

Causal impact study

GB

February 5, 2017

```
myfile<-file.choose()

mydata<-read.csv(myfile, header = TRUE)
mydata$ga.date<-as.Date(mydata$ga.date)
sprintf('start date in file: %s', min(mydata$ga.date))

## [1] "start date in file: 2015-04-01"
sprintf('end date in file: %s', max(mydata$ga.date))

## [1] "end date in file: 2016-11-21"

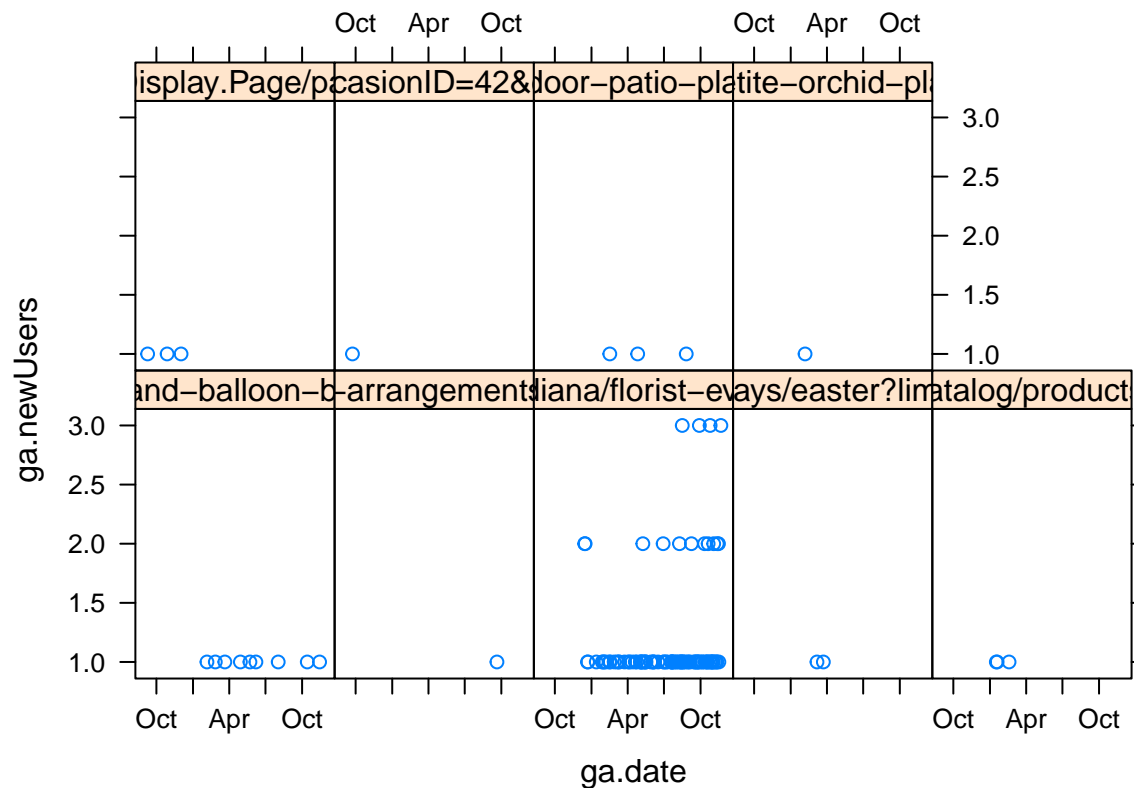
library(CausalImpact)
library(dplyr)
library(lattice)

paths<-unique(mydata$ga.landingPagePath)
sprintf('total different landing pages: %s', length(paths))

## [1] "total different landing pages: 7226"
```

