141B - SQL

Benjamin Jewell

2023-05-19

Load Libraries & make Connection

```
library(RSQLite)
library(DBI)
library(knitr)

db = dbConnect(SQLite(), "E:\\College\\UC Davis\\STA141B\\HW3\\stats.stackexchange.db")
```

Data Exploration:

Here we print out the schema of tables and their columns, since it's different than the website schema.

```
#list all tables for easy viewing
dbListTables(db)
```

```
##  [1] "BadgeClassMap"     "Badges"            "CloseReasonMap"
##  [4] "Comments"          "LinkTypeMap"       "PostHistory"
##  [7] "PostHistoryTypeId" "PostLinks"         "PostTypeIdMap"
## [10] "Posts"             "TagPosts"          "Users"
## [13] "VoteTypeMap"       "Votes"
```

```
db_summary = as.data.frame(do.call(cbind, lapply(dbListTables(db), function(x) dbListFields(db, x)[1:22])))
colnames(db_summary) = dbListTables(db)
db_summary
```

```
##    BadgeClassMap   Badges CloseReasonMap        Comments LinkTypeMap
## 1             id       Id             id              Id          id
## 2          value   UserId          value          PostId       value
## 3           <NA>     Name           <NA>           Score        <NA>
## 4           <NA>     Date           <NA>            Text        <NA>
## 5           <NA>    Class           <NA>    CreationDate        <NA>
## 6           <NA> TagBased           <NA>          UserId        <NA>
## 7           <NA>     <NA>           <NA>  ContentLicense        <NA>
## 8           <NA>     <NA>           <NA> UserDisplayName        <NA>
## 9           <NA>     <NA>           <NA>            <NA>        <NA>
## 10          <NA>     <NA>           <NA>            <NA>        <NA>
## 11          <NA>     <NA>           <NA>            <NA>        <NA>
## 12          <NA>     <NA>           <NA>            <NA>        <NA>
## 13          <NA>     <NA>           <NA>            <NA>        <NA>
## 14          <NA>     <NA>           <NA>            <NA>        <NA>
## 15          <NA>     <NA>           <NA>            <NA>        <NA>
## 16          <NA>     <NA>           <NA>            <NA>        <NA>
## 17          <NA>     <NA>           <NA>            <NA>        <NA>
## 18          <NA>     <NA>           <NA>            <NA>        <NA>
## 19          <NA>     <NA>           <NA>            <NA>        <NA>
## 20          <NA>     <NA>           <NA>            <NA>        <NA>
## 21          <NA>     <NA>           <NA>            <NA>        <NA>
## 22          <NA>     <NA>           <NA>            <NA>        <NA>
##          PostHistory PostHistoryTypeId     PostLinks PostTypeIdMap
## 1                 Id                Id            Id            id
## 2  PostHistoryTypeId             Label  CreationDate         value
## 3             PostId       Description        PostId          <NA>
## 4       RevisionGUID              <NA> RelatedPostId          <NA>
## 5       CreationDate              <NA>    LinkTypeId          <NA>
## 6             UserId              <NA>          <NA>          <NA>
## 7               Text              <NA>          <NA>          <NA>
## 8     ContentLicense              <NA>          <NA>          <NA>
## 9            Comment              <NA>          <NA>          <NA>
## 10   UserDisplayName              <NA>          <NA>          <NA>
## 11              <NA>              <NA>          <NA>          <NA>
## 12              <NA>              <NA>          <NA>          <NA>
## 13              <NA>              <NA>          <NA>          <NA>
## 14              <NA>              <NA>          <NA>          <NA>
## 15              <NA>              <NA>          <NA>          <NA>
## 16              <NA>              <NA>          <NA>          <NA>
## 17              <NA>              <NA>          <NA>          <NA>
## 18              <NA>              <NA>          <NA>          <NA>
## 19              <NA>              <NA>          <NA>          <NA>
## 20              <NA>              <NA>          <NA>          <NA>
## 21              <NA>              <NA>          <NA>          <NA>
## 22              <NA>              <NA>          <NA>          <NA>
##                       Posts TagPosts          Users VoteTypeMap        Votes
## 1                        Id       Id             Id          id           Id
## 2                PostTypeId      Tag     Reputation       value       PostId
## 3          AcceptedAnswerId     <NA>   CreationDate        <NA>   VoteTypeId
## 4              CreationDate     <NA>    DisplayName        <NA> CreationDate
## 5                     Score     <NA> LastAccessDate        <NA>       UserId
## 6                 ViewCount     <NA>     WebsiteUrl        <NA> BountyAmount
## 7                      Body     <NA>       Location        <NA>         <NA>
## 8               OwnerUserId     <NA>        AboutMe        <NA>         <NA>
## 9          LastActivityDate     <NA>          Views        <NA>         <NA>
## 10                    Title     <NA>        UpVotes        <NA>         <NA>
## 11                     Tags     <NA>      DownVotes        <NA>         <NA>
## 12              AnswerCount     <NA>      AccountId        <NA>         <NA>
## 13             CommentCount     <NA>           <NA>        <NA>         <NA>
## 14           ContentLicense     <NA>           <NA>        <NA>         <NA>
## 15 LastEditorDisplayName     <NA>           <NA>        <NA>         <NA>
## 16             LastEditDate     <NA>           <NA>        <NA>         <NA>
## 17           LastEditorUserId     <NA>           <NA>        <NA>         <NA>
## 18        CommunityOwnedDate     <NA>           <NA>        <NA>         <NA>
## 19                 ParentId     <NA>           <NA>        <NA>         <NA>
## 20         OwnerDisplayName     <NA>           <NA>        <NA>         <NA>
## 21               ClosedDate     <NA>           <NA>        <NA>         <NA>
## 22            FavoriteCount     <NA>           <NA>        <NA>         <NA>
```

General Note: I will occasionally leave sections of commented code out here. These are highlighting the tables I used to "build up" to the final query I used.

Answered Questions: {1,2,3,4,5,8,9,10,11,12,14,15,16}(13/13)

QUESTION 1: How many users are there?

It was found that there are 321,677 users, by counting the number of Ids from the Users table.

```
kable(dbGetQuery(db, 'SELECT COUNT(Id) FROM Users'), col.names = 'Number of users:')
```

Number of users:
321677

QUESTION 2: How many users joined since 2020? (Hint: Convert the CreationDate to a year.)

We found that 100796 users have joined from after the first second of January 1st, 2020. This was found by counting the CreationDate from the Users table where their CreationDate was greater than the start of 2020.

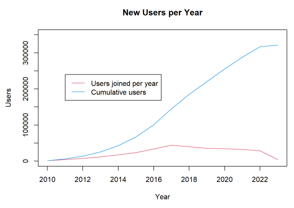
```
kable(dbGetQuery(db, 'SELECT COUNT(CreationDate) FROM Users WHERE JULIANDAY(CreationDate) > JULIANDAY("2020-01-01T00:00:0
1")'), col.names = "Number of users that have joined since the start of 2020")
```

Number of users that have joined since the start of 2020
100796

QUESTION 3: How many users joined each year? Describe this with a plot, commenting on any anomalies.

The CreationDate of each Users was selected and grouped by, counting how many joined each year. I calculated both the amount that joined each year and the cumulative sum so we can see the total number of users per year. It seems like the peak growth was in 2017 and has slowly been declining. I don't know enough about the statistics/stackoverflow community to comment on why it's slowly been going down other than to guess that as the product ages less people join. As for the large dip at the end however that is like explained because our table only has data up through March 5th of 2023, so it doesn't have a full year's worth of users joining. I imagine if you redid this with the data on December 31st this year there would be much less of a drastic decrease.

```
join_df = dbGetQuery(db, 'SELECT strftime("%Y", CreationDate) AS yr, COUNT(*) AS yr_count FROM Users GROUP BY yr')

plot(join_df, type='l', xlab = 'Year', ylab = 'Users', main='New Users per Year', ylim = c(0,350000), col=2)
join_df[,"Cumulative Users"] = cumsum(join_df$yr_count) #[2] Cumsum new column
lines(join_df$yr, join_df$`Cumulative Users`, col=4)
legend(2011,240000, legend=c('Users joined per year', 'Cumulative users'), col=c(2,4), lty=c(1,1))
```



```
kable(join_df, col.names = c('Year', 'New Users', 'Cumm. New Users'))
```

Year	New Users	Cumm. New Users
2010	1668	1668
2011	4396	6064
2012	7450	13514
2013	11846	25360
2014	17809	43169
2015	24012	67181
2016	33753	100934
2017	44416	145350
2018	40040	185390
2019	35491	220881
2020	34617	255498
2021	32765	288263
2022	28801	317064
2023	4613	321677

QUESTION 4: How many different types of posts are there in the Posts table?

According to the PostTypeIdMap table there are 8 kinds of posts. However here where we sorted the posts by their PostTypeId and used COUNT to count them, only 7 of the 8 are present in this data. We used a join with PostTypeIdMap to print their appropriate types.

```
kable(dbGetQuery(db, 'SELECT PostTypeID, COUNT(*) As Count, PostTypeIdMap.value FROM Posts INNER JOIN PostTypeIdMap ON Post
s.PostTypeId = PostTypeIdMap.id GROUP BY PostTypeID'), col.names = c('Id', 'Count', 'Value'))
```

Id	Count	Value
1	204370	Question
2	197928	Answer
3	6	Orphaned tag wiki
4	1444	Tag wiki excerpt
5	1444	Tag wiki
6	23	Moderator nomination
7	5	Wiki placeholder (seems to only be the election description)

QUESTION 5: How many posted questions are there?

There are 204,370 questions asked in our data. We counted all posts where the PostTypeId = 1, which indicates it's a question and not something like an answer. Here we demonstrate that using PostTypeId = 1 mapped to the 'Question' Id from PostTypeIdMap gives the same results.

```
#PostTypeId = 1 : Question Type
kable(dbGetQuery(db, 'SELECT COUNT(PostTypeID) FROM Posts WHERE PostTypeID = 1'), col.names = 'Number of Questions Asked')
```

Number of Questions Asked
204370

```
kable(dbGetQuery(db, 'SELECT COUNT(PostTypeID) FROM Posts WHERE PostTypeID IN (SELECT id FROM PostTypeIdMap WHERE value = "Q
uestion")'),
      col.names = 'Number of Questions Asked via PostTypeIdMap')
```

Number of Questions Asked via PostTypeIdMap
204370

QUESTION 8: How many answers are there?

There are 197,928 answers in our data. We counted all posts where the PostTypeId = 2, which indicates it's an answer and not something like a question.

```
kable(dbGetQuery(db, 'SELECT COUNT(PostTypeID) FROM Posts WHERE PostTypeID = 2'), col.names = 'Number of Answers')
```

Number of Answers
197928

```
kable(dbGetQuery(db, 'SELECT COUNT(PostTypeID) FROM Posts WHERE PostTypeID IN (SELECT id FROM PostTypeIdMap WHERE value = "A
nswer")'), col.names = 'Number of Answers via PostTypeIdMap')
```

Number of Answers via PostTypeIdMap
197928

```
kable(dbGetQuery(db, 'SELECT SUM(AnswerCount) FROM Posts'), col.names = 'Verifying via Summing AnswerCount of all posts')
```

Verifying via Summing AnswerCount of all posts
197928

QUESTION 9: What's the most recent question (by date-time) in the Posts table?

URL: https://stats.stackexchange.com/questions/608405 (https://stats.stackexchange.com/questions/608405)

ALT URL: https://stats.stackexchange.com/questions/608458/are-there-any-methods-that-combine-mcmc-and-vi (https://stats.stackexchange.com/questions/608458/are-there-any-methods-that-combine-mcmc-and-vi)

How would we map a question in posts table to it's URL:

Normally we could append the Post ID to https://stats.stackexchange.com/questions/%7BId%7D (https://stats.stackexchange.com/questions/%7BId%7D) but it seems like not all posts exist. For example, https://stats.stackexchange.com/questions/608405 (https://stats.stackexchange.com/questions/608405) doesn't exist and actually exist. However using Google with the contents of the post I was able to find a post that matched half of the new post and decided that would be more useful to us. You can't do much if the newest post just throws a 404 error at you.

To find this we found the largest CreationDate to find the newest post is at 5:10:18 AM on March 5th, 2023, with ID 608405. I wasn't certain how to link them to the URL but figured there must be a way. In fact it turns out if you type in 'https://stats.stackexchange.com/questions/ (https://stats.stackexchange.com/questions/)' and then append the Post Id it will take you to that post. While this works for a lot of the Question IDS, it doesn't work for post 608405, as it just gives you a '404 Page not found' error. However if you search Google using the text from the body you can find a post that occurs later after our data set, Post #608458 which matches half of the body of Post #608405. Ultimately together these two posts give you the newest post in our data set, with one answer sticking to the "letter of the law", and another giving an answer that is more the "spirit of the law".

```
kable(dbGetQuery(db, 'SELECT MAX(CreationDate) FROM Posts'), col.names = 'Newest Post Time')
```

Newest Post Time
2023-03-05T05:10:18.393

```
kable(dbGetQuery(db, 'SELECT Id FROM Posts WHERE CreationDate = (SELECT MAX(CreationDate) FROM Posts)'), col.names = 'Newest
Post ID')
```

Newest Post ID
608405

```
dbGetQuery(db, 'SELECT Body FROM Posts WHERE CreationDate = "2023-03-05T05:10:18.393"')[[1]]
```

```
## [1] "<p>Are there any methods that combine VI and MCMC? If it exists, why isn’t it used prominently over techniques such
as NUTS or other VIs.</p>\\\\n<p>Another question that just popped up in my head why is HMC sampling parallelizable after st
ationary? Is it related with some kind of sequential bottlenecks?</p>\\\\n"
```

QUESTION 10: Top 10 Questions Posted Users: How many questions did they post, usernames, when did they join, reputation, country?

For Username we use Users.DisplayName, for their join date we use Users.CreationDate, for their reputation we use Users.Reputation and for country we use Users.Location (not all people list a country, but many do. As there's no Country value this seemed most appropriate.)

We make a table with the top 10 users and how many posts they make, using an Inner Join with the Users table to filter to only having the to 10 question posters.

```
#Double Checking the way we select the number of posts per user is correct
ct = dbGetQuery(db, 'SELECT OwnerDisplayName, OwnerUserId, COUNT(*) FROM Posts GROUP BY OwnerUserId ')
```

```
## Warning: Column `OwnerUserId`: mixed type, first seen values of type integer,
## coercing other values of type string
```

```
#There are 405220 posts, and we have that many here!
sum(ct$`COUNT(*)`)
```

```
## [1] 405220
```

```
#We ignore user28 since they have no ID so we can't look up any results on them
dbGetQuery(db, 'SELECT OwnerDisplayName, OwnerUserId, COUNT(*) FROM Posts WHERE OwnerUserId <>"" GROUP BY OwnerUserId ORDER
BY COUNT(OwnerUserId) DESC LIMIT 10;')
```

```
##    OwnerDisplayName OwnerUserId COUNT(*)
## 1                           805     4850
## 2                           919     3152
## 3                           686     3080
## 4                         28500     2703
## 5                         11887     2638
## 6                         35989     2546
## 7                        173082     2297
## 8                          1352     2238
## 9                         85665     2190
## 10                         7290     1907
```

```
#dbGetQuery(db, 'SELECT Id, DisplayName, CreationDate, Reputation, Location FROM Users WHERE Id IN (SELECT OwnerUserId FROM
Posts WHERE OwnerUserId <>"" GROUP BY OwnerUserId ORDER BY COUNT(OwnerUserId) DESC LIMIT 10);')

kable(dbGetQuery(db, 'SELECT Id, DisplayName, CreationDate, Reputation, Location, Freq
                 FROM Users
                 INNER JOIN (SELECT OwnerUserId, COUNT(*) as Freq FROM Posts WHERE OwnerUserId <>"" GROUP BY OwnerUserId ORD
ER BY COUNT(OwnerUserId) DESC LIMIT 10) AS Table2 ON Table2.OwnerUserId = Users.Id
                 WHERE Id IN (SELECT OwnerUserId FROM Posts WHERE OwnerUserId <>"" GROUP BY OwnerUserId ORDER BY COUNT(Owner
UserId) DESC LIMIT 10);'),
      col.names = c('Id', 'Username', 'Join Date', 'Reputation', 'Location', 'Post Count'))
```

Id	Username	Join Date	Reputation	Location	Post Count
805	Glen_b	2010-08-07T08:40:07.287	268966	I'm right here	4850
919	whuber	2010-08-13T15:29:47.140	304878		3152
686	Peter Flom	2010-08-03T19:42:40.907	97091	New York, NY	3080
28500	EdM	2013-07-26T15:11:03.380	76431		2703
11887	kjetil b halvorsen	2012-06-09T22:52:37.473	70519	Bolivia	2638
35989	Tim	2013-12-10T21:19:06.223	126510	Warsaw, Poland	2546
173082	Ben	2017-08-10T03:27:26.793	108435	Canberra, Australia	2297
1352	Stephan Kolassa	2010-09-18T10:55:08.240	107905	Switzerland	2238
85665	BruceET	2015-08-11T17:22:01.590	53074	San Francisco Bay Area	2190
7290	gung - Reinstate Monica	2011-11-09T04:43:15.613	138999	Kingdom of Zhao	1907

QUESTION 11: Following from the previous questions, for the 10 users who posted the most questions, how many gold, silver and bronze badges does each of these 10 individuals have?

