

# Donation behaviour after press events

## Setup

```
library(readr)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(tidyr)
library(ggplot2)
library(lubridate)

##
## Attaching package: 'lubridate'

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

library(stringr)
```

## Reading the data

Contains cleaning...

```
donations = read_csv('Donations.csv')

## Parsed with column specification:
## cols(
##   Received = col_date(format = ""),
##   `Contribution (in euro)` = col_double(),
##   frequency = col_character(),
##   resource = col_character(),
##   personal_ID = col_integer(),
##   company = col_integer()
## )

max_donation_date = max(donations$Received)
min_donation_date = min(donations$Received)
donations

## # A tibble: 32,224 x 6
##   Received `Contribution (in euro)` frequency resource
##   <date>          <dbl>      <chr>      <chr>
## 1 2015-01-02      75.0      once      Geldspende
```

```

## 2 2015-01-02      10.0      once Mitgliedsbeitrag
## 3 2015-01-02      22.5      once      Geldspende
## 4 2015-01-02      50.0      once      Geldspende
## 5 2015-01-02       3.0      once Mitgliedsbeitrag
## 6 2015-01-02       5.0      once Mitgliedsbeitrag
## 7 2015-01-02      50.0      once      Geldspende
## 8 2015-01-02      50.0      once      Dauerspende
## 9 2015-01-02      30.0      once      Dauerspende
## 10 2015-01-02     40.0      once      Geldspende
## # ... with 32,214 more rows, and 2 more variables: personal_ID <int>,
## #   company <int>

m = c('Januar', 'Februar', 'März', 'April', 'Mai', 'Juni', 'Juli', 'August', 'September', 'Oktober', 'November', 'Dezember')
media = read_delim('Media.csv', delim=';') %>%
  mutate(monthN=match(Month, m), date=ISOdate(Year, monthN, Day)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('abendblatt', ignore_case = T)), 'Abendblatt', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('Acht Tonnen', ignore_case = T)), 'Acht Tonnen', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('eins mehr', ignore_case = T)), 'Eins mehr!', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('3sat', ignore_case = T)), '3sat', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('arcor', ignore_case = T)), 'arcor', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('Allgemeine Hotel', ignore_case = T)), 'Allgemeine Hotel', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('b4b', ignore_case = T)), 'b4b', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('berlin.de', ignore_case = T)), 'berlin.de', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('Berliner Morgenpost', ignore_case = T)), 'Berliner Morgenpost', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('Tagesspiegel', ignore_case = T)), 'Tagesspiegel', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('Morgenpost', ignore_case = T)), 'Morgenpost', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('Berliner Kurier', ignore_case = T)), 'Berliner Kurier', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('Berliner Rundfunk', ignore_case = T)), 'Berliner Rundfunk', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('Berliner Woche', ignore_case = T)), 'Berliner Woche', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('Berlinerwoche', ignore_case = T)), 'Berliner Woche', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('Deutschlandfunk', ignore_case = T)), 'Deutschlandfunk', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('Berliner Zeitung', ignore_case = T)), 'Berliner Zeitung', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, regex('^BZ', ignore_case = T)), 'Berliner Zeitung', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, regex('^Bild', ignore_case = T)), 'Bild', Agency)) %>%
  mutate(Agency=ifelse(str_detect(Agency, coll('Meine Berliner Woche', ignore_case = T)), 'Meine Berliner Woche', Agency)) %>%

## Parsed with column specification:
## cols(
##   Year = col_integer(),
##   Month = col_character(),
##   Day = col_integer(),
##   Agency = col_character(),
##   Article = col_character(),
##   Type = col_character(),
##   MediaType = col_character(),
##   CFD = col_integer(),
##   CFV = col_integer()
## )

media

## # A tibble: 930 x 11
##   Year Month Day Agency
##   <int> <chr> <int> <chr>
## 1 2015 Januar 1 Berliner Woche
## 2 2015 Januar 1 Morgenpost