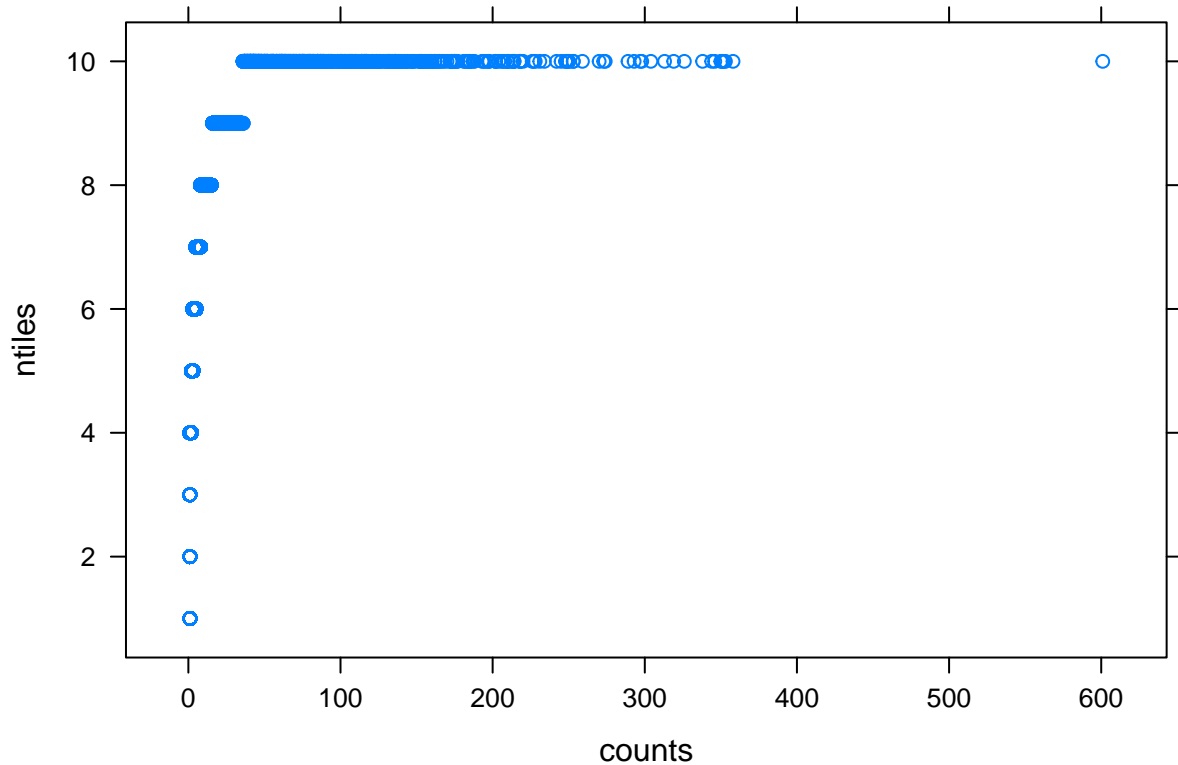
plotting some times series for the random 9 landing Pages, the number can be changed, but keep <12

```
testpaths<-sample(length(paths), 9)
data2<-mydata[mydata$ga.landingPagePath %in% paths[testpaths], ]
xyplot(ga.newUsers~ga.date|ga.landingPagePath, data2)
```



generating ntiles and plotting ntiles versus counts. counts here are total NewUsers per page. Number of ntiles is 10. can be changed.

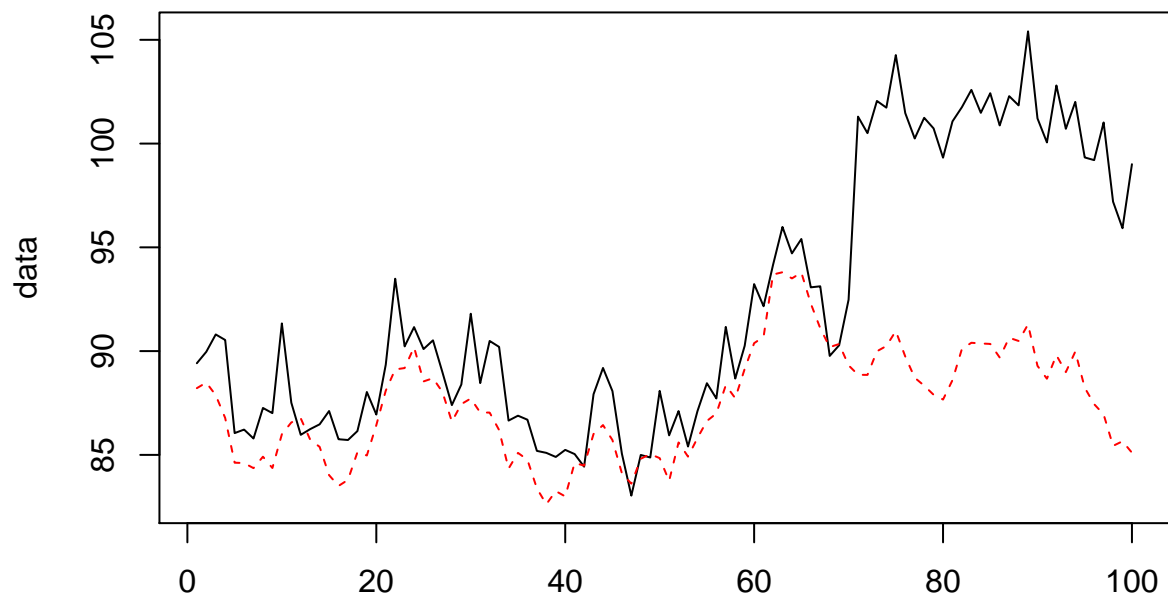
```
kn=10 # number of ntiles, can be changed
sumdata<-summarize(group_by(mydata, ga.landingPagePath), counts=n())
sumdata<-sumdata[order(-sumdata$counts), ]
sumdata$ntiles<-ntile(sumdata$counts, kn)
xyplot(ntiles~counts, sumdata)
```



```
kg=7 # kg = to number of test groups
```

