Video Analytics

INFO 5200 Learning Analytics Homework

[[Kimberly Williamson, khw44]]

In this homework, you will learn how to analyze video interaction data to generate video analytics.

Learning Objectives:

1. Exploring the structure of video interaction data
2. Identifying parts of the video with increased activity
3. Deciding what video analytics to report back to learners and instructors

You are given timestamped video interaction events for this video from an online course: https://youtu.be/qKNb8YQYTZg.

Dataset columns:

- id is a hashed learner id
- time is the exact time of the event
- order counts observed events in temporal for each id
- seconds measures time between events for the same person
- event_type
- video_current_time is the time in the video the event occured
- video_new/old_speed is for tracking speed_change events
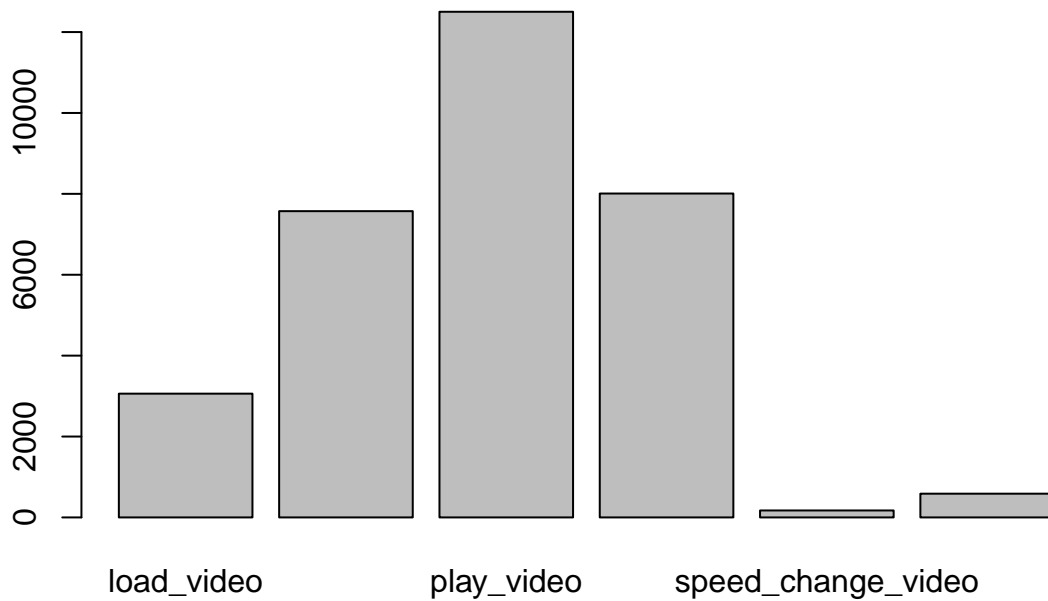- video_old/new_time is for tracking seek events

Part 1: Explore video data

Question 1: What event types are in the dataset and how many of each?

```
#########################################
####### BEGIN INPUT: Question 1 #######
#########################################
vid %>% group_by(event_type) %>% summarise(n=n())
```

```
## # A tibble: 6 x 2
##   event_type              n
##   <chr>               <int>
## 1 load_video           3060
## 2 pause_video          7572
## 3 play_video          12502
## 4 seek_video           8008
## 5 speed_change_video    172
## 6 stop_video            587
```

```
barplot(table(vid$event_type))
```

```
#######################################
#######################################
```

Question 2: How many users watched the video at all? How many video events do users have on average?

```
#######################################
####### BEGIN INPUT: Question 2 #######
#######################################
length(unique(vid$id))
```

```
## [1] 1544
```

```
nrow(vid)/length(unique(vid$id))
```

```
## [1] 20.66127
```

```
#######################################
#######################################
```

Question 3: Using a histogram, plot how much activity occurred throughout the video in total (x-axis = time in video; y-axis = event count). Tip: Set `breaks=300` in the hist() function for better resolution. Note that `seek_video` events have the `video_current_time` field missing so you should use video_old_time instead.

