Activity log

Jing Zhao

27 January, 2021

Activity Log of UNSIAP Data Visulization E-Learning Course

```
library(stringr)
library(dplyr)
library(tidyr)
library(ggplot2)
library(forcats)
```

Import data

```
data_log <- read.csv("Data/logs_DTV20_20210121-1615.csv")
```

Clean data

```
data_log <- data_log[!(data_log$User.full.name=="Christophe Bontemps" | data_log$User.full.name=="Ni Ni
```

Extract number and text from Description

- Extract course ID

```
CourseID <- str_match(data_log$Description, "(?i)\\bcourse with id '?\\s*(\\d+)")[,2]
head(CourseID)
```

```
## [1] "97" NA   NA   "97" "97" NA
```

- Extract users ID

```
UserID <- str_match(data_log$Description, "(?i)\\buser with id '?\\s*(\\d+)")[,2]
head(UserID)
```

```
## [1] "7523" "6283" "7523" "7673" "6283" "7005"
```

Extract module ID

```
ModuleID <- str_match(data_log$Description, "(?i)\\bcourse module with id '?\\s*(\\d+)")[,2]
head(ModuleID)
```

```
## [1] NA     "3235" NA     NA     NA     "3197"
```

- New dataframe with extracted ID

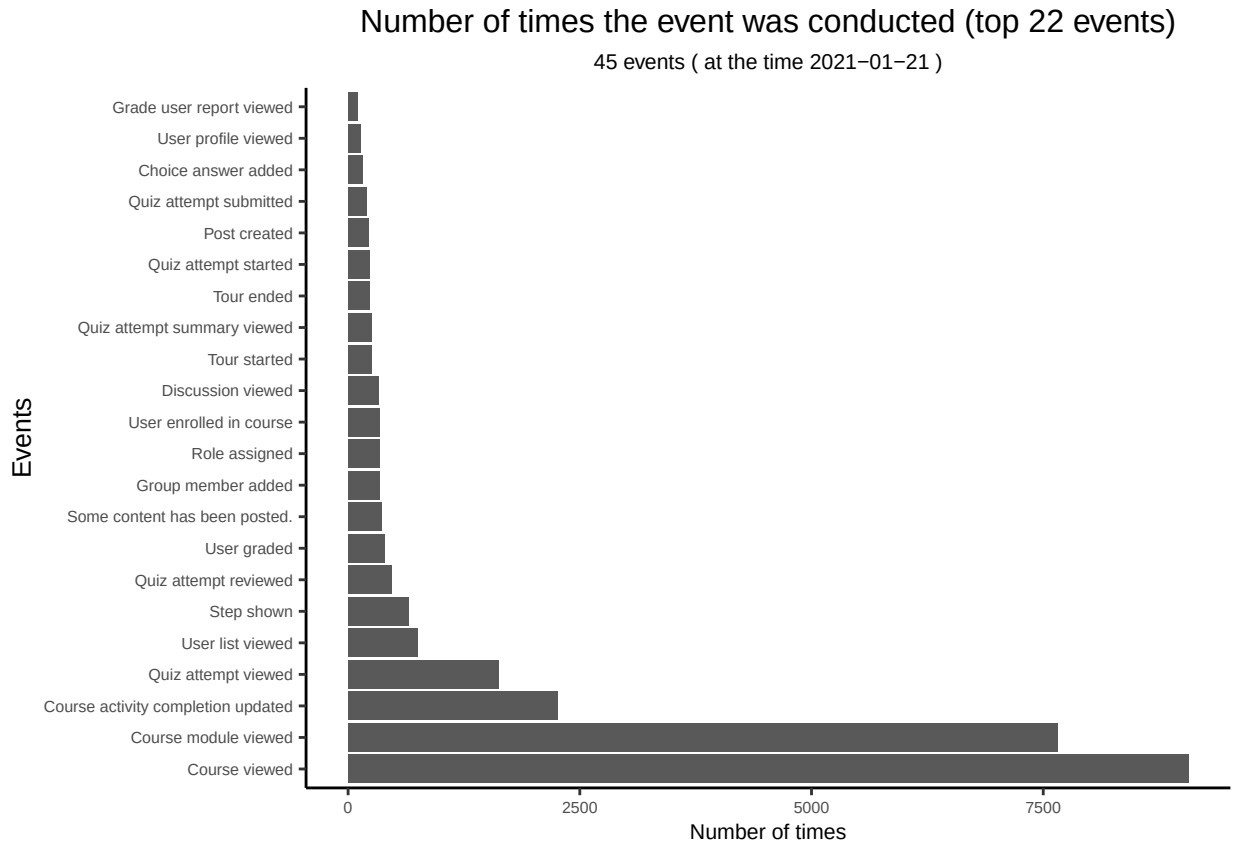
```
#CourseID is not added since it means this datavis course
log_new <- data.frame(data_log, ModuleID, UserID)
```

```
# log_new <- data_log %>%
#   select(Description) %>%
#   extract(Description, 'UserID', regex = "([\\d.]+)[^\\d.]", remove = FALSE) %>%
#   mutate(data_log)
```

