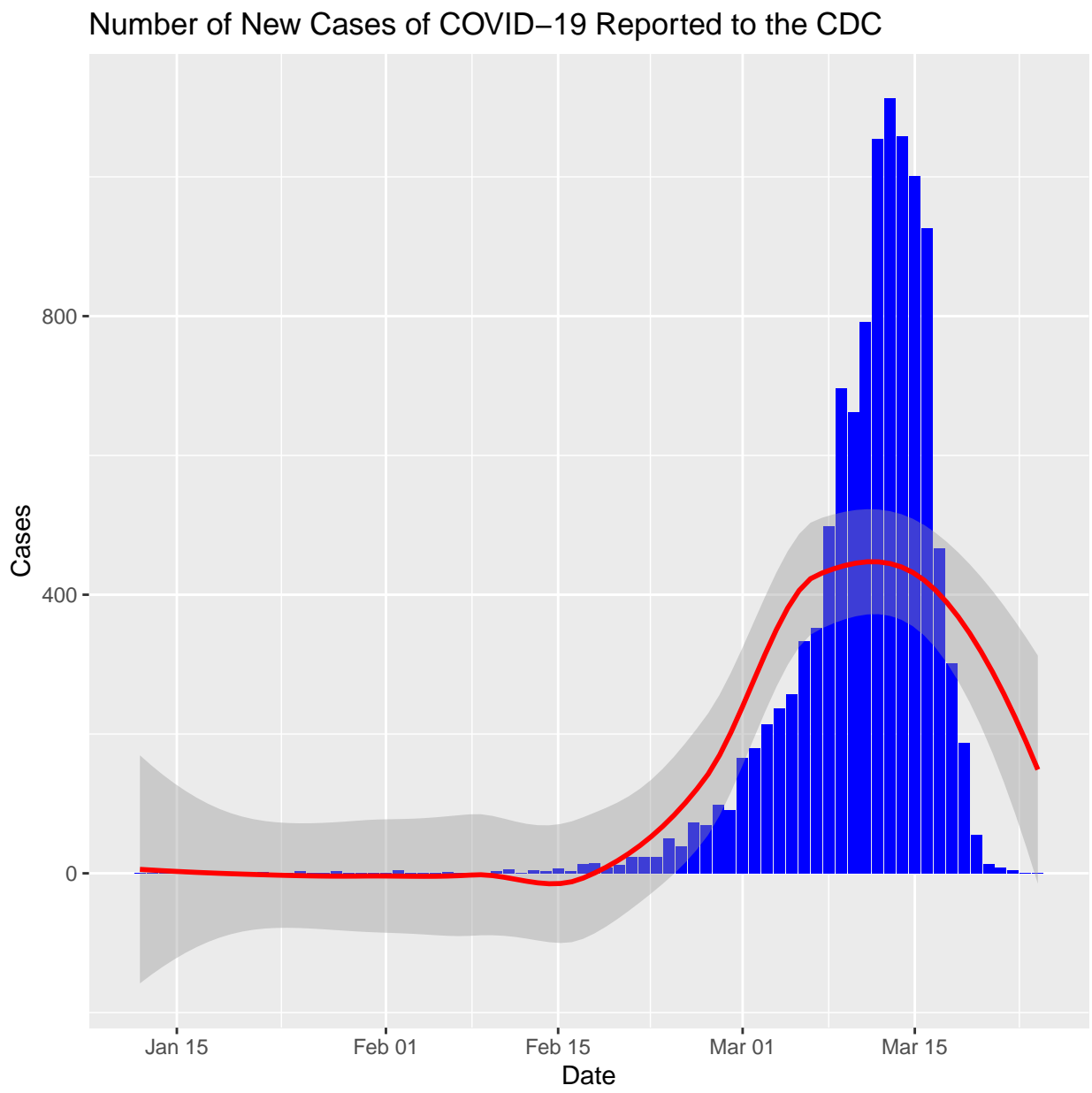


Figure 2: Epi curve 2, traditional

```

epi + geom_point(aes(x = Date,
  y = cum))+
  # geom_line(aes(x = Date,
  #y = `cum`)) +
  geom_smooth(aes(x = Date,
    y = `cum`,
    color = "red",
    fill = "blue") +
  labs(y = "Cumulative number of cases",
    title = "Cumulative Number of Cases of COVID-19 Reported to the CDC") +
  geom_hline(yintercept = mean(cdc$cum)) +
  geom_hline(yintercept = median(cdc$cum),
    lty = 2)