Here we made an Inner Join between Badges.UserId and the Id of the top 10 users, as we found them in the previous question. We group by their UserId and the Medal Classes to give our table, and use Count() to see how many of each they have.

```
kable(dbGetQuery(db, 'SELECT UserId, Class, COUNT(*)
                FROM Badges
                 INNER JOIN (SELECT OwnerUserId, COUNT(*) as Freq FROM Posts WHERE OwnerUserId <>"" GROUP BY OwnerUserId ORD
ER BY COUNT(OwnerUserId) DESC LIMIT 10) AS tp10 ON Badges.UserId = tp10.OwnerUserId
                 GROUP BY UserID, Class'), col.names = c('User Id', 'Medal Class', 'Frequency'))
```

User Id	Medal Class	Frequency
686	1	35
686	2	153
686	3	293
805	1	36
805	2	587
805	3	982
919	1	56
919	2	693
919	3	1193
1352	1	17
1352	2	215
1352	3	406
7290	1	85
7290	2	379
7290	3	676
11887	1	30
11887	2	160
11887	3	522
28500	1	9
28500	2	79
28500	3	230
35989	1	21
35989	2	240
35989	3	456
85665	1	2
85665	2	32
85665	3	88
173082	1	4
173082	2	193
173082	3	452

QUESTION 12: For each of the following terms, how many questions contain that term: Regression, ANOVA, Data Mining, Machine Learning, Deep Learning, Neural Network.

There are 54567 questions with these terms, either in the body or title. Here we use the LIKE term to look for how many posts contain the above key terms. The text of a question is found in Posts.Body, and the title via Posts.Title and we filter only using questions by only look at posts where PostTypeId = 1, marking it as a question.

```
#PostTypeId = 1 indicates that it is a question
kable(dbGetQuery(db, "SELECT COUNT(Body)
                 FROM Posts
                 WHERE (
                 (Body LIKE '%Regression%') OR (Body LIKE '%ANOVA%') OR (Body LIKE '%Data Mining%') OR (Body LIKE '%Machine
Learning%') OR (Body LIKE '%Deep Learning%') OR (Body LIKE '% Neural Network%')
                 OR
                 (Title LIKE '%Regression%') OR (Title LIKE '%ANOVA%') OR (Title LIKE '%Data Mining%') OR (Title LIKE '%Mach
ine Learning%') OR (Title LIKE '%Deep Learning%') OR (Title LIKE '% Neural Network%'))
                 AND Posts.PostTypeId = 1"),
      col.names = 'Number of Questions Containing the terms')
```

Number of Questions Containing the terms
58567

QUESTION 14: What is the date range for the questions and answers in this database?

Here we use MIN and MAX to look at the oldest and newest questions in our database, filtering by PostTypeId as appropriate. Because we are using JULIANDAY to calculate this, our differences are given in Days.

```
#Questions
kable(dbGetQuery(db, 'SELECT MAX(CreationDate) FROM Posts WHERE PostTypeId = 1'), col.names = 'Newest Question Time')
```

Newest Question Time
2023-03-05T05:10:18.393

```
kable(dbGetQuery(db, 'SELECT MIN(CreationDate) FROM Posts WHERE PostTypeId = 1'), col.names = 'Oldest Question Time')
```

Oldest Question Time
2009-02-02T14:21:12.103

```
#Answers
kable(dbGetQuery(db, 'SELECT MAX(CreationDate) FROM Posts WHERE PostTypeId = 2'), col.names = 'Newest Answer Time')
```

Newest Answer Time
2023-03-05T04:48:34.853

```
kable(dbGetQuery(db, 'SELECT MIN(CreationDate) FROM Posts WHERE PostTypeId = 1'), col.names = 'Oldest Answer Time')
```

Oldest Answer Time
2009-02-02T14:21:12.103

```
#Difference [1]
kable(dbGetQuery(db, 'SELECT JULIANDAY(MAX(CreationDate)) - JULIANDAY(MIN(CreationDate)) FROM Posts WHERE PostTypeId = 1'),
col.names = 'Number of days between oldest and newest question')
```

Number of days between oldest and newest question
5143.617

```
kable(dbGetQuery(db, 'SELECT JULIANDAY(MAX(CreationDate)) - JULIANDAY(MIN(CreationDate)) FROM Posts WHERE PostTypeId = 2'),
col.names = 'Number of days between oldest and newest answer')
```

Number of days between oldest and newest answer
5143.6

QUESTION 15: What Question has the most comments associated with it? How many answers are there?

Here we look at the Post (#386853) with the MAX CommentCount associated with it. We do a small amount of fiddling in R to make our Answer Count display as 0 below rather than NA, as this post has no answer.

```
q15 = dbGetQuery(db, 'SELECT Id, CommentCount, AnswerCount FROM Posts WHERE CommentCount = (SELECT MAX(CommentCount) FROM Po
sts)')
#Formatting:
if (q15[1,3] == ''){q15[1,3] = 0}
q15$AnswerCount = as.numeric(q15$AnswerCount)
kable(q15, col.names = c('Post Id', 'Comment Count', 'Answer Count'))
```

Post Id	Comment Count	Answer Count
386853	66	0

QUESTION 16: How many comments are there across all posts? How many posts have a comment? What is the distribution of comments per question?

This graph looks like it is likely Poisson or Exponentially distributed. This makes sense as it's more likely a post has a few comments then a huge amount of comments. We looked at the total CommentCount from the Posts table to find how many comments there are, making sure to filter out empty comments (aka no comments).

```
kable(dbGetQuery(db, 'SELECT SUM(CommentCount) FROM Posts WHERE CommentCount IS NOT NULL'), col.names = 'Total Comments')
```

Total Comments
768069

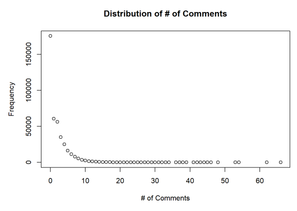
```
kable(dbGetQuery(db, 'SELECT COUNT(PostId) FROM Comments'), col.names = 'Total Comments via Comment Table')
```

Total Comments via Comment Table
768069

```
kable(dbGetQuery(db, 'SELECT COUNT(CommentCount) FROM Posts WHERE CommentCount IS NOT NULL'), col.names = 'Total Posts with
1+ Comment(s)')
```

Total Posts with 1+ Comment(s)
405220

```
plot(dbGetQuery(db, 'SELECT CommentCount, COUNT(CommentCount) FROM Posts WHERE CommentCount IS NOT NULL GROUP BY CommentCoun
t'), xlab = '# of Comments', ylab= 'Frequency', main='Distribution of # of Comments')
```



=== **REQUIRE QUESTIONS** ===

QUESTION 21: Compute the table that contains: Question, Username of poster, when that user joined, their location, the date the question was first posted, accepted answer, when accepted answer was posted, name of the user who posted accepted answer

For this question we had to combine several variables together to create our table. Here is a summary of what I values I thought we'd need: - Posts.Title - OwnerDisplayName - Users.CreationDate - Users.Location - CreationDate - AcceptedAnswerId –> Posts.Body - AcceptedAnswerId –> CreationDate - AcceptedAnswerId –> OwnerDisplayName

NOTE: Due to the size of table I used head() here to limit the size and allow my computer to actually knit the file. The query will work without it though if you want to see the whole table!

```
#Title, OwnerDisplayName, Users.CreationDate, Users.Location, CreationDate, AcceptedAnswerId-->Posts.Body, AcceptedAnswerId
-->CreationDate, AcceptedAnswerId-->OwnerDisplayName

# dbGetQuery(db, 'SELECT p1.OwnerUserId, PUser.Id, p1.AcceptedAnswerId, ans.Id, ans.OwnerUserId, AUser.Id
#           FROM Posts as p1
#           INNER JOIN Users As PUser ON p1.OwnerUserId = PUser.Id
#           INNER JOIN Users AS AUser ON ans.OwnerUserId = AUser.Id
#           INNER JOIN Posts AS ans ON p1.AcceptedAnswerId = ans.Id
#           ')

head(dbGetQuery(db, 'SELECT p1.Title, PUser.DisplayName AS PosterUsername, PUser.CreationDate AS PosterJoinDate, PUser.Locat
ion AS PosterLocation, p1.CreationDate AS PostDate, ans.Body AS Answer, ans.CreationDate AS AnswerDate, AUser.DisplayName AS
AnswerUsername
          FROM Posts as p1
          INNER JOIN Users As PUser ON p1.OwnerUserId = PUser.Id
          INNER JOIN Users AS AUser ON ans.OwnerUserId = AUser.Id
          INNER JOIN Posts AS ans ON p1.AcceptedAnswerId = ans.Id
          ORDER BY PostDate
          '))
```

```
##                                                              Title
## 1       What do you call an average that does not include outliers?
## 2                                     Eliciting priors from experts
## 3                                                What is normality?
## 4 What are some valuable Statistical Analysis open source projects?
## 5        Assessing the significance of differences in distributions
## 6                             Locating freely available data samples
##   PosterUsername          PosterJoinDate             PosterLocation
## 1         Tawani 2016-04-19T11:19:46.257
## 2    csgillespie 2010-07-19T19:04:52.280   Newcastle, United Kingdom
## 3         A Lion 2010-07-19T19:09:32.157
## 4         grokus 2010-07-19T19:08:29.070              United States
## 5    Jay Stevens 2010-07-19T19:09:16.917       Jacksonville, FL, USA
## 6         EAMann 2010-07-19T19:11:57.393 Tualatin, OR, United States
##                  PostDate
## 1 2009-02-02T14:21:12.103
## 2 2010-07-19T19:12:12.510
## 3 2010-07-19T19:12:57.157
## 4 2010-07-19T19:13:28.577
## 5 2010-07-19T19:13:31.617
## 6 2010-07-19T19:15:59.303
##
Answer
## 1
<p>It's called the <a href="https://wikipedia.org/wiki/Truncated_mean" rel="noreferrer">trimmed mean</a>.  Basically what yo
u do is compute the mean of the middle 80% of your data, ignoring the top and bottom 10%. Of course, these numbers can vary,
but that's the general idea.</p>\\\\n
## 2
<p>John Cook gives some interesting recommendations. Basically, get percentiles/quantiles (not means or obscure scale parame
ters!) from the experts, and fit them with the appropriate distribution.</p>\\\\n\\\\n<p><a href="http://www.johndcook.com/b
log/2010/01/31/parameters-from-percentiles/">http://www.johndcook.com/blog/2010/01/31/parameters-from-percentiles/</a></p
>\\\\n
## 3 <p>The assumption of normality is just the supposition that the underlying <a href="http://en.wikipedia.org/wiki/Random
_variable" rel="nofollow noreferrer">random variable</a> of interest is distributed <a href="http://en.wikipedia.org/wiki/No
rmal_distribution" rel="nofollow noreferrer">normally</a>, or approximately so.  Intuitively, normality may be understood as
the result of the sum of a large number of independent random events.</p>\\\\n<p>More specifically, normal distributions are
defined by the following function:</p>\\\\n<p><span class="math-container">$$ f(x) =\\frac{1}{\\sqrt{2\\pi\\sigma^2}}e^{ -\\
frac{(x-\\mu)^2}{2\\sigma^2} },$$</span></p>\\\\n<p>where <span class="math-container">$\\mu$</span> and <span class="math-c
ontainer">$\\sigma^2$</span> are the mean and the variance, respectively, and which appears as follows:</p>\\\\n<p><img src
="https://upload.wikimedia.org/wikipedia/commons/thumb/7/74/Normal_Distribution_PDF.svg/300px-Normal_Distribution_PDF.svg.pn
g" alt="alt text" /></p>\\\\n<p>This can be checked in <a href="http://en.wikipedia.org/wiki/Normality_test" rel="nofollow n
oreferrer">multiple ways</a>, that may be more or less suited to your problem by its features, such as the size of n.  Basic
ally, they all test for features expected if the distribution were normal (e.g. expected <a href="http://en.wikipedia.org/wi
ki/Q-Q_plot" rel="nofollow noreferrer">quantile distribution</a>).</p>\\\\n
## 4
<p>The R-project</p>\\\\n\\\\n<p><a href="http://www.r-project.org/">http://www.r-project.org/</a></p>\\\\n\\\\n<p>R is valu
able and significant because it was the first widely-accepted Open-Source alternative to big-box packages.  It's mature, wel
l supported, and a standard within many scientific communities.</p>\\\\n\\\\n<ul>\\\\n<li><a href="http://www.inside-r.org/w
hy-use-r">Some reasons why it is useful and valuable</a> </li>\\\\n<li>There are some nice tutorials <a href="http://getting
geneticsdone.blogspot.com/search/label/ggplot2">here</a>.</li>\\\\n</ul>\\\\n
## 5
<p>I believe that this calls for a <a href="http://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/ks2samp.htm">t
wo-sample Kolmogorov-Smirnov test</a>, or the like.  The two-sample Kolmogorov-Smirnov test is based on comparing difference
s in the <a href="http://en.wikipedia.org/wiki/Empirical_distribution_function">empirical distribution functions</a> (ECDF)
of two samples, meaning it is sensitive to both location and shape of the the two samples.  It also generalizes out to a mul
tivariate form.</p>\\\\n\\\\n<p>This test is found in various forms in different packages in R, so if you are basically prof
icient, all you have to do is install one of them (e.g. <a href="http://cran.r-project.org/web/packages/fBasics/fBasics.pdf"
>fBasics</a>), and run it on your sample data.</p>\\\\n
## 6
<p>Also see the UCI machine learning Data Repository.</p>\\\\n\\\\n<p><a href="http://archive.ics.uci.edu/ml/">http://archiv
e.ics.uci.edu/ml/</a></p>\\\\n
##                AnswerDate AnswerUsername
## 1 2009-02-02T14:24:31.740        dsimcha
## 2 2010-07-19T19:19:46.160         Harlan
## 3 2010-07-19T19:43:20.423 John L. Taylor
## 4 2010-07-19T19:14:43.050    Jay Stevens
## 5 2010-07-19T21:36:12.850 John L. Taylor
## 6 2010-07-19T19:24:18.580 Stephen Turner
```

QUESTION 22: Determine the users that have only posted questions and never answered a question? (Compute the table containing the number of questions, number of answers and the user's login name for this group.) How many are there?

This question made me think I would need to compute a table of everyone who has posted a question. Then I would find the table of everyone who answered a question and drop everyone on the answer table from the questions table. The variables I saw as pertinent to this were: - OwnerUserId (Via GROUP BY OwnerUserId) - OwnerDisplayName - COUNT(Questions per User) - COUNT(Answers per User) <– If we did things right this should be NA or 0.

