Analyse Badges in StackOverflow

Peter Baumgartner 2018-05-20

Contents

```
## Loading required package: tidyverse
## -- Attaching packages ------ tidyverse 1.2.1 --
## √ ggplot2 2.2.1
             √ purrr
                    0.2.4
## \sqrt{\text{tibble } 1.4.2}
              √ dplyr
                    0.7.5
## √ tidyr
       0.8.1
              √ stringr 1.3.1
## √ readr
       1.1.1
             √ forcats 0.3.0
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()
            masks stats::lag()
## Loading required package: lubridate
##
## Attaching package: 'lubridate'
## The following object is masked from 'package:base':
##
##
    date
## Loading required package: stackr
## Loading required package: jsonlite
##
## Attaching package: 'jsonlite'
## The following object is masked from 'package:purrr':
##
##
    flatten
## Loading required package: dlstats
## Loading required package: printr
```

0.1 Download badges information

I need to download all types of badges with their description and for further investigation to categorize them.

0.1.1 Categorization of Badges

The different types of badges can be seen on help page for badges. These types are:

- 1) Question badges
- 2) Answer badges
- 3) Partication badges
- 4) Moderation badges
- 5) Other badges
- 6) Documentation badges

Another differentiation is the responsibility of awarding badges. There are badges you can get by just doing a certain action once of specified times. And there are other badges you get only after some community action. I call the first type user and the second type community badges.

Examples for user badges:

Name: "Civic Duty" (Vote 300 times). Name: "Enthusiast" (Visit the site each day for 30 consecutive day)

Some of the user badges are only available after some threshold value is reached and a privileges is granted. For instance voting up needs 15 and voting down 125 reputation points.

Examples for community badges:

Name: "Favorite Question" (Question favored by 25 users) Name: "Popular Question" (Question with 1000 views).

There are also badges tied with certain events, like moderator election, working with the beta version, meeting employees at an event etc. And there are some retired badges (e.g. "Analytical") but also some badges already withdrawn (e.g. all three documenation badges). I call these two groups event and dead badges. In summary there are:

- 1) User badges
- 2) Community badges
- 3) Event badges
- 4) Dead badges

The user badges are of utmost importance for my research. Most of these badges are granted for quality assurance work on the website and are not linked with reputation points. They are – from a systemic perspective – important for the dynamic of the website and maintain respective raise the quality of the platform and their ecological sustainability.

When I can show that people are striving to get user badges – even when they are not linked with reputation points – than it is evident that badges are some additional (motivational) factors for the community development.

```
> badges <- readRDS("data/badges-raw.rds")
> activity <-
                 c(1,5,5,4,3,1,3,5,3,5,
                   4,4,4,3,4,3,3,4,4,1,
+
                   4,4,4,6,6,6,4,4,2,3,
                   3,4,2,1,3,1,2,2,1,2,
+
                   1,2,2,5,1,1,3,4,3,2,
                   2,1,5,1,4,3,4,1,2,3,
+
                   1,4,5,3,3,2,4,2,4,2,
                   1,2,4,1,4,1,4,4,1,4,
                   4,4,4,3,4,2,2,1,2,4,
> badges <- cbind(badges, activity)</pre>
> badge_labels <- c(</pre>
                         # 1
         "Question",
         "Answer",
+
+
         "Participation", #3
+
         "Moderation",
                         # 4
                         # 5
         "Other",
+
         "Documentation" # 6
> badges$activity <- factor(badges$activity, levels = c(1:6), labels = badge_labels)
> # add who's initiative is responsible for awarding the badge?
>
> who <- c(1,4,2,1,1,1,3,2,3,3,
          1,1,1,1,3,3,2,1,1,1,
+
          1,1,1,4,4,4,1,1,2,1,
          2,1,1,2,1,2,2,2,2,2,
          2,2,1,1,1,1,2,1,2,2,
+
          2,2,3,2,1,2,1,2,2,4,
+
          1,1,2,2,2,1,1,2,1,1,
          1,1,3,1,1,2,1,1,1,1,
+
          1,2,1,2,2,1,1,2,2,1,
          3)
> badges <- cbind(badges, who)</pre>
> badge_labels <- c(</pre>
         "User",
                         # 1
+
         "Community",
+
                         # 2
         "Event",
                         # 3
+
         "Retired"
                         # 4
+ )
> badges$who <- factor(badges$who, levels = c(1:4), labels = badge_labels)
> # Count the number of who badges and how many there are per type
> by activity <- badges %>%
         group_by(activity) %>%
+
         summarise(n = n(),
+
                  total = sum(award_count)) %>%
```