```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

## Filter to remove incomplete reporting

remove dates on or after 15 March as this data may not be completely reported

```

cdc <- cdc %>%
  filter(Date < as.Date("2020-03-15"))

```

## Visualize

```
epi <- ggplot(data = cdc)
```

```

epi + geom_point(aes(x = Date,
  y = `Number of new cases`))+
  # geom_line(aes(x = Date,
  # y = `Number of new cases`,
  # linetype = 2) +
  geom_smooth(aes(x = Date,
    y = `Number of new cases`,
    color = "red",
    fill = "blue") +
  labs(y = "Cases",
    title = "Number of New Cases of COVID-19 Reported to the CDC")

```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```

epi + geom_col(aes(x=Date,
  y=`Number of new cases`,
  fill= "blue") +
  geom_smooth(aes(x=Date,
    y=`Number of new cases`,
    color = "red") +
  labs(y = "Cases",
    title = "Number of New Cases of COVID-19 Reported to the CDC")

```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```

epi + geom_point(aes(x = Date,
  y = cum))+
  # geom_line(aes(x = Date,

```



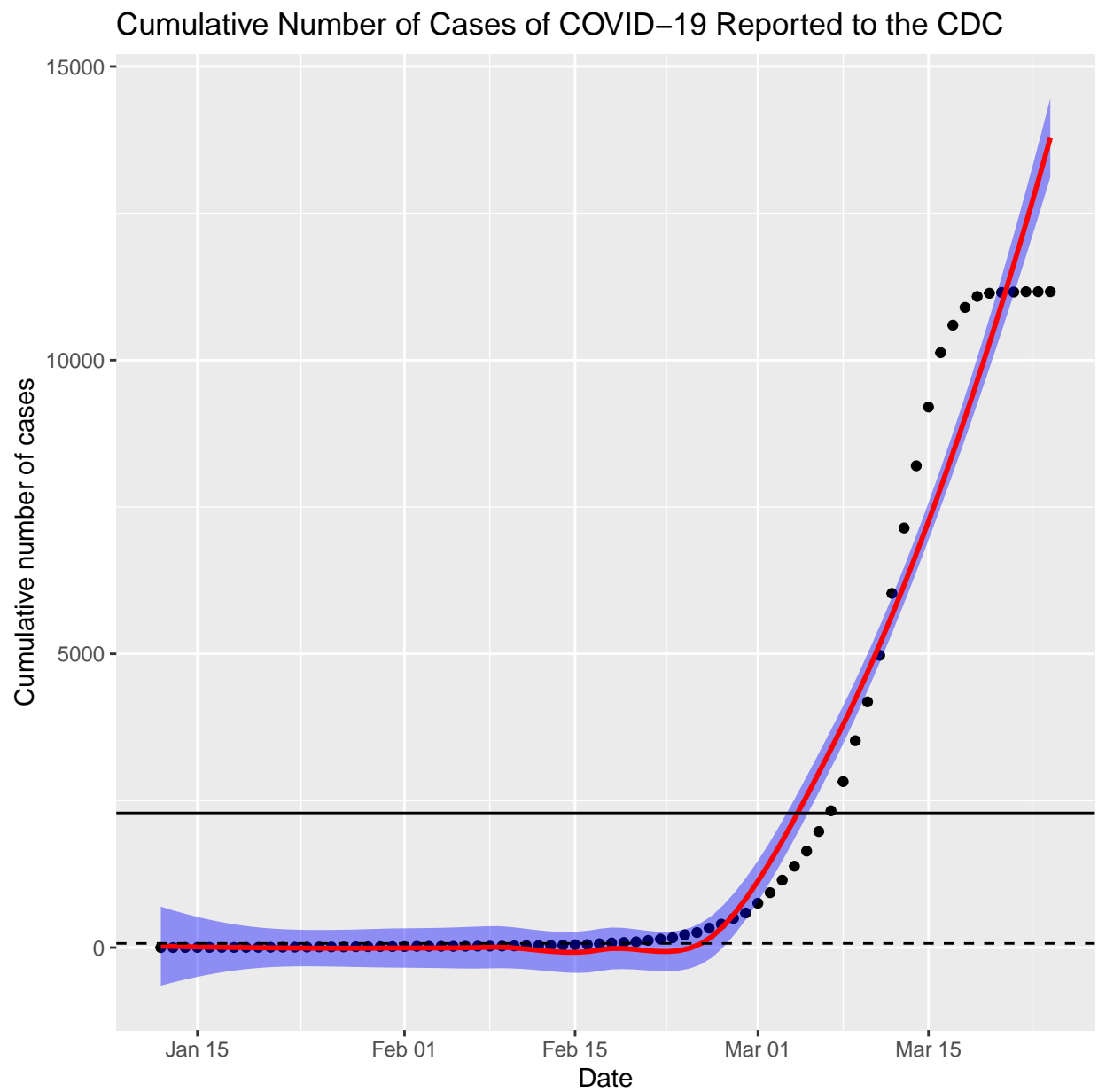


