

Dropout dynamics

Dominik Glandorf

2022-11-29

```
source('read_data.R')

## Lade nötiges Paket: tidyverse

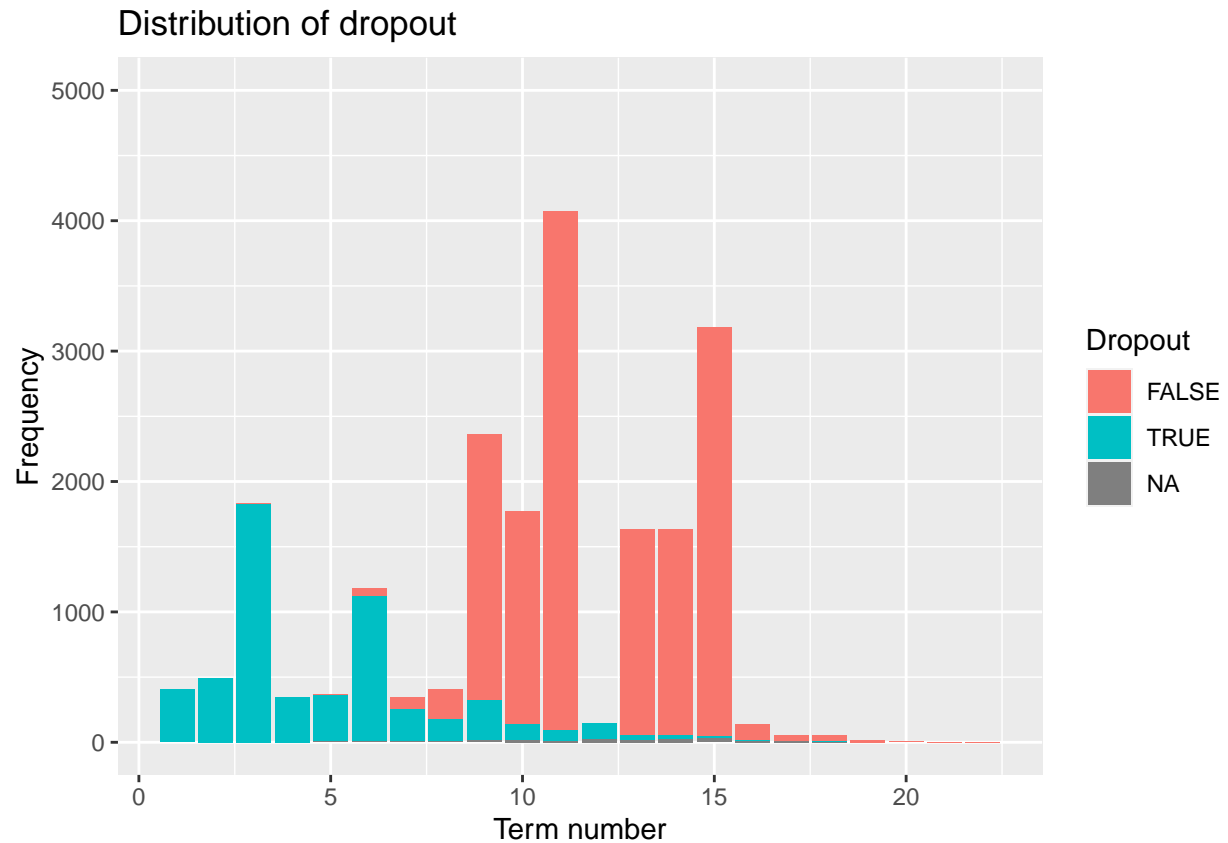
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

students = get_student_sub()
```

When does dropout happen?

```
ggplot(students, aes(fill=dropout, x=num_terms)) +
  geom_bar(position="stack", stat="count") +
  labs(title = "Distribution of dropout",
       x = "Term number",
       y = "Frequency", fill="Dropout") +
  ylim(0, 5000)
```

```
## Warning: Removed 1 rows containing missing values (‘geom_bar()’).
```



```
library(scales)
```

```
##
## Attache Paket: 'scales'
```

```
## Das folgende Objekt ist maskiert 'package:purrr':
##
##   discard
```

```
## Das folgende Objekt ist maskiert 'package:readr':
##
##   col_factor
```

```
stud = students[!is.na(students$dropout),]
dropped_out = data.frame(term=c(0:18), number=apply(c(0:18), FUN=function(i) nrow(stud[stud$dropout==T
ggplot(dropped_out,aes(x=term, y=number/max(number)))+
  geom_bar(stat="identity") +
  scale_y_continuous(labels=percent) +
  labs(title = "Cumulative histogram of dropouts",
        x = "Term number",
        y = "Share")
```

Cumulative histogram of dropouts