Table 1: Type of Badges in Stack

badge_id	name	award_count	rank	description
222	Altruist	8151	bronze	First bounty you manually award on another person's quest
1306	Analytical	43744	bronze	Visited every section of the FAQ (retired)
260	Announcer	130900	bronze	Share a link to a question later visited by 25 unique IP addr
1286	Archaeologist	1883	silver	Edit 100 posts that were inactive for 6 months
9	Autobiographer	798602	bronze	Complete "About Me" section of user profile
221	Benefactor	38849	bronze	First bounty you manually award on your own question
30	Beta	2510	silver	Voted 10 times, added 3 posts score > 0, and visited the sit
261	Booster	3224	silver	Share a link to a question later visited by 300 unique IP add
1973	Caucus	573686	bronze	Visit an election during any phase of an active election and
6644	Census	71168	silver	Completed the annual Stack Overflow survey; your response
8	Citizen Patrol	177353	bronze	First flagged post
32	Civic Duty	80224	silver	Vote 300 or more times
4	Cleanup	40815	bronze	First rollback
31	Commentator	799142	bronze	Leave 10 comments
3108	Constable	0	gold	Served as a pro-tem moderator for at least 1 year or throug
1974	Constituent	186474	silver	Vote for a candidate in the final phase of an election
901	Convention	2041	silver	10 posts with score of 2 on meta
223	Copy Editor	2807	gold	Edit 500 posts (excluding own or deleted posts and tag edit
7	Critic	337565	bronze	First down vote
4127	Curious	283242	bronze	Ask a well-received question on 5 separate days, and mainta
2278	Custodian	659074	bronze	Complete at least one review task. This badge is awarded o
1002	Deputy	10646	silver	Raise 80 helpful flags
37	Disciplined	10857	bronze	Delete own post with score of 3 or higher
6157	Documentation Beta	294	silver	Contributed 3+ substantive pieces of documentation during
6158	Documentation Pioneer	2359	silver	Contributed 3+ substantive pieces of documentation in the
7358	Documentation User	43293	silver	Earned at least one badge for contributing to Stack Overflow
3	Editor	1939014	bronze	First edit
155	Electorate	17035	gold	Vote on 600 questions and 25% or more of total votes are or
19	Enlightened	326524	silver	First to answer and accepted with score of 10 or more
71	Enthusiast	184595	silver	Visit the site each day for 30 consecutive days. (Days are co
145	Epic	659	silver	Earn 200 daily reputation 50 times
1287	Excavator	116090	bronze	Edit first post that was inactive for 6 months