You should see a spike at the start, the end, one at around 300, and another around 700.

```
#######################################
####### BEGIN INPUT: Question 3 #######
#######################################
vidSeekFix = vid %>% mutate(video_current_time = ifelse(is.na(video_current_time),
                                    video_old_time,video_current_time))
hist(vidSeekFix$video_current_time, breaks = 300)
```

Histogram of vidSeekFix$video_current_time

Frequency

vidSeekFix$video_current_time

```
#########################################
#########################################
```
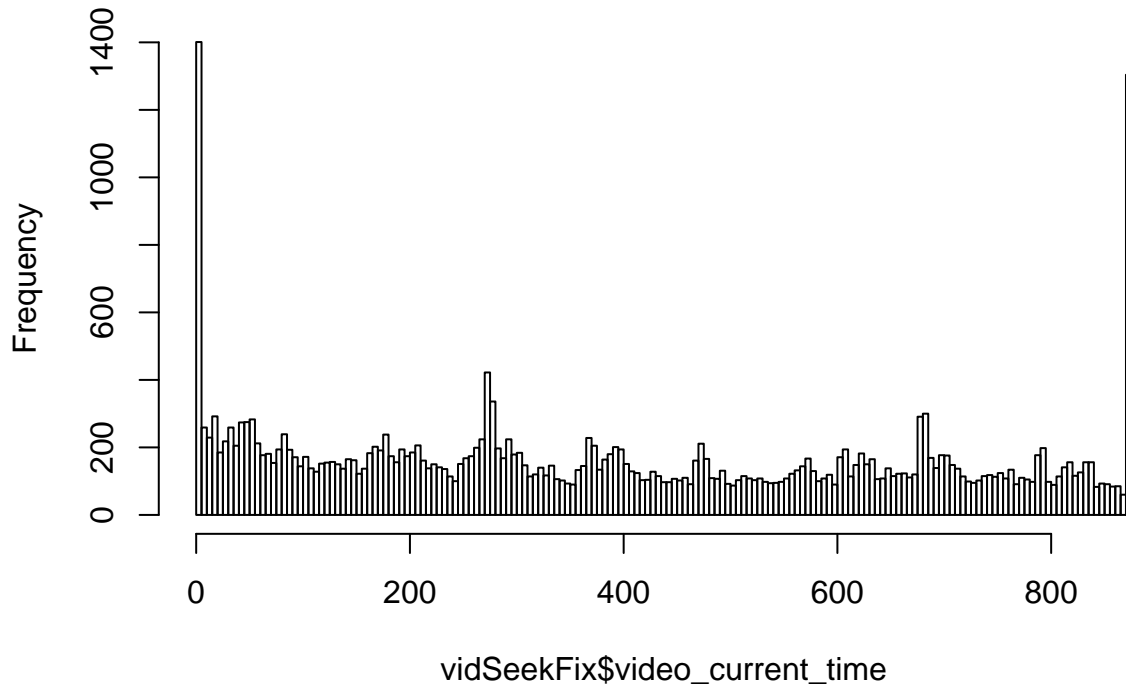
Question 4: At what point in the video do many learners pause? Using a precision of 10 seconds (meaning that you should round numbers to the nearest 10 seconds, so e.g. 123 becomes 120 and 47 becomes 40), report the 3 most common times in seconds. Then go to the video lecture and look up what happens at those times (note that you can type in the seconds in this URL so you don't need to convert it). Say why you think learners paused there.

```
#########################################
####### BEGIN INPUT: Question 4 #######
#########################################
pause = vid %>% filter(event_type=="pause_video")
pause %>% mutate(pauseRound = round(video_current_time, digits=-1)) %>%
  group_by(pauseRound) %>% summarise(n=n()) %>% arrange(-n) %>% head(3)


## # A tibble: 3 x 2
##    pauseRound       n
##         <dbl>  <int>
## 1         870    608
## 2         270    210
## 3         680    199
# 870: Is the end of the video. I'm guessing people paused so autovideo would not take over.

# 270: There is a query ran. There were pauses while the students following along
# could execute the query.

# 680: There is a query ran. There were pauses while the students following along could
# execute the query.
```