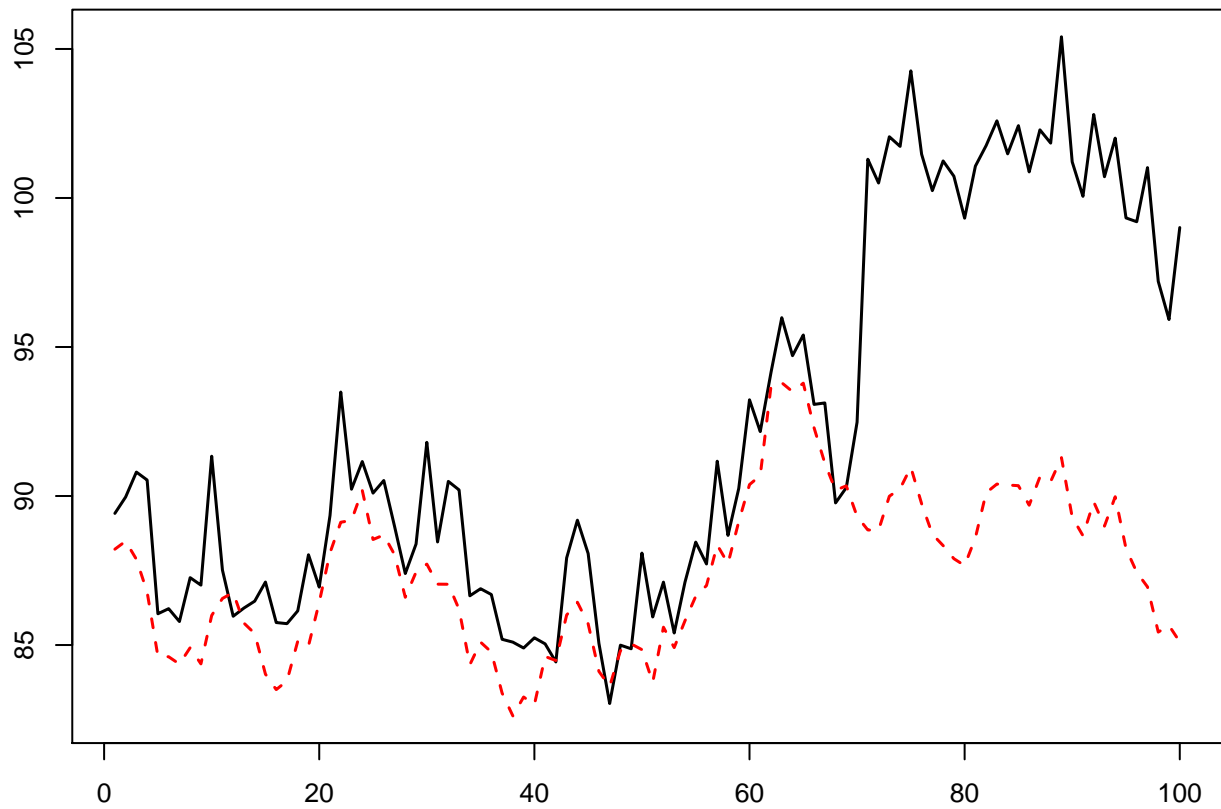
below is testing the package with synthetic data from the tutorial. this was done to assure that all dependences (supporting sub-packages) are in working condition

```
matplot(data, type = "l")
```



```
#par(cex = 0.85, oma = c(0, 0, 0, 0), mar = c(3, 2, 1, 1))  
#matplot(data, type = "l", lwd = 1.5)
```

```
par(cex = 0.85, oma = c(0, 0, 0, 0), mar = c(3, 2, 1, 1))  
matplot(data, type = "l", lwd = 1.5)
```