55923 bronze Edit and answer 1 question (both actions within 12 hours, ϵ

4368 Explainer

badge_id	name	award_count	rank	description
				-
28	Famous Question	530351	gold	Question with 10,000 views
83	Fanatic	27995	gold	Visit the site each day for 100 consecutive days. (Days are d
33	Favorite Question	42062	silver	Question favorited by 25 users
15	Generalist	1018	silver	Provide non-wiki answers of 15 total score in 20 of top 40 ta
24	Good Answer	338308	silver	Answer score of 25 or more
21	Good Question	158284	silver	Question score of 25 or more
25	Great Answer	58302	gold	Answer score of 100 or more
22	Great Question	26443	gold	Question score of 100 or more
18	Guru	107275	silver	Accepted answer and score of 40 or more
4370	Illuminator	92	gold	Edit and answer 500 questions (both actions within 12 hour
2600	Informed	1725866	bronze	Read the entire tour page
4128	Inquisitive	28291	silver	Ask a well-received question on 30 separate days, and maint
219	Investor	16361	bronze	First bounty you offer on another person's question
146	Legendary	250	gold	Earn 200 daily reputation 150 times
1298	Marshal	2508	gold	Raise 500 helpful flags
144	Mortarboard	29138	bronze	Earn at least 200 reputation (the daily maximum) in a single
17	Necromancer	450470	silver	Answer a question more than 60 days later with score of 5 d
23	Nice Answer	1019824	bronze	Answer score of 10 or more
20	Nice Question	490456	bronze	Question score of 10 or more
6381	Not a Robot	831	silver	Met a Stack Overflow employee at an event where Stack Ov
27	Notable Question	1857309	silver	Question with 2,500 views
5	Organizer	110957	bronze	First retag
998	Outspoken	1155	silver	Post 10 messages in chat starred by 10 different users
38	Peer Pressure	186189	bronze	Delete own post with score of -3 or lower
26	Popular Question	3720433	bronze	Question with 1,000 views
62	Populist	16748	gold	Highest scoring answer that outscored an accepted answer v
892	Precognitive	0	bronze	Followed the Area 51 proposal for this site before it entered
220	Promoter	69006	bronze	First bounty you offer on your own question
1276	Proofreader	16453	bronze	Approve or reject 100 suggested edits
262	Publicist	951	gold	Share a link to a question later visited by 1000 unique IP ac
94	Pundit	9197	silver	Leave 10 comments with score of 5 or more
900	Quorum	25210	bronze	One post with score of 2 on meta
4369	Refiner	1772	silver	Edit and answer 50 questions (both actions within 12 hours
1656	Research Assistant	280	silver	Edit 50 tag wikis
95	Reversal	291	gold	Provide an answer of $+20$ score to a question of -5 score
1478	Reviewer	33387	silver	Complete at least 250 review tasks. This badge is awarded of
837	Revival	323943	bronze	Answer more than 30 days after a question was asked as firs
10	Scholar	1595685	bronze	Ask a question and accept an answer
14	Self-Learner	121988	bronze	Answer your own question with score of 3 or more
3109	Sheriff	36	gold	Served as an elected moderator for at least 1 year
4129	Socratic	3357	gold	Ask a well-received question on 100 separate days, and mair
805	Sportsmanship	2572	silver	Up vote 100 answers on questions where an answer of yours
36	Stellar Question	6035	gold	Question favorited by 100 users
2279	Steward	11882	gold	Complete at least 1,000 review tasks. This badge is awarded
12	Strunk & White	14694	silver	Edit 80 posts
2	Student	1935264	bronze	First question with score of 1 or more
804	Suffrage	43757	bronze	Use 30 votes in a day
6	Supporter	1167482	bronze	First up vote
1224	Synonymizer	899	bronze	First approved tag synonym
254	Tag Editor	21962	bronze	First tag wiki edit
884	Talkative	9608	bronze	Post 10 messages, with 1 or more starred, in chat
11	Taxonomist	11741	silver	Create a tag used by 50 questions

$\underline{\mathrm{badge_id}}$	name	$award_count$	rank	description
1	Teacher	1265454	bronze	Answer a question with score of 1 or more
225	Tenacious	48154	silver	Zero score accepted answers: more than 5 and 20% of total
63	Tumbleweed	997382	bronze	Asked a question with zero score, no answers, no comments,
226	Unsung Hero	18844	gold	Zero score accepted answers: more than 10 and 25% of tota
1108	Vox Populi	28439	bronze	Use the maximum 40 votes in a day
13	Yearling	1327576	silver	Active member for a year, earning at least 200 reputation

> knitr::kable(by_activity, caption = "Main Types of Badges in Stack Overflow")

Table 2: Main Types of Badges in Stack Overflow

activity	n	total
Question	18	11806961
Answer	17	4154930
Participation	17	3977838
Moderation	29	5046601
Other	7	1976684
Documentation	3	45946

> knitr::kable(by_who, caption = "User and Community Badges in Stack Overflow")

Table 3: User and Community Badges in Stack Overflow

n	total
46	14365657
32	10391332
8	2162281
5	89690
	46 32 8

> saveRDS(badges, file = "data/badges.rds")

0.2 Groups of badges with same task but different difficulty

There are some badges which differ only in their difficulty. According to their level of difficulty they are awared with bronze, silver or gold badges. I will focus on those badges with who = 1 (user is responsible).

Visit the site each day for [...] consecutive days. (Days are counted in UTC.)