NOTE: Due to the size of table I used head() here to limit the size and allow my computer to actually knit the file. The query will work without it though if you want to see the whole table!

```
# q22 = dbGetQuery(db, 'SELECT Users.Id, Users.DisplayName, COUNT(Posts.PostTypeId) AS AnswerCount
#            FROM Posts
#            INNER JOIN Users ON Posts.OwnerUserId = Users.Id
#            WHERE Posts.PostTypeId = 2
#            GROUP BY Users.Id')
#q22

# q22b = dbGetQuery(db, 'SELECT Users.Id, Users.DisplayName, COUNT(Posts.PostTypeId) AS QCount
#            FROM Posts
#            INNER JOIN Users ON Posts.OwnerUserId = Users.Id
#            WHERE Posts.PostTypeId = 1
#            GROUP BY Users.Id
#            ')
#q22b

#Test to make sure we are selecting overlapping Users
#intersect(q22$Id, q22b$Id)

q22c = dbGetQuery(db, 'SELECT QstC.Id, QstC.DisplayName, QstC.QCount AS QuestionsAnswered, AnsC.AnswerCount
               FROM (SELECT Users.Id, Users.DisplayName, COUNT(Posts.PostTypeId) AS AnswerCount
          FROM Posts
          INNER JOIN Users ON Posts.OwnerUserId = Users.Id
          WHERE Posts.PostTypeId = 2
          GROUP BY Users.Id) AS AnsC
               RIGHT JOIN (SELECT Users.Id, Users.DisplayName, COUNT(Posts.PostTypeId) AS QCount
          FROM Posts
          INNER JOIN Users ON Posts.OwnerUserId = Users.Id
          WHERE Posts.PostTypeId = 1
          GROUP BY Users.Id) AS QstC
               ON AnsC.Id = QstC.Id
           WHERE AnsC.AnswerCount IS NULL
           ORDER BY QstC.Id
               ')
head(q22c)
```

```
##   Id DisplayName QuestionsAnswered AnswerCount
## 1 18      grokus                 2          NA
## 2 24      A Lion                 2          NA
## 3 38      EAMann                 1          NA
## 4 52     Alan H.                13          NA
## 5 53        kyle                 2          NA
## 6 58      Preets                 1          NA
```

```
#sum(q22c[which(is.na(q22c$AnswerCount)),])
```

QUESTION 23: Compute the table with information for the 75 users with the most accepted answers. This table should include: User display name, creation date, location, number of badges they have won (and names of badges), the dates of the earliest and most recent accepted answer (as two fields)(the unique tags for all questions for which they had the accepted answer (as a single string))

This was probably the table that took me the longest to understand how to group together. You can see multiple smaller tables I made below to try to figure out how to group the data. I decided to use Users as my "main" table, since a lot of the values seem to relate to that and we are finding the top 75 users. The hardest part for me was selecting the unique tags per user across questions they answered, as the DISTINCT value seems to apply to all selected columns. I'm sure there are more efficient ways to do this, but creating tables that acted as "keys" to INNER JOIN on made the most sense to me. As you can see by my notes, while some of these values seemed rather easy to calculate, others were much harder:

#DisplayName, CreationDate, Location, COUNT(Badges.UserId = Id), [???NAMES???], MIN(AcceptedAnswerId.OwnerUserId.CreationDate), MAX(AcceptedAnswerId.OwnerUserId.CreationDate), [???UNIQUE TAGS???]

For Names I ended up using GROUP_CONCAT to bring together all the badge names from a separate table I made of all the badges each user won.

For Unique Tags I had to use a similar strategy as Names but with the tags.

```
#MAIN TABLE: Users
#DisplayName, CreationDate, Location, COUNT(Badges.UserId = Id), [???NAMES???], MIN(AcceptedAnswerId.OwnerUserId.CreationDat
e), MAX(AcceptedAnswerId.OwnerUserId.CreationDate), [???UNIQUE TAGS???]

#User 6:Harlan has 2 accepted answers via manual checking

kable(dbGetQuery(db, 'SELECT Users.Id, Users.DisplayName, Users.CreationDate, Users.Location, Badges.BadgesWon, Badges.Badge
List, Top75.EarliestAnswer, Top75.LatestAnswer, TT.UnqTags
          FROM Users
          INNER JOIN Posts AS p1 ON Users.Id = p1.OwnerUserId
          INNER JOIN Posts AS p2 ON p1.Id = p2.AcceptedAnswerId
          INNER JOIN (SELECT OwnerUserId, COUNT(*) AS Freq, MIN(CreationDate) AS EarliestAnswer, MAX(CreationDate) AS Lates
tAnswer
               FROM Posts
               WHERE Id IN (SELECT AcceptedAnswerId FROM Posts WHERE (NOT AcceptedAnswerId = "") OR (NOT AcceptedAnswerId I
S NULL)) AND OwnerUserId <> ""
               GROUP BY OwnerUserId
               ORDER BY Freq DESC
               LIMIT 75
               ) AS Top75 ON Users.Id = Top75.OwnerUserId
         INNER JOIN (
               SELECT u.Id, COUNT(*) AS BadgesWon, GROUP_CONCAT(DISTINCT Badges.Name) AS BadgeList
               FROM Users AS u
               INNER JOIN Badges ON u.Id = Badges.UserId
               GROUP BY u.Id
               ) AS Badges ON Users.Id = Badges.Id
       LEFT JOIN (
               SELECT DISTINCT p.Id, t.Id, p.AcceptedAnswerId, a.Id, a.OwnerUserId, GROUP_CONCAT(DISTINCT t.Tag) AS UnqTags
               FROM Posts AS p
               INNER JOIN TagPosts AS t ON t.Id = p.Id
               INNER JOIN Posts AS a ON p.AcceptedAnswerId = a.Id
               WHERE p.AcceptedAnswerId <> ""
               GROUP BY a.OwnerUserId
               ) AS TT ON TT.OwnerUserId = Users.Id
       GROUP BY Users.Id
       ORDER BY Users.Id
'))
```

Id	DisplayName	CreationDate	Location	BadgesWon	BadgeList	EarliestAnswer	LatestAnswer	Un
159	Rob Hyndman	2010-07-19T23:05:39.653	Melbourne, Australia	346	Autobiographer,Supporter,Organizer,Teacher,Student,Editor,Mortarboard,Scholar,Critic,Self-Learner,Commentator,Nice Answer,Cleanup,Beta,Nice Question,Good Answer,Enthusiast,Disciplined,Popular Question,Civic Duty,Enlightened,Announcer,Revival,Talkative,Precognitive,Convention,Quorum,Strunk & White,Tag Editor,time-series,r,Citizen Patrol,Synonymizer,Yearling,Excavator,Notable Question,Taxonomist,Good Question,regression,Pundit,Caucus,Constituent,forecasting,Custodian,Famous Question,Great Answer,Favorite Question,Booster,arima,Constable,Generalist,Fanatic,Curious,Informed,Explainer,Guru,Great Question,Stellar Question,Publicist,Necromancer,seasonality,exponential-smoothing	2010-07-19T23:37:43.807	2022-12-08T21:17:24.037	sm thu tes rho squ net fun ser dat ave cer reg
686	Peter Flom	2010-08-03T19:42:40.907	New York, NY	481	Teacher,Supporter,Organizer,Student,Scholar,Quorum,Editor,Commentator,Nice Answer,Autobiographer,Necromancer,Yearling,Critic,Enthusiast,Good Answer,Nice Question,regression,Popular Question,Civic Duty,Sportsmanship,Citizen Patrol,Cleanup,r,Strunk & White,Custodian,logistic,Revival,Analytical,Suffrage,Vox Populi,Tag Editor,Excavator,correlation,Pundit,Convention,Reviewer,Generalist,Fanatic,Good Question,Electorate,Proofreader,Mortarboard,Copy Editor,multiple-regression,statistical-significance,hypothesis-testing,Disciplined,Enlightened,Caucus,interaction,Constituent,anova,t-test,Talkative,Promoter,Benefactor,Steward,categorical-data,distributions,Informed,data-transformation,multicollinearity,model-selection,Guru,probability,mean,self-study,modeling,Curious,Inquisitive,Great Answer,Refiner,Explainer,Sheriff,Taxonomist,normal-distribution,data-visualization,Favorite Question,Notable Question,outliers,spss,p-value,pca,interpretation,standard-deviation,generalized-linear-model,Investor,Altruist,descriptive-statistics,mixed-model,Synonymizer,least-squares,Research Assistant,Tumbleweed,regression-coefficients,normality-assumption,inference,mathematical-statistics,Famous Question,linear-model,variance,skewness,aic,nonparametric,ordinal-data,Populist,factor-analysis,machine-learning,likert,terminology,Announcer,median,feature-selection,ancova,sample-size,Deputy,chi-squared-test,Lifejacket	2010-12-28T15:30:54.830	2021-07-06T14:43:40.547	log log ser mo dat dis equ fun dat stra sta sar ind err dat var tab tes the bay allo mo pop rate rec cor sar lea dat mo

Id	DisplayName	CreationDate	Location	BadgesWon	BadgeList	EarliestAnswer	LatestAnswer	Un
805	Glen_b	2010-08-07T08:40:07.287	I'm right here	1605	Teacher,Editor,Supporter,Yearling,Commentator,Critic,Student,Scholar,Nice Answer,Analytical,Custodian,Excavator,Organizer,Enthusiast,Citizen Patrol,Tag Editor,Enlightened,Informed,Suffrage,Vox Populi,Civic Duty,Quorum,Electorate,Mortarboard,Generalist,Sportsmanship,Strunk & White,Synonymizer,regression,Fanatic,hypothesis-testing,Good Answer,r,Revival,Necromancer,Investor,Cleanup,Announcer,Booster,Proofreader,self-study,Pundit,Copy Editor,Altruist,Convention,Reviewer,Research Assistant,distributions,histogram,Self-Learner,Talkative,Autobiographer,statistical-significance,normal-distribution,data-visualization,t-test,probability,Caucus,Constituent,data-transformation,Tumbleweed,Promoter,Benefactor,Nice Question,p-value,Deputy,anova,mathematical-statistics,variance,Popular Question,sample-size,confidence-interval,nonparametric,Guru,mean,Taxonomist,generalized-linear-model,Outspoken,linear-model,Epic,residuals,time-series,correlation,sampling,least-squares,Curious,estimation,Good Question,maximum-likelihood,nonlinear-regression,assumptions,standard-error,Archaeologist,standard-deviation,simulation,Refiner,Explainer,Notable Question,logistic,Steward,Populist,categorical-data,inference,multiple-regression,heteroscedasticity,bayesian,random-variable,modeling,skewness,goodness-of-fit,random-generation,Legendary,median,Great Answer,expected-value,central-limit-theorem,covariance,outliers,gamma-distribution,terminology,proof,matlab,Disciplined,proportion,independence,machine-learning,t-distribution,Famous Question,kurtosis,interpretation,qq-plot,Sheriff,references,regression-coefficients,moments,lm,robust,boxplot,small-sample,Peer Pressure,Illuminator,fitting,computational-statistics,Favorite Question,econometrics,bootstrap,permutation-test,pearson-r,degrees-of-freedom,descriptive-statistics,model-selection,wilcoxon-signed-rank,discrete-data,poisson-regression,unbiased-estimator,prior,spss,python,likelihood,intuition,excel,poisson-distribution,sample,dice,normality-assumption,logarithm,f-test,count-data,conditional-probability,kernel-smoothing,copula,forecasting,monte-carlo,spearman-rho,linear,Lifejacket,paired-data,predictive-models,notation,Publicist,multivariate-analysis,curve-fitting,quantiles,extreme-value,statistical-power,lognormal-distribution,density-function,wilcoxon-mann-whitney-test,exponential-distribution,cumulative-distribution-function,uniform-distribution,negative-binomial-distribution,binomial-distribution,chi-squared-distribution,kolmogorov-smirnov-test,chi-squared-test,kruskal-wallis-test,fishers-exact-test,optimization,estimators,scatterplot,Marshal,error,power-law,z-test,beta-distribution	2010-08-09T00:37:45.090	2023-02-23T22:11:02.110	dis dis dis like dis tra the wh pov me dat sel reg dis dis fun alg dis mo line dis val dis tes dat hoc reg law reg fn, sm info tra reg cur ma line poi inte wa inte ser dis var nlm ma dis tre dev sea des hac
887	Dikran Marsupial	2010-08-12T07:45:05.553		331	Teacher,Supporter,Editor,Student,Commentator,Enthusiast,Critic,Nice Question,Self-Learner,Nice Answer,Necromancer,Enlightened,Popular Question,Yearling,machine-learning,Quorum,Revival,Good Question,cross-validation,Caucus,svm,Good Answer,regression,classification,bayesian,Custodian,Guru,Notable Question,Favorite Question,Sportsmanship,Promoter,Tumbleweed,Benefactor,neural-networks,Curious,Autobiographer,Generalist,feature-selection,Pundit,predictive-models,model-selection,kernel-trick,Civic Duty,logistic,Great Answer,Taxonomist,Disciplined,Famous Question,optimization,Mortarboard,unbalanced-classes,Excavator,r,Fanatic,overfitting,regularization	2010-08-16T09:07:50.237	2022-09-11T13:16:29.027	ma car mo cla cor sar reg inte sar dis lea fun cor kap
919	whuber	2010-08-13T15:29:47.140		1942	Autobiographer,Teacher,Student,Supporter,Editor,Commentator,Organizer,Scholar,Enthusiast,Civic Duty,Tag Editor,Critic,Nice Answer,Mortarboard,Necromancer,Suffrage,Revival,Citizen Patrol,Talkative,Convention,Quorum,distributions,Sportsmanship,Strunk & White,Enlightened,Electorate,Pundit,Cleanup,regression,Nice Question,probability,Fanatic,hypothesis-testing,Good Answer,Synonymizer,r,Yearling,Excavator,normal-distribution,time-series,Investor,Altruist,Analytical,correlation,Proofreader,Disciplined,Copy Editor,sampling,Announcer,data-visualization,Taxonomist,Guru,variance,Caucus,Constituent,confidence-interval,estimation,Reviewer,Research Assistant,data-transformation,Custodian,Self-Learner,Generalist,Good Question,Informed,Popular Question,self-study,mathematical-statistics,Outspoken,Sheriff,standard-deviation,Steward,excel,Great Answer,random-variable,Populist,mean,logistic,statistical-significance,Refiner,Explainer,Notable Question,maximum-likelihood,algorithms,expected-value,conditional-probability,clustering,covariance,spatial,random-generation,multiple-regression,least-squares,Favorite Question,Famous Question,multivariate-analysis,simulation,monte-carlo,Curious,standard-error,Vox Populi,Epic,order-statistics,convergence,combinatorics,machine-learning,Great Question,pca,t-test,quantiles,bayesian,moments,intuition,Archaeologist,Booster,nonparametric,p-value,poisson-distribution,independence,beta-distribution,conditional-expectation,linear-model,regression-coefficients,stochastic-processes,Publicist,curve-fitting,terminology,joint-distribution,covariance-matrix,asymptotics,Lifejacket,bernoulli-distribution,gamma-distribution,dice,markov-process,optimization,median,matrix,descriptive-statistics,outliers,sample,extreme-value,central-limit-theorem,mixture-distribution,density-function,cumulative-distribution-function,uniform-distribution,entropy,binomial-distribution,nonlinear-regression,games,chi-squared-test,exponential-distribution,integral,lognormal-distribution,references,stationarity,inference,linear-algebra,residuals,generalized-linear-model,python	2010-08-19T13:07:22.083	2023-03-05T00:12:24.030	pro dis poi wa sta dis tes ana dis dat dis ass cor ger me cer pro reg des the bas fun dat me the pro reg dis inte alg alg pro mis ma diff sar nur cor est sco dat net enc var fac sto dat ana pre dis dat dis sca par
930	chl	2010-08-13T20:50:47.397	Paris, France	608	Autobiographer,Student,Teacher,Supporter,Editor,Critic,Commentator,Organizer,Civic Duty,Citizen Patrol,Mortarboard,Nice Question,Promoter,Enthusiast,Scholar,Benefactor,Tag Editor,Investor,Self-Learner,Nice Answer,Altruist,Enlightened,Suffrage,Sportsmanship,Cleanup,Electorate,Strunk & White,Revival,Necromancer,Announcer,r,Disciplined,Fanatic,Talkative,Quorum,Convention,regression,data-visualization,Popular Question,Good Question,anova,clustering,Synonymizer,Copy Editor,Yearling,Vox Populi,Excavator,Analytical,Proofreader,Pundit,Favorite Question,Good Answer,Taxonomist,Caucus,Constituent,correlation,Notable Question,logistic,Reviewer,Custodian,Generalist,Informed,psychometrics,Sheriff,Steward,Outspoken,Guru,Curious,Refiner,Explainer,Famous Question,Booster,Stellar Question,Publicist,Great Answer,factor-analysis,reliability,pca,Archaeologist,Deputy,scales	2010-08-19T10:00:00.370	2020-11-20T19:50:00.997	ord dat sou lea tes rota ma pos fun red cor plo cat sup like me the mo size
1352	Stephan Kolassa	2010-09-18T10:55:08.240	Switzerland	638	Student,Teacher,Supporter,Autobiographer,Editor,Scholar,Nice Answer,Commentator,Organizer,Enthusiast,Suffrage,Yearling,Popular Question,Good Answer,Analytical,Critic,Nice Question,Custodian,Disciplined,Notable Question,Necromancer,Citizen Patrol,Civic Duty,Tag Editor,Fanatic,Revival,Promoter,Pundit,time-series,Explainer,Informed,forecasting,Quorum,Caucus,Constituent,r,Excavator,regression,Enlightened,Generalist,Strunk & White,Convention,Cleanup,Electorate,Talkative,Investor,Refiner,Altruist,Sportsmanship,arima,Outspoken,Mortarboard,Great Answer,Guru,Taxonomist,machine-learning,Synonymizer,Announcer,Booster,data-visualization,Self-Learner,Curious,Reviewer,distributions,classification,Famous Question,accuracy,hypothesis-testing,multiple-regression,Proofreader,Steward,seasonality,probability,Publicist,predictive-models,prediction,confidence-interval,logistic,modeling,python,model-selection,count-data,normal-distribution,neural-networks,Good Question,categorical-data,mathematical-statistics,mean,prediction-interval,exponential-smoothing,random-forest,model-evaluation,Favorite Question,statistical-significance,error,Research Assistant,Epic,terminology,poisson-regression,scoring-rules,cross-validation,correlation,r-squared,Deputy,variance,mse,binary-data,quantiles,loss-functions,Benefactor,references,autocorrelation,self-study,expected-value,generalized-linear-model,p-value,unbalanced-classes	2010-09-23T20:42:45.613	2023-03-02T07:47:57.410	cou ser sel dat mo nlm tes sta sea dat inte est wir ma val dec est mo sta dis dis val risk tau spe lea ana cla dat me cor mo pro priv dis allo ma dis val det reg
1390	Gavin Simpson	2010-09-21T18:58:34.843	Denmark	296	Editor,Teacher,Supporter,Nice Answer,Enlightened,Critic,Commentator,Enthusiast,Autobiographer,Organizer,Quorum,r,Student,Scholar,Talkative,Citizen Patrol,Promoter,Yearling,Necromancer,Mortarboard,Good Answer,Caucus,Constituent,Custodian,Pundit,Civic Duty,Nice Question,Benefactor,Guru,Popular Question,Informed,Convention,Excavator,regression,Strunk & White,Outspoken,Revival,Fanatic,Curious,Explainer,Notable Question,Great Answer,Cleanup,Taxonomist,Announcer,Generalist,mgcv,Electorate,Tag Editor,time-series,generalized-linear-model,Booster,mixed-model,generalized-additive-model,splines,smoothing,multivariate-analysis	2010-09-24T15:36:42.200	2023-03-02T11:42:04.713	r,bi sel ser reg dis mo ger ma lea cor dis dis pro allo dis reg nor fre