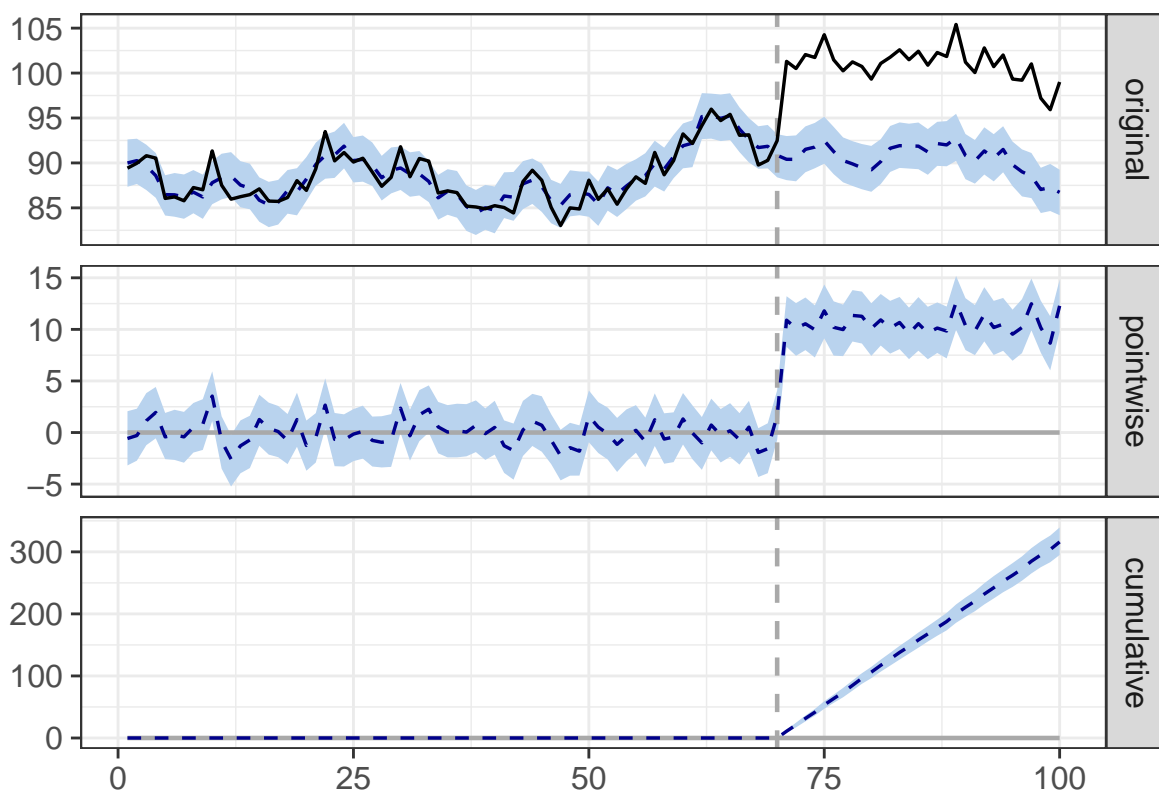
```
pre.period <- c(1, 70)
post.period <- c(71, 100)
```

```
impact <- CausalImpact(data, pre.period, post.period)
```

```
plot(impact)
```