3

```
##########################################
##########################################
```

Question 5: At what point in the video do many learners seek forwards and when do they seek backwards in the video (ignore for now where they seek to)? Using a precision of 10 seconds, report the 3 most common times in seconds for moving back, 3 most common for moving forward. Then go to the video lecture and look up what happens at those times. Why might learners move?

```
##########################################
####### BEGIN INPUT: Question 5 #######
##########################################
seek = vid %>% filter(event_type=="seek_video")
seek %>% mutate(seekRound = round(video_old_time, digits=-1)) %>%
  group_by(seekRound) %>% summarise(n=n()) %>% arrange(-n) %>% head(3)


## # A tibble: 3 x 2
##   seekRound     n
##       <dbl> <int>
## 1       280   229
## 2         0   173
## 3        50   171
# 280: Query ran and students who needed more information on how the query was composed
# were using the seek.

# 0: Students started the video and seeked to the content they wanted.

# 50: Deleting existing data. Students are not intersted in delete.


##########################################
##########################################
```

Question 6: At what point in the video do many learners seek backwards in the video and where do they go? Using a precision of 10 seconds, report the 3 most common pairs of times in seconds like this <from, to>. Then go to the video lecture and look up what happens at those times. Why do you think learners moved there?

```
##########################################
####### BEGIN INPUT: Question 6 #######
##########################################
seek %>% filter(video_old_time>video_new_time) %>%
  mutate(from=round(video_old_time, digits = -1), to=round(video_new_time, digits = -1)) %>%
  group_by(from, to) %>% summarise(n=n()) %>% arrange(-n) %>% head(3)


## # A tibble: 3 x 3
## # Groups:   from [3]
##    from    to     n
##   <dbl> <dbl> <int>
## 1   280   270    68
## 2   690   680    46
## 3    50    40    38
# 280-270: Query ran and students who needed more information on how the query was composed
# were using the seek.

# 690-680: Query ran and students who needed more information on how the query was composed
# were using the seek.
```

```
# 50-40: Students seem to want more information about inserting new data.
#########################################
#########################################
```

Question 7: Are students more likely to speed the video up or slow it down? Report the proportion speed-ups relative to all speed changes. How do you interpret this?

```
#########################################
####### BEGIN INPUT: Question 7 #######
#########################################
speedChange = vid %>% filter(event_type=="speed_change_video")
nrow(filter(speedChange,video_old_speed<video_new_speed))/nrow(speedChange)
```

```
## [1] 0.5755814
```

```
# I would interpret that students felt the video speed was slower than they wanted.
#########################################
#########################################
```

Question 8: Based on the video analytics you have done (or any additional analyses you'd like to do), what information would you give (a) a learner about to watch this lecture video, and (b) the instructor of the lecture video? Make emprically grounded recommendations for each stakeholder.

- **Recommendation to learners:** Watch this lecuture video with a double monitor so you can follow along with a screen showing the lecture and another screen showing your own SQL IDE.

- **Recommendation to instructor:** Spend more time constructing the queries to allow students to follow along and not have to pause or go back.

Self-reflection

Briefly summarize your experience on this homework. What was easy, what was hard, what did you learn?

This assignment was straight foward and a nice compliment to the lecutures for the week.

Submit Homework

This is the end of the homework. Please **Knit a PDF report** that shows both the R code and R output and upload it on the EdX platform. Alternatively, you can Knit it as a "doc", open it in Word, and save that as a PDF.

Important: Be sure that all your code is visible. If the line is too long, it gets cut off. If that happens, organize your code on several lines.