Id	DisplayName	CreationDate	Location	BadgesWon	BadgeList	EarliestAnswer	LatestAnswer	Un
1934	Wolfgang	2010-11-09T18:15:27.370	Netherlands	132	Student,Scholar,Nice Question,Supporter,Teacher,Nice Answer,Enlightened,Commentator,Editor,Critic,Enthusiast,Suffrage,Revival,Yearling,Caucus,Good Answer,Organizer,Citizen Patrol,Popular Question,Custodian,Civic Duty,r,meta-analysis,Informed,Necromancer,Explainer,Constituent,Fanatic,Notable Question,Proofreader,Good Question,Curious,Autobiographer,Announcer,Guru,meta-regression,effect-size,Booster	2010-11-25T23:37:14.657	2022-11-09T13:14:16.193	anc dat ser plo net tra mo tau mo mo hoc sta
2116	mpiktas	2010-11-24T09:52:34.957	Vilnius, Lithuania	231	Student,Editor,Scholar,Teacher,Supporter,Commentator,Revival,Nice Answer,Enthusiast,Organizer,Quorum,Enlightened,Disciplined,Civic Duty,Suffrage,Critic,Citizen Patrol,Tag Editor,Cleanup,Strunk & White,Fanatic,Investor,Convention,Vox Populi,Electorate,Nice Question,Excavator,Analytical,Autobiographer,Sportsmanship,Yearling,Good Answer,Announcer,Copy Editor,Caucus,Constituent,Custodian,Promoter,Benefactor,Proofreader,Informed,Popular Question,Taxonomist,Generalist,Necromancer,Guru,Synonymizer,Pundit,Mortarboard,Reviewer,Curious,Refiner,Explainer,Deputy,Good Question,Great Answer,Notable Question,Self-Learner,Booster,time-series,r,stationarity,econometrics,probability,forecasting,arima,regression,distributions,correlation	2010-12-08T10:22:15.833	2022-12-27T09:34:47.107	r,bo mo me ser dis val leib tra mo aut fun equ wa reg des
2126	Ben Bolker	2010-11-24T20:42:46.860	Hamilton, Canada	248	Teacher,Supporter,Nice Answer,Enlightened,Commentator,Editor,Yearling,Citizen Patrol,Critic,Caucus,Excavator,r,Necromancer,Organizer,Constituent,Civic Duty,Enthusiast,Good Answer,Custodian,Explainer,mixed-model,Pundit,Autobiographer,Informed,Investor,Altruist,Revival,Strunk & White,Guru,random-effects-model,generalized-linear-model,lme4-nlme,regression,Electorate,glmm,Quorum,Generalist,Cleanup,Refiner,logistic,Sportsmanship,Student,Mortarboard,anova,Announcer,repeated-measures,Scholar,hypothesis-testing,Curious,Promoter,Benefactor,Nice Question,multilevel-analysis	2010-12-11T20:40:51.063	2023-02-04T03:30:28.870	dis dat tes des dis err infl mo tra me log dev dis dec mo val inte tria
2958	Henry	2011-01-28T07:59:59.930		189	Teacher,Editor,Supporter,Commentator,Enthusiast,Quorum,Revival,Student,Scholar,Fanatic,Nice Answer,Enlightened,Critic,Necromancer,Yearling,r,Caucus,Constituent,Organizer,Custodian,Pundit,probability,Nice Question,Citizen Patrol,Explainer,Excavator,distributions,Popular Question,self-study,Generalist,Promoter,Benefactor,Civic Duty,Good Answer,regression,Taxonomist,Mortarboard,Guru,Cleanup,Notable Question,Lifejacket,bayesian,standard-deviation,Curious,normal-distribution,mean,mathematical-statistics,variance,Strunk & White,hypothesis-testing	2011-02-01T00:44:41.830	2023-02-15T17:59:58.590	pro hoc vis net dat tes rec pro sta like dis dat sel roc dis ana sar mo squ mo dis dis sm inte
3277	ttnphns	2011-02-16T19:33:34.657	Moscow, Russia	799	Student,Teacher,Commentator,Editor,Organizer,Supporter,Scholar,Enthusiast,Critic,Nice Answer,Fanatic,Tag Editor,Enlightened,Yearling,Popular Question,pca,Custodian,Promoter,Strunk & White,Civic Duty,Necromancer,Good Answer,Self-Learner,Excavator,Quorum,Notable Question,clustering,spss,regression,Caucus,Constituent,correlation,Pundit,Citizen Patrol,Nice Question,Generalist,Proofreader,factor-analysis,Curious,Benefactor,Autobiographer,Refiner,Explainer,Taxonomist,Famous Question,r,Disciplined,Guru,Electorate,Announcer,Tumbleweed,Investor,Altruist,Copy Editor,Revival,Good Question,Cleanup,Favorite Question,k-means,Booster,Deputy,Publicist,categorical-data,Great Answer,Archaeologist,Convention,Informed	2011-07-08T07:13:43.410	2022-11-13T01:42:10.793	r,fa me sca sim rho tes ma lea reg ass cor tes ma tes ana var
3382	IrishStat	2011-02-23T13:29:56.240	Warminster, PA, United States	96	Teacher,Autobiographer,Supporter,Editor,Revival,Commentator,Critic,Enthusiast,Necromancer,time-series,Yearling,Nice Answer,Tenacious,r,arima,forecasting,Enlightened,Explainer,Informed,Custodian,Caucus,regression,Good Answer,Civic Duty,Fanatic,Citizen Patrol,Guru,autocorrelation,Electorate,data-transformation,Unsung Hero,Peer Pressure,distributions,seasonality,Mortarboard,Great Answer,stationarity,Cleanup	2011-03-14T18:12:36.827	2022-04-23T18:41:27.197	me tes tra lea me sm sea line fore pro roc aut cha inte dat res reg wir
4253	Frank Harrell	2011-04-20T12:59:07.093	Nashville, TN	511	Teacher,Editor,Commentator,Supporter,Enthusiast,Revival,Nice Answer,Quorum,Critic,Good Answer,regression,Necromancer,Student,Yearling,Enlightened,Excavator,Promoter,r,Nice Question,Pundit,logistic,Generalist,Guru,Custodian,Fanatic,Popular Question,Caucus,Tag Editor,Organizer,model-selection,Informed,cross-validation,machine-learning,Autobiographer,Explainer,Curious,classification,Taxonomist,hypothesis-testing,multiple-regression,Constituent,Notable Question,Good Question,survival,cox-model,interaction,regression-strategies,Civic Duty,nonparametric,roc,feature-selection,predictive-models,generalized-linear-model,Scholar,Self-Learner,validation,Mortarboard,bayesian,Great Answer,modeling,ordered-logit,bootstrap,auc,splines,confidence-interval,statistical-significance,data-transformation,lasso,Lifejacket,categorical-data,anova,unbalanced-classes,ordinal-data,correlation,data-visualization,Citizen Patrol,t-test,regularization,wilcoxon-mann-whitney-test,Sportsmanship,least-squares,sample-size,Famous Question,ridge-regression,Talkative,random-forest,p-value,chi-squared-test,mathematical-statistics	2011-04-20T13:03:50.337	2023-01-25T14:29:24.820	epi pov ass len dat dat tes tau cor tria sig fore reg d,s rati ma dev the me effe mo cla sco opt gro ind
4505	Greg Snow	2011-05-07T13:44:25.593	Utah	261	Teacher,Editor,Supporter,Nice Answer,Mortarboard,Commentator,Revival,Critic,Enthusiast,Organizer,Enlightened,r,Yearling,hypothesis-testing,regression,Necromancer,Custodian,Civic Duty,Good Answer,Generalist,Caucus,anova,statistical-significance,Sportsmanship,distributions,confidence-interval,p-value,Explainer,logistic,multiple-regression,data-visualization,Cleanup,probability,t-test,correlation,Guru,normal-distribution,Populist,bayesian,Lifejacket,nonparametric,mathematical-statistics,statistical-power,regression-coefficients,Great Answer	2011-05-10T19:45:10.343	2023-02-06T18:36:27.193	p-v ana sig fun sta tes vis mo tes coe val dis pro exp pro size squ lea aut tes
4598	cbeleites unhappy with SX	2011-05-13T10:02:56.203	Germany	218	Editor,Teacher,Supporter,Commentator,Student,Citizen Patrol,Nice Answer,Enlightened,Quorum,Yearling,Critic,Self-Learner,Custodian,Organizer,Tumbleweed,Revival,Necromancer,Enthusiast,Tag Editor,Good Answer,Scholar,Caucus,cross-validation,classification,machine-learning,Talkative,Mortarboard,Civic Duty,Taxonomist,Announcer,Guru,r,Nice Question,Explainer,Autobiographer,pca,Generalist,regression,model-selection,Popular Question,feature-selection,Populist,validation,Pundit,predictive-models,Sportsmanship,bootstrap,svm,logistic,Notable Question,Great Answer,scikit-learn	2011-07-22T13:09:00.340	2023-01-15T20:11:53.877	rep vis lea net tric lea reg reg dec cla sar dev pre err err
5739	StasK	2011-08-08T17:31:44.413	Columbia, MO	255	Autobiographer,Teacher,Commentator,Supporter,Organizer,Editor,Tag Editor,Nice Answer,Citizen Patrol,Announcer,Student,Scholar,Critic,Quorum,Booster,Mortarboard,Necromancer,Revival,Good Answer,Yearling,Enlightened,Civic Duty,Excavator,Custodian,Convention,Nice Question,Generalist,Informed,Strunk & White,Promoter,Caucus,Constituent,Taxonomist,Guru,Popular Question,Curious,Explainer,Cleanup,Pundit,Notable Question,Suffrage,Investor,Sportsmanship,Talkative,Altruist,Proofreader,Self-Learner,Reviewer,Great Answer,logistic,stata,sampling,mixed-model,r,bootstrap,survey,regression,distributions,hypothesis-testing,structural-equation-modeling,Publicist,Good Question	2011-08-08T21:29:26.753	2020-07-21T22:43:03.630	cor ser leib dis dis equ pro car filte est equ dis mo ass me sta dat sta me reg