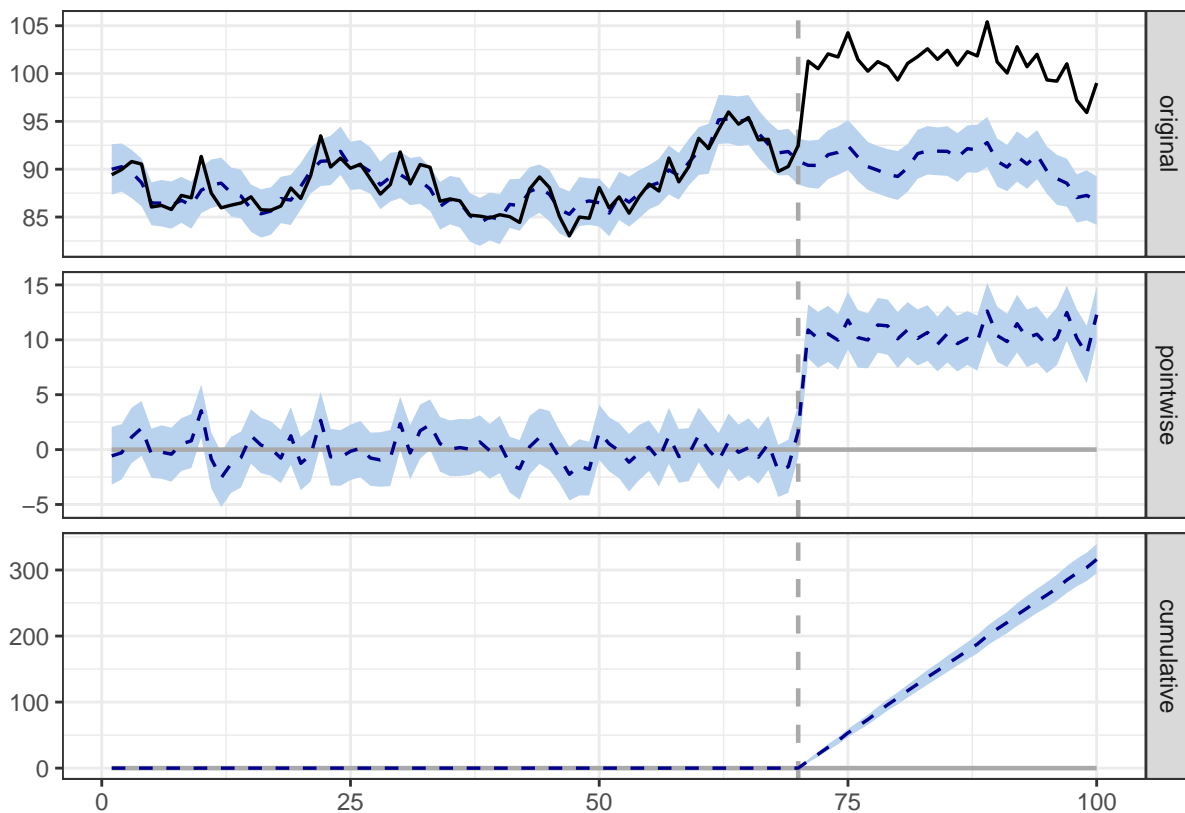
```
## Warning: Removed 100 rows containing missing values (geom_path).
```

```
## Warning: Removed 200 rows containing missing values (geom_path).
```



```
library(ggplot2)
```

```
q <- plot(impact) + theme_bw(base_size = 11)
suppressWarnings(plot(q))
```



```
time.points <- seq.Date(as.Date("2014-01-01"), by = 1, length.out = 100)
data <- zoo(cbind(y, x1), time.points)
head(data)
```

```
##           y      x1
## 2014-01-01 89.41626 88.21513
## 2014-01-02 89.96716 88.48415
## 2014-01-03 90.80304 87.87684
## 2014-01-04 90.53689 86.77954
## 2014-01-05 86.04914 84.62243
## 2014-01-06 86.21926 84.60650
```

```
pre.period <- as.Date(c("2014-01-01", "2014-03-11"))
post.period <- as.Date(c("2014-03-12", "2014-04-10"))
```

```
impact <- CausalImpact(data, pre.period, post.period)
plot(impact)
```

```
## Warning: Removed 100 rows containing missing values (geom_path).
```

```
## Warning: Removed 200 rows containing missing values (geom_path).
```