Visualize some variables

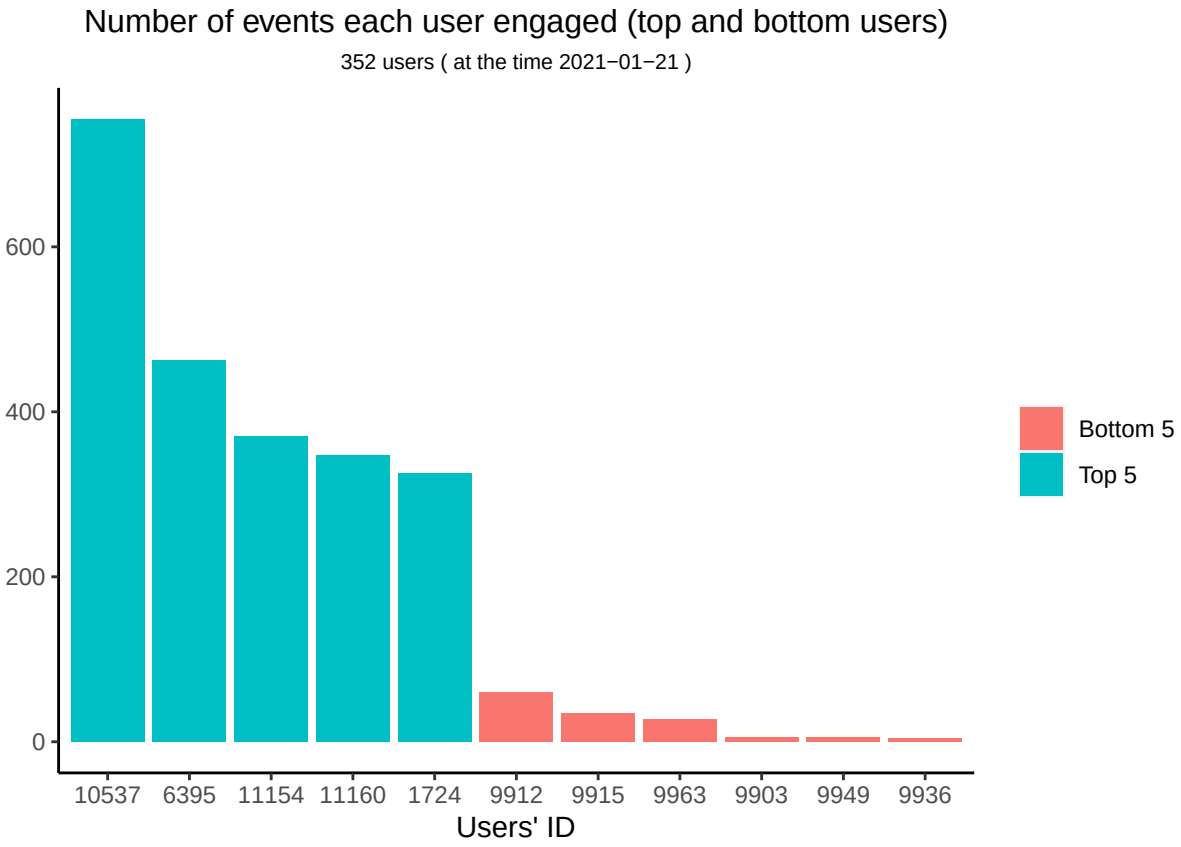
Number of times the event was conducted

```
log_new %>%
  filter(!is.na(Event.name)) %>%
  count(Event.name) %>%
  mutate(rank = dense_rank(desc(n))) %>%
  filter(rank <= 20) %>%
  ggplot()+
  aes(x=reorder(Event.name, -n), y=n)+
  geom_bar(stat='identity')+
  labs(x="Events", y="Number of times",
       subtitle = paste(length(unique(log_new$Event.name)),"events ( at the time", format(file.info("Da
                        ,"%Y-%m-%d"),")"))+
  ggtitle("Number of times the event was conducted (top 22 events)")+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        plot.title = element_text(hjust = 0.5, size=12),
        panel.background = element_blank(), text = element_text(size=8),
        axis.line = element_line(colour = "black"),
        plot.subtitle = element_text(hjust = 0.5, size=8),
        axis.title.y = element_text(size = 10))+
  coord_flip()
```



Number of activity events each user engaged (top and bottom users)

```
log_new %>%
  filter(!is.na(UserID)) %>%
  count(UserID) %>%
  mutate(rank = dense_rank(desc(n))) %>%
  filter(row_number() >= max(row_number()) - 5 | rank <= 5) %>%
  ggplot()+
  aes(x=reorder(UserID, -n), y=n,
      fill=factor(ifelse(rank <= 5,"Top 5","Bottom 5")))+
      geom_bar(stat='identity')+
  labs(x="Users' ID", y="", fill="",
       subtitle = paste(length(unique(log_new$UserID)),"users ( at the time", format(file.info("Data/log
                        ,"%Y-%m-%d"),")"))+
  ggtitle("Number of events each user engaged (top and bottom users)")+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        plot.title = element_text(hjust = 0.5, size=12),
        plot.subtitle = element_text(hjust = 0.5, size=8),
        panel.background = element_blank(), axis.line = element_line(colour = "black"))
```


Number of events each user engaged (top and bottom users)
352 users (at the time 2021−01−21)
600
400
200
0
10537
6395
11154
11160
1724
9912
9915
9963
9903
9949
9936
Users' ID
Bottom 5
Top 5