Id	DisplayName	CreationDate	Location	BadgesWon	BadgeList	EarliestAnswer	LatestAnswer	Un
6193	frank	2011-09-06T14:28:52.203		48	Scholar,Student,Popular Question,Custodian,Supporter,Notable Question,Famous Question,Caucus,Editor,Teacher,Yearling,Critic,Commentator,Nice Answer,Citizen Patrol,Tenacious,Enthusiast,Unsung Hero,Revival,Civic Duty,Mortarboard,Organizer,Explainer,machine-learning,Fanatic,Enlightened,Generalist,Disciplined	2022-02-08T07:52:25.710	2022-12-02T15:28:09.620	reg det mir me mo var dis dia pro imp clu trai opt mo net ma dis ma val sep pro pro sta
6633	Dilip Sarwate	2011-10-03T16:59:55.347	Urbana, IL	315	Teacher,Editor,Supporter,Commentator,Citizen Patrol,Critic,Student,Scholar,Enthusiast,Quorum,Fanatic,Nice Answer,Enlightened,Necromancer,Excavator,Revival,Yearling,Custodian,Pundit,Good Answer,probability,Popular Question,Nice Question,Caucus,Constituent,distributions,normal-distribution,Guru,self-study,mathematical-statistics,Autobiographer,Explainer,Notable Question,random-variable,Talkative,Civic Duty,Announcer,Organizer,Informed,Disciplined,correlation,independence,Taxonomist,Self-Learner,conditional-probability,Mortarboard,variance,Great Answer,Famous Question,covariance,Lifejacket,expected-value,bayesian,Booster,density-function,Cleanup,conditional-expectation,time-series	2011-10-04T12:03:25.833	2023-01-11T04:00:03.027	pro dis var mo pov dis fun fun dis alg cal the ger dim dis
7071	dimitriy	2011-10-26T15:10:36.547	Oakland, CA, USA	225	Student,Teacher,Scholar,Commentator,Editor,Supporter,Organizer,Critic,Promoter,Enthusiast,Tumbleweed,Quorum,Revival,Yearling,Citizen Patrol,Nice Answer,Enlightened,Custodian,Informed,Analytical,Civic Duty,Necromancer,Caucus,Constituent,Tag Editor,regression,Curious,Autobiographer,Explainer,Strunk & White,Cleanup,econometrics,stata,Good Answer,Fanatic,Popular Question,Excavator,Notable Question,Nice Question,Inquisitive,Generalist,Taxonomist,Announcer,time-series,interpretation,Pundit,multiple-regression,Booster,Publicist,r,Good Question,causality,Electorate,Benefactor,hypothesis-testing,Great Answer,logistic,panel-data,Guru,instrumental-variables,statistical-significance,Deputy	2012-01-25T20:03:37.290	2023-02-28T21:18:49.587	reg mo in-c coe dis imp sm me dis mo dis wa sta ass scc stu mo cor tes cor
7224	Xi'an	2011-11-05T07:56:15.903	Paris, France	793	Teacher,Editor,Supporter,Critic,Commentator,Analytical,Autobiographer,Organizer,Student,Announcer,Scholar,Nice Answer,Enlightened,Enthusiast,Disciplined,Tag Editor,Citizen Patrol,Civic Duty,Promoter,Self-Learner,Benefactor,Revival,Excavator,Nice Question,Yearling,bayesian,Mortarboard,Good Answer,Taxonomist,Explainer,Custodian,Strunk & White,Necromancer,Informed,Reviewer,distributions,Electorate,Fanatic,Caucus,Constituent,Pundit,mathematical-statistics,self-study,Popular Question,Refiner,Proofreader,Steward,sampling,Curious,simulation,probability,monte-carlo,Guru,Cleanup,r,Sportsmanship,Investor,normal-distribution,gibbs,Copy Editor,Notable Question,estimation,Suffrage,Booster,random-generation,prior,Deputy,conditional-probability,metropolis-hastings,posterior,maximum-likelihood,inference,likelihood,random-variable,variance,machine-learning,markov-process,Vox Populi,unbiased-estimator,sufficient-statistics,Famous Question,gamma-distribution,expectation-maximization,decision-theory,importance-sampling,conjugate-prior,exponential-family,regression,Populist,Census,Generalist,convergence,hypothesis-testing,expected-value,integral,mean,joint-distribution,Great Answer,Good Question,beta-distribution,independence,mixture-distribution,gaussian-mixture-distribution,markov-chain-montecarlo,density-function,uniform-distribution,binomial-distribution,cumulative-distribution-function,correlation,order-statistics,marginal-distribution,references,conditional-expectation,uninformative-prior	2011-11-05T21:11:34.147	2023-02-23T11:40:34.633	bay tes pric the pro mo aug inte dis dis the dis phe dis infl bay the pro lsd diff inte fun tes cal des fun ine sto pro dis pro pro reg lea
7290	gung - Reinstate Monica	2011-11-09T04:43:15.613	Kingdom of Zhao	1140	Editor,Scholar,Teacher,Supporter,Commentator,Analytical,Critic,Organizer,Enthusiast,Nice Answer,Enlightened,Quorum,Civic Duty,Fanatic,regression,Tag Editor,Revival,Student,Necromancer,Sportsmanship,Excavator,Citizen Patrol,Cleanup,Convention,Synonymizer,r,Good Answer,Strunk & White,Suffrage,Electorate,Disciplined,Mortarboard,Pundit,correlation,Vox Populi,hypothesis-testing,Reviewer,Custodian,Copy Editor,Promoter,Proofreader,Nice Question,Benefactor,Populist,Deputy,Yearling,multiple-regression,Generalist,logistic,Informed,Steward,generalized-linear-model,Guru,anova,Taxonomist,Self-Learner,t-test,data-visualization,Research Assistant,Archaeologist,Caucus,Investor,Talkative,Popular Question,Constituent,Altruist,Good Question,Great Answer,Favorite Question,Marshal,statistical-significance,categorical-data,interpretation,distributions,interaction,assumptions,residuals,Curious,Notable Question,self-study,Autobiographer,Refiner,Explainer,heteroscedasticity,linear-model,data-transformation,Announcer,probability,terminology,p-value,regression-coefficients,normal-distribution,Outspoken,confidence-interval,Illuminator,variance,multiple-comparisons,Booster,Famous Question,mean,multicollinearity,mixed-model,random-generation,survival,machine-learning,nonparametric,Publicist,clustering,simulation,effect-size,normality-assumption,mcnemar-test,Sheriff,contingency-tables,Lifejacket,predictive-models,r-squared,stepwise-regression,statistical-power,binomial-distribution,chi-squared-test,references,sample-size,ancova,repeated-measures	2011-11-10T04:08:06.260	2023-03-02T22:01:38.573	inte sig reg inte sta tes tes ass des squ mo pro dat sig tes reg cor sta var dis ind tes sta me dis dis dis fiel len eve lea spe
7486	Robert Long	2011-11-20T14:30:04.120	Leeds, United Kingdom	257	Teacher,Student,Supporter,Editor,Scholar,Commentator,Citizen Patrol,Critic,Yearling,Promoter,Tumbleweed,Tag Editor,Organizer,Nice Question,Revival,Popular Question,Nice Answer,Curious,Autobiographer,Enlightened,Notable Question,Taxonomist,Custodian,Explainer,Excavator,Enthusiast,Quorum,Civic Duty,Informed,Cleanup,Mortarboard,Good Answer,mixed-model,r,Famous Question,Guru,Necromancer,lme4-nlme,Great Answer,regression,Self-Learner,Fanatic,Suffrage,multilevel-analysis,Reviewer,Electorate,random-effects-model,repeated-measures,Generalist,logistic,Strunk & White,glmm,multiple-regression,missing-data,interaction,Refiner,anova,causality,correlation,Pundit,Caucus,Constituent,generalized-linear-model,python,linear-model,machine-learning,Epic,hypothesis-testing,Sportsmanship,inference,Vox Populi,interpretation,linear	2012-09-16T21:51:25.363	2021-10-24T18:45:04.263	mix like me stu reg dat siz dat mo dis ma eff squ tes vis ana fac mo rati mo squ for dat dis
7555	jbowman	2011-11-23T02:34:06.680	Berkeley, CA, USA	189	Teacher,Editor,Revival,Supporter,Commentator,Autobiographer,Nice Answer,Enlightened,Enthusiast,Organizer,Good Answer,Mortarboard,Critic,r,Civic Duty,Yearling,regression,Necromancer,Generalist,Custodian,Fanatic,Citizen Patrol,Pundit,Explainer,Reviewer,bayesian,Guru,Excavator,Proofreader,Informed,Steward,distributions,hypothesis-testing,Census,Electorate,self-study,estimation,probability,maximum-likelihood,poisson-distribution,Great Answer,optimization,Strunk & White,markov-chain-montecarlo,Deputy	2011-11-27T23:28:43.820	2023-02-25T18:00:47.460	lea dat squ lea nlm val inte like dat dis val vis squ dev cor infl des dat tes dis mo sar dim lea cha dis dis ma pro dis nor inte mu pro
7828	Has QUIT--Anony-Mousse	2011-12-06T08:56:19.207	disappeared into oblivion	174	Supporter,Teacher,Editor,Organizer,Commentator,Autobiographer,Student,Revival,Nice Question,Tag Editor,Critic,Citizen Patrol,Tumbleweed,Scholar,Nice Answer,clustering,Excavator,Yearling,Popular Question,Quorum,Custodian,k-means,Taxonomist,Good Answer,data-mining,Notable Question,Necromancer,machine-learning,Explainer,Good Question,Populist,Tenacious,Curious,Caucus,Constituent,Enlightened,r,Civic Duty,Informed,Great Answer,Enthusiast,distance-functions,Favorite Question,Guru,classification,Electorate,Announcer,Disciplined,Pundit,distance,hierarchical-clustering,Cleanup,Strunk & White,time-series,Booster,unsupervised-learning,pca,Famous Question	2011-12-21T08:22:42.870	2019-12-18T22:35:37.650	clu imp trai lea dat ana vis dis ana bas inte cor dat we est ma me

Id	DisplayName	CreationDate	Location	BadgesWon	BadgeList	EarliestAnswer	LatestAnswer	Un
7972	Peter Ellis	2011-12-13T02:20:43.653	Melbourne, Australia	136	Teacher,Scholar,Autobiographer,Editor,Supporter,Organizer,Commentator,Critic,Analytical,Revival,Suffrage,Vox Populi,Nice Answer,Quorum,Student,Self-Learner,Enlightened,Excavator,Civic Duty,Promoter,Nice Question,Citizen Patrol,Enthusiast,Necromancer,Custodian,Yearling,Generalist,Good Answer,Caucus,Constituent,Popular Question,Curious,Explainer,Guru,Electorate,Notable Question,Pundit,Good Question,regression,r	2012-01-17T01:06:11.930	2018-02-03T00:14:07.023	sta sar dis alg dis mo cla car red stu dat siz law inte rati
8013	AdamO	2011-12-14T21:46:36.197	Nakoja Abad	343	Teacher,Revival,Supporter,Editor,Student,Commentator,Critic,Analytical,Scholar,Yearling,Excavator,Organizer,Tumbleweed,Citizen Patrol,Nice Answer,Enthusiast,Quorum,Promoter,Necromancer,Custodian,Caucus,Constituent,Enlightened,Nice Question,Tag Editor,Strunk & White,Curious,regression,Autobiographer,Explainer,Inquisitive,Talkative,r,Good Answer,Popular Question,Proofreader,Generalist,Self-Learner,Taxonomist,logistic,Notable Question,Refiner,Pundit,Civic Duty,hypothesis-testing,Convention,generalized-linear-model,Informed,Investor,Suffrage,Vox Populi,Fanatic,Guru,Altruist,Reviewer,bayesian,Electorate,bootstrap,Benefactor,p-value,statistical-significance,survival,machine-learning,confidence-interval,mathematical-statistics,mixed-model,probability,correlation,Mortarboard,cox-model,multiple-regression,Deputy,predictive-models,Announcer,Copy Editor,distributions,Socratic,Famous Question,maximum-likelihood,estimation,causality,inference,t-test	2011-12-15T00:35:15.920	2023-02-27T19:12:49.380	chi squ sys sta reg mo dat ser siz equ mo mo tria dis sel est reg sel cor mo cov cor tes log eva alg we ten filte infl par dat ma gro
8336	Taylor	2012-01-04T21:10:36.243	Virginia, United States	109	Editor,Student,Teacher,Supporter,Analytical,Citizen Patrol,Commentator,Scholar,Promoter,Benefactor,Tumbleweed,Yearling,Autobiographer,Curious,Critic,Excavator,Explainer,Mortarboard,Organizer,Nice Answer,Enlightened,Revival,Custodian,Enthusiast,Informed,Quorum,Civic Duty,Necromancer,Taxonomist,Proofreader,Caucus,Constituent,Popular Question,Fanatic,Disciplined,Strunk & White,Cleanup,Self-Learner,Inquisitive,Good Answer,Electorate,Guru,self-study,mathematical-statistics,time-series,probability,bayesian,Notable Question,Nice Question,Generalist	2012-03-29T21:08:02.050	2023-01-12T16:38:42.003	sel ser var for dis des val est nur reg has squ ma sel info cor alg nor dis fun cor
9964	Danica	2012-03-19T19:49:40.507	Vancouver, BC, Canada	190	Autobiographer,Critic,Teacher,Supporter,Yearling,Commentator,Editor,Citizen Patrol,Caucus,Announcer,Nice Answer,Revival,Necromancer,Informed,Explainer,Constituent,Custodian,Student,Organizer,Quorum,Mortarboard,Civic Duty,Enthusiast,Excavator,Fanatic,Talkative,Tag Editor,Scholar,Strunk & White,Enlightened,machine-learning,Cleanup,Good Answer,svm,probability,Generalist,Guru,kernel-trick,Self-Learner,Pundit,Taxonomist,neural-networks,classification,optimization,distributions,Census,Booster,python,Popular Question,regression,Notable Question,Nice Question,normal-distribution	2013-05-21T02:10:09.063	2022-04-22T06:06:07.783	der fun tes fac pro cla des reg ma opt net pre mo sel cor exp ent dis ma rec rule
11032	Michael R. Chernick	2012-05-02T14:04:04.697	Holland, Pennsylvania, United States	254	Teacher,Revival,Student,Mortarboard,Commentator,Editor,Supporter,Critic,Necromancer,Quorum,Vox Populi,Suffrage,Nice Answer,Civic Duty,Enthusiast,Scholar,Enlightened,Sportsmanship,Peer Pressure,Fanatic,regression,r,Electorate,Convention,Strunk & White,Organizer,Proofreader,Nice Question,time-series,Custodian,statistical-significance,hypothesis-testing,Generalist,Yearling,Popular Question,Good Question,normal-distribution,Autobiographer,Explainer,Notable Question,bootstrap,Good Answer,Taxonomist,distributions,bayesian,Citizen Patrol,Pundit,Tag Editor,Informed,Excavator,Reviewer,references,standard-deviation,probability,Copy Editor,Steward,machine-learning,mathematical-statistics,Caucus,t-test,Constituent,Refiner,Favorite Question,self-study,outliers,confidence-interval,correlation,Lifejacket,variance,Guru	2012-05-02T15:41:40.813	2019-01-18T01:47:56.287	mix dis tes fac the me var rec thu des dis imp for mo dis line inte var reg dis fun trai dis sta
11852	usεr11852	2012-06-08T10:20:40.383	Manchester, UK	219	Teacher,Supporter,Analytical,Editor,Necromancer,Critic,Cleanup,Commentator,Organizer,Suffrage,Enthusiast,Vox Populi,Informed,Civic Duty,Yearling,Nice Answer,Enlightened,Custodian,Revival,Quorum,Tag Editor,Citizen Patrol,Autobiographer,Explainer,Caucus,Constituent,Proofreader,r,regression,Talkative,Good Answer,Reviewer,Excavator,Peer Pressure,Pundit,mixed-model,Generalist,machine-learning,pca,lme4-nlme,Convention,Electorate,Outspoken,Guru,Mortarboard,classification,Fanatic,optimization,boosting,Taxonomist,mathematical-statistics,python,random-forest	2012-08-22T01:47:05.583	2023-02-24T14:00:03.003	pca pro wa ana ten stu lan ma sig pre fac sta me tra dat cur tes lea enc cor net mo dis tes sm dev dat
11887	kjetil b halvorsen	2012-06-09T22:52:37.473	Bolivia	712	Teacher,Supporter,Critic,Commentator,Editor,Student,Enthusiast,Citizen Patrol,Revival,Suffrage,Yearling,Custodian,Organizer,Civic Duty,Tumbleweed,Quorum,Autobiographer,Explainer,Nice Answer,Enlightened,Reviewer,Informed,Tag Editor,Caucus,Constituent,Proofreader,Electorate,Vox Populi,Excavator,Steward,Strunk & White,Curious,Pundit,Necromancer,Fanatic,Deputy,Sportsmanship,Convention,regression,Good Answer,Promoter,Copy Editor,Scholar,Archaeologist,Generalist,distributions,Taxonomist,mathematical-statistics,variance,intuition,probability,model,logistic,Mortarboard,bayesian,Self-Learner,Nice Question,Guru,maximum-likelihood,Refiner,normal-distribution,Research Assistant,generalized-linear-model,r,Marshal,expected-value,Popular Question,correlation,estimation,random-variable,Talkative,modeling,poisson-regression,Cleanup,machine-learning,categorical-data,references,inference,hypothesis-testing,likelihood,Good Question,definition,terminology,poisson-distribution,confidence-interval,experiment-design,distance,categorical-encoding,self-study,Synonymizer,residuals,data-transformation,moment-generating-function,binomial-distribution,Notable Question,descriptive-statistics,mean,anova,Sheriff,matrix,regularization,time-series,Great Answer,independence,covariance-matrix,statistical-significance,linear-algebra,Investor,exponential-family,Altruist,Outspoken	2012-06-26T18:14:28.970	2023-02-16T14:22:14.383	ref like stu ma ser seg tes reg fun dat val sta dis dis dev pro sta pre las net risk net cor tes inte reg wa coe ger tail sco dat hoo dis ana dis pro ser dis mo var
14076	Kodiologist	2012-09-13T23:44:42.007	Mount Sinai Health System	112	Informed,Student,Scholar,Teacher,Supporter,Editor,Commentator,Yearling,Citizen Patrol,Caucus,Promoter,Benefactor,Explainer,Critic,Investor,Excavator,Altruist,Autobiographer,Nice Answer,Good Answer,Talkative,Tenacious,Mortarboard,Organizer,Quorum,Custodian,Guru,Curious,Unsung Hero,Enthusiast,Cleanup,Revival,Enlightened,Tag Editor,Fanatic,Outspoken,Civic Duty,Generalist,Pundit,Taxonomist,Constituent,Strunk & White,Tumbleweed,Popular Question,Announcer,Self-Learner,Nice Question,probability,regression,correlation,Notable Question,machine-learning	2015-02-08T18:10:36.170	2020-01-14T14:22:49.630	pri dis like inte spe siz mo vis sig dev tail lea reg reg los imp fun bay tes ana est

