```
readRDS("data/rem_data.rds") -> data
data %>% filter(!is.na(snd_da) & !is.na(rec_da)) %>% glimpse()
```

```
## Rows: 18,106
## Columns: 9
## $ session    <int> 2102, 2102, 2102, 2102, 2102, 2102, 2102, 2102, 2102, 2102,~
## $ sender     <chr> "Igor", "Ashley", "Will", "Igor", "Ashley", "Igor", "Will",~
## $ receiver   <chr> "Ashley", "Will", "Igor", "Ashley", "Igor", "Will", "Katya"~
## $ snd_gender <chr> "Female", "Male", "Male", "Female", "Male", "Male", "Female~
## $ rec_gender <chr> "Male", "Female", "Male", "Male", "Female", "Male", "Male",~
## $ snd_da     <chr> "disruption", "statement", "question", "statement", "statem~
## $ rec_da     <chr> "statement", "question", "statement", "statement", "stateme~
## $ snd_id     <int> 1, 2, 3, 1, 2, 1, 3, 4, 2, 5, 3, 1, 6, 2, 3, 2, 1, 3, 2, 1,~
## $ rec_id     <int> 2, 3, 1, 2, 1, 3, 4, 2, 5, 3, 1, 6, 2, 3, 2, 1, 3, 2, 1, 5,~
```

```
head(data)
```

```
## # A tibble: 6 x 9
##   session sender receiver snd_gender rec_gender snd_da     rec_da  snd_id rec_id
##     <int> <chr>  <chr>    <chr>      <chr>      <chr>      <chr>    <int>  <int>
## 1    2102 Igor   Ashley   Female     Male       disruption statem~      1      2
## 2    2102 Ashley Will     Male       Female     statement  questi~      2      3
## 3    2102 Will   Igor     Male       Male       question   statem~      3      1
## 4    2102 Igor   Ashley   Female     Male       statement  statem~      1      2
## 5    2102 Ashley Igor     Male       Female     statement  statem~      2      1
## 6    2102 Igor   Will     Male       Male       statement  questi~      1      3
```

```
# set 0 if snd_gender == Female and 0 if snd_gender = Male
data$rec_gender <- ifelse(data$rec_gender == "Female", 0, 1)
data$snd_gender <- ifelse(data$snd_gender== "Female", 0, 1)

data$sender <- as.factor(data$snd_id)
data$receiver <- as.factor(data$rec_id)
data$snd_gender <- as.factor(data$snd_gender)
data$rec_gender <- as.factor(data$rec_gender)
data$snd_da <- as.factor(data$snd_da)
data$rec_da <- as.factor(data$rec_da)

# convert to faactors
data %>% select(session, snd_id, rec_id, snd_gender, rec_gender, snd_da, rec_da) -> rem_data
```

```
head(rem_data)
```

```
## # A tibble: 6 x 7
##   session snd_id rec_id snd_gender rec_gender snd_da     rec_da
##     <int>  <int>  <int> <fct>      <fct>      <fct>      <fct>
## 1    2102      1      2 0          1          disruption statement
## 2    2102      2      3 1          0          statement  question
## 3    2102      3      1 1          1          question   statement
## 4    2102      1      2 0          1          statement  statement
## 5    2102      2      1 1          0          statement  statement
## 6    2102      1      3 1          1          statement  question
```

```r
# group by session and build REM for each session
rem_data %>% group_by(session) %>% summarise(num_of_dialogs = n()) %>% filter(num_of_dialogs > 1) -> ses
```

```r
print(sessions)
```

```
## # A tibble: 18 x 2
##    session num_of_dialogs
##      <int>          <int>
##  1    2102            676
##  2    2103           3799
##  3    2104            988
##  4    2105           1175
##  5    2106           1303
##  6    2107           1004
##  7    2108            608
##  8    2109            998
##  9    2110           1055
## 10    2111            527
## 11    2112            965
## 12    2113            618
## 13    2114            828
## 14    2115            821
## 15    2116            576
## 16    2117            910
## 17    2118            704
## 18    2119            551
```

```r
# for each session, create the timing system time using the order like 1, 2, 3 (ordinal)
rem_data %>% group_by(session) %>% mutate(event_order = row_number())
```

```
## # A tibble: 18,106 x 8
## # Groups:   session [18]
##    session snd_id rec_id snd_gender rec_gender snd_da     rec_da    event_order
##      <int>  <int>  <int> <fct>      <fct>      <fct>      <fct>           <int>
##  1    2102      1      2 0          1          disruption statement           1
##  2    2102      2      3 1          0          statement  question            2
##  3    2102      3      1 1          1          question   statement           3
##  4    2102      1      2 0          1          statement  statement           4
##  5    2102      2      1 1          0          statement  statement           5
##  6    2102      1      3 1          1          statement  question            6
##  7    2102      3      4 0          1          question   statement           7
##  8    2102      4      2 0          0          statement  statement           8
##  9    2102      2      5 1          0          statement  statement           9
## 10    2102      5      3 1          1          statement  statement          10
## # i 18,096 more rows
```