Figure 3: Cumulative cases

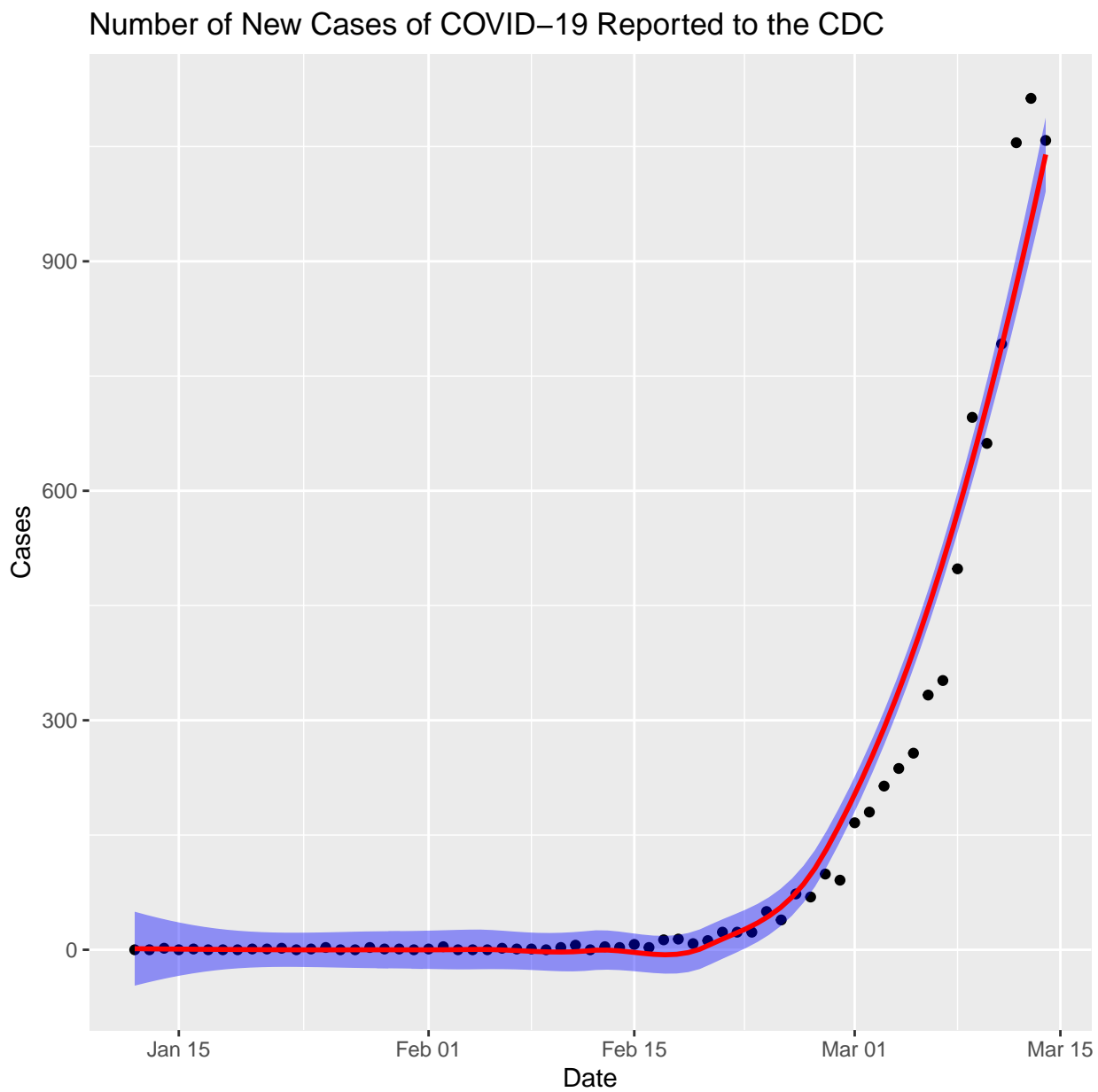


Figure 4: Epi curve 1

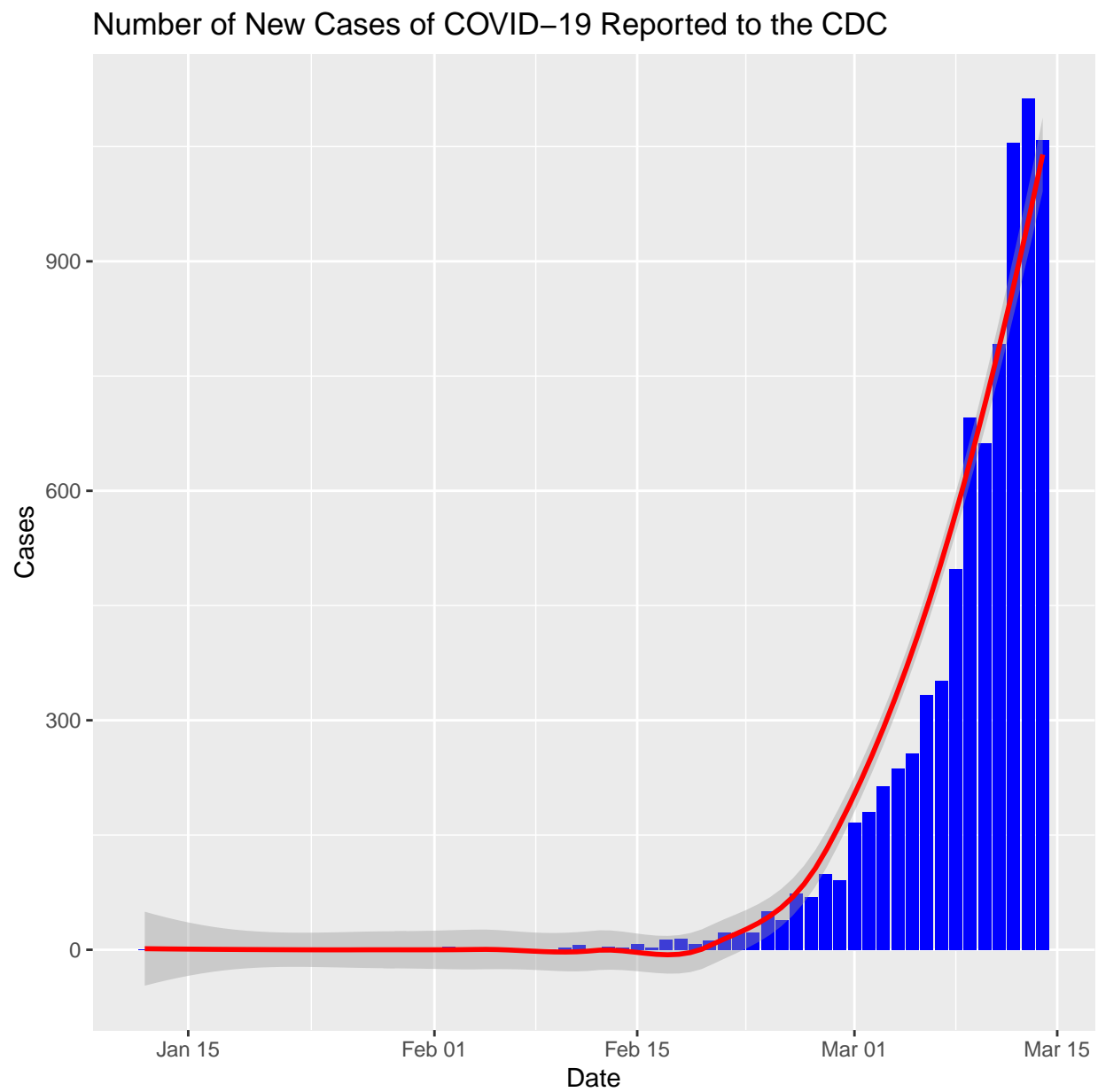


Figure 5: Epi curve 2, traditional

```

#y = `cum`)) +
geom_smooth(aes(x = Date,
                 y = `cum`,
                 color = "red",
                 fill = "blue") +
labs(y = "Cumulative number of cases",
      title = "Cumulative Number of Cases of COVID-19 Reported to the CDC") +
geom_hline(yintercept = mean(cdc$cum)) +
geom_hline(yintercept = median(cdc$cum),
           lty = 2)