```

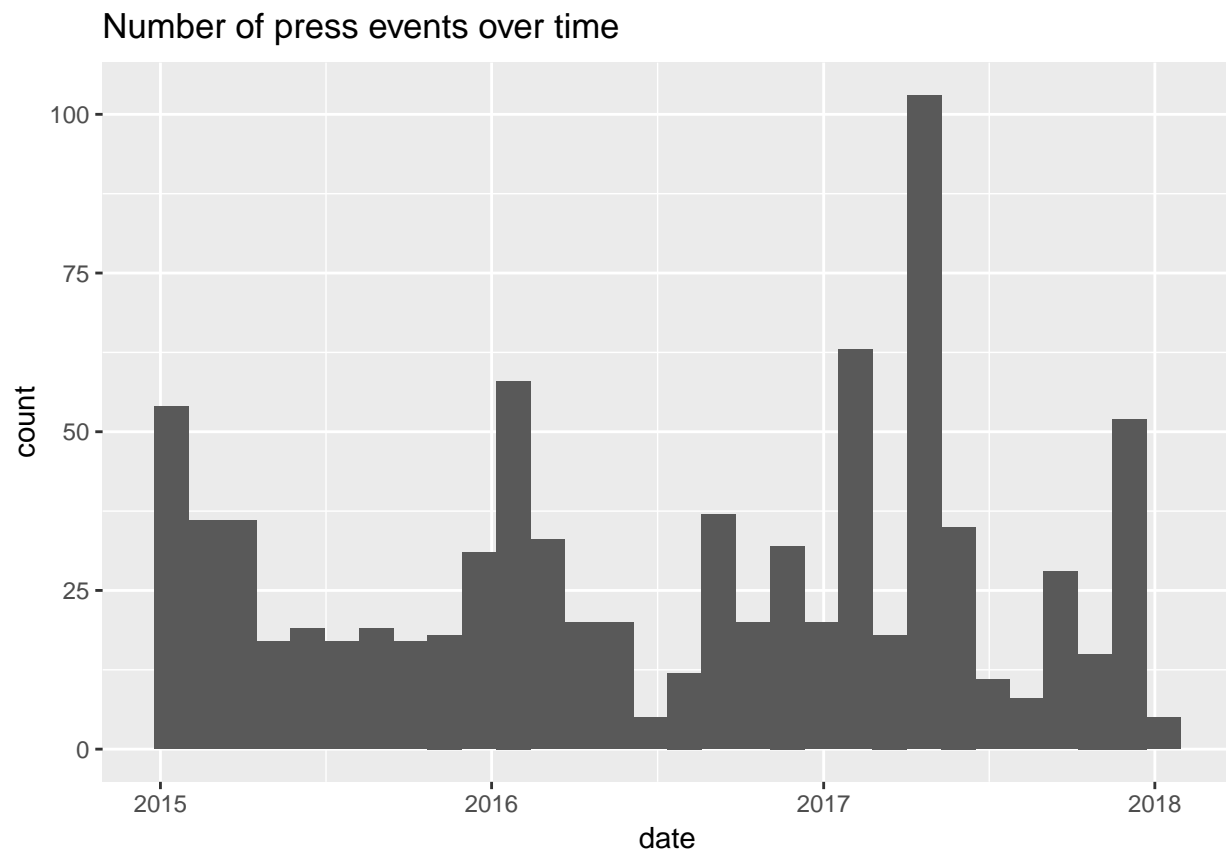
```
## 3 2015 Januar 3 taz
## 4 2015 Januar 3 rbb
## 5 2015 Januar 3 deutschlandradiokultur
## 6 2015 Januar 3 neues Deutschland
## 7 2015 Januar 4 Berliner Kurier
## 8 2015 Januar 4 Tagesspiegel
## 9 2015 Januar 5 neues Deutschland
## 10 2015 Januar 6 Berliner Zeitung
## # ... with 920 more rows, and 7 more variables: Article <chr>, Type <chr>,
## #   MediaType <chr>, CFD <int>, CFV <int>, monthN <int>, date <dtm>
```

## Descriptive exploration (just sanity check)

```
media %>% ggplot() + geom_histogram(aes(date)) + ggtitle('Number of press events over time')
```