Id	DisplayName	CreationDate	Location	BadgesWon	BadgeList	EarliestAnswer	LatestAnswer	Un
17072	Jeremy Miles	2012-11-21T21:55:31.517	Los Angeles, CA	109	Teacher,Supporter,Commentator,Editor,Student,Critic,Nice Question,Citizen Patrol,Scholar,Enthusiast,Excavator,Organizer,Informed,Caucus,Constituent,Civic Duty,Yearling,Taxonomist,Popular Question,Revival,Curious,Autobiographer,Explainer,Tag Editor,Notable Question,Custodian,Strunk & White,Nice Answer,Quorum,Famous Question,Enlightened,Reviewer,Tumbleweed,Steward,Proofreader,Electorate,r,confirmatory-factor,structural-equation-modeling,Good Question,Self-Learner,Good Answer,factor-analysis	2013-03-20T00:49:38.207	2023-01-25T11:47:22.057	cor tes sta mo ana coe visi des rati mo a-v
17230	Scortchi - Reinstate Monica	2012-11-27T12:24:58.070	England	360	Editor,Teacher,Analytical,Supporter,Excavator,Commentator,Critic,Citizen Patrol,Suffrage,Organizer,Revival,Custodian,Tag Editor,Vox Populi,Strunk & White,Informed,Civic Duty,Cleanup,Enthusiast,Electorate,Student,Synonymizer,Quorum,Proofreader,Investor,Necromancer,Caucus,Reviewer,Constituent,Archaeologist,Nice Answer,Enlightened,Yearling,regression,Copy Editor,Mortarboard,Generalist,Altruist,Pundit,Good Answer,r,Fanatic,Sportsmanship,logistic,Convention,Explainer,Populist,hypothesis-testing,Refiner,Talkative,confidence-interval,Steward,generalized-linear-model,Nice Question,Scholar,Sheriff,Guru,Outspoken,statistical-significance,Great Answer,Announcer,Popular Question,categorical-data,Booster,Autobiographer,self-study,terminology,Notable Question,p-value,Publicist,probability	2012-11-29T12:33:39.650	2022-04-10T17:49:14.653	sta inte dis like fitti fit,a tes stu noi err dis reg est p-v eng i-ar prin ana
21054	COOLSerdash	2013-02-20T08:48:36.937		225	Student,Editor,Teacher,Supporter,Informed,Scholar,Commentator,Critic,Analytical,Mortarboard,Suffrage,Custodian,Organizer,Nice Answer,Enlightened,Civic Duty,Excavator,Strunk & White,Enthusiast,Vox Populi,Talkative,Self-Learner,Promoter,r,Citizen Patrol,Proofreader,Electorate,Reviewer,Caucus,Constituent,Popular Question,Good Answer,Yearling,Curious,Pundit,Notable Question,Explainer,Fanatic,Nice Question,Copy Editor,Necromancer,Cleanup,Generalist,Guru,Great Answer,Famous Question,Steward,Investor,Altruist,Announcer,Tumbleweed,Benefactor,regression,Revival,Booster,distributions,Good Question,Deputy,Inquisitive,probability,logistic,Publicist,confidence-interval	2013-04-14T14:03:15.153	2023-02-17T14:34:40.123	sel dis coe reg cor visi risk info fit,c dis pov fun ana ana dis i-ar
22047	Nick Cox	2013-03-15T13:57:54.467	Durham, UK	296	Teacher,Supporter,Critic,Editor,Commentator,Citizen Patrol,Enthusiast,Organizer,Informed,Nice Answer,Enlightened,Excavator,Custodian,Strunk & White,Tenacious,Pundit,Mortarboard,Copy Editor,regression,Quorum,Tag Editor,Revival,Fanatic,r,Cleanup,Proofreader,Civic Duty,Caucus,Reviewer,Constituent,Good Answer,Generalist,Necromancer,Convention,Yearling,Archaeologist,data-visualization,Deputy,Autobiographer,Steward,Refiner,Explainer,distributions,data-transformation,Talkative,correlation,Sportsmanship,Guru,Outspoken,categorical-data,outliers,Electorate,logistic,normal-distribution,mean,descriptive-statistics,terminology,generalized-linear-model,hypothesis-testing,multiple-regression,Marshal,Great Answer,time-series,histogram,Lifejacket,chi-squared-test,mathematical-statistics,stata,boxplot,python	2013-03-22T02:41:59.837	2023-02-08T22:44:47.703	fac mo dis squ mo dat dis dat rec ind the pol reg dis var sta me pro inte var eng cat pac mo
22311	Sycorax	2013-03-20T23:59:56.610	Washington, DC, United States	565	Student,Editor,Autobiographer,Scholar,Supporter,Teacher,Commentator,Critic,Citizen Patrol,Informed,Organizer,Excavator,Caucus,Constituent,Promoter,Tag Editor,Strunk & White,Civic Duty,Enthusiast,Yearling,Custodian,Nice Answer,Mortarboard,Good Answer,Guru,Fanatic,Curious,Explainer,Proofreader,Deputy,Enlightened,bayesian,Benefactor,Quorum,Electorate,Self-Learner,Investor,Altruist,Copy Editor,Pundit,Reviewer,Popular Question,Refiner,regression,r,machine-learning,Cleanup,Nice Question,Talkative,probability,Revival,Convention,Generalist,Outspoken,Great Answer,Steward,Disciplined,Notable Question,Good Question,Necromancer,Taxonomist,random-forest,neural-networks,classification,optimization,logistic,deep-learning,distributions,Favorite Question,Suffrage,Vox Populi,lasso,Sportsmanship,Famous Question,Marshal,Stellar Question,Great Question,regularization,Inquisitive,gradient-descent,feature-selection,loss-functions,tensorflow,auc,Census,roc,python,scikit-learn,hyperparameter,Archaeologist,Synonymizer,boosting,conv-neural-network,keras,Sheriff,cross-entropy,recurrent-neural-network,cart,svm	2013-03-24T21:49:36.957	2023-02-22T13:47:21.993	nor des dis tab lea squ nlm bug idf, sel ser ger ana fitti red fun aug des rec lan err lea bay lea sta allo the dis sta val
23853	Maarten Buis	2013-04-03T10:24:26.900	Konstanz, Germany	92	Informed,Teacher,Editor,Autobiographer,Supporter,Custodian,Commentator,Citizen Patrol,Critic,Revival,Mortarboard,Organizer,Caucus,Constituent,Nice Answer,Yearling,Explainer,Enlightened,Enthusiast,Necromancer,Pundit,Good Answer,Generalist,Civic Duty,regression,logistic,interaction,r,stata,multiple-regression,hypothesis-testing,categorical-data	2013-04-05T09:19:30.130	2022-09-28T08:37:38.007	ger dis dis var dat dat larg coe tes sta ser cov reg like sco dis visi
25433	Marc Claesen	2013-05-09T09:38:54.573	Belgium	126	Editor,Teacher,Autobiographer,Supporter,Informed,Custodian,Citizen Patrol,Critic,Commentator,Enthusiast,Student,Promoter,Benefactor,Scholar,Revival,Nice Answer,Caucus,Fanatic,Investor,Civic Duty,Tenacious,Organizer,Yearling,Explainer,Quorum,Enlightened,Constituent,Pundit,Necromancer,Good Answer,Mortarboard,Popular Question,Guru,Notable Question,Announcer,Booster,svm,machine-learning,kernel-trick,roc,classification,libsvm	2013-05-19T21:59:35.793	2016-08-05T07:58:57.917	svr dat cla sel rec cla sup the des
26338	Andy	2013-05-31T11:45:45.470		199	Scholar,Student,Teacher,Supporter,Informed,Promoter,Editor,Benefactor,Commentator,Enthusiast,Suffrage,Vox Populi,Civic Duty,Citizen Patrol,Organizer,Critic,Autobiographer,Quorum,Custodian,Revival,Excavator,Tag Editor,Caucus,Constituent,Popular Question,Electorate,Deputy,Reviewer,Strunk & White,Announcer,Notable Question,Fanatic,Research Assistant,Steward,Nice Answer,Enlightened,Mortarboard,Marshal,Proofreader,Necromancer,Convention,Archaeologist,Pundit,Synonymizer,Yearling,Cleanup,regression,Sportsmanship,Copy Editor,Explainer,econometrics,Curious,Nice Question,Talkative,panel-data,instrumental-variables,fixed-effects-model,Famous Question,Refiner,stata,difference-in-difference,Good Answer,Taxonomist,Favorite Question,Guru	2013-06-24T20:39:06.227	2018-05-28T22:00:19.767	clu ma diff pro sm effe sel sci err des
26948	shimao	2013-06-16T15:24:48.493		142	Tumbleweed,Editor,Teacher,Student,Scholar,Yearling,Citizen Patrol,Supporter,Commentator,Critic,Nice Answer,Mortarboard,Good Answer,Autobiographer,Enthusiast,Tenacious,Custodian,Revival,machine-learning,neural-networks,Enlightened,Tag Editor,Fanatic,deep-learning,Necromancer,Organizer,Quorum,reinforcement-learning,Promoter,Curious,conv-neural-network,Caucus,Guru,Great Answer,autoencoders,variational-bayes,Popular Question,Taxonomist,loss-functions,Nice Question	2017-03-11T23:51:05.680	2022-06-12T16:16:08.920	ma sys val ana aug car pro ent cor pro crit tra the pro enc dis
28500	EdM	2013-07-26T15:11:03.380		318	Teacher,Supporter,Informed,Commentator,Yearling,Revival,Editor,Organizer,Nice Answer,Enlightened,Tenacious,Caucus,Custodian,Unsung Hero,Explainer,Enthusiast,Civic Duty,Cleanup,Fanatic,regression,Citizen Patrol,r,Mortarboard,Necromancer,Excavator,Generalist,Critic,Good Answer,Announcer,survival,Constituent,logistic,Booster,Pundit,cox-model,machine-learning,lasso,multiple-regression,bootstrap,Publicist,hypothesis-testing,feature-selection,statistical-significance,Sportsmanship,confidence-interval,mixed-model,anova,p-value,classification,Strunk & White,modeling,interaction,lme4-nlme,Electorate,proportional-hazards,generalized-linear-model,biostatistics,predictive-models,categorical-data,pca,Guru,multicollinearity,cross-validation,censoring,hazard,statistical-power	2013-08-13T20:01:37.963	2023-03-02T21:10:15.797	reg dis imp dat dis dat mo sel squ sta imp lsd squ spe dis dis net rati eve ana diff lea mo kap mo eng hsc pre los sca dis lea me ind val pro altr 19, ser cat stu dis log

Id	DisplayName	CreationDate	Location	BadgesWon	BadgeList	EarliestAnswer	LatestAnswer	Un
28666	amoeba	2013-07-31T11:19:00.203	St. Petersburg ⇨ Freiburg ⇨ London ⇨ Lisboa ⇨ Tübingen	652	Editor,Student,Scholar,Commentator,Informed,Teacher,Supporter,Custodian,Organizer,Nice Question,Citizen Patrol,Quorum,Critic,Excavator,Enthusiast,Revival,Civic Duty,Self-Learner,Curious,Yearling,Necromancer,Tag Editor,Autobiographer,Explainer,Popular Question,Strunk & White,Promoter,Benefactor,Nice Answer,pca,Good Answer,Good Question,Convention,Talkative,Suffrage,Enlightened,Electorate,Archaeologist,Refiner,Mortarboard,Deputy,dimensionality-reduction,Notable Question,Fanatic,Caucus,Copy Editor,Constituent,Favorite Question,Investor,Altruist,r,Synonymizer,Taxonomist,hypothesis-testing,regression,Famous Question,Outspoken,Pundit,Generalist,discriminant-analysis,Great Answer,Guru,svd,factor-analysis,Cleanup,Proofreader,Great Question,Stellar Question,Sportsmanship,Populist,Research Assistant,Vox Populi,p-value,Inquisitive,Marshal,Announcer,Booster,Publicist,Favorite Answer	2013-07-31T14:19:32.060	2019-04-12T22:30:19.857	r,sp val reg net ana lea ana cor phe sel sta squ lea r,sp pro infe line
28746	Alecos Papadopoulos	2013-08-02T14:24:21.923		398	Teacher,Editor,Supporter,Commentator,Student,Enthusiast,Citizen Patrol,Caucus,Constituent,Quorum,Nice Answer,Enlightened,Talkative,Fanatic,Tenacious,Custodian,Organizer,Tag Editor,Revival,regression,Mortarboard,Critic,Necromancer,self-study,Excavator,Scholar,distributions,probability,Good Answer,Nice Question,Promoter,maximum-likelihood,Benefactor,mathematical-statistics,Self-Learner,Curious,Tumbleweed,Yearling,normal-distribution,Convention,Autobiographer,Explainer,time-series,Popular Question,Generalist,Strunk & White,asymptotics,Notable Question,least-squares,correlation,Pundit,variance,econometrics,Guru,conditional-expectation,central-limit-theorem,estimation,random-variable,Inquisitive,convergence,Famous Question,consistency,multiple-regression,Civic Duty,Announcer,Good Question,logistic,Booster,expected-value,hypothesis-testing,unbiased-estimator,Great Answer,Favorite Question,Publicist	2013-08-06T23:18:34.317	2023-02-17T22:16:49.770	est fun ana roc cor cor val the dis coe ass like dis sar dis fan tes reg pro wa ser cau tes
30005	jld	2013-09-06T18:51:41.153	United States	124	Editor,Student,Scholar,Teacher,Supporter,Critic,Caucus,Yearling,Constituent,Autobiographer,Commentator,Enthusiast,Nice Answer,Enlightened,Civic Duty,Informed,Good Answer,Custodian,Guru,Citizen Patrol,Organizer,Explainer,Electorate,Fanatic,Excavator,Revival,machine-learning,Mortarboard,Tag Editor,mathematical-statistics,regression,Generalist,Self-Learner,Investor,Altruist,self-study,Pundit,probability	2014-06-23T18:53:12.960	2022-02-17T17:43:33.990	svr dis car mo pro squ squ lea mo fun inv bin nor reg inte sta nlm reg sm
31978	Greenparker	2013-10-28T02:13:22.950	Kanpur, Uttar Pradesh, India	127	Student,Teacher,Tumbleweed,Editor,Supporter,Commentator,Yearling,Critic,Organizer,Custodian,Explainer,Scholar,Curious,Informed,Autobiographer,Citizen Patrol,Investor,Nice Answer,Excavator,Enthusiast,Strunk & White,Revival,Suffrage,Vox Populi,Civic Duty,Cleanup,Proofreader,Promoter,Nice Question,Benefactor,Reviewer,Fanatic,Enlightened,Mortarboard,Refiner,Necromancer,Good Answer,Deputy,Steward,Disciplined,Electorate,Caucus,Popular Question,Pundit,Altruist,Notable Question,probability,sampling,bayesian,normal-distribution,self-study,markov-chain-montecarlo,monte-carlo	2013-11-16T21:11:29.867	2019-12-20T20:28:10.490	refe sar pro erro dis eve fun var ana val bla dis dis infe fre
35989	Tim	2013-12-10T21:19:06.223	Warsaw, Poland	717	Editor,Autobiographer,Supporter,Teacher,Custodian,Student,Scholar,Excavator,Commentator,Citizen Patrol,Critic,Explainer,Revival,Nice Answer,Tumbleweed,Quorum,Informed,Yearling,Enlightened,Curious,Enthusiast,Civic Duty,Organizer,Suffrage,Nice Question,Strunk & White,Fanatic,Caucus,Proofreader,Cleanup,Constituent,Convention,Deputy,Necromancer,Electorate,Popular Question,Tag Editor,Reviewer,Pundit,Taxonomist,Good Answer,Mortarboard,r,Generalist,bayesian,Self-Learner,Notable Question,Good Question,Promoter,Benefactor,Sportsmanship,Favorite Question,regression,distributions,normal-distribution,Great Question,Famous Question,Guru,probability,Refiner,Inquisitive,terminology,Investor,machine-learning,Copy Editor,Synonymizer,Altruist,Stellar Question,random-generation,Disciplined,prior,Talkative,Announcer,time-series,logistic,estimation,Sheriff,maximum-likelihood,correlation,generalized-linear-model,classification,mixed-model,neural-networks,references,inference,optimization,Booster,simulation,predictive-models,variance,multiple-regression,Great Answer,self-study,conditional-probability,Populist,hypothesis-testing,forecasting,notation,mathematical-statistics,beta-distribution,mean,posterior,density-function,binomial-distribution,bootstrap,definition,kernel-smoothing,model-selection,cross-validation,likelihood,markov-chain-montecarlo,clustering,regularization,Lifejacket,statistical-significance,frequentist,loss-functions,python,conjugate-prior,random-forest,confidence-interval,natural-language,monte-carlo,standard-deviation,overfitting,random-variable,sampling,naive-bayes	2014-10-31T11:05:38.017	2023-02-25T07:53:52.917	tim mo cor coe stu mo var me dat siz cor reg ma tes val var dat lea fan dis alg em bin del tes sel cor net reg net pro net cor diri net rec dat dis reg reg dis ass stu wa exp crit
36041	Aksakal	2013-12-11T15:34:17.843	VA, United States	286	Editor,Teacher,Commentator,Critic,Supporter,Revival,Student,Enthusiast,Organizer,Tenacious,Custodian,Nice Answer,Enlightened,Yearling,Necromancer,Explainer,Quorum,Unsung Hero,Nice Question,Citizen Patrol,Good Answer,Caucus,Constituent,Tumbleweed,Curious,Scholar,time-series,Generalist,distributions,regression,Pundit,Mortarboard,probability,mathematical-statistics,Autobiographer,Talkative,machine-learning,Guru,correlation,r,normal-distribution,Excavator,Disciplined,Fanatic,terminology,Great Answer,neural-networks,bayesian,Populist,multiple-regression,Popular Question,hypothesis-testing,random-variable,self-study,Civic Duty,Notable Question,Promoter,forecasting,least-squares,stationarity,optimization,intuition,arima,Synonymizer,Announcer,pca,linear-model	2014-02-26T15:16:38.457	2023-02-10T17:33:22.003	tim sm stu dis val squ nei dis exp err fun trai val sm r,st dat sar sel fun des net lea des squ del opt lea diff alg
44269	Alexis	2014-04-22T15:44:33.497		240	Teacher,Editor,Revival,Necromancer,Student,Critic,Supporter,Commentator,Informed,Excavator,Custodian,Scholar,Promoter,Tag Editor,Organizer,Citizen Patrol,Benefactor,Enthusiast,Curious,Nice Answer,Self-Learner,Nice Question,Quorum,Popular Question,Enlightened,Strunk & White,Civic Duty,Good Question,Fanatic,Autobiographer,Explainer,Notable Question,Good Answer,Caucus,Constituent,Yearling,Taxonomist,Famous Question,Mortarboard,hypothesis-testing,Cleanup,Reviewer,Pundit,r,regression,Inquisitive,Tumbleweed,Generalist,Convention,Electorate,Guru,Copy Editor,statistical-significance,Proofreader,Archaeologist,Populist,Talkative,time-series,Refiner,kruskal-wallis-test,multiple-comparisons,Investor,Altruist,Sportsmanship,nonparametric,confidence-interval	2014-04-23T18:00:43.053	2023-02-01T17:32:31.337	sur tes imp dis dat squ tes reg squ dis eng a-v r,sp reg visi diff
52554	Russ Lenth	2014-07-22T15:40:09.550	Iowa City, IA, USA	83	Student,Editor,Teacher,Informed,Supporter,Commentator,Revival,Critic,Quorum,Custodian,Promoter,Nice Question,Scholar,Necromancer,Benefactor,Self-Learner,Enthusiast,Citizen Patrol,Mortarboard,Organizer,Nice Answer,Enlightened,Curious,Tag Editor,Yearling,Explainer,r,Good Question,Popular Question,mixed-model,post-hoc,lme4-nlme,lsmeans,regression,anova,multiple-comparisons,Favorite Question,Notable Question	2014-07-23T20:19:03.463	2023-02-22T03:41:50.710	hyp reg inte pro aut siz hoo sta reg ana log mo diff
53690	Richard Hardy	2014-08-08T10:57:13.613	Europe	363	Student,Scholar,Editor,Commentator,Supporter,Excavator,Teacher,Revival,Autobiographer,Curious,Informed,Tumbleweed,Enthusiast,Organizer,Nice Answer,Enlightened,Fanatic,Quorum,Explainer,Caucus,Constituent,Tenacious,Custodian,Strunk & White,Critic,Unsung Hero,Yearling,Civic Duty,Nice Question,time-series,Popular Question,Cleanup,Inquisitive,Citizen Patrol,Refiner,Proofreader,Pundit,Copy Editor,Necromancer,r,Tag Editor,regression,Notable Question,Reviewer,arima,Convention,Electorate,Suffrage,Vox Populi,garch,forecasting,Synonymizer,Good Answer,Mortarboard,Talkative,Self-Learner,Generalist,Taxonomist,Famous Question,cointegration,Good Question,Archaeologist,Promoter,Benefactor,stationarity,Deputy,aic,Socratic,autocorrelation,model-selection,Favorite Question,Research Assistant,unit-root,vector-autoregression,least-squares,augmented-dickey-fuller,machine-learning,hypothesis-testing,Sportsmanship,granger-causality,lags,vector-error-correction-model	2014-10-19T11:52:22.160	2023-03-02T17:22:20.020	r,es aut sta tes roc line sel cor squ cor sel tes inte las erro reg dat lea net ana bia eng the sar var inte me dis