```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

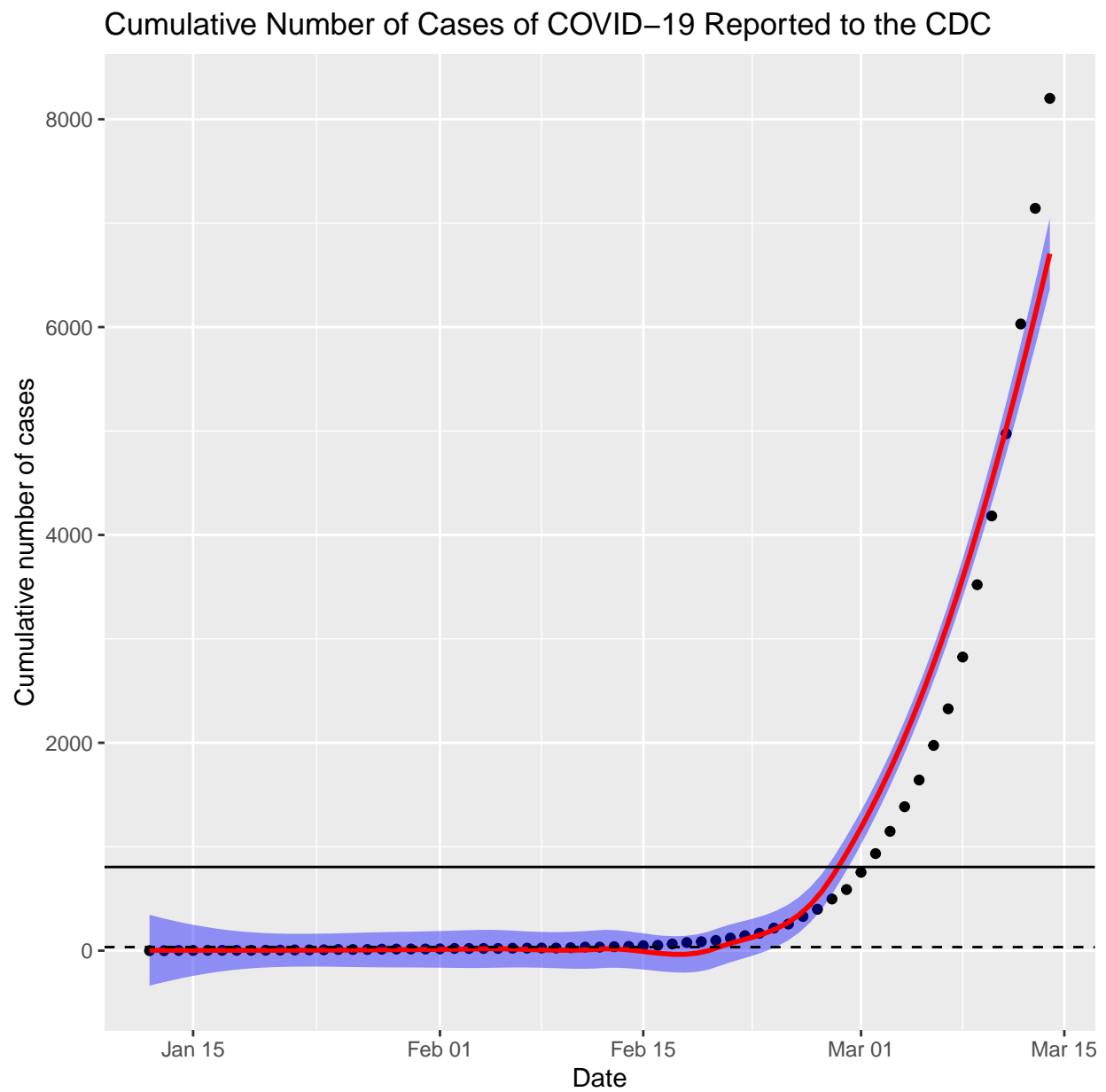


Figure 6: Cumulative cases