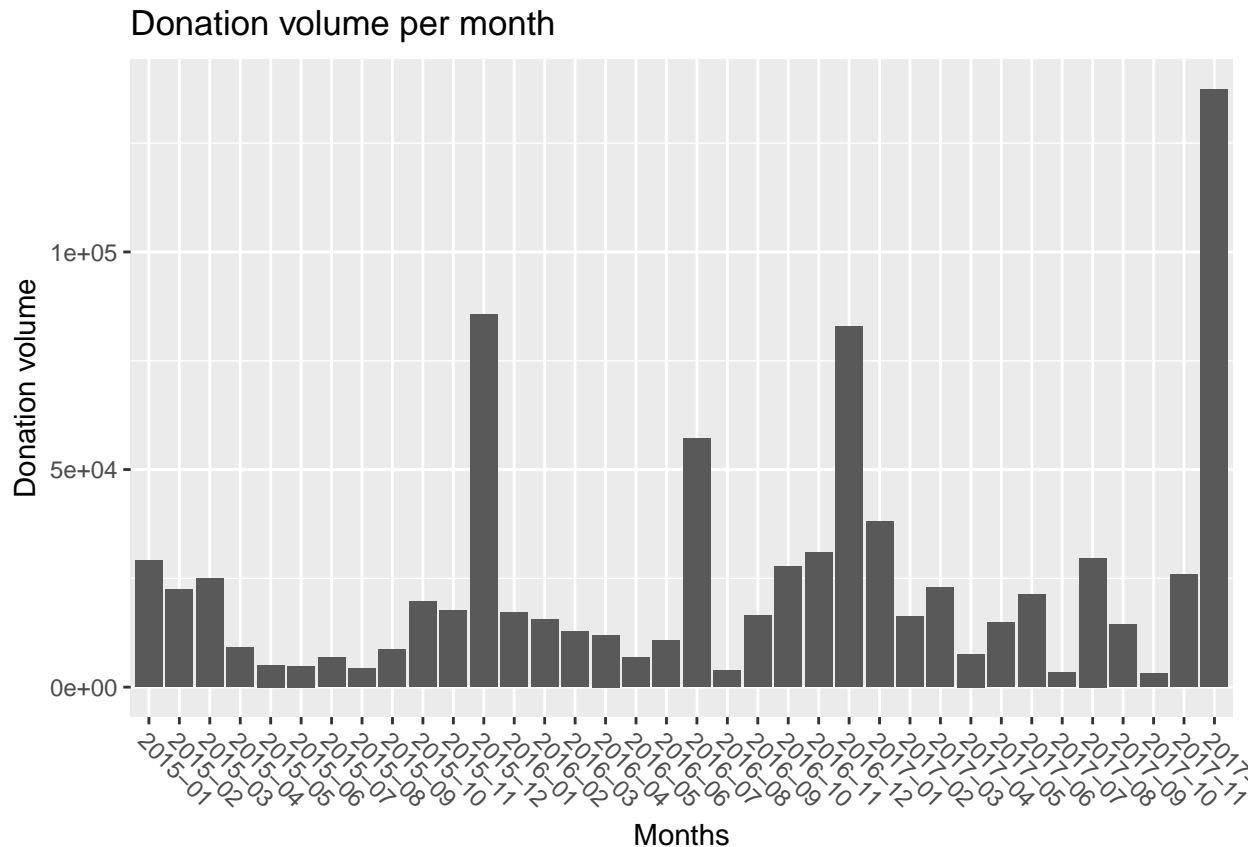
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 71 rows containing non-finite values (stat_bin).
```



```
donations %>% filter(resource == 'Geldspende') %>% ggplot() + geom_histogram(aes(format(Received, '%Y-%m')))
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



## Analysis

```
window = 60
```

I am interested in donation behavior after press events. My methodology is looking at the time period of 60 days after each press event. Averaging over all of these (possibly overlapping) periods I get an average view of donation behavior after press events.