- 30 (Enthusiast)
- 100 (Fanatic)

Ask a well-received question [...] and maintain a positive question record

- on 5 separate days (Curious)
- on 30 separate days (Inquisitativ)
- on 100 separate days (Socratic)

Complete at least [...]. This badge is awarded once per review type

• one review task (Custodian)

- 250 review tasks (Reviewer)
- 1000 review tasks (Steward)

Edit and answer [...] (both actions within 12 hours, answer score > 0)

- 1 question (Explainer)
- 50 questions (Refiner)
- 500 questions (Illuminator)

Edit

- first post (Editor)
- 80 posts (Strunk & White)
- 500 posts (Copy Editor)

Edit [...] that was inactive for 6 months

- first post (Excavator)
- 100 posts (Archeologist)

0.3 Infos on Tags

What are those tags related with the r-tag?

name	count
r	238811
ggplot2	22019
dataframe	15099
shiny	11812
plot	10107
dplyr	9061
data.table	7232
matrix	5461
loops	4065
regex	3982
function	3904
for-loop	3758
list	3689
rstudio	3651
time-series	3352
statistics	3252
knitr	3131
r-markdown	2957
subset	2925
csv	2906
python	2743
date	2519
string	2489
vector	2422
regression	2250
apply	2110
plyr	1992
merge	1922
lapply	1810
if-statement	1733
igraph	1675

name	count
plotly	1535
machine-learning	1509
xts	1490
rcpp	1461
aggregate	1425
graph	1392
excel	1338
web-scraping	1323
parallel-processing	1311
reshape	1307
bar-chart	1295
package	1260
na	1248
performance	1222
histogram	1197
xml	1187
html	1171
tidyr	1164
latex	1154

0.4 Get info for specified user

```
> # # user_info <- stack_users(id = 7322615,
 > # #
                                                                                                                         special = "badges",
 > # #
                                                                                                                          pagesize = 100,
 > # #
                                                                                                                         page = 1,
 > # #
                                                                                                                         num_pages = 1000
 > # # )
> # # user_timeline <- stack_users(id = 7322615,
 > # #
                                                                                                                         special = "reputation-history"",
 > # #
                                                                                                                         pagesize = 100,
 > # #
                                                                                                                         page = 1,
 > # #
                                                                                                                         num_pages = 1000
 > # # )
> #
> # # user_timeline1 <- combine_url(base = "users",
> # # id = 7322615,
> # #
                                                      special = "timeline")
>
 > # x1 <- stack_GET("users/id", id = 7322615)
> #
 \verb| > \# x2 <- stack\_GET("http://api.stackexchange.com/docs/timeline-on-users\#ids=7322615\&filter=default\&sit]| | + \# x2 <- \# x
 > #
 >
> # privileges <- stack_privileges()
> # how to find answers
 > # https://stackoverflow.com/search?q=%5Br%5D+lastactive%3A..6m+is%3Aa
 >
 > # what type of actions are allowed with combine_url?
```

```
> action_type <- c("answers", "badges", "comments", "favorites", "mentioned", "network-activity", "post
> url <- stackr:::combine_url(base = "users",
                               id = c(9371451),
                               special = "timeline",
+
                               special_ids = action_type)
> df_user <- stackr:::stack_GET(url, num_pages = 15)</pre>
> badges_consecutive <- stack_badges(c(71,83), "recipients")</pre>
> library("jsonlite")
>
> # I need a loop,
> # whenever json_list[["has_more"]] == TRUE than run annother page
> # this new database rbind with the previous one
> # maybe there is also a direct method after looking at the programm code of stackr
> json_file1 <- "https://api.stackexchange.com/2.2/users/7322615;3566854;1556049;2545523/timeline?page=
> json_list1 <- jsonlite::fromJSON(json_file1, flatten = TRUE)</pre>
> df1 <- as.tibble(json_list1[[1]])</pre>
> df1 <- df1 %>% mutate(created = with tz(as.POSIXct(creation date, origin = "1970-01-01"), "UTC"))
> # df2 <-
> #
            data.frame(matrix(unlist(json_list2[[2]]), nrow = 10, byrow = T), stringsAsFactors =
> #
            FALSE)
> # json_data <- json_data[[1]] <- NULL
> # df <- as.tibble(json_data)
```