Id	DisplayName	CreationDate	Location	BadgesWon	BadgeList	EarliestAnswer	LatestAnswer	Un
60261	Achim Zeileis	2014-11-08T08:45:50.373	Innsbruck, Austria	96	Teacher,Editor,Supporter,Caucus,Commentator,Constituent,Revival,Enthusiast,Citizen Patrol,Autobiographer,Critic,Informed,Explainer,Fanatic,Custodian,Yearling,Nice Answer,Enlightened,Organizer,Tag Editor,Synonymizer,Quorum,Good Answer,Suffrage,r,cart,regression,zero-inflation,beta-regression,negative-binomial-distribution,Guru,Generalist,generalized-linear-model,Populist,Civic Duty	2014-12-11T09:15:31.540	2023-01-06T08:48:25.507	poi reg err reg tes full ana wh mo squ ana reg pag mo ana fun
60613	Firebug	2014-11-13T14:13:15.340	Switzerland	210	Editor,Tumbleweed,Scholar,Student,Supporter,Citizen Patrol,Teacher,Informed,Autobiographer,Revival,Critic,Suffrage,Vox Populi,Yearling,Commentator,Enthusiast,Explainer,Civic Duty,Custodian,Disciplined,Necromancer,Promoter,Curious,Tenacious,Organizer,Investor,Excavator,Nice Answer,Mortarboard,Good Answer,Fanatic,Altruist,Quorum,Guru,Enlightened,Nice Question,Self-Learner,Deputy,Electorate,Strunk & White,Proofreader,Convention,Cleanup,Popular Question,Caucus,Constituent,Reviewer,machine-learning,Talkative,Notable Question,svm,Benefactor,Pundit,r,Good Question,Announcer,Booster,regression,classification,Publicist,random-forest,Synonymizer,Famous Question,Generalist,Taxonomist,pca,Favorite Question,neural-networks,optimization	2015-05-02T13:38:19.193	2023-02-11T14:15:42.127	ma dim mo bay fore sar net sel des mo est dis
67799	Christoph Hanck	2015-01-30T09:08:44.590	Essen	185	Teacher,Editor,Supporter,Student,Revival,Autobiographer,Scholar,Informed,Commentator,Explainer,Necromancer,Promoter,Excavator,Critic,Organizer,Benefactor,Caucus,Custodian,Citizen Patrol,Constituent,Suffrage,Enthusiast,Civic Duty,Vox Populi,Electorate,Fanatic,Self-Learner,Quorum,Deputy,Curious,Proofreader,Nice Answer,Enlightened,Tag Editor,Mortarboard,Reviewer,Investor,Altruist,Yearling,Nice Question,Steward,Sportsmanship,Good Answer,Generalist,Popular Question,Strunk & White,Pundit,Refiner,regression,r,bayesian,hypothesis-testing,self-study,time-series,econometrics,least-squares,Announcer,multiple-regression,stationarity,variance,unit-root,distributions,panel-data,mathematical-statistics	2015-02-18T13:39:17.883	2023-03-01T13:07:38.477	bay pro sel tes coe sm like tes mo fun var sta ana err ger dis dat larg lim cov opt sar dis
74500	Matthew Drury	2015-04-24T22:41:05.543	Seattle, WA	244	Student,Editor,Teacher,Supporter,Commentator,Nice Answer,Custodian,Explainer,Informed,Enlightened,Enthusiast,Critic,Autobiographer,Good Answer,Civic Duty,Fanatic,Mortarboard,Talkative,Organizer,Revival,regression,Citizen Patrol,Nice Question,Pundit,Scholar,r,Yearling,Outspoken,Quorum,Electorate,machine-learning,Necromancer,Guru,logistic,Curious,probability,Generalist,Caucus,Constituent,Excavator,Strunk & White,Proofreader,Popular Question,distributions,Good Question,boosting,Notable Question,Sportsmanship,Announcer,Favorite Question,generalized-linear-model,classification,Famous Question	2015-05-03T05:25:13.417	2020-10-19T18:45:52.417	reg mo mo trac visi dis sta mo fit,r dis the dis dis lea tes nei cat effe
85665	BruceET	2015-08-11T17:22:01.590	San Francisco Bay Area	122	Autobiographer,Teacher,Editor,Revival,Supporter,Commentator,Caucus,Yearling,Critic,Explainer,Citizen Patrol,Nice Answer,Informed,Enthusiast,Custodian,Organizer,Enlightened,Fanatic,Necromancer,Civic Duty,hypothesis-testing,normal-distribution,distributions,Generalist,probability,r,t-test,Sportsmanship,confidence-interval,statistical-significance,mathematical-statistics,Student,Self-Learner,Mortarboard,Good Answer,self-study,Electorate,Pundit,estimation,mean,Strunk & White,anova,bayesian,Constituent,variance,inference,binomial-distribution,Refiner,p-value,chi-squared-test,sample-size,standard-deviation,Cleanup,nonparametric,descriptive-statistics,Excavator,sampling,poisson-distribution,Disciplined,wilcoxon-mann-whitney-test,Nice Question,proportion	2016-10-04T05:59:47.963	2022-07-02T05:16:55.040	sta dis dis tes sta ass trai ii-e trac diff tim dis dat tes mo dis scc dis dat eng me ana tes pro sta dis est fitti rate cor
86652	Björn	2015-08-23T14:39:03.497	Europe	109	Informed,Student,Teacher,Supporter,Revival,Autobiographer,Commentator,Critic,Editor,Nice Answer,Enlightened,Enthusiast,Scholar,Custodian,Fanatic,Citizen Patrol,Civic Duty,Yearling,Promoter,Benefactor,Curious,Tumbleweed,Suffrage,Caucus,Organizer,Constituent,Popular Question,regression,bayesian,Census,Explainer,Vox Populi,Electorate,hypothesis-testing,logistic,Necromancer,Notable Question,Generalist,Mortarboard,Good Answer,Announcer,machine-learning,r,neural-networks,Famous Question	2015-08-24T19:13:07.693	2023-03-02T20:06:45.603	ma cov squ rati cor stu reg dat inte trei ana net cor visi scc aug eva cla ma ass fitti net opt dis sta visi
97925	Matthew Gunn	2015-12-11T19:40:42.483		141	Teacher,Editor,Supporter,Critic,Commentator,Nice Answer,Mortarboard,Informed,Good Answer,Enthusiast,Autobiographer,Quorum,Organizer,Custodian,Civic Duty,Excavator,Disciplined,Enlightened,Pundit,Revival,Fanatic,Yearling,Sportsmanship,Necromancer,Explainer,Citizen Patrol,Guru,Generalist,Announcer,regression,least-squares,time-series,mathematical-statistics,probability,Lifejacket,machine-learning	2015-12-11T08:30:43.697	2023-02-23T16:10:06.643	reg dis err val ma sta lea lea coe ana pro squ dev like enc
101426	mdewey	2016-01-21T14:04:37.277	United Kingdom	114	Informed,Teacher,Autobiographer,Commentator,Editor,Citizen Patrol,Tenacious,Enthusiast,Supporter,Custodian,Critic,Fanatic,Organizer,Unsung Hero,Excavator,Quorum,Explainer,Revival,Proofreader,Tag Editor,Reviewer,Deputy,Civic Duty,Nice Answer,Enlightened,meta-analysis,Mortarboard,Steward,regression,Electorate,Yearling,Taxonomist,Convention,Good Answer,Caucus,Constituent,Strunk & White,r,Necromancer,Suffrage,Pundit,Census,Cleanup,logistic,Generalist	2016-01-21T14:19:02.217	2023-02-08T12:05:02.193	me dat cor allc dis mo var plo trai tes squ wir tes cor ma dis
111259	Demetri Pananos	2016-04-06T21:32:38.250	Toronto, ON, Canada	162	Autobiographer,Student,Editor,Supporter,Tumbleweed,Scholar,Commentator,Critic,Curious,Informed,Nice Question,Yearling,Custodian,Popular Question,Teacher,Nice Answer,Enlightened,Citizen Patrol,Enthusiast,Inquisitive,Mortarboard,Good Answer,Explainer,Revival,Organizer,regression,Announcer,Notable Question,Fanatic,Necromancer,Pundit,r,Generalist,Civic Duty,Booster,Caucus,Constituent,Excavator,bayesian,hypothesis-testing,machine-learning,Promoter,Benefactor,logistic,statistical-significance,generalized-linear-model,t-test,confidence-interval,mathematical-statistics,probability	2018-11-11T21:12:41.223	2023-02-23T03:49:44.907	tim dat tab val visi reg of-t nei dis infl des dat squ cor cor inte trac cur fun risk squ diff
113777	Haitao Du	2016-04-27T20:51:38.203		377	Teacher,Revival,Editor,Supporter,Informed,Critic,Commentator,Autobiographer,Custodian,Student,Nice Question,Promoter,Scholar,Enthusiast,Citizen Patrol,Quorum,Popular Question,Excavator,Curious,Disciplined,Nice Answer,Organizer,Investor,Explainer,Altruist,Civic Duty,Suffrage,Vox Populi,Necromancer,Fanatic,Mortarboard,Good Question,Notable Question,Favorite Question,Self-Learner,Tumbleweed,Proofreader,Inquisitive,Benefactor,Tenacious,Enlightened,Strunk & White,machine-learning,Deputy,regression,Electorate,Reviewer,Yearling,Good Answer,logistic,Famous Question,Caucus,Constituent,neural-networks,r,classification,Socratic,Generalist,Sportsmanship,optimization,Great Question,Announcer,regularization,Convention,Pundit,boosting,gradient-descent,svm	2016-04-27T20:57:46.047	2021-11-14T13:52:20.913	ma ma net reg mix dim ana dat inv rec cla mo sar cor
115634	Isabella Ghement	2013-12-12T06:44:52.040	Richmond, BC, Canada	89	Editor,Teacher,Enthusiast,Fanatic,Autobiographer,Popular Question,Scholar,Student,Supporter,Custodian,Nice Answer,Good Answer,Civic Duty,Commentator,Yearling,Revival,Explainer,Electorate,Curious,Critic,Enlightened,Generalist,Mortarboard,Guru,Caucus,regression,multiple-regression,generalized-linear-model,r,mixed-model,lme4-nlme,Necromancer,logistic,Nice Question	2018-03-18T22:15:08.510	2021-04-15T02:34:15.507	exp dat reg me dat effe like stu dat tes wa pac dat squ

Id	DisplayName	CreationDate	Location	BadgesWon	BadgeList	EarliestAnswer	LatestAnswer	Un
116195	Noah	2016-05-19T15:46:14.703	Cambridge, MA, United States	109	Teacher,Student,Editor,Organizer,Supporter,Citizen Patrol,Commentator,Critic,Tumbleweed,Scholar,Tenacious,Yearling,Autobiographer,Unsung Hero,Custodian,Revival,Curious,Necromancer,Nice Answer,Excavator,propensity-scores,Enthusiast,regression,r,matching,causality,Census,Explainer,Mortarboard,Fanatic,Enlightened,treatment-effect,Announcer,Synonymizer,Good Answer,logistic,Popular Question,Civic Duty,Guru,Booster,confounding,Notable Question	2016-07-01T23:26:52.873	2023-02-16T22:33:43.547	inte mo stu risk tes res sta log fac enc alg like tran effe var reg bia tes sta sep pro reg
116440	user20160	2016-05-22T03:26:26.563		178	Teacher,Supporter,Revival,Editor,Informed,Student,Commentator,Citizen Patrol,Nice Answer,Custodian,Organizer,Critic,Tenacious,Tag Editor,Explainer,Enthusiast,machine-learning,Scholar,Necromancer,Popular Question,Enlightened,Yearling,neural-networks,Quorum,Good Answer,Fanatic,Civic Duty,Nice Question,Curious,Guru,regression,optimization,Generalist,Excavator,classification,Caucus,Constituent,Great Answer,Electorate,pca,regularization,Notable Question,Pundit	2016-05-22T03:39:12.343	2022-03-07T16:29:40.450	cro tran lea visu sta dis neu cur net ave mo dis squ mo pro cla lea sta pro info mo ana cla pro
158565	user158565	2017-04-24T04:49:47.677		31	Teacher,Editor,Informed,Commentator,Critic,Nice Answer,Enthusiast,Supporter,Tenacious,Autobiographer,Unsung Hero,Yearling,Explainer,Custodian,Quorum,Citizen Patrol,Reviewer,Steward,Enlightened	2017-04-24T07:03:17.630	2019-08-12T14:13:19.590	sel squ mo mo squ mo dis mo dis var mo tes exp tria fre
164061	Sextus Empiricus	2017-06-05T10:39:01.763	Sion, Switzerland	298	Autobiographer,Editor,Teacher,Supporter,Informed,Commentator,Critic,Revival,Enthusiast,Citizen Patrol,Caucus,Constituent,Custodian,Explainer,Nice Answer,Necromancer,Organizer,Student,Quorum,Excavator,Promoter,Mortarboard,Benefactor,Investor,Altruist,Civic Duty,Enlightened,Talkative,Cleanup,Yearling,Scholar,Disciplined,probability,Good Answer,Convention,regression,Generalist,r,Sportsmanship,distributions,mathematical-statistics,Census,Curious,Nice Question,Pundit,Good Question,Populist,Self-Learner,Strunk & White,Popular Question,normal-distribution,hypothesis-testing,Guru,bayesian,self-study,optimization,confidence-interval,epidemiology,Lifejacket,Inquisitive,least-squares,machine-learning,p-value,multiple-regression,Announcer,time-series,terminology,estimation,inference,Fanatic,statistical-significance,linear-model,expected-value,density-function,correlation,conditional-probability,anova,regularization,maximum-likelihood,lasso,random-variable	2017-06-07T21:08:17.853	2023-03-03T13:53:54.643	hyp fish dis dis red sar var exp ana tes dis tran val dis inte trac pro reg reg equ pro reg in-c sm info lea mu reg dec inte dis me dis mir ma dis mo pro
166526	Sal Mangiafico	2017-06-24T13:11:05.510	Greenwich, NJ 08323, United States	49	Teacher,Autobiographer,Supporter,Commentator,Editor,Enthusiast,Caucus,Tumbleweed,Critic,Constituent,Scholar,Student,Tenacious,Fanatic,Custodian,Informed,Unsung Hero,Revival,Yearling,Citizen Patrol,Civic Duty,Announcer,Self-Learner,Census,r,Necromancer,Popular Question,Nice Answer,Excavator,Enlightened,Talkative,chi-squared-test,hypothesis-testing	2017-06-24T13:40:10.947	2023-02-21T16:24:45.147	chi tes mo cor ser rate tria visu d,ir sta reg rati est
173082	Ben	2017-08-10T03:27:26.793	Canberra, Australia	649	Teacher,Autobiographer,Editor,Revival,Mortarboard,Supporter,Commentator,Critic,Citizen Patrol,Custodian,Informed,Student,Nice Question,Necromancer,Scholar,Quorum,Enthusiast,Explainer,Nice Answer,Fanatic,Enlightened,Good Answer,probability,Civic Duty,bayesian,Suffrage,Generalist,Curious,Promoter,Guru,Benefactor,Yearling,Excavator,Organizer,Cleanup,regression,Vox Populi,Disciplined,Electorate,self-study,distributions,Strunk & White,Self-Learner,mathematical-statistics,normal-distribution,maximum-likelihood,hypothesis-testing,random-variable,Sportsmanship,conditional-probability,r,time-series,inference,expected-value,variance,Popular Question,estimation,Refiner,machine-learning,statistical-significance,sampling,least-squares,terminology,Notable Question,Good Question,Inquisitive,Favorite Question,Pundit,central-limit-theorem,likelihood,Caucus,Tag Editor,mean,uniform-distribution,density-function,confidence-interval,convergence,Convention,correlation,intuition,Epic,unbiased-estimator,binomial-distribution,linear-model,stationarity,multiple-regression,frequentist,markov-chain-montecarlo,independence,arima,Deputy,generalized-linear-model,moments,p-value,references,modeling,notation,econometrics,asymptotics,poisson-distribution,residuals,Famous Question,estimators,cumulative-distribution-function,categorical-data,random-generation,r-squared	2018-01-17T05:37:49.037	2023-03-01T01:20:44.000	var sar the val me bay dis pro for rule mo pric car reg dis dis law sta ma dis dis est fitti nlm app pric diff exa fn,s ine var tail noi dis sys pro mo dis dis mo sim
204068	gunes	2018-04-12T10:42:43.307	Cambridge, UK	129	Teacher,Editor,Supporter,Critic,Commentator,Nice Answer,Custodian,Explainer,Organizer,Enthusiast,Autobiographer,Fanatic,Mortarboard,Tenacious,Yearling,probability,Enlightened,Citizen Patrol,Student,Scholar,self-study,machine-learning,Good Answer,mathematical-statistics,Generalist,Civic Duty,normal-distribution,distributions,regression,expected-value,neural-networks,conditional-probability,random-variable,variance,bayesian,Electorate,Guru,independence,Strunk & White,classification,covariance,Caucus,Constituent,Refiner,time-series,cross-validation,conditional-expectation,optimization,scikit-learn,Pundit	2018-09-09T17:27:12.530	2023-01-22T13:32:10.323	pro sta ent lea sar ser lea red lea alg leib mir dis nur mo ind pro tric cla mo the bin dec sar mo dis dis dis cor inte dis dat inte dat r,ps lea dis tim rec val
219012	Dimitris Rizopoulos	2018-08-28T20:43:20.297	Rotterdam, Netherlands	58	Teacher,Editor,Autobiographer,Commentator,Informed,Supporter,Critic,Organizer,Enthusiast,Custodian,Explainer,Revival,Tenacious,mixed-model,Excavator,Fanatic,r,lme4-nlme,Nice Answer,Enlightened,Strunk & White,Refiner,glmm,generalized-linear-model,regression,Yearling,random-effects-model,Citizen Patrol,logistic,Caucus,Constituent,Good Answer	2018-08-29T03:24:59.473	2023-02-09T13:41:20.213	ma mo mo me hoo red ana lea alg dat ass bay mo