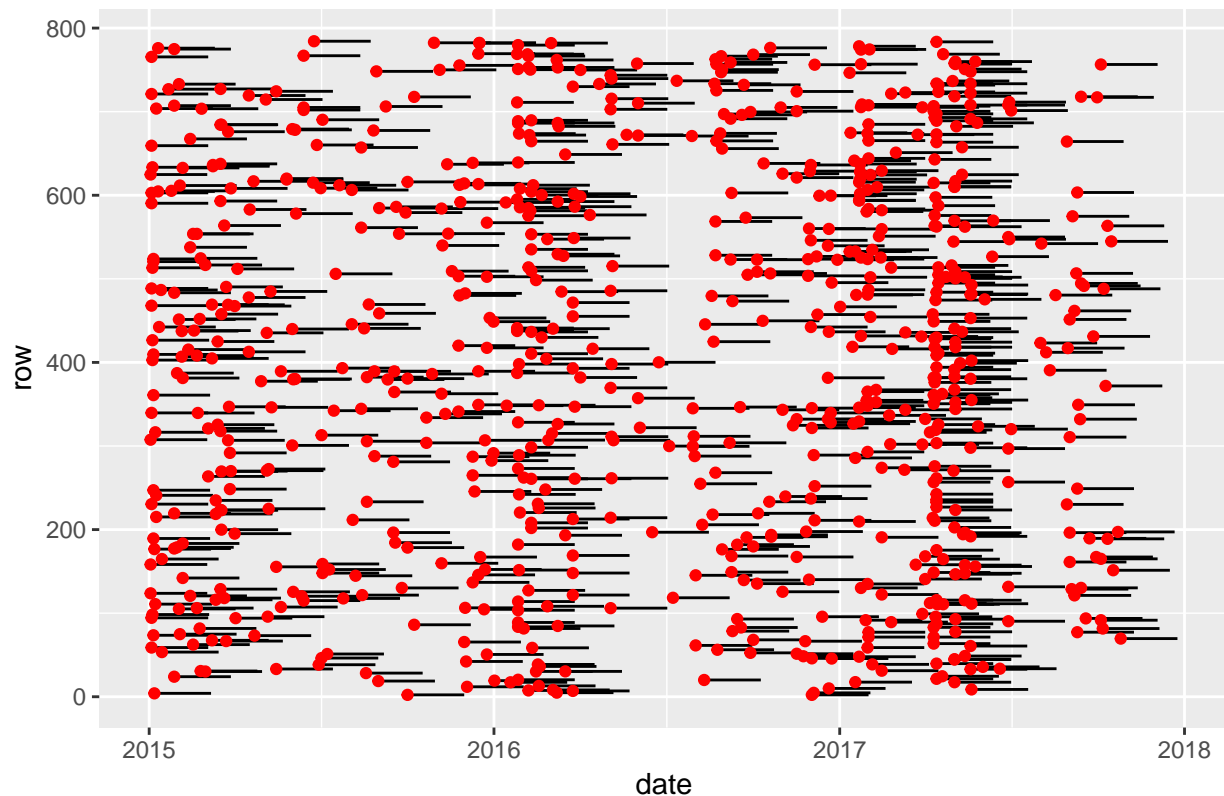
First I am filtering to press events where I have 60 days of donation history *after* the event. Then I am attaching to each event the 60 day history of donations after the event.

```
relevant_media = media %>%
  filter(lubridate::date(date) <= lubridate::date(max_donation_date) - days(window)) %>%
  filter(lubridate::date(date) >= lubridate::date(min_donation_date)) # only media events that fall bet
combi = donations %>% filter(resource == 'Geldspende') %>% crossing(relevant_media) %>% mutate(day_after
  filter(day_after_press >= 0 & day_after_press <= window) # keep donations window days after press eve
```

In the following plot shows each press event with it's 60 days observation period. These periods are overlapping each other as can be seen.

```
relevant_media %>% arrange(date) %>% mutate(row=runif(n(), max=n())) %>% ggplot() +
  geom_linerange(aes(x=row, ymin=date, ymax=date+days(60)), linetype=1) + coord_flip() +
  geom_point(aes(x=row, y=date), color='red') + ggtitle(paste('Press events with', window, 'days observ
```

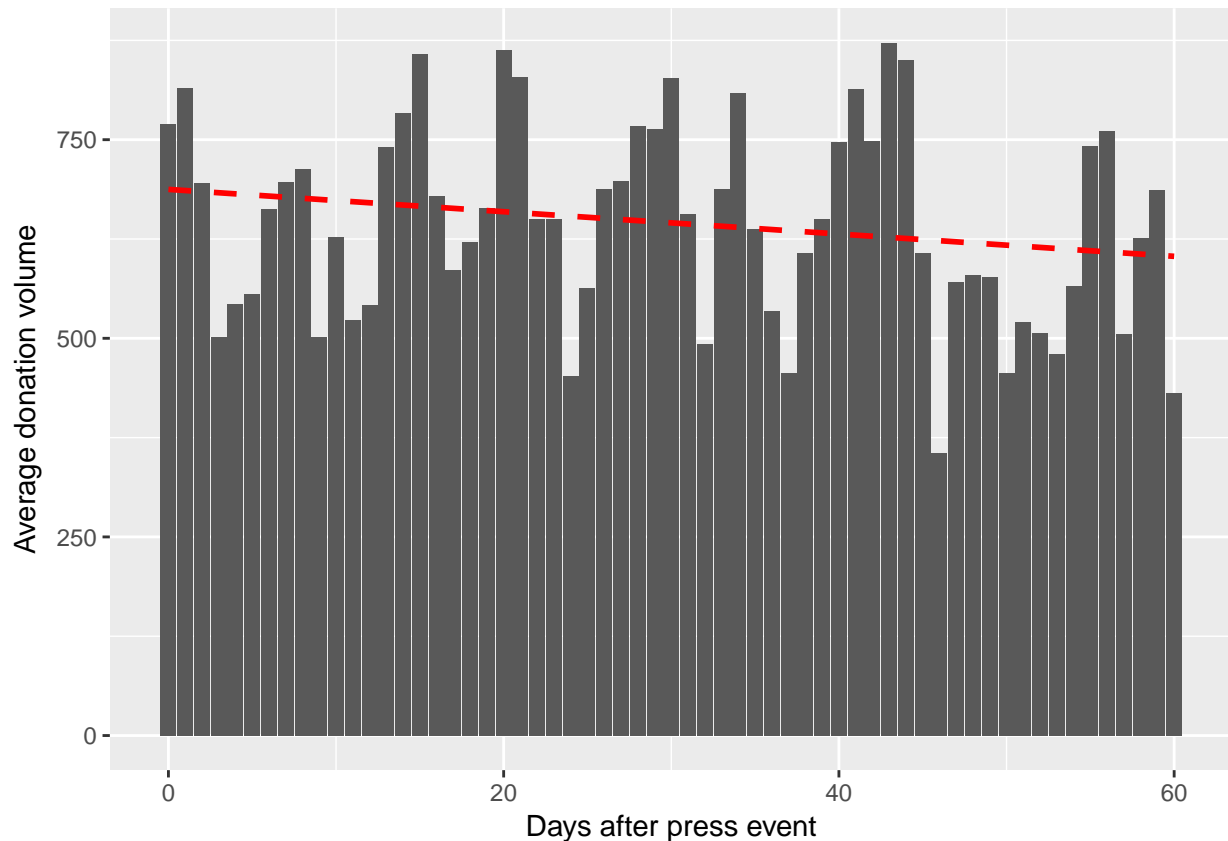
Press events with 60 days observation period



**Analysing the obvious: There actually is a link between press events and donation volume**

While this is the expected result it's really hard to see this from the data without the aggregations I made here because of all the effects from overlapping events. The next graph shows that the average donation volume is highest directly after press events and then steadily declines although it fluctuates heavily most likely from weekly periodicity and overlapping observation periods.

```
combi %>%
  group_by(day_after_press) %>%
  summarise(avgDonationVal=sum(`Contribution (in euro)`)/nrow(relevant_media)) %>%
  ggplot() + geom_bar(aes(day_after_press, avgDonationVal), stat='identity') + geom_smooth(aes(day_after_press, avgDonationVal))
  labs(x='Days after press event', y='Average donation volume')
```



## Delayed effects

This becomes even more interesting when looking at selected news outlets (with a higher number of articles on Berliner Tafel). Especially **dpa** exhibits an interesting delayed effect which can be explained when assuming that most people don't get the news from **dpa** but instead from other news outlets that pick up **dpa**'s press release. Interestingly **Tagesspiegel** displays a similar delayed effect while other daily newspapers do not and it's unclear why. This needs more investigation.

```
select_media = media %>%
  filter(Agency %in% c('dpa', 'Tagesspiegel', 'Berliner Zeitung', 'Berliner Woche')) %>% count(Agency)

combi %>%
  inner_join(select_media) %>%
  group_by(Agency, day_after_press) %>%
  summarise(avgDonationVal=sum(`Contribution (in euro)`)/first(n)) %>%
  ggplot() +
  geom_bar(aes(day_after_press, avgDonationVal), stat='identity') +
  geom_smooth(aes(day_after_press, avgDonationVal), se=F, linetype=2, color='red') +
  facet_wrap(~ Agency) +
  labs(x='Days after press event', y='Average donation volume')

## Joining, by = "Agency"
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