Id	DisplayName	CreationDate	Location	BadgesWon	BadgeList	EarliestAnswer	LatestAnswer	Un
247165	Christian Hennig	2019-05-07T15:59:22.940	Bologna, Italy	69	Teacher,Informed,Revival,Supporter,Commentator,Editor,Custodian,Student,Nice Answer,Yearling,Critic,Talkative,Scholar,Caucus,Constituent,Enthusiast,Explainer,Enlightened,Citizen Patrol,Autobiographer,hypothesis-testing,Excavator,Generalist,Necromancer,Organizer,regression,Popular Question,p-value,Self-Learner,Pundit,probability	2019-05-07T16:23:32.463	2023-02-16T14:05:56.513	wil sar dis ent dis sta est cor sel the prc diff scc a-v dat rati mo equ
247274	Dave	2019-05-08T13:23:32.777		234	Informed,Teacher,Editor,Supporter,Tumbleweed,Commentator,Scholar,Student,Critic,Curious,Custodian,Citizen Patrol,Enthusiast,Revival,Explainer,Nice Answer,Organizer,Quorum,Enlightened,Good Answer,Mortarboard,Civic Duty,Excavator,Guru,Fanatic,Pundit,Talkative,Yearling,Electorate,Tag Editor,Disciplined,Nice Question,Popular Question,Inquisitive,Good Question,Notable Question,hypothesis-testing,regression,mathematical-statistics,Generalist,Populist,machine-learning,Caucus,Constituent,statistical-significance,Strunk & White,Deputy,t-test,Investor,Altruist,Cleanup,r,probability,Self-Learner,normal-distribution,variance,Promoter,Benefactor,distributions,mean,least-squares,Favorite Question,Necromancer,classification,logistic,correlation,Convention,neural-networks,p-value,confidence-interval,Socratic,r-squared,python,Announcer,Sportsmanship,Autobiographer,linear-model	2019-05-08T21:36:54.920	2023-03-04T21:21:17.200	reg eva sta var dis var r,sp me reg net tes dis des lea a-v dis sta dis reg nei dat ana sca alg fn,r dat fac
249135	Thomas Lumley	2019-05-26T21:58:47.913	New Zealand	131	Teacher,Editor,Revival,Autobiographer,Supporter,Yearling,Commentator,Mortarboard,Custodian,Critic,Nice Answer,Necromancer,Enthusiast,Explainer,Student,Excavator,r,Enlightened,Proofreader,Good Answer,Guru,Informed,Reviewer,Generalist,distributions,regression,probability,Populist,Citizen Patrol,mathematical-statistics,Promoter,Disciplined,hypothesis-testing	2019-05-29T00:15:48.097	2023-03-04T07:49:01.590	r,de like mo sta err red coe mo sm squ spe dis dis dis ser bay dis dis reg nea ran prc free var car mo tes dis reg sep err reg fan

```
# The following queries are what I used to "build up" to the final answer
# They are commented out to avoid unnecessarily filling the output, but to give you an idea of what I did :)

#Select Users with the top 75 Accepted Answers
# dbGetQuery(db, 'SELECT OwnerUserId, COUNT(*) AS Freq, MIN(CreationDate) AS EarliestAnswer, MAX(CreationDate) AS LatestAnsw
er
#                 FROM Posts
#                 WHERE Id IN (SELECT AcceptedAnswerId FROM Posts WHERE (NOT AcceptedAnswerId = "") OR (NOT AcceptedAnswerId
IS NULL)) AND OwnerUserId <> ""
#                 GROUP BY OwnerUserId
#                 ORDER BY Freq DESC
#                 LIMIT 75;
#                 ')

#User Badges
#[3] Using Group_Concat to make Strings of columns
# dbGetQuery(db, 'SELECT u.Id, COUNT(*) AS BadgesWon, GROUP_CONCAT(Badges.Name) AS BadgeList
#                 FROM Users AS u
#                 INNER JOIN Badges ON u.Id = Badges.UserId
#                 GROUP BY u.Id')

#Get Unique tags per post
#[4] Distinct Tags for questions where they have accepted answer
# dbGetQuery(db, 'SELECT p.Id, t.Id, p.AcceptedAnswerId, a.Id, a.OwnerUserId, group_concat(t.Tag)
#           FROM Posts AS p
#           INNER JOIN TagPosts AS t ON t.Id = p.Id
#           INNER JOIN Posts AS a ON p.AcceptedAnswerId = a.Id
#           WHERE p.AcceptedAnswerId <> ""
#           GROUP BY a.OwnerUserId
#            ')
```

QUESTION 24: How many questions received no answers (accepted or unaccepted)? How many questions had no accepted answer?

For this question I used the AnswerCount to check which questions had no answers. Answers without an accepted answer was rather easy, I just checked which posts had no AcceptedAnswerId.

```
#Posts WHERE AnswerCount == 0
kable(dbGetQuery(db, "SELECT COUNT(AnswerCount) FROM Posts WHERE AnswerCount = 0"), col.names = 'Questions without Answer')
```

Questions without Answer
66970

```
#Verification:
nonzero = dbGetQuery(db, 'SELECT p.Id, COUNT(*) AS f
                 FROM Posts AS p
                 INNER JOIN Posts AS a ON p.Id = a.ParentId
                 WHERE a.PostTypeId = 2 AND p.PostTypeId = 1
                 GROUP BY p.Id
                 HAVING f > 0
                ')
allposts = dbGetQuery(db, 'SELECT Id FROM Posts WHERE PostTypeId = 1')
print(paste('The Number of no answer posts, without using AnswerCount:', nrow(allposts) - nrow(nonzero)))
```

```
## [1] "The Number of no answer posts, without using AnswerCount: 66970"
```

```
kable(dbGetQuery(db, 'SELECT COUNT(AcceptedAnswerID) FROM Posts WHERE AcceptedAnswerID =""'), col.names = 'Questions with no
accepted answer')
```

Questions with no accepted answer
337215

```
#Posts where there is a matching AcceptedAnswerID? Linked somehow
```

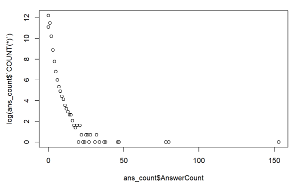
QUESTION 25: What is the distribution of answers per posted question?

This appears to be roughly a Poisson distribution, or following the shape of a negative exponential function. The AnswerCount was used for these values as in Question 24 we found that values was trustworthy.

```
ans_count = dbGetQuery(db, 'SELECT AnswerCount, COUNT(*) FROM Posts GROUP BY AnswerCount')
```

```
## Warning: Column `AnswerCount`: mixed type, first seen values of type integer,
## coercing other values of type string
```

```
plot(ans_count$AnswerCount, log(ans_count$`COUNT(*)`))
```



QUESTION 26: What is the length of time for a question to receive an answer? to obtaining an accepted answer? (Instructor Note: Make plots!)

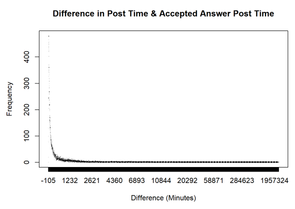
I used the CreationDate here to calculate these differences.

Note that we have some negative times in Question vs Accepted Answer. I went into the database to confirm this, and it seems like there are two posts where the answer is noted before the actual post somehow (Post 129091->129077 and Post 108295->109065)

```
#How long until accepted answer
q26a = dbGetQuery(db, 'SELECT p1.Id, p1.AcceptedAnswerId, p1.CreationDate, p2.CreationDate, (strftime("%s", p2.CreationDate)
- strftime("%s", p1.CreationDate))/60 as AnsDif FROM Posts AS p1
          INNER JOIN Posts AS p2 ON p2.Id = p1.AcceptedAnswerId
          ORDER BY AnsDif;')
head(q26a)
```

```
##       Id AcceptedAnswerId            CreationDate            CreationDate
## 1 129091           129077 2014-12-14T20:27:18.950 2014-12-14T18:41:59.110
## 2 108295           109065 2014-07-17T11:36:45.187 2014-07-17T11:09:39.077
## 3  24681            24682 2012-03-15T06:59:55.263 2012-03-15T07:00:17.697
## 4  29653            29654 2012-06-01T18:18:19.507 2012-06-01T18:18:19.707
## 5  29746            29747 2012-06-03T20:55:45.227 2012-06-03T20:55:45.337
## 6  34165            34166 2012-08-12T03:36:55.910 2012-08-12T03:36:55.910
##   AnsDif
## 1   -105
## 2    -27
## 3      0
## 4      0
## 5      0
## 6      0
```

```
q26a_table = as.data.frame(table(q26a$AnsDif))
plot(q26a_table$Var1, q26a_table$Freq, xlab = 'Difference (Minutes)', ylab = 'Frequency', main = 'Difference in Post Time &
Accepted Answer Post Time')
```

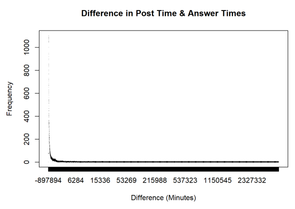


Note that we have many posts here with negative post to answer times, I'm wondering if this is due to editing or something.

```
q26p = dbGetQuery(db, 'SELECT p1.Id, p2.ParentId, p2.Id, p1.CreationDate, p2.CreationDate, (strftime("%s", p2.CreationDate)
- strftime("%s", p1.CreationDate))/60 as AnsDif
         FROM Posts as p1
         INNER JOIN Posts AS p2 ON p2.ParentId = p1.Id
         WHERE p2.PostTypeId = 2
         ORDER BY AnsDif')
head(q26p)
```

```
##       Id ParentId     Id            CreationDate            CreationDate
## 1 112451   112451  45056 2014-08-19T12:46:29.950 2012-12-03T23:52:18.377
## 2 112451   112451  45091 2014-08-19T12:46:29.950 2012-12-04T13:09:31.343
## 3  16921    16921    893 2011-10-12T20:16:52.280 2010-07-28T12:48:14.233
## 4  16921    16921    894 2011-10-12T20:16:52.280 2010-07-28T12:49:31.780
## 5 116804   116804  79863 2014-09-25T21:49:50.467 2013-12-16T21:13:01.827
## 6 239431   239431 190181 2016-10-10T15:24:14.047 2016-01-11T15:19:13.797
##    AnsDif
## 1 -897894
## 2 -897096
## 3 -635488
## 4 -635487
## 5 -407556
## 6 -393125
```

```
q26p_table = as.data.frame(table(q26p$AnsDif))
plot(q26p_table$Var1, q26p_table$Freq, xlab = 'Difference (Minutes)', ylab = 'Frequency', main = 'Difference in Post Time &
Answer Times')
```



QUESTION 27: How many answers are typically received before the accepted answer?

Here I used the seconds to calculate the difference, as allowing it to all be in the same unit allows for easy math. I used lots of verification using DB Browser for SQLite and comparing to the actual posts themselves.

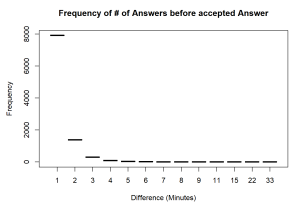
```
q27 = dbGetQuery(db, 'SELECT p1.Id, p2.Id AS EarlyPost, ((strftime("%s", p3.CreationDate) - strftime("%s", p2.CreationDat
e))/60) as Dif, COUNT(*) AS Freq
         FROM Posts as p1
         LEFT JOIN Posts AS p2 ON p1.Id = p2.ParentId
         LEFT JOIN Posts AS p3 ON p1.AcceptedAnswerId = p3.Id
         WHERE ((strftime("%s", p3.CreationDate) - strftime("%s", p2.CreationDate))/60) > 0
             GROUP BY p1.Id')

head(q27)
```

```
##   Id EarlyPost   Dif Freq
## 1  2        20    18    1
## 2  4       133     4    1
## 3  7        12     5    1
## 4 10       153 43880    2
## 5 26        41    11    2
## 6 35        38    18    1
```

```
q27_table = as.data.frame(table(q27$Freq))
head(q27_table)
```

```
##   Var1 Freq
## 1    1 7909
## 2    2 1390
## 3    3  293
## 4    4   88
## 5    5   38
## 6    6   11
```

```
plot(q27_table$Var1, q27_table$Freq, xlab = 'Difference (Minutes)', ylab = 'Frequency', main = 'Frequency of # of Answers be
fore accepted Answer')
```



[1] https://learnsql.com/cookbook/how-to-calculate-the-difference-between-two-timestamps-in-sqlite/ (https://learnsql.com/cookbook/how-to-calculate-the-difference-between-two-timestamps-in-sqlite/)

[2] https://rveryday.wordpress.com/2016/11/17/create-a-cumulative-sum-column-in-r/ (https://rveryday.wordpress.com/2016/11/17/create-a-cumulative-sum-column-in-r/)

[3] https://www.sqlitetutorial.net/sqlite-group_concat/ (https://www.sqlitetutorial.net/sqlite-group_concat/)

[4] https://www.sqlitetutorial.net/sqlite-distinct/ (https://www.sqlitetutorial.net/sqlite-distinct/)